ESAIM: PS 18 (2014) 277–307
DOI: 10.1051/ps/2013037

ESAIM: Probability and Statistics
www.esaim-ps.org

ESTIMATION IN AUTOREGRESSIVE MODEL WITH MEASUREMENT ERROR

Jérôme Dedecker[1], Adeline Samson[1] and Marie-Luce Taupin[2]

Abstract. Consider an autoregressive model with measurement error: we observe $Z_i = X_i + \varepsilon_i$, where the unobserved X_i is a stationary solution of the autoregressive equation $X_i = g_{\theta^0}(X_{i-1}) + \xi_i$. The regression function g_{θ^0} is known up to a finite dimensional parameter θ^0 to be estimated. The distributions of ξ_1 and X_0 are unknown and g_θ belongs to a large class of parametric regression functions. The distribution of ε_0 is completely known. We propose an estimation procedure with a new criterion computed as the Fourier transform of a weighted least square contrast. This procedure provides an asymptotically normal estimator $\hat{\theta}$ of θ^0, for a large class of regression functions and various noise distributions.

Mathematics Subject Classification. 62J02, 62F12, 62G05, 62G20.

Received October 24, 2011. Revised February 14, 2013.

1. Introduction

We consider an autoregressive model with measurement error satisfying

$$\begin{cases} Z_i = X_i + \varepsilon_i, \\ X_i = g_{\theta^0}(X_{i-1}) + \xi_i \end{cases} \tag{1.1}$$

where one observes $Z_0, \ldots, Z_n$ and the random variables $\xi_i, X_i, \varepsilon_i$ are unobserved. The regression function g_{θ^0} is known up to a finite dimensional parameter θ^0, belonging to the interior Θ° of a compact set $\Theta \subset \mathbb{R}^d$.

The assumptions on the random variables $(\xi_i)_{i\geq 1}$ and $(\varepsilon_i)_{i\geq 0}$ are the following. The innovations $(\xi_i)_{i\geq 1}$ and the errors $(\varepsilon_i)_{i\geq 0}$ are centered, independent and identically distributed (i.i.d.) random variables with finite variances $\mathrm{Var}(\xi_1) = \sigma_\xi^2$ and $\mathrm{Var}(\varepsilon_0) = \sigma_\varepsilon^2$. The variable ε_0 admits a known density f_ε with respect to the Lebesgue measure, and the random variables X_0, $(\xi_i)_{i\geq 1}$ and $(\varepsilon_i)_{i\geq 0}$ are independent. The Markov chain $(X_i)_{i\geq 0}$ admits an invariant distribution.

The main originalities of the paper are: 1/ the distribution of ξ_1 is completely unknown and we do not even assume that it admits a density with respect to the Lebesgue measure; 2/ we do not assume that the variable X_0 admits a density; 3/ we consider a general non−linear regression function g_θ.

The distribution of the innovations being unknown, this model belongs to the family of semi−parametric models.

Keywords and phrases. Autoregressive model, Markov chain, mixing, deconvolution, semi–parametric model.

[1] Laboratoire MAP5 UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, Paris cedex 6, France

[2] Laboratoire Statistique et Génome, UMR CNRS 8071-USC INRA, Université d'Évry Val d'Essonne, Évry, France.
marie-luce.taupin@genopole.cnrs.fr

Our aim is to estimate θ^0 for a large class of functions g_θ, whatever the known error distribution f_ε, and without the knowledge of the ξ_i's distribution.

Previously known results

Let us start with the special case of linear regression function $g_\theta(x) = \theta_1 x + \theta_2$ (linear in both θ and x), see *e.g.* Andersen and Deistler [1], Nowak [24], Chanda [7,8], Staudenmayer and Buonaccorsi [27], and Costa *et al.* [10]. The model (1.1) is also an ARMA model (see Sect. 5.1.1 for further details) and consequently, all previously known estimation procedures for ARMA models can be applied. It is noteworthy that, in this specific case, the knowledge of the error distribution f_ε is not required.

For a general regression function, the model (1.1) is a Hidden Markov Model with possibly a non compact continuous state space.

When the distribution of the innovation ξ_1 is known up to a finite dimensional parameter, model (1.1) becomes fully parametric. In this parametric context, various results are already stated: among others, 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 [21], Bickel *et al.* [3], Jensen and Petersen [20], Douc and Matias [14], Douc *et al.* [16], Fuh [17], Genon−Catalot and Laredo [18], Na *et al.* [23], and Douc *et al.* [15].

In this paper, the distribution of ξ_1 is unknown, and model (1.1) is a semi−parametric Hidden Markov Model.

To our knowledge, the only paper which gives a consistent estimator of θ^0 is Comte and Taupin [9]. They propose an estimation procedure based on a modified least squares minimization. They give an upper bound for the rate of convergence of their estimator, that depends on the smoothness of the regression function and on the smoothness of f_ε. Those results are obtained by assuming that the distribution P_X of 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).

Comte and Taupin [9] state that their estimator achieves the parametric rate only for very few couples of regression functions/error distribution. Lastly their dependency conditions are quite restrictive, and the assumption that X admits a density is not natural in this context.

Our results

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

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

where w is a weight function to be chosen and $\mathbb{E}$ is the expectation $\mathbb{E}_{\theta^0, P_X}$.

Firstly, we assume that one can exhibit a weight function w such that $(wg_\theta)^*/f_\varepsilon^*$ and $(wg_\theta^2)^*/f_\varepsilon^*$ are integrable, where φ^* is the Fourier transform of an integrable function φ. This holds for a large class of regression functions. Examples are given in Section 2.5.

We estimate θ^0 by $\widehat{\theta} = \arg\min_{\theta\in\Theta} S_n(\theta)$, where

$$S_n(\theta) = \frac{1}{2\pi n} \sum_{k=1}^{n} \mathbb{R}e \int \frac{\left(\left(Z_k - g_\theta\right)^2 w\right)^* (t) \, \mathrm{e}^{-itZ_{k-1}}}{f_\varepsilon^*(-t)} \mathrm{d}t, \tag{1.2}$$

where $\mathbb{R}e(u)$ is the real part of u. This criteria is simple to minimize in most situations. We prove that the resulting estimator $\widehat{\theta}$ is consistent and asymptotically Gaussian. Those results hold under weak dependency conditions as introduced in Dedecker and Prieur [12]. Compared to Comte and Taupin [9], this procedure is clearly simpler and its main advantage is that it yields the parametric rate of convergence for a larger class of regression functions.

Secondly, when it is not possible to exhibit a weight function w such that $(wg_\theta)^*/f_\varepsilon^*$ and $(wg_\theta^2)^*/f_\varepsilon^*$ are integrable, we propose a more general estimator. We establish a consistency result, and we give an upper bound for the quadratic risk, that relates the smoothness properties of the regression function to that of f_ε. These last results are proved under α–mixing conditions.

Finally, 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).

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 the weight function w. The influence of this weight function is also studied in the simulations.

The paper is organized as follows.

In Section 2 we present our notations and assumptions. In Section 3 we propose general conditions (on the couple w/g_θ), under which the estimator is consistent and asymptotically Gaussian. The Section 4 concerns regression functions for which it seems not easy to exhibit a weight function w that satisfies the conditions given in Section 2. In this case, we propose a more general estimator which remains consistent and we derive its asymptotic rate of convergence. Those theoretical results are illustrated by a simulation study, which is presented in Section 5.

The proofs are gathered in Appendix.

2. Notations and assumptions

We first give some preliminary notations and assumptions to define more rigorously our criterion. Examples of model (1.1) for which assumptions are satisfied are given in Section 2.5.

2.1. Notations

For $\theta \in \mathbb{R}^d$, $\| \theta \|_{\ell^2}^2 = \sum_{k=1}^d \theta_k^2$, and $\theta^\top$ is the transpose vector of θ. Let

$$\| \varphi \|_1 = \int |\varphi(x)|\mathrm{d}x, \ \| \varphi \|_2^2 = \int \varphi^2(x)\mathrm{d}x, \ \text{ and } \ \| \varphi \|_\infty = \sup_{x\in\mathbb{R}} |\varphi(x)|.$$

The convolution product of two square integrable functions p and q is denoted by $p \star q(z) = \int p(z-x)q(x)\mathrm{d}x$. The Fourier transform φ^* of a function φ is defined by

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

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

$$\varphi_\theta^{(1)}(\cdot) = \left(\varphi_{\theta,j}^{(1)}(\cdot)\right)_{1\leq j\leq d}, \ \text{ with } \varphi_{\theta,j}^{(1)}(\cdot) = \frac{\partial\varphi_\theta(\cdot)}{\partial\theta_j} \text{ for } j \in \{1,\dots,d\},$$

$$\varphi_\theta^{(2)}(\cdot) = \left(\varphi_{\theta,j,k}^{(2)}(\cdot)\right)_{1\leq j,k\leq d}, \ \text{ with } \varphi_{\theta,j,k}^{(2)}(\cdot) = \frac{\partial^2\varphi_\theta(\cdot)}{\partial\theta_j\partial\theta_k}, \text{ for } j,k \in \{1,\dots,d\},$$

$$\text{and} \quad \varphi_\theta^{(3)}(\cdot) = \left(\varphi_{\theta,i,j,k}^{(3)}(\cdot)\right)_{1\leq i,j,k\leq d}, \ \text{ with } \varphi_{\theta,i,j,k}^{(3)}(\cdot) = \frac{\partial^3\varphi_\theta(\cdot)}{\partial\theta_i\partial\theta_j\partial\theta_k}, \text{ for } i,j,k \in \{1,\dots,d\}.$$

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 $\mathrm{Var}_{\theta^0,P_X}$, when the underlying and unknown true parameters are θ^0 and P_X.

2.2. Assumptions

We consider three types of assumptions. The two firsts are usual in least squares regression function estimation. The assumption ($\mathbf{N_1}$) is quite usual when considering additive errors. As in deconvolution, it ensures the existence of the estimation criterion.

- **Smoothness and moment assumptions**

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

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

- **Identifiability assumptions**

$$S(\theta) = \mathbb{E}[(g_{\theta^0}(X) - g_\theta(X))^2 w(X)] \text{ admits one unique minimum at } \theta = \theta^0. \qquad (\mathbf{I1_1})$$

$$\text{For all } \theta \in \Theta^\circ, \text{ the matrix } S^{(2)}(\theta) = \left(\frac{\partial^2 S(\theta)}{\partial\theta_i \partial\theta_j}\right)_{1\le i,j\le d} \text{ exists and the matrix} \qquad (\mathbf{I1_2})$$

$$S^{(2)}(\theta^0) = 2\,\mathbb{E}\left[w(X)\left(g^{(1)}_{\theta^0}(X)\right)\left(g^{(1)}_{\theta^0}(X)\right)^\top\right] \text{ is positive definite.}$$

- **Assumptions on f_ε**

$$\text{The density } f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \text{ and for all } x \in \mathbb{R},\ f^*_\varepsilon(x) \neq 0. \qquad (\mathbf{N_1})$$

2.3. Conditions on the weight function

Let us now detail some conditions on the weight function w which appears in the contrast function.

The functions (wg_θ) and (wg^2_θ) belong to $\mathbb{L}_1(\mathbb{R}),$ and the functions w^*/f^*_ε, $(g_\theta w)^*/f^*_\varepsilon$, $(g^2_\theta w)^*/f^*_\varepsilon$ belong to $\mathbb{L}_1(\mathbb{R})$. $(\mathbf{C_1})$

$$\text{For any } i \in \{1,\dots,d\},\ \sup_{\theta\in\Theta}\left|\left(g^{(1)}_{\theta,i}w\right)^*/f^*_\varepsilon\right| \text{ and } \sup_{\theta\in\Theta}\left|\left(g_\theta g^{(1)}_{\theta,i}w\right)^*/f^*_\varepsilon\right| \text{ belong to } \mathbb{L}_1(\mathbb{R}). \qquad (\mathbf{C_2})$$

$$\text{For } i,j \in \{1,\dots,d\},\ \sup_{\theta\in\Theta}\left|\left(g^{(2)}_{\theta,i,j}w\right)^*/f^*_\varepsilon\right|, \sup_{\theta\in\Theta}\left|\left((g^2_\theta w)^{(2)}_{\theta,i,j}\right)^*/f^*_\varepsilon\right|, \sup_{\theta\in\Theta}\left|\left(g^{(3)}_{\theta,i,j,k}w\right)^*/f^*_\varepsilon\right| \qquad (\mathbf{C_3})$$

$$\text{and } \sup_{\theta\in\Theta}\left|\left((g^2_\theta w)^{(3)}_{\theta,i,j,k}\right)^*/f^*_\varepsilon\right| \text{ belong to } \mathbb{L}_1(\mathbb{R});$$

$$\text{For } k \in \{1,\dots,d\},\ \int |t(g_{\theta^0}w)^*(t)|\mathrm{d}t \text{ and } \int |t(g_{\theta^0}g^{(1)}_{\theta^0,k}w)^*(t)|\mathrm{d}t \text{ are finite.} \qquad (\mathbf{C_4})$$

The first part of Condition ($\mathbf{C_1}$) essentially ensures that the estimation criterion $S_n(\theta)$ exists for all θ through the existence of $(wg_\theta)^*$ and is not restrictive. The second part of Condition ($\mathbf{C_1}$) can be heuristically expressed as "one can find a weight function w such that wg_θ is smooth enough compared to f_ε". This is satisfied for a large class of functions. Examples are given hereafter. Conditions ($\mathbf{C_2}$)–($\mathbf{C_3}$) are similar to ($\mathbf{C_1}$) (and not more restrictive than ($\mathbf{C_1}$)). Condition ($\mathbf{C_4}$), not restrictive at all, is just technical. It is introduced to ensure the asymptotic normality of the estimator under τ–dependency of the chain $(X_i)_{i\ge 0}$.

2.4. Dependency conditions

The asymptotic properties are stated under dependency properties of the Markov chain $(X_i)_{i\geq 0}$. Those dependency properties are described through the coefficient $\alpha(\mathcal{M}, \sigma(Y))$ which is the usual strong mixing coefficient defined by Rosenblatt [26], and through the coefficient $\tau(\mathcal{M}, Y)$ which has been introduced by Dedecker and Prieur [12].

Definition 2.1. 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 $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. The dependence coefficients α and τ are defined by

$$\alpha(\mathcal{M}, \sigma(Y)) = \frac{1}{2} \sup_{A \in \mathcal{B}(\mathbb{B})} \mathbb{E}(|\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)|),$$

$$\text{and if } \mathbb{E}(\|Y\|_{\mathbb{B}}) < \infty, \quad \tau(\mathcal{M}, Y) = \mathbb{E}\Big(\sup_{f \in \Lambda_1(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(f) - \mathbb{P}_Y(f)| \Big).$$

Definition 2.2. 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

$$\alpha_{\mathbf{X}}(k) = \alpha(\sigma(X_0), \sigma(X_k))$$

$$\text{and if } \mathbb{E}(|X_0|) < \infty, \quad \tau_{\mathbf{X},2}(k) = \sup \left\{ \tau(\sigma(X_0), (X_{i_1}, X_{i_2})), k \leq i_1 \leq i_2 \right\}.$$

Dependency assumptions. We consider the two following conditions:

(α–mixing) The inverse cadlag $Q_{|X_1|}$ of the tail function $t \to \mathbb{P}(|X_1| > t)$ is such that **($\mathbf{D_1}$)**

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

(τ–dependence) Let $G(t) = t^{-1}\mathbb{E}(X_1^2 \mathbf{1}_{X_1^2 > t})$.The inverse cadlag G^{-1} of G is such that **($\mathbf{D_2}$)**

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

Note that, if $\mathbb{E}(|X_0|^p) < \infty$ for some $p > 2$, then Condition ($\mathbf{D_1}$) is satisfied provided that $\sum_{k>0} k^{2/(p-2)} \alpha_{\mathbf{X}}(k) < \infty$, and Condition ($\mathbf{D_2}$) holds provided that $\sum_{k>0} (\tau_{\mathbf{X},2}(k))^{(p-2)/p} < \infty$.

2.5. Examples of models (1.1) satisfying all the assumptions

We give some examples of regression functions for which a weight function w satisfying Conditions ($\mathbf{C_1}$)−($\mathbf{C_4}$) can be exhibited with different noise f_ε (Gaussian or Laplace for instance):

(A-1) Linear function $g_\theta(x) = \theta_1 x + \theta_2$ (see Sect. 5),
(A-2) Cauchy function $g_\theta(x) = \theta/(1 + x^2)$ (see Sect. 5),
(A-3) Gauss function $g_\theta(x) = \exp(-\theta x^2)$,
(A-4) Sinusoidal function $g_\theta(x) = \sum_{j=1}^p \theta_j \sin(jx)$.

Now, assume that $\mathbb{E}(|\xi_0|^s) < \infty$ for some $s > 1$. For θ^0 such that g_{θ^0} is ρ–Lipschitz, for some $\rho < 1$ (for instance if $|\theta_1^0| < 1$ in (A.1), $|\theta^0| < 8\sqrt{3}/9$ in (A.2), $\theta_0 \in]0, e/2[$ in (A.3), and $\sum_{j=1}^p j|\theta_j^0| < 1$ in (A.4)), there exists a unique invariant probability measure π such that $\int |x|^s \pi(\mathrm{d}x) < \infty$, and the coefficient $\tau_{\mathbf{X},2}(k)$ decreases with an exponential rate. Hence, Condition ($\mathbf{D_2}$) holds (see Appendix A for more details).

We could also consider polynomial regression function, but the existence of stationary distribution for X would be ensured only under more restrictive assumptions on the distribution of ξ (compactly supported distribution for instance).

3. Estimation procedure and asymptotic properties

Consider here that the Markov chain $(X_i)_i$ is strictly stationary, with π the invariant probability measure. An extension to π–almost all starting points is given in Section 3.3.1.

3.1. Definition of the estimator

Our method starts from the least square criterion

$$S(\theta) = \mathbb{E}[(Z_1 - g_\theta(X_0))^2\, w(X_0)], \tag{3.3}$$

which is minimum when $\theta = \theta^0$ under the identifiability assumption (**I1₁**).

If (**C₁**) holds, $\mathbb{E}(w(X))$, $\mathbb{E}(w(X)g_\theta(X))$ and $\mathbb{E}(w(X)g_\theta^2(X))$ can be easily estimated. Indeed, for φ such that φ and $\varphi^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})$, by the independence between ε_0 and X_0, we have

$$\mathbb{E}[\varphi(X_0)] = \mathbb{E}\left(\frac{1}{2\pi}\int \varphi^*(t)\mathrm{e}^{-itX_0}\mathrm{d}t\right) = \mathbb{E}\left(\frac{1}{2\pi}\int \frac{\varphi^*(t)\mathrm{e}^{-itZ_0}}{f_\varepsilon^*(-t)}\mathrm{d}t\right). \tag{3.4}$$

Hence, $\mathbb{E}[\varphi(X_0)]$ is estimated by

$$\frac{1}{2\pi}\mathbb{R}e\int \frac{\varphi^*(t)n^{-1}\sum_{j=1}^n \mathrm{e}^{-itZ_j}}{f_\varepsilon^*(-t)}\mathrm{d}t.$$

Following this general idea, we propose to estimate $S(\theta)$ by

$$S_n(\theta) = \frac{1}{2\pi n}\sum_{k=1}^n \mathbb{R}e\int \frac{\left(\left(Z_k - g_\theta\right)^2 w\right)^*(t)\ \mathrm{e}^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}\mathrm{d}t. \tag{3.5}$$

According to (3.4), $\mathbb{E}(S_n(\theta)) = \mathbb{E}[(Z_1 - g_\theta(X_0))^2 w(X_0)]$, and $S_n(\theta)$ is an unbiased estimator of $S(\theta)$. We propose to estimate θ^0 by minimizing the empirical criterion $S_n(\theta)$:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} S_n(\theta). \tag{3.6}$$

The choice of the weight function w is crucial to ensure that Conditions (**C₁**)–(**C₄**) are satisfied. Often, for a specific regression function, various weight functions can handle with Conditions (**C₁**)–(**C₄**). The numerical properties of the resulting estimators will differ from one choice to another. See the simulation study in Section 5.

3.2. Consistency and $\sqrt{n}$–asymptotic normality

We present the asymptotic properties of our estimator. The first result to mention is the consistency of our estimator.

Theorem 3.1. *Under assumptions* (**A₁**)–(**A₂**), (**I1₁**), (**I1₂**), (**N₁**), *and conditions* (**C₁**)–(**C₂**), $\widehat{\theta}$ *defined by* (3.6) *converges in probability to* θ^0.

Now, we state the asymptotic normality of $\widehat{\theta}$ when the Markov chain (X_i) is α–mixing.

Theorem 3.2. *Let Σ_1 be the covariance matrix defined in* (B.4). *Under assumptions* (**A₁**), (**A₂**), (**I1₁**), (**I1₂**), (**N₁**), *conditions* (**C₁**)–(**C₃**) *and α–mixing condition* (**D₁**), *then $\widehat{\theta}$ defined by* (3.6) *is a $\sqrt{n}$–consistent estimator of θ^0 which satisfies*

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

Next, we state the asymptotic normality of $\widehat{\theta}$ when the Markov chain (X_i) is τ–dependent.

Theorem 3.3. *Let Σ_1 be the covariance matrix defined in* (B.4). *Under assumptions* (**A₁**), (**A₂**), (**I1₁**), (**I1₂**), (**N₁**), *Conditions* (**C₁**)–(**C₄**) *and τ–dependence condition* (**D₂**), *then $\widehat{\theta}$ defined by* (3.6) *is a $\sqrt{n}$–consistent estimator of θ^0 which satisfies*

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

Our estimation procedure allows to achieve the parametric rate for a large class of regression functions, larger than what was previously proposed in the literature. A first example is the sinusoidal regression function. Whatever the error distribution, the rate proposed in the literature was always slower than the parametric rate, that we obtain here. For example, with Gaussian errors, the best proposed rate was of order $\exp\{p\sqrt{\log n}\}/\sqrt{n}$ (see Comte and Taupin [9]). A second example is the Cauchy regression function. To our knowledge, $\hat{\theta}$ is the first consistent estimator proposed in the literature.

Finally, Theorems 3.2 and 3.3 do not require the Markov chain to be absolutely regular. Consequently they apply to autoregressive models with weak dependency conditions (see examples in Sect. 2.5). This was not the case in Comte and Taupin [9].

3.3. Extensions

3.3.1. Results for almost all starting points

We still denote by X_i the strictly stationary Markov chain with invariant distribution π, and by X_i^x the chain starting from the point x. We observe $Z_k^x = X_k^x + \varepsilon_k$. We define then the empirical contrast with initial condition x by

$$S_n^x(\theta) = \frac{1}{2\pi n}\sum_{k=1}^{n} \mathbb{R}e \int \frac{\left(\left(Z_k^x - g_\theta\right)^2 w\right)^*(t)\, \mathrm{e}^{-itZ_{k-1}^x}}{f_\varepsilon^*(-t)}\mathrm{d}t. \tag{3.7}$$

We propose to estimate θ^0 by

$$\widehat{\theta}(x) = \arg\min_{\theta\in\Theta} S_n^x(\theta). \tag{3.8}$$

The following results hold

- Under the assumptions of Theorem 3.1, for π–almost all starting point x, the estimator $\widehat{\theta}(x)$ defined by (3.8) converges in probability to θ^0.
- Under the assumptions of Theorem 3.2 or of Theorem 3.3, where Σ_1 is defined in (B.4), for π–almost every starting point x, the estimator $\widehat{\theta}(x)$ defined by (3.8) is a $\sqrt{n}$–consistent estimator of θ^0 which satisfies

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

The proof of the consistency follows exactly that of Theorem 3.1 (see Appendix B.1), by replacing $S_n(\theta)$ by $S_n^x(\theta)$ and by noting that, by the ergodic theorem: for π–almost all starting point x, $S_n^x(\theta)$ converges in $\mathbb{L}^1$ to $S(\theta)$, and $\mathbb{E}(\sup_{\theta\in\Theta^0} \parallel (S_n^x)^{(1)}(\theta) \parallel_{\ell^2})$ is bounded thanks to assumption (**C₂**). The asymptotic normality for π–almost all starting points is true because the proofs of Theorems 3.2 and 3.3 are based on condition (B.6) of Appendix B.2. Now, under Condition (B.6), the central limit theorem holds for π–almost starting points, as proved in Theorem 2.1 of Dedecker *et al.* [11].

3.3.2. Results for martingale differences innovations ξ

We consider now a more general model: $(X_i)_{i\geq 0}$ is a Markov chain which admits an ergodic invariant probability, and the autoregression function is known up to a finite dimensional parameter θ, that is

$$g_{\theta^0}(x) = \mathbb{E}(X_1|X_0 = x).$$

We observe $Z_i = X_i + \varepsilon_i$ for $i = 0, \dots, n$, where $(\varepsilon_i)_{i\geq 0}$ is i.i.d. and independent of $(X_i)_{i\geq 0}$.

Clearly, this model is a generalization of (1.1), since we do not assume, that $(\xi_i)_{i\geq 1}$ is i.i.d., but only that $(\xi_i)_{i\geq 1}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_i = \sigma(X_k, k \leq i)$. Under the same assumptions, and following the same proofs, the estimator $\hat{\theta}$ defined by (3.6) has exactly the same asymptotic properties as those described in Section 3.2. This is mainly due to the fact that, since $\mathbb{E}(\xi_1|X_0) = 0$, we have $S(\theta) = \mathbb{E}[(Z_1 - g_\theta(X_0))^2\, w(X_0)] = \mathbb{E}[(g_{\theta^0}(X_0) - g_\theta(X_0))^2\, w(X_0)] + \mathbb{E}[(\xi_1^2 + \varepsilon_1^2)w(X_0)]$. And then $S(\theta)$ is minimal for $\theta = \theta^0$.

Let us give an example of this more general model. Consider the Markov chain

$$X_i = A_i f(X_{i-1}) + B_i,$$

where the coefficients $(A_i, B_i)_{i\geq 1}$ are random, i.i.d. and independent of X_0. Here, for $a = \mathbb{E}(A_1)$ and $b = \mathbb{E}(B_1)$, $\mathbb{E}(X_i|X_{i-1} = x) = af(x) + b$. Thus, $X_i = af(X_{i-1}) + b + \xi_i$ with non independent ξ_i's. An interesting example is the random AR(1) model $X_i = A_i X_{i-1} + B_i$, where we want to estimate $(\mathbb{E}(A_1), \mathbb{E}(B_1))$ from the observations $Z_k = X_k + \varepsilon_k$.

If $\theta^0 = (a, b)$, under the assumptions of Theorem 3.2 or 3.3, $\hat{\theta}$ defined by (3.6) is a consistent and asymptotically normal estimator of θ^0. Note that if $\mathbb{E}(|A_1|) < \infty$ and $\mathbb{E}(|B_1|) < \infty$, and if $|f(x) - f(y)| \leq \kappa|x - y|$ for $\kappa < 1/\mathbb{E}(|B_1|)$, then the chain is τ–dependent with $\tau_{\mathbf{X},2}(k) = O(\kappa^k)$.

4. A more general estimator

For some specific regression functions it seems not straightforward to find a weight function satisfying conditions ($\mathbf{C_1}$)–($\mathbf{C_4}$), for instance if $g_\theta(x) = \theta\mathbb{1}_{[0,1]}$.

In this section we propose a generalization of our estimator under weaker conditions than Conditions ($\mathbf{C_1}$)–($\mathbf{C_4}$). Here the Markov chain (X_i) is assumed to be strictly stationary.

4.1. Definition of the general estimator

The key idea for this construction relies on deconvolution tools and more specifically, it relies to a truncation of integrals in (3.5). Let K_{n,C_n} be a density deconvolution kernel defined *via* its Fourier transform

$$K^*_{n,C_n}(t) = \frac{K^*(t/C_n)}{f^*_\varepsilon(-t)} := \frac{K^*_{C_n}(t)}{f^*_\varepsilon(-t)}, \tag{4.9}$$

where K^* is the Fourier transform of K and C_n is a sequence which tends to infinity with n. The kernel K is chosen to belong to $\mathbb{L}^2(\mathbb{R})$ with a compactly supported Fourier transform K^* and satisfying $|1 - K^*(t)| \leq \mathbb{1}_{|t|\geq 1}$. Then, for any integrable function Φ, one has $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(Z_i) = \mathbb{E}(\Phi(X))$. Hence we estimate $\mathbb{E}(\Phi(X))$ by $n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(Z_i)$.

We propose to estimate θ^0 by

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \quad \text{with} \quad S_n(\theta) = \frac{1}{n}\sum_{i=1}^n \mathbb{R}\mathrm{e}\int (Z_i - g_\theta(x))^2\, w(x)\, K_{n,C_n}(Z_{i-1} - x)\mathrm{d}x. \tag{4.10}$$

This procedure still works under ($\mathbf{C_1}$)–($\mathbf{C_4}$) by choosing $K^*(t/C_n) = \mathbb{1}_{|t|\leq C_n}$ with $C_n = +\infty$.

4.2. Asymptotic properties under general conditions

Under milder conditions than (**C₁**)–(**C₄**), the estimator $\widehat{\theta}$ defined in (4.10) is still consistent, but its rate is not necessarily the parametric rate. For the sake of simplicity we only consider the case of α–mixing Markov chains.

We assume that

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

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

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

X_0 admits a density f_X with respect to the Lebesgue measure and there exist two constants $C_1(g^2_{\theta^0})$ and $C_2(g_{\theta^0})$ such that $\| g_{\theta^0} f_X \|_2^2 \leq C_1(g_{\theta^0})$, and $\| g^2_{\theta^0} f_X \|_2^2 \leq C_2(g^2_{\theta^0})$. $\quad (\mathbf{A_6})$

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

We say that a function $\psi\in\mathbb{L}_1(\mathbb{R})$ satisfies (4.11) if for a sequence C_n we have

$$\min_{q=1,2} \| \psi^*(K^*_{C_n}-1) \|_q^2 + n^{-1}\min_{q=1,2}\left\|\frac{\psi^* K^*_{C_n}}{f^*_\varepsilon}\right\|_q^2 = o(1). \tag{4.11}$$

Theorem 4.1. *Under assumptions* (**I1₁**), (**I1₂**), (**N₁**), (**A₁**) (**A₃**)–(**A₅**), *let $\widehat{\theta}$ defined in* (4.10) *with C_n such that* (4.11) *holds for w, wg_θ and wg_θ^2 and their first derivatives with respect to θ. Assume that the sequence (X_k) is α–mixing that is*

$$\alpha_{\mathbf{X}}(k) \underset{n\to\infty}{\longrightarrow} 0, \ \textit{as} \quad k \underset{n\to\infty}{\longrightarrow} \infty.$$

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 now give upper bounds for the rate of convergence under assumptions (**A₆**)–(**A₇**). The following theorem still holds when X_0 does not admit a density, under a slightly different moment assumption.

Theorem 4.2. *Suppose that the assumptions of Theorem 4.1 hold. Assume moreover that the sequence $(X_k)_{k\geq 0}$ is α–mixing with $\sum_{k\geq1}\sqrt{\alpha_{\mathbf{X}}(k)}<\infty$, and that, for all $\theta\in\Theta$, the functions w, $g_\theta w$ and $g_\theta^2 w$ and their derivatives up to order 3 with respect to θ satisfy* (4.11).

1) *Assume that the sequence X_0 admits a density with respect to the Lebesgue measure and that assumption* (**A₆**) *holds. Then $\widehat{\theta}-\theta^0 = O_p(\varphi_n^2)$ with $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi^2_{n,j} = B^2_{n,j} + V_{n,j}/n$, $j=1\dots,d$, where*

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

and for $q=1,2$

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

and

$$V^{[q]}_{n,j}{=}\left\|(wf^{(1)}_{\theta^0,j})^*\frac{K^*_{C_n}}{f^*_\varepsilon}\right\|_q^2 + \left\|(wg_{\theta^0}f^{(1)}_{\theta^0,j})^*\frac{K^*_{C_n}}{f^*_\varepsilon}\right\|_q^2.$$

2) *Under* ($\mathbf{A_7}$), $\widehat{\theta}-\theta^0=O_p(\varphi_n^2)$ *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, g_θ and f_ε. We refer to Table 1 in Butucea and Taupin [6] for more details on the resulting rates.

5. Simulation study

We investigate the numerical properties of our estimator for different regression functions and error distributions, on simulated data. We consider two error distributions: the Laplace distribution and the Gaussian distribution. When ε_1 is centered with variance σ_ε^2 and the Laplace distribution, its density and Fourier transform are

$$f_\varepsilon(x)=\frac{1}{\sigma_\varepsilon\sqrt{2}}\exp\Big(-\frac{\sqrt{2}}{\sigma_\varepsilon}|x|\Big),\ \text{and}\ f_\varepsilon^*(x)=\frac{1}{1+\sigma_\varepsilon^2x^2/2}\cdot \tag{5.12}$$

When ε_1 is centered with variance σ_ε^2 and Gaussian, its density and Fourier transform are

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

For each of these error distributions, we consider linear and Cauchy regression functions.

5.1. Linear regression function

Consider Model (1.1) with $g_\theta(x)=ax+b$, where $|a|<1$ and $\theta=(a,b)^T$. Here, we choose to illustrate the numerical properties of our estimator under the weakest of the dependency conditions, that is τ–dependency. As recalled in Appendix 2.5, when g_{θ^0} is linear with $|a|<1$, if ξ_0 has a density bounded from below in a neighborhood of the origin, then the Markov chain $(X_i)_{i\geq 0}$ is α–mixing. When ξ_0 does not have a density, then the chain may not be α–mixing (and not even irreducible), but it is always τ–dependent.

Here, we consider the case of discrete innovation distribution, in such a way that the stationary Markov Chain is τ–dependent but not α–mixing. We also consider two distinct values of θ_0. For the first value, the stationary distribution of X_i is absolutely continuous with respect to the Lebesgue measure. For the second value, the stationary distribution is singular with respect to the Lebesgue measure. In both cases Theorem 3.3 applies, and $\hat{\theta}$ is asymptotically Gaussian.

• **Case A** (absolutely continuous stationary distribution). If $\theta^0=(1/2,1/4)^T$, X_0 is uniformly distributed over $[0,1]$, and $(\xi_i)_{i\geq 1}$ is 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 strictly stationary Markov chain is defined for $i>0$ by

$$X_i=\frac{1}{4}+\frac{1}{2}X_{i-1}+\xi_i. \tag{5.14}$$

Its stationary distribution is the uniform distribution over $[0,1]$, with $\sigma_{X_0}^2=1/12$. This chain is non−irreducible, and the dependency coefficients are such that $\alpha_{\mathbf{X}}(k)=1/4$ (see for instance Bradley [4], p. 180) and $\tau_{\mathbf{X},2}(k)=O(2^{-k})$. Thus the Markov chain is not α–mixing, but it is τ–dependent. We start the simulation with X_0 uniformly distributed over $[0,1]$, so the simulated chain is stationary.

• **Case B** (singular stationary distribution). If $\theta^0=(1/3,1/3)^T$, X_0 is uniformly distributed over the Cantor set, and $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, independent of X_0 and such that $\mathbb{P}(\xi_1=-1/3)=\mathbb{P}(\xi_1=1/3)=1/2$. Hence, the strictly stationary Markov chain is defined for $i>0$ by

$$X_i=\frac{1}{3}+\frac{1}{3}X_{i-1}+\xi_i. \tag{5.15}$$

Its stationary distribution is the uniform distribution over the Cantor set, with $\sigma_X^2 = 1/8$. This chain is non−irreducible, and the dependency coefficients satisfy $\alpha_{\mathbf{X}}(k) = 1/4$ and $\tau_{\mathbf{X},2}(k) = O(3^{-k})$. The Markov chain is not α–mixing, but is τ–dependent. For the simulation, we start with X_0 uniformly distributed over $[0,1]$, and set $X_i = X_{i+1000}$ since we consider that the chain is close to the stationary chain after 1000 iterations.

We first give the explicit expression of $\hat{\theta}$ for two choices of weight functions w, satisfing conditions ($\mathbf{C_1}$)−($\mathbf{C_4}$), and recall the classic estimator when X is directly observed, the ARMA estimator, and the so-called naive estimator.

5.1.1. Expression of the estimator.

Set

$$w(x) = N(x) = \exp\{-x^2/(4\sigma_\varepsilon^2)\} \quad \text{and} \quad w(x) = SC(x) = \frac{1}{2\pi}\left(\frac{2\sin(x)}{x}\right)^4 \cdot \tag{5.16}$$

Conditions ($\mathbf{C_1}$)−($\mathbf{C_4}$) hold for both weight functions N and SC, and the two associated estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ are $\sqrt{n}$–consistent estimator of θ^0. 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 error variance as it appears in the simulations (see Fig. 1). Second, SC has strong smoothness properties since its Fourier transform is compactly supported.

The two associated estimators are based on $S_n(\theta)$ expressed as

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}[(Z_k^2 + b^2 - 2Z_k b)I_0(Z_{k-1}) + a^2 I_2(Z_{k-1}) - 2a(Z_k - b)I_1(Z_{k-1})],$$

$$\text{with} \quad I_j(Z) = \frac{1}{2\pi}\mathbb{R}e\int (p_j w)^*(u)\frac{\mathrm{e}^{-iuZ}}{f_\varepsilon^*(-u)}\mathrm{d}u, \tag{5.17}$$

where $p_j(x) = x^j$ for $j = 0,1,2$, w being either $w = N$ or $w = SC$. Hence, $\widehat{\theta} = (\widehat{a},\widehat{b})^T$ satisfies

$$\widehat{a} = \frac{\sum_{k=1}^n Z_k I_1(Z_{k-1})\sum_{k=1}^n I_0(Z_{k-1}) - \sum_{k=1}^n Z_k I_0(Z_{k-1})\sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_2(Z_{k-1})\sum_{k=1}^n I_0(Z_{k-1}) - \big(\sum_{k=1}^n I_1(Z_{k-1})\big)^2}, \tag{5.18}$$

$$\widehat{b} = \frac{\sum_{k=1}^n Z_k I_0(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})} - \widehat{a}\frac{\sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})}\cdot \tag{5.19}$$

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 give the details of the calculations for the two error distributions (Laplace and Gaussian), which 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

$$I_{0,N}(Z) = \frac{1}{2\pi}\mathbb{R}e\int\sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\frac{\mathrm{e}^{-itZ}}{f_\varepsilon^*(-t)}\mathrm{d}t,$$
$$I_{1,N}(Z) = \frac{1}{2\pi}\mathbb{R}e\int\sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 t/i\big)\frac{\mathrm{e}^{-itZ}}{f_\varepsilon^*(-t)}\mathrm{d}t,$$
$$I_{2,N}(Z) = \frac{1}{2\pi}\mathbb{R}e\int\sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(2\sigma_\varepsilon^2 - 4\sigma_\varepsilon^4 t^2\big)\frac{\mathrm{e}^{-itZ}}{f_\varepsilon^*(-t)}\mathrm{d}t.$$

If f_ε is the Laplace distribution (5.12), replacing f_ε^* by its expression we get

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

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

If f_ε is the Gaussian distribution (5.13), replacing f_ε^* by its expression we obtain

$$I_{0,N}(Z) = \sqrt{2}\mathrm{e}^{-Z^2/(2\sigma_\varepsilon^2)}, \quad I_{1,N}(Z) = 2\sqrt{2}Z\mathrm{e}^{-Z^2/(2\sigma_\varepsilon^2)} \quad \text{and} \quad I_{2,N}(Z) = \sqrt{2}(4Z^2 - 2\sigma_\varepsilon^2)\mathrm{e}^{-Z^2/(2\sigma_\varepsilon^2)}.$$

Hence we deduce the expression of $\widehat{a}_N$ and $\widehat{b}_N$ by applying (5.18) and (5.19).

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

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

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

$$I_{j,SC}(Z) = \frac{1}{2\pi}\mathbb{R}e\int(SCp_j)^*(t)\frac{\mathrm{e}^{-itZ}}{f_\varepsilon^*(-t)}\mathrm{d}t, \tag{5.20}$$

have no explicit form and have to be numerically computed. 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 results are taken as approximations of $I_{j,SC}(Z)$, in (5.18) and (5.19) to get $\widehat{a}_{SC}$ and $\widehat{b}_{SC}$.

5.1.2. Comparison with classical estimators

We compare the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ with three classical estimators, the usual least square estimator when there is no observation noise, the ARMA estimator, and the so-called naive estimator.

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

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

• ARMA estimator. When the regression function is linear, the model may be written as

$$Z_i - aZ_{i-1} - b = \xi_i + \varepsilon_i - a\varepsilon_{i-1}.$$

The auto-covariance function γ_Y of the stationary sequence $Y_i = \xi_i + \varepsilon_i - a\varepsilon_{i-1}$ is given by

$$\gamma_Y(0) = (1+a^2)\sigma_\varepsilon^2 + \sigma_\xi^2, \quad \gamma_Y(1) = -a\sigma_\varepsilon^2, \quad \text{and } \gamma_Y(k) = 0 \text{ for } k > 1.$$

It follows that Y_i is an MA(1) process, which may be written as

$$Y_i = \eta_i - \beta\eta_{i-1},$$

where η_i is the innovation, and $|\beta| < 1$ (note that $|\beta| \neq 1$ because $\gamma_Y(0) - 2|\gamma_Y(1)| > 0$). Moreover, one can give the explicit expression of β and σ_η^2 in terms of a, σ_ξ^2 and σ_ε^2. It follows that, if $|a| < 1$, $(Z_i)_{i\geq 0}$ is the causal invertible ARMA(1,1) process

$$Z_i - aZ_{i-1} = b + \eta_i - \beta\eta_{i-1}. \tag{5.21}$$

Note that $a \neq \beta$ except if $a = 0$. Hence, if $|a| < 1$ and $a \neq 0$, one can estimate the parameters (a, b, β) by maximizing the so-called Gaussian likelihood. These estimators are consistent and asymptotically Gaussian. Moreover they are efficient when both the innovations and the errors ε are Gaussian (see Hannan [19] or Brockwell and Davis [5]). Note that this well−known approach does not require the knowledge of the error distribution, but of course it works only in the particular case where the regression function g_θ is linear. For the computation of the ARMA estimator we use the function *arma* from the R *tseries* package (see Trapletti and Hornik [28]). The resulting estimators are denoted by $\widehat{a}_{\text{arma}}$ and $\widehat{b}_{\text{arma}}$.

• Naive estimator. The naive estimator is constructed by replacing the unobserved X_i by the observation Z_i in $\widehat{a}_X$ and $\widehat{b}_X$:

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

As confirmed by the simulation study, $\widehat{\theta}_{\text{naive}}$ is an asymptotically biased estimator of θ^0.

5.1.3. Simulation results

For each error distribution, we simulate 100 samples with size n, $n = 500$, 5000 and 10 000. We consider different values of σ_ε such that the ratio signal to noise $s2n = \sigma_\varepsilon^2/\text{Var}(X)$ is 0.5, 1.5 or 3. The comparison of the five estimators is based on the bias, the Mean Squared Error (MSE), and the box plots. 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:

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

Results are presented in Figures 1−2 and Tables 1−4.

The first thing to notice is that, not surprisingly, $\widehat{\theta}_{\text{naive}}$ 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 $\widehat{\theta}_{\text{arma}}$ has good asymptotic properties. However its bias is often larger than the biases of $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$, except when $s2n = 0.5$ and ε is Gaussian.

We now consider the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$. 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. But their numerical behaviors are not the same. For not too large $s2n$, $\widehat{\theta}_{SC}$ has a MSE smaller than $\widehat{\theta}_N$ (see Fig. 1 and Tables 1−4, when $s2n \leq 3$). With large $s2n$, the estimator $\widehat{\theta}_N$ seems to have better properties (see Fig. 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 error 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 with the theoretical properties in density deconvolution. In that context

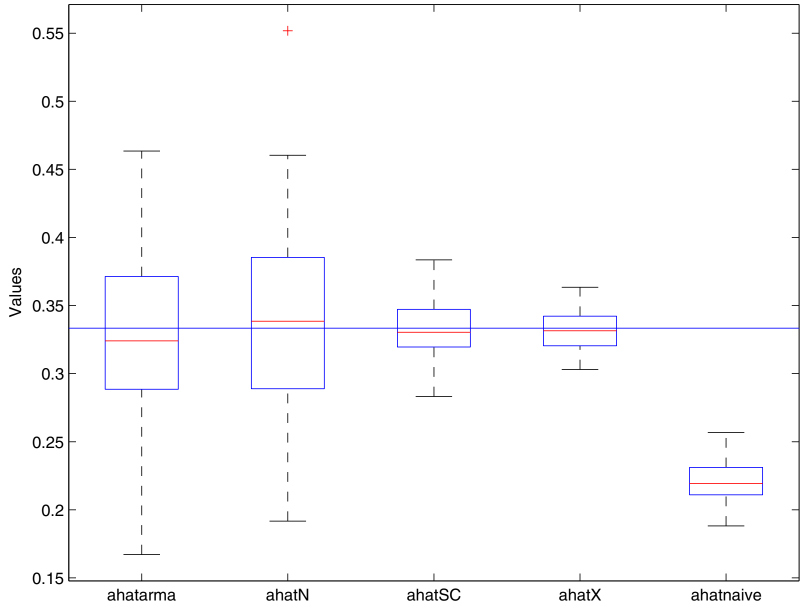

FIGURE 1. Results for linear Case B and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 0.5$. Box plots of the five estimators $\widehat{a}_{\mathrm{arma}}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{\mathrm{naive}}$, from left to right, based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

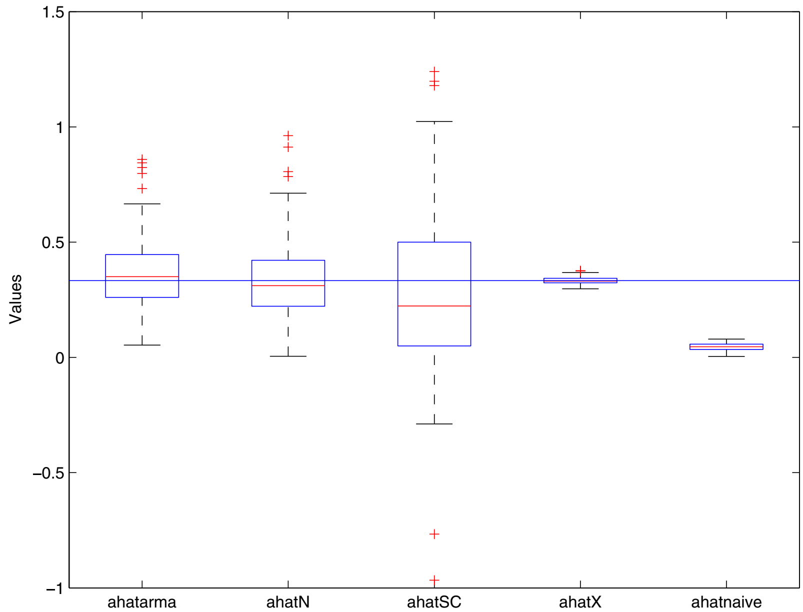

FIGURE 2. Results for linear Case B and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 6$. Box plots of the five estimators $\widehat{a}_{\mathrm{arma}}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{\mathrm{naive}}$, from left to right, based on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

TABLE 1. Estimation results for Linear Case A, Laplace error. Mean estimated values of the five estimators $\widehat{\theta}_{\text{arma}}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$ are presented for various values of n (1000, 5000 or 10 000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

n	ratio s2n		$\widehat{\theta}_{\text{arma}}$(MSE)	$\widehat{\theta}_N$(MSE)	Estimator $\widehat{\theta}_{SC}$(MSE)	$\widehat{\theta}_X$(MSE)	$\widehat{\theta}_{\text{naive}}$(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)
10 000	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)

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 $\widehat{\theta}_{\text{arma}}$ 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.

5.2. Cauchy regression model

We consider the model (1.1) with $g_\theta(x) = \theta/(1+x^2) = \theta f(x)$. The true parameter is $\theta^0 = 1.5$. For the law of ξ_0 we take $\xi_0 \sim \mathcal{N}(0, 0.01)$. In this case, an empirical study shows that σ_X^2 is about 0.1. Moreover $\alpha_{\mathbf{X}}(k) = O(\kappa^k)$ for some $\kappa \in]0,1[$ and the Markov chain is α–mixing (see Appendix 2.5). For w suitably chosen, Theorem 3.2 applies and states that $\hat{\theta}$ is asymptotically normal. 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 $X_i = X_{i+1000}$.

To our knowledge, the estimator $\widehat{\theta}$ is the first consistent estimator in the literature for this regression function. We first detail the 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.

TABLE 2. Estimation results for Linear Case A, Gaussian error. Mean estimated values of the five estimators $\widehat{\theta}_{\text{arma}}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$ are presented for various values of n (1000, 5000 or 10 000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

	ratio		Estimator				
n	s2n		$\widehat{\theta}_{\text{arma}}$(MSE)	$\widehat{\theta}_N$(MSE)	$\widehat{\theta}_{SC}$(MSE)	$\widehat{\theta}_X$(MSE)	$\widehat{\theta}_{\text{naive}}$(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)
10 000	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)

5.2.1. Expression of the estimator

We consider the two following weight functions:

$$N_c(x) = (1+x^2)^2 \exp\{-x^2/(4\sigma_\varepsilon^2)\} \text{ and } SC_c(x) = (1+x^2)^2 \frac{1}{2\pi}\left(\frac{2\sin(x)}{x}\right)^4, \tag{5.22}$$

These choices of w ensure that conditions (**C₁**)–(**C₄**) hold and $\hat{\theta}$ is asymptotically Gaussian. As in the linear case, these two weight functions differ by their dependence on σ_ε^2 and their smoothness properties. The two estimators are based on the expression of $S_n(\theta)$,

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

where

$$I_w(Z) = \frac{1}{2\pi}\mathbb{R}e\int (w)^*(u)\frac{\mathrm{e}^{-iuZ}}{f_\varepsilon^*(-u)}\mathrm{d}u, \qquad I_{wf}(Z) = \frac{1}{2\pi}\mathbb{R}e\int (wf)^*(u)\frac{\mathrm{e}^{-iuZ}}{f_\varepsilon^*(-u)}\mathrm{d}u$$

$$\text{and} \quad I_{wf^2}(Z) = \frac{1}{2\pi}\mathbb{R}e\int (wf^2)^*(u)\frac{\mathrm{e}^{-iuZ}}{f_\varepsilon^*(-u)}\mathrm{d}u.$$

The estimator can be expressed as

$$\widehat{\theta} = \frac{\sum_{k=1}^n Z_k I_{wf}(Z_{k-1})}{\sum_{k=1}^n I_{wf^2}(Z_{k-1})}. \tag{5.23}$$

TABLE 3. Estimation results for Linear Case B, Laplace error. Mean estimated values of the five estimators $\widehat{\theta}_{\text{arma}}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$ are presented for various values of n (1000, 5000 or 10 000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

n	ratio s2n		$\widehat{\theta}_{\text{arma}}$(MSE)	$\widehat{\theta}_N$(MSE)	$\widehat{\theta}_{SC}$(MSE)	$\widehat{\theta}_X$(MSE)	$\widehat{\theta}_{\text{naive}}$(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)
10 000	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)

We 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 w is either $w = N_c$ or $w = SC_c$, and set $\widehat{\theta}_{N_c}$ and $\widehat{\theta}_{SC_c}$ the corresponding estimators of θ^0.

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

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

If f_ε is the Laplace distribution (5.12), replacing f_ε^* by its expression we obtain

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

If f_ε is the Gaussian distribution (5.13), replacing f_ε^* by its expression we obtain

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

- When $w = SC_c$, $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ defined by (5.20) satisfy

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

As before, $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ have no explicit form, and are numerically approximated *via* the IFFT function.

Table 4. Estimation results for Linear Case B, Gaussian error. Mean estimated values of the five estimators $\widehat{\theta}_{\mathrm{arma}}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\mathrm{naive}}$ are presented for various values of n (1000, 5000 or 10 000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

n	ratio s2n		$\widehat{\theta}_{\mathrm{arma}}$(MSE)	$\widehat{\theta}_N$(MSE)	$\widehat{\theta}_{SC}$(MSE)	$\widehat{\theta}_X$(MSE)	$\widehat{\theta}_{\mathrm{naive}}$(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)
10 000	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)

5.2.2. Comparison with classical estimators.

We compare $\widehat{\theta}_{N_c}$ and $\widehat{\theta}_{SC_c}$ with two estimators, the usual least square estimator without observation noise, and the naive estimator.

• Estimator without noise. When $\varepsilon_i = 0$, $(X_0, \dots, X_n)$ is observed without errors, and θ^0 is simply estimated by the usual least square estimator

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

• Naive estimator. As confirmed by the simulation study, the naive estimator which is an asymptotically biased estimator of θ^0, can be expressed as

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

5.2.3. Simulations results

For each error distribution, we simulate 100 samples with size n, $n = 500$, 5000 and 10 000. We consider different values of σ_ε 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 four estimators is based on the bias, the Mean Squared Error (MSE), and the box plots, presented in Figure 3 and Tables 5–6.

TABLE 5. Estimation results for Cauchy, Laplace error. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$ are presented for various values of n (1000, 5000 or 10 000) 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)	$\widehat{\theta}_{\text{naive}}$(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)
10 000	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 6. Estimation results for Cauchy, Gaussian error. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$ are presented for various values of n (1000, 5000 or 10 000) 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)	$\widehat{\theta}_{\text{naive}}$(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)
10 000	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)

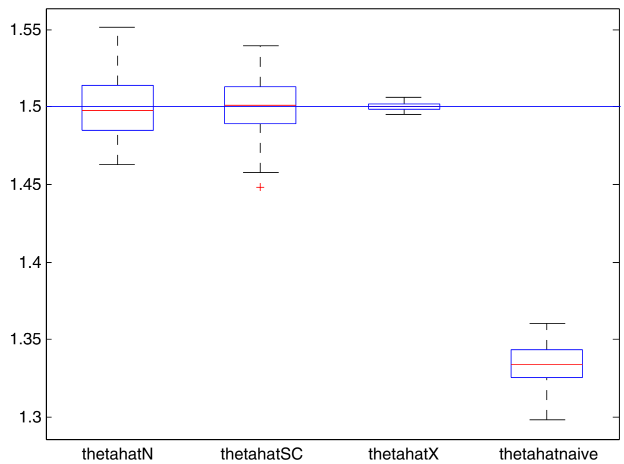

FIGURE 3. Results for Cauchy and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\text{Var}(X) = 1.5$. Box plots of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{\text{naive}}$, from left to right, based on 100 replications.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

Not surprisingly, $\widehat{\theta}_{\text{naive}}$ presents a bias and converges to (false) values which are different according to $s2n$ (see Tabs. 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.

The two estimators $\widehat{\theta}_{SC_c}$ and $\widehat{\theta}_{N_c}$ 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 $s2n$.

Appendix A. Covariance inequalities and coupling

The following results are the key arguments to prove the asymptotic normality of $\widehat{\theta}$.

With the same notations as in Definition (2.1), we first recall a covariance inequality due to Rio [25]. 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. The following inequality holds

$$|\text{Cov}(Y, X)| \leq 4 \int_0^{\alpha(\sigma(Y),\sigma(X))} Q_{|X|}(u) Q_{|Y|}(u) \mathrm{d}u. \tag{A.1}$$

Next, we recall the coupling properties of τ (see Dedecker and Prieur [12]): enlarging Ω if necessary, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$\tau(\mathcal{M}, X) = \mathbb{E}(\|X - X^*\|_{\mathbb{B}}). \tag{A.2}$$

Let us now give some specific criteria for autoregressive models illustrating dependence conditions (**$\mathbf{D_1}$**) or (**$\mathbf{D_2}$**), as described in particular in Mokkadem [22] and Ango−Nzé [2].

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 $(|\xi_0|^s) < \infty$.
- g_{θ^0} is continuous and there exist $r \geq 1$ and $\rho \in]0, 1[$ such that: for any $|x| \geq r$, $|g_{\theta^0}(x)| \leq \rho|x|$.

The stationary Markov chain $(X_i)_{i\geq 0}$ admits an unique invariant distribution measure, and it satisfies $\alpha_{\mathbf{x}}(k) = o(\kappa^k)$ for any $\kappa \in]\rho, 1[$ and is α–mixing.
Now, if the second point is weakened to

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

then there exists a unique invariant probability measure, and the stationary markov chain $(X_i)_{i\geq 0}$ satisfies $\alpha_{\mathbf{x}}(k) = o(k^{1-s/\delta})$ and is α–mixing.

If we do not assume that ξ_0 has a density, then the chain may not be α–mixing (and not even irreducible). However, under appropriate assumptions on g_{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 $(|\xi_0|^s) < \infty$.
- $|g_{\theta^0}(x) - g_{\theta^0}(y)| \leq \rho|x - y|$ 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 $\tau_{\mathbf{x},2}(k) = o(\rho^k)$ and is τ–dependent. Now if the second point is weakened to

- there exist δ in $[0, 1[$ and c in $]0, 1]$ such that $|g'_{\theta^0}(t)| \leq 1 - c(1 + |t|)^{-\delta}$ 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 $\tau_{\mathbf{x},2}(n) = o(n^{(\delta+1-s)/\delta})$ and is τ–dependent.

Appendix B. Proofs of theorems

B.1. Proof of Theorem 3.1

The proof mainly consists in showing the two following points

i) for any θ in Θ, $S_n(\theta) \xrightarrow[n\to\infty]{\mathbb{L}^1} S(\theta)$, with $S(\theta)$ admitting a unique minimum in $\theta = \theta^0$.
ii) For $\omega_2(n,\rho) = \sup\{|S_n(\theta) - S_n(\theta')| : \|\theta-\theta'\|_{\ell^2} \leq \rho\}$, there exists a sequence ρ_k tending to 0, such that

$$\mathbb{E}(\omega_2(n,\rho_k)) = O(\rho_k). \tag{B.1}$$

Let us start with the proof of i) by writing $S_n(\theta)$ as a function of a strictly stationary and ergodic sequence of random variables

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

We apply the ergodic theorem under Assumption (**A_2**) and conclude that for any $\theta \in \Theta$,

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

Now, combining Assumption (**C_2**) and the fact that

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

we infer that there exists a sequence ρ_k tending to 0, such that (B.1) holds.

B.2. Proof of Theorem 3.2

From the smoothness properties of $\theta \mapsto wg_\theta$ and the consistency of $\widehat{\theta}$, we have $S_n^{(1)}(\widehat{\theta}) = S_n^{(1)}(\theta^0) + S_n^{(2)}(\theta^0)(\widehat{\theta} - \theta^0) + R_n(\widehat{\theta}-\theta^0) = 0$, with $R_n = \int_0^1 [S_n^{(2)}(\theta^0 + s(\widehat{\theta}-\theta^0)) - S_n^{(2)}(\theta^0)]ds$. This implies that

$$\widehat{\theta} - \theta^0 = -[S_n^{(2)}(\theta^0) + R_n]^{-1} S_n^{(1)}(\theta^0). \tag{B.3}$$

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) $S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S^{(2)}(\theta^0)$;
iii) $R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Note that the covariance matrix $\Sigma_{0,1}$ in i) satisfies $\Sigma_{0,1} = \Sigma/4\pi^2$, with Σ defined by the equation (B.5) below. According to ii) and iii), the covariance matrix Σ_1 satisfies

$$\Sigma_1 = \frac{1}{4\pi^2}[S^{(2)}(\theta^0)]^{-1}\Sigma[S^{(2)}(\theta^0)]^{-1}, \quad \text{with } \Sigma \text{ defined by (B.5)}. \tag{B.4}$$

We only detail the proof of i). The proofs of ii) and iii) follow by the same arguments.

Under Assumption (**C_2**),

$$\left(\sqrt{n}S_n^{(1)}(\theta^0)\right)_i = \frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n \mathbb{R}e\int\left(\frac{\partial}{\partial\theta_i}((Z_k-g_\theta)^2)w\Big|_{\theta=\theta^0}\right)^*(t)\frac{\mathrm{e}^{-itZk-1}}{f_\varepsilon^*(-t)}\mathrm{d}t.$$

We have thus to prove that

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

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

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

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

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

Let $\mathcal{M}_1=\sigma(X_0,X_1,\varepsilon_0,\varepsilon_1)$. According to Dedecker and Rio [13], $n^{-1/2}\sum_{k=1}^{n}T_k$ converges to a centered Gaussian vector with covariance matrix

$$\Sigma=\mathrm{Cov}(T_1,T_1)+2\sum_{k>1}\mathrm{Cov}(T_1,T_k), \tag{B.5}$$

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

$$\sum_{k=3}^{\infty}\mathbb{E}|(T_1)_p\mathbb{E}((T_k)_q|\mathcal{M}_1)|<\infty. \tag{B.6}$$

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

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

The sequence $(\varepsilon_k,\varepsilon_{k-1})$ is independent of $\mathcal{M}_1\vee\sigma(X_k,X_{k-1})$, for $i,j\in\{1,2\}$, and thus

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

$$\begin{aligned}\text{with }(\tilde{W}_{k,1})_q&=X_k\int\big(g_{\theta^0,q}^{(1)}w\big)^*(t)\mathrm{e}^{-itX_{k-1}}\mathrm{d}t-\mathbb{E}\left[X_k\int\big(g_{\theta^0,q}^{(1)}w\big)^*(t)\mathrm{e}^{-itX_{k-1}}\mathrm{d}t\right]\\ (\tilde{W}_{k,2})_q&=\int\big(g_{\theta^0}g_{\theta^0,q}^{(1)}w\big)^*(t)\mathrm{e}^{-itX_{k-1}}\mathrm{d}t-\mathbb{E}\left[\int\big(g_{\theta^0}g_{\theta^0,q}^{(1)}w\big)^*(t)\mathrm{e}^{-itX_{k-1}}\mathrm{d}t\right].\end{aligned}$$

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

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

Next we use that under condition (**C₂**),

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

In the same way we get that $|(W_{1,2})_p| \le C_2$. Now, since ε_1 is independent of X_1, for $j \in \{1,2\}$

$$\mathbb{E}\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right| \le C_1 \mathbb{E}\left[(|Z_1| + \mathbb{E}(|Z_1|))\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|\right] \le C\mathbb{E}\left[(|X_1| + \mathbb{E}(|X_1|))\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|\right]. \quad \text{(B.7)}$$

In the same way

$$\mathbb{E}\left|(W_{1,2})_p \mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right| \le C\mathbb{E}\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|. \quad \text{(B.8)}$$

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 (A.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 $(X_i)_{i\ge 0}$ is a strictly stationary Markov chain, it is well known that

$$\alpha(\sigma(X_1), \sigma(X_{k-1}, X_k)) = \alpha(\sigma(X_1), \sigma(X_{k-1})) = \alpha_{\mathbf{X}}(k-2). \quad \text{(B.9)}$$

Hence, applying (A.1),

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

We conclude that

$$\sum_{k\ge 3} \mathbb{E}\left|(W_{1,1})_p \mathbb{E}((W_{k,1})_q|\mathcal{M}_1)\right| \le C\sum_{k\ge 3}\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X_1|}(u)\mathrm{d}u.$$

Finally, using similar arguments for the three quantities

$$\sum_{k\ge 3} \mathbb{E}\left|(W_{1,2})_p \mathbb{E}((W_{k,1})_q|\mathcal{M}_1)\right|, \quad \sum_{k\ge 3} \mathbb{E}\left|(W_{1,1})_p \mathbb{E}((W_{k,2})_q|\mathcal{M}_1)\right| \quad \text{and} \quad \sum_{k\ge 3} \mathbb{E}\left|(W_{1,2})_p \mathbb{E}((W_{k,2})_q|\mathcal{M}_1)\right|$$

we conclude that

$$\sqrt{n}S_n^{(1)}(\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma/(4\pi^2)) \text{ as soon as } \sum_{k\ge 1}\int_0^{\alpha_{\mathbf{X}}(k)} Q^2_{|X_1|}(u)\mathrm{d}u < \infty.$$

B.3. Proof of Theorem 3.3

Following the proof of Theorem 3.2, we check that (B.6) holds. Starting from inequalities (B.7) and (B.8), let

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

Let ψ_M be the truncating function defined by $\psi_M(x) = (x \wedge M) \vee (-M)$. Applying (A.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

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

Define the constants K_1 and K_2 by

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

Clearly

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

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

$$\begin{aligned}|(\tilde{W}_{k,1})_q(X_k, X_{k-1}) - (\tilde{W}_{k,1})_q(X_k^*, X_{k-1}^*) - ((\tilde{W}_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde{W}_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*))| \\ \le C(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M}).\end{aligned}$$

Hence

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

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 (**C₄**), one has

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

It follows that

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

Set $L(t) = \mathbb{E}(X_0^2 \mathbf{1}_{X_0^2>t})$, $G(t) = t^{-1}L(t)$, and G^{-1} be the inverse cadlag of G. Using that

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

we infer from (B.7) with $j = 1$ that there exists a positive constant K such that

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

By choosing $M^2 = G^{-1}(\tau_{\mathbf{X},2}(k-2))$ we obtain that

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

It follows that

$$\sum_{k\ge3} \mathbb{E}\left[\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right|\right] < \infty \quad \text{as soon as} \quad \sum_{k>0} G^{-1}(\tau_{\mathbf{X},2}(k))\tau_{\mathbf{X},2}(k) < \infty.$$

Easier controls hold for the other terms in (B.7) and (B.8), hence (B.6) holds as soon as (**D₂**) holds, and the proof is complete.

B.4. Proof of Theorem 4.1

The consistency proof of Theorem 4.1 is quite different from the one under (**C₁**)–(**C₂**) in Theorem 3.1, since $S_n(\theta)$ is now a triangular array of the form

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} \Psi_n(Z_k, Z_{k-1}) \text{ with } \Psi_n(Z_1, Z_0) = \frac{1}{2\pi}\mathbb{R}e \int \frac{\left(\left(Z_1 - g_\theta\right)^2 w\right)^* (t) \mathrm{e}^{-itZ_0} K^*_{C_n}(t)}{f^*_\varepsilon(-t)} \mathrm{d}t.$$

In this context we show that

i) For all θ in Θ, $\mathbb{E}[(S_n(\theta) - S(\theta))^2] = o(1)$ as $n \to \infty$.
ii) The control (B.1) holds.

Note first that ii) follows from the upper bound (B.2) and assumption ($\mathbf{A_4}$).

For the proof of i) we check that for all $\theta \in \Theta$,

$$\mathbb{E}[S_n(\theta)] - S(\theta) = o(1) \quad \text{ and } \mathrm{Var}(S_n(\theta)) = o(1), \quad \text{ as } n \to \infty. \tag{B.10}$$

Proof of the first part of (B.10). Since $Z_0 = X_0 + \varepsilon_0$, with ε_0 independent of (Z_1, X_0),

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

hence

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

Now, arguing as in Butucea and Taupin [6] we get that $|\mathbb{E}[S_n(\theta)] - S(\theta)|^2 = o(1)$.

Proof of the second part of (B.10). Using that the Z_i's are strictly stationary we get that

$$\begin{aligned}\mathrm{Var}[S_n(\theta)] &\le \frac{1}{n}\ \mathrm{Var}\ (A_{1,0}) + \frac{2}{n}\sum_{i=2}^{n} |\ \mathrm{Cov}\big(A_{1,0}, A_{i,i-1}\big)|\\ &\le \frac{3}{n}\ \mathrm{Var}\ (A_{1,0}) + \frac{2}{n}\sum_{k=3}^{n} |\ \mathrm{Cov}\big(A_{1,0}, A_{k,k-1}\big)|\end{aligned}$$

with $A_{k,k-1} = \mathbb{R}e\Big[\big((Z_k - g_\theta)^2\, w\big) \star K_{n,C_n}(Z_{k-1})\Big]$. As in Butucea and Taupin [6] we obtain that $\lim_{n\to\infty} n^{-1}$ Var $(A_{1,0}) = 0$ and $n^{-1}\sum_{k=3}^{n} |\operatorname{Cov}\big(A_{1,0}, A_{k,k-1}\big)|$ is studied using the lemma:

Lemma B.1. *Let Ψ such that $\mathbb{E}(|\Psi(Z)|) < \infty$ and let Φ be an integrable function. Let*

$$B_{k,k-1} = \mathbb{R}\left[e\Psi(Z_k)\Phi \star K_{n,C_n}(Z_{k-1})\right].$$

Then for $k \geq 3$, $\operatorname{Cov}(B_{k,k-1}, B_{1,0}) = \operatorname{Cov}[\Psi(Z_k)\Phi \star K_{C_n}(X_{k-1}), \Psi(Z_1)\Phi \star K_{C_n}(X_0)]$ equals

$$\frac{1}{(2\pi)^2}\iint \Phi^*(t)\Phi^*(s)\operatorname{Cov}\big(\Psi(Z_k)\mathrm{e}^{-itX_{k-1}}, \Psi(Z_1)\mathrm{e}^{-isX_0}\big)K^*_{C_n}(t)K^*_{C_n}(s)\mathrm{d}t\mathrm{d}s.$$

It follows from Lemma B.1 that for $k \geq 3$,

$$\operatorname{Cov}\big(A_{k,k-1}, A_{1,0}\big) = \operatorname{Cov}\Big[\big((Z_k - g_\theta)^2\, w\big) \star K_{C_n}(X_{k-1}), \big((Z_1 - g_\theta)^2\, w\big) \star K_{C_n}(X_0)\Big] = \sum_{i=1}^{9} C_i,$$

with

$$\begin{aligned}
C_1 &= \frac{1}{(2\pi)^2}\iint \operatorname{Cov}(\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0})(wg_\theta^2)^*(t)(wg_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_2 &= \frac{1}{\pi^2}\iint \operatorname{Cov}(X_k\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0})(wg_\theta)^*(t)(wg_\theta)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)\mathrm{d}t\mathrm{d}s,\\
C_3 &= \frac{1}{(2\pi)^2}\iint \operatorname{Cov}[(X_k^2+\varepsilon_k^2)\mathrm{e}^{-itX_{k-1}}, (X_1^2+\varepsilon_1^2)\mathrm{e}^{-isX_0}]w^*(t)w^*(s)K^*_{C_n}(t)K^*_{C_n}(s)\mathrm{d}t\mathrm{d}s,\\
C_4 &= \frac{-1}{2\pi^2}\iint \operatorname{Cov}(X_k\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0})(wg_\theta)^*(t)(wg_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_5 &= \frac{-1}{2\pi^2}\iint \operatorname{Cov}(\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0})(wg_\theta)^*(s)(wg_\theta^2)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_6 &= \frac{1}{(2\pi)^2}\iint \operatorname{Cov}[(X_k^2+\varepsilon_k^2)\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0}]w^*(t)(wg_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_7 &= \frac{1}{(2\pi)^2}\iint \operatorname{Cov}[\mathrm{e}^{-itX_{k-1}}, (X_1^2+\varepsilon_1^2)\mathrm{e}^{-isX_0}]w^*(s)(wg_\theta^2)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_8 &= \frac{-1}{2\pi^2}\iint \operatorname{Cov}[(X_k^2+\varepsilon_k^2)\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0}]w^*(t)(wg_\theta)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
C_9 &= \frac{-1}{2\pi^2}\iint \operatorname{Cov}[X_k\mathrm{e}^{-itX_{k-1}}, (X_1^2+\varepsilon_1^2)\mathrm{e}^{-isX_0}]w^*(s)(wg_\theta)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s
\end{aligned}$$

Easy computations give

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

In the same way,

$$\operatorname{Cov}[(X_k^2+\varepsilon_k^2)\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0}] = \operatorname{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0}) + \sigma_\varepsilon^2\operatorname{Cov}(\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0}),$$

and

$$\operatorname{Cov}[(X_k^2+\varepsilon_k^2)\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0}] = \operatorname{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0}) + \sigma_\varepsilon^2\operatorname{Cov}(\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0}).$$

We can rewrite $\sum_{i=1}^{9} C_i = \sum_{i=1}^{9} E_i$, with

$$\begin{aligned}
E_1 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0})K^*_{C_n}(t)K^*_{C_n}(s)\\
&\quad\times[(wg_\theta^2)^*(t)(wg_\theta^2)^*(s) + \sigma_\varepsilon^4 w^*(t)w^*(s) + \sigma_\varepsilon^2 w^*(t)(wg_\theta)^*(s) + \sigma_\varepsilon^2 w^*(s)(wg_\theta)^*(t))]\mathrm{d}t\mathrm{d}s,\\
E_2 &= C_2 = \frac{1}{\pi^2}\iint \mathrm{Cov}(X_k\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{isX_0})(wg_\theta)^*(t)(wg_\theta)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)\mathrm{d}t\mathrm{d}s,\\
E_3 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, X_1^2\mathrm{e}^{-isX_0})w^*(t)w^*(s)K^*_{C_n}(t)K^*_{C_n}(s)\mathrm{d}t\mathrm{d}s,\\
E_4 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0})K^*_{C_n}(s)K^*_{C_n}(t)(wg_\theta)^*(t)((wg_\theta^2)^*(s) + \sigma_\varepsilon^2 w^*(s))\mathrm{d}t\mathrm{d}s,\\
E_5 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0})K^*_{C_n}(s)K^*_{C_n}(t)(wg_\theta)^*(s)((wg_\theta^2)^*(t) + \sigma_\varepsilon^2 w^*(t))\mathrm{d}t\mathrm{d}s,\\
E_6 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, \mathrm{e}^{-isX_0})K^*_{C_n}(t)K^*_{C_n}(s)w^*(t)(\sigma_\varepsilon^2 w^*(s) + (wg_\theta^2)^*(s))\mathrm{d}t\mathrm{d}s,\\
E_7 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(\mathrm{e}^{-itX_{k-1}}, X_1^2\mathrm{e}^{isX_0})K^*_{C_n}(t)K^*_{C_n}(s)w^*(s)(\sigma_\varepsilon^2 w^*(t) + (wg_\theta^2)^*(t))\mathrm{d}t\mathrm{d}s,\\
E_8 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, X_1\mathrm{e}^{-isX_0})w^*(t)(wg_\theta)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s,\\
E_9 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k\mathrm{e}^{-itX_{k-1}}, X_1^2\mathrm{e}^{-isX_0})w^*(s)(wg_\theta)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)\mathrm{d}t\mathrm{d}s.
\end{aligned}$$

We now apply the following lemma.

Lemma B.2. *In model 1.1, with $\mathbb{E}(X_1^4) < \infty$, we have the upper bounds*

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

$$\text{and } |\mathrm{Cov}(X_k^2\mathrm{e}^{-itX_{k-1}}, X_1^2\mathrm{e}^{-isX_0})| \le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^4_{|X|}(u)\mathrm{d}u.$$

The proof of Lemma B.2 which follows from (A.1) and (B.9) is ommitted.

Apply Lemma B.2 to $E_1, \ldots, E_9$. Since $\mathbb{E}(X_1^4) < \infty$ and $\lim_{k\to\infty}\alpha_{\mathbf{X}}(k) = 0$, it follows that $\lim_{k\to\infty}|\mathrm{Cov}\big(A_{k,k-1}, A_{1,0}\big)| = 0$, and we conclude by applying Cesaro's mean convergence theorem to get $\lim_{n\to\infty}\frac{1}{n}\sum_{k=3}^{n}|\mathrm{Cov}\big(A_{1,0}, A_{k,k-1}\big)| = 0$.

B.5. Proof of Theorem 4.2

Proof of 1) *in Theorem 4.2.* Starting from (B.3) we shall check the three following points.

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

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

iii) $R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

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

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

We only give the proof of i), ii) and iii) following by the same arguments. We first write

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

Study of the bias. As in Butucea and Taupin [6], we get that

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

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

$$\begin{aligned}\operatorname{Var}\left(\left(S_n^{(1)}(\theta^0)\right)_j\right) &\le \frac{3}{n}\operatorname{Var}(D_{1,0}) + \frac{2}{n}\sum_{k=3}^n|\operatorname{Cov}(D_{1,0}, D_{k,k-1})|,\\ \text{with } D_{k,k-1} &= \mathbb{R}e\Big(\big(-2Z_k f_{\theta^0,j}^{(1)} + 2g_\theta f_{\theta^0,j}^{(1)}\big)w\Big)\star K_{n,C_n}(Z_{k-1}).\end{aligned} \tag{B.11}$$

Arguing as in Butucea and Taupin [6] we infer that

$$\frac{1}{n}\operatorname{Var}(D_{1,0}) \le \frac{C(\sigma_\xi^2, g_{\theta^0}, f_{\theta^0,j}^{(1)}, w, f_\varepsilon)}{n}\min\{V_{n,j}^{[1]}(\theta^0), V_{n,j}^{[2]}(\theta^0)\} \tag{B.12}$$

with $V_{n,j}^{[q]}$, $q = 1, 2$ defined in Theorem 4.2. We now control the term

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

Applying again Lemma B.1, we obtain that $\operatorname{Cov}(D_{1,0}, D_{k,k-1}) = F_1 + F_2 + F_3 + F_4$ with

$$\begin{aligned}F_1 &= \frac{1}{\pi^2}\mathbb{R}e\iint \operatorname{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{-isX_0})\big(f_{\theta^0,j}^{(1)}w\big)^*(t)\big(f_{\theta^0,j}^{(1)}w\big)^*(s)K_{C_n}^*(t)K_{C_n}^*(s)\mathrm{d}t\mathrm{d}s\\ F_2 &= \frac{1}{\pi^2}\mathbb{R}e\iint \operatorname{Cov}(e^{-itX_{k-1}}, e^{-isX_0})\big(g_{\theta^0}f_{\theta^0,j}^{(1)}w\big)^*(t)\big(g_{\theta^0}f_{\theta^0,j}^{(1)}w\big)^*(s)K_{C_n}^*(t)K_{C_n}^*(s)\mathrm{d}t\mathrm{d}s\\ F_3 &= \frac{-1}{\pi^2}\mathbb{R}e\iint \operatorname{Cov}(X_k e^{-itX_{k-1}}, e^{-isX_0})\big(f_{\theta^0,j}^{(1)}w\big)^*(t)\big(g_{\theta^0}f_{\theta^0,j}^{(1)}w\big)^*(s)K_{C_n}^*(t)K_{C_n}^*(s)\mathrm{d}t\mathrm{d}s\\ F_4 &= \frac{-1}{\pi^2}\mathbb{R}e\iint \operatorname{Cov}(e^{-itX_{k-1}}, X_1 e^{-isX_0})\big(f_{\theta^0,j}^{(1)}w\big)^*(t)\big(f_{\theta^0,j}^{(1)}w\big)^*(s)K_{C_n}^*(t)K_{C_n}^*(s)\mathrm{d}t\mathrm{d}s.\end{aligned}$$

Now, we apply Lemma B.2 to each term. Since $\mathbb{E}(X_1^4) < \infty$, $Q_{|X|}(u) \le Cu^{-1/4}$, 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

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

This, together with (B.12), implies that

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

Proof of 2) *in Theorem 4.2.* The proof of 2) in Theorem 4.2 is quite similar to the proof of 1) with differences appearing in the bias and variance of $S_n^{(1)}(\theta^0)$. More precisely, we start from

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

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

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

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

$$-2\mathbb{R}e\iint g_{\theta^0}(x)\mathrm{e}^{-iux}(g_{\theta^0}^{(1)}w)^*(u)(K_{C_n}^*(u)-1)P_X(\mathrm{d}x)\,\mathrm{d}u + 2\mathbb{R}e\iint \mathrm{e}^{-iux}(g_{\theta^0}g_{\theta^0}^{(1)}w)^*(u)(K_{C_n}^*(u)-1)P_X(\mathrm{d}x)\,\mathrm{d}u.$$

It follows that for $j=1,\ldots,d$, $\left|\mathbb{E}[(S_n^{(1)}(\theta^0))_j] - (S^{(1)}(\theta^0))_j\right|$ is less than

$$\mathbb{E}|g_{\theta^0}(X_0)|\int|(f_{\theta^0,j}^{(1)}w)^*(u)(K_{C_n}^*(u)-1)|\mathrm{d}u + \int|(g_{\theta^0}f_{\theta^0,j}^{(1)}w)^*(u)(K_{C_n}^*(u)-1)|\mathrm{d}u.$$

Study of the variance. We combine the proof in Butucea and Taupin [6] and the proof of 1) of Theorem 4.2. For these reasons we only give a sketch of the proof, with details only for specific parts. For $D_{k,k-1}$ defined in (B.11), we start from

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

The control of $(2/n^2)\sum_{1\leq j<k\leq n}\mathrm{Cov}(D_{k,k-1}, D_{j,j-1})$ is almost the same as in the proof of **1)**. First

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

Now, write that

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

with

$$\begin{aligned}II_1 &= \mathbb{R}e\iint f_\varepsilon(z-x)(g_{\theta^0}^2(x)+\sigma_\xi^2)\left(\int(g_{\theta^0}^{(1)}w)(u)K_{n,C_n}(z-u)\mathrm{d}u\right)^2 P_X(\mathrm{d}x)\mathrm{d}z\\ II_2 &= \mathbb{R}e\iint f_\varepsilon(z-x)\left(\int(g_{\theta^0}g_{\theta^0}^{(1)}w)(u)K_{n,C_n}(z-u)\mathrm{d}u\right)^2 P_X(\mathrm{d}x)\mathrm{d}z.\end{aligned}$$

We apply Hölder Inequality and obtain that

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

and that $|II_1|$ is also less than

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

In the same way we have

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

Consequently we have

$$\mathrm{Var}\big[(S^{(1)}_n(\theta^0))_j\big] \leq \frac{C(\sigma^2_\xi, g_{\theta^0}, f_\varepsilon)}{n}\Big[\parallel (g^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2 + \parallel (g_{\theta^0}g^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2\Big] \tag{B.13}$$

and

$$\mathrm{Var}\big[(S^{(1)}_n(\theta^0))_j\big] \leq \frac{C_1(g_{\theta^0})}{n}\Big[\parallel (g^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2 + \parallel (g_{\theta^0}g^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_1^2\Big]. \tag{B.14}$$

By combining (B.13) and (B.14), for $V^{[q]}_{n,j}$, $q=1,2$ defined in Theorem 4.2, we get that

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

Proof of Lemma B.1. By stationarity we write

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

The sequences $(X_k)_{k\in\mathbb{Z}}$ and $(\varepsilon_k)_{k\in\mathbb{Z}}$ being independent, (Z_1, X_0) is independent of ε_0 and thus

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

In the same way, we conclude the proof by writing that for $k\geq 3$,

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

References

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

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

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

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

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

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

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

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

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

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

[11] J. Dedecker F. Merlevède and M. Peligrad, A quenched weak invariance principle. Technical report, to appear in *Ann. Inst. Henri Poincaré Probab. Statist.* (2012). http://fr.arxiv.org/abs/math.ST/arxiv:1204.4554

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

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

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

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

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

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

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

[19] E.J. Hannan, The asymptotic theory of linear time−series models. *J. Appl. Probab.* **10** (1973) 130–145.

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

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

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

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

[24] E. Nowak, Global identification of the dynamic shock-error model. *J. Econom.* **27** (1985) 211–219.

[25] E. Rio, Covariance inequalities for strongly mixing processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **29** (1993) 587–597.

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

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

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