Stable limit theory for the variance targeting estimator

Igor Vaynman and Brendan K. Beare*

Department of Economics, University of California, San Diego

November 1, 2014

Published as ch. 18 in Advances in Econometrics vol. 33: Essays in Honor of Peter C. B. Phillips, Emerald Group Publishing, 2014. The final published version may be found here: www.emeraldinsight.com/series/aeco. Permission has been granted for this accepted version to appear here: econweb.ucsd.edu/~bbeare/pdfs/stablevte.pdf. This article is © Emerald Group Publishing and Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission of Emerald Group Publishing Limited.

Abstract

The variance targeting estimator (VTE) for generalized autoregressive conditionally heteroskedastic (GARCH) processes has been proposed as a computationally simpler and misspecification-robust alternative to the quasi-maximum likelihood estimator (QMLE). In this paper we investigate the asymptotic behavior of the VTE when the stationary distribution of the GARCH process has infinite fourth moment. Existing studies of historical asset returns indicate that this may be a case of empirical relevance. Under suitable technical conditions, we establish a stable limit theory for the VTE, with the rate of convergence determined by the tails of the stationary distribution. This rate is slower than that achieved by the QMLE. The limit distribution of the VTE is nondegenerate but singular. We investigate the use of subsampling techniques for inference, but find that finite sample performance is poor in empirically relevant scenarios.

*We thank the anonymous referee and seminar participants at Southern Methodist University for helpful comments.
1 Introduction

It is now widely recognized that the distributions of many economic and financial variables are heavy tailed. The relevance of heavy tailed laws for the study of price series was emphasized in early contributions by Mandelbrot (1963) and Fama (1963). Major theoretical advances in the study of heavy tailed phenomena were made over the following decades in various areas of applied mathematics, but it was perhaps not until a series of papers by P. C. B. Phillips and M. Loretan in the 1990s that these advances were brought into the mainstream of the econometric literature. Phillips (1990) studied the impact of infinite variance innovations on his now well-known nonparametric correction for serial dependence (Phillips, 1987) when testing for a unit root. Phillips and Loretan (1991) considered the behavior of the Durbin-Watson statistic in regression models with infinite variance residuals, while Phillips and Loretan (1995) considered the behavior of various diagnostic tests for covariance stationarity in the presence of heavy tails. Loretan and Phillips (1994) estimated the tail decay rates for various stock market and exchange rate return series. Their findings were broadly consistent with the return series having finite second moment but infinite fourth moment. In subsequent years there has been an enormous effort by economists and econometricians to better understand the impact of heavy tails on economic decision making and econometric procedures. Ibragimov (2009) provides a helpful discussion of much of this literature, and examines the profound implications of heavy tails for portfolio diversification strategies.

In this paper we are concerned with the impact of heavy tails on a popular estimator for the generalized autoregressive conditionally heteroskedastic (GARCH) model of Bollerslev (1986). GARCH models have been widely used to capture the time variation in volatility found in financial return series. Estimation of GARCH models is typically achieved by quasi-maximum likelihood estimation (QMLE) under a synthetic assumption of Gaussian innovations. There is a large literature examining the asymptotic properties of the QMLE when innovations have finite fourth moment; key papers include Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996), Berkes et al. (2003), Francq and Zakoïan (2004) and Robinson and Zaffaroni (2006). In the more delicate case where the innovations have infinite fourth moment, Hall and Yao (2003) and Mikosch and Straumann (2006) established a stable limit theory for the QMLE. The monograph of Straumann (2005) provides a high
level discussion of the asymptotic behavior of the QMLE in both the light and heavy tailed cases.

The variance targeting estimator (VTE) is a two step estimation procedure for GARCH models that is sometimes used as an alternative to the QMLE. The method originates with Engle and Mezrich (1996) and was studied briefly by Kristensen and Linton (2004) and in more depth by Francq et al. (2011). It is based on a reparametrization of GARCH in which the unconditional variance takes the place of the volatility equation intercept. In the first step the unconditional variance is estimated by the sample variance. In the second step the remaining parameters are estimated by maximizing the restricted quasi-likelihood function given the unconditional variance estimated in the first step. The resulting estimator is therefore a hybrid of QMLE and the method of moments. While inefficient relative to the QMLE, the VTE has two useful properties: it is computationally faster than the QMLE, and it provides a consistent estimate of the unconditional variance even in the presence of model misspecification. This latter property is especially important, since accurate estimation of the unconditional variance is critical for policy relevant measures such as value-at-risk at medium to long horizons.

A useful analysis of the VTE in the case where the stationary distribution of the GARCH process has finite fourth moment is presented in Francq et al. (2011). However, there is to our knowledge no published paper that characterizes the asymptotic distribution of the VTE when the stationary distribution has infinite fourth moment. We provide such a characterization in this paper. Empirical results of Loretan and Phillips (1994) and others suggest that this case may be the relevant one for many financial applications. Moreover, the dynamic structure of GARCH processes is such that, even when innovations are Gaussian and therefore have finite moments of all orders, the fourth moment of the stationary distribution may be infinite. The link between the tail behavior of innovations and the tail behavior of the stationary distribution is provided by Kesten’s theorem, as we will see later.

The primary difficulty in characterizing the behavior of the VTE when the fourth moment of the stationary distribution is infinite is that the central limit theorem is no longer applicable to the sample variance. Instead, a stable limit theory for the sample variance may be derived using point process techniques developed by Resnick (1987) for iid data,
extended by Davis and Hsing (1995) and Davis and Mikosch (1998) to weakly dependent univariate and multivariate processes respectively, and applied by Mikosch and Stărică (2000) and Basrak et al. (2002) to GARCH processes. We find that the stable limit of the sample variance computed in the first step of the VTE dominates the asymptotic behavior of the parameter estimates computed in the second step of the VTE. The asymptotic distribution of the full VTE is nondegenerate but singular, and consists of a stable law distributed over a line in multidimensional space. The rate of convergence of the VTE to this distribution is determined by the tail behavior of the stationary distribution, and is necessarily slower than the rate of convergence of the QMLE, which is governed by the tail behavior of the innovations. This contrasts with the light tailed case, where Francq et al. (2011) have shown that the VTE and QMLE converge at the same rate. We investigate the use of subsampling to approximate the limiting stable law of the VTE, as in Politis et al. (1999), but find that finite sample performance is poor in empirically relevant scenarios.

A recent working paper of Hill and Renault (2012) demonstrates that a tail trimmed version of the VTE may be asymptotically normal even in the presence of heavy tails. Since our results reflect somewhat negatively on the performance of the VTE under heavy tails, they point to the potential importance of techniques such as tail trimming that may reduce the sensitivity of the VTE to extreme realizations.

The remainder of the paper is organized as follows. In Section 2 we review known properties of GARCH processes, devoting particular attention to multivariate regular variation and to the weak convergence of GARCH-based point processes. This material has been developed largely outside the pages of econometrics journals, and we hope that some readers may find our brief review a useful entry point to the literature. Limit theory for the variance targeting estimator is presented in Section 3: we review known results for the light tailed case, and present new results for the heavy tailed case. We investigate the use of subsampling for inference in Section 4, and conclude in Section 5. Some proofs and technical lemmas are collected together in the Mathematical Appendix.
2 GARCH: Definition and properties

Let \((\eta_t)_{t \in \mathbb{Z}}\) be an iid sequence of centered random variables, and let \(p, q \in \mathbb{N}\). For notational convenience, we write \(\eta\) for \(\eta_t\) where appropriate. A stationary sequence of pairs of random variables \((y_t, h_t)_{t \in \mathbb{Z}}\), with each \(h_t\) nonnegative, is said to be a GARCH(p, q) process with innovations \((\eta_t)\) if it satisfies the equations

\[
\begin{align*}
  y_t &= h_t \eta_t \quad \text{(2.1)} \\
  h_t^2 &= c + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{j=1}^{q} b_j h_{t-j}^2 \quad \text{(2.2)}
\end{align*}
\]

for all \(t \in \mathbb{Z}\), and if it is nonanticipatory, meaning that each \(y_t\) can be represented as a measurable function of the current and past innovations \((\eta_s)_{s \leq t}\). We refer to \((y_t)\) as the return process and \((h_t)\) as the volatility process. The GARCH process is parameterized by \(\theta = (c, \lambda, P_\eta)\), where \(\lambda = (a_1, \ldots, a_p, b_1, \ldots, b_q)\), and \(P_\eta\) is the law of \(\eta\). The parameter space is \(\Gamma = \Gamma_1 \times \Gamma_2 \times \Gamma_3\), where \(c \in \Gamma_1\), \(\lambda \in \Gamma_2\), and \(P_\eta \in \Gamma_3\). We define \(\Gamma_1\), \(\Gamma_2\) and \(\Gamma_3\) as follows.

(i) \(\Gamma_1 = (0, \infty)\).

(ii) \(\Gamma_2 = \{\lambda \in [0, \infty)^{p+q} : \sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1 \text{ and } \sum_{i=1}^{p} a_i > 0\}\).

(iii) \(\Gamma_3 = \{P_\eta : \mathbb{E}\eta = 0, \mathbb{E}\eta^2 = 1, P_\eta\text{ has positive density on } \mathbb{R}, \text{ and there exists } \kappa \in (1, \infty] \text{ such that } \mathbb{E}(\eta^2)^\kappa = \infty \text{ and } \mathbb{E}(\eta^2)^{\kappa'} < \infty \text{ for all } \kappa' \in [0, \kappa)\}\}.

We next review known properties of GARCH(p, q) processes with \(\theta \in \Gamma\). More extensive reviews may be found in Lindner (2009) and Davis and Mikosch (2009).

2.1 Stochastic recurrence equation representation

The GARCH(p, q) model can be nested in the more general framework of stochastic recurrence equations (SREs) in the following way. If \(q\) or \(p\) is less than 2, then embed the GARCH(p, q) process in a GARCH(2 ∨ p, 2 ∨ q) process with \(a_i = 0\) for \(i > p\) and \(b_j = 0\)
for \( j > q \). Let \( d = p + q - 1 \), and define the sequence of \( d \times 1 \) random vectors \((X_t)_{t \in \mathbb{Z}}\) by

\[
X_t = (h_{t+1}^2, h_t^2, \ldots, h_{t-q+2}^2, y_t^2, y_{t-1}^2, \ldots, y_{t-p+2}^2)',
\]  

(2.3)

and the iid sequence of random \( d \times d \) matrices \((A_t)_{t \in \mathbb{Z}}\) and constant \( d \times 1 \) vector \( B \) by

\[
A_t = \begin{bmatrix}
    a_1 \eta_t^2 + b_1 & b_2 & \cdots & b_{q-1} & b_q & a_2 & \cdots & a_{p-1} & a_p \\
    1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
    0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\
    \eta_t^2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
    0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
    c \\
    0 \\
    0 \\
    \vdots \\
    0 \\
    0 \\
    \vdots \\
    0 \\
    0 \\
\end{bmatrix}.
\]  

(2.4)

Then \( X_t \) satisfies the SRE

\[
X_t = A_t X_{t-1} + B.
\]  

(2.5)

The SRE representation of the GARCH\((p, q)\) model has been instrumental in providing general conditions for the existence of a stationary nonanticipatory solution to (2.1) and (2.2), as well as establishing mixing conditions and describing the behavior of the tails of the marginal distributions of the squared returns \(y_t^2\) and squared volatilities \(h_t^2\). Building on results of Nelson (1990), Bougerol and Picard (1992) showed that if \( c > 0 \), there exists a unique stationary nonanticipatory solution to the SRE (2.5) if and only if

\[
\gamma := \inf \left\{ \frac{1}{n} \mathbb{E} \ln \|A_1 \cdots A_n\| : n \in \mathbb{N} \right\} < 0,
\]  

(2.6)

where \( \| \cdot \| \) is the operator norm induced by some norm on \( \mathbb{R}^d \). \( \gamma \) is referred to as the top Lyapunov exponent of \((A_t)\), and its sign is invariant to the choice of norm on \( \mathbb{R}^d \). When \( \theta \in \Gamma \) we have \( \gamma < 0 \), and the solution to (2.5) is given by

\[
X_t = B + \sum_{k=1}^{\infty} A_t \cdots A_{t-k+1} B.
\]  

(2.7)
As \((X_t)\) is stationary, for convenience we will write \(X\) for \(X_t\) when appropriate. Boussama (1998, Theorem 3.4.2) showed that when \(P_\eta\) has positive density in a neighborhood of the origin, as it must when \(\theta \in \Gamma\), \((X_t)\) is \(\beta\)-mixing (absolutely regular), with an exponential decay rate of mixing coefficients. See Carrasco and Chen (2002) for further results on the mixing properties of GARCH\((p,q)\) processes.

### 2.2 Parametrization by unconditional variance

When \(\theta \in \Gamma\) the unconditional variance of \(y_t\) is finite and given by

\[
\sigma^2 := \text{Var}(y_t) = \mathbb{E}y_t^2 = c \left( 1 - \sum_{i=1}^{p} a_i - \sum_{j=1}^{q} b_j \right)^{-1}.
\]

(2.8)

It is therefore equivalent to parameterize the GARCH\((p,q)\) model by \(\theta = (\sigma^2, \lambda, P_\eta)\), and let \(c(\sigma^2, \lambda) = \sigma^2(1 - \sum_{i=1}^{p} a_i - \sum_{j=1}^{q} b_j)\). The existence of a finite unconditional variance is the primary reason to restrict the parameter space by requiring \(\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1\); this condition is not in fact necessary for the top Lyapunov exponent \(\gamma\) to be negative. A necessary but insufficient condition for \(\gamma < 0\) is \(\sum_{j=1}^{q} b_j < 1\) (Bougerol and Picard, 1992, Corollary 2.3). However, the variance targeting estimator is clearly not appropriate when \(\mathbb{E}y_t^2 = \infty\), and hence we restrict the parameter space for \(\lambda\) to \(\Gamma_2\).

### 2.3 Multivariate regular variation

Loosely, a random vector \(X\) exhibits multivariate regular variation if the tail probability decays like a power law in any direction. In the univariate case, regular variation is often defined in terms of the tail behavior of distribution functions, but in the multivariate case it is more natural to use convergence of measures.

We now briefly define vague convergence of Radon measures. A more detailed account of this subject may be found in Kallenberg (1983, pp. 168–171) or Resnick (1987, pp. 139–149). Let \(E\) be a locally compact second countable Hausdorff space equipped with a \(\sigma\)-algebra \(\mathcal{E}\), and let \(C^+_K(E)\) denote the class of all continuous functions \(f : E \rightarrow \mathbb{R}_+\) with
compact support. A measure $\mu$ on $E$ is Radon if $\mu(A) < \infty$ for all relatively compact $A \in \mathcal{E}$. Let $M_+(E)$ denote the collection of Radon measures on $E$. A topology on $M_+(E)$ can be generated by using as a base the collection of finite intersections of sets $\{\mu \in M_+(E) : \int_E f d\mu \in (s,t)\}$, $f \in C_c^+(E)$, $0 < s < t < \infty$. This construction maps the usual topology of open intervals on $\mathbb{R}$ into a topology on $M_+(E)$ via the Lebesgue integral. The topology generated is known as the vague topology, and under this topology $M_+(E)$ is separable and completely metrizable (Kallenberg, 1983, Theorem 15.7.7). We write $\vto$ to signify convergence in the vague topology, or vague convergence.

Note that vague convergence is a more general notion than weak convergence, which we denote by $\pto$. The latter deals with classes of bounded measures on $E$, while measures in $M_+(E)$ need only be finite on relatively compact sets. Indeed, for bounded measures $\mu, \mu_1, \mu_2, \ldots \in M_+(E)$, $\mu_n \pto \mu$ if and only if $\mu_n \vto \mu$ and $\mu_n(E) \to \mu(E)$ (Kallenberg, 1983, Theorem 15.7.6).

We can now define multivariate regular variation in terms of the vague convergence of a sequence of Radon measures on the state space $E = \mathbb{R}^d \setminus \{0\}$ equipped with the Borel $\sigma$-algebra $\mathcal{E} = \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, where $\mathbb{R} = \mathbb{R} \cup \{-\infty, \infty\}$. In this space, Borel sets bounded away from 0 are relatively compact; this is the purpose of puncturing the usual space $\mathbb{R}^d$ at the origin. An $\mathbb{R}^d$ valued random vector $X$ is said to be multivariate regularly varying with index $\alpha \in (0, \infty)$ if the following condition is satisfied.

(i) There exists a sequence $(a_n) \uparrow \infty$ and a nonnull $\mu \in M_+(\mathbb{R}^d \setminus \{0\})$ such that

$$n \mathbb{P}(a_n^{-1}X \in \cdot) \vto \mu(\cdot) \quad \text{as } n \to \infty, \quad (2.9)$$

and $\mu(sB) = s^{-\alpha} \mu(B)$ for all $s > 0$ and all relatively compact $B \in \mathcal{B}(\mathbb{R}^d_+ \setminus \{0\})$.

Alternatively, given a norm $|\cdot|$ on $\mathbb{R}^d$, we may define multivariate regular variation of $X$ in terms of the weak convergence of a sequence of bounded measures on the unit sphere $S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$.

(ii) There exists a probability measure $\Theta$ on $S^{d-1}$ such that, for all $s > 0$,

$$\frac{\mathbb{P}(|X| > sx, X/|X| \in \cdot)}{\mathbb{P}(|X| > x)} \pto s^{-\alpha} \Theta(\cdot) \quad \text{as } x \to \infty. \quad (2.10)$$
Conditions (i) and (ii), taken from Resnick (1986, p. 69), are equivalent. Further equivalent characterizations are given by Resnick (2004). Condition (i) illustrates nicely the utility of vague convergence when characterizing the tail behavior of $X$: if $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ is relatively compact, it is bounded away from 0, and so for large $n$ the probability of the event $a_n^{-1}X \in A$ depends only on the tail behavior of $X$. The Radon measure $\mu$ in condition (i) satisfies $\mu\{x : |x| > s, x/|x| \in \cdot\} = cs^{-\alpha}\Theta(\cdot)$ for all $s > 0$, where $c > 0$ and $\Theta$ is the probability measure in condition (ii). $\Theta$ is called the spectral measure of $X$, and captures the tail dependency between the components of $X$, as well as the relative likelihood of positive and negative extremes. We see from condition (ii) that when $X$ is multivariate regularly varying with index $\alpha$ we have $\mathbb{P}(|X| > x) = x^{-\alpha}L(x)$, where $L$ is a slowly varying function.

The representation of the GARCH($p, q$) model as a SRE provides insights into the tail behavior of $X_t$ via the theory developed by Kesten (1973), as demonstrated in Basrak et al. (2002, Theorem 3.1). Although Kesten’s theorems are stated in terms of univariate regular variation of linear combinations of $X_t$, results of Basrak et al. (2002) and Boman and Lindskog (2009, Corollary 2) show the connection with multivariate regular variation. These results are summarized in the following theorem, adapted from Davis and Mikosch (2009, Theorem 1). The binary relation $\sim$ indicates that the ratio of two quantities tends to one as $x$ (or $n$, depending on context) tends to infinity.

**Theorem 2.1.** Suppose $\theta \in \Gamma$, and let $(X_t)$ be the unique stationary solution to the SRE (2.5). Then, writing $X$ for $X_t$, the following statements are true.

(i) $X$ is multivariate regularly varying with index $\alpha \in (1, \kappa)$.

(ii) For all $a \in \mathbb{R}^d \setminus \{0\}$, we have $\mathbb{P}(a'X > x) \sim w(a)x^{-\alpha}$ for some $w(a) \in [0, \infty)$. Moreover, if $a \in [0, \infty)^d \setminus \{0\}$, then $w(a) > 0$.

(iii) $\mathbb{P}(|X| > x) \sim cx^{-\alpha}$ for some $c \in (0, \infty)$.

Note that the regular variation of $|X|$ in Theorem 2.1(iii) is of the special kind where the slowly varying function $L(x) = x^\alpha \mathbb{P}(|X| > x)$ converges to a constant. Let the sequence $(a_n)$ satisfy

$$n\mathbb{P}(|X| > a_n) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty. \quad (2.11)$$
Then $\mathbb{P}(|X| > x) \sim cx^{-\alpha}$ implies that $a_n \sim (cn)^{1/\alpha}$. Letting $a$ in Theorem 2.1(iii) be the unit vector for the $(q + 1)^{th}$ or $2^{nd}$ coordinate, we see that $n\mathbb{P}(y_t^2 > a_n) \to c_y$ and $n\mathbb{P}(h_t^2 > a_n) \to c_h$ for constants $c_y, c_h \in (0, \infty)$. The sequence $(a_n)$ therefore controls the tails of $y_t^2$ and $h_t^2$, and in fact by Karamata’s theorem (Bingham et al., 1989, Theorem 1.5.11) we have

$$\frac{n}{a_n} \mathbb{E}((y_t^2)^\gamma 1(y_t^2 \leq a_n \epsilon)) \to \frac{c_y \alpha \epsilon^{\gamma - \alpha}}{\gamma - \alpha}, \quad \frac{n}{a_n} \mathbb{E}((h_t^2)^\gamma 1(h_t^2 \leq a_n \epsilon)) \to \frac{c_h \alpha \epsilon^{\gamma - \alpha}}{\gamma - \alpha}$$

(2.12)
as $n \to \infty$, for all $\gamma > \alpha$ and $\epsilon > 0$.

In general, the index of regular variation $\alpha$ is given by Kesten’s theorem as the unique solution to the equation

$$\lim_{n \to \infty} n^{-1} \ln \mathbb{E}\|A_n \cdots A_1\|^\alpha = 0.$$  

(2.13)

For the GARCH(1, 1) model (2.13) simplifies to $\mathbb{E}(a_1 \eta^2 + b_1)^\alpha = 1$, which can be solved by simulation or direct computation of the expectation if $(a_1, b_1, P_\eta)$ are known. However, it is difficult to determine $\alpha$ in the more general GARCH($p$, $q$) model. That $\alpha > 1$ is clear since $\mathbb{E}y_t^2 < \infty$, while $\alpha < \kappa$ follows from the fact that $\mathbb{E}\|A_n \cdots A_1\|^\kappa' = \infty$ for $\kappa' \geq \kappa$.

### 2.4 Point process convergence

Point process theory deals with the weak convergence of distributions over spaces of Radon counting measures, and is a powerful tool for establishing limit theory for weakly dependent sequences of random variables whose finite dimensional distributions are multivariate regularly varying. Here we review a few basic elements of this theory that are key to establishing certain results on the variance targeting estimator presented in the following section. For general information on point process theory, see Kallenberg (1983) and Resnick (1987).

Return for a moment to the general setting where our state space $E$ is some locally compact second countable Hausdorff space equipped with a $\sigma$-algebra $\mathcal{E}$. A Radon measure $\mu \in M_+(E)$ is said to be a Radon counting measure on $E$ if $\mu(A) \in \mathbb{N} \cup \{0, \infty\}$ for all $A \in \mathcal{E}$. Let $M_p(E)$ be the set of Radon counting measures on $E$, and equip $M_p(E)$ with the
σ-algebra $\mathcal{M}_p(E)$ generated by the vague topology. Every $\mu \in \mathcal{M}_p(E)$ can be represented as $\mu = \sum_{i \in I} \delta_{x_i}$, a countable sum of Dirac measures concentrated at the points $(x_i)_{i \in I}$ in $E$. These points need not be distinct: $x_i$ has multiplicity $\mu(\{x_i\})$. If $I$ is empty then $\mu$ is the null measure. Since $\mu$ is Radon, $(x_i)_{i \in I}$ can have no accumulation point in $E$.

A point process on $E$ is a random element of $\mathcal{M}_p(E)$; that is, a measurable map $N$ from a probability space $(\Omega, \mathcal{F}, P)$ to the measurable space $(\mathcal{M}_p(E), \mathcal{M}_p(E))$. One very important class of point processes is the family of Poisson processes. A point process $N$ is said to be a Poisson process if both of the following statements are true.

(i) For any $k \in \mathbb{N}$, if $A_1, \ldots, A_k \in \mathcal{E}$ are mutually disjoint sets, then $N(A_1), \ldots, N(A_k)$ are mutually independent random variables.

(ii) There is a Radon measure $\mu \in \mathcal{M}_+(E)$, called the intensity measure of $N$, such that, for any $A \in \mathcal{E}$ and any $k \in \mathbb{N} \cup \{0\}$,

$$
P(N(A) = k) = \begin{cases} 
e^{-\mu(A)}\mu(A)^k/k! & \text{if } \mu(A) < \infty \\ 0 & \text{if } \mu(A) = \infty. \end{cases} \quad (2.14)$$

Given any $\mu \in \mathcal{M}_+(E)$, we may construct a Poisson process on $E$ with intensity measure $\mu$, and its distribution is uniquely determined by $\mu$ (Resnick, 1987, Proposition 3.6).

Let us now consider the more concrete setting where we have state space $E = \mathbb{R}^d_+ \setminus \{0\}$ equipped with the Borel $\sigma$-algebra $\mathcal{E} = \mathcal{B}(\mathbb{R}^d_+ \setminus \{0\})$, where $\mathbb{R}^d_+ := [0, \infty]$. The unique stationary solution $(X_t)$ to the SRE (2.5) generates a sequence of point processes on $\mathbb{R}^d_+ \setminus \{0\}$ given by

$$
N_n = \sum_{t=1}^n \delta_{X_t/a_n}, \quad n \in \mathbb{N}, \quad (2.15)
$$

where $(a_n)$ is a sequence satisfying (2.11). Basrak et al. (2002) have shown by appealing to Davis and Mikosch (1998, Theorem 2.8), itself a multivariate extension of Davis and Hsing (1995, Theorem 2.7), that the sequence of point processes $N_n$ converges weakly to a limiting point process $N$ on $\mathbb{R}^d_+ \setminus \{0\}$. We summarize their results in the following theorem.

**Theorem 2.2.** Suppose $\theta \in \Gamma$, and let $(X_t)$ be the unique stationary solution to the SRE (2.5). Then the sequence of point processes $N_n$ defined in (2.15) satisfies $N_n \stackrel{w}{\rightarrow} N$ in the
vague topology, where $N$ is a point process on $\mathbb{R}_+^d \setminus \{0\}$ admitting the representation

$$N = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}. \tag{2.16}$$

Here,

(i) $\sum_{i=1}^{\infty} \delta_{P_i}$ is a Poisson process on $(0, \infty)$ with absolutely continuous intensity measure $\nu(dy) = \zeta \alpha y^{-\alpha-1} dy$, where $\alpha \in (1, \kappa)$ is the index of multivariate regular variation of $X$, and $\zeta \in (0, 1]$.

(ii) $\sum_{j=1}^{\infty} \delta_{Q_{ij}}, \ i \in \mathbb{N}$, is an iid sequence of point processes on $\mathbb{R}_+^d \setminus \{0\}$ taking values in the set $\{\mu \in M_p(\mathbb{R}_+^d \setminus \{0\}) : \mu(\{x : |x| > 1\}) = 0$ and $\mu(S^{d-1}) > 0\}$.

(iii) The point processes $\sum_{i=1}^{\infty} \delta_{P_i}$ and $\sum_{j=1}^{\infty} \delta_{Q_{ij}}, \ i \in \mathbb{N}$, are mutually independent.

(2.16) is known as the cluster representation of $N$ (Davis and Hsing, 1995, Corollary 2.4), while $\zeta$ is the extremal index of $(|X_t|)$ (Leadbetter and Rootzen, 1988, pp. 438–440). The point process convergence $N_n \xrightarrow{w} N$ plays a key role in a number of studies of the asymptotic behavior of statistics based on heavy tailed GARCH processes. Davis and Mikosch (1998), Mikosch and Stărică (2000) and Basrak et al. (2002) used point process methods to establish a stable limit theory for the sample autocovariances and autocorrelations of ARCH(1), GARCH(1, 1) and GARCH$(p, q)$ processes respectively, while Mikosch and Straumann (2006) used point process methods to study the asymptotic behavior of the QMLE for GARCH$(p, q)$ models when innovations are heavy tailed. In the following section we will use Theorem 2.2 to develop limit theory for a self-normalized version of the VTE for GARCH$(p, q)$ models with heavy tailed marginal distributions.

### 3 Limit theory for the VTE

In this section we investigate the asymptotic behavior of the VTE for the GARCH$(p, q)$ model. We review known results on the VTE applicable when $E y_t^4 < \infty$, and present new results covering the more problematic case $E y_t^4 = \infty$. 

3.1 Estimator and assumptions

The VTE is a two step estimator of the parameters $\nu = (\sigma^2, \lambda)$ in the GARCH($p, q$) model. In the first step, the mean square of the sample $y_1, \ldots, y_n$ is computed and used as an estimator for the unconditional variance, $\sigma^2$. Define the first step estimator as

$$\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} y_t^2. \quad (3.1)$$

We will refer to $\hat{\sigma}_n^2$ as the sample variance. In the second step, the remaining parameters $\lambda$ are estimated by maximizing the Gaussian quasi-likelihood of the sample, with $\sigma^2$ restricted to the estimate $\hat{\sigma}_n^2$ computed in the first step. For $\nu \in \Gamma_1 \times \Gamma_2$ and $t = 1, \ldots, n$, define the feasible conditional volatility $\tilde{h}_t^2(\nu)$ as

$$\tilde{h}_t^2(\nu) = \begin{cases} 
    c(\nu) + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{j=1}^{} b_j \tilde{h}_{t-j}^2(\nu) & \text{if } p \lor q < t \leq n \\
    \sigma^2 & \text{if } 1 \leq t \leq p \lor q.
\end{cases} \quad (3.2)$$

The second step Gaussian quasi-likelihood is

$$L_n(\lambda) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \tilde{h}_t^2(\hat{\sigma}_n^2, \lambda)}} \exp \left( -\frac{y_t^2}{2\tilde{h}_t^2(\hat{\sigma}_n^2, \lambda)} \right), \quad (3.3)$$

and the quasi-log-likelihood is

$$\log L_n(\lambda) = -\frac{1}{2} \sum_{t=1}^{n} \left( \log 2\pi \tilde{h}_t^2(\hat{\sigma}_n^2, \lambda) + \frac{y_t^2}{\tilde{h}_t^2(\hat{\sigma}_n^2, \lambda)} \right) = \frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \tilde{\ell}_t(\hat{\sigma}_n^2, \lambda), \quad (3.4)$$

where

$$\tilde{\ell}_t(\nu) = \log \tilde{h}_t^2(\nu) + \frac{y_t^2}{\tilde{h}_t^2(\nu)}. \quad (3.5)$$

The second step estimator $\hat{\lambda}_n$ is given by any measurable solution of

$$\log L_n(\hat{\lambda}_n) = \max_{\lambda \in \Lambda} \log L_n(\lambda), \quad (3.6)$$

where the solution space $\Lambda \subset \Gamma_2$ satisfies Assumption 3.1 below. The variance targeting estimator is then $\hat{\nu}_n = (\hat{\sigma}_n^2, \hat{\lambda}_n)$. 

13
Let \( c_0, \sigma^2_0, (a_{0i}), (b_{0j}) \) and \( P_{n,0} \) denote the true values of the unknown parameters \( c, \sigma^2, (a_i), (b_j) \) and \( P_n \) respectively, and let \( \lambda_0 = (a_{01}, \ldots, a_{0p}, b_{01}, \ldots, b_{0q}), \nu_0 = (\sigma^2_0, \lambda_0) \) and \( \theta_0 = (\nu_0, P_{n,0}) \). We impose the following conditions on \( \theta_0 \) and \( \Lambda \).

**Assumption 3.1.** (i) \( \theta_0 \in \Gamma \). (ii) \( \Lambda \) is a compact subset of \( \Gamma_2 \) whose interior contains \( \lambda_0 \). (iii) \( A(x) = \sum_{i=1}^{p} a_{0i} x^i \) and \( B(x) = 1 - \sum_{j=1}^{q} b_{0j} x^j \) are coprime in the set of polynomials with real coefficients.

Part (i) of Assumption 3.1 suffices for the various nice properties of GARCH processes discussed in Section 2, while conditions resembling (ii) are a standard ingredient for the study of extremum estimators. Condition (iii) is an identification condition used by Berkes et al. (2003, Theorem 2.5) to rule out observationally equivalent submodels of the GARCH\((p,q)\) specification.

### 3.2 Weak convergence of the VTE

Strong consistency of the VTE has been shown by Francq et al. (2011, Theorem 2.1) under conditions more general than Assumption 3.1. We will study weak convergence of suitable normalizations of the VTE, paying particular attention to the self-normalized quantity

\[
S_n = \frac{\sqrt{n}(\hat{\nu}_n - \nu_0)}{\tau_n}, \quad \text{where} \quad \tau_n^2 = \frac{1}{n} \sum_{t=1}^{n} y_t^4. \tag{3.7}
\]

Since \( \theta_0 \in \Gamma \), Theorem 2.1 implies that \( y_t^2 \) is regularly varying with index \( \alpha \in (1, \infty) \). Normalization of \( n^{1/2}(\hat{\nu}_n - \nu_0) \) by \( \tau_n \) is useful because the rate at which \( \hat{\nu}_n \) converges to \( \nu_0 \) depends on \( \alpha \). If \( \alpha \in (2, \infty) \) we have \( \mathbb{E}y_t^4 < \infty \). In this case, \( \hat{\nu}_n \) converges to \( \nu_0 \) at the rate \( n^{1/2} \) and \( \tau_n \) converges to a nonzero finite limit, so that \( S_n \) has a nondegenerate weak limit. On the other hand, if \( \alpha \in (1, 2) \) we have \( \mathbb{E}y_t^4 = \infty \). We will see that in this case \( \hat{\nu}_n \) converges to \( \nu_0 \) at a rate slower than \( n^{1/2} \), but \( \tau_n \) diverges to infinity at just the right rate to ensure that \( S_n \) continues to have a nondegenerate weak limit. This facilitates the use of subsampling techniques for inference (Politis et al., 1999) that do not require \( \alpha \) to be estimated, as we will see in Section 4. Following related papers on heavy tailed GARCH processes by Davis and Mikosch (1998), Mikosch and Stˇ aricˇ a (2000) and Basrak et al. (2002), we exclude from consideration the borderline case \( \alpha = 2 \), where \( y_t^2 \) has
The partition of the parameter space is valid for $E\eta^4 = 3$ in panel (a), and for $E\eta^4 \approx 7.83$ in panel (b). Parameters for daily or weekly returns on the Dow Jones Industrial Average index are the QMLEs reported by Francq et al. (2011, Tables 6-7).

The case $E\eta^4 = \infty$ is not merely an intellectual curiosity, and occurs for empirically plausible parameter values even if the innovations ($\eta_t$) are assumed to be Gaussian. In Figure 3.1 we partition the parameter space of the stationary GARCH(1,1) model according to whether $E\eta^4 < \infty$ or $E\eta^4 = \infty$; the former is true if and only if $E(a_1\eta^2 + b_1)^2 < 1$. The partition displayed in panel (a) obtains when $\eta \sim \mathcal{N}(0, 1)$ (or $E\eta^4 = 3$), while the partition in panel (b) is valid for $\eta \sim 0.774 \times t_5$ (or $E\eta^4 \approx 7.83$). We identify points in the parameter space corresponding to estimates reported by Francq et al. (2011) for daily and weekly returns on the Dow Jones Industrial Average (DJIA) index. Even in panel (a), where the innovation density is standard normal, we see that $E\eta^4 = \infty$ over a large region of the parameter space. Moreover, the parameter estimates for daily and weekly DJIA returns lie extremely close to the threshold of the regions of finite and infinite fourth moment. In panel (b), where the innovation distribution is heavier tailed, the region in
which \( \mathbb{E}y_t^4 = \infty \) takes up a larger share of the parameter space. The parameter estimates for weekly DJIA returns now lie squarely within the region where \( \mathbb{E}y_t^4 = \infty \), while the estimates for daily returns remain close to the threshold.

### 3.2.1 Preliminary results

We shall review some useful preliminary results that hold regardless of whether \( \mathbb{E}y_t^4 < \infty \) or \( \mathbb{E}y_t^4 = \infty \). Since \( \hat{\lambda}_n \) is strongly consistent for \( \lambda_0 \), and \( \lambda_0 \) lies in the interior of \( \Lambda \), we know that, for sufficiently large \( n \), \( \hat{\lambda}_n \) solves the first order condition

\[
0 = \frac{\partial}{\partial \lambda} \log L_n(\hat{\lambda}_n) = -\frac{1}{2} \sum_{t=1}^{n} \frac{\partial \tilde{\ell}_t(\hat{\nu}_n)}{\partial \lambda} .
\]

(3.8)

For each \( i = 1, \ldots, p + q \), a Taylor expansion of \( \frac{\partial \tilde{\ell}_t(\hat{\nu}_n)}{\partial \lambda_i} \) around \( \nu = \nu_0 \) yields

\[
0 = \sum_{t=1}^{n} \left( \frac{\partial \tilde{\ell}_t(\nu_0)}{\partial \lambda_i} + \frac{\partial^2 \tilde{\ell}_t(\nu^*_i)}{\partial \lambda_i \partial \nu} (\hat{\nu}_n - \nu_0) \right) \\
= \sum_{t=1}^{n} \frac{\partial \tilde{\ell}_t(\nu_0)}{\partial \lambda_i} + J_n^{(i)} n(\hat{\lambda}_n - \lambda_0) + K_n^{(i)} n(\hat{\sigma}_n^2 - \sigma_0^2),
\]

(3.10)

where \( \nu^*_i \) lies on the line segment between \( \hat{\nu}_n \) and \( \nu_0 \), and

\[
J_n^{(i)} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{\ell}_t(\nu^*_i)}{\partial \lambda_i \partial \lambda'}, \quad K_n^{(i)} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{\ell}_t(\nu^*_i)}{\partial \lambda_i \partial \sigma^2}.
\]

(3.11)

Let \( J_n \) denote the \( (p + q) \times (p + q) \) matrix with \( i \)th row \( J_n^{(i)} \), and let \( K_n \) denote the \( (p + q) \times 1 \) vector with \( i \)th element \( K_n^{(i)} \).

To examine the asymptotic behavior of \( \hat{\lambda}_n \), it is useful to approximate the sequences \( (\tilde{h}_t(\nu)) \) and \( (\tilde{\ell}_t(\nu)) \) by stationary ergodic sequences. To that end, define the infeasible conditional volatility as the unique stationary solution to

\[
\tilde{h}_t^2(\nu) = c(\nu) + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{j=1}^{q} b_j \tilde{h}_{t-j}^2(\nu),
\]

(3.12)
and note that \( h_t(\nu_0) = h_t^2 \). Also define \( \ell_t(\nu) \), the infeasible analog to \( \tilde{\ell}_t(\nu) \), as
\[
\ell_t(\nu) = \log h_t^2(\nu) + \frac{y_t^2}{h_t^2(\nu)}.
\]  
(3.13)

**Lemma 3.1.** Under Assumption 3.1, \( J_n \xrightarrow{a.s.} J \) and \( K_n \xrightarrow{a.s.} K \) as \( n \to \infty \), where
\[
J = \mathbb{E} \left( \frac{\partial^2 \ell_t(\nu_0)}{\partial \lambda \partial \lambda'} \right), \quad K = \mathbb{E} \left( \frac{\partial^2 \ell_t(\nu_0)}{\partial \lambda \partial \sigma^2} \right).
\]  
(3.14)

Further, the matrix \( J \) is nonsingular.

Lemma 3.1 follows from the proof of Francq and Zakoïan (2004, Theorem 2.2), specifically items (ii), (iv), and (vi). Although Francq and Zakoïan assume that \( \mathbb{E} \eta_t^4 < \infty \), this assumption is not used to show the claims of Lemma 3.1. The parametrization used by Francq and Zakoïan is \((c, \lambda)\), but their arguments apply directly to our equivalent parametrization \((\sigma^2, \lambda)\). Nonsingularity of \( J \) depends critically on the identification condition given in Assumption 3.1(iii).

### 3.2.2 Weak convergence with finite fourth moment

When \( \mathbb{E} y_t^4 < \infty \), a central limit theorem for martingale difference sequences may be used to demonstrate that the VTE converges at the rate \( n^{1/2} \) to a Gaussian limit. The following theorem is a slightly modified version of Francq et al. (2011, Theorem 2.1), and uses the fact that \( \tau_n^2 \xrightarrow{a.s.} \mathbb{E} y_t^4 = (\mathbb{E} h_t^4)(\mathbb{E} \eta_t^4) \) by the ergodic theorem.

**Theorem 3.1.** Under Assumption 3.1, if \( \mathbb{E} y_t^4 < \infty \) then
\[
\sqrt{n} \left( \hat{\nu}_n - \nu_0 \right) \xrightarrow{w} \mathcal{N} \left( 0, (\mathbb{E} \eta_t^4 - 1) \Sigma \right) \quad \text{and} \quad S_n \xrightarrow{w} \mathcal{N} \left( 0, \left( \frac{1}{\mathbb{E} h_t^4} - \frac{1}{\mathbb{E} y_t^4} \right) \Sigma \right),
\]
where the nonsingular matrix \( \Sigma \) is given by
\[
\Sigma = \begin{bmatrix}
D & -DK'J^{-1} \\
-DJ^{-1}K & J^{-1} + DJ^{-1}KK'J^{-1}
\end{bmatrix}, \quad D = C^2\mathbb{E} h_t^4, \quad C = \frac{\sigma_0^2}{c_0} \left( 1 - \sum_{j=1}^{q} b_{0j} \right).
\]
For details of the proof, see Francq et al. (2011). There it is shown further that the VTE is asymptotically less efficient than the QMLE, which also converges at the rate \( n^{1/2} \) to a Gaussian limit.

### 3.2.3 Weak convergence with infinite fourth moment

When \( \mathbb{E} y_t^4 = \infty \), the results of Francq et al. (2011) establishing asymptotic normality of the VTE are no longer applicable. We will show that a suitable normalization of the VTE nevertheless converges weakly to a stable limit. As discussed earlier, we focus on the case where the index of regular variation \( \alpha \) appearing in Theorem 2.1 satisfies \( \alpha \in (1, 2) \), and rule out the troublesome borderline case \( \alpha = 2 \).

Let \( (a_n) \) be a sequence satisfying (2.11), and define

\[
U_n = a_n^{-1} \sum_{t=1}^{n} (y_t^2 - \mathbb{E} y_t^2), \quad V_n = a_n^{-2} \sum_{t=1}^{n} y_t^4.
\]

(3.15)

Recall that \( a_n \sim cn^{1/\alpha} \) for some \( c > 0 \). A random variable \( X \) is said to be \( \alpha \)-stable if it has characteristic function

\[
\mathbb{E} \exp(itX) = \begin{cases} 
\exp \left( i\mu t - |\varsigma t|^{\alpha} \left( 1 - i\beta \text{sgn}(t) \tan(\pi \alpha / 2) \right) \right) & \text{if } \alpha \neq 1 \\
\exp \left( i\mu t - |\varsigma t| \left( 1 + i\beta(2/\pi)\text{sgn}(t) \ln |t| \right) \right) & \text{if } \alpha = 1.
\end{cases}
\]

(3.16)

The parameters \( \mu, \varsigma \) and \( \beta \) signify location, scale and skew respectively. See Samoridnitsky and Taqqu (1994) for further discussion of stable distributions.

**Theorem 3.2.** Under Assumption 3.1, if \( \alpha \in (1, 2) \) then

\[
(U_n, V_n) \overset{w}{\to} (U, V),
\]

(3.17)

where \( U \) is a nondegenerate \( \alpha \)-stable random variable and \( V \) is a strictly positive \( \alpha/2 \)-stable random variable.

Results similar to Theorem 3.2 have appeared in prior literature. Basrak et al. (2002, Theorem 3.6) proved that \( U_n \overset{w}{\to} U \) under the additional requirement that \( \mathbb{E} \eta^4 < \infty \),
which is assumed in equation (3.5) in their paper. Similar results were proved by Davis and Mikosch (1998) and Mikosch and Stărică (2000) for the ARCH(1) and GARCH(1,1) models respectively. Here we relax the requirement that $\mathbb{E}e^4 < \infty$ by appealing to a moment inequality for martingales due to von Bahr and Esseen (1965). Another result similar to Theorem 3.2 appeared in Kokoszka and Wolf (2004, Theorem 2), but imposes a symmetry condition that makes it inapplicable here due to the nonnegativity of $y_t^2$.

We now give a brief proof of Theorem 3.2, relegating details to technical lemmas in the Mathematical Appendix. The main ideas are due to Davis and Hsing (1995), and provide a nice illustration of the utility of the point process methods discussed in Section 2.4.

**Proof of Theorem 3.2.** Following Basrak et al. (2002, pp. 110–111), we begin by noting that $U_n = U_n^* + o_p(1)$, where

$$U_n^* = Ca_n^{-1} \sum_{t=1}^{n} h_t^2 (\eta_t^2 - 1), \quad (3.18)$$

and $C$ was defined in Theorem 3.1. To isolate the effect of the extreme realizations of $y_t^2$ and $h_t^2$ on $(U_n^*, V_n)$, we consider the censored partial sums

$$U_{ne}^* = Ca_n^{-1} \sum_{t=1}^{n} h_t^2 (\eta_t^2 - 1)1(\eta_t^2 > \varepsilon a_n), \quad V_{ne} = a_n^{-2} \sum_{t=1}^{n} y_t^4 1(y_t^2 > \varepsilon a_n), \quad (3.19)$$

where $\varepsilon > 0$. Observe that $(U_{ne}^*, V_{ne}) = T_\varepsilon(N_n)$, where $N_n$ is the point process defined in (2.15) and $T_\varepsilon : M_p(\mathbb{R}^d_+ \setminus \{0\}) \rightarrow \mathbb{R}^2$ is the map

$$T_\varepsilon \left( \sum_{i=1}^{\infty} \delta_{x_i} \right) = \left( C \sum_{i=1}^{\infty} (x_{i,q+1} - x_{i,2}) 1(x_{i,2} > \varepsilon), \sum_{i=1}^{\infty} x_{i,q+1}^2 1(x_{i,q+1} > \varepsilon) \right). \quad (3.20)$$

Theorem 2.2 states that $N_n \xrightarrow{w} N$, where $N$ satisfies the cluster representation (2.16). Thanks to the $\varepsilon$-trimming in (3.20), it can be shown (see Lemma A.2) that $T_\varepsilon$ is continuous on a subset of $M_p(\mathbb{R}^d_+ \setminus \{0\})$ inhabited by $N$ with probability one, and so an application of the continuous mapping theorem yields

$$(U_{ne}^*, V_{ne}) \xrightarrow{w} T_\varepsilon(N) \quad \text{as } n \rightarrow \infty. \quad (3.21)$$
Careful inspection of the characteristic function of \( T_\epsilon(N) \) (see Lemma A.3) reveals that
\[
T_\epsilon(N) \xrightarrow{w} (U, V) \quad \text{as } \epsilon \rightarrow 0. \tag{3.22}
\]
A result of Billingsley (1968, Theorem 3.2) allows us to combine (3.21) and (3.22) to obtain \((U_n, V_n) \xrightarrow{w} (U, V)\) as \( n \rightarrow \infty \), provided that the censoring of partial sums is asymptotically negligible in the sense that, for any \( \delta > 0 \),
\[
\lim_{\epsilon \to 0} \limsup_{n \to \infty} \mathbb{P}(|U^*_n - U^*_\epsilon| \geq \delta) = 0 \tag{3.23}
\]
and
\[
\lim_{\epsilon \to 0} \limsup_{n \to \infty} \mathbb{P}(|V_n - V_{\epsilon}| \geq \delta) = 0. \tag{3.24}
\]
(3.23) and (3.24) can be proved (see Lemma A.1) using Karamata’s theorem and a moment inequality of von Bahr and Esseen (1965).

The following corollary to Theorem 3.2 may be obtained by writing \( n^{1/2}(\hat{\sigma}^2_n - \sigma^2_0)/\tau_n = U_n/V_n^{1/2} \) and applying the continuous mapping theorem.

**Corollary 3.1.** Under Assumption 3.1, if \( \alpha \in (1, 2) \) then
\[
\frac{\sqrt{n}(\hat{\sigma}^2_n - \sigma^2_0)}{\tau_n} \xrightarrow{w} W := U/V^{1/2}. \tag{3.25}
\]

Comparing Theorem 3.2 and Corollary 3.1, we see that normalizing by \( \tau_n \) has the effect of eliminating the unknown scaling sequence \((a_n)\). The limiting distribution \( W \) is unknown, but may in principle be approximated using subsampling, as we will see in the next section.

Having established the limiting behavior of the sample variance computed in the first step of the variance targeting method, we proceed to the second step in which the Gaussian quasi-likelihood is maximized subject to \( \sigma^2 = \hat{\sigma}^2_n \). Rearranging the Taylor expansion (3.10), we obtain
\[
na_n^{-1}(\hat{\lambda}_n - \lambda_0) = -J_n^{-1}K_nna_n^{-1}(\hat{\sigma}^2_n - \sigma^2_0) - J_n^{-1}na_n^{-1}Y_n, \tag{3.26}
\]

20
where $J_n$ is a.s. invertible for large enough $n$, and

$$Y_n = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \hat{\ell}_t(\nu_0)}{\partial \lambda}.$$  \hspace{1cm} (3.27)

Lemma 3.1 ensures that $J_n^{-1} K_n \overset{a.s.}{\rightarrow} J^{-1} K$ and $J_n^{-1} \overset{a.s.}{\rightarrow} J^{-1}$, while Theorem 3.2 ensures that, when $\alpha \in (1,2)$, $n a^{-1}_n (\hat{\sigma}_n^2 - \sigma_0^2) \overset{w}{\rightarrow} U$. It may also be shown that $Y_n$ converges to zero at a rate determined by the index $\kappa$ characterizing the tail behavior of the squared innovations $\eta^2$. This rate turns out to be faster than $n a^{-1}_n$, so $n a^{-1}_n Y_n$ vanishes, and the limit distribution of the normalized VTE concentrates on a line in $\mathbb{R}^{p+q+1}$.

**Theorem 3.3.** Under Assumption 3.1, if $\alpha \in (1,2)$ then

$$n a^{-1}_n (\hat{\nu}_n - \nu_0) \overset{w}{\rightarrow} \begin{bmatrix} 1 \\ -J^{-1} K \end{bmatrix} U \quad \text{and} \quad S_n \overset{w}{\rightarrow} \begin{bmatrix} 1 \\ -J^{-1} K \end{bmatrix} W,$$ \hspace{1cm} (3.28)

where $U$ and $W$ are the weak limits in Theorem 3.2 and Corollary 3.1.

Theorem 3.3 is proved in the Mathematical Appendix. A consequence of Theorem 3.3 is that, if $\alpha \in (1,2)$ and $J^{-1} K$ has nonzero elements, then the VTE is consistent at the rate $n^{1-1/\alpha}$. This compares unfavorably to the rate of consistency achieved by the QMLE, shown by Mikosch and Straumann (2006) under similar technical conditions to be $n^{1/2}$ if $\kappa > 2$ or $n^{-1/\kappa} L(n)$ if $\kappa \in (1,2)$, for some slowly varying $L$. No such discrepancy arises when $\alpha > 2$: as discussed earlier, Francq et al. (2011) showed that the VTE and QMLE are both consistent at the rate $n^{1/2}$. Since $\kappa > \alpha$, our results indicate that the efficiency loss associated with the VTE may be more severe in the heavy tailed case.

### 4 Inference via subsampling

We have seen that the Gaussian limit theory established for the VTE by Francq et al. (2011) breaks down in the heavy tailed case $\alpha \in (1,2)$. In this section we investigate the use of subsampling methods to assign robust confidence intervals to the VTE. We describe the construction of subsampling confidence intervals in Section 4.1, and numerically assess their finite sample performance in Section 4.2.
4.1 Construction of subsampling confidence intervals

Theorem 3.1 and Theorem 3.3 indicate that the elements of the self-normalized quantity $S_n$ converge weakly to what are typically nondegenerate limit distributions. As discussed by Politis et al. (1999, ch. 11), subsampling can be used to estimate these limit distributions without having to first estimate the rate of convergence of the sample variance. For $i = 1, \ldots, p + q + 1$, let $\nu^{(i)}_0$ denote the $i$th element of $\nu_0$, let $S^{(i)}_n$ denote the $i$th element of $S_n$, and let $F^{(i)}_n$ denote the cdf of $S^{(i)}_n$. Let $m = m_n$ be a sequence of block sizes satisfying $m \to \infty$ and $m/n \to 0$ as $n \to \infty$, and let $\hat{\nu}_{nt}$ be the VTE of $\nu_0$ computed using the block of observations $(y_t, \ldots, y_{t+m-1})$. The subsampled version of $S^{(i)}_n$ for this block of observations is given by

$$S^{(i)}_{nt} = \frac{1}{m} \left( \hat{\nu}^{(i)}_{nt} - \hat{\nu}^{(i)}_n \right) / \tau_{nt},$$

where $\tau_{nt}^2 = \frac{1}{m-1} \sum_{j=t}^{t+m-1} y_j^4$, and $\hat{\nu}^{(i)}_{nt}$ is the $i$th element of $\hat{\nu}_{nt}$. Our subsampling approximation to $F^{(i)}_n$ is

$$\hat{F}^{(i)}_n(x) = \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} \mathbf{1}(S^{(i)}_{nt} \leq x), \quad x \in \mathbb{R}. \quad (4.1)$$

For $s \in (0, 1)$, let $c^{(i)}_n(s) = \inf \{ x : \hat{F}^{(i)}_n(x) \geq s \}$, the $s$-quantile of $\hat{F}^{(i)}_n$. The next result provides an asymptotic justification for using $c^{(i)}_n(s)$ to form confidence intervals for $\nu^{(i)}_0$.

**Theorem 4.1.** Under Assumption 3.1, if $\alpha \in (1, 2) \cup (2, \infty)$ then for $i = 1, \ldots, p + q + 1$ and $s \in (0, 1)$ we have

$$\lim_{n \to \infty} \mathbb{P} \left( S^{(i)}_n \leq c_n(1-s) \right) = 1 - s,$$

provided that the $i-1$th element of $J^{-1}K$ is nonzero in the case that $i \geq 2$ and $\alpha \in (1, 2)$.

Theorem 4.1 is proved in the Mathematical Appendix. The condition on $J^{-1}K$ is made to rule out the possibility that the line on which the asymptotic distribution of the VTE concentrates is perfectly orthogonal to one of the coordinate axes. In this case we would have a degenerate limit for the corresponding element of $S_n$, making the usual justification for subsampling inapplicable. Excluding this possibility, Theorem 4.1 implies that the equal tailed confidence interval for $\nu^{(i)}_0$,

$$[\hat{\nu}^{(i)}_n - n^{-1/2} \tau_n c^{(i)}_n(1-s/2), \hat{\nu}^{(i)}_n + n^{-1/2} \tau_n c^{(i)}_n(s/2)], \quad (4.3)$$

has limiting coverage probability equal to the nominal level $1 - s$. A simple modification
of Theorem 4.1 may be used to show that the symmetric confidence interval for $\nu_0^{(i)}$,

$$[\hat{\nu}_n^{(i)} - n^{-1/2}\tau_n\vec{c}_n^{(i)}(1-s), \hat{\nu}_n^{(i)} + n^{-1/2}\tau_n\vec{c}_n^{(i)}(1-s)],$$

(4.4)

where $\vec{c}_n^{(i)}(1-s) = \inf\{x : \tilde{F}_n^{(i)}(x) \geq 1-s\}$ and $\tilde{F}_n^{(i)}(x) = \frac{1}{n-m+1}\sum_{t=1}^{n-m+1} 1(|S_t^{(i)}| \leq x),$

also has limiting coverage probability $1-s$ when $S_n^{(i)}$ has nondegenerate limit.

### 4.2 Finite sample performance

Here we present the results of a simulation study used to assess the finite sample performance of the subsampling confidence intervals discussed in the previous subsection. We focus on tests of the null hypothesis $\sigma^2 = \sigma_0^2$ based on equal tailed and symmetric confidence intervals for $\sigma_0^2$. This parameter is easier to deal with because we need only compute the estimator $\hat{\sigma}_n^2$, which is a simple average. Confidence intervals for the other GARCH parameters require numerical optimization in the second step quasi-maximum likelihood estimation, which can be burdensome in the context of a Monte-Carlo study of subsampling approximation.

We consider a variety of configurations of sample sizes, block sizes, nominal sizes, parameter values, and error distributions in our simulations. Specifically, we consider sample sizes $n \in \{500, 2500\}$, block sizes $m \in \{10, 25, 50, 100\}$, and nominal sizes $s \in \{0.05, 0.1\}$. We confine ourselves to the GARCH(1,1) model with $\sigma_0^2 = 1$, $b_{01} = 0.85$, and $a_{01} \in \{0.06, 0.1, 0.14\}$; note that this implies $a_{01} + b_{01} \in \{0.91, 0.95, 0.99\}$. We consider three innovation distributions: the $\mathcal{N}(0,1)$ distribution and the Student $t_p$ distribution with $p \in \{3, 5\}$ degrees of freedom, scaled to have unit variance. The $t_3$ distribution has infinite fourth moment, the $t_5$ distribution has finite fourth moment but infinite absolute fifth moment, and the $\mathcal{N}(0,1)$ distribution has finite moments of all orders. The sample sizes and parameter values we use are chosen to be relevant for typical financial applications: in Figure 3.1, the permitted values of $a_{01}$ and $b_{01}$ lie in roughly the same part of the parameter space as the estimates using daily DJIA returns. Automated selection rules for the block size $m$ have been proposed by Politis et al. (1999) and Kokoszka and Wolf (2004), but these are difficult to implement within a large Monte-Carlo study of subsampling approximation, so we instead consider subsampling performance over a range of fixed block sizes.
The results of our simulations are reported in Tables 5.1, 5.2 and 5.3. Each table includes results for a single innovation distribution. Alongside the parameter value \( a_{01} \), we report the corresponding value of \( \alpha \) implied by Kesten’s theorem. The numbers reported in the last four columns of each table are the null rejection rates for testing \( \sigma^2 = \sigma_0^2 \); that is, the rate at which our \((1 - s)\)-level subsampling confidence intervals for \( \sigma_0^2 \) do not include \( \sigma_0^2 \). These rates are computed over 2000 experimental replications.

We draw four main conclusions from the results reported in Tables 5.1, 5.2 and 5.3.

1. For many parameter configurations, null rejection rates are far above nominal size.

2. Null rejection rates are much greater at larger values of \( a_{01} \). When \( a_{01} = 0.14 \), so that \( a_{01} + b_{01} = 0.99 \), the null rejection rate is always above 0.25 even when innovations are Gaussian, and often rises above 0.5 when innovations have the (rescaled) \( t_3 \) distribution.

3. Null rejection rates are generally greater at larger values of \( \alpha \), but the value of \( a_{01} \) seems to be the more important factor determining rejection rates. This can be seen by comparing the null rejection rates in Table 5.1 with \( a_{01} = 0.14 \) and in Table 5.3 with \( a_{01} = 0.06 \). The former group of rejection rates are in all cases substantially greater than the latter group, but \( \alpha \) is equal to 1.54 in the former group and 1.44 in the latter group.

4. The tests based on symmetric confidence intervals tend to control size better than those based on equal tailed confidence intervals.

In view of the above findings, it is difficult for us to recommend the use of subsampling confidence intervals in practical implementations of the VTE for which the sample sizes and parameter values used in our simulations appear to be relevant. We did find more encouraging results for the symmetric confidence intervals at sample sizes of \( n = 50000 \) observations, but this seems of limited interest from a practical perspective.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Nominal size</th>
<th>Test type</th>
<th>$m = 10$</th>
<th>$m = 25$</th>
<th>$m = 50$</th>
<th>$m = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0 = 0.06$, $\alpha \approx 10.70$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.15</td>
<td>0.12</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.02</td>
<td>0.05</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.20</td>
<td>0.17</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.05</td>
<td>0.11</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.15</td>
<td>0.11</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.00</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.19</td>
<td>0.16</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.04</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>$a_0 = 0.10$, $\alpha \approx 4.54$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.23</td>
<td>0.20</td>
<td>0.20</td>
<td>0.26</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.06</td>
<td>0.10</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.29</td>
<td>0.26</td>
<td>0.25</td>
<td>0.30</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.14</td>
<td>0.16</td>
<td>0.19</td>
<td>0.23</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.23</td>
<td>0.17</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.27</td>
<td>0.22</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>$a_0 = 0.14$, $\alpha \approx 1.54$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.46</td>
<td>0.45</td>
<td>0.44</td>
<td>0.47</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.54</td>
<td>0.51</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.44</td>
<td>0.43</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.37</td>
<td>0.35</td>
<td>0.32</td>
<td>0.29</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.27</td>
<td>0.26</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.47</td>
<td>0.44</td>
<td>0.39</td>
<td>0.36</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.37</td>
<td>0.34</td>
<td>0.32</td>
<td>0.32</td>
</tr>
</tbody>
</table>

**Table 5.1:** Finite sample size of subsampling tests of $H_0 : \sigma^2 = \sigma_0^2$ for the GARCH(1,1) model with $\sigma_0^2 = 1$, $a_{01} \in \{0.06, 0.1, 0.14\}$, $b_{01} = 0.85$, and $\eta \sim \mathcal{N}(0,1)$.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Nominal size</th>
<th>Test type</th>
<th>$m = 10$</th>
<th>$m = 25$</th>
<th>$m = 50$</th>
<th>$m = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.01</td>
<td>0.05</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.19</td>
<td>0.18</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.05</td>
<td>0.10</td>
<td>0.14</td>
<td>0.19</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.16</td>
<td>0.12</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.20</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.02</td>
<td>0.05</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.20</td>
<td>0.20</td>
<td>0.21</td>
<td>0.29</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.07</td>
<td>0.11</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.25</td>
<td>0.25</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.14</td>
<td>0.18</td>
<td>0.21</td>
<td>0.27</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.22</td>
<td>0.19</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.03</td>
<td>0.05</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.27</td>
<td>0.25</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.08</td>
<td>0.10</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.43</td>
<td>0.45</td>
<td>0.47</td>
<td>0.52</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.41</td>
<td>0.43</td>
<td>0.45</td>
<td>0.49</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.55</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.49</td>
<td>0.50</td>
<td>0.50</td>
<td>0.53</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.32</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Symmetric</td>
<td>0.30</td>
<td>0.32</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Symmetric</td>
<td>0.40</td>
<td>0.41</td>
<td>0.40</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 5.2: Finite sample size of subsampling tests of $H_0: \sigma^2 = \sigma_0^2$ for the GARCH(1, 1) model with $\sigma_0^2 = 1$, $a_{01} \in \{0.06, 0.1, 0.14\}$, $b_{01} = 0.85$, and $\eta \sim t_5$.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Nominal size</th>
<th>Test type</th>
<th>Subsampling block size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$m = 10$</td>
</tr>
<tr>
<td>$a_0 = 0.06$, $\alpha \approx 1.44$</td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.05</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.16</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.10</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.14</td>
</tr>
<tr>
<td>$a_0 = 0.10$, $\alpha \approx 1.31$</td>
<td></td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.15</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.25</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.12</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.18</td>
</tr>
<tr>
<td>$a_0 = 0.14$, $\alpha \approx 1.08$</td>
<td></td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.53</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.57</td>
</tr>
<tr>
<td>2500</td>
<td>0.05</td>
<td>Equal tailed</td>
<td>0.36</td>
</tr>
<tr>
<td>2500</td>
<td>0.10</td>
<td>Equal tailed</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 5.3: Finite sample size of subsampling tests of $H_0 : \sigma^2 = \sigma_0^2$ for the GARCH(1,1) model with $\sigma_0^2 = 1$, $a_{01} \in \{0.06, 0.1, 0.14\}$, $b_{01} = 0.85$, and $\eta \sim t_3$.  

27
5 Conclusion

In this paper we have studied the VTE for the GARCH\((p, q)\) model, devoting particular attention to the heavy tailed case where \(E y_t^4 = \infty\). Prior research by Francq et al. (2011) has demonstrated that in the light tailed case the VTE typically has a larger asymptotic variance than the QMLE, but enjoys a degree of robustness toward model misspecification. Our results indicate that in the heavy tailed case, the loss of efficiency associated with the VTE can be more severe, with the rate of consistency slower than the rate achieved by the QMLE. Moreover, the limit distribution of the VTE is non-Gaussian, invalidating the use of confidence intervals based on Gaussian approximation. Though subsampling techniques can in principle be used to generate asymptotically valid confidence intervals for the VTE, our numerical simulations indicate that the finite sample performance of these confidence intervals can be extremely poor at realistic sample sizes and parameter values. Taken together, our results indicate that the VTE should be used with caution in applications where heavy tails are suspected, particularly if valid statistical inference (as opposed to mere consistency) is a primary concern.

A Mathematical appendix

Lemma A.1. Under Assumption 3.1, if \(\alpha \in (1, 2)\) then (3.23) and (3.24) hold for \(\delta > 0\).

Proof of Lemma A.1. To prove (3.23), we fix \(\gamma \in (\alpha, \kappa \wedge 2)\) and use Markov’s inequality to write

\[
\Pr \left( a_n^{-1} \sum_{t=1}^n h_t^2 (\eta_t^2 - 1) 1 \left( h_t^2 \leq \epsilon a_n \right) > \delta \right) \leq \left( \delta a_n \right)^{-\gamma} \mathbb{E} |M_n|^{\gamma},
\]

(A.1)

where \(M_n = \sum_{t=1}^n h_t^2 (\eta_t^2 - 1) 1 \left( h_t^2 \leq \epsilon a_n \right)\). Since \(\mathbb{E}(M_{n+1}|M_n) = M_n\) a.s. for each \(n \in \mathbb{N}\), and \(\gamma \in (1, 2)\), we may apply the moment inequality of von Bahr and Esseen (1965, Theorem 2) to obtain

\[
\mathbb{E} |M_n|^{\gamma} \leq 2 \sum_{t=1}^n \mathbb{E} \left| h_t^2 (\eta_t^2 - 1) 1 \left( h_t^2 \leq \epsilon a_n \right) \right|^{\gamma} = 2n \mathbb{E} \left( h_t^{2\gamma} 1 \left( h_t^2 \leq \epsilon a_n \right) \right) \mathbb{E} |\eta_t^2 - 1|^{\gamma}. \quad (A.2)
\]
Combining (A.1), (A.2) and (2.12), a consequence of Karamata’s theorem, we find that

$$
\limsup_{n \to \infty} \mathbb{P}(|U_n^* - U_{n\epsilon}^*| > \delta) = \limsup_{n \to \infty} \mathbb{P}
\left(C \left|a_n^{-1} \sum_{t=1}^{n} h_t^2 (\eta_t^2 - 1) 1 (h_t^2 \leq \epsilon a_n)\right| > \delta\right) \\
\leq \frac{2(\delta/C)^{-\gamma} c_h \alpha \epsilon^{\gamma-\alpha}}{\gamma - \alpha} \mathbb{E} |\eta^2 - 1|^{\gamma}.
$$

(A.3)

Since $\gamma < \kappa$ we have $\mathbb{E} |\eta^2 - 1|^{\gamma} < \infty$, and so we obtain (3.23) by letting $\epsilon \to 0$.

The proof of (3.24) is similar, but uses a more elementary inequality in place of von Bahr and Esseen (1965, Theorem 2). We fix $\gamma \in (\alpha/2, 1)$ and use Markov’s inequality to write

$$
\mathbb{P}\left(a_n^{-2} \sum_{t=1}^{n} y_t^4 1 (y_t^2 \leq \epsilon a_n) > \delta\right) \leq (\delta a_n^{-2})^{-\gamma} \mathbb{E} \left|\sum_{t=1}^{n} y_t^4 1 (y_t^2 \leq \epsilon a_n)\right|^{\gamma}.
$$

(A.4)

The inequality $(\sum_{i=1}^{n} c_i)^s \leq \sum_{i=1}^{n} c_i^s$, valid for $c_i \geq 0$ and $s \in [0, 1]$, yields

$$
\mathbb{E} \left|\sum_{t=1}^{n} y_t^4 1 (y_t^2 \leq \epsilon a_n)\right|^{\gamma} \leq n \mathbb{E} ((y_t^4)^{\gamma} 1 (y_t^2 \leq \epsilon a_n)) .
$$

(A.5)

Combining (A.4), (A.5) and (2.12), we obtain

$$
\limsup_{n \to \infty} \mathbb{P}(|V_n - V_{n\epsilon}| > \delta) = \limsup_{n \to \infty} \mathbb{P}\left(a_n^{-2} \sum_{t=1}^{n} y_t^4 1 (y_t^2 \leq \epsilon a_n) > \delta\right) \\
\leq \frac{\delta^{-\gamma} c_y \alpha \epsilon^{\gamma-\alpha}}{2\gamma - \alpha},
$$

(A.6)

and (3.24) follows by letting $\epsilon \to 0$.

\[\square\]

**Lemma A.2.** Under Assumption 3.1, if $\alpha \in (1, 2)$ then, for any $\epsilon > 0$, $T_\epsilon$ is continuous on a subset of $M_p(\mathbb{R}_+^d \setminus \{0\})$ inhabited by $N$ with probability one.

**Proof of Lemma A.2.** Let $B_\epsilon = \{x \in \mathbb{R}_+^d : \max(x_2, x_{p+1}) > \epsilon\}$ and $A_\epsilon = \{m \in M_p(\mathbb{R}_+^d \setminus \{0\}) : m(\partial B_\epsilon) = 0\}$, and consider a sequence $m_1, m_2, \ldots \in \mathbb{R}_+^d \setminus \{0\}$ with $m_n \xrightarrow{v} m \in A_\epsilon$. $B_\epsilon$ is bounded away from the origin and hence relatively compact, so a result of Resnick (1987, Proposition 3.13) ensures that, for all $n$ sufficiently large, we may label the points of $m_n$ and $m$ in $B_\epsilon$ by $(x_1^{(n)}, \ldots, x_k^{(n)}), (x_1, \ldots, x_k) \in B_\epsilon^k$, with $k = m(B_\epsilon)$, and for each
\[ i = 1, \ldots, k \text{ we have } x_i^{(n)} \to x_i. \] Since \( T_\epsilon(m_n) \) and \( T_\epsilon(m) \) depend only on the points \((x_1^{(n)}, \ldots, x_k^{(n)})\) and \((x_1, \ldots, x_k)\) respectively, it is clear from (3.20) that \( T_\epsilon(m_n) \to T_\epsilon(m) \). Thus \( T_\epsilon \) is continuous on \( A_\epsilon \). The limiting point process \( N \) satisfies the cluster representation (2.16) given in Theorem 2.2, so we may write \( P(N \notin A_\epsilon) = P(P_i Q_{ij} \in \partial B_\epsilon) \leq \sum_{i,j} P(P_i Q_{ij} \in \partial B_\epsilon) \). Since \( \sum_i \delta_{P_i} \) is Poisson with absolutely continuous intensity, each \( P_i \) must be a continuous random variable. Further, each \( P_i \) is independent of each \( Q_{ij} \). We thus have \( P(P_i Q_{ij} \in \partial B_\epsilon) = 0 \), and hence \( P(N \notin A_\epsilon) = 0 \).

**Lemma A.3.** Under Assumption 3.1, if \( \alpha \in (1, 2) \) then \( T_\epsilon(N) \overset{w}{\to} (U, V) \) as \( \epsilon \to 0 \).

**Proof of Lemma A.3.** Let \( \phi_\epsilon : \mathbb{R}^2 \to \mathbb{C} \) denote the characteristic function of \( (U_\epsilon, V_\epsilon) := T_\epsilon(N) \). We will show that \( \phi_\epsilon \) converges to some \( \phi : \mathbb{R}^2 \to \mathbb{C} \) that is continuous at the origin by showing that, as \( \epsilon \to 0 \), \( \phi_\epsilon(t) \) is Cauchy at all fixed \( t \in \mathbb{R}^2 \), and uniformly Cauchy in a neighborhood of the origin. It then follows from the Lévy continuity theorem that \( T_\epsilon(N) \) converges weakly to a pair of random variables with characteristic function \( \phi \).

Let \( A = \{ t \in \mathbb{R}^2 : |t_1| \vee |t_2| \leq 1 \} \). We will show that \( \phi_\epsilon \) is uniformly Cauchy on \( A \): for any \( \eta > 0 \), there exists an \( \epsilon > 0 \) such that

\[
\sup_{0 < u < v < \epsilon} \sup_{t \in A} |\phi_\epsilon(t) - \phi_u(t)| < \eta. \tag{A.8}
\]

Observe that, for any \( \delta > 0 \),

\[
|\phi_\epsilon(t) - \phi_u(t)| \leq \mathbb{E} (g(t, u, v)1 (|U_v - U_u| \vee |V_v - V_u| \leq \delta)) \\
+ \mathbb{E} (g(t, u, v)1 (|U_v - U_u| \vee |V_v - V_u| > \delta)), \tag{A.9}
\]

where we have defined

\[
g(t, u, v) = |\exp(i(t_1 U_v + t_2 V_v)) - \exp(i(t_1 U_u + t_2 V_u))| \tag{A.10}
\]

\[
= \sqrt{2 - 2 \cos(t_1(U_v - U_u) + t_2(V_v - V_u))}. \tag{A.11}
\]

For \( t \in A \) and small enough \( \delta > 0 \) we have \( g(t, u, v) \leq \sqrt{2 - 2 \cos(2\delta)} < \eta/2 \) when \( |U_v - U_u| \vee |V_v - V_u| \leq \delta \). Thus the first term on the right-hand side of (A.9) is less than \( \eta/2 \) for small enough \( \delta \). Fix such a \( \delta \); it remains to show that the second term is less than

30
\[ \eta/2 \] for small enough \( \epsilon > 0 \). Since \( 0 \leq g \leq 2 \), it suffices to find \( \epsilon > 0 \) such that
\[
\sup_{0 < u < v < \epsilon} \mathbb{P} (|U_v - U_u| \vee |V_v - V_u| > \delta) < \eta/4. \tag{A.12}
\]

Let \( \tilde{T}_{uv} = |T_v^{(1)} - T_u^{(1)}| \vee |T_v^{(2)} - T_u^{(2)}| \), where \( T_v^{(1)} \) and \( T_v^{(2)} \) denote the first and second components of \( T_v \). Lemma A.2 ensures that \( \tilde{T}_{uv} \) is continuous on a subset of \( M_p(\mathbb{R}_+^d \setminus \{0\}) \) inhabited by \( N \) with probability one. Thus, from the continuous mapping theorem and the weak convergence \( N_n \overset{w}{\rightarrow} N \) given by Theorem 2.2, we have \( \tilde{T}_{uv}(N_n) \overset{w}{\rightarrow} \tilde{T}_{uv}(N) \). Consequently,
\[
\mathbb{P} (|U_v - U_u| \vee |V_v - V_u| > \delta) = \lim_{n \rightarrow \infty} \mathbb{P} (\tilde{T}_{uv}(N_n) > \delta). \tag{A.13}
\]

Observe that
\[
\tilde{T}_{uv}(N_n) = |U_{nv}^* - U_{nu}^*| \vee |V_{nv} - V_{nu}| \leq |U_{nv}^* - U_{n}^*| + |U_{n}^* - U_{nu}^*| + |V_{n} - V_{nv}| + |V_{nv} - V_{nu}|. \tag{A.14}
\]

It follows from (A.14) and Lemma A.1 that \( \sup_{0 < u < v < \epsilon} \lim_{n \rightarrow \infty} \mathbb{P} (\tilde{T}_{uv}(N_n) > \delta) = 0 \).

In view of (A.13), this establishes the validity of (A.12), and we conclude that \( \phi_\epsilon \) is uniformly Cauchy on \( A \). A slight modification of the foregoing argument establishes that \( \phi_\epsilon(t) \) is Cauchy at each fixed \( t = (t_1, t_2) \): first choose \( \delta = \delta(t) \) small enough that
\[
g(t, u, v) \leq \sqrt{2 - 2 \cos(2\delta(|t_1| \vee |t_2|))} < \eta/2 \quad \text{when} \quad |U_v - U_u| \vee |V_v - V_u| \leq \delta,
\]
and then choose \( \epsilon \) small enough that (A.12) is valid. We conclude that \( T_\epsilon(N) \) converges weakly to a pair of random variables. One can show that the first of these random variables is \( \alpha \)-stable by arguing as in Davis and Hsing (1995, p. 898), and that the second is \( \alpha/2 \)-stable and positive by arguing as in Kokoszka and Wolf (2004, pp. 231–232).

\[ \square \]

\textit{Proof of Theorem 3.3.} In view of Theorem 3.2, Corollary 3.1, and the remarks preceding the statement of Theorem 3.3, it suffices for us to show that \( na_n^{-1} Y_n = o_p(1) \). Begin by writing
\[
na_n^{-1} Y_n = a_n^{-1} \sum_{t=1}^n \frac{\partial \ell_t(\nu_0)}{\partial \lambda} + a_n^{-1} \sum_{t=1}^n \left( \frac{\partial \ell_t(\nu_0)}{\partial \lambda} - \frac{\partial \tilde{\ell}_t(\nu_0)}{\partial \lambda} \right). \tag{A.15}
\]

The second term on the right-hand side of (A.15) is \( o_p(1) \) by item (iv) in the proof of Theorem 2.2 of Francq and Zakoïan (2004). It remains to show that the first term is
\( o_p(1) \). Fix \( i \in \{1, \ldots, p + q\} \). Differentiating (3.13) with respect to \( \lambda_i \) we see that

\[
\frac{\partial \ell_t(\nu_0)}{\partial \lambda_i} = (1 - \eta_i^2) \left( \frac{1}{h_t^2} \frac{\partial h_t^2(\nu_0)}{\partial \lambda_i} \right),
\]

(A.16)

and differentiating (3.12) we see further that \( 0 \leq h_t^{-2} \partial h_t^2(\nu_0)/\partial \lambda_i < \lambda_0^{-1} < \infty \). Fix \( \gamma \in (\alpha, \kappa \wedge 2) \), so that \( \mathbb{E}|1 - \eta|^\gamma < \infty \) and \( na_n^{-\gamma} \rightarrow 0 \) as \( n \rightarrow \infty \). Proceeding as in the proof of Lemma A.1 and applying the inequalities of Markov and of von Bahr and Esseen to the martingale difference sequence \( (\partial \ell_t(\nu_0)/\partial \lambda_i) \), we obtain

\[
P \left( \left| \sum_{t=1}^n \frac{\partial \ell_t(\nu_0)}{\partial \lambda_i} \right| > \delta \right) \leq 2(\delta \lambda_0)^{-\gamma} na_n^{-\gamma} \mathbb{E} |1 - \eta|^\gamma \overset{p}{\rightarrow} 0 \quad (A.17)
\]

as \( n \rightarrow \infty \), for any \( \delta > 0 \). Thus the first term on the right-hand side of (A.15) is also \( o_p(1) \), and \( na_n^{-1} Y_n = o_p(1) \) as claimed.

Proof of Theorem 4.1. This follows immediately from a result of Kokoszka and Wolf (2004, Theorem 1) if we can find sequences \( u_n \) and \( v_n \) with \( u_n/v_n = n^{1/2} \) such that \( S_n(i), u_n(\nu_n(i) - \nu_0(i)) \) and \( v_n \tau_n \) have atomless weak limits. In the light tailed case \( \alpha \in (2, \infty) \) we set \( u_n = n^{1/2} \) and \( v_n = 1 \), and the desired weak convergence then follows from Theorem 3.1 and the ergodic theorem. In the heavy tailed case \( \alpha \in (1, 2) \) we set \( u_n = na_n^{-1} \) and \( v_n = n^{1/2} a_n^{-1} \), and the desired weak convergence then follows from Theorem 3.2 and Theorem 3.3. The requirement that the \( (i - 1) \)th element of \( J^{-1}K \) is nonzero when \( i \geq 2 \) ensures that the weak limits of \( S_n(i) \) and \( u_n(\nu_n(i) - \nu_0(i)) \) given by Theorem 3.3 are not degenerate at zero.

\[ \square \]

References


