Risk-parameter estimation in volatility models

Christian Francq

CREST
and
University Lille 3 (EQUIPPE), BP 60149
59653 Villeneuve d'Ascq cedex, France
e-mail: christian.francq@univ-lille3.fr

and

Jean-Michel Zakoïan

CREST, 15 Boulevard Gabriel Peri
92245 Malakoff Cedex, France
and
University Lille 3 (EQUIPPE)
e-mail: zakoian@ensae.fr
[*]

Abstract: This paper introduces the concept of risk parameter in conditional volatility models of the form $\epsilon_t = \sigma_t(\theta_0)\eta_t$ and develops statistical procedures to estimate this parameter. For a given risk measure r, the risk parameter is expressed as a function of the volatility coefficients θ_0 and the risk, $r(\eta_t)$, of the innovation process. A two-step method is proposed to successively estimate these quantities. An alternative one-step approach, relying on a reparameterization of the model and the use of a non Gaussian QML, is proposed. Asymptotic results are established for smooth risk measures as well as for the Value-at-Risk (VaR). Asymptotic comparisons of the two approaches for VaR estimation suggest a superiority of the one-step method when the innovations are heavy-tailed. For standard GARCH models, the comparison only depends on characteristics of the innovations distribution, not on the volatility parameters. Monte-Carlo experiments and an empirical study illustrate these findings.

Keywords and phrases: GARCH, Quantile Regression, Quasi-Maximum Likelihood, Risk measures, Value-at-Risk.

[*]We are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project ECONOM&RISK (ANR 2010 blanc 1804 03).

1. Introduction

Modern financial risk management generally focuses on risks measures based on distributional information. Compared to traditional approaches relying on the marginal distribution of returns, more sophisticated approaches view risk as a stochastic process. For instance, conditional Value-at-Risk (VaR) - arguably the most widely used measure since the 1996 amendment of the Basel Capital Accord - is defined as the opposite of a quantile of the returns (or profit & losses, P&L, variables) conditional distribution. Another popular risk measure is the conditional Expected Shortfall which, conditional on the past returns, measures the average loss when the loss is above the VaR[1]. Many econometric approaches have been proposed in the finance and statistical literatures for measuring conditional risk.

A crucial issue that arises in this context is how to evaluate the performance of conditional risk estimators. Comparison of the performances of estimators of parameters based on the asymptotic theory is standard. But comparing the performances of VaR estimators, for instance, is more intricate because the conditional VaR is a random process, not a parameter.

The first objective of this paper is to introduce a concept of *risk parameter* in conditional volatility models. The risk parameter can be interpreted as a summary of conditional risk. Summaries of unconditional risk (such as the VaR based on historical simulation) are commonly used but they do not account for the dynamics of risk. By contrast, risk parameters are vector coefficients which take into account the returns dynamics and for which an asymptotic theory of estimation can be derived.

To be more specific, consider a conditional volatility model of the form

$$\epsilon_t = \sigma_t(\theta_0)\eta_t, \tag{1.1}$$

where ϵ_t denotes the log-return, σ_t is a volatility process, that is a positive measurable function of the past log-returns, θ_0 is a finite-dimensional parameter and (η_t) is a sequence of independent and identically distributed (iid) random variables, η_t being also independent of the past returns. Consider a risk measure, r, satisfying the assumption of positive homogeneity, such as the VaR or the Expected Shortfall. Then the conditional risk of ϵ_t is given by

$$r_{t-1}(\epsilon_t) = \sigma_t(\theta_0)r(\eta_t),$$

[1] In the risk management literature, the term "conditional VaR" sometimes refer to what many authors, including us in this article, call Expected Shortfall. In this paper, we call conditional risks the risks computed conditional on the past returns.

where $r(\eta_t)$ is a constant. In most parametric volatility models, multiplying the volatility by a constant amounts to modifying the parameter value. Under this assumption, we have

$$r_{t-1}(\epsilon_t) = \sigma_t(\theta_0^*), \quad \text{where} \quad \theta_0^* = H\{\theta_0, r(\eta_t)\} \tag{1.2}$$

for some function H which is specific to the model under consideration. In this setting, we call θ_0^* the *risk parameter* associated to the risk function r. It incorporates not only the volatility parameters but also the (unconditional) risk of the innovation process (η_t). When r is the risk associated with the VaR at some level $\alpha \in (0, 1)$, the vector θ_0^* is referred to as the *VaR parameter* at level α.

Deriving an asymptotic theory for estimators of risk parameters is the second objective of this article. Two estimation procedures will be studied and compared. A two-step approach relies on the expression of θ_0^* in (1.2). Under the identifiability assumption

$$E\eta_t^2 = 1, \tag{1.3}$$

a consistent and asymptotically normal (CAN) estimator $\hat{\theta}$ of the parameter θ_0 can be obtained by standard methods for conditional volatility models, the most widely used being the Gaussian Quasi-Maximum Likelihood (QML). In a second step, an estimator $\hat{r}$ of the innovation risk $r(\eta_t)$ can be constructed, under conditions to be discussed, from the residuals $\hat{\eta}_t = \epsilon_t/\sigma_t(\hat{\theta})$ of the first step. A consistent estimator $H\{\hat{\theta}, \hat{r}\}$ of the risk parameter, θ_0^*, will be deduced (under smoothness assumptions on the function H). The asymptotic distribution of this estimator will follow from the joint asymptotic distribution of $\{\hat{\theta}, \hat{r}\}$.

An alternative strategy of estimation introduced in this article relies on a reparameterization of the conditional volatility model. The multiplicative form of model (1.1) generally allows us to rewrite it as

$$\epsilon_t = \sigma_t(\theta_0^*)\eta_t^*, \quad \text{with} \quad r(\eta_t^*) = 1.$$

The latter equality replaces the standard assumption (1.3). The interest of such a representation is that, if a consistent estimator $\hat{\theta}_0^*$ of θ_0^* can be obtained, the conditional risk $r_{t-1}(\epsilon_t)$ of ϵ_t can be estimated in one step by $\sigma_t(\hat{\theta}_0^*)$.

Estimation of conditional volatility models under moment conditions different from (1.3) has been studied by Berkes and Horváth (2004), Francq, Lepage and Zakoian (2011), Zhu and Ling (2011), Francq and Zakoian (2012).

In the framework of this paper, the condition $r(\eta_t^*) = 1$ is not necessarily a moment condition. We propose a QML approach based on non-Gaussian densities depending on the risk function r. A case of particular importance is the VaR at a given level α: the identifiability condition consists in setting an appropriate quantile of the distribution of η_t^* to unity. It turns out that the only asymptotically valid QML criterion, that is, ensuring the consistency of the QML estimator of θ_0^* whatever the distribution of η_t^*, takes the form of a non linear quantile regression criterion.

The third objective of this article is to compare the one-step and two-step estimators of the VaR parameter. As we will see, the assumptions required for the CAN of the two estimators are quite different. When such assumptions are met, the asymptotic variances can be compared. Surprisingly, for important subclasses of conditional volatility models the ranking of the two methods, in term of asymptotic efficiency, depends on α and on simple characteristics of the law of η_t, but not on the volatility parameter θ_0.

Most of previous work on statistical inference for GARCH-type models dealt exclusively with the estimation of volatility parameters. The asymptotic theory of the QML estimation for volatility parameters has been extensively studied, in particular for the GARCH(1,1) by Lee and Hansen (1994), Lumsdaine (1996), for the GARCH(p, q) by Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004), for general models by Mikosch and Straumann (2006), Straumann and Mikosch (2006), Bardet and Wintenberger (2009). For the VaR parameter, it turns out that the QML criterion can be written under the form of a M-estimation criterion which is similar to those introduced in the quantile regression literature (see Koenker (2005) for a comprehensive book on quantile regression, and see Xiao and Koenker (2009), Xiao and Wan (2010) for recent applications to linear GARCH models) and in the least-absolute deviations (LAD) time series literature (see Davis, Knight and Liu (1992), Davis and Dunsmuir (1997), Breidt, Davis and Trindade (2001), Ling (2005)).

The paper is organized as follows. In Section 2 we introduce the concept of risk parameter in a general conditional volatility model, and we discuss identifiability issues. Section 3 is devoted to the asymptotic properties of non-Gaussian QML estimators for general smooth risk measures r. Section 4 is devoted to the estimation of the VaR parameter. The smoothness assumptions introduced in Section 3 being non satisfied by the VaR, the asymptotic properties of the one-step estimator are established in a completely different manner. The asymptotic properties of the two-step method are also established, and are compared with those of the one-step estimator. In Section 5, we consider two extensions. In particular, we consider estimation of the conditional Expected Shortfall. A Monte-Carlo study and applications on real financial data are provided in Section 6. Section 7 concludes. Proofs are collected in the Appendix.

2. Risk parameter in volatility models

Most conditional volatility models are of the form

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \end{cases} \tag{2.1}$$

where (η_t) is a sequence of iid random variables, η_t being independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$. When $E\eta_t = 0$ and $E\eta_t^2 = 1$, the variable σ_t^2 is generally referred to as the volatility of ϵ_t. However, we will not make such moment assumptions in this section and the following ones. A leading model, the most widely used among practitioners, is the GARCH(1,1) model of Engle (1982) and Bollerslev (1986), defined by

$$\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{2.2}$$

where $\theta_0 = (\omega_0, \alpha_0, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, 1)$. For this model we have $\sigma_t^2 = \sum_{i=1}^{\infty} \beta_0^{i-1} (\omega_0 + \alpha_0 \epsilon_{t-i}^2)$, which is of the form (2.1).

It is assumed throughout that

A0: There exists a function H such that for any $\theta \in \Theta$, for any $K > 0$, and any sequence $(x_i)_i$

$$K\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; \theta^*), \quad \text{where } \theta^* = H(\theta, K).$$

Most conditional volatility models are such that for $K \geq 1$, $\theta^* \geq \theta$ componentwise. For instance, in the GARCH(1,1) case we have $\theta^* = (K^2\omega, K^2\alpha, \beta)'$ with standard notation. The parameter θ_0 can thus be interpreted as a *volatility parameter* in the sense that the larger θ_0 the larger the volatility.

Now we define the notion of *risk parameter*. Following the terminology of Artzner, Delbaen, Eber, and Heath (1999), let r denote a risk measure, that is, a mapping from the set of the real random variables to $\mathbb{R}$. Assume that r is nonnegative, positively homogenous[2] and law-invariant [3]. Then the risk of ϵ_t conditional on $\{\epsilon_u, u < t\}$ is given by

$$r_{t-1}(\epsilon_t) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) r(\eta_t). \tag{2.3}$$

[2] that is, $r(\lambda X) = \lambda r(X)$ for any risk variable X and any $\lambda > 0$.

[3] that is, the risk $r(X)$ of any risk variable X only depends on the distribution of X.

Now, assuming $r(\eta_t) \neq 0$, let $\eta_t^* = \eta_t / r(\eta_t)$ and let $\theta_0^* = H(\theta_0, r(\eta_t))$. Under **A0**, the model can be reparameterized as

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, & r(\eta_t^*) = 1, \\ \sigma_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0^*). \end{cases} \tag{2.4}$$

Because the conditional risk of ϵ_t is now simply

$$r_{t-1}(\epsilon_t) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0^*),$$

θ_0^* will be called the risk parameter.

Example 2.1 (**VaR parameter**)**.** An important example is the VaR, which is the most standard risk measure used in the current regulations. For a continuous risk variable X with quantile function F_X^{-1}, the VaR at level α, with $\alpha \in (0, 1)$, is given by $r(X) = -F_X^{-1}(\alpha)$. The *conditional* VaR of the process (ϵ_t) at risk level $\alpha \in (0, 1)$, denoted by $\mathrm{VaR}_t(\alpha)$, is defined by

$$P_{t-1}[\epsilon_t < -\mathrm{VaR}_t(\alpha)] = \alpha,$$

where P_{t-1} denotes the historical distribution conditional on $\{\epsilon_u, u < t\}$. When (ϵ_t) satisfies (2.1), the theoretical VaR is then given by

$$\mathrm{VaR}_t(\alpha) = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) F_\eta^{-1}(\alpha)$$

where F_η is the probability distribution function of η_t. Let α be small enough so that $F_\eta^{-1}(\alpha) < 0$. Thus (2.3) holds with $r_{t-1}(\epsilon_t) = \mathrm{VaR}_t(\alpha)$ and $r(\eta_t) = -F_\eta^{-1}(\alpha)$. Now suppose that the volatility model is the GARCH(1,1) model (2.2). Then the VaR parameter at level α is given by $\theta_0^* = (K^2\omega_0, K^2\alpha_0, \beta_0)'$ with $K = -F_\eta^{-1}(\alpha)$. This coefficient takes into account the dynamics of the GARCH process through the volatility parameters, but also the lower tail of the innovations distribution.

Example 2.2 (**Expected Shortfall parameter**)**.** Another popular measure of financial risk is the expected shortfall (ES). One reason for its attractiveness is that, in contrast to the VaR, the ES satisfies the sub-additivity property (see Acerbi and Tasche (2002)). For a continuous risk variable X such that $E(X^-) < \infty$, the ES at level $\alpha \in (0, 1)$ is given by $r(X) = -E[X \mid X \leq F_X^{-1}(\alpha)]$. The *conditional* ES of the process (ϵ_t) at risk level α, denoted by $\mathrm{ES}_t(\alpha)$, is defined by

$$\mathrm{ES}_t(\alpha) = -E_{t-1}[\epsilon_t \mid \epsilon_t < -\mathrm{VaR}_t(\alpha)],$$

TABLE 1
VaR and ES parameters at the 1% level for GARCH(1,1) models

Errors distribution	$\eta_t \sim \mathcal{N}(0,1)$	$\eta_t \sim \frac{1}{\sqrt{2}}St_4$
Volatility parameter	$(1, 0.05, 0.9)$	$(1, 0.04, 0.9)$
VaR parameter	$(5.41, 0.27, 0.9)$	$(7.01, 0.28, 0.9)$
ES parameter	$(7.10, 0.36, 0.9)$	$(13.63, 0.55, 0.9)$

where E_{t-1} denotes the expectation conditional on $\{\epsilon_u, u < t\}$. When (ϵ_t) satisfies (2.1), the theoretical ES is then given by

$$\mathrm{ES}_t(\alpha) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0)\mathrm{ES}_\eta(\alpha), \tag{2.5}$$

where $\mathrm{ES}_\eta(\alpha)$ is the ES at level α of η_t, which is of the form (2.3). For the GARCH(1,1) model (2.2), the ES parameter at level α is $\theta_0^* = (K^2\omega_0, K^2\alpha_0, \beta_0)'$ with $K = \mathrm{ES}_\eta(\alpha)$.

Example 2.3 (**VaR and ES parameters for two GARCH(1,1)**)**.** For the sake of illustration, consider two GARCH(1,1) models with, respectively, standard Gaussian and standardized Student(4) innovations. The volatility parameter, as displayed in Table 1, is larger for the Gaussian-innovation model than for the Student-innovation model. In contrast, the VaR parameter at level 1% is slightly larger for the second model. In other words, the first model is more volatile but less risky than the second one for the VaR at 1%. The difference between the two models is even more pronounced when ES-parameters at the level 1% are considered. In particular, the coefficient α_0^*, measuring the impact of a large squared return on the risk of the next period, is 1.5 larger in the model with student errors than in the conditionally Gaussian model.

We consider estimating θ_0^* by an appropriate QML method in the next section.

3. QML estimators of general risk parameters

In this section, we consider QML estimation of Model (2.4). The usual identifiability condition $E\eta_t^{*2} = 1$ being replaced by $r(\eta_t^*) = 1$, we will define a non Gaussian QML estimator.

Given observations $\epsilon_1, \dots, \epsilon_n$, and arbitrary initial values $\tilde{\epsilon}_i$ for $i \leq 0$, we define, under assumptions given below

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

This random variable will be used to approximate

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

We choose an arbitrary positive density h which can be called *instrumental* density, and define the QML criterion

$$\tilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \tilde{\sigma}_t(\theta)), \qquad g(x,\sigma) = \log \frac{1}{\sigma} h\left(\frac{x}{\sigma}\right). \tag{3.1}$$

Let the QMLE

$$\hat{\theta}_n^* = \arg\max_{\theta\in\Theta} \tilde{Q}_n(\theta) \tag{3.2}$$

for some compact subspace Θ of $\mathbb{R}^m$. This estimator is the standard Gaussian QMLE when h is the standard Gaussian density ϕ. However, the Gaussian QMLE is in general an inconsistent estimator of θ_0^*, unless if, for instance, the risk measure is $r(X) = \sqrt{E(X^2)}$.

To derive the asymptotic properties of $\hat{\theta}_n^*$ we introduce the following assumptions.

A1: (ϵ_t) is a strictly stationary and ergodic solution of Model (2.4).

A2: For any real sequence (x_i), the function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ is continuous. Almost surely, $\sigma_t(\theta) \in (\underline{\omega}, \infty]$ for any $\theta \in \Theta$ and for some $\underline{\omega} > 0$. Moreover, $\sigma_t(\theta_0^*)/\sigma_t(\theta) = 1$ *a.s.* iff $\theta = \theta_0^*$.

In addition, we assume that the function $\sigma \to Eg(\eta_0^*, \sigma)$ is valued in $[-\infty, +\infty)$ and has a unique maximum at 1:

A3: $Eg(\eta_0^*, \sigma) < Eg(\eta_0^*, 1), \qquad \forall\sigma > 0, \quad \sigma \neq 1.$

A4: h is continuous on $\mathbb{R}$, differentiable except on a finite set A, and there exist constants $\delta \geq 0$ and $C_0 > 0$ such that for all $u \in A^c$, $|uh'(u)/h(u)| \leq C_0(1 + |u|^\delta)$ with $E|\eta_0^*|^\delta < \infty$. Moreover, $E|\epsilon_0|^s < \infty$ for some $s > 0$.

A5: There exist a random variable C_1 measurable with respect to $\{\epsilon_u, u < 0\}$ and a constant $\rho \in (0,1)$ such that $\sup_{\theta\in\Theta} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \leq C_1\rho^t$.

Theorem 3.1 **(Consistency of the risk parameter estimator).** *If* **A0-A5** *hold, then the QMLE defined by (3.1) and (3.2) satisfies*

$$\hat{\theta}_n^* \to \theta_0^*, \quad a.s.$$

The condition on h in **A4** is mild; it vanishes for instance when the instrumental density has the form $h(u) = K_1|u|^\lambda \exp\{K_2|u|^r\}$, for some constants r, λ, K_1, K_2. In this case, the inequality is satisfied with $\delta = r$. Assumptions **A2** and **A5** can be simplified for specific forms of σ_t: for instance if the model is a standard GARCH, **A2** reduces to standard assumptions on the lag polynomials of the volatility and **A5** can be directly verified. Note also that the only moment assumption on the observed process is the existence of a small moment in **A4**, which is automatically satisfied under **A1** in standard GARCH models.

We now discuss Assumption **A3**. Many risks measures involve a moment of a function of the risk variable X, in the sense that

$$r(X) = 1 \quad \text{iff} \quad E\{\psi(X)\} = 0 \tag{3.3}$$

for some measurable function $\psi : \mathbb{R} \to \mathbb{R}$. The following result shows that, for such risk measures, **A3** can be omitted provided the QML instrumental density is appropriately chosen.

Proposition 3.1 (Choice of the instrumental density). *Let $r(\cdot)$ satisfying (3.3). Assume* **A4** *holds with $A = \emptyset$. Then* **A3** *holds for any distribution of η_0^* satisfying $r(\eta_0^*) = 1$ iff the density h is such that*

$$x\{\log h(x)\}' = [\lambda\psi(x) - 1], \quad \textit{for all } x \tag{3.4}$$

and for some constant $\lambda \neq 0$.

This result provides a practical way to choose the QML density h, as illustrated in the next examples.

Example 3.1. Let $r(X) = \sqrt{E(X^2)}$. Then we have $\psi(X) = 1 - X^2$ and, for any constant $\lambda > 0$, by solving (3.4) we find

$$h(x) \propto |x|^{-(1-\lambda)} \exp(-\lambda x^2/2). \tag{3.5}$$

For $\lambda = 1$ the assumption **A4** is satisfied and h is the density of the standard Gaussian distribution. Thus, we retrieve the Gaussian likelihood for the standard identifiability condition $E(\eta_t^{*2}) = 1$. It can be seen that any density h of the form (3.5) provides the same QMLE, it is thus not restrictive to take $\lambda = 1$.

Example 3.2. More generally, let $r(X) = \|X\|_s = (E|X|^s)^{1/s}$ where s is a positive number. This risk measure has interest, for $s < 2$, when the variable

X has an infinite second-order moment. In this case $\psi(X) = 1 - |X|^s$ and we find , for $\lambda > 0$ and for some constant c,

$$h(x) \propto |x|^{-(1-\lambda)} \exp(-\lambda|x|^s/s). \tag{3.6}$$

For $\lambda = 1$, the assumption **A4** is satisfied.

Example 3.3 (Example 2.1 continued). When the measure of risk is the VaR at level α, we have $r(X) = -F_X^{-1}(\alpha)$. Suppose that the distribution of the risk variable X is symmetric and $\alpha \in (0, 0.5)$. Thus (3.3) is satisfied with $\psi(X) = \mathbf{1}_{\{|X|>1\}} - 2\alpha$. Solving (3.4) yields

$$h(x) = h_\alpha(x) = \lambda\alpha(1-2\alpha)|x|^{2\lambda\alpha-1}\{|x|^{-\lambda}\mathbf{1}_{\{|x|>1\}} + \mathbf{1}_{\{|x|\le 1\}}\} \tag{3.7}$$

for some positive constant λ. The choice of λ does not matter, any value leading to the same $\hat{\theta}_n^*$. By choosing $\lambda = (2\alpha)^{-1}$, the density is defined on $\mathbb{R}$ and satisfies **A4**. Since h is not differentiable everywhere, the assumptions of Proposition 3.1 are not satisfied. However, it will be shown in the next section that, under mild additional assumptions on the distribution of η_0, Assumptions **A3** and **A4** of Theorem 3.1 hold true.

To show the asymptotic normality of $\hat{\theta}_n^*$ we need additional assumptions which, for the reader's convenience, are deferred to the appendix (see **A6-A10** in Appendix A.1). Note that, for most classical GARCH formulations, **A7** reduces to standard assumptions on lag polynomials and that **A8** and **A10** can be directly verified. Let $g_1(x,\sigma) = \partial g(x,\sigma)/\partial\sigma$, $g_2(x,\sigma) = \partial g_1(x,\sigma)/\partial\sigma$.

Theorem 3.2 (Asymptotic normality). *Under* **A0-A10** *and if* $Eg_2(\eta_0^*, 1) \neq 0$ *then*

$$\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\tau_{h,f}^2 I^{-1}),$$

where

$$I = I(\theta_0^*) = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0^*)\right) \quad \textit{and} \quad \tau_{h,f}^2 = \frac{Eg_1^2(\eta_0^*, 1)}{\{Eg_2(\eta_0^*, 1)\}^2}.$$

Theorem 3.2 does not apply to the VaR because the differentiability assumption in **A9** is not satisfied for densities h of the form (3.7).

4. Estimating the conditional VaR

4.1. One-step VaR estimation

To estimate in one step the conditional VaR at level α, as defined in Example 2.1, we first need to reparameterize Model (2.1). If α is not too large (more precisely $\alpha < P(\eta_0 > 0)$), from $P[\eta_t < F^{-1}(\alpha)] = \alpha$ we deduce $P[\eta_t^* < -1] = \alpha$ where $\eta_t^* = -\eta_t/F^{-1}(\alpha)$. Letting $\theta_{0,\alpha} = \theta_0^* = H(\theta_0, -F^{-1}(\alpha))$, under **A0**, the model can be reparameterized as

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, \quad P[\eta_t^* < -1] = \alpha, \\ \sigma_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_{0,\alpha}). \end{cases} \tag{4.1}$$

The theoretical VaR is now given by

$$\text{VaR}_t(\alpha) = \sigma_t(\theta_{0,\alpha}). \tag{4.2}$$

We will thus call $\theta_{0,\alpha}$ the *VaR parameter* at level α.

Define a QMLE of $\theta_{0,\alpha}$ by

$$\hat{\theta}_{n,\alpha} = \arg\max_{\theta \in \Theta} \sum_{t=1}^{n} \log \frac{1}{\tilde{\sigma}_t(\theta)} h_\alpha \left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right) \tag{4.3}$$

where h_α is defined by (3.7). The following corollary of Theorem 3.1 shows that a one-step consistent estimator of the VaR parameter, not requiring any estimation of the quantile function of the innovations η_t, is given by

$$\widehat{\text{VaR}}_t(\alpha) = \tilde{\sigma}_t(\hat{\theta}_{n,\alpha}). \tag{4.4}$$

Corollary 4.1 (Consistency of the VaR parameter estimator). *Let (ϵ_t) be a strictly stationary and ergodic solution of (4.1), where the distribution of η_0^* is symmetric, satisfies the moment condition $E|\log|\eta_0^*|| < \infty$, and admits a density in a neighborhood of 1. If* **A0**, **A2** *and* **A5** *hold and if $E|\epsilon_0|^s < \infty$ for some $s > 0$, then, for all $\alpha \in (0, 1/2)$, the QMLE defined by (4.3) satisfies*

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

The non-Gaussian QML estimator of the VaR parameter is also related to estimators introduced in the quantile regression literature (see references in the introduction). Let $\rho_\alpha(u) = u(\alpha - \mathbf{1}_{\{u \le 0\}})$. Then, from the proof of

Corollary 4.1 it can be seen that

$$\begin{aligned}\hat{\theta}_{n,\alpha} &= \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \log\left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)}\right) \mathbf{1}_{\{|\epsilon_t|>\tilde{\sigma}_t(\theta)\}} - 2\alpha \log\left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)}\right) \\ &= \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \rho_{1-2\alpha}\left\{\log\left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)}\right)\right\}. \qquad (4.5)\end{aligned}$$

To interpret this expression, note that the first equation in Model (4.1) can be equivalently written as

$$\log|\epsilon_t| = \log\sigma_t^* + \log|\eta_t^*|, \qquad P[\log|\eta_0^*| < 0] = 1-2\alpha \qquad (4.6)$$

under the assumption of a symmetric distribution for η_0^*. Model (4.6) resembling a quantile regression model, it is not surprising to get an estimator of the form (4.5). An important difference with the quantile regression or autoregression, however, is that $\tilde{\sigma}_t(\theta)$ is not assumed to be a linear combination of explanatory variables, or past observables.

To study the asymptotic distribution of $\hat{\theta}_{n,\alpha}$ we need the following additional assumption.

A11: The density f^* of η_0^* is continuous at 1 and satisfies $f^*(1) > 0$ and $M = \sup_{x\in\mathbb{R}} |x|f^*(x) < \infty$.

Theorem 4.1 (Asymptotic normality). *Under the assumptions of Corollary 4.1,* **A6-A8** *and* **A10-A11**, *there exists a sequence of local minimizers $\hat{\theta}_{n,\alpha}$ of the criterion defined in (4.5) satisfying*

$$\sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha}) \xrightarrow{d} \mathcal{N}\left(0, \Xi_\alpha := \frac{2\alpha(1-2\alpha)}{4f^{*2}(1)} J_\alpha^{-1}\right),$$

where $J_\alpha = ED_t(\theta_{0,\alpha})D_t'(\theta_{0,\alpha})$ and $D_t(\theta) = \sigma_t^{-1}(\theta)\partial\sigma_t(\theta)/\partial\theta$.

Let $\widehat{\Xi}_\alpha$ denote a consistent estimator of the asymptotic variance Ξ_α. The delta method thus suggests a $(1-\alpha_0)\%$ confidence interval for $\mathrm{VaR}_t(\alpha)$ whose bounds are

$$\tilde{\sigma}_t(\hat{\theta}_{n,\alpha}) \pm \frac{\Phi^{-1}_{1-\alpha_0/2}}{\sqrt{n}} \left\{\frac{\partial\tilde{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial\theta'} \widehat{\Xi}_\alpha \frac{\partial\tilde{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial\theta}\right\}^{1/2}, \qquad (4.7)$$

where $\Phi^{-1}_{\alpha_0}$ denotes the α_0-quantile of the standard Gaussian distribution. Drawing such confidence intervals allows to underline that the VaR evaluation is subject to estimation risk. Even when the model is correctly specified, the market risk, as measured by the theoretical VaR defined by (4.2), is not known with exactness, but is likely to belong to the confidence interval (4.7).

4.2. Two-step VaR estimation

In this section, we consider the usual approach for estimating the VaR in Model (2.1) under the identifiability condition

$$E\eta_t^2 = 1. \tag{4.8}$$

This approach involves two steps. In a first step, the model is estimated by the standard QMLE and, in a second step, the theoretical quantile $\xi_\alpha := F_\eta^{-1}(\alpha)$ is estimated using the estimated rescaled innovations. More precisely, let $\hat{\theta}_n$ denote the Gaussian QMLE of θ_0 in Model (2.1) under the constraint (4.8), let

$$\hat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t(\hat{\theta}_n)},$$

and let $\xi_{n,\alpha}$ denote the empirical α-quantile of $\hat{\eta}_1, \dots, \hat{\eta}_n$.

An estimator of the VaR at level α is then given by

$$\widetilde{\mathrm{VaR}}_t(\alpha) = -\tilde{\sigma}_t(\hat{\theta}_n)\xi_{n,\alpha} = \tilde{\sigma}_t\{H(\hat{\theta}_n, -\xi_{n,\alpha})\}$$

under **A0** and provided $-\xi_{n,\alpha} > 0$. A comparison of the VaR estimators $\widetilde{\mathrm{VaR}}_t(\alpha)$ and $\widehat{\mathrm{VaR}}_t(\alpha)$ defined in (4.4) can then be based on the asymptotic accuracies of the estimators $\hat{\theta}_{n,\alpha}$ and $H(\hat{\theta}_n, -\xi_{n,\alpha})$ of $\theta_{0,\alpha}$.

Contrary to the one-step estimator, the resulting two step estimator of the VaR does not take advantage of the (hypothesized) symmetry of the errors distribution. An estimator exploiting this additional information is

$$\widetilde{\widetilde{\mathrm{VaR}}}_t(\alpha) = \tilde{\sigma}_t(\hat{\theta}_n)\tilde{\xi}_{n,1-2\alpha} = \tilde{\sigma}_t\{H(\hat{\theta}_n, \tilde{\xi}_{n,1-2\alpha})\}$$

where $\tilde{\xi}_{n,1-2\alpha}$ is the empirical $(1-2\alpha)$-quantile of $|\hat{\eta}_1|, \dots, |\hat{\eta}_n|$.

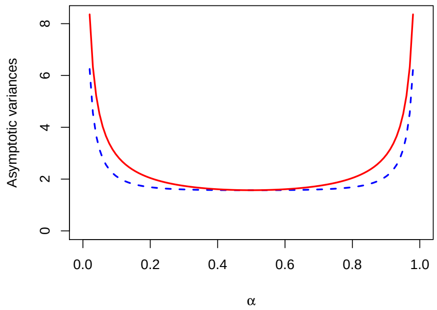
The next result gives the joint asymptotic distributions of $(\hat{\theta}_n', -\xi_{n,\alpha})$ and $(\hat{\theta}_n', \tilde{\xi}_{n,1-2\alpha})$.

Theorem 4.2. *Assume $\xi_\alpha < 0$, $E\eta_t^2 = 1$ and $\kappa_4 := E\eta_t^4 < \infty$. Suppose that η_1 admits a density f in a neighborhood of ξ_α. Let* **A1**, **A5**, **A8** *hold. Let* **A2**, **A6**, **A7** *and* **A10** *hold with $\delta = 2$ and θ_0^* replaced by θ_0. Then*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \\ \sqrt{n}(\xi_\alpha - \xi_{n,\alpha}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_\alpha), \qquad \Sigma_\alpha = \begin{pmatrix} \frac{\kappa_4 - 1}{4} J^{-1} & \lambda_\alpha J^{-1}\Omega \\ \lambda_\alpha \Omega' J^{-1} & \zeta_\alpha \end{pmatrix},$$

where $\Omega = E(D_t)$, $J = E(D_t D_t')$ with $D_t = D_t(\theta_0)$, and

$$\lambda_\alpha = \xi_\alpha \frac{\kappa_4 - 1}{4} + \frac{p_\alpha}{2f(\xi_\alpha)}, \qquad \zeta_\alpha = \xi_\alpha^2 \frac{\kappa_4 - 1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}.$$



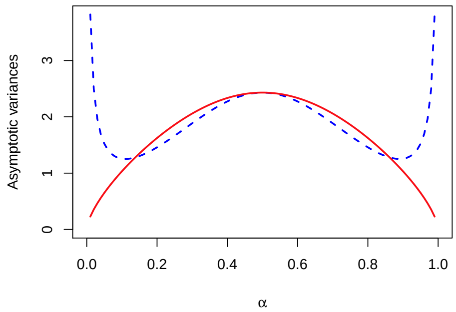


FIG 1. *Asymptotic variances ζ_α in dotted lines, and $\alpha(1-\alpha)/f^2(\xi_\alpha)$ in full line, for a standard Gaussian distribution (left panel) and the standardized GED(ν) with $\nu = 0.25$ (right panel).*

with $p_\alpha = E\left(\eta_1^2 \mathbf{1}_{\{\eta_1<\xi_\alpha\}}\right) - \alpha$.

Under the additional assumption that η_1 is symmetrically distributed, we have

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \tilde{\xi}_{n,1-2\alpha} + \xi_\alpha \end{pmatrix} \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, \tilde{\Sigma}_\alpha), \quad \tilde{\Sigma}_\alpha = \begin{pmatrix} \frac{\kappa_4-1}{4}J^{-1} & -\lambda_\alpha J^{-1}\Omega \\ -\lambda_\alpha \Omega' J^{-1} & \tilde{\zeta}_\alpha \end{pmatrix}$$

where

$$\tilde{\zeta}_\alpha = \xi_\alpha^2 \frac{\kappa_4-1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{2\alpha(1-2\alpha)}{4f^2(\xi_\alpha)} = \zeta_\alpha - \frac{\alpha}{2f^2(\xi_\alpha)}.$$

Remark: The asymptotic variance ζ_α of the empirical quantile of the standardized residuals $\hat{\eta}_t$ is the sum of $\alpha(1-\alpha)/f^2(\xi_\alpha)$, which can be interpreted as the asymptotic variance of the empirical α-quantile of the η_t's, and a term due to the estimation of $\theta_{0,\alpha}$. The same additional term appears in $\tilde{\zeta}_\alpha$. Unexpectedly, this term measuring the effect of estimation can be negative. For instance, for a standard Gaussian distribution we have

$$\xi_\alpha^2 \frac{\kappa_4-1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} = \frac{-1}{2}\xi_\alpha^2 \leq 0.$$

This is illustrated in the left panel of Figure 1. For fat tailed distribution, on the contrary, this term can be positive and arbitrarily large. For thin tailed distributions, the comparison can depend on the value of α. This is illustrated in the right panel of Figure 1 for a standardized GED (ν) (Generalized Error Distribution of parameter ν) of density $f(x) \propto \exp\{-0.5|x|^{1/\nu}\}$.

We now deduce the asymptotic distributions of the *two-step* and the *symmetric two-step* estimators $\hat{\theta}^{2S}_{n,\alpha} := H(\hat{\theta}_n, -\xi_{n,\alpha})$ and $\hat{\theta}^{S2S}_{n,\alpha} := H(\hat{\theta}_n, \tilde{\xi}_{n,1-2\alpha})$ of $\theta_{0,\alpha} = H(\theta_0, -\xi_\alpha)$.

Corollary 4.2. *Under the assumptions of Theorem 4.2, the two-step estimators of the VaR-parameter at level α satisfy*

$$\sqrt{n}\left(\hat{\theta}^{2S}_{n,\alpha} - \theta_{0,\alpha}\right) \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, \Upsilon_\alpha), \qquad \sqrt{n}\left(\hat{\theta}^{S2S}_{n,\alpha} - \theta_{0,\alpha}\right) \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, \tilde{\Upsilon}_\alpha),$$

where

$$\Upsilon_\alpha = \left[\frac{\partial H(\theta,\xi)}{\partial(\theta',\xi)}\right]_{(\theta_0,-\xi_\alpha)} \Sigma_\alpha \left[\frac{\partial H(\theta,\xi)}{\partial(\theta',\xi)'}\right]_{(\theta_0,-\xi_\alpha)},$$

and $\tilde{\Upsilon}_\alpha$ is obtained by replacing Σ_α by $\tilde{\Sigma}_\alpha$ in Υ_α.

It can be noted that, because the matrices $\tilde{\Sigma}_\alpha$ and Σ_α only differ by their lower-right term, with $\xi_\alpha > \tilde{\xi}_\alpha$, we have, in the sense of positive definite matrices,

$$\tilde{\Upsilon}_\alpha \preceq \Upsilon_\alpha.$$

Thus, the symmetric two-step estimator is asymptotically more accurate than the two-step estimator. But the former estimator is inconsistent if the errors distribution is not symmetric.

4.3. Comparing the one-step and two-step estimators in the standard GARCH case

The results of Theorem 3.1 are now applied to the standard GARCH(p,q) model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t, \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^2, \end{cases} \tag{4.9}$$

where $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \beta_{0p})'$ satisfies $\omega_0 > 0, \alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. Let

$$\overline{\theta}_0 = \begin{pmatrix} \theta_0^{[1:q+1]} \\ 0_p \end{pmatrix}, \quad \theta_0^{[1:q+1]} = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q})', \quad A = \begin{pmatrix} \xi_\alpha^2 I_{q+1} & 0 \\ 0 & I_p \end{pmatrix}.$$

Corollary 4.3. *Under the assumptions of Corollary 4.2, for the standard GARCH model (4.9) the asymptotic variances of the two-step estimators of the VaR parameter take the form*

$$\begin{aligned} \Upsilon_\alpha &= \frac{\kappa_4 - 1}{4} A\{J^{-1} - 4\overline{\theta}_0\overline{\theta}_0'\}A + 4\xi_\alpha^2 \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}\overline{\theta}_0\overline{\theta}_0', \\ \tilde{\Upsilon}_\alpha &= \frac{\kappa_4 - 1}{4} A\{J^{-1} - 4\overline{\theta}_0\overline{\theta}_0'\}A + \xi_\alpha^2 \frac{2\alpha(1-2\alpha)}{f^2(\xi_\alpha)}\overline{\theta}_0\overline{\theta}_0'. \end{aligned}$$

To compare the asymptotic variances of the two estimators of $\theta_{0,\alpha}$ note that

$$J_\alpha^{-1} = AJ^{-1}A, \quad f_{\eta^*}(1) = -\xi_\alpha f(\xi_\alpha).$$

Hence

$$\mathrm{Var}_{as}\{\sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha})\} = \frac{2\alpha(1-2\alpha)}{4\xi_\alpha^2 f^2(\xi_\alpha)} AJ^{-1}A.$$

It follows that

$$\mathrm{Var}_{as}\{\sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha})\} - \Upsilon_\alpha \;=\; \Delta_\alpha A\left\{\frac{J^{-1}}{4} - \overline{\theta}_0\overline{\theta}_0'\right\}A - \frac{2\alpha}{\xi_\alpha^2 f^2(\xi_\alpha)} A\overline{\theta}_0\overline{\theta}_0'A',$$

$$\mathrm{Var}_{as}\{\sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha})\} - \tilde{\Upsilon}_\alpha \;=\; \Delta_\alpha A\left\{\frac{J^{-1}}{4} - \overline{\theta}_0\overline{\theta}_0'\right\}A,$$

where

$$\Delta_\alpha = \frac{2\alpha(1-2\alpha)}{\xi_\alpha^2 f^2(\xi_\alpha)} - (\kappa_4 - 1). \tag{4.10}$$

The following result is a consequence of the fact that $4\overline{\theta}_0\overline{\theta}_0' \preceq J^{-1}$. It allows to compare the asymptotic variances of the one-step estimator $\hat{\theta}_{n,\alpha}$ and two-step estimator $\hat{\theta}_{n,\alpha}^{S2S}$.

Corollary 4.4. *Under the assumptions of Corollary 4.2, for the standard GARCH model (4.9) with symmetric innovations, we have*

$$Var_{as}\{\sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha})\} \preceq Var_{as}\left\{\sqrt{n}\left(\hat{\theta}_{n,\alpha}^{S2S} - \theta_{0,\alpha}\right)\right\} \quad \textit{iff} \quad \Delta_\alpha \leq 0.$$

Interestingly, comparing the asymptotic variance matrices of the estimators amounts to determining the sign of a real coefficient, which solely depends on the distribution of η_t. None of the methods is superior in every situation. If the fourth-order moment is large, *a fortiori* if it does not exist, the one-step estimator will be better. Conversely, for distributions admitting moments at any order (such as the Gaussian) the two-step estimator may be superior. Figure 2 shows the surface $\Delta_\alpha \leq 0$, as a function of α and ν, for Student distributions with ν degrees of freedom. It can be seen that for small and moderate values of ν, the one-step estimator is asymptotically more efficient than the two-step estimator for the values of α which are used in practice. In Figure 3, Δ_α is drawn as a function of ν, for GED(ν) and $\alpha \in \{0.01, 0.05\}$. The comparison is in favor of the one-step estimator for $\nu < 0.15$, and also for sufficiently large values of ν.

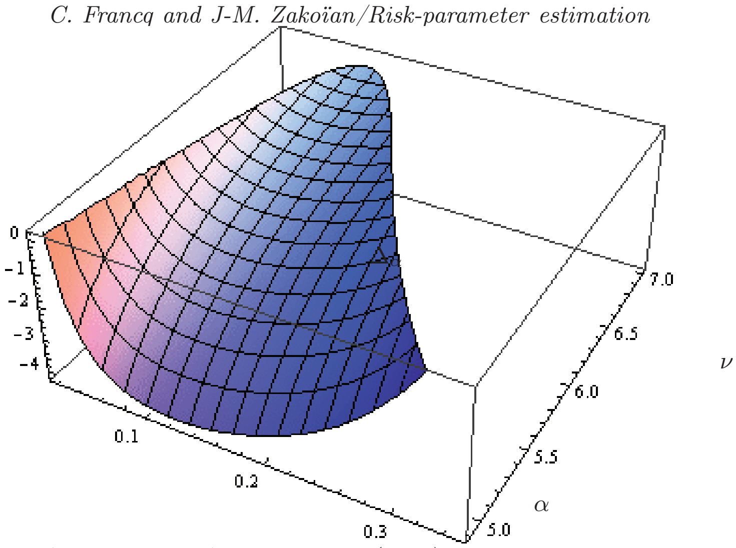


FIG 2. *Surface* $\Delta_\alpha \leq 0$, *with* Δ_α *defined in (4.10), for which the one-step estimator is asymptotically more efficient than the symmetric two-step estimator of* $\theta_{0,\alpha}$ *when the distribution of* η_t *is a standardized Student with* ν *degrees of freedom,* $\nu \in [4.9, 7]$ *and* $\alpha \in [0.01, 0.35]$.

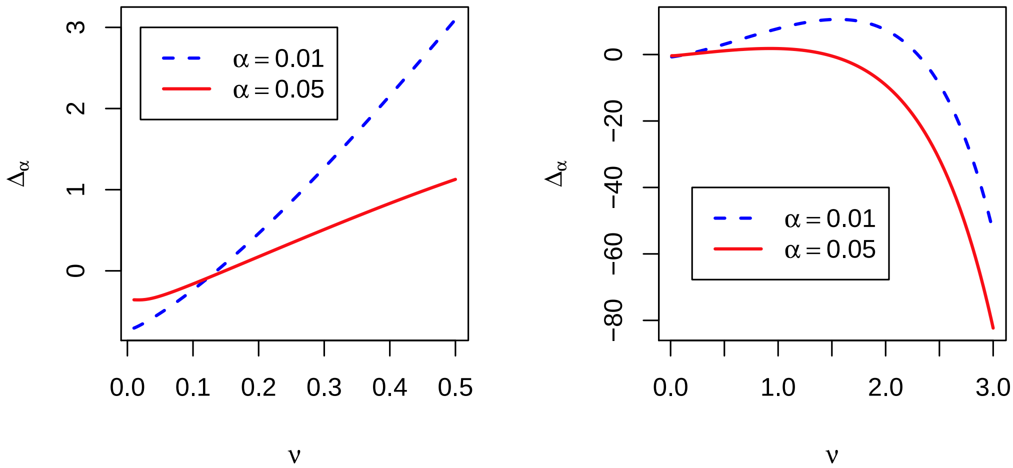


FIG 3. Δ_α, *defined in (4.10), for* $\alpha \in \{0.01, 0.05\}$, *when the distribution of* η_t *follows a GED with* ν *degrees of freedom.*

5. Extensions

5.1. One-step estimation of the VaR parameter without the symmetry assumption

We have seen that, under the assumption that the distribution of η_0^* is symmetric and under some regularity assumptions, the VaR parameter $\theta_{0,\alpha}$ defined by (4.2) is consistently estimated by a QMLE, if and only if the instrumental density h is that defined by (3.6). This estimator has the form of the quantile estimator $\hat{\theta}_{n,\alpha}$ defined by (4.5) in the nonlinear regression model (4.6), but may be inconsistent when the distribution of η_0^* is asymmetric.

It is however possible to define a one-step quantile estimator of the VaR parameter without assuming that η_0^* is symmetric. To this aim, note that similarly to (4.6) we have for $\epsilon_t < 0$

$$\log \epsilon_t^- = \log \sigma_t^* + \log \eta_t^{*-}, \qquad P[\log \eta_t^{*-} < 0 \mid \epsilon_t < 0] = \tau_0$$

where $x^- = \max\{0, -x\}$ and $\tau_0 = 1 - \{\alpha / P(\eta_t^* < 0)\}$. This leads us to consider the quantile estimator

$$\breve{\theta}_{n,\alpha} = \arg\min_{\theta \in \Theta} \sum_{t\,:\epsilon_t < 0} \rho_{\hat{\tau}} \left\{ \log \left(\frac{\epsilon_t^-}{\tilde{\sigma}_t(\theta)} \right) \right\}, \quad \hat{\tau} = 1 - \left(\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\epsilon_t < 0\}} \right)^{-1} \alpha.$$

It is known (see Lemma 2.3 in Berkes, Horváth and Kokoszka, 2003) that any strictly stationarity GARCH model possesses a fractional moment of order $s \in (0, 1)$. It is then easy to check that for these models the following assumption holds true.

A13: $E \sup_{\theta \in \Theta} |\log \sigma_1(\theta)| < \infty$.

Theorem 5.1. *Let (ϵ_t) be a strictly stationary and ergodic solution of (4.1), where the distribution of η_0^* satisfies $E \log^+ \eta_0^{*-} < \infty$, and admits a density in a neighborhood of -1. If* **A0**, **A2**, **A5** *and* **A13** *hold, then, for all $\alpha \in (0, P(\eta_0^* < 0))$, we have*

$$\breve{\theta}_{n,\alpha} \to \theta_{0,\alpha}, \quad a.s.$$

where $\theta_{0,\alpha}$ is the VaR parameter satisfying (4.2).

The asymptotic distribution of this estimator is left for future research.

5.2. Two-step estimation of the ES parameter

Computing the ES involves two-steps, VaR computation in a first step followed by the computation of a conditional expectation in a second step. For this reason, ES estimation is not amenable to the one-step QML estimation method developed in Section 3. However, the two step approach can be derived as follows.

Let $\mu_\alpha = -E(\eta_0 \mid \eta_0 < \xi_\alpha)$ denote the ES of the distribution of η_0. By (2.5) and **A0**, the theoretical ES is given by

$$\mathrm{ES}_t(\alpha) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0)\mathrm{ES}_\eta(\alpha) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta^*_{0,\alpha}), \qquad (5.1)$$

where $\theta^*_{0,\alpha} = H(\theta_0, \mu_\alpha)$ is the ES parameter at level α.

An estimator of $\theta^*_{0,\alpha}$ is then given by

$$\widetilde{\mathrm{ES}}_t(\alpha) = \tilde{\sigma}_t(\hat{\theta}_n)\mu_{n,\alpha} = \tilde{\sigma}_t\{H(\hat{\theta}_n, \mu_{n,\alpha})\}$$

where $\mu_{n,\alpha}$ is the ES of the errors:

$$\mu_{n,\alpha} = -\frac{\sum_{t=1}^n \hat{\eta}_t \mathbb{1}_{\hat{\eta}_t < \xi_{n,\alpha}}}{\sum_{t=1}^n \mathbb{1}_{\hat{\eta}_t < \xi_{n,\alpha}}} = \frac{-1}{[n\alpha]+1} \sum_{t=1}^n \hat{\eta}_t \mathbb{1}_{\hat{\eta}_t < \xi_{n,\alpha}},$$

and $H(\hat{\theta}_n, \mu_{n,\alpha})$ is an estimator of the ES parameter.

The next result gives the joint asymptotic distribution of $(\hat{\theta}'_n, \mu_{n,\alpha})$, and the asymptotic distribution of $H(\hat{\theta}_n, \mu_{n,\alpha})$ in the standard GARCH case.

Theorem 5.2. *Assume $\mu_\alpha > 0$, $E\eta_t^2 = 1$ and $\kappa_4 := E\eta_t^4 < \infty$. Let* **A1**, **A5**, **A8** *hold. Let* **A2**, **A6**, **A7** *and* **A10** *hold with $\delta = 2$ and θ^*_0 replaced by θ_0. Then*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \\ \sqrt{n}(\mu_{n,\alpha} - \mu_\alpha) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma_\alpha), \qquad \Gamma_\alpha = \begin{pmatrix} \frac{\kappa_4 - 1}{4} J^{-1} & \varphi_\alpha J^{-1}\Omega \\ \varphi_\alpha \Omega' J^{-1} & \nu_\alpha \end{pmatrix},$$

where J is as in Theorem 4.2 and

$$\nu_\alpha = \sigma_\alpha^2 + \frac{\kappa_4 - 1}{4}\mu_\alpha^2 + \mu_\alpha x_\alpha, \qquad \varphi_\alpha = -\frac{1}{2}x_\alpha - \mu_\alpha \frac{\kappa_4 - 1}{4},$$

$$\sigma_\alpha^2 = \frac{1}{\alpha^2} var\left\{(\eta_t - \xi_\alpha)\mathbb{1}_{\eta_t < \xi_\alpha}\right\}, \qquad x_\alpha = \frac{1}{\alpha} cov\left\{1 - \eta_t^2, (\eta_t - \xi_\alpha)\mathbb{1}_{\eta_t < \xi_\alpha}\right\}.$$

It follows that the asymptotic distribution of the ES parameter is given by

$$\sqrt{n}(H(\hat{\theta}_n, \mu_{n,\alpha}) - \theta^*_{0,\alpha}) \xrightarrow{d} \mathcal{N}\left(0, \left[\frac{\partial H(\theta, \mu)}{\partial(\theta', \mu)}\right]_{(\theta_0, \mu_\alpha)} \Gamma_\alpha \left[\frac{\partial H(\theta, \mu)}{\partial(\theta', \mu)'}\right]_{(\theta_0, \mu_\alpha)}\right).$$

For the standard GARCH model (4.9), we have

$$\sqrt{n}(H(\hat{\theta}_n, \mu_{n,\alpha}) - \theta^*_{0,\alpha}) \xrightarrow{d} \mathcal{N}\left(0, \Upsilon^*_\alpha\right),$$

where

$$\Upsilon^*_\alpha = \frac{\kappa_4 - 1}{4} A^*\{J^{-1} - 4\overline{\theta}_0\overline{\theta}_0'\}A^* + 4\mu_\alpha^2\sigma_\alpha^2\overline{\theta}_0\overline{\theta}_0',$$

and A^* is obtained by replacing ξ_α by μ_α in A.

The asymptotic distribution of the ES-parameter estimator thus depends on the GARCH coefficients and simple characteristics of the innovations distribution. In contrast with the VaR parameter, the asymptotic variance does not involve the density of the errors distribution and thus can be more straightforwardly estimated. In the asymptotic variance ν_α of the errors ES estimator, the term σ_α^2 can be interpreted as the asymptotic variance of the ES estimator if the η_t's were observed (see Chen, 2008). The additional term, $\frac{\kappa_4-1}{4}\mu_\alpha^2 + \mu_\alpha x_\alpha$, thus reflects the effect of estimating the GARCH coefficients.

6. Numerical experiments

A Monte Carlo study was conducted in order to throw light on the performance of the VaR-parameter estimators in finite sample. We also report an empirical application to stock indices.

6.1. On simulated data

We begin by considering one of the simplest volatility model, the ARCH(1) where η_t follows the Student distribution with ν degrees of freedom:

$$\epsilon_t = \sigma_t\eta_t, \quad \sigma_t^2 = \omega_0 + \alpha_0\epsilon_{t-1}^2, \qquad \eta_t \sim St_\nu. \tag{6.1}$$

In this model, the VaR parameter is equal to $\theta_{0,\alpha} = (\omega_{0,\alpha}, \alpha_{0,\alpha})$ where $\omega_{0,\alpha} = \omega_0 K_\nu^2$, $\alpha_{0,\alpha} = \alpha_0 K_\nu^2$, and K_ν denoted the α-quantile of the St_ν. Simulating $N = 1,000$ independent trajectories of size $n = 500$ and $n = 5,000$ of this model, we compared the estimations of $\theta_{0,\alpha}$ obtained by the one-step estimator $\hat{\theta}_{n,\alpha}$ and the two-step estimator $\hat{\theta}_{n,\alpha}^{S2S}$. We considered the VaR levels $\alpha = 0.01$ and $\alpha = 0.05$. The intercept was fixed to $\omega_0 = 1$ and we took $\alpha_0 = \exp(-E\log\eta_1^2)/5$, which guarantees strict stationarity of the model for any value of ν (the necessary and sufficient strict stationarity condition being $E\log\alpha_0\eta_1^2 < 0$). Over the N simulations, and for both components of $\theta_{0,\alpha}$, the root mean squared error of estimation of the two methods are respectively denoted by RMSE_1 and RMSE_2. We then defined the empirical relative efficiency (ERE) of the one-step method with respect to the two-step method by the ratio $\mathrm{RMSE}_2/\mathrm{RMSE}_1$. For instance, with an ERE of two the

Table 2
ERE of the one-step method with respect to the two-step method for estimating the VaR parameter for the ARCH(1) model with Student innovations (6.1). The number of replication is $N = 1,000$, the level of the VaR is $\alpha = 5\%$ or $\alpha = 1\%$ and the length of each simulation is $n = 500$ or $n = 5,000$.

	$n = 500$							$n = 5,000$						
	ν							ν						
	1	2	3	4	5	6	∞	1	2	3	4	5	6	∞
$\alpha = 5\%$														
$\omega_{0,\alpha}$	7.5	2.8	1.7	1.3	1	0.9	0.9	13.9	6.6	2.7	1.3	1.1	1	0.8
$\alpha_{0,\alpha}$	7.3	3.6	1.7	1.3	1	1.0	0.8	22.2	8.7	3.2	1.3	1.1	1	0.9
$\alpha = 1\%$														
$\omega_{0,\alpha}$	6.1	1.6	1.0	0.8	0.7	0.7	0.7	41.1	3.6	1.6	0.9	0.8	0.8	0.7
$\alpha_{0,\alpha}$	3.8	1.8	2.6	0.8	0.7	0.7	0.7	13.7	6.0	2.1	0.9	0.8	0.7	0.7

one-step method can be considered as twice more efficient than the two-step method. Table 6.1 shows that, as expected from the asymptotic theory, the estimator $\hat{\theta}_{n,\alpha}$ is much more accurate than $\hat{\theta}^{2S}_{n,\alpha}$ when the distribution of η_t is heavy-tailed (*i.e.* ν is small) whereas $\hat{\theta}^{2S}_{n,\alpha}$ is slightly more accurate than $\hat{\theta}_{n,\alpha}$ when the distribution of η_t is close to the normal (*i.e.* ν is large). Similar conclusions were obtained for other distributions of the noise and for more elaborate volatility models. In particular, we considered the Threshold GARCH model introduced by Zakoïan (1994)

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t = \omega_0 + \alpha_{0+}\epsilon^+_{t-1} + \alpha_{0-}\epsilon^-_{t-1} + \beta_0 \sigma_{t-1}, \qquad \eta_t \sim \text{GED}(\nu). \tag{6.2}$$

The VaR parameter is now $\theta_{0,\alpha} = (\omega_{0,\alpha}, \alpha_{0+,\alpha}, \alpha_{0-,\alpha}, \beta_0)$ where $\omega_{0,\alpha} = -\omega_0 K_\nu$, $\alpha_{0+,\alpha} = -\alpha_{0+} K_\nu$, $\alpha_{0-,\alpha} = -\alpha_{0-} K_\nu$, and K_ν denotes the α-quantile of the GED(ν). In our simulation experiments, we chose $\omega_0 = 0.001$, $\beta_0 = 0.87$, $\alpha_{0+} = \overline{\alpha}/4$ and $\alpha_{0-} = \overline{\alpha}$ where $\overline{\alpha}$ is such that $E \log(\overline{\alpha}|\eta_t| + \beta_0) = 0$, which ensures a strict stationary solution to (6.2). The results presented in Table 3 are in accordance with Figure 3. The one-step estimator of the VaR parameter is more accurate then the two-step estimator when $\nu \leq 0.1$ or when ν is large (depending on the value of α). In the Gaussian case ($\nu = 0.5$), the two-step method should be preferred.

6.2. On real data

We considered nine major world stock indices covering the period from January, 2 1991 to August, 26 2011 : CAC (Paris), DAX (Frankfurt), FTSE

Table 3

ERE of the one-step method with respect to the two-step method for estimating the VaR parameter for the TGARCH(1,1) with GED(ν) innovations (6.2). The number of replication is $N = 1,000$, the level of the VaR is $\alpha = 5\%$ or $\alpha = 1\%$ and the length of each simulation is $n = 500$ or $n = 1,000$.

	$n = 500$						$n = 1,000$					
	ν						ν					
	0.01	0.1	0.5	1	2	3	0.01	0.1	0.5	1	2	3
$\alpha = 5\%$												
$\omega_{0,\alpha}$	1.5	1.2	0.8	2.0	4.7	2.3	1.3	1.2	0.8	2.0	6.3	2.4
$\alpha_{0+,\alpha}$	1.3	1.1	0.9	0.9	2.5	2.2	1.5	1.2	0.8	0.9	2.0	2.7
$\alpha_{0-,\alpha}$	1.4	1.1	0.8	1.0	2.2	2.1	1.4	1.2	0.9	0.9	1.5	2.7
β_0	1.5	1.1	0.9	1.6	4.0	1.9	1.4	1.2	0.8	1.6	4.6	2.4
$\alpha = 1\%$												
$\omega_{0,\alpha}$	2.1	1.2	0.6	0.7	3.0	4.7	2.2	1.3	0.6	0.6	4.2	4.2
$\alpha_{0+,\alpha}$	2.6	1.3	0.8	0.7	1.5	1.7	3.2	1.3	0.6	0.7	1.2	1.8
$\alpha_{0-,\alpha}$	2.6	1.3	0.8	0.8	1.3	1.6	2.7	1.2	0.7	0.7	1.0	2.2
β_0	2.2	1.2	0.7	0.7	2.4	2.3	2.7	1.3	0.7	0.7	3.0	2.8

(London), Nikkei (Tokyo), NSE (Bombay), SMI (Switzerland), SP500 (New York), SPTSX (Toronto), and SSE (Shanghai). For each series of log-returns, $\epsilon_t = \log(p_t/p_{t-1})$ where p_t denotes the value of the index, we estimated the VaR parameter $\theta_{0,\alpha}$ of GARCH(1,1) models, for $\alpha = 5\%$ and 1%. We report in Table 4 our estimates of $\theta_{0,\alpha}$ obtained by the one-step and the symmetric two-step methods, along with standard deviations. We also report two estimates of Δ_α based on residuals $\hat{\eta}_t$ or $\hat{\eta}_t^*$ of the two-step or the one-step method. Such estimates are negative for 7 out of 9 indices indicating, by Corollary 4.4, a superiority in accuracy of the one-step method for the risk level 5%. The conclusions are quite different when more extreme risks are considered. Table 5 shows that, for $\alpha = 1\%$, the two-step method is probably more efficient than the one-step method for large n (except perhaps for the SMI).

Table 6 reports estimation results for the ES parameter. While the estimated VaR parameters over the nine stocks were very similar, large differences between stocks appear when risk is measured by the ES. In particular, the Nikkei, NSE and SMI display much larger estimated ARCH coefficients $\alpha_{0,5\%}$ and $\alpha_{0,1\%}$.

Figure 4 displays the returns, estimated -VaR (at the 5% and 1% levels) and VaR accuracy intervals for the SP index from April, 6, 2011 to August,

Table 4

One-step estimator $\hat{\theta}_{n,\alpha}$ and two-step estimator $\hat{\theta}^{S2S}_{n,\alpha}$ for the VaR parameter $\theta_{0,\alpha}$ at level $\alpha = 5\%$ of GARCH(1,1) models for 9 stock market indices. The estimated standard deviations are given into brackets. The last two columns displays estimations of Δ_α ($\hat{\theta}_{n,\alpha}$ should be asymptotically more efficient than $\hat{\theta}^{S2S}_{n,\alpha}$ if and only if $\Delta_\alpha < 0$) based on residuals of the two-step and one-step methods.

Index	Estimator	$\omega_{0,5\%}$	$\alpha_{0,5\%}$	$\beta_{0,5\%}$	$\hat{\Delta}^{S2S}_{5\%}$	$\hat{\Delta}_{5\%}$
CAC	$\hat{\theta}^{S2S}_{n,5\%}$	0.08 (0.02)	0.23 (0.03)	0.90 (0.01)	-0.43	-0.70
	$\hat{\theta}_{n,5\%}$	0.05 (0.01)	0.23 (0.03)	0.90 (0.01)		
DAX	$\hat{\theta}^{S2S}_{n,5\%}$	0.09 (0.03)	0.22 (0.04)	0.90 (0.02)	-4.68	-6.84
	$\hat{\theta}_{n,5\%}$	0.04 (0.01)	0.22 (0.02)	0.91 (0.01)		
FTSE	$\hat{\theta}^{S2S}_{n,5\%}$	0.04 (0.01)	0.25 (0.02)	0.89 (0.01)	0.29	0.15
	$\hat{\theta}_{n,5\%}$	0.03 (0.01)	0.25 (0.02)	0.90 (0.01)		
Nikkei	$\hat{\theta}^{S2S}_{n,5\%}$	0.08 (0.02)	0.33 (0.05)	0.87 (0.02)	-3.86	-4.54
	$\hat{\theta}_{n,5\%}$	0.04 (0.01)	0.30 (0.03)	0.88 (0.01)		
NSE	$\hat{\theta}^{S2S}_{n,5\%}$	0.16 (0.06)	0.26 (0.06)	0.87 (0.03)	-3.11	-3.30
	$\hat{\theta}_{n,5\%}$	0.18 (0.05)	0.31 (0.05)	0.85 (0.02)		
SMI	$\hat{\theta}^{S2S}_{n,5\%}$	0.12 (0.03)	0.31 (0.05)	0.84 (0.03)	-3.05	-5.00
	$\hat{\theta}_{n,5\%}$	0.07 (0.02)	0.30 (0.04)	0.87 (0.02)		
SP500	$\hat{\theta}^{S2S}_{n,5\%}$	0.02 (0.00)	0.20 (0.02)	0.92 (0.01)	-2.10	-2.31
	$\hat{\theta}_{n,5\%}$	0.02 (0.00)	0.19 (0.01)	0.92 (0.01)		
SPTSX	$\hat{\theta}^{S2S}_{n,5\%}$	0.02 (0.01)	0.17 (0.03)	0.93 (0.01)	-0.06	-0.52
	$\hat{\theta}_{n,5\%}$	0.04 (0.01)	0.23 (0.03)	0.90 (0.01)		
SSE	$\hat{\theta}^{S2S}_{n,5\%}$	0.07 (0.03)	0.17 (0.03)	0.93 (0.01)	0.58	0.07
	$\hat{\theta}_{n,5\%}$	0.12 (0.04)	0.19 (0.04)	0.91 (0.02)		

26, 2011. The $(1-\alpha_0)\%$ confidence intervals (for $\alpha_0 = 5\%$) are obtained from formula (4.7). We preferred to report the opposite of the VaR, because the aim of such risk measures is to determine the capital reserve which acts as a protection against big losses (*i.e.* large negative values of ϵ_t). Obviously the estimated VaR's increase in module during the recent hectic period, while the confidence intervals are generally larger in such periods. In terms of risks, this can be interpreted as follows: unsurprisingly, the market risk increases in turbulent periods; but the estimation risk also increases and the magnitude of the confidence intervals allow to quantify the level of estimation risk. Similar conclusions can be drawn from Figure 5, displaying accuracy intervals for the opposite of the ES.

Table 5
As Table 4 but for the risk level $\alpha = 0.01$

Index	Estimator	$\omega_{0,1\%}$	$\alpha_{0,1\%}$	$\beta_{0,1\%}$	$\hat{\Delta}^{S2S}_{1\%}$	$\hat{\Delta}_{1\%}$
CAC	$\hat{\theta}^{S2S}_{n,1\%}$	0.18 (0.04)	0.52 (0.07)	0.90 (0.01)	3.29	3.31
	$\hat{\theta}_{n,1\%}$	0.17 (0.06)	0.54 (0.09)	0.89 (0.02)		
DAX	$\hat{\theta}^{S2S}_{n,1\%}$	0.19 (0.06)	0.48 (0.09)	0.90 (0.02)	2.38	0.61
	$\hat{\theta}_{n,1\%}$	0.23 (0.07)	0.68 (0.13)	0.87 (0.02)		
FTSE	$\hat{\theta}^{S2S}_{n,1\%}$	0.09 (0.02)	0.53 (0.05)	0.89 (0.01)	5.22	4.63
	$\hat{\theta}_{n,1\%}$	0.08 (0.03)	0.58 (0.08)	0.89 (0.02)		
Nikkei	$\hat{\theta}^{S2S}_{n,1\%}$	0.17 (0.05)	0.76 (0.11)	0.87 (0.02)	2.07	1.20
	$\hat{\theta}_{n,1\%}$	0.24 (0.06)	0.88 (0.12)	0.84 (0.02)		
NSE	$\hat{\theta}^{S2S}_{n,1\%}$	0.42 (0.15)	0.69 (0.16)	0.87 (0.03)	11.84	11.0
	$\hat{\theta}_{n,1\%}$	0.36 (0.23)	0.70 (0.26)	0.87 (0.04)		
SMI	$\hat{\theta}^{S2S}_{n,1\%}$	0.27 (0.07)	0.72 (0.12)	0.84 (0.03)	0.31	-0.93
	$\hat{\theta}_{n,1\%}$	0.24 (0.06)	0.80 (0.13)	0.83 (0.02)		
SP500	$\hat{\theta}^{S2S}_{n,1\%}$	0.04 (0.01)	0.47 (0.04)	0.92 (0.01)	1.17	0.51
	$\hat{\theta}_{n,1\%}$	0.04 (0.01)	0.44 (0.04)	0.92 (0.01)		
SPTSX	$\hat{\theta}^{S2S}_{n,1\%}$	0.05 (0.02)	0.42 (0.07)	0.93 (0.01)	2.92	3.18
	$\hat{\theta}_{n,1\%}$	0.03 (0.02)	0.42 (0.09)	0.93 (0.01)		
SSE	$\hat{\theta}^{S2S}_{n,1\%}$	0.18 (0.07)	0.43 (0.08)	0.93 (0.01)	9.33	6.47
	$\hat{\theta}_{n,1\%}$	0.30 (0.15)	0.60 (0.16)	0.90 (0.02)		

Table 6
Estimation of the ES parameter of GARCH(1,1) models at level $\alpha = 5\%$ and $\alpha = 1\%$ for 9 stock market indices. The estimated standard deviations are given into brackets.

Index	$\omega^*_{0,5\%}$	$\alpha^*_{0,5\%}$	$\beta^*_{0,5\%}$	$\omega^*_{0,1\%}$	$\alpha^*_{0,1\%}$	$\beta^*_{0,1\%}$
CAC	0.16 (0.04)	0.44 (0.06)	0.90 (0.01)	0.31 (0.08)	0.87 (0.14)	0.90 (0.01)
DAX	0.17 (0.05)	0.43 (0.08)	0.90 (0.02)	0.35 (0.12)	0.90 (0.21)	0.90 (0.02)
FTSE	0.08 (0.02)	0.47 (0.05)	0.89 (0.01)	0.17 (0.03)	0.93 (0.11)	0.89 (0.01)
Nikkei	0.15 (0.04)	0.68 (0.10)	0.87 (0.02)	0.34 (0.10)	1.50 (0.27)	0.87 (0.02)
NSE	0.36 (0.13)	0.59 (0.14)	0.87 (0.03)	0.91 (0.36)	1.50 (0.41)	0.87 (0.03)
SMI	0.25 (0.06)	0.65 (0.11)	0.84 (0.03)	0.52 (0.15)	1.37 (0.29)	0.84 (0.03)
SP500	0.04 (0.01)	0.40 (0.04)	0.92 (0.01)	0.08 (0.02)	0.91 (0.10)	0.92 (0.01)
SPTSX	0.05 (0.02)	0.37 (0.06)	0.93 (0.01)	0.10 (0.04)	0.81 (0.16)	0.93 (0.01)
SSE	0.15 (0.06)	0.36 (0.07)	0.93 (0.01)	0.33 (0.12)	0.79 (0.16)	0.93 (0.01)

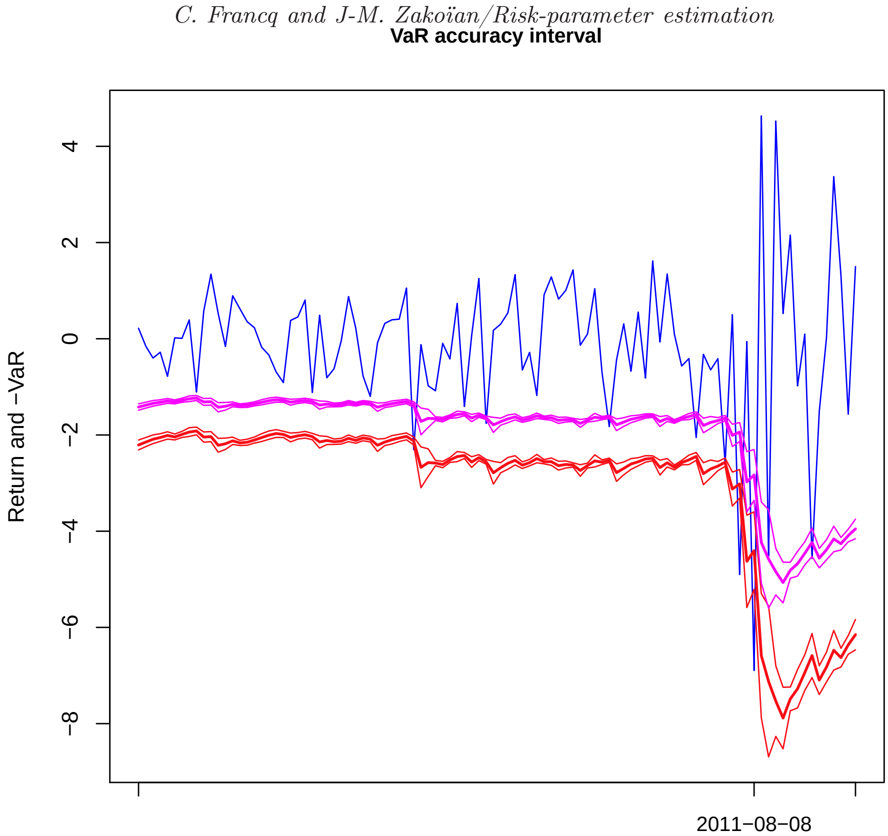


FIG 4. *Returns, estimated -VaR (at the 5% and 1% levels) and VaR accuracy intervals for the SP index from April, 6, 2011 to August, 26, 2011. Estimation of the VaR parameter is based on the 500 previous values.*

7. Conclusion

This paper presented two methods for estimating the risk parameter in conditionally heteroskedastic models. Asymptotic results were established for general risk measure, and a particular attention was devoted to the VaR and the ES.

The introduction of a VaR parameter facilitates the asymptotic comparison of the risk evaluation procedures. In particular, for the standard GARCH models the ranking of the two methods, in terms of asymptotic relative efficiency, only depends on the sign of the scalar Δ_α defined by (4.10), involving the risk level α and characteristics of the innovations distribution. Estimation of the ES parameter can be achieved by the two-step method, and it

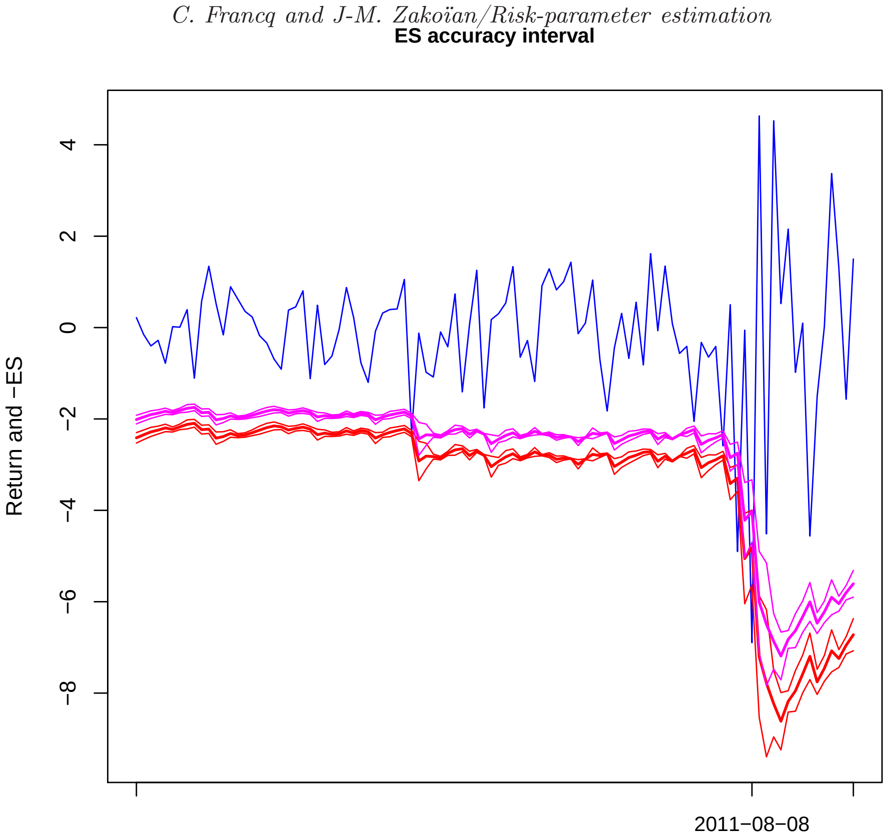


FIG 5. *Returns, estimated ES (at the 5% and 1% levels) and ES accuracy intervals for the SP index from April, 6, 2011 to August, 26, 2011. Estimation of the ES parameter is based on the 500 previous values.*

turns out that the asymptotic distribution of the estimator depends on the GARCH coefficients and other simple characteristics of the innovations distribution.

Finally, the estimation risk, namely the effect of the inaccuracy of the parameter estimation on the risk evaluation, can be explicitly taken into account, leading to confidence bounds for the VaR and the ES.

A natural extension of this work would consider heteroskedastic models including a conditional mean. For instance, Ling (2004) introduced a class of double-autoregressive models and studied the properties of the QMLE. In addition, the concept of risk parameter, as well as the proposed estimation procedures, undoubtedly can be generalized to multivariate conditional volatility models. These extensions are left for future research.

Appendix A: Technical assumptions and proofs

A.1. Technical assumptions

A6: θ_0^* belongs to the interior of Θ.

A7: There exist no non-zero $x \in \mathbb{R}^m$ such that $x' \frac{\partial \sigma_t(\theta_0^*)}{\partial \theta} = 0$, *a.s.*

A8: The function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives, and

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

where C_1 and ρ are as in **A5**.

A9: h is twice differentiable with $|u^2 \left(h'(u)/h(u)\right)'| \leq C_0(1 + |u|^\delta)$ for all $u \in \mathbb{R}$.

A10: There exists a neighborhood $V(\theta_0^*)$ of θ_0^* such that the following variables have finite expectation:

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

A.2. Proofs for the results of Section 3

We start by a lemma. For $x/\sigma \in A^c$, let

$$g_1(x, \sigma) = \frac{\partial g(x, \sigma)}{\partial \sigma} = -\frac{1}{\sigma} - \frac{h'(x/\sigma)}{h(x/\sigma)} \frac{x}{\sigma^2}.$$

Lemma A.1. *Under Assumption* **A4***, for $\sigma, \tilde{\sigma} > \underline{\omega} > 0$ we have, for $\sigma^\star$ between σ and $\tilde{\sigma}$ such that $x/\sigma^\star \in A^c$,*

$$|g(x, \sigma) - g(x, \tilde{\sigma})| \leq \begin{cases} |g_1(x, \sigma^\star)||\sigma - \tilde{\sigma}| & \text{if } x \neq 0, \\ \frac{1}{\underline{\omega}} |\sigma - \tilde{\sigma}| & \text{if } x = 0. \end{cases}$$

Proof. It is not restrictive to assume $\sigma > \tilde{\sigma}$. For $x \neq 0$, write $A \cap (x/\sigma, x/\tilde{\sigma}) = \{x/\sigma_1, \dots, x/\sigma_j\}$ when this set is non empty. By convention, $j = 0$ when this set is empty. Assume

$$\sigma = \sigma_0 > \sigma_1 > \ldots > \sigma_{j+1} = \tilde{\sigma}.$$

By applying Rolle's theorem on the sets $(x/\sigma_i, x/\sigma_{i+1})$ we get, for $\sigma_i^\star \in (\sigma_i, \sigma_{i+1})$,

$$|g(x, \sigma) - g(x, \tilde{\sigma})| = \left| \sum_{i=0}^{j} g_1(x, \sigma_i^\star)(\sigma_{i+1} - \sigma_i) \right| \leq \sup_i |g_1(x, \sigma_i^\star)|(\sigma - \tilde{\sigma}).$$

We also have $|g(0, \sigma) - g(0, \tilde{\sigma})| = \log \sigma - \log \tilde{\sigma}$ and the conclusion follows. □

Proof of Theorem 3.1. The consistency is a consequence of the following intermediate results (see *e.g.* the proofs of Theorem 7.1 in Francq and Zakoian, 2010):

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

ii) $\mathbb{E}|g(\epsilon_t, \sigma_t(\theta_0^*))| < \infty$ and if $\theta \neq \theta_0^*$, $\mathbb{E}g(\epsilon_t, \sigma_t(\theta)) < \mathbb{E}g(\epsilon_t, \sigma_t(\theta_0^*))$,

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

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

where

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

Let K be a generic positive constant, allowed to be a random variable, measurable with respect to $\{\epsilon_u, u \le 0\}$, whose values will be modified along the proofs.

We begin by showing *i)*. Using a Taylor expansion, almost surely

$$\begin{aligned}
&\sup_{\theta\in\Theta}|Q_n(\theta) - \tilde{Q}_n(\theta)| \\
\le\ & n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|g_1(\epsilon_t, \sigma_t^{\star}(\theta))||\tilde{\sigma}_t(\theta) - \sigma_t(\theta)|\mathbf{1}_{\{\epsilon_t\neq 0\}} \\
&+n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\frac{1}{\underline{\omega}}|\tilde{\sigma}_t(\theta) - \sigma_t(\theta)|\mathbf{1}_{\{\epsilon_t=0\}} \\
\le\ & n^{-1}K\sum_{t=1}^{n}\rho^t\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^{\star}}\frac{\epsilon_t}{\sigma_t^{\star}}\frac{h'}{h}\left(\frac{\epsilon_t}{\sigma_t^{\star}}\right)\right|\mathbf{1}_{\{\epsilon_t\neq 0\}} + \frac{K}{\underline{\omega}}\,n^{-1}\sum_{t=1}^{n}\rho^t \\
\le\ & Kn^{-1}\sum_{t=1}^{n}\rho^t(1+|\epsilon_t|^{\delta}),
\end{aligned}$$

where g_1 is defined in Theorem 3.2 and $\sigma_t^{\star}(\theta)$ is between $\tilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. The first inequality is a consequence of Lemma A.1. The last two inequalities rest on Assumptions **A2**, **A4** and **A5**. By the Markov inequality and **A4**, we deduce

$$\sum_{t=1}^{\infty}\mathbb{P}(\rho^t|\epsilon_t|^{\delta} > \varepsilon) \ \le\ \sum_{t=1}^{\infty}\frac{\rho^{ts/\delta}\mathbb{E}|\epsilon_t|^s}{\varepsilon^{\frac{s}{\delta}}} < \infty$$

and thus $\rho^t|\epsilon_t|^{\delta} \to 0$ $a.s$ by the Borel-Cantelli lemma. Thus, *i)* follows by the Cesàro lemma.

Condition *ii)* is a consequence of **A2**-**A3**. Indeed,

$$\mathbb{E}\{g(\epsilon_t, \sigma_t(\theta)) - g(\epsilon_t, \sigma_t(\theta_0^*))\} = \mathbb{E}\left\{g\left(\eta_t^*, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0^*)}\right) - g(\eta_t^*, 1)\right\} \le 0,$$

with equality iff $\theta = \theta_0^*$.

The proof of *iii)* is omitted because it uses more standard arguments. □

Proof of Proposition 3.1. First assume that h satisfies (3.4). By Assumptions **A4** and **A2**, the function $\sigma \mapsto \partial \log h\,(\eta_0^*/\sigma)\partial\sigma$ is bounded by an integrable random variable, uniformly in a neighborhood of any $\sigma > 0$. By the dominated convergence theorem, we thus have

$$\frac{\partial}{\partial\sigma}Eg(\eta_0^*, \sigma) = \frac{-1}{\sigma} - \frac{1}{\sigma}E\left\{\lambda\psi\left(\frac{\eta_0^*}{\sigma}\right) - 1\right\},$$

which is equal to zero if and only if $r(\eta_0^*/\sigma) = 1$. By $r(\eta_0^*) = 1$ and by the positive homogeneity of r, $r(\eta_0^*/\sigma) = 1$ is equivalent to $\sigma = 1$. Thus, for h satisfying (3.4), **A3** is a consequence of (3.3).

To prove the "only if" part, note that

$$\mathbf{A3} \quad \Rightarrow \quad E\left(\frac{h'(\eta_0^*)}{h(\eta_0^*)}\eta_0^*\right) = -1. \tag{A.1}$$

Note also that ψ cannot be null on $\mathbb{R}$ because this would imply $r(X) = 1$ for any variable X, by (3.3), in contradiction with the positive homogeneity of r. Now if (3.4) does not hold, then for some x_1, x_2 with $\psi(x_1) \neq 0$, and $\lambda_1 \neq \lambda_2$

$$\frac{h'(x_i)}{h(x_i)} x_i + 1 = \lambda_i \psi(x_i), \quad i = 1, 2.$$

Let η such that $P(\eta = x_i) = p_i > 0$ with $p_1 + p_2 = 1$, and $\psi(x_1)p_1 + \psi(x_2)p_2 = 0$. Then $E\psi(\eta) = 0$ and

$$E\left(\frac{h'(\eta)}{h(\eta)}\eta\right) + 1 = \lambda_1\psi(x_1)p_1 + \lambda_2\psi(x_2)p_2 = (\lambda_1 - \lambda_2)\psi(x_1)p_1 \neq 0.$$

Then, in view of (A.1), Assumption **A3** is not satisfied. We have found a distribution of η_0^* such that $r(\eta_0^*) = 1$ but **A3** is not satisfied. The proposition follows. □

Proof of Theorem 3.2. The proof is based on a Taylor expansion of the criterion $\tilde{Q}_n$ at θ_0^*. Since $\hat{\theta}_n^*$ converges to θ_0^*, which stands in the interior of the parameter space by **A6**, for n large enough the derivative of the criterion is equal to zero at $\hat{\theta}_n^*$. We thus have

$$0 \quad = \quad \sqrt{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta} g(\epsilon_t, \tilde{\sigma}_t(\theta_0^*)) + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j} g(\epsilon_t, \tilde{\sigma}_t(\theta_{ij}^*))\right)\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right)$$

where the θ_{ij}^*'s are between $\hat{\theta}_n^*$ and θ_0^*. The asymptotic normality is proven by means of the following intermediate results: for some neighboorhood $V(\theta_0^*)$ of θ_0^*,

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

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

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

vii) I is nonsingular,

for any θ^* between $\hat{\theta}_n^*$ and θ_0^*. For brevity, we will skip the proof of *v)-vii)* which is available from the authors. To prove *iv)* we note that

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0^*)}\sqrt{n}\left\|\frac{\partial}{\partial\theta}Q_n(\theta) - \frac{\partial}{\partial\theta}\tilde{Q}_n(\theta)\right\| \\
\leq\ & \sup_{\theta\in V(\theta_0^*)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|g_1(\epsilon_t, \sigma_t(\theta)) - g_1(\epsilon_t, \tilde{\sigma}_t(\theta))|\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| \\
&+ \sup_{\theta\in V(\theta_0^*)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|g_1(\epsilon_t, \tilde{\sigma}_t(\theta))|\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta} - \frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}\right\|.
\end{aligned} \tag{A.2}$$

Note that $|g_1(\epsilon_t, \tilde{\sigma}_t(\theta))| \leq K(1 + |\epsilon_t|^\delta)$ by **A4** and the first part of **A2**. Thus, using **A8**, the last term in (A.2) is bounded by

$$\frac{K}{\sqrt{n}}\sum_{t=1}^{n}\rho^t(1 + |\epsilon_t|^\delta),$$

It is not restrictive to assume $s < \delta$ in **A4**. We have, by the c_r-inequality,

$$E\left(\sum_{t=1}^{\infty} \rho^t (1 + |\epsilon_t|^\delta)\right)^{s/\delta} \leq \sum_{t=1}^{\infty} \rho^{st/\delta}(1 + |\epsilon_t|^s) < \infty.$$

It follows that

$$\frac{K}{\sqrt{n}} \sum_{t=1}^{n} \rho^t (1 + |\epsilon_t|^\delta) \leq \frac{K}{\sqrt{n}} \sum_{t=1}^{\infty} \rho^t (1 + |\epsilon_t|^\delta) \to 0, \quad a.s.$$

Thus the second term in the right-hand side of (A.2) goes to 0 $a.s.$

The first term in (A.2) is bounded by

$$\sup_{\theta \in V(\theta_0^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_2(\epsilon_t, \sigma_t^*(\theta))| K \rho^t \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \tag{A.3}$$

where $g_2(x, \sigma) = \partial g_1(x, \sigma)/\partial \sigma$ and $\sigma_t^*(\theta)$ is between $\tilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. Noting that

$$g_2(x, \sigma) = \frac{1}{\sigma^2} \left[1 + \frac{x}{\sigma} \left\{2\frac{h'}{h} + \frac{x}{\sigma}\left(\frac{h'}{h}\right)'\right\} \left(\frac{x}{\sigma}\right)\right],$$

we have

$$|g_2(\epsilon_t, \sigma_t^*(\theta))| \leq \frac{K}{\sigma_t^*(\theta)} \left(1 + |\epsilon_t|^\delta\right)$$

by **A4**, **A9** and the first part of **A2**. The term in (A.3) is thus bounded by

$$\frac{K}{\sqrt{n}} \sum_{t=1}^{n} \rho^t (1 + |\epsilon_t|^\delta) \sup_{\theta \in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|.$$

By the Cauchy-Schwarz inequality and **A10** we have, for $s < 2\delta$,

$$\begin{aligned} & E\left(\sum_{t=1}^{\infty} \rho^t (1 + |\epsilon_t|^\delta) \sup_{\theta \in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|\right)^{s/2\delta} \\ \leq & \left(\sum_{t=1}^{\infty} \rho^{st/2\delta} \{E(1 + 2|\epsilon_t|^{s/2} + |\epsilon_t|^s)\}^{1/2} \left\{E \sup_{\theta \in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^{s/\delta}\right\}^{1/2}\right) < \infty. \end{aligned}$$

We can conclude that the first term in the right-hand side of (A.2) goes to 0 $a.s.$, which completes the proof of *iv)*. □

A.3. Proofs for the results of Section 4

Proof of Corollary 4.1. The distribution of η_t^* being symmetric, first note that Model (4.1) is in the form (2.4) with $r(\eta_t^*) = F_{\eta^*}^{-1}(1 - \alpha)$. Thus **A1** is satisfied. Now note that $\hat{\theta}_{n,\alpha} = \hat{\theta}_n^*$, where $\hat{\theta}_n^*$ is defined by (3.1)-(3.2) with, up to an additive constant, $g(\epsilon_t, \tilde{\sigma}_t(\theta))$ equal to

$$(2\lambda\alpha - 1) \log |\epsilon_t| - \lambda \mathbf{1}_{\{|\epsilon_t| > \tilde{\sigma}_t(\theta)\}} \log |\epsilon_t| - \lambda(2\alpha - \mathbf{1}_{\{|\epsilon_t| > \tilde{\sigma}_t(\theta)\}}) \log \tilde{\sigma}_t(\theta).$$

Assumption **A4** being satisfied with $A = \{-1, 0, 1\}$ and $\delta = 0$, it remains to show that **A3** holds true and the conclusion will follow from Theorem 3.1.

Note that $E|g(\eta_0^*, 1)| < \infty$ under the moment condition on $\log|\eta_0^*|$. Moreover, for any $\sigma > 0, \sigma \neq 1$, the distribution of η_0^* being symmetric,

$$Eg(\eta_0^*, \sigma) - Eg(\eta_0^*, 1) = -\lambda E(\log|\eta_0^*| - \log\sigma)(\mathbf{1}_{\{|\eta_0^*|>\sigma\}} - \mathbf{1}_{\{|\eta_0^*|>1\}}) \leq 0.$$

Note that the inequality is strict because η_0^* admits a density in a neighborhood of 1, which completes the proof. □

Proof of Theorem 4.1. First note that $\hat{\nu}_n := \sqrt{n}(\hat{\theta}_{n,\alpha} - \theta_{0,\alpha})$ is such that

$$\hat{\nu}_n = \arg\min_{\nu\in\Lambda_n} \tilde{S}_n(\nu),$$

where $\Lambda_n := \sqrt{n}(\Theta - \theta_{0,\alpha})$ and

$$\tilde{S}_n(\nu) = \sum_{t=1}^n \rho_{1-2\alpha}\left\{\log\left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta_{0,\alpha} + n^{-1/2}\nu)}\right)\right\} - \rho_{1-2\alpha}\left\{\log\left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta_{0,\alpha})}\right)\right\}.$$

Showing that the initial values are asymptotically negligible, and linearizing $\log\sigma_t(\theta_{0,\alpha} + n^{-1/2}\nu)$ around $\nu = 0$, Lemma A.2 below demonstrates that $\tilde{S}_n(\nu)$ can be approximated by

$$S_n(\nu) = \sum_{t=1}^n \rho_{1-2\alpha}\left\{\log|\eta_t^*| - \frac{\nu'}{\sqrt{n}}D_t(\theta_{0,\alpha})\right\} - \rho_{1-2\alpha}\left(\log|\eta_t^*|\right).$$

Let $f_{\log|\eta^*|}$ denote the density of the variable $\log|\eta_t^*|$. Using a convexity argument, Lemma A.3 then shows that $S_n(\nu)$ weakly converges to the process

$$S(\nu) = \sqrt{2\alpha(1-2\alpha)}\nu' N + f_{\log|\eta^*|}(0)\nu' J_\alpha \nu/2, \quad N \sim \mathcal{N}(0, J_\alpha).$$

The process is minimized at

$$\hat{\nu} = -\frac{\sqrt{2\alpha(1-2\alpha)}}{f_{\log|\eta^*|}(0)} J_\alpha^{-1} N \sim \mathcal{N}\left(0, \frac{2\alpha(1-2\alpha)}{f^2_{\log|\eta^*|}(0)} J_\alpha^{-1}\right)$$

The conclusion follows from Remark 1 and Lemma 2.2 in Davis, Knight and Liu (1992).□

Lemma A.2. *Let $\mathcal{C}_K = \{\nu \in \mathbb{R}^m : \|\nu\| \leq K\}$. Under the assumptions of Theorem 4.1, for all $K > 0$, we have*

$$\sup_{\nu\in\mathcal{C}_K} |S_n(\nu) - \tilde{S}_n(\nu)| = o_P(1).$$

Proof: Let $\dot{D}_t(\theta) = \frac{\partial}{\partial\theta}D_t'(\theta)$, and let $\tilde{D}_t(\theta)$ (resp. $\dot{\tilde{D}}_t(\theta)$) be obtained by replacing $\sigma_t(\theta)$ by $\tilde{\sigma}_t(\theta)$ in $D_t(\theta)$ (resp. $\dot{D}_t(\theta)$). A Taylor expansion yields

$$\log\tilde{\sigma}_t\left(\theta_{0,\alpha} + \frac{\nu}{\sqrt{n}}\right) = \log\tilde{\sigma}_t(\theta_{0,\alpha}) + \sum_{i=0}^1 \tilde{v}_{t,n}^{(i)}(\nu),$$

with

$$\tilde{v}_{t,n}^{(0)}(\nu) = \frac{1}{\sqrt{n}}\nu'\tilde{D}_t(\theta_{0,\alpha}), \qquad \tilde{v}_{t,n}^{(1)}(\nu) = \frac{1}{2n}\nu'\dot{\tilde{D}}_t(\tilde{\theta}_t^*)\nu,$$

where $\tilde{\theta}_t^*$ is between $\theta_{0,\alpha}$ and $\theta_{0,\alpha} + n^{-1/2}\nu$. Let $\tilde{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\theta_{0,\alpha})$. Using the identity

$$\begin{aligned}
& \rho_{1-2\alpha}(u-v) - \rho_{1-2\alpha}(u) \\
= & -v(1-2\alpha - \mathbf{1}_{\{u<0\}}) + (u-v)\left\{\mathbf{1}_{\{0>u>v\}} - \mathbf{1}_{\{0<u<v\}}\right\} \\
= & -v(1-2\alpha - \mathbf{1}_{\{u<0\}}) + \int_0^v \left\{\mathbf{1}_{\{u\leq s\}} - \mathbf{1}_{\{u<0\}}\right\} ds \qquad \text{(A.4)}
\end{aligned}$$

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

$$\tilde{S}_n(\nu) = \sum_{i=0}^{1} \tilde{T}_n^{(i)}(\nu) + \tilde{U}_n^{(i)}(\nu),$$

with

$$\tilde{T}_n^{(i)}(\nu) = -\sum_{t=1}^{n} \tilde{v}_{t,n}^{(i)}(\nu)(1-2\alpha - \mathbf{1}_{\{|\tilde{\eta}_t|<1\}}),$$

and

$$\tilde{U}_n^{(0)}(\nu) = \sum_{t=1}^{n} \tilde{\xi}_{t,n}^{(0)}(\nu), \qquad \tilde{\xi}_{t,n}^{(0)}(\nu) = \int_0^{\tilde{v}_{t,n}^{(0)}(\nu)} \left\{\mathbf{1}_{\{\log|\tilde{\eta}_t|\leq s\}} - \mathbf{1}_{\{\log|\tilde{\eta}_t|<0\}}\right\} ds,$$

$$\tilde{U}_n^{(1)}(\nu) = \sum_{t=1}^{n} \tilde{\xi}_{t,n}^{(1)}(\nu), \qquad \tilde{\xi}_{t,n}^{(1)}(\nu) = \int_{\tilde{v}_{t,n}^{(0)}(\nu)}^{\sum_{i=0}^{1}\tilde{v}_{t,n}^{(i)}(\nu)} \left\{\mathbf{1}_{\{\log|\tilde{\eta}_t|\leq s\}} - \mathbf{1}_{\{\log|\tilde{\eta}_t|<0\}}\right\} ds.$$

Define $T_n^{(i)}(\nu)$ by replacing $\tilde{\sigma}_t(\cdot)$ by $\sigma_t(\cdot)$ in $\tilde{T}_n^{(i)}(\nu)$. Define $U_n^{(i)}(\nu)$, $\xi_{t,n}^{(i)}(\nu)$ and $v_{t,n}^{(i)}(\nu)$ similarly. Noting that $T_n^{(1)}(\nu)$ is centered and has a variance of order $O(1/n)$ uniformly in ν, in view of **A10**, we obtain $\sup_{\nu\in\mathcal{C}_K} |T_n^{(1)}(\nu)| = o_P(1)$. Now we have

$$\begin{aligned}
\left|\tilde{T}_n^{(1)}(\nu) - T_n^{(1)}(\nu)\right| \leq & \frac{1}{2n}\sum_{t=1}^{n}\left|\nu'\dot{\tilde{D}}_t(\tilde{\theta}_t^*)\nu\right| \left|\mathbf{1}^*_{\{|\epsilon_t|\in(\sigma_t(\theta_{0,\alpha}),\tilde{\sigma}_t(\theta_{0,\alpha}))\}}\right| \\
& + \frac{1}{2n}\sum_{t=1}^{n}\left|\nu'\left\{\dot{\tilde{D}}_t(\tilde{\theta}_t^*) - \dot{D}_t(\tilde{\theta}_t^*)\right\}\nu\right|
\end{aligned}$$

with $\mathbf{1}^*_{\{X\in(a,b)\}} = \mathbf{1}_{\{X<b\}} - \mathbf{1}_{\{X<a\}}$ for any real numbers a, b and any real random variable X. The second term of the right-hand side of the previous inequality is bounded by $K\sum_{t=1}^{n}\rho^t/n = o(1)$ by **A2**, **A5** and **A8**. Using the Hölder inequality, **A10**, and

$$E\left|\mathbf{1}^*_{\{|\epsilon_t|\in(\sigma_t(\theta_{0,\alpha}),\tilde{\sigma}_t(\theta_{0,\alpha}))\}}\right| \leq P\left\{|\eta_t^*| \in \left(1, \frac{\sigma_t(\theta_{0,\alpha})}{\tilde{\sigma}_t(\theta_{0,\alpha})}\right)\right\} \leq K\rho^t \qquad \text{(A.5)}$$

the first term also tends to zero. Thus $\sup_{\nu\in\mathcal{C}_K} |\tilde{T}_n^{(1)}(\nu)| = o_P(1)$. Furthermore

$$\begin{aligned}
\left|\tilde{T}_n^{(0)}(\nu) - T_n^{(0)}(\nu)\right| \leq & \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\nu'\tilde{D}_t(\theta_{0,\alpha})\right| \left|\mathbf{1}^*_{\{|\epsilon_t|\in(\sigma_t(\theta_{0,\alpha}),\tilde{\sigma}_t(\theta_{0,\alpha}))\}}\right| \\
& + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\nu'\left\{\tilde{D}_t(\theta_{0,\alpha}) - D_t(\theta_{0,\alpha})\right\}\right|
\end{aligned}$$

is also of order $o_P(1)$ uniformly in ν, by the Markov inequality and already used arguments.

Now, using the elementary relation

$$\int_0^{\tilde{a}} \tilde{f}(x)dx - \int_0^a f(x)dx = \int_a^{\tilde{a}} \tilde{f}(x)dx + \int_0^a \left\{\tilde{f}(x) - f(x)\right\} dx$$

with standard notations, we have

$$|\tilde{\xi}_{t,n}^{(0)}(\nu) - \xi_{t,n}^{(0)}(\nu)| = \left| \int_{v_{t,n}^{(0)}(\nu)}^{\tilde{v}_{t,n}^{(0)}(\nu)} \left\{\mathbf{1}_{\{\log|\tilde{\eta}_t|\le s\}} - \mathbf{1}_{\{\log|\tilde{\eta}_t|<0\}}\right\} ds \right.$$
$$\left. + \int_0^{v_{t,n}^{(0)}(\nu)} \left\{\mathbf{1}^*_{\{|\eta_t^*|\in\left(e^s, e^s\frac{\tilde{\sigma}_t(\theta_{0,\alpha})}{\sigma_t(\theta_{0,\alpha})}\right)\}} - \mathbf{1}^*_{\{|\eta_t^*|\in\left(1,\frac{\tilde{\sigma}_t(\theta_{0,\alpha})}{\sigma_t(\theta_{0,\alpha})}\right)\}}\right\} ds \right|.$$

Using the inequalities $\sqrt{n}|v_{t,n}^{(0)}(\nu) - \tilde{v}_{t,n}^{(0)}(\nu)| \le K\rho^t$ and (A.5), splitting the latter integral into two parts, and taking conditional expectation, we then obtain

$$E_{t-1}|\tilde{\xi}_{t,n}^{(0)}(\nu) - \xi_{t,n}^{(0)}(\nu)| \quad \le \quad \frac{K\rho^t}{\sqrt{n}} + \int_0^{\frac{1}{n^{1/4}}} K\rho^t ds + 2|v_{t,n}^{(0)}(\nu)|\mathbf{1}_{\{|v_{t,n}^{(0)}(\nu)|\ge n^{-1/4}\}}$$

where E_{t-1} denotes the expectation conditional on $\{\eta_u : u < t\}$. Using in particular

$$\begin{aligned} E|v_{t,n}^{(0)}(\nu)|\mathbf{1}_{\{|v_{t,n}^{(0)}(\nu)|\ge n^{-1/4}\}} \quad &\le \quad \left\|\nu' D_t(\theta_{0,\alpha})/\sqrt{n}\right\|_4 \left\{P(|\nu' D_t(\theta_{0,\alpha})| \ge n^{1/4})\right\}^{3/4} \\ &\le \quad K/n^{11/16}, \end{aligned}$$

the Markov inequality shows that $\sup_{\nu\in\mathcal{C}_K} |\tilde{U}_n^{(0)}(\nu) - U_n^{(0)}(\nu)| = o_P(1)$. Similarly, one can show that $\sup_{\nu\in\mathcal{C}_K} |\tilde{U}_n^{(1)}(\nu) - U_n^{(1)}(\nu)| = o_P(1)$.

Now note that

$$\begin{aligned} \left|E_{t-1}\xi_{t,n}^{(1)}(\nu)\right| \quad &\le \quad \left| \int_{v_{t,n}^{(0)}(\nu)}^{\sum_{i=0}^1 v_{t,n}^{(i)}(\nu)} \int_0^s f_{\log|\eta^*|}(s)ds \right| \\ &\le \quad \left| \int_{v_{t,n}^{(0)}(\nu)}^{\sum_{i=0}^1 v_{t,n}^{(i)}(\nu)} s \max_{[0,s]} |f_{\log|\eta^*|}(x)|ds. \right| \end{aligned}$$

By **A11** and arguments already given, the expectation of the previous variable is of order $O_P(n^{-3/2})$. It follows that $\sup_{\nu\in\mathcal{C}_K} |U_n^{(1)}(\nu)| = o_P(1)$.

We have thus shown that

$$\sup_{\nu\in\mathcal{C}_K} |\tilde{S}_n(\nu) - T_n^{(0)}(\nu) - U_n^{(0)}(\nu)| = o_P(1).$$

The conclusion follows by noting that, in view of (A.4), $S_n(\nu) = T_n^{(0)}(\nu) + U_n^{(0)}(\nu)$. □

Lemma A.3. *Under the assumptions of Theorem 4.1 we have*

$$S_n(\cdot) \xrightarrow{d} S(\cdot)$$

on the space $C(\mathbb{R}^m)$ of the continuous functions on $\mathbb{R}^m$ where convergence means uniform convergence on every compact set.

Proof: We have shown in Lemma A.2 that (A.4) entails

$$S_n(\nu) = T_n(\nu) + U_n(\nu),$$

with

$$T_n(\nu) = -\frac{\nu'}{\sqrt{n}} \sum_{t=1}^{n} D_t(\theta_{0,\alpha})(1 - 2\alpha - \mathbf{1}_{\{|\eta_t^*|<1\}})$$

and

$$\begin{aligned} U_n(\nu) &= \sum_{t=1}^{n} \xi_{t,n}(\nu), \\ \xi_{t,n}(\nu) &= \left(\log|\eta_t^*| - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) \\ &\quad \times \left(\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})<\log|\eta_t^*|<0\}} - \mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>\log|\eta_t^*|>0\}}\right). \end{aligned}$$

We have

$$\begin{aligned} & E_{t-1}\left(\log|\eta_t^*| - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) \mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>\log|\eta_t^*|>0\}} \\ = \; & \mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>0\}} \int_0^{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})} \left(x - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) f_{\log|\eta^*|}(x)dx \\ = \; & f_{\log|\eta^*|}(0)\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>0\}} \int_0^{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})} \left(x - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) dx + R_{n,t} \\ = \; & \frac{-1}{2n} f_{\log|\eta^*|}(0)\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>0\}} \nu' D_t(\theta_{0,\alpha}) D_t'(\theta_{0,\alpha})\nu + R_{n,t} \end{aligned}$$

where $R_{n,t}$ is equal to

$$\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>0\}} \int_0^{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})} \left(x - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) (f_{\log|\eta^*|}(x) - f_{\log|\eta^*|}(0))dx.$$

By **A11**, for any $\epsilon > 0$ there exists $\tau > 0$ such that $|x| < \tau$ entails $|f_{\log|\eta^*|}(x) - f_{\log|\eta^*|}(0)| < \epsilon$. It follows that

$$|R_{n,t}| \leq \frac{1}{2n} \{\nu' D_t(\theta_{0,\alpha})\}^2 (\epsilon + 2M\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})>\tau\}}).$$

By the Hölder and Markov inequalities and **A10**, we then show that $\sum_{t=1}^n |R_{n,t}| = o(1)$ $a.s.$ Similarly,

$$\begin{aligned} & E_{t-1}\left(\log|\eta_t^*| - \frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})\right) \mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})<\log|\eta_t^*|<0\}} \\ = \; & \frac{1}{2n} f_{\log|\eta^*|}(0)\mathbf{1}_{\{\frac{\nu'}{\sqrt{n}} D_t(\theta_{0,\alpha})<0\}} \nu' D_t(\theta_{0,\alpha}) D_t'(\theta_{0,\alpha})\nu + R_{n,t}^* \end{aligned}$$

where $R_{n,t}^*$ is analogous to $R_{n,t}$. We thus have

$$\begin{aligned} \sum_{t=1}^{n} E_{t-1}\xi_{t,n}(\nu) &= f_{\log|\eta^*|}(0) \frac{1}{2n} \sum_{t=1}^{n} \nu' D_t(\theta_{0,\alpha}) D_t'(\theta_{0,\alpha})\nu + o(1) \\ &= f_{\log|\eta^*|}(0)\nu' J\nu/2 + o(1) \;\; a.s. \end{aligned}$$

Now, note that if X is an integrable random variable and c is a real constant, then the random variable

$$Y = (X - c)(\mathbf{1}_{\{c<X<0\}} - \mathbf{1}_{\{0<X<c\}})$$

satisfies $\mathrm{Var}Y \leq -cEY$. Using this elementary inequality and the previous results, the martingale difference $\overline{\xi}_{t,n}(\nu) := \xi_{t,n}(\nu) - E_{t-1}\xi_{t,n}(\nu)$ satisfies $\sum_{t=1}^{n} E_{t-1}\overline{\xi}_{t,n}^{2}(\nu) = o_P(1)$.

It follows that $U_n(\nu) \to f_{\log|\eta^*|}(0)\nu' J_\alpha \nu/2$ in probability as $n \to \infty$. The martingale CLT entails that $T_n(\nu)$ converges in distribution to the Gaussian vector $\sqrt{2\alpha(1-2\alpha)}\nu' N$ where $N \sim \mathcal{N}(0, J_\alpha)$. We thus have shown that the finite dimensional distributions of $S_n(\cdot)$ converge to that of the process $S(\cdot)$. Since the process $S_n(\cdot)$ has convex sample paths, the convexity lemmas of Knight (1989) and Pollard (1991) show that S_n converges weakly to the process S. □

A.4. Proofs for the results of Section 4.2

Proof of Theorem 4.2. It will be sufficient to prove the first asymptotic normality result. The symmetric case can be handled similarly and the proof is available from the authors. We have

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

Thus

$$\sqrt{n}(\xi_{n,\alpha} - \xi_\alpha) = \arg\min_{z\in\mathbb{R}} Q_n(z)$$

where

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

Let $\eta_t(\theta) = \epsilon_t/\tilde{\sigma}_t(\theta)$. A Taylor expansion around θ_0 yields

$$\hat{\eta}_t = \eta_t - \eta_t D_t'(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' \frac{\partial^2 \eta_t(\theta_0)}{\partial\theta\partial\theta'}(\hat{\theta}_n - \theta_0) + o_P(n^{-1}).$$

We thus have

$$\begin{aligned} Q_n(z) &= \sum_{t=1}^{n} \rho_\alpha\left(\eta_t - \xi_\alpha - \eta_t D_t'(\hat{\theta}_n - \theta_0) - \frac{z}{\sqrt{n}} + o_P(n^{-1/2})\right) \\ &\quad - \rho_\alpha(\eta_t - \xi_\alpha) \\ &= zX_n + Y_n + I_n(z) + J_n(z) \end{aligned}$$

where

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

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

By the change of variable $u = s - z/\sqrt{n}$, we have $J_n(z) = \sum_{i=1}^2 J_n^{(i)}(z)$ with $J_n^{(i)}(z) = \sum_{t=1}^n J_{n,t}^{(i)}$ where

$$\begin{aligned} J_{n,t}^{(1)} &= \int_0^{R_{t,n}/\sqrt{n}} \left(\mathbf{1}_{\{\eta_t - \xi_\alpha - z/\sqrt{n} \le u\}} - \mathbf{1}_{\{\eta_t - \xi_\alpha - z/\sqrt{n} < 0\}} \right) du, \\ J_{n,t}^{(2)} &= \int_0^{R_{t,n}/\sqrt{n}} \left(\mathbf{1}_{\{\eta_t - \xi_\alpha - z/\sqrt{n} < 0\}} - \mathbf{1}_{\{\eta_t - \xi_\alpha < 0\}} \right) du \\ &= \left\{ D_t'(\hat{\theta}_n - \theta_0) + o_P(n^{-1/2}) \right\} \eta_t \mathbf{1}^*_{\{\eta_t - \xi_\alpha \in (0, z/\sqrt{n})\}}. \end{aligned}$$

By arguments already used, it follows that

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

Note that, for $z > 0$,

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

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

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

It follows that

$$\sum_{t=1}^n J_{n,t}^{(2)} = z \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1).$$

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

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

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 $J_{n,t}^{(2)}$ it can therefore be shown that $J_n^{(1)}(z)$ converges in distribution to a random variable which does not depend on z. Note also that Y_n can be subtracted to the objective function $Q_n(z)$ because it does not depend on z. Moreover $I_n(z) \to \frac{z^2}{2} f(\xi_\alpha)$ in probability as $n \to \infty$. Finally,

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

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

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

We now derive the joint asymptotic distribution of $(\sqrt{n}(\hat{\theta}_n - \theta_0)', \sqrt{n}(\xi_\alpha - \xi_{n,\alpha}))$. The following Taylor expansion holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-J^{-1}}{2\sqrt{n}} \sum_{t=1}^{n} (1 - \eta_t^2) D_t + o_P(1).$$

Hence

$$\mathrm{Cov}_{as}\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbf{1}_{\{\eta_t < \xi_\alpha\}} - \alpha)\right) = \frac{1}{2} p_\alpha J^{-1} \Omega$$

and thus

$$\begin{aligned}&\mathrm{Var}_{as}\{\sqrt{n}(\xi_{n,\alpha} - \xi_\alpha)\} = \left\{\xi_\alpha^2 \frac{\kappa_4 - 1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)}\right\} \Omega' J^{-1} \Omega + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}, \\ &\mathrm{Cov}_{as}\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \sqrt{n}(\xi_\alpha - \xi_{n,\alpha})\right) = \lambda_\alpha J^{-1} \Omega.\end{aligned}$$

We have $\Omega' J^{-1} \Omega = 1$ and thus we obtain

$$\mathrm{Var}_{as}\{\sqrt{n}(\xi_\alpha - \xi_{n,\alpha})\} = \zeta_\alpha.$$

By the CLT for martingale differences, we get the announced result. □

Proof of Corollary 4.2. The asymptotic normality of the two-step estimator follows from Theorem 4.2 and the following Taylor expansion of H around $(\theta_0, -\xi_\alpha)$

$$\sqrt{n}\left(\hat{\theta}_{n,\alpha}^{2S} - \theta_{0,\alpha}\right) = \left[\frac{\partial H(\theta, \xi)}{\partial(\theta', \xi)}\right]_{(\theta_0, -\xi_\alpha)} \begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \\ \sqrt{n}(\xi_\alpha - \xi_{n,\alpha}) \end{pmatrix} + o_P(1).$$

A similar expansion holds for $\sqrt{n}\left(\hat{\theta}_{n,\alpha}^{S2S} - \theta_{0,\alpha}\right)$. □

Proof of Corollary 4.3. For the standard GARCH model, we have $J^{-1}\Omega = 2\overline{\theta}_0$, and the asymptotic variance in Theorem 4.2 takes the form

$$\Sigma_\alpha = \begin{pmatrix} \frac{\kappa_4 - 1}{4} J^{-1} & 2\lambda_\alpha \overline{\theta}_0 \\ 2\lambda_\alpha \overline{\theta}_0' & \zeta_\alpha \end{pmatrix}.$$

Moreover,

$$H(\theta, \xi) = \xi^2 \overline{\theta}' + (0'_{[1,q+1]}, \beta_1, \ldots, \beta_p)', \quad \frac{\partial H(\theta, \xi)}{\partial(\theta', \xi)} = [A \quad 2\xi\overline{\theta}].$$

Therefore

$$\begin{aligned}\Upsilon_\alpha &= \frac{\kappa_4-1}{4}AJ^{-1}A+4\xi_\alpha^2(2\lambda_\alpha\xi_\alpha+\zeta_\alpha)\overline{\theta}_0\overline{\theta}_0' \\ &= \frac{\kappa_4-1}{4}AJ^{-1}A+4\xi_\alpha^2\left(\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}-\xi_\alpha^2\frac{\kappa_4-1}{4}\right)\overline{\theta}_0\overline{\theta}_0' \\ &= \frac{\kappa_4-1}{4}A\{J^{-1}-4\overline{\theta}_0\overline{\theta}_0'\}A+4\xi_\alpha^2\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}\overline{\theta}_0\overline{\theta}_0'.\end{aligned}$$

Similarly

$$\begin{aligned}\tilde{\Upsilon}_\alpha &= \frac{\kappa_4-1}{4}AJ^{-1}A+4\xi_\alpha^2\left(-\xi_\alpha^2\frac{\kappa_4-1}{4}+\frac{2\alpha(1-2\alpha)}{4f^2(\xi_\alpha)}\right)\overline{\theta}_0\overline{\theta}_0' \\ &= \frac{\kappa_4-1}{4}A\{J^{-1}-4\overline{\theta}_0\overline{\theta}_0'\}A+\frac{2\alpha(1-2\alpha)}{\xi_\alpha^2 f^2(\xi_\alpha)}A\overline{\theta}_0\overline{\theta}_0'A.\end{aligned}$$

□

A.5. Proofs for the results of Section 5.1

Proof of Theorem 5.1. We have $\breve{\theta}_{n,\alpha}=\arg\min_{\theta\in\Theta}\mathcal{Q}_{n,\tau}(\theta)$ where

$$\mathcal{Q}_{n,\tau}(\theta)=\frac{1}{n}\sum_{t=1}^n q_{t,\tau}(\theta),\quad q_{t,\tau}(\theta)=\rho_\tau\left\{\log\left(\frac{\epsilon_t^-}{\sigma_t(\theta)}\right)\right\}\mathbf{1}_{\{\epsilon_t<0\}}.$$

We follow the line of proof of Theorem 3.1, replacing the point *ii*) by

$$\begin{aligned} &ii') \qquad \lim_{\tau\to\tau_0}\mathbb{E}\sup_{\theta\in\Theta}|q_{1,\tau}(\theta)-q_{1,\tau_0}(\theta)|=0; \\ &ii'') \qquad \mathbb{E}q_{1,\tau_0}(\theta)>\mathbb{E}q_{1,\tau_0}(\theta_0),\quad \forall\theta\neq\theta_0.\end{aligned}$$

The point *ii′*) follows from **A13**, since

$$q_{1,\tau}(\theta)-q_{1,\tau_0}(\theta)=(\tau-\tau_0)\log\left(\frac{\epsilon_1^-}{\sigma_1(\theta)}\right)\mathbf{1}_{\{\epsilon_1<0\}}.$$

Note that when $\eta_1<0$ we have

$$\rho_\tau\left\{\log\left(\frac{\epsilon_1^-}{\sigma_1(\theta)}\right)\right\}-\rho_\tau\left(\log\eta_1^-\right)=d\left(\eta_1^-,\frac{\sigma_1(\theta)}{\sigma_1(\theta_0)}\right),$$

where $d(\eta_1^-,\sigma)=(\log\eta_1^-)(\mathbf{1}_{\{\eta_1^-\le 1\}}-\mathbf{1}_{\{\eta_1^-\le\sigma\}})-(\log\sigma)(\tau-\mathbf{1}_{\{\eta_1^-\le\sigma\}})$. Since $\tau=E\left(\mathbf{1}_{\{\eta_1^-\le 1\}}\mid\eta_1<0\right)$, we have

$$E\left(d(\eta_1^-,\sigma)\mid\eta_1<0\right)=E\left\{(\log\eta_1^--\log\sigma)\left(\mathbf{1}_{\{\eta_1^-\le 1\}}-\mathbf{1}_{\{\eta_1^-\le\sigma\}}\right)\mid\eta_1<0\right\}$$

which is strictly positive when $\sigma\neq 1$. Point *ii")* follows. The rest of the proof is similar to that of Theorem 3.1. □

A.6. Proofs for the results of Section 5.2

Proof of Theorem 5.2.

We have already shown that

$$\begin{aligned}\hat{\eta}_t &= \eta_t - \eta_t D_t'(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' \frac{\partial^2 \eta_t(\theta^*)}{\partial\theta\partial\theta'}(\hat{\theta}_n - \theta_0) + o_P(n^{-1}) \qquad \text{(A.6)}\\ &:= \eta_t + v_{t,n}\end{aligned}$$

and

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

We have

$$\sqrt{n}(\mu_{n,\alpha} - \mu_\alpha) = -\frac{1}{\alpha}\frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\eta}_t \mathbf{1}_{\hat{\eta}_t < \xi_{n,\alpha}} + \alpha\mu_\alpha) + o_P(1)$$

and

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\eta}_t \mathbf{1}_{\hat{\eta}_t < \xi_{n,\alpha}} + \alpha\mu_\alpha) - \frac{1}{\sqrt{n}}\sum_{t=1}^n (\eta_t - \xi_\alpha)\mathbf{1}_{\eta_t < \xi_\alpha} - \frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\eta}_t - \eta_t)\mathbf{1}_{\eta_t < \xi_\alpha}\\ =\; & V_n + \frac{1}{\sqrt{n}}\sum_{t=1}^n \xi_\alpha \mathbf{1}_{\hat{\eta}_t < \xi_{n,\alpha}} + \sqrt{n}\alpha\mu_\alpha = V_n + \sqrt{n}\alpha\{\xi_\alpha + \mu_\alpha\},\end{aligned}$$

where

$$V_n := \frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\eta}_t - \xi_\alpha)\left(\mathbf{1}_{\hat{\eta}_t < \xi_{n,\alpha}} - \mathbf{1}_{\eta_t < \xi_\alpha}\right) = o_P(1).$$

The later equality can be established following the lines of proof of Lemma 2 in Chen (2008). Note also that, by (A.6) and the ergodic theorem,

$$-\frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\eta}_t - \eta_t)\mathbf{1}_{\eta_t < \xi_\alpha} \overset{o_P(1)}{=} \sqrt{n}(\hat{\theta}_n - \theta_0)\frac{1}{n}\sum_{t=1}^n \eta_t D_t' \mathbf{1}_{\eta_t < \xi_\alpha} \overset{o_P(1)}{=} -\mu_\alpha \alpha \Omega' \sqrt{n}(\hat{\theta}_n - \theta_0),$$

writing $a \overset{c}{=} b$ for $a = b + c$.

It follows that

$$\begin{aligned}\sqrt{n}(\mu_{n,\alpha} - \mu_\alpha) &\overset{o_P(1)}{=} -\frac{1}{\alpha\sqrt{n}}\sum_{t=1}^n \{(\eta_t - \xi_\alpha)\mathbf{1}_{\eta_t < \xi_\alpha} + \alpha(\xi_\alpha + \mu_\alpha)\} - \mu_\alpha \Omega' \sqrt{n}(\hat{\theta}_n - \theta_0),\\ \sqrt{n}(\hat{\theta}_n - \theta_0) &\overset{o_P(1)}{=} \frac{-J^{-1}}{2\sqrt{n}}\sum_{t=1}^n (1 - \eta_t^2) D_t.\end{aligned}$$

We have

$$\begin{aligned}\mathrm{var}\left\{\frac{1}{\alpha\sqrt{n}}\sum_{t=1}^n (\eta_t - \xi_\alpha)\mathbf{1}_{\eta_t < \xi_\alpha}\right\} &= \sigma_\alpha^2,\\ \mathrm{var}_{as}\left\{\mu_\alpha \Omega' \sqrt{n}(\hat{\theta}_n - \theta_0)\right\} &= \frac{\kappa_4 - 1}{4}\mu_\alpha^2\\ \mathrm{cov}_{as}\left\{\frac{1}{\alpha\sqrt{n}}\sum_{t=1}^n (\eta_t - \xi_\alpha)\mathbf{1}_{\eta_t < \xi_\alpha}, \mu_\alpha \Omega' \sqrt{n}(\hat{\theta}_n - \theta_0)\right\} &= -\frac{\mu_\alpha}{2} x_\alpha.\end{aligned}$$

The asymptotic distribution of $(\sqrt{n}(\hat{\theta}_n - \theta_0)', \sqrt{n}(\mu_{n,\alpha} - \mu_\alpha))$ follows.

Finally, we turn to the asymptotic variance of the ES-parameter in the standard GARCH. We have

$$H(\theta,\mu) = \mu^2\overline{\theta}' + (0'_{[1,q+1]}, \beta_1, \ldots, \beta_p)', \quad \frac{\partial H(\theta,\mu_\alpha)}{\partial(\theta',\mu)} = [A^* \quad 2\mu_\alpha\overline{\theta}]$$

where $A^* = \begin{pmatrix} \mu_\alpha^2 I_{q+1} & 0 \\ 0 & I_p \end{pmatrix}$. Therefore the asymptotic variance of $H(\hat{\theta}_n, -\mu_{n,\alpha})$ is

$$\begin{aligned} \Upsilon_\alpha^* &= \frac{\kappa_4 - 1}{4} A^* J^{-1} A^* + 4\mu_\alpha^2(2\varphi_\alpha\mu_\alpha + \nu_\alpha)\overline{\theta}_0\overline{\theta}_0' \\ &= \frac{\kappa_4 - 1}{4} A^* J^{-1} A^* + 4\mu_\alpha^2 \left(\sigma_\alpha^2 - \mu_\alpha^2 \frac{\kappa_4 - 1}{4}\right) \overline{\theta}_0\overline{\theta}_0' \\ &= \frac{\kappa_4 - 1}{4} A^* \{J^{-1} - 4\overline{\theta}_0\overline{\theta}_0'\} A^* + 4\mu_\alpha^2\sigma_\alpha^2\overline{\theta}_0\overline{\theta}_0'. \end{aligned}$$

References

Acerbi, C. and D. Tasche (2002) On the coherence of expected shortfall *Journal of Banking & Finance* 26, 1487–1503.

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.

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.

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

Breidt, F.J., Davis, R.A. and A. A. Trindade (2001) Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics* 29, 919–946.

Chen, S.X. (2008) Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* 6, 87–107.

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

Davis, R.A. and W.T.M. Dunsmuir (1997) Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497.

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

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 (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Francq, C. and J-M. Zakoïan (2012) Optimal predictions of powers of conditionally heteroskedastic processes. Forthcoming in the *Journal of the Royal Statistical Society - Series B.*

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.

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

Ling, S. (2004) Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society B*, 66, 63–78.

Ling, S. (2005). Self-weighted LAD estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society B*, 67, 381–393.

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.

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.

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.