ADAPTIVE DENSITY ESTIMATION FOR GENERAL ARCH MODELS

F. COMTE*,[1], J. DEDECKER[2], AND M. L. TAUPIN [3]

ABSTRACT. We consider a model $Y_t = \sigma_t \eta_t$ in which (σ_t) is not independent of the noise process (η_t), but σ_t is independent of η_t for each t. We assume that (σ_t) is stationary and we propose an adaptive estimator of the density of $\ln(\sigma_t^2)$ based on the observations Y_t. Under various dependence structures, the rates of this nonparametric estimator coincide with the minimax rates obtained in the i.i.d. case when (σ_t) and (η_t) are independent, in all cases where these minimax rates are known. The results apply to various linear and non linear ARCH processes. They are illustrated by simulations and a real data application.

MSC 2000 Subject Classifications. 62G07 - 62M05. March 5, 2007
Keywords and phrases. Adaptive density estimation. Deconvolution. General ARCH models. Model selection. Penalized contrast.

1. INTRODUCTION

In this paper, we consider the following general ARCH-type model: $((Y_t, \sigma_t))_{t\in\mathbb{N}}$ is a strictly stationary sequence of $\mathbb{R} \times \mathbb{R}^+$-valued random variables, satisfying the equation

$$Y_t = \sigma_t \eta_t \tag{1.1}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and finite variance, and for each $t \geq 0$, the random vector $(\sigma_i, \eta_{i-1})_{0\leq i\leq t}$ is independent of the sequence $(\eta_i)_{i\geq t}$.

Such models are classically encountered in financial models, when $(\sigma_t^2)_{t\in\mathbb{Z}}$, the volatility process of interest, is unobserved. The only available data are the demeaned or detrended log-return process (Y_t) of an asset. A large variety of parametric models have been proposed since the first ARCH(1) model of Engle (1982), such as the GARCH(p, q) models of Bollerslev (1986), and other extensions to be found in Duan (1997). In general, it is not possible to compute, even in those parametric cases, the stationary density of σ_t. Here, we want to use flexible and powerful nonparametric tools to obtain information on the properties of the hidden process. More precisely, we shall build an estimator of the density

[1] Université Paris 5, MAP5, UMR CNRS 8145.
[2] Université Paris 6, Laboratoire de Statistique Théorique et Appliquée.
[3] IUT de Paris 5 et Université Paris Sud, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628.

of $\ln(\sigma_t^2)$ based on the data $(Y_t)_{1\leq t\leq n}$. This allows to look for its possible bimodality properties or to precise the localization of the peaks. Such a representation is of importance for modelization purpose.

Model (1.1) is classically re-written *via* a logarithmic transformation:

$$Z_t = X_t + \varepsilon_t, \tag{1.2}$$

where $Z_t = \ln(Y_t^2)$, $X_t = \ln(\sigma_t^2)$ and $\varepsilon_t = \ln(\eta_t^2)$. In the context derived from the model (1.1), X_t and ε_t are independent for a given t, whereas the processes $(X_t)_{t\geq 0}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are not independent.

Our aim is the adaptive estimation of g, the common distribution of the unobserved variables $X_t = \ln(\sigma_t^2)$, when the density f_ε of $\varepsilon_t = \ln(\eta_t^2)$ is known. More precisely we shall build an estimator of g without any prior knowledge on its smoothness, using the observations $Z_t = \ln(Y_t^2)$ and the knowledge of the convolution kernel f_ε. Since X_t and ε_t are independent for each t, the common density f_Z of the Z_t's is given by the convolution equation $f_Z = g * f_\varepsilon$, which justifies the term "deconvolution" for the estimation of g.

It is often assumed that the noise process (η_t) is Gaussian (e.g. in van Es *et al.* (2005)), but general distributions can be considered. This may be of interest if heavier or thinner tails are suspected to be relevant. Nevertheless, for identifiability of the statistical problem, the density of ε_t is required to be known. This assumption cannot be easily removed: even if the density of ε_t is completely known up to a scale parameter, Model (1.2) may be non-identifiable as soon as the unknown density g of X_t is smoother than the density of ε_t (see Butucea and Matias (2005), Section 1). For instance, the model in which g is an unknown normal distribution and ε_t is a Gaussian random variable with unknown variance is non-identifiable.

In density deconvolution of i.i.d variables the X_t's and the ε_t's are i.i.d. and the sequences $(X_t)_{t\geq 0}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent (for short we shall refer to this case as the i.i.d. case). In the i.i.d case, the slowest rates of convergence for estimating g are obtained for most regular error densities. For instance, when ε_t is Gaussian or the log of a squared Gaussian and g belongs to some Sobolev class, the minimax rates are negative powers of $\ln(n)$ (see Fan (1991)). Nevertheless, it has been noticed by several authors (see Pensky and Vidakovic (1999), Butucea (2004), Butucea and Tsybakov (2005), Comte *et al.* (2006)) that the rates are improved if g has stronger smoothness properties.

In the setting of Model (1.2), the classical assumptions of independence between the processes $(X_t)_{t\geq 0}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are no longer satisfied and the tools for deconvolution have to be revisited. Our estimator of g is constructed by minimizing an appropriate penalized contrast function only depending on the observations and on f_ε. It is chosen in a purely data-driven way among a collection of non-adaptive estimators. We start by the study of those non-adaptive estimators and show that their mean integrated squared error (MISE) has the same order as in the i.i.d. case. Next we prove that the MISE of our adaptive estimator is of the same order as the MISE of the best non-adaptive estimator in the collection, up to some possible negligible logarithmic loss in one case.

In their 2005 paper, van Es *et al.* (2005) have considered the case where η_t is Gaussian, the density g of X_t is twice differentiable, and the process (Z_t, X_t) is α-mixing. Here we show that, if g happens to be more regular (e.g. if g is a gaussian density, or the log of a squared gaussian exactly like the noise process ε_t), then their procedure is suboptimal and the bandwidth they propose is the not the best one. This is the reason why we do not make any assumption on the smoothness of g: this is the advantage of the adaptive procedure.

We also consider two types of dependence properties, which are satisfied by many ARCH processes. First we shall use the classical β-mixing properties of general ARCH models, as recalled in Doukhan (1994) and described in more details in Carrasco and Chen (2002). But we also illustrate that new recent coefficients can be used in our context, which allow an easy characterization of the dependence properties in function of the parameters of the models. Those new dependence coefficients, recently defined and studied in Dedecker and Prieur (2005), are interesting and powerful because they require much lighter conditions on the models. Such ideas have been popularized by Ango Nzé and Doukhan (2004) and Doukhan *et al.* (2006). For instance, these coefficients allow to deal with the general ARCH(∞) processes defined by Giraitis *et al.* (2000).

To be complete, we also give some simulation study: first in the i.i.d. case when ε_t is the log of the square of a standard gaussian, we study the effect of misspecification of the error distribution in that case. Next we present two simulation study for two distinct ARCH processes. This simulation step, which is useful to calibrate the numerical constant in the penalty term, is followed by an application to real data.

The paper is organized as follows. The estimator is defined in Section 2. The MISE bounds are given in Section 3 under some dependence properties. Many examples and simulation results are described in Section 4, together with an application to real data. All the proofs are given in Section 5.

2. The estimators

For two complex-valued functions u and v in $\mathbb{L}_2(\mathbb{R}) \cap \mathbb{L}_1(\mathbb{R})$, let $u^*(x) = \int e^{itx}u(t)dt$, $u * v(x) = \int u(y)v(x-y)dy$, and $\langle u, v\rangle = \int u(x)\overline{v}(x)dx$ with $\overline{z}$ the conjugate of a complex number z. We also denote by $\|u\|_1 = \int |u(x)|dx$, $\|u\|^2 = \int |u(x)|^2dx$, and $\|u\|_\infty = \sup_{x\in\mathbb{R}} |u(x)|$.

In the sequel, we consider Model (1.1) and its equivalent representation (1.2) and we describe the estimation strategy for g the density of the X_i's based on observations Z_i, $i = 1, \dots, n$ and on the knowledge of f_ε, the density of the ε_i's.

2.1. **Construction of the minimum contrast estimators.** First, the heuristics for the construction of the estimators are the following. Mean-square type contrasts are determined by following the idea that we look for a function t such that its $\mathbb{L}_2$-distance to the function g is minimum. Since $\|t-g\|^2 = \|t\|^2 - 2\langle t, g\rangle + \|g\|^2$, this amounts to minimize $\|t\|^2 - 2\langle t, g\rangle$, in a well chosen class of functions t. Now, if f_Z is the common density of the Z_i's, then $f_Z = f_\varepsilon * g$, so that $f_Z^* = f_\varepsilon^* g^*$ and $g^* = f_Z^*/f_\varepsilon^*$, where f_ε^* is assumed to be known and $f_Z^*(x) = \mathbb{E}(e^{ixZ})$ has an empirical counterpart. We can now define the contrast function

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^n \left[\|t\|^2 - 2u_t^*(Z_i)\right], \quad \text{with} \quad u_t(x) = \frac{1}{2\pi}\left(\frac{t^*(-x)}{f_\varepsilon^*(x)}\right),$$

under the assumption

$$f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \text{ and is such that } f_\varepsilon^*(x) \neq 0 \text{ for any } x \text{ in } \mathbb{R}. \tag{2.1}$$

We have $\mathbb{E}\left[u_t^*(Z_i)\right] = (2\pi)^{-1}\langle f_Z^*/f_\varepsilon^*, t^*\rangle = \langle t, g\rangle$, where the last equality follows from the Parseval identity. It follows that $\mathbb{E}(\gamma_n(t)) = \|t-g\|^2 - \|g\|^2$ is minimal when $t = g$.

Clearly the functions t must be chosen such that $u_t^*(x)$ is well defined. Since $1/f_\varepsilon^*$ can be non integrable (think of a Gaussian density), a solution is to choose the functions t such that t^* exists and has compact support.

The most simple spaces that suit to that aim are often studied in preliminary wavelet courses (see e.g. Meyer (1990), p.22) and are the following. Let $\varphi(x) = \sin(\pi x)/(\pi x)$. The Fourier transform of φ is obtained by noticing that the Fourier transform of $\mathbf{I}_{[-\pi,\pi]}(x)/2\pi$ is equal to φ and by using the inverse Fourier formula, and thus $\varphi^*(x) = \mathbf{I}_{[-\pi,\pi]}(x)$. For $m \in \mathbb{N}$ and $j \in \mathbb{Z}$, set $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$. The functions $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$: indeed $\varphi_{m,j}^*(x) = e^{ijx/m}\varphi^*(x/m)/\sqrt{m}$ and $\langle \varphi_{m,j}, \varphi_{m,k}\rangle = (2\pi)^{-1}\langle \varphi_{m,j}^*, \varphi_{m,k}^*\rangle = \delta_{j,k}$. Therefore, if we define

$$S_m = \overline{\text{span}}\{\varphi_{m,j},\ j \in \mathbb{Z}\}, m \in \mathbb{N},$$

the space S_m is exactly the subspace of $\mathbb{L}_2(\mathbb{R})$ of functions having a Fourier transform with compact support contained in $[-\pi m, \pi m]$. The orthogonal projection of g on S_m is $g_m = \sum_{j\in\mathbb{Z}} a_{m,j}(g)\varphi_{m,j}$ where $a_{m,j}(g) =< \varphi_{m,j}, g >$. Now, we can not describe practical algorithms involving infinite representations. Thus, to obtain representations having a finite number of "coordinates", we introduce

$$S_m^{(n)} = \overline{\text{span}}\left\{\varphi_{m,j}, |j| \leq n\right\}.$$

The family $\{\varphi_{m,j}\}_{|j|\leq n}$ is an orthonormal basis of $S_m^{(n)}$ and the orthogonal projection of g on $S_m^{(n)}$ is given by $g_m^{(n)} = \sum_{|j|\leq n} a_{m,j}(g)\varphi_{m,j}$. Subsequently a space $S_m^{(n)}$ will be referred to as a "model" as well as a "projection space".

Then, for an arbitrary fixed integer m, an estimator of g belonging to $S_m^{(n)}$ is defined by

$$\hat{g}_m^{(n)} = \arg \min_{t \in S_m^{(n)}} \gamma_n(t). \tag{2.2}$$

Moreover, it is easy to see that

$$\hat{g}_m^{(n)} = \sum_{|j| \leq n} \hat{a}_{m,j} \varphi_{m,j} \text{ with } \hat{a}_{m,j} = \frac{1}{n} \sum_{i=1}^{n} u^*_{\varphi_{m,j}}(Z_i), \text{ and } \mathbb{E}(\hat{a}_{m,j}) =< g, \varphi_{m,j} >= a_{m,j}(g).$$

2.2. **Minimum penalized contrast estimator.** For the sake of completeness in the description of our estimation strategy, let us give also the last step. We show in the following that a "good" choice of m exists, but results from a squared bias - variance compromise in which the bias is unknown. The end of the procedure is the analogue of a bandwidth selection strategy in kernel estimation and gives a criterion to select m.

The minimum penalized estimator of g is defined as $\tilde{g} = \hat{g}_{\hat{m}_g}^{(n)}$ where $\hat{m}_g$ is chosen in a purely data-driven way. The main point of the estimation procedure lies in the choice of $m = \hat{m}$ (or equivalently in the choice of a model $S_{\hat{m}}^{(n)}$) involved in the estimators $\hat{g}_m^{(n)}$ given by (2.2), in order to mimic the oracle parameter

$$\breve{m}_g = \arg \min_m \mathbb{E} \parallel \hat{g}_m^{(n)} - g \parallel^2 . \tag{2.3}$$

The model selection is performed in an automatic way, using the following penalized criteria

$$\tilde{g} = \hat{g}_{\hat{m}}^{(n)} \text{ with } \hat{m} = \arg \min_{m \in \{1, \cdots, m_n\}} \left[\gamma_n(\hat{g}_m^{(n)}) + \text{ pen}(m) \right], \tag{2.4}$$

where $\text{pen}(m)$ is a penalty function that depends on $f_\varepsilon^*(\cdot)$ through $\Delta(m)$ defined by

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|^2} dx. \tag{2.5}$$

The key point in the dependent context is to find a penalty function not depending on the dependence coefficients such that

$$\mathbb{E} \parallel \tilde{g} - g \parallel^2 \leq C \inf_{m \in \{1, \cdots, m_n\}} \mathbb{E} \parallel \hat{g}_m^{(n)} - g \parallel^2 .$$

In that way, the estimator $\tilde{g}$ is adaptive since it achieves the best rate among the estimators $\hat{g}_m^{(n)}$, without any prior knowledge on the smoothness on g.

3. Density estimation bounds

3.1. **Mixing assumptions.** Clearly, the process of interest after the logarithmic transformation ($Z_t = \ln(Y_t^2)$, $X_t = \ln(\sigma_t^2)$) is

$$(W_t)_{t \in \mathbb{Z}} = ((Z_t, X_t))_{t \in \mathbb{Z}} \tag{3.1}$$

and involves dependent variables. We want to use two types of representation of this dependence: first the classical β-mixing properties, and then a new τ-dependence notion which reveals useful for a large class of models. Let us give now the definition of the corresponding coefficients.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let W be a random vector with values in a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, and let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let $\mathbb{P}_{W|\mathcal{M}}$ be a conditional distribution of W given $\mathcal{M}$, and let $\mathbb{P}_W$ be the distribution of W. Let $\mathcal{B}(\mathbb{B})$ be the Borel σ-algebra on $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, and let $\Lambda_1(\mathbb{B})$ be the set of 1-Lipschitz functions from $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ to $\mathbb{R}$. Define now $\beta(\mathcal{M}, \sigma(W)) = \mathbb{E}\Big(\sup_{A\in\mathcal{B}(\mathbb{B})} |\mathbb{P}_{W|\mathcal{M}}(A) - \mathbb{P}_W(A)|\Big)$ and if $\mathbb{E}(\|W\|_{\mathbb{B}}) < \infty$, $\tau(\mathcal{M}, W) = \mathbb{E}\Big(\sup_{f\in\Lambda_1(\mathbb{B})} |\mathbb{P}_{W|\mathcal{M}}(f) - \mathbb{P}_W(f)|\Big)$. The coefficient $\beta(\mathcal{M}, \sigma(W))$ is the usual mixing coefficient, introduced by Rozanov and Volkonskii (1960). The coefficient $\tau(\mathcal{M}, W)$ has been introduced by Dedecker and Prieur (2005).

Let $(W_t)_{t\geq 0}$ be a strictly stationary sequence of $\mathbb{R}^2$-valued random variables. On $\mathbb{R}^2$, we put the norm $\|x-y\|_{\mathbb{R}^2} = |x_1 - y_1| + |x_2 - y_2|$. For any $k \geq 0$, define the coefficients

$$\beta_1(k) = \beta(\sigma(W_0), \sigma(W_k)), \quad \text{and if } \mathbb{E}(\|W_0\|_{\mathbb{R}^2}) < \infty, \quad \tau_1(k) = \tau(\sigma(W_0), W_k). \tag{3.2}$$

On $(\mathbb{R}^2)^l$, we put the norm $\|x-y\|_{(\mathbb{R}^2)^l} = l^{-1}(\|x_1 - y_1\|_{\mathbb{R}^2} + \cdots + \|x_l - y_l\|_{\mathbb{R}^2})$. Let $\mathcal{M}_i = \sigma(W_k, 0 \leq k \leq i)$. The coefficients $\beta_\infty(k)$ and $\tau_\infty(k)$ are defined by

$$\beta_\infty(k) = \sup_{i\geq 0}\sup_{l\geq 1} \{\beta(\mathcal{M}_i, \sigma(W_{i_1}, \ldots, W_{i_l})), i + k \leq i_1 < \cdots < i_l\}, \tag{3.3}$$

and if $\mathbb{E}(\|W_1\|_{\mathbb{R}^2}) < \infty$,

$$\tau_\infty(k) = \sup_{i\geq 0}\sup_{l\geq 1} \{\tau(\mathcal{M}_i, (W_{i_1}, \ldots, W_{i_l})), i + k \leq i_1 < \cdots < i_l\}. \tag{3.4}$$

We say that the process $(W_t)_{t\geq 0}$ is β-mixing (resp. τ-dependent) if the coefficients $\beta_\infty(k)$ (resp. $\tau_\infty(k)$) tend to zero as k tends to infinity. We say that it is geometrically β-mixing (resp. geometrically τ-dependent), if there exist $a < 1$ and $C > 0$ such that $\beta_\infty(k) \leq Ca^k$ (resp. $\tau_\infty(k) \leq Ca^k$) for all $k \geq 1$.

From now on, the dependence coefficients are defined as in (3.2), (3.3) and (3.4) with $(W_t)_{t\in\mathbb{Z}} = ((Z_t, X_t))_{t\in\mathbb{Z}}$. Moreover, we summarize the dependency assumptions for Model (1.2):

- The ε_i's are i.i.d.,
- The random vector $(X_i, \varepsilon_{i-1})_{0\leq i\leq t}$ is independent of the sequence $(\varepsilon_i)_{i\geq t}$,
- The process $(W_t)_{t\in\mathbb{Z}}$ is strictly stationary and either β-mixing or τ-dependent.

3.2. **Risk bound of the minimum contrast estimators $\hat{g}_m^{(n)}$.** Subsequently, the density g is assumed to satisfy the following assumption:

$$g \in \mathbb{L}_2(\mathbb{R}), \text{and there exists } M_2 > 0, \int x^2 g^2(x)dx \leq M_2 < \infty. \tag{3.5}$$

For instance, (3.5) is fulfilled if g is bounded by M_0 and $\mathbb{E}(X_1^2) \leq M_1 < +\infty$, with $M_2 = M_0 M_1$. Assumption (3.5) is due to the construction of the estimator on $S_m^{(n)}$ instead of S_m, and is not very restrictive.

The order of the MISE of $\hat{g}_m^{(n)}$ is given in the following proposition.

Proposition 3.1. *If (2.1) and (3.5) hold, then* $\hat{g}_m^{(n)}$ *defined by (2.2) satisfies*

$$\mathbb{E}\|g-\hat{g}_m^{(n)}\|^2 \le \|g-g_m\|^2 + \frac{m^2(M_2+1)}{n} + \frac{2\Delta(m)}{n} + \frac{2R_m}{n},$$

where $\Delta(m)$ *is defined by (2.5) and*

$$R_m = \frac{1}{\pi}\sum_{k=2}^{n}\int_{-\pi m}^{\pi m}\Big|\frac{\mathrm{Cov}\left(e^{ixZ_1}, e^{ixX_k}\right)}{f_\varepsilon^*(-x)}\Big|dx. \tag{3.6}$$

Moreover, $R_m \le \min(R_{m,\beta}, R_{m,\tau})$*, where*

$$R_{m,\beta} = 4\Delta_{1/2}(m)\sum_{k=1}^{n-1}\beta_1(k) \quad \text{and} \quad R_{m,\tau} = \pi m\Delta_{1/2}(m)\sum_{k=1}^{n-1}\tau_1(k)\,,$$

with β_1*,* τ_1 *defined by (3.2), and where*

$$\Delta_{1/2}(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{1}{|f_\varepsilon^*(x)|}dx. \tag{3.7}$$

This proposition requires several comments.

The order of the risk is given by a bias term $\| g_m - g \|^2 + m^2(M_2+1)/n$ and a variance term $2\Delta(m)/n + 2R_m/n$.

The main part of the bias term is $\|g-g_m\|^2$. It is noteworthy that

$$\|g-g_m\|^2 = \int_{|x|\ge\pi m}|g^*(x)|^2dx, \tag{3.8}$$

so that the order of the bias depends on the rate of decay of $g^*(x)$. The additional term $m^2(M_2+1)/n$ is negligible with respect to the variance term.

The variance term $2\Delta(m)/n + 2R_m/n$ depends on the rate of decay of the Fourier transform of f_ε. It is the sum of the variance term appearing in density deconvolution for i.i.d. variables $2\Delta(m)/n$ and of an additional term $2R_m/n$. This last term R_m involves the dependency coefficients and the quantity $\Delta_{1/2}(m)$, which is specific to the ARCH problem. The point is that the main order term in the variance part is $\Delta(m)/n$, which does not involve the dependency coefficients. In other words, the dependency coefficients only appear in front of the additional and negligible term, specific to ARCH models, $\Delta_{1/2}(m)/n$.

3.3. **Discussion about the rates.** Now, we need to parameterize the rate of decay of f_ε^* and g^* to evaluate the order of the variance term $\Delta(m)/n$ and of the main part of the bias term given by (3.8).

More precisely, we assume that f_ε is such that: there exist nonnegative numbers κ_0, γ, μ, and δ such that the fourier transform f_ε^* of f_ε satisfies

$$\kappa_0(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\} \le |f_\varepsilon^*(x)| \le \kappa_0'(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\}. \tag{3.9}$$

Since f_ε is known, the constants μ, δ, κ_0, and γ defined in (3.9) are known. The class described by (3.9) is very general. When $\delta = 0$ in (3.9), the errors are called "ordinary smooth" errors. When $\mu > 0$ and $\delta > 0$, they are called "super smooth". An example of ordinary smooth density is the Laplace distribution, for which $f_\varepsilon^*(x) = 1/(1+x^2)$ so that $\delta = \mu = 0$ and $\gamma = 2$. The standard examples for super smooth densities are Gaussian ($f_\varepsilon^*(x) = e^{-x^2/2}$ so that $\gamma = 0$, $\delta = 2$, $\mu = 1/2$) or Cauchy ($f_\varepsilon^*(x) = e^{-|x|}$, so that $\gamma = 0, \delta = 1$, $\mu = 1$) distributions. When $\varepsilon_t = \ln(\eta_t^2)$ with $\eta_t \sim \mathcal{N}(0,1)$ as in van Es *et al.* (2005), then $f_\varepsilon^*(x) = 2^{ix}\Gamma(1+ix)/\sqrt{\pi}$ and with Stirling formula $|f_\varepsilon^*(x)| \sim_{+\infty} \sqrt{2/e}e^{-\pi|x|/2}$. Then ε_t is super-smooth with $\delta = 1, \gamma = 0$ and $\mu = \pi/2$. Note that, in that case, $\mathbb{E}(\varepsilon_1) = -\ln(2) - \gamma$ where γ is the Euler constant and $\mathrm{Var}(\varepsilon_1) = \pi^2/2$.

Moreover, the square integrability of f_ε and (3.9) require that $\gamma > 1/2$ when $\delta = 0$.

Now, concerning the main variance term, if f_ε^* satisfies (3.9), then $\Delta(m)$ given by (2.5) has the same order as

$$\Gamma(m) = (1 + (\pi m)^2)^\gamma (\pi m)^{1-\delta} \exp\left\{2\mu(\pi m)^\delta\right\},$$

up to some constant bounded by

$$\lambda_1(f_\varepsilon, \kappa_0) = \frac{1}{\kappa_0^2 \pi R(\mu,\delta)}, \text{ where } R(\mu,\delta) = \mathbb{1}_{\{\delta=0\}} + 2\mu\delta\mathbb{1}_{\{\delta>0\}}. \tag{3.10}$$

Also, we describe the smoothness properties of g by the set

$$\mathcal{S}_{s,r,b}(C_1) = \left\{\psi \text{ such that } \int_{-\infty}^{+\infty} |\psi^*(x)|^2 (x^2+1)^s \exp\{2b|x|^r\} dx \le C_1\right\} \tag{3.11}$$

for s, r, b unknown non negative numbers. When $r = 0$, the class $\mathcal{S}_{s,r,b}(C_1)$ corresponds to a Sobolev ball. Heuristically, a function ψ admitting k continuous derivatives is such that $|\psi^*(x)| = O(|x|^{-k})$ when x tends to infinity and thus belongs to $\mathcal{S}_{s,0,0}(C_1)$ for $s < k - 1/2$. When $r > 0, b > 0$ functions belonging to $\mathcal{S}_{s,r,b}(C_1)$ are infinitely many times differentiable. For instance $\psi(x) = e^{-x^2/2}$ belongs to $\mathcal{S}_{s,b,2}(C_1)$ for any $b < 1/2$ and any $s \ge 0$.

Now, concerning the main bias term, it follows from (3.8) that if g belongs to a space $\mathcal{S}_{s,r,b}(C_1)$, then

$$\|g - g_m\|^2 \le \frac{C}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\},$$

where C is a constant depending on C_1.

As a consequence, the rates resulting from Proposition 3.1 under (3.9) and (3.11) are deduced from the following proposition.

Corollary 3.1. *Assume that (3.9), (2.1), and (3.5) hold, that g belongs to $\mathcal{S}_{s,r,b}(C_1)$ defined by (3.11). Assume either that*

(1) $\sum_{k\ge1}\beta_1(k) < +\infty$
(2) *or* $\delta = 0$, $\gamma > 1$ *in (3.9) and* $\sum_{k\ge1}\tau_1(k) < +\infty$
(3) *or* $\delta > 0$ *in (3.9) and* $\sum_{k\ge1}\tau_1(k) < +\infty$.

Then $\hat g_m^{(n)}$ defined by (2.2) satisfies

$$\mathbb{E}\|g - \hat g_m^{(n)}\|^2 \le \frac{C_1}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\} + \frac{2\lambda_1(f_\varepsilon,\kappa_0)\Gamma(m)}{n} + \frac{C_2}{n}\Gamma(m)o_m(1), \tag{3.12}$$

where C_1 and C_2 are finite constants. The constant C_2 depends on $\sum_{k\geq 1}\beta_1(k)$ (respectively on $\sum_{k\geq 1}\tau_1(k)$).

The rate of convergence of $\hat{g}^{(n)}_{\breve{m}}$ is the same as the rate for density deconvolution for i.i.d. sequences. Moreover, our context encompasses the particular case considered by van Es *et al.* (2005). Table 1 gives a summary of these rates obtained when minimizing the right hand of (3.12). The $\breve{m}_g$ denotes the corresponding minimizer (see 2.3).

TABLE 1. Choice of $\breve{m}_g$ and corresponding rates under Assumptions (3.9) and (3.11).

		f_ε: $\delta = 0$ ordinary smooth	f_ε: $\delta > 0$ supersmooth
g	$r = 0$ Sobolev(s)	$\pi\breve{m}_g = O(n^{1/(2s+2\gamma+1)})$ rate $= O(n^{-2s/(2s+2\gamma+1)})$	$\pi\breve{m}_g = [\ln(n)/(2\mu+1)]^{1/\delta}$ rate $= O((\ln(n))^{-2s/\delta})$
g	$r > 0$ $\mathcal{C}^\infty$	$\pi\breve{m}_g = [\ln(n)/2b]^{1/r}$ rate $= O\left(\dfrac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$	$\breve{m}_g$ solution of $\breve{m}_g^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m}_g)^\delta + 2b\pi^r\breve{m}_g^r\}$ $= O(n)$

If g belongs to a classical Sobolev class ($r = 0$), the rates range from negative powers of n when the errors are ordinary smooth, to negative powers of $\ln(n)$ if the errors are super-smooth. But if the function to estimate becomes super-smooth also (case $r > 0$), the rates become much better in both cases. When $r > 0, \delta > 0$ the value of $\breve{m}_g$ is not explicitly given. It is obtained as the solution of the equation

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

Consequently, the rate of $\hat{g}^{(n)}_{\breve{m}_g}$ is not easy to give explicitly and depends on the ratio r/δ. We refer to Comte *et al.* (2006) for further discussions about those rates. We refer to Lacour (2006b) for explicit formulae for the rates in the special case $r > 0$ and $\delta > 0$.

Example. Consider the case of van Es *et al* (2005)'s paper, where f_ε^* satisfies (3.9) with $\delta = 1$, $\mu = \pi/2$ and $\gamma = 0$. They consider that the regularity of g is such that g belongs to $\mathcal{S}_{s,r,b}(C_1)$ with s at most $3/2$ and $r = 0$. This is the up-right case of Table 1 and leads to a logarithmic rate. Now, if for instance $g^*(x)$ is of order $e^{-\pi|x|/2}$ like the noise, then the squared bias-variance compromise $C_1e^{-\pi m} + C_2e^{\pi m}/n$ leads to choose $\breve{m} = \ln(n)/(2\pi)$ and the rate is of order $1/\sqrt{n}$, which is better than logarithmic. If g is Gaussian, then the squared bias - variance compromise $C_1m^{-1}e^{-\pi^2m^2/2} + C_2e^{\pi m}/n$ leads to choose $\breve{m} = 2\sqrt{\ln(n)}/\pi$ and the rate if of order $e^{2\sqrt{\ln(n)}}/n = o(1/n^{1-\epsilon})$ for any $\epsilon > 0$, which is even better than the previous one. This example is another way to see the interest of the adaptive procedure, which enables to face any regularity for g, without requiring to know it.

3.4. **Adaptive bound.** Theorem 3.1 below gives a general bound which holds under weak dependence conditions, for ε being either ordinary or super smooth.

For $a > 1$, let $\text{pen}(m)$ be defined by

$$\text{pen}(m) = \begin{cases} 48a\dfrac{\Delta(m)}{n} & \text{if } 0 \le \delta < 1/3, \\ 16a\lambda_3\dfrac{\Delta(m)\, m^{\min((3\delta/2-1/2)_+,\delta))}}{n} & \text{if } \delta \ge 1/3, \end{cases} \tag{3.13}$$

where $\Delta(m)$ is defined by (2.5). The constant $\lambda_1(f_\varepsilon, \kappa_0)$ is defined in (3.10) and

$$\lambda_3 = 1 + \frac{192\mu\pi^\delta}{\lambda_1(f_\varepsilon, \kappa_0')}\Big((\sqrt{2}+8)\|f_\varepsilon\|_\infty \kappa_0^{-1}\sqrt{\lambda_1(f_\varepsilon,\kappa_0)}\,\mathbb{I}_{0\le\delta\le 1} + 2\lambda_1(f_\varepsilon,\kappa_0)\mathbb{I}_{\delta>1}\Big). \tag{3.14}$$

The important point here is that λ_3 is known. Hence the penalty is explicit up to a numerical multiplicative constant. This procedure has already been practically studied for independent sequences $(X_t)_{t\ge1}$ and $(\varepsilon_t)_{t\ge1}$ in Comte *et al.* (2005, 2006). In particular, the practical implementation of the penalty functions, and the calibration of the constants have been studied in the two previously mentioned papers. Moreover, it is shown therein that the estimation procedure is robust to various types of dependence, whether the errors ε_i's are ordinary or super smooth (see Tables 4 and 5 in Comte *et al.* (2005)).

In order to bound up $\text{pen}(m)$, we impose that

$$\pi m_n \le \begin{cases} n^{1/(2\gamma+1)} & \text{if } \delta = 0 \\ \left[\dfrac{\ln(n)}{2\mu} + \dfrac{2\gamma+1-\delta}{2\delta\mu}\ln\left(\dfrac{\ln(n)}{2\mu}\right)\right]^{1/\delta} & \text{if } \delta > 0. \end{cases} \tag{3.15}$$

Subsequently we set

$$C_a = \max(\kappa_a^2, 2\kappa_a) \text{ where } \kappa_a = (a+1)/(a-1). \tag{3.16}$$

Theorem 3.1. *Assume that f_ε satisfies (3.9) and 2.1, that g satisfies (3.5), and that m_n satisfies (3.15). Let* $\text{pen}(m)$ *be defined by (3.13). Consider the collection of estimators $\hat{g}_m^{(n)}$ defined by (2.2) with $1 \le m \le m_n$. Let β_∞ and τ_∞ be defined as in (3.3) and (3.4) respectively. Assume either that*

1. *$\beta_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3$*
2. *or $\delta = 0$, $\gamma \ge 3/2$ in (3.9) and $\tau_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3 + 2/(1+2\gamma)$*
3. *or $\delta > 0$ in (3.9) and $\tau_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3$.*

Then the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by (2.4) satisfies

$$\mathbb{E}(\|g - \tilde{g}\|^2) \le C_a \inf_{m\in\{1,\cdots,m_n\}}\left[\|g - g_m\|^2 + \text{pen}(m) + \frac{m^2(M_2+1)}{n}\right] + \frac{\overline{C}}{n}, \tag{3.17}$$

where C_a *is defined in (3.16) and* $\overline{C}$ *is a constant depending on* f_ε*,* a*, and* $\sum_{k\geq 1}\beta_\infty(k)$ *(respectively on* $\sum_{k\geq 1}\tau_\infty(k)$*).*

The estimator $\tilde{g}$ is adaptive in the sense that it is purely data-driven. This is due to the fact that pen(.) is explicitly known. In particular, its construction does not require any prior smoothness knowledge on the unknown density g and does not use the dependency coefficients. This point is important since all quantities involving dependency coefficients are usually not tractable in practice.

Moreover, one can compare the order of the penalty to the variance order $\Delta(m)/n$ in the light of Table 1. They are equal for $0 \leq \delta \leq 1/3$: this means that, asymptotically, the right-hand side of (3.17) realizes an automatic squared-bias variance compromise. For $\delta > 1/3$, the order penalty has the order of the variance increased by a power of m, but this does not change the choice of $\pi\breve{m}_g$ if $r = 0$ and thus the rate is unchanged then; for $r > 0$, the optimal choice of $\breve{m}$ in the case $\delta > 0$ is of order $\ln(n)$ and the rate is always faster than logarithmic. Therefore, if a loss occurs, it is negligible, when compared to the rate.

To summarize, the main result in Theorem 3.1 shows that the MISE of $\tilde{g}$ automatically achieves the best squared-bias variance compromise (possibly up to some logarithmic factor) among the estimators $\hat{g}_m^{(n)}$. Consequently, it achieves the best rate among the rates of the $\hat{g}_m^{(n)}$, even from a non-asymptotical point of view. This non-asymptotic feature is important since the m's selected in practice are small and far away from asymptotic. For practical illustration of this point in the case of density deconvolution of i.i.d. variables, we refer to Comte *et al.* (2005, 2006).

As a conclusion, the estimator $\tilde{g}$ has the same rate as in the i.i.d. case, with an explicit penalty function not depending on the dependence coefficients.

3.5. **Further comments.** We first show how the density f of σ_t^2 can be estimated. Since for $u > 0$, $f(u) = g(\ln(u))/u$, we choose the estimator

$$\hat{f}(u) = \frac{\tilde{g}(\ln(u))}{u}, \quad \text{for } u > 0.$$

A change of variables gives the equality

$$\mathbb{E}\Big(\int_0^\infty |\hat{f}(t) - f(t)|^2 t dt\Big) = \mathbb{E}(\|\tilde{g} - g\|^2)\,.$$

Note that the term on left hand is a MISE for $\hat{f}$ with respect to the measure tdt. In particular, it follows that for any $a > 0$,

$$\mathbb{E}\Big(\int_a^\infty |\hat{f}(t) - f(t)|^2 dt\Big) \leq \frac{1}{a}\mathbb{E}(\|\tilde{g} - g\|^2)\,,$$

which shows that the usual MISE for $\hat{f}$ on any interval $[a, \infty[$ tends to 0 at least with the same rate as $\mathbb{E}(\|\tilde{g} - g\|^2)$.

Secondly, we can mention that in the context where the process X is mixing, but globally independent of the noise process (ε_t), it is possible to build in the same way as here an estimator of the joint distribution of $(X_t, X_{t-1})_t$, see Lacour (2006a). It is likely that the multivariate procedure would work in our context as well, but the technical price sets this question beyond the scope of the present paper.

Lastly, it is possible to deal with dependent errors, but again if the processes $(X_t)_t$ and $(\varepsilon_t)_t$ are independent. This context is not developed because then, the presence of the unknown mixing coefficients in the variance bound and then in the penalty function is unavoidable. The procedure is then not theoretically satisfactory.

4. Example of β and τ-mixing processes and illustration of the method

4.1. **Examples of mixing models.** In this section, we give conditions under which classical ARCH models satisfy the Assumptions of Corollary 3.1 or Theorem 3.1.

A particular case of model (1.1) is

$$Y_t = \sigma_t \eta_t, \text{ with } \sigma_t = f(\eta_{t-1}, \eta_{t-2}, \dots) \tag{4.1}$$

for some measurable function f. Another important case is

$$Y_t = \sigma_t \eta_t, \text{ with } \sigma_t = f(\sigma_{t-1}, \eta_{t-1}) \text{ and } \sigma_0 \text{ independent of } (\eta_t)_{t\geq 0}, \tag{4.2}$$

that is σ_t is a stationary Markov chain. For the sake of simplicity, we shall always assume here that $\mathbb{E}(\eta_0^2) = 1$.

We begin with models satisfying a recursive equation, whose stationary solution satisfies (4.1). The original ARCH model was introduced by Engle (1982) and generalized by Bollerslev (1986) with the class of GARCH(p, q) models defined by $Y_t = \sigma_t \eta_t$ and

$$\sigma_t^2 = a + \sum_{i=1}^{p} a_i Y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \tag{4.3}$$

where the coefficients $a, a_i, i = 1, \dots, p$ and $b_j, j = 1, \dots, q$ are all positive real numbers. Those processes were studied from the point of view of existence and stationarity of solutions by Bougerol and Picard (1992a, 1992b) and Ango Nzé (1992). Under the condition $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1$, this model has a unique stationary solution of the form (4.1).

Many extensions have been proposed since then. A general linear example of model is given by the ARCH(∞) model described by Giraitis *et al.* (2000):

$$\sigma_t^2 = a + \sum_{j=1}^{\infty} a_j Y_{t-j}^2, \tag{4.4}$$

where $a \geq 0$ and $a_j \geq 0$. Again if $\sum_{j\geq 1} a_j < 1$, then there exists a unique strictly stationary solution to (4.4) of the form (4.1).

For the models satisfying (4.2), let us cite first the so-called augmented GARCH$(1,1)$ models introduced by Duan (1997):

$$\Lambda(\sigma_t^2) = c(\eta_{t-1})\Lambda(\sigma_{t-1}^2) + h(\eta_{t-1}), \tag{4.5}$$

where Λ is an increasing and continuous function on $\mathbb{R}^+$. We refer to Duan (1997) for numerous examples of more standard models belonging to this class. There exists a stationary solution to (4.5), provided c satisfies the condition A_2^* given in Carrasco and Chen (2002) (this condition is satisfied as soon as $\mathbb{E}(|c(\eta_0)|^s) < 1$ and $\mathbb{E}(|h(\eta_0)|^s) < \infty$ for integer $s \geq 1$, see the condition A_2 of the same paper). An example of the model (4.5) is the threshold ARCH model (see Zakoïan (1993)):

$$\sigma_t = a + b\sigma_{t-1}\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}>0\}} - c\sigma_{t-1}\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}<0\}},\ a, b, c > 0 \tag{4.6}$$

for which $c(\eta_{t-1}) = b\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}>0\}} - c\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}<0\}}$ and $h = a$. In particular, the condition for the stationarity is satisfied as soon as $b \vee c < 1$.

Other models satisfying (4.2) are the non linear ARCH models (see Doukhan (1994), p. 106-107), for which:

$$\sigma_t = f(\sigma_{t-1}\eta_{t-1}). \tag{4.7}$$

There exists a stationary solution to (4.7) provided that the density of η_0 is positive on a neighborhood of 0 and $\limsup_{|x|\to\infty} |f(x)/x| < 1$.

Now, we give the properties of the models (4.3)-(4.7) in terms of the dependence coefficients.

First, concerning the β-mixing, the following results can be deduced from the literature. If we assume that in all cases the η_t's are centered with unit variance and admit a density with respect to the Lebesgue measure, then

- The process $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ defined by Model (4.3) is geometrically β-mixing, as soon as $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$ (see Carrasco and Chen (2000, 2002)).
- The process $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ defined by Model (4.5) is geometrically β-mixing as soon as: the density of η_0 is positive on an open set containing 0; c and h are polynomial functions; there exists an integer $s \geq 1$ such that $|c(0)| < 1$, $\mathbb{E}(|c(\eta_0)|^s) < 1$, and $\mathbb{E}(|h(\eta_0)|^s) < \infty$. See Proposition 5 in Carrasco and Chen (2002).
- The process $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ defined by Model (4.6) is geometrically β-mixing as soon as $0 < b \vee c < 1$.
- The process $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ defined by Model (4.7) is geometrically β-mixing as soon as the density of η_0 is positive on a neighborhood of 0 and $\limsup_{|x|\to+\infty} |f(x)/x| < 1$ (see Doukhan (1994), Proposition 6 page 107).

Note that some other extensions to nonlinear models having stationarity and dependency properties can be found in Lee and Shin (2005). Note also that, for the models (1.1) and

(1.2), the β-mixing coefficients of the process $(W_t)_{t\in\mathbb{Z}}$ are smaller than that of $((Y_t,\sigma_t))_{t\in\mathbb{Z}}$ (because of the inclusion of σ-algebras).

Concerning the τ-dependence, we can prove the following results that rely on a more general result given in Section 5.6.

Proposition 4.1. *Let Y_t and σ_t satisfy either (4.1) or (4.2). Assume that σ_0^2 and η_0^2 admit bounded densities and that $\mathbb{E}(\eta_0^2)=1$.*

1) For the ARCH(∞) Model (4.4), the following rates for $((X_t,Z_t))_{t\geq 0}$ hold:

- *If $a_j=0$, for $j\geq J$, then $((X_t,Z_t))_{t\geq 0}$ is geometrically τ-dependent.*
- *If $a_j=O(b^j)$ for some $b<1$ then $\tau_\infty(n)=O(\kappa^{\sqrt{n}})$ for some $\kappa<1$.*
- *If $a_j=O(j^{-b})$ for some $b>1$ then $\tau_\infty(n)=O(n^{-b/2}(\ln(n))^{b+2})$.*

2) For Model (4.2), if there exists $\kappa<1$ such that

$$\mathbb{E}(|(f(x,\eta_0))^2-(f(y,\eta_0))^2|)\leq \kappa|x^2-y^2|\,, \tag{4.8}$$

then $((X_t,Z_t))_{t>0}$ is geometrically τ dependent with $\tau_\infty(n)=O(n(\sqrt{\kappa})^n)$.

For more general models than (4.4), we refer to Doukhan *et al.* (2006). An example of Markov chain satisfying (4.8) is the autoregressive model $\sigma_t^2=h(\sigma_{t-1}^2)+r(\eta_{t-1})$ for some κ-lipschitz function h.

4.2. **Illustration of the method.** In this Section, we use Matlab programs that can be found on Yves Rozenholc's page: http://www.math-info.univ-paris5.fr/ rozen/.

In their 2005 paper, Comte *et al.* (2005) provide a simulation study for deconvolution in the i.i.d. case. Here are studied various signal densities g to estimate, and two types of errors: Gaussian and Laplace. We add here the case of $\varepsilon_t \rightsquigarrow \ln[(\mathcal{N}(0,1))^2]$. We refer to Comte *et al.* (2005) for the values of the constants in the penalty and similarly, we take for the last case

$$\mathrm{pen}(m)=\min\Big\{\max\Big[2.5\Big(1-\frac{1}{s2n}\Big)^2,0.6\Big],1\Big\}\Big(1+\frac{8\ln(\pi m)^{2.5}}{\pi m}+\pi m\Big)\int_{-\pi m}^{\pi m}\frac{dx}{|f_\varepsilon^*(x)|^2}$$

where $s2n=\widehat{\mathrm{var}}(Z)/\sigma_\varepsilon^2-1$, σ_ε^2 is the (known) variance of ε_1 and $\widehat{\mathrm{var}}(Z)$ is the empirical variance of the observations.

First, we study the problem of misspecification of the noise density. Figure 1-left plots the deconvolution "kernels" $x\mapsto|f_\varepsilon(x)|$ corresponding to the three cases. The similarity of the curves explains why choosing a wrong type of noise may not do much harm. Figure 1-right shows the true density and the estimates obtained with the three kernels in a model $Z_i=X_i+\ln(\eta_i^2)$ where the variables are i.i.d. with $\eta_i\sim\mathcal{N}(0,1)$ and X_i has density $0.6f_1+0.4f_2$ with f_1 a $\mathcal{N}(-2,4)$ density and f_2 a $\mathcal{N}(6,4)$. The signal to noise ratio s2n is about 5.6.

Figure 2-left shows the same picture when f_1 is $\mathcal{N}(-2,1)$ density and f_2 a $\mathcal{N}(4,1)$. Then the estimated s2n falls to 1.94 and the result is much less satisfactory. Nevertheless,

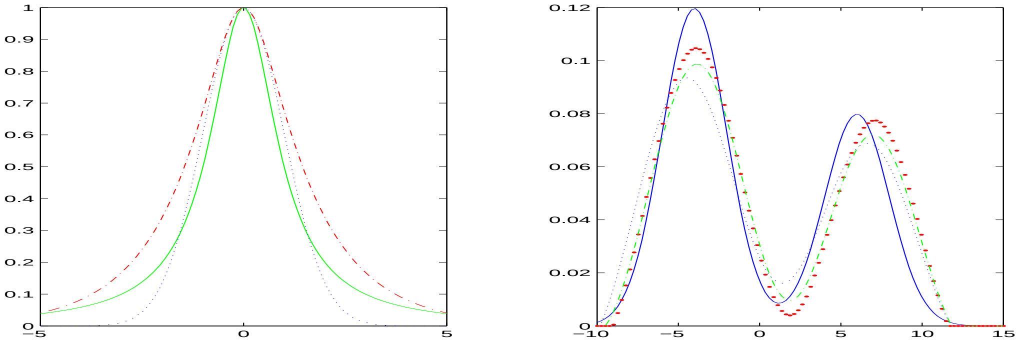

FIGURE 1. Left: $x \mapsto 1/(1+x^2)$ full line (green), $x \mapsto e^{-x^2/2}$ dotted line (blue), $x \mapsto |\Gamma(1/2+ix)/\sqrt{\pi}|$ dashed-dotted line (red). Right: Deconvolution of mixed normals with different noises, $n = 1000$ data. True density, full line. Estimation with the right noise density, small dotted line (blue). Estimation with Gaussian noise density, big dotted line (red). Estimation with Laplace noise density, dashed-dotted line (green).

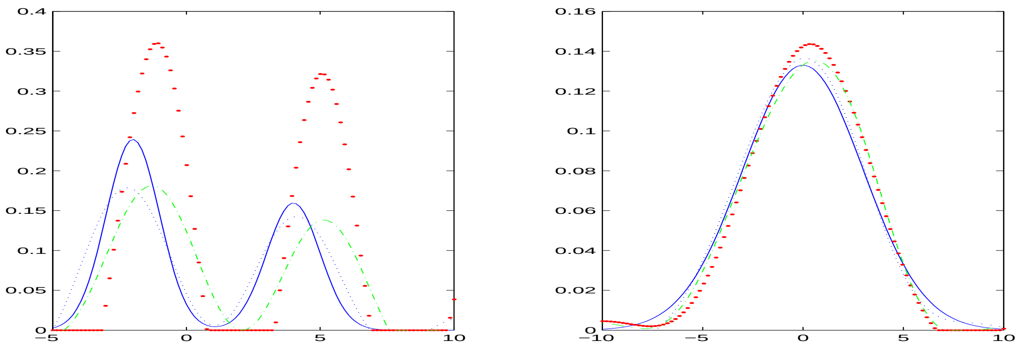

FIGURE 2. Deconvolution of a mixed normal (right) and of a normal (left) density with different noises, $n = 1000$ data. True density, full line. Estimation with the right noise density, small dotted line (blue). Estimation with Gaussian noise density, big dotted line (red). Estimation with Laplace noise density, dashed-dotted line (green).

the bimodal feature is captured in all cases. Figure 2-right shows the same picture for Gaussian $\mathcal{N}(0, 9)$ X_i's, giving here an estimated s2n equal to 1.66.

We also simulated two dependent processes. First, a GARCH(1,1) process, that is we take $p = q = 1$ in (4.3). For the parameter values $a = 1, a_1 = 0.7, b_1 = 0.2$ experimented by van Es *et al.* (2005), we obtain very small signal to noise ratio. Therefore, the deconvolution does not work well, and their Figure 1 shows it. Figure 3-top-left shows the estimated curve on the $\ln(Y_t^2)$ data (the estimated s2n is here equal to 0.13), and compares with the optimal histogram selected with Birgé and Rozenholc's (2006) procedure using the direct data $\ln(\sigma_t^2)$. We can see that the peak is cut and the estimated density is not null for negative x's as it should be. If we increase the GARCH variance by choosing $a = 5$,

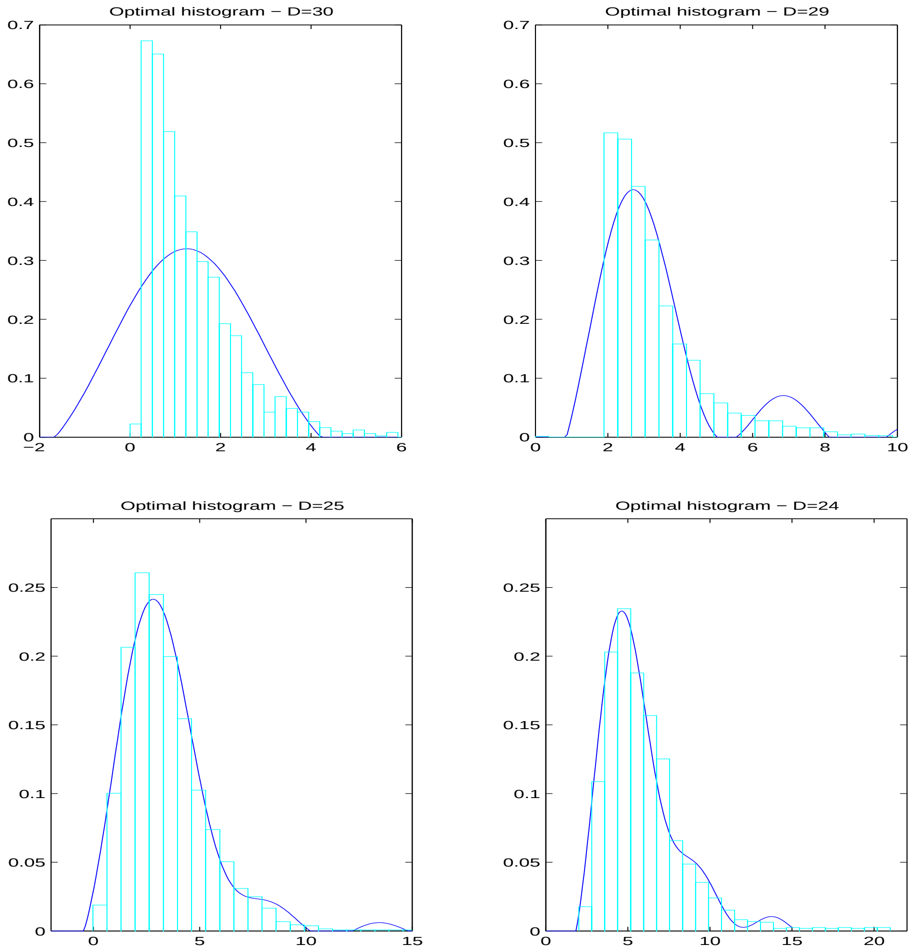

FIGURE 3. Deconvolution of two GARCH processes (top) and of two de Vries model (bottom). Estimated curve and optimal histogram on the direct data $\ln(\sigma_t^2)$ for $n = 2000$ simulated data.

$a_1 = 0.79$ and $b_1 = 0.2$, the result is much better, see Figure 3-top-right, where s2n reaches now 0.35. The second simulated process is de Vries's (1991) example, equation (10), given by $Y_t = \sigma_t \eta_t$ and $\sigma_t^2 = \tau^2 \sigma_{t-1}^2 + 1/\eta_{t-1}^2$, with i.i.d. $\mathcal{N}(0, 1)$ variables η_t. We choose $\tau^2 = 0.5$ (Figure 3-bottom-left) and $\tau^2 = 0.8$ (Figure 3-bottom-right) and obtain a signal to noise ratio which is around 1 in general, and equal to 0.6 (left) and 1.34 (right) in Figure 3-bottom. Here, the procedure works when we compare with the histograms of the true $\ln(\sigma_t^2)$ data. This shows that the procedure is well calibrated and can be experimented with real data.

We downloaded from Franses and van Dijk (2000) stock market data corresponding to four daily indices of stock markets in Amsterdam (EOE), Paris (CAC40), Tokyo (Nikkei) and New York (S&P 500) from January 6, 1986 until December 31, 1997, except for the CAC 40 which was created on July, 9, 1987. This gives 3128 data for each index, except the CAC 40 (2735 data). We compute as variables Y_t the centered log-returns of the indexes

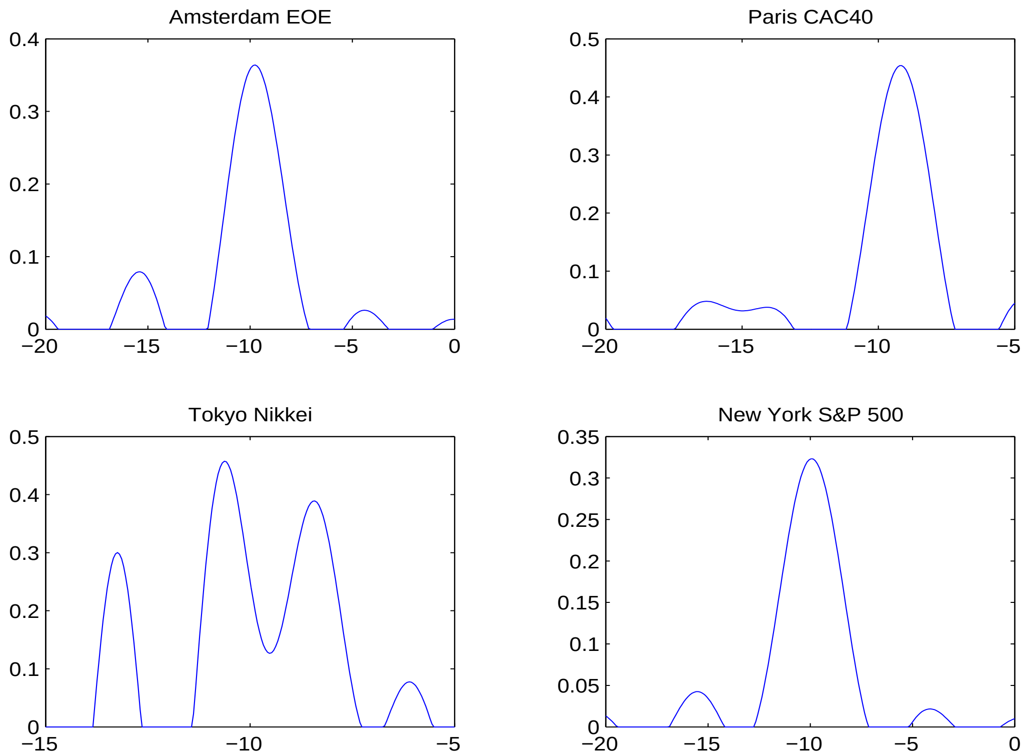

FIGURE 4. Deconvolution with real index returns data.

(i.e. $\ln(S_t/S_{t-1})$ minus its mean, if S_t denotes the value at t of the index). We plot in Figure 4 the deconvoluted density of $\ln(\sigma_t^2)$, assuming that $Y_t = \sigma_t\eta_t$ with Gaussian η_t's. The extremal left bump for the CAC 40 is certainly an artefact. They all look Gaussian, with mean around -10, and different variances, except the Nikkei, which exhibits an interesting bimodality (again, the two extremal left and right bumps are likely to be artefact, maybe due to small nonstationarity). Clearly, models must not be the same for this place.

5. PROOFS

5.1. **Two useful tools in the dependent context.** We first recall the coupling properties associated with the dependence coefficients. Assume that Ω is rich enough, which means that there exists U uniformly distributed over $[0,1]$ and independent of $\mathcal{M} \vee \sigma(W)$. There exist two $\mathcal{M} \vee \sigma(U) \vee \sigma(W)$-measurable random variables $W_1^\star$ and $W_2^\star$ distributed as W and independent of $\mathcal{M}$ such that

$$\beta(\mathcal{M}, \sigma(W)) = \mathbb{P}(W \neq W_1^\star) \quad \text{and} \quad \tau(\mathcal{M}, W) = \mathbb{E}(\|W - W_2^\star\|_{\mathbb{B}})\,. \tag{5.1}$$

The first equality in (5.1) is due to Berbee (1979), and the second one has been established in Dedecker and Prieur (2005), Section 7.1.

As consequences of the coupling properties (5.1), we have the following covariance inequalities. Let $\|\cdot\|_{\infty,\mathbb{P}}$ be the $\mathbb{L}^\infty(\Omega,\mathbb{P})$-norm. For two measurable functions f, h from $\mathbb{R}$

to $\mathbb{C}$, we have

$$|\mathrm{Cov}(f(Y), h(X))| \leq 2\|f(Y)\|_{\infty,\mathbb{P}}\|h(X)\|_{\infty,\mathbb{P}}\, \beta(\sigma(X), \sigma(Y))\,. \tag{5.2}$$

Moreover, if $\mathrm{Lip}(h)$ is the Lipschitz coefficient of h,

$$|\mathrm{Cov}\ (f(Y), h(X))| \leq \|f(Y)\|_{\infty,\mathbb{P}}\mathrm{Lip}(h)\, \tau(\sigma(Y), X)\,. \tag{5.3}$$

Thus, using that $t \to e^{ixt}$ is $|x|$-Lipschitz, we obtain the bounds

$$|\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})| \leq 2\beta_1(k-1) \quad \text{and} \quad |\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})| \ \leq\ |x|\tau_1(k-1). \tag{5.4}$$

5.2. **Proof of Proposition 3.1.** The proof of Proposition 3.1 follows the same lines as in the independent framework (see Comte *et al.* (2006)). The main difference lies in the control of the variance term. We keep the same notations as in Section 2.1. According to (2.2), for any given m belonging to $\{1, \cdots, m_n\}$, $\hat{g}_m^{(n)}$ satisfies, $\gamma_n(\hat{g}_m^{(n)}) - \gamma_n(g_m^{(n)}) \leq 0$. For a random variable T with density f_T, and any function ψ such that $\psi(T)$ is integrable, set $\nu_{n,T}(\psi) = n^{-1}\sum_{i=1}^n [\psi(T_i) - \langle \psi, f_T\rangle]$. In particular,

$$\nu_{n,Z}(u_t^*) = \frac{1}{n}\sum_{i=1}^n [u_t^*(Z_i) - \langle t, g\rangle]\,. \tag{5.5}$$

Since

$$\gamma_n(t) - \gamma_n(s) = \|t-g\|^2 - \|s-g\|^2 - 2\nu_{n,Z}(u_{t-s}^*), \tag{5.6}$$

we infer that

$$\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m^{(n)}\|^2 + 2\nu_{n,Z}\left(u^*_{\hat{g}_m^{(n)} - g_m^{(n)}}\right)\,. \tag{5.7}$$

Writing that $\hat{a}_{m,j} - a_{m,j} = \nu_{n,Z}(u^*_{\varphi_{m,j}})$, we obtain that

$$\nu_{n,Z}\left(u^*_{\hat{g}_m^{(n)} - g_m^{(n)}}\right) = \sum_{|j|\leq k_n} (\hat{a}_{m,j} - a_{m,j})\nu_{n,Z}(u^*_{\varphi_{m,j}}) = \sum_{|j|\leq k_n} [\nu_{n,Z}(u^*_{\varphi_{m,j}})]^2.$$

Consequently, $\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m^{(n)}\|^2 + 2\sum_{j\in\mathbb{Z}} \mathbb{E}[(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2]$. According to Comte *et al.* (2006),

$$\|g - g_m^{(n)}\|^2 = \| g - g_m \|^2 + \|g_m - g_m^{(n)}\|^2 \leq \| g - g_m \|^2 + \frac{(\pi m)^2(M_2+1)}{k_n}. \tag{5.8}$$

The variance term is studied by using first that for $f \in \mathbb{L}_1(\mathbb{R})$,

$$\nu_{n,Z}(f^*) = \int \nu_{n,Z}(e^{ix\cdot})f(x)dx. \tag{5.9}$$

Now, we use (5.9) and apply Parseval's formula to obtain

$$\begin{aligned} \mathbb{E}\Big(\sum_{j\in\mathbb{Z}} (\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2\Big) &= \frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}} \mathbb{E}\Big(\int \frac{\varphi^*_{m,j}(-x)}{f^*_\varepsilon(x)}\nu_{n,Z}(e^{ix\cdot})dx\Big)^2 \\ &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2}{|f^*_\varepsilon(x)|^2}dx. \end{aligned} \tag{5.10}$$

Since $\nu_{n,Z}$ involves centered and stationary variables, we have

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 = \mathrm{Var}|\nu_{n,Z}(e^{ix\cdot})| = \frac{1}{n}\mathrm{Var}(e^{ixZ_1}) + \frac{1}{n^2}\sum_{1\le k\neq l\le n}\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}). \tag{5.11}$$

It follows from the structure of the model that, for $k < l$, ε_l is independent of (X_l, Z_k), so that $\mathbb{E}(e^{ixZ_k}) = f_\varepsilon^*(x)g^*(x)$ and $\mathbb{E}(e^{ix(Z_l - Z_k)}) = f_\varepsilon^*(x)\mathbb{E}(e^{ix(X_l - Z_k)})$. Thus, for $k < l$,

$$\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}) = f_\varepsilon^*(x)\mathrm{Cov}(e^{ixZ_k}, e^{ixX_l}). \tag{5.12}$$

From (5.11) and the stationarity of $(X_i)_{i\ge1}$, we obtain that

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 \le \frac{1}{n} + \frac{2}{n}\sum_{k=2}^{n}\left|\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})\right| |f_\varepsilon^*(x)|. \tag{5.13}$$

The first part of Proposition 3.1 follows from the stationarity of the X_i's, and from (5.7), (5.8), (5.10) and (5.13).

The proof of $R_m \le \min(R_{m,\beta}, R_{m,\tau})$, where $R_{m,\beta}$ and $R_{m,\tau}$ are defined in Proposition 3.1, comes from the inequalities (5.4) in Section 5.1. Hence we get the result.□

5.3. **Proof of Corollary 3.1.** According to Butucea and Tsybakov (2005), under (3.9), we have

$$\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m)(1 + o_m(1)) \le \Delta(m) \le \lambda_1(f_\varepsilon, \kappa_0)\Gamma(m)(1 + o_m(1)) \quad \text{as } m \to \infty, \text{ where}$$

$$\Gamma(m) = (1 + (\pi m)^2)^\gamma(\pi m)^{1-\delta}\exp\left\{2\mu(\pi m)^\delta\right\}, \tag{5.14}$$

where λ_1 is defined in (3.10). In the same way

$$\overline{\lambda_1}(f_\varepsilon, \kappa_0')\overline{\Gamma}(m)(1 + o_m(1)) \le \Delta_{1/2}(m) \le \overline{\lambda_1}(f_\varepsilon, \kappa_0)\overline{\Gamma}(m)(1 + o_m(1)) \quad \text{as } m \to \infty,$$

where

$$\begin{aligned}\overline{\Gamma}(m) &= (1 + (\pi m)^2)^{\gamma/2}(\pi m)^{1-\delta}\exp(\mu(\pi m)^\delta)\\ \overline{\lambda_1}(f_\varepsilon, \kappa_0) &= \left[\kappa_0^2\pi(1\!\mathrm{I}_{\{\delta=0\}} + \mu\delta 1\!\mathrm{I}_{\{\delta>0\}})\right]^{-1}.\end{aligned}$$

It is easy to see that $\Delta_{1/2}(m) \le \sqrt{m\Delta(m)}$ and hence $\Delta_{1/2}(m) = \Gamma(m)o_m(1)$. Now, as soon as $\gamma > 1$ when $\delta = 0$, $m\Delta_{1/2}(m) = \Gamma(m)o_m(1)$. Set m_1 such that for $m \ge m_1$ we have

$$0.5\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m) \le \Delta(m) \le 2\lambda_1(f_\varepsilon, \kappa_0)\Gamma(m), \tag{5.15}$$

and

$$0.5\overline{\lambda_1}(f_\varepsilon, \kappa_0')\overline{\Gamma}(m) \le \Delta_{1/2}(m) \le 2\overline{\lambda_1}(f_\varepsilon, \kappa_0)\overline{\Gamma}(m). \tag{5.16}$$

If $\sum_{k\geq 1}\beta_1(k)<+\infty$, (3.9) and (3.5) hold, and if $k_n\geq n$, then we have the upper bounds: for $m\geq m_1$, $\lambda_1=\lambda_1(f_\varepsilon,\kappa_0)$ and $\overline{\lambda_1}=\overline{\lambda_1}(f_\varepsilon,\kappa_0)$,

$$\begin{aligned}\mathbb{E}\|g-\hat g_m^{(n)}\|^2 &\leq \|g-g_m\|^2+\frac{m^2(M_2+1)}{n}+\frac{2\lambda_1\Gamma(m)}{n}+8\overline{\lambda_1}\sum_{k\geq 1}\beta_1(k)\frac{\overline{\Gamma}(m)}{n}\\ &\leq \|g-g_m\|^2+\frac{m^2(M_2+1)}{n}+\frac{2\lambda_1\Gamma(m)}{n}+\frac{C(\sum_{k\geq 1}\beta_1(k))\Gamma(m)}{n}o_m(1).\end{aligned}$$

In the same way, if $\sum_{k\geq 1}\tau_1(k)<+\infty$, if $\gamma>1$ when $\delta=0$, if (3.9) and (3.5) hold, and if $k_n\geq n$, then we have the upper bound: for $m\geq m_1$,

$$\begin{aligned}\mathbb{E}\|g-\hat g_m^{(n)}\|^2 &\leq \|g-g_m\|^2+\frac{m^2(M_2+1)}{n}+\frac{2\lambda_1\Gamma(m)}{n}+2\pi\overline{\lambda_1}\sum_{k\geq 1}\tau_1(k)\frac{m\overline{\Gamma}(m)}{n}\\ &\leq \|g-g_m\|^2+\frac{m^2(M_2+1)}{n}+\frac{2\lambda_1\Gamma(m)}{n}+\frac{C(\sum_{k\geq 1}\tau_1(k))\Gamma(m)}{n}o_m(1).\end{aligned}$$

Since $\gamma>1$ when $\delta=0$, the residual term $n^{-1}m^2(M_2+1)$ is negligible with respect to the variance term.

Finally, g_m being the orthogonal projection of g on S_m, we get $g_m^*=g^*\mathbb{1}_{[-m\pi,m\pi]}$ and therefore

$$\|g-g_m\|^2=\frac{1}{2\pi}\|g^*-g_m^*\|^2=\frac{1}{2\pi}\int_{|x|\geq\pi m}|g^*|^2(x)dx.$$

If g belongs to the class $\mathcal{S}_{s,r,b}(C_1)$ defined in (3.11), then

$$\|g-g_m\|^2\leq\frac{C_1}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\}.$$

The corollary is proved. □

5.4. **Proof of Theorem 3.1.** By definition, $\tilde g$ satisfies that for all $m\in\{1,\cdots,m_n\}$,

$$\gamma_n(\tilde g)+\mathrm{pen}(\hat m)\leq\gamma_n(g_m)+\mathrm{pen}(m).$$

Therefore, by using (5.6) we get

$$\|\tilde g-g\|^2\leq\|g_m^{(n)}-g\|^2+2\nu_{n,Z}(u^*_{\tilde g-g_m^{(n)}})+\mathrm{pen}(m)-\mathrm{pen}(\hat m),$$

where $\nu_{n,Z}$ is defined in (5.5). If $t=t_1+t_2$ with t_1 in $S_m^{(n)}$ and t_2 in $S_{m'}^{(n)}$, t^* has its support in $[-\pi\max(m,m'),\pi\max(m,m')]$ and t belongs to $S^{(n)}_{\max(m,m')}$. Set $B_{m,m'}(0,1)=\{t\in S^{(n)}_{\max(m,m')}\ /\ \|t\|=1\}$ and write

$$|\nu_{n,Z}(u^*_{\tilde g-g_m^{(n)}})|\leq\|\tilde g-g_m^{(n)}\|\sup_{t\in B_{m,\hat m}(0,1)}|\nu_{n,Z}(u_t^*)|.$$

Using that $2uv\leq a^{-1}u^2+av^2$ for any $a>1$, leads to

$$\|\tilde g-g\|^2\ \leq\ \|g_m^{(n)}-g\|^2+a^{-1}\|\tilde g-g_m^{(n)}\|^2+a\sup_{t\in B_{m,\hat m}(0,1)}(\nu_{n,Z}(u_t^*))^2+\mathrm{pen}(m)-\mathrm{pen}(\hat m).$$

Therefore, we have

$$\|\tilde g-g\|^2\leq\kappa_a^2\|g_m^{(n)}-g\|^2+a\kappa_a\sup_{t\in B_{m,\hat m}(0,1)}|\nu_{n,Z}(u_t^*)|^2+\kappa_a(\mathrm{pen}(m)-\mathrm{pen}(\hat m)),\tag{5.17}$$

where κ_a is defined in (3.16).

The main point is to control $\sup_{t\in B_{m,\hat{m}}(0,1)}|\nu_{n,Z}(u_t^*)|^2$, the supremum of a centered empirical process $\nu_{n,Z}(u_t^*)$. To handle this supremum, we use coupling methods to replace the dependent random variables (r.v.'s) involved in this process, by block-independent r.v.'s in order to apply Talagrand's Inequality. Heuristically, we replace the expectation of the supremum by the (negligible) price of coupling plus the expectation of a term $p(m,\hat{m})$ imposed by Talagrand's Inequality, which in turn will fix the choice of the penalty function.

Proof in the β-mixing case.

We use the coupling methods recalled in Section 5.1 to build approximating variables for the $W_i=(Z_i,X_i)$'s. More precisely, we build variables $W_i^\star$ such that if $n=2p_nq_n+r_n$, $0\le r_n<q_n$, and $\ell=0,\cdots,p_n-1$

$$E_\ell=(W_{2\ell q_n+1},...,W_{(2\ell+1)q_n}),\ F_\ell=(W_{(2\ell+1)q_n+1},...,W_{(2\ell+2)q_n}),$$

$$E_\ell^\star=(W^\star_{2\ell q_n+1},...,W^\star_{(2\ell+1)q_n}),\ F_\ell^\star=(W^\star_{(2\ell+1)q_n+1},...,W^\star_{(2\ell+2)q_n}).$$

The variables $E_\ell^\star$ and $F_\ell^\star$ are such that

- $E_\ell^\star$ and E_ℓ are identically distributed. $F_\ell^\star$ and F_ℓ are identically distributed.
- $\mathbb{P}(E_\ell\neq E_\ell^*)\le\beta_\infty(q_n)$ and $\mathbb{P}(F_\ell\neq F_\ell^*)\le\beta_\infty(q_n)$,
- $E_\ell^\star$ and $\mathcal{M}_0\vee\sigma(E_0,E_1,...,E_{\ell-1},E_0^\star,E_1^\star,\cdots,E_{\ell-1}^\star)$ are independent, and therefore independent of $\mathcal{M}_{(\ell-1)q_n}$ and the same holds for the blocks $F_\ell^\star$.

For the sake of simplicity we assume that $r_n=0$. We denote by $(Z_i^\star,X_i^\star)=W_i^\star$ the new couple of variables. Using the notation (5.5), we denote by $\nu_{n,Z}^\star(u_t^*)$ the empirical contrast computed on the $Z_i^\star$. Then we write

$$\begin{aligned}\|\tilde{g}-g\|^2 &\le \kappa_a^2\|g-g_m^{(n)}\|^2+2a\kappa_a\sup_{t\in B_{m,\hat{m}}(0,1)}|\nu_{n,Z}^\star(u_t^*)|^2+\kappa_a(\mathrm{pen}(m)-\mathrm{pen}(\hat{m}))\\ &\quad +2a\kappa_a\sup_{t\in B_{m,\hat{m}}(0,1)}|\nu_{n,Z}^\star(u_t^*)-\nu_{n,Z}(u_t^*)|^2.\end{aligned}$$

Set

$$T_n^\star(m,m'):=\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,Z}^\star(t)|^2-p(m,m')\Big]_+,\tag{5.18}$$

where $p(m,m')$ will defined further. Hence

$$\begin{aligned}\|\tilde{g}-g\|^2 &\le \kappa_a^2\|g-g_m^{(n)}\|^2+2a\kappa_aT_n^\star(m,\hat{m})+\kappa_a\left(2ap(m,\hat{m})+\mathrm{pen}(m)-\mathrm{pen}(\hat{m})\right)\\ &\quad +2a\kappa_a\sup_{t\in B_{m,\hat{m}}(0,1)}|\nu_{n,Z}(u_t^*)-\nu_{n,Z}^\star(u_t^*)|^2\\ &\le \kappa_a^2\|g-g_m^{(n)}\|^2+2\kappa_a\mathrm{pen}(m)+2a\kappa_a\sup_{t\in B_{m,\hat{m}}(0,1)}|\nu_{n,Z}(u_t^*)-\nu_{n,Z}^\star(u_t^*)|^2\\ &\quad +2a\kappa_aT_n^\star(m,\hat{m})\end{aligned}\tag{5.19}$$

where $\mathrm{pen}(m)$ is chosen such that

$$2ap(m,m') \le \mathrm{pen}(m) + \mathrm{pen}(m'). \tag{5.20}$$

Now write

$$\begin{aligned}\nu_{n,Z}(u_t^*) - \nu_{n,Z}^\star(u_t^*) &= \frac{1}{2\pi}\frac{1}{n}\sum_{k=1}^n \int [e^{ixZ_k} - e^{ixZ_k^\star}]\frac{t^*(-x)}{f_\varepsilon^*(x)}dx \\ &= \frac{1}{2\pi}\int [\nu_{n,Z}(e^{ix\cdot}) - \nu_{n,Z}^\star(e^{ix\cdot})]\frac{t^*(-x)}{f_\varepsilon^*(x)}dx.\end{aligned}$$

Consequently,

(5.21)

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu_{n,Z}^\star(u_t^*)|^2\Big] \le \int_{-\pi m_n}^{\pi m_n} \mathbb{E}[|\nu_{n,Z}(e^{ix\cdot}) - \nu_{n,Z}^\star(e^{ix\cdot})|^2]\frac{1}{|f_\varepsilon^*(x)|^2}dx.$$

Since

$$\begin{aligned}\mathbb{E}[|\nu_{n,Z}(e^{ix\cdot}) - \nu_{n,Z}^\star(e^{ix\cdot})|^2] &= \mathbb{E}[|\nu_{n,Z}(e^{ix\cdot}) - \nu_{n,Z}^\star(e^{ix\cdot})1\!\!1_{Z_k\neq Z_k^\star}|^2] \\ &\le 4\mathbb{E}\Big[\frac{1}{n}\sum_{k=1}^n |1\!\!1_{Z_k\neq Z_k^\star}|^2\Big] \le 4\beta_\infty(q_n),\end{aligned}$$

we obtain that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu_{n,Z}^\star(u_t^*)|^2\Big] \le 4\beta_\infty(q_n)\Delta(m_n). \tag{5.22}$$

By gathering (5.19) and (5.22) we get

$$\mathbb{E}\|\tilde g - g\|^2 \le \kappa_a^2\|g - g_m^{(n)}\|^2 + 2a\kappa_a\sum_{m'=1}^{m_n}\mathbb{E}\big[T_n^\star(m,m')\big] + 2\kappa_a\mathrm{pen}(m) + 2a\kappa_a\beta_\infty(q_n)\Delta(m_n).$$

Therefore we infer that, for all $m\in\{1,\cdots,m_n\}$,

$$\mathbb{E}\|g - \tilde g\|^2 \le C_a\left[\|g - g_m^{(n)}\|^2 + \mathrm{pen}(m)\right] + 2a\kappa_a(C_1+C_2)/n, \tag{5.23}$$

provided that

$$\Delta(m_n)\beta_\infty(q_n) \le C_1/n \quad \text{and} \quad \sum_{m'=1}^{m_n}\mathbb{E}(T_n^\star(m,m')) \le C_2/n. \tag{5.24}$$

Using (5.15), we conclude that the first part of (5.24) is fulfilled as soon as

$$m_n^{2\gamma+1-\delta}\exp\{2\mu\pi^\delta m_n^{\ \delta}\}\beta_\infty(q_n) \le C_1'/n. \tag{5.25}$$

In order to ensure that our estimators converge, we only consider models with bounded penalty, and therefore (5.25) requires that $\beta_\infty(q_n) \le C_1'/n^2$. For $q_n = [n^c]$ and $\beta_\infty(k) = O(n^{-1-\theta})$, we obtain the condition $n^{-c(1+\theta)} = O(n^{-2})$. If $\theta > 3$, one can find $c \in]0,1/2[$, such that this condition is satisfied. Consequently, (5.25) holds.

To prove the second part of (5.24), we split $T_n^\star(m,m')$ into two terms

$$T_n^\star(m,m') \le (T_{n,1}^\star(m,m') + T_{n,2}^\star(m,m'))/2,$$

where $p_1(m,m') = p_2(m,m') = p(m,m')/2$, and for $k=1,2$

(5.26)
$$T^\star_{n,k}(m,m') = \Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\big(u_t^*(Z^\star_{(2\ell+k-1)q_n+i}) - \langle t,g\rangle\big)\Big|^2 - p_k(m,m')\Big]_+.$$

We only study $T^\star_{n,1}(m,m')$ and conclude for $T^\star_{n,2}(m,m')$ analogously. The study of $T^\star_{n,1}(m,m')$ consists in applying a concentration inequality to $\nu^\star_{n,1}(t)$ defined by

$$\nu^\star_{n,1}(t) = \frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\big(u_t^*(Z^\star_{2\ell q_n+i}) - \langle t,g\rangle\big) = \frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu^\star_{q_n,\ell}(u_t^*). \tag{5.27}$$

The random variable $\nu^\star_{n,1}(u_t^*)$ is considered as the sum of the p_n independent random variables $\nu^\star_{q_n,\ell}(t)$ defined as

$$\nu^\star_{q_n,\ell}(u_t^*) = (1/q_n)\sum_{j=1}^{q_n} u_t^*(Z^\star_{2\ell q_n+j}) - \langle t,g\rangle. \tag{5.28}$$

Let $m^* = \max(m,m')$. Let $M_1^\star(m^*)$, $v^\star(m^*)$ and $H^\star(m^*)$ be some terms such that $\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u_t^*) \|_\infty \le M_1^\star(m^*)$, $\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell}(u_t^*)) \le v^\star(m)$ and lastly $\mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u_t^*)|) \le H^\star(m^*)$. According to Lemma 5.2 we take

$$(H^\star(m^*))^2 = \frac{2\Delta(m^*)}{n},\ M_1^\star(m^*) = \sqrt{\Delta(m^*)} \text{ and } v^\star(m^*) = \frac{2\sqrt{\Delta_2(m^*,f_Z)}}{2\pi q_n},$$

where

$$\Delta_2(m,f_Z) = \int_{-\pi m}^{\pi m}\int_{-\pi m}^{\pi m} \frac{|f_Z^*(x-y)|^2}{|f_\varepsilon^*(x)f_\varepsilon^*(y)|^2}dxdy. \tag{5.29}$$

From the definition of $T^\star_{n,1}(m,m')$, by taking $p_1(m,m') = 2(1+2\xi^2)(H^\star)^2(m^*)$, we get

$$\mathbb{E}(T^\star_{n,1}(m,m')) \le \mathbb{E}\big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u_t^*) - 2(1+2\xi^2)(H^\star)^2(m^*)\big]_+. \tag{5.30}$$

According to the condition (5.20), we thus take

$$\begin{aligned} \mathrm{pen}(m) &= 2ap(m,m) = 2a(p_1(m,m)+p_2(m,m)) = 4ap_1(m,m) \\ &= 8a(1+2\xi^2)\big(2n^{-1}\Delta(m)\big) = 16a(1+2\xi^2)n^{-1}\Delta(m). \end{aligned} \tag{5.31}$$

where ξ^2 is suitably chosen. Set m_2 and m_3 as defined in Lemma 5.2, and set m_1 such that for $m^* \ge m_1$, $\Delta(m^*)$ satisfies (5.15). Take $m_0 = m_1 \vee m_2 \vee m_3$. We split the sum over m' in two parts and write

$$\sum_{m'=1}^{m_n} \mathbb{E}(T^\star_{n,1}(m,m')) = \sum_{m'|m^*\le m_0} \mathbb{E}(T^\star_{n,1}(m,m')) + \sum_{m'|m^*\ge m_0} \mathbb{E}(T^\star_{n,1}(m,m')). \tag{5.32}$$

By applying Lemma 5.4, we get $\mathbb{E}(T^\star_{n,1}(m,m')) \le K[I(m^*)+II(m^*)]$, where

$$I(m^*) = \frac{\sqrt{\Delta_2(m^*,f_Z)}}{p_n}\exp\left\{-K_1\xi^2\frac{\Delta(m^*)}{q_nv^\star(m^*)}\right\},\ II(m^*) = \frac{\Delta(m^*)}{p_n^2}\exp\left\{-\frac{2K_1\xi C(\xi)^2}{7}\frac{\sqrt{n}}{q_n}\right\}.$$

When $m^* \le m_0$, with m_0 finite, we get that, for all $m \in \{1, \cdots, m_n\}$,

$$\sum_{m'|m^* \le m_0} \mathbb{E}(R^\star_{n,1}(m,m')) \le \frac{C(m_0)}{n}.$$

We now come to the sum over m' such that $m^* \ge m_0$. It follows from Comte *et al.* (2006) that

$$v^\star(m^*) = \frac{2\sqrt{\Delta_2(m^*, f_Z)}}{2\pi q_n} \quad \le \quad 2\lambda^\star_2(f_\varepsilon, \kappa_0)\frac{\Gamma_2(m^*)}{q_n}, \tag{5.33}$$

with

$$\lambda^\star_2(f_\varepsilon, \kappa_0) = \kappa_0^{-1}\sqrt{2\pi\lambda_1}\|f_{\varepsilon^*}\|\mathbb{I}_{\delta\le 1} + \mathbb{I}_{\delta>1} \tag{5.34}$$

where $\lambda_1 = \lambda_1(f_\varepsilon, \kappa_0)$ is defined in (3.10) and

(5.35)
$$\Gamma_2(m) = (1+(\pi m)^2)^\gamma (\pi m)^{\min((1/2-\delta/2),(1-\delta))} \exp(2\mu(\pi m)^\delta) = (\pi m)^{-(1/2-\delta/2)_+}\Gamma(m).$$

By combining the left hand-side of (5.15) and (5.33), we get that, for $m^* \ge m_0$,

$$I(m^*) \le \frac{\lambda^\star_2(f_\varepsilon, \kappa_0)\Gamma_2(m^*)}{n} \exp\left\{-\frac{K_1\xi^2\lambda_1(f_\varepsilon, \kappa'_0)}{2\lambda^\star_2(f_\varepsilon,\kappa_0)}(\pi m^*)^{(1/2-\delta/2)_+}\right\}$$

$$\text{and} \qquad II(m^*) \le \frac{\Delta(m^*)q_n^2}{n^2}\exp\left\{-\frac{2K_1\xi C(\xi^2)}{7}\frac{\sqrt{n}}{q_n}\right\}.$$

• Study of $\sum_{m'|m^*\ge m_0} II(m^*)$. According to the choices for $v^\star(m^*)$, $(H^\star(m^*))^2$ and $M_1^\star(m^*)$, we have

$$\begin{aligned}\sum_{m'|m^*\ge m_0} II(m^*) \quad &\le \quad \sum_{m'\in\{1,\cdots,m_n\}} \frac{\Delta(m^*)q_n^2}{n^2}\exp\left\{-\frac{2K_1\xi C(\xi^2)}{7}\frac{\sqrt{n}}{q_n}\right\} \\ &= \quad O\left[m_n \exp\left\{-\frac{2K_1\xi C(\xi^2)}{7}\frac{\sqrt{n}}{q_n}\right\}\frac{\Delta(m_n)q_n^2}{n^2}\right].\end{aligned}$$

Since $\Delta(m_n)/n$ is bounded, then $q_n = [n^c]$ with c in $]0, 1/2[$ ensures that

$$\sum_{m'=1}^{m_n} m_n \exp\left\{-\frac{2K_1\xi C(\xi^2)}{7}\frac{\sqrt{n}}{q_n}\right\}\frac{\Delta(m_n)q_n^2}{n^2} \le \frac{C}{n}. \tag{5.36}$$

Consequently

$$\sum_{m'|m^*\ge m_0} II^\star(m^*) \le \frac{C}{n}. \tag{5.37}$$

• Study of $\sum_{m'|m^*\ge m_0} I(m^*)$. Denote by $\psi = 2\gamma + \min(1/2-\delta/2, 1-\delta)$, $\omega = (1/2-\delta/2)_+$, and $K' = K_1\lambda_1(f_\varepsilon, \kappa'_0)/(2\lambda^\star_2(f_\varepsilon,\kappa_0))$. For $a, b \ge 1$, we use that

$$\begin{aligned}\max(a,b)^\psi e^{2\mu\pi^\delta \max(a,b)^\delta} e^{-K'\xi^2\max(a,b)^\omega} &\le (a^\psi e^{2\mu\pi^\delta a^\delta} + b^\psi e^{2\mu\pi^\delta b^\delta})e^{-(K'\xi^2/2)(a^\omega+b^\omega)} \\ &\le a^\psi e^{2\mu\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega} e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2\mu\pi^\delta b^\delta} e^{-(K'\xi^2/2)b^\omega}. \end{aligned} \tag{5.38}$$

Consequently,

$$\sum_{m'|m^*\geq m_0} I(m^*) \leq \sum_{m'=1}^{m_n} \frac{\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m^*)}{n} \exp\left\{-\frac{K_1\xi^2\lambda_1(f_\varepsilon,\kappa_0')}{2\lambda_2^\star(f_\varepsilon,\kappa_0)}(\pi m^*)^{(1/2-\delta/2)_+}\right\}$$

$$\leq \frac{2\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m)}{n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m)^{(1/2-\delta/2)_+}\right\}\sum_{m'=1}^{m_n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+}\right\}$$

$$+\sum_{m'=1}^{m_n}\frac{2\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m')}{n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+}\right\}. \tag{5.39}$$

Case $0\leq\delta<1/3$. In that case, since $\delta<(1/2-\delta/2)_+$, the choice $\xi^2=1$ ensures that $\Gamma_2(m)\exp\{-(K'\xi^2/2)(m)^{(1/2-\delta/2)}\}$ is bounded and thus the first term in (5.39) is bounded by C/n. Since $1\leq m\leq m_n$ with m_n such that $\Delta(m_n)/n$ is bounded, the term $\sum_{m'=1}^{m_n}\Gamma_2(m')\exp\{-(K'/2)(m')^{(1/2-\delta/2)}\}/n$ is bounded by C'/n, and hence

$$\sum_{m'|m^*\geq m_0} I(m^*)\leq\frac{C}{n}.$$

According to (5.20), the result follows by choosing $\mathrm{pen}(m)=2ap(m,m)=48a\Delta(m)/n$.

Case $\delta=1/3$. According to the inequality (5.38), ξ^2 is such that $2\mu\pi^\delta(m)^\delta-(K'\xi^2/2)m^\delta=-2\mu(\pi m^*)^\delta$ that is

$$\xi^2=\frac{16\mu\pi^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{K_1\lambda_1(f_\varepsilon,\kappa_0')}.$$

Arguing as for the case $0\leq\delta<1/3$, this choice ensures that $\sum_{m'|m^*\geq m_0} I(m^*)\leq C/n$. The result follows by taking $p(m,m')=2(1+2\xi^2)\Delta(m^*)/n$, and

$$\mathrm{pen}(m) = 16a(1+2\xi^2)\frac{\Delta(m)}{n}=16a\left(1+\frac{192\mu\pi^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{\lambda_1(f_\varepsilon,\kappa_0')}\right)\frac{\Delta(m)}{n}.$$

Case $\delta>1/3$. In that case $\delta>(1/2-\delta/2)_+$. We choose ξ^2 such that

$$2\mu\pi^\delta(m^*)^\delta-(K'\xi^2/2)(m^*)^\omega=-2\mu\pi^\delta(m^*)^\delta.$$

In other words

$$\xi^2=\xi^2(m^*)=\frac{16\mu(\pi)^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{K_1\lambda_1(f_\varepsilon,\kappa_0')}(\pi m^*)^{\min((3\delta/2-1/2)_+,\delta)}.$$

Consequently, we have $\sum_{m'|m^*\geq m_0} I(m^*)\leq C/n$. The result follows by choosing $p(m,m')=2(1+2\xi^2(m,m'))\Delta(m)/n$, associated to

$$\begin{aligned}\mathrm{pen}(m) &= 16a(1+2\xi^2(m))\frac{\Delta(m)}{n}\\ &= 16a\left(1+\frac{192\mu\pi^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{\lambda_1(f_\varepsilon,\kappa_0')}(\pi m^*)^{\min((3\delta/2-1/2)_+,\delta)}\right)\frac{\Delta(m)}{n}\end{aligned}$$ □

Proof in the τ-dependent case.

We use the coupling properties recalled in Section 5.1 to build approximating variables for the $W_i = (Z_i, X_i)$'s. More precisely, we build variables $W_i^\star$ such that if $n = 2p_nq_n + r_n$, $0 \le r_n < q_n$, and $\ell = 0, \cdots, p_n - 1$

$$E_\ell = (W_{2\ell q_n+1}, ..., W_{(2\ell+1)q_n}),\ F_\ell = (W_{(2\ell+1)q_n+1}, ..., W_{(2\ell+2)q_n}),$$
$$E_\ell^\star = (W^\star_{2\ell q_n+1}, ..., W^\star_{(2\ell+1)q_n}),\ F_\ell^\star = (W^\star_{(2\ell+1)q_n+1}, ..., W^\star_{(2\ell+2)q_n}).$$

The variables $E_\ell^\star$ and $F_\ell^\star$ are such that

- $E_\ell^\star$ and E_ℓ are identically distributed, $F_\ell^\star$ and F_ℓ are identically distributed,
- $\sum_{i=1}^{q_n} \mathbb{E}(\|W_{2\ell q_n+i} - W^\star_{2\ell q_n+i}\|_{\mathbb{R}^2}) \le q_n\tau_\infty(q_n)$, $\sum_{i=1}^{q_n} \mathbb{E}(\|W_{(2\ell+1)q_n+i} - W^\star_{(2\ell+1)q_n+i}\|_{\mathbb{R}^2}) \le q_n\tau_\infty(q_n)$,
- $E_\ell^\star$ and $\mathcal{M}_0 \vee \sigma(E_0, E_1, ..., E_{\ell-1}, E_0^\star, E_1^\star, \cdots, E_{\ell-1}^\star)$ are independent, and therefore independent of $\mathcal{M}_{(\ell-1)q_n}$ and the same holds for the blocks $F_\ell^\star$.

For the sake of simplicity we assume that $r_n = 0$. We denote by $(Z_i^\star, X_i^\star) = W_i^\star$ the new couple of variables.

As for the proof in the β-mixing framework, we start from (5.19) with $R_n^\star(m, \hat m)$ defined by (5.18) and pen(m) chosen such that (5.20) holds. Next we use (5.21) and the bound $|e^{-ixt} - e^{-ixs}| \le |x||t-s|$. Hence we conclude that

$$\sum_{i=1}^{q_n} \mathbb{E}(|e^{-iX_{2\ell q_n+i}} - e^{-iX^\star_{2\ell q_n+i}}|) \le q_n|x|\tau_{\mathbf{X},\infty}(q_n)$$

It follows that

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu^\star_{n,Z}(u_t^*)|^2\Big] &\le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} \mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|dx \\ &\le \frac{\tau_{\mathbf{X},\infty}(q_n)}{\pi}\int_{-\pi m_n}^{\pi m_n} \frac{|x|}{|f_\varepsilon^*(x)|^2}dx \\ &\le \tau_{\mathbf{X},\infty}(q_n)m_n\Delta(m_n). \end{aligned} \tag{5.40}$$

By gathering (5.19) and (5.40) we get

$$\mathbb{E}\|\tilde g - g\|^2 \le \kappa_a^2\|g - g_m^{(n)}\|^2 + 2a\kappa_a\sum_{m'=1}^{m_n} \mathbb{E}\big[T_n^\star(m, m')\big] + 2\kappa_a\mathrm{pen}(m) + 2a\kappa_a\tau_\infty(q_n)m_n\Delta(m_n).$$

Therefore we infer that, for all $m \in \{1, \cdots, m_n\}$, (5.23) holds provided that

$$\Delta(m_n)m_n\tau_\infty(q_n) \le C_1/n \quad \text{and} \quad \sum_{m'=1}^{m_n} \mathbb{E}(T_n^\star(m, m')) \le C_2/n. \tag{5.41}$$

Using (5.15), we conclude that the first part of (5.41) is fulfilled as soon as

$$m_n{}^{2\gamma+2-\delta}\exp\{2\mu\pi^\delta m_n{}^\delta\}\tau_\infty(q_n) \le C_1'/n. \tag{5.42}$$

In order to ensure that our estimators converge, we only consider models with bounded penalty, that is $\Delta(m_n) = O(n)$. Therefore (5.42) requires that $m_n\tau_\infty(q_n) \le C_1'/n^2$. For $q_n = [n^c]$ and $\tau_\infty(k) = O(n^{-1-\theta})$, we obtain the condition

$$m_n n^{-c(1+\theta)} = O(n^{-2}). \tag{5.43}$$

If f_ε satisfies (3.9) with $\delta > 0$, and if $\theta > 3$, one can find $c \in]0, 1/2[$, such that (5.43) is satisfied. Now, if $\delta = 0$ and $\gamma \geq 3/2$ in (3.9) and if $\theta > 3 + 2/(1 + 2\gamma)$, then one can find $c \in]0, 1/2[$, such that (5.43) is satisfied. These conditions ensure that (5.25) holds.

In order to prove the second part of (5.41), we proceed as for the proof of the second part of (5.24) and split $T_n^\star(m, m')$ into two terms

$$T_n^\star(m, m') \leq (T_{n,1}^\star(m, m') + T_{n,2}^\star(m, m'))/2,$$

where the $T_{n,k}^\star(m, m')$'s are defined in (5.26). We only study $T_{n,1}^\star(m, m')$ and conclude for $T_{n,2}^\star(m, m')$ analogously. As in the β-mixing framework, the study of $T_{n,1}^\star(m, m')$ consists in applying a concentration inequality to $\nu_{n,1}^\star(t)$ defined in (5.27) and considered as the sum of the p_n independent random variables $\nu_{q_n,\ell}^\star(t)$ defined as in (5.28). Once again, set $m^* = \max(m, m')$, and denote by $M_1^\star(m^*)$, $v^\star(m^*)$ and $H^\star(m^*)$ the terms such that $\sup_{t\in B_{m,m'}(0,1)} \| \nu_{q_n,\ell}^\star(u_t^*) \|_\infty \leq M_1^\star(m^*)$, $\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu_{q_n,\ell}^\star(u_t^*)) \leq v^\star(m)$ and lastly $\mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(u_t^*)|) \leq H^\star(m^*)$. According to Lemma 5.3, we take

$$(H^\star(m^*))^2 = \frac{2\Delta(m^*)}{n},\ M_1^\star(m^*) = \sqrt{\Delta(m^*)} \text{ and } v^\star(m^*) = \frac{C_{v^*}\sqrt{\Delta_2(m^*, f_Z)}}{2\pi q_n},$$

where $\Delta_2(m, f_Z)$ is defined in (5.29) and where

$$C_{v^*} = 2\Big[\mathbb{I}_{\delta>0} + \frac{\sqrt{2}\pi^{3/2}(2\pi)^{3/2}}{\sqrt{3}} \sum_{k\geq 1} \tau_1(k)\mathbb{I}_{\delta=0}\Big]. \tag{5.44}$$

From the definition of $T_{n,1}^\star(m, m')$, by taking $p_1(m, m') = 2(1 + 2\xi^2)(H^\star)^2(m^*)$, we get

$$\mathbb{E}(T_{n,1}^\star(m, m')) \leq \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(u_t^*) - 2(1 + 2\xi^2)(H^\star)^2(m^*)\Big]_+. \tag{5.45}$$

As in the β-mixing framework we take $\mathrm{pen}(m) = 16a\Delta(m)(1+2\xi^2)/n$ where ξ^2 is suitably chosen (see (5.45)). Set m_2 and m_3 as defined in Lemma 5.3, and set m_1 such that for $m^* \geq m_1$ (5.15) holds. Take $m_0 = m_1 \vee m_2 \vee m_3$ and $K' = K_1\lambda_1(f_\varepsilon, \kappa_0')/(C_{v^*}\lambda_2^\star(f_\varepsilon, \kappa_0))$. The end of the proof is the same as in β-mixing framework, up to possible multiplicative constants.□

5.5. **Technical lemmas.**

Lemma 5.1.

$$\| \sum_{j\in\mathbb{Z}} |u_{\varphi_{m,j}}^*|^2 \|_\infty \leq \Delta(m). \tag{5.46}$$

The proof of Lemma 5.1 can be found in Comte *et al.* (2006).

Lemma 5.2. *Assume that* $\sum_{k\geq 1}\beta_1(k) < +\infty$. *Then we have*

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu_{q_n,\ell}^\star(u_t^*) \|_\infty \leq \sqrt{\Delta(m^*)} \tag{5.47}$$

Moreover, there exist m_2 and m_3 such that

$$\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u^*_t)|] \leq \sqrt{2\Delta(m^*)/n} \text{ for } m^* \geq m_2,$$

$$\text{and} \quad \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell}(u^*_t)) \leq 2\sqrt{\Delta_2(m^*, f_Z)}/(2\pi q_n) \text{ for } m^* \geq m_3,$$

where $\Delta(m)$ and $\Delta_2(m, f_Z)$ are defined by (2.5) and (5.29).

Proof of Lemma 5.2. Arguing as in Lemma 5.1 and by using Cauchy-Schwartz Inequality and Parseval formula, we obtain that the first term $\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u^*_t) \|_\infty$ is bounded by

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u^*_t) \|_\infty \leq \sqrt{\sum_{j\in\mathbb{Z}} \int \left| \frac{\varphi^*_{m^*,j}(x)}{f^*_\varepsilon(x)} \right|^2 dx} = \sqrt{\Delta(m^*)}.$$

Next

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] &= \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\frac{1}{p_n q_n} \sum_{\ell=1}^{p_n} \sum_{i=1}^{q_n} u^*_t(Z^\star_{2\ell q_n+i}) - \langle t, g\rangle\Big|\Big] \\ &\leq \sqrt{\sum_{j\in\mathbb{Z}} \mathrm{Var}(\nu^\star_{n,1}(u^*_{\varphi_{m^*,j}}))}.\end{aligned}$$

By using (5.10) we obtain

$$\begin{aligned}\sqrt{\sum_{j\in\mathbb{Z}} \mathrm{Var}(\nu^\star_{n,1}(u^*_{\varphi_{m^*,j}}))} &= \sqrt{\sum_{j\in\mathbb{Z}} \frac{1}{p_n^2} \sum_{\ell=1}^{p_n} \mathrm{Var}\left(\nu^\star_{q_n,\ell}(u^*_{\varphi_{m^*,j}})\right)} = \sqrt{\sum_{j\in\mathbb{Z}} \frac{1}{p_n^2} \sum_{\ell=1}^{p_n} \mathrm{Var}\left(\nu_{q_n,\ell}(u^*_{\varphi_{m^*,j}})\right)} \\ &= \sqrt{\sum_{j\in\mathbb{Z}} \frac{1}{p_n} \mathrm{Var}\left(\nu_{q_n,1}(u^*_{\varphi_{m^*,j}})\right)} = \sqrt{\frac{1}{2\pi p_n} \int_{-\pi m^*}^{\pi m^*} \frac{\mathbb{E}|\nu_{q_n,1}(e^{ix\cdot})|^2}{|f^*_\varepsilon(x)|^2} dx}.\end{aligned}$$

Now, according to (5.13) and (5.2)

$$\mathbb{E}|\nu_{q_n,1}(e^{ix\cdot})|^2 \leq \frac{1}{q_n} + \frac{2}{q_n} \sum_{k=1}^{n-1} \beta_1(k) |f^*_\varepsilon(x)|.$$

This implies that

$$\mathbb{E}^2\Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] \leq \frac{1}{p_n}\Big(\frac{1}{q_n}\Delta(m^*) + \frac{2}{q_n} \sum_{k=1}^{n-1} \beta_1(k) \Delta_{1/2}(m^*)\Big).$$

Since $2\sum_{k\geq 1} \beta_1(k)\Delta_{1/2}(m) \leq \Delta(m)$ for m large enough, we get that, for m^* large enough,

$$\mathbb{E}^2\Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] \leq 2\Delta(m^*)/n.$$

Now, for $t \in B_{m,m'}(0,1)$ we write

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n} u_t^*(Z^\star_{2\ell q_n+i})\Big) &= \mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n} u_t^*(Z_i)\Big)\\ &= \frac{1}{q_n^2}\Big[\sum_{k=1}^{q_n}\mathrm{Var}(u_t^*(Z_k)) + 2\sum_{1\le k<l\le q_n}\mathrm{Cov}(u_t^*(Z_k),u_t^*(Z_l))\Big].\end{aligned}$$

According to (5.9), (5.12) and (5.2) we have

$$\begin{aligned}|\mathrm{Cov}(u_t^*(Z_k),u_t^*(Z_l))| &= \Big|\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{\mathrm{Cov}(e^{ixZ_k},e^{iyZ_l})t^*(x)t^*(y)}{f_\varepsilon^*(x)f_\varepsilon^*(-y)}dxdy\Big|\\ &= \Big|\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{f_\varepsilon^*(-y)\mathrm{Cov}(e^{ixZ_k},e^{iyX_l})t^*(x)t^*(y)}{f_\varepsilon^*(x)f_\varepsilon^*(-y)}dxdy\Big|\\ &\le \int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{2\beta_1(k)|t^*(x)t^*(y)|}{|f_\varepsilon^*(x)|}dxdy.\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n} u_t^*(Z^\star_{2\ell q_n+i})\Big) &\le \frac{1}{q_n}\Big(\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{f_Z^*(u-v)t^*(u)t^*(-v)}{f_\varepsilon(u)f_\varepsilon(-v)}dudv\\ &\quad +2\sum_{k=1}^{q_n}\beta_1(k)\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{t^*(u)t^*(v)}{f_\varepsilon^*(v)}\Big|dudv\Big).\end{aligned}$$

Following Comte *et al.* (2006) and applying Parseval's formula, the first integral is less that $\sqrt{\Delta_2(m^*,f_Z)}/2\pi$. For the second one, write

$$\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{t^*(u)t^*(v)}{f_\varepsilon^*(v)}\Big|dudv \le \sqrt{2\pi m^*}\|t^*\|\sqrt{\int |t^*(v)|^2dv\int_{-\pi m^*}^{\pi m^*}\frac{dv}{|f_\varepsilon^*(v)|^2}},$$

that is

$$\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{t^*(u)t^*(v)}{f_\varepsilon^*(v)}\Big|dudv \le (2\pi)^2\sqrt{m^*\Delta(m^*)}.$$

Using that $\gamma > 1/2$ if $\delta = 0$, we get that $\sqrt{m^*\Delta(m^*)} = o_m(\sqrt{\Delta_2(m^*,f_Z)})$ and hence the result follows for m large enough. □

Lemma 5.3. *Assume that* $\sum_{k\ge 1}\tau_1(k) < +\infty$. *Assume either that*

(1) $\delta = 0$, $\gamma \ge 3/2$ *in (3.9)*
(2) *or* $\delta > 0$ *in (3.9).*

Then we have

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u_t^*) \|_\infty \le \sqrt{\Delta(m^*)} \tag{5.48}$$

Moreover, there exist m_2 *and* m_3 *such that*

$$\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1}(u_t^*)|] \le \sqrt{2\Delta(m^*)/n} \text{ for } m^*\ge m_2,$$

$$\text{and}\quad \sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(\nu^\star_{q_n,\ell}(u_t^*)) \le C_{v^*}\sqrt{\Delta_2(m^*,f_Z)}/(2\pi q_n) \text{ for } m^*\ge m_3,$$

where $\Delta(m)$ and $\Delta_2(m, f_Z)$ are defined by (2.5) and (5.29) and where C_{v^*} is defined in (5.44).

Proof of Lemma 5.3. The proof of (5.48) is the same as the proof of (5.47). Next, again as for the proof of Lemma 5.2

$$\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}\Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] \quad \leq \quad \sqrt{\sum_{j\in\mathbb{Z}}\mathrm{Var}(\nu^\star_{n,1}(u^*_{\varphi_{m^*,j}}))}$$

with

$$\sqrt{\sum_{j\in\mathbb{Z}}\mathrm{Var}(\nu^\star_{n,1}(u^*_{\varphi_{m^*,j}}))} \quad = \quad \sqrt{\frac{1}{2\pi p_n}\int_{-\pi m^*}^{\pi m^*}\frac{\mathbb{E}|\nu_{q_n,1}(e^{ix\cdot})|^2}{|f^*_\varepsilon(x)|^2}dx}.$$

Now, according to (5.13) and (5.3)

$$\mathbb{E}|\nu_{q_n,1}(e^{ix\cdot})|^2 \leq \frac{1}{q_n} + \frac{1}{q_n}\sum_{k=1}^{n-1}\tau_1(k)|x||f^*_\varepsilon(x)|.$$

This implies that

$$\mathbb{E}^2\Big[\sup_{t\in B_{m,m'}(0,1)}\Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] \quad \leq \quad \frac{1}{p_n}\Big(\frac{1}{q_n}\Delta(m^*) + \frac{2\pi}{q_n}\sum_{k=1}^{n-1}\tau_1(k)m\Delta_{1/2}(m^*)\Big).$$

Since $2\pi\sum_{k\geq 1}\tau_1(k)m\Delta_{1/2}(m) \leq \Delta(m)$ for m large enough, we get that for m^* large enough

$$\mathbb{E}^2\Big[\sup_{t\in B_{m,m'}(0,1)}\Big|\nu^\star_{n,1}(u^*_t)\Big|\Big] \leq 2\Delta(m^*)/n.$$

Now, for $t \in B_{m,m'}(0,1)$ we write

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n}u^*_t(Z^\star_{2\ell q_n+i})\Big) \quad &= \quad \mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n}u^*_t(Z_i)\Big)\\ &= \quad \frac{1}{q_n^2}\Big[\sum_{k=1}^{q_n}\mathrm{Var}(u^*_t(Z_k)) + 2\sum_{1\leq k<l\leq q_n}\mathrm{Cov}(u^*_t(Z_k), u^*_t(Z_l))\Big].\end{aligned}$$

According to (5.9), (5.12) and (5.3) and by applying the same arguments as for the proof of Lemma 5.2 we have

$$\begin{aligned}|\mathrm{Cov}(u^*_t(Z_k), u^*_t(Z_l))| \quad &= \quad \Big|\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{f^*_\varepsilon(-y)\mathrm{Cov}(e^{ixZ_k}, e^{iyX_l})t^*(x)t^*(y)}{f^*_\varepsilon(x)f^*_\varepsilon(-y)}dxdy\Big|\\ &\leq \quad \int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{|y|\tau_1(k)|t^*(x)t^*(y)|}{|f^*_\varepsilon(x)|}dxdy.\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{q_n}\sum_{i=1}^{q_n}u^*_t(Z^\star_{2\ell q_n+i})\Big) \quad \leq \quad &\frac{1}{q_n}\Big(\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\frac{f^*_Z(u-v)t^*(u)t^*(-v)}{f_\varepsilon(u)f_\varepsilon(-v)}dudv\\ &+2\sum_{k=1}^{q_n}\tau_1(k)\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{ut^*(u)t^*(v)}{f^*_\varepsilon(v)}\Big|dudv\Big).\end{aligned}$$

Once again the first integral is less that $\sqrt{\Delta_2(m^*, f_Z)}/2\pi$. For the second one, write

$$\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{ut^*(u)t^*(v)}{f_\varepsilon^*(v)}\Big|dudv \leq \frac{\sqrt{2}\pi^{3/2}}{\sqrt{3}}(m^*)^{3/2}\|t^*\|\sqrt{\int |t^*(v)|^2dv\int_{-\pi m^*}^{\pi m^*}\frac{dv}{|f_\varepsilon^*(v)|^2}},$$

that is

$$\int_{-\pi m^*}^{\pi m^*}\int_{-\pi m^*}^{\pi m^*}\Big|\frac{t^*(u)t^*(v)}{f_\varepsilon^*(v)}\Big|dudv \leq \frac{\sqrt{2}\pi^{3/2}}{\sqrt{3}}(2\pi)^{3/2}\sqrt{(m^*)^3\Delta(m^*)}.$$

If $\delta > 0$, then $\sqrt{(m^*)^3\Delta(m^*)} = o_m\sqrt{\Delta_2(m^*, f_Z)}$. If $\gamma > 3/2$ and $\delta = 0$, we get that $\sqrt{(m^*)^3\Delta(m^*)} = o_m\sqrt{\Delta_2(m^*, f_Z)}$. Lastly, if $\gamma = 3/2$ and $\delta = 0$, we get that $\sqrt{(m^*)^3\Delta(m^*)} \leq \sqrt{\Delta_2(m^*, f_Z)}$ and the result follows for m large enough. □

Lemma 5.4. *Let $Y_1, \ldots, Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\xi^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2-2(1+2\xi^2)H^2\Big]_+ \leq \frac{2}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}} + \frac{49M_1^2}{4K_1n^2C^2(\xi^2)}e^{-\frac{2\sqrt{2}K_1C(\xi^2)\xi}{7}\frac{nH}{M_1}}\right),$$

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

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

This inequality comes from a concentration Inequality in Klein and Rio (2005) and arguments that can be found in Birgé and Massart (1998) (see the proof of Corollary 2 page 354). Usual density arguments show that this result can be applied to the class of functions $\mathcal{F} = B_{m,m'}(0,1)$.

5.6. **Proof of Proposition 4.1.** We describe here a general method to handle the models (4.1) and (4.2) and prove the following result that implies Proposition 4.1 (see Ango Nzé and Doukhan (2004) and Doukhan *et al.* (2006) for related results).

Proposition 5.1. *Let Y_t and σ_t satisfy either (4.1) or (4.2). For Model (4.1), let $(\eta'_t)_{t\in\mathbb{Z}}$ be an independent copy of $(\eta_t)_{t\in\mathbb{Z}}$, and for $t > 0$, let $\sigma_t^* = f(\eta_{t-1}, \ldots, \eta_1, \eta'_0, \eta'_{-1}, \ldots)$. For Model (4.2), let σ_0^* be a copy of σ_0 independent of $(\sigma_0, \eta_t)_{t\in\mathbb{Z}}$, and for $t > 0$ let $\sigma_t^* = f(\sigma_{t-1}^*, \eta_{t-1})$. Let δ_n be a non increasing sequence such that*

$$2\mathbb{E}(|\sigma_n^2 - (\sigma_n^*)^2|) \leq \delta_n\,. \tag{5.49}$$

Then

1. *The process $((Y_t^2, \sigma_t^2))_{t\geq 0}$ is τ-dependent with $\tau_\infty(n) \leq \delta_n$.*
2. *Assume that Y_0^2, σ_0^2 have densities satisfying $\max(f_{\sigma^2}(x), f_{Y^2}(x)) \leq C|\ln(x)|^\alpha x^{-\rho}$ in a neighborhood of 0, for some $\alpha \geq 0$ and $0 \leq \rho < 1$. The process $((X_t, Z_t))_{t\geq 0}$ is τ-dependent with $\tau_\infty(n) = O((\delta_n)^{(1-\rho)/(2-\rho)}|\ln(\delta_n)|^{(1+\alpha)/(2-\rho)})$.*

Consider Model (4.4) with $\mathbb{E}(\eta_0^2) = 1$*, and assume that* $c = \sum_{j\geq 1} a_j < 1$*. Let then* $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ *be the unique strictly stationary solution of the form (4.1). Then (5.49) holds with*

$$\delta_n = O\Big(\inf_{1\leq k\leq n}\Big\{c^{n/k} + \sum_{i=k+1}^{\infty} a_i\Big\}\Big).$$

Let us first explain how Proposition 5.1 implies Proposition 4.1. First, if σ_0^2 and η_0^2 have bounded densities, then $f_{Y^2}(x)) \leq C|\ln(x)|$ in a neighborhood of 0, so that Proposition 4.1(2) holds with $\rho = 0$ and $\alpha = 1$.

Under the assumptions of Proposition 5.1(2), we obtain straightforwardly the rates given in the first two cases for Model (4.4) and in the third case, the general rate $\tau_\infty(n) = O(n^{-b(1-\rho)/(2-\rho)}(\ln(n))^{(b+2)(1+\alpha)/2})$. Taking here $\rho = 0$ and $\alpha = 1$ gives the result.

For Model (4.2), if there exists $\kappa < 1$ such that (4.8) is satisfied, then one can take $\delta_n = 4\mathbb{E}(\sigma_0^2)\kappa^n$. Hence, under the assumptions of Proposition 4.1(2), $((X_t, Z_t))_{t>0}$ is geometrically τ dependent, and substituting δ_n gives the order of $\tau_\infty(n)$.

Proof of Proposition 5.1. To prove (1), let for $t > 0$, $Y_t^* = \eta_t\sigma_t^*$. Note that the sequence $((Y_t^*, \sigma_t^*))_{t\geq 1}$ is distributed as $((Y_t, \sigma_t))_{t\geq 1}$ and independent of $\mathcal{M}_i = \sigma(\sigma_j, Y_j, 0 \leq j \leq i)$. Hence, by the coupling properties of τ (see (5.1)), we have that, for $n + i \leq i_1 < \cdots < i_l$,

$$\tau(\mathcal{M}_i, (Y_{i_1}^2, \sigma_{i_1}^2), \ldots, (Y_{i_l}^2, \sigma_{i_l}^2)) \leq \frac{1}{l}\sum_{j=1}^{l} \|(Y_{i_j}^2, \sigma_{i_j}^2) - ((Y_{i_j}^*)^2, (\sigma_{i_j}^*))^2\|_{\mathbb{R}^2} \leq \delta_n\,,$$

and (1) follows.

To prove (2), define the function $f_\epsilon(x) = \ln(x)\mathbb{1}_{x>\epsilon} + 2\ln(\epsilon)\mathbb{1}_{x\leq\epsilon}$ and the function $g_\epsilon(x) = \ln(x) - f_\epsilon(x)$. Clearly, for any $\epsilon > 0$ and any $n + i \leq i_1 < \ldots < i_l$, we have

$$\begin{aligned}(5.50)\quad \tau(\mathcal{M}_i, (Z_{i_1}, X_{i_1}), \ldots, (Z_{i_l}, X_{i_l})) &\leq 2\mathbb{E}(|g_\epsilon(Y_0^2)| + |g_\epsilon(\sigma_0^2)|)\\ &\quad + \tau(\mathcal{M}_i, (f_\epsilon(Y_{i_1}^2), f_\epsilon(\sigma_{i_1}^2)), \ldots, (f_\epsilon(Y_{i_l}^2), f_\epsilon(\sigma_{i_l}^2)))\end{aligned}$$

For $0 < \epsilon < 1$, the function f_ϵ is $1/\epsilon$-Lipschitz. Hence, applying (1),

$$\tau(\mathcal{M}_i, (f_\epsilon(Y_{i_1}^2), f_\epsilon(\sigma_{i_1}^2)), \ldots, (f_\epsilon(Y_{i_l}^2), f_\epsilon(\sigma_{i_l}^2))) \leq \frac{\delta_n}{\epsilon}\,.$$

Since $\max(f_{\sigma^2}(x), f_{Y^2}(x)) \leq C|\ln(x)|^\alpha x^{-\rho}$ in a neighborhood of 0, we infer that for small enough ϵ,

$$\mathbb{E}(|g_\epsilon(Y_0^2)| + |g_\epsilon(\sigma_0^2)|) \leq K_1\epsilon^{1-\rho}|\ln(\epsilon)|^{1+\alpha}\,,$$

for K_1 a positive constant. From (5.50), we infer that there exists a positive constant K_2 such that, for small enough ϵ,

$$\tau(\mathcal{M}_i, (Z_{i_1}, X_{i_1}), \ldots, (Z_{i_l}, X_{i_l})) \leq K_2\Big(\frac{\delta_n}{\epsilon} + \epsilon^{1-\rho}|\ln(\epsilon)|^{1+\alpha}\Big).$$

The result follows by taking $\epsilon = (\delta_n)^{1/(2-\rho)}|\ln(\delta_n)|^{-(1+\alpha)/(2-\rho)}$.

Now, we go back to the model (4.4). If $\sum_{j=1}^{\infty} a_j < 1$, the unique stationary solution to (4.4) is given by Giraitis *et al.* (2000):

$$\sigma_t^2 = a + a \sum_{\ell=1}^{\infty} \sum_{j_1,\dots,j_l=1}^{\infty} a_{j_1} \dots a_{j_l} \eta_{t-j_1}^2 \dots \eta_{t-(j_1+\dots+j_l)}^2.$$

for any $1 \le k \le n$, let

$$\sigma_t^2(k,n) = a + a \sum_{\ell=1}^{[n/k]} \sum_{j_1,\dots,j_l=1}^{k} a_{j_1} \dots a_{j_l} \eta_{t-j_1}^2 \dots \eta_{t-(j_1+\dots+j_l)}^2.$$

Clearly

$$\mathbb{E}(|\sigma_n^2 - (\sigma_n^*)^2|) \le 2\mathbb{E}(|\sigma_0^2 - \sigma_0^2(k,n)|)\,.$$

Now

$$\mathbb{E}(|\sigma_0^2 - \sigma_0^2(k,n)|) \le \Big(\sum_{l=[n/k]+1}^{\infty} c^l + \sum_{l=1}^{\infty} c^{l-1} \sum_{j>k} a_j \Big)\,.$$

This being true for any $1 \le k \le n$, the proof of Proposition 5.1 is complete.

Acknowledgements.

We thank Yves Rozenholc for his precious help in adding a new type of noise in the deconvolution Matlab programs. We are also grateful to the Associate editor and the referees who helped us to improve the paper. Their questions and remarks were most helpful.

References

Ango Nzé, P. (1992). Critères d'ergodicité de quelques modèles à représentation markovienne. *C. R. Acad. Sci. Paris Sér. I Math. 315*(12), 1301–1304.

Ango Nzé, P. and P. Doukhan (2004). Weak dependence: models and applications to econometrics. *Econometric Theory 20*(6), 995–1045.

Berbee, H. (1979). *Random walks with Stationary Increments and Renewal Theory.* Mathematical Centre Tracts 112. Amsterdam: Mathematisch Centrum.

Birgé, L. and P. Massart (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli 4*(3), 329–375.

Birgé, L. and Y. Rozenholc (2006). How many bins should be put in a regular histogram. *ESAIM Probab. Stat. 10*, 24–45 (electronic).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics 31*(3), 307–327.

Bougerol, P. and N. Picard (1992a). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics 52*(1-2), 115–127.

Bougerol, P. and N. Picard (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Probab. 20*(4), 1714–1730.

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics 32*, 181–192.

Butucea, C. and C. Matias (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli 11*(2), 309–340.

Butucea, C. and A. B. Tsybakov (2005). Sharp optimality for density deconvolution with dominating bias. *Theor. Probab. Appl. To appear*.

Carrasco, M. and X. Chen (2000). β-mixing and moment properties of RCA models with application to GARCH(p, q). *C. R. Acad. Sci. Paris Sér. I Math. 331*(1), 85–90.

Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory 18*(1), 17–39.

Comte, F., Y. Rozenholc, and M.-L. Taupin (2005). Finite sample penalization in adaptive density deconvolution. *Preprint MAP5 2005-11. To appear in Journal of Statistical Computation and Simulation*.

Comte, F., Y. Rozenholc, and M. L. Taupin (2006). Penalized contrast estimator for adaptive density deconvolution. *The Canadian Journal of Statistics 34*(3), 431–452.

de Vries, C. G. (1991). On the relation between GARCH and stable processes. *J. Econometrics 48*(3), 313–324.

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

Doukhan, P. (1994). *Mixing: Properties and examples.* Lecture Notes in Statistics (Springer). New York: Springer-Verlag.

Doukhan, P., G. Teyssière, and P. Winant (2006). A LARCH(∞) vector valued process. *Lectures Notes in Statistics 187*, 245–258.

Duan, J.-C. (1997). Augmented GARCH(p, q) process and its diffusion limit. *J. Econometrics 79*(1), 97–127.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Stat. 19*(3), 1257–1272.

Franses, P. and D. van Dijk (2000). *Nonlinear Time Series Models in Empirical Finance.* Cambridge University Press. Cambridge.

Giraitis, L., P. Kokoszka, and R. Leipus (2000). Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory 16*(1), 3–22.

Klein, T. and E. Rio (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab. 33*(3), 1060–1077.

Lacour, C. (2006a). Adaptive estimation of the transition density of a particular hidden markov chain. *Preprint MAP5 2006-8*.

Lacour, C. (2006b). Rates of convergence for nonparametric deconvolution. *Comptes Rendus Acad. Sc. Ser. I Math. 342*(11), 877–882.

Lee, O. and D. W. Shin (2005). On stationarity and β-mixing property of certain nonlinear GARCH(p, q) models. *Statist. Probab. Lett. 73*(1), 25–35.

Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Statist. 32*(4), 439–449.

Meyer, Y. (1990). *Ondelettes et opérateurs I: Ondelettes.* Actualités Mathématiques. Paris: Hermann, éditeurs des Sciences et des Arts.

Pensky, M. and B. Vidakovic (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat. 27*(6), 2033–2053.

van Es, B., P. Spreij, and H. van Zanten (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat. 17*(2), 237–251.

Volkonskij, V. and Y. Rozanov (1960). Some limit theorems for random functions. I. *Theor. Probab. Appl. 4*, 178–197.

Zakoian, J.-M. (1993). Modèle autoregressif à un seuil. *Publ. Inst. Statist. Univ. Paris 37*(3-4), 85–113.