Qml inference for volatility models with covariates

Francq, Christian and Thieu, Le Quyen

Crest and Universté de Lille, Université Pierre et Marie Curie

March 2015

Qml inference for volatility models with covariates

Christian Francq   Le Quyen Thieu*

CREST and Université de Lille, Université Pierre et Marie Curie.

Abstract

The asymptotic distribution of the Gaussian quasi-maximum likelihood estimator (QMLE) is obtained for a wide class of asymmetric GARCH models with exogenous covariates. The true value of the parameter is not restricted to belong to the interior of the parameter space, which allows us to derive tests for the significance of the parameters. In particular, the relevance of the exogenous variables can be assessed. The results are obtained without assuming that the innovations are independent, which allows conditioning on different information sets. Monte Carlo experiments and applications to financial series illustrate the asymptotic results. In particular, an empirical study demonstrates that the realized volatility is an helpful covariate for predicting squared returns, but does not constitute an ideal proxy of the volatility.

Keywords: APARCH model augmented with explanatory variables, Boundary of the parameter space, Consistency and asymptotic distribution of the Gaussian quasi-maximum likelihood estimator, GARCH-X models, Power-transformed and Threshold GARCH with exogenous covariates.

1 Introduction

The GARCH-type models are of the form

$$\varepsilon_t = \sigma_t \eta_t, \tag{1}$$

*Corresponding author: Le Quyen Thieu, Université Pierre et Marie Curie, France. Telephone: (+33)7 81 95 25 89. E-mail: thieulequyen1411@gmail.com

where the squared volatility σ_t^2 is the best predictor of ε_t^2 given a certain information set $\mathcal{F}_{t-1}$ available at time t. More precisely, it is assumed that $E(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2 > 0$, or equivalently that $\sigma_t > 0$, $\sigma_t \in \mathcal{F}_{t-1}$ and $E(\eta_t^2 \mid \mathcal{F}_{t-1}) = 1$. For the usual GARCH models, $\mathcal{F}_{t-1}$ is simply the sigma-field generated by the past returns $\{\varepsilon_u, u < t\}$, and the volatility has a parametric form $\sigma_t = \sigma(\varepsilon_u, u < t; \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is a vector of parameters. It is however often the case that some extra information is available, under the form of a vector $\boldsymbol{x}_{t-1}$ of exogenous covariates, such as the daily volume of transactions, or high frequency intraday data, or even series of other returns. It is natural to try to take advantage of the extra information, in order to improve the prediction of the squares. To incorporate the information conveyed by $\{\boldsymbol{x}_u, u < t\}$ into $\mathcal{F}_{t-1}$, researchers have considered GARCH models augmented with additional explanatory variables, the so-called GARCH-X models, which are of the form $\sigma_t = \sigma(\varepsilon_u, \boldsymbol{x}_u, u < t; \boldsymbol{\vartheta}_0)$, where $\boldsymbol{\vartheta}_0$ is a vector of parameters including a parameter $\boldsymbol{\theta}_0$ specific to the past returns and a parameter $\boldsymbol{\pi}_0$ related to the exogenous covariates (see *e.g.* Engle and Patton (2001) and the references therein).

In practice, the difficulties are the choice of the parametric form (as illustrated by Bollerslev (2008), there exists a plethora of GARCH formulations) and the estimation of the parameter $\boldsymbol{\vartheta}_0$. The two problems are closely related. For GARCH, as well as for GARCH-X models, the coefficients are generally positively constrained, and tests of nullity of some components of $\boldsymbol{\vartheta}_0$ help to find a parsimonious GARCH-X formulation. The usual estimator of the GARCH models is the quasi-maximum likelihood estimator (QMLE), which does not require to specify a particular distribution for the error term η_t. The consistency of the QMLE does even not require that (η_t) be iid, which is particularly relevant for GARCH-X models (see Remarks 3 and 4 below). The asymptotic normality however requires that the true value of the parameter belongs to the interior of the parameter space, which is generally not the case when components of $\boldsymbol{\vartheta}_0$ are equal to zero.

Questions that seem particularly relevant in the GARCH-X framework are: is it really useful to introduce covariates in the volatility? which covariates should we add to $\mathcal{F}_{t-1}$? how many lagged values should we consider in the GARCH formulation? Some researchers and practitioners even reject any GARCH model, and consider that the realized volatility is a sufficiently good proxy of the volatility. In the GARCH-X framework, that leads to the following question: is it necessary to include the past returns $\{\varepsilon_u, u < t\}$ in the volatility when the sequence $(\boldsymbol{x}_t)$ of the realized volatilities is available?

Each of these questions can be discussed by testing the nullity of certain components of $\boldsymbol{\vartheta}_0$. It is thus of interest to study the behaviour of the estimator $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}_0$ when this parameter may stand at the boundary of the parameter space. To our knowledge, this problem has not yet been explicitly considered for GARCH-X models. This will be the focus of this paper. We now present the class of GARCH-X that we will consider, and then we detail the main objectives of the paper.

1.1 The model

Let $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. We consider the model defined by

$$\begin{cases} \varepsilon_t = h_t^{1/\delta}\eta_t \\ h_t = \omega_0 + \sum_{i=1}^q \alpha_{0i+}(\varepsilon_{t-i}^+)^\delta + \alpha_{0i-}(\varepsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j}h_{t-j} + \boldsymbol{\pi}_0'\boldsymbol{x}_{t-1} \end{cases} \tag{2}$$

where $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{r,t})'$ is a vector of r exogenous covariates. To ensure that $h_t > 0$ with probability one, assume that the covariates are almost surely positive and that the coefficients satisfy $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$, $\delta > 0$ and $\boldsymbol{\pi}_0 = (\pi_{01}, \dots, \pi_{0r})' \geq 0$ componentwise.

In absence of covariates, *i.e.* when $\boldsymbol{\pi}_0 = 0$, this equation corresponds to the Asymmetric Power GARCH (APARCH) model introduced by Ding et al. (1993). Model (2) can thus be called APARCH-X. The APARCH is rather general, since it nests numerous ARCH-type parameterizations used by the practitioners. The standard GARCH is obtained with $\delta = 2$ and $\alpha_{0i-} = \alpha_{0i+}$. Motivated by the fact that the autocorrelations are often larger for $|\varepsilon_t|$ than for ε_t^2, Taylor (1986) proposed the model with $\delta = 1$ and $\alpha_{0i-} = \alpha_{0i+}$. When $\alpha_{0i-} > \alpha_{0i+}$, a negative return has a higher impact on the future volatility than a positive return of the same magnitude. This is a well-documented stylized fact that is called "leverage effect". Two widely used models that allow for the leverage effect are the TARCH of Zakoïan (1994), obtained with $\delta = 1$, and the GJR of (Glosten et al., 1993), obtained with $\delta = 2$. One popular ARCH formulation that is not nested by the APARCH is the EGARCH model of Nelson (1991). The inference of the EGARCH is however quite difficult, and the behaviour of the QMLE is still partially unknown for this model (see Wintenberger (2013)). Another exponential formulation that is not encompassed by (2) is the log-GARCH model (see Sucarrat and Escribano (2010)).

1.2 The objectives

The most comprehensible results concerning the inference of the APARCH model can be found in Pan et al. (2008) and in Hamadeh and Zakoïan (2011) (HZ hereafter).[1] To our knowledge, there exists no general result concerning the estimation of the APARCH-X model. Actually, even if practitioners often add exogenous variables to volatility models, the probabilistic properties and the statistical inference of ARCH models with exogenous variables have not been yet extensively studied. Notable exceptions are the papers of Han (2013), Han and Kristensen (2014) and Han and Park (2012, 2014), which studied the inference of the GARCH(1,1) model augmented by an additional covariate which can be persistent. A common assumption to all the references previously given in this section, is that the true value of the parameter belongs to the interior of the parameter space. Under this assumption, and other regularity conditions, the QMLE is asymptotically normally distributed. When the parameter belongs to the boundary of the parameter space, the asymptotic distribution of the QMLE may be non standard (see Andrews (2001) for a general reference, and Francq and Zakoïan (2007) for applications to GARCH models). An important consequence of the non normality of the QMLE is that the standard t-ratio or the Wald tests used to identify the order p and q are also non standard (see *e.g.* Francq and Zakoïan (2009) and the reference therein).

Our first objective is thus to study the asymptotic distribution of the QMLE of the APARCH-X model when the parameter is not restricted to belong to the interior of the parameter space. For the applications we have in mind, the covariates can be for instance lagged values of other squares returns, or realized volatilities, or positive and negative parts of relative volume increments. The covariates will be supposed to be positive and stationary, but they are allowed to be strongly correlated, and also correlated with η_t. Therefore, the covariates will not be weakly or strongly exogenous in the sense of Engle et al. (1983), but we can say that the $x_{i,t}$'s are exogenous variables in the sense that their

[1] Note that the APARCH model is called Power-Transformed Threshold GARCH in these two papers

dynamics is not specified by the APARCH-X model.

Our second objective is to propose tests of nullity for one or several components of $\boldsymbol{\vartheta}_0$. This is closely related to the first objective because, due to the positivity constraints on the components of $\boldsymbol{\vartheta}_0$, under the null hypothesis, the true parameter stands at the boundary of the parameter space. This allows us to determine the asymptotic distribution of the QMLE.

The remainder of the paper is organized as follows. In Section 2, we first discuss the strict stationarity. We then introduce the Gaussian quasi-maximum likelihood estimator for APARCH-X model (2) and derive conditions for its consistency. The asymptotic distribution of the QMLE is studied conditioning on different information sets. We also consider the problem of testing the nullity of certain coefficients. The simulation results and two real data applications are presented in Section 3. Section 4 concludes the paper. All the proofs are collected in Section 5.

2 Main results

We first discuss the strict stationarity, which will be the main condition for the consistency of the QMLE.

2.1 Strict stationarity

Assuming that $p \geq 2$ and $q \geq 2$, let the vector of dimension $2q + p - 2$

$$\boldsymbol{Y}_t = \left(h_{t+1}, \ldots, h_{t-p+2}, \left(\varepsilon_t^+\right)^\delta, \left(\varepsilon_t^-\right)^\delta, \ldots, \left(\varepsilon_{t-q+2}^+\right)^\delta, \left(\varepsilon_{t-q+2}^-\right)^\delta\right)'.$$

It is easy to see that (ε_t) satisfies (2) if and only if

$$\boldsymbol{Y}_t = \boldsymbol{C}_{0t}\boldsymbol{Y}_{t-1} + \boldsymbol{B}_{0t}, \tag{3}$$

where $\boldsymbol{B}_{0t} = (\omega_0 + \boldsymbol{\pi}_0'\boldsymbol{x}_t, 0, \ldots, 0)'$ is a vector of dimension $2q+p-2$ and $\boldsymbol{C}_{0t}$ is a matrix depending on $(\eta_t^+)^\delta$, $(\eta_t^-)^\delta$ and

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0', \boldsymbol{\pi}_0')', \qquad \boldsymbol{\theta}_0 = (\omega_0, \alpha_{01+}, \alpha_{01-}, \ldots, \alpha_{0q+}, \alpha_{0q-}, \beta_{01}, \ldots, \beta_{0p})'.$$

The explicit form of $\boldsymbol{C}_{0t}$ can be found on page 507 in HZ. By modifying slightly the definitions of $\boldsymbol{Y}_t$ and $\boldsymbol{C}_{0t}$, we still have the representation (3) when $p < 2$ or $q < 2$.

Now assume that

A1: $(\eta_t, \boldsymbol{x}_t')$ is a strictly stationary and ergodic process, and there exists $s > 0$ such that $E|\eta_1|^s < \infty$ and $E\|\boldsymbol{x}_1\|^s < \infty$.

Note that, for GARCH-type models of the form (2), the sequence (η_t) is usually assumed to be a white noise, but this assumption is not necessary. Following Brandt (1986) and Bougerol and Picard (992b), the stationarity relies on the top Lyapunov

$$\gamma := \lim_{t\to\infty} \frac{1}{t} \log \|\boldsymbol{C}_{0t}\boldsymbol{C}_{0,t-1}\cdots\boldsymbol{C}_{01}\| \qquad \text{a.s.},$$

which is well defined in $[-\infty, +\infty)$ because $E\log^+\|\boldsymbol{C}_{01}\| < \infty$ under the condition $E|\eta_1|^s < \infty$ (see (A.5) in Pan et al. (2008)). It is showed in the previous reference that when (η_t) is iid and satisfies some regularity conditions, there exists a unique strictly stationary solution to the APARCH model if and only if $\gamma < 0$. The following lemma shows that the condition is the same for the APARCH-X.

Lemma 1 *Suppose that* **A1** *is satisfied. If $\gamma < 0$, the APARCH-X equation* (2) *(or equivalently* (3)*) admits a unique strictly stationary, non anticipative and ergodic solution. The solution of* (3) *is given by*

$$\boldsymbol{Y}_t = \boldsymbol{B}_{0t} + \sum_{k=1}^{\infty}\left(\prod_{i=1}^{k}\boldsymbol{C}_{0,t-i-1}\right)\boldsymbol{B}_{0,t-k}. \tag{4}$$

When $\gamma \geq 0$, there exists no stationary solution to (2) *and to* (3)*.*

Remark 1 *In the case $p = q = 1$, the top Lyapunov takes the explicit form*

$$\gamma = E\log\{\alpha_{0+}\left(\eta_1^+\right)^\delta + \alpha_{0-}\left(\eta_1^-\right)^\delta + \beta_0\} \tag{5}$$

with the simplified notations $\alpha_{0+} = \alpha_{01+}$, $\alpha_{0-} = \alpha_{01-}$ and $\beta_0 = \beta_{01}$. Under **A1** *and $\gamma < 0$, the volatility is given by*

$$h_t = \sum_{k=0}^{\infty}\prod_{i=1}^{k} a\left(\eta_{t-i}\right)\varpi_{t-k-1}, \tag{6}$$

with $a(z) = \alpha_{0+}(z^+)^\delta + \alpha_{0-}(z^-)^\delta + \beta_0$, the convention $\prod_{i=1}^{k} a\left(\eta_{t-i}\right) = 1$ when $k = 0$, and $\varpi_t = \omega_0 + \boldsymbol{\pi}_0'\boldsymbol{x}_t$. The stationary solution of the APARCH-X model is

$$\varepsilon_t = \left\{\sum_{k=0}^{\infty}\prod_{i=1}^{k} a\left(\eta_{t-i}\right)\varpi_{t-k-1}\right\}^{1/\delta}\eta_t. \tag{7}$$

Remark 2 *It has to be noted that the strict stationarity condition $\gamma < 0$ given in Lemma 1 does not involve the exogenous variables $\boldsymbol{x}_t$. Taking $\boldsymbol{x}_t = \varepsilon_t$ is not forbidden, but of course Assumption* **A1** *entails that $(\boldsymbol{x}_t)$ is stationary, and in this case, the lemma becomes trivial.*

2.2 Strong consistency of the QMLE

Hamadeh and Zakoïan (2011) showed that, for APARCH models, the power parameter δ is difficult to be estimated in practice. The quasi-likelihood being very flat in the direction of δ, estimating this parameter leads to imprecise results and considerably slows down the optimization routine. We therefore consider that δ is fixed. In many applications, $\delta = 1$ (as in the TARCH) or $\delta = 2$ (as in the GJR model). Let $d = 2q + p + r + 1$ be the remaining number of unknown parameters. A generic element of the parameter space $\Theta \subseteq (0, +\infty) \times [0, +\infty)^{d-1}$ is denoted by

$$\boldsymbol{\vartheta} = \left(\omega, \alpha_{1+}, \alpha_{1-}, \ldots, \alpha_{q+}, \alpha_{q-}, \beta_1, \ldots, \beta_p, \boldsymbol{\pi}'\right)'.$$

Let $(\varepsilon_1, \ldots, \varepsilon_n)$ be a realization of length n of the stationary solution (ε_t) to the APARCH-X model (2), and let $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be the corresponding observations of the exogenous variables. Given initial values $\varepsilon_{1-q}, \ldots, \varepsilon_0$, $\widetilde{\sigma}_{1-p} \geq 0, \ldots, \widetilde{\sigma}_0 \geq 0$, $\boldsymbol{x}_0 \geq 0$, the Gaussian quasi-likelihood is given by

$$L_n(\boldsymbol{\vartheta}) = L_n(\boldsymbol{\vartheta}, \varepsilon_1, \ldots, \varepsilon_n, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\widetilde{\sigma}_t^2}} \exp\left\{\frac{-\varepsilon_t^2}{2\widetilde{\sigma}_t^2}\right\}$$

where the $\widetilde{\sigma}_t$ are defined recursively, for $t \geq 1$, by

$$\widetilde{\sigma}_t^{\delta} = \widetilde{\sigma}_t^{\delta}(\boldsymbol{\vartheta}) = \omega + \sum_{i=1}^{q} \alpha_{i+}\left(\varepsilon_{t-i}^{+}\right)^{\delta} + \alpha_{i-}\left(\varepsilon_{t-i}^{-}\right)^{\delta} + \sum_{j=1}^{p} \beta_j \widetilde{\sigma}_{t-j}^{\delta} + \boldsymbol{\pi}'\boldsymbol{x}_{t-1}.$$

The QMLE of $\boldsymbol{\vartheta}_0$ is defined as any measurable solution $\widehat{\boldsymbol{\vartheta}}_n$ of

$$\widehat{\boldsymbol{\vartheta}}_n = \underset{\boldsymbol{\vartheta}\in\Theta}{\arg\max}\, L_n(\boldsymbol{\vartheta}) = \underset{\boldsymbol{\vartheta}\in\Theta}{\arg\min}\, \widetilde{Q}_n(\boldsymbol{\vartheta}) \tag{8}$$

where

$$\widetilde{Q}_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\ell}_t, \qquad \widetilde{\ell}_t = \widetilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} + \ln \widetilde{\sigma}_t^2. \tag{9}$$

Let $\mathcal{A}_{\boldsymbol{\vartheta}+}(z) = \sum_{i=1}^{q} \alpha_{i+} z^i$, $\mathcal{A}_{\boldsymbol{\vartheta}-}(z) = \sum_{i=1}^{q} \alpha_{i-} z^i$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$. To show the strong consistency of the QMLE, we need the following assumptions.

A2: $E\left(\eta_t \mid \mathcal{F}_{t-1}\right) = 0$ and $E\left(\eta_t^2 \mid \mathcal{F}_{t-1}\right) = 1$, where $\mathcal{F}_{t-1}$ denotes the σ-field generated by $\{\varepsilon_u, \boldsymbol{x}_u, u < t\}$.

A3: $\boldsymbol{\vartheta}_0 \in \Theta$, Θ is compact.

A4: for all $i \geq 1$, the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i}$, where $\mathcal{F}_{t,i}$ is a $\sigma-$field generated by $\{\eta_{t-j}, j > i, \boldsymbol{x}_{t-k}, k > 0\}$, is not included in $[0, \infty)$ or in $(-\infty, 0]$ and contains at least three points.

Assumption **A4** is an identifiability condition which prevents taking redundant explanatory variables in the volatility, for instance $\boldsymbol{x}_{t-1} = \left(\varepsilon_{t-i}^+\right)^\delta$ (see Remark 5 below).

A5: $\gamma < 0$ and $\sum_{j=1}^p \beta_j < 1$ for all $\boldsymbol{\vartheta} \in \Theta$.

A6: there exists $s > 0$, such that $Eh_t^s < \infty$ and $E|\varepsilon_t|^s < \infty$.

A7: if $p > 0$, $\mathcal{B}_{\boldsymbol{\vartheta}_0}(z)$ has no common root with $\mathcal{A}_{\boldsymbol{\vartheta}_0+}(z)$ and $\mathcal{A}_{\boldsymbol{\vartheta}_0-}(z)$; $\mathcal{A}_{\boldsymbol{\vartheta}_0+}(1) + \mathcal{A}_{\boldsymbol{\vartheta}_0-}(1) \neq 0$ and $\alpha_{0q+} + \alpha_{0q-} + \beta_{0p} \neq 0$ (with the notation $\alpha_{00+} = \alpha_{00-} = \beta_{00} = 1$).

A8: If $\boldsymbol{d}$ is a non zero vector of $\mathbb{R}^r$ then $\boldsymbol{d}'\boldsymbol{x}_1$ is not degenerated.

Assumptions **A3**, **A5** and **A7** have already been used to show the consistency of the QMLE for GARCH models. Assumption **A8** is an identifiability condition which is obviously necessary to avoid multicollinearity of the explanatory variables. The following remarks concern respectively **A2** and **A4**.

Remark 3 *Assumptions* **A1** *and* **A2** *entail that* $(\eta_t, \mathcal{F}_t)$ *is a conditionally homoscedastic martingale difference. For the GARCH-type models, it is usual to assume the stronger assumption that* (η_t) *is iid* $(0, 1)$*. Note, however, that Escanciano (2009) and Han and Kristensen (2014) employed* **A2***. The advantage of using* **A2** *is that* (2) *becomes a semi-strong model, that can be satisfied for different* σ*-fields* $\mathcal{F}_t$*, corresponding for example to different sequences of exogenous variables* $(\boldsymbol{x}_t)$*. Indeed,* **A2** *is satisfied for a model of the form* (1) *whenever* $E\left(\varepsilon_t \mid \mathcal{F}_{t-1}\right) = 0$ *and* $E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) = \sigma_t^2 > 0$*. With APARCH-X models, for which several information sets* $\mathcal{F}_t$ *can be naturally investigated, Assumption* **A2** *seems thus more flexible than the iid assumption.*

Remark 4 *Let us give an example of a data generated process for which several GARCH-X models of the form* (2) *coexist under the semi-strong noise Assumption* **A2**. *Assume that* $\boldsymbol{X}_t = (\varepsilon_t, y_t)'$ *follows the bivariate GARCH model* $\boldsymbol{X}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\eta}_t$, *where* $\boldsymbol{\Sigma}_t = diag\left(\sigma_{1,t}^2, \sigma_{2,t}^2\right)$ *with* $\boldsymbol{\eta}_t$ *iid* $\mathcal{N}(0, \boldsymbol{I}_2)$, *and* $\sigma_{i,t}^2 = \omega_i + \alpha_i\varepsilon_{t-1}^2 + \beta_i\sigma_{i,t-1}^2 + \pi_i y_{t-1}^2$ *for* $i = 1, 2$. *The process* (ε_t) *thus follows a (strong) GARCH-X(1,1) model with exogenous variable* $x_t = y_t^2$. *Nijman and Sentana (1996) showed that* (ε_t) *also follows a GARCH(2,2) model, without exogenous variable, but with a semi-strong noise satisfying* **A2**, *which is not independent in general.*

Remark 5 *Note that when there is no covariate and when* (η_t) *is iid, Assumption* **A4** *reduces to*

$$P[\eta_1 > 0] \in (0, 1) \text{ and the support of the distribution of } \eta_1 \text{ contains at least 3 points,}$$

which is exactly the identifiability condition **A2** *of HZ. When there exist covariates,* **A4** *rules out the existence of collinearities between the exogenous variables and the functions of the past returns involved in the volatility. For example, the assumption precludes that* $\boldsymbol{d}'\boldsymbol{x}_{t-1} = (\varepsilon_{t-i}^+)^\delta$ *with* $\boldsymbol{d} \in \mathbb{R}^r$ *(otherwise the variable* $(\eta_{t-i}^+)^\delta$ *given* $\mathcal{F}_{t,i}$ *would be degenerated, and thus almost surely equal to 0, which is impossible under* **A4***).*

The following lemma shows that **A6** can be suppressed when (η_t) is iid.

Lemma 2 *If* $\gamma < 0$ *and Assumptions* **A1**-**A2** *hold with* (η_t) *iid (0,1), then* **A6** *is satisfied.*

It will be convenient to approximate the sequence $\left(\widetilde{\ell}_t(\boldsymbol{\vartheta})\right)$ by an ergodic stationary sequence. Therefore, denote by $\left(\sigma_t^\delta\right)_t = \left\{\sigma_t^\delta(\boldsymbol{\vartheta})\right\}_t$ the strictly stationary, ergodic and non-anticipative solution of

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_{i+}\left(\varepsilon_{t-i}^+\right)^\delta + \alpha_{i-}\left(\varepsilon_{t-i}^-\right)^\delta + \sum_{j=1}^p \beta_j\sigma_{t-j}^\delta + \boldsymbol{\pi}'\boldsymbol{x}_{t-1}. \tag{10}$$

Note that $\sigma_t^\delta(\boldsymbol{\vartheta}_0) = h_t$. Let $Q_n(\boldsymbol{\vartheta})$ and ℓ_t be obtained by replacing $\widetilde{\sigma}_t^\delta$ with σ_t^δ in $\widetilde{Q}_n(\boldsymbol{\vartheta})$ and $\widetilde{\ell}_t$.

Theorem 1 *Let* $\widehat{\boldsymbol{\vartheta}}_n$ *be a sequence of QMLE satisfying* (8). *Then, under* **A1**–**A8**,

$$\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0 \text{ a.s. as } n \to \infty.$$

2.3 Asymptotic distribution of the QMLE

For the computation of (9), it is necessary to have $\widetilde{\sigma}_t(\boldsymbol{\vartheta}) > 0$ almost surely, for any $\boldsymbol{\vartheta} \in \Theta$. This is why the components of $\boldsymbol{\vartheta} \in \Theta$ are constrained to be non negative. More precisely, it can be assumed that, for $i = 2, \ldots, d$, the i-th section of Θ is $[0, K_i]$ with $K_i > 0$ (the first section being $[\underline{\omega}, \overline{\omega}]$ with $0 < \underline{\omega} < \overline{\omega}$). If Θ is of this form and is large enough (to avoid, for instance, that the i-th component $\boldsymbol{\vartheta}_{0i}$ of $\boldsymbol{\vartheta}_0$ be less than or equal to K_i), the following assumption is satisfied.

A9: $\mathcal{C} := \lim_{n\to\infty} \sqrt{n}(\Theta - \boldsymbol{\vartheta}_0) = \prod_{i=1}^{d} \mathcal{C}_i$, where $\mathcal{C}_i = [0, +\infty)$ when $\boldsymbol{\vartheta}_{0i} = 0$ and $\mathcal{C}_i = \mathbb{R}$ otherwise.

The set $\mathcal{C}$ will be called the local parameter space. This is a convex cone, which is equal to $\mathbb{R}^d$ if and only if $\boldsymbol{\vartheta}_0$ belongs to the interior of Θ, *i.e.* if all the components of $\boldsymbol{\vartheta}_0$ are non zero, under **A9**.

For standard GARCH models, without covariates and with (η_t) iid, note that η_t is independent of $\mathcal{F}_{t-1}$. In that situation, it is known that no moment condition on ε_t is needed for the consistency and asymptotic normality (CAN) of the QMLE when the GARCH parameter belongs to the interior of the parameter space, whereas moments conditions are required when the parameter stands at the boundary of the parameter space (see the example given in Section 3.1 of Francq and Zakoïan (2007)). When the model is semi-strong, *i.e.* when η_t is not independent of $\mathcal{F}_{t-1}$, stronger moment conditions will be required. We thus distinguish four cases:

Case A : η_t is independent of $\mathcal{F}_{t-1}$ and all the components of $\boldsymbol{\vartheta}_0$ are strictly positive;

Case B : η_t is independent of $\mathcal{F}_{t-1}$ and at least one component of $\boldsymbol{\vartheta}_0$ is equal to zero;

Case C : η_t is not independent of $\mathcal{F}_{t-1}$ and all the components of $\boldsymbol{\vartheta}_0$ are strictly positive;

Case D : η_t is not independent of $\mathcal{F}_{t-1}$ and at least one component of $\boldsymbol{\vartheta}_0$ is equal to zero.

For simplicity, these four cases are referred to respectively as strong in the interior, strong at the boundary, semi-strong in the interior and semi-strong at the boundary. We assume that

A10: $E\eta_t^4 < \infty$ in Cases A and B, and $E|\eta_t|^{4+\nu} < \infty$ for some $\nu > 0$ in Cases C and D.

A11: $E\left|\varepsilon_t\right|^{2\delta} < \infty$ and $E\left\|\boldsymbol{x}_t\right\|^2 < \infty$ in Case B, and $E\left|\varepsilon_t\right|^{2\delta+8\delta/\nu} < \infty$ and $E\left\|\boldsymbol{x}_t\right\|^{2+8/\nu} < \infty$ in Case D.

Under the previous assumptions, Lemma 4 in Section 5 below shows that the matrix

$$\boldsymbol{J} := E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right) = \frac{4}{\delta^2} E\left(\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} \frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right) \tag{11}$$

is positive definite. Let us thus consider the norm $\|\boldsymbol{x}\|_{\boldsymbol{J}}^2 = \boldsymbol{x}'\boldsymbol{J}\boldsymbol{x}$ and the scalar product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{J}} = \boldsymbol{x}'\boldsymbol{J}\boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. In the sense of this scalar product, the orthogonal projection of a vector $\boldsymbol{Z} \in \mathbb{R}^d$ on $\mathcal{C}$ is defined by

$$\boldsymbol{Z}^{\mathcal{C}} = \arg\inf_{\boldsymbol{C}\in\mathcal{C}} \|\boldsymbol{C} - \boldsymbol{Z}\|_{\boldsymbol{J}}$$

or equivalently by

$$\boldsymbol{Z}^{\mathcal{C}} \in \mathcal{C} \quad \text{and} \quad \left\langle \boldsymbol{Z} - \boldsymbol{Z}^{\mathcal{C}}, \boldsymbol{C} - \boldsymbol{Z}^{\mathcal{C}}\right\rangle_{\boldsymbol{J}} \leq 0, \quad \forall \boldsymbol{C} \in \mathcal{C}. \tag{12}$$

When $\boldsymbol{\vartheta}_0$ is allowed to lie at the boundary of the parameter space, we also need the following moment assumption.

A12: in Cases B and D, there exist Hölder conjugate numbers p and $q > 1$ such that

$$p^{-1} + q^{-1} = 1 \quad \text{and} \quad E|\varepsilon_t|^{2\delta q} < \infty, \quad E|\varepsilon_t|^{2p} < \infty, \quad E\|\boldsymbol{x}_t\|^{2q} < \infty.$$

Theorem 2 *Under the assumptions of Theorem 1 and* **A9**–**A12**, *as* $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \boldsymbol{Z}^{\mathcal{C}}, \text{ where } \boldsymbol{Z} \sim \mathcal{N}\left\{0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right\}, \tag{13}$$

$\boldsymbol{J}$ *is defined by* (11) *and*

$$\boldsymbol{I} = \frac{4}{\delta^2} E\left[\left\{E\left(\eta_t^4 \mid \mathcal{F}_{t-1}\right) - 1\right\} \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} \frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right].$$

Remark 6 *The previous theorem provides the asymptotic distribution of the QMLE in each of the Cases A-D. In all cases, Assumptions* **A1-A9** *are required. Note that, in Cases A and B, we have* $\boldsymbol{I} = (E\eta_1^4 - 1)\boldsymbol{J}$. *In Cases A and C, the local parameter space is* $\mathcal{C} = \mathbb{R}^d$, *and the asymptotic distribution of the QMLE is thus normal:*

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\left\{0, (E\eta_1^4 - 1)\boldsymbol{J}^{-1}\right\} \qquad \textit{in Cases A} \tag{14}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\left\{0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right\} \qquad \text{in Cases C.} \tag{15}$$

This result is obtained under the assumption that $E\eta_t^4 < \infty$ *in Case A and a slightly stronger condition in Case C (see* **A10***), but without moment condition on the observed process* ε_t*. When there is no covariate* ($r = 0$)*, we retrieve the results obtained by Francq (2004) in the GARCH case* ($\delta = 2$ *and* $\alpha_{0i+} = \alpha_{0i-}$) *and when* (η_t) *is iid, by Escanciano (2009) in the GARCH case when* (η_t) *is a conditionally homoscedastic martingale difference, and by HZ in the general APARCH case. In the presence of covariates,* (15) *allows to retrieve some of the results obtained by Han and Kristensen (2014) for the GARCH-X(1,1) model, under slightly different assumptions.*

When $\boldsymbol{\vartheta}_0$ *stands at the boundary of the parameter space (Cases B and D), it seems that there existed no result similar to* (13) *for GARCH models with covariates. It is however worth considering Cases B and D, in particular, because this gives the asymptotic distribution of the QMLE under the null that* $\boldsymbol{\pi}_0 = 0$*. When there is no covariate, note that* **A12** *is satisfied when* $E|\varepsilon_t|^6 < \infty$ *(by taking* $p = 3$ *and* $q = 3/2$*). We thus retrieve* (13) *under the conditions given by Francq and Zakoïan (2007) in the particular case of GARCH models with* (η_t) *iid and* $r = 0$ *(see also Andrews (1999) and the references therein for the boundary problem in a more general estimation framework). Even when* $r = 0$*, the authors are not aware of the existence of conditions entailing* (13) *for the general APARCH model, or even for the subclass of the GARCH model in Case D.*

The next proposition provides estimations for the matrices $\boldsymbol{I}$ and $\boldsymbol{J}$ required to apply Theorem 2. Assumption **A12** needs to be slightly reinforced as follow

A12': in Cases B and D, there exist Hölder conjugate numbers p and $q > 1$ such that

$$p^{-1} + q^{-1} = 1 \quad \text{and} \quad E|\varepsilon_t|^{2\delta q} < \infty, \quad E|\varepsilon_t|^{4p} < \infty, \quad E\|\boldsymbol{x}_t\|^{2q} < \infty.$$

Proposition 1 *Under the assumptions of Theorem 2 with* **A12** *replaced by* **A12'***, strongly consistent estimators of* $\boldsymbol{J}$ *and* $\boldsymbol{I}$ *are given by*

$$\widehat{\boldsymbol{J}}_n = \frac{4}{\delta^2}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)}\frac{\partial\widetilde{\sigma}_t^{\delta}(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}}\frac{\partial\widetilde{\sigma}_t^{\delta}(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}'}, \tag{16}$$

and

$$\widehat{\boldsymbol{I}}_n = \frac{4}{\delta^2}\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\eta}_t^4 - 1\right)\frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)}\frac{\partial\widetilde{\sigma}_t^{\delta}(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}}\frac{\partial\widetilde{\sigma}_t^{\delta}(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}'}. \tag{17}$$

with $\widehat{\eta}_t = \varepsilon_t/\widetilde{\sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)$.

Remark 7 *In Cases A and B, in view of Remark 6, the estimator defined by (17) can be replaced by*

$$\widehat{\boldsymbol{I}}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\widehat{\eta}_t^4 - 1\right)\widehat{\boldsymbol{J}}_n. \tag{18}$$

Theorem 2 and Proposition 1 allow to test if one or several GARCH coefficients are equal to zero, which is important for identifying the orders of the model and the relevant covariates. For simplicity, we concentrate on the case of testing the nullity of only one coefficient. Let $\boldsymbol{e}_k$ be the k-th element of the canonical basis of $\mathbb{R}^d$. We will test the hypothesis that the k-th element of $\boldsymbol{\vartheta}_0$ is equal to zero, assuming the other elements are positive:

$$H_0 : \boldsymbol{e}_k'\boldsymbol{\vartheta}_0 = 0 \text{ and } \boldsymbol{e}_\ell'\boldsymbol{\vartheta}_0 > 0 \quad \forall \ell \neq k \qquad \text{against} \qquad H_1 : \boldsymbol{e}_k'\boldsymbol{\vartheta}_0 > 0. \tag{19}$$

For this testing problem, the Student t-test statistic is defined by

$$t_n(k) = \frac{\boldsymbol{e}_k'\widehat{\boldsymbol{\vartheta}}_n}{\sqrt{\boldsymbol{e}_k'\widehat{\boldsymbol{\Sigma}}\boldsymbol{e}_k}}, \qquad \widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{J}}_n^{-1}\widehat{\boldsymbol{I}}_n\widehat{\boldsymbol{J}}_n^{-1}.$$

Denote by $\chi^2_\ell(\alpha)$ the α-quantile of the chi-squared distribution with ℓ degrees of freedom. As a corollary of Theorem 2 and Proposition 1, we obtain the following result.

Corollary 1 *Under the assumptions of Theorem 2, the test of rejection region*

$$\{t_n^2(k) > \chi^2_1(1 - 2\alpha)\}$$

has the asymptotic level α *under* H_0 *and is consistent under* H_1 *defined in (19).*

Remark 8 *Because the asymptotic distribution of the QMLE is non Gaussian under the null, the standard t-ratio test of rejection region* $\{t_n^2(k) > \chi^2_1(1 - \alpha)\}$ *would have the wrong asymptotic level* $\alpha/2$ *instead of* α.

3 Numerical illustrations

We now illustrate our asymptotic results on Monte Carlo simulations and on financial series of daily returns and volumes, as well as high frequency intraday data.

3.1 Simulation experiments

The aim of this section is to study the finite sample behavior of the QMLE and of tests of significance of the form (19), in the different frameworks corresponding to Cases A-D. We thus simulated the following TARCH-X(1,1) model with 2 lagged values of an exogenous variable

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t = \omega_0 + \alpha_{0+}(\varepsilon_{t-1}^+) + \alpha_{0-}(\varepsilon_{t-1}^-) + \beta_0 \sigma_{t-1} + \pi_{01} x_{t-1} + \pi_{02} x_{t-2}. \end{cases} \tag{20}$$

The exogenous variable x_t is assumed to be the stochastic volatility defined by

$$x_t = e^{y_t}, \quad y_t = a y_{t-1} + e_t \tag{21}$$

where (e_t) is independently and $\mathcal{N}(0,1)$ distributed. In Cases A and B, we also assume that η_t is iid $\mathcal{N}(0,1)$ and independent of (e_t). In Cases C and D, we assume that, given $\mathcal{F}_{t-1}$, the variable $\sqrt{\nu_t/(\nu_t-2)}\eta_t$ follows a Student distribution with ν_t degrees of freedom, where $\nu_t = 5 + x_{t-1}$. This specification of ν_t guarantees that **A10** is satisfied. Indeed, the fourth order moment exists because

$$E\eta_t^4 = E\left(\eta_t^4 \middle| \mathcal{F}_{t-1}\right) = E\left(\frac{3(\nu_t-2)}{\nu_t-4}\right) \leq E\left(3(\nu_t-2)\right) < \infty$$

and, by a similar argument, it can be shown that a moment of order larger than 4 also exists. We took the parameter $\boldsymbol{\vartheta}_0 = (0.046, 0.027, 0.092, 0.843, 0.089, \pi_{02})$ where $\pi_{02} = 0$ in Cases B and D (this value of parameter corresponds to the estimated value of the parameter for the series BA studied in Section 3.3 below) and $\pi_{02} = \pi_{01} = 0.089$ in Cases A and C. In each of the four cases, we simulated 500 independent replications of model (20) for the two sample sizes $n = 1,000$ and $n = 2,000$. To attenuate the effect of the initial values, the first 200 values of each simulation have been eliminated. Figures 1, 2, 3, 4 display the boxplots of the estimation errors of the QMLE corresponding to the four cases. As expected, the accuracy of the estimators always increases with n. It can be noted that the estimators are more accurate when the model is strong (Cases A and B) than when it is semi-strong (Cases C and D). The boxplots also display more frequent outliers in the semi-strong case. Also, in accordance with the asymptotic theory, the distribution of the errors is clearly non Gaussian, especially for the estimation of π_{02}, when the true value of the parameter stands at the boundary of the parameter space (Cases B and D). Table 1 gives the empirical frequencies of rejection of the hypotheses $\pi_{01} = 0$ and $\pi_{02} = 0$. The test of the null hypothesis $\pi_{01} = 0$ (which is false in the four cases) is more powerful in Cases A and B (corresponding to a strong model) than C and D (corresponding to a semi-strong model). This is not surprising since, as shown by the boxplots, the semi-strong model is less accurately estimated than the strong one. Less obviously, the test is slightly more powerful in Cases B and D (when $\pi_{02} = 0$) than in Cases A and C (when $\pi_{02} = 0.089$). Turning to the test of the null hypothesis $\pi_{02} = 0$ (which is true in Cases B and D), one can see that the type 1 errors are well controlled when $n = 2,000$. Indeed, when the nominal level is 5%, the empirical relative frequency of rejection over the 500 independent replications should vary between 2.2% and 6.6% with probability of approximately 95%. When the nominal level is 1%, it varies from 0.4% to 2.0% with the same probability. All the relative rejection frequencies displayed in Table 1 are within these 95% limits when $n = 2,000$.

Table 1: Relative frequencies (in %) of rejection of the assumptions that the first and second lagged values of the exogenous variable do not appear in the conditional variance

	$H_0^{\pi_{01}} : \pi_{01} = 0$				$H_0^{\pi_{02}} : \pi_{02} = 0$			
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$	
	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$
A	83.00	99.40	96.00	99.80	66.80	91.00	85.80	98.40
B	99.40	100.00	100.00	100.00	**3.60**	1.40	**7.60**	5.20
C	72.80	92.00	88.00	98.00	50.80	77.20	70.60	92.60
D	96.40	98.80	99.20	98.80	**2.20**	2.00	6.40	5.80

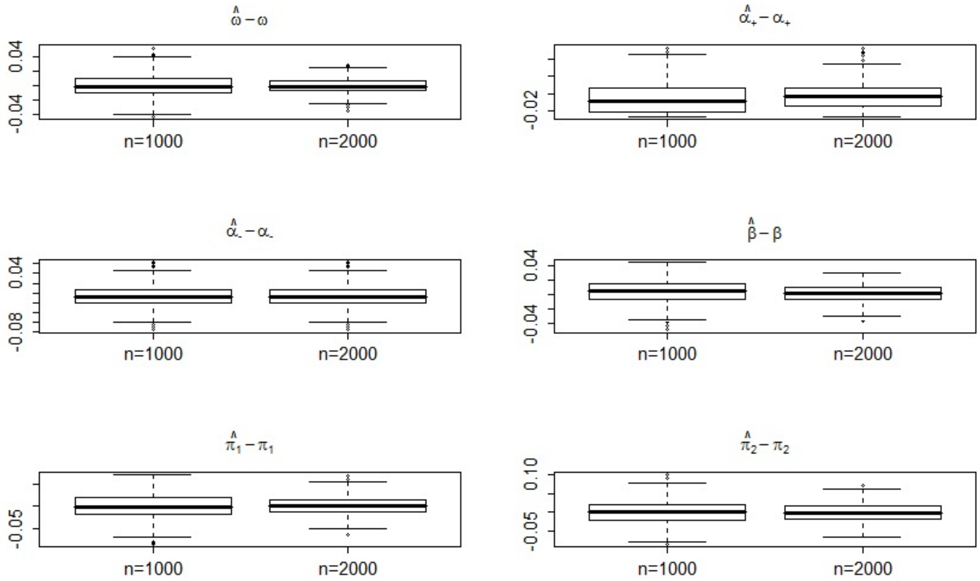

Figure 1: Boxplots of 500 estimation errors for the QMLE of the parameter ϑ_0 of a TARCH-X(1,1) in Case A (strong in the interior) for the two sample sizes $n = 1,000$ and $n = 2,000$.

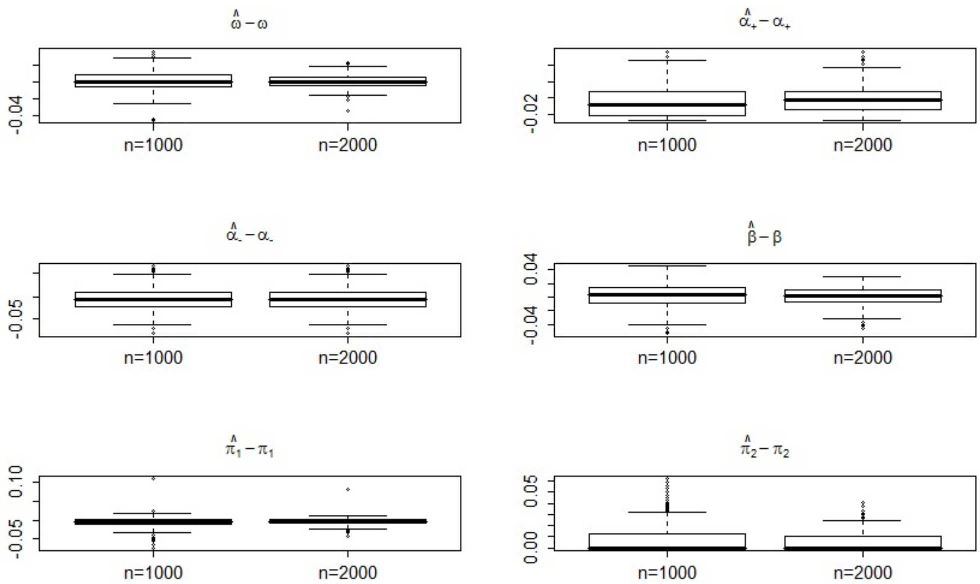

Figure 2: As Figure 1 but in Case B (strong at the boundary)

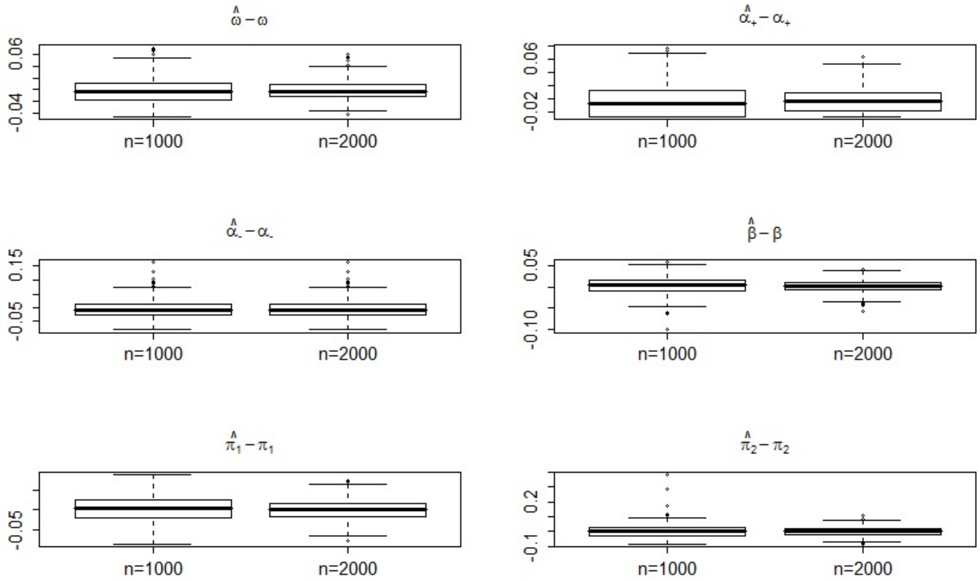

Figure 3: As Figure 1 but in Case C (semi-strong in the interior)

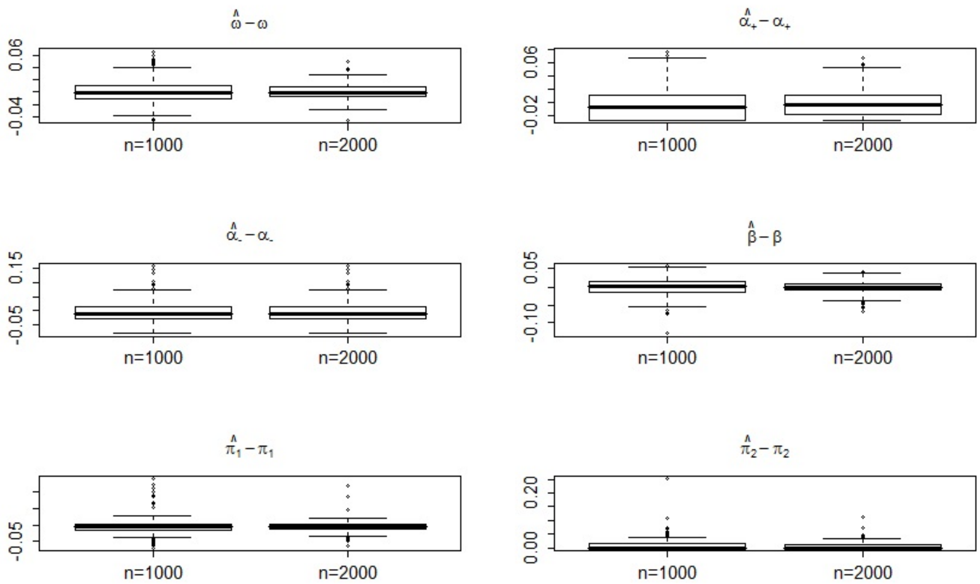

Figure 4: As Figure 1 but in Case D (semi-strong at the boundary)

3.2 SP500 with realized range, volume and other indices

In this section, we built a model which aims to explain the volatility of the daily returns of the SP500 index by its past values, the realized range, the volume and other stock returns. The data set has been downloaded from http://finance.yahoo.com/ and covers the period from January 4, 1985 to August 26, 2011. We considered the series of the relative range $rr_t = (high_t - low_t)/low_t$, where $high_t$ and low_t denote respectively the highest and lowest prices of the day. We also measured the relative volume by the formula $v_t = \left| \frac{vol_t}{\frac{1}{20}\sum_{i=1}^{20} vol_{t-i}} - 1 \right|$, where vol_t denotes the daily number of shares traded. We did not consider directly (vol_t) as covariate because this series is non stationary. The indicator v_t compares the present volume with the averaged volume over the past 20 days, which is a technique used by some traders. Figure 3.2 displays the series of the returns ε_t, the ranges rr_t and the relative volumes v_t, which look stationary. We also added the returns of the Nikkei, Nik_t, and of the FTSE, Ft_t, as potential explanatory variables for the SP500 volatility. We fitted APARCH-X(1,1) models with $\delta \in \{0.5, 1, 1.5, 2\}$. The model with the largest likelihood is obtained for $\delta = 1$, and is given by

$$\begin{cases} \varepsilon_t = & h_t \eta_t \\ h_t = & \underset{(0.006)\,0.002}{0.018} + \underset{(0.020)\,0.500}{0.000}\,\varepsilon_{t-1}^{+} + \underset{(0.035)\,0.001}{0.110}\,\varepsilon_{t-1}^{-} + \underset{(0.020)\,0.000}{0.879}\,h_{t-1} + \underset{(1.493)\,0.002}{4.331}\,rr_{t-1} \\ & + \underset{(0.026)\,0.010}{0.061}\,v_{t-1} + \underset{(0.007)\,0.500}{0.000}\,Nik_{t-1}^2 + \underset{(0.007)\,0.500}{0.000}\,Ft_{t-1}^2. \end{cases}$$

Under the estimated value of each coefficient, the estimated standard deviation is given into brackets, followed by the p-value of the test that the coefficient is equal to zero. One can see that the range rr_{t-1} and the volume v_{t-1} are significant covariates, whereas the returns Nik_{t-1}^2 and Ft_{t-1}^2 are not. This is in accordance with several empirical studies showing that the realized range, and to a lesser extent, the volume can help to predict the volatility (see *e.g.* Fuertes et al. (2009)). This is also consistent with other studies showing that the volatility spillover effects between stock markets are mainly instantaneous.

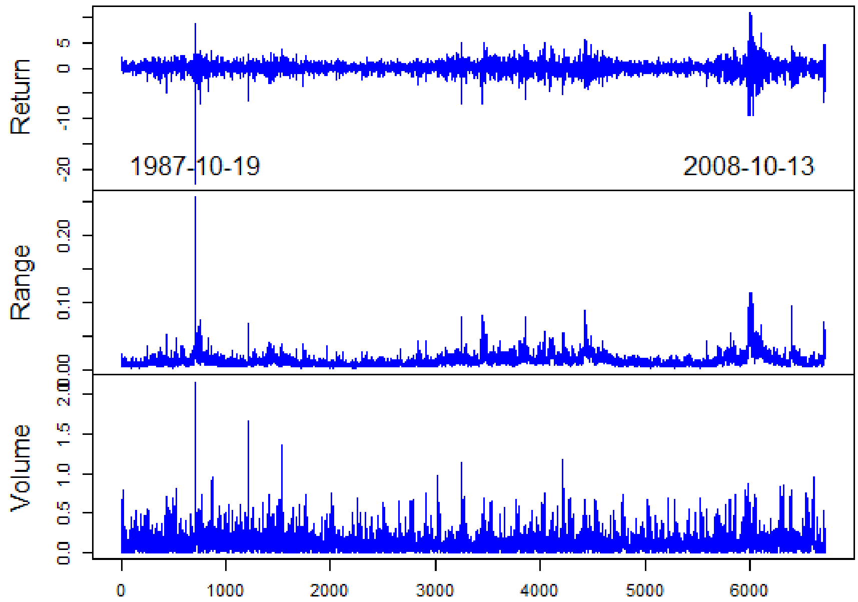

Figure 5: Return, range and relative volume of the SP500 index from January 4, 1985 to August 26, 2011 (October 19, 1987 corresponds to the black Monday, and October 13, 2008 corresponds to the beginning of the stock market crash of 2008).

3.3 US stocks with realized volatility

The data used in this section come from Section 4.2 of Laurent et al. (2014)[2] and concern 49 large capitalization stocks of american stock exchanges, covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days). At the end of each trading day t, the log-return in percentage ε_t and the realized volatility rv_t (computed as the sum of intraday squared 5-minute log-returns) are available.

The first question that we are interested in is whether the realized volatility is useful to predict the squared returns or not. More precisely, we would like to know how many lagged values of the realized volatility have to be considered in the volatility equation.

In order to answer this question, we estimated APARCH-X(1,1) models of the form

$$\begin{cases} \varepsilon_t = h_t^{1/\delta}\eta_t \\ h_t = \omega + \alpha_+(\varepsilon_{t-1}^+)^\delta + \alpha_-(\varepsilon_{t-1}^-)^\delta + \beta h_{t-1} + \pi_1\, rv_{t-1}^{\delta/2} + \pi_2\, rv_{t-2}^{\delta/2}, \end{cases} \tag{22}$$

with $\delta \in \{0.5, 1, 1.5, 2\}$. The variables rv_{t-1} and rv_{t-2} are raised to the power $\delta/2$ in order to have the same unit of measure for ε_t^2, the squared volatility $h_t^{2/\delta}$ and the realized volatility rv_t, regardless of δ. The selected value of δ is that which leads to the maximum value of the quasi-likelihood. Table 2 displays the fitted model on each of the 49 stocks. For all the estimated models, except 3 over the 49, one observes that $\alpha_- > \alpha+$, which is in accordance with the leverage effect (*i.e.* the fact that the volatility tends to increase more after a negative return than after a positive return of the same magnitude). We mostly find π_1 significantly non zero and π_2 close to zero. From this table, it is clear that yesterday's realized volatility often helps in predicting today's squared return.

Another question that we would like to investigate is whether the realized volatility is a good proxy of the volatility or not. Of course, the answer depends on what the precise meaning of "volatility" is. Here, we define the volatility as the best predictor of the squared return given all the information available $\mathcal{F}_{t-1}$, consisting in the past returns and the past realized volatilities. We thus consider the model

$$\begin{cases} \varepsilon_t = h_t^{1/\delta}\eta_t \\ h_t = \omega + \alpha_+(\varepsilon_{t-1}^+)^\delta + \alpha_-(\varepsilon_{t-1}^-)^\delta + \beta h_{t-1} + \pi_0\, rv_t^{\delta/2}. \end{cases} \tag{23}$$

[2]The authors are grateful to Sébastien Laurent who has kindly provided them with the data set.

Note that this model is considered for explanatory purposes only, but can not be used for predicting ε_t^2 since it involves the unavailable realized volatility at time t. The null hypothesis that the realized volatility rv_t is the best proxy of the volatility h_t can be formally written as

$$H_0: \quad \alpha_+ = \alpha_- = \beta = 0. \tag{24}$$

No need to use a formal test, the null hypothesis (24) is clearly rejected on all the estimated models for the 49 stocks (see Table 3), in particular, because the persistent parameter $\widehat{\beta}$ is always highly significant. From this study, we can draw the conclusion that the realized volatility is far from being an ideal proxy of the actual volatility. It is thus questionable to compare volatility forecasts with realized volatilities, a practice which is however becoming common in finance since the celebrated paper of Hansen and Lunde (2005).

Table 2: APARCH-X(1,1) models (22) fitted by QMLE on daily returns of US stock with two lagged values of realized volatilities as covariates. The estimated standard deviations are displayed into parentheses. For the estimated values of π_1 and π_2, one star (*) means a p-value $p \in [0.01, 0.05)$ for testing the nullity of the coefficient, two stars (**) means $p \in [0.001, 0.01)$, and three stars (***) means $p < 0.001$. The last column gives the selected value of the power δ.

	ω	α_+	α_-	β	π_1	π_2	δ
AAPL	0.080 (0.029)	0.042 (0.015)	0.055 (0.013)	0.796 (0.056)	0.120^{*} (0.072)	0.000 (0.079)	0.5
ABT	0.046 (0.027)	0.023 (0.024)	0.019 (0.027)	0.661 (0.100)	0.285^{***}(0.071)	0.000 (0.102)	0.5
AXP	0.027 (0.010)	0.000 (0.019)	0.074 (0.019)	0.809 (0.038)	0.155^{**} (0.061)	0.000 (0.069)	1
BA	0.046 (0.022)	0.027 (0.020)	0.092 (0.023)	0.843 (0.037)	0.084 (0.065)	0.000 (0.071)	2
BAC	0.007 (0.010)	0.009 (0.026)	0.090 (0.029)	0.813 (0.045)	0.151^{**} (0.063)	0.000 (0.083)	1
BMY	0.000 (0.020)	0.051 (0.018)	0.077 (0.025)	0.880 (0.031)	0.072 (0.106)	0.000 (0.108)	1
BP	0.017 (0.015)	0.010 (0.020)	0.043 (0.017)	0.682 (0.068)	0.191^{*} (0.086)	0.106 (0.092)	0.5
C	0.013 (0.011)	0.019 (0.025)	0.123 (0.028)	0.744 (0.056)	0.178^{**} (0.059)	0.007 (0.077)	1
CAT	0.045 (0.022)	0.000 (0.016)	0.011 (0.017)	0.780 (0.066)	0.183^{**} (0.074)	0.000 (0.087)	0.5
CL	0.127 (0.054)	0.032 (0.021)	0.212 (0.065)	0.424 (0.105)	0.156^{**} (0.057)	0.193^{**} (0.073)	2
CSCO	0.013 (0.015)	0.000 (0.020)	0.054 (0.020)	0.848 (0.034)	0.131 (0.081)	0.000 (0.084)	1
CVX	0.082 (0.030)	0.014 (0.022)	0.068 (0.025)	0.716 (0.071)	0.102^{*} (0.062)	0.088 (0.088)	2
DELL	0.000 (0.006)	0.030 (0.012)	0.055 (0.012)	0.874 (0.034)	0.093 (0.080)	0.000 (0.085)	0.5
DIS	0.042 (0.021)	0.000 (0.016)	0.070 (0.022)	0.807 (0.056)	0.132^{*} (0.063)	0.002 (0.069)	2
EK	0.182 (0.091)	0.071 (0.029)	0.111 (0.036)	0.583 (0.161)	0.224^{*} (0.103)	0.000 (0.155)	0.5
EXC	0.092 (0.033)	0.056 (0.032)	0.157 (0.039)	0.677 (0.060)	0.196^{**} (0.072)	0.000 (0.077)	1.5
F	0.060 (0.046)	0.068 (0.037)	0.075 (0.025)	0.740 (0.050)	0.091 (0.088)	0.103 (0.097)	1
FDX	0.033 (0.018)	0.012 (0.021)	0.025 (0.021)	0.803 (0.070)	0.162^{*} (0.074)	0.000 (0.081)	0.5
GE	0.005 (0.010)	0.000 (0.019)	0.052 (0.022)	0.802 (0.050)	0.180^{*} (0.085)	0.000 (0.076)	1
GM	0.026 (0.023)	0.018 (0.015)	0.047 (0.023)	0.881 (0.031)	0.094 (0.093)	0.000 (0.099)	2
HD	0.007 (0.010)	0.000 (0.015)	0.030 (0.013)	0.850 (0.037)	0.135^{*} (0.070)	0.000 (0.071)	0.5
HNZ	0.007 (0.009)	0.050 (0.018)	0.084 (0.022)	0.840 (0.038)	0.105^{*} (0.059)	0.000 (0.064)	1
HON	0.015 (0.012)	0.000 (0.023)	0.108 (0.019)	0.860 (0.030)	0.092 (0.068)	0.000 (0.065)	1
IBM	0.011 (0.007)	0.000 (0.015)	0.044 (0.016)	0.858 (0.033)	0.116^{*} (0.063)	0.000 (0.070)	0.5
INTC	0.013 (0.009)	0.000 (0.012)	0.031 (0.013)	0.862 (0.029)	0.119^{*} (0.061)	0.000 (0.063)	0.5
JNJ	0.022 (0.011)	0.004 (0.023)	0.176 (0.033)	0.757 (0.049)	0.144^{**} (0.061)	0.000 (0.070)	1.5
KO	0.009 (0.016)	0.010 (0.023)	0.084 (0.028)	0.702 (0.057)	0.248^{**} (0.082)	0.000 (0.093)	1
LLY	0.021 (0.089)	0.071 (0.029)	0.065 (0.036)	0.245 (0.146)	0.108 (0.097)	0.599^{*} (0.316)	0.5
MCD	0.015 (0.010)	0.031 (0.015)	0.043 (0.016)	0.863 (0.029)	0.096 (0.065)	0.000 (0.070)	0.5

Table 2: (continued)

	ω	α_+	α_-	β	π_1	π_2	δ
MMM	0.035 (0.020)	0.011 (0.024)	0.015 (0.024)	0.777 (0.057)	0.181* (0.104)	0.000 (0.104)	0.5
MOT	0.011 (0.010)	0.004 (0.014)	0.066 (0.015)	0.888 (0.025)	0.081 (0.067)	0.000 (0.070)	1
MRK	0.022 (0.013)	0.017 (0.017)	0.085 (0.024)	0.904 (0.026)	0.046 (0.073)	0.000 (0.065)	1
MS	0.015 (0.016)	0.011 (0.019)	0.058 (0.022)	0.720 (0.080)	0.251*** (0.078)	0.000 (0.102)	0.5
MSFT	0.000 (0.011)	0.046 (0.019)	0.038 (0.015)	0.731 (0.066)	0.237*** (0.073)	0.000 (0.100)	0.5
ORCL	0.000 (0.010)	0.001 (0.014)	0.050 (0.016)	0.888 (0.024)	0.095 (0.063)	0.000 (0.065)	1
PEP	0.011 (0.010)	0.042 (0.017)	0.070 (0.021)	0.842 (0.035)	0.084 (0.058)	0.000 (0.066)	2
PFE	0.005 (0.006)	0.014 (0.010)	0.041 (0.010)	0.956 (0.010)	0.014 (0.031)	0.000 (0.029)	2
PG	0.032 (0.021)	0.000 (0.027)	0.134 (0.035)	0.649 (0.074)	0.269*** (0.080)	0.000 (0.100)	1
QCOM	0.051 (0.027)	0.029 (0.019)	0.110 (0.024)	0.819 (0.038)	0.116* (0.065)	0.000 (0.070)	1.5
SLB	0.116 (0.049)	0.003 (0.017)	0.015 (0.019)	0.827 (0.045)	0.121* (0.067)	0.000 (0.073)	2
T	0.008 (0.009)	0.003 (0.013)	0.050 (0.018)	0.881 (0.023)	0.087 (0.055)	0.000 (0.058)	2
TWX	0.000 (0.030)	0.041 (0.028)	0.150 (0.033)	0.564 (0.063)	0.211** (0.071)	0.166* (0.081)	1.5
UN	0.020 (0.010)	0.039 (0.021)	0.108 (0.039)	0.705 (0.064)	0.189** (0.072)	0.000 (0.078)	2
VZ	0.012 (0.011)	0.050 (0.019)	0.054 (0.020)	0.787 (0.051)	0.162** (0.063)	0.000 (0.074)	0.5
WFC	0.000 (0.011)	0.020 (0.024)	0.091 (0.029)	0.734 (0.056)	0.120* (0.069)	0.100 (0.073)	1
WMT	0.002 (0.006)	0.010 (0.012)	0.047 (0.013)	0.916 (0.016)	0.050 (0.061)	0.000 (0.061)	2
WYE	0.000 (0.008)	0.012 (0.013)	0.042 (0.013)	0.877 (0.029)	0.099 (0.063)	0.005 (0.069)	0.5
XOM	0.073 (0.030)	0.021 (0.021)	0.066 (0.023)	0.742 (0.073)	0.118* (0.060)	0.048 (0.085)	2
XRX	0.000 (0.017)	0.010 (0.019)	0.012 (0.020)	0.828 (0.049)	0.170** (0.056)	0.000 (0.084)	0.5

Table 3: "Unusable" APARCH-X(1,1) model (23) with contemporaneous realized volatility as covariate (extract).

	ω	α_+	α_-	β	π_0	δ
AAPL	0.078 (0.026)	0.031 (0.015)	0.042 (0.014)	0.794 (0.048)	0.133 (0.039)	0.5
ABT	0.037 (0.023)	0.005 (0.023)	0.000 (0.027)	0.664 (0.066)	0.304 (0.064)	0.5
AXP	0.029 (0.010)	0.000 (0.016)	0.028 (0.016)	0.746 (0.039)	0.227 (0.039)	0.5
BA	0.080 (0.034)	0.012 (0.029)	0.069 (0.029)	0.702 (0.059)	0.231 (0.054)	1.5
BAC	0.000 (0.013)	0.000 (0.028)	0.027 (0.027)	0.648 (0.053)	0.349 (0.062)	0.5
BMY	0.000 (0.025)	0.035 (0.019)	0.047 (0.021)	0.821 (0.029)	0.147 (0.043)	0.5
BP	0.009 (0.012)	0.000 (0.017)	0.021 (0.016)	0.727 (0.049)	0.269 (0.053)	0.5
C	0.008 (0.013)	0.000 (0.024)	0.076 (0.024)	0.679 (0.058)	0.276 (0.057)	1
UN	0.023 (0.011)	0.018 (0.023)	0.085 (0.039)	0.626 (0.059)	0.274 (0.052)	2
VZ	0.008 (0.010)	0.043 (0.019)	0.041 (0.020)	0.801 (0.042)	0.159 (0.043)	0.5
WFC	0.000 (0.011)	0.000 (0.022)	0.051 (0.028)	0.748 (0.050)	0.228 (0.050)	1
WMT	0.000 (0.008)	0.013 (0.013)	0.017 (0.014)	0.835 (0.028)	0.152 (0.028)	0.5
WYE	0.000 (0.009)	0.007 (0.014)	0.034 (0.014)	0.868 (0.025)	0.118 (0.032)	0.5
XOM	0.066 (0.026)	0.014 (0.022)	0.036 (0.024)	0.757 (0.052)	0.170 (0.046)	2
XRX	0.000 (0.019)	0.000 (0.019)	0.000 (0.025)	0.792 (0.044)	0.216 (0.059)	0.5

4 Conclusion

In this paper, we studied the asymptotic behavior of the QMLE for the versatile class of the semi-strong PGARCH models augmented with exogenous variables. The main assumptions on the exogenous variables are the stationarity and non-colinearity with the other explanatory variables of the volatility. This allows to incorporate some additional covariates for predicting the volatility of the financial returns, such as the volumes, the realized ranges, other past squared returns or intraday realized volatilities. Since the true value of the parameter is not constrained to belong to the interior of the parameter space, we were able to derive tests for the significance of the exogenous variables. For the asymptotic distribution of the QMLE, we investigated four different situations corresponding to strong or semi-strong models, and to parameters inside or at the boundary of the parameter space. When the GARCH-X parameter belongs to the interior of the parameter space, the asymptotic distribution of the QMLE is normal, whereas it is the projection of a normal distribution on a convex cone when one or several coefficients are equal to zero. For models with positive coefficients, the asymptotic distribution is obtained under very mild conditions, in particular, without any moment condition on the observed process. When the parameter stands at the boundary, moment conditions are required, and the extra assumptions are stronger for semi-strong than for strong models. Moreover, the sandwich form of the variance involved in the asymptotic distribution becomes simpler in the strong case.

The asymptotic theory developed in the paper has been applied to simulations and real series. Our empirical results are in accordance with numerous applied studies, and complement them by providing a formal test for the significance of the exogenous variables. In particular, we generally find useful the volume, the realized range and the intraday realized volatility for predicting the squares of the financial returns, but none of these variables can be considered as a perfect proxy of the volatility.

5 Proofs and technical lemmas

Proof of Lemma 1. The arguments being quite standard, we just give a sketch of proof (see *e.g.* Theorem 2.4 in Francq and Zakoïan (2010) for a similar result with a more detailed proof). Because all the components of $\boldsymbol{C}_{0t}$ and $\boldsymbol{B}_{0t}$ are non-negative, the components of $\boldsymbol{Y}_t$ defined by (4) are always well defined in $[0, +\infty]$. Since $E\left|\log \|\boldsymbol{B}_{01}\|\right| < \infty$, by the Cauchy rule, when $\gamma < 0$, the components of $\boldsymbol{Y}_t$ are shown to be almost surely finite. The process $\boldsymbol{Y}_t$ satisfies (3) and, being a measurable function of $(\eta_t, \boldsymbol{x}_t)$, it is also stationary and ergodic under **A1**. The uniqueness of the stationary solution is shown as in the case $\boldsymbol{\pi}_0 = 0$. For the converse, note that when a (finite) stationary solution exists for (3), then

$$\lim_{k\to\infty} \left(\prod_{i=1}^{k} \boldsymbol{C}_{0,t-i-1} \right) \boldsymbol{B}_{0,t-k} = 0 \quad \text{a.s.}$$

Note also that

$$\left(\prod_{i=1}^{k} \boldsymbol{C}_{0,t-i-1} \right) \boldsymbol{B}_{0,t-k} \geq \left(\prod_{i=1}^{k} \boldsymbol{C}_{0,t-i-1} \right) \boldsymbol{B}$$

where $\boldsymbol{B}$ is obtained by replacing $\boldsymbol{\pi}_0$ with 0 in $\boldsymbol{B}_{01}$. Pan et al. (2008) have shown that $\left(\prod_{i=1}^{k} \boldsymbol{C}_{0,t-i-1} \right) \boldsymbol{B} \to 0$ a.s. entails that $\gamma < 0$.[3] □

Proof of Lemma 2. The result is known when $\boldsymbol{\pi}_0 = 0$ (see Proposition A.1 in HZ). For notational simplicity, we give the proof for general $\boldsymbol{\pi}_0$ when $p = q = 1$. If $\gamma < 0$, h_t is then given by (6). Using the elementary inequality $\left(\sum_i u_i\right)^s \leq \sum_i u_i^s$ for any sequence of positive numbers u_i and any $s \in (0, 1]$, and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} E h_t^s &\leq \sum_{k=0}^{\infty} E \left(\prod_{i=1}^{k} a(\eta_{t-i}) \varpi_{t-k-1} \right)^s = \sum_{k=0}^{\infty} E \left(\prod_{i=1}^{k} a^s(\eta_{t-i}) \varpi_{t-k-1}^s \right) \\ &\leq \sum_{k=0}^{\infty} \left(\prod_{i=1}^{k} E a^{2s}(\eta_{t-i}) E \varpi_{t-k-1}^{2s} \right)^{1/2}. \end{aligned}$$

By Assumption **A1**, there exists $s > 0$ such that $E\varpi_t^{2s} < \infty$. Moreover, the fact that $\gamma = E \log a(\eta_t) < 0$ and that $Ea^r(\eta_t) < \infty$ for some $r > 0$ entails that there exists $s > 0$ such that $Ea^{2s}(\eta_t) < 1$ (see *e.g.* Lemma 2.2 in (Francq and Zakoïan, 2010)). It follows

[3] Pan *et al.* (2008) employed different assumptions, but a careful examination of their proof and of Lemma 3.4 in Bougerol and Picard (992a) reveals that the assumption that (η_t) is iid, as well as their regularity condition **A1**, are useless for that result.

that $Eh_t^s < \infty$, and thus $Eh_t^{s/\delta} < \infty$ for some $s > 0$. By the Cauchy-Schwarz inequality and **A1**, we deduce that $E|\varepsilon_t|^{s/2} < \infty$. $\square$

The following lemma is useful to show the identifiability of the parameters under **A4**.

Lemma 3 *Let X be a random variable which takes at least three values and $P(X > 0) \in (0,1)$. If $a(X^+)^\delta + b(X^-)^\delta = c$ a.s., with $a, b, c \in \mathbb{R}$, then $a = b = 0$.*

Proof. If $a = 0$ and $b \neq 0$ then $b(X^-)^\delta = c$ with probability one. Since $P(X > 0) > 0$, we must have $c = 0$. It follows that $X^- = 0$ a.s. which is in contradiction with the assumption that $P(X > 0) < 1$.

If $a \neq 0$ and $b \neq 0$, we obtain $X = \left(\frac{c}{a}\right)^{1/\delta}$ when $X > 0$ and $X = -\left(\frac{c}{b}\right)^{1/\delta}$ when $X < 0$ which is in contradition with assumption that X has at least three values. $\square$

Proof of Theorem 1. The consistency can be shown by establishing the following intermediate results:

i) $\lim_{n\to\infty} \sup_{\boldsymbol{\vartheta}\in\Theta} \left| Q_n(\boldsymbol{\vartheta}) - \widetilde{Q}_n(\boldsymbol{\vartheta}) \right| = 0$ a.s.

ii) If $\sigma_t(\boldsymbol{\vartheta}) = \sigma_t(\boldsymbol{\vartheta}_0)$ a.s. then $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$.

iii) $E|\ell_t(\boldsymbol{\vartheta})| < \infty$ and if $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$, $E|\ell_t(\boldsymbol{\vartheta})| > E|\ell_t(\boldsymbol{\vartheta}_0)|$.

iv) For any $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$, there exists a neighborhood $V(\boldsymbol{\vartheta})$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\vartheta}^*\in V(\boldsymbol{\vartheta})} Q_n(\boldsymbol{\vartheta}^*) > E\ell_1(\boldsymbol{\vartheta}_0) \text{ a.s.}$$

The proofs of i), iii) and iv) being essentially the same as when the parameter $\boldsymbol{\pi}$ is not present, we only give the proof of ii).

Assume that $\sigma_t(\boldsymbol{\vartheta}) = \sigma_t(\boldsymbol{\vartheta}_0)$ a.s. By the second part of **A5**, the polynomials $\mathcal{B}_{\boldsymbol{\vartheta}}(B)$ and $\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)$ are invertible. We thus have

$$\begin{aligned}&\left\{\frac{\mathcal{A}_{\boldsymbol{\vartheta}+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} - \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}\right\}\left(\varepsilon_t^+\right)^\delta + \left\{\frac{\mathcal{A}_{\boldsymbol{\vartheta}-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} - \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}\right\}\left(\varepsilon_t^-\right)^\delta \\ &+ \left\{\frac{\boldsymbol{\pi}'B}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} - \frac{\boldsymbol{\pi}_0'B}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}\right\}\boldsymbol{x}_t = \frac{\omega_0}{\mathcal{B}_{\boldsymbol{\vartheta}}(1)} - \frac{\omega}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(1)} \text{ a.s.}\end{aligned} \tag{25}$$

If $\frac{\mathcal{A}_{\boldsymbol{\vartheta}+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} \neq \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}$ or $\frac{\mathcal{A}_{\boldsymbol{\vartheta}-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} \neq \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}$, then there exist $c_{i_0}^+ \neq 0$ or $c_{i_0}^- \neq 0$, a constant e and a sequence of vectors $(\boldsymbol{d}_i)$ such that

$$\sum_{i=i_0}^{\infty} c_i^+ B^i \left(\varepsilon_t^+\right)^\delta + \sum_{i=i_0}^{\infty} c_i^- B^i \left(\varepsilon_t^-\right)^\delta + \sum_{i=1}^{\infty} \boldsymbol{d}_i' \boldsymbol{x}_{t-i} = e.$$

Since $\left(\varepsilon_t^+\right)^\delta = \sigma_t^\delta\left(\eta_t^+\right)^\delta$ and $\left(\varepsilon_t^-\right)^\delta = \sigma_t^\delta\left(\eta_t^-\right)^\delta$, there exists $(a,b)' \in \mathbb{R}^2\backslash(0,0)$ such that

$$a\left(\eta_{t-i_0}^+\right)^\delta + b\left(\eta_{t-i_0}^-\right)^\delta = c_{t,i_0}$$

where $c_{t,i}$ is a measurable function of $\{\eta_{t-j}, j > i, \boldsymbol{x}_{t-k}, k > 0\}$. It follows that

$$\mathcal{L}\left(a\left(\eta_{t-i_0}^+\right)^\delta + b\left(\eta_{t-i_0}^-\right)^\delta \middle| \mathcal{F}_{t,i_0}\right) = \mathcal{L}\left(c_{t,i_0} \middle| \mathcal{F}_{t,i_0}\right)$$

where $\mathcal{L}(X|Y)$ denotes the distribution of X given Y. By lemma 3 and **A4**, we obtain $a = b = 0$. Therefore, we have

$$\frac{\mathcal{A}_{\boldsymbol{\vartheta}+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)} \text{ and } \frac{\mathcal{A}_{\boldsymbol{\vartheta}-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)} \tag{26}$$

By **A7**, we obtain $\mathcal{A}_{\boldsymbol{\vartheta}+}(B) = \mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)$, $\mathcal{A}_{\boldsymbol{\vartheta}-}(B) = \mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(B) = \mathcal{B}_{\boldsymbol{\vartheta}_0}(B)$. Then (25) becomes

$$(\boldsymbol{\pi} - \boldsymbol{\pi}_0)'\, \boldsymbol{x}_{t-1} = \omega_0 - \omega,$$

which entails $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ and $\omega = \omega_0$ under **A8**. Hence, (ii) is proved. □

The proof of the asymptotic distribution of the QMLE is split into several technical lemmas.

Lemma 4 *Under the assumptions of Theorem 2,*

(i) $\quad E\left\|\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'}\right\| < \infty, \quad E\left\|\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\| < \infty.$

(ii) $\quad \boldsymbol{J}$ *is non-singular and* $\operatorname{var}\left\{\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right\} = \boldsymbol{I}.$

Proof of Lemma 4. First note that the derivatives of $\ell_t(\boldsymbol{\vartheta}) = \frac{\varepsilon_t^2}{\sigma_t^2} + \ln \sigma_t^2$ are

$$\frac{\partial \ell_t\left(\boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\vartheta}} = \frac{\partial}{\partial \boldsymbol{\vartheta}}\left\{\frac{\varepsilon_t^2}{(\sigma_t^\delta)^{2/\delta}} + \frac{2}{\delta}\ln \sigma_t^\delta\right\} = \frac{2}{\delta}\left\{1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}}\right\} \tag{27}$$

and

$$\frac{\partial \ell_t^2(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} = \frac{2}{\delta}\left\{1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^\delta}\frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\} + \frac{2}{\delta}\left\{\frac{\delta+2}{\delta}\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}}\right\}\left\{\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}'}\right\}. \tag{28}$$

In Cases A and B (strong model), one can thus prove (i) by showing that

$$E\left\|\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right\|^2 < \infty \tag{29}$$

and

$$E\left\|\frac{1}{\sigma_t^\delta}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\|<\infty. \tag{30}$$

In Cases C and D (semi-strong model), using the Hölder inequality and $E|\eta_t|^{4+\nu}<\infty$, the existence of the first expectation in (i) can be proven by showing that

$$E\left\|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\|^{2+8/\nu}<\infty. \tag{31}$$

Thanks to **A2**, the second expectation of (i) is still obtained by showing (30).

The existence of the moments in (29)–(31) is already known when $\boldsymbol{\pi}_0$ is absent, (η_t) is iid and $\boldsymbol{\vartheta}_0$ belongs to the interior of Θ (see the proof of (i) and (ii) of Theorem 2.2 in HZ). We now explain the changes in the proof induced by the our particular framework in the case $p=q=1$. The proof can be easily extended to the general case.

Since $\sigma_t^\delta=\sum_{j=0}^\infty\beta^jc_{t,j}$ with $c_{t,j}=\omega+\alpha_+\left(\varepsilon_{t-j-1}^+\right)^\delta+\alpha_-\left(\varepsilon_{t-j-1}^-\right)^\delta+\boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}$, we have

$$\begin{aligned}
\frac{\partial\sigma_t^\delta}{\partial\omega}&=\sum_{j=0}^\infty\beta^j=\frac{1}{1-\beta},\\
\frac{\partial\sigma_t^\delta}{\partial\alpha_+}&=\sum_{j=0}^\infty\beta^j\left(\varepsilon_{t-j-1}^+\right)^\delta,\quad \frac{\partial\sigma_t^\delta}{\partial\alpha_-}=\sum_{j=0}^\infty\beta^j\left(\varepsilon_{t-j-1}^-\right)^\delta,\\
\frac{\partial\sigma_t^\delta}{\partial\beta}&=\sum_{j=1}^\infty j\beta^{j-1}c_{t,j},\\
\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\pi}}&=\sum_{j=0}^\infty\beta^j\boldsymbol{x}_{t-j-1}.
\end{aligned}$$

Similar expressions hold for the second order derivatives. Noting also that

$$\underline{\omega}:=\inf_{\boldsymbol{\vartheta}\in\Theta}\sigma_t^\delta>0, \tag{32}$$

under the moment conditions in **A11**, we have (29) and (30) in Case B, and (30) and (31) in Case D. In Cases A and C, we have

$$\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\omega}\leq\frac{1}{\underline{\omega}},\quad\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\alpha_+}\leq\frac{1}{\alpha_+},\quad\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\alpha_-}\leq\frac{1}{\alpha_-},\quad\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\pi_i}\leq\frac{1}{\pi_i},$$

with the notation $\boldsymbol{\pi}=(\pi_1,\ldots,\pi_r)'$. Using the inequalities $\frac{x}{1+x}\leq x^s$ and $(x+y)^s\leq x^s+y^s$, for $x,y\geq0$ and $s\in(0,1]$, we also have

$$\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\beta}\leq\frac{1}{\beta\omega^s}\sum_{i=1}^\infty i\beta^{is}\left\{\omega^s+\alpha_+^s\left(\varepsilon_{t-i-1}^+\right)^{\delta s}+\alpha_-^s\left(\varepsilon_{t-i-1}^-\right)^{\delta s}+\left(\boldsymbol{\pi}'\boldsymbol{x}_{t-i-1}\right)^s\right\}.$$

Note that, for any $q \geq 1$, under the moments assumptions in **A1** and **A6**, we have

$$\left\| \left(\varepsilon_{t-i-1}^{+}\right)^{\delta s} \right\|_q < \infty, \quad \left\| \left(\varepsilon_{t-i-1}^{-}\right)^{\delta s} \right\|_q < \infty \text{ and } \left\| (\boldsymbol{\pi}' \boldsymbol{x}_{t-i-1})^s \right\|_q < \infty$$

for sufficiently small $s > 0$. It follows that (29) and (31), for any $\nu > 0$, hold true when all the components of $\boldsymbol{\vartheta}_0$ are non-zero. By the same arguments, we can even show the stronger result that for any $s_0 > 0$ there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ included in Θ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^{s_0} < \infty \quad \text{in Cases A and C.} \tag{33}$$

We obtain (30) by similar arguments, which completes the proof of (i) in the two remaining cases A and C.

We now turn to the proof of (ii). Note that the second equality in (11) comes from (28) and the fact that $E(\eta_t^2 \mid \mathcal{F}_{t-1}) = 1$. Using (27), we also have $\operatorname{var}\left(\partial \ell_t\left(\boldsymbol{\vartheta}_0\right) / \partial \boldsymbol{\vartheta}\right) = \boldsymbol{I}$. It remains to show that $\boldsymbol{J}$ is invertible. If it is not the case, then there exists $\boldsymbol{c} \in \mathbb{R}^d$ such that

$$\boldsymbol{c}' \boldsymbol{J} \boldsymbol{c} = \frac{4}{\delta^2} E \left\{ \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \left(\boldsymbol{c}' \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right)^2 \right\} = 0.$$

Then a.s., $\boldsymbol{c}' \dfrac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = 0$. In view of (10), this implies

$$\boldsymbol{c}' \left(1, (\varepsilon_{t-1}^{+})^\delta, (\varepsilon_{t-1}^{-})^\delta, \ldots, (\varepsilon_{t-q}^{+})^\delta, (\varepsilon_{t-q}^{-})^\delta, \sigma_{t-1}^\delta(\boldsymbol{\vartheta}_0) \ldots, \sigma_{t-p}^\delta(\boldsymbol{\vartheta}_0), \boldsymbol{x}_{t-1}\right) = 0.$$

By the arguments used to prove (ii) of Theorem 1, this is impossible with $\boldsymbol{c} \neq 0$, which completes the proof of (ii). □

Lemma 5 *Under the assumptions of Theorem 2, as $n \to \infty$, we have*

$$\sqrt{n} \left\| \frac{\partial}{\partial \boldsymbol{\vartheta}} \widetilde{Q}_n(\boldsymbol{\vartheta}_0) - \frac{\partial}{\partial \boldsymbol{\vartheta}} Q_n(\boldsymbol{\vartheta}_0) \right\| = o(1) \text{ a.s.,} \tag{34}$$

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \widetilde{Q}_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) \right\| = o(1) \text{ a.s.} \tag{35}$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$,

$$\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \widetilde{Q}_n(\boldsymbol{\vartheta}_n) \to \boldsymbol{J} \text{ in probability when } \boldsymbol{\vartheta}_n \to \boldsymbol{\vartheta}_0 \text{ in probability,} \tag{36}$$

$$\sqrt{n} \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \xrightarrow{d} \mathcal{N}\left\{0, \boldsymbol{I}\right\}. \tag{37}$$

Proof of Lemma 5. In this proof, K and ρ denote generic constants whose values can be modified and such that $K > 0$ and $\rho \in (0, 1)$.

By the definition of $Q_n(\boldsymbol{\vartheta})$ and $\widetilde{Q}_n(\boldsymbol{\vartheta})$, (34) and (35) are entailed by

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} - \frac{\partial \widetilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\| \to 0 \text{ a.s. }, \tag{38}$$

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \widetilde{\ell}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \to 0 \text{ a.s.} \tag{39}$$

By the arguments used to show (7.60) in Francq and Zakoïan (2010), we have

$$\left| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \widetilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right| \leqslant K\rho^t \left(1 + K\eta_t^2\right) \left| 1 + \frac{1}{\sigma_t^\delta(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right|.$$

Under **A10-A11**, we have $E\eta_t^4 < \infty$ and (29), and thus the expectation of the right-hand side of the inequality is bounded by $K\rho^t$. It follows that

$$\sum_{t=1}^{\infty} \left| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \widetilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right|$$

has a finite expectation, and thus is finite almost surely, which entails (38). The convergence (39) is shown by arguments which follow the scheme of the proof of the last part of (d) on Page 167 in Francq and Zakoïan (2010).

To establish (36), first note that

$$P\left(\left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \widetilde{Q}_n(\boldsymbol{\vartheta}_n) - \boldsymbol{J} \right\| \geq \epsilon \right) \leq a_1 + a_2 + a_3 + a_4,$$

where

$$a_1 = P\left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \widetilde{Q}_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) \right\| \geq \frac{\epsilon}{3} \right),$$

$$a_2 = P\left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}_0) \right\| \geq \frac{\epsilon}{3} \right),$$

$$a_3 = P\left(\left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}_0) - \boldsymbol{J} \right\| \geq \frac{\epsilon}{3} \right), \qquad a_4 = P\left\{ \boldsymbol{\vartheta}_n \notin V(\boldsymbol{\vartheta}_0) \right\}$$

for any $\epsilon > 0$ and any neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$. By the assumption that $\boldsymbol{\vartheta}_n \to \boldsymbol{\vartheta}_0$ in probability, we have $a_4 \to 0$ as $n \to \infty$. By (35), for any $\epsilon > 0$ and when $V(\boldsymbol{\vartheta}_0)$ is sufficiently small, $a_1 \to 0$. The ergodic theorem and Lemma 4 imply that $a_3 \to 0$ for any $\epsilon > 0$. To prove that $a_2 \to 0$, it suffices to show that, for all $\epsilon > 0$, there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ satisfying

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \left\| \frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\ell_t(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\ell_t(\boldsymbol{\vartheta}_0) \right\| \leq \epsilon \quad \text{a.s.}$$

The result follows from the ergodic theorem, the dominated convergence theorem, the uniform continuity of the second order derivatives of $\ell_t(\boldsymbol{\vartheta})$, and by showing that

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \left\| \frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\ell_t(\boldsymbol{\vartheta}) \right\| < \infty \tag{40}$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$. Let us begin to prove (40) in Cases B and D by using **A12**. In view of (28) and since σ_t is bounded away from zero, to show (40), it suffices to establish that

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \left\| \varepsilon_t^2 \frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\vartheta}} \frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}) \right\| < \infty, \quad E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \left\| \varepsilon_t^2 \frac{\partial^2\sigma_t^\delta}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}) \right\| < \infty. \tag{41}$$

Now, (10), the second part of **A5** and the compactness of Θ entail that

$$\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\vartheta}_i} = d_{i,0}(\boldsymbol{\vartheta}) + \sum_{k=1}^{\infty} d_{i,k}^+(\boldsymbol{\vartheta}) \left(\varepsilon_{t-k}^+\right)^\delta + d_{i,k}^-(\boldsymbol{\vartheta}) \left(\varepsilon_{t-k}^-\right)^\delta + \boldsymbol{\pi}_{i,k}'(\boldsymbol{\vartheta})\boldsymbol{x}_{t-k},$$

with

$$\sup_{\boldsymbol{\vartheta}\in\Theta} |d_{i,0}(\boldsymbol{\vartheta})| \leq K, \quad \sup_{\boldsymbol{\vartheta}\in\Theta} \max\left\{|d_{i,k}^+(\boldsymbol{\vartheta})|, |d_{i,k}^-(\boldsymbol{\vartheta})|, \|\boldsymbol{\pi}_{i,k}(\boldsymbol{\vartheta})\|\right\} \leq K\rho^k.$$

The first moment condition in (41) thus follows from the Hölder inequality and **A12**. The second moment condition is obtained by doing similar developments for the second order derivatives. We thus have shown (40) in Cases B and D.

To establish (40) in the two other cases, let us first show that, for any $s_0 > 0$, there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that $V(\boldsymbol{\vartheta}_0) \subset \Theta$ and

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right|^{s_0} < \infty \quad \text{in Cases A and C.} \tag{42}$$

By the arguments used to show (7.51) in Francq and Zakoïan (2010), for all $\xi > 0$ and $s \in (0,1)$, there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta(\boldsymbol{\vartheta})} \leq K + K\sum_{j=0}^{\infty} (1+\xi)^j \rho^{js} \left|\varepsilon_{t-j-1}\right|^{\delta s} + K\sum_{j=0}^{\infty} (1+\xi)^j \rho^{js} \left\|\boldsymbol{x}_{t-j-1}\right\|^s.$$

Without loss of generality, it can be assumed that $2s_0/\delta \geq 1$. By the Minskowski inequality, choosing s such that $E|\varepsilon_1|^{2ss_0} < \infty$ and $E\|\boldsymbol{x}_1\|^{2ss_0/\delta} < \infty$ and choosing for instance $\xi = \frac{1-\rho^s}{2\rho^s}$, we have

$$\begin{aligned}\left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2}\right\|_{s_0}^{\delta/2} &= \left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta}\right\|_{2s_0/\delta} \\ \leq & K + K\sum_{j=0}^{\infty}(1+\xi)^j\rho^{js}\left\||\varepsilon_1|^{\delta s}\right\|_{2s_0/\delta} + K\sum_{j=0}^{\infty}(1+\xi)^j\rho^{js}\left\|\|\boldsymbol{x}_1\|^s\right\|_{2s_0/\delta} < \infty.\end{aligned}$$

Now, in view of (28), **A2** and the Hölder inequality entail

$$\begin{aligned}E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\ell_t(\boldsymbol{\vartheta})\right\| \leq K\left\|1+\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left|\frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})}\right|\right\|_2 &\left\{\left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^\delta}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\|\right\|_2\right. \\ &\left.+\left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\|\right\|_4^2\right\}.\end{aligned}$$

Thus, in Cases A and C, (40) comes from (33) and (42), and the analog (33) for second order derivatives.

By (27), Lemma 4 and **A2**, the last result, (37), is a consequence of the central limit theorem for square integrable martingale difference of Billingsley (1961). □

Proof of Theorem 2. A Taylor expansion of $\widetilde{Q}_n(\boldsymbol{\vartheta})$ around $\boldsymbol{\vartheta}_0$ gives

$$\widetilde{Q}_n(\boldsymbol{\vartheta}) - \widetilde{Q}_n(\boldsymbol{\vartheta}_0) = \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0) + \frac{1}{2}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)'\boldsymbol{J}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0) + R_n(\boldsymbol{\vartheta}),$$

where

$$R_n(\boldsymbol{\vartheta}) = \left\{\frac{\partial\widetilde{Q}_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'} - \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0) + \frac{1}{2}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)'\left(\frac{\partial^2\widetilde{Q}_n(\boldsymbol{\vartheta}^*)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \boldsymbol{J}\right)(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0),$$

and $\boldsymbol{\vartheta}^*$ is between $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}_0$. In view of (34), (35) and (36), as $n \to \infty$, we have

$$nR_n(\boldsymbol{\vartheta}_n) = o_P\left\{\sqrt{n}(\boldsymbol{\vartheta}_n-\boldsymbol{\vartheta}_0)\right\} + o_P\left\{n\|\boldsymbol{\vartheta}_n-\boldsymbol{\vartheta}_0\|^2\right\} \tag{43}$$

when $\boldsymbol{\vartheta}_n - \boldsymbol{\vartheta}_0 = o_P(1)$. Therefore

$$nR_n(\boldsymbol{\vartheta}_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\boldsymbol{\vartheta}_n-\boldsymbol{\vartheta}_0) = O_P(1). \tag{44}$$

Letting

$$\boldsymbol{Z}_n = -\boldsymbol{J}^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}},$$

we obtain the following quadratic approximation of the objective function

$$\widetilde{Q}_n(\boldsymbol{\vartheta}) - \widetilde{Q}_n(\boldsymbol{\vartheta}_0) = \frac{1}{2n} \left\| \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}}^2 - \frac{1}{2n} \left\| \boldsymbol{Z}_n \right\|_{\boldsymbol{J}}^2 + R_n(\boldsymbol{\vartheta}). \tag{45}$$

Let

$$\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} = \arg \inf_{\boldsymbol{\vartheta} \in \Theta} \left\| \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}}.$$

Note that

$$\sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) = \boldsymbol{Z}_n^{\mathcal{C}} \text{ for } n \text{ large enough,} \tag{46}$$

where $\boldsymbol{Z}_n^{\mathcal{C}}$ denotes the projection of $\boldsymbol{Z}_n$ on $\mathcal{C}$. We have

$$\begin{aligned} 0 &\leq \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \\ &= 2n \left\{ \widetilde{Q}_n(\widehat{\boldsymbol{\vartheta}}_n) - \widetilde{Q}_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) \right\} + 2n \left\{ R_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) - R_n(\widehat{\boldsymbol{\vartheta}}_n) \right\} \\ &\leq 2n \left\{ R_n(\widehat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) \right\}, \end{aligned}$$

where the first inequality comes from the definition of $\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}$, the equality from (45), and the second inequality from the definition of $\widehat{\boldsymbol{\vartheta}}_n$. By (46), it follows that

$$\left| \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \right| \leq 2n \left\{ R_n(\widehat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) \right\}. \tag{47}$$

We now show that $R_n(\widehat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) = o_P(1)$. By definition of $\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}$ and since $\boldsymbol{\vartheta}_0 \in \Theta$, we also have

$$\left\| \sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}} \leq \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}.$$

The Minskowski inequality then entails that

$$\left\| \sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) \right\|_{\boldsymbol{J}} \leq \left\| \sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}} \leq 2 \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}.$$

By (37), we have $\|\boldsymbol{Z}_n\|_{\boldsymbol{J}} = O_P(1)$, and thus $\sqrt{n}(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n} - \boldsymbol{\vartheta}_0) = O_P(1)$. In view of (44), this entails $nR_n(\boldsymbol{\vartheta}_{\boldsymbol{Z}_n}) = o_P(1)$. By definition of $\widehat{\boldsymbol{\vartheta}}_n$ and (45), we have

$$0 \leq 2n\widetilde{Q}_n(\boldsymbol{\vartheta}_0) - 2n\widetilde{Q}_n(\widehat{\boldsymbol{\vartheta}}_n) = \|\boldsymbol{Z}_n\|_J^2 - \left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_J^2 - 2nR_n(\widehat{\boldsymbol{\vartheta}}_n).$$

It follows that, by the c_r-inequality,

$$\begin{aligned} \left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \right\|_{\boldsymbol{J}}^2 &\leq 2 \left(\left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \right) \\ &\leq 4 \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - 4nR_n(\widehat{\boldsymbol{\vartheta}}_n). \end{aligned}$$

We now show that $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = O_P(1)$, arguing by contradiction. The consistency of $\widehat{\boldsymbol{\vartheta}}_n$ and (43) then entail that $nR_n(\widehat{\boldsymbol{\vartheta}}_n) = o_P\left(\left\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)\right\|_{\boldsymbol{J}}^2\right)$. It follows that $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = O_P(1)$. From (44), we deduce that $nR_n(\widehat{\boldsymbol{\vartheta}}_n) = o_P(1)$. Thus, (47) can be rewritten as

$$\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 = o_P(1). \tag{48}$$

Using the characterization (12) of the projection on a convex cone, we have

$$\begin{aligned}\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 &= \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \\ &\quad + 2\left\langle \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n^{\mathcal{C}}, \boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n \right\rangle_{\boldsymbol{J}} \\ &\geq \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2.\end{aligned}$$

Using (48), we thus obtain

$$\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 \leq \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 = o_P(1).$$

The continuous mapping theorem and (37) entail $\boldsymbol{Z}_n^{\mathcal{C}} \xrightarrow{d} \boldsymbol{Z}^{\mathcal{C}}$, and the conclusion follows. □

Proof of Proposition 1. We only show (17) because (16) and (18) are obtained directly from arguments used in the proof of Theorem 1. Let

$$\boldsymbol{I}_n(\boldsymbol{\vartheta}) = \frac{4}{\delta^2}\frac{1}{n}\sum_{t=1}^n S_t(\boldsymbol{\vartheta}), \quad S_t(\boldsymbol{\vartheta}) = \left(\frac{\varepsilon_t^4}{\sigma_t^4(\boldsymbol{\vartheta})} - 1\right)\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\sigma_t^\delta(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}'}. \tag{49}$$

The difference between $\boldsymbol{I}_n(\widehat{\boldsymbol{\vartheta}}_n)$ and $\widehat{\boldsymbol{I}}_n$ is due to the initial values used to compute $\widetilde{\sigma}_t^\delta(\boldsymbol{\vartheta})$. Because the difference $\sigma_t^\delta(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t^\delta(\boldsymbol{\vartheta})$ does not depend on the covariates $(\boldsymbol{x}_t)$, one can use standard arguments to show that

$$\lim_{n\to\infty}\left\|\boldsymbol{I}_n(\widehat{\boldsymbol{\vartheta}}_n) - \widehat{\boldsymbol{I}}_n\right\| = 0 \quad \text{a.s.}$$

The ergodic theorem shows that

$$\lim_{n\to\infty}\|\boldsymbol{I}_n(\boldsymbol{\vartheta}_0) - \boldsymbol{I}\| = 0 \quad \text{a.s.}$$

Using the strong convergence of $\widehat{\boldsymbol{\vartheta}}_n$ to $\boldsymbol{\vartheta}_0$, it remains to show that, for any $\epsilon > 0$, there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$\lim_{n\to\infty}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\|\boldsymbol{I}_n(\boldsymbol{\vartheta}) - \boldsymbol{I}_n(\boldsymbol{\vartheta}_0)\| \leq \epsilon.$$

It suffices to show that there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \|S_t(\boldsymbol{\vartheta})\| < \infty, \tag{50}$$

and the result will follow from the ergodic theorem applied to $\left\{\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \|S_t(\boldsymbol{\vartheta}) - S_t(\boldsymbol{\vartheta}_0)\|\right\}_t$, the dominated convergence theorem, and the uniform continuity of $\boldsymbol{\vartheta} \mapsto S_t(\boldsymbol{\vartheta})$ in the neighborhood of $\boldsymbol{\vartheta}_0$.

In Cases A and C, (50) follows from **A10**, (33) and (42). In Cases B and D, (50) follows from **A12** and the Hölder inequality. □

References

Andrews, D.W. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.

Andrews, D.W. (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.

Billingsley, P. (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Bollerslev, T. (2008) Glossary to ARCH (GARCH). *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (eds. T. Bollerslev, J. R. Russell and M. Watson), Oxford University Press, Oxford, UK.

Bougerol, P. and N. Picard (1992a) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.

Bougerol, P. and N. Picard (1992b) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advance in Applied Probability* 18, 221–254.

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

Engle, R.F., Hendry, D.F. and J.F. Richard (1983) Exogeneity. *Econometrica* 51, 277–304.

Engle, R.F. and A.J. Patton (2001) What good is a volatility model. *Quantitative finance* 1, 237–245.

Escanciano, J.C. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

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 (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications* 117, 1265–1284.

Francq, C. and J-M. Zakoian (2009) Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.

Fuertes, A.M., Izzeldin, M. and E. Kalotychou (2009) On forecasting daily stock volatility: the role of intraday information and market conditions. *International Journal of Forecasting* 25, 259–281.

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

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

Han, H. (2013) Asymptotic Properties of GARCH-X Processes. *Journal of Financial Econometrics* Forthcoming.

Han, H. and D. Kristensen (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.

Han, H. and J.Y. Park (2012) ARCH/GARCH with persistent covariates: Asymptotic theory of MLE. *Journal of Econometrics* 167, 95–112.

Han, H. and J.Y. Park (2014) GARCH with omitted persistent covariate. *Economics Letters* 124, 248–254.

Hansen, P.R. and A. Lunde (2005) A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of applied econometrics* 20, 873–889.

Laurent, S., Lecourt, C. and F.C. Palm (2014) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis.* http://dx.doi.org/10.1016/j.csda.2014.05.015

Nelson, D. B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.

Nijman, T. and E. Sentana (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71–87.

Pan, J., Wang, H. and H. Tong (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.

Sucarrat, G. and A. Escribano (2010) The Power Log-GARCH Model. *Working document*, Economic Series 10-13, University Carlos III, Madrid.

Taylor, S.J. (1986) *Modelling Financial Time Series*, New-York : Wiley.

Wintenberger, O. (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.

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