Merits and drawbacks of variance targeting in GARCH models

Christian Francq*, Lajos Horvath† and Jean-Michel Zakoïan‡

April 13, 2009

Variance targeting estimation is a technique used to alleviate the numerical difficulties encountered in the quasi-maximum likelihood (QML) estimation 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 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.

Keywords. Consistency and Asymptotic Normality, GARCH, Heteroskedastic Time Series, Quasi Maximum Likelihood Estimation, Value-at-Risk, Variance Targeting Estimator.

1 Introduction

More than two decades after the introduction of ARCH models and their generalization (Engle (1982), Bollerslev (1986)), the properties of GARCH type sequences are well understood and general statistical methods have been established to work with this type of sequences. 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 recent 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 iid process, and ii) no moment condition has to be imposed on the observations to obtain consistency and asymptotic normality.

However, practitioners are often reluctant to directly apply the QMLE to their data. They generally 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*

*Université Lille III, GREMARS-EQUIPPE Universités de Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-Mail: christian.francq@univ-lille3.fr

†University of Utah, Department of Mathematics, 155 South 1400 East, Salt Lake City UT 84112-0090, USA. E-mail: horvath@math.utah.edu

‡GREMARS-EQUIPPE and CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr

(VTE) method. The method is based on a reparamerization 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 QML in a second step.

To our knowledge, the asymptotic properties of the VTE have not been established, and they are the main aim of this paper. 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 iid errors. The magnitude of the accuracy loss, however, cannot be intuited. Moreover, for non Gaussian iid 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, *i.e.* 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 a first step and then plugged in the quasi-likelihood in a second step. Two-step estimators are quite common in econometrics but, in general, these estimators are given in closed form. From a technical point of view, the main difficulty here is that the first step estimator is plugged in a criterion, not directly in a formula giving the second-step estimator. This particularity makes the proof of the asymptotic properties of the VTE non standard.

The paper is organized as follows. Section 2 describes the reparameterization of the standard GARCH(1,1) model and provides the asymptotic properties of the VTE. Section 3 proposes an extension to the GARCH(p,q) model. Section 4 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 stock indices is also proposed. Section 5 concludes and outlines topics for future research. Proofs are relegated to an appendix.

2 Asymptotic distribution of the VTE in the GARCH(1,1) case

Consider the GARCH(1,1) model

$$\begin{cases} \epsilon_t = \sqrt{h_t}\eta_t \\ h_t = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0 h_{t-1}, \quad \forall t \in \mathbb{Z} \end{cases} \tag{2.1}$$

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

$$(\eta_t) \text{ is a sequence of independent and identically distributed (i.i.d) random variables} \tag{2.2}$$

such that

$$E\eta_t^2 = 1, \tag{2.3}$$

and

$$\omega_0 > 0, \qquad \alpha_0 \geq 0, \qquad \beta_0 \geq 0. \tag{2.4}$$

Under the condition

$$\alpha_0 + \beta_0 < 1 \tag{2.5}$$

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

$$\epsilon_t = \sqrt{h_t}\eta_t, \qquad h_t = h_{t-1} + \kappa_0(\gamma_0 - h_{t-1}) + \alpha_0(\epsilon_{t-1}^2 - h_{t-1}), \tag{2.6}$$

which allows us to interpret κ_0 as the speed of mean reversion in variance (see Christoffersen (2008)). 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 reparametrization, constraints (2.4) and (2.5) become

$$\kappa_0, \gamma_0 > 0, \quad \alpha_0 \geq 0, \quad \kappa_0 + \alpha_0 \leq 1. \tag{2.7}$$

Let $(\epsilon_1, \dots, \epsilon_n)$ be a realization of length n of the unique nonanticipative second-order stationary solution (ϵ_t) to model (2.1) which satisfies (2.3) and (2.7). In this framework, VTE involves (i) reparametrizing the model as in (2.6), (ii) estimating 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 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 difference $\hat{\sigma}_n^2 - \gamma_0$ 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 $\Lambda \subset \{(\alpha, \kappa) \mid \alpha \geq 0, \kappa > 0, \alpha + \kappa \leq 1\}$. 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 the 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 quasi-likelihood 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 $t \geq 1$, by

$$\tilde{\sigma}_t^2(\boldsymbol{\vartheta}) = \kappa\gamma + \alpha\epsilon_{t-1}^2 + (1 - \kappa - \alpha)\tilde{\sigma}_{t-1}^2(\boldsymbol{\vartheta}) \tag{2.8}$$

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

$$\hat{\boldsymbol{\lambda}}_n = \arg\max_{\boldsymbol{\lambda} \in \Lambda} L_n(\boldsymbol{\lambda}) = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \tilde{\mathbf{l}}_n(\boldsymbol{\lambda}). \tag{2.9}$$

where

$$\tilde{\mathbf{l}}_n(\boldsymbol{\lambda}) = n^{-1}\sum_{t=1}^n \ell_{t,n}, \quad \text{and} \quad \ell_{t,n} := \ell_{t,n}(\boldsymbol{\lambda}) = \frac{\epsilon_t^2}{\sigma_{t,n}^2} + \log\sigma_{t,n}^2. \tag{2.10}$$

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 $0 \leq 1 - \kappa - \alpha < 1$, and one can define the strictly stationary and ergodic process

$$\sigma_t^2(\boldsymbol{\vartheta}) = \kappa\gamma + \alpha\epsilon_{t-1}^2 + (1 - \kappa - \alpha)\sigma_{t-1}^2(\boldsymbol{\vartheta}) = \sum_{i=0}^\infty (1 - \kappa - \alpha)^i(\kappa\gamma + \alpha\epsilon_{t-i-1}^2). \tag{2.11}$$

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 non-degenerate 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 2.1 *Under assumptions* **A1**-**A2**, $\alpha_0 \neq 0$, *and (2.2)–(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**-**A4**, *we have*

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

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 non-singular with

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

and

$$\boldsymbol{J} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}'}\right)_{2\times 2}, \qquad \boldsymbol{K} = E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \gamma}\right)_{2\times 1}. \tag{2.12}$$

This result 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 $\hat{\omega}_n = \hat{\lambda}_{2n}\hat{\sigma}_n^2$, $\quad \hat{\alpha}_n = \hat{\lambda}_{1n}$, $\quad \hat{\beta}_n = 1 - \hat{\lambda}_{1n} - \hat{\lambda}_{2n}$. Theorem 2.1 yields the following result.

Corollary 2.1 *Under the assumptions of Theorem 2.1, the VTE of* $\boldsymbol{\theta}_0$ *satisfies*

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

It is important to note that the asymptotic normality of the VTE requires the existence of $E(\epsilon_t^4)$, whereas the strict stationarity is sufficient for the asymptotic normality of the QMLE (see Berkes, Horváth and Kokoszka (2003) and 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 2.2 *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 of* $\boldsymbol{\vartheta}_0$ *satisfy*

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

where

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

Note that $a>0$ because $\det\boldsymbol{\Sigma}^*=a\det\boldsymbol{J}^{-1}$. It will be shown that $b-a\geq 0$ (in a more general setting, see Proposition 3.1 below), which implies that the VTE cannot be asymptotically more accurate than the QMLE. Corollary 2.2 shows that, as expected, the VTE becomes much less accurate than the QMLE when $\alpha_0^2\left(E\eta_t^4-1\right)+(1-\kappa_0)^2$ approaches 1. More interestingly, the relative loss of efficiency of the VTE is the same for all 3 parameters γ_0, α_0 and κ_0. In general the asymptotic variances of the GARCH coefficients which are estimated by the two methods do not coincide. This point will be illustrated in Section 4.

3 Extension to the general GARCH(p,q) case

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

$$\begin{cases}\epsilon_t=\sqrt{h_t}\eta_t\\ h_t=\omega_0+\sum_{i=1}^q\alpha_{0i}\epsilon_{t-i}^2+\sum_{j=1}^p\beta_{0j}h_{t-j},\quad\forall t\in\mathbb{Z}\end{cases}\tag{3.1}$$

where (η_t) satisfies (2.2)-(2.3) and where the coefficients satisfy:

$$\omega_0>0,\qquad \alpha_{0i}\geq 0\ \ \forall i\in\{1,\dots,q\},\qquad \beta_{0j}\geq 0\ \ \forall j\in\{1,\dots,p\}.\tag{3.2}$$

Under the condition

$$\sum_{i=1}^q\alpha_{0i}+\sum_{j=1}^p\beta_{0j}<1.\tag{3.3}$$

the observations have finite variance $\gamma_0=\omega_0\left\{1-\sum_{i=1}^q\alpha_{0i}-\sum_{j=1}^p\beta_{0j}\right\}^{-1}$. 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 included in the simplex $\left\{\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_{p+q}):\ \ \lambda_i\geq 0\,\forall i\in\{1,\dots,p+q\},\ \ \sum_{i=1}^{p+q}\lambda_i<1\right\}$.

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

$$\hat{\boldsymbol{\lambda}}_n=\arg\min_{\boldsymbol{\lambda}\in\Lambda}\tilde{\mathbf{l}}_n(\boldsymbol{\lambda}),\qquad \tilde{\mathbf{l}}_n(\boldsymbol{\lambda})=n^{-1}\sum_{t=1}^n\tilde\ell_t(\hat\sigma_n^2,\boldsymbol{\lambda}),\qquad \tilde\ell_t(\boldsymbol{\vartheta})=\frac{\epsilon_t^2}{\tilde\sigma_t^2(\boldsymbol{\vartheta})}+\log\tilde\sigma_t^2(\boldsymbol{\vartheta}),$$

and the $\tilde\sigma_t^2(\boldsymbol{\vartheta})$'s are defined recursively, for $t\geq 1$, 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 $\epsilon_0,\dots,\epsilon_{1-q}$ and $\tilde\sigma_0^2(\boldsymbol{\vartheta}),\dots,\tilde\sigma_{1-p}^2(\boldsymbol{\vartheta})$. Define $\mathcal{A}_{\boldsymbol{\vartheta}}(z)=\sum_{i=1}^q\lambda_iz^i$ 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 2.1 and Corollary 2.2.

Theorem 3.1 *Under assumptions* **A1** *with (3.2) and (3.3),* **A2** *with (2.2) and (2.3), and* **A5***, the VTE of the GARCH(p,q) model (3.1) is strongly consistent. Under the additional assumptions $E\epsilon_t^4 < \infty$ and* **A4***, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \xrightarrow{d} \mathcal{N}\left\{0, (E\eta_0^4 - 1)\boldsymbol{\Sigma}\right\}, \tag{3.4}$$

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 non-singular 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

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

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

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n^* - \boldsymbol{\vartheta}_0\right) \xrightarrow{d} \mathcal{N}\left\{0, (E\eta_0^4 - 1)\boldsymbol{\Sigma}^*\right\}, \tag{3.5}$$

where

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - (c - d)\boldsymbol{C}\boldsymbol{C}', \tag{3.6}$$

with

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

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 3.1 *Under the assumptions of Theorem 3.1, 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*

$$\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^* \quad \textit{is positive semidefinite, but not positive definite.}$$

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

Corollary 3.1 *Let the assumptions of Theorem 3.1 be satisfied, and let $\boldsymbol{\phi}$ be a mapping from $\mathbb{R}^{p+q+1}$ to $\mathbb{R}$, which is continuously differentiable in a neighborhood of $\boldsymbol{\vartheta}_0$. If*

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

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

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

where

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

4 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) 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 4.3 studies the estimation of the parameters of a set of typical financial time series using both methods. Subsection 4.4 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.

4.1 Asymptotic variances of the QMLE and VTE for ARCH(1) models

For an ARCH(1) model, the asymptotic variances $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$ of the VTE and QMLE are given by Theorem 3.1 with

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

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

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

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

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

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

The results presented in Table 1 are obtained from simulations of the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$ above, with expectations replaced by empirical means. More precisely, the table displays the mean of 1,000 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 3.1 that there exist transformations of $\boldsymbol{\vartheta}_0$ which are estimated by VTE and QMLE with the same asymptotic accuracy.

4.2 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 1,000 independent simulated trajectories of length $n = 500$, $n = 5,000$ and $n = 10,000$ of three ARCH(1) models. 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, 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 when $n = 500$. This finite-sample result counterbalances the result of Proposition 3.1 showing that the VTE can not be asymptotically more efficient than the QMLE. 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 when n is large (*i.e.* $n = 5,000$ and $n = 10,000$) and $\alpha = 0.55$ or $\alpha = 0.9$.

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

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	$\begin{pmatrix} 3.5 & -1.4 \\ -1.4 & 1.7 \end{pmatrix}$	$\begin{pmatrix} 4.2 & -1.8 \\ -1.8 & 2.9 \end{pmatrix}$	$\begin{pmatrix} 4.9 & -1.9 \\ -1.9 & 6.1 \end{pmatrix}$	$\begin{pmatrix} 5.1 & -2.1 \\ -2.1 & 9.3 \end{pmatrix}$	∞

4.3 Comparison of the QMLE and VTE on daily stock market returns

In this section, we consider daily returns of 11 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. For each series, a GARCH(1,1) model was estimated, by QMLE and by VTE. Table 4 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 3.1 to obtain the asymptotic normality are likely to be unsatisfied. Nevertheless, it is seen from Table 4 that the parameters estimated by VTE are always very close to those estimated by QMLE.

As expected, the VTE is more successful than the QMLE in terms of amount of computation time. Table 5 compares the computation time of the QMLE and VTE for estimating the models of the 11 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 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

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

Table 3: 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
			$n = 500$						
ω	1.0	QMLE	0.000	0.092	0.755	0.934	0.997	1.059	1.343
		VTE	-0.001	0.092	0.759	0.932	0.995	1.062	1.335
α	0.3	QMLE	-0.004	0.076	0.085	0.242	0.299	0.348	0.582
		VTE	-0.006	0.075	0.084	0.243	0.295	0.346	0.552
ω	1.0	QMLE	0.013	0.102	0.712	0.944	1.011	1.079	1.332
		VTE	0.012	0.102	0.719	0.943	1.010	1.078	1.327
α	0.55	QMLE	-0.012	0.092	0.233	0.474	0.536	0.601	0.789
		VTE	-0.026	0.088	0.236	0.463	0.524	0.579	0.895
ω	1.0	QMLE	0.012	0.114	0.623	0.928	1.010	1.087	1.435
		VTE	0.036	0.111	0.637	0.955	1.032	1.108	1.428
α	0.9	QMLE	-0.012	0.110	0.491	0.814	0.884	0.961	1.283
		VTE	-0.103	0.089	0.505	0.740	0.800	0.860	0.998
			$n = 5000$						
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
			$n = 10000$						
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

RMSE is the Root Mean Square Error, Q_i, $i = 1, 3$, denote the quartiles.

Table 4: Comparison of the QMLE and VTE of GARCH(1,1) models for 11 daily stock market returns. The estimated standard deviation are displayed into brackets. 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	ρ_4
CAC	QMLE	0.033 (0.009)	0.090 (0.014)	0.893 (0.015)	1.0067
	VTE	0.033 (0.009)	0.090 (0.014)	0.893 (0.015)	
DAX	QMLE	0.037 (0.014)	0.093 (0.023)	0.888 (0.024)	1.0622
	VTE	0.036 (0.013)	0.095 (0.022)	0.888 (0.024)	
DJA	QMLE	0.019 (0.005)	0.088 (0.014)	0.894 (0.014)	0.9981
	VTE	0.019 (0.005)	0.089 (0.012)	0.894 (0.007)	
DJI	QMLE	0.017 (0.004)	0.085 (0.013)	0.901 (0.013)	1.002
	VTE	0.016 (0.004)	0.085 (0.012)	0.901 (0.013)	
DJT	QMLE	0.040 (0.013)	0.089 (0.016)	0.894 (0.018)	1.0183
	VTE	0.042 (0.013)	0.086 (0.016)	0.894 (0.018)	
DJU	QMLE	0.021 (0.005)	0.118 (0.016)	0.865 (0.014)	1.0152
	VTE	0.021 (0.004)	0.119 (0.013)	0.865 (0.013)	
FTSE	QMLE	0.013 (0.004)	0.091 (0.014)	0.899 (0.014)	1.0228
	VTE	0.013 (0.004)	0.090 (0.013)	0.899 (0.014)	
Nasdaq	QMLE	0.025 (0.006)	0.072 (0.009)	0.922 (0.009)	1.0021
	VTE	0.025 (0.006)	0.072 (0.009)	0.922 (0.009)	
Nikkei	QMLE	0.053 (0.012)	0.100 (0.013)	0.880 (0.014)	0.9985
	VTE	0.054 (0.012)	0.098 (0.013)	0.880 (0.015)	
SMI	QMLE	0.049 (0.014)	0.127 (0.028)	0.835 (0.029)	1.0672
	VTE	0.048 (0.014)	0.133 (0.025)	0.834 (0.029)	
SP500	QMLE	0.014 (0.004)	0.084 (0.012)	0.905 (0.012)	1.0072
	VTE	0.014 (0.003)	0.084 (0.011)	0.905 (0.012)	

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 2 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 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; iii) the VTE provides good initial values for the QMLE and may avoid that this estimator be trapped in local optima.

4.4 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 convergence to hold true, it suffices that the observed process be stationary and

Table 5: Comparison of the computation time of the QMLE and VTE (in seconds of CPU time), for estimating the models of the 11 indices of Table 4. The method VTE+QMLE consists in using the VTE as initial values for the QMLE. Design 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

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.

4.4.1 Prediction over long horizons with models estimated by VTE

We will study the asymptotic behavior 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 (2.1). 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, $\hat{F}_\eta(\underline{\alpha})$ 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

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

when the GARCH model is estimated by QMLE, and by

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

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 pseudo true value $\tilde{\boldsymbol{\vartheta}}^*$ is defined by

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

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 pseudo-true 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

$$\left[\sqrt{\tilde{\gamma}^*}F_\eta^{-1}(\underline{\alpha}/2),\ \sqrt{\tilde{\gamma}^*}F_\eta^{-1}(1-\underline{\alpha}/2)\right], \tag{4.1}$$

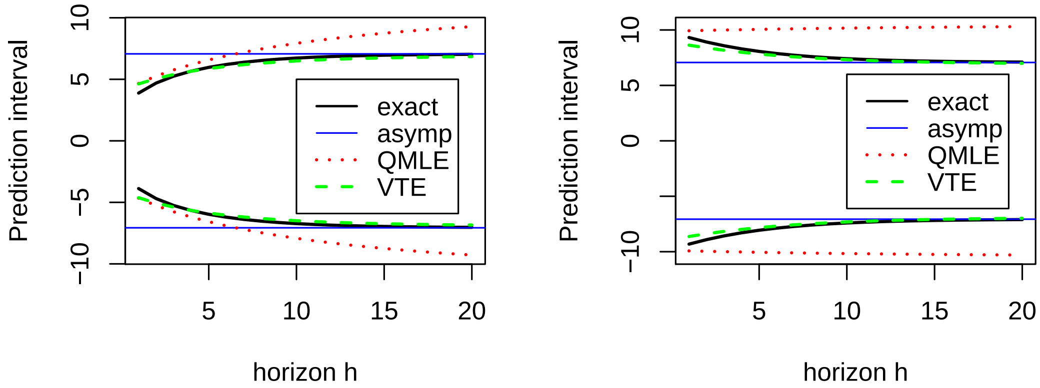

Figure 1: Asymptotic prediction intervals based on the true model (between the full lines), for a GARCH(1,1) estimated by QMLE (dotted lines) and a GARCH(1,1) estimated by VTE (dashed lines). The horizontal full lines are the bounds of the large-horizon prediction intervals (4.1). The DGP is the Markov-switching model (4.3). 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.

with the QMLE, and equivalent to

$$\left[\sqrt{E\epsilon_1^2}F_\eta^{-1}(\underline{\alpha}/2),\ \sqrt{E\epsilon_1^2}F_\eta^{-1}(1-\underline{\alpha}/2)\right], \tag{4.2}$$

with the VTE. Note that the long horizon prediction intervals (4.1) 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 (4.2) tends to the nominal probability $1-\underline{\alpha}$ as the horizon h increases.

Example 4.1 To give an elementary illustration, consider the Markov-switching model

$$\epsilon_t = \omega(\Delta_t)\eta_t, \tag{4.3}$$

where η_t is an iid noise, (Δ_t) is a stationary irreducible and aperiodic Markov chain, independent of (η_t), with state-space $\{1,\dots,d\}$. For Figure 1, we took $\eta_t \sim \mathcal{N}(0,1)$, $d=2$ 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 (4.3). In our experiments, we fitted a GARCH(1,1) model by the two methods, on a simulation of size 1,000 of Model (4.3). 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 when the horizon h is large (indicating a pseud-true value $\tilde{\gamma}^*$ larger than $E\epsilon_1^2$, for this particular model).

4.4.2 Estimating long horizon Value-at-Risk

Value-at-Risk is one of the most important market-risk measurement tool (see *e.g.* the web site `http://www.gloriamundi.org/` which 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:

$$\text{VaR}_{t,h}(\underline{\alpha}) = \inf\left\{x \in \mathbb{R} \mid P\left(L_{t,t+h} \leq x \mid V_u, u \leq t\right) \geq 1-\underline{\alpha}\right\}.$$

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

$$\text{VaR}_{t,h}(\underline{\alpha}) = \left[1 - \exp\left\{q_{t,h}(\underline{\alpha})\right\}\right] V_t, \tag{4.4}$$

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 $\text{VaR}_{t,h}(\underline{\alpha})$ for large h, under some α-mixing condition on the process (ϵ_t).

Lemma 4.1 *Assume that (ϵ_t) is a strictly stationary process such that $E\epsilon_t = 0$, $\sum_{h=1}^{\infty}\left\{\alpha_\epsilon(h)\right\}^{\nu/(2+\nu)} < \infty$ and $E|\epsilon_t|^{2+\nu} < \infty$ for some $\nu > 0$. Let $Var(\epsilon_t) = \overline{\omega}^2$. We have*

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

Remark 4.1 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) and Francq and Zakoïan (2006). This condition is also satisfied for the Markov-switching process (4.3). 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 $\alpha_\Delta(k) \leq K\rho^k$ 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 $\alpha_\epsilon(k) \leq K\rho^k$.

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

$$\text{VaR}_{t,1}(\underline{\alpha}) = \left[1 - \exp\left\{\sigma_t(\boldsymbol{\theta}_0)F_\eta^{-1}(1-\underline{\alpha})\right\}\right] V_t,$$

in view of (4.4). Hence, if $\hat{\boldsymbol{\theta}}_n$ is an estimator of $\boldsymbol{\theta}_0$, an obvious estimator of $\text{VaR}_{t,1}(\underline{\alpha})$ is obtained by plugging. 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 $\widehat{\overline{\omega}}$ of $\overline{\omega}$, one can take

$$\widehat{\text{VaR}}_{t,h}(\underline{\alpha}) = \left[1 - \exp\left\{\sqrt{h}\,\Phi^{-1}(\underline{\alpha})\,\widehat{\overline{\omega}}\right\}\right] V_t. \tag{4.5}$$

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

Example 4.2 (Example 4.1 continued) We shall consider VaR at horizons $h = 1$ and $h = 10$ obtained from estimated GARCH(1,1) models, when the observations are drawn from the Markov-switching process (4.3). For the sake of comparison, we shall also consider the theoretical VaR's of the true model, obtained at horizon 1 as the solution of

$$1 - \underline{\alpha} = \sum_{j=1}^{d} F_\eta\left\{\frac{\text{VaR}_{t,1}(\underline{\alpha})}{\sigma(j)}\right\} P(\Delta_{t+1} = j \mid \epsilon_u, u \leq t)$$

and approximated for large h by

$$\text{VaR}_{t,h}(\underline{\alpha}) = \left[1 - \exp\left\{\sqrt{h}\,\Phi^{-1}(\underline{\alpha})\,\overline{\omega}\right\}\right] V_t, \qquad \overline{\omega} = \sum_{j=1}^{d} \omega(j) P(\Delta_t = j).$$

Figure 2 shows samples paths of $L_{t,t+h}$, for $h = 1$ and $h = 10$, 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 $h = 10$. 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 $h = 10$. Standard evaluation of the performance of VaR estimation methods relies on comparing the percentages of exceptions (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 6 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 2: the VaR at the horizon $h = 10$ computed with the misspecified GARCH(1,1) is more satisfactory in terms of backtesting when the model is estimated by VTE than by QMLE.

Table 6: Backtesting comparison of the VaR estimations given by the true HMM model (4.3), the GARCH(1,1) model estimated by QMLE, and the GARCH(1,1) estimated by VTE on $n = 1,000$ observations. The comparison is made out-of-sample, on a simulation of size $N = 50,000$ of the profit and loss (P&L) function, for the two horizons $h = 1$ and $h = 10$ and the three levels $\underline{\alpha} = 1\%$, $\underline{\alpha} = 5\%$ and $\underline{\alpha} = 10\%$.

$\underline{\alpha} = 1\%$						
		$h = 1$			$h = 10$	
	HMM	QMLE	VTE	HMM	QMLE	VTE
Relative VaR average (in %)	4.64	4.01	3.78	12.42	17.77	12.17
Exceptions (in %)	1.01	2.28	2.57	1.52	0.13	1.67
$\underline{\alpha} = 5\%$						
		$h = 1$			$h = 10$	
	HMM	QMLE	VTE	HMM	QMLE	VTE
Relative VaR average (in %)	2.65	2.86	2.69	8.95	12.92	8.77
Exceptions (in %)	4.82	4.84	5.49	5.17	1.25	5.47
$\underline{\alpha} = 10\%$						
		$h = 1$			$h = 10$	
	HMM	QMLE	VTE	HMM	QMLE	VTE
Relative VaR average (in %)	1.87	2.23	2.10	7.05	10.22	6.90
Exceptions (in %)	9.99	7.32	8.17	9.08	3.39	9.42

5 Conclusion

VTE is a two-step estimation method which reduces the computational complexity of the optimization procedure and guarantees that the implied variance 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

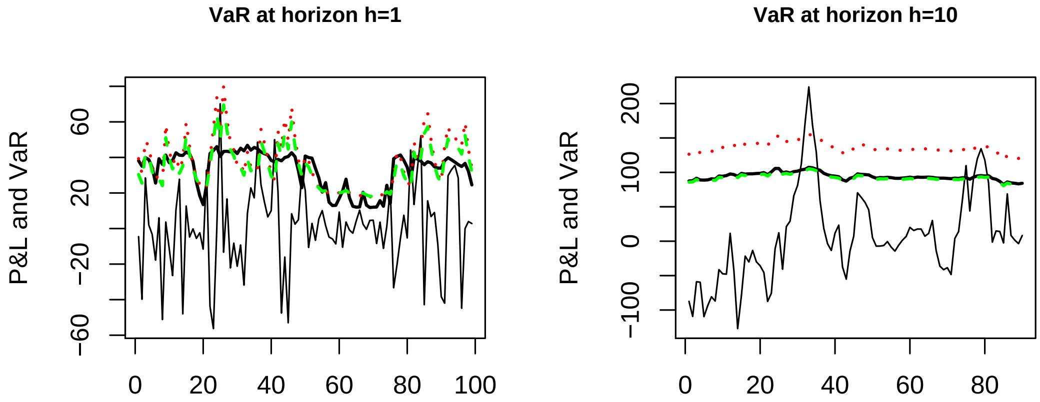

Figure 2: Sample paths of the P&L process generated by the Markov-switching model (4.3) and VaR at the confidence level 5%. The full line corresponds to the exact VaR, the doted (resp. dashed) line to the asymptotic approximation obtained from Lemma 4.1 applied to a GARCH model estimated by QMLE (resp. VTE).

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.

Appendix: proofs

Let

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

For $t \geq 1$ 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 2.1

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 which 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.

$i)$ $\lim_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\Lambda} |\mathbf{l}_n(\boldsymbol{\lambda}) - \tilde{\mathbf{l}}_n(\boldsymbol{\lambda})| = 0, \quad a.s.$

$ii)$ if $\sigma_t^2(\gamma_0, \boldsymbol{\lambda}) = \sigma_t^2(\gamma_0, \boldsymbol{\lambda}_0) \quad a.s.,$ then $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0,$

$iii)$ if $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0,$ then $E\ell_t(\gamma_0, \boldsymbol{\lambda}) > E\ell_t(\gamma_0, \boldsymbol{\lambda}_0),$

$iv)$ any $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$ has a neighborhood $V(\boldsymbol{\lambda})$ such that $\liminf_{n\to\infty} \inf_{\boldsymbol{\lambda}^*\in V(\boldsymbol{\lambda})} \tilde{\mathbf{l}}_n(\boldsymbol{\lambda}^*) > E\ell_1(\gamma_0, \boldsymbol{\lambda}_0) \quad a.s.$

We first show *i)*. 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

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

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

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

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

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

and the arguments used in the proof in FZ.

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

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

Suppose that $\sigma_t^2(\boldsymbol{\vartheta}) = \sigma_t^2(\boldsymbol{\vartheta}_0)$ *a.s.* Then, in view of (A.1),

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

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

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

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

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

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

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 *ii)*.

Let us show *iv)*. Let $V_k(\boldsymbol{\lambda})$ be the open ball with center $\boldsymbol{\lambda}$ and radius $1/k$. Using successively *i)*, the ergodic process, the monotone convergence theorem and *iii)*, we obtain almost surely

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

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

A.2 Proof of asymptotic normality in Theorem 2.1

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 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 (2.10), $\ell_{t,n}(\boldsymbol{\lambda}) = \tilde{\ell}_t(\hat{\sigma}_n^2, \boldsymbol{\lambda})$, we have

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

where the $\boldsymbol{\vartheta}_i^*$ are between $\hat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$,

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

We will show that

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

ii) $\boldsymbol{A} := E\left(\frac{1}{\sigma_t^4(\boldsymbol{\vartheta}_0)}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right)$ is non-singular and $\operatorname{Var}\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\} = \left\{E\eta_0^4 - 1\right\}\boldsymbol{A},$

iii) there exists a neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that, for all $i,j,k \in \{1,\dots,p+q+1\}$,

$$E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{\partial^3\ell_t(\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j\partial\vartheta_k}\right| < \infty,$$

iv) $\left\|n^{-1/2}\sum_{t=1}^n\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} - \frac{\partial\tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\}\right\| \to 0$ and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|n^{-1}\sum_{t=1}^n\left\{\frac{\partial^2\ell_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \frac{\partial^2\tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\}\right\| \to 0$

in probability when $n \to \infty$,

v) $n^{-1}\sum_{t=1}^n \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ell_t(\boldsymbol{\vartheta}_k^*) \to \boldsymbol{A}(i,j)$ *a.s.*

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

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

$$\frac{\partial \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}(\boldsymbol{\vartheta}), \tag{A.3}$$

$$\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}) + \left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}). \tag{A.4}$$

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 *i)* it will therefore be sufficient to show that

$$E\left\|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0)\right\| < \infty, \quad E\left\|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0)\right\| < \infty, \quad E\left\|\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0)\right\| < \infty. \tag{A.5}$$

The following expansions hold

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

Recall that **A3** implies $E\epsilon_t^4 < \infty$. Moreover we have $\sigma_t^{-2}(\boldsymbol{\vartheta}_0) \leq \kappa_0^{-1}\gamma_0^{-1} < \infty$. This allows us to prove the first and third inequalities in (A.5). The second inequality is proved by exactly the same arguments, and *i)* is proved. Note that we made use of the moment assumption $E\epsilon_t^4 < \infty$ to facilitate the proof of (A.5). This moment assumption is actually unnecessary. Indeed, we will show, without this assumption, that for any integer d, there exists a neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \vartheta_i}\right|^d < \infty, \qquad E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \vartheta_i\partial\vartheta_j}\right|^d < \infty \qquad E\sup_{\theta\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial^3 \sigma_t^2}{\partial \vartheta_i\partial\vartheta_j\partial\vartheta_k}\right|^d < \infty. \tag{A.6}$$

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 $x/(1+x) \leq x^s$ for all $x \geq 0$ and all $s \in (0,1]$, for all $\boldsymbol{\vartheta} \in \mathcal{V}(\boldsymbol{\vartheta}_0)$ we have

$$\begin{aligned}
\left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \gamma}(\boldsymbol{\vartheta})\right| &\leq K, \\
\left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \alpha}(\boldsymbol{\vartheta})\right| &\leq K + \sum_{\ell=0}^{\infty}\frac{\beta^\ell\epsilon_{t-\ell-1}^2}{K + \alpha\beta^\ell\epsilon_{t-\ell-1}^2} + \alpha\sum_{\ell=1}^{\infty}\frac{\ell\beta^{\ell-1}\epsilon_{t-\ell-1}^2}{K + \alpha\beta^\ell\epsilon_{t-\ell-1}^2} \\
&\leq K + K\sum_{\ell=0}^{\infty}\beta^{\ell s}\epsilon_{t-\ell-1}^{2s} + K\sum_{\ell=0}^{\infty}\ell\beta^{\ell s}\epsilon_{t-\ell-1}^{2s}.
\end{aligned}$$

Similarly

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

Under 2.5 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 (A.6) comes from the Hölder inequality. The other results of (A.6) are obtained with the same arguments.

Now we prove *ii)*. 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

$$\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}} = \frac{\partial \kappa\gamma}{\partial \boldsymbol{\vartheta}} + \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}}\epsilon_{t-1}^2 + \frac{\partial \beta}{\partial \boldsymbol{\vartheta}}\sigma_{t-1}^2 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \boldsymbol{\vartheta}}, \tag{A.7}$$

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

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

We thus have

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

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 non-singularity of $\boldsymbol{A}$ is proved.

Let us prove *iii)*. Differentiating (A.4), we obtain

$$\begin{aligned}\frac{\partial^3\ell_t(\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j\partial\vartheta_k} &= \left\{1-\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\vartheta_i\partial\vartheta_j\partial\vartheta_k}\right\}(\boldsymbol{\vartheta}) + \left\{2\frac{\epsilon_t^2}{\sigma_t^2}-1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\vartheta_j\partial\vartheta_k}\right\}(\boldsymbol{\vartheta}) \\ &+ \left\{2\frac{\epsilon_t^2}{\sigma_t^2}-1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_j}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\vartheta_i\partial\vartheta_k}\right\}(\boldsymbol{\vartheta}) + \left\{2\frac{\epsilon_t^2}{\sigma_t^2}-1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_k}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\vartheta_i\partial\vartheta_j}\right\}(\boldsymbol{\vartheta}) \\ &+ \left\{2-6\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_j}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_k}\right\}(\boldsymbol{\vartheta}).\end{aligned} \quad \text{(A.8)}$$

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 (A.6) with $d=1,2,3$, the Hölder inequality entails *iii)*.

We now turn to the proof of *iv)*. In view of (2.8) and (2.11), we have

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

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

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

and

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

In view of (A.3), we then obtain

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|n^{-1}\sum_{t=1}^n\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}-\frac{\partial\tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\}\right\|\le Kn^{-1}\sum_{t=1}^n\rho^t\Upsilon_t,\quad a.s., \quad \text{(A.9)}$$

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 (A.6), the Cauchy-Schwarz inequality shows that $E\Upsilon_t<\infty$. By the Borel-Cantelli lemma, it follows that $\rho^t\Upsilon_t\to 0$ $a.s.$ and thus the right-hand side of (A.9) converges to 0 $a.s.$ Using the same arguments, and replacing first derivatives with second derivatives, *iv)* is shown.

Similarly to the proof of (4.36) in FZ, *v)* 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 *iii)*.

The proof of *vi)* relies on a Central Limit Theorem for martingale differences. From Horváth, Kokoszka and Zitikis (2006, Proof of Theorem 1) (see (A.15) below) we have the representation

$$\hat{\sigma}_n^2 = \gamma_0 + \frac{1-\beta_0}{\kappa_0}\frac{1}{n}\sum_{t=1}^n h_t(\eta_t^2-1)+o_P(n^{-1/2}). \quad \text{(A.10)}$$

Moreover, in view of (A.3)

$$n^{-1/2}\sum_{t=1}^n\frac{\partial}{\partial\boldsymbol{\lambda}}\ell_t(\boldsymbol{\vartheta}_0) = n^{-1/2}\sum_{t=1}^n\frac{1-\eta_t^2}{h_t}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\lambda}}(\boldsymbol{\vartheta}_0). \quad \text{(A.11)}$$

We then have

$$\boldsymbol{X}_n = n^{-1/2}\sum_{t=1}^{n}(1-\eta_t^2)\boldsymbol{Z}_t + o_P(1), \qquad \boldsymbol{Z}_t = \begin{pmatrix} -(1-\beta_0)\kappa_0^{-1}h_t \\ h_t^{-1}\partial\sigma_t^2(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\lambda} \end{pmatrix}. \tag{A.12}$$

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

$$E\left\{\frac{\partial\sigma_t^2}{\partial\boldsymbol{\lambda}}(\boldsymbol{\vartheta}_0)\right\} = 0. \tag{A.13}$$

It follows that

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

Notice that b is a positive real number and that the matrix $\boldsymbol{J}$ in the right-lower block of $\boldsymbol{A}$ is non-singular, in view of the non-singularity of $\boldsymbol{A}$. By assumptions **A2** and **A3**, we get $0 < E\eta_0^4 - 1 < \infty$, and thus the matrix $\mathrm{Var}\left\{(1-\eta_t^2)\boldsymbol{Z}_t\right\}$ is nondegenerate. Hence for any $\boldsymbol{\lambda} \in \mathbb{R}^3$, the sequence $\left\{(1-\eta_t^2)\boldsymbol{\lambda}'\boldsymbol{Z}_t, \mathcal{F}_t\right\}_t$ is a square integrable stationary martingale difference. By (A.12), the central limit theorem of Billingsley (1961) and the Wold-Cramer device we obtain the asymptotic normality of $\boldsymbol{X}_n$, which proves *v)*.

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

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

It follows that, using *iv)*,

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

Thus, by *v)*, *vi)* and Slutsky's lemma, $\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)$ 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 2.1 is complete.

A.3 Proof of Corollary 2.1

This corollary of Theorem 2.1 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 $\boldsymbol{\phi}$ which 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 Corollary 2.2

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

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

whereas

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

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 3.1

The consistency can be obtained as in FZ, by a direct extension of Theorem 2.1. Let us concentrate on the asymptotic normality. Similarly to (A.2), we have

$$\begin{aligned} 0_{p+q} &= n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\hat{\boldsymbol{\vartheta}}_n) \\ &= n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \lambda_i \partial \vartheta_j} \tilde{\ell}_t(\boldsymbol{\vartheta}_i^*) \right)_{(p+q)\times(p+q+1)} \sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \\ &= n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{J}_n \sqrt{n}\left(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0\right) + \boldsymbol{K}_n \sqrt{n}\left(\hat{\sigma}_n^2 - \gamma_0\right), \end{aligned} \tag{A.14}$$

where the $\boldsymbol{\vartheta}_i^*$ are between $\hat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$,

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

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 (A.15) of Horváth, Kokoszka and Zitikis (2006)

$$\hat{\sigma}_n^2 = \gamma_0 + \frac{1-\sum_{j=1}^p \beta_{0j}}{1-\sum_{j=1}^q \alpha_{0i} - \sum_{j=1}^p \beta_{0j}} \frac{1}{n}\sum_{t=1}^n h_t(\eta_t^2 - 1) + o_P(n^{-1/2}). \tag{A.15}$$

Using (A.14), (A.15), and direct extensions of *iv)* and *vi)*, we obtain

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = \begin{pmatrix} 1 & 0 \\ -\boldsymbol{J}_n^{-1}\boldsymbol{K}_n & -\boldsymbol{J}_n^{-1} \end{pmatrix} \boldsymbol{X}_n + o_P(1), \tag{A.16}$$

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 (A.13) still holds. The conclusion follows.

A.6 Proof of Proposition 3.1

Let

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

Using (A.13), we observe that

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

We thus have

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

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

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

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

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

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 3.1

The first convergence in distribution is a direct consequence of (3.4) and of the delta method. In view of (3.5), (3.6) and the delta method we have

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

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 4.1

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

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

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

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

and the Borel-Cantelli lemma entails that

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \le x \mid I_t \right\} - P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \le x \right\} \right| = 0 \quad a.s. \tag{A.17}$$

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) and Dehling and Philipp (1982)). By the law of the iterated logarithm, we have

$$\frac{1}{\sqrt{h}} \sum_{i=1}^{hu_h} \epsilon_{t+i} = \frac{\sqrt{u_h}}{\sqrt{hu_h}} \sum_{i=1}^{hu_h} \epsilon_{t+i} \to 0 \quad a.s. \tag{A.18}$$

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

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=1}^{h} \epsilon_{t+i} \le x \mid I_t \right\} - P\left\{ \frac{1}{\sqrt{h}} \sum_{i=hu_h+1}^{h} \epsilon_{t+i} \le x \mid I_t \right\} \right| = 0 \quad a.s., \tag{A.19}$$

The central limit theorem entails $h^{-1/2} \sum_{i=1}^{h} \epsilon_{t+i} \xrightarrow{d} \mathcal{N}\left(0, \varpi^2\right)$ as $h \to \infty$. In view of (A.17)-(A.19) we then obtain

$$\lim_{h\to\infty} \left| P\left\{ \frac{1}{\sqrt{h}} \sum_{i=1}^{h} \epsilon_{t+i} \le x \mid I_t \right\} - \Phi(x/\varpi) \right| = 0 \quad a.s., \tag{A.20}$$

and the conclusion follows.

References

Bauwens, L. and J. Rombouts (2007) Bayesian clustering of many GARCH models. *Econometric Reviews* 26, 365–386.

Berkes, I., Horváth, L. 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. (2008) Value-at-Risk Models. *Handbook of Financial Time Series, T.G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (eds.)*, 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* (Univ. California, Berkeley, Calif., 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.

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

Horváth, L., Kokoszka, P. 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.

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

Straumann, D. (2005) *Estimation in conditionally heteroscedastic time series models.* Lecture Notes in Statistics, Springer, Berlin.

van der Vaart, A.W. (1998) *Asymptotic statistics.* Cambridge University Press, United Kingdom.

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.