Adaptive density estimation for general ARCH models

Fabienne Comte, Jérôme Dedecker, Marie-Luce Taupin

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.

MSC 2000 Subject Classifications. 62G07 - 62M05. September 27, 2006
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: $((Y_t, \sigma_t))_{t\geq 0}$ 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 $(\eta_i)_{i\geq t}$.

The model 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 $Z_t = \ln(Y_t^2)_t$ 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 $f_Z = g * f_\varepsilon$.

In many papers dealing with ARCH models, ε_t is assumed to be Gaussian or the log of a squared Gaussian (when η_t is Gaussian, see van Es *et al.* (2005) or in slightly different

---

[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.

contexts van Es *et al.* (2003), Comte and Genon-Catalot (2005)). Our setting is more general since we consider various type of error densities. More precisely, we assume that f_ε belongs to some class of smooth functions described below: there exist nonnegative numbers κ_0, γ, μ, and δ such that the fourier transform f_ε^* of f_ε satisfies

$$\kappa_0(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\} \le |f_\varepsilon^*(x)| \le \kappa_0'(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\}. \tag{1.3}$$

Since f_ε is known, the constants μ, δ, κ_0, and γ defined in (1.3) are known. When $\delta = 0$ in (1.3), the errors are called "ordinary smooth" errors. When $\mu > 0$ and $\delta > 0$, they are called "super smooth". The standard examples for super smooth densities are Gaussian or Cauchy distributions (super smooth of order $\gamma = 0, \delta = 2$ and $\gamma = 0, \delta = 1$ respectively). When $\varepsilon_t = \ln(\eta_t^2)$ with $\eta_t \sim \mathcal{N}(0,1)$ as in van Es *et al.* (2003, 2005), then ε_t is super-smooth with $\delta = 1, \gamma = 0$ and $\mu = \pi/2$. An example of ordinary smooth density is the Laplace distribution, for which $\delta = \mu = 0$ and $\gamma = 2$.

In density deconvolution of i.i.d variables the X_t's and the ε_t's are i.i.d. and the sequences $(X_t)_{t\ge0}$ 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 setting of Model (1.2), the classical assumptions of independence between the processes $(X_t)_{t\ge0}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are no longer satisfied and the tools for deconvolution have to be revisited.

As in density deconvolution for i.i.d. variables, the slowest rates of convergence for estimating g are obtained for super smooth error densities. For instance, in the i.i.d case, 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. So, we describe the smoothness properties of g by the set

$$\mathcal{S}_{s,r,b}(C_1) = \left\{\psi \text{ such that } \int_{-\infty}^{+\infty} |\psi^*(x)|^2(x^2+1)^s\exp\{2b|x|^r\}dx \le C_1\right\} \tag{1.4}$$

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. When $r > 0, b > 0$ functions belonging to $\mathcal{S}_{s,r,b}(C_1)$ are infinitely many times differentiable.

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. In particular they reach the minimax rates of the i.i.d. case in all cases where they are known (see Fan (1991), Butucea (2004) and Butucea and Tsybakov (2005)). Next we prove that the MISE of our adaptive estimator is of the same order as the MISE of the best non-adaptive estimator, 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 consider various types of error density, and we do not make any assumption on the smoothness of g: this is the advantage of the adaptive procedure. We shall 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).

The paper is organized as follows. Many examples are described in Section 2, together with their dependence properties. The estimator is defined in Section 3. The MISE bounds are given in Section 4, and the proofs are given in Section 5.

2. The model and its dependence properties

2.1. **Models and examples.** A particular case of model (1.1) is

$$Y_t = \sigma_t \eta_t, \text{ with } \sigma_t = f(\eta_{t-1}, \eta_{t-2}, \ldots) \tag{2.1}$$

for some measurable function f. Another important case is

$$Y_t = \sigma_t \eta_t, \text{ with } \sigma_t = f(\sigma_{t-1}, \eta_{t-1}) \text{ and } \sigma_0 \text{ independent of } (\eta_t)_{t \geq 0}, \tag{2.2}$$

that is σ_t is a stationary Markov chain.

We begin with models satisfying a recursive equation, whose stationary solution satisfies (2.1). The original ARCH model as introduced by Engle (1982) was given by

$$Y_t = \sqrt{a + bY_{t-1}^2}\,\eta_t, \; a \geq 0, b \geq 0 \tag{2.3}$$

It has been generalized by Bollerslev (1986) with the class of GARCH(p, q) models defined by $Y_t = \sigma_t \eta_t$ and

$$\sigma_t^2 = a + \sum_{i=1}^{p} a_i Y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \tag{2.4}$$

where the coefficients $a, a_i, i = 1, \ldots, p$ and $b_j, j = 1, \ldots, q$ 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 (2.1).

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):

$$\sigma_t^2 = a + \sum_{j=1}^{\infty} a_j Y_{t-j}^2, \tag{2.5}$$

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 (2.5) of the form (2.1).

For the models satisfying (2.2), let us cite first the so-called augmented GARCH(1, 1) models introduced by Duan (1997):

$$\Lambda(\sigma_t^2) = c(\eta_{t-1})\Lambda(\sigma_{t-1}^2) + h(\eta_{t-1}), \tag{2.6}$$

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 (2.6), 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 (2.6) is the threshold ARCH model (see Zakoïan (1993)):

$$\sigma_t = a + b\sigma_{t-1}\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}>0\}} - c\sigma_{t-1}\eta_{t-1}\mathbf{1}_{\{\eta_{t-1}<0\}}, \ a, b, c > 0 \tag{2.7}$$

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

$$\sigma_t = f(\sigma_{t-1}\eta_{t-1}). \tag{2.8}$$

There exists a stationary solution to (2.8) provided that the density of η_0 is positive on a neighborhood of 0 and $\limsup_{|x|\to\infty} |f(x)/x| < 1$.

In the next section, we define the dependence coefficients that we shall use in this paper, and we give the dependence properties of the models (2.3)-(2.8) in terms of these coefficients.

2.2. **Measures of dependence.** 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 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

$$\begin{aligned} \beta(\mathcal{M}, \sigma(W)) &= \mathbb{E}\Big(\sup_{A\in\mathcal{B}(\mathcal{X})} |\mathbb{P}_{W|\mathcal{M}}(A) - \mathbb{P}_W(A)|\Big), \\ \text{and if } \mathbb{E}(\|W\|_{\mathbb{B}}) < \infty, \quad \tau(\mathcal{M}, W) &= \mathbb{E}\Big(\sup_{f\in\Lambda_1(\mathbb{B})} |\mathbb{P}_{W|\mathcal{M}}(f) - \mathbb{P}_W(f)|\Big). \end{aligned}$$

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

$$\beta_1(k) = \beta(\sigma(W_0), \sigma(W_k)), \qquad \text{and if } \mathbb{E}(\|W_0\|_{\mathbb{R}^2}) < \infty, \quad \tau_1(k) = \tau(\sigma(W_0), W_k). \tag{2.9}$$

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

$$\beta_\infty(k) = \sup_{i\geq 0}\sup_{l\geq 1} \{\beta(\mathcal{M}_i, \sigma(W_{i_1}, \ldots, W_{i_l})), i + k \leq i_1 < \cdots < i_l\}, \tag{2.10}$$

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

$$\tau_\infty(k) = \sup_{i\geq 0}\sup_{l\geq 1} \{\tau(\mathcal{M}_i, (W_{i_1}, \ldots, W_{i_l})), i + k \leq i_1 < \cdots < i_l\}. \tag{2.11}$$

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. 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 $k \geq 1$.

We now recall the coupling properties associated with the dependency 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

$$\beta(\mathcal{M}, \sigma(W)) = \mathbb{P}(W \neq W_1^\star) \quad \text{and} \quad \tau(\mathcal{M}, W) = \mathbb{E}(\|W - W_2^\star\|_{\mathbb{B}})\,. \tag{2.12}$$

The first equality in (2.12) 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 (2.12), 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

$$|\mathrm{Cov}(f(Y), h(X))| \leq 2\|f(Y)\|_{\infty,\mathbb{P}}\|h(X)\|_{\infty,\mathbb{P}}\, \beta(\sigma(X), \sigma(Y))\,. \tag{2.13}$$

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

$$|\mathrm{Cov}\ (f(Y), h(X))| \leq \|f(Y)\|_{\infty,\mathbb{P}}\mathrm{Lip}(h)\, \tau(\sigma(Y), X)\,. \tag{2.14}$$

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

$$|\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})| \leq 2\beta_1(k-1) \quad \text{and} \quad |\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})| \ \leq\ |x|\tau_1(k-1). \tag{2.15}$$

2.3. **Application to ARCH models.** For the models (1.1) and (1.2), the β-mixing coefficients of the process

$$(W_t)_{t\in\mathbb{Z}} = ((Z_t, X_t))_{t\in\mathbb{Z}} \tag{2.16}$$

are smaller than that of $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ (because of the inclusion of σ-algebras). 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 (2.3) is geometrically β-mixing as soon as $0 < b < 1$.
- The process $((Y_t, \sigma_t))_{t\in\mathbb{Z}}$ defined by Model (2.4) 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 (2.6) 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 (2.7) 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 (2.8) 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).

Concerning the τ-dependence, here is a general method to handle the models (2.1) and (2.2). The following Proposition will be proved in appendix (see Ango Nzé and Doukhan (2004) and Doukhan *et al.* (2006) for related results).

Proposition 2.1. *Let Y_t and σ_t satisfy either (2.1) or (2.2). For Model (2.1), let $(\eta'_t)_{t\in\mathbb{Z}}$ be an independent copy of $(\eta_t)_{t\in\mathbb{Z}}$, and for $t>0$, let $\sigma^*_t = f(\eta_{t-1},\dots,\eta_1,\eta'_0,\eta'_{-1},\dots)$. For Model (2.2), 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*

$$2\mathbb{E}(|\sigma_n^2-(\sigma_n^*)^2|)\le \delta_n\,. \tag{2.17}$$

Then

1. *The process $((Y_t^2,\sigma_t^2))_{t\ge 0}$ is τ-dependent with $\tau_\infty(n)\le\delta_n$.*
2. *Assume that Y_0^2, σ_0^2 have densities satisfying $\max(f_{\sigma^2}(x),f_{Y^2}(x))\le C|\ln(x)|^\alpha x^{-\rho}$ in a neighborhood of 0, for some $\alpha\ge 0$ and $0\le\rho<1$. The process $((X_t,Z_t))_{t\ge0}$ is τ-dependent with $\tau_\infty(n)=O((\delta_n)^{(1-\rho)/(2-\rho)}|\ln(\delta_n)|^{(1+\alpha)/(2-\rho)})$.*

Consider Model (2.5), and assume that $c=\sum_{j\ge1}a_j<1$. Let then $((Y_t,\sigma_t))_{t\in\mathbb{Z}}$ be the unique strictly stationary solution of the form (2.1). Then (2.17) holds with

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

Note that if σ_0^2 and η_0^2 have bounded densities, then $f_{Y^2}(x))\le C|\ln(x)|$ in a neighborhood of 0, so that Proposition 2.1(2) holds with $\rho=0$ and $\alpha=1$.

Under the assumptions of Proposition 2.1(2), we obtain for Model (2.5) the following rates for $((X_t,Z_t))_{t\ge0}$:

- If $a_j=0$, for $j\ge J$, then $((X_t,Z_t))_{t\ge0}$ 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 $\tau_\infty(n)=O(n^{-b(1-\rho)/(2-\rho)}(\ln(n))^{(b+2)(1+\alpha)/2})$.

For more general models than (2.5), we refer to Doukhan *et al.* (2006).

For Model (2.2), if there exists $\kappa<1$ such that

$$\mathbb{E}(|(f(x,\eta_0))^2-(f(y,\eta_0))^2|)\le\kappa|x^2-y^2|\,, \tag{2.18}$$

then one can take $\delta_n=4\mathbb{E}(\sigma_0^2)\kappa^n$. Hence, under the assumptions of Proposition 2.1(2), $((X_t,Z_t))_{t>0}$ is geometrically τ dependent. An example of Markov chain satisfying (2.18) is the autoregressive model $\sigma_t^2=h(\sigma_{t-1}^2)+r(\eta_{t-1})$ for some κ-lipschitz function h.

3. 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 $\|u\|_\infty=\sup_{x\in\mathbb{R}}|u(x)|$.

3.1. **The projection spaces.** Let $\varphi(x) = \sin(\pi x)/(\pi x)$. 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 $\mathbb{L}^2(\mathbb{R})$ (see e.g. Meyer (1990), p.22). Let us 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 $a_{m,j}(g) =< \varphi_{m,j}, g >$. To obtain representations having a finite number of "coordinates", we introduce

$$S_m^{(n)} = \overline{\text{span}}\,\{\varphi_{m,j}, |j| \le k_n\}$$

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

3.2. **Construction of the minimum contrast estimators.** We subsequently assume that

$$f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \text{ and is such that } \forall x \in \mathbb{R},\ f_\varepsilon^*(x) \neq 0. \tag{3.1}$$

Note that the square integrability of f_ε and (1.3) require that $\gamma > 1/2$ when $\delta = 0$. Under Condition (3.1) and for or t in $S_m^{(n)}$, we define the contrast function

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

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

$$\hat{g}_m^{(n)} = \arg\min_{t\in S_m^{(n)}} \gamma_n(t). \tag{3.2}$$

By using Parseval and inverse Fourier formulae we obtain that $\mathbb{E}\left[u_t^*(Z_i)\right] = \langle t, g\rangle$, so that $\mathbb{E}(\gamma_n(t)) = \|t - g\|^2 - \|g\|^2$ is minimal when $t = g$. This shows that $\gamma_n(t)$ suits well for the estimation of g. It is easy to see that

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

3.3. **Minimum penalized contrast estimator.** The minimum penalized estimator of g is defined as $\tilde{g} = \hat{g}^{(n)}_{\hat{m}_g}$ 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 model $S_{\hat{m}}^{(n)}$) involved in the estimators $\hat{g}_m^{(n)}$ given by (3.2), in order to mimic the oracle parameter

$$\breve{m}_g = \arg\min_m \mathbb{E} \parallel \hat{g}_m^{(n)} - g \parallel_2^2 . \tag{3.3}$$

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

$$\tilde{g} = \hat{g}_{\hat{m}}^{(n)} \text{ with } \hat{m} = \arg \min_{m \in \{1, \cdots, m_n\}} \left[\gamma_n(\hat{g}_m^{(n)}) + \text{ pen}(m) \right], \tag{3.4}$$

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

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|^2} dx. \tag{3.5}$$

The key point in the dependent context is to find a penalty function not depending on the dependency 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.

4. Density estimation bounds

¿From now on, the dependence coefficients are defined as in (2.9), (2.10) and (2.11) with $(W_t)_{t \in \mathbb{Z}} = ((Z_t, X_t))_{t \in \mathbb{Z}}$.

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

$$g \in \mathbb{L}_2(\mathbb{R}), \text{and there exists } M_2 > 0, \int x^2 g^2(x) dx \leq M_2 < \infty. \tag{4.1}$$

Assumption (4.1), which is due to the construction of the estimator, already appears in density deconvolution in the independent framework in Comte *et al.* (2005, 2006). It is important to note that Assumption (4.1) is very unrestrictive. In particular, all densities having tails of order $|x|^{-(s+1)}$ as x tends to infinity satisfy (4.1) only if $s > 1/2$. One can cite for instance the Cauchy distribution or all stable distributions with exponent $r > 1/2$ (see Devroye (1986)). The Lévy distribution, with exponent $r = 1/2$ does not satisfies (4.1).

Note that (4.1) is fulfilled if g is bounded by M_0 and $\mathbb{E}(X_1^2) \leq M_1 < +\infty$, with $M_2 = M_0 M_1$.

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

Proposition 4.1. *If (3.1) and (4.1) hold, then $\hat{g}_m^{(n)}$ defined by (3.2) satisfies*

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

where

$$R_m = \frac{1}{\pi} \sum_{k=2}^{n} \int_{-\pi m}^{\pi m} \left| \frac{\text{Cov}\left(e^{ixZ_1}, e^{ixX_k}\right)}{f_\varepsilon^*(-x)} \right| dx. \tag{4.2}$$

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

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

with β_1*,* τ_1 *defined by (2.9), and where*

$$\Delta_{1/2}(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|} dx. \tag{4.3}$$

This proposition requires several comments.

As usual, the order of the risk is given by a bias term $\| g_m - g \|^2 + m^2(M_2+1)/k_n$ and a variance term $2\Delta(m)/n + 2R_m/n$. As in density deconvolution for i.i.d. variables, 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, as in the i.i.d. case, 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 $\Delta_{1/2}(m)/n$, specific to ARCH models.

The bias term is the sum of the usual bias term $\| g_m - g \|^2$, depending on the smoothness properties of g, and on an additional term $m^2(M_2+1)/k_n$. With a suitable choice of k_n, not depending on g, this last term is negligible with respect to the variance term.

Concerning the main variance term, $\Delta(m)$ given by (3.5) 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

$$\lambda_1(f_\varepsilon, \kappa_0) = \frac{1}{\kappa_0^2 \pi R(\mu,\delta)}, \text{ where } R(\mu,\delta) = \mathbb{1}_{\{\delta=0\}} + 2\mu\delta \mathbb{1}_{\{\delta>0\}}. \tag{4.4}$$

The rates resulting from Proposition 4.1 under (1.3) and (1.4) are given in the following proposition.

Corollary 4.1. *Assume that (1.3), (3.1), and (4.1) hold, that* g *belongs to* $\mathcal{S}_{s,r,b}(C_1)$ *defined by (1.4), and that* $k_n \geq n$*. Assume either that*

1. $\sum_{k\geq 1} \beta_1(k) < +\infty$
2. *or* $\delta = 0$*,* $\gamma > 1$ *in (1.3) and* $\sum_{k\geq 1} \tau_1(k) < +\infty$
3. *or* $\delta > 0$ *in (1.3) and* $\sum_{k\geq 1} \tau_1(k) < +\infty$*.*

Then $\hat{g}_m^{(n)}$ *defined by (3.2) satisfies*

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \frac{C_1}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\} + \frac{2\lambda_1(f_\varepsilon,\kappa_0)\Gamma(m)}{n} + \frac{C_2}{n}\Gamma(m)o_m(1), \tag{4.5}$$

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

If $\gamma = 1$ when $\delta = 0$, then the bound 4.5 becomes

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \frac{C_1}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\} + \frac{(2+C_2)\lambda_1(f_\varepsilon,\kappa_0)\Gamma(m)}{n}, \tag{4.6}$$

with C_2 depending on $\sum_{k\geq 1}\beta_1(k)$ (respectively on $\sum_{k\geq 1}\tau_1(k)$).

The rate of convergence of $\hat{g}_{\breve{m}}^{(n)}$ is the same as the rate for density deconvolution for i.i.d. sequences. Our context here encompasses the particular case considered by van Es *et al.* (2005).

Table 1 below gives a summary of these rates obtained when minimizing the right hand of (4.5). The $\breve{m}_g$ denotes the corresponding minimizer (see 3.3).

TABLE 1. Choice of $\breve{m}_g$ and corresponding rates under Assumptions (1.3) and (1.4).

		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 $= O\left(\frac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$	$\breve{m}_g$ solution of $\breve{m}_g^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m}_g)^\delta + 2b\pi^r\breve{m}_g^r\} = O(n)$

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 $\hat{g}_{\breve{m}_g}^{(n)}$ is not easy to give explicitly and depends on the ratio r/δ. If r/δ or δ/r belongs to $]k/(k+1);(k+1)/(k+2)]$ with k integer, the rate of convergence can be expressed as a function of k. We refer to Comte *et al.* (2006) for further discussions about those rates. We refer to Lacour (2006) for explicit formulae for the rates in the special case $r > 0$ and $\delta > 0$.

4.2. **Adaptive bound.** Theorem 4.1 below gives a general bound which holds under weak dependency conditions, for ε being either ordinary or super smooth.

For $a > 1$, let pen(m) be defined by

$$\text{pen}(m) = \begin{cases} 192a\dfrac{\Delta(m)}{n} & \text{if } 0 \leq \delta < 1/3, \\ 64a\lambda_3\dfrac{\Delta(m)\, m^{\min((3\delta/2-1/2)_+,\delta))}}{n} & \text{if } \delta \geq 1/3, \end{cases} \tag{4.7}$$

where $\Delta(m)$ is defined by (3.5). The constant $\lambda_1(f_\varepsilon,\kappa_0)$ is defined in (4.4) and

$$\lambda_3 = 1 + \frac{32\mu\pi^\delta}{\lambda_1(f_\varepsilon,\kappa_0')}\left((\sqrt{2}+8)\|f_\varepsilon\|_\infty\kappa_0^{-1}\sqrt{\lambda_1(f_\varepsilon,\kappa_0)}\,\mathbb{I}_{0\leq\delta\leq 1} + 2\lambda_1(f_\varepsilon,\kappa_0)\mathbb{I}_{\delta>1}\right). \tag{4.8}$$

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 $(X_t)_{t\geq 1}$ and $(\varepsilon_t)_{t\geq 1}$ 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

$$\pi m_n \leq \begin{cases} n^{1/(2\gamma+1)} & \text{if } \delta = 0 \\ \left[\dfrac{\ln(n)}{2\mu} + \dfrac{2\gamma+1-\delta}{2\delta\mu}\ln\left(\dfrac{\ln(n)}{2\mu}\right)\right]^{1/\delta} & \text{if } \delta > 0. \end{cases} \tag{4.9}$$

Subsequently we set

$$C_a = \max(\kappa_a^2, 2\kappa_a) \text{ where } \kappa_a = (a+1)/(a-1). \tag{4.10}$$

Theorem 4.1. *Assume that f_ε satisfies (1.3) and 3.1, that g satisfies (4.1), and that m_n satisfies (4.9). Let* $\text{pen}(m)$ *be defined by (4.7). Consider the collection of estimators $\hat{g}_m^{(n)}$ defined by (3.2) with $k_n \geq n$ and $1 \leq m \leq m_n$. Let β_∞ and τ_∞ be defined as in (2.10) and (2.11) respectively. Assume either that*

1. *$\beta_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3$*
2. *or $\delta = 0$, $\gamma \geq 3/2$ in (1.3) and $\tau_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3 + 2/(1+2\gamma)$*
3. *or $\delta > 0$ in (1.3) and $\tau_\infty(k) = O(k^{-(1+\theta)})$ for some $\theta > 3$.*

Then the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by (3.4) satisfies

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m\in\{1,\cdots,m_n\}} \left[\|g - g_m\|^2 + \text{pen}(m) + \frac{m^2(M_2+1)}{n}\right] + \frac{\overline{C}}{n}, \tag{4.11}$$

where C_a is defined in (4.10) and $\overline{C}$ is a constant depending on f_ε, a, and $\sum_{k\geq 1}\beta_\infty(k)$ (respectively on $\sum_{k\geq 1}\tau_\infty(k)$).

Remark 4.1. In case (2), when $\delta = 0$ in (1.3), the condition on θ is weaker as γ increases and f_ε gets smoother.

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.

The main result in Theorem 4.1 shows that the MISE of $\tilde{g}$ automatically achieves the best squared-bias variance compromise (possibly up to some logarithmic factor). Consequently, it achieves the best rate among the rates of the $\hat{g}_m^{(n)}$, even from a non-asymptotical point of view. This last point is of most importance since the m 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). Another important point is that, if we consider the asymptotic trade-off, then the rates given in

Table 1 are automatically reached in most cases by the adaptive estimator $\tilde{g}$. Only in the case $\delta > 1/3$ and $r > 0$, a loss may occur in the rate of $\tilde{g}$. This comes from the additional power of m in the penalty for $\delta \geq 1/3$ with respect to the variance order $\Delta(m)$. Nevertheless, the resulting loss in the rate has an order which is negligible compared to the main order rate.

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

5. Proofs

5.1. **Proof of Proposition 4.1.** The proof of Proposition 4.1 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 3.2. According to (3.2), 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,

$$\nu_{n,Z}(u_t^*) = \frac{1}{n}\sum_{i=1}^n \left[u_t^*(Z_i) - \langle t, g\rangle\right]. \tag{5.1}$$

Since

$$\gamma_n(t) - \gamma_n(s) = \|t-g\|^2 - \|s-g\|^2 - 2\nu_{n,Z}(u_{t-s}^*), \tag{5.2}$$

we infer that

$$\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m^{(n)}\|^2 + 2\nu_{n,Z}\left(u^*_{\hat{g}_m^{(n)} - g_m^{(n)}}\right). \tag{5.3}$$

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),

$$\|g - g_m^{(n)}\|^2 = \| g - g_m \|^2 + \|g_m - g_m^{(n)}\|^2 \leq \| g - g_m \|^2 + \frac{(\pi m)^2(M_2+1)}{k_n}. \tag{5.4}$$

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

$$\nu_{n,Z}(f^*) = \int \nu_{n,Z}(e^{ix\cdot})f(x)dx. \tag{5.5}$$

Now, we use (5.5) and apply Parseval's formula to obtain

$$\begin{aligned}\mathbb{E}\Big(\sum_{j\in\mathbb{Z}}(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2\Big) &= \frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}}\mathbb{E}\Big(\int \frac{\varphi^*_{m,j}(-x)}{f_\varepsilon^*(x)}\nu_{n,Z}(e^{ix\cdot})dx\Big)^2 \\ &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2}{|f_\varepsilon^*(x)|^2}dx. \end{aligned} \tag{5.6}$$

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

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 = \mathrm{Var}|\nu_{n,Z}(e^{ix\cdot})| = \frac{1}{n}\mathrm{Var}(e^{ixZ_1}) + \frac{1}{n^2}\sum_{1\leq k\neq l\leq n}\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}). \tag{5.7}$$

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 $k < l$,

$$\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}) = f_\varepsilon^*(x)\mathrm{Cov}(e^{ixZ_k}, e^{ixX_l}). \tag{5.8}$$

¿From (5.7) and the stationarity of $(X_i)_{i\geq 1}$, we obtain that

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 \leq \frac{1}{n} + \frac{2}{n}\sum_{k=2}^{n}\left|\mathrm{Cov}(e^{ixZ_1}, e^{ixX_k})\right| |f_\varepsilon^*(x)|. \tag{5.9}$$

The first part of Proposition 4.1 follows from the stationarity of the X_i's, and from (5.3), (5.4), (5.6) and (5.9).

The proof of $R_m \leq \min(R_{m,\beta}, R_{m,\tau})$, where $R_{m,\beta}$ and $R_{m,\tau}$ are defined in Proposition 4.1, comes from the inequalities (2.15) in Section 2.2. Hence we get the result.□

5.2. **Proof of Corollary 4.1.** According to Butucea and Tsybakov (2005), under (1.3), we have

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

$$\Gamma(m) = (1 + (\pi m)^2)^\gamma (\pi m)^{1-\delta} \exp\left\{2\mu(\pi m)^\delta\right\}, \tag{5.10}$$

where λ_1 is defined in (4.4). In the same way

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

where

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

It is easy to see that $\Delta_{1/2}(m) \leq \sqrt{m\Delta(m)}$ 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 $m \geq m_1$ we have

$$0.5\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m) \leq \Delta(m) \leq 2\lambda_1(f_\varepsilon, \kappa_0)\Gamma(m), \tag{5.11}$$

and

$$0.5\overline{\lambda_1}(f_\varepsilon, \kappa_0')\overline{\Gamma}(m) \leq \Delta_{1/2}(m) \leq 2\overline{\lambda_1}(f_\varepsilon, \kappa_0)\overline{\Gamma}(m). \tag{5.12}$$

If $\sum_{k\geq 1}\beta_1(k) < +\infty$, (1.3) and (4.1) 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 (1.3) and (4.1) 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 $g_m^*=g^*1\!\!\mathrm{I}_{[-m\pi,m\pi]}$ 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 (1.4), then

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

The corollary is proved. □

5.3. **Proof of Theorem 4.1.** 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 (5.2) 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 (5.1). 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

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

Proof in the β-mixing case.

We use the coupling methods recalled in Section 2.2 to build approximating variables for the $W_i=(Z_i,X_i)$'s. More precisely, we build variables $W_i^\star$ such that if $n=2p_nq_n+r_n$, $0\leq r_n<q_n$, 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.
- $\mathbb{P}(E_\ell\neq E_\ell^*)\leq\beta_\infty(q_n)$ and $\mathbb{P}(F_\ell\neq F_\ell^*)\leq\beta_\infty(q_n)$,

- $E^\star_\ell$ and $\mathcal{M}_0 \vee \sigma(E_0, E_1, ..., E_{\ell-1}, E^\star_0, E^\star_1, \cdots, E^\star_{\ell-1})$ are independent, and therefore independent of $\mathcal{M}_{(\ell-1)q_n}$ and the same holds for the blocks $F^\star_\ell$.

For the sake of simplicity we assume that $r_n = 0$. We denote by $(Z^\star_i, X^\star_i) = W^\star_i$ the new couple of variables. We start from

$$\| \tilde g - g \|^2 \leq \kappa_a^2 \| g_m^{(n)} - g \|^2 + a\kappa_a \sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*)|^2 + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat m)), \tag{5.13}$$

where κ_a is defined in (4.10). Using the notation (5.1), we denote by $\nu^\star_{n,Z}(u_t^*)$ the empirical contrast computed on the $Z_i^\star$. Then we write

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

Set

$$T_n^\star(m, m') := \Big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,Z}(t)|^2 - p(m,m')\Big]_+. \tag{5.14}$$

Hence

$$\begin{aligned}\|\tilde g - g\|^2 &\leq \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2a\kappa_a T_n^\star(m,\hat m) + \kappa_a\left(2ap(m,\hat m) + \mathrm{pen}(m) - \mathrm{pen}(\hat m)\right) \\ &\quad + 2a\kappa_a \sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu^\star_{n,Z}(u_t^*)|^2 \\ &\leq \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \mathrm{pen}(m) + 2a\kappa_a \sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu^\star_{n,Z}(u_t^*)|^2 \\ &\quad + 2a\kappa_a T_n^\star(m,\hat m)\end{aligned} \tag{5.15}$$

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

$$2ap(m,m') \leq \mathrm{pen}(m) + \mathrm{pen}(m'). \tag{5.16}$$

Now write

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

Consequently,

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*) - \nu^\star_{n,Z}(u_t^*)|^2\Big] \leq \int_{-\pi m_n}^{\pi m_n} \mathbb{E}[|\nu_{n,Z}(e^{ix\cdot}) - \nu^\star_{n,Z}(e^{ix\cdot})|^2] \frac{1}{|f_\varepsilon^*(x)|^2} dx. \tag{5.17}$$

Since

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

we obtain that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*)-\nu^\star_{n,Z}(u_t^*)|^2\Big] \;\leq\; 4\beta_\infty(q_n)\Delta(m_n). \tag{5.18}$$

By gathering (5.15) and (5.18) we get

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

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

$$\mathbb{E}\|g-\tilde g\|^2\leq C_a\left[\|g-g_m^{(n)}\|^2+\text{pen}(m)\right]+2a\kappa_a(C_1+C_2)/n, \tag{5.19}$$

provided that

$$\Delta(m_n)\beta_\infty(q_n)\leq C_1/n \quad\text{ and }\quad \sum_{m'=1}^{m_n}\mathbb{E}(T_n^\star(m,m'))\leq C_2/n. \tag{5.20}$$

Using (5.11), we conclude that the first part of (5.20) is fulfilled as soon as

$${m_n}^{2\gamma+1-\delta}\exp\{2\mu\pi^\delta {m_n}^\delta\}\beta_\infty(q_n)\leq C_1'/n. \tag{5.21}$$

In order to ensure that our estimators converge, we only consider models with bounded penalty, and therefore (5.21) requires that $\beta_\infty(q_n)\leq C_1'/n^2$. 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, (5.21) holds.

To prove the second part of (5.20), we split $T_n^\star(m,m')$ into two terms

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

where, for $k=1,2$,

$$T_{n,k}^\star(m,m')=\Big[\sup_{t\in B_{m,m'}(0,1)}\Big|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\big(u_t^*(Z^\star_{(2\ell+k-1)q_n+i})-\langle t,g\rangle\big)\Big|^2-p_k(m,m')\Big]_+. \tag{5.22}$$

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

$$\nu^\star_{n,1}(t) \;=\; \frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\big(u_t^*(Z^\star_{2\ell q_n+i})-\langle t,g\rangle\big)=\frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu^\star_{q_n,\ell}(u_t^*). \tag{5.23}$$

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

$$\nu^\star_{q_n,\ell}(u_t^*)=(1/q_n)\sum_{j=1}^{q_n}u_t^*(Z^\star_{2\ell q_n+j})-\langle t,g\rangle. \tag{5.24}$$

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

$$\Delta_2(m, f_Z) = \int_{-\pi m}^{\pi m}\int_{-\pi m}^{\pi m} \frac{|f_Z^*(x-y)|^2}{|f_\varepsilon^*(x) f_\varepsilon^*(y)|^2} dxdy. \tag{5.25}$$

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

$$\mathbb{E}(T^\star_{n,1}(m,m')) \leq \mathbb{E}\big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u_t^*) - 2(1+2\xi^2)(H^\star)^2(m^*)\big]_+. \tag{5.26}$$

According to the condition (5.16), we thus take

$$\begin{aligned}\mathrm{pen}(m) &= 4ap(m,m) = 4a(2p_1(m,m) + 2p_2(m,m)) = 16ap_1(m,m)\\ &= 32a(1+2\xi^2)\big(2n^{-1}\Delta(m)\big) = 64a(1+2\xi^2)n^{-1}\Delta(m).\end{aligned} \tag{5.27}$$

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 $m^* \geq m_1$, $\Delta(m^*)$ satisfies (5.11). Take $m_0 = m_1 \vee m_2 \vee m_3$. We split the sum over m' in two parts and write

$$\sum_{m'=1}^{m_n} \mathbb{E}(T^\star_{n,1}(m,m')) = \sum_{m'|m^*\leq m_0} \mathbb{E}(T^\star_{n,1}(m,m')) + \sum_{m'|m^*\geq m_0} \mathbb{E}(T^\star_{n,1}(m,m')). \tag{5.28}$$

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

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

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

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

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

$$v^\star(m^*) = \frac{2\sqrt{\Delta_2(m^*, f_Z)}}{2\pi q_n} \quad \leq \quad 2\lambda_2^\star(f_\varepsilon, \kappa_0)\frac{\Gamma_2(m^*)}{q_n}, \tag{5.29}$$

with

$$\lambda_2^\star(f_\varepsilon, \kappa_0) = \kappa_0^{-1}\sqrt{2\pi\lambda_1}\|f_{\varepsilon^*}\|\mathbf{I}_{\delta\leq 1} + \mathbf{I}_{\delta>1} \tag{5.30}$$

where $\lambda_1 = \lambda_1(f_\varepsilon, \kappa_0)$ is defined in (4.4) and

$$\Gamma_2(m) = (1+(\pi m)^2)^\gamma (\pi m)^{\min((1/2-\delta/2),(1-\delta))} \exp(2\mu(\pi m)^\delta) = (\pi m)^{-(1/2-\delta/2)_+}\Gamma(m). \tag{5.31}$$

By combining the left hand-side of (5.11) and (5.29), we get that, for $m^* \geq m_0$,

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

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

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

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

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

$$\sum_{m'=1}^{m_n} m_n\exp\left\{-\frac{2K_1\xi C(\xi)}{7}\frac{\sqrt{n}}{q_n}\right\}\frac{\Delta(m_n)q_n^2}{n^2}\leq\frac{C}{n}. \tag{5.32}$$

Consequently

$$\sum_{m'|m^*\geq m_0} II^\star(m^*)\leq\frac{C}{n}. \tag{5.33}$$

• Study of $\sum_{m'|m^*\geq m_0} I(m^*)$. Denote by $\psi=2\gamma+\min(1/2-\delta/2,1-\delta)$, $\omega=(1/2-\delta/2)_+$, and $K'=K_1\lambda_1(f_\varepsilon,\kappa_0')/(2\lambda_2^\star(f_\varepsilon,\kappa_0))$. For $a,b\geq 1$, we use that

$$\begin{aligned}\max(a,b)^\psi e^{2\mu\pi^\delta\max(a,b)^\delta}e^{-K'\xi^2\max(a,b)^\omega} &\leq (a^\psi e^{2\mu\pi^\delta a^\delta}+b^\psi e^{2\mu\pi^\delta b^\delta})e^{-(K'\xi^2/2)(a^\omega+b^\omega)}\\ &\leq a^\psi e^{2\mu\pi^\delta a^\delta}e^{-(K'\xi^2/2)a^\omega}e^{-(K'\xi^2/2)b^\omega}+b^\psi e^{2\mu\pi^\delta b^\delta}e^{-(K'\xi^2/2)b^\omega}.\end{aligned}\tag{5.34}$$

Consequently,

$$\begin{aligned}\sum_{m'|m^*\geq m_0} I(m^*) &\leq\sum_{m'=1}^{m_n}\frac{\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m^*)}{n}\exp\left\{-\frac{K_1\xi^2\lambda_1(f_\varepsilon,\kappa_0')}{2\lambda_2^\star(f_\varepsilon,\kappa_0)}(\pi m^*)^{(1/2-\delta/2)_+}\right\}\\ &\leq \frac{2\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m)}{n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m)^{(1/2-\delta/2)_+}\right\}\sum_{m'=1}^{m_n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+}\right\}\\ &\quad+\sum_{m'=1}^{m_n}\frac{2\lambda_2^\star(f_\varepsilon,\kappa_0)\Gamma_2(m')}{n}\exp\left\{-\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+}\right\}.\end{aligned}\tag{5.35}$$

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 (5.35) 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 (5.16), the result follows by choosing $\mathrm{pen}(m) = 4ap(m,m) = 192a\Delta(m)/n$.

Case $\delta = 1/3$. According to the inequality (5.34), 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 $0 \le \delta < 1/3$, this choice ensures that $\sum_{m'|m^*\ge m_0} I(m^*) \le C/n$. The result follows by taking $p(m,m') = 2(1+2\xi^2)\Delta(m^*)/n$, and

$$\mathrm{pen}(m) \;=\; 64a(1+2\xi^2)\frac{\Delta(m)}{n} = 64a\left(1+\frac{32\mu\pi^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{K_1\lambda_1(f_\varepsilon,\kappa_0')}\right)\frac{\Delta(m)}{n}.$$

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

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

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

$$\begin{aligned}\mathrm{pen}(m) &= 64a(1+2\xi^2(m))\frac{\Delta(m)}{n}\\ &= 64a\left(1+\frac{32\mu\pi^\delta\lambda_2^\star(f_\varepsilon,\kappa_0)}{K_1\lambda_1(f_\varepsilon,\kappa_0')}(\pi m^*)^{\min((3\delta/2-1/2)_+,\delta)}\right)\frac{\Delta(m)}{n} \qquad \square\end{aligned}$$

Proof in the τ-dependent case.
We use the coupling properties recalled in Section 2.2 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 (5.15) with $R_n^\star(m,\hat m)$ defined by (5.14) and pen(m) chosen such that (5.16) holds. Next we use (5.17) 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

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)}|\nu_{n,Z}(u_t^*)-\nu^\star_{n,Z}(u_t^*)|^2\Big] &\le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n}\mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot})-\nu_{n,X}(e^{ix\cdot})|dx\\ &\le \frac{\tau_{\mathbf{X},\infty}(q_n)}{\pi}\int_{-\pi m_n}^{\pi m_n}\frac{|x|}{|f_\varepsilon^*(x)|^2}dx\\ &\le \tau_{\mathbf{X},\infty}(q_n)m_n\Delta(m_n).\end{aligned}\tag{5.36}$$

By gathering (5.15) and (5.36) we get

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

Therefore we infer that, for all $m\in\{1,\cdots,m_n\}$, (5.19) holds provided that

$$\Delta(m_n)m_n\tau_\infty(q_n)\le C_1/n \quad\text{and}\quad \sum_{m'=1}^{m_n}\mathbb{E}(T_n^\star(m,m'))\le C_2/n.\tag{5.37}$$

Using (5.11), we conclude that the first part of (5.37) is fulfilled as soon as

$${m_n}^{2\gamma+2-\delta}\exp\{2\mu\pi^\delta {m_n}^\delta\}\tau_\infty(q_n)\le C_1'/n.\tag{5.38}$$

In order to ensure that our estimators converge, we only consider models with bounded penalty, that is $\Delta(m_n)=O(n)$. Therefore (5.38) 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

$$m_n n^{-c(1+\theta)}=O(n^{-2}).\tag{5.39}$$

If f_ε satisfies (1.3) with $\delta>0$, and if $\theta>3$, one can find $c\in]0,1/2[$, such that (5.39) is satisfied. Now, if $\delta=0$ and $\gamma\ge 3/2$ in (1.3) and if $\theta>3+2/(1+2\gamma)$, then one can find $c\in]0,1/2[$, such that (5.39) is satisfied. These conditions ensure that (5.21) holds.

In order to prove the second part of (5.37), we proceed as for the proof of the second part of (5.20) and split $T_n^\star(m,m')$ into two terms

$$T_n^\star(m,m')=(T^\star_{n,1}(m,m')+T^\star_{n,2}(m,m'))/2,$$

where the $T^\star_{n,k}(m,m')$'s are defined in (5.22). We only study $T^\star_{n,1}(m,m')$ and conclude for $T^\star_{n,2}(m,m')$ analogously. As in the β-mixing framework, the study of $T^\star_{n,1}(m,m')$ consists in applying a concentration inequality to $\nu^\star_{n,1}(t)$ defined in (5.23) and considered as the sum of the p_n independent random variables $\nu^\star_{q_n,\ell}(t)$ defined as in (5.24). 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

$\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u^*_t) \|_\infty \leq M^\star_1(m^*)$, $\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell}(u^*_t)) \leq v^\star(m)$ and lastly $\mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u^*_t)|) \leq H^\star(m^*)$. According to Lemma 5.3, we take

$$(H^\star(m^*))^2 = \frac{2\Delta(m^*)}{n},\ M^\star_1(m^*) = \sqrt{\Delta(m^*)} \text{ and } v^\star(m^*) = \frac{C_{v^*}\sqrt{\Delta_2(m^*, f_Z)}}{2\pi q_n},$$

where $\Delta_2(m, f_Z)$ is defined in (5.25) and where

$$C_{v^*} = 2\Big[\mathbb{I}_{\delta>0} + \frac{\sqrt{2}\pi^{3/2}(2\pi)^{3/2}}{\sqrt{3}}\sum_{k\geq 1}\tau_1(k)\mathbb{I}_{\delta=0}\Big]. \tag{5.40}$$

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

$$\mathbb{E}(T^\star_{n,1}(m, m')) \leq \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(u^*_t) - 2(1 + 2\xi^2)(H^\star)^2(m^*)\Big]_+. \tag{5.41}$$

As in the β-mixing framework we take $\mathrm{pen}(m) = 64a\Delta(m)(1+2\xi^2)/n$ where ξ^2 is suitably chosen (see (5.41)). Set m_2 and m_3 as defined in Lemma 5.3, and set m_1 such that for $m^* \geq m_1$ (5.11) holds. Take $m_0 = m_1 \vee m_2 \vee m_3$ and $K' = K_1\lambda_1(f_\varepsilon, \kappa'_0)/(C_{v^*}\lambda^\star_2(f_\varepsilon, \kappa_0))$. The end of the proof is the same as in β-mixing framework, up to possible multiplicative constants.□

5.4. **Technical lemmas.**

Lemma 5.1.

$$\| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \|_\infty \leq \Delta(m). \tag{5.42}$$

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*

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u^*_t) \|_\infty \leq \sqrt{\Delta(m^*)} \tag{5.43}$$

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

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

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

where $\Delta(m)$ *and* $\Delta_2(m, f_Z)$ *are defined by (3.5) and (5.25).*

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

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

By using (5.6) we obtain

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

Now, according to (5.9) and (2.13)

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

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

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

According to (5.5), (5.8) and (2.13) we have

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

Hence,

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

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

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

that is

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

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

(1) $\delta = 0$, $\gamma \geq 3/2$ *in (1.3)*
(2) *or* $\delta > 0$ *in (1.3).*

Then we have

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu^\star_{q_n,\ell}(u_t^*) \|_\infty \leq \sqrt{\Delta(m^*)} \tag{5.44}$$

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

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

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

where $\Delta(m)$ *and* $\Delta_2(m, f_Z)$ *are defined by (3.5) and (5.25) and where* C_{v^*} *is defined in (5.40).*

Proof of Lemma 5.3. The proof of (5.44) is the same as the proof of (5.43). Next, again as for the proof of Lemma 5.2

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

with

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

Now, according to (5.9) and (2.14)

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

This implies that

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

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

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

According to (5.5), (5.8) and (2.14) and by applying the same arguments as for the proof of Lemma 5.2 we have

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

Hence,

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

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 $\xi^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1+2\xi^2)H^2\Big]_+ \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\frac{nH}{M_1}}\right),$$

with $C(\xi) = \sqrt{1+\xi^2} - 1$, $K_1 = 1/6$, 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). Usual density arguments show that this result can be applied to the class of functions $\mathcal{F} = B_{m,m'}(0,1)$.

Proof of Proposition 2.1. To prove (1), let for $t > 0$, $Y_t^* = \eta_t\sigma_t^*$. Note that the sequence $((Y_t^*, \sigma_t^*))_{t\geq1}$ is distributed as $((Y_t, \sigma_t))_{t\geq1}$ and independent of $\mathcal{M}_i = \sigma(\sigma_j, Y_j, 0 \leq j \leq i)$. Hence, by the coupling properties of τ (see (2.12)), 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

$$\begin{aligned}(5.45)\quad \tau(\mathcal{M}_i, (Z_{i_1}, X_{i_1}), \ldots, (Z_{i_l}, X_{i_l})) &\leq 2\mathbb{E}(|g_\epsilon(Y_0^2)| + |g_\epsilon(\sigma_0^2)|) \\ &\quad + \tau(\mathcal{M}_i, (f_\epsilon(Y_{i_1}^2), f_\epsilon(\sigma_{i_1}^2)), \ldots, (f_\epsilon(Y_{i_l}^2), f_\epsilon(\sigma_{i_l}^2)))\end{aligned}$$

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 (5.45), 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 (2.5). If $\sum_{j=1}^{\infty} a_j < 1$, the unique stationary solution to (2.5) is given by Giraitis *et al.* (2000):

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

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

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

Clearly

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

Now

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

This being true for any $1 \le k \le n$, the proof of Proposition 2.1 is complete.

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

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

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

Bougerol, P. and Picard, N. (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 Tsybakov, A. B. (2005). Sharp optimality for density deconvolution with dominating bias. *Theor. Probab. Appl. To appear*.

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

Comte, F. and Genon-Catalot, V. (2005). Penalized projection estimator for volatility density. *Preprint MAP5 2005-9. To appear in Scandinavian Journal of Statistics*.

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

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

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

Devroye, L. (1986). *Non-uniform random variate generation.* New York etc.: Springer-Verlag.

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

Doukhan, P., Teyssière, G., and Winant, P. (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.

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

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

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

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

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

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

van Es, B., Spreij, P., and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli 9*(3), 451–465.

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

Volkonskij, V. and Rozanov, Y. (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.