Submitted to Bernoulli

Root$-n$ consistent estimation of the marginal density in semiparametric autoregressive time series models

LIONEL TRUQUET ,

ENSAI - Campus de Ker-Lann
Rue Blaise Pascal - BP 37203
35712 BRUZ cedex
FRANCE
E-mail: lionel.truquet@ensai.fr

In this paper, we consider the problem of estimating the marginal density in some autoregressive time series models for which the conditional mean and variance have a parametric specification. Under some regularity conditions, we show that a kernel type estimate based on the residuals can be root$-n$ consistent even if the noise density is unknown. Our results substantially extend those existing in the literature. Our assumptions are carefully checked for some standard time series models such as ARMA or GARCH processes. Asymptotic expansion of our estimator is obtained by combining some martingale type arguments and a coupling method for time series which is of independent interest. We also study the uniform convergence of our estimator on compact intervals.

MSC 2010 subject classifications: Primary 62G07, 62M10; secondary 62G20.
Keywords: Kernel density estimation, Nonlinear time series.

1. Introduction

Nonparametric density estimation is an important tool for the analysis of time dependent data, especially for economic and financial time series. For instance, estimating the marginal density often provides important information on the shape and tail behavior of the distribution of stock prices or market indices. A discussion of more specific applications in econometrics or finance can be found in Liao and Stachurski (2015) and the references therein. In time series analysis, short-term prediction is an important issue and most of the statistical models used by practitioners are based on the conditional density which is the main quantity of interest in this context. The so-called ARCH processes introduced in the literature to model the dynamics of conditionally heteroscedastic time series, illustrate this approach quite well. However, as pointed out in Francq and Zakoan (2013), the marginal density of time series becomes the relevant quantity for long-term prediction or long-term value-at-risk evaluation. We can also mention that an important application of marginal density estimation is in model checking. For some interest rates series, Aït-Sahalia (1996) tested model adequacy by comparing a nonparametric estimator of the marginal density and the corresponding estimator implied by a parametric diffusion model and shows that the linearity of the drift imposed by the literature is the main source of misspecification. See also Gao and King (2004) and Corradi and Swanson (2005) for similar test procedures for diffusion processes.

Nonparametric density estimation has been extensively studied in the literature. Kernel density estimation is probably one of the most popular methods used for this problem and the properties of the so-called Parzen-Rosenblatt estimator have been investigated under various mixing type conditions. For instance by Robinson (1983), Ango Nze and Doukhan (1998),Doukhan and Louhichi (2001) or Roussas (2000). See also the monograph of Bosq (1998) for the kernel density estimation for strong mixing sequences and Dedecker et al. (2007) for numerous weak dependence conditions ensuring consistency properties of this estimator.

When additional structure is assumed however, for the stochastic process of interest, kernel density estimation can be used more cleverly to obtain sharper rates of convergence, in particular $\sqrt{n}-$consistency. This atypical rate of convergence in the nonparametric density estimation has been first noticed for the estimation of the density of some functionals of independent random variables. See Frees (1994), Schick and Wefelmeyer (2004b) and Giné and Mason (2007). In time series, existing contributions exploit the representation of the marginal density as a convolution product between the innovation density and the marginal density of a predictable process. Such an approach has been used by Saavedra and Cao (1999), Schick and Wefelmeyer (2007) and Schick and Wefelmeyer (2004a) for estimating the marginal density of invertible, moving average processes. In the latter contribution, sharper results are obtained for possibly infinite moving average processes. More recently, Kim et al. (2015) obtained some results for nonlinear and homoscedastic autoregressive processes of order 1 for which the conditional mean has a parametric specification. However, the authors did not study the case of conditionally heteroscedastic time series and used a different approach to test model adequacy in the presence of a time-varying conditional variance. Another recent contribution has been made by Delaigle et al. (2015) who constructed a $\sqrt{n}-$consistent estimator of the density of the log-volatility for a GARCH$(1,1)$ process. Note however that the purpose of this latter contribution was not the estimation of the marginal distribution and the volatility process was not directly observed. Moreover, the approach used seems specific to the autoregressive equation followed by the GARCH$(1,1)$.

In the literature, $\sqrt{n}-$consistent estimation of the marginal density in conditionally heteroscedastic time series models has not been considered. Moreover, even in the homoscedastic case, a general approach has not been studied for obtaining this convergence rate. In this paper, we consider the problem of estimating the marginal density with the $\sqrt{n}$ rate of convergence in some autoregressive time series models, conditionally homoscedastic or heteroscedastic. We will restrict our study to short memory models with a location-scale formulation

$$X_t = m_t(\theta) + \sigma_t(\theta)\varepsilon_t, \quad t \in \mathbb{Z}.$$

where the conditional mean $m_t(\theta)$ and volatility $\sigma_t(\theta)$ depend smoothly on a finite-dimensional parameter θ and are random variables measurable with respect to $\sigma(X_{t-1}, X_{t-2}, \ldots)$. With respect to the existing results, our approach covers many cases, from the ARMA processes with independent and identically distributed innovations to ARMA processes with a GARCH noise. Let us also mention that our contribution provides an answer to a question addressed in Zhao (2010), a paper in which an estimator similar to ours was suggested for density estimation in autoregressive time series models. It also suggests that such a convergence rate is not due to a convolution representation of the marginal density but on a $U-$statistics representation of our estimator, in the spirit of the method proposed by Frees (1994) or Giné and Mason (2007). For heteroscedastic parametric regression models, the root−n rate of a similar estimator is also conjectured in Li and Tu (2016), Section 6.1 and with straightforward modifications, our results apply in this case. However, apart from some classical smoothness conditions, the root-n consistency of our estimator is only guaranteed under the square integrability, with respect to the noise distribution, of the conditional density of the marginal X_t given the noise component ε_t. For independent data, this kind of condition is classical for deriving the asymptotic properties of Free's estimators. See Giné and Mason (2007), Theorem 1, assumption (b). Such an assumption is not always satisfied and has to be checked for the model under study. In this paper, we show in particular that ARMA processes with GARCH errors generally satisfy this integrability condition. A similar integrability condition is discussed by Müller (2012) for estimating the marginal density in some homoscedastic regression models. See also Schick and Wefelmeyer (2009) who showed that estimating a convolution of some powers of independent random variables can lead to a slower rate of convergence when this condition fails to hold.

This convergence rate is not only interesting for theoretical reasons. A classical application is to test model adequacy using the marginal density in the spirit of Aït-Sahalia (1996) who popularized this approach for diffusion processes. For time series, Kim et al. (2015) have recently used a statistics based on a comparison between the standard kernel estimator and a convolution estimator. However, the method given in Kim et al. (2015) for conditionally heteroscedastic time series models, is not based directly on the estimation of the marginal density. Our contribution has the benefit of providing precise assumptions under which such a root$-n$ consistent density estimation can be obtained in some semiparametric time series models. This estimator can be used in turn, to construct adequation tests which naturally extends the test derived by Kim et al. (2015) for conditionally homoscedastic time series models. We discuss such a test in Section 2.3.

The paper is organized as follows. In Section 2, we define our estimator and give its asymptotic properties. In Section 3, we check the assumptions of our theorems for some standard examples of time series models. We also compare our assumptions with that used in the aforementioned references. In Section 4, we provide a simulation study which shows that our estimator outperforms the standard kernel estimator when the data generating process is an ARMA/GARCH process. Proofs of some of our results are postponed to the last section of the paper. Supplementary material available online provides additional technical lemmas as well as a proof of some technical points.

2. Marginal density estimation of a time series

2.1. Model and marginal density estimator

We first introduce the general model used in the following. Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d square integrable random variables. If Θ denotes a Borel subset of $\mathbb{R}^d$, we consider two measurable functions $H, G : \Theta \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$. We assume that for a $\theta_0 \in \Theta$, $(X_t)_{t\in\mathbb{Z}}$ is a stationary process such that

$$X_t = H\left(\theta_0; X_{t-1}, X_{t-2}, \ldots\right) + \varepsilon_t G(\theta_0; X_{t-1}, X_{t-2}, \ldots). \tag{1}$$

Note that the two functions H and G will be more precisely defined $\lambda_d \otimes \mathbb{P}_X$ almost everywhere, where λ_d denotes the Lebesgue measure on $\mathbb{R}^d$ and $\mathbb{P}_X$ the probability distribution of $(X_{t-1}, X_{t-2}, \ldots)$. We also assume that $X_t \in \sigma\left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right)$, i.e

$$X_t = E\left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right),$$

for a suitable measurable function $E : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ defined $\mathbb{P}_\varepsilon$ almost everywhere. We also set

$$m_t(\theta) = H\left(\theta; X_{t-1}, X_{t-2}, \ldots\right), \quad \sigma_t(\theta) = G\left(\theta; X_{t-1}, X_{t-2}, \ldots\right).$$

The generation of all the past values are not available. We then assume that there exist measurable functions, $H_t, G_t : \Theta \times \mathbb{R}^t \to \mathbb{R}$ such that $H_t\left(\theta; X_{t-1}, \ldots, X_1\right)$ (and $G_t\left(\theta; X_{t-1}, \ldots, X_1\right)$) is an approximation of $m_t(\theta)$ (and $\sigma_t(\theta)$). We then use the notations

$$\overline{m}_t(\theta) = H_t\left(\theta; X_{t-1}, \ldots, X_1\right), \quad \overline{\sigma}_t(\theta) = G_t\left(\theta; X_{t-1}, \ldots, X_1\right).$$

In general, these approximations are obtained by replacing X_{t-i} by 0 for $i \geq t$ in the expressions of H and G. For instance, for an invertible moving average of order 1, $X_t = \varepsilon_t - \theta_0 \varepsilon_{t-1}$, one can set

$$m_t(\theta) = -\sum_{j\geq 1} \theta^j X_{t-j}, \quad \overline{m}_t(\theta) = -\sum_{j=1}^{t-1} \theta^j X_{t-j}, \quad t \geq 2.$$

Our estimator is based on the representation of the marginal density f_X of the stationary process $(X_t)_{t\in\mathbb{Z}}$ as a smooth functional of the noise density f_ε. More precisely, setting $X_t^- = (X_{t-1}, X_{t-2}, \ldots)$ and denoting by $f\left(\cdot|X_t^-\right)$ the conditional density of X_t given X_t^-, we have for $v \in \mathbb{R}$,

$$f_X(v) = \mathbb{E}\left[f\left(v|X_t^-\right)\right] = \mathbb{E}\left[\frac{1}{\sigma_i(\theta_0)} f_\varepsilon\left(\frac{v - m_i(\theta_0)}{\sigma_i(\theta_0)}\right)\right]. \tag{2}$$

Imagine first that a sample $(X_i, m_i(\theta_0), \sigma_i(\theta_0))_{1\leq i\leq n}$ is available. Then the vector of innovations $(\varepsilon_1, \ldots, \varepsilon_n)$ is also observed. The noise density f_ε can be estimated by the classical Parzen-Rosenblatt kernel estimator. If $K : \mathbb{R} \to \mathbb{R}_+$ is a symmetric probability density with compact support $[-1,1]$, which will be assumed to be continuously differentiable in the following, we set

$$\hat{f}_\varepsilon(z) = \frac{1}{n}\sum_{i=1}^n K_b\left(z-\varepsilon_i\right), \quad K_b(x) = \frac{1}{b}K\left(\frac{x}{b}\right).$$

Then, using the expression (2), we define the following unfeasible estimator

$$\begin{aligned}\check{f}_X(v) &= \frac{1}{n}\sum_{i=1}^n \frac{1}{\sigma_i(\theta_0)}\hat{f}_\varepsilon\left(L_{v,i}(\theta_0)\right)\\ &= \frac{1}{n^2}\sum_{i,j=1}^n \frac{1}{\sigma_i(\theta_0)}K_b\left[L_{v,i}(\theta_0)-\varepsilon_j\right],\end{aligned}$$

with $L_{v,i}(\theta) = \frac{v-m_i(\theta)}{\sigma_i(\theta)}$ for $(v,\theta)\in\mathbb{R}\times\Theta$. In practice, the parameter θ_0 has to be estimated and only the vector $(X_1,X_2,\ldots,X_n)$ is observed. Let us introduce additional notations. For $(v,\theta)\in\mathbb{R}\times\Theta$, we set $\varepsilon_j(\theta) = \frac{X_j-m_j(\theta)}{\sigma_j(\theta)}$. We call the process $\left(\varepsilon_j(\cdot)\right)_{j\in\mathbb{Z}}$ the residual process. We also denote by $\overline{\varepsilon}_j(\theta)$ and $\overline{L}_{v,i}(\theta)$ the truncated versions of $\varepsilon_j(\theta)$ and $L_{v,i}(\theta)$ respectively, e.g $\overline{\varepsilon}_j(\theta) = \frac{X_j-\overline{m}_j(\theta)}{\overline{\sigma}_j(\theta)}$.

Then, if $\hat{\theta}$ denotes an estimator of θ_0, the feasible estimator of $f_X(v)$ is defined by

$$\hat{f}_X(v) = \frac{1}{n^2}\sum_{i,j=1}^n \frac{1}{\overline{\sigma}_i(\hat{\theta})}K_b\left[\overline{L}_{v,i}(\hat{\theta})-\overline{\varepsilon}_j(\hat{\theta})\right].$$

Note that $\left(\overline{\varepsilon}_j(\hat{\theta})\right)_{1\leq j\leq n}$ are the residuals obtained after the estimation step.

In the homosecedastic case, i.e there exists $\sigma>0$ such that $\sigma_t(\theta)=\sigma$ for all $(t,\theta)\in\mathbb{Z}\times\Theta$, our estimator is simply defined by

$$\hat{f}_X(v) = \frac{1}{n^2}\sum_{i,j=1}^n K_b\left[v-\overline{m}_i(\hat{\theta})-X_j+\overline{m}_j(\hat{\theta})\right]. \tag{3}$$

Note that the estimation of the variance σ^2 is unnecessary in the homoscedastic case. An estimator of type (3) already appears in the literature but using a convolution approach. See for instance Schick and Wefelmeyer (2007) for linear processes, Müller (2012) for homoscedastic regression models and Kim et al. (2015) for some non linear conditionally homoscedastic time series). In this case, the kernel K is a convolution product of type $k*k$ and the estimator (3) is obtained as a convolution product of two kernel estimators: the Parzen-Rosenblatt estimator, with kernel k, of the density of $m_t(\theta_0)$ and that of f_ε with the same kernel. In this paper, we will consider an arbitrary continously differentiable and symmetric kernel K and the homoscedastic case as a special case of the conditionally heteroscedastic case, by in this case, setting the two quantities σ_t and $\overline{\sigma}_t$ to 1 in all our statements.

2.2. Assumptions and asymptotic behavior of the marginal density estimate

We now give our assumptions for deriving the asymptotic behavior of the unfeasible estimator $\check{f}_X$ and the feasible estimator $\hat{f}_X$. In the following, we will denote by $\|\cdot\|$ a norm on $\mathbb{R}^d$ whatever the value of the integer d. We will still denote by $\|\cdot\|$ the corresponding operator norm. For a family $\{A(\theta); \theta \in \Theta\}$ of matrices and a family $\{B(\theta) : \theta \in \Theta\}$ of real numbers, we set $|A|_{\infty,\epsilon} = \sup_{\theta \in \Theta_{0,\epsilon}} \|A(\theta)\|$ and $|A, B|_{\infty,\epsilon} = \sup_{\theta,\theta' \in \Theta_{0,\epsilon}} \|\frac{A(\theta)}{B(\theta')}\|$, where $\Theta_{0,\epsilon} = \{\theta \in \Theta : \|\theta - \theta_0\| < \epsilon\}$.

Finally, since for $i \in \mathbb{Z}$, $m_i(\theta)$ and $\sigma_i(\theta)$ are measurable functions of $Y_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots)$, we define some coupling versions of these two quantities (m_i, σ_i do not depend on ε_i but we keep this additional variable for simplicity of notations). For an integer $\ell \geq 1$, we denote by $\left(\varepsilon_j^{(i)}\right)_{j \in \mathbb{Z}}$, a copy of $(\varepsilon_j)_{j \in \mathbb{Z}}$ and we denote by $m_{i\ell}(\theta)$ and $\sigma_{i\ell}(\theta)$ the two random variables defined as $m_i(\theta)$ and $\sigma_i(\theta)$ but for which Y_i is replaced with

$$Y_{i\ell} = \left(\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-\ell+1}, \varepsilon_{i-\ell}^{(i)}, \varepsilon_{i-\ell-1}^{(i)}, \ldots\right).$$

One can note that $(m_{i\ell}(\theta), \sigma_{i\ell}(\theta))$ has the same distribution as $(m_i(\theta), \sigma_i(\theta))$. The interest of such a coupling method will be explained in Section 5. The following assumptions will be needed.

A1 The parameter $\theta \in \Theta$ where Θ is a subset of $\mathbb{R}^d$.

A2 The volatility is bounded away from zero, i.e there exists $\gamma > 0$ such that $\inf_{\theta \in \Theta} \sigma_i(\theta) \geq \gamma$ a.s. We also assume $\inf_{\theta \in \Theta} \overline{\sigma}_i(\theta) \geq \gamma$ a.s. Moreover, there exists $s, a \in (0,1)$ and $\kappa > 0$ such that

$$\mathbb{E}\left[\sup_{\theta \in \Theta_{0,\epsilon}} |m_t(\theta)|^s + \sup_{\theta \in \Theta_{0,\epsilon}} |\sigma_t(\theta)|^s\right] < \infty,$$

$$\mathbb{E}\left[\sup_{\theta \in \Theta_{0,\epsilon}} |m_i(\theta) - m_{i\ell}(\theta)|^s + \sup_{\theta \in \Theta_{0,\epsilon}} \left|\sigma_i^2(\theta) - \sigma_{i\ell}^2(\theta)\right|^s\right] \leq \kappa a^\ell.$$

and

$$\mathbb{E}\left[\sup_{\theta \in \Theta_{0,\epsilon}} |m_i(\theta) - \overline{m}_i(\theta)|^s + \sup_{\theta \in \Theta_{0,\epsilon}} \left|\sigma_i^2(\theta) - \overline{\sigma}_i^2(\theta)\right|^s\right] \leq \kappa a^i.$$

A3 The two functions $\theta \mapsto \sigma_t(\theta)$ and $\theta \mapsto m_t(\theta)$ are twice differentiable over Θ. Moreover, there exists $\epsilon > 0$ such that the following random variables are integrable.

$$|\dot{m}_i, \sigma_i|_{\infty,\epsilon}^3, \quad \left|\dot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}^3, \quad |\dot{m}_i, \sigma_i|_{\infty,\epsilon}^2 \cdot \left|\dot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}^2, \quad |\sigma_i, \sigma_i|_{\infty,\epsilon} \cdot |\ddot{m}_i, \sigma_i|_{\infty,\epsilon},$$

$$|m_i, \sigma_i|_{\infty,\epsilon} \cdot \left|\ddot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}, \quad \left|\ddot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}^{6/5},$$

where for a function $g : \Theta \to \mathbb{R}$, $\dot{g}$ and $\ddot{g}$ denote the gradient and the Hessian matrix of g respectively.

A4 There exists an estimator $\hat{\theta}$ of θ_0 such that $\hat{\theta} - \theta_0 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$.

A5 The noise density f_ε is twice differentiable and its second derivative f''_ε is bounded (hence f_ε and f'_ε are also bounded).

A6 Let I a compact interval on the real line. For all $v \in I$, we assume that the ratio $\frac{v-m_t(\theta_0)}{\sigma_t(\theta_0)}$ has a density denoted by h_v and for $x \in \mathbb{R}$, we set

$$g_v(x) = \mathbb{E}\left[\frac{1}{\sigma_t(\theta_0)} \mid \frac{v - m_t(\theta_0)}{\sigma_t(\theta_0)} = x\right] \cdot h_v(x).$$

We assume that the application $(x, v) \mapsto g_v(x)$ is jointly measurable and also that there exists $s_0 > 0$ such that for all $v \in I$,

$$\int g_v(x)^2 d\mu(x) < \infty, \quad d\mu(x) = \sup_{|s| \leq s_0} f_\varepsilon(x + s)dx.$$

A7 The envelope function G defined by $G(x) = \sup_{v \in I} g_v(x)$ satisfies $\int G(x)^{2+o} d\mu(x) < \infty$ for some $o \in (0, 1)$. Moreover, there exist some constants $\eta, C > 0$ such that

$$N_{[]}\left(\epsilon, \mathcal{G}, \mathbb{L}^2(\mu)\right) \leq C\epsilon^{-\eta},$$

where $N_{[]}\left(\epsilon, \mathcal{G}, \mathbb{L}^2(\mu)\right)$ denotes the bracketing numbers of the family $\mathcal{G}_I = \{g_v : v \in I\}$.

Notes

1. Different constants a, s and κ can be found for the three bounds given assumptions **A2**. However, we can always take the minimal value of the exponents s, the maximum value of the constant a and the maximum value of the constants $\kappa \geq 1$. There is therefore, no loss of generality in assuming the same constants for the three bounds.
2. Assumption **A2** imposes a restriction on the dependence structure of the time series models. These conditions, which are usually referred to as short-memory properties, are satisfied for the standard ARMA or GARCH processes. Roughly speaking, this weak dependence condition means that a perturbation of initial conditions in the data generating process is forgotten exponentially fast. This type of dependence condition is also used by Zhao (2010) or Kim et al. (2015).

Discussion of Assumption A6. Our results require some regularity conditions for the family of functions $\{g_v : v \in I\}$. The integrability condition assumed on g_v is necessary for root n consistency, as shown in Theorem 1 stated below. One can also relate this function to the conditional density of $X_t|\varepsilon_t$. Indeed, if $h : \mathbb{R}^2 \to \mathbb{R}_+$ is a measurable function, we have, setting for simplicity of notations $m_t = m_t(\theta_0)$ and $\sigma_t = \sigma_t(\theta_0)$,

$$\begin{aligned}
\mathbb{E}\left[h(X_t, \varepsilon_t)\right] &= \mathbb{E}\int h(m_t + \sigma_t x, x) f_\varepsilon(x)dx \\
&= \mathbb{E}\int h\left(v, \frac{v - m_t}{\sigma_t}\right)\frac{1}{\sigma_t} f_\varepsilon\left(\frac{v - m_t}{\sigma_t}\right)dv \\
&= \int\int h(v, x) g_v(x) f_\varepsilon(x)dxdv.
\end{aligned}$$

This shows that $v \mapsto g_v(x)$ can be seen as a version of the conditional density of X_t given that $\varepsilon_t = x$. In what follows, we discuss alternative expressions for the function g_v as well as some sufficient conditions ensuring the square integrability of g_v required in **A6**.

1. For the homoscedastic model, the volatility $\sigma_t(\cdot)$ equals a constant σ. Then if $m_t(\theta_0)$ has a density denoted by f_m, we have $g_v(x) = f_m(v - \sigma x)$. In this case, estimation of parameter σ will be unnecessary. Assumption **A6** holds true as soon as f_ε is bounded and f_m is square-integrable.
2. For a pure heteroscedastic model, i.e the conditional mean $m_t(\cdot)$ being reduced to a constant m, we assume that $\sigma_t(\theta_0)$ has a density f_σ. Assumption **A6** can hold only if $v \neq m$. For $v \neq m$, one can show that for $x \neq 0$, $g_v(x) = \frac{1}{|x|} f_\sigma\left(\frac{v-m}{x}\right)$. In this case, we have for $v \neq m$,
$$\int g_v(x)^2 \mu(dx) \leq \frac{\|f_\varepsilon\|_\infty}{|v-m|} \int_0^\infty f_\sigma^2(y) dy.$$
Then if for instance $f_{\sigma^2}(y) = \frac{1}{2\sqrt{y}} f_\sigma(\sqrt{y})$ is bounded, the integrability condition given in **A3** is guaranteed as soon as $\mathbb{E}\sigma_t(\theta_0) < \infty$, which is not a strong restriction.
3. In the location-scale case with a non degenerate conditional mean, we assume that the distribution of the pair $(m_t(\theta_0), \sigma_t(\theta_0))$ has a density denoted by $f_{m,\sigma}$. Then if $v \in \mathbb{R}$, the distribution of the couple $\left(\frac{v - m_t(\theta_0)}{\sigma_t(\theta_0)}, \frac{1}{\sigma_t(\theta_0)}\right)$ has a density ω given by
$$\omega(x,y) = \frac{1}{y^3} f_{m,\sigma}\left(v - \frac{x}{y}, \frac{1}{y}\right).$$
We deduce that
$$g_v(x) = \int_0^{1/\gamma} \frac{1}{y^2} f_{m,\sigma}\left(v - \frac{x}{y}, \frac{1}{y}\right) dy. \tag{4}$$
Using Jensen inequality and a change of variables, one can show that
$$\int g_v(x)^2 dx \leq \frac{1}{\gamma} \cdot \mathbb{E}\left[\sigma_t(\theta_0) f_{m,\sigma}\left(m_t(\theta_0), \sigma_t(\theta_0)\right)\right]. \tag{5}$$
Then the integrability condition given in **A3** follows if f_ε is bounded and if the integral $\int y f_{m,\sigma}^2(x,y) dx dy$ is finite.

Discussion of Assumption A3 Assumption **A3** imposes some moment restrictions and smoothness conditions. In the pure heteroscedastic case, we only have to check the integrability of
$$\left|\dot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}^3 \quad \left|\ddot{\sigma_i^2}, \sigma_i^2\right|_{\infty,\epsilon}^{6/5}.$$
In the homoscedastic case, these conditions reduce to the integrability of
$$|m_i|_{\infty,\epsilon}^2, \quad |\dot{m}_i|_{\infty,\epsilon}^3, \quad |\ddot{m}_i|_{\infty,\epsilon}.$$

These moment restrictions are explained by the technique used for the proof of Theorem 2 given below and which consists in studying the derivative of $\hat{f}$ with respect to $\hat{\theta}$, the estimator of θ_0. For a general heteroscedastic time series model, other conditions could be possible, the single requirement is to get the conclusions of Lemma 4 and Lemmas 5 and 6 given in the supplementary material.

We now give the asymptotic behavior of our estimates. We first start with the unfeasible estimator $\check{f}_X$.

Theorem 1. *Let $nb^{2+\delta} \to \infty$ and $nb^4 \to 0$ for some $\delta \in (0,1)$.*

1. *Assume that assumptions* **A2**, **A5** *and* **A6** *hold true. Then for all $v \in I$, we have*
$$\begin{aligned}&\sqrt{n}\left[\check{f}_X(v) - f_X(v)\right] \\ = \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g_v(\varepsilon_i) + \frac{1}{\sigma_i(\theta_0)}f_\varepsilon\left(\frac{v-m_i(\theta_0)}{\sigma_i(\theta_0)}\right) - 2f_X(v)\right] + o_{\mathbb{P}}(1). \end{aligned} \tag{6}$$
In particular, for all $v \in I$, we have $\sqrt{n}\left[\check{f}_X(v) - f_X(v)\right] = O_{\mathbb{P}}(1)$
2. *If in addition Assumption* **A7** *holds true, the approximation (6) is uniform in $v \in I$. In particular*
$$\sqrt{n}\sup_{v\in I}\left|\check{f}_X(v) - f_X(v)\right| = O_{\mathbb{P}}(1).$$

Note. The proof of Theorem 1 relies on the decomposition of a $U-$statistic for which the degenerate part is shown to be negligible under our bandwidth conditions. The bandwidth condition $\sqrt{n}b^2 \to 0$ is a bias condition, the bias of our estimator has to decrease faster than the rate $1/\sqrt{n}$. However, our estimator will be first approximated by a U-statistic involving $\ell-$dependent random variables in order to facilitate the study of its asymptotic behavior.

In the next result, we compare the asymptotic behavior of the feasible estimator $\hat{f}_X$ with that of the unfeasible one. For $\theta \in \Theta$, we denote by f_θ the density of $\varepsilon_i(\theta)$. We also set
$$h_\theta(v) = \mathbb{E}\left(\frac{1}{\sigma_i(\theta)}f_\theta\left(\frac{v-m_i(\theta)}{\sigma_i(\theta)}\right)\right).$$
In the following $\dot{h}_\theta(v)$ will denote the partial derivative with respect to θ of the function $(\theta, v) \mapsto h_\theta(v)$. By convention, we represent $\dot{h}_\theta(v)$ by a column vector. Moreover, A^T will denote the transpose of a matrix A.

Theorem 2. *Let $nb^{3+\delta} \to \infty$ and $nb^4 \to 0$ for some $\delta \in (0,1)$ and assume that the assumptions* **A1-A6** *hold true. We then have*
$$\sqrt{n}\sup_{v\in I}\left|\hat{f}_X(v) - \check{f}_X(v) - \dot{h}_{\theta_0}(v)^T\left(\hat{\theta}-\theta_0\right)\right| = o_{\mathbb{P}}(1).$$

Note. For the comparison of the two estimators $\hat{f}$ and $\check{f}$, we do not use $U-$statistics arguments. We wanted to avoid additional regularity conditions on the function g_v and its local approximation when $\theta \to \theta_0$, these conditions being difficult to check for practical examples. For the proof of Theorem 2, we first show that the effect of truncations of m_i, σ_i is negligible. We then use $\ell-$dependent approximations of these quantities. Finally, we study a Taylor expansion of order 1 and control the local approximation of the derivatives using martingale tools and integration with respect to the residual process $\theta \mapsto \frac{X_j - m_j(\theta)}{\sigma_j(\theta)}$. Then regularity conditions exclusively concern the densities $\theta \mapsto f_\theta$ of this residual process. One can note that the range of bandwidths allowed in Theorem 2 is reduced compared to Theorem 1.

In the next result, we provide the asymptotic distribution of the feasible estimator $\hat{f}_X$ of f_X. To this end, it is necessary to use a particular representation of the estimator $\hat{\theta}$. Similar representations are used in Zhao (2010) or Kim et al. (2015).

A8 There exists a square integrable process $Z_i(\theta_0) = H_{\theta_0}(\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$ taking values in $\mathcal{M}_{d,\bar{d}}(\mathbb{R})$, the space of real matrices of size $d \times \bar{d}$ and a measurable function $F : \mathbb{R} \to \mathbb{R}^{\bar{d}}$ such that $\mathbb{E}F(\varepsilon_0) = 0$, $Y_i(\theta_0)$ and $F(\varepsilon_0)$ are square integrable, $\mathbb{E}\|Z_i(\theta_0) - Z_{i\ell}(\theta_0)\|^2 \leq \kappa a^\ell$ (where $\kappa > 0$ and $a \in (0,1)$) and

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(\theta_0) F(\varepsilon_i) + o_{\mathbb{P}}(1).$$

For stating our next result, we define for $v \in I$ and $i \in \mathbb{Z}$,

$$M_{i,v} = g_v(\varepsilon_i) + \frac{1}{\sigma_i(\theta_0)} f_\varepsilon\left(\frac{v - m_i(\theta_0)}{\sigma_i(\theta_0)}\right) + \dot{h}_{\theta_0}(v)^T Z_i(\theta_0) F(\varepsilon_i).$$

The proof of the following corollary is given in the supplementary material.

Corollary 1. *Under the assumptions* **A1-A6** *and* **A8***, the process* $\left(\sqrt{n}\left[\hat{f}_X(v) - f_X(v)\right]\right)_{v \in I}$ *converges in the sense of finite dimensional distributions towards a centered Gaussian process* $(W_v)_{v \in I}$ *such that for* $v_1, v_2 \in I$,

$$\begin{aligned} Cov(W_{v_1}, W_{v_2}) &= Cov(M_{0,v_1}, M_{0,v_2}) \\ &+ \sum_{i \geq 1} \left[Cov(M_{0,v_1}, M_{i,v_2}) + Cov(M_{0,v_2}, M_{i,v_1})\right]. \end{aligned}$$

Moreover if assumption **A7** *also holds, the convergence occurs in* $\ell^\infty(I)$.

Notes

1. In the homoscedastic case, one can check that the covariance structure of the Gaussian process $(W_v)_{v \in I}$ is the same as in Theorem 2 of Kim et al. (2015). Thus, our result can be seen as an extension of the result obtained in their paper, allowing for an arbitrary number of lags in the conditional mean and also conditional heteroscedasticity.

2. The proofs of Theorem 1, Theorem 2 and Corollary 1 extensively make use of the coupling method discussed at the beginning of Section 5. This method allows an approximation to be made of some processes by $\ell-$dependent processes which has the same marginal distribution.

2.3. Adequation test for semiparametric time series models

In this section, we show that our results can be used to derive an adequation test in the spirit of the tests studied in Kim et al. (2015) for semiparametric time series models. For ARCH-type processes, the test considered by these authors is based on the maximum deviation between the kernel density estimator of $X_t/\sigma_t(\theta_0)$ and the convolution estimator of the sum $m_t(\theta_0)/\sigma_t(\theta_0)+\varepsilon_t$. This strategy was used to circumvent the nonparametric estimation of the marginal density under the model assumption whose consistency was found more difficult to get. Alternately, it is possible to use our estimator to define test statistics based on the marginal distribution of the process and which naturally extends the homoscedastic case considered in Kim et al. (2015). See also Gao and King (2004) for a test based on the marginal density for diffusion processes.

Let $f_X^{\#}$ be the kernel density estimator of f_X, i.e

$$f_X^{\#}(v)=\frac{1}{n}\sum_{i=1}^n K_b(v-X_i),\quad v\in\mathcal{I}.$$

We set

$$\Delta_n=\sqrt{nb}\sup_{v\in\mathcal{I}}\frac{\left|f_X^{\#}(v)-\hat{f}_X(v)\right|}{\sqrt{\hat{f}_X(v)\int K^2(u)du}}.$$

The proof of the following result is straightforward, using Theorem 2.2 in Liu and Wu (2010). Details are given in the supplementary material. The length of a compact interval I is denoted by $|I|$.

Corollary 2. *Assume that the assumptions* **A1-A7** *hold true and in addition that f_ε is positive everywhere. Setting*

$$\bar{b}=b/|I| \text{ and } C_K=\frac{1}{2}\log\left\{\int\left[K'(u)\right]^2du/\int K^2(u)du\right\}-\log(2\pi),$$

we have

$$\mathbb{P}\left(\left[2\log\bar{b}^{-1}\right]^{1/2}\Delta_n-2\log\bar{b}_n^{-1}-C_K\le z\right)\to e^{-2e^{-z}}.$$

Testing the hypothesis $H_0:(m_t,\sigma_t)=(m_t(\theta_0),\sigma_t(\theta_0))$ can be based on the statistics Δ_n. In Section 11 of the supplementary material, we discuss the implementation of such a test as well as a numerical comparison with the test of Kim et al. (2015) for testing ARCH type structures.

3. Examples

In this section, we explain how to check the assumptions **A1** − **A8** for some classical time series models.

3.1. Conditionally homoscedastic times series

Let us assume that

$$X_t = m_t(\theta_0) + \varepsilon_t, \quad t \in \mathbb{Z}.$$

We recall that $g_v(x) = f_m(v - x)$ where f_m denotes the density of m_t. In this case, assumption **A6** holds true as soon as f_ε is bounded and $\int f_m^2(x)dx < \infty$. The latter condition holds for instance if f_m is bounded which is the case when $m_t(\theta_0)$ is linear (see below the example of ARMA processes). However, f_m can be bounded even if $m_t(\theta_0)$ is also nonlinear with for instance $m_t(\theta_0) = \theta_{01} X_{t-1}^+ + \theta_{02} X_{t-1}^-$ $(a, b \neq 0)$ which corresponds to a threshold model discussed in Tong (1990). Note also that a semi-linear autoregressive model $m_t(\theta_0) = \theta_{01} X_{t-1} + h_{\theta_{02}}(X_{t-2}, \ldots, X_{t-p})$ checks this condition if $\theta_{01} \neq 0$. As pointed out by a referee, the boundedness of f_m is very restrictive as it excludes many functions $H(\theta_0; \cdot)$ whose gradient can vanish. However for a model with one lag, the square integrability of f_m can still hold but if a is a critical point, $x \mapsto |\partial_x H(\theta_0; x)|$ should be bounded from below by something such as $|x - a|^\alpha$, for some $\alpha \in (0, 1)$. This is a serious restriction because it excludes the existence of higher order derivatives and then two "flat" regression functions. In this case, Schick and Wefelmeyer (2009) have shown that Free's estimator has a slower convergence rate. For a detailed discussion of the required shape for H in the context of regression models, see Müller (2012), Section 5. However, for a model with several lags, it is more difficult to give a clear sufficient condition. Note also that our result can be applied to more complex autoregressive structures. ()See in particular the notes in Section 3.2 with the example of $\log X_t^2$ where $(X_t)_t$ is a GARCH process().

Assumption **A7** is satisfied if, in addition, f_m is Lipschitz and $\int |f_\varepsilon'(x)|dx < \infty$. Indeed, the latter condition entails that the measure μ of assumption **A7** has a finite mass and in this case, the polynomial decay of the bracketing number is classical. (See van der Vaart (1998), Example 19.7). If f_m is of bounded variation, **A7** is also satisfied. ()See van der Vaart (1998), Example 19.11).

For homoscedastic and autoregressive time series with one lag, Kim et al. (2015) obtained a root$-n$ consistent estimation of the marginal density by using a representation of the density of $m_t(\theta_0) + \varepsilon_t$ as a convolution product. In that paper, similar regularity assumptions are used for the noise distribution. These authors use bandwidth conditions similar to ours (see Theorem 2 of their paper). Their moment conditions for $\theta \mapsto m_t(\theta_0)$ and its derivative are less restrictive than ours but at the same time more regularity conditions on the density of $m_t(\theta_0)$ have to be checked for their non-linear models. See Assumption 4 and Assumption 6 of that paper for a precise statement of their regularity conditions. One advantage of our approach is to present a unified approach for homoscedastic and heteroscedastic time series and for which the dynamic can depend on an arbitrary and possibly infinite number of lags.

The case of ARMA processes. Let us now consider the case of ARMA processes, i.e there exist two integers p and q such that

$$X_t = \eta_0 + \sum_{i=1}^{p} a_{0i}\left(X_{t-i} - \eta_0\right) + \varepsilon_t - \sum_{j=1}^{q} b_{0j}\varepsilon_{t-j}, \quad t \in \mathbb{Z}.$$

As usual, we assume that for $\theta = (\eta, a_1, \ldots, a_p, b_1, \ldots, b_q) \in \Theta$, the roots of the two polynomials $\mathcal{P}(z) = 1 - \sum_{i=1}^{p} a_i z^i$ and $\mathcal{Q}(z) = 1 - \sum_{j=1}^{q} b_j z^j$ are outside the unit disc. Then defining

$$Z_t(\theta) = X_t - \eta - \sum_{j=1}^{p} a_j\left(X_{t-j} - \eta\right) + \sum_{j=1}^{q} b_j Z_{t-j}(\theta)$$

and $\underline{Z}_t(\theta) = \left(Z_t(\theta), Z_{t-1}(\theta), \ldots, Z_{t-q+1}(\theta)\right)^T$, we have

$$\underline{Z}_t(\theta) = A_1(\theta)\underline{Z}_{t-1}(\theta) + B_{1,t}(\theta),$$

where $A_1(\theta)$ denotes the companion matrix associated to $a_1, \ldots, a_p$ and

$$B_{1,t}(\theta) = \left(X_t - \eta - \sum_{j=1}^{p} a_j(X_{t-j} - \eta), 0, \ldots, 0\right)^T.$$

We then have

$$\underline{Z}_t(\theta) = \sum_{j=0}^{\infty} A_1(\theta)^j B_{1,t-j}(\theta), m_t(\theta) = \eta + \sum_{j=1}^{p} a_j\left(X_{t-j} - \eta\right) - \sum_{j=1}^{q} b_j Z_{t-j}(\theta). \quad (7)$$

Moreover $\overline{m}_t(\theta)$ can be defined by setting $X_0, X_{-1}, \ldots$ to 0 in (7). We now explain how to check assumptions **A3** − **A6**. Assumption **A2** will be checked directly for ARMA-GARCH processes.

- Assumption **A3** holds if $\mathbb{E}|\varepsilon_t|^3 < \infty$. Indeed, using the well-known infinite moving-average representation

$$X_t = \zeta_0 + \varepsilon_t + \sum_{j=1}^{\infty} \zeta_j \varepsilon_{t-j}, \quad \sum_{j=1}^{\infty} |\zeta_j| < \infty,$$

  we have $\mathbb{E}|X_t|^3 < \infty$. Assumption **A3** then holds true using (7) (the order of the derivative can be arbitrary in this example).
- Now assumptions **A6** − **A7** follow from the fact that f_m is bounded and Lipschitz. Indeed, using the infinite moving average representation, we have $m_t(\theta_0) = \zeta_0 +$

$\sum_{j=1}^{\infty} \zeta_j \varepsilon_{t-j}$. If we assume, without loss of generality, that $\zeta_1 \neq 0$ and also $\zeta_1 = 1$ for simplicity, we have

$$f_m(z) = \int f_\varepsilon(z-x)\nu(dx),$$

where ν denotes the probability distribution of the random variable $\zeta_0+\sum_{j=2}^{\infty} \zeta_j \varepsilon_{t-j}$. Hence the boundedness and Lipschitz property of f_m follows from assumption **A5**.

- Finally, assumptions **A4**, **A8** hold true using for instance, conditional maximum likelihood estimators. See Brockwell and Davis (1991), Chapter 8, for some asymptotic results for different inference methods of ARMA parameters.

Let us now compare our results with that of Schick and Wefelmeyer (2007). The results obtained by these authors are very general for applications to linear processes which contain ARMA processes as a special case. Their results are sharper than ours because they obtained uniform convergence of their convolution estimate on the real line whereas we consider uniformity only on compact intervals. However, our results apply if ε_0 has a moment of order 3, whereas Schick and Wefelmeyer (2007) use a moment of order 4 (see assumption F of the paper). Moreover, the kernels used in Schick and Wefelmeyer (2007) cannot be non-negative (see the assumption K applied with an order $m \geq 2$ for the kernel) and then the estimator of the density can take negative values. This excludes some classical kernels often used by the practitioners.

3.2. Pure GARCH models

In this subsection, we consider the process

$$X_t = m_0 + \varepsilon_t \sigma_t(\theta_0), \quad \sigma_t^2(\theta_0) = \alpha_{00} + \sum_{j=1}^{Q} \alpha_{0j}(X_{t-j} - m_0)^2 + \sum_{j=1}^{P} \beta_{0j}\sigma_{t-j}^2(\theta_0),$$

with $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}\varepsilon_0^2 = 1$. We set $\theta = (m, \alpha_0, \dots, \alpha_Q, \beta_1, \dots, \beta_P)$. Moreover let

$$\sigma_t^2(\theta) = \alpha_0 + \sum_{j=1}^{Q} \alpha_j (X_{t-j} - m)^2 + \sum_{j=1}^{Q} \beta_j \sigma_{t-j}^2(\theta).$$

Then $(X_t)_{t\in\mathbb{Z}}$ is (up to parameter m_0) a GARCH(p, q) process. We set

$$\underline{Y}_t = \left((X_t - m_0)^2, \dots, (X_{t-Q+1} - m_0)^2, \sigma_t^2(\theta_0), \dots, \sigma_{t-P+1}^2(\theta_0)\right)'.$$

There exist a sequence of i.i.d random matrices $(A_t)_{t\in\mathbb{Z}}$ of size $(p+q)\times(p+q)$ and a sequence of random vectors $(B_t)_{t\in\mathbb{Z}}$ of dimension $p+q$ such that

$$\underline{Y}_t = A_t \underline{Y}_{t-1} + B_t, \quad t \in \mathbb{Z}. \tag{8}$$

We refer the reader to Francq and Zakoïan (2010), p.29, for a precise expression of (A_t, B_t) as well as the definition of the Lyapunov exponent $\gamma(A)$ of the sequence $(A_t)_{t\in\mathbb{Z}}$. The following assumptions will be needed.

G1 $\gamma(A) < 0$ and for all $\theta \in \Theta$, $\sum_{j=1}^{P} \beta_j < 1$.
G2 We have $\alpha_{0,j} > 0$ and $\beta_{0,j'} > 0$ for $0 \leq j \leq Q$ and $1 \leq j' \leq P$.

In the following, we denote by C a generic positive constant. Under the assumption **G1**, there exist $s > 0$ and an integer $k \geq 1$ such that $c = \mathbb{E}^{1/k}\left(\|A_k A_{k-1} \cdots A_1\|^s\right) < 1$ and $\mathbb{E}\sigma_t^{2s}(\theta_0) < \infty$. Using the representation

$$\underline{Y}_t = B_t + \sum_{j=1}^{\infty} A_t \cdots A_{t-j+1} B_{t-j}$$

and the fact that $A_t, B_t \in \sigma(\varepsilon_t)$, we get

$$\begin{aligned} \mathbb{E}\|\underline{Y}_t - \underline{Y}_{t\ell}\|^s &\leq 2\sum_{j\geq \ell} \mathbb{E}\|A_t \cdots A_{t-j+1} B_{t-j}\|^s \\ &\leq C\sum_{j\geq\ell} c^j \\ &\leq Cc^\ell. \end{aligned}$$

We also then have $\mathbb{E}|X_t^2 - X_{t\ell}^2|^s \leq Cc^\ell$. We now check the assumptions **A2**, **A3**, **A4** and **A8**, **A5** and **A6**.

1. We first check **A2**. Setting for $t \in \mathbb{Z}$,

$$\underline{\sigma}_t^2(\theta) = \left(\sigma_t^2(\theta), \sigma_{t-1}^2(\theta), \ldots, \sigma_{t-P+1}^2(\theta)\right)^T$$

we have the recursive equations $\underline{\sigma}_t^2(\theta) = A_2(\theta)\underline{\sigma}_t^2(\theta) + B_{2,t}(\theta)$ with

$$B_{2,t}(\theta) = \left(\alpha_0 + \sum_{j=1}^{Q} \alpha_j (X_{t-j} - m)^2, 0, \ldots, 0\right)^T, \quad A_2(\theta) = \begin{pmatrix} \beta_1 & \cdots & \beta_P \\ & & 0 \\ & I_{P-1} & \vdots \\ & & 0 \end{pmatrix}.$$

$A_2(\theta)$ is then the companion matrix associated with $\mathcal{P}(z) = 1 - \sum_{j=1}^{P} \beta_j z^j$. Since $\sum_{j=1}^{P} \beta_{0j} < 1$, the spectral radius of $A_2(\theta)$ is less than 1 and there exists a positive integer k such that $\|A_2(\theta_0)^k\| < 1$ and then $\rho^k = \left|A_2^k\right|_{\infty,\epsilon} = \sup_{\theta\in\Theta_{0,\epsilon}} \|A_2(\theta)^k\| < 1$ if $\epsilon > 0$ is sufficiently small, using the continuity of $\theta \mapsto A_2(\theta)$. In particular, we have the expansion

$$\underline{\sigma}_t^2(\theta) = \sum_{j\geq 0} A_2(\theta)^j B_{2,t-j}(\theta).$$

Then, using the fact that the sequence $\left(\left|A_2^j\right|_{\infty,\epsilon}\right)_{j\geq 1}$ is bounded and $\mathbb{E}X_t^{2s} < \infty$, we deduce that

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta_{0,\epsilon}}\|\underline{\sigma}_t^2(\theta)-\underline{\sigma}_{t\ell}^2(\theta)\|^s &\leq C\left[\sum_{j=1}^{\frac{\ell}{2}}c^{(\ell-j)s}+\sum_{j>\frac{\ell}{2}+1}\left|A_2^j\right|_{\infty,\epsilon}^s\right]\\ &\leq C\left(c^{\frac{\ell s}{2}}+\rho^{\frac{s\ell}{2}}\right)\\ &\leq C\left[c^{\frac{s}{2}}\vee\rho^{\frac{s}{2}}\right]^{\ell}.\end{aligned}$$

This shows that $\mathbb{E}\left[\max_{\theta\in\Theta_{0,\epsilon}}\left|\sigma_t^2(\theta)-\sigma_{t\ell}^2(\theta)\right|^s\right]\leq Ca^{\ell}$ for some $a\in(0,1)$. The proof of

$$\mathbb{E}\left[\max_{\theta\in\Theta_{0,\epsilon}}\sigma_t^{2s}(\theta)\right]<\infty,\quad \mathbb{E}\left[\max_{\theta\in\Theta_{0,\epsilon}}\left|\sigma_t^2(\theta)-\overline{\sigma}_t^2(\theta)\right|^s\right]<\infty$$

is similar, using the expansion of $\underline{\sigma}_t^2(\theta)$ and the fact that $\mathbb{E}X_t^{2s}<\infty$.

2. The assumption **A3** follows from the fact that the random variables $\left|\dot{\sigma_t^2},\sigma_t^2\right|_{\infty,\epsilon}$ and $\left|\ddot{\sigma_t^2},\sigma_t^2\right|_{\infty,\epsilon}$ have moments of any order if ϵ is sufficiently small (see the proof Theorem 7.2 in Francq and Zakoïan (2010), part c for the main arguments used for showing these properties).
3. For checking **A4** and **A8**, one can use the Gaussian QML estimator. When all the GARCH coefficients are assumed to be positive, the representation **A8** holds for the corresponding estimator (see Francq and Zakoïan (2010), p. 159 − 160). Note that the representation **A8** requires the assumptions **G1** and **G2**.
4. Finally we check the assumptions **A5** or **A6**. First, we note that assumption **A5** does not hold when $v=m_0$. In this case, the ratio $\frac{v-m_0}{\sigma_t(\theta_0)}$ is degenerate and does not have a density. For estimating $f(m_0)$, one can use the classical kernel estimate, the approach proposed in this paper has no interest because the convergence rate will be similar. In the following, we assume that $m_0\notin I$. Using Lemma 7 given in the supplementary material and the representation of GARCH processes as an ARCH(∞) process, one can see that $\mathbb{E}\sigma_t(\theta_0)<\infty$ is sufficient for **A5**. Moreover, using the additional assumptions $\mathbb{E}\sigma_t(\theta_0)^e<\infty$ and $u\mapsto|u|^e f_\varepsilon(u)$ is bounded for $e>\frac{3}{2}$, assumption **A6** also holds if $I\subset(m_0,\infty)$ or $I\subset(-\infty,m_0)$. Note that the moment condition $\mathbb{E}\sigma_t(\theta_0)^e<\infty$ is satisfied under the classical condition $\sum_{j=1}^Q\alpha_{0,j}+\sum_{j=1}^P\beta_{0,j}<1$ which implies $\mathbb{E}\sigma_t(\theta_0)^2<\infty$.

Notes

1. If we assume that $m=0$ in the model, one can use the logarithm to get

$$\log\left(X_t^2\right)=\log\left(\sigma_t^2(\theta_0)\right)+\log\left(\varepsilon_t^2\right).$$

One can apply our results for directly estimating the density of $\log\left(X_t^2\right)$. Setting $Z_t=\log\left(\varepsilon_t^2\right)$ and $m_t(\theta_0)=\log\left(\sigma_t^2(\theta_0)\right)$, one can show that the density of Z_t satisfies the assumption **A5** if we also assume that $u\mapsto u^3 f_\varepsilon^{''}(u)$ is bounded. Moreover, we have $g_v(x) = f_m(v - x) = e^{v-x} f_{\sigma^2}\left(e^{v-x}\right)$ where f_m denotes the density of $m_t(\theta_0)$. One can then show that assumption **A6** is satisfied if $\mathbb{E}\sigma_t^2(\theta) < \infty$, which is a classical condition found in practice in using GARCH models. Moreover, it is also possible to show that assumption **A7** is satisfied under the additional condition: $u \mapsto u^{\frac{3}{2}+\delta} f_\varepsilon(u)$ is bounded. The proof is omitted since one can use the third point of Lemma 7 as well as some arguments used in the proof of Lemma 7. All the other assumptions are automatically satisfied if **G1** and **G2** hold true. Note that the root n consistent estimation of the density of $\log \sigma_t^2(\theta_0)$ is studied in Delaigle et al. (2015), for a GARCH$(1,1)$ process. We consider here, the estimation of the density of $\log X_t^2$ which is a different problem but this convergence rate also holds. Note also that we consider a more general GARCH(P,Q) model in this work.

2. One can prove similar results for pure ARCH processes (i.e $\beta_1 = \cdots = \beta_p = 0$ in the GARCH model), assuming $\alpha_1, \ldots, \alpha_q > 0$. However assumption **A6** (resp. assumption **A7**) requires $q \geq 2$ (resp. $q \geq 3$). See Lemma 7 for details. Let us show that assumption **A6** is not satisfied in the case of $q = 1$. In the case of $q = 1$, we have $\sigma_t^2(\theta_0) = \alpha_{00} + \alpha_{01} X_{t-1}^2$. Then if f_X is the marginal density of the process, we have
$$f_{\sigma^2}(y) = \mathbb{1}_{(\alpha_{00},\infty)}(y) \frac{\sqrt{\alpha_{01}}}{2\sqrt{y - \alpha_{00}}} f_X\left(\sqrt{\frac{y - \alpha_{00}}{\alpha 01}}\right).$$
Let us assume that $0 < \alpha_{01} < 1$, v is positive and $f(0) > 0$ (the last condition holds true when $f_\varepsilon(0) > 0$). We then have $\int g_v(x)^2 f_\varepsilon(x) dx = \infty$ when $f_\varepsilon\left(\frac{v}{\sqrt{\alpha_{00}}}\right) > 0$. However, one can check that $\int g_v(x)^{2-\delta} dx < \infty$ for any $\delta \in (0,1)$. Indeed, we have
$$\int g_v^{2-\delta}(x) dx \leq C \int_{\alpha_{00}}^{\infty} \sqrt{y} f_{\sigma^2}(y)^{2-\delta} dy$$
which is finite (the integrability holds around the singularity $y = \alpha_{00}$, f_{σ^2} is bounded outside a neighborhood of $\alpha_{0,0}$ and $\int \sqrt{y} f_{\sigma^2}(y) dy = \mathbb{E}\left[\sigma_t(\theta_0)\right] < \infty$). We recover a phenomenon described by Schick and Wefelmeyer (2009) when Frees estimator is applied for estimating the density of a sum of powers of two independent random variables. In general, a slower convergence rate is obtained when square integrability of the density fails. See Schick and Wefelmeyer (2009), Theorem 2, where the rate $\frac{n}{\log(n)}$ is obtained in the estimation of the density of a sum of squares $X_1^2 + X_2^2$.
3. When some parameters of the GARCH process are equal to zero, assumption **A3** is not always guaranteed unless assuming $\mathbb{E}\left(X_t^6\right) < \infty$. In this case assumption **A3** is automatic.

3.3. ARMA processes with GARCH noises

In this subsection, we consider the model

$$X_t - \eta_0 = \sum_{j=1}^{p} a_{0j} \left(X_{t-j} - \eta_0 \right) + Z_t - \sum_{j=1}^{q} b_{0j} Z_{t-j}, \quad Z_t = \varepsilon_t \sigma_t(\theta_0),$$

$$\sigma_t^2(\theta_0) = \alpha_{00} + \sum_{j=1}^{Q} \alpha_{0j} Z_{t-j}^2 + \sum_{j=1}^{P} \beta_{0j} \sigma_{t-j}^2(\theta_0).$$

We define for $\theta = (\eta, a_1, \ldots, a_p, b_1, \ldots, b_q, \alpha_0, \ldots, \alpha_Q, \beta_1, \ldots, \beta_P)$, As for ARMA processes, we define

$$Z_t(\theta) = X_t - \eta - \sum_{j=1}^{p} a_j \left(X_{t-j} - \eta \right) + \sum_{j=1}^{q} b_j Z_{t-j}(\theta),$$

$$\sigma_t^2(\theta) = \alpha_0 + \sum_{j=1}^{Q} \alpha_j Z_{t-j}^2(\theta) + \sum_{j=1}^{P} \beta_j \sigma_{t-j}^2(\theta).$$

In addition to assumption **G1** for the Garch parameters $(\alpha_0, \ldots, \alpha_Q, \beta_1, \ldots, \beta_P)$, we consider the following classical assumption which guarantees causality and invertibility of the ARMA part.

AG1 The roots of the two polynomials $\mathcal{P}$ and $\mathcal{Q}$ defined by

$$\mathcal{P}(z) = 1 - \sum_{j=1}^{p} a_{0,j} z^j, \quad \mathcal{Q}(z) = 1 - \sum_{j=1}^{q} b_{0,j} z^j$$

are outside the unit disc.

AG2 We have $\mathbb{E} Z_t^6 < \infty$.

Note the if $\epsilon > 0$ is sufficiently small, the assumption **AG1** will also be valid for all $\theta \in \Theta_{0,\epsilon}$. Assumption **AG2** is restrictive but we do not find a way to avoid this moment condition for checking assumption **A3**. This restriction is due to the technique used for the proof of Theorem 2, with a control of the derivative of our estimator with respect to θ when θ is close to θ_0. Note that, under the assumptions **AG1-AG2**, we have $\mathbb{E} X_t^6 < \infty$.

We now check the assumptions **A2** $-$ **A8**, except **A5** and **A7** which we were able to check only in the pure GARCH case.

1. For the assumption **A2**, one can use the following expansions for $\theta \in \Theta_{0,\epsilon}$. If

$$\underline{X}_t = \left(X_t, X_{t-1}, \ldots, X_{t-P+1} \right)^T, \quad \underline{Z}_t(\theta) = \left(Z_t(\theta), \ldots, Z_{t-Q+1}(\theta) \right)^T,$$

$$\underline{\sigma}_t^2(\theta) = \left(\sigma_t^2(\theta), \ldots, \sigma_{t-q+1}^2(\theta) \right)^T,$$

we have

$$\underline{X}_t = A_3(\theta)\underline{X}_{t-1} - \eta A_3(\theta)\mathbb{1} + \eta\mathbb{1} + B_{3,t}(\theta), \quad \underline{Z}_t(\theta) = A_1(\theta)\underline{Z}_{t-1}(\theta) + B_{1,t}(\theta),$$
$$\underline{\sigma}_t^2(\theta) = A_2(\theta)\underline{\sigma}_{t-1}^2(\theta) + B_{2,t}(\theta),$$

where $A_1(\theta)$, $A_2(\theta)$ and $A_3(\theta)$ are the companion matrices associated with $(a_1, \ldots, a_p)$, $(b_1, \ldots, b_q)$ and $(\beta_1, \ldots, \beta_P)$ respectively and

$$B_{2,t}(\theta) = \begin{pmatrix} \alpha_0 + \sum_{j=1}^{Q} \alpha_j Z_{t-j}^2(\theta) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad B_{3,t}(\theta) = \begin{pmatrix} Z_t(\theta) - \sum_{j=1}^{q} b_j Z_{t-j}(\theta) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$B_{1,t}(\theta) = \begin{pmatrix} X_t - \eta - \sum_{j=1}^{p} a_j (X_{t-j} - \eta) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then assumption **A2** follows from the fact that if $\epsilon > 0$ is sufficiently small, there exists three positive integers k_1, k_2, k_3 such that $\sup_{\theta \in \Theta_{0,\epsilon}} \|A_i(\theta)^{k_i}\| < 1$ for $i = 1, 2, 3$. The proof uses the same arguments as in the pure GARCH case and is omitted.

2. The assumption **A3** holds true if we assume **AG2**. In this case, using the expansions

$$\underline{\sigma}_t^2(\theta) = \sum_{j=0}^{\infty} A_2(\theta)^j B_{2,t-j}(\theta), \quad \underline{Z}_t(\theta) = \sum_{j=0}^{\infty} A_1(\theta)^j B_{1,t-j}$$

and the equation for $\underline{X}_t$ to show that on can show that $\left|\sigma_t^2\right|_{\infty,\epsilon}$, $\left|\dot{\sigma}_t^2\right|_{\infty,\epsilon}$ and $\left|\ddot{\sigma}_t^2\right|_{\infty,\epsilon}$ have a moment of order 3. Moreover, one can show that $|m_i|_{\infty,\epsilon}$, $|\dot{m}_i|_{\infty,\epsilon}$ and $|\ddot{m}_i|_{\infty,\epsilon}$ have a moment of order 6. This is sufficient for checking **A3**.

3. Assumptions **A4** and **A8** hold true if we consider the Gaussian quasi-maximum likelihood. See the proof of Theorem 7.5 in Francq and Zakoïan (2010). Note that the expansion given in **A8** requires that θ_0 lies in the interior of Θ and that $\mathbb{E}Z_t^4 < \infty$.

4. Assumption **A6** is a consequence of Lemma 8 given in the supplementary material. Indeed, if f_ε is bounded, the conditional density of $Z_t|Z_{t-1}, Z_{t-2}\ldots$ is bounded. Since the pair $(m_t(\theta_0), \sigma_t(\theta_0))$ can be expressed as $\left(\sum_{j=1}^{\infty} \psi_j Z_{t-j}, \sqrt{\alpha_0 + \sum_{j=1}^{\infty} \alpha_j Z_{t-j}^2}\right)$ for some summable sequences of coefficients $(\psi_j)_{j\geq 1}$ and $(\alpha_j)_{j\geq 1}$, Lemma 8 guarantees that the density $f_{m,\sigma}$ of this pair can be bounded as follows: $f_{m,\sigma}(x, y) \leq Cy$ for a positive constant C. One can then conclude using inequality (5) and assumption **AG2** which entails the condition $\mathbb{E}\sigma_t(\theta_0)^2 < \infty$.

4. Simulation study

In this section, by means of a simulation, we compare the mean-square error of our estimator with that of the classical kernel density estimate. Our estimator is implemented using the quadratic kernel. The standard kernel density estimate is computed using the function *Density* of the software R. Bandwidth selection for our estimator is beyond the scope of this paper. However, we use the simple approach proposed in Kim et al. (2015) which consists in multiplying the bandwidth selected for the kernel density estimate by a factor $n^{\frac{1}{5}-\kappa}$ where κ is compatible with our theoretical results. For a bandwidth $\hat{b} = \hat{C}n^{-\frac{1}{5}}$ with the optimal rate, we then keep the constant $\hat{C}$ and simply modify the rate. In our simulations, we found that the exponent $\kappa = 2/7$ provides good results.

In the following, we consider three simulation scenarios.

1. In the first scenario, we consider the conditionally homoscedastic case, with the AR process $X_t = 0.5X_{t-1} + \varepsilon_t$ such that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is i.i.d with $\varepsilon_1 \sim t(5)$ (the Student distribution with 5 degrees of freedom).
2. In the second scenario, we consider the pure ARCH case with a GARCH$(1,1)$ process $X_t = \varepsilon_t\sigma_t$ such that $\sigma_t^2 = 0.1 + 0.1X_{t-1}^2 + 0.8\sigma_{t-1}^2$. The noise component ε still follows a $t(5)$ distribution.
3. In the last scenario, we consider an AR process with a GARCH$(1,1)$ noise,

$$X_t = 0.5X_{t-1} + Z_t, \quad Z_t = \varepsilon_t\sigma_t, \quad \sigma_t^2 = 0.1 + 0.1X_{t-1}^2 + 0.8\sigma_t^2.$$

   We assume that ε_t follows a standard Gaussian. One can show that the moment condition $\mathbb{E}Z_t^6 < \infty$ required for applying our results, is not satisfied in this example.

Note that GARCH parameters are chosen so that the expectation of the square equals 1 and lag coefficients have typical values encountered in practice.

The marginal density is evaluated at 10 equally spaced points, starting from $v = 0$ to $v = 5$. The true density is approximated by

$$f_X(v) \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i}f_\varepsilon\left(\frac{v-m_i}{\sigma_i}\right), \quad N = 500000.$$

The (normalized) RMSE is $\sqrt{\mathbb{E}\left[\hat{f}_X(v) - f_X(v)\right]^2}/f_X(v)$. This RMSE is approximated by its empirical counterpart using 10^3 samples. GARCH parameters are estimated using the function *garchFit* of the package *fGarch*. The R code is available upon request from the author.

In Figure 1, 2 and 3, the blue curve represents the normalized RMSE for kernel density estimation and the red curve that for our method. Whatever the original bandwidth selection, our estimator performs better for estimating accurately GARCH processes, even if the sample size n is small. A notable exception is the second scenario when v is in a neighborhood of 0. In this case, the standard method performs better. This is not surprising because of the singularity at point $v = 0$, a point for which our method is almost equivalent to the standard one and our bandwidth parameter does not have the optimal convergence rate. This problem is less perceptible for the larger sample size $n = 500$. A general finding is the notable superiority of our method for estimating the tails, which have an important role in financial time series.

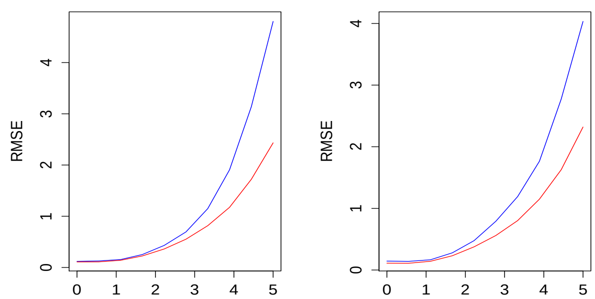

Figure 1. RMSE for the AR-GARCH process. The bandwidth is obtained using an optimal rate with constant 1 (left, $n = 200$) or cross-validation (right, $n = 200$)

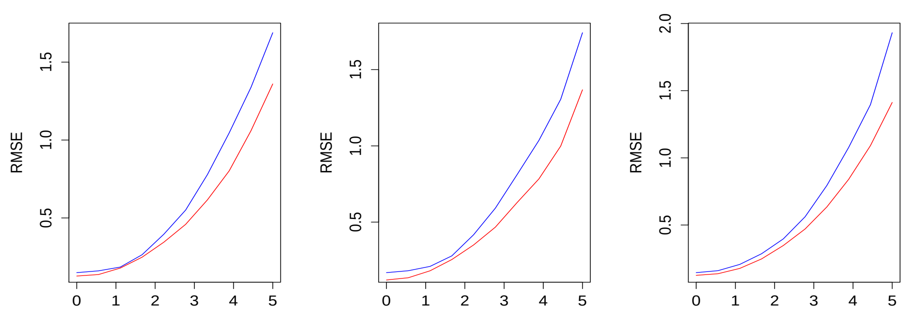

Figure 2. RMSE for the AR process. The bandwidth for the kernel estimate is obtained by Silverman's rule of thumb (left, $n = 100$), cross-validation (middle, $n = 100$) and optimal rate with constant 1 (right, $n = 100$)

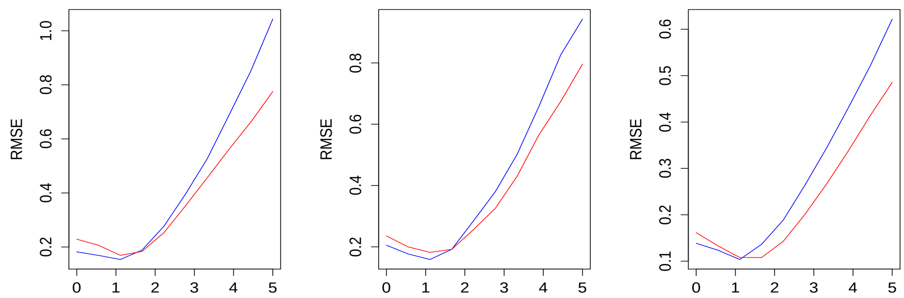

Figure 3. RMSE for the GARCH process. The bandwidth is obtained using an optimal rate with constant 1 (left, $n = 200$), cross-validation (middle, $n = 200$) and cross-validation (right, $n = 500$)

5. Proofs of the results

In the subsequent proofs, $C > 0$ will denote a generic constant that can change from line to line. Moreover, if X is a random variable, we set $\overline{X} = X - \mathbb{E}(X)$.

For deriving our results, we first introduce a coupling method which will be very useful. The goal of this coupling method is to construct $\ell-$dependent random sequences which approximate some weakly dependent random sequences. We then show that our initial estimator is asymptotically equivalent to an estimator involving an $\ell-$dependent random sequence, provided that $\ell = \ell_n$ grows with a polynomial rate.

5.1. From dependence to $\ell-$dependence via coupling

Let $F : \mathbb{R}^{\mathbb{N}} \to G$ be a measurable application taking values to an arbitrary measurable space $(G, \mathcal{G})$. If $Z_i = F(\varepsilon_i, \varepsilon_{i-1}, \ldots)$, we set

$$Z_{i\ell} = F\left(\varepsilon_i, \ldots, \varepsilon_{i-\ell+1}, \varepsilon^{(i)}_{i-\ell}, \varepsilon^{(i)}_{i-\ell-1}, \ldots\right),$$

where $\left\{\varepsilon^{(i)}_t : (i,t) \in \mathbb{Z}^2\right\}$ is a family of i.i.d random variables independent of $(\varepsilon_t)_{t\in\mathbb{Z}}$ and such that for all $(i,t) \in \mathbb{Z}^2$, $\varepsilon^{(i)}_t$ has the same distribution as ε_0. In this case, the sequence $(Z_{i\ell})_{i\in\mathbb{Z}}$ is $\ell-$dependent. This means that for all $i \in \mathbb{Z}$, the two $\sigma-$algebra $\sigma(Z_{j\ell} : j \geq i)$ and $\sigma(Z_{j\ell} : j \leq i - \ell)$ are independent. We call this new sequence an $\ell-$dependent approximation of $(Z_i)_{i\in\mathbb{Z}}$. Note that $Z_{i\ell}$ has the same distribution as Z_i. Note that the two processes $(m_i)_{i\in\mathbb{Z}}$ and $(\sigma_i)_{i\in\mathbb{Z}}$ are of this form, if G denotes the set of real-valued functions defined on the set Θ. We will denote their corresponding $\ell-$dependent approximations by $m_{i\ell}$ and $\sigma_{i\ell}$. These coupling versions of the conditional mean/variance of the process will be central to our proofs. One can note that $m_{i\ell}, \sigma_{i\ell}$ and their derivatives with respect to θ have the same distribution as the original quantities.

5.2. A martingale decomposition

The control of the derivative of our estimator will be done using appropriate martingale differences. In this subsection, we set for $i \in \mathbb{Z}$,

$$Y_{i\ell} = \left(\varepsilon_i, \varepsilon_{i-1} \ldots, \varepsilon_{i-\ell+1}, \varepsilon^{(i)}_{i-\ell}, \varepsilon^{(i)}_{i-\ell-1}, \ldots\right).$$

Let $n' = k\ell$ and $\mathcal{I}_n = \{1 \leq i, j \leq n' : i \leq j - \ell \text{ or } i \geq j + \ell\}$. Here k denotes the integer part of the ratio n/ℓ and $\ell \in (0, n)$ is an integer. For $s = 1, 2, \ldots, \ell$, we set $[s] = \{s + g\ell : 0 \leq g \leq k - 1\}$. For each s, we define two filters. We set for $g = 0, 1, \ldots, k - 1$,

$$\mathcal{G}_{s,g} = \sigma\left(\varepsilon_i, \varepsilon^{(i)} : i \leq s + g\ell\right), \quad \overline{\mathcal{G}}_{s,g} = \sigma\left(\varepsilon_i, \varepsilon^{(i)} : i > n' - s - g\ell\right).$$

Now if $T_{ij}(v, \theta)$ is a random variable measurable with respect to $\sigma\left(Y_{i\ell}, Y_{j\ell}\right)$, we set

$$M_s(T)_g(v, \theta) = \sum_{i=1}^{s+(g-1)\ell} \left[T_{i,s+g\ell}(v, \theta) - \mathbb{E}\left(T_{i,s+g\ell}(v, \theta)|\mathcal{G}_{s,g-1}\right)\right],$$

and $\overline{M}_s(T)_g(v, \theta) = \sum_{i=n'-s-(g-1)\ell}^{n'} \left[T_{i,n'-s-g\ell}(v, \theta) - \mathbb{E}\left(T_{i,n'-s-g\ell}(v, \theta)|\overline{\mathcal{G}}_{s,g}\right)\right]$. Then for each $s = 1, \ldots, \ell$, $\{(M_s(T)_g, \mathcal{G}_{s,g}) : 0 \leq g \leq k - 1\}$ and $\left\{\left(\overline{M}_s(T)_g, \overline{\mathcal{G}}_{s,g}\right) : 0 \leq g \leq k - 1\right\}$ are two martingale differences. Moreover, if $\mathbb{E}_{Y_{i\ell}}$ denotes integration with respect to the distribution of $Y_{i\ell}$, we have

$$\sum_{(i,j)\in\mathcal{I}_n} \left[T_{ij}(v, \theta) - \mathbb{E}_{Y_{j\ell}}\left(T_{ij}(v, \theta)\right)\right] = \sum_{s=1}^{\ell} \left[\sum_{g=0}^{k-1} M_s(T)_g(v, \theta) + \sum_{g=0}^{k-1} \overline{M}_s(T)_g(v, \theta)\right]. \quad (9)$$

5.3. Proof of Theorem 1

For simplicity of notations, we drop the parameter θ_0 and simply write for instance σ_i instead of $\sigma_i(\theta_0)$. Let $0 < \delta_1 < \delta$ and $t \in (0, 1/2)$ sufficiently small such that $\frac{n^t}{nb^{2+\delta_1}} \to 0$ and $\frac{n^t}{\sqrt{n}b} \to 0$. We denote by ℓ the integer part of n^t and by k the integer part of the ratio n/ℓ. We then set $n' = k\ell$ and $\mathcal{I}_n = \{1 \leq i, j \leq n' : i \leq j - \ell \text{ or } i \geq j + \ell\}$. Note that the cardinal $|\mathcal{I}_n|$ of the set $\mathcal{I}_n$ satisfies $(n - \ell)(n - 3\ell - 1) \leq |\mathcal{I}_n| \leq n^2$. We set $\check{f}_\ell(v) = \frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n} A_{v,ij}, \quad A_{v,ij} = \frac{1}{\sigma_{i\ell}} K_b\left(\frac{v-m_{i\ell}}{\sigma_{i\ell}} - \varepsilon_j\right)$. Note first that

$$\sqrt{n} \sup_{v\in I} \left|\check{f}_X(v) - \check{f}_\ell(v)\right| = o_{\mathbb{P}}(1). \quad (10)$$

From Lemma 2, $\sqrt{n}\sup_{v\in I}\left|\check{f}_X(v) - \frac{1}{n^2}\sum_{1\leq i,j\leq n} A_{v,ij}\right| = o_{\mathbb{P}}(1)$ and there exists a constant $C > 0$ such that $\sqrt{n}\sup_{v\in I}\left|\check{f}_\ell(v) - \frac{1}{n^2}\sum_{1\leq i,j\leq n} A_{v,ij}\right| \leq \frac{C\ell}{\sqrt{n}b} \to 0$. Hence (10) follows. In the following, we study the behavior of the estimator $\check{f}_\ell(v)$.

5.3.1. Bias part

We recall that $f_X(v) = \mathbb{E}\left[\frac{1}{\sigma_i} f_\varepsilon\left(\frac{v-m_i}{\sigma_i}\right)\right]$. Since $f_\varepsilon^{''}$ is bounded, there exists a $C > 0$ such that for all $x, h \in \mathbb{R}$,

$$|f_\varepsilon(x+h) - f_\varepsilon(x) - hf_\varepsilon'(x)| \leq Ch^2. \tag{11}$$

From (11), we deduce that

$$\begin{aligned} \mathbb{E}\left[A_{v,ij}\right] - f(v) &= \mathbb{E}\int \frac{1}{\sigma_i} K(w)\left[f_\varepsilon\left(\frac{v-m_i}{\sigma_i} - bw\right) - f_\varepsilon\left(\frac{v-m_i}{\sigma_i}\right)\right] dw \\ &= O\left(b^2\right). \end{aligned}$$

Using the condition $\sqrt{n}b^2 \to 0$, we get $\sqrt{n}\left(\mathbb{E}\left[\check{f}_\ell(v)\right] - f(v)\right) = o(1)$.

5.3.2. The variance part

The main points of this part concern the limiting behavior of two $U-$statistics

$$\frac{1}{n^{3/2}} \sum_{j=\ell}^{n'-1} \sum_{i=n'-j+\ell}^{n'} A_{v,i(n'-j)} \text{ and } \frac{1}{n^{3/2}} \sum_{j=\ell+1}^{n'} \sum_{i=1}^{j-\ell} A_{v,i(n'-j)}. \tag{12}$$

We focus mainly on the degenerate part of the first one. We set

$$\begin{aligned} S_{v,ij} &= A_{v,i(n'-j)} - \frac{1}{\sigma_{i\ell}} \int K(w) f_\varepsilon\left[\frac{v-m_{i\ell}}{\sigma_{i\ell}} - bw\right] dw \\ &- \int K(h) g_v\left(\varepsilon_{n'-j} + bh\right) dh + \mathbb{E}\left[A_{v,i(n'-j)}\right], \end{aligned}$$

one can show (see the supplementary material) that

$$\sup_{v\in I} \frac{1}{n^{3/2}} \left|\sum_{j=\ell}^{n'-1} \sum_{i=n'-j-\ell}^{n} S_{v,ij}\right| = o_{\mathbb{P}}(1). \tag{13}$$

Using the same type of arguments, one can also show that the degenerate part of the second U statistics in (12) satisfies $\sup_{v\in I} \frac{1}{n^{3/2}} \left|\sum_{j=\ell+1}^{n'} \sum_{i=1}^{j-\ell} \bar{S}_{v,ij}\right| = o_{\mathbb{P}}(1)$, with

$$\bar{S}_{v,ij} = A_{v,ij} - \int \frac{K(w)}{\sigma_{i\ell}} f_\varepsilon\left[\frac{v-m_{i\ell}}{\sigma_{i\ell}} - bw\right] dw - \int K(h) g_v\left(\varepsilon_j + bh\right) dh + \mathbb{E}\left[A_{v,ij}\right].$$

Next, we discuss the behavior of the two empirical process parts of the $U-$statistics (12). Since f_ε is $\mathcal{C}^2$, we have (see also the control of the bias)

$$\frac{1}{n^{3/2}} \sup_{v\in I} \left|\sum_{(i,j)\in\mathcal{I}_n} \int \frac{K(w)}{\sigma_{i\ell}} \left[f_\varepsilon\left[\frac{v-m_{i\ell}}{\sigma_{i\ell}} - bw\right] - f_\varepsilon\left[\frac{v-m_{i\ell}}{\sigma_{i\ell}}\right]\right] dw\right| = O_{\mathbb{P}}\left(\sqrt{n}b^2\right).$$

The next tedious step of the proof consists in showing that

$$\frac{1}{n^{3/2}} \sup_{v\in I} \left| \sum_{(i,j)\in\mathcal{I}_n} \int K(h) \left[g_v(\varepsilon_j + bh) - g_v(\varepsilon_j) \right] \right| = o_{\mathbb{P}}(1). \tag{14}$$

A proof of (14) is also given in the supplementary material.

5.4. End of the proof of Theorem 1

Collecting the results of the two previous subsections, we have shown that for $c_{n,j} = n^{-1}\sum_{i=1}^{n'} \mathbb{1}_{|i-j|\geq \ell}$,

$$\sqrt{n}\left(\check{f}_X(v) - f_X(v)\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n'} c_{n,j}\overline{g_v\left(\varepsilon_j\right)} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n'} c_{n,i}\overline{\frac{1}{\sigma_{i\ell}} f_\varepsilon\left(\frac{v - m_{i\ell}}{\sigma_{i\ell}}\right)} + o_{\mathbb{P}}(1),$$

uniformly on I and with a uniform convergence on I for the partial sum involving g_v if assumption **A7** holds true. It is then straightforward to show that one can replace $c_{n,i}$ and $c_{n,j}$ by 1 in this asymptotic expansion. Indeed, we have

$$\sqrt{n}\left|c_{n,i} - 1\right| \leq \frac{n - n' + 2\ell - 1}{\sqrt{n}} \to 0$$

and for the uniform convergence over I, one can use the bounds

$$g_v(\varepsilon_j) \leq G(\varepsilon_j), \quad \frac{1}{\sigma_{i\ell}} f_\varepsilon\left(\frac{v - m_{i\ell}}{\sigma_{i\ell}}\right) \leq \frac{1}{\gamma}\|f_\varepsilon\|_\infty.$$

Using the same arguments, n' can be replaced with n. Finally, using the arguments given in the proof of Lemma 2, we have

$$\frac{1}{\sqrt{n}} \sup_{v\in I} \sum_{i=1}^{n} \left| \sigma_{i\ell}^{-1} f_\varepsilon\left(\frac{v - m_{i\ell}}{\sigma_{i\ell}}\right) - \sigma_i^{-1} f_\varepsilon\left(\frac{v - m_i}{\sigma_i}\right) \right| = o_{\mathbb{P}}(1).$$

The proof of the tightness of $v \mapsto \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \sigma_i^{-1}\left[f_\varepsilon\left(\frac{v-m_i}{\sigma_i}\right) - f_X(v)\right]$ will be studied in detail in the proof of Corollary 1.□

5.5. Proof of Theorem 2

Using Lemma 1 and our bandwidth conditions, we have $\sup_{v\in I}\left|\hat{f}_X(v) - \widetilde{f}_X(v)\right| = o_{\mathbb{P}}(1/\sqrt{n})$, where $\widetilde{f}_X$ is defined as $\hat{f}_X$ but the quantities $\overline{m}_i$ and $\overline{\sigma}_i$ being replaced with m_i and σ_i respectively. This shows that possible truncations of the conditional mean/variance only using $X_1, \dots, X_n$ are asymptotically negligible for our estimator. Now, for $\theta \in \Theta$, we recall that $(m_{i\ell}(\theta))_{i\in\mathbb{Z}}$ and $(\sigma_{i\ell}(\theta))_{i\in\mathbb{Z}}$ denote the $\ell-$dependent approximations of $(m_i(\theta))_{i\in\mathbb{Z}}$ and $(\sigma_i(\theta))_{i\in\mathbb{Z}}$ respectively. Then setting $X_{j\ell} = m_{j\ell}(\theta_0) + \varepsilon_j \sigma_{j\ell}(\theta_0)$, we define

$$L_{v,i\ell}(\theta) = \frac{v - m_{i\ell}(\theta)}{\sigma_{i\ell}(\theta)}, \quad \varepsilon_{j\ell}(\theta) = \frac{X_{j\ell} - m_{j\ell}(\theta)}{\sigma_{j\ell}(\theta)}$$

and $\widetilde{f}_\ell(v) = \frac{1}{n^2}\sum_{1\le i,j\le n} \frac{1}{\sigma_{i\ell}(\hat{\theta})} K_b\left[L_{v,i\ell}(\hat{\theta}) - \varepsilon_{j\ell}(\hat{\theta})\right]$. In the rest of the proof, we fix a real number t such that $0 < t < \frac{\delta}{2(3+\delta)}$ and we denote by ℓ the integer part of n^t. Using Lemma 2, we have

$$\sup_{v\in I}\left|\widetilde{f}_X(v) - \widetilde{f}_\ell(v)\right| = o_{\mathbb{P}}\left(1/\sqrt{n}\right).$$

We will also suppress some terms in the estimator $\widetilde{f}_\ell$ in order to get stochastic independence between the two pairs of random functions $(m_{i\ell}, \sigma_{i\ell})$ and $(m_{j\ell}, \sigma_{j\ell})$ involved in the U-statistic. To this end, for $\theta \in \Theta$, $v \in I$ and $1 \le i, j \le n$, we set

$$A_{v,ij}(\theta) = \frac{1}{\sigma_{i\ell}(\theta)} K_b\left[L_{v,i\ell}(\theta) - \varepsilon_{j\ell}(\theta)\right].$$

Using the condition $\frac{\ell}{\sqrt{n}b} = o(1)$, we have $\sup_{v\in I}\left|\widetilde{f}_\ell(v) - \frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n} A_{v,ij}(\hat{\theta})\right| = o_{\mathbb{P}}(1)$.

5.5.1. *Outline of the proof*

The goal of the proof is to show that

$$\sup_{v\in I}\left|\frac{1}{n^{3/2}}\sum_{1\le i,j\le n}\left[A_{v,ij}(\hat{\theta}) - A_{v,ij}(\theta_0) - \dot{A}_{v,ij}(\theta_0)^T(\hat{\theta} - \theta_0)\right]\right| = o_{\mathbb{P}}(1) \tag{15}$$

and in a second step that

$$\sup_{v\in I}\left|\frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n}\dot{A}_{v,ij}(\theta_0) - \dot{h}_{\theta_0}(v)\right| = o_{\mathbb{P}}(1). \tag{16}$$

In the proof of Theorem 1, we have already shown that

$$\sqrt{n}\sup_{v\in I}\left|\check{f}_X(v) - \frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n} A_{v,ij}(\theta_0)\right| = o_{\mathbb{P}}(1).$$

Note that from assumption **A4**, assertion (15) will hold if for all $M > 0$ and integers n such that $M/\sqrt{n} < \epsilon$, we have

$$\sup_{v\in I,\theta\in\Theta_{0,n}}\left|\frac{1}{n^2}\sum_{1\le i,j\le n}\left[A_{v,ij}(\theta) - A_{v,ij}(\theta_0) - \dot{A}_{v,ij}(\theta_0)^T(\theta - \theta_0)\right]\right| = o_{\mathbb{P}}(1/\sqrt{n}),$$

where $\Theta_{0,n}$ is a short notation for $\Theta_{0,M/\sqrt{n}}$. We will show the following sufficient condition

$$\sup_{v\in I,\theta\in\Theta_{0,n}}\left\|\frac{1}{n^2}\sum_{1\le i,j\le n}\left[\dot{A}_{v,ij}(\theta)-\dot{A}_{v,ij}(\theta_0)\right]\right\|=o_{\mathbb{P}}(1).$$

The two assertions (15) and (16) (and Theorem 2) will then follow if we show that

$$\frac{1}{n^2}\sup_{v\in I,\theta\in\Theta_{0,n}}\left\|\sum_{(i,j)\in\mathcal{I}_n}\left[\dot{A}_{v,ij}(\theta)-\mathbb{E}_{Y_{j\ell}}\dot{A}_{v,ij}(\theta)\right]\right\|=o_{\mathbb{P}}(1),\tag{17}$$

$$\frac{1}{n^2}\sup_{v\in I,\theta\in\Theta_{0,n}}\left\|\sum_{(i,j)\in\mathcal{I}_n}\left[\mathbb{E}_{Y_{j\ell}}\dot{A}_{v,ij}(\theta)-\mathbb{E}_{Y_{j\ell}}\dot{A}_{v,ij}(\theta_0)\right]\right\|=o_{\mathbb{P}}(1)\tag{18}$$

and

$$\sup_{v\in I}\left\|\frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n}\mathbb{E}_{Y_{j\ell}}\dot{A}_{v,ij}(\theta_0)-\dot{h}_{\theta_0}(v)\right\|=o_{\mathbb{P}}(1),\tag{19}$$

where the function h_θ is defined before the statement of Theorem 2. Assertion (17) will be studied using martingale properties (see the subsection 5.2). Proofs of assertions (18) and (19) follow from more tedious arguments and are given in the supplementary material. Note that we have the following expression.

$$\dot{A}_{v,ij}(\theta)=\dot{\sigma^{-1}}_{i\ell}(\theta)K_b\left[L_{v,i\ell}(\theta)-\varepsilon_{j\ell}(\theta)\right]+\frac{\dot{L}_{v,i\ell}(\theta)-\dot{\varepsilon}_{j\ell}(\theta)}{\sigma_{i\ell}(\theta)}K_b'\left[L_{v,i\ell}(\theta)-\varepsilon_{j\ell}(\theta)\right].$$

Proof of assertion (17) We set

$$S_n(v,\theta)=\frac{1}{n^2}\sum_{(i,j)\in\mathcal{I}_n}\left[\dot{A}_{v,ij}(\theta)-\mathbb{E}_{Y_{j\ell}}\left(\dot{A}_{v,ij}(\theta)\right)\right].$$

Let $\eta=\eta_n$ a sequence of positive real numbers such that $\eta/b^3=o_{\mathbb{P}}(1)$. We take for instance $\eta=n^{-\frac{3}{3+\delta}}$. Let $\{(v_h,\theta_h):h\in\mathcal{H}\}$ a family of points in $I\times\Theta_{0,n}$ such that for $(v,\theta)\in I\times\Theta_{0,n}$, there exists $h\in\mathcal{H}$ such that $\max\{|v-v_h|,\|\theta-\theta_h\|\}\le\eta$. The set $\mathcal{H}$ can be chosen such that $|\mathcal{H}|=O\left(\eta^{-d-1}\right)=O\left(n^{\frac{3(d+1)}{3+\delta}}\right)$. Using Lemma 4 (3), we first notice that

$$\sup_{(v,\theta)\in I\times\Theta_{0,n}}\|S_n(v,\theta)\|-\sup_{(v,\theta)\in\mathcal{H}}\|S_n(v,\theta)\|=O_{\mathbb{P}}\left(\frac{\eta}{b^3}\right)=o_{\mathbb{P}}(1).$$

To show (17), it remains to prove that

$$\sup_{(v,\theta)\in\mathcal{H}}\|S_n(v,\theta)\|=o_{\mathbb{P}}(1).$$

But this is a consequence of Lemma 3 applied coordinatewise to $S_n(\cdot,\cdot)$, using the martingale decomposition (9). The assumptions used in Lemma 3 can be checked using Lemma 4.

5.6. Auxiliary Lemmas

This subsection contains several auxiliary lemmas which are central to the proof of Theorem 2. Their proofs are given in the supplementary material.

The two first results assert that under our assumptions, truncated versions $\overline{m}_t$ and $\overline{\sigma}_t$ of m_t and σ_t have no effect in the asymptotic expansion of our estimator. Moreover, m_t and σ_t can be replaced with their $\ell-$dependent approximations, provided $\ell = \ell_n$ grows at an arbitrary small power of n.

In what follows, we set for $1 \leq i, j \leq n$ and $(v, \theta) \in I \times \Theta_{0,\epsilon}$,

$$\mathcal{L}_{v,ij}(\theta) = L_{v,i}(\theta) - \varepsilon_j(\theta), \quad \mathcal{L}_{v,ij\ell}(\theta) = L_{v,i\ell}(\theta) - \varepsilon_{j\ell}(\theta), \quad \overline{\mathcal{L}}_{v,ij}(\theta) = \overline{L}_{v,i}(\theta) - \overline{\varepsilon}_j(\theta).$$

Here $L_{v,ij\ell}(\theta) = \frac{v - m_{i\ell}(\theta)}{\sigma_{i\ell}(\theta)}$ and $\varepsilon_{j\ell}(\theta) = \frac{X_{j\ell} - m_{j\ell}(\theta)}{\sigma_{j\ell}(\theta)}$. We also set

$$\widetilde{f}_X(v) = \frac{1}{n^2} \sum_{i,j=1}^{n} \frac{1}{\sigma_i(\hat{\theta})} K_b \left[\mathcal{L}_{v,ij}(\hat{\theta}) \right].$$

Lemma 1. *Assume that there exists $\delta \in (0,1)$ such that $nb^{2+\delta} \to \infty$. We then have* $\sup_{v \in I} \left| \hat{f}_X(v) - \widetilde{f}_X(v) \right| = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right)$.

Lemma 2. *Assume that $\ell = n^t$ with $t > 0$ and that $nb \to \infty$. Then if*

$$\widetilde{f}_\ell(v) = \frac{1}{n^2} \sum_{i,j=1}^{n} \frac{1}{\sigma_{i\ell}(\hat{\theta})} K_b \left[L_{v,i\ell}(\hat{\theta}) - \varepsilon_{j\ell}(\hat{\theta}) \right].$$

We have $\sup_{v \in I} \left| \widetilde{f}_X(v) - \widetilde{f}_\ell(v) \right| = o_{\mathbb{P}} \left(n^{-1/2} \right)$.

The following lemma will be needed in the following for controlling triangular arrays of martingale differences. Its proof is based on the Freedman's inequality for martingales.

Lemma 3. *Assume that for a $\delta \in (0,1)$, $nb^{3+\delta} \to \infty$. Let $(\mathcal{H}_n)_n$ be a sequence of finite sets such that $|\mathcal{H}_n| = O(n^\gamma)$ for some $\gamma > 0$. For each $h \in \mathcal{H}_n$, let $\left(\xi_{n,g}^{(h,s)}, \mathcal{T}_{n,g}^{(s)} \right)_{0 \leq g \leq k_n}$ be a martingale difference. We assume that*

$$\max_{h \in \mathcal{H}_n} \sum_{s=1}^{\ell_n} \sum_{g=1}^{k_n} \mathbb{E} \left(\left| \xi_{n,g}^{(h,s)} \right|^2 | \mathcal{T}_{n,g}^{(s)} \right) = O_{\mathbb{P}} \left(\frac{1}{nb^3} \right)$$

and that

$$\max_{\substack{h \in \mathcal{H}_n \\ 1 \leq s \leq \ell_n \\ 1 \leq g \leq k_n}} \left| \xi_{n,g}^{(h,s)} \right| = O_{\mathbb{P}} \left(\frac{1}{n^{2/3} b^2} \right).$$

with $\ell_n = O\left(n^t\right)$, $0 < t < \frac{\delta}{2(3+\delta)}$. *We then get the conclusion:*

$$\max_{h\in\mathcal{H}_n}\left|\sum_{s=1}^{\ell_n}\sum_{g=1}^{k_n}\xi_{n,g}^{(h,s)}\right| = o_{\mathbb{P}}(1).$$

We now state a lemma which will be useful for checking the assumptions of the previous lemma in our context. We set $n' = k\ell$ where k is the integer part of n/ℓ. Moreover, we set

$$\mathcal{I}_n = \left\{1 \leq i, j \leq n' : i \leq j - \ell \text{ or } i \geq i + \ell\right\}.$$

Lemma 4. *Assume that assumptions* **A3***and* **A5** *hold true.*

1. *We set* $B_{v,ij}(\theta) = \dot{A}_{v,ij}(\theta) - \mathbb{E}_{Y_{j\ell}}\left[\dot{A}_{v,ij}(\theta)\right]$. *Then we have*

$$\sum_{(i,j)\in\mathcal{I}_n}\sup_{\substack{v\in I\\ \theta\in B(\theta_0,\epsilon)}}\mathbb{E}_{Y_{j\ell}}\|B_{v,ij}(\theta)\|^2 = O_{\mathbb{P}}\left(\frac{n^2}{b^3}\right).$$

2. *We have*

$$\max_{1\leq j\leq n}\sum_{\substack{i:(i,j)\in\mathcal{I}_n\\ 1\leq i\leq n}}^{n}\sup_{v\in I}|B_{v,ij}|_{\infty,\epsilon} = O_{\mathbb{P}}\left(\frac{n^{4/3}}{b^2}\right).$$

3. *Let* η *be a positive number. We have*

$$\sum_{(i,j)\in\mathcal{I}_n}\sup_{\substack{|v-w|\leq\eta\\ \|\theta-\zeta\|\leq\eta}}\|B_{v,ij}(\theta) - B_{w,ij}(\zeta)\| = O_{\mathbb{P}}\left(\frac{\eta n^2}{b^3}\right).$$

References

Y. Aït-Sahalia. Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9:385–426, 1996.

P. Ango Nze and P. Doukhan. Functional estimation for time series: Uniform convergence properties. *J. Statist. Plann. Inference*, 68:5–29, 1998.

D. Bosq. *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, volume 110. Lecture Notes in Statistics. Springer, 1998.

P. J. Brockwell and R.A. Davis. *Time Series: Theory and Methods. Second Edition.* Springer Series in Statistics, 1991.

V. Corradi and N.R. Swanson. Bootstrap specification tests for diffusion processes. *Journal of Econometrics*, 124:117–148, 2005.

J. Dedecker, P. Doukhan, J.-R. Leon, S. Louhichi, and C. Prieur. *Weak dependence, with examples and some applications.* Springer Berlin, 2007.

A. Delaigle, A. Meister, and J. Rombouts. Root-t consistent density estimation in garch models. *Journal of Econometrics*, page http://dx.doi.org/10.1016/j.jeconom.2015.10.009, 2015.

P. Doukhan and S. Louhichi. Functional estimation of a density under a new weak dependence condition. *Scand. J. Statist.*, 28:325–341, 2001.

C. Francq and J-M. Zakoan. Estimating the marginal law of a time series with applications to heavy-tailed distributions. *The Review of Financial Studies*, 31:412–425, 2013.

C. Francq and J-M. Zakoïan. *GARCH models: structure, statistical inference and financial applications*. Wiley, 2010.

E.W. Frees. Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89:517–525, 1994.

J. Gao and M. King. Adaptive testing in continuous-time diffusion models. *Econometric Theory*, 20:844–882, 2004.

E. Giné and D.M. Mason. On local u-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, 35:1105–1145, 2007.

K.H Kim, T. Zhang, and W.B. Wu. Parametric specification test for nonlinear autoregressive models. *Econometric Theory*, 31:1078–1101, 2015.

S. Li and Y. Tu. $\sqrt{n}-$consistent density estimation in semiparametric regression models. *Computational Statistics & Data Analysis*, 104:91–109, 2016.

Y. Liao and J. Stachurski. Simulation-based density estimation for time series using covariate data. *Journal of Business & Economic Statistics*, 33:595–606, 2015.

W. Liu and W.B. Wu. Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38:2388–2421, 2010.

U. U. Müller. Estimating the density of a possibly missing response variable in nonlinear regression. *Journal of Statistical Planning and Inference*, 142:1198–1214, 2012.

P.M. Robinson. Nonparametric estimators for time series. *J. Time Ser. Anal.*, 4:185–207, 1983.

G.G. Roussas. Asymptotic normality of the kernel estimate of a probability density function under association. *Statist. Probab. Lett.*, 50:1–12, 2000.

A. Saavedra and R. Cao. Rate of convergence of a convolution-type estimator of the marginal density of an ma(1) process. *Stochastic Process. Appl.*, 80:129–155, 1999.

A. Schick and W. Wefelmeyer. Root n consistent and optimal density estimators for moving average processes. *Scand. J. Statist.*, 31:63–78, 2004a.

A. Schick and W. Wefelmeyer. Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, 16:925–935, 2004b.

A. Schick and W. Wefelmeyer. Uniformly root$-n$ consistent density estimators for weakly dependent invertible linear processes. *Ann. Statist.*, 35:815–843, 2007.

A. Schick and W. Wefelmeyer. Convergence rates of density estimators for sums of powers of observations. *Metrika*, 69:249–264, 2009.

H. Tong. Non-linear time series. *A Dynamical System Approach*, 1990.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Z. Zhao. Density estimation for nonlinear parametric models with conditional heteroscedasticity. *Journal of Econometrics*, 155:71–82, 2010.