Journal of Financial Econometrics, 2011, Vol. 9, No. 4, 619–656

Merits and Drawbacks of Variance Targeting in GARCH Models

CHRISTIAN FRANCQ
Center for Research in Economics and Statistics and Université Lille 3

LAJOS HORVÁTH
University of Utah

JEAN-MICHEL ZAKOÏAN
Center for Research in Economics and Statistics and Université Lille 3

ABSTRACT

Variance targeting estimation (VTE) is a technique used to alleviate the numerical difficulties encountered in the quasi-maximum likelihood estimation (QMLE) of GARCH models. It relies on a reparameterization of the model and a first-step estimation of the unconditional variance. The remaining parameters are estimated by quasi maximum likelihood (QML) in a second step. This paper establishes the asymptotic distribution of the estimators obtained by this method in univariate GARCH models. Comparisons with the standard QML are provided and the merits of the variance targeting method are discussed. In particular, it is shown that when the model is misspecified, the VTE can be superior to the QMLE for long-term prediction or value-at-risk calculation. An empirical application based on stock market indices is proposed. (*JEL*: C13, C22)

KEYWORDS: Consistency and asymptotic normality, GARCH, heteroskedastic time series, quasi-maximum likelihood estimation, value-at-risk, variance targeting estimator

More than two decades after the introduction of ARCH models and their generalization (Engle 1982; Bollerslev 1986), the properties of GARCH-type sequences

We should like to thank the co-editor (Eric Renault), two anonymous referees and the participants of the NBER-NSF Time Series Conference 2009 for helpful comments in particular Dennis Kristensen for pointing out to us an important reference. This work is partially supported by the Agence Nationale de la Recherche (ANR) (Project ECONOM & RISKS, ANR 2010 blanc 1804 03). Address correspondence to Jean-Michel Zakoïan, CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France, or e-mail: zakoian@ensae.fr.

doi: 10.1093/jjfinec/nbr004
Advance Access publication July 18, 2011

are well understood and general statistical methods have been established to work with this type of sequences (see the recent book by Francq and Zakoïan 2010 for an overview of GARCH models). In recent years, special attention has been given to the asymptotic properties of the Gaussian quasi-maximum likelihood estimation (QMLE) (see Berkes, Horváth, and Kokoszka 2003; Francq and Zakoïan 2004, and the monograph by Straumann 2005, among others). While many other estimation methods have been proposed for GARCH-type models (for instance, the L_p-estimators of Horváth and Liese 2004, the self-weighted QMLE of Ling 2007), QMLE can be recommended for at least two reasons: (i) it is consistent under very mild conditions, in particular, it is robust to the distribution of the underlying independent and identically distributed (i.i.d.) process, and (ii) the moment conditions required for the asymptotic normality of the QMLE do not concern the observed process but only the underlying i.i.d. process (a finite fourth moment).

However, practitioners are sometimes reluctant to directly apply the QMLE to their data. They often make use of closed-form estimators to reduce the dimensionality of the parameter space or to speed-up the convergence of the optimization routines. Such estimators are particularly attractive for the estimation of multivariate GARCH models or when a large number of univariate GARCH models have to be estimated (see Bauwens and Rombouts 2007). In the framework of a scalar BEKK (Engle and Kroner 1995), Engle and Mezrich (1996) proposed a two-step estimation method, the so-called "variance targeting estimation" (VTE) method. The method is based on a reparameterization of the volatility equation in which the intercept is replaced by the returns unconditional variance (the "long-run" variance). A first-step estimator of the unconditional variance is computed while, conditioning on this estimate, the remaining parameters are estimated by quasi maximum likelihood (QML) in a second step. This method has been recommended in the practitioner-oriented textbooks by Hull (2003) and Christoffersen (2003). See Shephard and Sheppard (2010) for an example of application of the VTE method in univariate models.

The main aim of this paper is to derive the asymptotic properties of the VTE. This problem has been considered by Kristensen and Linton (2004), under conditions we will discuss below. While the VTE method facilitates the estimation of parameters in GARCH models, even in the simple univariate GARCH(1,1), it is not clear if this advantage is not paid for in terms of asymptotic accuracy loss, when the VTE is compared to the QMLE. Intuitively, even if the sample variance converges to the population variance, the use of a two-step procedure should deteriorate the asymptotic precision of the GARCH QML estimates for Gaussian i.i.d. errors. The magnitude of the accuracy loss, however, cannot be intuited. Moreover, for nonGaussian i.i.d. errors, the superiority of QMLE over the VTE cannot be taken for granted.

On the other hand, the potential merits of the VTE may not be limited to numerical simplicity. This procedure guarantees that the estimated unconditional variance of the GARCH model is equal to the sample variance. It is therefore possible that, in case of misspecification, that is, when the true underlying process is not a GARCH, the GARCH approximation provided by the VTE is superior, in some sense, to that obtained by QMLE. This issue will be examined through the problems of long-term prediction and value-at-risk (VaR) calculation.

The VTE is a two-step estimator, the marginal variance being estimated in the first step and then plugged into the quasilikelihood in a second step. Two-step estimators are quite common in econometrics; see, for instance, Pastorello, Patilea, and Renault (2003) and Newey and McFadden (1994). From a technical point of view, the main difficulty arises when, as is the case of the VTE, the first-step estimator is plugged into a criterion not directly into a formula giving the second-step estimator.

The paper is organized as follows. Section 1 describes the reparameterization of the standard GARCH(1,1) model and provides the asymptotic properties of the VTE. Section 2 proposes an extension to the GARCH(p,q) model. Section 3 examines the performances of the VTE, by comparison to the QMLE, in the cases of well specified and misspecified models. An empirical comparison based on eleven, daily, and weekly, stock indices is also provided. Section 5 concludes and outlines topics for future research. Proofs are relegated to the appendix.

1 ASYMPTOTIC DISTRIBUTION OF THE VTE IN THE GARCH(1,1) CASE

Consider the GARCH(1,1) model

$$\begin{cases} \epsilon_t = \sqrt{h_t}\eta_t, \\ h_t = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0 h_{t-1}, \quad \forall\, t \in \mathbb{Z}, \end{cases} \tag{1}$$

where $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0)'$ is an unknown parameter,

$$(\eta_t) \text{ is a sequence of i.i.d random variables} \tag{2}$$

such that

$$E\eta_t^2 = 1, \tag{3}$$

and

$$\omega_0 > 0, \quad \alpha_0 \geqslant 0, \quad \beta_0 \geqslant 0. \tag{4}$$

Under the condition

$$\alpha_0 + \beta_0 < 1, \tag{5}$$

Bollerslev (1986) showed that this model admits a second-order stationary solution (ϵ_t), whose unconditional variance is given by

$$\gamma_0 := \sigma^2(\omega_0, \alpha_0, \beta_0) = \frac{\omega_0}{1 - \alpha_0 - \beta_0} := \frac{\omega_0}{\kappa_0}.$$

A reparameterization of the model with $\boldsymbol{\vartheta}_0 = (\gamma_0, \alpha_0, \kappa_0)'$ yields

$$h_t = h_{t-1} + \kappa_0(\gamma_0 - h_{t-1}) + \alpha_0(\epsilon_{t-1}^2 - h_{t-1}), \tag{6}$$

which allows us to interpret κ_0 as the speed of mean reversion in variance (see Christoffersen 2008). Note that $1 - \kappa_0$ is often referred to as the volatility persistence (see, e.g., Meddahi and Renault 2004). Writing

$$h_t = \kappa_0\gamma_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0 h_{t-1}, \quad \kappa_0 + \alpha_0 + \beta_0 = 1,$$

one can also interpret the volatility at time t, h_t, as a weighted average of the long-run variance γ_0, of the square of the last return ϵ_{t-1}^2 and of the previous volatility h_{t-1}. In this average, κ_0 is the weight of the long-run variance. Note that in this reparameterization, constraints (4) and (5) become

$$\kappa_0, \gamma_0 > 0, \quad \alpha_0 \geqslant 0, \quad \kappa_0 + \alpha_0 \leqslant 1. \tag{7}$$

Let $(\epsilon_1, \dots, \epsilon_n)$ be a realization of length n of the unique nonanticipative second-order stationary solution (ϵ_t) to Model (1) that satisfies (3) and (7). In this framework, the VTE involves (i) reparameterization the model as in (6), (ii) estimation γ_0 by the sample variance and then $\boldsymbol{\lambda}_0 := (\alpha_0, \kappa_0)'$ by the QML estimator.

The QMLE of $\boldsymbol{\theta}_0$ is denoted by $\hat{\boldsymbol{\theta}}_n^* := (\hat{\omega}_n^*, \hat{\alpha}_n^*, \hat{\beta}_n^*)'$. Two asymptotically consistent estimators of γ_0 are the sample variance and the QML-based estimator, given by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 \quad \text{and} \quad \sigma^2(\hat{\boldsymbol{\theta}}_n^*) = \frac{\hat{\omega}_n^*}{1 - \hat{\alpha}_n^* - \hat{\beta}_n^*}.$$

Horváth, Kokoszka, and Zitikis (2006) showed that the differences $\hat{\sigma}_n^2 - \gamma_0$ and $\hat{\sigma}_n^2 - \sigma^2(\hat{\boldsymbol{\theta}}_n^*)$ are asymptotically normal with different asymptotic variances. They also use the latter difference as a statistic for testing that the model is correctly specified.

Consider a parameter space $\Lambda \subset \{(\alpha, \kappa) | \alpha \geqslant 0, \kappa > 0, \alpha + \kappa \leqslant 1\}$. Set $\boldsymbol{\lambda}_0 = (\alpha_0, \kappa_0)'$ and write $\boldsymbol{\lambda} = (\alpha, \kappa)'$ and $\beta = 1 - \alpha - \kappa$. At this stage, we use the convention that all vectors considered in the sequel are column vectors even when, for simplicity, they are written as row vectors. In particular, we write $\boldsymbol{\vartheta}_0 = (\gamma_0, \boldsymbol{\lambda}_0)$ instead of $\boldsymbol{\vartheta}_0 = (\gamma_0, \boldsymbol{\lambda}_0')'$. At the point $\boldsymbol{\vartheta} = (\gamma, \boldsymbol{\lambda}) \in (0, \infty) \times \Lambda$, the Gaussian quasilikelihood of the sample is given by

$$\tilde{L}_n(\boldsymbol{\vartheta}) = \tilde{L}_n(\gamma, \boldsymbol{\lambda}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\boldsymbol{\vartheta})}} \exp\left\{-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2(\boldsymbol{\vartheta})}\right\},$$

where the $\tilde{\sigma}_t^2(\boldsymbol{\vartheta})$'s are defined recursively, for $t \geqslant 1$, by

$$\tilde{\sigma}_t^2(\boldsymbol{\vartheta}) = \kappa\gamma + \alpha\epsilon_{t-1}^2 + (1 - \kappa - \alpha)\tilde{\sigma}_{t-1}^2(\boldsymbol{\vartheta}) \tag{8}$$

with the initial values ϵ_0 and $\tilde{\sigma}_0^2(\boldsymbol{\vartheta}) := \sigma_0^2$. Since the parameter γ_0 is estimated by the sample variance $\hat{\sigma}_n^2$, the variance targeting version of the Gaussian quasi-likelihood function is

$$L_n(\boldsymbol{\lambda}) = \tilde{L}_n(\hat{\sigma}_n^2, \boldsymbol{\lambda}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{t,n}^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_{t,n}^2}\right),$$

where

$$\sigma_{t,n}^2 := \sigma_{t,n}^2(\boldsymbol{\lambda}) = \kappa\hat{\sigma}_n^2 + \alpha\epsilon_{t-1}^2 + (1-\kappa-\alpha)\sigma_{t-1,n}^2$$

with $\sigma_{0,n}^2 = \sigma_0^2$. A variance targeting estimator (VTE) of $\boldsymbol{\lambda}_0$ is defined as any measurable solution $\hat{\boldsymbol{\lambda}}_n$ of

$$\hat{\boldsymbol{\lambda}}_n = \arg\max_{\boldsymbol{\lambda}\in\Lambda} L_n(\boldsymbol{\lambda}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \tilde{\mathbf{l}}_n(\boldsymbol{\lambda}), \tag{9}$$

where

$$\tilde{\mathbf{l}}_n(\boldsymbol{\lambda}) = n^{-1}\sum_{t=1}^{n} \ell_{t,n} \quad \text{and} \quad \ell_{t,n} := \ell_{t,n}(\boldsymbol{\lambda}) = \frac{\epsilon_t^2}{\sigma_{t,n}^2} + \log\sigma_{t,n}^2. \tag{10}$$

Note that $\sigma_{t,n}^2 = \tilde{\sigma}_t^2(\hat{\sigma}_n^2, \boldsymbol{\lambda})$. For any $\boldsymbol{\vartheta} = (\gamma, \boldsymbol{\lambda}) \in (0,\infty)\times\Lambda$, we have $0 \leqslant 1-\kappa-\alpha < 1$, and one can define the strictly stationary and ergodic process

$$\sigma_t^2(\boldsymbol{\vartheta}) = \kappa\gamma + \alpha\epsilon_{t-1}^2 + (1-\kappa-\alpha)\sigma_{t-1}^2(\boldsymbol{\vartheta}) = \sum_{i=0}^{\infty}(1-\kappa-\alpha)^i(\kappa\gamma + \alpha\epsilon_{t-i-1}^2). \tag{11}$$

Note that $h_t = \sigma_t^2(\gamma_0, \boldsymbol{\lambda}_0)$. We denote by $\hat{\boldsymbol{\vartheta}}_n = (\hat{\sigma}_n^2, \hat{\boldsymbol{\lambda}}_n)$ the VTE of $\boldsymbol{\vartheta}_0$.

To show the strong consistency and the asymptotic normality of the VTE, the following assumptions will be made:

A1: $\boldsymbol{\lambda}_0$ belongs to Λ and Λ is compact.
A2: η_t^2 has a nondegenerate distribution.
A3: $\alpha_0^2\left(E\eta_t^4 - 1\right) + (1-\kappa_0)^2 < 1$.
A4: $\boldsymbol{\lambda}_0$ belongs to the interior of Λ.

Note that **A3** is the necessary and sufficient condition for $E\epsilon_t^4 < \infty$.

Theorem 1.1. Under assumptions **A1** and **A2**, $\alpha_0 \neq 0$ and (2)–(5), the VTE satisfies

$$\hat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0$$

almost surely as $n \to \infty$ and, under the additional assumptions **A3** and **A4**, we have

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}(0, (E\eta_0^4 - 1)\boldsymbol{\Sigma}), \tag{12}$$

where the matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} b & -b\boldsymbol{K}'\boldsymbol{J}^{-1} \\ -b\boldsymbol{J}^{-1}\boldsymbol{K} & \boldsymbol{J}^{-1} + b\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} \end{pmatrix}$$

is nonsingular with

$$b = \frac{(\alpha_0+\kappa_0)^2}{\kappa_0^2} E(h_t^2) = \frac{(\alpha_0+\kappa_0)^2\gamma^2(2-\kappa_0)}{\kappa_0\{1-\alpha_0^2\left(E\eta_t^4-1\right)-(1-\kappa_0)^2\}}$$

and

$$\boldsymbol{J} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\lambda}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\lambda}'}\right)_{2\times 2}, \quad \boldsymbol{K} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\lambda}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\gamma}\right)_{2\times 1}. \tag{13}$$

The asymptotic normality in (12) was obtained by Kristensen and Linton (2004). In their paper, the asymptotic variance has a more complex form than our $\boldsymbol{\Sigma}$. It involves the term $\mathrm{var}(\epsilon_t^2) + 2\sum_{s=1}^{\infty}\mathrm{cov}(\epsilon_0^2, \epsilon_s^2)$ that requires a Bartlett-type estimator. On the contrary, the matrix $\boldsymbol{\Sigma}$ can be estimated using empirical means since $\boldsymbol{J}$ and $\boldsymbol{K}$ have the forms of expectations.

Theorem 1.1 also complements the paper of Horváth, Kokoszka, and Zitikis (2006), where the asymptotic distribution of $\sqrt{n}(\hat{\sigma}_n^2 - \gamma_0)$ was derived. Letting $\hat{\boldsymbol{\lambda}}_n = (\hat{\lambda}_{1n}, \hat{\lambda}_{2n})$, the VTE of the original parameter $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0)$ is defined by $\hat{\boldsymbol{\theta}}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)$, where $\hat{\omega}_n = \hat{\lambda}_{2n}\hat{\sigma}_n^2, \hat{\alpha}_n = \hat{\lambda}_{1n}, \hat{\beta}_n = 1 - \hat{\lambda}_{1n} - \hat{\lambda}_{2n}$. Theorem 1.1 yields the following result.

Corollary 1. Under the assumptions of Theorem 1.1, the VTE of $\boldsymbol{\theta}_0$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, (E\eta_0^4 - 1)\boldsymbol{L}'\boldsymbol{\Sigma}\boldsymbol{L}), \quad \boldsymbol{L} = \begin{pmatrix} 1-\alpha_0-\beta_0 & 0 & 0 \\ 0 & 1 & -1 \\ \omega_0(1-\alpha_0-\beta_0)^{-1} & 0 & -1 \end{pmatrix}.$$

It is important to note that the asymptotic normality of the VTE requires the existence of $E(\epsilon_t^4)$, whereas the strict stationarity and the existence of $E(\eta_t^4)$ are sufficient for the asymptotic normality of the QMLE (see Berkes, Horváth, and Kokoszka 2003; Francq and Zakoïan 2004). The QMLE of $\boldsymbol{\vartheta}_0$ is denoted by $\hat{\boldsymbol{\vartheta}}_n^* = (\hat{\gamma}_n^*, \hat{\alpha}_n^*, \hat{\kappa}_n^*)$. The asymptotic variance matrix of $\hat{\boldsymbol{\vartheta}}_n^*$ is $(E\eta_0^4 - 1)\boldsymbol{\Sigma}^*$, where

$$(\boldsymbol{\Sigma}^*)^{-1} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right) = \begin{pmatrix} \frac{\kappa_0^2}{(\alpha_0+\kappa_0)^2}E(1/h_t^2) & \boldsymbol{K}' \\ \boldsymbol{K} & \boldsymbol{J} \end{pmatrix}.$$

The following corollary allows us to compare the asymptotic variances of the QMLE and VTE.

Corollary 2. Under the assumptions of Theorem 1.1, the asymptotic variance $(E\eta_0^4 - 1)\mathbf{\Sigma}$ of the VTE and the asymptotic variance $(E\eta_0^4 - 1)\mathbf{\Sigma}^*$ of the QMLE of $\boldsymbol{\vartheta}_0$ satisfy

$$\mathbf{\Sigma} - \mathbf{\Sigma}^* = (b - a)\boldsymbol{C}\boldsymbol{C}',$$

where

$$\boldsymbol{C} = \begin{pmatrix} 1 \\ -\boldsymbol{J}^{-1}\boldsymbol{K} \end{pmatrix}, \quad a = \left\{ \frac{\kappa_0^2}{(\alpha_0 + \kappa_0)^2} E(1/h_t^2) - \boldsymbol{K}'\boldsymbol{J}^{-1}\boldsymbol{K} \right\}^{-1}.$$

Note that $a > 0$ because $\det \mathbf{\Sigma}^* = a \det \boldsymbol{J}^{-1}$. It will be shown that $b - a \geqslant 0$ (in a more general setting, see Proposition 3 below), which implies that the VTE cannot be asymptotically more accurate than the QMLE. Corollary 2 shows that, as expected, the VTE becomes much less accurate than the QMLE when $\alpha_0^2(E\eta_t^4 - 1) + (1 - \kappa_0)^2$ approaches 1. In general $b > a$, as will be illustrated in Section 3.

2 EXTENSION TO THE GENERAL GARCH(P, Q) CASE

In this section, we consider the general GARCH(p, q) model

$$\begin{cases} \epsilon_t = \sqrt{h_t}\eta_t, \\ h_t = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}h_{t-j}, \quad \forall t \in \mathbb{Z}, \end{cases} \tag{14}$$

where (η_t) satisfies (2) and (3) and the coefficients satisfy

$$\omega_0 > 0, \quad \alpha_{0i} \geqslant 0 \quad \forall\, i \in \{1, \dots, q\}, \quad \beta_{0j} \geqslant 0 \quad \forall\, j \in \{1, \dots, p\}. \tag{15}$$

Under the condition,

$$\sum_{i=1}^q \alpha_{0i} + \sum_{j=1}^p \beta_{0j} < 1. \tag{16}$$

the observations have finite variance $\gamma_0 = \omega_0\{1 - \sum_{i=1}^q \alpha_{0i} - \sum_{j=1}^p \beta_{0j}\}^{-1}$. In this section, we parameterize the model with

$$\boldsymbol{\vartheta}_0 = (\gamma_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p}) = (\gamma_0, \boldsymbol{\lambda}_0) \in (0, \infty) \times \Lambda,$$

where Λ is a subset of the simplex $\{\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{p+q}) : \lambda_i \geqslant 0\ \forall i \in \{1, \dots, p + q\}, \sum_{i=1}^{p+q} \lambda_i < 1\}$.

The VTE of $\boldsymbol{\vartheta}_0$ is $\hat{\boldsymbol{\vartheta}}_n = (\hat{\sigma}_n^2, \hat{\boldsymbol{\lambda}}_n)$, where $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \epsilon_t^2$,

$$\hat{\boldsymbol{\lambda}}_n = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \tilde{\mathrm{I}}_n(\boldsymbol{\lambda}), \quad \tilde{\mathrm{I}}_n(\boldsymbol{\lambda}) = n^{-1}\sum_{t=1}^n \tilde{\ell}_t(\hat{\sigma}_n^2, \boldsymbol{\lambda}), \quad \tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})} + \log \tilde{\sigma}_t^2(\boldsymbol{\vartheta}),$$

and the $\tilde{\sigma}_t^2(\boldsymbol{\vartheta})$'s are defined recursively, for $t \geqslant 1$, by

$$\tilde{\sigma}_t^2(\boldsymbol{\vartheta}) = \tilde{\sigma}_t^2(\gamma, \boldsymbol{\lambda}) = \gamma \left(1 - \sum_{i=1}^{p+q} \lambda_i\right) + \sum_{i=1}^{q} \lambda_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \lambda_{q+j} \tilde{\sigma}_{t-j}^2(\boldsymbol{\vartheta})$$

with fixed initial values for $\epsilon_0, \ldots, \epsilon_{1-q}$ and $\tilde{\sigma}_0^2(\boldsymbol{\vartheta}), \ldots, \tilde{\sigma}_{1-p}^2(\boldsymbol{\vartheta})$. Define $\mathcal{A}_{\boldsymbol{\vartheta}}(z) = \sum_{i=1}^{q} \lambda_i z^i$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = 1 - \sum_{j=1}^{p} \lambda_{q+j} z^j$, with the convention $\mathcal{A}_{\boldsymbol{\vartheta}}(z) = 0$ if $q = 0$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = 1$ if $p = 0$. We need the following additional identifiability assumption:

A5: if $p > 0$, $\mathcal{A}_{\boldsymbol{\vartheta}_0}(z)$ and $\mathcal{B}_{\boldsymbol{\vartheta}_0}(z)$ have no common root,$\mathcal{A}_{\boldsymbol{\vartheta}_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

We can now state the following extension of Theorem 1.1 and Corollary 2. The QMLE of $\boldsymbol{\theta}_0$ is denoted by $\hat{\boldsymbol{\theta}}_n^* := (\hat{\omega}_n^*, \hat{\alpha}_{1n}^*, \ldots, \alpha_{qn}^*, \hat{\beta}_{1n}^*, \ldots, \hat{\beta}_{pn}^*)'$.

Theorem 2.1. Under assumptions **A1** with (15) and (16), **A2** with (2) and (3), and **A5**, the VTE of the GARCH(p, q) model (14) is strongly consistent. Under the additional assumptions $E\epsilon_t^4 < \infty$ and **A4**, we have

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, (E\eta_0^4 - 1)\boldsymbol{\Sigma}\}, \tag{17}$$

where the matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} c & -c\boldsymbol{K}'\boldsymbol{J}^{-1} \\ -c\boldsymbol{J}^{-1}\boldsymbol{K} & \boldsymbol{J}^{-1} + c\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} \end{pmatrix}$$

is nonsingular with

$$c = \left(\frac{1 - \sum_{i=1}^{q} \beta_{0i}}{1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{j=1}^{p} \beta_{0j}}\right)^2 E(h_t^2)$$

and

$$\boldsymbol{J} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}'}\right)_{(p+q)\times(p+q)},$$

$$\boldsymbol{K} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \gamma}\right)_{(p+q)\times 1}.$$

Under these assumptions, the QMLE $\hat{\boldsymbol{\vartheta}}_n^*$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n^* - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, (E\eta_0^4 - 1)\boldsymbol{\Sigma}^*\}, \tag{18}$$

where

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - (c - d)\boldsymbol{C}\boldsymbol{C}', \tag{19}$$

with

$$\boldsymbol{C} = \begin{pmatrix} 1 \\ -\boldsymbol{J}^{-1}\boldsymbol{K} \end{pmatrix}, \quad d = \left\{ \left(\frac{1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{j=1}^{p} \beta_{0j}}{1 - \sum_{i=1}^{q} \beta_{0i}} \right)^2 E\left(\frac{1}{h_t^2}\right) - \boldsymbol{K}'\boldsymbol{J}^{-1}\boldsymbol{K} \right\}^{-1}.$$

The following result shows that the VTE can never be asymptotically more efficient than the QMLE, regardless of the values of the GARCH parameters and the distribution of η_t.

Proposition 3. Under the assumptions of Theorem 2.1, the asymptotic variance $(E\eta_0^4 - 1)\boldsymbol{\Sigma}$ of the VTE and the asymptotic variance $(E\eta_0^4 - 1)\boldsymbol{\Sigma}^*$ of the QMLE satisfy

$$\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^* \text{ is positive semidefinite but not necessarily positive definite.}$$

The following result characterizes the parameters that are estimated with the same asymptotic accuracy by the VTE and by the QMLE.

Corollary 4. Let the assumptions of Theorem 2.1 be satisfied, and let $\boldsymbol{\phi}$ be a mapping from $\mathbb{R}^{p+q+1}$ to $\mathbb{R}$, which is continuously differentiable in a neighbourhood of $\boldsymbol{\vartheta}_0$. If

$$\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0) \begin{pmatrix} 1 \\ -\boldsymbol{J}^{-1}\boldsymbol{K} \end{pmatrix} = 0,$$

then the asymptotic distribution of the VTE of the parameter $\boldsymbol{\phi}(\boldsymbol{\vartheta}_0)$ is the same as that of the QMLE, in the sense that

$$\sqrt{n}\{\boldsymbol{\phi}(\hat{\boldsymbol{\vartheta}}_n) - \boldsymbol{\phi}(\boldsymbol{\vartheta}_0)\} \xrightarrow{d} \mathcal{N}(0, s^2), \quad \sqrt{n}\{\boldsymbol{\phi}(\hat{\boldsymbol{\vartheta}}_n^*) - \boldsymbol{\phi}(\boldsymbol{\vartheta}_0)\} \xrightarrow{d} \mathcal{N}(0, s^2),$$

where

$$s^2 = (E\eta_0^4 - 1)\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}'}\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0).$$

For instance, suppose that $\boldsymbol{\phi}$ is a linear function, $\boldsymbol{\phi}(\boldsymbol{\vartheta}) = \sum_{i=1}^{p+q+1} a_i \boldsymbol{\vartheta}_i$ say. This result shows that the set of vectors $a = (a_i)_{i=1,\ldots,p+q+1}$ such that the VTE and QMLE of $a'\boldsymbol{\vartheta}_0$ have the same asymptotic accuracy is the hyperplan of $\mathbb{R}^{p+q+1}$ defined by $a'(1, -\boldsymbol{K}'\boldsymbol{J}^{-1})' = 0$.

3 COMPARISONS WITH THE QMLE

In this section, we compare the effective performance of the QMLE and VTE. In the first subsection, we numerically evaluate and compare the asymptotic variances of the two estimators. For simplicity, this comparison is made in ARCH(1) and GARCH(1,1) models. The second subsection presents simulation results with the aim to determine whether the ratio of the asymptotic variances gives a good idea of the ratio of accuracies of the two estimators in finite samples. Subsection 3.3 considers the situation where the GARCH model is misspecified. It will be shown that the fact that the VTE guarantees a consistent estimation of the long-run variance may be a crucial advantage of the VTE over the QMLE. Section 4 compares the QMLE and VTE for estimating GARCH(1,1) models and for computing VaR on set of daily and weekly financial returns.

3.1 Asymptotic Variances of the QMLE and VTE for ARCH(1) and GARCH(1,1) Models

For an ARCH(1) model, the asymptotic variances $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^*$ of the VTE and QMLE are given by Theorem 2.1 with

$$c = \frac{1}{(1-\alpha_0)^2} Eh_t^2, \quad \boldsymbol{J} = E\left\{\frac{(\epsilon_{t-1}^2 - \gamma_0)^2}{h_t^2}\right\}, \quad \boldsymbol{K} = (1-\alpha_0)E\left\{\frac{\epsilon_{t-1}^2 - \gamma_0}{h_t^2}\right\}.$$

In particular, the asymptotic variance of the VTE $\hat{\alpha}_n$ of α_0 is given by

$$\lim_{n\to\infty} \mathrm{Var}\{\sqrt{n}(\hat{\alpha}_n - \alpha_0)\} = \frac{E\eta_0^4 - 1}{E\{(\epsilon_{t-1}^2 - \gamma_0)^2/\sigma_t^4\}} \left[1 + \frac{E\sigma_t^4 (E\{(\epsilon_{t-1}^2 - \gamma_0)/\sigma_t^4\})^2}{E\{(\epsilon_{t-1}^2 - \gamma_0)^2/\sigma_t^4\}}\right].$$

For the QMLE $\hat{\alpha}_n^*$ of α_0, the asymptotic variance is

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{Var}\{\sqrt{n}(\hat{\alpha}_n^* - \alpha_0)\} \\
&= \frac{E\eta_0^4 - 1}{E\{(\epsilon_{t-1}^2 - \gamma_0)^2/\sigma_t^4\}} \\
&\quad \left[1 + \frac{(E\{(\epsilon_{t-1}^2 - \gamma_0)/\sigma_t^4\})^2}{E(1/\sigma_t^4)E\{(\epsilon_{t-1}^2 - \gamma_0)^2/\sigma_t^4\} - (E\{(\epsilon_{t-1}^2 - \gamma_0)/\sigma_t^4\})^2}\right] \\
&= \frac{(E\eta_0^4 - 1)E(1/\sigma_t^4)}{E(1/\sigma_t^4)E(\epsilon_{t-1}^4/\sigma_t^4) - \{E(\epsilon_{t-1}^2/\sigma_t^4)\}^2},
\end{aligned}$$

where the first equality is obtained with the parameterization $\boldsymbol{\vartheta}_0 = (\gamma_0, \alpha_0)$, and the second equality with the parameterization $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0)$.

The results presented in Table 1 are obtained from simulations of the matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^*$ above, with expectations replaced by empirical means and the standard

Table 1 Asymptotic variances of the QMLE and VTE of $\boldsymbol{\vartheta}_0$ for an ARCH(1) with $\gamma_0 = 1$ and $\eta_t \sim \mathcal{N}(0,1)$.

	$\alpha_0 = 0.1$	$\alpha_0 = 0.3$	$\alpha_0 = 0.5$	$\alpha_0 = 0.55$	$\alpha_0 = 0.7$
QMLE	$\begin{pmatrix} 2.52 & 0.51 \\ 0.51 & 1.69 \end{pmatrix}$	$\begin{pmatrix} 4.80 & 2.24 \\ 2.24 & 2.84 \end{pmatrix}$	$\begin{pmatrix} 12.01 & 5.71 \\ 5.71 & 3.93 \end{pmatrix}$	$\begin{pmatrix} 15.94 & 7.11 \\ 7.11 & 4.20 \end{pmatrix}$	$\begin{pmatrix} 45.27 & 14.32 \\ 14.32 & 5.02 \end{pmatrix}$
VTE	$\begin{pmatrix} 2.52 & 0.51 \\ 0.51 & 1.69 \end{pmatrix}$	$\begin{pmatrix} 5.06 & 2.36 \\ 2.36 & 2.90 \end{pmatrix}$	$\begin{pmatrix} 18.13 & 8.61 \\ 8.61 & 5.30 \end{pmatrix}$	$\begin{pmatrix} 28.78 & 12.82 \\ 12.82 & 6.74 \end{pmatrix}$	∞

∞ means that the asymptotic variance does not exist.

normal distribution for η_t. More precisely, the table displays the mean of 1000 independent estimates of the matrices

$$2\boldsymbol{\Sigma} = \lim_{n\to\infty} \mathrm{Var}\{\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)\} \quad \text{and} \quad 2\boldsymbol{\Sigma}^* = \lim_{n\to\infty} \mathrm{Var}\{\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n^* - \boldsymbol{\vartheta}_0^*)\},$$

where each estimation is obtained from empirical means based on a simulation of size $n = 10,000$ of the ARCH(1) model. It is seen that the variance targeting does not affect the asymptotic distribution of the estimator of $\boldsymbol{\vartheta}_0$ when α_0 is small but entails a dramatic loss of efficiency when α_0 approaches the limit implied by the existence of a fourth moment ($\alpha_0 < 0.57$ when η_t has a standard normal distribution).

Table 2 is the analog of Table 1 but gives the asymptotic variances of the QMLE and VTE for the standard ARCH parameter $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0)$. From this table, it is seen that the asymptotic distribution of the VTE of the parameter ω_0 should be close to that of the QMLE. This is not surprising because we know from Corollary 4 that there exist transformations of $\boldsymbol{\vartheta}_0$ that are estimated by VTE and QMLE with the same asymptotic accuracy.

Now, we turn to a GARCH(1,1) model given by

$$h_t = 0.0003 + 0.09\epsilon_{t-1}^2 + 0.89h_{t-1}.$$

The volatility coefficients have been chosen close to those usually estimated on real data (see Section 4). When $\eta_t \sim \mathcal{N}(0,1)$, the asymptotic variances of the VTE and QMLE of $\boldsymbol{\vartheta}_0$ are approximated, via an average over 1000 independent simulations of length 1000, by

$$\begin{pmatrix} 0.022 & 0.043 & -0.003 \\ 0.043 & 0.403 & -0.409 \\ -0.003 & -0.409 & 0.565 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0.018 & 0.034 & -0.001 \\ 0.034 & 0.384 & -0.405 \\ -0.001 & -0.405 & 0.564 \end{pmatrix},$$

Table 2 Asymptotic variances of the QMLE and VTE of $\boldsymbol{\theta}_0$ for an ARCH(1) with $\omega_0 = 1$ and $\eta_t \sim \mathcal{N}(0,1)$.

	$\alpha_0 = 0.1$	$\alpha_0 = 0.3$	$\alpha_0 = 0.5$	$\alpha_0 = 0.55$	$\alpha_0 = 0.7$
QMLE	$\begin{pmatrix} 3.5 & -1.4 \\ -1.4 & 1.7 \end{pmatrix}$	$\begin{pmatrix} 4.2 & -1.8 \\ -1.8 & 2.8 \end{pmatrix}$	$\begin{pmatrix} 4.9 & -2.2 \\ -2.2 & 3.9 \end{pmatrix}$	$\begin{pmatrix} 5.1 & -2.2 \\ -2.2 & 4.2 \end{pmatrix}$	$\begin{pmatrix} 5.6 & -2.4 \\ -2.4 & 5.1 \end{pmatrix}$
VTE	$\begin{pmatrix} 3.5 & -1.4 \\ -1.4 & 1.7 \end{pmatrix}$	$\begin{pmatrix} 4.2 & -1.8 \\ -1.8 & 2.9 \end{pmatrix}$	$\begin{pmatrix} 4.9 & -1.9 \\ -1.9 & 6.1 \end{pmatrix}$	$\begin{pmatrix} 5.1 & -2.1 \\ -2.1 & 9.3 \end{pmatrix}$	∞

respectively. Now when the errors follow a Student distribution, namely $\eta_t \sim t(9)$, the asymptotic variances of the VTE and QMLE of $\boldsymbol{\vartheta}_0$ are approximated by

$$\begin{pmatrix} 0.052 & 0.101 & -0.012 \\ 0.101 & 0.621 & -0.560 \\ -0.012 & -0.560 & 0.743 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0.031 & 0.059 & -0.004 \\ 0.059 & 0.536 & -0.543 \\ -0.004 & -0.543 & 0.740 \end{pmatrix}.$$

For this GARCH(1,1) model, we can conclude that the variance targeting does not entail an important loss of efficiency with respect to the QMLE.

3.2 Sampling Distribution of the QMLE and VTE

To compare the performance of the QMLE and VTE in finite samples, we computed the two estimators on 2000 independent simulated trajectories of length $n = 500$ of three ARCH(1) models, with errors distributed as Gaussian and Student $t(\nu)$ distributions (with $\nu = 3$ and $\nu = 9$ degrees of freedom). The three ARCH(1) models have already been considered in Table 2. From this table, we know the asymptotic variances of the QMLE and VTE. For the three models, the first parameter is fixed to $\omega = 1$ and the second varies from $\alpha = 0.3$, $\alpha = 0.55$ to $\alpha = 0.9$. Note that for the last value of α, and also for the $t(3)$ distribution, Assumption **A3** is not satisfied, so the asymptotic normality of the VTE is not guaranteed. Table 3 provides an overview of these simulations experiments. The most noticeable output is that the VTE performs remarkably well and even outperforms the (Q)MLE. This finite-sample result counterbalances the result of Proposition 3 showing that the VTE cannot be asymptotically more efficient than the QMLE. For the $t(9)$ distribution, it turns out that the VTE performs much better than the QMLE. More precisely, values of α bigger than one are avoided with this method.

To evaluate the impact of the sample size, Table 4 reports results for the two estimators on 1000 independent simulated trajectories of length $n = 5000$ and $n = 10,000$ of three ARCH(1) models with Gaussian errors. As expected from Table 2, the QMLE and VTE of ω have very similar accuracy, and the QMLE of α is slightly more accurate than the VTE and $\alpha = 0.55$ or $\alpha = 0.9$.

Table 3 Sampling distribution of the QMLE and VTE of $\boldsymbol{\theta}_0$ for ARCH(1) models with $\eta_t \sim t(\nu)$ and $n = 500$.

Parameter	True value	Estimator	Bias	RMSE	Minimum	Q_1	Q_2	Q_3	Maximum
				$\nu = \infty$					
ω	1.0	QMLE	0.005	0.083	0.738	0.948	1.000	1.060	1.283
		VTE	0.005	0.083	0.736	0.949	1.001	1.059	1.282
α	0.3	QMLE	−0.004	0.057	0.000	0.054	0.094	0.133	0.305
		VTE	−0.005	0.056	0.000	0.054	0.094	0.133	0.301
ω	1.0	QMLE	0.006	0.100	0.666	0.935	1.007	1.072	1.394
		VTE	0.005	0.099	0.678	0.934	1.006	1.071	1.397
α	0.55	QMLE	−0.010	0.092	0.257	0.480	0.538	0.600	0.854
		VTE	−0.024	0.091	0.246	0.467	0.523	0.582	0.942
ω	1.0	QMLE	0.011	0.114	0.655	0.936	1.002	1.082	1.537
		VTE	0.036	0.112	0.700	0.958	1.030	1.105	1.542
α	0.9	QMLE	−0.011	0.108	0.512	0.810	0.891	0.963	1.272
		VTE	−0.108	0.089	0.468	0.730	0.795	0.854	0.997
				$\nu = 9$					
ω	1.0	QMLE	0.006	0.101	0.708	0.940	1.002	1.071	1.453
		VTE	0.006	0.101	0.703	0.941	1.001	1.071	1.451
α	0.3	QMLE	−0.004	0.066	0.000	0.047	0.090	0.134	0.564
		VTE	−0.005	0.064	0.000	0.047	0.090	0.133	0.421
ω	1.0	QMLE	0.014	0.122	0.669	0.929	1.012	1.090	1.703
		VTE	0.015	0.120	0.706	0.929	1.011	1.088	1.698
α	0.55	QMLE	−0.014	0.119	0.205	0.457	0.530	0.611	1.149
		VTE	−0.040	0.106	0.211	0.442	0.506	0.575	0.922
ω	1.0	QMLE	0.006	0.137	0.618	0.913	0.992	1.092	1.611
		VTE	0.036	0.132	0.688	0.945	1.023	1.119	1.625
α	0.9	QMLE	−0.010	0.144	0.462	0.792	0.887	0.986	1.479
		VTE	−0.133	0.099	0.429	0.701	0.767	0.836	0.995
				$\nu = 3$					
ω	1.0	QMLE	−0.044	0.355	0.407	0.771	0.881	1.037	5.865
		VTE	−0.038	0.363	0.403	0.773	0.884	1.044	5.865
α	0.3	QMLE	0.022	0.248	0.000	0.008	0.053	0.134	5.824
		VTE	−0.006	0.118	0.000	0.008	0.054	0.130	0.906
ω	1.0	QMLE	−0.001	0.549	0.423	0.773	0.902	1.078	11.827
		VTE	0.032	0.576	0.467	0.787	0.921	1.107	11.754
α	0.55	QMLE	−0.007	0.569	0.000	0.278	0.414	0.621	6.949
		VTE	−0.154	0.180	0.000	0.272	0.374	0.506	0.989
ω	1.0	QMLE	−0.022	0.465	0.361	0.751	0.881	1.054	8.820
		VTE	0.036	0.582	0.468	0.782	0.913	1.103	11.428
α	0.9	QMLE	−0.062	0.605	0.043	0.540	0.732	0.983	10.951
		VTE	−0.304	0.175	0.043	0.478	0.599	0.721	0.997

RMSE is the root mean square error, Q_i, $i = 1, 3$, denote the quartiles.

3.3 Variance Targeting Estimator in Misspecified Models

The variance targeting technique ensures robust estimation of the marginal variance, provided that it exists. Indeed, the variance of a model estimated by VTE converges to the theoretical variance, even if the model is misspecified. For the

Table 4 Sampling distribution of the QMLE and VTE of $\boldsymbol{\theta}_0$ for ARCH(1) models with $\eta_t \sim \mathcal{N}(0,1)$.

Parameter	True value	Estimator	Bias	RMSE	Min	Q_1	Q_2	Q_3	Max
				$n = 5000$					
ω	1.0	QMLE	0.002	0.029	0.891	0.983	1.001	1.021	1.089
		VTE	0.002	0.029	0.892	0.983	1.001	1.022	1.089
α	0.3	QMLE	0.000	0.024	0.219	0.284	0.301	0.316	0.389
		VTE	0.000	0.024	0.218	0.285	0.301	0.317	0.413
ω	1.0	QMLE	0.002	0.032	0.910	0.981	1.003	1.022	1.104
		VTE	0.002	0.032	0.880	0.980	1.002	1.021	1.102
α	0.55	QMLE	−0.002	0.028	0.455	0.529	0.548	0.567	0.631
		VTE	−0.003	0.036	0.451	0.524	0.544	0.567	0.896
ω	1.0	QMLE	0.000	0.035	0.892	0.976	1.000	1.023	1.126
		VTE	0.015	0.036	0.904	0.991	1.014	1.040	1.134
α	0.9	QMLE	0.001	0.035	0.797	0.877	0.902	0.924	1.027
		VTE	−0.053	0.047	0.713	0.814	0.843	0.875	0.999
				$n = 10,000$					
ω	1.0	QMLE	0.001	0.009	0.972	0.994	1.000	1.007	1.032
		VTE	0.001	0.009	0.972	0.994	1.000	1.007	1.031
α	0.3	QMLE	−0.001	0.008	0.272	0.294	0.299	0.304	0.324
		VTE	−0.001	0.008	0.272	0.294	0.299	0.304	0.326
ω	1.0	QMLE	0.000	0.010	0.965	0.993	1.000	1.007	1.041
		VTE	0.000	0.010	0.966	0.993	1.000	1.007	1.040
α	0.55	QMLE	0.000	0.009	0.525	0.544	0.550	0.556	0.578
		VTE	0.000	0.013	0.521	0.542	0.549	0.557	0.695
ω	1.0	QMLE	0.000	0.012	0.966	0.992	1.000	1.008	1.043
		VTE	0.010	0.015	0.955	1.001	1.010	1.020	1.051
α	0.9	QMLE	0.000	0.011	0.863	0.892	0.900	0.907	0.943
		VTE	−0.032	0.032	0.815	0.847	0.863	0.883	0.998

convergence to hold true, it suffices that the observed process be stationary and ergodic with a finite second-order moment. This is generally not the case when the misspecified model is estimated by QMLE.

In the next sections, we consider two applications where this robustness feature of the VTE is particularly attractive.

3.3.1 Prediction over long horizons with models estimated by VTE.

We will study the asymptotic behaviour of the GARCH(1,1) predictions when the forecast horizon is large and when the data-generating process (DGP) may be different from the GARCH(1,1) model in (1). The results of this section can be extended to general GARCH(p,q) models, but the presentation will be simpler with

GARCH(1,1) models. With the (possibly misspecified) GARCH(1,1) model, h-step ahead prediction intervals for ϵ_{n+h} are given by

$$\left[\sqrt{\hat{\sigma}^2_{n+h|n}}\hat{F}_\eta^{-1}(\underline{\alpha}/2), \sqrt{\hat{\sigma}^2_{n+h|n}}\hat{F}_\eta^{-1}(1-\underline{\alpha}/2)\right],$$

where $1-\underline{\alpha}$ is the nominal asymptotic probability of the interval, $\hat{F}_\eta^{-1}(\underline{\alpha})$ denotes an estimate of the $\underline{\alpha}$-quantile of the distribution F_η of η_1, and $\hat{\sigma}^2_{n+h|n}$ is the estimate of the h-step ahead forecast error variance given by

$$\hat{\sigma}^2_{n+h|n} = \hat{\gamma}^*_n + \{\sigma^2_{n+1}(\hat{\boldsymbol{\vartheta}}^*_n) - \hat{\gamma}^*_n\}(1-\hat{\kappa}^*_n)^{h-1}$$

when the GARCH model is estimated by QMLE, and by

$$\hat{\sigma}^2_{n+h|n} = \hat{\sigma}^2_n + \{\sigma^2_{n+1}(\hat{\boldsymbol{\vartheta}}_n) - \hat{\sigma}^2_n\}(1-\hat{\kappa}_n)^{h-1}$$

when the GARCH model is estimated by VTE. When the GARCH(1,1) model is misspecified, the true parameter value $\boldsymbol{\vartheta}_0$ does not exist, but one can expect that the QMLE and VTE converge to some so-called "pseudo"true values. More precisely, under stationarity, ergodicity, and other general conditions, see White (1982), $\hat{\boldsymbol{\vartheta}}^*_n \to \tilde{\boldsymbol{\vartheta}}^* = (\tilde{\gamma}^*, \tilde{\boldsymbol{\lambda}}^*)$ almost surely as $n \to \infty$, where the pseudotrue value $\tilde{\boldsymbol{\vartheta}}^*$ is defined by

$$\tilde{\boldsymbol{\vartheta}}^* = \arg\min_{\boldsymbol{\vartheta}} E\ell_1(\boldsymbol{\vartheta}), \quad \ell_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\vartheta})} + \log\sigma_t^2(\boldsymbol{\vartheta}).$$

Similarly, one should generally have

$$\hat{\boldsymbol{\vartheta}}_n \to \tilde{\boldsymbol{\vartheta}} = (\tilde{\gamma}, \tilde{\boldsymbol{\lambda}}) \quad a.s. \quad \text{with } \tilde{\gamma} = E\epsilon_1^2 \text{ and } \tilde{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} E\ell_1(\tilde{\gamma}, \boldsymbol{\lambda}).$$

Assume that these pseudotrue values $\tilde{\boldsymbol{\lambda}} = (\tilde{\alpha}, \tilde{\kappa})$ and $\tilde{\boldsymbol{\lambda}}^* = (\tilde{\alpha}^*, \tilde{\kappa}^*)$ are such that $\tilde{\kappa} < 1$ and $\tilde{\kappa}^* < 1$. When the horizon h is large, the asymptotic prediction interval is thus equivalent to

$$\left[\sqrt{\tilde{\gamma}^*}F_\eta^{-1}(\underline{\alpha}/2), \sqrt{\tilde{\gamma}^*}F_\eta^{-1}(1-\underline{\alpha}/2)\right], \tag{20}$$

with the QMLE, and equivalent to

$$\left[\sqrt{E\epsilon_1^2}F_\eta^{-1}(\underline{\alpha}/2), \sqrt{E\epsilon_1^2}F_\eta^{-1}(1-\underline{\alpha}/2)\right], \tag{21}$$

with the VTE. Note that the long horizon prediction intervals (20) obtained with QMLE are not correct when $\tilde{\gamma}^* \neq E\epsilon_1^2$, which is generally the case for misspecified models. On the contrary, even when the model is misspecified, the probability that ϵ_{n+h} belongs to the VTE asymptotic prediction interval (21) tends to the nominal probability $1-\underline{\alpha}$ as the horizon h increases.

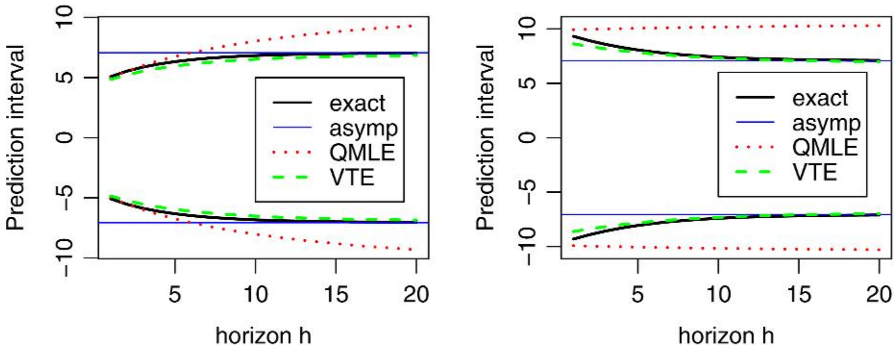

Figure 1 Asymptotic prediction intervals based on the true model (between the full lines), for a GARCH(1,1) estimated by QMLE (dotted lines) and a GARCH(1,1) estimated by VTE (dashed lines). The horizontal full lines are the bounds of the large-horizon prediction intervals (20). The DGP is the Markov-switching model (22). The figure on the left corresponds to predictions when the present volatility σ_n is low, and the figure on the right corresponds to predictions in the case when σ_n is large.

Example 1. To give an elementary illustration, consider the Markov-switching model

$$\epsilon_t = \omega(\Delta_t)\eta_t, \tag{22}$$

where η_t is an i.i.d. noise, (Δ_t) is a stationary irreducible and aperiodic Markov chain, independent of (η_t), with state-space $\{1, \ldots, d\}$. For Figure 1, we choose $\eta_t \sim \mathcal{N}(0,1)$, $d = 2$ regimes with $\omega(1) = 1$ and $\omega(2) = 5$, and the transition probabilities $P(\Delta_t = 1|\Delta_{t-1} = 1) = P(\Delta_t = 2|\Delta_{t-1} = 2) = 0.9$. It can be noted that (ϵ_t) is a white noise and that (ϵ_t^2) is an autocorrelated process. Therefore, it is not unrealistic to assume that an empirical researcher would fit a misspecified GARCH model to data generated by Model (22). In our experiments, we fitted a GARCH(1,1) model by the two methods, on a simulation of size 1000 of Model (22). Figure 1 shows the h-step ahead prediction intervals obtained from the true model and from the estimated GARCH models. It can be seen that, in particular, when the prediction horizon h is large, the prediction intervals based on the false GARCH(1,1) model estimated by VTE are close to those obtained with the right model. When the GARCH parameters are estimated by QMLE, the prediction intervals are clearly oversized for large horizons h (indicating a pseudotrue value $\tilde{\gamma}^*$ larger than $E\epsilon_1^2$, for this particular model).

3.3.2 Estimating long horizon VaR.

VaR is one of the most important market-risk measurement tool (see, e.g., the web site http://www.gloriamundi.org/ that is entirely devoted to VaR). For a portfolio whose value at time t is a random variable V_t, the profits and losses function at the horizon h is $L_{t,t+h} = -(V_{t+h} - V_t)$. At the confidence level $\underline{\alpha} \in (0,1)$, the horizon h and the date t, the (conditional) VaR is the $(1-\underline{\alpha})$-quantile of the conditional distribution of $L_{t,t+h}$ given the information available at time t:

$$\mathrm{VaR}_{t,h}(\underline{\alpha}) = \inf\{x \in \mathbb{R} | P(L_{t,t+h} \leqslant x | V_u, u \leqslant t) \geqslant 1 - \underline{\alpha}\}.$$

Introducing the log-returns $\epsilon_t = \log(V_t / V_{t-1})$, we have

$$\mathrm{VaR}_{t,h}(\underline{\alpha}) = [1 - \exp\{q_{t,h}(\underline{\alpha})\}]V_t, \tag{23}$$

where $q_{t,h}(\underline{\alpha})$ is the $\underline{\alpha}$-quantile of the conditional distribution of the future returns $\epsilon_{t+1} + \cdots + \epsilon_{t+h}$. The following lemma shows how to approximate $\mathrm{VaR}_{t,h}(\underline{\alpha})$ for large h, under some α-mixing condition on the process (ϵ_t).

Lemma 3.1. Assume that (ϵ_t) is a strictly stationary process such that $E\epsilon_t = 0$, $E|\epsilon_t|^{2+\nu} < \infty$ and $\sum_{h=1}^{\infty}\{\alpha_\epsilon(h)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$. Let $\mathrm{Var}(\epsilon_t) = \gamma$. We have

$$\lim_{h\to\infty} \sqrt{h}\,\sqrt{\gamma}\,\Phi^{-1}(\underline{\alpha}) / q_{t,h}(\underline{\alpha}) = 1 \quad \text{a.s.}$$

Remark 3.1. The mixing condition of the lemma is satisfied for a variety of processes, in particular GARCH-type processes (see, for instance, Carrasco and Chen 2002; Francq and Zakoïan 2006). This condition is also satisfied for the Markov-switching process (22). Indeed, the Markov chain (Δ_t) enjoys a number of mixing properties (see, e.g., Theorem 3.1 in Bradley 2005). In particular, there exist $K > 0$ and $\rho \in (0,1)$ such that $\alpha_\Delta(k) \leqslant K\rho^k$ for all $k \in \mathbb{N}$. Because (η_t) and (Δ_t) are independent, and ϵ_t is a measurable function of Δ_t and η_t, Theorem 5.2 in Bradley (2005) entails that $\alpha_\epsilon(k) \leqslant K\rho^k$. For any conditionally heteroscedatic process of the form $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$, where $\eta_t \sim F_\eta$, the VaR at Horizon 1 is given by

$$\mathrm{VaR}_{t,1}(\underline{\alpha}) = [1 - \exp\{\sigma_{t+1}(\boldsymbol{\theta}_0)F_\eta^{-1}(\underline{\alpha})\}]V_t,$$

in view of (23). Hence, if $\hat{\boldsymbol{\theta}}_n$ is an estimator of $\boldsymbol{\theta}_0$, an obvious estimator of $\mathrm{VaR}_{t,1}(\underline{\alpha})$ is obtained by the plug-in method. In general, exact VaR's at horizon $h > 1$ cannot be computed explicitly. It is therefore of interest to use the previous lemma to approximate the VaR at a long horizon h. Given an estimator $\hat{\gamma}$ of γ, one can use

$$\widehat{\mathrm{VaR}}_{t,h}(\underline{\alpha}) = [1 - \exp\{\sqrt{h}\,\Phi^{-1}(\underline{\alpha})\ \sqrt{\hat{\gamma}}\}]V_t. \tag{24}$$

If (ϵ_t) follows a GARCH model, both the VTE and the QMLE methods provide consistent estimators of γ. When the GARCH model is misspecified, only the VTE guarantees consistency of $\hat{\gamma}$, and thus asymptotically valid estimates for long horizon VaR's. This is illustrated in the next example.

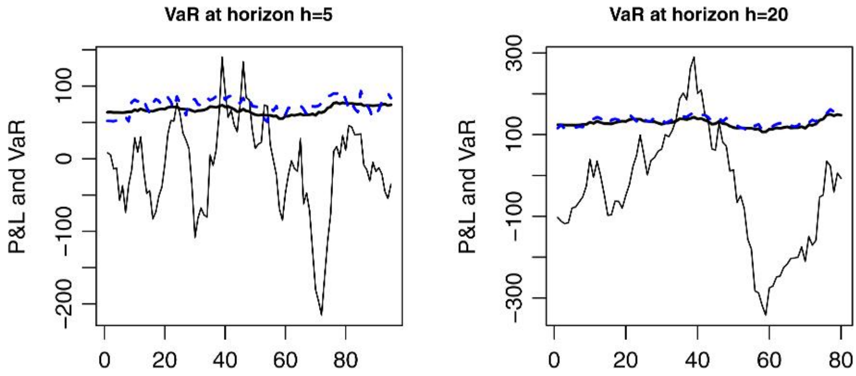

Figure 2 Sample paths of the profit and loss (P&L) process generated by the Markov-switching model (22) and VaR at the confidence level 5%. The full line corresponds to the asymptotic approximation obtained from Lemma 3.1, the dashed line is obtained by simulations.

Example 2 (Example 1 continued). We shall consider VaR at horizons $h = 1$ and $h = 20$ obtained from estimated GARCH(1,1) models, when the observations are drawn from the Markov-switching process (22). For the sake of comparison, we shall also consider the theoretical VaR's of the true model. At Horizon 1, the theoretical VaR is given by (23) with $h = 1$ and $q_{t,1}(\underline{\alpha})$ obtained as the solution of

$$\underline{\alpha} = \sum_{j=1}^{d} F_\eta \left\{ \frac{q_{t,1}(\underline{\alpha})}{\sigma(j)} \right\} P(\Delta_{t+1} = j | \epsilon_u, u \leqslant t).$$

For large h, the theoretical VaR can be approximated by

$$\mathrm{VaR}_{t,h}(\underline{\alpha}) = [1 - \exp\{\sqrt{h}\Phi^{-1}(\underline{\alpha})\sqrt{\gamma}\}]V_t, \quad \sqrt{\gamma} = \sum_{j=1}^{d} \omega(j) P(\Delta_t = j).$$

To illustrate Lemma 3.1, we computed the VaRs by simulations and compared it to the approximation above. It can be seen in Figure 2 that, whereas a noticeable difference appears for $h = 5$, this difference vanishes at $h = 20$. Simulations not reported here suggest that a discrepancy still exists at $h = 10$. For this model, the approximation can thus be considered as reliable for $h \geqslant 20$.

Figure 3 shows samples paths of $L_{t,t+h}$, for $h = 1$ and $h = 20$, obtained with $\eta_t \sim \mathcal{N}(0, 1)$, $d = 2$ regimes, $\omega(1) = 1/200$ and $\omega(2) = 5/200$. The full line indicates the VaR at the 5% level, computed with the true model, using the asymptotic approximation for $h = 20$. The VaR estimated by VTE from a misspecified GARCH(1,1) model, displayed in dashed line, appears to be very close to the correct VaR, especially when h is large. This is not the case for the VaR estimated by QMLE (dotted lines), which strongly overestimates for $h = 20$. Standard

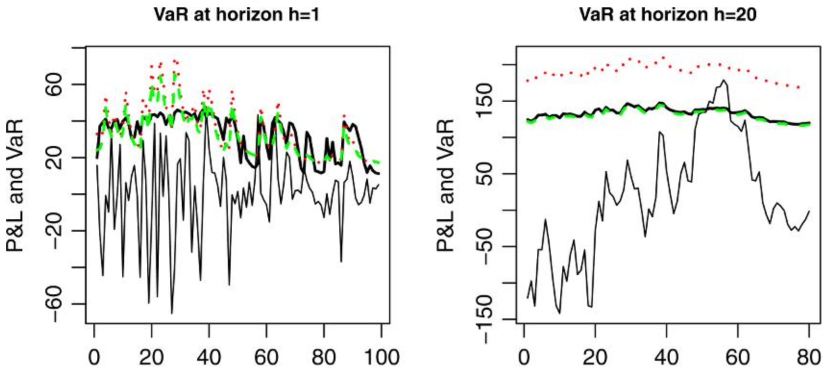

Figure 3 Sample paths of the P&L process generated by the Markov-switching model (22) and VaR at the confidence level 5%. The full line corresponds to the exact VaR on the left graph and to its asymptotic approximation provided by Lemma 3.1 on the right graph. The dotted (respectively dashed) line corresponds to the asymptotic approximation obtained from Lemma 3.1 applied to a GARCH model estimated by QMLE (respectively VTE).

Table 5 Backtesting comparison of the VaR estimations given by the true Hidden Markov Model model (22), the GARCH(1,1) model estimated by QMLE, and the GARCH(1,1) estimated by VTE on $n = 1000$ observations. The comparison is made out-of-sample, on a simulation of size $N = 50,000$ of the P&L function, for the two horizons $h = 1$ and $h = 20$ and the three levels $\underline{\alpha} = 1\%$, $\underline{\alpha} = 5\%$, and $\underline{\alpha} = 10\%$.

	$h = 1$			$h = 20$		
	HMM	QMLE	VTE	HMM	QMLE	VTE
$\underline{\alpha} = 1\%$						
Relative VaR average (in %)	4.64	4.01	3.78	17.10	24.18	16.77
Violations (in %)	1.01	2.28	2.57	1.15	0.09	1.24
$\underline{\alpha} = 5\%$						
Relative VaR average (in %)	2.65	2.86	2.69	12.42	17.77	12.17
Violations (in %)	4.82	4.84	5.49	4.81	0.89	5.16
$\underline{\alpha} = 10\%$						
Relative VaR average (in %)	1.87	2.23	2.10	9.82	14.14	9.62
Violations (in %)	9.99	7.32	8.17	9.11	2.86	9.47

evaluation of the performance of VaR estimation methods relies on comparing the percentages of violations (losses larger than the estimated VaR) with the nominal level $\underline{\alpha}$, on out-of-sample observations. Such a procedure is often referred to as "backtesting." Table 5 displays the average VaR (in percentages of the portfolio value V_t, that is, $100/N\sum_{t=1}^{N} \mathrm{VaR}_{t,h}(\underline{\alpha})/V_t$) together with the number of violations, over a very long period of time. This table confirms the conclusions drawn from Figure 3: the VaR at the horizon $h = 20$ computed with the misspecified GARCH (1,1) is more satisfactory in terms of backtesting when the model is estimated by VTE than by QMLE.

4 COMPARISON OF THE QMLE AND VTE ON STOCK MARKET RETURNS

In this section, we compare the VTE and QMLE on stock returns of eleven indices, namely the CAC, DAX, DJA, DJI, DJT, DJU, FTSE, Nasdaq,[1] Nikkei, SMI, and SP500. The samples extend from January 2, 1990, to January 22, 2009, except for the indices for which such historical data do not exist.

4.1 Daily and Weekly Estimations

For each daily series, a GARCH(1,1) model was estimated, by QMLE and by VTE. Table 6 displays the models estimated by the two procedures. For these series of daily returns, it seems that the moment assumption $E\epsilon_t^4 < \infty$ is questionable because $(\hat{\alpha} + \hat{\beta})^2 + (\widehat{E\eta_0^4} - 1)\hat{\alpha}^2$ is often close to or larger than 1, and it is known that $E\epsilon_t^4 < \infty$ if and only if $(\alpha_0 + \beta_0)^2 + (E\eta_0^4 - 1)\alpha_0^2 < 1$. Therefore, the assumptions given in Theorem 2.1 to obtain the asymptotic normality are likely to be unsatisfied. Nevertheless, it is seen from Table 6 that the parameters estimated by VTE are always very close to those estimated by QMLE. It is also interesting to note how close the asymptotic standard errors are in the two cases (except for the parameter β of the DJA). Table 7 reports results for the same series but for weekly date. The differences between the VTE and QMLE (parameters and standard errors) are more evident than for the daily frequency. This can be seen, in particular, in the annualized volatilities that highlight how much long-run forecasts are going to be different.

4.2 Computation Time

As expected, the VTE is more successful than the QMLE in terms of amount of computation time. Table 8 compares the computation time of the QMLE and VTE for estimating the models of the eleven indices. Two designs, corresponding to two different initial values, are considered. Design 1 corresponds to the initial values $\alpha = 0.05$, $\beta = 0.85$, and ω equal to $(1 - \alpha - \beta)$ times the empirical variance of

[1]One outlier has been eliminated since the Nasdaq index level was halved on January 3, 1994.

Table 6 Comparison of the QMLE and VTE of GARCH(1,1) models for eleven daily stock market returns. The estimated standard deviation are displayed into brackets. The column $\hat{\sigma}_{\text{ann}}$ displays the annualized volatilities. The last column corresponds to plug-in estimates of $\rho_4 = (\alpha+\beta)^2 + (E\eta_0^4 - 1)\alpha^2$. We have $E\epsilon_0^4 < \infty$ if and only if $\rho_4 < 1$.

Index	Estimator	ω	α	β	$\hat{\sigma}_{\text{ann}}$	ρ_4
CAC	QMLE	0.033 (0.009)	0.090 (0.014)	0.893 (0.015)	22.2	1.0067
	VTE	0.033 (0.009)	0.090 (0.014)	0.893 (0.015)	22.4	
DAX	QMLE	0.037 (0.014)	0.093 (0.023)	0.888 (0.024)	22.2	1.0622
	VTE	0.036 (0.013)	0.095 (0.022)	0.888 (0.024)	23.1	
DJA	QMLE	0.019 (0.005)	0.088 (0.014)	0.894 (0.014)	16.5	0.9981
	VTE	0.019 (0.005)	0.089 (0.012)	0.894 (0.007)	16.8	
DJI	QMLE	0.017 (0.004)	0.085 (0.013)	0.901 (0.013)	17.3	1.002
	VTE	0.016 (0.004)	0.085 (0.012)	0.901 (0.013)	17.6	
DJT	QMLE	0.040 (0.013)	0.089 (0.016)	0.894 (0.018)	24.2	1.0183
	VTE	0.042 (0.013)	0.086 (0.016)	0.894 (0.018)	22.8	
DJU	QMLE	0.021 (0.005)	0.118 (0.016)	0.865 (0.014)	17.7	1.0152
	VTE	0.021 (0.004)	0.119 (0.013)	0.865 (0.013)	18.1	
FTSE	QMLE	0.013 (0.004)	0.091 (0.014)	0.899 (0.014)	18.1	1.0228
	VTE	0.013 (0.004)	0.090 (0.013)	0.899 (0.014)	18.0	
Nasdaq	QMLE	0.025 (0.006)	0.072 (0.009)	0.922 (0.009)	35.4	1.0021
	VTE	0.025 (0.006)	0.072 (0.009)	0.922 (0.009)	34.3	
Nikkei	QMLE	0.053 (0.012)	0.100 (0.013)	0.880 (0.014)	26.0	0.9985
	VTE	0.054 (0.012)	0.098 (0.013)	0.880 (0.015)	25.0	
SMI	QMLE	0.049 (0.014)	0.127 (0.028)	0.835 (0.029)	18.1	1.0672
	VTE	0.048 (0.014)	0.133 (0.025)	0.834 (0.029)	19.1	
SP500	QMLE	0.014 (0.004)	0.084 (0.012)	0.905 (0.012)	18.2	1.0072
	VTE	0.014 (0.003)	0.084 (0.011)	0.905 (0.012)	18.2	

the series. Design 2 corresponds to the initial values $\alpha = 0$, $\beta = 0$, and $\omega = 1$. The initial values of Design 1 are much closer to the final estimates than those of Design 2. Thus, it is not surprising to observe longer computation times in Design 2 than in Design 1. In both designs, the QMLE is around 1.6 times slower than the VTE, and the time required for the two estimates (QMLE+VTE) is not much bigger than that taken by the QMLE. More interestingly, an examination of the estimated models shows that, in Design 2 (i.e., when the initial values are far from the final estimates) and for two indices (namely, the DJI and SP500), the QMLE is trapped in a local estimate for which the likelihood is less than for the solution obtained in Design 1. For the VTE, and also for VTE+QMLE method, the solutions obtained in the two designs are the same. From these experiments, one can conclude that (i) when the initial values are reasonably well chosen (in Design 1), there is no

Table 7 As Table 6 but for weekly returns.

Index	Estimator	ω	α	β	$\hat{\sigma}_{\text{ann}}$	ρ_4
CAC	QMLE	0.223 (0.156)	0.123 (0.045)	0.862 (0.052)	27.8	1.0863
	VTE	0.268 (0.198)	0.112 (0.046)	0.857 (0.058)	21.2	
DAX	QMLE	0.461 (0.292)	0.194 (0.071)	0.777 (0.081)	28.4	1.2768
	VTE	0.523 (0.334)	0.176 (0.073)	0.771 (0.088)	22.7	
DJA	QMLE	0.384 (0.187)	0.178 (0.059)	0.754 (0.075)	17.2	1.0501
	VTE	0.433 (0.209)	0.170 (0.059)	0.742 (0.082)	16.0	
DJI	QMLE	0.340 (0.144)	0.208 (0.055)	0.742 (0.062)	18.8	1.1088
	VTE	0.387 (0.161)	0.196 (0.056)	0.731 (0.069)	16.6	
DJT	QMLE	0.177 (0.155)	0.074 (0.036)	0.916 (0.041)	30.9	1.0467
	VTE	0.237 (0.219)	0.065 (0.035)	0.911 (0.048)	22.6	
DJU	QMLE	0.201 (0.125)	0.134 (0.047)	0.839 (0.054)	19.7	1.0789
	VTE	0.227 (0.137)	0.128 (0.048)	0.832 (0.058)	17.2	
FTSE	QMLE	0.152 (0.099)	0.140 (0.048)	0.846 (0.053)	24.3	1.1292
	VTE	0.194 (0.131)	0.128 (0.049)	0.837 (0.061)	17.0	
Nasdaq	QMLE	2.641 (4.515)	0.109 (0.227)	0.774 (0.349)	34.3	2.0578
	VTE	2.6160 (4.957)	0.086 (0.196)	0.777 (0.390)	31.5	
Nikkei	QMLE	1.497 (0.618)	0.169 (0.059)	0.683 (0.098)	23.0	0.8534
	VTE	1.507 (0.631)	0.166 (0.058)	0.682 (0.100)	22.7	
SMI	QMLE	0.535 (0.162)	0.203 (0.049)	0.705 (0.057)	17.4	0.9442
	VTE	0.535 (0.162)	0.203 (0.046)	0.705 (0.058)	17.3	
SP500	QMLE	0.119 (0.109)	0.158 (0.071)	0.832 (0.076)	25.5	1.3724
	VTE	0.152 (0.119)	0.149 (0.064)	0.823 (0.071)	16.6	

Table 8 Comparison of the computation time of the QMLE and VTE (in seconds of CPU time) for estimating the models of the eleven indices of Table 6. The method VTE+QMLE consists in using the VTE as initial values for the QMLE. Designs 1 and 2 correspond to different initial values (see the text).

	Design 1	Design 2
VTE	39.0	55.5
QMLE	61.6	88.1
VTE+QMLE	85.1	98.9

sensible differences between the estimated parameters of the two methods; (ii) the VTE is a little bit faster and seems more robust relatively to the choice of the initial values; and (iii) the VTE provides good initial values for the QMLE and may avoid that this estimator be trapped in local optima.

Table 9 Comparison of VaR predictions by QMLE and VTE of GARCH(1,1) models for eleven daily stock market returns. Column 4 gives the relative VaR averages in percentages. The last three columns display the p-values of the LR tests of the unconditional coverage hypothesis, the independence hypothesis, and the joint conditional coverage and independence.

Index	Estimator	% Violations	Average	p-value(LR_{uc})	p-value(LR_{ind})	p-value(LR_{cc})
CAC	QMLE	6.46	2.16	0.00	0.44	0.01
	VTE	6.38	2.17	0.00	0.30	0.01
DAX	QMLE	6.60	2.29	0.00	0.06	0.00
	VTE	6.60	2.28	0.00	0.06	0.00
DJA	QMLE	6.58	1.66	0.00	0.52	0.00
	VTE	6.71	1.65	0.00	0.28	0.00
DJI	QMLE	5.87	1.73	0.06	0.55	0.13
	VTE	6.04	1.72	0.02	0.72	0.07
DJT	QMLE	6.16	2.30	0.01	0.64	0.04
	VTE	6.16	2.29	0.01	0.64	0.04
DJU	QMLE	8.95	1.70	0.00	0.09	0.00
	VTE	9.16	1.69	0.00	0.04	0.00
FTSE	QMLE	6.53	1.79	0.00	0.21	0.00
	VTE	6.53	1.79	0.00	0.21	0.00
Nasdaq	QMLE	6.54	3.00	0.00	0.57	0.00
	VTE	6.54	3.00	0.00	0.57	0.00
Nikkei	QMLE	5.63	2.40	0.17	0.50	0.31
	VTE	5.68	2.39	0.14	0.48	0.26
SMI	QMLE	5.73	1.83	0.12	0.49	0.23
	VTE	5.73	1.83	0.12	0.49	0.23
SP500	QMLE	6.25	1.82	0.01	0.51	0.02
	VTE	6.54	1.79	0.00	0.50	0.00

4.3 Backtesting VaRs

We now study how the VaR performances are influenced by the choice of the estimator. To this aim, we re-estimated GARCH(1,1) models for the daily indices, but on half of each sample. The rest of the sample was kept for evaluating the VaR performances. The results are displayed in Table 9. The first columns indicates the percentage of the daily losses that are larger than the VaR(5%) predicted at Horizon 1. By the two methods, these percentages appear to be slightly above the nominal 5% level. The last three columns provide the p-values of the likelihood ratio (LR) tests developed by Christoffersen (1998) (see also Christoffersen 2003 for details) to test consequences of the null hypothesis that the "hit sequence" violations are i.i.d. and distributed as a Bernoulli of parameter 5%. While the independence assumption is generally not rejected, for most assets, the joint test lead to rejection of the null hypothesis. One can conclude that the GARCH(1,1) model is not an adequate model for VaR prediction on these series. One plausible explanation for the

Table 10 As Figure 9, but when a St_ν is assumed for the distribution of η_t. The column $\hat{\nu}$ gives the estimates of ν.

Index	Estimator	$\hat{\nu}$	% Violations	Average	p-value(LR_{uc})	p-value(LR_{ind})	p-value(LR_{cc})
CAC	MLE	10.06	4.57	2.39	33.34	56.99	53.30
	VTE	10.06	4.49	2.40	24.55	61.81	45.00
DAX	MLE	6.90	4.15	2.68	5.49	55.33	13.30
	VTE	6.90	4.15	2.68	5.49	55.33	13.30
DJA	MLE	6.24	3.96	1.98	1.51	25.03	2.69
	VTE	6.24	3.96	1.98	1.51	25.03	2.69
DJI	MLE	6.34	3.67	2.05	0.17	60.93	0.63
	VTE	6.34	3.67	2.05	0.17	60.93	0.63
DJT	MLE	7.41	4.04	2.68	2.58	0.62	0.20
	VTE	7.41	4.00	2.68	1.98	1.70	0.38
DJU	MLE	8.35	5.46	2.02	31.20	1.41	2.96
	VTE	8.35	5.50	2.01	27.04	1.62	3.02
FTSE	MLE	12.03	4.90	1.96	82.91	8.65	22.48
	VTE	12.03	4.90	1.96	82.91	8.65	22.48
Nasdaq	MLE	10.55	4.04	3.40	2.61	77.17	8.06
	VTE	10.62	4.04	3.38	2.61	77.17	8.06
Nikkei	MLE	6.88	3.33	2.75	0.01	63.39	0.04
	VTE	6.88	3.41	2.74	0.02	58.36	0.08
SMI	MLE	8.05	4.24	2.07	8.78	76.38	22.26
	VTE	8.05	4.24	2.08	8.78	76.38	22.26
SP500	MLE	6.25	3.67	2.16	0.17	60.93	0.63
	VTE	6.25	3.67	2.16	0.17	60.93	0.63

too high number of violations is that the VaR are computed under the Gaussian assumption for η_t, which is often considered as unrealistic for such series. For this reason, we re-estimated the GARCH(1,1) models with a Student distribution St_ν for η_t (normalized in such a way that $E\eta_t^2 = 1$). The three volatility parameters and the degree of freedom ν have been estimated by the maximum likelihood estimator (MLE) and by VTE (i.e., with a Student MLE under the constraint that the variance of the estimated model is equal to the marginal variance). The estimated values of ν are given in Table 10. The VaR computations are now more satisfactory, at least for the European indices. As far as our comparison is concerned, the two estimators lead to similar VaR evaluations, for the horizon $h = 1$ as well as for larger horizons (not reported here).

5 CONCLUSION

VTE is a two-step estimation method that reduces the computational complexity of the optimization procedure and guarantees that the marginal variance of the estimated model is equal to the sample variance. This paper provides asymptotic results for the VTE, allowing for valid inference procedures, such as tests or the construction of confidence intervals, based on this method. This paper also compares the asymptotic and empirical performances of the VTE to the standard QMLE.

One evident drawback of the VTE is that the existence of $E(\epsilon_t^4)$ is required for the asymptotic normality, whereas the strict stationarity suffices for the asymptotic normality of the QMLE. It was not immediately clear how the asymptotic distribution of the VTE would compare to the standard QMLE asymptotic distribution. In particular, one might have thought that (i) the VTE could asymptotically outperform the QMLE for some error distributions, (ii) the variance targeting procedure would not substantially affect the asymptotic precision of the GARCH coefficients since the sample variance converges to the population variance. Our results show that both claims are incorrect: (i) the asymptotic variance of the VTE can never be smaller than that of the QMLE; (ii) the variance targeting may result in a serious deterioration of the asymptotic precision when the moment condition is close to be violated. On the other hand, the finite-sample performance of the VTE seems quite satisfactory. Moreover, our experiments on daily stock returns do not show sensible differences between the estimated parameters of the two methods. Finally, we have shown that, for some specific purposes such as long-term prediction, the fact that the VTE guarantees a consistent estimation of the long-run variance may be a crucial advantage of the VTE over the QMLE.

While this paper has provided evidence, there is value in considering a VTE in GARCH models, there remain interesting questions in this area. Other moments could be targeted, not only the long-run variance, and it would be interesting to examine the asymptotic properties of the resulting estimators. In particular, in a multivariate framework, "correlation targeting" has been considered by Engle (2002) for the specification of the dynamic conditional correlation model. Another issue is the extension of the present results for models accommodating the "leverage effect" generally observed in financial series.

APPENDIX A: PROOFS

Let

$$\mathbf{l}_n(\boldsymbol{\lambda}) = n^{-1}\sum_{t=1}^{n} \ell_t(\gamma_0, \boldsymbol{\lambda}), \quad \ell_t(\gamma, \boldsymbol{\lambda}) = \ell_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\vartheta})} + \log \sigma_t^2(\boldsymbol{\vartheta}).$$

For $t \geqslant 1$, we define

$$\tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})} + \log \tilde{\sigma}_t^2(\boldsymbol{\vartheta}).$$

In this appendix, the letters K and ρ denote generic constants, whose values can vary along the text but always satisfy $K > 0$ and $0 < \rho < 1$.

A.1. Proof of consistency in Theorem 1.1

We will follow the proof that Francq and Zakoïan (2004), hereafter FZ, gave for the strong consistency of the QMLE $\hat{\boldsymbol{\theta}}_n^*$. This result also entails the consistency of $\hat{\boldsymbol{\vartheta}}_n^*$ but is not directly applicable to show the consistency of $\hat{\boldsymbol{\vartheta}}_n$ because the VTE is a two-step estimator that is not expressible as a function of the QMLE.

The almost sure convergence of $\hat{\sigma}_n^2$ to γ_0 is a direct consequence of the ergodic theorem. To show the strong consistency of $\hat{\boldsymbol{\lambda}}_n$, we employ the classical technique of Wald (1949). Following the lines of FZ, it suffices to establish the following results:

1. $\lim_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\Lambda} |\mathbf{l}_n(\boldsymbol{\lambda}) - \tilde{\mathbf{l}}_n(\boldsymbol{\lambda})| = 0, \quad \text{a.s.}$
2. If $\sigma_t^2(\gamma_0, \boldsymbol{\lambda}) = \sigma_t^2(\gamma_0, \boldsymbol{\lambda}_0)$ *a.s.*, then $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$,
3. If $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$, then $E\ell_t(\gamma_0, \boldsymbol{\lambda}) > E\ell_t(\gamma_0, \boldsymbol{\lambda}_0)$,
4. Any $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$ has a neighbourhood $V(\boldsymbol{\lambda})$ such that $\liminf_{n\to\infty} \inf_{\boldsymbol{\lambda}^*\in V(\boldsymbol{\lambda})} \tilde{\mathbf{l}}_n(\boldsymbol{\lambda}^*) > E\ell_1(\gamma_0, \boldsymbol{\lambda}_0)$ a.s.

We first show 1. Note that the difference between $\mathbf{l}_n(\boldsymbol{\lambda})$ and $\tilde{\mathbf{l}}_n(\boldsymbol{\lambda})$ is due to the replacement of γ_0 by $\hat{\sigma}_n^2$ and is also due to the initial values taken for ϵ_0 and $\sigma_0^2(\gamma_0, \boldsymbol{\lambda})$. To handle simultaneously the two sources of difference, note that

$$\begin{aligned}\sigma_{t,n}^2(\boldsymbol{\lambda}) - \sigma_t^2(\gamma_0, \boldsymbol{\lambda}) &= \kappa(\hat{\sigma}_n^2 - \gamma_0) + \beta\{\sigma_{t-1,n}^2(\boldsymbol{\lambda}) - \sigma_{t-1}^2(\gamma_0, \boldsymbol{\lambda})\} \\ &= \kappa(\hat{\sigma}_n^2 - \gamma_0)\frac{1-\beta^t}{1-\beta} + \beta^t\{\sigma_0^2 - \sigma_0^2(\gamma_0, \boldsymbol{\lambda})\}.\end{aligned}$$

Thus, since $\hat{\sigma}_n^2$ converges to γ_0 almost surely, we have

$$\sup_{\boldsymbol{\lambda}\in\Lambda} |\sigma_{t,n}^2 - \sigma_t^2(\gamma_0, \boldsymbol{\lambda})| \leqslant K\rho^t + o(1) \quad \text{a.s.}$$

Note that K is a measurable function of $\{\epsilon_u,\ u \leqslant 0\}$. For the almost sure consistency, the trajectory is fixed in a set of probability one and n tends to infinity. Thus, K can be considered as a constant, that is, K is almost surely invariant with n. The requirement in 1 follows from

$$\begin{aligned}\sup_{\boldsymbol{\lambda}\in\Lambda} |\mathbf{l}_n(\boldsymbol{\lambda}) - \tilde{\mathbf{l}}_n(\boldsymbol{\lambda})| &\leqslant n^{-1}\sum_{t=1}^n \sup_{\boldsymbol{\lambda}\in\Lambda}\left\{\left|\frac{\sigma_{t,n}^2 - \sigma_t^2(\gamma_0, \boldsymbol{\lambda})}{\sigma_{t,n}^2\sigma_t^2(\gamma_0, \boldsymbol{\lambda})}\right|\epsilon_t^2\right. \\ &\qquad\left. + \left|\log\left(1 + \frac{\sigma_t^2(\gamma_0, \boldsymbol{\lambda}) - \sigma_{t,n}^2}{\sigma_{t,n}^2}\right)\right|\right\} \\ &\leqslant \left\{\sup_{\boldsymbol{\lambda}\in\Lambda}\frac{1}{\kappa^2}\right\}\frac{1}{\gamma_0\hat{\sigma}_n^2}Kn^{-1}\sum_{t=1}^n \rho^t\epsilon_t^2 \\ &\quad + \left\{\sup_{\boldsymbol{\lambda}\in\Lambda}\frac{1}{\kappa}\right\}\frac{1}{\hat{\sigma}_n^2}Kn^{-1}\sum_{t=1}^n \rho^t + o(1)\end{aligned}$$

and the arguments used in the proof in FZ.

The requirements 2 and 3 have already been proven in FZ in a more general framework (see the proof of their Theorem 2.1). The proof of 4 is also a direct adaptation of the proof given in FZ. For the reader convenience, we briefly restate the proofs of 2–4 in our particular GARCH(1,1) framework. To show 2 note that

$$\sigma_t^2(\boldsymbol{\vartheta}) = \frac{\kappa\gamma}{1-\beta} + \alpha \sum_{i=0}^{\infty} \beta^i \epsilon_{t-i-1}^2. \tag{A1}$$

Suppose that $\sigma_t^2(\boldsymbol{\vartheta}) = \sigma_t^2(\boldsymbol{\vartheta}_0)$ a.s. Then, in view of (A1),

$$\frac{\kappa\gamma}{\alpha+\kappa} + \alpha \sum_{i=0}^{\infty} \beta^i \epsilon_{t-i-1}^2 = \frac{\kappa_0\gamma_0}{\alpha_0+\kappa_0} + \alpha_0 \sum_{i=0}^{\infty} \beta_0^i \epsilon_{t-i-1}^2.$$

Because the innovation of ϵ_t^2 is not a.s. equal to zero under **A2**, we must have

$$\frac{\kappa\gamma}{\alpha+\kappa} = \frac{\kappa_0\gamma_0}{\alpha_0+\kappa_0} \quad \text{and} \quad \alpha\beta^i = \alpha_0\beta_0^i \quad \forall\, i \geqslant 0.$$

This entails $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$.

To show 3, we argue that for $x > 0$, $\log x \leqslant x - 1$, with equality if and only if $x = 1$. We thus have

$$\begin{aligned} E\ell_t(\boldsymbol{\vartheta}) - E\ell_t(\boldsymbol{\vartheta}_0) &= E \log \frac{\sigma_t^2(\boldsymbol{\vartheta})}{\sigma_t^2(\boldsymbol{\vartheta}_0)} + E \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} - 1 \\ &\geqslant E\left\{ \log \frac{\sigma_t^2(\boldsymbol{\vartheta})}{\sigma_t^2(\boldsymbol{\vartheta}_0)} + \log \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right\} = 0 \end{aligned}$$

with equality if and only if $\sigma_t^2(\boldsymbol{\vartheta}_0)/\sigma_t^2(\boldsymbol{\vartheta}) = 1$ a.s., which is equivalent to $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ in view of 2.

Let us show 4. Let $V_k(\boldsymbol{\lambda})$ be the open ball with center $\boldsymbol{\lambda}$ and radius $1/k$. Using successively 1, the ergodicity of the process, the monotone convergence theorem and 3, we obtain almost surely

$$\begin{aligned} \liminf_{n\to\infty} \inf_{\boldsymbol{\lambda}^* \in V_k(\boldsymbol{\lambda})\cap\Lambda} \tilde{\mathbf{l}}_n(\boldsymbol{\lambda}^*) &\geqslant \liminf_{n\to\infty} \inf_{\boldsymbol{\lambda}^* \in V_k(\boldsymbol{\lambda})\cap\Lambda} \mathbf{l}_n(\boldsymbol{\lambda}^*) - \limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\Lambda} |\mathbf{l}_n(\boldsymbol{\lambda}) - \tilde{\mathbf{l}}_n(\boldsymbol{\lambda})| \\ &\geqslant \liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} \inf_{\boldsymbol{\lambda}^* \in V_k(\boldsymbol{\lambda})\cap\Lambda} \ell_t(\gamma_0, \boldsymbol{\lambda}^*) \\ &= E \inf_{\boldsymbol{\lambda}^* \in V_k(\boldsymbol{\lambda})\cap\Lambda} \ell_1(\gamma_0, \boldsymbol{\lambda}^*) \\ &> E\ell_1(\gamma_0, \boldsymbol{\lambda}_0) \end{aligned}$$

for k large enough, when $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$.

A.2. Proof of the asymptotic normality in Theorem 1.1

The proof of the asymptotic normality rests classically on a Taylor-series expansion of each component of the score vector around $\boldsymbol{\vartheta}_0$. In comparison to the proof given by FZ for the QMLE, additional difficulties come from the fact that the VTE is a two-step estimator. On the other hand, the proof of some technical parts will be facilitated by the assumption $E\epsilon_t^4 < \infty$. Although restrictive, this moment assumption is required for the asymptotic normality of the empirical variance $\hat{\sigma}_n^2$. Write $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \vartheta_3)$. Noting that, in (10), $\ell_{t,n}(\boldsymbol{\lambda}) = \tilde{\ell}_t(\hat{\sigma}_n^2, \boldsymbol{\lambda})$, we have

$$
\begin{aligned}
(0,0)' &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\ell_{t,n}(\hat{\boldsymbol{\lambda}}_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\tilde{\ell}_t(\hat{\boldsymbol{\vartheta}}_n)\\
&= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\lambda_i\partial\vartheta_j}\tilde{\ell}_t(\boldsymbol{\vartheta}_i^*)\right)_{2\times 3}\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)\\
&= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{J}_n\sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) + \boldsymbol{K}_n\sqrt{n}(\hat{\sigma}_n^2 - \gamma_0), \qquad \text{(A2)}
\end{aligned}
$$

where the $\boldsymbol{\vartheta}_i^*$ satisfy $\|\boldsymbol{\vartheta}_i^* - \boldsymbol{\vartheta}_0\| \leqslant \|\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\|$,

$$
\boldsymbol{J}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\lambda_i\partial\lambda_j}\tilde{\ell}_t(\boldsymbol{\vartheta}_i^*)\right)_{2\times 2}, \quad \boldsymbol{K}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\gamma\partial\lambda_1}\tilde{\ell}_t(\boldsymbol{\vartheta}_1^*), \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\gamma\partial\lambda_2}\tilde{\ell}_t(\boldsymbol{\vartheta}_2^*)\right)'.
$$

We will show that

1. $E\left\|\dfrac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\dfrac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right\| < \infty, \quad E\left\|\dfrac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\| < \infty,$
2. $\boldsymbol{A} := E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right)$ is nonsingular and $\mathrm{Var}\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\} = \left\{E\eta_0^4 - 1\right\}$ $\boldsymbol{A}$,
3. There exists a neighbourhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that, for all $i, j, k \in \{1, \ldots, p + q + 1\}$,
$$E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{\partial^3\ell_t(\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j\partial\vartheta_k}\right| < \infty,$$
4. $\left\|n^{-1/2}\sum_{t=1}^{n}\left\{\dfrac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} - \dfrac{\partial\tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\}\right\| \to 0$ and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|n^{-1}\sum_{t=1}^{n}\left\{\dfrac{\partial^2\ell_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \dfrac{\partial^2\tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\}\right\| \to 0$ in probability when $n \to \infty$,
5. $n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ell_t(\boldsymbol{\vartheta}_k^*) \to \boldsymbol{A}(i,j)$ a.s.
6. $\boldsymbol{X}_n := \begin{pmatrix} n^{1/2}\left(\hat{\sigma}_n^2 - \gamma_0\right) \\ n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\ell_t(\boldsymbol{\vartheta}_0)\end{pmatrix} \Rightarrow \mathcal{N}\left\{0, (E\eta_0^4 - 1)\begin{pmatrix} b & 0 \\ 0 & \boldsymbol{J}\end{pmatrix}\right\}.$

The derivatives of $\ell_t = \epsilon_t^2/\sigma_t^2 + \log\sigma_t^2$ are given by

$$\frac{\partial \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}(\boldsymbol{\vartheta}), \tag{A3}$$

$$\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}) + \left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}). \tag{A4}$$

The same formulas hold for the derivatives of $\tilde{\ell}_t$, with σ_t^2 replaced by $\tilde{\sigma}_t^2$.

For $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, $\epsilon_t^2/\sigma_t^2 = \eta_t^2$ is independent of the terms involving σ_t^2 and its derivatives. To prove 1, it will therefore be sufficient to show that

$$E\left\|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0)\right\| < \infty, \quad E\left\|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0)\right\| < \infty, \quad E\left\|\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0)\right\| < \infty. \tag{A5}$$

The following expansion holds:

$$\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) = \left(\frac{\kappa}{1-\beta}, \frac{-\kappa\gamma}{(1-\beta)^2} + \sum_{\ell=0}^{\infty}\beta^{\ell}\epsilon_{t-\ell-1}^2 - \alpha\sum_{\ell=1}^{\infty}\ell\beta^{\ell-1}\epsilon_{t-\ell-1}^2, \frac{\alpha\gamma}{(1-\beta)^2} - \alpha\sum_{\ell=1}^{\infty}\ell\beta^{\ell-1}\epsilon_{t-\ell-1}^2\right)'.$$

Recall that **A3** implies $E\epsilon_t^4 < \infty$. Moreover, we have $\sigma_t^{-2}(\boldsymbol{\vartheta}_0) \leqslant \kappa_0^{-1}\gamma_0^{-1} < \infty$. This allows us to prove the first and third inequalities in (A5). The second inequality is proved by exactly the same arguments, and 1 is proved. Note that we used the moment assumption $E\epsilon_t^4 < \infty$ to facilitate the proof of (A5). This moment assumption is actually unnecessary. Indeed, we will show, without this assumption, that for any integer d, there exists a neighbourhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \vartheta_i}\right|^d < \infty, \quad E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \vartheta_i\partial \vartheta_j}\right|^d < \infty,$$
$$E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial^3 \sigma_t^2}{\partial \vartheta_i\partial \vartheta_j\partial \vartheta_k}\right|^d < \infty. \tag{A6}$$

Choose $\mathcal{V}(\boldsymbol{\vartheta}_0)$ small enough, so that all the parameters γ, κ, α, and β be bounded away from zero. Using the elementary inequality $x/(1+x) \leqslant x^s$ for all $x \geqslant 0$ and all $s \in (0,1]$, for all $\boldsymbol{\vartheta} \in \mathcal{V}(\boldsymbol{\vartheta}_0)$, we have

$$\left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \gamma}(\boldsymbol{\vartheta}) \right| \leqslant K,$$

$$\left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha}(\boldsymbol{\vartheta}) \right| \leqslant K + \sum_{\ell=0}^{\infty} \frac{\beta^\ell \epsilon_{t-\ell-1}^2}{K + \alpha \beta^\ell \epsilon_{t-\ell-1}^2} + \alpha \sum_{\ell=1}^{\infty} \frac{\ell \beta^{\ell-1} \epsilon_{t-\ell-1}^2}{K + \alpha \beta^\ell \epsilon_{t-\ell-1}^2}$$
$$\leqslant K + K \sum_{\ell=0}^{\infty} \beta^{\ell s} \epsilon_{t-\ell-1}^{2s} + K \sum_{\ell=0}^{\infty} \ell \beta^{\ell s} \epsilon_{t-\ell-1}^{2s}.$$

Similarly,

$$\left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \kappa}(\boldsymbol{\vartheta}) \right| \leqslant K + K \sum_{\ell=0}^{\infty} \beta^{\ell s} \epsilon_{t-\ell-1}^{2s} + K \sum_{\ell=0}^{\infty} \ell \beta^{\ell s} \epsilon_{t-\ell-1}^{2s}.$$

Under (5), we have $E\epsilon_t^2 < \infty$ and $\sup_{\boldsymbol{\vartheta} \in \mathcal{V}(\boldsymbol{\vartheta}_0)} \beta < 1$. Thus, $\|\epsilon_t^{2s}\|_d < \infty$ for some $s \in (0,1]$, and the first result of (A6) comes from the Hlder inequality. The other results of (A6) are obtained with the same arguments.

Now, we prove 2. If $\boldsymbol{A}$ is singular, there exists $\boldsymbol{x} = (x_1, x_2, x_3) \neq 0$ such that $\boldsymbol{x}' \left\{ \partial \sigma_t^2(\boldsymbol{\vartheta}_0) / \partial \boldsymbol{\vartheta} \right\} = 0$ a.s. Since

$$\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}} = \frac{\partial \kappa \gamma}{\partial \boldsymbol{\vartheta}} + \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}} \epsilon_{t-1}^2 + \frac{\partial \beta}{\partial \boldsymbol{\vartheta}} \sigma_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \boldsymbol{\vartheta}}, \tag{A7}$$

the strict stationarity of $\sigma_t^2(\boldsymbol{\vartheta}_0)$ implies

$$\boldsymbol{x}' \frac{\partial \kappa \gamma}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0) + \boldsymbol{x}' \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0) \epsilon_{t-1}^2 + \boldsymbol{x}' \frac{\partial \beta}{\partial \boldsymbol{\vartheta}} \sigma_{t-1}^2(\boldsymbol{\vartheta}_0) = 0 \quad \text{a.s.}$$

We thus have

$$x_1 \kappa_0 + x_3 \gamma_0 + x_2 \epsilon_{t-1}^2 = (x_2 + x_3) \sigma_{t-1}^2(\boldsymbol{\vartheta}_0) \quad \text{a.s.}$$

This entails that $x_2 \epsilon_{t-1}^2$ is a function of $\{\epsilon_{t-i}^2,\ i > 1\}$, which is impossible under **A2**, unless $x_2 = 0$. We then obtain that $x_3 = 0$ because $\sigma_{t-1}^2(\boldsymbol{\vartheta}_0)$ is not almost surely constant. It then follows that $x_1 = 0$. Finally, we obtained a contradiction and the nonsingularity of $\boldsymbol{A}$ is proved.

Let us prove 3. Differentiating (A4), we obtain

$$\begin{aligned} \frac{\partial^3 \ell_t(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j \partial \vartheta_k} &= \left\{ 1 - \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \vartheta_i \partial \vartheta_j \partial \vartheta_k} \right\} (\boldsymbol{\vartheta}) + \left\{ 2 \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_i} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \vartheta_j \partial \vartheta_k} \right\} (\boldsymbol{\vartheta}) \\ &+ \left\{ 2 \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_j} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \vartheta_i \partial \vartheta_k} \right\} (\boldsymbol{\vartheta}) + \left\{ 2 \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_k} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \vartheta_i \partial \vartheta_j} \right\} (\boldsymbol{\vartheta}) \\ &+ \left\{ 2 - 6 \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_i} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_j} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_k} \right\} (\boldsymbol{\vartheta}). \end{aligned} \tag{A8}$$

Because $\inf_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\sigma_t^2(\boldsymbol{\vartheta})>0$ and $E\epsilon_t^4<\infty$, we have

$$E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\{1-\frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\vartheta})}\right\}^2<\infty,\quad E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\{2\frac{\epsilon_t^2}{\sigma_t^2}-1\right\}^2<\infty,\quad E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\{2-6\frac{\epsilon_t^2}{\sigma_t^2}\right\}^2<\infty.$$

In view of this result and of (A6) with $d=1,2,3$, the Hlder inequality entails 3.

We now turn to the proof of 4. In view of (8) and (11), we have

$$\sigma_t^2(\boldsymbol{\vartheta})-\tilde{\sigma}_t^2(\boldsymbol{\vartheta})=\beta^t\{\sigma_0^2(\boldsymbol{\vartheta})-\tilde{\sigma}_0^2\}.$$

Therefore, choosing $\mathcal{V}(\boldsymbol{\vartheta}_0)$ such that $\boldsymbol{\lambda}\in\Lambda$ for all $\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)$, we have

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}|\sigma_t^2(\boldsymbol{\vartheta})-\tilde{\sigma}_t^2(\boldsymbol{\vartheta})|\leqslant K\rho^t,\qquad\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}-\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\|\leqslant K\rho^t$$

and

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2(\boldsymbol{\vartheta})}-\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})}\right|=\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2(\boldsymbol{\vartheta})}\{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})-\sigma_t^2(\boldsymbol{\vartheta})\}\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})}\right|\leqslant K\rho^t.$$

In view of (A3), we then obtain

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|n^{-1}\sum_{t=1}^n\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}-\frac{\partial\tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\}\right\|\leqslant Kn^{-1}\sum_{t=1}^n\rho^t\Upsilon_t,\quad\text{a.s.,}\tag{A9}$$

where

$$\Upsilon_t=\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\{1+\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{1+\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\vartheta}}\right\}(\boldsymbol{\vartheta}).$$

Using $E\epsilon_t^4<\infty$, $\inf_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\sigma_t^2(\boldsymbol{\vartheta})>0$ and (A6), the Cauchy–Schwarz inequality shows that $E\Upsilon_t<\infty$. By the Borel–Cantelli lemma, it follows that $\rho^t\Upsilon_t\to0$ a.s. and thus the right-hand side of (A9) converges to 0 a.s. Using the same arguments and replacing first derivatives with second derivatives, 4 is shown.

Similarly to the proof of (4.36) in FZ, 5 follows from the Taylor expansion of $n^{-1}\sum_{t=1}^n\partial^2\ell_t(\boldsymbol{\vartheta}_k^*)/\partial\vartheta_i\partial\vartheta_j$ around θ_0, the convergence of $\boldsymbol{\vartheta}_k^*$ to θ_0, and 3.

The proof of 6 relies on a central limit theorem for martingale differences. From Horváth, Kokoszka, and Zitikis (2006, Proof of Theorem 1) (see (A15) below), we have the representation

$$\hat{\sigma}_n^2=\gamma_0+\frac{1-\beta_0}{\kappa_0}\frac{1}{n}\sum_{t=1}^n h_t(\eta_t^2-1)+o_P(n^{-1/2}).\tag{A10}$$

Moreover, in view of (A3)

$$n^{-1/2}\sum_{t=1}^n\frac{\partial}{\partial\boldsymbol{\lambda}}\ell_t(\boldsymbol{\vartheta}_0)=n^{-1/2}\sum_{t=1}^n\frac{1-\eta_t^2}{h_t}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\lambda}}(\boldsymbol{\vartheta}_0).\tag{A11}$$

We then have

$$\boldsymbol{X}_n = n^{-1/2}\sum_{t=1}^{n}(1-\eta_t^2)\mathbf{Z}_t + o_P(1), \quad \mathbf{Z}_t = \begin{pmatrix} -(1-\beta_0)\kappa_0^{-1}h_t \\ h_t^{-1}\partial\sigma_t^2(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\lambda} \end{pmatrix}. \tag{A12}$$

Notice that $E\left((1-\eta_t^2)\mathbf{Z}_t|\mathcal{F}_{t-1}\right) = 0$, where $\mathcal{F}_t$ is the σ-algebra generated by the random variables ϵ_{t-i}, $i \geqslant 0$. Moreover, we have

$$\frac{\partial\sigma_t^2}{\partial\alpha} = \epsilon_{t-1}^2 - \sigma_{t-1}^2 + \beta\frac{\partial\sigma_{t-1}^2}{\partial\alpha} = \sum_{i=0}^{\infty}\beta^i(\epsilon_{t-i-1}^2 - \sigma_{t-i-1}^2),$$
$$\frac{\partial\sigma_t^2}{\partial\kappa} = \gamma - \sigma_{t-1}^2 + \beta\frac{\partial\sigma_{t-1}^2}{\partial\kappa} = \sum_{i=0}^{\infty}\beta^i(\gamma - \sigma_{t-i-1}^2),$$

and thus

$$E\left\{\frac{\partial\sigma_t^2}{\partial\boldsymbol{\lambda}}(\boldsymbol{\vartheta}_0)\right\} = 0. \tag{A13}$$

It follows that

$$\mathrm{Var}\{(1-\eta_t^2)\mathbf{Z}_t\} = (E\eta_0^4 - 1)\begin{pmatrix} b & 0 \\ 0 & \boldsymbol{J} \end{pmatrix}.$$

Notice that b is a positive real number and that the matrix $\boldsymbol{J}$ in the right-lower block of $\boldsymbol{A}$ is nonsingular, in view of the nonsingularity of $\boldsymbol{A}$. By Assumptions **A2** and **A3**, we get $0 < E\eta_0^4 - 1 < \infty$, and thus the matrix $\mathrm{Var}\{(1-\eta_t^2)\mathbf{Z}_t\}$ is nondegenerate. Hence, for any $\boldsymbol{\lambda} \in \mathbb{R}^3$, the sequence $\{(1-\eta_t^2)\boldsymbol{\lambda}'\mathbf{Z}_t, \mathcal{F}_t\}_t$ is a square integrable stationary martingale difference. By (A12), the central limit theorem of Billingsley (1961) and the Wold–Cramer device, we obtain the asymptotic normality of $\boldsymbol{X}_n$, which proves 5.

To complete the proof of the theorem, note that, from 2, 4, and 5, it follows that the matrix $\boldsymbol{J}_n$ is a.s. invertible for sufficiently large n. Therefore, in view of (A2),

$$\sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = -\boldsymbol{J}_n^{-1}\left\{n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\lambda}}\tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{K}_n\sqrt{n}(\hat{\sigma}_n^2 - \gamma_0)\right\}.$$

It follows that, using 4,

$$\sqrt{n}\begin{pmatrix} \hat{\sigma}_n^2 - \gamma_0 \\ \hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\boldsymbol{J}_n^{-1}\boldsymbol{K}_n & -\boldsymbol{J}_n^{-1} \end{pmatrix}\boldsymbol{X}_n + o_P(1).$$

Thus, by 5, 6, and Slutsky's lemma, $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)$ is asymptotically $\mathcal{N}(0,\boldsymbol{\Sigma})$ distributed, with

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ -\boldsymbol{J}^{-1}\boldsymbol{K} & -\boldsymbol{J}^{-1} \end{pmatrix}\begin{pmatrix} b & 0 \\ 0 & \boldsymbol{J} \end{pmatrix}\begin{pmatrix} 1 & -\boldsymbol{K}'\boldsymbol{J}^{-1} \\ 0 & -\boldsymbol{J}^{-1} \end{pmatrix}.$$

The invertibility of $\boldsymbol{\Sigma}$ follows and the proof of Theorem 1.1 is complete.

A.3. Proof of Corollary 1

This corollary of Theorem 1.1 can be proven by a direct application of the delta method (see, e.g., Theorem 3.1 in van der Vaart 1998). Indeed, the map $\boldsymbol{\phi}$ that transforms $\boldsymbol{\vartheta}_0$ into $\boldsymbol{\theta}_0$ is differentiable at $\boldsymbol{\vartheta}_0$, and the Jacobian matrix of this map is

$$\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}_0'} = \begin{pmatrix} \partial(\gamma_0\kappa_0)/\partial\gamma_0 & \partial(\gamma_0\kappa_0)/\partial\alpha_0 & \partial(\gamma_0\kappa_0)/\partial\kappa_0 \\ \partial\alpha_0/\partial\gamma_0 & \partial\alpha_0/\partial\alpha_0 & \partial\alpha_0/\partial\kappa_0 \\ \partial(1-\kappa_0-\alpha_0)/\partial\gamma_0 & \partial(1-\kappa_0-\alpha_0)/\partial\alpha_0 & \partial(1-\kappa_0-\alpha_0)/\partial\kappa_0 \end{pmatrix} = \boldsymbol{L}'.$$

A.4. Proof of Corollary2

It is known that for an invertible partitioned matrix

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix},$$

if $\boldsymbol{A}_{11}$ is invertible, then we have

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{F} & -\boldsymbol{F}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{F} & \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{F}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \end{bmatrix},$$

where $\boldsymbol{F} = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}$. Using this classical result, we get

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} a & -a\boldsymbol{K}'\boldsymbol{J}^{-1} \\ -a\boldsymbol{J}^{-1}\boldsymbol{K} & \boldsymbol{J}^{-1} + a\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} \end{pmatrix}, \quad a = \left\{ \frac{\kappa_0^2}{(\alpha_0+\kappa_0)^2} E\left(\frac{1}{h_t^2}\right) - \boldsymbol{K}'\boldsymbol{J}^{-1}\boldsymbol{K} \right\}^{-1},$$

whereas

$$\boldsymbol{\Sigma} = \begin{pmatrix} b & -b\boldsymbol{K}'\boldsymbol{J}^{-1} \\ -b\boldsymbol{J}^{-1}\boldsymbol{K} & \boldsymbol{J}^{-1} + b\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} \end{pmatrix}, \quad b = \frac{(\alpha_0+\kappa_0)^2}{\kappa_0^2} E(h_t^2).$$

Thus,

$$\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^* = (b-a) \begin{pmatrix} 1 & -\boldsymbol{K}'\boldsymbol{J}^{-1} \\ -\boldsymbol{J}^{-1}\boldsymbol{K} & \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} \end{pmatrix},$$

and the result follows.

A.5. Proof of Theorem 2.1

The consistency can be obtained as in FZ by a direct extension of Theorem 1.1. Let us concentrate on the asymptotic normality. Similarly to (A2), we have

$$0_{p+q} = n^{-1/2} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\hat{\boldsymbol{\vartheta}}_n)$$

$$= n^{-1/2} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \lambda_i \partial \vartheta_j} \tilde{\ell}_t(\boldsymbol{\vartheta}_i^*) \right)_{(p+q)\times(p+q+1)} \sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)$$

$$= n^{-1/2} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{J}_n \sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) + \boldsymbol{K}_n \sqrt{n}(\hat{\sigma}_n^2 - \gamma_0), \tag{A14}$$

where the $\boldsymbol{\vartheta}_i^*$ satisfy $\|\boldsymbol{\vartheta}_i^* - \boldsymbol{\vartheta}_0\| \leqslant \|\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\|$,

$$\boldsymbol{J}_n = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \tilde{\ell}_t(\boldsymbol{\vartheta}_i^*) \right)_{(p+q)\times(p+q)},$$

$$\boldsymbol{K}_n = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \gamma \partial \lambda_1} \tilde{\ell}_t(\boldsymbol{\vartheta}_1^*), \ldots, \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \gamma \partial \lambda_{p+q}} \tilde{\ell}_t(\boldsymbol{\vartheta}_{p+q}^*) \right)'.$$

We now use the ARMA representation for ϵ_t^2:

$$\epsilon_t^2 = \omega_0 + \sum_{i=1}^{p\vee q} (\alpha_{0i} + \beta_{0i}) \epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{p} \beta_{0j} \nu_{t-j},$$

with $\nu_t = \epsilon_t^2 - h_t = h_t(\eta_t^2 - 1)$ and obvious conventions. Taking the average of both sides of the equality when t varies from 1 to n, we obtain

$$\hat{\sigma}_n^2 = \omega_0 + \sum_{i=1}^{p\vee q} (\alpha_{0i} + \beta_{0i}) \hat{\sigma}_n^2 + \left(1 - \sum_{j=1}^{p} \beta_{0j} \right) \frac{1}{n} \sum_{t=1}^{n} \nu_t + O_P(n^{-1}),$$

which allows us to extend the representation (A15) of Horváth, Kokoszka, and Zitikis (2006)

$$\hat{\sigma}_n^2 = \gamma_0 + \frac{1 - \sum_{j=1}^{p} \beta_{0j}}{1 - \sum_{j=1}^{q} \alpha_{0i} - \sum_{j=1}^{p} \beta_{0j}} \frac{1}{n} \sum_{t=1}^{n} \nu_t + o_P(n^{-1/2}). \tag{A15}$$

Using (A14), (A15), and direct extensions of 4 and 6, we obtain

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = \begin{pmatrix} 1 & 0 \\ -\boldsymbol{J}_n^{-1} \boldsymbol{K}_n & -\boldsymbol{J}_n^{-1} \end{pmatrix} \boldsymbol{X}_n + o_P(1), \tag{A16}$$

where

$$\boldsymbol{X}_n := \begin{pmatrix} n^{1/2} \left(\hat{\sigma}_n^2 - \gamma_0 \right) \\ n^{-1/2} \sum_{t=1}^{n} \frac{1-\eta_t^2}{h_t} \frac{\partial}{\partial \boldsymbol{\lambda}} \sigma_t^2(\boldsymbol{\vartheta}_0) \end{pmatrix} \Rightarrow \mathcal{N} \left\{ 0, (E\eta_0^4 - 1) \begin{pmatrix} c & 0 \\ 0 & \boldsymbol{J} \end{pmatrix} \right\},$$

noting that (A13) still holds. The conclusion follows.

A.6. Proof of Proposition 3

Let

$$\boldsymbol{S}_t = \begin{pmatrix} e h_t^{-1} \\ h_t^{-1} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\lambda}}(\boldsymbol{\vartheta}_0) \\ e^{-1} h_t \end{pmatrix}, \qquad e = \frac{1 - \sum_{i=1}^q \alpha_{0i} - \sum_{j=1}^p \beta_{0j}}{1 - \sum_{i=1}^q \beta_{0i}}.$$

Using (A13), we observe that

$$E\left(\boldsymbol{S}_t \boldsymbol{S}_t'\right) = \begin{pmatrix} e^2 E h_t^{-2} & \boldsymbol{K}' & 1 \\ \boldsymbol{K} & \boldsymbol{J} & 0 \\ 1 & 0 & c \end{pmatrix}.$$

We thus have

$$\boldsymbol{\Sigma}^* = \{E(\boldsymbol{G}\boldsymbol{S}_t\boldsymbol{S}_t'\boldsymbol{G}')\}^{-1}, \quad \boldsymbol{\Sigma} = E(\boldsymbol{H}\boldsymbol{S}_t\boldsymbol{S}_t'\boldsymbol{H}'),$$

where

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{I}_{p+q+1} & 0 \end{pmatrix}, \qquad \boldsymbol{H} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \boldsymbol{J}^{-1} & -\boldsymbol{J}^{-1}\boldsymbol{K} \end{pmatrix},$$

$\boldsymbol{I}_k$ denoting the identity matrix of size k. Note that $\boldsymbol{G}E\left(\boldsymbol{S}_t\boldsymbol{S}_t'\right)\boldsymbol{H}' = \boldsymbol{I}_{p+q}$. Letting $\boldsymbol{D}_t = \boldsymbol{\Sigma}^*\boldsymbol{G}\boldsymbol{S}_t - \boldsymbol{H}\boldsymbol{S}_t$, we then obtain

$$E(\boldsymbol{D}_t\boldsymbol{D}_t') = \boldsymbol{\Sigma}^* + \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*\boldsymbol{G}E(\boldsymbol{S}_t\boldsymbol{S}_t')\boldsymbol{H}' - \boldsymbol{H}E(\boldsymbol{S}_t\boldsymbol{S}_t')\boldsymbol{G}'\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*$$

which shows that $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*$ is positive semidefinite. It can be seen that the matrix

$$E\boldsymbol{D}_t\boldsymbol{D}_t' = (\boldsymbol{\Sigma}^*\boldsymbol{G} - \boldsymbol{H})E(\boldsymbol{S}_t\boldsymbol{S}_t')(\boldsymbol{\Sigma}^*\boldsymbol{G} - \boldsymbol{H})'$$

is not positive definite because

$$\boldsymbol{\Sigma}^*\boldsymbol{G} - \boldsymbol{H} = \begin{pmatrix} d & -d\boldsymbol{K}'\boldsymbol{J}^{-1} & -1 \\ -d\boldsymbol{J}^{-1}\boldsymbol{K} & d\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} & \boldsymbol{J}^{-1}\boldsymbol{K} \end{pmatrix} = \begin{pmatrix} d\boldsymbol{C}\boldsymbol{C}' & -\boldsymbol{C} \end{pmatrix}$$

is of Rank 1.

A.7. Proof of Corollary 4

The first convergence in distribution is a direct consequence of (17) and of the delta method. In view of (18), (19), and the delta method, we have

$$\sqrt{n}\{\boldsymbol{\phi}(\hat{\boldsymbol{\vartheta}}_n^*) - \boldsymbol{\phi}(\boldsymbol{\vartheta}_0)\} \xrightarrow{d} \mathcal{N}(0, s^{*2}),$$

where

$$s^{*2} = (E\eta_0^4 - 1)\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}'}\boldsymbol{\Sigma}^*\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}} = s^2 - (c - d)\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}'}CC'\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\vartheta}}.$$

The conclusion follows.

A.8. Proof of Lemma 3.1

Let $x \in \mathbb{R}$ and let u_h such that hu_h is a sequence of integers tending to ∞ and $u_h = o\{1/\log\log(hu_h)\}$ as $h \to \infty$. Let

$$D_h = D_h(I_t) = \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \leqslant x | I_t \right\} - P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \leqslant x \right\} \right|,$$

where $I_t = \sigma\{\epsilon_u, u \leqslant t\}$. By Lemma 5.2 in Dvoretzky (1972), we have $ED_h \leqslant 4\alpha_\epsilon(hu_h)$. Moreover, $\sum_h \alpha_\epsilon^{\nu/(2+\nu)} < \infty$ implies $\alpha_\epsilon(h) = o\{h^{-(1+\delta)}\}$ for $0 < \delta < 2/\nu$. Thus,

$$\sum_{h \geqslant 1} \alpha_\epsilon(hu_h) \leqslant \sum_{h \geqslant 1} \left\{ \frac{\log\log(hu_h)}{h} \right\}^{1+\delta} < \infty,$$

and the Borel–Cantelli lemma entails that

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \leqslant x | I_t \right\} - P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \leqslant x \right\} \right| = 0 \quad \text{a.s.} \tag{A17}$$

The α-mixing processes (ϵ_t) thus satisfies a central limit theorem (see Herrndorf 1984) and a law of the iterated logarithm (see Berkes and Philipp 1979; Dehling and Philipp 1982). By the law of the iterated logarithm, we have

$$\frac{1}{\sqrt{h}} \sum_{i=1}^{hu_h} \epsilon_{t+i} = \frac{\sqrt{u_h}}{\sqrt{hu_h}} \sum_{i=1}^{hu_h} \epsilon_{t+i} \to 0 \quad \text{a.s.} \tag{A18}$$

Therefore, for all $\varepsilon > 0$, $P(h^{-1/2} \sum_{i=1}^{hu_h} \epsilon_{t+i} > \varepsilon | I_t) \to 0$ a.s. It follows that

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=1}^{h} \epsilon_{t+i} \leqslant x | I_t \right\} - P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \leqslant x | I_t \right\} \right| = 0 \quad \text{a.s.,} \tag{A19}$$

The central limit theorem entails $h^{-1/2} \sum_{i=1}^{h} \epsilon_{t+i} \xrightarrow{d} \mathcal{N}(0, \gamma)$ as $h \to \infty$. In view of (A17) to (A19), we then obtain

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=1}^{h} \epsilon_{t+i} \leqslant x | I_t \right\} - \Phi(x/\sqrt{\gamma}) \right| = 0 \quad \text{a.s.,} \tag{A20}$$

and the conclusion follows.

Received October 2, 2009; revised June 3, 2011; accepted June 3, 2011.

REFERENCES

Bauwens, L., and J. Rombouts. 2007. Bayesian Clustering of Many GARCH Models. *Econometric Reviews* 26: 365–386.

Berkes, I., L. Horváth, and P. Kokoszka. 2003. GARCH Processes: Structure and Estimation. *Bernoulli* 9: 201–227.

Berkes, I., and W. Philipp. 1979. Approximation Theorems for Independent and Weakly Dependent Random Vectors. *The Annals of Probability* 7: 29–54.

Billingsley, P. 1961. The Lindeberg-Levy Theorem for Martingales. *Proceedings of the American Mathematical Society* 12: 788–792.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Bradley, R. C. 2005. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys* 2: 107–144.

Carrasco, M., and X. Chen. 2002. Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18: 17–39.

Christoffersen, P. F. 1998. Evaluating Interval Forecasts. *International Economic Review* 39: 841–862.

Christoffersen, P. F. 2003. *Elements of Financial Risk Management.* London: Academic Press.

Christoffersen, P. F. 2008. Value-at-Risk Models. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*. Berlin: Springer Verlag.

Dehling, H., and W. Philipp. 1982. Almost Sure Invariance Principles for Weakly Dependent Vector-Valued Random Variables. *The Annals of Probability* 10: 689–701.

Dvoretzky, A. 1972. Asymptotic Normality for Sums of Dependent Random Variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (University of California, Berkeley, California, 1970/1971), Vol. II: Probability Theory*, 513–535.

Engle, R. F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica* 50: 987–1007.

Engle, R. F. 2002. Dynamic Conditional Correlation : a Simple Class of Multivariate GARCH Models. *Journal of Business of Economic Statistics* 20: 339–350.

Engle, R. F., and K. Kroner. 1995. Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11: 122–150.

Engle, R. F., and J. Mezrich. 1996. GARCH for Groups. *Risk* 9: 36–40.

Francq, C., and J.-M. Zakoïan. 2004. Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* 10: 605–637.

Francq, C., and J.-M. Zakoïan. 2006. Mixing Properties of a General Class of GARCH(1,1) Models Without Moment Assumptions on the Observed Process. *Econometric Theory* 22: 815–834.

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

Herrndorf, N. 1984. A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12: 141–153.

Horváth, L., P. Kokoszka, and R. Zitikis. 2006. Sample and Implied Volatility in GARCH Models. *Journal of Financial Econometrics* 4: 617–635.

Horváth, L., and F. Liese. 2004. L_p Estimators in ARCH Models. *Journal of Statistical Planning and Inference* 119: 277–309.

Hull, J. 2003. *Options, Futures, and Other Derivatives.* New Jersey: Princeton Hall.

Kristensen, D., and O. Linton. 2004. Consistent Standard Errors for Target Variance Approach to GARCH Estimation. *Econometric Theory* 20: 990–993.

Ling, S. 2007. Self-Weighted and Local Quasi-Maximum Likelihood for ARMA-GARCH/IGARCH Models. *Journal of Econometrics* 140: 849–873.

Meddahi, N., and E. Renault. 2004. Temporal Aggregation of Volatility Models. *Journal of Econometrics* 119: 355–379.

Newey, W., and D. McFadden. 1994. Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (eds.) Handbook of Econometrics, Vol. IV. Amsterdam, The Netherlands: North Holland. 2111–2245.

Pastorello, S., V. Patilea, and E. Renault. 2003. Iterative and Recursive Estimation in Structural Nonadaptive Models. *Journal of Business & Economic Statistics* 21: 449–482.

Shephard, N., and K. Sheppard. 2010. Realising the Future: Forecasting With High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 25: 197–231.

Straumann, D. 2005. *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics. Berlin: Springer.

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

Wald, A. 1949. Note on the Consistency of the Maximum Likelihood Estimate. *Annals of Mathematical Statistics* 20: 595–601.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50: 1–26.