



# Adaptive estimation of the dynamics of a discrete time stochastic volatility model

Fabienne Comte, Claire Lacour, Yves Rozenholc

## ► To cite this version:

Fabienne Comte, Claire Lacour, Yves Rozenholc. Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *Econometrics*, 2010, 154 (1), pp.59-73. 10.1016/j.jeconom.2009.07.001 . hal-00170740

**HAL Id: hal-00170740**

**<https://hal.science/hal-00170740>**

Submitted on 10 Sep 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Adaptive estimation of the dynamics of a discrete time stochastic volatility model

F. Comte<sup>1</sup>

C. Lacour<sup>1</sup>

Y. Rozenholc<sup>1</sup>

<sup>(1)</sup> *Université Paris Descartes, MAP5, UMR CNRS 8145.*

*fabienne.comte@univ-paris5.fr, claire.lacour@univ-paris5.fr, yves.rozenholc@univ-paris5.fr.*

**Summary.** This paper is concerned with the particular hidden model:  $X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}$ ,  $Z_i = X_i + \varepsilon_i$ , where  $(\xi_i)$  and  $(\varepsilon_i)$  are independent sequences of i.i.d. noise. Moreover, the sequences  $(X_i)$  and  $(\varepsilon_i)$  are independent and the distribution of  $\varepsilon$  is known. Our aim is to estimate the functions  $b$  and  $\sigma^2$  when only observations  $Z_1, \dots, Z_n$  are available. We propose to estimate  $bf$  and  $(b^2 + \sigma^2)f$  and study the integrated mean square error of projection estimators of these functions on automatically selected projection spaces. By ratio strategy, estimators of  $b$  and  $\sigma^2$  are then deduced. The mean square risk of the resulting estimators are studied and their rates are discussed. Lastly, simulation experiments are provided: constants in the penalty functions defining the estimators are calibrated and the quality of the estimators is checked on several examples.

**KEYWORDS.** Adaptive Estimation; Autoregression; Deconvolution; Heteroscedastic; Hidden Markov Model; Nonparametric Projection Estimator.

## 1. Introduction

When price processes of assets are observed, generally in discrete time, the dynamics of the unobserved underlying volatility process proves very interesting. Therefore, the so-called discrete time stochastic volatility model has recently become most popular and widely studied, see Ghysels *et al.* (1996) or Shephard (1996). In this paper, we propose a statistical strategy corresponding to the following model:

$$\begin{cases} Y_i = \exp(X_i/2)\eta_i, \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}, \end{cases}$$

where  $(\eta_i)$  and  $(\xi_i)$  are two independent sequences of independent and identically distributed (i.i.d.) random variables (noise processes). The only available observations are  $Y_1, \dots, Y_n$ , the process of interest is the unobserved volatility  $V_i = \exp(X_i/2)$ . We describe an estimation method leading to nonparametric estimates of the functions  $b$  and  $\sigma^2$  driving the dynamics of the volatility process  $(V_i)$ .

To achieve this aim, we use a deconvolution strategy, which is made possible through the rewriting of the model as follows

$$\begin{cases} Z_i = X_i + \varepsilon_i \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1} \end{cases} \quad (1)$$

where  $\varepsilon_i = \ln(\eta_i^2) - \mathbb{E}(\ln(\eta_i^2))$  and  $Z_i = \ln(Y_i^2) - \mathbb{E}(\ln(\eta_i^2))$ .

In such a setting, regarding the identifiability of the model, it must be assumed that the distribution of  $\varepsilon$ ,  $f_\varepsilon$  (or equivalently of  $\eta$ ) is fully known. For instance, the process  $\eta$  is often

modelled as a standard Gaussian i.i.d. sequence, and then  $\varepsilon_i$  has the distribution of  $\ln(\mathcal{N}(0, 1)^2) + \ln(2) + C$  where  $C$  is the Euler constant. Van Es *et al.* (2005) specifically study this case in terms of density estimation, however more general distributions can also be considered (see Comte *et al.* (2006b)).

Model (1) may be considered as a non-linear autoregressive model observed with an additive noise with known (and general) distribution. In this case, the process is sometimes called autoregression with errors-in-variables. Such models have already been studied, but in parametric or semi-parametric context only (see Chanda (1995), Comte and Taupin (2001)).

Lastly, Model (1) belongs to the general class of hidden Markov models (HMM). These models constitute a very famous class of discrete time processes with applications in various areas (see Cappé, Moulines and Ryden (2005)). Here our model is simpler in the sense that our noise is additive, but in standard HMMs it is assumed that the joint density of  $(X_i, Z_i)$  has a parametric form.

To our knowledge, the question of estimating  $b$  and  $\sigma^2$  in Model (1) on the basis of observations  $Z_1, \dots, Z_n$  has not been studied yet. Only the following regressive model  $Z_i = X_i + \varepsilon_i$ ,  $Y_i = b(X_i) + \xi_i$ , in which  $(Y_i)$  and  $(Z_i)$  for  $i = 1, \dots, n+1$  are observed, has received attention. Then two processes are observed, all sequences  $(X_i)$ ,  $(\xi_i)$ ,  $(\varepsilon_i)$  can be supposed independent identically distributed (i.i.d.) and independent from each other, and  $(Y_i)$  is homoscedastic ( $\sigma(x) \equiv 1$ ). In this context, Fan and Truong (1990), and Comte and Taupin (2007) study the problem of the estimation of  $b$ . See also Fan *et al.* (1991), Fan and Masry (1992), Ioannides and Alevizos (1997), Koo and Lee (1998). Most authors propose estimators of  $b$  based on the ratio of two estimators and this quotient strategy is also adopted in our more general setting. More precisely, we assume that the process  $(X_i)$  is stationary, with stationary density denoted by  $f$ , and we estimate  $b$  (resp.  $b^2 + \sigma^2$ ) as a ratio of an estimator of  $b f$  (resp.  $(b^2 + \sigma^2) f$ ) divided by  $f$ .

Several papers develop estimation methods for  $f$ , see Fan (1991), Pensky and Vidakovic (1999), Comte *et al.* (2006b), and the optimality of the rates studied in Fan (1991), Butucea (2004) and Butucea and Tsybakov (2007). The adaptive estimator of Comte *et al.* (2006b) is used in this study. We adopt the same type of projection strategy on automatically selected projection spaces for the numerators. In this respect, this allows to consider general classes of noise density  $f_\varepsilon$  and also various classes of regularities for the functions to estimate ( $b f$ ,  $(b^2 + \sigma^2) f$ ,  $f$ ).

The proofs of our results involve the study of several centered empirical processes and are interesting more for their general schemes than for their technical details. It is nevertheless worth mentioning that, in the end, we obtain flexible tools that work in a satisfactory way. The programmes developed for density deconvolution in Comte *et al.* (2007) can indeed be generalized to the present framework.

The plan of the paper is the following: we first give the notations, the model assumptions and describe projection spaces in Section 2. Next, Section 3 explains the estimation strategy for  $b$  and gives bounds of the integrated mean square risk of the estimators. Section 4 develops the same study for the estimation of  $\sigma^2$ . Simulation experiments are conducted in Section 5 in order to illustrate the method and compare its performance with previous results. Lastly, proofs are gathered in Sections 6-7-8 and an appendix, namely section 9, describes auxiliary tools.

## 2. General setting and assumptions

### 2.1. The principle

Let us assume that the sequence  $(X_i)$  is stationary and let us denote by  $f$  their stationary density. The principle of the estimation methods relies in all cases on a “Nadaraya-Watson-strategy” in the sense that  $b$  or  $b^2 + \sigma^2$  are estimated as ratio of an estimator of  $\ell = b f$  (respectively  $\vartheta = (b^2 + \sigma^2) f$ ) and an estimator of  $f$ .

In all cases, we use the adaptive estimator of  $f$  described in Comte *et al.* (2006b, 2007) which studies independent and  $\beta$ -mixing contexts.

## 2.2. Notations and Assumptions

Subsequently we denote by  $u^*$  the Fourier transform of the function  $u$  defined as  $u^*(t) = \int e^{itx} u(x) dx$ , and by  $\|u\|$ ,  $\|u\|_\infty$ ,  $\|u\|_{\infty, K}$ ,  $\langle u, v \rangle$ ,  $u * v$  the quantities

$$\begin{aligned} \|u\|^2 &= \int u^2(x) dx, \quad \|u\|_\infty = \sup_{x \in \mathbb{R}} |u(x)|, \quad \|u\|_{\infty, K} = \sup_{x \in K} |u(x)|, \\ \langle u, v \rangle &= \int u(x) \bar{v}(x) dx \text{ with } z\bar{z} = |z|^2 \text{ and } u * v(x) = \int u(t) \bar{v}(x-t) dt. \end{aligned}$$

Moreover, we recall that for any integrable and square-integrable functions  $u, u_1, u_2$ ,

$$(u^*)^*(x) = 2\pi u(-x) \text{ and } \langle u_1, u_2 \rangle = (2\pi)^{-1} \langle u_1^*, u_2^* \rangle. \quad (2)$$

We consider the autoregressive model (1). The assumptions are the following:

- A1** (i) The  $\varepsilon_i$ 's are i.i.d. centered ( $\mathbb{E}(\varepsilon_1) = 0$ ) random variables with finite variance,  $\mathbb{E}(\varepsilon_1^2) = s_\varepsilon^2$ . The density of  $\varepsilon_1$ ,  $f_\varepsilon$ , belongs to  $\mathbb{L}_2(\mathbb{R})$ , and for all  $x \in \mathbb{R}$ ,  $f_\varepsilon^*(x) \neq 0$ .
- (ii) The  $\xi_i$ 's are i.i.d. centered with unit variance ( $\mathbb{E}(\xi_1^2) = 1$ ) and  $\mathbb{E}(\xi_1^3) = 0$ .
- A2** The  $X_i$ 's are stationary and absolutely regular.
- A3** The sequences  $(X_i)_{i \in \mathbb{N}}$  and  $(\varepsilon_i)_{i \in \mathbb{N}}$  are independent. The sequences  $(\xi_i)_{i \in \mathbb{N}}$  and  $(\varepsilon_i)_{i \in \mathbb{N}}$  are independent.

The  $Z_i$ 's are observed but the  $X_i$ 's are not, the stationary density  $f$  of the  $X_i$ 's is unknown and the density  $f_\varepsilon$  of the  $\varepsilon_i$ 's is known.

Standard assumptions on  $b$ ,  $\sigma$  and the  $\xi_i$ 's ensure that the sequence  $(X_i)_{i \in \mathbb{Z}}$  is stationary with stationary density denoted by  $f$ . This sequence is also absolutely regular, with  $\beta$ -mixing coefficients denoted by  $\beta(k)$ , see Doukhan (1994) or Comte and Rozenholc (2002) for precise sets of conditions. We shall consider that the mixing is at least arithmetical with rate  $\theta$ , i.e. that there exists  $\theta > 0$  such that

$$\forall k \in \mathbb{N}, \beta(k) \leq (1+k)^{-(1+\theta)}, \quad (3)$$

or, more often, geometrical, i.e.  $\exists \theta > 0, \forall k \in \mathbb{N}, \beta(k) \leq e^{-\theta k}$ . The definition of the  $\beta$ -mixing coefficients and related properties are recalled in Section 9.

Moreover, as we develop an  $\mathbb{L}^2$ -strategy, we require the target functions to be square-integrable.

- A4** The function to estimate ( $\ell = bf$ ,  $\vartheta = (b^2 + \sigma^2)f$ , or  $f$ ) is square-integrable.

According to Assumption **A3** the (unknown) density  $h$  of the  $Z_i$ 's equals  $f * f_\varepsilon$ . This implies that  $h^* = f^* f_\varepsilon^*$  and  $f^* = h^* / f_\varepsilon^*$ , a relation which explains the estimation strategy. It is well known that the rate of convergence for estimating  $f$  is strongly related to the rate of decrease of  $f_\varepsilon^*$ . More precisely, the smoother  $f_\varepsilon$ , the slower the rate of convergence for estimating  $f$  is and we shall see that the same happens for the estimation of  $bf$  or  $(b^2 + \sigma^2)f$ . Nevertheless, this rate of convergence can be improved by assuming some additional regularity conditions on  $f$ ,  $bf$  or  $(b^2 + \sigma^2)f$ . These regularity conditions are described by considering functions in the space:

$$\mathcal{S}_{s,a,r}(A) = \{u : \int_{-\infty}^{+\infty} |u^*(x)|^2 (x^2 + 1)^s \exp\{2a|x|^r\} dx \leq A\}, \quad (4)$$

for nonnegative constants  $s, a, r$  and  $A > 0$ . When  $r = 0$ , this corresponds to Sobolev spaces of order  $s$ ; when  $r > 0, a > 0$ , this corresponds to analytic functions, which are often called "super-smooth" functions.

In the following, we also assume that  $f_\varepsilon$  is such that

**A5** For all  $t$  in  $\mathbb{R}$ ,  $A_0(t^2 + 1)^{-\gamma/2} \exp\{-\mu|t|^\delta\} \leq |f_\varepsilon^*(t)| \leq A'_0(t^2 + 1)^{-\gamma/2} \exp\{-\mu|t|^\delta\}$ , with  $\gamma > 1/2$  if  $\delta = 0$ .

Under Assumption **A5**, when  $\delta = 0$ , the errors are usually called "ordinary smooth" errors, and "super smooth" errors when  $\delta > 0, \mu > 0$ . The standard examples are the following : Gaussian or Cauchy distributions are super smooth of order  $(\gamma = 0, \mu = 1/2, \delta = 2)$  and  $(\gamma = 0, \mu = 1, \delta = 1)$  respectively, and the Laplace (symmetric exponential) distribution is ordinary smooth ( $\delta = 0$ ) of order  $\gamma = 2$ . When  $\varepsilon = \ln(\eta^2) - \mathbb{E}(\ln(\eta^2))$  with  $\eta \sim \mathcal{N}(0, 1)$  as in Van Es *et al.* (2005), then  $\varepsilon$  is super-smooth with  $\gamma = 0, \mu = \pi/2$  and  $\delta = 1$ .

### 2.3. The projection spaces

As projection estimators are used in all cases, we hereby provide a description of the projection spaces. Let us define

$$\varphi(x) = \frac{\sin(\pi x)}{\pi x} \quad \text{and} \quad \varphi_{m,j}(x) = \sqrt{m} \varphi(mx - j),$$

where  $m$  can be replaced by  $2^m$ . It is well known (see Meyer (1990), p.22) that  $\{\varphi_{m,j}\}_{j \in \mathbb{Z}}$  is an orthonormal basis of the space of square integrable function having Fourier transform with compact support included into  $[-\pi m, \pi m]$ . Such a space is denoted by  $S_m$ .

$$S_m = \text{Span}\{\varphi_{m,j}, j \in \mathbb{Z}\} = \{f \in \mathbb{L}^2(\mathbb{R}), \text{supp}(f^*) \subset [-\pi m, \pi m]\}.$$

Moreover,  $(S_m)_{m \in \mathcal{M}_n}$ , with  $\mathcal{M}_n = \{1, \dots, m_n\}$ , denotes the collection of linear spaces.

In practice, we should consider the truncated spaces  $S_m^{(n)} = \text{Span}\{\varphi_{m,j}, j \in \mathbb{Z}, |j| \leq K_n\}$ , where  $K_n$  is an integer depending on  $n$ , and the associated estimators under the additional assumption:  $\int x^2 \psi^2(x) dx < A_\psi < \infty$ , where  $\psi = bf, (b^2 + \sigma^2)f$  or  $f$  is the function to estimate. This is done in Comte *et al.* (2006b) and does not change the main part of the study. For the sake of simplicity, we write in the theoretical part of the present study the sums over  $\mathbb{Z}$ .

## 3. Estimation of $b$

### 3.1. The steps of the estimation

#### 3.1.1. First step: the estimators of $\ell = bf$

The orthogonal projection of  $\ell = bf$  on  $S_m$ ,  $\ell_m$ , is given by

$$\ell_m = \sum_{j \in \mathbb{Z}} a_{m,j}(\ell) \varphi_{m,j} \quad \text{with} \quad a_{m,j}(\ell) = \int_{\mathbb{R}} \varphi_{m,j}(x) \ell(x) dx = \langle \varphi_{m,j}, \ell \rangle. \quad (5)$$

For  $t$  belonging to a space  $S_m$  of the collection  $(S_m)_{m \in \mathcal{M}_n}$ , let

$$\gamma_n(t) = \frac{1}{n} \sum_{i=1}^n (\|t\|^2 - 2Z_{i+1} u_t^*(Z_i)), \quad u_t(x) = \frac{1}{2\pi} \frac{t^*(-x)}{f_\varepsilon^*(x)}.$$

The following sequence of equalities, relying on the Fourier equalities (2), explains the choice of the contrast  $\gamma_n$ :

$$\begin{aligned}\mathbb{E}(Z_2 u_t^*(Z_1)) &= \mathbb{E}(b(X_1) u_t^*(Z_1)) = \langle u_t^* * f_\varepsilon(-\cdot), bf \rangle = \frac{1}{2\pi} \langle \frac{t^*}{f_\varepsilon^*(-\cdot)} f_\varepsilon^*(-\cdot), (bf)^* \rangle \\ &= \frac{1}{2\pi} \langle t^*, (bf)^* \rangle = \langle t, bf \rangle = \mathbb{E}(b(X_1)t(X_1)) = \int t(x)b(x)f(x)dx = \langle t, \ell \rangle.\end{aligned}$$

Therefore, we find that  $\mathbb{E}(\gamma_n(t)) = \|t\|^2 - 2\langle \ell, t \rangle = \|t - \ell\|^2 - \|\ell\|^2$  is minimal when  $t = \ell$ . Thus, we define

$$\hat{\ell}_m = \arg \min_{t \in S_m} \gamma_n(t) \quad (6)$$

The estimator can also be written

$$\hat{\ell}_m = \sum_{j \in \mathbb{Z}} \hat{a}_{m,j}(\ell) \varphi_{m,j}, \text{ with } \hat{a}_{m,j}(\ell) = \frac{1}{n} \sum_{i=1}^n Z_{i+1} u_{\varphi_{m,j}}^*(Z_i). \quad (7)$$

Now, to select an adequate value of  $m$ , we define  $\hat{\ell}_{\hat{m}}$ , by setting

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left\{ \gamma_n(\hat{\ell}_m) + \text{pen}(m) \right\},$$

where the penalty function is given by  $\text{pen}(m) = \kappa \mathbb{E}(Z_2^2) \Psi(m)$  where

$$\Psi(m) = \begin{cases} \frac{\Delta(m)}{m^{[(3\delta-1)/2] \wedge \delta}} & \text{if } 0 \leq \delta < 1/3 \\ \frac{\Delta(m)}{n} & \text{if } \delta \geq 1/3, \end{cases} \quad \text{and } \Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{dx}{|f_\varepsilon^*(x)|^2}, \quad (8)$$

where  $x \wedge y := \inf(x, y)$ . In practice  $\mathbb{E}(Z_2^2)$  is unknown and is replaced by its empirical version. The resulting penalty function,  $\widehat{\text{pen}}$ , then becomes random. We note that  $\gamma_n(\hat{\ell}_m) = -\sum_{j \in \mathbb{Z}} [\hat{a}_{m,j}(\ell)]^2$ , which explains (10) below.

### 3.1.2. Second step: the estimators of $f$

The second stage of the estimation procedure is to estimate  $f$ . In fact, Comte *et al.* (2006b, 2007) explain how to estimate  $f$  in an adaptive way and in a mixing context. The estimator of  $f$  on  $S_m$  is defined by

$$\hat{f}_m = \sum_{j \in \mathbb{Z}} \hat{a}_{m,j}(f) \varphi_{m,j} \text{ with } \hat{a}_{m,j}(f) = \frac{1}{n} \sum_{i=1}^n u_{\varphi_{m,j}}^*(Z_i). \quad (9)$$

Then we define  $\hat{f}_{\hat{m}}$ ,

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left\{ -\sum_{j \in \mathbb{Z}} [\hat{a}_{m,j}(f)]^2 + \widehat{\text{pen}}(m) \right\}, \quad (10)$$

where the penalty function is given by  $\widehat{\text{pen}}(m) = \kappa \Psi(m)$  with  $\Psi(m)$  given by (8). For the properties of  $\hat{f}_{\hat{m}}$  we refer to Comte *et al.* (2006b). Up to the multiplicative constants, the control of the mean square risk of the estimator is the same as the one obtained for  $\ell$  here.

### 3.1.3. Last step: the estimator of $b$ .

We estimate  $b$  on a compact set  $B$  only and the following additional assumption is required:

- A6** (i)  $\forall x \in B, f_0 \leq f(x) \leq f_1$  for two positive constants  $f_0$  and  $f_1$ .  
(ii)  $b$  is bounded on  $B$ .

Then we can define:

$$\tilde{b} = \hat{b}_{\hat{m}, \hat{m}} = \frac{\hat{\ell}_{\hat{m}}}{\hat{f}_{\hat{m}}} \text{ if } \|\hat{\ell}_{\hat{m}} / \hat{f}_{\hat{m}}\| \leq a_n, \quad \tilde{b} = \hat{b}_{\hat{m}, \hat{m}} = 0 \text{ else,} \quad (11)$$

where  $a_n$  is a sequence to be specified later.

### 3.2. Risk bound for $\hat{\ell}_m$ and $\hat{\ell}_{\hat{m}}$ .

We define the following empirical centered process

$$\nu_n(t) = \frac{1}{n} \sum_{k=1}^n (Z_{k+1} u_t^*(Z_k) - \langle t, \ell \rangle),$$

and with (5) and (7), we note that the following equalities hold

$$\begin{aligned} \|\ell - \hat{\ell}_m\|^2 &= \|\ell - \ell_m\|^2 + \|\ell_m - \hat{\ell}_m\|^2 = \|\ell - \ell_m\|^2 + \sum_{j \in \mathbb{Z}} (a_{m,j}(\ell) - \hat{a}_{m,j}(\ell))^2 \\ &= \|\ell - \ell_m\|^2 + \sum_{j \in \mathbb{Z}} \nu_n^2(\varphi_{m,j}). \end{aligned}$$

Therefore

$$\mathbb{E} \|\ell - \hat{\ell}_m\|^2 \leq \|\ell - \ell_m\|^2 + \sum_{j \in \mathbb{Z}} \text{Var}[\nu_n(\varphi_{m,j})].$$

The following decomposition will prove useful:  $\nu_n(t) = \nu_n^{(1)}(t) + \nu_n^{(2)}(t) + \nu_n^{(3)}(t)$  with

$$\nu_n^{(1)}(t) = \frac{1}{n} \sum_{k=1}^n \varepsilon_{k+1} u_t^*(Z_k), \quad \nu_n^{(2)}(t) = \frac{1}{n} \sum_{k=1}^n \xi_{k+1} \sigma(X_k) u_t^*(Z_k), \quad (12)$$

$$\nu_n^{(3)}(t) = \frac{1}{n} \sum_{k=1}^n (b(X_k) u_t^*(Z_k) - \langle t, \ell \rangle). \quad (13)$$

Here the terms  $\nu_n^{(1)}$  and  $\nu_n^{(2)}$  can be kept together and benefit from the uncorrelatedness of the variables involved in the sums. The term  $\nu_n^{(3)}$  involves dependent variables. Then we find

$$\text{Var}[\nu_n(\varphi_{m,j})] \leq 2\text{Var}[\nu_n^{(1)}(\varphi_{m,j}) + \nu_n^{(2)}(\varphi_{m,j})] + 2\text{Var}[\nu_n^{(3)}(\varphi_{m,j})].$$

The first variance involves uncorrelated and centered terms and leads to

$$\text{Var} \left[ \frac{1}{n} \sum_{i=1}^n (\varepsilon_{i+1} + \sigma(X_i) \xi_{i+1}) u_{\varphi_{m,j}}^*(Z_i) \right] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}[(s_\varepsilon^2 + \sigma^2(X_i)) |u_{\varphi_{m,j}}^*(Z_i)|^2]$$

so that

$$\sum_{j \in \mathbb{Z}} \text{Var} [\nu_n^{(1)}(\varphi_{m,j}) + \nu_n^{(2)}(\varphi_{m,j})] = \frac{1}{n} \sum_{j \in \mathbb{Z}} \mathbb{E}[(s_\varepsilon^2 + \sigma^2(X_1)) |u_{\varphi_{m,j}}^*(Z_1)|^2] = \frac{(s_\varepsilon^2 + \mathbb{E}(\sigma^2(X_1))) \Delta(m)}{n}.$$

We use here the following useful property of our basis (resulting from a Parseval's formula):

$$\forall x \in \mathbb{R}, \sum_j |u_{\varphi_{m,j}}^*(x)|^2 = \Delta(m),$$

where  $\Delta(m)$  is defined by (8) and the  $u_{\varphi_{m,j}}^*(x)$  are just rewritten as Fourier coefficients.

For the second term, we use the standard tools specific to the  $\beta$ -mixing context (namely Vienet's (1997) covariance Inequality) and we can easily prove the following Lemma:

LEMMA 3.1. *Under Assumptions **A1-A3**,*

$$\sum_{j \in \mathbb{Z}} \text{Var} \left( \nu_n^{(3)}(\varphi_{m,j}) \right) \leq \mathbb{E}(b^2(X_1)) \frac{\Delta(m)}{n} + \frac{4Km}{n},$$

where  $K = \sqrt{2 \sum_{k \geq 0} (k+1)\beta(k)\mathbb{E}(b^4(X_1))}$  if  $\mathbb{E}(b^4(X_1)) < \infty$  and  $\sum_k k\beta(k) < +\infty$ .

Therefore, the rate of the estimate  $\hat{\ell}_m$  is as follows:

PROPOSITION 3.1. *Consider the estimator  $\hat{\ell}_m$  of  $\ell$  defined by (6) where  $\ell = bf$  with  $b$  and  $f$  as in Model (1). Then under Assumptions **A1-A4**, if  $\mathbb{E}(b^4(X_1)) < +\infty$  and  $\theta > 1$  for arithmetical mixing (see (3)), we have*

$$\mathbb{E}(\|\ell - \hat{\ell}_m\|^2) \leq \|\ell - \ell_m\|^2 + 2\mathbb{E}(Z_2^2) \frac{\Delta(m)}{n} + 8K \frac{m}{n}.$$

In addition, assume that  $\ell$  belongs to a space  $\mathcal{S}_{s,a,r}(A)$  defined by (4) and that Assumption **A5** is fulfilled. Then the estimate  $\hat{\ell}_{\tilde{m}}$  with  $\tilde{m}$  as in Table 1, has the rates given in Table 1 in terms of its mean square integrated risk  $\mathbb{E}(\|\hat{\ell}_m - \ell\|^2)$ .

The orders given in Table 1 classically take into account that:

(a) When  $\ell$  belongs to a space  $\mathcal{S}_{s,a,r}(A)$  defined by (4), then the order of the squared bias is

$$\|\ell - \ell_m\|^2 = (2\pi)^{-1} \int_{|x| \geq \pi m} |f^*(x)|^2 dx \leq Cm^{-2s} \exp(-2a(\pi m)^r).$$

(b) When  $f_\varepsilon^*$  satisfies **A5** then the order of the variance term is bounded by:

$$C\Delta(m)/n \leq C'm^{2\gamma+1-\delta} \exp(2\mu(\pi m)^\delta)/n.$$

When  $r > 0, \delta > 0$  the value of  $\tilde{m}$  is not explicitly given. It is obtained as the solution of the equation

$$\tilde{m}^{2s+2\gamma+1-r} \exp\{2\mu(\pi\tilde{m})^\delta + 2a\pi^r \tilde{m}^r\} = O(n).$$

For explicit formulae for the rates, see Lacour (2006).

These rates enhance the interest of building an estimator for which the choice of the relevant model  $m$  is automatically performed. This is done with  $\hat{\ell}_{\tilde{m}}$ , and we can prove the following result:

THEOREM 3.1. *Assume that Assumptions **A1-A4** hold, that  $\mathbb{E}(b^8(X_1)), \mathbb{E}(\sigma^8(X_1))$  and  $\mathbb{E}(\xi_1^8)$  are finite and that  $\mathbb{E}(\varepsilon_1^6) < +\infty$ . Assume moreover that the process  $X$  is geometrically  $\beta$ -mixing, (or arithmetically  $\beta$ -mixing with  $\theta > 14$ ) and that the collection  $\mathcal{M}_n$  is such that for all  $m \in \mathcal{M}_n$ ,  $\text{pen}(m) \leq 1$ , then*

$$\mathbb{E}(\|\hat{\ell}_{\tilde{m}} - \ell\|^2) \leq C \inf_{m \in \mathcal{M}_n} (\|\ell - \ell_m\|^2 + \text{pen}(m)) + \frac{C'}{n}.$$

$\delta = 0$		$\delta > 0$
$f_\varepsilon$ ordinary smooth		$f_\varepsilon$ supersmooth
$r = 0$	$\pi\tilde{m} = O(n^{1/(2s+2\gamma+1)})$	$\pi\tilde{m} = [\ln(n)/(2\mu+1)]^{1/\delta}$
$\ell$ Sobolev( $s$ )	rate = $O(n^{-2s/(2s+2\gamma+1)})$	rate = $O((\ln(n))^{-2s/\delta})$
$r > 0$	$\pi\tilde{m} = [\ln(n)/2a]^{1/r}$	$\tilde{m}$ solution of
$\ell$ $C^\infty$	rate = $O\left(\frac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$	$\tilde{m}^{2s+2\gamma+1-r} \exp\{2\mu(\pi\tilde{m})^\delta + 2a\pi^r \tilde{m}^r\}$ = $O(n)$

**Fig. 1.** Choice of  $\tilde{m}$  and corresponding rates under **A1-A5** and if  $\ell$  belongs to  $\mathcal{S}_{s,a,r}(A)$  defined by (4).

The proof of Theorem 3.1 is based on the study of several empirical processes deduced from the preliminary decomposition of  $\nu_n$  given by (12)-(13) and is sketched in Section 6.

Theorem 3.1 shows that the estimator automatically selects the optimal  $m$  when  $\delta \leq 1/3$ , since in that case, the penalty has exactly the same order as the variance (namely  $\Delta(m)/n$ ). When  $\delta > 1/3$ , a compromise is still performed, but the penalty is slightly greater than the variance. In an asymptotic setting, this implies a loss in the rate of convergence of the estimator, but this loss can be shown to be negligible with respect to the rates. For discussions on this point, see Comte *et al.* (2006b).

### 3.3. Risk bounds for $\tilde{b}$

Comte *et al.* (2006a) that  $\hat{f}_{\tilde{m}}$  satisfies the same inequality as  $\hat{\ell}_{\tilde{m}}$  that follows.

**THEOREM 3.2.** *Assume that Assumptions **A1-A4** hold. Assume that the process  $X$  is geometrically  $\beta$ -mixing, (or arithmetically  $\beta$ -mixing with  $\theta > 3$  in (3)) and that the collection  $\mathcal{M}_n$  is such that for all  $m \in \mathcal{M}_n$ ,  $\text{pen}(m) \leq 1$ , then*

$$\mathbb{E}(\|\hat{f}_{\tilde{m}} - f\|^2) \leq C \inf_{m \in \mathcal{M}_n} (\|f - f_m\|^2 + \text{pen}(m)) + \frac{C'}{n}$$

where  $f_m$  denotes the orthogonal projection of  $f$  on  $S_m$ .

Then it is common (see e.g. Lacour (2005) or Comte and Taupin (2007)) to obtain that under the assumptions of Theorems 3.1 and 3.2, under **A5**, if  $f$  belongs to a space  $\mathcal{S}_{s,a,r}(A)$  with  $s > 1/2$  if  $r = 0$  and  $\ell$  to a space  $\mathcal{S}_{s',a',r'}(A')$ , if  $\ln(\ln(n)) \leq m_n \leq (n/\ln(n))^{1/(2\gamma+1)}$  for  $\hat{f}_{\tilde{m}}$ , under the additional assumption **A6**, and for  $n$  great enough, that

$$\mathbb{E}(\|\tilde{b} - b\|_B^2) \leq C_1 \mathbb{E}(\|\hat{f}_{\tilde{m}} - f\|^2) + C_2 \mathbb{E}(\|\hat{\ell}_{\tilde{m}} - \ell\|^2) + \frac{C_3}{n}$$

where  $a_n = n^\omega$  with  $\omega > 1/2$  and  $C_1, C_2, C_3$  are constants.

## 4. Estimation of $\sigma^2$

### 4.1. Steps of the estimation

We now aim at estimating  $\sigma^2$  and we also follow the strategy described in Section 2.1.

*First step.* We set  $\vartheta = (b^2 + \sigma^2)f$  and we first estimate  $\vartheta$ . To this end, we consider the following contrast:

$$\check{\gamma}_n(t) = \|t\|^2 - \frac{2}{n} \sum_{k=1}^n (Z_{k+1}^2 - \sigma_\varepsilon^2) u_t^*(Z_k). \quad (14)$$

As  $f_\varepsilon$  is assumed to be known, then so is the variance  $s_\varepsilon^2$ . Suggestions on how to estimate it can be drawn from Butucea and Matias (2004). Then we define

$$\hat{\vartheta}_m = \arg \min_{t \in S_m} \check{\gamma}_n(t) \text{ and } \check{m} = \arg \min_{m \in \mathcal{M}_n} \check{\gamma}_n(\hat{\vartheta}_m) + \text{pen}(m) \quad (15)$$

where  $\text{pen}(m)$  is a penalty function given by:  $\text{pen}(m) = \kappa \mathbb{E}((Z_2^2 - s_\varepsilon^2)^2) \Psi(m)$ , with  $\Psi(m)$  given by (8). Again, the expectation  $\mathbb{E}[(Z_2^2 - s_\varepsilon^2)^2]$  is replaced by its empirical version in practice.

*Second step.* As previously, we use as an estimator of  $f$ , the estimator  $\hat{f}_{\check{m}}$  as defined by (9)-(10). Its risk is controlled by Theorem 3.2.

*Third step.* We obtain, by defining, similarly to (11),

$$\widetilde{b^2 + \sigma^2} = \frac{\hat{\vartheta}_{\check{m}}}{\hat{f}_{\check{m}}} \text{ if } \|\hat{\vartheta}_{\check{m}}/\hat{f}_{\check{m}}\| \leq \check{a}_n \text{ and } 0 \text{ otherwise,}$$

and  $\check{a}_n$  is a sequence to be specified in the same way as  $a_n$  for the estimation of  $b$ . Clearly,  $\widetilde{b^2 + \sigma^2}$  is an estimator of  $b^2 + \sigma^2$ . For the study of steps 2 and 3, see Section 3.3.

*Fourth step.* The estimator of  $\sigma^2$  must be built by setting

$$\tilde{\sigma}^2 = \widetilde{b^2 + \sigma^2} - (\tilde{b})^2.$$

Clearly, as  $\|\tilde{\sigma}^2 - \sigma^2\|^2 \leq 2\|\widetilde{b^2 + \sigma^2} - (b^2 + \sigma^2)\|^2 + 2\|b + \tilde{b}\|^2\|b - \tilde{b}\|^2$ , the risk of the final estimator is the sum of the risks of the estimators of  $b^2 + \sigma^2$  and  $b$ , provided that  $b$  is bounded and  $\tilde{b}$  is bounded with high probability. The latter step is studied from an empirical point of view only.

#### 4.2. Risk bounds for $\hat{\vartheta}_m$ and $\hat{\vartheta}_{\check{m}}$ .

It is not difficult to check that  $\mathbb{E}(\check{\gamma}_n(t)) = \|t\|^2 - 2\langle \vartheta, t \rangle$  which justifies the choice of  $\check{\gamma}_n$  given in (14). We can also easily obtain the decomposition  $\check{\gamma}_n(t) - \check{\gamma}_n(s) = \|t - \vartheta\|^2 - \|s - \vartheta\|^2 - 2\check{\nu}_n(t - s)$  where

$$\check{\nu}_n(t) = \frac{1}{n} \sum_{k=1}^n [(Z_{k+1}^2 - s_\varepsilon^2)u_t^*(Z_k) - \langle t, \vartheta \rangle].$$

As for  $b$  previously, we can write that

$$\|\hat{\vartheta}_m - \vartheta\|^2 = \|\vartheta_m - \vartheta\|^2 + \sum_{j \in \mathbb{Z}} \check{\nu}_n^2(\varphi_{m,j}).$$

With the same tools as for the study of  $\ell$ , using a relevant decomposition of the empirical process  $\check{\nu}_n$ , we prove (see Section 7) that:

**PROPOSITION 4.1.** *Consider the estimator  $\hat{\vartheta}_m$  of  $\vartheta$  defined by (15) where  $\vartheta = (b^2 + \sigma^2)f$  with  $b$ ,  $\sigma$  and  $f$  as in Model (1). Then under Assumptions **A1-A4**, and if  $\xi_{2,\varepsilon_1}, b^2(X_1)$  and  $\sigma^2(X_1)$  admit moments of order 4, then*

$$\mathbb{E}(\|\vartheta - \hat{\vartheta}_m\|_2^2) \leq \|\vartheta - \vartheta_m\|^2 + 4\mathbb{E}[(Z_2^2 - s_\varepsilon^2)^2] \frac{\Delta(m)}{n} + \check{K} \frac{m}{n}$$

where  $\check{K} = 16\sqrt{2 \sum_{k \geq 0} (k+1)\beta(k)\mathbb{E}((b^2(X_1) + \sigma^2(X_1))^4)}$  if  $\sum_k k\beta(k) < +\infty$ .

Method	$n =$	Model $b_{01}$				Model $b_{02}$			
		Laplace error		Normal error		Laplace error		Normal error	
Oracle Kernel		400	800	400	800	400	800	400	800
Projection		2.92	2.30	6.72	5.86	3.07	2.22	5.15	3.92
		4.95	2.76	6.42	4.72	3.63	2.53	5.06	2.59

**Fig. 2.** Average Squared Errors (ASE) for estimating  $b_{01}$  (ASE  $(\times 10^{-6})$ ) and  $b_{02}$  (ASE  $(\times 10^{-2})$ ).

It appears from the details of the above study that the empirical processes involved in the decomposition of  $\check{\nu}_n$  are of the same type as the processes studied for the estimation of  $\ell$ . Therefore, we give the risk bound for  $\hat{\vartheta}_{\check{m}}$  but we omit the proof.

**THEOREM 4.1.** *Assume that Assumptions **A1-A4** hold, that  $\mathbb{E}(b^p(X_1))$ ,  $\mathbb{E}(\sigma^p(X_1))$  and  $\mathbb{E}(\xi_1^p)$  are finite for a  $p \geq 16$  and that  $\mathbb{E}(\varepsilon_1^{12}) < +\infty$ . Assume that the process  $X$  is geometrically  $\beta$ -mixing and that the collection  $\mathcal{M}_n$  is such that for all  $m \in \mathcal{M}_n$ ,  $\text{p}\check{\text{e}}\text{n}(m) \leq 1$ , then*

$$\mathbb{E}(\|\hat{\vartheta}_{\check{m}} - \vartheta\|^2) \leq C \inf_{m \in \mathcal{M}_n} (\|\vartheta - \vartheta_m\|^2 + \text{p}\check{\text{e}}\text{n}(m)) + \frac{C'}{n}.$$

## 5. Simulation results

For the simulations, we adapt the deconvolution algorithm described in Comte *et al.* (2006a,2007) to the new context here. The penalties are chosen as:

$$\frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^n W_i^2 \right) \begin{cases} \pi m + \frac{\ln^3(\pi m)}{1 + s_\varepsilon^2} + (\pi m)^3 s_\varepsilon^2 + \frac{3}{5} (\pi m)^5 s_\varepsilon^4 & \text{if } \varepsilon_1 \text{ Laplace.} \\ \left( \pi m + \frac{\ln^3(\pi m)}{1 + s_\varepsilon^2} + (\pi m)^3 s_\varepsilon^2 \right) \int_0^1 e^{(\pi m u)^2 s_\varepsilon^2} du & \text{if } \varepsilon_1 \text{ Gaussian.} \end{cases}$$

where  $W_i \equiv 1$  for the estimation of  $f$ ,  $W_i = Z_i$  for the estimation of  $b$  and  $W_i = Z_i^2 - s_\varepsilon^2$  for the estimation of  $b^2 + \sigma^2$ . In both cases, a (negligible) logarithmic term (namely  $\ln^3(\pi m)/(1 + s_\varepsilon^2)$ ) is added as a standard correction for small dimensions. The constant  $\kappa$  is chosen equal to  $\pi$  in the Gaussian case, the term  $(\pi m)^3 s_\varepsilon^2$  corresponds to the loss in the rate when  $\delta = 2$ . In the Laplace case, the constant  $\kappa$  is term by term modified when computing  $\Delta(m)$ . Indeed in that case, we find  $\Delta(m) = (1/\pi)(\pi m + (\pi m)^3 s_\varepsilon^2/3 + (\pi m)^5 s_\varepsilon^4/20)$ . We select  $m$  among 26 values ranging from  $10 \ln(n)/(\pi n)$  to 10. We refer to Comte *et al.* (2007) for stability properties of the estimation algorithm.

### 5.1. Two examples in a regression context

First we compare our method with the kernel strategy described by Fan and Truong (1993). The model here is  $Y_i = b(X_i) + \xi_i$ ,  $Z_i = X_i + \varepsilon_i$ , with observations  $((Y_1, Z_1), \dots, (Y_n, Z_n))$ , and unobserved i.i.d  $X_i$ 's. Two regression functions are considered:

$$b_{01}(x) = x_+^3(1-x)_+^3 \text{ and } b_{02}(x) = 1 + 4x.$$

The variance of  $\varepsilon$ ,  $s_\varepsilon^2$  is adjusted in all cases such that  $\text{Var}(X)/(s_\varepsilon^2 + \text{Var}(X)) = 0.70$ . The  $X_i$ 's are  $\mathcal{N}(0.5, 0.25^2)$  and the  $\xi_i$ 's are  $\mathcal{N}(0, 0.0015^2)$  with  $b_{01}$  and  $\mathcal{N}(0, 0.25^2)$  with  $b_{02}$ . The convolution noise  $\varepsilon$  is such that either

$$f_\varepsilon^*(x) = \exp(-\frac{1}{2}s_\varepsilon^2 x^2) \text{ or } f_\varepsilon^*(x) = \frac{1}{1 + \frac{1}{2}s_\varepsilon^2 x^2}, \quad (16)$$

corresponding to a Gaussian super-smooth case or a Laplace ordinary smooth case.

For each simulation, we compute the average squared error (ASE) at 101 grid points from 0.1 to 0.9 of our adaptive estimator and average these ASEs over 100 replications. We compare it with the “oracle” computed by Fan and Truong (1993): this is not an estimator but an oracle because they average ASEs obtained with the best bandwidth in terms of the (unknown in practice) ASE, for each replication: they choose *a posteriori* the bandwidth that minimizes the ASE. We report Fan and Truong’s kernel results and ours, in Table fig.2. We mention that if we had also computed oracles, we would have systematically had better results than theirs. With our true adaptive estimator, our results remain better than their oracles in the Gaussian case. They are slightly deteriorated in the Laplace case, but they keep the same order as Fan and Truong (1993)’s oracles. Note that we do not study the case  $n = 200$  because it is too small for our method to work in a satisfactory way.

## 5.2. Three examples in an heteroscedastic autoregressive context

Comte and Rozenholc (2002) provides a simulation study of the direct model  $X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}$  when the  $X_i$ ’s are observed. The strategy consists in a penalized mean-square contrast minimization which can not be applied to the present context. But we can keep some couples of functions  $(b, \sigma)$  and see how the deconvolution method behaves with respect to the estimation of these functions. More precisely, we borrow the following couples from Comte and Rozenholc (2002):

$$\begin{cases} b_1(x) = 0.25 \sin(2\pi x + \pi/3) & \sigma_1(x) = s_1(0.31 + 0.7 \exp(-5x^2)), s_1 = 0.4 \\ b_2(x) = -0.25(x + 2 * \exp(-16x^2)) & \sigma_2(x) = s_2(0.2 + 0.4 * \exp(-2x^2)), s_2 = 0.5 \\ b_3(x) = 1/(1 + e^{-x}) & \sigma_3(x) = s_3(\phi(x + 1.2) + 1.5\phi(x - 1.2)), s_3 = 0.32, \end{cases} \quad (17)$$

where  $\phi$  is the Gaussian  $\mathcal{N}(0, 1)$  probability distribution function. Moreover, we choose  $\xi \sim \mathcal{N}(0, 1)$ ,  $\varepsilon$  is either Laplace or Gaussian as given by (16) with  $s_\varepsilon = 0.1$ .

Figures 4-5-6 illustrate our results in the three cases of couples  $(b_i, \sigma_i^2)$  given in (17). For the estimation of  $f$ , the density of the  $X_i$ ’s, the true function is unknown, and the estimated function only is plotted. We can see that  $b$  and  $b^2 + \sigma^2$  are well estimated by the ratio strategy. The extraction of  $\sigma^2$  sometimes suffers from scale problems (if  $\sigma^2$  is much smaller than  $b^2$  or if both are very small). The relative order of both variance of  $\varepsilon$  and quantity  $s_i$  in the definition of  $\sigma_i$ ,  $i = 1, 2, 3$  seem to play an important role in the quality of the estimation.

Figures 4-5-6 (top right and bottom left) also plot the data sets generated, not only the  $Z_i$ ’s used for the estimation, but also the  $X_i$ ’s, to show the influence of the noise  $\varepsilon$ : a line joins  $X_i$  to  $Z_i$ , for each  $i$ , with a  $+$  for the true observation  $Z_i$ .

We also performed a Monte Carlo study which is reported in Table fig.3. We show that there is little difference between Laplace and Gaussian  $\varepsilon_i$ ’s, in spite of the difference between the theoretical rates, and that increasing the sample size leads to noticeable improvements of the results.

## 6. Proofs

### 6.1. Proof of Lemma 3.1

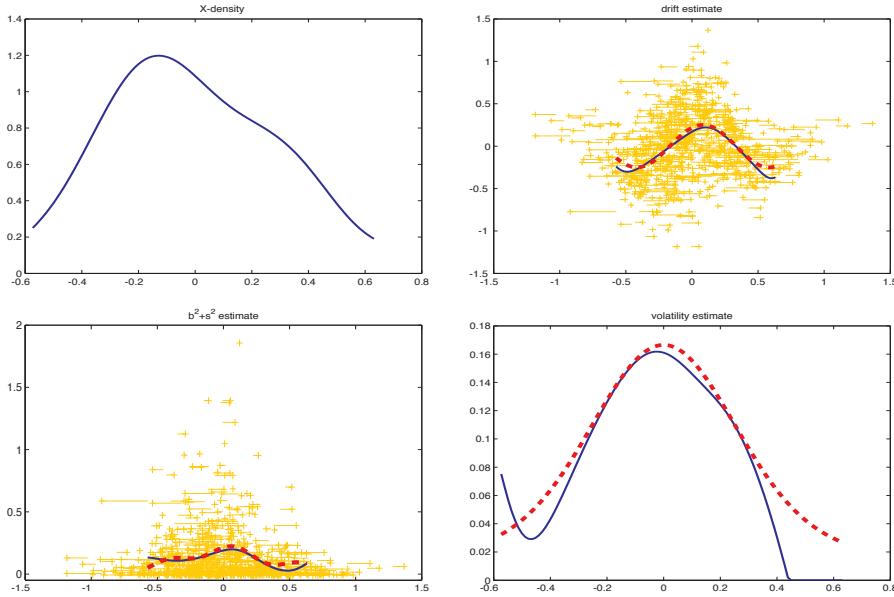
$$\text{Var} \left( \nu_n^{(3)}(\varphi_{m,j}) \right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var} \left( b(X_k) u_{\varphi_{m,j}}^*(Z_k) \right) + \frac{1}{n^2} \sum_{1 \leq k \neq l \leq n} \text{cov} \left( b(X_k) u_{\varphi_{m,j}}^*(Z_k), b(X_l) u_{\varphi_{m,j}}^*(Z_l) \right).$$

Then

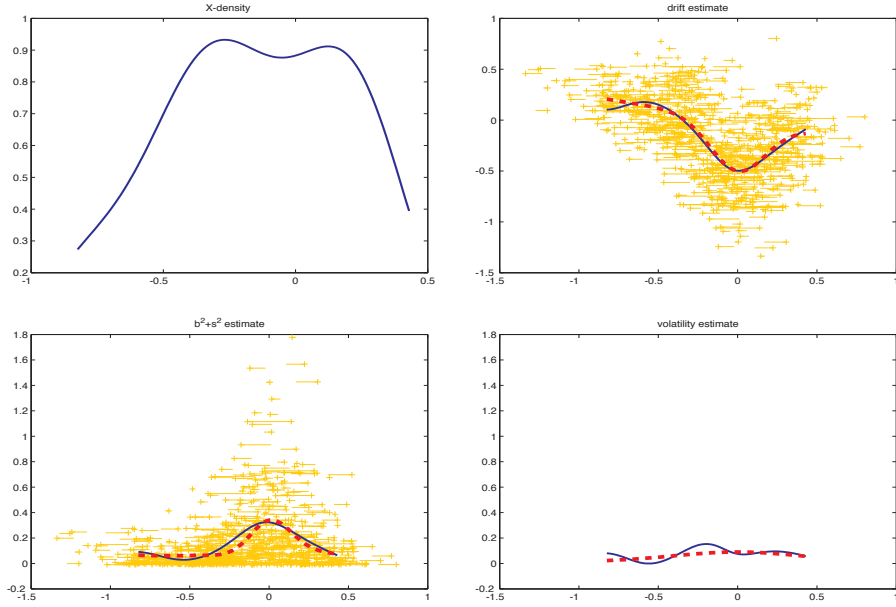
$$\mathbb{E}(b(X_k) u_{\varphi_{m,j}}^*(Z_k)) = \frac{1}{2\pi} \int \mathbb{E}(b(X_1) e^{ixX_1}) \varphi_{m,j}^*(-x) dx = \langle bf, \varphi_{m,j} \rangle = \mathbb{E}(b(X_1) \varphi_{m,j}(X_1))$$

$n =$	Laplace $\varepsilon$			Gaussian $\varepsilon$		
	400	800	1600	400	800	1600
$b_1$	0.7989 (0.5976)	0.3341 (0.1966)	0.1677 (0.0773)	0.8278 (0.7159)	0.4993 (0.3558)	0.2247 (0.1352)
$b_1^2 + \sigma_1^2$	0.1515 (0.1033)	0.1011 (0.0643)	0.0529 (0.0278)	0.1810 (0.1153)	0.1117 (0.0580)	0.0637 (0.0300)
$\sigma_1^2$	0.0946 (0.0426)	0.0823 (0.0330)	0.0534 (0.0242)	0.1112 (0.0471)	0.0880 (0.0345)	0.0620 (0.0215)
$b_2$	4.8038 (2.2205)	3.8121 (2.7124)	1.7513 (1.7207)	4.8138 (2.2205)	4.2150 (2.7124)	1.6728 (1.7207)
$b_2^2 + \sigma_2^2$	5.4090 (6.8739)	2.7263 (2.3491)	1.7643 (1.8050)	4.0527 (4.0048)	2.6609 (2.3839)	1.5938 (1.1365)
$\sigma_2^2$	4.8544 (5.2914)	2.6762 (1.7226)	1.8548 (1.5342)	3.5284 (4.0048)	2.8233 (2.3839)	1.6558 (1.1365)
$b_3$	0.0787 (0.1029)	0.0389 (0.0657)	0.0291 (0.0345)	0.0973 (0.1402)	0.0704 (0.1439)	0.0313 (0.0492)
$b_3^2 + \sigma_3^2$	0.1532 (0.1613)	0.0702 (0.0798)	0.0604 (0.0588)	0.1789 (0.1818)	0.1064 (0.1999)	0.0608 (0.0717)
$\sigma_3^2$	0.0276 (0.0528)	0.0145 (0.0260)	0.0132 (0.0169)	0.0516 (0.1117)	0.0206 (0.0337)	0.0135 (0.0217)

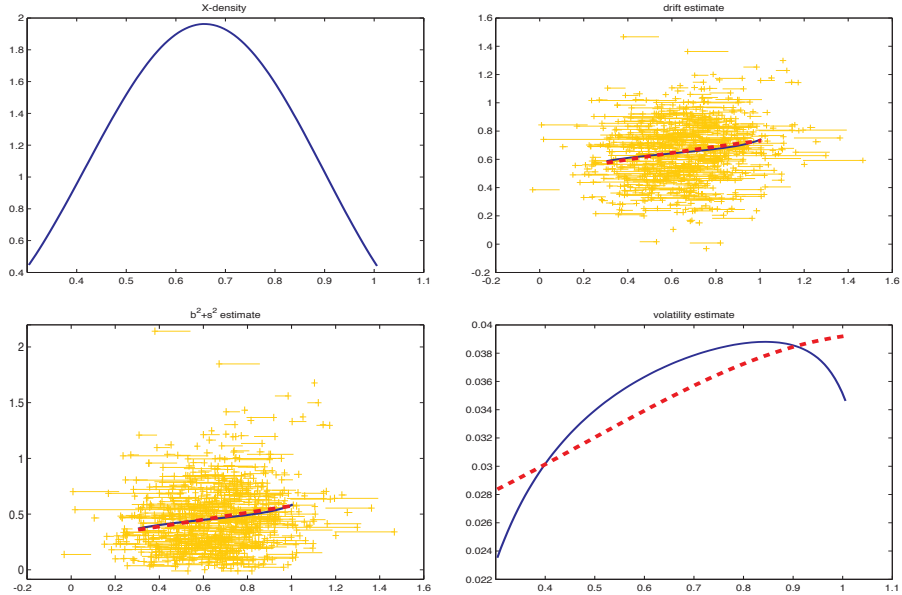
**Fig. 3.** ASE $\times 100$  (with standard deviation  $\times 100$  in parenthesis) for the estimation of  $b_i$  and  $\sigma_i^2$ ,  $i = 1, 2, 3$  given in (17), for 100 replications of the estimation procedure.



**Fig. 4.** Example 1.  $n = 800$ , dotted line: true function, full line: estimated function. Top left: estimated density of the  $X_i$ 's, top right: true and estimated  $b_1$  with data  $(X_i, Z_i)$ , bottom left: true and estimated  $b_1^2 + \sigma_1^2$  with data  $(X_i^2, Z_i^2 - s_\varepsilon^2)$ , bottom right: true and estimated  $\sigma_1^2$ .



**Fig. 5.** Example 2.  $n = 800$ , dotted line: true function, full line: estimated function. Top left: estimated density of the  $X_i$ 's, top right: true and estimated  $b_2$  with data  $(X_i, Z_i)$ , bottom left: true and estimated  $b_2^2 + \sigma_2^2$  with data  $(X_i^2, Z_i^2 - s_\varepsilon^2)$ , bottom right: true and estimated  $\sigma_2^2$ .



**Fig. 6.** Example 3.  $n = 800$ , dotted line: true function, full line: estimated function. Top left: estimated density of the  $X_i$ 's, top right: true and estimated  $b_1$  with data  $(X_i, Z_i)$ , bottom left: true and estimated  $b_3^2 + \sigma_3^2$  with data  $(X_i^2, Z_i^2 - s_\varepsilon^2)$ , bottom right: true and estimated  $\sigma_3^2$ .

and note that, for  $k \neq l$ ,

$$\begin{aligned}\mathbb{E}(b(X_k)u_{\varphi_{m,j}}^*(Z_k)b(X_l)u_{\varphi_{m,j}}^*(Z_l)) &= \frac{1}{4\pi^2} \iint \mathbb{E}(b(X_k)b(X_l)e^{ixX_k-iyX_l})\varphi_{m,j}^*(-x)\varphi_{m,j}^*(y)dxdy \\ &= \mathbb{E}[b(X_k)b(X_l)\varphi_{m,j}(X_k)\varphi_{m,j}(X_l)].\end{aligned}$$

Therefore,

$$\text{Var}\left(\frac{1}{n}\sum_{k=1}^nb(X_k)u_{\varphi_{m,j}}^*(Z_k)\right) \leq \frac{1}{n}\text{Var}(b(X_1)u_{\varphi_{m,j}}^*(Z_1)) + \text{Var}\left(\frac{1}{n}\sum_{k=1}^nb(X_k)\varphi_{m,j}(X_k)\right).$$

The last term requires a covariance inequality for mixing variables (Delyon (1990), Viennet (1997), Theorem 9.1 in the appendix) and uses the fact that the  $X_i$ 's are  $\beta$ -mixing with coefficients  $\beta(k)$ .

$$\sum_{j \in \mathbb{Z}} \text{Var}\left[\frac{1}{n}\sum_{i=1}^nb(X_i)\varphi_{m,j}(X_i)\right] \leq \sum_{j \in \mathbb{Z}} \frac{4}{n} \int \beta(x)b^2(x)|\varphi_{m,j}(x)|^2f(x)dx \leq \frac{4m}{n} \int \beta(x)b^2(x)f(x)dx$$

where  $\beta$  is a nonnegative function such that  $\mathbb{E}(\beta^p(X)) \leq p \sum_{k \geq 0} (k+1)^{p-1} \beta(k)$  and by using that  $\|\sum_j |\varphi_{m,j}|^2(\cdot)\|_\infty = m$ . Therefore if  $\mathbb{E}(b^4(X_1)) < \infty$  and  $\theta > 1$ , then

$$\sum_{j \in \mathbb{Z}} \text{Var}\left[\frac{1}{n}\sum_{i=1}^nb(X_i)\varphi_{m,j}(X_i)\right] \leq \frac{4m\sqrt{2\sum_{k \geq 0} (k+1)\beta(k)\mathbb{E}(b^4(X_1))}}{n}.$$

Moreover  $\sum_{j \in \mathbb{Z}} \text{Var}\left(b(X_1)u_{\varphi_{m,j}}^*(Z_1)\right) \leq \mathbb{E}\left(b^2(X_1)\sum_{j \in \mathbb{Z}}(u_{\varphi_{m,j}}^*(Z_1))^2\right)$  so that

$$\sum_{j \in \mathbb{Z}} \frac{1}{n} \text{Var}\left(b(X_1)u_{\varphi_{m,j}}^*(Z_1)\right) \leq \frac{\mathbb{E}(b^2(X_1))\Delta(m)}{n},$$

which gives the result.  $\square$

## 6.2. Proof of Theorem 3.1.

The proof could be sketched as follows. Let us define for  $m, m' \in \mathcal{M}_n$ ,  $B_m(0, 1) = \{t \in S_m, \|t\| = 1\}$  and  $B_{m,m'}(0, 1) = \{t \in S_m + S_{m'}, \|t\| = 1\}$ . Under the definition of  $\hat{m}$ ,  $\forall m \in \mathcal{M}_n$ ,  $\gamma_n(\hat{\ell}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(\ell_m) + \text{pen}(m)$ . For all functions  $s$  and  $t$ ,  $\gamma_n(t) - \gamma_n(s) = \|t - \ell\|^2 - \|s - \ell\|^2 - 2\nu_n(t - s)$ , and

$$2\nu_n(\hat{\ell}_{\hat{m}} - \ell_m) \leq \frac{1}{4}\|\hat{\ell}_{\hat{m}} - \ell_m\|^2 + 4 \sup_{t \in B_{m,\hat{m}}(0,1)} \nu_n^2(t).$$

Thus, we obtain, as  $\|\hat{\ell}_{\hat{m}} - \ell_m\|^2 \leq 2\|\hat{\ell}_{\hat{m}} - \ell\|^2 + 2\|\ell_m - \ell\|^2$  that

$$\frac{1}{2}\|\hat{\ell}_{\hat{m}} - \ell\|^2 \leq \frac{3}{2}\|\ell_m - \ell\|^2 + \text{pen}(m) + 4 \sup_{t \in B_{m,\hat{m}}(0,1)} \nu_n^2(t) - \text{pen}(\hat{m}). \quad (18)$$

Then we need to find a function  $p(m, m')$  such that

$$\mathbb{E}\left(\sup_{t \in B_{m,\hat{m}}(0,1)} \nu_n^2(t) - p(m, \hat{m})\right)_+ \leq \frac{C}{n} \quad (19)$$

which in turn will fix the penalty function through the requirement:  $\forall m, m' \in \mathcal{M}_n$ ,

$$4p(m, m') \leq \text{pen}(m) + \text{pen}(m'). \quad (20)$$

Gathering (18), (19) and (20) will lead to,  $\forall m \in \mathcal{M}_n$ ,

$$\frac{1}{2} \mathbb{E} \left( \|\hat{\ell}_{\hat{m}} - \ell\|^2 \right) \leq \frac{3}{2} \|\ell_m - \ell\|^2 + 2\text{pen}(m) + \frac{4C}{n}$$

which is the result.

Now, if  $\nu_n$  is split into several terms, deduced from the first decomposition given by (12)-(13), say  $\nu_n(t) = \sum_{i=1}^3 \sum_{j=1}^{p_i} \nu_n^{(i,j)}(t)$  where  $p_i \leq 3$ , then, up to some multiplicative constants, inequality (19) will follow from inequalities

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(i,j)}(t)]^2 - p_{i,j}(m, \hat{m}) \right)_+ \leq \frac{C_{i,j}}{n},$$

with  $C = 9 \sum_{i,j} C_{i,j}$  and  $p(m, m') = 9 \sum_{i,j} p_{i,j}(m, m')$ . The study of the  $\nu_n^{(i,j)}(t)$  is explained below.

First we split  $\nu_n^{(1)}$  in two parts, so that both expressions involve independent variables, conditionally to (X):  $\nu_n^{(1)} = \nu_n^{(1, \text{odd})} + \nu_n^{(1, \text{even})}$  where

$$\nu_n^{(1, \text{even})}(t) = \frac{1}{n} \sum_{1 \leq 2k \leq n} \varepsilon_{2k+1} u_t^*(Z_{2k}), \quad \nu_n^{(1, \text{odd})}(t) = \frac{1}{n} \sum_{1 \leq 2k+1 \leq n} \varepsilon_{2k+2} u_t^*(Z_{2k+1}).$$

Now, we shall study  $\nu_n^{(1, \text{even})}$  only since both terms lead to the same type of result. As Talagrand's Inequality requires the random variables involved to be bounded, we have an additional step that allows to obtain the result under a moment condition on the  $\varepsilon_i$ 's:  $\nu_n^{(1, \text{even})} = \nu_n^{(1,1)} + \nu_n^{(1,2)} + \nu_n^{(1,3)}$  with

$$\begin{aligned} \nu_n^{(1,1)}(t) &= \frac{1}{n} \sum_{1 \leq 2k \leq n} \left[ \varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4}} u_t^*(Z_{2k}) - \mathbb{E}_X(\varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4}} u_t^*(Z_{2k})) \right], \\ \nu_n^{(1,2)}(t) &= \frac{1}{n} \sum_{1 \leq 2k \leq n} \mathbb{E}(\varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4}} [t(X_{2k}) - \mathbb{E}(t(X_{2k}))]) \end{aligned}$$

and

$$\nu_n^{(1,3)}(t) = \frac{1}{n} \sum_{1 \leq 2k \leq n} \left[ \varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| > n^{1/4}} u_t^*(Z_{2k}) - \mathbb{E}(\varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| > n^{1/4}} u_t^*(Z_{2k})) \right],$$

where  $\mathbb{E}_X$  denotes the conditional expectation given  $(X_k)_{1 \leq k \leq n+1}$ . It is worth noticing that  $\nu_n^{(1,2)}$  vanishes if the  $\varepsilon$ 's are symmetric and  $\nu_n^{(1,3)}(t)$  is negligible under adequate moment conditions on the  $\varepsilon$ 's.

The following lemmas are proved below:

LEMMA 6.1.

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(1,1)}(t)]^2 - p_{1,1}(m, \hat{m}) \right)_+ \leq \frac{C}{n},$$

where  $p_{1,1}(m, m') = K \mathbb{E}(\varepsilon_1^2) \Psi(m \vee m')$  where  $K$  is a numerical constant and  $\Psi(m)$  is defined by (8).

LEMMA 6.2. *If the process  $(X_k)$  is geometrically  $\beta$ -mixing (or arithmetically with  $\theta > 3$ ), then*

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(1,2)}(t)]^2 - K \mathbb{E}(|\varepsilon_1|) \sum_k \beta(k) \frac{m + \hat{m}}{n} \right)_+ \leq \frac{C}{n}.$$

LEMMA 6.3. *If  $\mathbb{E}(\varepsilon_1^6) < +\infty$ , and  $m_n$  is the largest value of  $m$  such that  $\Delta(m_n)/n \leq 1$ , then,*

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(1,3)}(t)]^2 \right) \leq \mathbb{E} \left( \sup_{t \in B_{m_n}(0,1)} (\nu_n^{(1,3)}(t))^2 \right) \leq \frac{2\mathbb{E}(\varepsilon_1^6)}{n}.$$

For the study of  $\nu_n^{(2)}(t)$ , a result is given, whose proof is detailed in Section 8:

LEMMA 6.4. *Let  $\tau_n(t) = \nu_n^{(2)}(t) = (1/n) \sum_{k=1}^n \xi_{k+1} \sigma(X_k) u_t^*(Z_k)$ . Under the assumptions of Theorem 3.1,*

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\tau_n(t)]^2 - p_\tau(m, \hat{m}) \right)_+ \leq \frac{C}{n},$$

where  $p_\tau(m, m') = \kappa \mathbb{E}(\sigma^2(X_1)) \Psi(m \vee m')$ .

For  $\nu_n^{(3)}(t)$  we write  $\nu_n^{(3)}(t) = \nu_n^{(3,1)}(t) + \nu_n^{(3,2)}(t)$  with

$$\nu_n^{(3,1)}(t) = \frac{1}{n} \sum_{k=1}^n [b(X_k) u_t^*(Z_k) - b(X_k) t(X_k)], \quad \nu_n^{(3,2)}(t) = \frac{1}{n} \sum_{k=1}^n [b(X_k) t(X_k) - \langle t, \ell \rangle],$$

where  $b(X_k) t(X_k) = \mathbb{E}_{(X)} [b(X_k) u_t^*(Z_k)]$ . For  $\nu_n^{(3,1)}(t)$  we can apply Talagrand's Inequality conditionally to  $(X)$ , for  $\nu_n^{(3,2)}(t)$ , we can use approximation techniques. More precisely, using the same techniques as previously, we get

LEMMA 6.5. *If  $\mathbb{E}(b^8(X_1)) < +\infty$ , and  $(X_i)_{i \in \mathbb{N}}$  is arithmetically  $\beta$ -mixing with  $\theta > 14$ , then*

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(3,1)}(t)]^2 - p_{3,1}(m, \hat{m}) \right)_+ \leq \frac{C}{n},$$

where  $p_{3,1}(m, m') = K \mathbb{E}(b^2(X_1)) \Psi(m \vee m')$ , and

$$\mathbb{E} \left( \sup_{t \in B_{m, \hat{m}}(0,1)} [\nu_n^{(3,2)}(t)]^2 - K' (\mathbb{E}(b^4(X_1))) \sum_k (k+1) \beta(k))^{1/2} \frac{m + \hat{m}}{n} \right)_+ \leq \frac{C}{n},$$

where  $\kappa$  and  $\kappa'$  are numerical constants.

The proof of the result concerning  $\nu_n^{(3,1)}$  follows the same line as the proof of Lemma 6.4, which is detailed in section 8 and is therefore omitted here. For  $\nu_n^{(3,2)}$ , the bound can be obtained directly by applying Talagrand's inequality (see Theorem 9.2) to this process, if  $b$  is bounded. As this is not assumed, we write  $b = b_1 + b_2$  with  $b_1(x) = b(x) \mathbf{1}_{|b(x)| \leq n^{1/4}}$  and  $b_2(x) = b(x) \mathbf{1}_{|b(x)| > n^{1/4}}$ . This allows to split the process in two parts and consequently to obtain the result under  $\mathbb{E}(|b(X_1)|^8) < +\infty$  and  $m_n \leq \sqrt{n}$ , where  $m_n$  is the largest over the  $m \in \mathcal{M}_n$  (a condition which is fulfilled in our problem).

**Proof of Lemma 6.1.**

We apply Lemma 9.2 to process  $\nu_n^{(1,1)}(t)$  conditionally to the sequence  $(X_k)_{1 \leq k \leq n}$ . Given the  $X_i$ 's, the variables  $(Z_{2k}, \varepsilon_{2k+1})_{k \geq 1}$  are independent and we have, for  $m^* = m \vee m'$ ,

$$\begin{aligned}
& \mathbb{E}_X \left( \sup_{t \in B_{m,m'}(0,1)} (\nu_n^{(1,1)}(t))^2 \right) \\
& \leq \sum_{j \in \mathbb{Z}} \mathbb{E}_X \left[ \left( \frac{1}{n} \sum_{1 \leq 2k \leq n} \varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4} u_{\varphi_{m^*,j}}^*(Z_{2k})} - \mathbb{E}_X(\varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4} u_{\varphi_{m^*,j}}^*(Z_{2k})}) \right)^2 \right] \\
& \leq \sum_{j \in \mathbb{Z}} \frac{1}{n^2} \sum_{1 \leq 2k \leq n} \text{Var}_X \left( \varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4} u_{\varphi_{m^*,j}}^*(Z_{2k})} \right) \\
& \leq \sum_{j \in \mathbb{Z}} \frac{1}{n^2} \sum_{1 \leq 2k \leq n} \mathbb{E}_X \left[ \left( \varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4} u_{\varphi_{m^*,j}}^*(Z_{2k})} \right)^2 \right] \\
& = \frac{1}{n^2} \sum_{1 \leq 2k \leq n} \mathbb{E}_X \left( \varepsilon_{2k+1}^2 \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4}} \sum_{j \in \mathbb{Z}} (u_{\varphi_{m^*,j}}^*(Z_{2k}))^2 \right) \leq \mathbb{E}(\varepsilon_1^2) \frac{\Delta(m^*)}{n} := H^2,
\end{aligned}$$

as  $\|\sum_j |u_{\varphi_{m,j}}^*(\cdot)|^2\|_\infty \leq \Delta(m)$  by using Parseval's formula.

Next  $\sup_{x,y} |y \mathbf{1}_{|y| \leq n^{1/4}} u_t^*(x)| \leq n^{1/4} \|u_t^*(\cdot)\|_\infty \leq n^{1/4} \sqrt{\Delta(m^*)} := M_1$ . Lastly, following the same method as described in Comte *et al.* (2006b), Lemma 4, we get

$$\begin{aligned}
& \sup_{t \in B_{m,m'}(0,1)} \frac{1}{n} \sum_k \text{Var}_X(\varepsilon_{2k+1} \mathbf{1}_{|\varepsilon_{2k+1}| \leq n^{1/4}} u_t^*(Z_{2k})) \\
& \leq \sup_{t \in B_{m,m'}(0,1)} \frac{1}{n} \sum_k \mathbb{E}(\varepsilon_{2k+1}^2) \mathbb{E}_X[(u_t^*(Z_{2k}))^2] \leq \mathbb{E}(\varepsilon_1^2) \sqrt{\Delta_2(m^*)} / (2\pi), \text{ where}
\end{aligned}$$

$$\Delta_2(m) = m^2 \iint \left| \frac{\varphi^*(x) \varphi^*(y)}{f_\varepsilon^*(mx) f_\varepsilon^*(my)} f_\varepsilon^*(m(x-y)) \right|^2 dx dy. \quad (21)$$

Then the usual bounds for  $\Delta_2$  hold, namely,  $\sqrt{\Delta_2(m^*)} \leq \Delta(m^*)$  if  $\delta > 1$  and if  $\delta \leq 1$ ,  $\sqrt{\Delta_2(m^*)}/2\pi \leq \kappa \Delta(m^*) / (m^*)^{(1-\delta)/2}$ . This gives  $v = c \Delta(m^*) (m^*)^{-(1-\delta)/2}$ .

Given that the orders are the same as in Comte *et al.* (2006b) for  $v$  and  $H^2$  and inserting the slight difference on  $M_1$ , it can easily be checked that the conclusion still holds and therefore the result of Lemma 6.1 follows.  $\square$

**Proof of Lemma 6.2.**

The result given in Lemma 6.2 is a standard result of density estimation for mixing variables. We refer the reader to Tribouley and Viennet (1998) or Comte and Merlevède (2002), p.217.  $\square$

**Proof of Lemma 6.3.** Let  $e_k = \varepsilon_k \mathbf{1}_{|\varepsilon_k| > n^{1/4}}$ .

$$\begin{aligned}
& \mathbb{E} \left( \sup_{t \in B_{m_n}(0,1)} (\nu_n^{(1,3)}(t))^2 \right) \leq \sum_{j \in \mathbb{Z}} \text{Var} \left( \frac{1}{n} \sum_k e_{2k+1} u_{\varphi_{m_n,j}}^*(Z_{2k}) \right) \\
&= \frac{1}{n^2} \sum_{j \in \mathbb{Z}} \left[ \sum_k \text{Var} \left( e_{2k+1} u_{\varphi_{m_n,j}}^*(Z_{2k}) \right) + \sum_{k \neq l} \text{cov} \left( e_{2k+1} u_{\varphi_{m_n,j}}^*(Z_{2k}), e_{2l+1} u_{\varphi_{m_n,j}}^*(Z_{2l}) \right) \right] \\
&\leq \frac{\mathbb{E}(e_3^2) \Delta(m_n)}{n} + \frac{1}{n^2} \sum_{j \in \mathbb{Z}} \mathbb{E}(e_3)^2 \sum_{k \neq l} \text{cov}(\varphi_{m_n,j}(X_{2k}), \varphi_{m_n,j}(X_{2l})) \\
&\leq \frac{\mathbb{E}(e_3^2) \Delta(m_n)}{n} + \frac{1}{n^2} \sum_{j \in \mathbb{Z}} \mathbb{E}(e_3)^2 \text{Var} \left( \sum_k \varphi_{m_n,j}(X_{2k}) \right) \\
&\leq \mathbb{E}(e_3^2) \left( \frac{\Delta(m_n)}{n} + \frac{m_n \sum_k \beta_k}{n} \right) \leq \frac{2\Delta(m_n)}{n} \mathbb{E}(\varepsilon_1^2 \mathbf{1}_{|\varepsilon_1| \geq n^{1/4}}) \leq 2\mathbb{E}(\varepsilon_1^6/n). \square
\end{aligned}$$

## 7. Proof of Proposition 4.1.

We bound the expectations of the empirical processes involved in order to obtain the bound of  $\sum_{j \in \mathbb{Z}} \mathbb{E}(\check{\nu}_n(\varphi_{m,j}))$ , using the decomposition  $\check{\nu}_n = \sum_{i=1}^4 \check{\nu}_n^{(i)}$  with

$$\check{\nu}_n^{(1)}(t) = (1/n) \sum_{k=1}^n [(b^2(X_k) + \sigma^2(X_k)) u_t^*(Z_k) - \langle t, \vartheta \rangle],$$

$\check{\nu}_n^{(2)}(t) = (1/n) \sum_{k=1}^n (\xi_{k+1}^2 - 1) \sigma^2(X_k) u_t^*(Z_k)$ ,  $\check{\nu}_n^{(3)}(t) = \frac{1}{n} \sum_{k=1}^n (\varepsilon_{k+1}^2 - s_\varepsilon^2) u_t^*(Z_k)$ , and  $\check{\nu}_n^{(4)}(t) = (2/n) \sum_{k=1}^n [\varepsilon_{k+1} \xi_{k+1} \sigma(X_k) + b(X_k) \varepsilon_{k+1} + \sigma(X_k) b(X_k) \xi_{k+1}] u_t^*(Z_k)$ . But it is clear that  $\check{\nu}_n^{(1)}$  is the same process as  $\nu_n^{(3)}$  with  $b(X_k)$  replaced by  $(b^2 + \sigma^2)(X_k)$ , that  $\check{\nu}_n^{(2)}$  is of the same type as  $\nu_n^{(2)}$  with  $\sigma(X_k)$  replaced by  $\sigma^2(X_k)$  and  $\xi_{k+1}$  by  $\xi_{k+1}^2 - 1$ . Next,  $\check{\nu}_n^{(3)}$  corresponds to  $\nu_n^{(1)}$  with  $\varepsilon_{k+1}$  replaced by  $\varepsilon_{k+1}^2 - s_\varepsilon^2$ . Lastly

$$\begin{aligned}
\sum_{j \in \mathbb{Z}} \mathbb{E}[(\check{\nu}_n^{(4)}(\varphi_{m,j}))^2] &= \frac{4}{n} \sum_{j \in \mathbb{Z}} \mathbb{E}[(\varepsilon_2 \xi_2 \sigma(X_1) + b(X_1) \varepsilon_2 + (\sigma b)(X_1) \xi_2)^2 (u_{\varphi_{m,j}}^*(Z_k))^2] \\
&\leq \frac{4\Delta(m)}{n} [s_\varepsilon^2 \mathbb{E}(\sigma^2(X_1)) + \mathbb{E}(b^2(X_1)(s_\varepsilon^2 + \sigma^2(X_1)))].
\end{aligned}$$

The last step is to gather the terms.  $\square$

## 8. Proof of Lemma 6.4

If  $t = t_1 + t_2$  with  $t_1$  in  $S_m$  and  $t_2$  in  $S_{m'}$ , then  $t$  is such that  $t^*$  has its support included in  $[-\pi \max(m, m'), \pi \max(m, m')]$  and therefore  $t$  belongs to  $S_{m^*}$  where  $m^* = \max(m, m')$ . We recall that  $B_{m,m'}(0,1) = \{t \in S_{m^*} / \|t\| = 1\}$ . Denote by

$$\mathbb{H}_\tau^2(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)) \Delta(m^*)/n, \quad (22)$$

and let  $\sigma_\tau^2 = \mathbb{E}(\xi_2^2) \mathbb{E}(\sigma^2(X_1)) = \mathbb{E}(\sigma^2(X_1))$ . We have

$$\mathbb{H}_\tau^2(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2) \Delta(m^*)/n + \sigma_\tau^2 \Delta(m^*)/n,$$

which is bounded by  $\mathbb{H}_{\tau,1}(m, m') + \mathbb{H}_{\tau,2}(m, m')$  where

$$\mathbb{H}_{\tau,1}(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2) \mathbf{1}_{\{n^{-1} |\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2| > \sigma_\tau^2/2\}} \frac{\Delta(m^*)}{n}$$

and  $\mathbb{H}_{\tau,2}(m, m') = 3\sigma_\tau^2 \Delta(m^*)/(2n)$ . We infer that  $\tau_n(t) = \tau_n^{(1)}(t) + \tau_n^{(2)}(t)$  with

$$\tau_n^{(1)}(t) = \frac{1}{n} \sum_{k=1}^n \xi_{k+1} \sigma(X_k) [u_t^*(Z_k) - t(X_k)], \quad \tau_n^{(2)}(t) = \frac{1}{n} \sum_{k=1}^n \xi_{k+1} \sigma(X_k) t(X_k),$$

and

$$\begin{aligned} & \mathbb{E} \left[ \sup_{t \in B_{m,m'}(0,1)} |\tau_n(t)|^2 - p_\tau(m, m') \right]_+ \\ & \leq 2\mathbb{E} \left[ \sup_{t \in B_{m,m'}(0,1)} |\tau_n^{(1)}(t)|^2 - 2(1 + 2\epsilon^2(m, m')) \mathbb{H}_\tau^2(m, m') \right]_+ \\ & \quad + 2\mathbb{E} \left[ \sup_{t \in B_{m,m'}(0,1)} |\tau_n^{(2)}(t)|^2 - 6p_2(m, m') \right]_+ \\ & \quad + \mathbb{E} [4(1 + 2\epsilon^2(m, m')) \mathbb{H}_\tau^2(m, m') + 12p_2(m, m') - p_\tau(m, m')]_+, \end{aligned}$$

where  $\epsilon(m, m')$  is specified later and

$$p_2(m, m') = \mathbb{E}(\sigma^2(X_1)) \frac{m^*}{n}. \quad (23)$$

Clearly,  $p_2(m, m')$  is negligible with respect to  $p_\tau(m, m')$ , so that for simplicity we consider that  $(12p_2(m, m') - p_\tau(m, m')/2)_+ \leq C/n$ .

$$\begin{aligned} & \mathbb{E} [4(1 + 2\epsilon^2(m, m')) \mathbb{H}_\tau^2(m, m') + 12p_2(m, m') - p_\tau(m, m')]_+ \\ & \leq 4(1 + 2\epsilon^2(m, m')) \mathbb{E} |\mathbb{H}_{\tau,1}(m, m')| + \mathbb{E} [4(1 + 2\epsilon^2(m, m')) \mathbb{H}_{\tau,2}(m, m') - p_\tau(m, m')/2]_+. \end{aligned} \quad (24)$$

Since we only consider values of  $m$  such that the penalty are bounded by some constant  $K$ , we obtain that for some  $p \geq 2$ ,  $\mathbb{E} |\mathbb{H}_{\tau,1}(m, m')|$  is bounded by

$$\begin{aligned} & C \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2 \right| \mathbf{1}_{\{n^{-1} |\sum_{i=1}^n (\xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2)| > \sigma_\tau^2/2\}} \right] \\ & \leq C 2^{p-1} \mathbb{E} \left[ \left| n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2 \right|^p \right] / \sigma_\tau^{2(p-1)}. \end{aligned}$$

Moreover, we shall see below that  $\epsilon(m, m')$  is constant (if  $\delta = 0$  or  $0 < \delta < 1/3$ ) or at most of order  $(\ln(n))^\delta$  (if  $\delta > 1/3$ ). According to Rosenthal's inequality (see Rosenthal (1970)) generalized to the mixing case (see Doukhan (1994) and Inequality (27) recalled in Lemma 9.1), we find that,

$$\mathbb{E} |n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2|^p \leq C'(p, \xi, \sigma(X)) \left( n^{1-p} + n^{-p/2} \right).$$

Now, Assumption **A1(i)**-**A5** implies that  $\gamma > 1/2$ , therefore  $|\mathcal{M}_n| \leq \sqrt{n}$  if  $\delta = 0$  and has logarithmic order if  $\delta > 0$  and thus, choosing  $p = 3$  leads to  $\sum_{m' \in \mathcal{M}_n} \mathbb{E} |(1 + 2\epsilon^2(m, m')) \mathbb{H}_{\tau,1}(m, m')| \leq C(\xi, \sigma(X))/n$ , where  $C(\xi, \sigma(X))$  is a constant depending on the moments of  $\xi_1$  and  $\sigma(X_1)$ . In particular this requires that  $\xi$  admit a moment of order 8.

The last term of the inequality (24) vanishes as soon as

$$p_\tau(m, m') = 8(1 + 2\epsilon^2(m, m'))\mathbb{H}_{\tau,2}(m, m') = 12(1 + 2\epsilon^2(m, m'))\mathbb{E}(\sigma^2(X_1))\Delta(m^*)/n.$$

For this choice of  $p_\tau(m, m')$ , we obtain that

$$\begin{aligned} & \mathbb{E}\left[\sup_{t \in B_{m, \hat{m}}(0,1)} |\tau_n(t)|^2 - p_\tau(m, \hat{m})\right]_+ \\ & \leq 2 \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\sup_{t \in B_{m, m'}(0,1)} (n^{-1} \sum_{i=1}^n \xi_i \sigma(X_i)(u_t^*(Z_i) - t(X_i)))^2 - 2(1 + 2\epsilon^2(m, m'))\mathbb{H}_\tau^2(m, m')\right]_+ \\ & \quad + 2 \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\sup_{t \in B_{m, m'}(0,1)} |\tau_n^{(2)}(t)|^2 - 6p_2(m, m')\right]_+ + \frac{C}{n}. \end{aligned}$$

Then we apply the following Lemma.

LEMMA 8.1. *Under the assumptions on the model, if  $\mathbb{E}|\xi_1|^8 < \infty$  and  $\mathbb{E}(\sigma^8(X_1))$ , then for some given  $\epsilon > 0$ :*

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\sup_{t \in B_{m, m'}(0,1)} \left(\frac{1}{n} \sum_{i=1}^n \xi_{i+1} \sigma(X_i)(u_t^*(Z_i) - t(X_i))\right)^2 - 2(1 + 2\epsilon^2)\mathbb{H}_\tau^2(m, m')\right]_+ \\ & \leq K_1 \left\{ \sum_{m' \in \mathcal{M}_n} \left[ \frac{\sigma_\tau^2 \lambda_2 \Gamma_2(m^*)}{n} \exp\left(-K_2 \epsilon^2 \frac{\Delta(m^*)}{\lambda_2 \Gamma_2(m^*)}\right) \right] + \left(1 + \frac{\ln^4(n)}{\sqrt{n}}\right) \frac{1}{n} \right\}, \end{aligned}$$

where  $\lambda_2$  is a constant,  $\Gamma_2(m)$  is defined by

$$\Gamma_2(m) = (m)^{2\gamma + \min[(1/2 - \delta/2), (1 - \delta)]} \exp\{2\mu(\pi m)^\delta\} \quad (25)$$

and  $K_1$  and  $K_2$  are constants depending on the moments of  $\xi$  and  $\sigma(X)$ .

Moreover, it also follows from Baraud *et al.* (2001) and Comte and Rozenholc (2002), that the process  $\tau_n^{(2)}$  is a standard process of the auto-regressive context and satisfies, for  $p_2(m, m')$  defined by (23),

$$2 \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\sup_{t \in B_{m, m'}(0,1)} |\tau_n^{(2)}(t)|^2 - 6p_2(m, m')\right]_+ \leq \frac{c}{n}.$$

We denote by

$$A(m) = \frac{K_1 \sigma_\tau^2}{n} \lambda_2 \Gamma_2(m) \exp\left(-K_2 \epsilon^2 \frac{\Delta(m)}{\lambda_2 \Gamma_2(m)}\right) = \frac{K_1 \sigma_\tau^2 \lambda_2 \Gamma_2(m)}{n} \exp\left(-\kappa_2 \epsilon^2 m^{(1/2 - \delta/2)_+}\right).$$

The study of  $A(m^*)$  is standard in deconvolution (see Comte *et al.* (2006b)) and leads to choose  $\epsilon^2(m, m')$  as a constant if  $\delta \leq 1/3$  and of order  $m^{\delta - (1/2 - \delta/2)_+}$  if  $\delta > 1/3$ , to ensure that  $\sum_{m' \in \mathcal{M}_n} A(m^*)$  is less than  $C/n$ .

With  $p_\tau(m, m')$  given in Lemma 6.4, by gathering all terms we find the result.  $\square$

### Proof of Lemma 8.1.

We work conditionally to the  $(\xi_i, X_i)$ 's and  $\mathbb{E}_{X, \xi}$  and  $\mathbb{P}_{X, \xi}$  denote the conditional expectations and probability for fixed  $\xi_1, \dots, \xi_n, \xi_{n+1}, X_1, \dots, X_n$ .

We apply Lemma 9.2 with  $f_t(\xi_i, X_i, Z_i) = \xi_{i+1} \sigma(X_i) u_t^*(Z_i)$ , conditionally to the  $\xi_i$ 's and  $X_i$ 's to the random variables  $(\xi_2, X_1, Z_1), \dots, (\xi_{n+1}, X_n, Z_n)$  which are independent but non identically distributed since the  $\xi_i$ 's and the  $X_i$ 's are fixed constants.

Straightforward calculations give that for  $\mathbb{H}_\tau(m, m')$  defined in (22) we have

$$\mathbb{E}_{X, \xi}^2 \left[ \sup_{t \in B_{m, m'}(0, 1)} n^{-1} \sum_{l=1}^n \xi_{l+1} \sigma(X_l) (u_t^*(Z_l) - t(X_l)) \right] \leq \mathbb{H}_\tau^2(m, m').$$

Let  $P_{j,k}^{(l)}(m) = \mathbb{E}_{X, \xi} [u_{\varphi_{m,j}}^*(Z_l) u_{\varphi_{m,k}}^*(-Z_l)]$ . Write

$$\sup_{t \in B_{m, m'}(0, 1)} \frac{1}{n} \sum_{l=1}^n \text{Var}_{X, \xi}(\xi_{l+1} \sigma(X_l) u_t^*(Z_l)) \leq \frac{1}{n} \sum_{l=1}^n \xi_{l+1}^2 \sigma^2(X_l) \left( \sum_{j,k \in \mathbb{Z}} |P_{j,k}^{(l)}(m^*)|^2 \right)^{1/2}.$$

We argue as in Comte *et al.* (2006b). Let recall that  $\Delta_2(m)$  is defined by (21). We have  $\Delta_2(m) \leq \lambda_2^2 \Gamma_2^2(m)$ , with  $\Gamma_2$  defined by (25) and  $\lambda_2 = \lambda_2(\gamma, A_0, \delta, \mu, \|f_\varepsilon\|)$ . Now, write  $P_{j,k}^{(l)}$  as

$$P_{j,k}^{(l)}(m) = \frac{m}{4\pi^2} \iint \frac{e^{-ixj - iyk} e^{im(x-y)X_l} \varphi^*(-x) \varphi^*(-y)}{f_\varepsilon^*(mx) f_\varepsilon^*(my)} f_\varepsilon^*(m(x-y)) dx dy.$$

By applying Parseval's formula we get that  $\sum_{j,k} |P_{j,k}^{(l)}(m)|^2$  equals  $\Delta_2(m)$ . We now write that  $\sup_{t \in B_{m, m'}(0, 1)} (n^{-1} \sum_{i=1}^n \text{Var}_{X, \xi}(\xi_{i+1} \sigma(X_i) u_t^*(Z_i))) \leq (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)) \lambda_2 \Gamma_2(m^*)$ , and thus we take  $v_\tau(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)) \lambda_2 \Gamma_2(m^*)$ . Lastly, since  $\sup_{t \in B_{m, m'}(0, 1)} \|f_t\|_\infty \leq 2 \max_{1 \leq i \leq n} |\xi_{i+1} \sigma(X_i)| \sqrt{\Delta(m^*)}$ , we take  $M_{1,\tau}(m, m') = 2 \max_{1 \leq i \leq n} |\xi_{i+1} \sigma(X_i)| \sqrt{\Delta(m^*)}$ . By applying Lemma 9.2, we get for some constants  $\kappa_1, \kappa_2, \kappa_3$

$$\begin{aligned} & \mathbb{E}_{X, \xi} \left[ \sup_{t \in B_{m, m'}(0, 1)} \nu_{n,1}^2(t) - 2(1 + 2\epsilon^2) \mathbb{H}_\tau^2 \right]_+ \\ & \leq K_1 \left[ \frac{\lambda_2 \Gamma_2(m^*)}{n^2} \left( \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) \right) \exp \left\{ -K_2 \epsilon^2 \frac{\Delta(m^*)}{\lambda_2 \Gamma_2(m^*)} \right\} \right. \\ & \quad \left. + \frac{\Delta(m^*)}{n^2} \left( \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \right) \exp \left\{ -K_3 \epsilon C(\epsilon^2) \frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}}{\max_i |\xi_{i+1} \sigma(X_i)|} \right\} \right] \end{aligned}$$

To relax the conditioning, it suffices to integrate with respect to the law of the  $(\xi_{i+1}, X_i)$ 's the above expression. The first term in the bound simply becomes:

$$\sigma_\tau^2 \lambda_2 \Gamma_2(m^*) \exp[-\kappa_2 \epsilon \Delta(m^*) / (\lambda_2 \Gamma_2(m^*))] / n.$$

The second term is bounded by

$$\frac{\Delta(m^*)}{n^2} \mathbb{E} \left[ \left( \max |\xi_{i+1} \sigma(X_i)|^2 \right) \exp \left( -\kappa_3 \epsilon C(\epsilon^2) \frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}}{\max_{1 \leq i \leq n} |\xi_{i+1} \sigma(X_i)|} \right) \right]. \quad (26)$$

Since we only consider integers  $m$  such that the penalty term is bounded, we have  $\Delta(m)/n \leq K$  and the sum of the above terms for  $m' \in \mathcal{M}_n$  and  $|\mathcal{M}_n| \leq \sqrt{n}$  is less than

$$\frac{K}{\sqrt{n}} \mathbb{E} \left[ \left( \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \right) \exp \left( -\kappa_3 \epsilon C(\epsilon^2) \frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}}{\max_{1 \leq i \leq n} |\xi_{i+1} \sigma(X_i)|} \right) \right].$$

We need to study when such a term is less than  $c/n$  for some constant  $c$ . We bound  $\max_i |\xi_{i+1}\sigma(X_i)|$  by  $m_{\xi,\sigma}$  on the set  $\{\max_i |\xi_{i+1}\sigma(X_i)| \leq m_{\xi,\sigma}\}$  and the exponential by 1 on the set  $\{\max_i |\xi_{i+1}\sigma(X_i)| > m_{\xi,\sigma}\}$  and by denoting  $\mu_\epsilon = \kappa_3 \epsilon C(\epsilon^2)$ , this yields

$$\begin{aligned}
& \mathbb{E} \left[ \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \exp \left( -\mu_\epsilon \sqrt{\frac{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}{\max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i)}} \right) \right] \\
& \leq m_{\xi,\sigma}^2 \mathbb{E} \left( \exp \left( -\mu_\epsilon \frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}}{m_{\xi,\sigma}} \right) \right) + \mathbb{E} \left( \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \mathbf{1}_{\{\max_{1 \leq i \leq n} |\xi_{i+1}\sigma(X_i)| > m_{\xi,\sigma}\}} \right) \\
& \leq m_{\xi,\sigma}^2 \left[ \mathbb{E} \left( \exp \left( -\mu_\epsilon \sqrt{n\sigma_\tau^2 / (2m_{\xi,\sigma}^2)} \right) \right) + \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2 \right| \geq \sigma_\tau^2 / 2 \right) \right] \\
& \quad + m_{\xi,\sigma}^{-r} \mathbb{E} \left( \max_{1 \leq i \leq n} |\xi_{i+1}\sigma(X_i)|^{r+2} \right) \\
& \leq m_{\xi,\sigma}^2 e^{-\mu_\epsilon \sqrt{n}\sigma_\tau / (\sqrt{2}m_{\xi,\sigma})} + m_{\xi,\sigma}^2 2^p \sigma_\tau^{-2p} \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2 \right|^p \right) \\
& \quad + m_{\xi,\sigma}^{-r} \mathbb{E} \left( \max_{1 \leq i \leq n} |\xi_{i+1}\sigma(X_i)|^{r+2} \right).
\end{aligned}$$

Again by applying Rosenthal's inequality (see Lemma 9.1), we obtain that

$$\begin{aligned}
& \mathbb{E} \left[ \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \exp \left( -\mu_\epsilon \sqrt{\frac{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}{\max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i)}} \right) \right] \\
& \leq m_{\xi,\sigma}^2 e^{-\mu_\epsilon \sqrt{n}\sigma_\tau / (\sqrt{2}m_{\xi,\sigma})} + m_{\xi,\sigma}^2 \frac{C(p, \xi, \sigma(X))}{n^p} [n + n^{p/2}] + n \mathbb{E}(|\xi_2 \sigma(X_1)|^{r+2}) m_{\xi,\sigma}^{-r}
\end{aligned}$$

also bounded by

$$m_{\xi,\sigma}^2 e^{-\mu_\epsilon \sqrt{n}\sigma_\tau / (\sqrt{2}m_{\xi,\sigma})} + C'(p, \xi, \sigma(X)) m_{\xi,\sigma}^2 [n^{1-p} + n^{-p/2}] + n \mathbb{E}(|\xi_1|^{r+2}) \mathbb{E}(|\sigma(X_1)|^{r+2}) m_{\xi,\sigma}^{-r}.$$

Since  $\mathbb{E}|\xi_1|^8 < \infty$ , we take  $p = 3$ ,  $c = 4$  in Lemma 9.1,  $r = 4$ ,  $m_{\xi,\sigma} = \sigma_\tau \epsilon C(\epsilon^2) \kappa_3 \sqrt{n} / [2\sqrt{2} \ln(n)]$  and for any  $n \geq 3$ , and for  $C_1$  and  $C_2$  some constants depending on the moments of  $\xi$  and  $\sigma(X)$ , we find that

$$\mathbb{E} \left\{ \left( \max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i) \right) \exp \left( -\kappa_3 \epsilon C(\epsilon^2) \sqrt{\frac{\sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i)}{\max_{1 \leq i \leq n} \xi_{i+1}^2 \sigma^2(X_i)}} \right) \right\} \leq \frac{C_1}{\sqrt{n}} + C_2 \left( \frac{\ln^4(n)}{\sqrt{n}} \right) \frac{1}{\sqrt{n}}.$$

Then the sum over  $\mathcal{M}_n$  with cardinality less than  $\sqrt{n}$  of the terms in (26) is bounded by  $C(1 + \ln(n)^4 / \sqrt{n}) / n$  for some constant  $C$ , by using again that  $\Delta(m^*) / n$  is bounded.  $\square$

## 9. Appendix

As a reminder, some definitions and properties related to  $\beta$ -mixing sequences are given in this section. Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space. Let  $Y$  be a random variable with values in a Banach space  $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ , and let  $\mathcal{M}$  be a  $\sigma$ -algebra of  $\mathcal{A}$ . Let  $\mathbb{P}_{Y|\mathcal{M}}$  be a conditional distribution of  $Y$  given  $\mathcal{M}$ , and let  $P_Y$  be the distribution of  $Y$ . Let  $\mathcal{B}(\mathbb{B})$  be the borel  $\sigma$ -algebra on  $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ . Define now  $\beta(\mathcal{M}, \sigma(Y)) = \mathbb{E} \left( \sup_{A \in \mathcal{B}(\mathcal{X})} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)| \right)$  The coefficient  $\beta(\mathcal{M}, \sigma(Y))$  is the usual mixing

coefficient, introduced by Volkonskiĭ and Rozanov (1960). Let  $\mathbf{X} = (X_i)_{i \geq 1}$  be a strictly stationary sequence of real-valued random variables. For any  $k \geq 0$ , the coefficients  $\beta_{\mathbf{X},1}(k)$  are defined by  $\beta_{\mathbf{X},1}(k) = \beta(\sigma(X_1), \sigma(X_{1+k}))$ . Let  $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$ . The coefficients  $\beta_{\mathbf{X},\infty}(k)$  are defined by  $\beta_{\mathbf{X},\infty}(k) = \sup_{i \geq 1, l \geq 1} \sup \{ \beta(\mathcal{M}_i, \sigma(X_{i_1}, \dots, X_{i_l})), i+k \leq i_1 < \dots < i_l \}$ ,

In the paper, we do not distinguish between the two types of mixing and denote the coefficients of the process  $X$  by  $\beta(k)$  or  $\beta_X(k)$ . It is implicit that when only covariance inequality are involved, then the milder mixing  $\beta_{X,1}(k)$  is required, and we shall assume that stronger  $\beta_{X,\infty}(k)$  mixing coefficients are used in the general case.

Now, a Rosenthal-type inequality for mixing variables can be deduced from Doukhan (1994), Theorem 2 p.26 and the following result holds:

LEMMA 9.1. *Let  $(Y_k)_{1 \leq k \leq n}$  be a sequence of centered and stationary  $\beta$ -mixing variables with coefficients  $\beta(k)$ , admitting moments of order  $r+1$  and  $r \geq 2$ , then if*

$$\exists c \in 2\mathbb{N}, c \geq r, \text{ such that } \sum_{k \geq 1} (k+1)^{c-2} \beta(k)^{\frac{1}{c+1}} < +\infty,$$

*we have the bound*

$$\mathbb{E} \left| \frac{1}{n} \sum_{k=1}^n Y_k \right|^r \leq C(r) \left\{ n^{1-r} [\mathbb{E}|Y_1|^{r+1}]^{r/(r+1)} + n^{-r/2} [(\mathbb{E}|Y_1|^3)]^{r/3} \right\}. \quad (27)$$

We also use Delyon's (1990) covariance Inequality, successfully exploited by Viennet (1997) for partial sums of strictly stationary processes.

THEOREM 9.1. (Delyon (1990), Viennet (1997)) *Let  $P$  be the distribution of  $Z_0$  on a probability space  $\mathcal{X}$ ,  $\int f dP = \mathbb{E}_P(f)$  for any function  $f$   $P$ -integrable. For  $r \geq 2$ , let  $\mathcal{L}(r, \beta, P)$  be the set of functions  $b_Z : \mathcal{X} \rightarrow \mathbb{R}^+$  such that*

$$b_Z = \sum_{l \geq 0} (l+1)^{r-2} b_{l,Z} \text{ with } 0 \leq b_{l,Z} \leq 1 \text{ and } \mathbb{E}_P(b_{l,Z}) \leq \beta_Z(l)$$

*We define  $B_r$  as  $B_r = \sum_{l \geq 0} (l+1)^{r-2} \beta_Z(l)$ . Then for  $1 \leq p < \infty$  and any function  $b_Z$  in  $\mathcal{L}(2, \beta, P)$ ,  $\mathbb{E}_P(b_Z^p) \leq p B_{p+1}$ , as soon as  $B_{p+1} < \infty$ . The following result holds for a strictly stationary absolutely regular sequence,  $(Z_i)_{i \in \mathbb{Z}}$ , with  $\beta$ -mixing coefficients  $(\beta_Z(k))_{k \geq 0}$ : if  $B_2 < +\infty$ , there exists  $b_Z \in \mathcal{L}(2, \beta, \infty)$  such that for any positive integer  $n$  and any measurable function  $f \in \mathbb{L}_2(P)$ , we have*

$$\text{Var} \left( \sum_{i=1}^n f(Z_i) \right) \leq 4n \mathbb{E}_P(b_Z f^2) = 4n \int b_Z(x) f^2(x) dP(x).$$

Lastly, we recall the version of the Talagrand inequality that is required in the paper. Mention must be made that it is valid for independent but non necessarily identically distributed random variables, which is useful here when we work conditionally to one or two of the sequences.

LEMMA 9.2. *Let  $Y_1, \dots, Y_n$  be independent random variables, let  $\nu_{n,Y}(f) = (1/n) \sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$  and let  $\mathcal{F}$  be a countable class of uniformly bounded measurable functions. Then for  $\epsilon^2 > 0$*

$$\mathbb{E} \left[ \sup_{f \in \mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1 + 2\epsilon^2)H^2 \right]_+ \leq \frac{4}{K_1} \left( \frac{v}{n} e^{-K_1 \epsilon^2 \frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\epsilon^2)} e^{-\frac{2K_1 C(\epsilon^2) \epsilon}{7\sqrt{2}} \frac{nH}{M_1}} \right),$$

*with  $C(\epsilon^2) = \sqrt{1 + \epsilon^2} - 1$ ,  $K_1 = 1/6$ , and*

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \quad \mathbb{E} \left[ \sup_{f \in \mathcal{F}} |\nu_{n,Y}(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^n \text{Var}(f(Y_k)) \leq v.$$

This result follows from the concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

## References

- [1] Baraud, Y., Comte, F. and Viennet, G. (2001a). Adaptive estimation in an autoregressive and a geometrical  $\beta$ -mixing regression framework. *Annals of Statistics*, **39**, 839-875.
- [2] Barron, A.R., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Relat. Fields.* **97** 113-150.
- [3] Berbee, H.C.P. (1979). Random walks with Stationary Increments and Renewal Theory. Mathematical Centre Tracts 112. Amsterdam: Mathematisch 223p.
- [4] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
- [5] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11**, 309-340.
- [6] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32**, 181-192.
- [7] Butucea, C. and Tsybakov, A.B. (2007). Sharp optimality for density deconvolution with dominating bias. I. *Theor. Probab. Appl.* **52**, 1, to appear.
- [8] Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in hidden Markov models*. Springer series in Statistics. Springer, New-York.
- [9] Chanda, K. C. (1995). Large sample analysis of autoregressive moving average models with errors-in-variables. *J. Time Ser. Anal.*, **16**, 1, 1-15.
- [10] Comte, F., Dedecker, J. and Taupin, M.-L. (2006a). Adaptive density estimation for dependent inputs with measurement errors. Preprint MAP5 2006-4.
- [11] Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.* **6**, 211-238.
- [12] Comte, F. and Rozenholc, Y. (2002). Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.* **97**, 111-145.
- [13] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006b). Penalized contrast estimator for density deconvolution. *Canad. J. Statist.* **34**, 431-452.
- [14] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2007). Finite sample penalization in adaptive density deconvolution. To appear in *Journal of Statistical Computation and Simulation*. Working paper MAP5 2005-11.
- [15] Comte, F. and Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear errors-in-variables autoregressive model. *Mathematical Methods of Statistics* **10**, 121-160.
- [16] Comte, F. and Taupin, M.-L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statistica Sinica*, **17**, 3, to appear. Preprint MAP5 2004-20.
- [17] Delyon, B. (1990). Limit theorem for mixing processes. *Technical report IRISA Rennes*, **1**, 546.
- [18] van Es, B., Spreij, P. and van Zanten, H. (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.* **17**, 237-251.

- [19] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.
- [20] Fan, J. and Masry, M. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* **43**, 237-272.
- [21] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 4, 1900-1925.
- [22] Fan, J., Truong, Y. K. and Wang, Y. (1991). Nonparametric function estimation involving errors-in-variables. *Nonparametric Functional Estimation and Related Topics* 613-627.
- [23] Ghysels, E., Harvey, A. and Renault, E. (1996). Stochastic volatility. In: G.S. Maddala and C.R. Rao (Eds). *Handbook of Statistics, Statistical Methods in Finance*, Vol. 14. Amsterdam: North Holland, 119-191.
- [24] Ioannides, D.A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.* **32**, 35-43.
- [25] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.
- [26] Koo, J.-Y. and Lee K.-W. (1998). B-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* **40**, 57-66.
- [27] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877-882.
- [28] Lacour, C. (2005). Nonparametric estimation of the stationary density and the transition of a Markov chain. *Preprint MAP5 2005-8*, to appear in *Stochastic Process. Appl.*
- [29] Pensky, M. and Vidakovic, B (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** (6) 2033-2053.
- [30] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (Eds). *Time Series Models in Econometrics, Finance and Other Fields* London: Chapman & Hall, 1-67.
- [31] Taupin, M.L. (2001). Semi-parametric estimation in nonlinear structural errors-in-variables model. *Ann. Statist.* **29** (1) 66-93.
- [32] Tribouley, K. and Viennet, G. (1998).  $L_p$  adaptive density estimation in a  $\beta$  mixing framework. *Ann. Inst. H. Poincaré Probab. Statist.* **34**, 179-208.
- [33] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields* **107**, 467-492.
- [34] Volkonskiĭ, V. and Rozanov, Y. (1960). Some limit theorems for random functions. I. *Theor. Probab. Appl.* **4**, 178-197.