KERNEL DECONVOLUTION OF STOCHASTIC VOLATILITY MODELS

F. COMTE

ABSTRACT. In this paper, we study the problem of the non parametric estimation of the function m in a stochastic volatility model $h_t = \exp(X_t/2\lambda)\xi_t$, $X_t = m(X_{t-1}) + \eta_t$, where ξ_t is a gaussian white noise. We show that the model can be written as an autoregression with errors-in-variables. Then an adaptation of the deconvolution Kernel estimator proposed by Fan and Truong (1993) estimates the function m with the optimal rate, which depends on the distribution of the measurement error. The rates vary from powers of n to powers of $\ln(n)$ depending on the rate of decay near infinity of the characteristic function of this noise. Central Limit Theorems and multidimensional extensions are available. Some simulation experiments study the performances of the method.

Key Words and Phrases: Deconvolution Kernel estimator, Stochastic volatility model, Errors-in-variables model, Nonparametric optimal rate.

AMS Classifications (1991): Primary 62G05; Secondary 62J02.

1. Introduction

Stochastic volatility models have been widely studied in the last ten years, in particular in view of financial applications. In continuous time modelling the Hull and White (1987) model was a natural extension of the constant volatility Black and Scholes (1973) model. It develops the idea that a specific noise governs the volatility seen as a true process, and not only as a deterministic function. Later studies by Renault and Touzi (1996) proved the ability of such models to explain some observed behavior of option prices and implicit volatilities, and in particular the so-called smile effect well-known by practitioners. The idea of considering those models from a parametric statistical point of view as a hidden Markov chain has recently been exploited by Genon-Catalot *et al* (2000).

Discretisations of such models are then required both for estimation and simulation purposes. Indeed, even if the continuous time model gives an adequate account for the probabilistic properties, the sample of observations available for estimation is necessarily in discrete time.

Therefore, those models have also been developed in a discrete time framework, with first an AR(1) parametric specification of the volatility process proposed by Taylor (1986). But the question arises of appropriateness of such a simple parametric

University Paris VI, Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, email: comte@ccr.jussieu.fr.

description, see e.g. Gouriéroux and Monfort (1992) or Feldmann *et al* (1998). Indeed it is a natural idea to wonder why the volatility process would not be a general non-parametric AR(1) process instead of a simple linear autoregression: for instance Feldmann *et al* (1998) develop a testing procedure to determine is the null hypothesis of linearity can be accepted.

In this paper, we consider a non parametric stochastic volatility model for returns of asset prices. We show that these models can be viewed as the autoregressive version of regressive models with errors in variables studied in an independent framework by Fan and Truong (1993) and Taupin (1998). Dependence in the regression variables was studied by Masry (1993) and Fan and Masry (1992) in the nonparametric setting and by Chanda (1995) and Comte and Taupin (2000) in a parametric or semiparametric setting. Chanda (1995) deals with the particular case of ARMA(p, q) models with errors in variables and reaches the parametric rate $(\sqrt{n})$. Comte and Taupin (2000) prove the consistency of a parametric estimator and compute a bound for its rate of convergence for some classes of parametric regression functions; they work in a context of β-mixing (or absolutely regular) sequences. Masry (1993) and Fan and Masry (1992) consider α-mixing sequences in the same nonparametric setting as the one studied here. Fan and Masry (1992) prove some central limit theorems (CLT). Masry (1993) looks for a control of the supremum norm of the error: this implies a (standard) loss of order $\ln(n)$ in some rates of convergence of his estimate and some strong constraints on the mixing coefficients. We adapt the Kernel estimator proposed by Fan and Truong (1993) to the autoregressive case and we show that we can recover the optimal rates of convergence of the independent setting when considering the $\mathbb{L}_2$-risk (and not the $\mathbb{L}_\infty$-risk as done by Masry (1993)). Moreover, the techniques of proof in that case are simpler than those developed by Masry (1993). Our main tool is a variance inequality for sums of strongly mixing variables. Moreover, we show how the rates are varying when the distribution of the observation noise process changes. These rates can be standard powers of n depending on the regularity of the autoregressive function; or they can fall down until some power of $\ln(n)$ despite they remain optimal.

The plan of the paper is as follows. Section 2 presents the model, the assumptions and the discussions coming with. Section 3 gathers the results that can be obtained in our particular cases of interest as well as possible extensions. Those results are consequences of some theorems which are natural extensions of the independent framework; they are given in the final section 5 which gathers all the proofs. Lastly, Section 4 is devoted to a test of the method through a simulation study.

2. The Framework

2.1. **The model.** The well-known Black and Scholes' (1973) model assumes the following dynamics for an asset price S_t at time t:

$$dS_t = S_t(\mu dt + \sigma dB_t) \tag{1}$$

where B_t is a standard Brownian motion. The solution of this equation leads to the following model for $h_t = \ln(S_t/S_{t-1})$, the returns associated to the asset price S_t at time t:

$$h_t = (\mu - \sigma^2/2) + \sigma\xi_t$$

where $\xi_t = B_t - B_{t-1}$ is a sequence of i.i.d. Gaussian variables.
This model is widely used for option pricing and hedging, but it appears from financial data that the volatility σ is time varying. This is the reason why extensions of this model have been studied, following mainly two ideas:

(1) The first extension considers the model $h_t = \mu(t) + \sigma(t)\xi_t$ with i.i.d. ξ_t's or alternatively $h_t = \mu(h_{t-1}) + \sigma(h_{t-1})\xi_t$. The functions μ and σ map $\mathbb{R}$ into $\mathbb{R}$ and $\mathbb{R}^+$ respectively. Those models belong to the class of discrete time regressive or autoregressive models with varying volatility functions. The problem of the estimation of the functions μ and σ has been widely studied: see e.g. Hall and Carroll (1989), Fan and Yao (1998), Härdle and Tsybakov (1997) or Neumann (1994).

(2) The second extension starts from the same equation

$$h_t = \mu(t) + \sigma(t)\xi_t \tag{2}$$

but with $\sigma^2(t) = \exp(X_t)$ and $X_t = m(X_{t-1}) + \eta_t$ where both the ξ_t's and the η_t's are i.i.d. This model is studied for instance by Feldmann *et al.* (1998) who propose tests of linearity for m. This idea corresponds also to the one developed by the continuous time Hull and White (1987)'s model which extends (1) as follows:

$$\begin{cases} dS_t/S_t = \mu(t, S_t, Z_t)dt + \sigma_t dW_t^S, \\ \sigma_t^2 = f(Z_t),\ dZ_t = a(t, Z_t)dt + b(t, Z_t)dW_t^\sigma, \end{cases}$$

where $W_t = (W_t^S, W_t^\sigma)$ is a two-dimensional brownian motion. This model has been studied by Renault and Touzi (1996) who used it in particular to explain the so-called smile effect observed by practitioners on implicit volatilities. In particular, defining $Z_t = \ln(\sigma_t)$ as an Ornstein-Uhlenbeck process: $dZ_t = -kZ_t dt + \gamma dW_t^\sigma$ leads to an exact discretized Z_t as a gaussian AR(1) process: $Z_t = e^{-k}Z_{t-1} + \eta_t$ with i.i.d. η_t's with law $\mathcal{N}(0, s^2)$, s being a known function of k and γ.

The main difference between the two extensions is clear: in the first case a *deterministic* volatility is introduced whereas in the second, the volatility is seen as a *random* process governed by a specific noise. In the first case, one source of hazard only is considered whereas there are two of them in the second. The idea that the volatility is itself a random process is very natural since practitioners deal volatilities just as other assets. We must be aware – since it is both well-known and somehow logical – that the second type of models is also more difficult to estimate for the precise reason which makes its interest: there are two noise processes and only one observed process.

Nevertheless, we are interested in the second model defined by equation (2). For sake of simplicity, we first assume that $\mu(t) \equiv \mu$ is constant. In a second time, if X_0 is independent of ξ and η and if ξ and η are independent, then $\mu = \mathbb{E}(h_t)$ can easily be estimated by the empirical mean of h. Thus we assume $\mu = 0$. This leads to the following model

$$h_t = \exp(X_t/2\lambda)\xi_t \tag{3}$$
$$X_t = m(X_{t-1}) + \eta_t \tag{4}$$

where ξ and η are i.i.d. sequences and independent processes and where the ξ_t's are gaussian $\mathcal{N}(0,1)$.
Let us make some comments about the coefficient λ. The most standard case amounts to take $\lambda = 1$ and then to consider $\ln(h_t^2)$. But h_t is in general small, so that h_t^2 is even smaller and $\ln(h_t^2)$ is great and has too much variance; more specifically, it is easy to see that $\ln(\xi_t^2)$ has a quite high variance for standard laws. This means that it is wiser to take power smaller than one than usual powers larger than one. For instance, we chose $\lambda = 1/10$ in the empirical examples. This does not structurally change the autoregressive function; for instance in the case of a linear function m, the function is even exactly the same.

Taking the logarithm of $|h_t|^{2\lambda}$ transforms the first equation into $\ln(|h_t|^{2\lambda}) = X_t + \lambda\ln(\xi_t^2) = \lambda\omega + X_t + \lambda(\ln(\xi_t^2) - \omega)$ where $\omega = \mathbb{E}(\ln(\xi_1^2))$. Since ω is assumed to be known, setting $Y_t = \ln(h_t^2) - \lambda\omega$ and $\varepsilon_t = \lambda(\ln(\xi_t^2) - \omega)$ leads to the model:

$$X_t = m(X_{t-1}) + \eta_t \tag{5}$$
$$Y_t = X_t + \varepsilon_t \tag{6}$$

with

A1 (η_t) i.i.d., (ε_t) i.i.d. η and ε are independent, $\mathbb{E}(\eta_1) = \mathbb{E}(\varepsilon_1) = 0$, $\mathrm{Var}(\eta) = \sigma_\eta^2 < +\infty$, $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2 < +\infty$, f_η and f_ε denote the densities of η and ε respectively.

and

A2 $\varepsilon_t = \lambda(\ln(\xi_t^2) - \omega)$ where $\omega = \mathbb{E}(\ln(\xi_1^2))$ and the ξ_t's are i.i.d. $\mathcal{N}(0,1)$.

Only the Y_t's are observed and we aim at estimating m.

The model (5)-(6) can be considered as a particular case of the standard regression model with errors in variables:

$$Z_t = m(X_t) + \eta_t, \tag{7}$$
$$Y_t = X_t + \varepsilon_t \tag{8}$$

under the same assumptions on the noises and where the two processes (Y, Z) are observed for $t = 1, \dots, n$. Note that it is difficult to gather this model and the autoregression model in a unified writing since in the regression model, two processes are observed, whereas in the autoregression model, only one process is observed. But the estimation results for one model also hold for the other.

2.2. **The estimator.** Before being more precise about the assumptions, we need to explain how the estimator is built. Consider (X_t, Y_t) drawn from model (5)-(6). Then the following equality holds:

$$\mathbb{E}[Y_{j+1}|X_j = x] = m(x) = \frac{\int y f_{(X,Y)}(x,y)dy}{f_X(x)}$$

where $f_{(X,Y)}$ denotes the joint distribution of (X_1, Y_1) and $f_X(x)$ the distribution of X_1. Therefore, if both the X_i's and the Y_i's were observed, we would use the following standard kernel estimator

$$\hat{m}(x) = \frac{\dfrac{1}{nh_n}\displaystyle\sum_{j=1}^{n} Y_{j+1} K\left(\frac{x - X_j}{h_n}\right)}{\dfrac{1}{nh_n}\displaystyle\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)} \tag{9}$$

where K is a kernel and h_n a bandwidth parameter; the denominator

$$\hat{f}_X(x) = \frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)$$

is a standard estimator of $f_X(x)$.

Since the X_j's are not observable, the estimator must be built in another way. First, observe that from (6), $f_Y(y) = f_X * f_\varepsilon(y)$ where $*$ denotes the convolution product $g * h(x) = \int g(x-u)h(u)du$. Second, if $g^\star$ denotes the Fourier transform: $g^\star(x) = \int e^{itx} g(t)dt$, then we know: $f_Y^\star(y) = f_X^\star(y) f_\varepsilon^\star(y)$. Therefore, using the Fourier inverse formula: $(g^\star)^\star(x) = 2\pi g(-x)$, we find that

$$f_X(x) = \frac{1}{2\pi}\left(\frac{f_Y^\star}{f_\varepsilon^\star}\right)^\star(-x) = \frac{1}{2\pi}\int e^{-itx}\frac{f_Y^\star(t)}{f_\varepsilon^\star(t)}dt.$$

This explains why we consider, as suggested by Stefanski and Carroll (1990) and many others after them, the following estimator of $f_X(x)$:

$$\hat{f}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx} K^\star(th_n)\frac{\hat{\Phi}_n(t)}{f_\varepsilon^\star(t)}dt, \tag{10}$$

where $\hat{\Phi}_n$ is the empirical characteristic function of Y:

$$\hat{\Phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j}.$$

Defining

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{K^\star(t)}{f_\varepsilon^\star(t/h_n)}dt, \tag{11}$$

this can be re-written

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K_n\left(\frac{x-Y_i}{h_n}\right). \tag{12}$$

For the numerator of (9), we simply substitute the unknown X_j's by the Y_j's and the kernel K by the deconvolution kernel K_n. At last, following Fan and Truong (1993) and by analogy with (9), we consider the estimate of m, based on the observations $Y_1, \ldots, Y_{n+1}$:

$$\hat{m}_n(x) = \frac{\sum_{j=1}^n K_n((x-Y_j)/h_n)Y_{j+1}}{\sum_{i=1}^n K_n((x-Y_i)/h_n)} = \frac{1}{nh_n}\sum_{j=1}^{n}\frac{K_n((x-Y_j)/h_n)Y_{j+1}}{\hat{f}_n(x)} \tag{13}$$

where K_n is given by (11) and $\hat{f}_n(x)$ by (12).

It is worth emphasizing here that no knowledge about the distribution of the innovation η is required to build $\hat{m}_n(x)$. The price to pay for it is that the distribution of ε is assumed to be *completely* known. Under this assumption, $\hat{m}_n$ and $\hat{f}_n$ can be used to estimate σ_η^2:

$$\widehat{\sigma}_\eta^2 = \int (x^2 - \hat{m}_n^2(x))\hat{f}_n(x)dx$$

using that $\sigma_\eta^2 = Var(X) - Var(m(X)) = \mathbb{E}(X^2) - \mathbb{E}(m^2(X))$. Since $\mathbb{E}(Y^2) = \mathbb{E}(X^2) + \sigma_\varepsilon^2$, we can also propose

$$\widehat{\sigma}_\eta^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \sigma_\varepsilon^2 - \int \hat{m}_n^2(x)\hat{f}_n(x)dx.$$

2.3. **Stability and regularity assumptions.** In this section, we discuss briefly the assumptions required for the variables X_i's, the function m and the kernel K. We denote by k a given positive integer. Then we need the following assumptions:

A3(ν, k) (i) The X_i's are stationary with law $d\mu(x) = f_X(x)dx$,
(ii) $\mathbb{E}(|m|^\nu(X_1)) < +\infty$ for some $\nu > 2$.
(iii) The sequence $(X_i)_{i\in\mathbb{N}}$ is strongly mixing with α-mixing coefficients α_j.
(iv) f_X is bounded away from 0 on $[a, b]$ and admits k bounded derivatives.

A4(k) m admits continuous derivatives until order k on any interval $[a, b]$, $a < b$.

Moreover, the user-chosen kernel must be in accordance with the properties of the model:

A5(k) K is a kernel of order k i.e. $\int K(y)dy = 1, \int y^k K(y)dy \neq 0, \int y^j K(y)dy = 0, j = 1, \ldots, k-1$.

For instance, the gaussian kernel, $K(x) = e^{-x^2/2}/\sqrt{2\pi}$, satisfies **A5**(2), as well as the kernel K defined by $K^\star(t) = (1-t^2)_+^3$.

Let us recall the definition of the mixing coefficients (see Doukhan (1995), p.3-4.). Let $\mathcal{F}_i^n$ the σ-algebra of events generated by the random variables $\{X_j, i \le j \le n\}$. A stationary process $\{X_j\}$ is called strongly mixing or α-mixing (Rosenblatt (1956)) if

$$\sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_n^\infty} |\mathbb{P}(A\cap B - \mathbb{P}(A)\mathbb{P}(B)| = \alpha_n \to 0 \text{ as } n \to +\infty.$$

It is said to be absolutely regular or β-mixing (Kolmogorov and Rozanov 1960) if

$$\frac{1}{2}\sup\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}|\mathbb{P}(U_i\cap V_j - \mathbb{P}(U_i)\mathbb{P}(V_j)|\right\} = \beta_n \to 0 \text{ as } n \to +\infty.$$

where the above supremum is taken over all finite partitions $(U_i)_{1\le i\le I}$ and $(V_j)_{1\le j\le J}$ of Ω respectively $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_n^\infty$ measurable. Note that β-mixing implies α-mixing since the following inequality holds: $2\alpha_n \le \beta_n$.

Next, let us make some remarks concerning autoregressive systems. Assumption **A3** (i)-(ii) holds under standard assumptions (see Duflo (1990) p.178) as e.g. if

$$X_0 \text{ independent of } \eta,\ \ |m(x)| \le a|x| + b \text{ with } 0 < a < 1, \mathbb{E}(|\eta|^\nu) < +\infty. \tag{14}$$

Moreover, writing that $\mathbb{E}(g(X_{i+1})) = \mathbb{E}(g(m(X_i) + \varepsilon_{i+1}))$ implies

$$\int g(x)d\mu(x) = \int\int g(m(x)+y)d\mu(x)f_\eta(y)dy = \int g(u)\left[\int f_\eta(u-m(x))d\mu(x)\right]du.$$

Therefore if η admits a density with respect to the Lebesgue measure, then so does the stationary process X,

$$f_X(u) = \int f_\eta(u-m(x))f_X(x)dx.$$

and f_X admits bounded derivatives as soon as f_η does. For instance, for gaussian η_t's, f_η admits bounded derivatives of any order k and therefore, so does f_X. Since in that case f_X cannot be zero on a compact set $[a,b]$, this implies that for gaussian η_t's, assumption **A3** (iv) is fulfilled.

Remark 2.1. Under (14), we obtain geometrically β-mixing X_i's (see Meyn and Tweedie (1993)) i.e. with β_n satisfying $\beta_n \le Me^{-\theta n}$ for all n and for some given M and θ. Conditions on m implying some arithmetical β-mixing (i.e. $\beta_n \le Mn^{-\theta}$) are given by Tuominen and Tweedie (1994) or Ango Nze (1998). These properties imply geometrical or arithmetical α-mixing respectively.

2.4. **Assumptions on the measurement errors.**

2.4.1. *General Assumptions.* Looking at (11), it is obvious that the definition of the estimator is available provided that the following assumption holds:

A6 $f_\varepsilon^\star(t) = \mathbb{E}(e^{it\varepsilon}) \neq 0, \quad \forall t \in \mathbb{R}$.

Assumption **A6** rules out some very standard distributions as e.g. the uniform distribution. Take $\varepsilon \sim \mathcal{U}([-a,a])$ for some positive real number a, then $f_\varepsilon^\star(t) = \sin(at)/(at)$, $f_\varepsilon^\star(0) = 1$ is zero for all $t = k\pi/a$, $k \in Z^*$. This does not allow to compute K_n unless e.g. $K^\star$ is compactly supported and all the zeros of $f_\varepsilon^\star$ are outside the compact.

Moreover, two classes of error ε are generally considered (see Fan and Truong (1993) or Masry (1993)), depending on the rate of decay of the characteristic function of ε:

(1) Super Smooth of order δ:

$$\gamma_0|t|^{\delta_0}\exp\left(-|t|^\delta/\gamma_2\right) \leq |f_\varepsilon^\star(t)| \leq \gamma_1|t|^{\delta_1}\exp\left(-|t|^\delta/\gamma_2\right) \text{ as } t \to +\infty$$

(2) Ordinary Smooth of order δ:

$$\gamma_0|t|^{-\delta} \leq |f_\varepsilon^\star(t)| \leq \gamma_1|t|^{-\delta} \text{ as } t \to +\infty. \tag{15}$$

Observe that the faster the decay of the tail of the characteristic function, the smoother the corresponding distribution of ε.
The standard examples are the following: Gaussian or Cauchy distributions are super smooth of orders $\delta = 2$ and $\delta = 1$ respectively, the double exponential distribution is ordinary smooth of order $\delta = 2$.

2.4.2. *Some particular cases.* In our model, ξ is assumed to be gaussian $\mathcal{N}(0,1)$. Therefore, we need to compute $\omega = \mathbb{E}(\ln(\xi_1^2))$, possibly $\mathrm{Var}(\varepsilon_1) = \lambda^2\mathrm{Var}(\ln(\xi_1^2))$ and both $f_\varepsilon^\star(t)$ and the rate of decay of its modulus when t goes to infinity. We find the following results:

Proposition 2.1. *Let* $\xi_1 \rightsquigarrow \mathcal{N}(0,1)$, $\omega = \mathbb{E}(\ln(\xi_1^2))$, $\sigma^2 = \mathrm{Var}(\ln(\xi_1^2))$ *and* $\varepsilon_1 = \lambda(\ln(\xi_1^2) - \omega)$. *Then*

(1) $\omega = -\gamma - \ln(2)$ *where* γ *is the Euler constant, and* $\sigma^2 = \pi^2/2$.

(2) $f_\varepsilon^\star(t) = \dfrac{e^{it\lambda(\gamma+2\ln(2))}}{\sqrt{\pi}}\Gamma(\dfrac{1}{2} + i\lambda t)$,

(3) $|f_\varepsilon^\star(t)| \sim_{t\to+\infty} \sqrt{\frac{2}{e}}e^{-\pi\lambda t/2}$.

Therefore, for gaussian ξ_t's, $\varepsilon_t = \ln(\xi_t^2)$ is super smooth of order 1.
Since ξ is more directly implied in the observed log-return process, it is interesting to study which distribution for ξ leads to which class for ε. Indeed, the rate of convergence of $\hat{m}_n$ towards m depends on the class to which ε belongs. Moreover, fatter tails than gaussian ones have often been observed for financial data and in

particular log-returns (see Ding et al. (1993), Mikosch (2001)). For instance, some double exponential law for ξ with same mean and variance may be studied:

$$f_\xi(u) = (1/\sqrt{2})e^{\sqrt{2}|x|}$$

is centered with unit variance and gives for $\lambda = 1$

(1) $\omega = -2\gamma - \ln(2)$ and $\sigma^2 = 2\pi^2/3$.
(2) $f_\varepsilon^\star(t) = e^{2it\gamma}\Gamma(1+2it)$,
(3) $|f_\varepsilon^\star(t)| \sim_{t\to+\infty} \frac{2\sqrt{\pi}}{e} t^{1/2} e^{-\pi t}$.

In that case, the noise ε remains super smooth of order 1.

On the other hand, a simple computation shows that the following law for ξ:

$$\frac{1}{4}\mathbb{I}_{[-1,1]} + \frac{1}{4u^2}\mathbb{I}_{|u|>1}. \tag{16}$$

leads to a double exponential distribution for ε: this implies an ordinary smooth ε with $\delta = 2$. In that case, ξ does not admit second order moments. Moreover we find here $\omega = 0$, $\sigma_\varepsilon^2 = 8$ and $f_\varepsilon^\star(t) = \dfrac{1}{1+4t^2}$, implying $|f_\varepsilon^\star(t)| \sim_{t\to+\infty} (1/4)t^{-2}$.

More generally, densities as:

$$f_\xi(u) = \frac{1}{2a}(1-\frac{1}{k})\left(1_{[-a,a]}(u) + \frac{a^k}{|x|^k}\mathbb{I}_{|u|>a}\right) \tag{17}$$

lead to noises ε's such that $|f_\varepsilon^\star(t)|$ is of order $1/|t|^2$ when $t \to +\infty$. The choice of k determines the order of the finite moments of ξ.

3. Some results in $\mathbb{L}_2$-norm

3.1. **Results for Gaussian (or exponential) ξ_t's.** For super smooth errors, the optimal rates of convergence found by Fan and Truong (1993) are $\ln(n)^{2k/\delta}$. Masry (1993) provides results for the errors-in-variables model in an α-mixing regressive framework, but he studies the particular supremum norm and therefore works under slightly stronger assumptions. We study here an $\mathbb{L}_2$-risk for our autoregressive models and we keep the same rate. Since Fan and Truong (1993) proved that these rates were optimal in the independent framework, this means that we keep optimal rates in our autoregressive context even if this corresponds to rates of order $\ln(n)$.

Proposition 3.1. *Under Assumptions* **A1**, **A2**, **A3**$(\nu, 2)$, **A4**(2), *if the mixing coefficients satisfy*

$$\sum_j \alpha_j^{1-2/\nu} < +\infty, \tag{18}$$

and for $K^\star(x) = (1-x^2)_+^3$, *then for bandwidth* $h_n = c(\ln(n))^{-1}$ *with* $c > \pi\lambda$,

$$\mathbb{E}|(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f_X(x)|^2 = O\left((\ln(n))^{-4}\right)$$

and

$$\mathbb{E}\int_a^b |(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f_X(x)|^2 dx \quad = \quad O\left((\ln(n))^{-4})\right)$$

We use the same factor $\hat{f}_n(x)/f_X(x)$ which tends to 1 in probability, as Fan and Truong (1993). Note that we can prove that those rates are uniform over some classes of functions (m, f_X), see Theorem 5.2 in section 5.

From Remark 2.1, we know that most stable autoregressive systems drive to geometrically α-mixing variables, i.e. $\alpha_k \leq Me^{-\theta k}$; in that case, condition (18) is fulfilled. For arithmetical α-mixing, $\alpha_k \leq Mk^{-\theta}$, then condition (18) is fulfilled if $\theta > \nu/(\nu - 2)$. For instance if $\nu = 4$, then we require $\theta > 2$, and if $\nu \to +\infty$ (gaussian or bounded variables η_t's), then condition (18) simply requires $\theta > 1$. Masry (1993) finds for $\nu = 4$ a constraint $\theta > 11$. In the general case, he finds $\theta > (4 + 9\nu/2)/(\nu - 2)$.

Note that in the case described in Theorem 3.1, we have

$$K_n(t) = \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}(1-t^2)^3 \frac{1}{f_\varepsilon^\star(t/h_n)} dt \tag{19}$$

with

$$f_\varepsilon^\star(t) = \sqrt{\frac{2}{\pi}} e^{it\lambda(\gamma+\ln(2))} \int_{-\infty}^{+\infty} e^{2it\lambda\ln(u)} \exp\left(-\frac{1}{2}u^2\right) du.$$

Both integrals must be computed using approximations.

Those rates are often considered as slow and explain why it can be interesting to consider parametric or semi-parametric models. For instance Chanda (1995) considers an ARMA(p, q) model with errors in variables; using the particular structure of his model, he builds an estimator reaching the parametric rate for any distribution of the measurement error. Comte and Taupin (2000) provide an estimator for any kind of parametric family of functions m and give a general method to compute an upper bound for the rate of the estimator: they recover for well-specified families rates of order n^c, with $0 < c \leq 1/2$ even for super smooth errors. But it is worth saying that for $n = 100$, $\ln(n)^2 = 21.2$ whereas $\sqrt{n} = 10$ and for $n = 1000$, $\ln(n)^2 = 47.7$ whereas $\sqrt{n} = 31.6$! This explains why the simulation results are found as good for super smooth errors as for ordinary smooth ones.

3.2. **A result for an infinite variance ξ_t.** In the case of ordinary smooth ε's, we also recover the optimal rates of the independent regressive setting dealt with by Fan and Truong (1993), but the result is of true interest since in the case of α-mixing sequences and supremum norm, Masry (1993) has a loss of order $\ln(n)$. Moreover, the rates found in that case are powers of n namely $n^{-2k/[2(k+\delta)+1]}$. Let us give a result for ξ given by (16).

Proposition 3.2. *Under Assumptions* **A1**, **A3**$(\nu, 2)$, **A4**(2), *if the distribution of* ξ *is given by (16), if* K *is chosen as a gaussian kernel* ($K(x) = \exp(-x^2/2)/\sqrt{2\pi}$) *and if the* α*-mixing coefficients fulfill (18), then for bandwidth* $h_n = cn^{-1/9}$ *with* $c > 0$,

$$\mathbb{E}|(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f(x)|^2 = O\left(n^{-4/9}\right)$$

and

$$\mathbb{E}\int_a^b |(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f(x)|^2 dx = O\left(n^{-4/9}\right).$$

Again, we can check that this rate holds uniformly over some classes of couples (m, f_X), see Theorem 5.4 in section 5.

Then we can compute here that with $K^\star(t) = e^{-t^2/2}$, and $f_\varepsilon^\star(t) = 1/(1+4t^2)$,

$$\begin{aligned} K_n(x) &= \frac{1}{2\pi}\int e^{-itx}e^{-t^2/2}\left[1 + 4\left(\frac{t}{h_n}\right)^2\right]dt \\ &= \frac{1}{\sqrt{2\pi}}\left(1 + \frac{4(1+x^2)}{h_n^2}\right)e^{-x^2/2} \end{aligned} \tag{20}$$

3.3. **Some other results.** Let us mention moreover that Fan and Masry (1992) proved central limit theorems for the regression model with errors-in-variables in an α-mixing framework. Moreover, extensions to multivariate setting hold. For instance Masry (1993) and Fan and Masry (1992) work with multidimensional regressions in the case $k = 2$. If p is the dimension of the regression vector, then they aim at estimating a function m from $\mathbb{R}^p$ into $\mathbb{R}$. The order of the $\mathbb{L}_2$-risk for ordinary smooth ε's of order δ would become $n^{-2k/2k+(2\delta+1)p}$. For super smooth errors, the order would remain $\ln(n)^{-2k/\delta}$.

This framework leads in the autoregressive context to autoregressions of higher order provided that some stability and stationarity conditions are set on the autoregressive function m. More general conditions for nonlinear AR(p) models are given by Doukhan (1995, p. 102) requiring that the η_t's admit a density with respect to the Lebesgue measure and that there exist nonnegative constants $a_0, \ldots, a_p$ and positive constants x_0 and c such that

$$m(x) = m_(x_1, \ldots, x_p) \leq \sum_{i=1}^p a_i|x_i| \text{ if } |x| > x_0 \text{ and } \sup_{|x| \leq x_0} < \infty.$$

Moreover the unique nonnegative real zero ρ of the polynomial $P(z) = z^p - a_1 z^{p-1} - \cdots - a_p$ must satisfy $\rho < 1$.
The autoregressive process is then known to be geometrically β-mixing and thus geometrically α-mixing. Clearly, the orders found in the regressive case for dependent variables can be extended to the autoregressive case under the above conditions.

4. Simulation study

Simulations were performed using Matlab programming environment. More precisely, $T = 100$ replications of $Y_1, \dots, Y_n$ for $n = 400$, were drawn from the following model:

$$\begin{cases} X_{t+1} = m(X_t) + \eta_{t+1}, & \eta_t \text{ i.i.d. } \mathcal{N}(0,1), \\ Y_t = X_t + \varepsilon_t \end{cases}$$

with

$$\varepsilon_t = \lambda(\ln(\xi_t^2) - \omega), \omega = \mathbb{E}(\ln(\xi^2 - 1)),\ \lambda = 0.1,$$

and

(1) Linear (affine) models (L(a)): $m(x) = ax + 1$,
(2) Sinusoidal models (S(a)): $m(x) = \sin(ax) + 2$,
(3) Fan and Yao (1998)'s class of models (FY(a)): $m(x) = a(x + 5e^{-x^2})$

for values of a ensuring the stability of the autoregressive part of the model (generally a =0.3, 0.5, 0.7; in any case, we need for stability that $|a| < 1$). The variables X_i and Y_i are standardly geometrically β-mixing in that case.
We experimented two types of errors ξ:

(1) ξ gaussian $\mathcal{N}(0,1)$ associated with the kernel K_n given by (19)
(2) ξ with distribution given by (16) associated with the kernel K_n given (20).

Then the computation of $\hat{m}_n(x)$ (simply $\sum_{j=1}^n K_n((x - Y_j)/h_n)Y_{j+1}/\sum_{i=1}^n K_n((x - Y_i)/h_n)$) as defined by (13) becomes very easy provided that the bandwidth h_n is chosen. For each sample path, we compute the average squared error (ASE) at $P = 100$ grid points $(x_i)_{1\leq i\leq P}$ equally spaced in an interval with bounds the tenth and nineteenth percentiles of the data (to avoid rare observations on the borders). We report in the tables below the average of these ASE's over 100 replications:

$$\mathrm{MSE}^2 = \frac{1}{P}\sum_{i=1}^{P}[\hat{m}_n(x_i) - m(x_i)]^2$$

and the relative MSE:

$$\mathrm{RMSE}^2 = \frac{1}{P}\sum_{i=1}^{P}\frac{[\hat{m}_n(x_i) - m(x_i)]^2}{m^2(x_i)} = \frac{1}{P}\sum_{i=1}^{P}\left[1 - \frac{\hat{m}_n(x_i)}{m(x_i)}\right]^2.$$

For each set of replications, we have tested 10 values of the bandwidth (denoted by h in the tables below) equally spaced in $[0.12, 0.30]$. We report only the central values of this grid for Gaussian errors and all of them for infinite variance errors.
The results are reported in Tables 1 and 2 below. Since financial data are often long samples, some typical estimated curves for a simulated $FY(0.7)$ are pictured in Figure 1 in the two cases of noise and for two lengths of the sample, namely $n = 500$ ($h = 0.20$) and $n = 5000$ ($h = 0.12$).

From the tables, we can see that the estimation procedure performs very well. Table 2 allows to locate the optimal bandwidths. If we denote the optimal bandwidth by h_{opt} we can see that for functions $L(a)$, $h_{opt} = 0.28, 0.26, 0.22$ when

$a = 0.3, 0.5, 0.7$ and for functions $FY(a)$, $h_{opt} = 0.16, 0.14, 0.16$ when $a = 0.3, 0.5, 0.7$ for instance. A value near of 0.2 would certainly be an adequate choice for this length of sample. Besides, Figure 1 illustrates that the shape of the curve is well capture and that the estimation is almost perfect for very large samples, in both cases.

		L(a)			$S(a)$			$FY(a)$		
h	a	0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
	MSE	0.089	0.093	0.103	0.078	0.080	0.096	0.151	0.215	0.204
0.20	RMSE	0.065	0.052	0.035	0.029	0.028	0.036	0.160	0.126	0.098

Table 1. Mean Squared Errors for bandwidth $h = 0.20$ and different coefficients a. $T = 100$ samples of length $n = 400$ when the noise ξ is gaussian.

		L(a)			$S(a)$			$FY(a)$		
h	a	0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
	MSE	0.131	0.144	0.158	0.117	0.123	0.130	0.141	0.169	0.170
0.12	RMSE	0.095	0.077	0.051	0.044	0.042	0.047	0.166	0.114	0.087
	MSE	0.121	0.134	0.148	0.109	0.114	0.121	0.136	0.167	0.166
0.14	RMSE	0.088	0.071	0.048	0.040	0.039	0.045	0.157	0.110	0.084
	MSE	0.114	0.126	0.141	0.101	0.106	0.115	0.132	0.168	0.165
0.16	RMSE	0.083	0.068	0.046	0.037	0.036	0.042	0.151	0.108	0.083
	MSE	0.109	0.120	0.136	0.095	0.100	0.109	0.131	0.171	0.167
0.18	RMSE	0.079	0.064	0.045	0.036	0.035	0.040	0.148	0.108	0.084
	MSE	0.104	0.116	0.132	0.091	0.095	0.105	0.132	0.177	0.171
0.20	RMSE	0.076	0.064	0.044	0.033	0.033	0.039	0.147	0.109	0.085
	MSE	0.100	0.113	0.130	0.087	0.091	0.102	0.134	0.184	0.177
0.22	RMSE	0.070	0.061	0.042	0.032	0.032	0.037	0.147	0.111	0.087
	MSE	0.098	0.111	0.129	0.084	0.087	0.100	0.138	0.193	0.184
0.24	RMSE	0.069	0.060	0.042	0.032	0.030	0.037	0.149	0.115	0.090
	MSE	0.096	0.110	0.130	0.081	0.084	0.099	0.142	0.203	0.193
0.26	RMSE	0.069	0.059	0.042	0.030	0.028	0.037	0.153	0.119	0.094
	MSE	0.095	0.109	0.131	0.079	0.082	0.098	0.148	0.214	0.202
0.28	RMSE	0.069	0.059	0.044	0.030	0.028	0.037	0.157	0.124	0.098
	MSE	0.094	0.110	0.133	0.078	0.080	0.099	0.154	0.225	0.212
0.30	RMSE	0.069	0.060	0.045	0.028	0.028	0.037	0.162	0.130	0.102

Table 2. Mean Squared Errors for different bandwidths h and coefficients a. $T = 100$ samples of length $n = 400$ when the noise ξ has infinite variance.

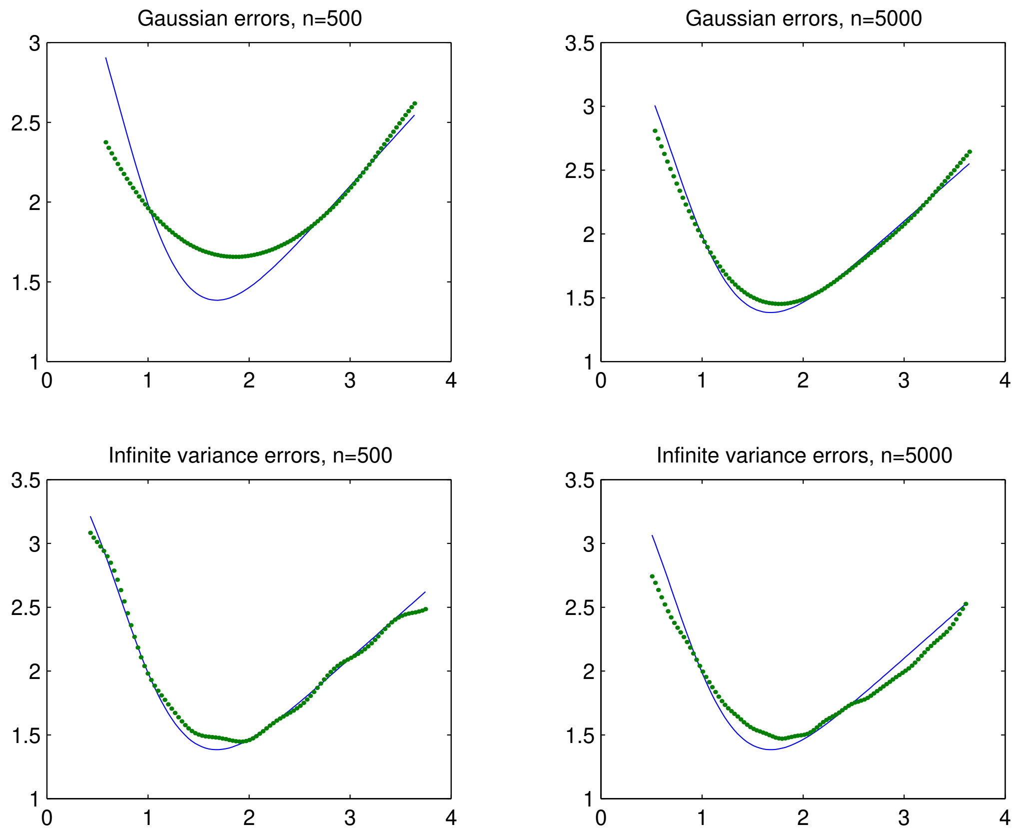

FIGURE 1. Function $FY(0.7)$ and one estimation (bold-dotted line) with $h = 0.2$ when $n = 500$ and $h = 0.12$ when $n = 5000$. Gaussian errors, MSE $= 0.171$ and 0.071, RMSE$= 0.090$ and 0.033. Infinite variance, MSE $= 0.133$ and 0.106, RMSE $= 0.053$ and 0.049.

5. Proofs

5.1. **Proofs of the results of section 2.4.2. Proof of Proposition 2.1.** We recall that the Gamma function is defined as $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}dt$ and that the digamma function is defined as $\psi(z) = d\ln(\Gamma(z))/dz = \Gamma'(z)/\Gamma(z)$, see Abramowitz and Stegun (1992), AS from now on. Then

$$\begin{aligned}
\mathbb{E}(\ln(\xi_1^2)) &= \int \ln(x^2)e^{-x^2/2}dx/\sqrt{2\pi} = 2\int_0^{+\infty} \ln(x^2)e^{-x^2/2}dx/\sqrt{2\pi} \\
&= \frac{2}{\sqrt{2\pi}}\int_0^{+\infty} \ln(2u)e^{-u}\frac{du}{\sqrt{2u}} \\
&= \frac{1}{\sqrt{\pi}}\left\{\int_0^{+\infty} u^{-1/2}e^{-u}\ln(u)du + \ln(2)\int_0^{+\infty} u^{-1/2}e^{-u}du\right\} \\
&= \frac{1}{\sqrt{\pi}}[\Gamma'(1/2) + \ln(2)\Gamma(1/2)] = \frac{1}{\sqrt{\pi}}\Gamma(1/2)[\psi(1/2) + \ln(2)].
\end{aligned}$$

Then we use that $\Gamma(1/2) = \sqrt{\pi}$ (eqn 6.1.8 in AS) and $\psi(1/2) = -\gamma - 2\ln(2)$ (eqn 6.3.3 in AS), which entails $\omega = -\gamma - \ln(2)$. The same kind of computations gives

$$\mathbb{E}(\ln(\xi_1^2)^2) = \ln(2)^2 + 2\ln(2)\psi(1/2) + \Gamma"(1/2)/\Gamma(1/2)$$

and since $\Gamma"(z)/\Gamma(z) = \psi'(z) + \psi^2(z)$ and $\psi'(1/2) = 3\zeta(2) = \pi^2/2$ (eqn 6.4.2 in AS), we find that $\sigma_\varepsilon^2 = \pi^2/2$.

For the characteristic function, we have

$$\begin{aligned}
\sqrt{2\pi}f_\varepsilon^\star(t) &= \int_{\mathbb{R}} e^{it(\ln(x^2)-\omega)}e^{-x^2/2}dx = 2e^{-it\omega}\int_0^{+\infty} e^{it\ln(x^2)}\exp(-x^2/2)dx \\
&= 2e^{-it\omega}\int_0^{+\infty} x^{2it}\exp(-x^2/2)dx \\
&= 2e^{-it\omega}\int_0^{+\infty} (2u)^{it}\exp(-u)\frac{du}{\sqrt{2u}},
\end{aligned}$$

which gives the announced formula. Lastly, the equivalent for t going to infinity is found by using Stirling's formula (eqn 6.1.37 in AS) which implies that

$$\Gamma(1/2 + it) \sim \sqrt{2\pi}e^{-1/2-it}(1/2 + it)^{it}.$$

Then writing that $\exp(it\ln(1/2 + it)) = \exp[it(\ln(\sqrt{1/4 + t^2}) + i\arctan(2t))]$ where $Arctg(2t) \sim \pi/2$ when t goes to infinity gives when taking the modulus $|\Gamma(1/2 + it)| \sim (\sqrt{2\pi/e})e^{-t\pi/2}$ and then the announced formula for $|f_\varepsilon^\star(t)|$. $\square$

The computations for the double exponential law being of the same type are not reported here.

Proof of (16). Let us compute the characteristic of ε function associated to this law for ξ and show it is the same as the characteristic function of the double exponential, namely $1/(1+4t^2)$. First

$$\begin{aligned}\omega &= \frac{1}{4}\int_{-1}^{1}\ln(x^2)dx+\frac{1}{4}\int_{|x|\geq 1}\frac{\ln(x^2)}{x^2}dx\\ &= \int_0^1\ln(x)dx+\int_1^{\infty}\frac{\ln(x)}{x^2}dx\\ &= [x\ln(x)-x]_0^1+\int_0^{+\infty}ue^{-2u}e^{u}du=-1+\Gamma(2)=0.\end{aligned}$$

Analogously

$$\mathbb{E}(\ln(\xi_1^2)^2)=2\int_0^1(\ln(x))^2dx+2\int_1^{+\infty}\frac{(\ln(x))^2}{x^2}dx=8$$

since both integrals are equal to $\Gamma(3)$ ($u=-\ln(x)$ in the first one, $u=\ln(x)$ in the second one).

$$\begin{aligned}\mathbb{E}(e^{it(\ln(\xi^2)-\omega)}) &= (1/2)\left(\int_0^1 e^{2it\ln(x)}dx+\int_1^{+\infty}\frac{e^{2it\ln(x)}}{x^2}dx\right)\\ &= (1/2)\left(\int_0^{+\infty}e^{-2itu}e^{-u}du+\int_0^{+\infty}e^{2itu}e^{-u}du\right)\\ &= (1/2)\left(\frac{1}{2it+1}+\frac{1}{1-2it}\right)=\frac{1}{1+4t^2}.\end{aligned}$$

For the density defined by (17), it is easy to compute $\omega=2\ln(a)-(k-1)(k-2)$ and

$$f_\varepsilon^\star(t)=\frac{(1-k)e^{-it\omega+2it\ln(a)}}{(2it+1)(2it-k+1)}=\frac{(1-k)e^{it(k-1)(k-2)}}{(2it+1)(2it-k+1)}.$$

5.2. **Proof of Propositions 3.1 and 3.2.** The results given here are the consequences of some generalizations of the results given by Fan and Truong (1993) in the independent case to the α-mixing case. There are two couples of theorems corresponding to super-smooth and ordinary smooth errors ε respectively. More precisely, Let

$$d_k(x)=(-1)^k\left[\frac{[m(x)f_X(x)]^{(k)}}{k!}-m(x)\frac{f_X^{(k)}(x)}{k!}\right]f_X^{-1}(x)\int_{-\infty}^{+\infty}u^kK(u)du. \tag{21}$$

Then for super smooth errors, we find:

Theorem 5.1. *Under Assumptions* **A1**, **A3**(ν,k), **A4**–**A6**, *if the mixing coefficients satisfy (18), if* $|f_\varepsilon^\star(t)|\geq\gamma_0|t|^{\delta_0}\exp(-|t|^\delta/\gamma)$ *and if* $K^\star$ *has a bounded support on* $|t|\leq M_0$, *then for bandwidth* $h_n=c(\ln(n))^{-1/\delta})$ *with* $c>M_0(2/\gamma_2)^{1/\delta}$,

$$\mathbb{E}|(\hat m_n(x)-m(x))\hat f_n(x)/f_X(x)|^2=[c^kd_k(x)]^2(\ln(n))^{-2k/\delta}(1+o(1))=O\left((\ln(n))^{-2k/\delta}\right)$$

and

$$
\begin{aligned}
\mathbb{E}\int_a^b |(\hat{m}_n(x)-m(x))\hat{f}_n(x)/f_X(x)|^2 dx &= \int_a^b [c^k d_k(x)]^2 dx (\ln(n))^{-2k/\delta}(1+o(1)) \\
&= O\left((\ln(n))^{-2k/\delta}\right)
\end{aligned}
$$

We also prove the uniformity of those rates over $\mathcal{F}_{k,B}$ where $\mathcal{F}_{k,B}$ is the following class of functions, all bounds being taken on a compact interval $[a,b]$:

$$
\mathcal{F}_{k,B} = \left\{(m, f_X) \;/\; |m^{(j)}(.)| \leq B, j=0,1,\ldots,k,\; |f_X^{(k)}(.)| \leq B\right\}.
$$

Theorem 5.2. *Under the assumptions of Theorem 5.1,*

$$
\lim_{d\to+\infty}\limsup_{n\to+\infty}\sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{P}_{(f,g)}\left\{|\hat{m}_n(x)-m(x)| \geq d(\ln(n))^{-k/\delta}\right\} = 0
$$

and if the weight function w has a support $[a,b]$, then

$$
\lim_{d\to+\infty}\limsup_{n\to+\infty}\sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{P}_{(f,g)}\left\{\|\hat{m}_n(.)-m(.)\|_{w2} \geq d(\ln(n))^{-k/\delta}\right\} = 0
$$

where $\|u(.)\|_{w2}^2 = \int |u(x)|^2 w(x)dx$.

On the other hand, if we consider ordinary smooth errors, we have the following result.

Theorem 5.3. *Under Assumptions* **A1**, **A3**(ν,k), **A4**-**A6**, *if ε admits a bounded density with ordinary smooth distribution (15) of order δ with*

$$
t^\delta f_\varepsilon^\star(t) \to \delta_1 \text{ and } t^{\delta+1}(f_\varepsilon^\star)'(t) = O(1) \text{ when } t\to+\infty,
$$

if the α-mixing coefficients fulfill $\sum_j (j+1)^{1-2/\nu}\alpha_j < +\infty$, and if

$$
\int_{-\infty}^{+\infty} |t^{\delta+1}|(|K^\star(t)| + |(K^\star)'(t)|)dt < +\infty,\quad \int_{-\infty}^{+\infty} |t^{\delta+1}K^\star(t)|^2 dt < \infty,
$$

then for bandwidth $h_n = \gamma n^{-1/(2k+2\delta+1)}$ with $\gamma > 0$,

$$
\mathbb{E}|(\hat{m}_n(x)-m(x))\hat{f}_n(x)/f(x)|^2 = O\left(n^{-2k/[2(k+\delta)+1]}\right)
$$

and

$$
\mathbb{E}\int_a^b |(\hat{m}_n(x)-m(x))\hat{f}_n(x)/f(x)|^2 dx = O\left(n^{-2k/[2(k+\delta)+1]}\right).
$$

Note that the condition on the mixing coefficients here is the same as for super smooth errors. Masry (1993) finds for the control of the supremum norm, the constraint $\alpha_k \leq Mk^{-\theta}$ with $\theta > (5+9\nu/2)/(\nu-2)$ i.e. $\theta > 11.5$ for $\nu = 4$ e.g.

Lastly, we can prove the uniformity of the above results with the following theorem:

Theorem 5.4. *Under the Assumptions of Theorem 5.3, then*

$$\lim_{d\to+\infty} \limsup_{n\to+\infty} \sup_{(f,g)\in\mathcal{F}_{k,B}} I\!P_{(f,g)}\left\{|\hat{m}_n(x)-m(x)| \geq dn^{-k/[2(k+\delta)+1]}\right\} = 0$$

and if the weight function w has a support $[a,b]$, then

$$\lim_{d\to+\infty} \limsup_{n\to+\infty} \sup_{(f,g)\in\mathcal{F}_{k,B}} I\!P_{(f,g)}\left\{\|\hat{m}(.)-m(.)\|_{w2} \geq dn^{-k/[2(k+\delta)+1]}\right\} = 0.$$

5.2.1. *The basic inequality.* Let us first recall a standard variance inequality for α-mixing variables that is extensively used in the following.

From Wolonski and Rozanov (1959) (for $p=q=\infty$) and Davydov (1970), we know that if U and V are measurable random variables with respect to $\mathcal{U}$ and $\mathcal{V}$ respectively then

$$|cov(U,V)| \leq \alpha^{1/r}(\mathcal{U},\mathcal{V})\|U\|_p\|V\|_q \text{ for any } p,\ q,\ r\geq 1 \text{ and } \frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1,$$

where $\|U\|_p = \mathbb{E}^{1/p}(|U|^p)$.
Therefore, if (U_i) is a stationary sequence of α-mixing variables, and g is a function such that $\mathbb{E}(|g(U_i)|^\nu) < +\infty$, this implies that

$$Var\left(\sum_{i=1}^n g(U_i)\right) \leq nVar(g(U_1)) + 2n\|g(U_1)\|_\nu^2\left(\sum_{k=1}^n \alpha_k^{1-2/\nu}\right) \tag{22}$$

5.2.2. *Proof of Theorem 5.1.* We follow the same line of proof as Fan and Truong (1993) and we extend their result to the α-mixing setting. We write

$$[\hat{m}_n(x)-m(x)]\frac{\hat{f}_n(x)}{f_X(x)} = \frac{A_n(x)}{f(x)} \tag{23}$$

with

$$A_n(x) = \frac{1}{nh_n}\sum_{j=1}^n K\left(\frac{x-Y_j}{h_n}\right)(Y_{j+1}-m(x)).$$

We can straightforwardly extend Fan and Truong's (1993) Lemma 1:

Lemma 5.1. *If $K^\star(.)$ vanishes outside the interval $[-M_0,M_0]$ then:*

$$I\!E(A_n(x)) = \frac{1}{h_n}\int_{-\infty}^{+\infty}[m(u)-m(x)]\,K\left(\frac{x-u}{h_n}\right)f_X(u)du.$$

Therefore the bias is given by

$$\mathbb{E}A_n(x) = f_X(x)d_k(x)h_n^k(1+o(1)). \tag{24}$$

where d_k is defined in (21).
Consequently we need only to prove that the variance of A_n is negligible with respect

to the square bias. The following bound

$$
\begin{aligned}
n^2h_n^2\mathrm{Var}(A_n(x)) &\leq 2\mathrm{Var}\left[\sum_j K_n\left(\frac{x-(X_j+\varepsilon_j)}{h_n}\right)(m(X_j)-m(x))\right] \\
&\quad +2\mathrm{Var}\left[\sum_j K_n\left(\frac{x-(X_j+\varepsilon_j)}{h_n}\right)(\varepsilon_{j+1}+\eta_{j+1})\right] \qquad (25)\\
&:= 2n^2h_n^2Var_1+2n^2h_n^2Var_2
\end{aligned}
$$

leads to study separately the two different terms appearing in (25), denoted by Var_1 and Var_2.

Since the ε_j's and the η_j's are centered and uncorrelated with past X_j's and past ε_j's, we have

(26)

$$
n^2h_n^2\mathrm{Var}_2=(\sigma_\eta^2+\sigma_\varepsilon^2)\sum_j \mathbb{E}K_n^2\left(\frac{x-(X_j+\varepsilon_j)}{h_n}\right)=n(\sigma_\eta^2+\sigma_\varepsilon^2)\mathbb{E}K_n^2\left(\frac{x-(X_1+\varepsilon_1)}{h_n}\right).
$$

We know from Fan and Truong (1993, eqn (7.8)) that

$$
\sup_x |K_n(x)|\leq \int_{-\infty}^{+\infty}\frac{|K^\star(t)|}{|f_\varepsilon^\star(t/h_n)|}dt\leq O(h_n)+O(h_n^{-1}\exp(|M_0/h_n|^\delta/\gamma). \qquad (27)
$$

Therefore Var_2 is of order $\sup|K_n(x)|^2/nh_n^2$ i.e. is of order $o(h_n^{2k})=o(\ln(n)^{-2k/\delta})$ for the choice $h_n=c(\ln(n))^{-1/\delta})$ with $c>M_0(2/\gamma)^{1/\delta}$.
To deal with Var_1, we use (22) since (X_i,ε_i) is α-mixing.

$$
\begin{aligned}
n^2h_n^2Var_1 &\leq n\mathbb{E}\left(K_n\left(\frac{x-(X_1+\varepsilon_1)}{h_n}\right)(m(X_1)-m(x))\right)^2 \\
&+ 2n\left(\sum_k \alpha_k^{1-2/\nu}\right)\mathbb{E}^{2/\nu}\left|K_n\left(\frac{x-(X_1+\varepsilon_1)}{h_n}\right)(m(X_1)-m(x))\right|^\nu \qquad (28)\\
&\leq n\left(\sup_x|K_n(x)|\right)^2\left(1+2\sum_k\alpha_k^{1-2/\nu}\right)\|m(X_1)-m(x)\|_\nu^2.
\end{aligned}
$$

Using (27) again gives the result as soon as $\sum_k \alpha_k^{1-2/\nu}<+\infty$. This ends the proof of Theorem 5.1 since the inequalities are uniform for x in $[a,b]$.□

5.2.3. *Proof of Theorem 5.2.*

Lemma 5.2. *Under the assumptions of Theorem 5.2, then*

$$
\sup_{(f,g)\in\mathcal{F}_{k,B}} I\!E_{(f,g)}\sup_x|\hat f_n(x)-f_X(x)|^2=o(1),\quad p\geq 1.
$$

Proof of Lemma 5.2. From (10), it follows that

$$\begin{aligned}\mathbb{E}(\hat{f}_n(x)) &= \frac{1}{2\pi}\int e^{-itx}K^\star(th_n)f_X^\star(t)dt = \frac{1}{2\pi h_n}\int e^{-itx}(K(./h_n)*f_X)^\star(t)dt\\ &= \frac{1}{h_n}K(./h_n)*f_X(x) = \int \frac{1}{h_n}K\left(\frac{x-u}{h_n}\right)f_X(u)du.\end{aligned}$$

Therefore

$$\sup_{(f,g)\in\mathcal{F}_{k,B}}\sup_x |\mathbb{E}(\hat{f}_n(x)) - f_X(x)| = o(1). \tag{29}$$

Next (10) also implies that

$$\sup_x |\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \leq \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{|K^\star(th_n)|}{|f_\varepsilon^\star(t)|}|\hat{\Phi}_n(t) - f_Y^\star(t)|dt.$$

Therefore, if we prove

$$\mathbb{E}\left[|\hat{\Phi}_n(t) - f_Y^\star(t)|^2\right] \leq c_2 n^{-1} \tag{30}$$

then

$$\begin{aligned}&(2\pi)^2\mathbb{E}\sup_x |\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)|^2\\ \leq\; & \left(\int_{-\infty}^{+\infty}\frac{|K^\star(th_n)|}{|f_\varepsilon^\star(t)|}dt\right)\int_{-\infty}^{+\infty}\frac{|K^\star(th_n)|}{|f_\varepsilon^\star(t)|}\mathbb{E}|\hat{\Phi}_n(t) - f_Y^\star(t)|^2dt\\ \leq\; & c_2\left(\frac{1}{\sqrt{n}h_n}\int_{-\infty}^{+\infty}\frac{|K^\star(t)|}{|f_\varepsilon^\star(t/h_n)|}dt\right)^2.\end{aligned}$$

With (27) and (29), this implies the result.
Inequality (30) is a consequence of (22) provided that $\sum_k \alpha_k < \infty$ since the Y_i's are α-mixing with the same coefficients as the X_i's, and

$$\begin{aligned}\mathbb{E}\left[|\hat{\Phi}_n(t) - f_Y^\star(t)|^2\right] &= Var\left(\frac{1}{n}\sum_j \cos(tY_j)\right) + Var\left(\frac{1}{n}\sum_j \sin(tY_j)\right)\\ &\leq \frac{2}{n}\left(1 + 2\sum_{k=1}^n \alpha_k\right),\end{aligned}$$

because the functions g involved here are bounded and we can take $\nu = +\infty$. $\Box$

Now we prove Theorem 5.2. The local rate follows from arguments similar to the previous one, therefore we focus on proving the global result. By Lemma 5.2 and (23), it suffices to prove

$$\sup_{(f,g)\in\mathcal{F}_{k,B}}\mathbb{E}\int_a^b |A_n(x)|^2 w(x)dx = O(\ln(n)^{-2k/\delta})$$

or also

$$\sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{E}\sup_x |A_n(x)|^2 = O(\ln(n)^{-2k/\delta}).$$

Since

$$\begin{aligned}\sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{E}\sup_x |A_n(x)|^2 &\leq 2 \sup_{(f,g)\in\mathcal{F}_{k,B}} \sup_x \mathbb{E}|A_n(x)|^2 \\ &\quad +2 \sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{E}\sup_x |A_n(x) - \mathbb{E}A_n(x)|^2,\end{aligned} \tag{31}$$

and from (24), the first term is of order $O(h_n^{2k})$, we study the order of the second term. Let

$$U_j(t) = e^{itY_j}Y_{j+1} - \mathbb{E}(e^{itY_j}Y_{j+1}) \text{ and } V_j(t) = e^{itY_j} - \mathbb{E}(e^{itY_j}),$$

we have

$$\begin{aligned}&(2\pi \sup_x |A_n(x) - \mathbb{E}A_n(x)|)^2 \\ \leq\ & \left[\int_{-\infty}^{+\infty} \left(\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right| + \frac{\sup_x |m(x)|}{n} \left|\sum_{j=1}^n V_j(t)\right| \right) \left|\frac{K^\star(th_n)}{f_\varepsilon^\star(t)}\right| dt\right]^2 \\ \leq\ & \left[\int_{-\infty}^{+\infty} \left|\frac{K^\star(th_n)}{f_\varepsilon^\star(t)}\right| dt\right] \int_{-\infty}^{+\infty} \left(\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right| + \frac{B}{n} \left|\sum_{j=1}^n V_j(t)\right| \right)^2 \left|\frac{K^\star(th_n)}{f_\varepsilon^\star(t)}\right| dt\end{aligned}$$

Therefore we need to bound

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|^2 \text{ and } \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n V_j(t)\right|^2$$

From Lemma 5.2, the second term is of order n^{-1}. For U_j, writing that $Y_{j+1} = m(X_i) + \varepsilon_{i+1} + \eta_{i+1}$ we can use that (X_i, ε_i) is α-mixing and apply the variance inequality (22) to $(1/n)\sum_j e^{itY_j}m(X_j) - \mathbb{E}(e^{itY_j}m(X_j))$. Then we find also for the variance of this term an order n^{-1} if $\sum_j \alpha_j^{1-2/\nu} < +\infty$. Uncorrelatedness properties ensure the same result for $Var[(1/n)\sum_j e^{itY_j}\eta_{j+1}]$ and $Var[(1/n)\sum_j e^{itY_j}\varepsilon_{j+1}]$. This implies that $\sup_{(f,g)\in\mathcal{F}_{k,B}} \sup_x |A_n(x) - \mathbb{E}A_n(x)|^2$ is of order

$$O\left(\frac{1}{n}\left[\int_{-\infty}^{\infty} \left|\frac{K^\star(th_n)}{f_\varepsilon^\star(t)}\right| dt\right]^2\right),$$

i.e. of order $o(\ln(n)^{-A})$ for any $A > 0$ using the choice of h_n and (27). This completes the proof of Theorem 5.2. □

5.2.4. *Proof of Theorem 5.3.* Using (23) again, we compute as in the proof of Theorem 5.1 the order of the bias and of the variance of $A_n(x)$. For the bias, (24) still holds. Thus we prove that $\mathrm{Var}A_n(x) = O(1/(nh_n^{1+2\delta}))$. First the decomposition (25) and the equality (26) for Var_2 still hold. Next, we use Fan's Lemma 2.1 (1991) that we recall below:

Lemma 5.3. *Suppose that $G_n(.)$ is a sequence of Borel functions satisfying*

$$G_n(y) \to G(y), \quad \text{and} \quad \sup_n |G_n(y)| \leq G^*(y),$$

where $G^*(y)$ satisfies

$$\int_{-\infty}^{+\infty} G^*(y)dy < \infty \text{ and } \lim_{y\to\infty} |yG^*(y)| = 0.$$

If x is a continuity point of a density $f(.)$, then for any sequence $h_n \to 0$, we have

$$\lim_{n\to\infty} \frac{1}{h_n} \int_{-\infty}^{+\infty} G_n\left(\frac{x-y}{h_n}\right) f(y)dy = f(x) \int_{-\infty}^{+\infty} G(y)dy.$$

We apply this result to $G_n(y) = [h_n^\delta K_n(y)]^2$. Since we know from Fan and Truong (1993) that

$$h_n^\delta K_n(y) \to \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ity} \frac{t^\delta}{\delta_1} K^\star(t)dt := J(y)$$

and that (Lemma 3 of Fan and Truong (1993) p.1918):

$$|h_n^\delta K_n(y)| \leq \frac{C}{1+|y|} \tag{32}$$

where C is some positive constant, we find that

$$\begin{aligned} Var_2 &= \frac{\sigma_\eta^2+\sigma_\varepsilon^2}{nh_n^2} \int\int K_n^2\left(\frac{x-(u+v)}{h_n}\right) f_X(u)f_\varepsilon(v)dudv \\ &= \frac{1}{nh_n^{1+2\delta}} \int_{-\infty}^{+\infty} J^2(u)du \int_{-\infty}^{+\infty} f_X(x-v)f_\varepsilon(v)dv[1+o(1)] \end{aligned} \tag{33}$$

where

$$\int_{-\infty}^{+\infty} J^2(u)du = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left|\frac{t^\delta}{\delta_1}\right|^2 |K^\star(t)|^2 dt.$$

Therefore, under the assumptions of Theorem 5.3, all integrals in (33) are finite. For Var_1, we start from (28) and find, denoting by $S(\alpha) = 1 + 2\sum_k \alpha_k^{1-2/\nu}$,

$$Var_1 \leq \frac{S(\alpha)}{1} nh_n^2 \mathbb{E}^{2/\nu}\left(\left|K_n\left(\frac{x-(X_1+\varepsilon_1)}{h_n}\right)(m(X_1)-m(x))\right|^\nu\right)$$

$$\leq \frac{S(\alpha)}{nh_n^2}\left(\int\int\left|K_n\left(\frac{x-(u+v)}{h_n}\right)\right|^\nu |m(u)-m(x)|^\nu f_X(u) f_\varepsilon(v)dudv\right)^{2/\nu}$$

$$\leq \frac{S(\alpha)}{nh_n^{1+2\delta}}\left(\int |J|^\nu(y)dy\int |m(u)-m(x)|^\nu f_X(u) f_\varepsilon(x-u)du\right)^{2/\nu}(1+o(1))$$

and the order follows from the boundedness of f_ε and the finiteness of $\mathbb{E}(|m(X_1)|^\nu)$.□

5.2.5. *Proof of Theorem 5.4.* The local rate follows from arguments similar to those given in Theorem 5.3. Thus, we focus on proving the global rate.
First, the result of Lemma 5.2 still holds under the assumptions of Theorem 5.4 because Fan and Truong (eqn. 7.16, 1993) proved that

$$\int_{-\infty}^{+\infty}\left|\frac{K^\star(t)}{f_\varepsilon^\star(t/h_n)}\right| dt = O(h_n^{-\delta})$$

for ordinary smooth errors.
From this result and relation (23), it suffices to show that

$$\sup_{(f,g)\in\mathcal{F}_{k,B}} \mathbb{E}\int_a^b |A_n(x)|^2 w(x)dx = O(n^{-2k/(2k+2\delta+1)}).$$

Since (31) still holds, it follows from (24) that the first term is of order $O(h_n^{2k})$. Therefore, we study the order of the second term of (31). Following Fan and Truong (1993), we must prove that

$$\sup_{\mathcal{F}_{k,B}}\sup_{x\in[a,b]} \mathbb{E}|A_n(x)-\mathbb{E}(A_n(x))|^2 = O\left(\frac{h_n^{1-2(\delta+1)}}{n}\right).$$

In other words, we must prove that the bounds and limits found in the proof of Theorem 5.3 are uniform with respect to x and the functions m and f_X taken in $\mathcal{F}_{k,B}$. First, Var_2 can simply be bounded by writing:

$$Var_2 = \frac{1}{nh_n}\int\int K_n^2(z) f_X(y) f_\varepsilon(x-y-h_n z)dydz \leq \frac{\|f_\varepsilon\|_\infty}{nh_n}\int K_n^2(z)dz.$$

Since from (32), $\int K_n^2(z)dz$ is bounded, we find that $Var_2 = O(1/(nh_n^{1+2\delta}))$ which is the aimed order and is uniform in x and over $\mathcal{F}_{k,B}$.
Moreover, let $T_j = K_n\left([x-(X_j+\varepsilon_j)]/h_n\right)(m(X_j)-m(x))$ and let $\mathcal{F}_n = \sigma(X_1,\ldots,X_n)$ be the σ-algebra generated by the X_i's, $i=1,\ldots,n$. We know that for any square

integrable random variables X, Y, we have the equality: $Var(X) = Var(\mathbb{E}(X|Y)) + \mathbb{E}(Var(X|Y))$. This implies here that

$$
\begin{aligned}
n^2 h_n^2 Var_1 &= Var(\sum_{j=1}^n T_j) = Var\left[\mathbb{E}\left(\sum_{j=1}^n T_j \middle| \mathcal{F}_n\right)\right] + \mathbb{E}\left[Var\left(\sum_{j=1}^n T_j \middle| \mathcal{F}_n\right)\right] \\
&= Var\left[\sum_{j=1}^n \mathbb{E}(T_j|\mathcal{F}_n)\right] + \mathbb{E}\left(\sum_{j=1}^n Var(T_j|\mathcal{F}_n)\right) := CV_1 + CV_2
\end{aligned}
$$

since the independency of the ε_j's and of the sequences X and ε imply that the T_j's are independent for given X_j's. Then we study separately the two terms CV_1 and CV_2 of the above decomposition.

$$
\begin{aligned}
CV_2 &\leq \sum_{j=1}^n \mathbb{E}\left[(m(X_j) - m(x))^2\, \mathbb{E}\left(K_n^2\left(\frac{x-(X_j+\varepsilon_j)}{h_n}\right)\middle| \mathcal{F}_n\right)\right] \\
&\leq \sum_{j=1}^n \mathbb{E}\left[(m(X_j) - m(x))^2 K_n^2\left(\frac{x-(X_j+\varepsilon_j)}{h_n}\right)\right] \\
&= nh_n \int\int K_n^2(z)(m(y)-m(x))^2 f_X(y) f_\varepsilon(x-y-zh_n)\,dydz \\
&\leq nh_n \|f_\varepsilon\|_\infty \int \frac{C^2 h_n^{-2\delta}}{(1+|z|)^2} dz \mathbb{E}[(m(X_1)-m(x))^2] \text{ with (32) again}
\end{aligned}
$$

and we find again the same order as for Var_1. Let us denote by

$$
\psi(X_j) = \mathbb{E}(T_j|\mathcal{F}_n) = (m(X_j) - m(x)) \int K_n\left(\frac{x-(X_j+v)}{h_n}\right) f_\varepsilon(v)dv.
$$

Then

$$
\begin{aligned}
CV_1 &= Var\left(\sum_{j=1}^n \psi(X_j)\right) \leq n\left(Var(\psi(X_1)) + 2\sum_j \alpha_j^{1-2/\nu}\|\psi(X_1)\|_\nu^2\right) \\
&\leq n\left(1 + 2\sum_j \alpha_j^{1-2/\nu}\right)\|\psi(X_1)\|_\nu^2 \qquad (34)
\end{aligned}
$$

and

$$\begin{aligned}
\|\psi(X_1)\|_\nu^2 &= \left(\int |\psi(u)|^\nu f_X(u)du\right)^{\nu/2} \\
&\leq \left(\int |m(u)-m(x)|^\nu \left|\int K_n^2\left(\frac{x-(u+v)}{h_n}\right) f_\varepsilon(v)dv\right|^{\nu/2} f_X(u)du\right)^{2/\nu} \\
&= h_n\left(\int |m(u)-m(x)|^\nu \left|\int K_n^2(z) f_\varepsilon(x-u-zh_n)dz\right|^{\nu/2} f_X(u)du\right)^{2/\nu} \\
&\leq h_n^{1-2\delta} D\|m(X_1)m(x)\|_\nu^2 \qquad (35)
\end{aligned}$$

where D is a positive constant deduced from (32) again. (34) and (35) imply the same uniform order for CV_1 which finally gives the result. □

References

[1] Abramowitz, M. and Stegun, I.A. (1992) *Handbook of Mathematical functions*, Dover Publications, Inc., New York.

[2] Ango Nze, P. (1992) Critères d'ergodicité de quelques modèles à représentation markovienne. *C. R. Acad. Sci. Paris Ser.I Math.*, 315, 1301-1304.

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

[4] Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities, *Journal of Political Economy* 3, 637-654.

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

[6] Comte, F. and Taupin, M.-L. (2000) Semiparametric estimation in the (auto)-regressive β-mixing model with errors-in-variables. *Preprint 606, University Paris 6.*

[7] Davydov, Yu A. (1970) The invariance principle for stationary processes. *Theory Probab. Appl.*, 15, 487-498.

[8] Ding, Z., Granger, C. W.J. and Engle, R.F. (1993) A long memory property of stock market returns and a new model. *J. Empirical Finance*, 1, 1, 83-106.

[9] Doukhan, P. (1995) *Mixing properties and examples.* Springer-Verlag.

[10] Duflo, M. (1990). *Méthodes récursives aléatoires.* Masson, Paris.

[11] Fan, J. (1991) Asymptotic normality for deconvolution kernel estimators. *Sankhya* 53, A, 97-110.

[12] Fan, J. and Masry, E. (1992) Multivariate regression estmation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.*, 43, 237-272.

[13] Fan, J. and Truong, Y.K.(1993) Nonparametric regression with errors in variables. *Ann. Statist.* 21, 4, 1900-1925.

[14] Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regressions. *Biometrika*, 85, 3, 645-660.

[15] Feldmann, D, Hardle, W., Hafner, C., Hoffmann, M. Lepski, O. and Tsybakov, A. (1998) Testing linearity in a stochastic volatility model, *Preprint SBF* 373, Humboldt Universitat Berlin.

[16] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000) Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6, 1051-1079.

[17] Gouriéroux, C. and Monfort, A. (1992) Qualitative Threshold models, *J. Econometrics* 52, 159-199.

[18] Hall, P. and Carroll, R.J. (1989) Variance function estimation in regression: the esffect of estimating the mean. *J. R. Statist. Soc.*B, 51, 3-14.

[19] Härdle, W. and Tsybakov, A. (1998) Local polynomial estimators of the volatility function in nonparametric regression. *J. Econometrics*, 81, 223-242.

[20] Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 3, 281-300.

[21] Kolmogorov, A.R. & Y.A. Rozanov (1960). On the strong mixing conditions for stationary gaussian sequences. *Theor. Probab. Appl.*, 5, 204-207.

[22] Masry, E. (1993) Multivariate regression estimation with errors-in-variables for stationary processes. *Nonparametric Statistics* 3, 13-26.

[23] Mikosch, T. (2001) Modeling Financial Time Series. *Notes on the CIRM course*, preprint.

[24] Meyn, S.P. and Tweedie, R.L.(1993) *Markov chains and stochastic stability*, Springer-Verlag, London.

[25] Neumann, M. (1994) Fully data-driven variance estimators. *Statistics*, 25, 189-212.

[26] Renault, E. and Touzi, N. (1996) Option hedging and implicit volatilities in a stochastic volatility model. *Mathematical Finance*, 6, 279-302.

[27] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, 4, 43-47.

[28] Stefanski, L.A. and Carroll, R.J. (1990) Deconvolution kernel density estimators. *Statistics*, 21, 169-184.

[29] Taupin, M.L. (1998) Semi-parametric estimation in nonlinear structural errors-in-variables model. *Preprint Université Paris-Sud* 98-35, accepted for publication in *Ann. Statist.*.

[30] Taylor, R (1986) *Modelling Financial Time Series*, John Wiley, Chichester.

[31] Tuominen, P. and Tweedie, R. L. (1994). Subgeometric rates of convergenceof f-ergodic Markov chains. *Adv . Appl. Prob.* 26, 775-798.

[32] Wolonski, V.A. and Rozanov, Y.A. (1959) Some limit theorems for random functions, PartI. *Theory Probab. Appl.*, 4, 178-197.