Econometric Theory, **35**, 2019, 37–72.
doi:10.1017/S0266466617000512

QML INFERENCE FOR VOLATILITY MODELS WITH COVARIATES

CHRISTIAN FRANCQ AND LE QUYEN THIEU
CREST and Université de Lille, RiskDesign and Université Pierre et Marie Curie

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 can be a helpful covariate for predicting squared returns.

1. INTRODUCTION

The GARCH-type models are of the form

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

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

We are grateful to the Editor, Associate Editor and anonymous reviewers for useful comments and suggestions. Christian Francq gratefully acknowledges the financial support from the labex ECODEC and the Agence Nationale de la Recherche (ANR) via the Project MultiRisk (ANR-16-CE26-0015-02). Address correspondence to Christian Francq, 5 Avenue Henry Le Chatelier 91120 Plaiseau, France; e-mail: christian.francq@univ-lille3.fr.

© Cambridge University Press 2018

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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 not even 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 the following. Is it really useful to introduce covariates in the volatility? Which covariates should we add to the filtration $\mathcal{F}_{t-1}$? How many lag values should we consider in the GARCH formulation?

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 $x^- = \max(-x, 0)$. For $\delta_0 > 0$, consider the model defined by

$$\begin{cases} \varepsilon_t = h_t^{1/\delta_0}\eta_t \\ h_t = \omega_0 + \sum_{i=1}^{q}\left\{\alpha_{0i+}(\varepsilon_{t-i}^+)^{\delta_0} + \alpha_{0i-}(\varepsilon_{t-i}^-)^{\delta_0}\right\} + \sum_{j=1}^{p}\beta_{0j}h_{t-j} + \boldsymbol{\pi}_0'\boldsymbol{x}_{t-1}, \end{cases} \tag{2}$$

where $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{r,t})'$ 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 $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$, $\delta_0 > 0$ and $\boldsymbol{\pi}_0 = (\pi_{01}, \ldots, \pi_{0r})' \geq 0$ componentwise.

In the absence of covariates, i.e., when $\boldsymbol{\pi}_0 = 0$, this equation corresponds to the Asymmetric Power GARCH (APARCH) model introduced by Ding, Granger, and Engle (1993). Model (2) can thus be called APARCH-X. The APARCH is rather general, since it nests numerous ARCH-type parameterizations used by

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

practitioners. The standard GARCH is obtained with $\delta_0 = 2$ 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 $\delta_0 = 1$ 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 $\delta_0 = 1$, and the GJR of Glosten, Jaganathan, and Runkle (1993), obtained with $\delta_0 = 2$. 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, Wang, and Tong (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. In particular, Han and Kristensen (2014) studied the inference of the standard GARCH(1,1) model augmented by an additional covariate. With the aim to detect simultaneously causality-in-mean and causality-in-variance, Guo, Ling, and Zhu (2014) studied the factor double autoregressive model, whose particular case is an ARCH-X model in which the volatility depends on lagged values of a covariate. Kristensen and Rahbek (2005) derived the asymptotic properties of the QMLE for a wide class of ARCH models in which the volatility of the observed process y_t depends on a covariate x_{t-1}. Since this covariate is assumed to be measurable with respect to the past observations (see their assumption C3(ii)), their model does not fall exactly into the class of the ARCH-X.

To summarize, the few existing references dealing with the estimation of volatility models with exogenous variables concern specific models. In particular, these models are unable to take into account the leverage effect that is usually observed in financial series. More importantly, a common assumption to the above-mentioned papers 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, Francq and Zakoïan, 2007 for an application to univariate GARCH, and Pedersen, 2017 for a recent application to multivariate GARCH models). An important consequence of the nonnormality of the QMLE is that the usual t-ratio or the Wald tests

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

employed to identify the order p and q are also nonstandard (see e.g., Francq and Zakoïan, 2009b 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 squared 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, Hendry, and Richard (1983), but we can say that the $x_{i,t}$'s are exogenous variables in the sense that their dynamics are not specified by the APARCH-X model.

Our second objective is to propose tests of nullity for one or several components of ϑ_0. This is closely related to the first objective because, due to the positivity constraints on the components of ϑ_0, under the null hypothesis, the true parameter stands at the boundary of the parameter space.

The remainder of the paper is organized as follows. In Section 2, we first discuss conditions for 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. Our asymptotic results are obtained under regularity conditions that are close to optimal. They even allow to improve the existing results for GARCH models without exogenous variables (see Remark 8 below). 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 condition, which is the main requirement for the consistency of the QMLE.

2.1. Strict Stationarity

Without loss of generality, assume that $p \geq 2$ and $q \geq 2$. Let the vector of dimension $d_0 = 2q + p - 2$

$$\boldsymbol{Y}_t = \left(h_{t+1}, \ldots, h_{t-p+2}, \left(\varepsilon_t^+\right)^{\delta_0}, \left(\varepsilon_t^-\right)^{\delta_0}, \ldots, \left(\varepsilon_{t-q+2}^+\right)^{\delta_0}, \left(\varepsilon_{t-q+2}^-\right)^{\delta_0}\right)'.$$

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 $\boldsymbol{B}_{0t} = \left(\omega_0 + \boldsymbol{\pi}_0'\boldsymbol{x}_t, 0, \ldots, 0\right)'$ is a vector of dimension d_0 and $\boldsymbol{C}_{0t}$ is a $d_0 \times d_0$ matrix. Denoting by $\boldsymbol{C}_{0t}(i, j)$ the element of the row i and column j of $\boldsymbol{C}_{0t}$,

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

we have

$$\boldsymbol{C}_{0t}(1,1) = \alpha_{01+}(\eta_t^+)^{\delta_0} + \alpha_{01-}(\eta_t^-)^{\delta_0} + \beta_{01}, \quad \boldsymbol{C}_{0t}(1,j) = \beta_{0j} \quad \text{for } j = 2,\dots,p,$$
$$\boldsymbol{C}_{0t}(1,p+2i-3) = \alpha_{0i+} \text{ and } \boldsymbol{C}_{0t}(1,p+2i-2)) = \alpha_{0i-} \quad \text{for } i = 2,\dots,q,$$
$$\boldsymbol{C}_{0t}(i,i-1) = 1 \quad \text{for } i = 2,\dots,p,$$
$$\boldsymbol{C}_{0t}(p+1,1) = (\eta_t^+)^{\delta_0}, \quad \boldsymbol{C}_{0t}(p+2,1) = (\eta_t^-)^{\delta_0},$$
$$\boldsymbol{C}_{0t}(p+2i-1,p+2i-3) = \boldsymbol{C}_{0t}(p+2i,p+2i-2) = 1 \quad \text{for } i = 2,\dots,q-1,$$

and all the other elements of $\boldsymbol{C}_{0t}$ are equal to zero. 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 (1992a), the stationarity relies on the top Lyapunov exponent

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

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 shown 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, nonanticipative and ergodic solution. The solution of* (3) *is given by*

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

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

The proof is omitted since it is quite similar to Bougerol and Picard (1992b). Pan et al. (2008) also obtained the same result for the APARCH without exogenous variables. They used different assumptions, but a careful examination of their proof and of Lemma 3.4 in Bougerol and Picard (1992b) reveals that the assumption that (η_t) is iid, as well as their regularity condition **A1**, are useless in our framework. Moreover, the introduction of exogenous variables does not complicate anything since, under our assumption **A1**, the proof is unchanged when the intercept ω_0 is replaced by $\varpi_t = \omega_0 + \boldsymbol{\pi}_0'\boldsymbol{x}_t$.

Remark 1. Note that Lemma 1 can also be used when $p < 2$ or $q < 2$. Consider for instance the APARCH-X(1,1) model, i.e., (2) with $p = q = 1$.

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

This is equivalent to considering a model with $p = q = 2$ and $\alpha_{02+} = \alpha_{02-} = \beta_{02} = 0$. In that case we have

$$\boldsymbol{C}_{0t} = \begin{pmatrix} a(\eta_t) & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \left(\eta_t^+\right)^{\delta_0} & 1 & 0 & 0 \\ \left(\eta_t^-\right)^{\delta_0} & 0 & 1 & 0 \end{pmatrix}, \quad a(z) = \alpha_{0+}(z^+)^{\delta_0} + \alpha_{0-}(z^-)^{\delta_0} + \beta_0,$$

with the simplified notations $\alpha_{0+} = \alpha_{01+}$, $\alpha_{0-} = \alpha_{01-}$ and $\beta_0 = \beta_{01}$. In view of the shape of $\boldsymbol{C}_{0t}$, the top Lyapunov then takes the explicit form

$$\gamma = E \log a(\eta_1). \tag{4}$$

The condition $\alpha_{0+} E(\eta_1^+)^{\delta_0} + \alpha_{0-} E(\eta_1^-)^{\delta_0} + \beta_0 < 1$ then implies the strict stationarity condition $\gamma < 0$. It can be checked directly that, under **A1** and $\gamma < 0$, the APARCH-X(1,1) model admits a strictly stationary solution, whose volatility is given by

$$h_t = \sum_{k=0}^{\infty} \prod_{i=1}^{k} a\left(\eta_{t-i}\right) \varpi_{t-k-1}, \tag{5}$$

with the convention $\prod_{i=1}^{k} a\left(\eta_{t-i}\right) = 1$ when $k = 0$, and $\varpi_t = \omega_0 + \boldsymbol{\pi}_0' \boldsymbol{x}_t$.

Remark 2. It has to be noted that the top Lyapunov coefficient γ does not depend on the exogenous variables $\boldsymbol{x}_t$. Therefore, as long as **A1** holds, the presence of exogenous variables does not modify the stationarity constraints. This is not surprising because, for the APARCH model without exogenous variables, it is known that the stationarity condition does not depend on the intercept ω_0 and, as illustrated by (5) in the APARCH(1,1) case, adding exogenous variables essentially amounts to replacing ω_0 by $\varpi_t = \omega_0 + \boldsymbol{\pi}_0' \boldsymbol{x}_t$. It is not forbidden to take $\boldsymbol{x}_t = \varepsilon_t$, 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_0 is difficult to estimate in practice. The quasi-likelihood being very flat in the direction of δ_0, estimating this parameter leads to imprecise results and considerably slows down the optimization routine. We therefore first consider that δ_0 is fixed by the practitioner. We postpone the data-driven choice of the power δ_0 in Section 2.5 below. 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

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

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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

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

where the $\widetilde{\sigma}_t$'s are defined recursively, for $t\geq 1$, by

$$\widetilde{\sigma}_t^{\delta_0}=\widetilde{\sigma}_t^{\delta_0}(\vartheta)=\omega+\sum_{i=1}^{q}\left\{\alpha_{i+}\left(\varepsilon_{t-i}^{+}\right)^{\delta_0}+\alpha_{i-}\left(\varepsilon_{t-i}^{-}\right)^{\delta_0}\right\}+\sum_{j=1}^{p}\beta_j\widetilde{\sigma}_{t-j}^{\delta_0}+\boldsymbol{\pi}'\boldsymbol{x}_{t-1}. \quad \textbf{(6)}$$

The QMLE of ϑ_0 is defined as any measurable solution $\widehat{\vartheta}_n$ of

$$\widehat{\vartheta}_n=\underset{\vartheta\in\Theta}{\arg\max}\,L_n(\vartheta)=\underset{\vartheta\in\Theta}{\arg\min}\,\widetilde{Q}_n(\vartheta), \quad \textbf{(7)}$$

where

$$\widetilde{Q}_n(\vartheta)=\frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t,\quad \widetilde{\ell}_t=\widetilde{\ell}_t(\vartheta)=\frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2}+\ln\widetilde{\sigma}_t^2. \quad \textbf{(8)}$$

Let $\mathcal{A}_{\vartheta+}(z)=\sum_{i=1}^{q}\alpha_{i+}z^i$, $\mathcal{A}_{\vartheta-}(z)=\sum_{i=1}^{q}\alpha_{i-}z^i$ and $\mathcal{B}_{\vartheta}(z)=1-\sum_{j=1}^{p}\beta_j z^j$. Denote by $\mathcal{F}_{t-1}$ the σ-field generated by $\{\varepsilon_u,\boldsymbol{x}_u,u<t\}$ and denote by $\mathcal{F}_{t,i}$ the $\sigma-$field generated by $\{\eta_{t-j},j>i,\boldsymbol{x}_{t-k},k>0\}$ (so that $\mathcal{F}_{t-1}=\mathcal{F}_{t,0}$). To show the strong consistency of the QMLE, we need the following assumptions.

A2: $E(\eta_t\mid\mathcal{F}_{t-1})=0$ and $E\left(\eta_t^2\mid\mathcal{F}_{t-1}\right)=1$.
A3: $\vartheta_0\in\Theta$, Θ is compact.
A4: for all $i\geq 1$ (and all $t\in\mathbb{Z}$), the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i}$ is not included in $[0,\infty)$ or in $(-\infty,0]$ and contains at least three points.

Assumption **A4** is an identifiability condition that prevents taking redundant explanatory variables in the volatility, for instance $\boldsymbol{x}_{t-1}=\left(\varepsilon_{t-i}^{+}\right)^{\delta_0}$ (see Remarks 5 and 6 below). If Assumption **A4** does not hold, then there exists $i>0$ such that the law of η_{t-i} given the information $\mathcal{F}_{t,i}$ contained in $\{\eta_{t-i-1},\eta_{t-i-1},\dots;\boldsymbol{x}_{t-1},\boldsymbol{x}_{t-2},\dots\}$ is "degenerated". In such a case, either ε_{t-i}^{+} or ε_{t-i}^{-} can be "predicted" from $\mathcal{F}_{t,i}$, and the volatility can be written in several ways.

A5: $\gamma<0$ and $\sum_{j=1}^{p}\beta_j<1$ for all $\vartheta\in\Theta$.
A6: there exists $s>0$, such that $Eh_t^s<\infty$ and $E|\varepsilon_t|^s<\infty$.
A7: if $p>0$, $\mathcal{B}_{\vartheta_0}(z)$ has no common root with $\mathcal{A}_{\vartheta_0+}(z)$ and $\mathcal{A}_{\vartheta_0-}(z)$; $\mathcal{A}_{\vartheta_0+}(1)+\mathcal{A}_{\vartheta_0-}(1)\neq 0$ and $\alpha_{0q+}+\alpha_{0q-}+\beta_{0p}\neq 0$ (with the notation $\alpha_{00+}=\alpha_{00-}=\beta_{00}=1$).

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

A8: if $\boldsymbol{c}$ is a non zero vector of $\mathbb{R}^r$ then $\boldsymbol{c}'\boldsymbol{x}_1$ 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 **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 **A2** is employed by Escanciano (2009) and Han and Kristensen (2014). 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 $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ 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 $\boldsymbol{\Sigma}_t = \operatorname{diag}\left(\sigma_{1,t}^2, \sigma_{2,t}^2\right)$ 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 variables, 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 exists 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 $\boldsymbol{c}'\boldsymbol{x}_{t-1} = (\varepsilon_{t-i}^+)^{\delta_0}$ with $\boldsymbol{c} \in \mathbb{R}^r$ (otherwise the variable $(\eta_{t-i}^+)^{\delta_0}$ given $\mathcal{F}_{t,i}$ would be degenerated, which is impossible under **A4**).

The following lemma, which was obtained by Berkes, Horváth, and Kokoszka (2003) for standard GARCH and by Ling (2007) for ARMA-GARCH models, 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.*

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

We are now in a position to show the consistency of the QMLE.

THEOREM 1. *Let* $\widehat{\vartheta}_n$ *be a sequence of QMLE satisfying* (7). *Then, under* **A1**–**A8**,

$\widehat{\vartheta}_n \to \vartheta_0$ *a.s. as* $n \to \infty$.

Remark 6. The case of the standard GARCH(1,1) with one covariate ($\delta_0 = 2$, $p = q = r = 1$ and $\alpha_{i+} = \alpha_{i-}$) has been considered by Han and Kristensen (2014), with possibly nonstationary (x_t). However, our regularity conditions do not coincide with those of that paper. In particular, the identifiability condition given in Han and Kristensen (2014) is, with the notation of the present paper, for any $(a, b) \in \mathbb{R}^2$

$$(a,b) \neq (0,0) \implies a\eta_t^2 + bx_t^2 \mid \mathcal{F}_{t-1} \text{ has a non degenerated distribution.} \quad \textbf{(9)}$$

One has to preclude the covariate $x_t^2 = \varepsilon_t^2$, since the parameters α and π of the volatility $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \pi x_{t-1}^2$ are clearly not identifiable in that case. Note that we then have $\mathcal{F}_{t,1} = \sigma\left(\eta_{t-1}^2, \eta_{t-2}, \eta_{t-3}, \dots\right)$, and the distribution of $\eta_{t-1} \mid \mathcal{F}_{t,1}$ is concentrated on two points. The condition **A4** thus excludes the covariate $x_t^2 = \varepsilon_t^2$. This is not the case of (9) because $a\eta_t^2 + bx_t^2 \mid \mathcal{F}_{t-1} = (a + b\sigma_t^2)\eta_t^2 \mid \mathcal{F}_{t-1}$ is not degenerated when $(a, b) \neq (0, 0)$ (and η_t^2 is not degenerated). Thus (9) is not sufficient to ensure the identifiability.

2.3. Asymptotic Distribution of the QMLE

For the computation of (8), it is necessary to have $\widetilde{\sigma}_t(\vartheta) > 0$ almost surely, for any $\vartheta \in \Theta$. This is why the components of $\vartheta \in \Theta$ are constrained to be non-negative. More precisely, it can be assumed that, for $i = 2, \dots, d$, 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 ϑ_{0i} of ϑ_0 be greater than K_i), the following assumption is satisfied.

A9: $\mathcal{C} := \cup_{n=1}^{\infty} \left\{\sqrt{n}(\vartheta - \vartheta_0) : \vartheta \in \Theta\right\} = \prod_{i=1}^{d} \mathcal{C}_i$, where $\mathcal{C}_i = [0, +\infty)$ when $\vartheta_{0i} = 0$ and $\mathcal{C}_i = \mathbb{R}$ otherwise.

The set $\mathcal{C} = \mathcal{C}_{\vartheta_0}$ will be called the local parameter space. Take the simple example of Model (2) with $q = 1$, $p = 0$ and without exogenous variables. Assume that the parameter space is of the form $\Theta = [\underline{\omega}, \overline{\omega}] \times [0, \overline{\alpha}_+] \times [0, \overline{\alpha}_-]$, with obvious notations. Assumption **A9** is satisfied if $\underline{\omega} < \omega_0 < \overline{\omega}$, $\overline{\alpha}_+ > \alpha_{01+}$, and $\overline{\alpha}_- > \alpha_{01-}$ (i.e., if Θ is large enough). We then have, for instance, $\mathcal{C} = \mathbb{R} \times [0, \infty) \times \mathbb{R}$ when $\alpha_{01+} = 0$ and $\alpha_{01-} > 0$. More generally, $\mathcal{C}$ is a convex cone, which is equal to $\mathbb{R}^d$ if and only if ϑ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ, i.e., if all the components of ϑ_0 are nonzero, 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

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

ε_t is needed for the consistency and asymptotic normality (CAN) of the QMLE when the GARCH parameter belongs to $\overset{\circ}{\Theta}$, whereas moments conditions are required when the parameter stands at the boundary of the parameter space (see the example given in Sect. 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 ϑ_0 are strictly positive;

Case B : η_t is independent of $\mathcal{F}_{t-1}$ and at least one component of ϑ_0 is equal to zero;

Case C : η_t is not independent of $\mathcal{F}_{t-1}$ and all the components of ϑ_0 are strictly positive;

Case D : η_t is not independent of $\mathcal{F}_{t-1}$ and at least one component of ϑ_0 is equal to zero.

It is however worth considering Cases B and D, in particular, because this gives the asymptotic distribution of the QMLE under the null that $\pi_0 = 0$. 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: For the power δ_0 of the APARCH model (2): in Case B we have $E|\varepsilon_t|^{2\delta_0+\nu_1} < \infty$ and $E\|\boldsymbol{x}_t\|^{2+\nu_1} < \infty$ for some $\nu_1 > 0$; in Case D we have $E|\varepsilon_t|^{2\delta_0+8\delta_0/\nu} < \infty$ and $E\|\boldsymbol{x}_t\|^{2+8/\nu} < \infty$ with ν satisfying **A10**.

It will be convenient to approximate the sequence $\{\widetilde{\ell}_t(\vartheta)\}$ by an ergodic stationary sequence. Therefore, denote by $\left(\sigma_t^{\delta_0}\right)_t = \left\{\sigma_t^{\delta_0}(\vartheta)\right\}_t$ the strictly stationary, ergodic and nonanticipative solution of

$$\sigma_t^{\delta_0} = \omega + \sum_{i=1}^{q} \left\{\alpha_{i+}\left(\varepsilon_{t-i}^{+}\right)^{\delta_0} + \alpha_{i-}\left(\varepsilon_{t-i}^{-}\right)^{\delta_0}\right\} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\delta_0} + \boldsymbol{\pi}'\boldsymbol{x}_{t-1}. \tag{10}$$

Note that $\sigma_t^{\delta_0}(\vartheta_0) = h_t$. Let $Q_n(\vartheta)$ and ℓ_t be obtained by replacing $\widetilde{\sigma}_t^{\delta_0}$ with $\sigma_t^{\delta_0}$ in $\widetilde{Q}_n(\vartheta)$ and $\widetilde{\ell}_t$.

Remark 7. In order to explain the need of the moment condition **A11**, let us consider the simple example of an ARCH(1) with one covariate: $\sigma_t^2 = \omega_0 + \alpha_0\varepsilon_{t-1}^2 + \pi_0 x_{t-1}$. The score is then equal to

$$\frac{\partial \ell_t(\vartheta_0)}{\partial \vartheta} = (1-\eta_t^2)\boldsymbol{D}_t, \quad \boldsymbol{D}_t = \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2(\vartheta_0)}{\partial \vartheta} = \frac{1}{\omega_0 + \alpha_0\varepsilon_{t-1}^2 + \pi_0 x_{t-1}} \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ x_{t-1} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

In Cases A and C, the components of $\boldsymbol{D}_t$ are bounded, respectively, by $1/\omega_0$, $1/\alpha_0$ and $1/\pi_0$. For more general GARCH models, the components of $\boldsymbol{D}_t$ are not bounded, but they admit moments of any order. This explains why, when $\vartheta_0 \in \overset{\circ}{\Theta}$, no moment assumption is required, except those of **A10**, for the existence of the variance of the score. On the contrary, when $\alpha_0 = \pi_0 = 0$, the existence of the variance of the score requires $E\eta_t^4\varepsilon_{t-1}^4 < \infty$ and $E\eta_t^4 x_{t-1}^2 < \infty$, which gives $E\eta_t^4 < \infty$, $E\varepsilon_t^4 < \infty$ and $Ex_t^2 < \infty$ in Case B, and a stronger condition in Case D.

Hall and Yao (2003) showed that the asymptotic distribution of the QMLE is in general nonstandard when the score does not admit a second order moment. Even in the case where the score belongs to the domain of attraction of the Gaussian law, they found a rate of convergence different from $\sqrt{n}$.

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

$$\boldsymbol{J} := E\left(\frac{\partial^2 \ell_t(\vartheta_0)}{\partial\vartheta\partial\vartheta'}\right) = \frac{4}{\delta_0^2} E\left(\frac{1}{\sigma_t^{2\delta_0}(\vartheta_0)} \frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta} \frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta'}\right) \tag{11}$$

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

$$\boldsymbol{Z}^{\mathcal{C}} = \arg \inf_{\boldsymbol{C}\in\mathcal{C}} \|\boldsymbol{C} - \boldsymbol{Z}\|_{\boldsymbol{J}} \tag{12}$$

or equivalently by

$$\boldsymbol{Z}^{\mathcal{C}} \in \mathcal{C} \quad \text{and} \quad \left\langle \boldsymbol{Z} - \boldsymbol{Z}^{\mathcal{C}}, \boldsymbol{C} - \boldsymbol{Z}^{\mathcal{C}} \right\rangle_{\boldsymbol{J}} \le 0, \quad \forall \boldsymbol{C} \in \mathcal{C}. \tag{13}$$

THEOREM 2. *Under the assumptions of Theorem 1 and* **A9**–**A11**, *as* $n \to \infty$,

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} \boldsymbol{Z}^{\mathcal{C}}, \text{ where } \boldsymbol{Z} \sim \mathcal{N}\left\{0, \boldsymbol{\Sigma} := \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right\}, \tag{14}$$

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

$$\boldsymbol{I} = \frac{4}{\delta_0^2} E\left[\left\{E\left(\eta_t^4 \mid \mathcal{F}_{t-1}\right) - 1\right\} \frac{1}{\sigma_t^{2\delta_0}(\vartheta_0)} \frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta} \frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta'}\right].$$

Remark 8. 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:

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} \mathcal{N}\left\{0, (E\eta_1^4 - 1)\boldsymbol{J}^{-1}\right\} \qquad \text{in Case A,} \tag{15}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

and

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} \mathcal{N}\left\{0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right\} \qquad \text{in Case C.} \tag{16}$$

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 with no moment condition on the observed process ε_t in Cases A and C (note that **A11** does not concern these two cases). When there is no covariate ($r = 0$), we retrieve the results obtained by Francq and Zakoïan (2004) in the GARCH case ($\delta_0 = 2$ 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.

Under the moment condition $E|\varepsilon_t|^6 < \infty$, and other regularity conditions similar to ours, Francq and Zakoïan (2007) obtained (14) 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). Note that, when $\delta_0 = 2$, **A11** in Case B requires $E|\varepsilon_t|^{4+\nu_1} < \infty$ for some $\nu_1 > 0$, which is a milder condition than $E|\varepsilon_t|^6 < \infty$. Therefore Theorem 2 slightly improves the results available for the standard GARCH models. Even when $r = 0$, the authors are not aware of the existence of conditions entailing (14) for the general APARCH model, or even for the subclass of the GARCH model in Case D.

To the best of our knowledge, when ϑ_0 stands at the boundary of the parameter space (Cases B and D), a result similar to (14) for GARCH models with covariates is not available in the literature.

The next proposition provides estimates of the matrices $\boldsymbol{I}$ and $\boldsymbol{J}$ required to apply Theorem 2.

PROPOSITION 1. *Under the assumptions of Theorem 2, strongly consistent estimators of* $\boldsymbol{J}$ *and* $\boldsymbol{I}$ *are given by*

$$\widehat{\boldsymbol{J}}_n = \frac{4}{\delta_0^2}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\sigma}_t^{2\delta_0}(\widehat{\vartheta}_n)}\frac{\partial\widetilde{\sigma}_t^{\delta_0}(\widehat{\vartheta}_n)}{\partial\vartheta}\frac{\partial\widetilde{\sigma}_t^{\delta_0}(\widehat{\vartheta}_n)}{\partial\vartheta'}, \tag{17}$$

and

$$\widehat{\boldsymbol{I}}_n = \frac{4}{\delta_0^2}\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\eta}_t^4 - 1\right)\frac{1}{\widetilde{\sigma}_t^{2\delta_0}(\widehat{\vartheta}_n)}\frac{\partial\widetilde{\sigma}_t^{\delta_0}(\widehat{\vartheta}_n)}{\partial\vartheta}\frac{\partial\widetilde{\sigma}_t^{\delta_0}(\widehat{\vartheta}_n)}{\partial\vartheta'}, \tag{18}$$

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

Remark 9. Note that, for Theorem 2 and Proposition 1, a moment assumption is required on the observed process in Cases B and D. To see what may happen when this moment condition is not met, consider again the example of the ARCH(1)-X of Remark 7 with $\alpha_0 = \pi_0 = 0$ (so that the data generating process is

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

a noise). In Case B, the law of large numbers for nonintegrable positive random variables (Billingsley, 1995, pp. 284 and 495) and the arguments of the proof of Proposition 1 entail that

$$\widehat{\boldsymbol{I}}_n(2,2) = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\eta}_t^4 - 1\right)\frac{1}{\widetilde{\sigma}_t^4(\widehat{\boldsymbol{\vartheta}}_n)}\varepsilon_{t-1}^4 \to +\infty \quad \text{a.s.}$$

when $E\varepsilon_t^4 = +\infty$. This divergence was expected because, in the same framework, the information matrix is such that

$$\boldsymbol{I}(2,2) = \left(E\eta_t^4 - 1\right)\frac{1}{\omega_0^2}E\varepsilon_t^4 = +\infty.$$

When the information is not finite, the asymptotic distribution of the (Q)MLE may be nonstandard (Hall and Yao, 2003), although the estimator remains consistent (Theorem 1 does not require moment conditions). We made simulation experiments showing that the inference based on $\widehat{\boldsymbol{\vartheta}}_n$ can however work nicely even if the moment conditions of Assumption **A11** do not hold true.

Remark 10. In Cases A and B, in view of Remark 8, the estimator defined by (18) can be replaced by

$$\widehat{\boldsymbol{I}}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\widehat{\eta}_t^4 - 1\right)\widehat{\boldsymbol{J}}_n. \tag{19}$$

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. The Wald, Rao score (or Lagrange multiplier) and likelihood ratio tests are the most widely used parametric tests in econometrics. The Rao score test presents the advantage that its asymptotic distribution remains generally unchanged when, under the null, the parameter is at the boundary of the parameter space or not (Guo et al., 2014). In the sequel, we however concentrate on the Wald test because it has generally a higher local asymptotic power than the score test (see Francq and Zakoïan, 2009b for a comparison of the Bahadur slopes and local asymptotic powers of different tests on standard GARCH), and reduces to the popular t-test when testing the nullity of one coefficient.

We first concentrate on this one-dimensional testing problem. 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:

$$H_0 : \boldsymbol{e}_k'\boldsymbol{\vartheta}_0 = 0 \text{ and } \boldsymbol{e}_\ell'\boldsymbol{\vartheta}_0 > 0 \quad \forall \ell \neq k \qquad \text{against} \qquad H_1 : \boldsymbol{e}_k'\boldsymbol{\vartheta}_0 > 0. \tag{20}$$

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

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

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

Denote, respectively, by $\Phi^{-1}(\alpha)$ and $\chi^2_\ell(\alpha)$ the α-quantiles of the standard normal and 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 Proposition 1, for $\alpha \in (0, 1/2)$ the test of rejection region*

$$\{t_n(k) > \Phi^{-1}(1-\alpha)\} \quad \Leftrightarrow \quad \{t_n^2(k) > \chi^2_1(1-2\alpha)\} \tag{21}$$

has the asymptotic level α under H_0 and is consistent under H_1 defined in (20).

Remark 11. Under H_0, it can be shown that the QMLE saturates the positivity constraint (i.e., that $\boldsymbol{e}_k'\widehat{\boldsymbol{\vartheta}}_n = 0$) with probability $1/2$. The Wald statistic $t_n^2(k)$ then follows the chi-bar-square distribution $\underline{\delta}_0/2 + \chi^2_1/2$, which is the mixture of the Dirac at 0 (denoted $\underline{\delta}_0$) and of the χ^2_1 with equal weights. The p-value of the test of rejection region (21) is then equal to

$$P\left(\frac{\underline{\delta}_0}{2} + \frac{\chi^2_1}{2} > t_n^2(k)\right) = \frac{1}{2}P\left(\chi^2_1 > t_n^2(k)\right).$$

When $t_n^2(k) = 0$, the p-value is then equal to 0.5. The standard two-sided t-ratio test of rejection region $\{|t_n^2(k)| > \Phi^{-1}(1-\alpha/2)\}$ (i.e., $\{t_n^2(k) > \chi^2_1(1-\alpha)\}$) would have the wrong asymptotic level $\alpha/2$ instead of α. By contrast, the standard one-sided test has a correct asymptotic level.

Now consider a more general Wald test of the form

$$H_0^W : \boldsymbol{R}\boldsymbol{\vartheta}_0 = \boldsymbol{r} \qquad \text{against} \qquad H_1^W : \boldsymbol{R}\boldsymbol{\vartheta}_0 \neq \boldsymbol{r}, \tag{22}$$

where $\boldsymbol{R}$ is a $r_0 \times d$ full-rank matrix and $\boldsymbol{r}$ is a vector of dimension r_0. Noting that, under H_0^W, we have $\boldsymbol{R}\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{r} = \boldsymbol{R}\left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)$, Theorem 2 and Proposition 1 immediately show that the standard Wald test statistic

$$W_n = n\left(\boldsymbol{R}\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{r}\right)'\left(\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}_n\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{r}\right)$$

has the following asymptotic behavior.

COROLLARY 2. *Under the assumptions of Proposition 1 and H_0^W,*

$$W_n \xrightarrow{d} W = \left\|\boldsymbol{R}\boldsymbol{Z}^{\mathcal{C}}\right\|^2_{(\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}')^{-1}}, \qquad \boldsymbol{Z} \sim \mathcal{N}(0, \boldsymbol{\Sigma}).$$

The previous result is essentially useful to test that a given subset of the components of $\boldsymbol{\vartheta}_0$ are zero and that the other components are nonzero, because in this case the local parameter space $\mathcal{C}$ is determined by the null hypothesis. At the asymptotic level $\alpha \in (0, 1)$, the critical value q_α of the Wald test $\{W_n \geq q_\alpha\}$ can be obtained by taking the empirical $(1-\alpha)$-quantile of the set

$$\left\{\left\|\boldsymbol{R}\widehat{\boldsymbol{Z}}_i^{\mathcal{C}}\right\|^2_{(\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}_n\boldsymbol{R}')^{-1}}, i = 1, \ldots, N\right\},$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

where the $\widehat{\boldsymbol{Z}}_i$'s are independent and $\mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}_n)$-distributed random vectors, and N is a large number, say $N = 10,000$ or more (see Algorithm 1 of Pedersen, 2017). An R code of the algorithm can be found on the web pages of the authors.

2.4. Testing the Relevance of the Covariates

For GARCH-X models, the most interesting tests concern the relevance of the covariates. When the GARCH orders are small, for instance in the case of a GARCH-X(1,1) model, the GARCH coefficients are typically far from zero. In that case, it is reasonable to assume that the GARCH parameter $\boldsymbol{\theta}_0$ is strictly positive component-wise. Under this assumption, no moment condition on the observed process ε_t is needed for the QMLE to have the asymptotic distribution given by (14). We only assume

A12: in Case B we have $E\,\|\boldsymbol{x}_t\|^{2+\nu_1} < \infty$ for some $\nu_1 > 0$, and in Case D we have $E\,\|\boldsymbol{x}_t\|^{2+8/\nu} < \infty$ with ν as in **A10**.

THEOREM 3. *Under* **A1**–**A10** *and* **A12**, *when* $\boldsymbol{\theta}_0 > 0$, *the conclusion of Theorem 2 holds true.*

Without loss of generality, one can assume that we test that the last $r_0 \geq 1$ exogenous variables are irrelevant and the others are relevant. Let

$$\boldsymbol{\pi}_0 = (\boldsymbol{\pi}_{01}', \boldsymbol{\pi}_{02}')', \qquad \boldsymbol{\pi}_{01} = \big(\pi_{01}, \ldots, \pi_{0,r-r_0}\big)', \quad \boldsymbol{\pi}_{02} = \big(\pi_{0,r-r_0+1}, \ldots, \pi_{0r}\big)'. \tag{23}$$

Under the maintained assumption $\boldsymbol{\theta}_0 > 0$ and $\boldsymbol{\pi}_{01} > 0$ element-wise, we are interested in testing

$$H_0^{\boldsymbol{\pi}} : \boldsymbol{\pi}_{02} = 0_{r_0} \qquad \text{against} \qquad H_1^{\boldsymbol{\pi}} : \boldsymbol{\pi}_{02} \neq 0_{r_0}. \tag{24}$$

Let $\boldsymbol{K} = (0_{r_0 \times (d-r_0)} : I_{r_0})$ where I_k denotes the identity matrix of size k and $0_{k_1 \times k_2}$ denotes the null matrix of size $k_1 \times k_2$, with $0_{k_1 \times k_2} = 0_{k_1}$ when $k_2 = 1$. Notice that under H_0,

$$\sqrt{n}\hat{\boldsymbol{\pi}}_{2n} \xrightarrow{d} \boldsymbol{K}\boldsymbol{Z}^{\mathcal{C}}, \quad \text{as } n \to \infty,$$

where $\hat{\boldsymbol{\pi}}_{2n} = \boldsymbol{K}\hat{\boldsymbol{\vartheta}}_n$. The usual Wald test statistic is defined by

$$W_n = n\,\big\|\hat{\boldsymbol{\pi}}_{2n}\big\|^2_{(\boldsymbol{K}\hat{\boldsymbol{\Sigma}}\boldsymbol{K}')^{-1}}. \tag{25}$$

COROLLARY 3. *Under the assumptions of Theorem 3 and* $H_0^{\boldsymbol{\pi}}$,

$$W_n \xrightarrow{d} W = \big\|\boldsymbol{K}\boldsymbol{Z}^{\mathcal{C}}\big\|^2_{(\boldsymbol{K}\boldsymbol{\Sigma}\boldsymbol{K}')^{-1}}, \tag{26}$$

where $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}$ *and* $\mathcal{C} = \mathbb{R}^{d-r_0} \times [0,\infty)^{r_0}$.

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

2.5. Choosing the Power δ_0

So far δ_0 was assumed to be known. In practice, the choice of δ_0 has little influence on the fitted volatility (see e.g., Ali, 2013). However, a practitioner could hesitate between, for instance, fitting a TARCH (corresponding to $\delta_0 = 1$) or a GJR model (corresponding to $\delta_0 = 2$). Since the different models, obtained from several choices of δ_0, have the same number of unknown parameters, it seems natural to select the model that has the highest (quasi-)likelihood.

To justify this practice, assume there are k candidates for δ_0:

$$\delta_0 \in \Delta = \{\delta_1, \ldots, \delta_k\}, \qquad \delta_i > 0, \quad i = 1, \ldots, k. \tag{27}$$

In order to underline that $\widetilde{\sigma}_t(\vartheta)$ defined by (6) actually depends on δ_0, we now write $\widetilde{\sigma}_{\delta_0,t}(\vartheta)$ instead of $\widetilde{\sigma}_t(\vartheta)$. Now, let

$$(\widehat{\delta}_n, \widehat{\vartheta}_n) = \underset{\delta\in\Delta, \vartheta\in\Theta}{\arg\min}\, \widetilde{Q}_{\delta,n}(\vartheta), \tag{28}$$

where

$$\widetilde{Q}_{\delta,n}(\vartheta) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\ell}_{\delta,t}(\vartheta), \quad \widetilde{\ell}_{\delta,t}(\vartheta) = \frac{\varepsilon_t^2}{\widetilde{\sigma}_{\delta,t}^2(\vartheta)} + \ln \widetilde{\sigma}_{\delta,t}^2(\vartheta). \tag{29}$$

We need to slightly reinforce the identifiability condition **A4**.

A4*: for all $i \geq 1$, the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i}$ contains at least three points in $[0, \infty)$ and three points in $(-\infty, 0]$.

THEOREM 4. *Let* $(\widehat{\delta}_n, \widehat{\vartheta}_n)$ *be a measurable solution of* (28). *Under* **A1**–**A3**, **A4***, **A5**–**A8**, *and* (27), *almost surely* $\widehat{\vartheta}_n \to \vartheta_0$ *as* $n \to \infty$ *and*

$$\widehat{\delta}_n = \delta_0 \quad \textit{for } n \textit{ large enough.} \tag{30}$$

Moreover, under the additional assumptions **A9**–**A11**, *the limiting distribution of* $\widehat{\vartheta}_n$ *is still given by* (14).

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 (20), 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

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t = \omega_0 + \alpha_{0+}(\varepsilon_{t-1}^+) + \alpha_{0-}(\varepsilon_{t-1}^-) + \beta_0 \sigma_{t-1} + \pi_{01} x_{t-1} + \pi_{02} x_{t-2}. \end{cases} \tag{31}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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

$$x_t = e^{y_t}, \quad y_t = a y_{t-1} + e_t, \tag{32}$$

where $a = 0.8$ and (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 $\sqrt{\nu_t/(\nu_t-2)}\eta_t$ is generated according to the Student t distribution with ν_t degrees of freedom, where $\nu_t = 5 + x_{t-1}$. This specification of ν_t, in particular the existence of the constant 5, guarantees that **A10** is satisfied. Indeed, the fourth order moment exists because

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

and, by a similar argument, it can be shown that a moment of order larger than 4 also exists. We took the parameter $\vartheta_0 = (0.046, 0.027, 0.092, 0.843, 0.089, \pi_{02})$ 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 $\pi_{01} = \pi_{02} = 0.089$ in Cases A and C. In each of the four cases, we simulated 1,000 independent replications of model (31) 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. Table 1 summarizes the distribution of the estimation errors in the four cases. As expected, the Root Mean Square Error (RMSE) of estimation always decreases when n increases. It can be noted that the estimators are more accurate when the model is strong (Cases A and B) than when

TABLE 1. Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Cases A–D

	Case A		Case B		Case C		Case D	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
				$n = 1,000$				
ω	0.0007	0.0157	0.0012	0.0119	−0.0011	0.0227	−0.0002	0.0180
α_+	−0.0056	0.0182	−0.0055	0.0185	−0.0027	0.0240	−0.0028	0.0243
α_-	−0.0063	0.0232	−0.0056	0.0236	−0.0065	0.0326	−0.0056	0.0336
β	0.0042	0.0158	0.0020	0.0167	0.0022	0.0215	−0.0005	0.0230
π_1	0.0000	0.0271	−0.0053	0.0180	0.0021	0.0328	−0.0051	0.0229
π_2	0.0001	0.0307	0.0078	0.0118	0.0029	0.0683	0.0112	0.0250
				$n = 2,000$				
ω	0.0000	0.0108	0.0009	0.0076	−0.0009	0.0154	0.0005	0.0111
α_+	−0.0022	0.0147	−0.0021	0.0150	−0.0028	0.0187	−0.0027	0.0191
α_-	−0.0030	0.0162	−0.0025	0.0166	−0.0026	0.0229	−0.0020	0.0234
β	0.0018	0.0117	0.0004	0.0121	0.0019	0.0144	0.0001	0.0153
π_1	−0.0005	0.0195	−0.0048	0.0087	0.0008	0.0244	−0.0057	0.0120
π_2	0.0010	0.0225	0.0060	0.0082	−0.0005	0.0296	0.0073	0.0115

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

TABLE 2. 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	83.2	98.7	95.7	99.6	67.1	91.4	84.5	97.5
B	99.9	100	100	100	2.8	2.6	7.3	6.3
C	71.7	94.7	89.2	98.9	48.0	77.2	70.3	91.4
D	95.4	99.8	99.4	99.9	1.8	1.4	7.1	4.7

it is semi-strong (Cases C and D). A more detailed examination of the estimation errors reveals the presence of more frequent outliers in the semi-strong case. Also, in accordance with the asymptotic theory, the distribution of $\widehat{\pi}_2$ has a mass at zero in Cases B and D, and thus the distribution of the estimation errors is nonGaussian when the true value of the parameter stands at the boundary of the parameter space.

Table 2 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) and the test of the null hypothesis $\pi_{02} = 0$ (which is false in Cases A and C) are more powerful for the strong model than for the semi-strong model. This is not surprising since the semi-strong model is less accurately estimated than the strong one. Less obviously, the test that $\pi_{01} = 0$ 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 slightly too high. Indeed, when the nominal level is 5%, the empirical relative frequency of rejection over the 1000 independent replications should vary between 3.7% and 6.3% with probability of approximately 95%. When the nominal level is 1%, it varies from 0.5% to 1.6% with the same probability. The relative rejection frequencies displayed in Table 2 are slightly larger than the theoretical level, but they improve when n increases.

3.2. S&P 500 with Realized Range, Volume and Other Indices

In this section, we build a model to explain the volatility of the daily returns of the S&P 500 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 3, 1985 to June 29, 2007. 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 t. 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 nonstationary (the Augmented Dickey-Fuller (ADF) test does not reject the unit root hypothesis, whereas the KPSS test clearly rejects

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

the stationarity). 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 1 displays the series of the returns ε_t, the ranges rr_t and the relative volumes v_t, for which the output of the ADF and KPSS tests is compatible with stationarity. As potential explanatory variables for the S&P 500 volatility, we also added the returns of the Nikkei index, Nik_t, and of the FTSE index, Ft_t (in absolute value and at the power δ_0, so that all the returns involved in the volatility h_t have the same unit of measurement). Following Theorem 4, we fitted APARCH-X(1,1) models with $\delta_0 \in \{0.5, 1, 1.5, 2\}$. The model with the largest likelihood is obtained for $\delta_0 = 1$, and is given by $\varepsilon_t = h_t \eta_t$ with

$$h_t = 0.019 + \underset{(0.022)\,0.500}{0.000}\,\varepsilon_{t-1}^{+} + \underset{(0.042)\,0.003}{0.115}\,\varepsilon_{t-1}^{-} + \underset{(0.028)\,0.000}{0.872}\,h_{t-1} + \underset{(1.684)\,0.040}{2.943}\,rr_{t-1} + \underset{(0.032)\,0.055}{0.051}\,v_{t-1} + \underset{(0.004)\,0.500}{0.000}\,|Nik_{t-1}| + \underset{(0.013)\,0.017}{0.028}\,|Ft_{t-1}|.$$

The estimated standard deviation is given into brackets (under the estimated value of each coefficient), followed by the p-value of the test that the coefficient is equal to zero. The estimated standard deviation of the i-th parameter, defined by $\sqrt{\widehat{\boldsymbol{\Sigma}}_n(i,i)/n}$ where $\widehat{\boldsymbol{\Sigma}}_n = \widehat{\boldsymbol{J}}_n^{-1}\widehat{\boldsymbol{I}}_n\widehat{\boldsymbol{J}}_n^{-1}$ with (18)–(17), is consistent in Cases A and C. Each p-value is obtained from an application of Corollary 1 and corresponds to a test of the form (20). Since two estimated parameters are equal to zero, a test of nullity of a single coefficient of the form (20) does not seem appropriate. Denoting by π_{04} the theoretical value of the coefficient in front of the variable $|Nik_{t-1}|$, we thus performed the test

$$\alpha_{0+} = \pi_{04} = 0 \qquad \text{against} \qquad \alpha_{0+} \neq 0 \text{ or } \pi_{04} \neq 0,$$

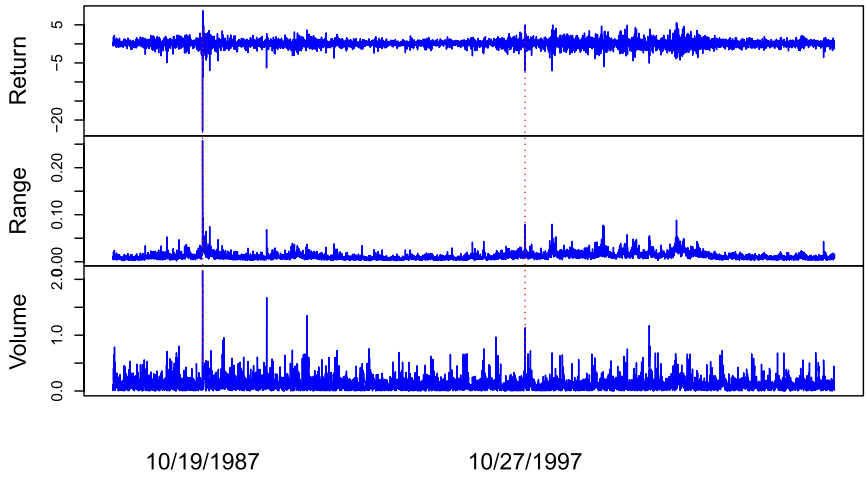


FIGURE 1. Return, range and volume of the S&P 500 index from January 4, 1985 to August 26, 2011 (October 19, 1987 corresponds to Black Monday, and October 27, 1997 is the date of a mini-crash that was caused by an economic crisis in Asia).

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

under the maintained assumption that the other coefficients are nonzero. This test is of the form (22). The estimated p-value of this test being 0.7722, we re-estimated the model under the null, and we found

$$h_t = 0.019 + \underset{(0.033)\,0.000}{0.115}\,\varepsilon_{t-1}^{-} + \underset{(0.028)\,0.000}{0.872}\,h_{t-1} + \underset{(1.150)\,0.005}{2.943}\,rr_{t-1} + \underset{(0.034)\,0.063}{0.051}\,v_{t-1} + \underset{(0.012)\,0.012}{0.028}\,|Ft_{t-1}|. \tag{33}$$

One can see that the range rr_{t-1} and the return series $|Ft_{t-1}|$ are significant covariates, whereas there exists a doubt concerning the significance of the volume v_{t-1}. 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, Izzeldin, and Kalotychou, 2009).

To assess the adequacy of the model, we also plotted in Figure 2 the correlogram of the series ε_t of the S&P 500 returns. Many sample autocorrelations lie outside the significance bounds $\pm 1.96/\sqrt{n}$, but these bounds, derived from the standard Bartlett's formula for linear processes, are only valid when the observations are iid. The significance bounds delimitated by the full lines are based on a generalized Bartlett's formula (see Francq and Zakoïan, 2009a) which is valid for GARCH and more general nonlinear processes. In view of these significance bounds, there is no evidence against the weak white noise assumption. Figure 2 also displays the autocorrelations of the squared residuals of the APARCH-X model, which are all very close to zero. The portmanteau test of Li and Mak (1994) and the weighted version of Fisher and Gallagher (2012) give, respectively, p-values of 0.1826 and 0.1742, and thus do not reject the adequacy of the model.

3.3. US Stocks with Realized Volatility

The data used in this section come from Section 4.2 of Laurent, Lecourt, and Palm (2016)[2] and concern 49 large capitalization stocks of American stock exchanges,

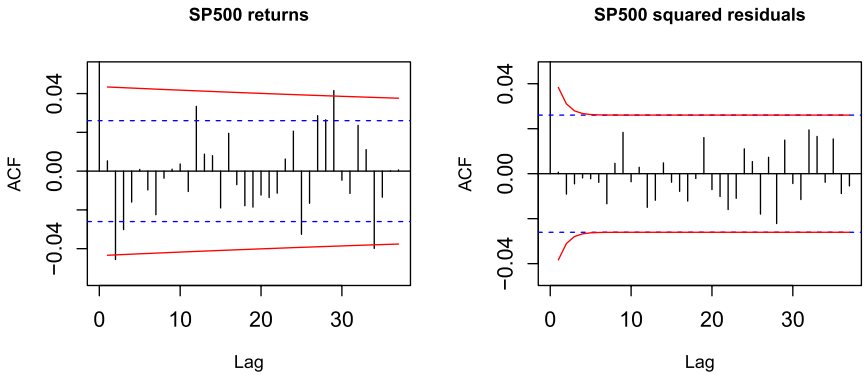


FIGURE 2. Autocorrelations of the S&P 500 returns and of the squared residuals of model (33), with the significance bands obtained from the standard Bartlett's formula (in dotted lines) and the generalized Bartlett's formula (in full lines).

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

TABLE 3. APARCH-X(1,1) models (34) fitted by QMLE on daily returns of US stock with two lagged values of realized volatilities as covariates. The estimated standard deviations are displayed in 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_0.

	ω_0	α_{0+}	α_-	β_0	π_{01}	π_{02}	δ_0
AAPL	0.080	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	0.023 (0.024)	0.019 (0.027)	0.661 (0.100)	0.285*** (0.071)	0.000 (0.102)	0.5
AXP	0.027	0.000 (0.019)	0.074 (0.019)	0.809 (0.038)	0.155** (0.061)	0.000 (0.069)	1
BA	0.046	0.027 (0.020)	0.092 (0.023)	0.843 (0.037)	0.084 (0.065)	0.000 (0.071)	2
BAC	0.007	0.009 (0.026)	0.090 (0.029)	0.813 (0.045)	0.151** (0.063)	0.000 (0.083)	1
BMY	0.000	0.051 (0.018)	0.077 (0.025)	0.880 (0.031)	0.072 (0.106)	0.000 (0.108)	1
BP	0.017	0.010 (0.020)	0.043 (0.017)	0.682 (0.068)	0.191* (0.086)	0.106 (0.092)	0.5
C	0.013	0.019 (0.025)	0.123 (0.028)	0.744 (0.056)	0.178** (0.059)	0.007 (0.077)	1
CAT	0.045	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	0.032 (0.021)	0.212 (0.065)	0.424 (0.105)	0.156** (0.057)	0.193** (0.073)	2
CSCO	0.013	0.000 (0.020)	0.054 (0.020)	0.848 (0.034)	0.131 (0.081)	0.000 (0.084)	1
CVX	0.082	0.014 (0.022)	0.068 (0.025)	0.716 (0.071)	0.102* (0.062)	0.088 (0.088)	2
DELL	0.000	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	0.000 (0.016)	0.070 (0.022)	0.807 (0.056)	0.132* (0.063)	0.002 (0.069)	2
EK	0.182	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	0.056 (0.032)	0.157 (0.039)	0.677 (0.060)	0.196** (0.072)	0.000 (0.077)	1.5
F	0.060	0.068 (0.037)	0.075 (0.025)	0.740 (0.050)	0.091 (0.088)	0.103 (0.097)	1
FDX	0.033	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	0.000 (0.019)	0.052 (0.022)	0.802 (0.050)	0.180* (0.085)	0.000 (0.076)	1
GM	0.026	0.018 (0.015)	0.047 (0.023)	0.881 (0.031)	0.094 (0.093)	0.000 (0.099)	2
HD	0.007	0.000 (0.015)	0.030 (0.013)	0.850 (0.037)	0.135* (0.070)	0.000 (0.071)	0.5
HNZ	0.007	0.050 (0.018)	0.084 (0.022)	0.840 (0.038)	0.105* (0.059)	0.000 (0.064)	1
HON	0.015	0.000 (0.023)	0.108 (0.019)	0.860 (0.030)	0.092 (0.068)	0.000 (0.065)	1
IBM	0.011	0.000 (0.015)	0.044 (0.016)	0.858 (0.033)	0.116* (0.063)	0.000 (0.070)	0.5
INTC	0.013	0.000 (0.012)	0.031 (0.013)	0.862 (0.029)	0.119* (0.061)	0.000 (0.063)	0.5
JNJ	0.022	0.004 (0.023)	0.176 (0.033)	0.757 (0.049)	0.144** (0.061)	0.000 (0.070)	1.5
KO	0.009	0.010 (0.023)	0.084 (0.028)	0.702 (0.057)	0.48** (0.082)	0.000 (0.093)	1
LLY	0.021	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	0.031 (0.015)	0.043 (0.016)	0.863 (0.029)	0.096 (0.065)	0.000 (0.070)	0.5
MMM	0.035	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	0.004 (0.014)	0.066 (0.015)	0.888 (0.025)	0.081 (0.067)	0.000 (0.070)	1
MRK	0.022	0.017 (0.017)	0.085 (0.024)	0.904 (0.026)	0.046 (0.073)	0.000 (0.065)	1
MS	0.015	0.011 (0.019)	0.058 (0.022)	0.720 (0.080)	0.251*** (0.078)	0.000 (0.102)	0.5
MSFT	0.000	0.046 (0.019)	0.038 (0.015)	0.731 (0.066)	0.237*** (0.073)	0.000 (0.100)	0.5
ORCL	0.000	0.001 (0.014)	0.050 (0.016)	0.888 (0.024)	0.095 (0.063)	0.000 (0.065)	1
PEP	0.011	0.042 (0.017)	0.070 (0.021)	0.842 (0.035)	0.084 (0.058)	0.000 (0.066)	2
PFE	0.005	0.014 (0.010)	0.041 (0.010)	0.956 (0.010)	0.014 (0.031)	0.000 (0.029)	2
PG	0.032	0.000 (0.027)	0.134 (0.035)	0.649 (0.074)	0.269*** (0.080)	0.000 (0.100)	1
QCOM	0.051	0.029 (0.019)	0.110 (0.024)	0.819 (0.038)	0.116* (0.065)	0.000 (0.070)	1.5
SLB	0.116	0.003 (0.017)	0.015 (0.019)	0.827 (0.045)	0.121* (0.067)	0.000 (0.073)	2
T	0.008	0.003 (0.013)	0.050 (0.018)	0.881 (0.023)	0.087 (0.055)	0.000 (0.058)	2
TWX	0.000	0.041 (0.028)	0.150 (0.033)	0.564 (0.063)	0.211** (0.071)	0.166* (0.081)	1.5
UN	0.020	0.039 (0.021)	0.108 (0.039)	0.705 (0.064)	0.189** (0.072)	0.000 (0.078)	2

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

TABLE 3. (continued)

	ω_0	α_{0+}	α_{0-}	β_0	π_{01}		π_{02}		δ_0
VZ	0.012	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	0.020 (0.024)	0.091 (0.029)	0.734 (0.056)	0.120*	(0.069)	0.100	(0.073)	1
WMT	0.002	0.010 (0.012)	0.047 (0.013)	0.916 (0.016)	0.050	(0.061)	0.000	(0.061)	2
WYE	0.000	0.012 (0.013)	0.042 (0.013)	0.877 (0.029)	0.099	(0.063)	0.005	(0.069)	0.5
XOM	0.073	0.021 (0.021)	0.066 (0.023)	0.742 (0.073)	0.118*	(0.060)	0.048	(0.085)	2
XRX	0.000	0.010 (0.019)	0.012 (0.020)	0.828 (0.049)	0.170**	(0.056)	0.000	(0.084)	0.5

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 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

$$\begin{cases} \varepsilon_t = h_t^{1/\delta_0}\eta_t \\ h_t = \omega_0 + \alpha_{0+}(\varepsilon_{t-1}^+)^{\delta_0} + \alpha_{0-}(\varepsilon_{t-1}^-)^{\delta_0} + \beta_0 h_{t-1} + \pi_{01} rv_{t-1}^{\delta_0/2} + \pi_{02} rv_{t-2}^{\delta_0/2}, \end{cases} \tag{34}$$

with $\delta_0 \in \{0.5, 1, 1.5, 2\}$. The variables rv_{t-1} and rv_{t-2} are raised to the power $\delta_0/2$ in order to have the same unit of measure for ε_t^2, the squared volatility h_t^{2/δ_0} and the realized volatility rv_t, regardless of δ_0. In view of Theorem 4, the selected value of δ_0 is that which leads to the maximum value of the quasi-likelihood. Table 3 displays the fitted model on each of the 49 stocks. For all the estimated models, except 3 over the 49, one observes that $\hat{\alpha}_- > \hat{\alpha}_+$, 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 $\hat{\pi}_1$ significantly nonzero and $\hat{\pi}_2$ close to zero. Similarly to what was done in Section 3.2, we also performed tests for the null that $\alpha_{0+} = \pi_{02} = 0$, and we reestimated constrained models. From these results, we find that yesterday's realized volatility often helps in predicting today's squared return.

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 stationarity and noncollinearity with the other explanatory variables of volatility. This framework allows additional covariates for predicting the volatility of the financial returns,

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

such as volumes, 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 GARCH coefficients, the asymptotic distribution is obtained under very mild conditions, in particular, without any moment condition on the observed process. When the GARCH parameter stands at the boundary, fourth-order moment conditions are required for the information matrix to be finite. Our asymptotic results are obtained under conditions that are only marginally stronger than these optimal moment conditions, which extends and improves the results that existed for GARCH models without covariates.

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.

5. PROOFS AND TECHNICAL LEMMAS

Proof of Lemma 2. The result is known when $\pi_0 = 0$ (see Proposition A.1 in HZ). For notational simplicity, we give the proof for general π_0 when $p = q = 1$. If $\gamma < 0$, h_t is given by (5). 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

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

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 that $Eh_t^s < \infty$, and thus $Eh_t^{s/\delta_0} < \infty$ for some $s > 0$. By the Cauchy-Schwarz inequality and **A1**, we deduce that $E|\varepsilon_t|^{s/2} < \infty$. ■

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

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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 $\delta\neq 0$ and $a,b,c\in\mathbb{R}$, 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 contradicts 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 contradicts the assumption that X has at least three values. ■

Proof of Theorem 1. In the sequel, a neighborhood of ϑ is denoted by $V(\vartheta)$. The consistency can be shown by establishing the following intermediate results:

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

ii) If $\sigma_t(\vartheta)=\sigma_t(\vartheta_0)$ a.s. then $\vartheta=\vartheta_0$.

iii) $E\left|\ell_t(\vartheta_0)\right|<\infty$ and if $\vartheta\neq\vartheta_0$, $E\left|\ell_t(\vartheta)\right|>E\left|\ell_t(\vartheta_0)\right|$.

iv) For any $\vartheta\neq\vartheta_0$, there exists $V(\vartheta)$ such that

$$\liminf_{n\to\infty}\inf_{\vartheta^*\in V(\vartheta)}Q_n\left(\vartheta^*\right)>E\ell_1(\vartheta_0)\text{ a.s.}$$

The proofs of i), iii) and iv) being essentially the same as when the parameter π is not present, we only give the proof of ii).

Assume that $\sigma_t(\vartheta)=\sigma_t(\vartheta_0)$ a.s. By the second part of **A5**, the polynomials $\mathcal{B}_\vartheta(B)$ and $\mathcal{B}_{\vartheta_0}(B)$ are invertible. We thus have

$$\left\{\frac{\mathcal{A}_{\vartheta+}(B)}{\mathcal{B}_\vartheta(B)}-\frac{\mathcal{A}_{\vartheta_0+}(B)}{\mathcal{B}_{\vartheta_0}(B)}\right\}\left(\varepsilon_t^+\right)^{\delta_0}+\left\{\frac{\mathcal{A}_{\vartheta-}(B)}{\mathcal{B}_\vartheta(B)}-\frac{\mathcal{A}_{\vartheta_0-}(B)}{\mathcal{B}_{\vartheta_0}(B)}\right\}\left(\varepsilon_t^-\right)^{\delta_0}$$
$$+\left\{\frac{\pi'B}{\mathcal{B}_\vartheta(B)}-\frac{\pi_0'B}{\mathcal{B}_{\vartheta_0}(B)}\right\}\boldsymbol{x}_t=\frac{\omega_0}{\mathcal{B}_\vartheta(1)}-\frac{\omega}{\mathcal{B}_{\vartheta_0}(1)}\text{ a.s.}\qquad(35)$$

If $\frac{\mathcal{A}_{\vartheta+}(B)}{\mathcal{B}_\vartheta(B)}\neq\frac{\mathcal{A}_{\vartheta_0+}(B)}{\mathcal{B}_{\vartheta_0}(B)}$ or $\frac{\mathcal{A}_{\vartheta-}(B)}{\mathcal{B}_\vartheta(B)}\neq\frac{\mathcal{A}_{\vartheta_0-}(B)}{\mathcal{B}_{\vartheta_0}(B)}$, 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

$$\sum_{i=i_0}^{\infty}c_i^+B^i\left(\varepsilon_t^+\right)^{\delta_0}+\sum_{i=i_0}^{\infty}c_i^-B^i\left(\varepsilon_t^-\right)^{\delta_0}+\sum_{i=1}^{\infty}\boldsymbol{d}_i'\boldsymbol{x}_{t-i}=e.$$

Since $\left(\varepsilon_t^+\right)^{\delta_0}=\sigma_t^{\delta_0}\left(\eta_t^+\right)^{\delta_0}$ and $\left(\varepsilon_t^-\right)^{\delta_0}=\sigma_t^{\delta_0}\left(\eta_t^-\right)^{\delta_0}$, there exists $(a,b)'\in\mathbb{R}^2\backslash(0,0)$ such that

$$a\left(\eta_{t-i_0}^+\right)^{\delta_0}+b\left(\eta_{t-i_0}^-\right)^{\delta_0}=c_{t,i_0},$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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

$$\mathcal{L}\left(a\left(\eta_{t-i_0}^{+}\right)^{\delta_0} + b\left(\eta_{t-i_0}^{-}\right)^{\delta_0} \middle| \mathcal{F}_{t,i_0}\right) = \mathcal{L}\left(c_{t,i_0} \middle| \mathcal{F}_{t,i_0}\right),$$

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

$$\frac{\mathcal{A}_{\boldsymbol{\vartheta}+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)} \text{ and } \frac{\mathcal{A}_{\boldsymbol{\vartheta}-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)}. \tag{36}$$

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 (35) 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,*

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

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

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

$$\frac{\partial \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \frac{\partial}{\partial \boldsymbol{\vartheta}} \left\{ \frac{\varepsilon_t^2}{(\sigma_t^{\delta_0})^{2/\delta_0}} + \frac{2}{\delta_0} \ln \sigma_t^{\delta_0} \right\} = \frac{2}{\delta_0} \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}}{\partial \boldsymbol{\vartheta}} \right\}, \tag{37}$$

and

$$\frac{\partial \ell_t^2(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} = \frac{2}{\delta_0} \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^{\delta_0}} \frac{\partial^2 \sigma_t^{\delta_0}}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\} + \frac{2}{\delta_0} \left\{ \frac{\delta_0 + 2}{\delta_0} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}}{\partial \boldsymbol{\vartheta}} \right\} \left\{ \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}}{\partial \boldsymbol{\vartheta}'} \right\}. \tag{38}$$

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

$$E\left\| \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\|^2 < \infty, \tag{39}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

and

$$E\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial^2\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta\partial\vartheta'}\right\|<\infty. \tag{40}$$

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

$$E\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta}\right\|^{2+8/\nu}<\infty. \tag{41}$$

Thanks to $E(\eta_t^2\mid\mathcal{F}_{t-1})=1$ in **A2**, the second expectation of (i) is still obtained by showing (39) and (40).

The existence of the moments in (39)–(41) is already known when π_0 is absent, (η_t) is iid and $\vartheta_0\in\overset{\circ}{\Theta}$ (see the proof of (i) and (ii) of Theorem 2.2 in HZ). We now explain the changes in the proof induced by our particular framework in the case $p=q=1$. The proof can be easily extended to the general case.

Since $\sigma_t^{\delta_0}=\sum_{j=0}^{\infty}\beta^j c_{t,j}$ with $c_{t,j}=\omega+\alpha_+\left(\varepsilon_{t-j-1}^+\right)^{\delta_0}+\alpha_-\left(\varepsilon_{t-j-1}^-\right)^{\delta_0}+\boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}$, we have

$$\begin{aligned}
\frac{\partial\sigma_t^{\delta_0}}{\partial\omega}&=\sum_{j=0}^{\infty}\beta^j=\frac{1}{1-\beta},\\
\frac{\partial\sigma_t^{\delta_0}}{\partial\alpha_+}&=\sum_{j=0}^{\infty}\beta^j\left(\varepsilon_{t-j-1}^+\right)^{\delta_0},\quad \frac{\partial\sigma_t^{\delta_0}}{\partial\alpha_-}=\sum_{j=0}^{\infty}\beta^j\left(\varepsilon_{t-j-1}^-\right)^{\delta_0},\\
\frac{\partial\sigma_t^{\delta_0}}{\partial\beta}&=\sum_{j=1}^{\infty}j\beta^{j-1}c_{t,j},\\
\frac{\partial\sigma_t^{\delta_0}}{\partial\boldsymbol{\pi}}&=\sum_{j=0}^{\infty}\beta^j\boldsymbol{x}_{t-j-1}.
\end{aligned} \tag{42}$$

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

$$\underline{\omega}:=\inf_{\vartheta\in\Theta}\sigma_t^{\delta_0}>0, \tag{43}$$

under the moment conditions in **A11**, we have (39) and (40) in Case B, and (40) and (41) in Case D. In Cases A and C, we have

$$\frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}}{\partial\omega}\le\frac{1}{\omega},\quad \frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}}{\partial\alpha_+}\le\frac{1}{\alpha_+},\quad \frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}}{\partial\alpha_-}\le\frac{1}{\alpha_-},\quad \frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}}{\partial\pi_i}\le\frac{1}{\pi_i},$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

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

$$\frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}}{\partial \beta} \leq \frac{1}{\beta \omega^s} \sum_{i=1}^{\infty} i \beta^{is} \left\{ \omega^s + \alpha_+^s \left(\varepsilon_{t-i-1}^+\right)^{\delta_0 s} + \alpha_-^s \left(\varepsilon_{t-i-1}^-\right)^{\delta_0 s} + (\boldsymbol{\pi}' \boldsymbol{x}_{t-i-1})^s \right\}.$$

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

$$\left\| \left(\varepsilon_{t-i-1}^+\right)^{\delta_0 s} \right\|_q < \infty, \quad \left\| \left(\varepsilon_{t-i-1}^-\right)^{\delta_0 s} \right\|_q < \infty \text{ and } \left\| (\boldsymbol{\pi}' \boldsymbol{x}_{t-i-1})^s \right\|_q < \infty,$$

for sufficiently small $s > 0$. It follows that (39) and (41), for any $\nu > 0$, hold true when all the components of $\boldsymbol{\vartheta}_0$ are nonzero. By the same arguments, we can even show the stronger result that for any $s_0 > 0$ there exists $V(\boldsymbol{\vartheta}_0)$ included in Θ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^{s_0} < \infty \quad \text{in Cases A and C.} \tag{44}$$

We obtain (40) 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 (38) and the fact that $E(\eta_t^2 \mid \mathcal{F}_{t-1}) = 1$. Using (37), we also have $\operatorname{var}(\partial \ell_t(\boldsymbol{\vartheta}_0)/\partial \boldsymbol{\vartheta}) = \boldsymbol{J}$. 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

$$\boldsymbol{c}' \boldsymbol{J} \boldsymbol{c} = \frac{4}{{\delta_0}^2} E \left\{ \frac{1}{\sigma_t^{2\delta_0}(\boldsymbol{\vartheta}_0)} \left(\boldsymbol{c}' \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right)^2 \right\} = 0.$$

Then a.s., $\boldsymbol{c}' \dfrac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = 0$. In view of (10), this implies

$$\boldsymbol{c}' \left(1, \left(\varepsilon_{t-1}^+\right)^{\delta_0}, \left(\varepsilon_{t-1}^-\right)^{\delta_0}, \ldots, \left(\varepsilon_{t-q}^+\right)^{\delta_0}, \left(\varepsilon_{t-q}^-\right)^{\delta_0}, \sigma_{t-1}^{\delta_0}(\boldsymbol{\vartheta}_0) \ldots, \sigma_{t-p}^{\delta_0}(\boldsymbol{\vartheta}_0), \boldsymbol{x}_{t-1} \right) = 0.$$

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*

$$\sqrt{n} \left\| \frac{\partial}{\partial \boldsymbol{\vartheta}} \widetilde{Q}_n(\boldsymbol{\vartheta}_0) - \frac{\partial}{\partial \boldsymbol{\vartheta}} Q_n(\boldsymbol{\vartheta}_0) \right\| = o(1) \quad \text{a.s.,} \tag{45}$$

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \widetilde{Q}_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) \right\| = o(1) \quad \text{a.s.} \tag{46}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

for some neighborhood $V(\vartheta_0)$ *of* ϑ_0,

$$\frac{\partial^2}{\partial\vartheta\partial\vartheta'}\widetilde{Q}_n(\vartheta_n)\to \boldsymbol{J} \text{ in probability when } \vartheta_n\to\vartheta_0 \text{ in probability,} \tag{47}$$

$$\sqrt{n}\frac{\partial Q_n(\vartheta_0)}{\partial\vartheta}\xrightarrow{d}\mathcal{N}\{0,\boldsymbol{I}\}. \tag{48}$$

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 $Q_n(\vartheta)$ and $\widetilde{Q}_n(\vartheta)$, (45) and (46) are entailed by

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{\partial\ell_t(\vartheta_0)}{\partial\vartheta}-\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta}\right\|\to 0 \quad \text{a.s.,} \tag{49}$$

$$\sup_{\vartheta\in V(\vartheta_0)\cap\Theta}\left\|\frac{\partial^2\ell_t(\vartheta)}{\partial\vartheta\partial\vartheta'}-\frac{\partial^2\widetilde{\ell}_t(\vartheta)}{\partial\vartheta\partial\vartheta'}\right\|\to 0 \quad \text{a.s.} \tag{50}$$

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

$$\left|\frac{\partial\ell_t(\vartheta_0)}{\partial\vartheta_i}-\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta_i}\right|\leqslant K\rho^t\left(1+K\eta_t^2\right)\left|1+\frac{1}{\sigma_t^{\delta_0}(\vartheta_0)}\frac{\partial\sigma_t^{\delta_0}(\vartheta_0)}{\partial\vartheta_i}\right|.$$

Under **A10**–**A11**, we have $E\eta_t^4<\infty$ and (39), and thus the expectation of the right-hand side of the inequality is bounded by $K\rho^t$. It follows that

$$\sum_{t=1}^{\infty}\left|\frac{\partial\ell_t(\vartheta_0)}{\partial\vartheta_i}-\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta_i}\right|$$

has a finite expectation, and thus is finite almost surely, which entails (49). The convergence (50) 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 (47), first note that

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

where

$$a_1=P\left(\sup_{\vartheta\in V(\vartheta_0)\cap\Theta}\left\|\frac{\partial^2}{\partial\vartheta\partial\vartheta'}\widetilde{Q}_n(\vartheta)-\frac{\partial^2}{\partial\vartheta\partial\vartheta'}Q_n(\vartheta)\right\|\geq\frac{\epsilon}{3}\right),$$

$$a_2=P\left(\sup_{\vartheta\in V(\vartheta_0)\cap\Theta}\left\|\frac{\partial^2}{\partial\vartheta\partial\vartheta'}Q_n(\vartheta)-\frac{\partial^2}{\partial\vartheta\partial\vartheta'}Q_n(\vartheta_0)\right\|\geq\frac{\epsilon}{3}\right),$$

$$a_3=P\left(\left\|\frac{\partial^2}{\partial\vartheta\partial\vartheta'}Q_n(\vartheta_0)-\boldsymbol{J}\right\|\geq\frac{\epsilon}{3}\right),\qquad a_4=P\left\{\vartheta_n\notin V(\vartheta_0)\right\},$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

for any $\epsilon > 0$ and any neighborhood $V(\vartheta_0)$. By the assumption that $\vartheta_n \to \vartheta_0$ in probability, we have $a_4 \to 0$ as $n \to \infty$. By (46), for any $\epsilon > 0$ and when $V(\vartheta_0)$ 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 $V(\vartheta_0)$ satisfying

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \sup_{\vartheta\in V(\vartheta_0)\cap\Theta} \left\| \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\ell_t(\vartheta) - \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\ell_t(\vartheta_0) \right\| \le \epsilon \quad \text{a.s.}$$

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

$$E \sup_{\vartheta\in V(\vartheta_0)\cap\Theta} \left\| \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\ell_t(\vartheta) \right\| < \infty, \tag{51}$$

for some neighborhood $V(\vartheta_0)$.

We begin by showing that, for any $s_0 > 0$, there exists $V(\vartheta_0)$ such that $V(\vartheta_0) \subset \Theta$ and

$$E \sup_{\vartheta\in V(\vartheta_0)} \left| \frac{\sigma_t^2(\vartheta_0)}{\sigma_t^2(\vartheta)} \right|^{s_0} < \infty. \tag{52}$$

Let us first consider the case $p = q = 1$. When $\vartheta_0 \in \overset{\circ}{\Theta}$, there exists $V(\vartheta_0)$ such that for all $\vartheta \in V(\vartheta_0)$

$$\frac{\sigma_t^{\delta_0}(\vartheta_0)}{\sigma_t^{\delta_0}(\vartheta)} \le K + K\sum_{j=0}^{\infty} \frac{\beta_0^j}{\beta^j} \left\{ \frac{\frac{\alpha_+}{\omega}\beta^j \varepsilon_{t-j-1}^{+\delta_0}}{1+\frac{\alpha_+}{\omega}\beta^j \varepsilon_{t-j-1}^{+\delta_0}} + \frac{\frac{\alpha_-}{\omega}\beta^j \varepsilon_{t-j-1}^{+\delta_0}}{1+\frac{\alpha_-}{\omega}\beta^j \varepsilon_{t-j-1}^{-\delta_0}} + \frac{\frac{1}{\omega}\beta^j \boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}}{1+\frac{1}{\omega}\beta^j \boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}} \right\}.$$

Now, consider the case where ϑ_0 belongs to the boundary of Θ. Assume for instance that $\alpha_{0+} = 0$, $\alpha_{0-} > 0$, $\beta_0 > 0$ and $\boldsymbol{\pi}_0 > 0$. Then there exists $V(\vartheta_0)$ such that for all $\vartheta \in V(\vartheta_0)$

$$\frac{\sigma_t^{\delta_0}(\vartheta_0)}{\sigma_t^{\delta_0}(\vartheta)} \le K + K\sum_{j=0}^{\infty} \frac{\beta_0^j}{\beta^j} \left\{ \frac{\frac{\alpha_-}{\omega}\beta^j \varepsilon_{t-j-1}^{+\delta_0}}{1+\frac{\alpha_-}{\omega}\beta^j \varepsilon_{t-j-1}^{-\delta_0}} + \frac{\frac{1}{\omega}\beta^j \boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}}{1+\frac{1}{\omega}\beta^j \boldsymbol{\pi}'\boldsymbol{x}_{t-j-1}} \right\}.$$

A similar inequality holds in the other cases. For all $\xi > 0$, there exists $V(\vartheta_0)$ such that $\beta_0 \le (1+\xi)\beta$ for all $\vartheta \in V(\vartheta_0)$. It follows that for all $\xi > 0$ and $s \in (0,1)$, using the inequality $x/(1+x) \le x^s$ for all $x \ge 0$, there exists $V(\vartheta_0)$ such that

$$\sup_{\vartheta\in V(\vartheta_0)} \frac{\sigma_t^{\delta_0}(\vartheta_0)}{\sigma_t^{\delta_0}(\vartheta)} \le K + K\sum_{j=0}^{\infty} (1+\xi)^j \rho^{js} \left| \varepsilon_{t-j-1} \right|^{\delta_0 s} + K\sum_{j=0}^{\infty} (1+\xi)^j \rho^{js} \left\| \boldsymbol{x}_{t-j-1} \right\|^{s}.$$

The same inequality can be obtained for general orders p and q. Without loss of generality, it can be assumed that $2s_0/\delta_0 \ge 1$. By the Minskowski inequality,

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

choosing s such that $E|\varepsilon_1|^{2ss_0} < \infty$ and $E\|\boldsymbol{x}_1\|^{2ss_0/\delta_0} < \infty$ and choosing for instance $\xi = \dfrac{1-\rho^s}{2\rho^s}$, we have

$$\left\| \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right\|_{s_0}^{\delta_0/2} = \left\| \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\sigma_t^{\delta_0}(\boldsymbol{\vartheta})} \right\|_{2s_0/\delta_0}$$

$$\leq K + K\sum_{j=0}^{\infty}(1+\xi)^j\rho^{js}\left\||\varepsilon_1|^{\delta_0 s}\right\|_{2s_0/\delta_0} + K\sum_{j=0}^{\infty}(1+\xi)^j\rho^{js}\left\|\|\boldsymbol{x}_1\|^s\right\|_{2s_0/\delta_0} < \infty.$$

We thus have shown (52).

Now, let us prove (51) in Cases A and C. In view of (38), **A2** and the Cauchy-Schwarz inequality entail

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\ell_t(\boldsymbol{\vartheta})\right\| \leq K\left\|1+\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left|\frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})}\right|\right\|_2\left\{\left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial^2\sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\|\right\|_2 + \left\|\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\|\right\|_4^2\right\}.$$

Hence, in Cases A and C, (51) comes from (44) and (52), and the analog (44) for second order derivatives.

It remains to prove (51) in Cases B and D. Using (38), **A2**, (52) and the Hölder inequality it suffices to show that

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial^2\sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\|^{1+s_1} < \infty, \quad E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^{\delta_0}}\frac{\partial\sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\|^{2+s_1} < \infty, \tag{53}$$

for some $s_1 > 0$. Since σ_t is bounded away from zero, to show (53), it suffices to establish that

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta}\left\|\frac{\partial^2\sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\|^{1+s_1} < \infty, \quad E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta}\left\|\frac{\partial\sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\|^{2+s_1} < \infty. \tag{54}$$

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

$$\frac{\partial\sigma_t^{\delta_0}}{\partial\boldsymbol{\vartheta}_i} = d_{i,0}(\boldsymbol{\vartheta}) + \sum_{k=1}^{\infty} d_{i,k}^+(\boldsymbol{\vartheta})\left(\varepsilon_{t-k}^+\right)^{\delta_0} + d_{i,k}^-(\boldsymbol{\vartheta})\left(\varepsilon_{t-k}^-\right)^{\delta_0} + \boldsymbol{\pi}_{i,k}'(\boldsymbol{\vartheta})\boldsymbol{x}_{t-k},$$

with

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

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

The second moment condition in (54) thus follows from **A11**. The first moment condition is obtained by doing similar developments for the second order derivatives. We thus have shown (51) in all cases, which completes the proof of (47).

By (37), Lemma 4 and **A2**, the last result, (48), is a consequence of the central limit theorem for square integrable martingale difference of Billingsley (1961). ■

Proof of Theorem 2. A Taylor expansion of $\widetilde{Q}_n(\vartheta)$ around ϑ_0 gives

$$\widetilde{Q}_n(\vartheta) - \widetilde{Q}_n(\vartheta_0) = \frac{\partial Q_n(\vartheta_0)}{\partial \vartheta'}(\vartheta - \vartheta_0) + \frac{1}{2}(\vartheta - \vartheta_0)' \boldsymbol{J}(\vartheta - \vartheta_0) + R_n(\vartheta),$$

where

$$R_n(\vartheta) = \left\{ \frac{\partial \widetilde{Q}_n(\vartheta_0)}{\partial \vartheta'} - \frac{\partial Q_n(\vartheta_0)}{\partial \vartheta'} \right\}(\vartheta - \vartheta_0) + \frac{1}{2}(\vartheta - \vartheta_0)' \left(\frac{\partial^2 \widetilde{Q}_n(\vartheta^*)}{\partial \vartheta \partial \vartheta'} - \boldsymbol{J} \right)(\vartheta - \vartheta_0),$$

and ϑ^* is between ϑ and ϑ_0. In view of (45), (46) and (47), as $n \to \infty$, we have

$$n R_n(\vartheta_n) = o_P\left\{\sqrt{n}(\vartheta_n - \vartheta_0)\right\} + o_P\left\{n\|\vartheta_n - \vartheta_0\|^2\right\}, \tag{55}$$

when $\vartheta_n - \vartheta_0 = o_P(1)$. Therefore

$$n R_n(\vartheta_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\vartheta_n - \vartheta_0) = O_P(1). \tag{56}$$

Letting

$$\boldsymbol{Z}_n = -\boldsymbol{J}^{-1}\sqrt{n}\frac{\partial Q_n(\vartheta_0)}{\partial \vartheta},$$

we obtain the following quadratic approximation of the objective function

$$\widetilde{Q}_n(\vartheta) - \widetilde{Q}_n(\vartheta_0) = \frac{1}{2n}\left\|\sqrt{n}(\vartheta - \vartheta_0) - \boldsymbol{Z}_n\right\|_{\boldsymbol{J}}^2 - \frac{1}{2n}\|\boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 + R_n(\vartheta). \tag{57}$$

Let

$$\vartheta_{\boldsymbol{Z}_n} = \arg\inf_{\vartheta \in \Theta}\left\|\sqrt{n}(\vartheta - \vartheta_0) - \boldsymbol{Z}_n\right\|_{\boldsymbol{J}}.$$

Note that

$$\sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) = \boldsymbol{Z}_n^{\mathcal{C}} \text{ for } n \text{ large enough}, \tag{58}$$

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

$$\begin{aligned} 0 &\le \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \\ &= 2n\left\{\widetilde{Q}_n(\widehat{\vartheta}_n) - \widetilde{Q}_n(\vartheta_{\boldsymbol{Z}_n})\right\} + 2n\left\{R_n(\vartheta_{\boldsymbol{Z}_n}) - R_n(\widehat{\vartheta}_n)\right\} \\ &\le 2n\left\{R_n(\widehat{\vartheta}_n) - R_n(\vartheta_{\boldsymbol{Z}_n})\right\}, \end{aligned}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

where the first inequality comes from the definition of $\vartheta_{\boldsymbol{Z}_n}$, the equality from (57), and the second inequality from the definition of $\widehat{\vartheta}_n$. By (58), it follows that

$$\left| \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \right| \leq 2n \left\{ R_n(\widehat{\vartheta}_n) - R_n(\vartheta_{\boldsymbol{Z}_n}) \right\}. \tag{59}$$

We now show that $R_n(\widehat{\vartheta}_n) - R_n(\vartheta_{\boldsymbol{Z}_n}) = o_P(1)$. By definition of $\vartheta_{\boldsymbol{Z}_n}$ and since $\vartheta_0 \in \Theta$, we also have

$$\left\| \sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}} \leq \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}.$$

The Minskowski inequality then entails that

$$\left\| \sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) \right\|_{\boldsymbol{J}} \leq \left\| \sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}} \leq 2\|\boldsymbol{Z}_n\|_{\boldsymbol{J}}.$$

By (48), we have $\|\boldsymbol{Z}_n\|_{\boldsymbol{J}} = O_P(1)$, and thus $\sqrt{n}(\vartheta_{\boldsymbol{Z}_n} - \vartheta_0) = O_P(1)$. In view of (56), this entails $nR_n(\vartheta_{\boldsymbol{Z}_n}) = o_P(1)$. By definition of $\widehat{\vartheta}_n$ and (57), we have

$$0 \leq 2n\widetilde{Q}_n(\vartheta_0) - 2n\widetilde{Q}_n(\widehat{\vartheta}_n) = \|\boldsymbol{Z}_n\|_J^2 - \left\| \sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n \right\|_J^2 - 2nR_n(\widehat{\vartheta}_n).$$

It follows that, by the c_r-inequality,

$$\begin{aligned} \left\| \sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \right\|_{\boldsymbol{J}}^2 &\leq 2\left(\left\| \sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n \right\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \right) \\ &\leq 4\|\boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - 4nR_n(\widehat{\vartheta}_n). \end{aligned}$$

We now show that $\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) = O_P(1)$, arguing by contradiction. The consistency of $\widehat{\vartheta}_n$ and (55) then entail that $nR_n(\widehat{\vartheta}_n) = o_P\left(\left\| \sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \right\|_{\boldsymbol{J}}^2 \right)$. It follows that $\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) = O_P(1)$. From (56), we deduce that $nR_n(\widehat{\vartheta}_n) = o_P(1)$. Thus, (59) can be rewritten as

$$\|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 = o_P(1). \tag{60}$$

Using the characterization (13) of the projection on a convex cone, we have

$$\begin{aligned} \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 &= \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 \\ &\quad + 2\left\langle \sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n^{\mathcal{C}}, \boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n \right\rangle_{\boldsymbol{J}} \\ &\geq \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 + \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2. \end{aligned}$$

Using (60), we thus obtain

$$\|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n^{\mathcal{C}}\|_{\boldsymbol{J}}^2 \leq \|\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 - \|\boldsymbol{Z}_n^{\mathcal{C}} - \boldsymbol{Z}_n\|_{\boldsymbol{J}}^2 = o_P(1).$$

The continuous mapping theorem and (48) entail $\boldsymbol{Z}_n^{\mathcal{C}} \xrightarrow{d} \boldsymbol{Z}^{\mathcal{C}}$, and the conclusion follows. ■

Proof of Proposition 1. We only show (18) because (17) and (19) are obtained directly from arguments used in the proof of Theorem 1. Let

$$\boldsymbol{I}_n(\vartheta) = \frac{4}{{\delta_0}^2} \frac{1}{n} \sum_{t=1}^{n} S_t(\vartheta), \quad S_t(\vartheta) = \left(\frac{\varepsilon_t^4}{\sigma_t^4(\vartheta)} - 1 \right) \frac{1}{\sigma_t^{2\delta_0}(\vartheta)} \frac{\partial \sigma_t^{\delta_0}(\vartheta)}{\partial \vartheta} \frac{\partial \sigma_t^{\delta_0}(\vartheta)}{\partial \vartheta'}. \tag{61}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

The difference between $\boldsymbol{I}_n(\widehat{\boldsymbol{\vartheta}}_n)$ and $\widehat{\boldsymbol{I}}_n$ is due to the initial values used to compute $\widetilde{\sigma}_t^{\delta_0}(\boldsymbol{\vartheta})$. Because the difference $\sigma_t^{\delta_0}(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t^{\delta_0}(\boldsymbol{\vartheta})$ 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 $V(\boldsymbol{\vartheta}_0)$ 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 $V(\boldsymbol{\vartheta}_0)$ such that

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \|S_t(\boldsymbol{\vartheta})\| < \infty, \tag{62}$$

and the result will follow from the ergodic theorem applied to the sequence $\left\{\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \|S_t(\boldsymbol{\vartheta}) - S_t(\boldsymbol{\vartheta}_0)\|\right\}_t$, 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, using the Hölder inequality, (62) follows from **A10**, (44) and (52). In Case B, (62) follows from **A10**, (52) and the second part of (53). In Case D, (62) follows from **A10**, **A11**, (52) and the second part of (53) with $s_1 = 8/\nu$. ■

Proof of Theorem 3. The proof is the same as that of Theorem 2, except that the proofs of Lemmas 4 and 5 have to be modified in Cases B and D. First note that $\boldsymbol{\theta}_0 > 0$ entails that for any $s_0 > 0$ there exists $V(\boldsymbol{\vartheta}_0)$ included in Θ such that

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^{\delta_0}} \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} \right\|^{s_0} < \infty \text{ and } E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^{\delta_0}} \frac{\partial^2 \sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \right\|^{s_0} < \infty. \tag{63}$$

To show (39), it remains to show that

$$E \left\| \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\pi}} \right\|^2 < \infty,$$

which follows from (42) and the existence of $E\|\boldsymbol{x}_t\|^2$ implied by **A12**. Similarly the condition $E\|\boldsymbol{x}_t\|^{2+8/\nu} < \infty$ in **A12** entails (41) in Case D, and **A12** entails (40) in the two Cases B and D. The results of Lemma 4 thus continue to hold. Turning to the proof of Lemma 5, in view of (38), one can prove (51) by showing that

$$E \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \frac{1}{\sigma_t^{\delta_0}(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} \frac{1}{\sigma_t^{\delta_0}(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^{\delta_0}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}'} \right\| < \infty, \tag{64}$$

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

$$E \sup_{\vartheta \in V(\vartheta_0)\cap\Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\vartheta_0)}{\sigma_t^2(\vartheta)} \frac{1}{\sigma_t^{\delta_0}(\vartheta)} \frac{\partial^2 \sigma_t^{\delta_0}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\vartheta) \right\| < \infty, \tag{65}$$

$$E \sup_{\vartheta \in V(\vartheta_0)\cap\Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\vartheta_0)}{\sigma_t^2(\vartheta)} \frac{\partial \sigma_t^{\delta_0}(\vartheta)}{\partial\boldsymbol{\pi}} \frac{\partial \sigma_t^{\delta_0}(\vartheta)}{\partial\boldsymbol{\pi}'} \right\| < \infty,$$

$$E \sup_{\vartheta \in V(\vartheta_0)\cap\Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\vartheta_0)}{\sigma_t^2(\vartheta)} \frac{\partial^2 \sigma_t^{\delta_0}(\vartheta)}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}'} \right\| < \infty, \tag{66}$$

and similar moments for cross derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. Using Hölder's inequality, the moment conditions (64)–(65) come from (52), (63) and $E(\eta_t^2 \mid \mathcal{F}_{t-1}) = 1$, and the moment conditions (66) come from the same arguments and **A12**. ■

Proof of Corollaries 1, 2 and 3. As in Francq and Zakoïan (2009b). ■

Proof of Theorem 4. It suffices to show (30), the other results being obvious. Under **A1** and **A5**, let $\{\sigma_{\delta,t}(\vartheta)\}_t$ be the stationary, ergodic and nonanticipative solution of

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

We have

$$\begin{aligned} \sigma_{\delta,t}^{\delta}(\vartheta) &= \omega_{t-1} + \sum_{i=1}^{\infty} d_{i+} \left(\varepsilon_{t-i}^{+}\right)^{\delta} + d_{i-} \left(\varepsilon_{t-i}^{-}\right)^{\delta} \\ &= c_{t-1} + c_{t-2,+} \left(\eta_{t-1}^{+}\right)^{\delta} + c_{t-2,-} \left(\eta_{t-1}^{-}\right)^{\delta}, \end{aligned}$$

where $(d_{i+})_{i\geq 1} = \{d_{i+}(\vartheta)\}_{i\geq 1}$ and $(d_{i-})_{i\geq 1} = \{d_{i-}(\vartheta)\}_{i\geq 1}$ are summable sequences of positive constants, $\omega_{t-1} = \omega_{t-1}(\vartheta)$ is measurable with respect to $\{\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \dots\}$, $c_{t-1} = c_{t-1}(\vartheta)$ is measurable with respect to $\mathcal{F}_{t,1}$ defined in **A4** (i.e., the sigma-field generated by $\{\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \dots; \eta_{t-2}, \eta_{t-3}, \dots\}$), and where

$$\begin{aligned} c_{t-2,+} &= c_{t-2,+}(\vartheta) = d_{1+}(\vartheta)\, \sigma_{\delta_0,t-1}^{\delta}(\vartheta_0) \quad \text{and} \\ c_{t-2,-} &= c_{t-2,-}(\vartheta) = d_{1-}(\vartheta)\, \sigma_{\delta_0,t-1}^{\delta}(\vartheta_0) \end{aligned}$$

are measurable with respect to $\mathcal{F}_{t-2}$. We first show the identifiability condition

$$\frac{\sigma_{\delta,t}(\vartheta)}{\sigma_{\delta_0,t}(\vartheta_0)} = 1 \quad \text{a.s.} \qquad \Longrightarrow \qquad \delta = \delta_0. \tag{67}$$

For $\theta = (a, c, r, b, d) \in (0,\infty)^3 \times [0,\infty)^2$, let the function $\varphi_\theta : [0,\infty) \to (0,\infty)$ defined by $\varphi_\theta(x) = \{c + dx^r\}^{1/r} / \{a + bx\}$. We have $\varphi_\theta'(x) = 0$ if and only if $adx^{r-1} = bc$. Therefore the equation $\varphi_\theta(x) = 1$ cannot have more than two

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

solutions in $[0,\infty)$, except if i) $r=1$, $a=c$ and $b=d$ or ii) $b=d=0$ and $c=a^r$. Note that, conditional on $\mathcal{F}_{t-1}$, we have

$$\left(\frac{\sigma_{\delta,t}(\vartheta)}{\sigma_{\delta_0,t}(\vartheta_0)}\right)^{\delta_0}=\varphi_{\theta_+}(|\eta_{t-1}|^{\delta_0})1_{\{\eta_{t-1}\geq 0\}}+\varphi_{\theta_-}(|\eta_{t-1}|^{\delta_0})1_{\{\eta_{t-1}\leq 0\}}, \tag{68}$$

where $\theta_+=(a,c,r,b_+,d_+)$ and $\theta_-=(a,c,r,b_-,d_-)$, with $a=c_{t-1}(\vartheta_0)$, $c=c_{t-1}(\vartheta)$, $b_+=c_{t-2,+}(\vartheta_0)$, $d_+=c_{t-2,+}(\vartheta)$, $b_-=c_{t-2,-}(\vartheta_0)$, $d_-=c_{t-2,-}(\vartheta)$ and $r=\delta/\delta_0$. In view of the cases i) and ii), under **A4*** the first equality in (67) entails either $\delta=\delta_0$ or $d_{1+}(\vartheta)=d_{1+}(\vartheta_0)=d_{1-}(\vartheta)=d_{1-}(\vartheta_0)=0$. In the latter case, (68) holds with η_{t-1} replaced by η_{t-2}. Iterating the arguments, under **A4***, the first equality in (67) entails either $\delta=\delta_0$ or $d_{i+}(\vartheta_0)=d_{i-}(\vartheta_0)=0$ for all $i\geq 1$. The latter case is precluded by **A7**. We thus have shown (67), which entails (30) by Theorem 1 and the arguments of its proof. ■

NOTES

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

2. The authors are grateful to Sébastien Laurent who has kindly provided them with the data set, and who has provided very useful comments on a preliminary version of this paper.

REFERENCES

Ali, G. (2013) EGARCH, GJR-GARCH, TGARCH, AVGARCH, NGARCH, IGARCH and APARCH models for pathogens at marine recreational sites. *Journal of Statistical and Econometric Methods* 2, 57–73.

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.

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

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

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

Bollerslev, T. (2008) Glossary to ARCH (GARCH). In T. Bollerslev, J.R. Russell, & M. Watson (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press, 137–163.

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

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

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

Ding, Z., C. Granger, & 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., D.F. Hendry, & J.F. Richard (1983) Exogeneity. *Econometrica* 51, 277–304.

Engle, R.F. & 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.

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press

Fisher, T.J. & C.M. Gallagher (2012) New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association* 107, 777–787.

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

Francq, C. & 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. & J.-M. Zakoian (2009a) Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis* 30, 449–465.

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

Francq, C. & J.-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Fuertes, A.M., M. Izzeldin, & 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., R. Jaganathan, & 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.

Guo, S., S. Ling, & K. Zhu (2014) Factor double autoregressive models with application to simultaneous causality testing. *Journal of Statistical Planning and Inference* 148, 82–94.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

Hamadeh, T. & 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. & 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.

Kristensen, D. & A. Rahbek (2005) Asymptotics of the QMLE for a class of ARCH(q) models. *Econometric Theory* 21, 946–961.

Laurent, S., C. Lecourt, & F.C. Palm (2016) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis* 100, 383–400.

Li, W.K. & T.K. Mak (1994) On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.

Ling, S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.

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

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

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

Pedersen, R.S. (2017) Inference and testing on the boundary in extended constant conditional correlation GARCH models. *Journal of Econometrics* 196, 23–36.

Sucarrat, G. & 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*. 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.

https://doi.org/10.1017/S0266466617000512 Published online by Cambridge University Press