Consistent estimation of the Value-at-Risk when the error distribution of the volatility model is misspecified

El Ghourabi, Mohamed and Francq, Christian and Telmoudi, Fedya

University of Tunis, CREST and University Lille 3, University Lille 3

October 2013

Consistent estimation of the Value-at-Risk when the error distribution of the volatility model is misspecified [*]

Mohamed El Ghourabi, Christian Francq and Fedya Telmoudi [†]

University of Tunis, CREST and University Lille 3, University Lille 3

Abstract : A two-step approach for conditional Value at Risk (VaR) estimation is considered. In the first step, a generalized-quasi-maximum likelihood estimator (gQMLE) is employed to estimate the volatility parameter, and in the second step the empirical quantile of the residuals serves to estimate the theoretical quantile of the innovations. When the instrumental density h of the gQMLE is not the Gaussian density utilized in the standard QMLE, or is not the true distribution of the innovations, both the estimations of the volatility and of the quantile are asymptotically biased. The two errors however counterbalance each other, and we finally obtain a consistent estimator of the conditional VaR. For a wide class of GARCH models, we derive the asymptotic distribution of the VaR estimation based on gQMLE. We show that the optimal instrumental density h depends neither on the GARCH parameter nor on the risk level, but only on the distribution of the innovations. A simple adaptive method based on empirical moments of the residuals makes it possible to infer an optimal element within a class of potential instrumental densities. Important asymptotic efficiency gains are achieved by using gQMLE instead of the usual Gaussian QML when the innovations are heavy-tailed. We extended our approach to Distortion Risk Measure parameter estimation, where consistency of the gQMLE-based method is also proved. Numerical illustrations are provided, through simulation experiments and an application to financial stock indexes.

Jel Classification : C22 and C58

Keywords : APARCH, Conditional VaR, Distortion Risk Measures, GARCH, Generalized Quasi Maximum Likelihood Estimation, Instrumental density.

---

[*]The authors gratefully acknowledge financial support of the ANR via the ECONOM&RISK Project (ANR 2010 blanc 1804 03).

[†]Corresponding author: Fedya Telmoudi, University Lille 3 (EQUIPPE), BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: fedya.telmoudi@etu.univ-lille3.fr

1 Introduction

Financial market risk is usually perceived as the exposure to potential losses of portfolios of risky assets. To assess the risk level, practitioners rest on risk management tools, such as the notorious Value-at-Risk (VaR). In the late 1980, financial firms began the use of VaR, defined as the loss that should not be reached for a given position over a holding time period and at a certain confidence level.

The VaR is often estimated by a simple quantile of the historical returns. This practice implicitly assumes that the sequence of the returns is stationary, and neglects the dynamics, in particular this does not account for the existence of clusters of extreme returns. It is preferable to take into account the information available, by reasoning on the conditional distribution of the returns (see *e.g.* McNeil, Frey and Embrechts (2005) and Kuester, Mittnik and Paolella (2006), who clearly showed that unconditional models of VaR are outperformed by conditional ones). The VaR conditional on past observations will be called the conditional VaR [1].

More precisely, at the risk level $\alpha \in (0,1)$, the (conditional) VaR of a sequence of returns (ϵ_t) is the opposite of the α-quantile of the conditional distribution :

$$\mathrm{VaR}_t(\alpha) = -\inf\left\{x : P(\epsilon_{t+1} \leq x \mid \epsilon_u, u \leq t) \geq \alpha\right\}. \tag{1.1}$$

Assume that the returns follow the general conditionally heteroscedastic model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \end{cases} \tag{1.2}$$

where (η_t) is a sequence of independent and identically distributed (iid) random variables, η_t is independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a compact parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0,\infty)$. The variable σ_t^2 is generally referred to as the volatility of ϵ_t. For this GARCH-type volatility model, we have

$$\mathrm{VaR}_t(\alpha) = -\sigma_t(\theta_0)\xi_\alpha, \tag{1.3}$$

where ξ_α is the α-quantile of the distribution P_η of the innovations. Note that the model (1.2) is not identifiable without a scaling assumption on P_η. The standard identifiability assumption is $E\eta_t^2 = 1$, but we do not need to make this assumption in the present paper.

A simple and widely used example of the form (1.2) is the GARCH(p,q) model

1. Sometimes the conditional VaR refers to another risk measure called the expected shortfall.

of Engle (1982) and Bollerslev (1986), defined by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j} \sigma_{t-j}^2 \end{cases} \tag{1.4}$$

where $\omega_0 > 0, \alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. For the GARCH(1,1) model, we have $\sigma_t^2 = \sum_{i=1}^{\infty} \beta_{01}^{i-1}(\omega_0 + \alpha_{01}\epsilon_{t-i}^2)$, provided $\beta_{01} < 1$.

The most widely used estimator of ARCH-type models is arguably the Gaussian QMLE. The consistency and asymptotic normality (CAN) of this estimator requires only few regularity assumptions, and the standard identifiability condition $E\eta_t^2 = 1$ (see Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004) for the case of standard GARCH and ARMA-GARCH models, Mikosch and Straumann (2006), Straumann and Mikosch (2006), Bardet and Wintenberger (2009) for more general models). In the framework of standard GARCH models, Berkes and Horváth (2004) introduced generalized non-Gaussian QMLE (gQMLE) and established their CAN under alternative identifiability conditions. For the general model (1.2), Francq and Zakoian (2013) (hereafter FZ) showed that particular gQMLE lead to convenient one-step predictions of the powers $|\epsilon_t|^r$, $r \in \mathbb{R}$. Francq, Lepage and Zakoian (2011) constructed a two-step procedure based on a particular class of gQMLE for estimating standard GARCH(p, q) models. Independently, Fan *et al.* (2013) proposed, for the same problem, a three-step quasi maximum likelihood procedure, allowing for the use of a vast class of non-Gaussian likelihood functions. Francq and Zakoïan (2012) propose a gQMLE which allows for estimating a conditional VaR in one step, and compare this method with the more standard two-step method which consists in estimating the volatility parameter by Gaussian QMLE and the quantile of the innovations by the empirical quantile of the residuals.

In the present paper, we extend the above-mentioned conditional VaR two-step evaluation method by investigating the use of gQMLE's based on a generic instrumental density h. It is well known that the standard Gaussian QMLE, which is based on the instrumental density $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, converges to the volatility parameter θ_0, under mild regularity conditions. Moreover the empirical α-quantile of the Gaussian QMLE residuals converges to ξ_α. Section 2.1 shows that, in a very general framework, the gQMLE converges to some parameter θ_0^*, which depends on h, P_η and θ_0. When $h \neq \phi$ or $h \neq P_\eta$, we have $\theta_0^* \neq \theta_0$, and the empirical α-quantile of the gQMLE residuals converges to $\xi_\alpha^* \neq \xi_\alpha$. The conditional VaR two-step estimator is however consistent because $\sigma_t(\theta_0)\xi_\alpha = \sigma_t(\theta_0^*)\xi_\alpha^*$. Section 2.2 studies the asymptotic distribution of this estimator, for the general model (1.2). Section 3 makes explicit the asymptotic distributions for an extension of the GARCH model (1.4). It is shown that the optimal instrumental density, *i.e.* the function h which minimizes the asymptotic variance of the VaR estimator, depends neither on the GARCH parameter θ_0 nor on the risk level α, but only on simple characteristics of P_η. It follows that a simple adaptive method based on empirical moments of the residuals makes it possible to infer which h is optimal. Section 4 extends some of the results to conditional Distortion Risk Measures (DRM). The numerical illustrations are displayed in Section 5. Section 6 concludes. The proofs are collected in the Appendix.

2 Estimating the conditional VaR by gQMLE

For the standard volatility models, the following assumption is satisfied.

A1 : There exists a continuous function H such that for any $\theta \in \Theta$, for any $K > 0$, and any sequence $(x_i)_i$

$$K\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; H(\theta, K)).$$

In the case of the GARCH(1,1) model, we have

$$K\sigma_t(\theta) = \sqrt{K^2\omega + K^2\alpha + \beta\sigma_{t-1}^2 \{H(\theta_0, K)\}} = \sigma_t \{H(\theta_0, K)\}$$

where $H(\theta_0, K) = (K^2\omega_0, K^2\alpha_{01}, \beta_{01})'$. Assumption **A1** means that the parametric form of the volatility is stable by scaling, which is a highly desirable property for an ARCH model.

In view of (1.3) and **A1**, when $\xi_\alpha < 0$ we have

$$\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0)\xi_\alpha = \sigma_{t+1}(\theta_{0,\alpha})$$

where $\theta_{0,\alpha} = H(\theta_0, -\xi_\alpha)$. The parameter $\theta_{0,\alpha}$ is called the VaR parameter in Francq and Zakoian (2012).

In the next section, we show that the gQMLE generally converges to a parameter θ_0^* such that $\sigma_t(\theta_0^*) = \sigma^*\sigma_t(\theta_0)$, where $\sigma^* > 0$ depends on h and P_η. The residuals of the gQMLE are thus approximations of η_t/σ^*. Consequently, the gQMLE of the volatility converges to $\sigma^*\sigma_t(\theta_0)$ and the empirical quantile of the gQMLE residuals converges to $\xi_\alpha^* = \xi_\alpha/\sigma^*$. The gQMLE of the VaR thus gives a consistent estimator of $\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0^*)\xi_\alpha^*$.

2.1 Estimating the volatility parameter

Given observations $\epsilon_1, \dots, \epsilon_n$, and arbitrary initial values $\widetilde{\epsilon}_i$ for $i \leq 0$, let

$$\widetilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \widetilde{\epsilon}_0, \widetilde{\epsilon}_{-1}, \dots; \theta).$$

This random variable can be seen as a proxy of

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta).$$

Given an *instrumental* density $h > 0$, consider the QML criterion

$$\widetilde{Q}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} g(\epsilon_t, \widetilde{\sigma}_t(\theta)), \qquad g(x, \sigma) = \log \frac{1}{\sigma} h\left(\frac{x}{\sigma}\right), \tag{2.1}$$

and the (generalized) QMLE

$$\hat{\theta}_n^* = \arg\max_{\theta \in \Theta} \widetilde{Q}_n(\theta).$$

Throughout the text, starred symbols are used to designate quantities which depend on the instrumental density h. This estimator is the standard Gaussian QMLE if h is the standard Gaussian density ϕ. To establish the CAN of $\hat{\theta}_n^*$, we make the following assumptions.

A2 : (ϵ_t) is a strictly stationary and ergodic solution of (1.2), and there exists $s > 0$ such that $E|\epsilon_1|^s < \infty$.

A3 : For some $\underline{\omega} > 0$, almost surely, $\sigma_t(\theta) \in (\underline{\omega}, \infty]$ for any $\theta \in \Theta$. Moreover, for $\theta_1, \theta_2 \in \Theta$, we have $\sigma_t(\theta_1) = \sigma_t(\theta_2)$ *a.s.* if and only if $\theta_1 = \theta_2$.

A4 : The function $\sigma \to Eg(\eta_0, \sigma)$ takes its values in $[-\infty, +\infty)$ and has a unique maximum at some point $\sigma_* \in (0, \infty)$.

A5 : The instrumental density h is continuous on $\mathbb{R}$, it is also differentiable, except possibly in 0, and there exist constants $\delta \geq 0$ and $C_0 > 0$ such that, for all $u \in \mathbb{R} \setminus \{0\}$, $|uh'(u)/h(u)| \leq C_0(1 + |u|^\delta)$ and $E|\eta_0|^{2\delta} < \infty$.

A6 : There exist a random variable C_1 measurable with respect to $\{\epsilon_u, u < 0\}$ and a constant $\rho \in (0, 1)$ such that $\sup_{\theta \in \Theta} |\sigma_t(\theta) - \widetilde{\sigma}_t(\theta)| \leq C_1 \rho^t$.

Under **A1** and **A4**, define the parameter

$$\theta_0^* = H(\theta_0, \sigma_*). \tag{2.2}$$

A7 : The parameter θ_0^* belongs to the compact parameter space Θ.

A8 : The parameter θ_0^* belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

A9 : There exists no non-zero $x \in \mathbb{R}^m$ such that $x' \frac{\partial \sigma_t(\theta_0^*)}{\partial \theta} = 0, \quad a.s.$

A10 : The function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives, and

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial \theta} \right\| + \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \widetilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho^t,$$

where C_1 and ρ are as in **A6**.

A11 : h is twice continuously differentiable, except possibly at 0, with $|u^2\,(h'(u)/h(u))'\,| \leq C_0(1+|u|^\delta)$ for all $u \in \mathbb{R}\setminus\{0\}$ and $E|\eta_0|^\delta < \infty$, where C_0 and δ are as in **A5**.

A12 : There exists a neighborhood $V(\theta_0^*)$ of θ_0^* such that

$$\sup_{\theta\in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\|^4, \quad \sup_{\theta\in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial\theta\partial\theta'} \right\|^2, \quad \sup_{\theta\in V(\theta_0^*)} \left| \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)} \right|^{2\delta}$$

have finite expectations.

Most of these assumptions are similar to those of Berkes and Horváth (2004) and FZ.

Remark 2.1 *Note that* **A4** *is much less restrictive than the analog assumption in FZ, which requires a maximum at* $\sigma_* = 1$ *(see* **A3** *in FZ). Note also that we do not need any identifiability condition on* η_t *(such that* $E\eta_t^2 = 1$*). We need weaker assumptions because, in our framework, it will only be necessary to define the volatility up to an unknown multiplicative constant. Actually,* **A4** *is the same as Assumption 2 made by Fan* et al. *(2013) for their three-step estimation procedure.*

Remark 2.2 *In view of (A.7) below, under* **A5** *the parameter* σ_* *defined in* **A4** *is such that*

$$E\left\{\frac{\eta_0}{\sigma_*}\frac{h'}{h}\left(\frac{\eta_0}{\sigma_*}\right)\right\} = -1. \tag{2.3}$$

For the standard GARCH case, several assumptions can be made more explicit. The true value of the parameter is $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \beta_{0p})'$ and the generic element of Θ is denoted by $\theta = (\omega, \alpha_1, \ldots, \beta_p)'$. It is well-known that a necessary and sufficient condition for the existence of a strictly stationary solution to (1.4) is $\gamma < 0$, where γ denotes the top-Lyapunov exponent of the model (see e.g. Francq and Zakoïan (2004)). Write $\gamma = \gamma(\theta_0)$ to emphasize that γ depends on θ_0 (and also on the law of η_1). Let $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^j$. In that framework the assumptions **A2**, **A3**, **A6**, **A9**, **A10** and **A12** reduce to :

C : $\gamma(\theta_0) < 0$; $\forall\theta \in \Theta$, $\sum_{j=1}^p \beta_j < 1$ and $\omega > \underline{\omega}$ for some $\underline{\omega} > 0$; $|\eta_0|$ has a non degenerate distribution ; if $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common root, $\mathcal{A}_{\theta_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

The following lemma is similar to results given by Berkes and Horváth (2004) and FZ.

Lemma 2.1 (Asymptotic behavior of generalized QMLE) *If* **A1-A7** *are satisfied, then*

$$\hat\theta_n^* \to \theta_0^*, \quad a.s.$$

where θ_0^* is defined by (2.2). If, in addition, **A8-A12** *are satisfied and* $Eg_2(\eta_0, 1) \neq 0$ *then*

$$\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau_h J_*^{-1})$$

where

$$J_* = 4ED_t(\theta_0^*)D_t'(\theta_0^*) \quad \textit{and} \quad \tau_h = \frac{4Eg_1^2(\sigma_*^{-1}\eta_0, 1)}{\left\{Eg_2(\sigma_*^{-1}\eta_0, 1)\right\}^2}, \tag{2.4}$$

in which

$$D_t(\theta) = \frac{1}{\sigma_t(\theta)}\frac{\partial \sigma_t(\theta)}{\partial \theta}, \quad g_1(x, \sigma) = \frac{\partial g(x, \sigma)}{\partial \sigma} \quad \textit{and} \quad g_2(x, \sigma) = \frac{\partial g_1(x, \sigma)}{\partial \sigma}.$$

Example 2.1 (GED instrumental density) Consider the case in which h belongs to the class of the Generalized Error Distributions of shape parameter $\kappa > 0$, defined by

$$h_\kappa(x) = \frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}} e^{-\frac{|x|^\kappa}{2}},$$

which will be denoted by GED(κ). We then have, for $x \neq 0$,

$$\frac{h'}{h}(x) = -\frac{\kappa |x|^\kappa}{2x}.$$

In view of (2.3), we obtain

$$\sigma_* = \left(\frac{\kappa}{2}E|\eta_1|^\kappa\right)^{1/\kappa}.$$

By (A.1) and (A.5) given in the proof of Lemma 2.1,

$$g_1\left(\frac{\eta_1}{\sigma_*}, 1\right) = -1 + \frac{|\eta_1|^\kappa}{E|\eta_1|^\kappa}, \quad g_2\left(\frac{\eta_1}{\sigma_*}, 1\right) = 1 - (1+\kappa)\frac{|\eta_1|^\kappa}{E|\eta_1|^\kappa}$$

and

$$\tau_h := \tau_{GED} = \frac{4}{\kappa^2}\left(\frac{E|\eta_1|^{2\kappa}}{(E|\eta_1|^\kappa)^2} - 1\right). \tag{2.5}$$

To give a more explicit example, assume that we have a standard GARCH(1,1) with parameter $\theta_0 = (\omega_0, \alpha_0, \beta_0)$ and $\eta_t \sim \mathcal{N}(0, 1)$. For this distribution we have $E|\eta_1| = \sqrt{2/\pi}$. If we take the double exponential distribution $(1/4)e^{-|x|/2}$ as instrumental density h, which corresponds to the GED(1) , then $\hat{\theta}_n^*$ thus converges to $\theta_0^* = (2\omega_0/\pi, 2\alpha_0/\pi, \beta_0)$. Moreover the asymptotic variance is obtained with $\tau_h = 2\pi - 4$.

Example 2.2 (Double Generalized Gamma instrumental density) Now consider a larger class of densities, which contains, in particular, the GED, the Laplace, the double Weibull, Rayleigh and Maxwell, and the Gaussian distributions. Assume that h follows a double generalized Gamma (dgG) distribution $\Gamma(b, p, d)$ with parameters $b > 0$, $p > 0$ and $d > 0$, defined by the density

$$h(x) = h_{dgG}(x) = \frac{db^p}{2\Gamma(\frac{p}{d})}|x|^{p-1}e^{-|bx|^d}.$$

For $x \neq 0$, we have

$$x\frac{h'}{h}(x) = p - 1 - d|bx|^d.$$

In view of (2.3), we have $\sigma_* = \left(\frac{db^d}{p}E|\eta_1|^d\right)^{1/d}$. Thus,

$$g_1\left(\frac{\eta_1}{\sigma_*}, 1\right) = p\left(\frac{|\eta_1|^d}{E\,|\eta_1|^d} - 1\right), \quad g_2\left(\frac{\eta_1}{\sigma_*}, 1\right) = p\left(1 - (d+1)\frac{|\eta_1|^d}{E\,|\eta_1|^d}\right).$$

We then have

$$\tau_h = \tau_{dgG} = \frac{4}{d^2}\left(\frac{E\,|\eta_1|^{2d}}{\left(E\,|\eta_1|^d\right)^2} - 1\right).$$

Note that τ_{dgG} is equal to τ_{GED} when $\kappa = d$.

Therefore, compared to the GED, the introduction of the more complicated class of the dgG distributions is useless, because it does not lead to any efficiency gain.

Example 2.3 (Student instrumental density) Now consider the case where the instrumental density h is the Student distribution with ν degrees of freedom

$$h(x) = h_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

By (A.1) and (A.5), we have

$$g_1(x,\sigma) = \frac{\nu(x-\sigma)(x+\sigma)}{\sigma(x^2+\nu\sigma^2)}, \quad g_2(x,\sigma) = -\frac{\nu\left\{x^4 + x^2(1+3\nu)\sigma^2 - \nu\sigma^4\right\}}{\sigma^2(x^2+\nu\sigma^2)^2}.$$

In view of (2.3), the parameter σ_* satisfies

$$E\frac{\eta_1^2}{\nu\sigma_*^2 + \eta_1^2} = \frac{1}{\nu+1}.$$

Contrary to what happens in Example 2.1, the parameters σ_* and τ_h do not have simple expressions as a function of ν and of the distribution of η_1, but can be obtained by numerical algorithms.

2.2 Estimating the VaR parameter

For the general volatility model (1.4), we have

$$\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0^*)\xi_\alpha^*,$$

where ξ_α^* denotes the α-quantile of $\eta_t^* := \eta_t/\sigma_*$. Note that, when $\xi_\alpha^* < 0$, **A1** entails

$$\mathrm{VaR}_t(\alpha) = \sigma_{t+1}(\theta_{0,\alpha}) \quad \text{where} \quad \theta_{0,\alpha} = H(\theta_0^*, -\xi_\alpha^*).$$

The parameter $\theta_{0,\alpha}$ is called the VaR parameter in Francq and Zakoian (2012). Note that $\xi_\alpha := \sigma_* \xi_\alpha^*$ is the α-quantile of η_t. Thus we have $\theta_{0,\alpha} = H(\theta_0^*, -\xi_\alpha^*) = H(\theta_0, -\xi_\alpha)$.

Let $\hat{\xi}_{\alpha,n}^*$ be the empirical quantile of the residuals $\hat{\eta}_t^* := \epsilon_t / \widetilde{\sigma}_t(\hat{\theta}_n^*)$ for $t = 1, \dots, n$. We now give an intermediate result that will be used to obtain the asymptotic distribution of two-step estimators of the VaR parameter.

Theorem 2.1 *Assume η_1 has a density f, continuous at ξ_α, such as $f(\xi_\alpha) > 0$. Under the assumptions of Lemma 2.1, we have*

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n^* - \theta_0^* \\ \hat{\xi}_{\alpha,n}^* - \xi_\alpha^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left\{ 0, \Sigma^* := \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\prime} & \Sigma_{22}^* \end{pmatrix} \right\},$$

where

$$\begin{aligned} \Sigma_{11}^* &= \tau_h J_*^{-1}, \\ \Sigma_{12}^* &= -\left\{ \xi_\alpha^* \tau_h - \frac{4c_\alpha}{\sigma_* f(\xi_\alpha) E g_2(\eta_0^*, 1)} \right\} J_*^{-1} \Omega_*, \\ \Sigma_{22}^* &= \frac{\tau_h (\xi_\alpha^*)^2}{4} - \frac{2c_\alpha \xi_\alpha^*}{\sigma_* f(\xi_\alpha) E g_2(\eta_0^*, 1)} + \frac{\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)}, \end{aligned}$$

with $\Omega_* = E D_t(\theta_0^*)$, $c_\alpha = Cov(\mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}}, g_1(\eta_t^*, 1))$.

In the case $h = \phi$ we retrieve Theorem 4.2 in Francq and Zakoian (2012).

Note that $\hat{\theta}_{n,\alpha}^*$ converges to the VaR parameter $\theta_{0,\alpha}$. The star symbol is used to emphasize that, contrary to the parameter, the estimator depends on h.

The delta method immediately gives the following result.

Corollary 2.1 *Under the assumptions of Theorem 2.1 and if H is differentiable at $(\theta_0^*, -\xi_\alpha^*)$, with $\xi_\alpha^* < 0$, we have*

$$\sqrt{n} \left(\hat{\theta}_{n,\alpha}^* - \theta_{0,\alpha} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, G_* \Sigma^* G_*' \right),$$

where

$$G_* = \left[\frac{\partial H(\theta, K)}{\partial(\theta', K)} \right]_{(\theta_0^*, -\xi_\alpha^*)}.$$

By empirically estimating the asymptotic variance, this corollary makes it possible to obtain a confidence interval at an asymptotic statistical estimation-risk level α_1 for the risk parameter at the market-risk level α. Using again the delta method, confidence intervals for $\mathrm{VaR}_t(\alpha) = \sigma_{t+1}(\theta_{0,\alpha})$ at a given estimation-risk level can be deduced, exactly as Francq and Zakoian (2012) did for the VaR estimation method based on the Gaussian QMLE.

The following result shows that the estimator of the VaR parameter is not sensitive to a scaling of the instrumental density.

Corollary 2.2 *Under the assumptions of Corollary 2.1, and if* **A1** *holds true when* σ_t *is replaced by* $\tilde{\sigma}_t$, i.e. *if*

$$K\tilde{\sigma}_t(\theta) = \tilde{\sigma}_t(\theta)\left\{H(\theta, K)\right\}, \tag{2.6}$$

then the estimator $\hat{\theta}^*_{n,\alpha}$ *is not changed if* $h(x)$ *is replaced by* $h_s(x) = s^{-1}h(s^{-1}x)$, *for any* $s > 0$.

In the standard GARCH(1,1) case, it is easy to see that (2.6) is satisfied when the initial value $\tilde{\sigma}_0(\theta)$ is chosen equal to zero.

3 Application to GARCH models

For particular GARCH models, we now verify the regularity conditions of Lemma 2.1, and we give a more explicit expression for the asymptotic variance of Corollary 2.1. We begin with the GARCH(1,1) model, and extend the result for a much wider class.

3.1 The first-order GARCH model

First begin with the GARCH(1,1) case, under Assumption **C**. In that case, the matrix G_* of Corollary 2.1 is given by

$$G_* = \begin{pmatrix} (\xi^*_\alpha)^2 & 0 & 0 & -2\xi^*_\alpha\omega^*_0 \\ 0 & (\xi^*_\alpha)^2 & 0 & -2\xi^*_\alpha\alpha^*_0 \\ 0 & 0 & 1 & 0 \end{pmatrix} := \begin{pmatrix} A_* & -2\xi^*_\alpha \begin{pmatrix} \omega^*_0 \\ \alpha^*_0 \\ 0 \end{pmatrix} \end{pmatrix}.$$

Note also that, for any $\theta^*_0 = (\omega^*_0, \alpha^*_0, \beta^*_0) \in \Theta$, we have

$$\begin{aligned}(\omega^*_0, \alpha^*_0, 0)\frac{\partial \sigma^2_t(\theta^*_0)}{\partial \theta} &= \omega^*_0 + \alpha^*_0\epsilon^2_{t-1} + \beta^*_0\left\{(\omega^*_0, \alpha^*_0, 0)\frac{\partial \sigma^2_{t-1}(\theta^*_0)}{\partial \theta}\right\} \\ &= \sum_{i=0}^{\infty} \beta^{*i}_0\left\{\omega^*_0 + \alpha^*_0\epsilon^2_{t-i}\right\} = \sigma^2_t(\theta^*_0).\end{aligned}$$

It follows that

$$\frac{1}{\sigma_t(\theta^*_0)}\frac{\partial \sigma_t(\theta^*_0)}{\partial \theta'}\begin{pmatrix} \omega^*_0 \\ \alpha^*_0 \\ 0 \end{pmatrix} = \frac{1}{2} \quad a.s.,$$

and thus

$$\Omega'_*\begin{pmatrix} \omega^*_0 \\ \alpha^*_0 \\ 0 \end{pmatrix} = \frac{1}{2}, \quad J_*\begin{pmatrix} \omega^*_0 \\ \alpha^*_0 \\ 0 \end{pmatrix} = 2\Omega_*, \quad J^{-1}_*\Omega_* = \frac{1}{2}\begin{pmatrix} \omega^*_0 \\ \alpha^*_0 \\ 0 \end{pmatrix}, \quad \Omega'_*J^{-1}_*\Omega_* = \frac{1}{4}.$$

The second equality of the previous line shows that

$$\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right) = J_* - 4\Omega_*\Omega_*' = J_*(J_*^{-1} - \Psi_*)J_*,$$

where

$$\Psi_* = \begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} \begin{pmatrix} \omega_0^* & \alpha_0^* & 0 \end{pmatrix} = \begin{pmatrix} \omega_0^{*2} & \omega_0^*\alpha_0^* & 0 \\ \omega_0^*\alpha_0^* & \alpha_0^{*2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Under **A9**, which is implied by the identifiability condition in Assumption **C**, the matrix $\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right)$ is positive definite. It follows that

$$J_*^{-1} - \Psi_* \quad \text{is positive definite.} \tag{3.1}$$

Moreover we have

$$\begin{aligned} G_*\Sigma^*G_*{}' &= \tau_h A_* J_*^{-1} A_* + \left(\frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)} - \tau_h(\xi_\alpha^*)^4\right)\Psi_* \\ &= \tau_h A_*(J_*^{-1} - \Psi_*)A_* + \frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)}\Psi_*. \end{aligned}$$

For the last equality we used that $A_*\Psi_*A_* = (\xi_\alpha^*)^4\Psi_*$.

Now we introduce analogs of the starred symbols, which are independent of the instrumental density h, using the matrix transformation

$$M_* = \begin{pmatrix} \frac{1}{\sigma_*^2}I_2 & 0_2 \\ 0_2' & 1 \end{pmatrix}.$$

We thus define $A = M_*^{-1}A_*$ and $\Psi = M_*\Psi_*M_* = \sigma_*{}^{-4}\Psi_*$. Note also that

$$\theta_0 \;=\; M_*\theta_0^*, \quad D_t(\theta_0^*) = M_*D_t(\theta_0) \quad \text{and} \quad J_* = M_*JM_*.$$

With this notation, we have

$$G_*\Sigma^*G_*{}' = \tau_h A(J^{-1} - \Psi)A + \frac{4\xi_\alpha^2\alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi. \tag{3.2}$$

The instrumental density h_1 is said to be more efficient than h_2, which is denoted by $h_1 \succ h_2$, if the difference of the asymptotic variances given by (3.2) is positive definite. In the asymptotic variance, only τ_h depends on h. In view of (3.1), this shows that $h_1 \succ h_2$ if and only if $\tau_{h_1} < \tau_{h_2}$.

3.2 The Asymmetric Power GARCH model

Ding, Granger and Engle (1993) introduced the so-called Asymmetric Power GARCH (APARCH) models, which include the standard GARCH of Bollerslev (1991), the TARCH of Zakoian (1994), the GJR of Glosten, Jagannathan and

Runkle (1993) and many other popular specifications of the volatility. Letting $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$, the model is defined by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i+}(\epsilon_{t-i}^+)^\delta + \alpha_{0i-}(-\epsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^\delta \end{cases} \tag{3.3}$$

where the coefficients satisfy $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$ and $\delta > 0$. The standard GARCH is obtained with $\delta = 2$ and $\alpha_{0i-} = \alpha_{0i+}$. When $\alpha_{0i-} > \alpha_{0i+}$, a negative return has a higher impact on the future volatility than a positive return of the same magnitude, which is a well-documented stylized fact that is called "leverage effect".

Hamadeh and Zakoïan (2011) showed that the power parameter δ is not easily estimated. We therefore consider that δ is fixed. In many applications, $\delta = 1$ (as in the TARCH) or $\delta = 2$ (as in the GJR model). As in Assumption **C**, let $\gamma(\theta_0)$ be the top-Lyapunov exponent associated with (3.3). Hamadeh and Zakoïan (2011) showed the CAN of the Gaussian QMLE of $\theta_0 = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-}, \beta_{01}, \ldots, \beta_{0p})'$ under the assumption :

D : $\gamma(\theta_0) < 0$; θ_0 belongs to the interior of Θ ; there exists $\underline{\omega} > 0$ such that, $\forall \theta \in \Theta$, $\omega > \underline{\omega}$ and $\sum_{j=1}^p \beta_j < 1$; the support of the distribution of η_1 contains at least 3 points ; $P[\eta_t > 0] \in (0, 1)$; if $p > 0$, $\mathcal{B}_{\theta_0}(z)$ has no common root with $\mathcal{A}_{\theta_0+}(z) = 1 - \sum_{i=1}^q \alpha_{0i+} z^i$ and $\mathcal{A}_{\theta_0-}(z) = 1 - \sum_{i=1}^q \alpha_{0i-} z^i$; $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ and $\alpha_{0q,+} + \alpha_{0q,-} + \beta_{0p} \neq 0$ (with the notation $\alpha_{00,+} = \alpha_{00,-} = \beta_{00} = 1$)

and under the identifiability condition $E\eta_1^2 = 1$ (that we do not assume in our framework).

The following theorem extends the results obtained in the previous section.

Theorem 3.1 *Consider the APARCH(p, q) model (3.3) under Assumption* **D**. *Assume η_1 has a density f, continuous at $\xi_\alpha < 0$, such as $f(\xi_\alpha) > 0$. If the instrumental density h satisfies* **A4**, **A5**, **A7**, **A8** *and* **A11**, *then the two-step estimator of the VaR parameter at the confidence level $\alpha \in (0, 1)$ satisfies*

$$\sqrt{n}\left\{\hat{\theta}_{n,\alpha}^* - H(\theta_0^*, -\xi_\alpha^*)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_*\Sigma^* G_*{}'\right),$$

where, for $\xi > 0$,

$$H(\omega, \alpha_{1+}, \ldots, \alpha_{q-}, \beta_1, \ldots, \beta_p, \xi) = \left(\xi^\delta\omega, \xi^\delta\alpha_{1+}, \ldots, \xi^\delta\alpha_{q-}, \beta_1, \ldots, \beta_p\right)$$

and

$$G_*\Sigma^* G_*{}' = \tau_h A(J^{-1} - \Psi)A + \frac{4\xi_\alpha^2 \alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi,$$

where $\overline{\theta}_0' = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-}, 0, \ldots, 0)$,

$$A = \text{diag}\left\{(-\xi_\alpha)^\delta I_{2q+1}, I_p\right\}, \quad \Psi = \overline{\theta}_0\overline{\theta}_0', \quad J = 4ED_1(\theta_0)D_1'(\theta_0).$$

For the instrumental densities h_1 *and* h_2 *, we have* $h_1 \succ h_2$ *if and only if* $\tau_{h_1} < \tau_{h_2}$*.*

Remark 3.1 (On the optimal instrumental density) This theorem shows that an instrumental density h with the smallest value of τ_h is optimal. It is worth noting that the knowledge of the distribution of η_1, up to some (unknown) scaling constant, is sufficient to determine if h is optimal within the class of the two-step estimators introduced in this paper. In particular the optimality of h : 1) does not depend on θ_0^*, or even on the volatility model ; 2) does not depend on α.

Francq and Zakoïan (2013) compared the two-step estimator based on ϕ with a one step estimator. As in 1), the ranking of the two estimators is the same regardless of the model. However the relative efficiency of their two methods varies with α.

Note also that the optimal instrumental density for estimating the VaR parameter is the same as that obtained by Fan et al. (2013) for their three-step estimator of the volatility parameter.

3.3 Optimal choice of the instrumental density

In view of Theorem 3.1, the optimal h (within a given class of instrumental densities satisfying the assumptions of the theorem) has the smallest τ_h. We first give an example of density h for which τ_h is a function of moments of η_1 that can be empirically estimated. We then give an example in which τ_h is not explicit, but can however be easily estimated.

3.3.1 GED instrumental distribution

Consider the case in which h is the GED(κ) distribution of Example 2.1. The value κ_0 of κ which minimizes (2.5) is considered as optimal. An empirical estimator of κ_0 can then be obtained as follows. Let $\hat{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1, \ldots, n$, be the residuals obtained from a first-step estimation procedure, which is consistent but not necessarily optimal, for example the Gaussian QMLE. An estimator of the parameter κ_0 for the optimal instrumental density is defined by

$$\hat{\kappa} = \arg\min_{\kappa \in \mathcal{K}} \frac{1}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2} - 1\right), \quad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t|^r,$$

where $\mathcal{K}$ is a bounded interval containing κ_0. Note that it is important to minimize over a bounded interval because, by Lemma 3.1 in Francq et al. (2011), for any fixed n, we have

$$\frac{1}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2}-1\right)\to 0, \qquad \text{as } \kappa\to\infty.$$

3.3.2 Student instrumental distribution

As in Example 2.3, let us take the Student distribution with ν degrees of freedom as instrumental density h. The parameters σ_* and τ_h can be estimated as follows. Let $\hat{\eta}_1,\dots,\hat{\eta}_n$ be the residuals of a first-step estimation procedure. Let C and S be compact subsets of $]0,\infty[$. For any value of $\nu\in C$, σ_* can be estimated by

$$\hat{\sigma}_* = \arg\max_{\sigma\in S}\sum_{t=1}^n g(\hat{\eta}_t,\sigma).$$

An estimator of the optimal value of ν is then obtained as

$$\hat{\nu} = \arg\min_{\nu\in C}\frac{n^{-1}\sum_{t=1}^n g_1^2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t,1\right)}{\left\{n^{-1}\sum_{t=1}^n g_2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t,1\right)\right\}^2}. \tag{3.4}$$

3.4 Suboptimality of the naive adaptive approach

Assume a parametric form $h_\kappa(x)$, $\kappa\in\mathcal{K}$ for the instrumental density. We know that the optimal instrumental density is the (unknown) distribution f of η_1, or equivalently any scaled version $\sigma^{-1}f(x/\sigma)$, $\sigma>0$, of this density (see Corollary 2.2). If some scaled version of f belongs to the chosen class of parametric instrumental densities, *i.e.* if $f(x)=\sigma_0^{-1}h_{\kappa_0}(x/\sigma_0)$ for some $\kappa_0\in\mathcal{K}$ and some $\sigma_0>0$, then the optimal instrumental density can be found by the (quasi-)maximum likelihood procedure

$$(\hat{\kappa},\hat{\sigma}) = \arg\max_{(\kappa,\sigma)\in\mathcal{K}\times(0,\infty)}\sum_{t=1}^n \log\sigma^{-1}h_\kappa(\hat{\eta}_t/\sigma),$$

where $\hat{\eta}_t=\epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t=1,\dots,n$, are the residuals obtained from a Gaussian QMLE, or any other consistent first-step estimation procedure. Even if f does not belong to the class of densities, the procedure makes sense and converges, under general regularity conditions (see White 1982), to a minimizer of a Kullback-Leibler divergence, solution to

$$(\kappa^*,\sigma^*) = \arg\max_{(\kappa,\sigma)\in\mathcal{K}\times(0,\infty)} E\log\sigma^{-1}h_\kappa(\eta_1/\sigma).$$

For example, consider the class of the Generalized Error Distributions of shape parameter $\kappa>0$, defined by

$$h_\kappa(x)=\frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}}e^{-\frac{|x|^\kappa}{2}},$$

which will be denoted by GED(κ). We then have,

$$\sigma^*=\left(\frac{\kappa^* E|\eta_1|^{\kappa^*}}{2}\right)^{1/\kappa^*},$$

where

$$\kappa^* = \arg\max_{\kappa \in \mathcal{K}} \log\left(\frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}}\right) - \frac{1}{\kappa}\left\{\log\left(\frac{\kappa E|\eta_1|^\kappa}{2}\right) + 1\right\}.$$

Let τ_0 be the optimal value of τ_h when h belongs to the class of the GED(κ) instrumental densities. In view of (2.5), we have

$$\tau_0 = \frac{4}{\kappa_0^2}\left(\frac{E\left|\eta_1\right|^{2\kappa_0}}{\left(E\left|\eta_1\right|^{\kappa_0}\right)^2} - 1\right), \quad \kappa_0 = \arg\min_{\kappa} \frac{4}{\kappa^2}\left(\frac{E\left|\eta_1\right|^{2\kappa}}{\left(E\left|\eta_1\right|^{\kappa}\right)^2} - 1\right).$$

Let τ^* be the value of τ_h when h is the GED(κ^*). This τ^* is optimal (*i.e.* minimal) when the density f of η_1 is a rescaled GED, and in this case we have $\tau^* = \tau_0$. In general, there is no guarantee that τ^* be optimal in the class of the GED instrumental density, *i.e.* that $\tau^* = \tau_0$.

4 Extension to other conditional risk measures

VaR is used by academics to define more sophisticated risk measures and VaR constitutes a powerful tool for professional risk managers, but it has been criticized for giving a too limited view of the actual risk level. In particular, VaR says nothing on what happens when losses exceed VaR. The expected shortfall (ES) is a popular alternative risk measure which circumvents this problem by measuring the average loss in the case of losses exceeding VaR. Another argument often given against VaR is that it does not satisfy the subadditivity property (see *e.g.* Artzner, Delbaen, Eber and Heath (1999), Wirch and Hardy (1999)). That means that the VaR of an average of risky assets can be larger than the average of the VaR of the individual assets. [2]

ES satisfies the subadditivity property and constitues a leading example of the wide class of the Distortion Risk Measures (DRM) (see Wang (2000) and the references therein). Assuming that $E|\eta_1| < \infty$, the function $u \mapsto \mathrm{VaR}_t(u)$ is a.s. integrable, and a conditional DRM is defined by

$$\mathrm{DRM}_t = \int_0^1 \mathrm{VaR}_t(u) dG(u), \tag{4.1}$$

2. That the risk of an average must be less than the average of the risks is however questionable. The usual central limit theorem (CLT) leads us to think that the answer should be positive, but this is not the case when considering generalized CLT's for variables without second order moments. Indeed, the risk of an average of iid Cauchy variables is the risk of a single Cauchy variable. More generally, an average of iid alpha-stable random variables with tail index smaller than 1 remains alpha-stable, but its scale increases, and thus the average should have a larger risk.

where G is a cumulative distribution function (cdf) on $[0,1]$ that is called the distortion function. The DRM can be interpreted as a weighted sum of VaR's, where the weights are the increases of the distortion function. ES is obtained with $G(u) = (u/\alpha)1_{[0,\alpha[}(u) + 1_{[\alpha,\infty[}(u)$. Other examples of DRM are the proportional hazard DRM, obtained with $G(u) = u^r$, and the exponential DRM, obtained with $G(u) = (1-e^{ru})/(1-e^r)$, $r > 0$. Assuming $\int_0^1 \xi_u dG(u) < 0$, under **A1** we have

$$\mathrm{DRM}_t = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \int_0^1 \xi_u dG(u) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_{0,G}),$$

where

$$\theta_{0,G} = H\left(\theta_0, -\int_0^1 \xi_u dG(u)\right)$$

can be called the conditional DRM risk parameter. A natural estimator of that parameter is

$$\hat{\theta}^*_{n,G} = H\left(\hat{\theta}^*_n, -\int_0^1 \hat{\xi}^*_{n,u} dG(u)\right).$$

Theorem 4.1 (Consistency of the DRM conditional parameter estimator) *If* **A1-A7** *are satisfied,* $E|\eta_1| < \infty$, $\int_0^1 \xi_u dG(u) < 0$*, and the cdf* F_η *of* η_1 *is invertible on* $(0,1)$*, then, as* $n \to \infty$,

$$\hat{\theta}^*_{n,G} \to \theta_{0,G} \quad a.s.$$

For estimating the conditional VaR, $-\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0)\xi_u$, the optimal instrumental density h does not depend on u (see Remark 3.1). For estimating the weighted VaR, $\mathrm{DRM}_t = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \int_0^1 \xi_u dG(u)$, it is natural to chose the same optimal instrumental density h, which minimizes τ_h, at least in the APARCH case (see Theorem 3.1).

5 Numerical illustrations

We first consider a theoretical framework in which the distribution of η_t is assumed to be known. Considering two classes of instrumental densities, the GED(κ) and the Student St_ν distributions, we determined the best instrumental densities within each class, and we compared them with the standard Gaussian density in term of asymptotic relative efficiency. In the second subsection, Monte Carlo experiments are used to compare the finite sample performance of the different VaR estimation procedures. The last subsection proposes illustrations on financial series.

5.1 Theoretical comparison of the asymptotic efficiencies

Assume that η_1 follows the double generalized Gamma distribution $\Gamma(b,p,d)$ considered in Example 2.2. We then have $E|\eta_1|^r = b^{-r}\Gamma((p+r)/d)/\Gamma(p/d)$. In view of (2.5), the minimal value of τ_h, which is obtained when $h \sim \Gamma(b,p,d)$, is given by

$$\tau_{opt} = \frac{4}{pd}.$$

With the standard approach based on the Gaussian QMLE, we have

$$\tau_\phi = \left(\frac{E\left|\eta_1\right|^4}{\left(E\left|\eta_1\right|^2\right)^2} - 1 \right) = \left(\frac{\Gamma\left(\frac{p}{d}\right)\Gamma\left(\frac{p+4}{d}\right)}{\left\{\Gamma\left(\frac{p+2}{d}\right)\right\}^2} - 1 \right).$$

The asymptotic relative efficiency (ARE) of the generalized QMLE based on the instrumental density h with respect to the standard Gaussian QMLE can be measured by the ratio

$$\mathrm{ARE} = \frac{\tau_\phi}{\tau_h}.$$

In view of (2.5), the method based on the instrumental density GED(κ) is optimal (*i.e.* $\tau_{\mathrm{GED}(\kappa)} = \tau_{opt}$) when $\kappa = d$. Figure 1 shows that, even if the instrumental densities GED(d) and $\Gamma(b,p,d)$ are asymptotically equivalent, they can be surprisingly different.

Figure 2 shows that the GED instrumental density can be much more efficient than the Gaussian one (indeed its ARE is much greater than 1 when d is small). The ARE reaches 1 for $d = \kappa = 2$. This was expected because the GED(2) and $\Gamma(\sqrt{1/2},1,2)$ distributions both coincide with the standard Gaussian distribution. This figure also displays the ARE of the best Student instrumental density with respect to the Gaussian distribution. Even if the Student is generally not optimal when $\eta_t \sim \Gamma(b,p,d)$, it can also be much more efficient than the gaussian.

5.2 Simulation experiments

In the previous section, the selection of the optimal instrumental density, GED or Student, is accomplished by assuming that the distribution of η_t is known, which is obviously unrealistic in practice. In this section, we first study if the selection of the optimal procedures can be satisfactorily done by using the estimated residuals. We thus simulate $N = 100$ independent trajectories of size $n = 1,000$ of a

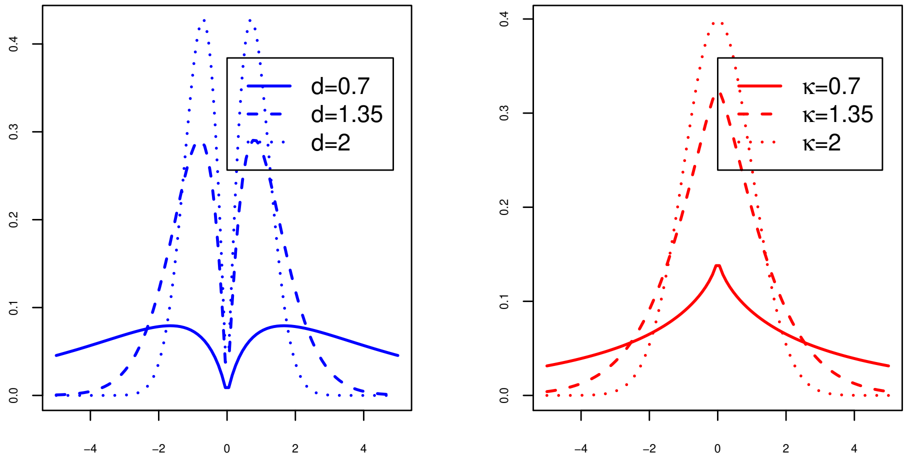

FIGURE 1 – Density $\Gamma(1, 2, d)$ for $d = 0.7$, $d = 1.35$ and $d = 2$ (left panel) and density GED(κ) for $\kappa = 0.7$, $\kappa = 1.35$ and $\kappa = 2$ (right panel). The asymptotic distribution of the generalized QMLE based on $\Gamma(b, p, d)$ is the same as that based on GED(κ) when $\kappa = d$.

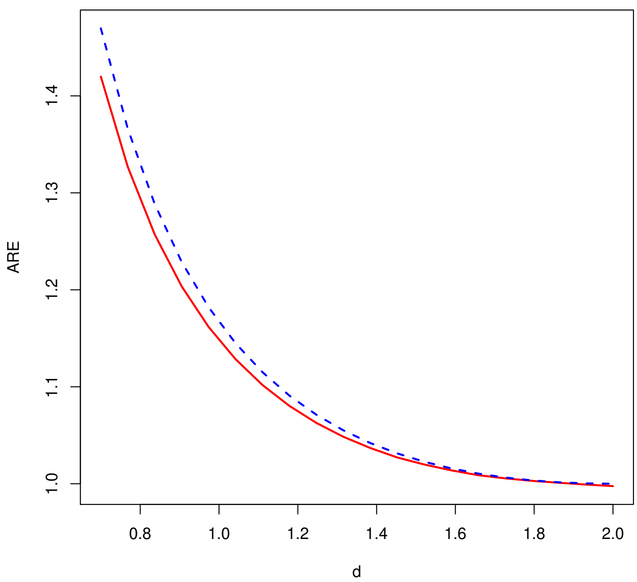

FIGURE 2 – ARE of the generalized QMLE based on the optimal GED (dotted line), or based on the optimal Student instrumental density (full line), with respect to the Gaussian QMLE, when $\eta_t \sim \Gamma(1, 2, d)$ and d varies from $d = 0.7$ to $d = 2$.

GARCH(1,1) model with $\theta_0 = (0.02, 0.002, 0.8)$ and $\eta_t \sim \Gamma(1, 2, d)$, where d takes 20 values between $d = 0.7$ and $d = 2$, as in Figure 2. For each simulation and each value of d, the parameter τ_ϕ is estimated by

$$\hat{\tau}_\phi = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 1, \qquad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t^r, \quad \hat{\eta}_t = \frac{\epsilon_t}{\widetilde{\sigma}_t(\widehat{\theta})},$$

where $\widehat{\theta}$ denotes the Gaussian QMLE. We then obtain an estimate of the optimal value of τ_{GED} by taking the minimum of

$$\frac{4}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2} - 1\right)$$

over $\kappa \in [0.1, 5]$. An estimate of the optimal value of τ_{St} is similarly obtained from (3.4). The curves of Figure 3 correspond to the average estimated ARE's over the N replications. The curves have very similar shapes to those of Figure 2, and lead to the same ranking of the estimation methods. This shows that one can actually select the asymptotically optimal method by choosing the method which minimizes the estimated value of τ computed from the residuals.

Table 1 compares the actual accuracies of the different methods for estimating the VaR parameter at the 5% risk level. For clarity reasons, the results are only given for the 4 values of $d \in \{0.7, 0.97, 1.66, 2\}$. The columns "mean" and "median" give the average and the median of the absolute value of the N estimation errors. The column RMSE gives the root mean square error of estimation. As expected from the asymptotic results (see Figure 2), the estimators based on the GED and Student instrumental densities are always very close, and they are much more efficient than the usual two-step estimator based on the Gaussian QMLE when the density of η_t is far from the Gaussian (*i.e.* when $d = 0.7$ or $d = 0.97$), whereas all the estimators are equivalent when d is close to 2 (which corresponds to the Gaussian case). Table 2 shows that, as expected from the theory, the ranking of the method is the same for the risk level of 1%.

5.3 Application to daily stock indices

We now consider the estimation of the VaR parameter for daily returns of 7 world stock market indices : CAC, DAX, FTSE, Nikkei, SMI (Swiss Market Index), SP500 and TSX (Toronto Stock Exchange). The data set comes from Yahoo Finance and covers the period from early January 1990 to the end of June 2013, when these

TABLE 1 – Distribution of the estimation errors for the 5%-VaR parameter of a GARCH(1,1) model with $\eta_t \sim \Gamma(1, 2, d)$, using the standard Gaussian QMLE, the generalized QMLE based on the optimal GED instrumental density, or that based on the Student density. The smallest errors are displayed in bold.

	Gaussian-QMLE			GED-QMLE			Student-QMLE		
VaR parameter ω									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	0.956	2.306	4.147	**0.814**	**1.349**	**2.673**	0.847	1.582	3.278
0.97	1.344	1.108	1.226	**0.234**	**0.440**	**0.625**	0.613	0.694	0.848
1.66	**0.041**	**0.085**	**0.121**	0.045	0.091	0.125	0.042	0.088	**0.122**
2	**0.025**	**0.053**	**0.076**	0.027	**0.053**	**0.075**	0.027	0.054	**0.077**
VaR parameter α									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	0.062	0.079	0.105	0.060	0.072	0.094	**0.053**	**0.068**	**0.092**
0.97	**0.034**	0.049	0.065	**0.034**	**0.037**	**0.053**	**0.034**	0.045	0.060
1.66	**0.006**	0.029	**0.048**	**0.006**	0.029	**0.049**	**0.006**	**0.028**	**0.048**
2	**0.005**	**0.026**	**0.044**	**0.005**	**0.026**	**0.045**	**0.005**	**0.026**	**0.045**
VaR parameter β									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	0.071	0.142	0.243	0.057	**0.087**	**0.154**	**0.054**	0.102	0.193
0.97	0.800	0.622	0.683	**0.134**	**0.251**	**0.350**	0.350	0.390	0.475
1.66	**0.126**	**0.268**	**0.381**	0.133	0.286	0.395	0.133	0.278	0.387
2	**0.112**	**0.235**	0.338	0.115	**0.235**	**0.334**	0.116	0.238	0.342

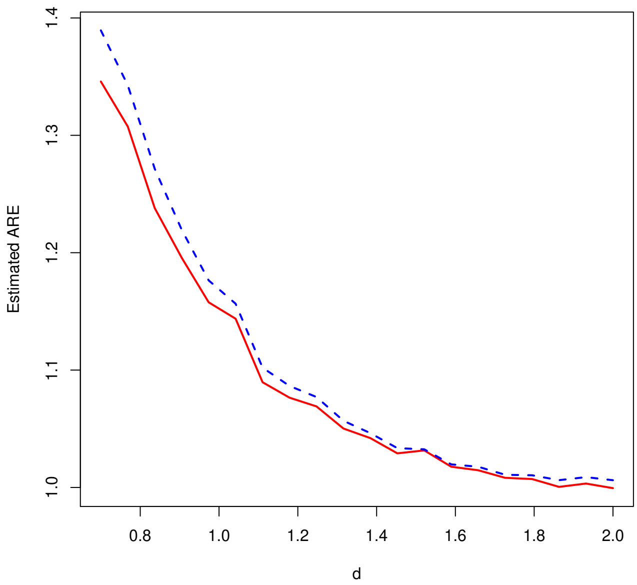

FIGURE 3 – As figure 2, but the ARE's are estimated from the residuals of a GARCH(1,1) with innovations $\eta_t \sim \Gamma(1, 2, d)$.

TABLE 2 – As Table 2, but for the 1% risk level.

	Gaussian-QMLE			GED-QMLE			Student-QMLE		
VaR parameter ω									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	2.557	6.076	11.025	2.14	3.562	6.433	2.218	4.301	8.407
0.97	3.033	2.534	2.818	0.519	1.017	1.458	1.464	1.577	1.929
1.66	0.074	0.151	0.212	0.08	0.163	0.223	0.080	0.159	0.218
2	0.042	0.091	0.132	0.044	0.091	0.131	0.045	0.092	0.134
VaR parameter α									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	0.163	0.225	0.297	0.168	0.198	0.258	0.143	0.187	0.253
0.97	0.077	0.111	0.148	0.077	0.084	0.119	0.077	0.102	0.138
1.66	0.012	0.051	0.086	0.012	0.051	0.087	0.012	0.05	0.085
2	0.008	0.044	0.075	0.008	0.044	0.076	0.008	0.044	0.076
VaR parameter β									
d	median	mean	RMSE	median	mean	RMSE	median	mean	RMSE
0.7	0.071	0.142	0.243	0.057	0.087	0.154	0.054	0.102	0.193
0.97	0.800	0.622	0.683	0.134	0.251	0.35	0.350	0.390	0.475
1.66	0.126	0.268	0.381	0.133	0.286	0.395	0.133	0.278	0.387
2	0.112	0.235	0.338	0.115	0.235	0.334	0.116	0.238	0.342

historical data exist. The number of observations varies from 5721 (for the DAX) to 5934 (for FTSE).

For each series of log-returns ϵ_t, we estimated the VaR parameter $\theta_{0,\alpha}$ of GARCH(1, 1) models. Tables 3 and 4 report the estimated VaR parameters, their related standard deviations and the estimated τ_h's for three different instrumental densities h, namely the Gaussian, Student(ν) and GED(κ) distributions. For the last two instrumental densities, we chose the parameters ν and κ which minimize the τ_h's that are estimated from the QMLE residuals (as explained in Section 5.2). The estimated values of the τ_h's are thus the same for $\alpha = 5\%$ and $\alpha = 1\%$, which is in concordance with the asymptotic theory, since the τ_h's do not depend on α, nor on the volatility parameter θ_0. Recall that the most accurate estimator is that with the smallest τ_h. Therefore, the estimators based on the GED and Student distributions should be much more accurate than that based on the Gaussian density. This not surprising because the Student and GED laws can have thicker tails than the normal distribution, and the financial series are known to have Leptokurtic conditional distributions. Thus, we addressed the issue of Leptokurticity through the use of Student and GED distributions. Over the 7 indices, it is clear to note that $\hat{\theta}^*_{n,\alpha}$ based on the GED and Student distributions are quite similar, with always a slight advantage (*i.e.* a smaller estimated τ_h) for the Student. The same conclusion can be drawn by looking at the estimated standard deviations, which are almost equal for the GED and Student distributions, and are clearly larger for the Gaussian instrumental density.

6 Conclusion

To conclude, we first summarize the outputs of the paper. We have considered a general volatility model with an unknown volatility parameter θ_0, and an unknown distribution P_η for the iid noise. We did not make any identifiability assumption, such as $E\eta_t^2 = 1$, and we considered a generalized QMLE based on an arbitrary instrumental density h. We are thus in a misspecified framework, where the volatility parameter is not well identified and the instrumental density is not the density of P_η in general. We have shown that, under mild regularity conditions, the gQMLE converges however to some "pseudo-true" value θ_0^* which depends on θ_0 and on some scale parameter depending on P_η and h.

Simply noting that, for any reasonable ARCH-type model, the ratio $\sigma_t(\theta_0^*)/\sigma_t(\theta_0)$ is constant, the conditional VaR at the level α can be obtained by multiplying $\sigma_t(\theta_0^*)$

TABLE 3 – Comparison of estimators of the 5% level VaR parameter for 7 daily stock market returns. The estimated standard deviation are displayed in brackets.

Index	h	$\omega_{5\%}$	$\alpha_{5\%}$	$\beta_{5\%}$	τ_h
CAC	ϕ	0.091 (0.021)	0.247 (0.030)	0.899 (0.011)	3.711
	GED	0.071 (0.015)	0.221 (0.024)	0.912 (0.008)	2.699
	St	0.065 (0.014)	0.220 (0.023)	0.914 (0.008)	2.537
DAX	ϕ	0.089 (0.026)	0.231 (0.041)	0.902 (0.016)	7.707
	GED	0.048 (0.011)	0.225 (0.024)	0.914 (0.008)	2.952
	St	0.045 (0.011)	0.230 (0.023)	0.913 (0.008)	2.676
FTSE	ϕ	0.037 (0.008)	0.243 (0.025)	0.906 (0.009)	2.780
	GED	0.035 (0.007)	0.230 (0.023)	0.911 (0.008)	2.513
	St	0.033 (0.007)	0.231 (0.023)	0.911 (0.008)	2.454
Nikkei	ϕ	0.153 (0.031)	0.286 (0.034)	0.878 (0.013)	3.517
	GED	0.110 (0.022)	0.249 (0.026)	0.897 (0.010)	2.803
	St	0.103 (0.020)	0.246 (0.025)	0.900 (0.009)	2.659
SMI	ϕ	0.137 (0.033)	0.353 (0.058)	0.845 (0.023)	7.429
	GED	0.076 (0.014)	0.319 (0.033)	0.877 (0.011)	2.908
	St	0.073 (0.013)	0.321 (0.032)	0.878 (0.010)	2.659
SP500	ϕ	0.028 (0.007)	0.204 (0.024)	0.918 (0.009)	3.777
	GED	0.020 (0.005)	0.192 (0.020)	0.926 (0.007)	2.997
	St	0.019 (0.005)	0.188 (0.019)	0.928 (0.007)	2.890
TSX	ϕ	0.021 (0.006)	0.230 (0.028)	0.914 (0.010)	4.347
	GED	0.016 (0.004)	0.204 (0.021)	0.924 (0.007)	2.887
	St	0.017 (0.004)	0.207 (0.021)	0.923 (0.007)	2.735

Table 4 – As Table 4, but for the risk level 1%.

Index	h	$\omega_{1\%}$	$\alpha_{1\%}$	$\beta_{1\%}$	τ_h
CAC	ϕ	0.198 (0.045)	0.537 (0.067)	0.899 (0.011)	3.711
	GED	0.153 (0.032)	0.478 (0.053)	0.912 (0.008)	2.699
	St	0.140 (0.030)	0.474 (0.050)	0.914 (0.008)	2.537
DAX	ϕ	0.203 (0.059)	0.525 (0.093)	0.902 (0.016)	7.707
	GED	0.112 (0.026)	0.526 (0.057)	0.914 (0.008)	2.952
	St	0.107 (0.024)	0.540 (0.056)	0.913 (0.008)	2.676
FTSE	ϕ	0.084 (0.018)	0.550 (0.061)	0.906 (0.009)	2.780
	GED	0.079 (0.017)	0.523 (0.057)	0.911 (0.008)	2.513
	St	0.076 (0.016)	0.528 (0.057)	0.911 (0.008)	2.454
Nikkei	ϕ	0.332 (0.068)	0.622 (0.076)	0.878 (0.013)	3.517
	GED	0.234 (0.047)	0.528 (0.058)	0.897 (0.010)	2.803
	St	0.221 (0.044)	0.527 (0.056)	0.900 (0.009)	2.659
SMI	ϕ	0.316 (0.077)	0.814 (0.137)	0.845 (0.023)	7.429
	GED	0.174 (0.032)	0.726 (0.079)	0.877 (0.011)	2.908
	St	0.165 (0.030)	0.729 (0.075)	0.878 (0.010)	2.659
SP500	ϕ	0.070 (0.017)	0.506 (0.061)	0.918 (0.009)	3.777
	GED	0.048 (0.012)	0.462 (0.05)	0.926 (0.007)	2.997
	St	0.047 (0.012)	0.455 (0.049)	0.928 (0.007)	2.890
TSX	ϕ	0.056 (0.015)	0.602 (0.083)	0.914 (0.01)	4.347
	GED	0.043 (0.011)	0.541 (0.064)	0.924 (0.007)	2.887
	St	0.044 (0.011)	0.546 (0.064)	0.923 (0.007)	2.735

by the α-quantile of $\eta_t^* = \epsilon_t/\sigma_t(\theta^*)$. This shows that the natural two-step method leads to a consistent estimation of the VaR, even the instrumental density h does not coincide with P_η. The result extends to the Expected Shortfall and to other DRM. The asymptotic and finite-sample accuracy of the method however depends on θ_0, h and P_η. We have shown that, for a large class of standard GARCH models, the optimal choice of h only depends on P_η and can be estimated easily. It is shown that, compared to the usual two-step method based on the Gaussian QMLE, important efficiency gains can be achieved by appropriately choosing the instrumental density.

Future extensions of this work could be the following. Firstly, it could be interesting to extend Corollary 2.1 in the case of a DRM parameter. Such a result could be used to obtain confidence intervals for DRM that would integrate the estimation risk. This extension is however far from being trivial because it should involve the limit distribution of the random function $\sqrt{n}\left(\hat{\theta}^*_{n,\alpha} - \theta_{0,\alpha}\right)$ where α varies in $[0, 1]$. Another potential extension would be to consider conditional risk measures for a time horizon larger than 1. Existing techniques are based on scenario simulations. The question of interest would be to determine whether such simulation techniques are more efficient at any horizon when they are based on models estimated by an optimal gQMLE than when they are based on the Gaussian QMLE.

References

Artzner, P., Delbaen, F., Eber, J-M. and D. Heath (1999) Coherent measures of risk. *Mathematical Finance* 9, 203–228.

Bardet, J-M. and O. Wintenberger (2009) Asymptotic normality of the Quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* 37, 2730–2759.

Bassett, G-W. and R-W. Koenker (1986) Strong Consistency of Regression Quantiles and Related Empirical Processes. *Econometric Theory* 2, 191–201.

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH processes : structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P. (1961) *Statistical inference for Markov processes.* Chicago : University of Chicago Press.

Billingsley, P. (1995) *Probability and Measure.* Third edition John Wiley and Sons

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

Davis, R.A., Knight, K. and J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145–180.

Ding, Z., Granger C. et R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Engle, R-F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987–1008.

Fan, J., Qi, L. and D. Xiu (2013) Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. Available at SSRN : `http://ssrn.com/abstract=1540363orhttp://dx.doi.org/10.2139/ssrn.1540363`

Francq, C., Lepage, G. and J-M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. and J-M. Zakoïan (2012) Risk-parameter estimation in volatility models. MPRA Preprint No. 41713.

Francq, C. and J.M. Zakoïan (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society - Series B* 75, 345–367.

Glosten, L., Jagannathan, R. and D. Runkle (1993) Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Hamadeh, T., and J-M. Zakoïan (2011)Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *Journal of Statistical Planning and Inference* 141, 488Ũ-507.

Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics* 26, 755–770.

Koenker, R. (2005)*Quantile Regression.* Cambridge : Cambridge University Press.

Koenker, R. and Z. Xiao (2006) Quantile autoregression. *Journal of the American Statistical Association* 101, 980–990.

Kuester, K., Mittnik, S. and M.S. Paolella (2006) Value-at-Risk predictions : A comparison of alternative strategies. *Journal of Financial Econometrics* 4, 53–89.

Lee, S.W. and B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10, 29–52.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

McNeil, A. J., Frey, R. and P. Embrechts (2005) *Quantitative Risk Management.* Princeton University Press.

Mikosch, T. and D. Straumann (2006) Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics* 34, 493–522.

Pollard, D. (1991) Asymptotics for Least Absolute Deviation Regression Estimators *Econometric Theory* 7, 186–199.

Straumann, D. and T. Mikosch (2006) Quasi-maximum likelihood estimation in conditionally heteroscedastic Time Series : a stochastic recurrence equations approach. *The Annals of Statistics* 5, 2449–2495.

Wang, S. (2000) A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67, 15–36.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

Wirch, J. L. and M. R. Hardy (1999) A Synthesis of Risk Measures for Capital Adequacy. *Insurance : Mathematics and Economics* 25, 337–347.

Xiao, Z. and R. Koenker (2009) Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 104, 1696–1712.

Xiao, Z. and C. Wan (2010) A robust estimator of conditional volatility. Unpublished document.

Zakoïan J.M. (1994) Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18, 931–955.

Zhu, K. and S. Ling (2011) Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *The Annals of Statistics* 32, 2131–2163.

A Proofs

A.1 Proof of Lemma 2.1

The proof is similar to that of Theorem 2.1 in FZ. It rests on the following intermediate results :

$$i)\quad \lim_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| = 0 \ , \quad a.s.$$

$$ii)\ \text{ if } \theta \neq \theta_0^* \ , \quad \mathbb{E}g(\epsilon_1, \sigma_1(\theta)) < \mathbb{E}g(\epsilon_1, \sigma_1(\theta_0^*)) \ ,$$

$iii)$ any $\theta \neq \theta_0^*$ has a neighborhood $V(\theta)$ such that

$$\limsup_{n\to\infty} \sup_{\theta^*\in V(\theta)} \widetilde{Q}_n(\theta^*) < \limsup_{n\to\infty} \widetilde{Q}_n(\theta_0^*) \ , \quad a.s.$$

where

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \sigma_t(\theta)),$$

$$iv)\quad \lim_{n\to\infty} \sqrt{n} \sup_{\theta\in V(\theta_0^*)} \left\| \frac{\partial}{\partial\theta} Q_n(\theta) - \frac{\partial}{\partial\theta}\widetilde{Q}_n(\theta) \right\| = 0 \ , \quad \text{in probability,}$$

for some neighborhood $V(\theta_0^*)$ of θ_0^*,

$$v)\quad J_* \text{ invertible and } \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta^*) \to \frac{Eg_2(\sigma_*^{-1}\eta_0, 1)}{4} J_* \ , \quad \text{in probability,}$$

for any θ^* between $\hat{\theta}_n^*$ and θ_0^*,

$$vi)\quad \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{Eg_1^2(\sigma_*^{-1}\eta_0, 1)}{4} J_*\right).$$

We begin to show *i)*. First note that a Taylor expansion and **A5** show that

$$g(\epsilon_t, \widetilde{\sigma}_t(\theta)) - g(\epsilon_t, \sigma_t(\theta)) = g_1(\epsilon_t, \sigma_t^*(\theta)) \left\{\widetilde{\sigma}_t(\theta) - \sigma_t(\theta)\right\}$$

where

$$g_1(\epsilon, \sigma) = -\frac{1}{\sigma}\left\{1 + \frac{\epsilon}{\sigma}\frac{h'}{h}\left(\frac{\epsilon}{\sigma}\right) 1_{\epsilon\neq 0}\right\} \qquad \text{(A.1)}$$

and $\sigma_t^*(\theta)$ is between $\widetilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. Using **A3** and **A5**, we then have almost surely

$$\begin{aligned} \sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| &\leq C_1 n^{-1} \sum_{t=1}^{n} \sup_{\theta\in\Theta} |g_1(\epsilon_t, \sigma_t^*(\theta))| \rho^t \\ &\leq \frac{C_1}{n\underline{\omega}} \sum_{t=1}^{n} \rho^t \left\{1 + C_0\left(1 + \left|\frac{\epsilon_t}{\underline{\omega}}\right|^{\delta}\right)\right\}. \end{aligned}$$

The Markov inequality and **A2** entail

$$\sum_{t=1}^{\infty} \mathbb{P}(\rho^t |\epsilon_t|^\delta > \varepsilon) \ \leq \ \sum_{t=1}^{\infty} \frac{\rho^{st/\delta} \mathbb{E}|\epsilon_t|^s}{\varepsilon^s} < \infty \tag{A.2}$$

and thus the proof of *i)* is completed by the Borel-Cantelli lemma.

To prove *ii)*, first note that by **A2**

$$g(\epsilon_t, \sigma_t(\theta)) = g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - \log \sigma_t(\theta_0).$$

Moreover, by **A1** and (2.2), we have

$$\frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta_0)} = \sigma_*,$$

where σ_* is defined in **A4**. In view of **A3**-**A4**, we thus have

$$\mathbb{E}\{g(\epsilon_1, \sigma_1(\theta)) - g(\epsilon_1, \sigma_1(\theta_0^*))\} \ = \ \mathbb{E}\left\{g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - g(\eta_t, \sigma_*)\right\} \leq 0,$$

with equality if and only if $\theta = \theta_0^*$, which shows *ii)*.

We now turn to the proof of *iii)*. For any $\theta \in \Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with center θ and radius $1/k$. We have,

$$\begin{aligned} & \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \widetilde{Q}_n(\theta^*) \\ \leq \ & \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} Q_n(\theta^*) + \limsup_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| \\ \leq \ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) \quad a.s. \end{aligned}$$

where the second inequality comes from *i)*. Note that since h is integrable and continuous, h is bounded by some constant C. It follows, by **A3**, that

$$\mathbb{E} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) < \log \frac{1}{\underline{\omega}} + \log C < \infty. \tag{A.3}$$

Using **A2** and an ergodic theorem for stationary and ergodic processes (X_t) such that $\mathbb{E}(X_t)$ exists in $\mathbb{R} \cup \{-\infty\}$ (see Billingsley, 1995, p. 284 and 495), it follows that

$$\limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \widetilde{Q}_n(\theta^*) \leq \mathbb{E}X_{t,k}(\theta), \qquad X_{t,k}(\theta) = \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) \ .$$

When k tends to infinity, the sequence $\{X_{t,k}(\theta)\}_k$ decreases to $X_t(\theta) = g(\epsilon_t, \sigma_t(\theta))$. Thus $\{X_{t,k}^-(\theta)\}_k$ increases to $X_t^-(\theta)$. By the Beppo-Levi theorem,

$\mathbb{E}X^-_{t,k}(\theta) \uparrow \mathbb{E}_{\theta_0} X^-_t(\theta)$ when $k \uparrow +\infty$. By (A.3), the fact that the sequence $\{X^+_{t,k}(\theta)\}_k$ is decreasing, and the Lebesgue theorem, $\mathbb{E}X^+_{t,k}(\theta) \downarrow \mathbb{E}X^+_t(\theta)$ when $k \uparrow +\infty$. Thus we have shown that $\mathbb{E}X_{t,k}$ converges to $\mathbb{E}\{X_t(\theta)\}$ when $k \to \infty$. By *ii)*, *iii)* is proved.

The consistency is a consequence of **A7**, a standard compactness argument and of the intermediate results *i)-iii)*.

Now we prove *iv)*. We have

$$\begin{aligned}
\frac{\partial}{\partial\theta} Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \sigma_t(\theta)) \frac{\partial \sigma_t(\theta)}{\partial\theta}, \\
\frac{\partial}{\partial\theta} \widetilde{Q}_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \widetilde{\sigma}_t(\theta)) \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial\theta}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0^*)} \sqrt{n} \left\| \frac{\partial}{\partial\theta} Q_n(\theta) - \frac{\partial}{\partial\theta} \widetilde{Q}_n(\theta) \right\| \\
\leq\ & \sup_{\theta\in V(\theta_0^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_1(\epsilon_t, \sigma_t(\theta)) - g_1(\epsilon_t, \widetilde{\sigma}_t(\theta))| \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \\
&+ \sup_{\theta\in V(\theta_0^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_1(\epsilon_t, \widetilde{\sigma}_t(\theta))| \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} - \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial\theta} \right\|. \qquad \text{(A.4)}
\end{aligned}$$

In view of **A5** and **A10**, the last term is bounded by

$$\frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} \rho^t \left\{ 1 + C_0 \left(1 + \left| \frac{\epsilon_t}{\underline{\omega}} \right|^\delta \right) \right\},$$

which is a.s. an $O(1/\sqrt{n})$ by arguments used to show *i)*. Thus it remains to show that the first term on the right-hand side of the inequality (A.4) converges also to zero a.s. as n tends to infinity. Noting that

$$g_2(x,\sigma) := \frac{\partial g_1(x,\sigma)}{\partial\sigma} = \frac{1}{\sigma^2}\left[1 + \frac{x}{\sigma}\left\{ 2\frac{h'}{h} + \frac{x}{\sigma}\left(\frac{h'}{h}\right)' \right\} \left(\frac{x}{\sigma}\right) 1_{x\neq 0} \right], \qquad \text{(A.5)}$$

and using **A5**, **A6** and **A11**, this term is bounded by

$$\begin{aligned}
&\frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} |g_2(\epsilon_t, \sigma_t^*)| \rho^t \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \\
\leq\ & \frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} \rho^t \left\{ 1 + 3C_0 \left(1 + \left| \frac{\epsilon_t}{\underline{\omega}} \right|^\delta \right) \right\} \sup_{\theta\in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \qquad \text{(A.6)}
\end{aligned}$$

where $\sigma_t^* = \sigma_t^*(\theta)$ is between $\widetilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. Using the Cauchy-Schwarz inequality, **A12**, and already given arguments, it can be show that the right-hand side of (A.6) is a.s. equal to $O(1/\sqrt{n})$. It follows that the right-hand side of (A.4) tends to zero, which completes the proof of *iv)*.

Now we establish *v)*. The invertibility of J_* follows from **A9**. Using **A5** and **A11**, we have

$$\begin{aligned}\left\|\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta))}{\partial\theta\partial\theta'}\right\| &= \left\|g_2(\epsilon_t,\sigma_t(\theta))\frac{\partial\sigma_t(\theta)}{\partial\theta}\frac{\partial\sigma_t(\theta)}{\partial\theta'}+g_1(\epsilon_t,\sigma_t(\theta))\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\| \\ &\leq \left\{1+3C_0\left(1+\left|\frac{\eta_t}{\sigma_*}\right|^{\delta}\left|\frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)}\right|^{\delta}\right)\right\}\left(\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\|\right. \\ &\qquad \left.+\left\|\frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\frac{\partial\sigma_t(\theta)}{\partial\theta'}\right\|\right).\end{aligned}$$

Hence

$$E\sup_{\theta\in V(\theta_0^*)}\left\|\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta))}{\partial\theta\partial\theta'}\right\|<\infty$$

by the Hölder inequality, **A5** and **A12**. The ergodic theorem then implies that

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\theta\in V(\theta_0^*)}\left\|\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'}\right\| \\ \leq\ & E\sup_{\theta\in V(\theta_0^*)}\left\|\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta))}{\partial\theta\partial\theta'}-\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta_0^*))}{\partial\theta\partial\theta'}\right\| \qquad a.s.\end{aligned}$$

By the dominated convergence theorem, the last expectation tends to zero when the neighborhood $V(\theta_0^*)$ tends to the singleton $\{\theta_0^*\}$. The consistency of $\hat{\theta}_n^*$ thus entails

$$\lim_{n\to\infty}\left|\frac{\partial^2 Q_n(\theta^*)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'}\right|=0,\quad a.s.$$

Now, note that by **A4**, **A5** and the dominated convergence theorem

$$Eg_1(\eta_0,\sigma_*)=0,\qquad \text{and thus } Eg_1(\sigma_*^{-1}\eta_0,1)=0. \tag{A.7}$$

Moreover, we have

$$g_1\left\{\epsilon_t,\sigma_t(\theta_0^*)\right\}=g_1\left\{\sigma_t(\theta_0)\eta_t,\sigma_*\sigma_t(\theta_0)\right\}=\sigma_t^{-1}(\theta_0^*)g_1(\sigma_*^{-1}\eta_t,1).$$

It follows that

$$Eg_1\left\{\epsilon_t,\sigma_t(\theta_0^*)\right\}\frac{\partial^2\sigma_t(\theta_0^*)}{\partial\theta\partial\theta'}=0.$$

Similarly, in view of (A.5), $g_2(\epsilon_t,\sigma_t(\theta_0^*))=\sigma_t^{-2}(\theta_0^*)g_2(\sigma_*^{-1}\eta_t,1)$. We also have $\partial\sigma_t^2(\theta)/\partial\theta=2\sigma_t(\theta)\partial\sigma_t(\theta)/\partial\theta$. By the ergodic theorem, we then have

$$\lim_{n\to\infty}\frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'}=\frac{Eg_2(\sigma_*^{-1}\eta_0,1)}{4}J_*,\quad a.s.$$

and *v)* is established.

To prove *vi)* it suffices to note that, using arguments used to show *v)*,

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0^*) = \frac{1}{\sqrt{n}}\sum_{t=1}^n g_1\left(\sigma_*^{-1}\eta_t, 1\right)\frac{1}{2\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} \qquad \text{(A.8)}$$

and to apply a CLT for square integrable stationary martingale differences (see Billingsley (1961)).

Now, from **A8** and the consistency of $\hat{\theta}_n^*$, a Taylor expansion shows that for n large enough

$$\begin{aligned} 0 &= \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n^*) + \sqrt{n}\frac{\partial}{\partial\theta}\widetilde{Q}_n(\hat{\theta}_n^*) - \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n^*) \\ &= \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0^*) + \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta^*)\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\ &\quad +\sqrt{n}\left(\frac{\partial}{\partial\theta}\widetilde{Q}_n(\hat{\theta}_n^*) - \frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n^*)\right), \end{aligned}$$

where θ^* is between $\hat{\theta}_n^*$ and θ_0^*. Applying *iv)* and *v)* we obtain

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) = \frac{-4}{Eg_2\left(\sigma_*^{-1}\eta_t, 1\right)}J_*^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0^*) + o_P(1). \qquad \text{(A.9)}$$

and the proof of the asymptotic normality comes from *vi)*. □

A.2 Proof of Theorem 2.1

Following Koenker (2006), we have

$$\hat{\xi}_{\alpha,n}^* = \arg\min_{z\in\mathbb{R}}\sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z),$$

where $\rho_\alpha(u) = u(\alpha - \mathbf{1}_{u<0})$. Thus

$$\sqrt{n}(\hat{\xi}_{\alpha,n}^* - \xi_\alpha^*) = \arg\min_{z\in\mathbb{R}} O_n(z)$$

where

$$O_n(z) = \sum_{t=1}^n \rho_\alpha\left(\hat{\eta}_t^* - \xi_\alpha^* - \frac{z}{\sqrt{n}}\right) - \sum_{t=1}^n \rho_\alpha(\eta_t^* - \xi_\alpha^*).$$

A Taylor expansion around θ_0^* and **A3**, **A6** yield

$$\hat{\eta}_t^* = \eta_t^* - \eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) + \frac{1}{2}(\hat{\theta}_n^* - \theta_0^*)'\frac{\partial^2\eta_t(\theta^*)}{\partial\theta\partial\theta'}(\hat{\theta}_n^* - \theta_0^*) + \eta_t^* O(\rho^t)$$

where $D_t = D_t(\theta_0^*)$ and θ^* is between $\hat{\theta}_n^*$ and θ_0^*. Using (2.1), we thus have

$$\begin{aligned} O_n(z) &= \sum_{t=1}^n \rho_\alpha\left(\eta_t^* - \xi_\alpha^* - \eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) - \frac{z}{\sqrt{n}} + o_P(n^{-1/2}) + O_P(\rho^t)\right) \\ &\quad -\rho_\alpha(\eta_t^* - \xi_\alpha^*). \end{aligned}$$

Using the identity

$$\rho_\alpha(u-v) - \rho_\alpha(u) = -v(\alpha - \mathbf{1}_{\{u<0\}}) + \int_0^v \left\{\mathbf{1}_{\{u\le s\}} - \mathbf{1}_{\{u<0\}}\right\} ds$$

for $u \neq 0$ (see Equation (A.3) in Koenker and Xiao, 2006), we then obtain

$$O_n(z) = zX_n + Y_n + Z_n(z) + W_n(z)$$

where

$$\begin{aligned} X_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}} - \alpha), \quad Y_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n R_{t,n}(\mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}} - \alpha), \\ Z_n(z) &= \sum_{t=1}^n \int_0^{z/\sqrt{n}} (\mathbf{1}_{\{\eta_t^* \le \xi_\alpha^* + s\}} - \mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}}) ds, \\ W_n(z) &= \sum_{t=1}^n \int_{z/\sqrt{n}}^{(z+R_{t,n})/\sqrt{n}} (\mathbf{1}_{\{\eta_t^* \le \xi_\alpha^* + s\}} - \mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}}) ds \end{aligned}$$

with $R_{t,n} = \eta_t^* D_t' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1) + \sqrt{n} O_P(\rho^t)$.

By the change of variable $w = s - z/\sqrt{n}$, we have $W_n(z) = \sum_{i=1}^2 W_n^{(i)}(z)$ in which $W_n^{(i)}(z) = \sum_{t=1}^n W_{n,t}^{(i)}$ where

$$\begin{aligned} W_{n,t}^{(1)} &= \int_0^{R_{t,n}/\sqrt{n}} (\mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} \le w\}} - \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} < 0\}}) dw, \\ W_{n,t}^{(2)} &= \int_0^{R_{t,n}/\sqrt{n}} \left(\mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} < 0\}} - \mathbf{1}_{\{\eta_t^* - \xi_\alpha^*\} < 0}\right) dw. \end{aligned}$$

Note that the integrand in $W_{n,t}^{(2)}$ does not depend on w. Therefore, we have

$$W_{n,t}^{(2)} = \left\{\eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) + o_P(n^{-1/2}) + O_P(\rho^t)\right\} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}}$$

when $z \ge 0$, and

$$W_{n,t}^{(2)} = -\left\{\eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) + o_P(n^{-1/2}) + O_P(\rho^t)\right\} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in (z/\sqrt{n}, 0)\}}$$

when $z < 0$.

First consider the case $z \geq 0$. Note that

$$\begin{aligned}\sum_{t=1}^{n} W_{n,t}^{(2)} &= \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t^* \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} D_t' \right) \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\ &\quad + o_P(n^{-1/2}) \sum_{t=1}^{n} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} \\ &\quad + O_P(1) \sum_{t=1}^{n} \rho^t \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}}. \qquad \text{(A.10)}\end{aligned}$$

The term $\sum_{t=1}^{n} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} = O_P(\sqrt{n})$ because its expectation is $O(\sqrt{n})$ and its variance is $O(\sqrt{n})$. It follows that the second term on the right-hand side of (A.10) tends to zero in probability. By the same arguments, we show that the third term has the same behavior. Now, noting that $\xi_\alpha^* f^*(\xi_\alpha^*) = \xi_\alpha f(\xi_\alpha)$ when f^* is the density of $\eta_1^* = \eta_1/\sigma_*$, we have

$$\begin{aligned}E(\eta_t^* \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} &= \int_0^{z/\sqrt{n}} (x + \xi_\alpha^*) f^*(x + \xi_\alpha^*) dx \\ &= \xi_\alpha f(\xi_\alpha) \frac{z}{\sqrt{n}} + o(1/\sqrt{n}).\end{aligned}$$

Thus, in view of the independence of η_t^* and D_t, we have

$$E\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t^* \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} D_t' \right) = z \xi_\alpha f(\xi_\alpha) \Omega_*' + o(1).$$

By similar computations we find

$$\text{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t^* \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} D_t' \right) = o(1).$$

It follows that

$$\sum_{t=1}^{n} W_{n,t}^{(2)} = z \xi_\alpha f(\xi_\alpha) \Omega_*' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o(1) \;\; a.s.$$

The same equality holds for $z \leq 0$.

We now denote by $E_{t-1}X$ the expectation of some variable X conditional on $\{\hat{\theta}_n^* - \theta_0^*, (\eta_u : u < t)\}$. We have, by the change of variable $w = \eta_t^* v$,

$$\begin{aligned}E_{t-1} W_{n,t}^{(1)} &= \int_0^{D_t'(\hat{\theta}_n^* - \theta_0^*) + o(n^{-1/2})} E_{t-1}(\eta_t^* \mathbf{1}^*_{\{\eta_t^* \in (\xi_\alpha^* + z/\sqrt{n}, (\xi_\alpha^* + z/\sqrt{n})(1-v)^{-1})\}}) dv \\ &= \frac{(\xi_\alpha^*)^2}{2} f_{n,t}^*(\xi_\alpha^*)(\hat{\theta}_n^* - \theta_0^*)' D_t D_t'(\hat{\theta}_n^* - \theta_0^*) + o(n^{-1}) \quad a.s.\end{aligned}$$

where $f_{n,t}^*$ denotes the density of η_t^* conditional on $\{\hat{\theta}_n^* - \theta_0^*, (\eta_u : u < t)\}$ and $o(n^{-1})$ is a function of $(\hat{\theta}_n^* - \theta_0^*)$ and the past values of η_t^*. By the arguments used for $X_{n,t}^{(2)}$ it can therefore be shown that $W_n^{(1)}(z)$ converges in distribution to a random variable which does not depend on z. Note also that Y_n can be subtracted from the objective function $O_n(z)$ because it does not depend on z. Moreover $Z_n(z) \to \frac{z^2}{2} f^*(\xi_\alpha^*)$ in probability as $n \to \infty$. Finally,

$$\tilde{O}_n(z) := O_n(z) - Y_n = \frac{z^2}{2} f^*(\xi_\alpha^*) + z\{X_n + \xi_\alpha^* f(\xi_\alpha)\Omega_*' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*)\} + O_P(1).$$

Since the process $\tilde{O}_n(\cdot)$ has convex sample paths, the convexity Lemmas of Knight (1989) and Pollard (1991) show that $\tilde{O}_n$ converges weakly to some convex process. By Lemma 2.2 in Davis et al. (1992), we can conclude that

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{\alpha,n}^* - \xi_\alpha^*) &= -\xi_\alpha^* \Omega_*' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\ &\quad - \frac{1}{f^*(\xi_\alpha^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}} - \alpha) + o_P(1).\end{aligned}$$

In view of (A.8) and (A.9), we have

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) = \frac{-4}{Eg_2(\eta_0^*, 1)} J_*^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n g_1(\eta_t^*, 1) D_t(\theta_0^*) + o_P(1).$$

By the CLT for martingale differences, we get

$$S_n := \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} g_1(\eta_t^*, 1) D_t(\theta_0^*) \\ \mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}} - \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \begin{pmatrix} \frac{Eg_1^2(\eta_1^*, 1)}{4} J_* & c_\alpha \Omega_* \\ c_\alpha \Omega_*' & \alpha(1-\alpha) \end{pmatrix}\right\}.$$

The result follows from

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n^* - \theta_0^* \\ \hat{\xi}_{\alpha,n}^* - \xi_\alpha^* \end{pmatrix} = \begin{pmatrix} \frac{-4}{Eg_2(\eta_0^*,1)} J_*^{-1} & 0_m \\ \frac{4\xi_\alpha^*}{Eg_2(\eta_0^*,1)} \Omega_*' J_*^{-1} & \frac{-1}{f^*(\xi_\alpha^*)} \end{pmatrix} S_n,$$

using the relation $\Omega_*' J_*^{-1} \Omega_* = 1/4$ by Remark 3.1 in FZ. □

A.3 Proof of Corollary 2.2

In view of (2.1), when h is replaced by h_s, then $\hat{\theta}_n^*$ is replaced by $\hat{\theta}_n^{(s)}$ such that $\hat{\theta}_n^* = H(\hat{\theta}_n^{(s)}, s)$. It is then clear that $\tilde{\sigma}_t(\hat{\theta}_n^*)$ and η_t^* are replaced by respectively $\tilde{\sigma}_t(\hat{\theta}_{n,s}^*) = s^{-1}\tilde{\sigma}_t(\hat{\theta}_n^*)$ and $s\eta_t^*$, and thus the VaR estimator is unchanged. □

A.4 Proof of Theorem 4.1

If G is the cdf of a discrete distribution, that puts the masses $p_1, \ldots, p_k$ at the points $0 < \alpha_1 < \ldots < \alpha_k < 1$, then

$$\int_0^1 \xi_u dG(u) = \sum_{i=1}^{k} \xi_{\alpha_i} p_i \quad \text{and} \quad \int_0^1 \hat{\xi}^*_{n,u} dG(u) = \sum_{i=1}^{k} \hat{\xi}^*_{n,\alpha_i} p_i.$$

By Basset and Koenker (1986), $\hat{\xi}^*_{n,\alpha_i} \to \xi_{\alpha_i}$ a.s. for $i = 1, \ldots, k$, and the result follows from Lemma 2.1 in the discrete case.

For a general distortion function G and for all $\varepsilon > 0$, one can define discrete distributions G_1 and G_2 such that

$$\int_0^1 \xi_u dG_1(u) \leq \int_0^1 \xi_u dG(u) \leq \int_0^1 \xi_u dG_2(u)$$

and

$$\left| \int_0^1 \xi_u dG_1(u) - \int_0^1 \xi_u dG_2(u) \right| < \varepsilon.$$

The conclusion then follows from the consistency in the discrete case. $\square$