**ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS**

**DEPARTMENT OF ECONOMICS**

**WORKING PAPER SERIES** **7-2015**

# A CLT For Martingale Transforms With Infinite Variance

*Stelios Arvanitis Alexandros Louka*

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

# A CLT For Martingale Transforms With Infinite Variance

*Stelios Arvanitis*[*] *Alexandros Louka*[†]

Dept. of Economics, AUEB, Patision str. 76, Athens P.O. Box 10434, GREECE.

*Acknowledgement.* This research was funded by the Research Centre of the Athens University of Economics and Business, in the framework of "Research Funding at AUEB for Excellence and Extroversion".

**Abstract**

We provide a CLT for martingale transforms that holds even when the second moments are infinite. Compared to an analogous result in Hall and Yao [*Econometrica* **71** (2003) 285-317] we impose minimal assumptions and utilize the Principle of Conditioning to verify a modified version of the Lindeberg's condition. When the variance is infinite, the rate of convergence, which we allow to be matrix valued, is slower than $\sqrt{n}$ and depends on the rate of divergence of the truncated second moments. In many cases it can be consistently estimated. A major application concerns the characterization of the rate and the limiting distribution of the Gaussian QMLE in the case of GARCH type models with infinite fourth moments for the innovation process. The results are particularly useful in the case of the EGARCH(1,1) model as we show that the usual limit theory is still valid without any further parameter restrictions when we relax the assumption for finite fourth moments of the innovation process.

KEYWORDS: CLT , Generalized Domain of Attraction, Martingale Transform, Slowly Varying Second Moment, Stationarity, Ergodicity, Matrix Normalization, Linear Model, OLSE, Self-Normalized Wald Tests, Robustness, Conditional heteroskedasticity, Gaussian Quasi Likelihood, QMLE, Infinite Fourth Moments, Leptokurtosis, GARCH, EGARCH.

JEL: C10, C13.

## 1 Introduction

It is known that asymptotic normality with the usual $\sqrt{n}$ rate of the Gaussian QMLE for GARCH-type models breaks down when the fourth moment of the error process is infinite. Hall and Yao [3] obtained the asymptotic distribution of the QMLE in GARCH models by examining the asymptotic behavior of sums of the form $\sum_{t=1}^{n} \xi_t v_t$ where $(\xi_t)_{t\in\mathbb{N}}$ is an i.i.d. sequence and $(v_t)_{t\in\mathbb{N}}$ is a stationary ergodic sequence of essentially bounded random variables and the distribution of $\xi_1$ lies in the domain of attraction of an $\alpha$-stable distribution with $\alpha \in [1, 2]$. Specifically for the case where $\alpha = 2$, they derive asymptotic normality of the QMLE by deriving a CLT for the aforementioned sums assuming the finiteness of $\mathbb{E}\left\|v_1\right\|^{2+\epsilon}$ for some $\epsilon > 0$. While this holds trivially in the case of the GARCH$(p, q)$ case, it can be restrictive if used for other models such as the EGARCH$(1, 1)$ one.

The aim of this note is to derive a CLT for sums of martingale transforms allowing for infinite second moments, not only for the $(\xi_t)_{t\in\mathbb{Z}}$ process but also for $(v_t)_{t\in\mathbb{Z}}$ , for which we also allow for matrix normalization, thus extending the results of Hall and Yao [3]. Our method of proof relies on the Principle of Conditioning which allows us to easily extend the classical CLT's for sums of dependent variables (see e.g. Jakubowski [5]). Thus it suffices to show that a modified version of Lindeberg's condition, where expectations are replaced by conditional expectations with respect to the past, holds. We show that by relaxing the assumption of finite variance for $\xi_1$ or $v_1$ asymptotic normality can still be obtained albeit at a different rate that depends on the rate of divergence of the truncated second moments of each process. The latter can be readily estimated in many cases.

[†] Email address: stelios@aueb.gr
[†] Email address: loukaalex@aueb.gr

An easy application concerns the case of the linear model where the innovation process is a martingale transform and the regressor process is regularly varying with index $-2$. We obtain asymptotic normality for the OLSE in the context of this model, while allowing for matrix normalization rates. We show that in this context the standard self-normalized Wald test remains asymptotically robust.

A second application concerns the limit theory of the QMLE in the EGARCH$(1,1)$ model. The results in Wintenberger and Cai [12][1] show that the assumption that $\mathbb{E}\left\|v_1\right\|^{2+\epsilon}<+\infty$ would imply restrictions on the parameter space which can be avoided if the CLT we propose is used. Furthermore, it can be seen that the results in Wintenberger and Cai [12] will practically hold entirely if we only replace the assumption for finite fourth moments of the innovation process by slowly varying truncated fourth moments and as a consequence Wald-type tests are again asymptotically robust under this setting.

The structure of the remaining note is organized as follows. In the next section, the main result is presented regarding the CLT involving martingale transforms with slowly varying truncated second moments that could diverge slowly enough. In the second section we derive the aforementioned applications, and in the last section we briefly describe a potential extension.

## 2 A CLT with infinite variance

In this section we present the main probabilistic result which is a direct consequence of the Lindeberg–Feller Central Limit Theorem for triangular arrays and the Principle of Conditioning (see Jakubowski [5]). Given an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we are interested in the asymptotic behavior of the partial sums of the martingale transform process $\left(\xi_t u_t v_t\right)_{t\in\mathbb{N}}$. In what follows $\rightsquigarrow$ denotes weak convergence, $\|\cdot\|$ either the Euclidean or any matrix norm dependening on the context, $Tr$ transposition. Furthermore, the abbreviation slowly varying function implies slow variation at infinity.

We immediately specify our assumption framework for both the "innovation" process $\left(\xi_t\right)_{t\in\mathbb{N}}$ as well as the "scaling" processes $\left(v_t\right)_{t\in\mathbb{N}}$ and $\left(u_t\right)_{t\in\mathbb{N}}$. Consider the filtration $\left(\mathcal{F}_t\right)_{t\in\mathbb{N}}$ defined by $\mathcal{F}_t \triangleq \sigma\left(\xi_{t-j}u_{t-j}v_{t-j}, j>0\right)$. Our first assumption specifies the basic measurability properties of the aforementioned processes.

**Assumption 1.** *$(\xi_t)_{t\in\mathbb{N}}$ is a stationary sequence of random variables where $\xi_t$ is independent of $\mathcal{F}_t$ for all $t\in\mathbb{N}$. $(u_t)_{t\in\mathbb{N}}$ is a stationary and ergodic sequence of random variables adapted to $(\mathcal{F}_t)_{t\in\mathbb{N}}$. $(v_t)_{t\in\mathbb{N}}$ is a stationary ergodic $\mathbb{R}^q$-valued process adapted to $(\mathcal{F}_t)_{t\in\mathbb{N}}$. $u_t$ is independent of $v_t$ for all $t\in\mathbb{N}$.*

*Remark* 1. The first condition is obviously satisfied if the $\xi$ sequence is i.i.d. The remaining ones preclude non-trivial mixtures as weak limits for the martingale transform. Such cases cannot be handled in the present context. In the context of a linear model, with martingale difference innovations, $(u_t^2)_{t\in\mathbb{N}}$ could be perceived as a conditional volatility process in the context of a stationary ergodic GARCH type model with finite second moment. $(v_t)_{t\in\mathbb{N}}$ could specify the regressors process. In the context of the quasi-likelihood process for a stationary ergodic GARCH type model, $\xi_t$ would be related to the squared innovations, $u_t^2$ would be degenerate to 1, while $v_t$ would be associated to the stationary and ergodic solutions of the derivatives of the stochastic recursions that specify the conditional variance process.

**Assumption 2.** *$\mathbb{E}\left(\xi_0\right)=0$ and $H_\xi:\mathbb{R}_{++}\to\mathbb{R}_{++}$, $H_\xi(x):=\mathbb{E}\left[\xi_1^2 1\left\{|\xi_1|\leq x\right\}\right]$, is a slowly varying function at infinity. Furthermore, $\mathbb{E}\left(u_1^2\right)<+\infty$. For any $\lambda\in\mathbb{R}^q$ and $\lambda\neq 0_q$ suppose that $H_{\lambda v}(x):=\mathbb{E}\left[\left(\lambda^{Tr}v_1\right)^2 1\left\{\left|\lambda^{Tr}v_1\right|\leq x\right\}\right]$ is a slowly varying function at infinity with $\inf_{x\in\mathbb{R}_{++},\lambda} H_{\lambda v}(x)>0$.*

*Remark* 2.

(a) The second assumption directly implies that $\left(\xi_t\right)_{t\in\mathbb{N}}$ lies in the *generalized* domain of attraction to the normal distribution (see Theorem 2.6.5 of Ibragimov and Linnik [4]). The classical cases are recovered when $H_\xi$ converges to a positive constant. For a simple alternative example consider the case where the distribution of $\xi_0$ is the Student's $t$-distribution with 2 degrees of freedom, where a

[1] See also Wintenberger [11].

simple calculation yields $H_{\xi}(x) \backsim \frac{1}{2}\log x$ as $x \to \infty$. Analogous qualifications hold for the distributions of $u_0$ and $\lambda^{Tr}v_0$. More specifically the former lies in the *normal* domain of attraction to the normal.

(b) For the definition of $H_{\lambda v}(x)$ we have that due to Meerschaert [6] (see Section 2) this is equivalent to the existence of a slowly varying operator valued-function $K : \mathbb{R}_{++} \to \mathrm{GL}(\mathbb{R}^q)$ such that $H_{\lambda v}(x) = \|K(x)\lambda\|^2$ for any $\lambda$ as above, hence $K(x) = \mathbb{E}\left[v_1 v_1^{Tr} 1\left\{\left|\left(v_1 v_1^{Tr}\right)_{i,j}\right| \leq x,\ i,j = 1,\dots,q\right\}\right]$. The condition on the infimum of $H_{\lambda v}$ implies that $K$ remains asymptotically positive definite.

(c) Section VII.7 of Feller [2] (specifically Theorem 2) implies that for any $\lambda$ as above, there exist slowly varying sequences $r_n$, $\kappa_{\lambda n}$ such that $\begin{cases} r_n \backsim H_{\xi}\left(\sqrt{n r_n}\right) \\ \kappa_{\lambda n} \backsim H_{\lambda v}\left(\sqrt{n \kappa_{\lambda n}}\right) \end{cases}$ and that $n\mathbb{P}\left(\left|\lambda^{Tr}v_1\right| > \sqrt{n\kappa_n}\right)$ and $n\mathbb{P}\left(|\xi_1| > \sqrt{n r_n}\right)$ are asymptotically negligible. E.g. for the aforementioned case of the $t_2$ distribution we can easily see that $r_n = 2\log n$. Hence there exists a sequence of slowly varying and positive definite integer valued matrices $(K_n)_{n\in\mathbb{N}}$, $K_n \in \mathrm{GL}(\mathbb{N}^q)$, such that due to the finiteness of $q$, Assumptions 1, 2 and Lemmata 1, 2 in the Appendix we have that

$$\left\|\frac{1}{n}\sum_{t=1}^{n} u_t^2 v_t v_t^{Tr} - \mathbb{E}\left(u_1^2\right) K_n\right\| = o_p(1).$$

**Assumption 3.** *Consider the slowly varying sequences* $(r_n)_{n\in\mathbb{N}}$*,* $(\kappa_{\lambda n})_{n\in\mathbb{N}}$ *defined by* $\begin{cases} r_n = H_{\xi}\left(\sqrt{n r_n}\right) \\ \kappa_{\lambda n} = H_{\lambda v}\left(\sqrt{n\kappa_{\lambda n}}\right) \end{cases}$.

*If for some* $\lambda$, $\kappa_{\lambda n} \to +\infty$ *while* $r_n$ *converges, then assume that*

$$\sum_{t=1}^{n} \mathbb{C}ov\left(v_{1,\lambda n}^2, v_{t+1,\lambda n}^2\right) = o\left(n\kappa_{\lambda n}^2\right),$$

*where* $v_{t,\lambda n} \doteqdot \lambda^{Tr}v_t 1_{\{|\lambda^{Tr}v_t| \leq \sqrt{n r_n}\}}$.

*If for some* $\lambda$, $\kappa_{\lambda n} \to +\infty$, *while* $r_n$ *diverges then assume that* $H_{\xi}$ *lies in the Zygmund class of slowly varying functions with* $xH'_{\xi}(x)$ *Lebesgue almost everywhere bounded. Also as* $n \to \infty$ *there exists a* $K \geq 0$ *such that*

$$\log\sqrt{\kappa_{\lambda n}} = o(r_n),$$

$$\frac{\log^2\sqrt{n\kappa_{\lambda n}}\, h_{\lambda v}\left(\sqrt{n\kappa_{\lambda n}}\right)}{\kappa_{\lambda n} r_n^2} + \frac{\int_K^{\sqrt{n\kappa_{\lambda n}}} x\left(2\left|\log|x|\right| + 1\right) h_{\lambda v}(x)\,dx}{n\kappa_{\lambda n}^2 r_n^2} \to 0.$$

*Also,*

$$\sum_{t=1}^{n} \mathbb{C}ov\left(v_{1,\lambda n}^2\left|\log|v_{1,\lambda n}|\right|, v_{t+1,\lambda n}^2\left|\log|v_{i+1,\lambda n}|\right|\right) = o\left(n\kappa_{\lambda n}^2 r_n^2\right),$$

*where* $h_{\lambda v}$ *is defined by* $h_{\lambda v}(x) := x^2\mathbb{P}\left(\left|\lambda^{Tr}v_1\right| > x\right)$ *and* $\mathbb{E}\left(u_1^2\left|\log|v_1|\right|\right) < +\infty$.

*Remark* 3.

(a) Both the covariance summability conditions above are implied by uniform mixing with summable coefficients for $(v_t)_{t\in\mathbb{N}}$ via Theorem 1.4.(b) of Rio [7].

(b) The Zygmund class hypothesis means that the convergent function in the Karamata representation $H_{\xi}$ is constant (see Theorem 1.5.5 of Bingham et. al. [1]). Examples that satisfy the almost everywhere boundness $xH'_{\xi}(x)$ are iterated logarithms or appropriate quotients of iterated logarithms e.t.c.

(c) The conditions describing relations between between $r_n$ and $\kappa_n$ ensure that when both diverge, the former does so appropriately faster than the former so that quantities such as $\mathbb{E}\left[\left(\lambda^{Tr}v_0\right)^2\left|\log\left|\lambda^{Tr}v_0\right|\right|1\left\{\left|\lambda^{Tr}v_0\right|\leq\sqrt{n\kappa_n}\right\}\right]^2$ do not diverge very *quickly*.

(d) The final condition is readily implied by $\mathbb{E}\left(u_1^{2+\epsilon}\right)<+\infty$ for some $\epsilon>0$.

We are now ready to present the CLT.

**Theorem 1.** *Under Assumptions 1-3*

$$\frac{K_n^{-\frac{1}{2}}}{\sqrt{nr_n\mathbb{E}\left(u_1^2\right)}}\sum_{i=1}^{n}\xi_iu_iv_i\rightsquigarrow N\left(0_q,Id_q\right).$$

*Remark* 4.

(a) Due to the fact that $q\geq 1$ the result essentially employs a matrix rate that contains information on both (possibly divergence rates of) $H_\xi$ and $K$. Hence in its generality the rate cannot be factorized as $\sqrt{nr_n}A_n$ where $A_n$ is a convergence real matrix. In such cases and/or in cases where $r_n$ diverges, the rate is slower than the classical $\sqrt{n}$.

(b) The classical result is obtained only when $r_n$ and $K_n$ converge.

(c) When $r_n$ diverges the result can be preceived as a generalization of the one in Theorem 2.1.(b) of Hall and Yao [3] the proof of which uses the assumption that $\mathbb{E}\left\|u_1v_1\right\|^{2+\epsilon}$ for some $\epsilon>0$, that implies a uniform integrability property, in order to obtain a similar result at least in the context of the limit theory of the QMLE in GARCH$(p,q)$ model. This is due to the fact that we are able to obtain the same limit even in cases where $\mathbb{E}\left\|u_1v_1\right\|^2=+\infty$. Even when $K_n$ converges the result is essentially obtained under the weaker condition $\mathbb{E}\left\|u_1v_1\right\|^2<+\infty$, something that could be very useful for the weakening of restrictions enforced on parameter spaces in order to obtain the limit theory of the QMLE is such type of models. We prove the result using a modified version of the classical martingale CLT where expectations are replaced with conditional expectations based on the Principle of Conditioning, as in Theorem 1.1 of Jakubowski [5].

(d) Finally note that due to Lemmata 1 and 2 the rates $r_n$ and $K_n$ that appear in the rate of convergence are asymptotically equivalent to $\frac{1}{n}\sum_{t=1}^{n}\xi_t^2$ and $\frac{1}{n}\sum_{t=1}^{n}v_tv_t^{Tr}$, a fact that could be very useful in the construction of robust inferential procedures based on self-normalization.

*Proof.* Given Remark 2 it suffices to prove the result for $q=1$, whence $K_n=\kappa_n$. Assume without loss of generality that $\mathbb{E}\left(u_1^2\right)=1$. Define $\xi_{i,n}\doteqdot\xi_i1_{\{|\xi_i|\leq\sqrt{nr_n}\}}$ and $v_{i,n}\doteqdot v_i1_{\{|v_i|\leq\sqrt{n\kappa_n}\}}$. Next notice that

$$\mathbb{P}\left(\frac{1}{\sqrt{nr_n\kappa_n}}\left|\sum_{i=1}^{n}\left[\xi_iu_iv_i-\xi_{i,n}u_iv_{i,n}\right]\right|>\delta\right)$$

$$\leq\mathbb{P}\left(\frac{1}{\sqrt{nr_n\kappa_n}}\sum_{i=1}^{n}\left|\xi_i-\xi_{i,n}\right|\left|u_i\right|\left|v_i\right|>\frac{\delta}{2}\right)+\mathbb{P}\left(\frac{1}{\sqrt{nr_n\kappa_n}}\sum_{i=1}^{n}\left|v_i-v_{i,n}\right|\left|u_i\right|\left|\xi_{i,n}\right|>\frac{\delta}{2}\right)$$

$$\leq n\left[\mathbb{P}\left(\left|\xi_1\right|>\sqrt{nr_n}\right)+\mathbb{P}\left(\left|v_1\right|>\sqrt{n\kappa_n}\right)\right]=o\left(1\right)$$

due to Assumption 2. Furthermore, $\frac{\mathbb{E}(\xi_{i,n})}{\sqrt{nr_n\kappa_n}}\sum_{i=1}^{n}\left|u_iv_{i,n}\right|\overset{L_1}{\to}0$, due to the fact that

$$\frac{1}{\sqrt{\kappa_n}}\mathbb{E}\left|u_iv_{i,n}\right|\leq\mathbb{E}\left|u_1\right|\mathbb{E}\max\left\{1,\frac{1}{\kappa_n}v_{i,n}^2\right\}\leq 1,$$

so that $\frac{1}{\sqrt{\kappa_n}}\sum_{i=1}^{n}\mathbb{E}\left|u_i v_{i,n}\right|\leq n$ and $\frac{\sqrt{n}}{\sqrt{r_n}}\mathbb{E}\left(\xi_{1,n}\right)\to 0$ which is implied by Theorem 2 in Section VIII.9 of Feller [2]. Hence it suffices to show that $\sum_{i=1}^{n}X_{i,n}\rightsquigarrow N\left(0,1\right)$, where $X_{i,n}\doteqdot\frac{1}{\sqrt{nr_n\kappa_n}}\left(\xi_{i,n}-\mathbb{E}\xi_{i,n}\right)u_i v_{i,n}$ is a martingale difference array w.r.t. $\left(\mathcal{F}_{i,n}\right)_{(i,n)\in\mathbb{N}^2,i\leq n}\doteqdot\left(\mathcal{F}_i\right)_{i\in\mathbb{N}}$. Since $\mathbb{E}\left(\xi_1\right)=0$,

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left(X_{i,n}^2|\mathcal{F}_{i-1}\right) &= \frac{1}{nr_n\kappa_n}\sum_{i=1}^{n}u_i^2v_{1,n}^2\mathbb{E}\left[\left(\xi_{1,n}-\mathbb{E}\xi_{1,n}\right)^2\right]\\ &= \left[\frac{1}{r_n}\mathbb{E}\xi_{1,n}^2-\frac{1}{r_n}\left(\mathbb{E}\xi_{1,n}\right)^2\right]\frac{1}{n\kappa_n}\sum_{i=1}^{n}u_i^2v_{1,n}^2\\ &= 1+o_p(1)\end{aligned}$$

by Lemma 1. Consider $V_{n,\varepsilon}:=\sum_{i=1}^{n}\mathbb{E}\left(X_{i,n}^2 1\left\{\left|X_{i,n}\right|>\varepsilon\right\}/\mathcal{F}_{i-1}\right)$. Suppose first that $\kappa_n$ converges. We have that

$$\frac{1}{r_n\kappa_n}\mathbb{E}\left(u_1^2v_{1,n}^2\mathbb{E}\left[\left(\xi_{1,n}-\mathbb{E}\xi_{i,n}\right)^2 1\left\{\left|\xi_{1,n}-\mathbb{E}\xi_{1,n}\right|>\varepsilon\frac{\sqrt{nr_n\kappa_n}}{|u_1|\left|v_{1,n}\right|}\right\}/\mathcal{F}_1\right]\right)$$

$$\leq\frac{1}{r_n\kappa_n}\mathbb{E}\left(u_1^2v_{1,n}^2 1\left\{|u_1|\left|v_{1,n}\right|>\varepsilon\sqrt{\kappa_n}\right\}\mathbb{E}\left[\xi_1^2 1\left\{\varepsilon\frac{\sqrt{nr_n\kappa_n}}{|u_1|\left|v_{1,n}\right|}<|\xi_1|\leq\sqrt{nr_n}\right\}/\mathcal{F}_1\right]\right)+c_n,$$

and the expression inside the expectation in the last display is $\mathbb{P}$ a.s. less than or equal to $\frac{u_1^2v_1^2}{\kappa_{[n]}}$ where $[n]\left(\omega\right)\doteqdot\inf_{n\in\mathbb{N}}\left\{u_1^2\left(\omega\right)v_1^2\left(\omega\right)\leq n\kappa_n\right\}$, and this last inequality holds, since $\frac{1}{r_n}\left|H_\xi\left(\sqrt{nr_n}\right)-H_\xi\left(\varepsilon\frac{\sqrt{nr_n\kappa_n}}{|u_1|\left|v_{1,n}\right|}\right)\right|\leq 1$ for any $\varepsilon\sqrt{\kappa_n}<|u_1|\left|v_1\right|\leq\sqrt{n\kappa_n}$. Obviously $\mathbb{E}\left(\frac{u_1^2v_1^2}{\kappa_{[n]}}\right)$ exists due to the dominated convergence theorem and the fact that $H_\xi$ is slowly varying. The latter also implies that $c_n=o\left(1\right)$ and thereby we have that $V_{n,\varepsilon}\overset{L_1}{\to}0$. When $\kappa_n$ diverges yet $r_n$ converges, then a similar argument to the above, without the use of the $1\left\{|u_1|\left|v_{1,n}\right|>\varepsilon\sqrt{\kappa_n}\right\}$ truncation also implies that $V_{n,\varepsilon}\overset{L_1}{\to}0$. When $\kappa_n$ and $r_n$ both diverge, then due to Assumption 3 we have that for any $x\in\mathbb{R}$

$$\left|\frac{H_\xi(\sqrt{nr_n}\varepsilon\exp(x))}{H_\xi(\sqrt{nr_n}\varepsilon)}-1\right|\leq\left|H_\xi'(\sqrt{nr_n}\varepsilon\exp(x^*))\sqrt{nr_n}\varepsilon\exp(x^*)x\right|$$

where $x^*$ lies between $x$ and $0$. Now letting $x=\log\sqrt{\kappa_n}-\log|u_1|-\log\left|v_{1,n}\right|$ we obtain

$$\frac{1}{n\kappa_n}\sum_{i=1}^{n}u_i^2v_i^2 1\left\{|v_i|\leq\sqrt{n\kappa_n}\right\}\left|1-\frac{H_\xi\left(\varepsilon\frac{\sqrt{nr_n\kappa_n}}{|u_i|\left|v_{i,n}\right|}\right)}{H_\xi\left(\sqrt{nr_n}\right)}\right|$$

$$\leq\frac{1}{n\kappa_n r_n}(1+o(1))C\sum_{i=1}^{n}u_i^2v_i^2\left(\log\sqrt{\kappa_n}+\left|\log|v_i|\right|+\left|\log|u_i|\right|\right)1\left\{|v_i|\leq\sqrt{n\kappa_n}\right\}.$$

We have first that $\log\sqrt{\kappa_n}=o\left(r_n\right)$, while for arbitrary $\delta>0$ due to Chebyshev's inequality

$$\mathbb{P}\left(\frac{1}{n\kappa_n r_n}\sum_{i=1}^{n}v_i^2\left|\log|v_i|\right|1\left\{|v_i|\leq\sqrt{n\kappa_n}\right\}>\delta\right)$$

$$\leq\frac{1}{n\kappa_n^2r_n^2\delta^2}\mathbb{E}\left[v_0^2\left|\log|v_0|\right|1\left\{|v_0|\leq\sqrt{n\kappa_n}\right\}\right]^2-\frac{\left(\mathbb{E}\left[v_0^2\left|\log|v_0|\right|1\left\{|v_0|\leq\sqrt{n\kappa_n}\right\}\right]\right)^2}{n\kappa_n^2r_n^2\delta^2}$$
$$+\frac{2\sum_{i=1}^{n}\left(n-i+1\right)\mathbb{C}ov\left(v_{1,n}^2\left|\log\left|v_{1,n}\right|\right|,v_{i+1,n}^2\left|\log\left|v_{i+1,n}\right|\right|\right)}{n^2\kappa_n^2r_n^2\delta^2}$$

and

$$\frac{1}{n\kappa_n^2 r_n^2}\mathbb{E}\left[v_0^2\left|\log|v_0|\right|1\left\{|v_0|\leq\sqrt{n\kappa_n}\right\}\right]^2$$

$$=\frac{1}{n\kappa_n^2 r_n^2}\int_0^{\sqrt{n\kappa_n}}\left(x^2\left|\log|x|\right|\right)^2 d\mathbb{P}\left(|v_0|\leq x\right)$$

$$=\frac{\log^2\sqrt{n\kappa_n}h_v\left(\sqrt{n\kappa_n}\right)}{\kappa_n r_n^2}+\frac{2}{n\kappa_n^2 r_n^2}\left[C+\int_K^{\sqrt{n\kappa_n}}x\left(2\left|\log|x|\right|+1\right)h_v(x)\,dx\right]\to 0$$

due to Assumption 2. This along with repeated use of Lemma 2 imply that $V_{n,\varepsilon}\xrightarrow{p}0$. In all cases the result follows by Theorem 1.1 of Jakubowski [5]. □

## 3 Application: Examples of Asymptotic Normality for Cases of M-Estimators and Robustness for Self-Normalized Wald Tests

In the present section we are occupied with two simple examples that utilize Theorem (1). Those concern the limit theory of the OLSE and the QMLE estimator, as well as the asymptotic behavior of the relevant Wald tests, in the context of a linear model and a GARCH type one respectively.

Consider first the process $(y_t)_{t\in\mathbb{N}}$ defined by

$$y_t = v_t^{Tr}\theta_0+\xi_t u_t,$$

where $\theta_0\in\mathbb{R}^q$. The sample is the random element $(y_t, v_t)_{t=1}^n$, and we are interested in the asymptotic behavior of the OLSE for $\theta_0$, i.e.

$$\theta_n=\left(\sum_{t=1}^n v_t v_t^{Tr}\right)^{-1}\sum_{t=1}^n v_t y_t.$$

We readily obtain the following result.

**Proposition 1.** *Suppose that Assumptions 1 and 2 hold. Then*

$$\frac{K_n}{\sqrt{nr_n\mathbb{E}\left(u_1^2\right)}}\left(\theta_n-\theta_0\right)\rightsquigarrow N\left(0_q, Id_q\right)$$

*and*

$$n\left(\theta_n-\theta_0\right)^{Tr}\frac{\sum_{t=1}^n v_t v_t^{Tr}}{\sum_{t=1}^n\varepsilon_t^2}\left(\theta_n-\theta_0\right)\rightsquigarrow\chi_q^2,$$

*where* $\varepsilon_t := y_t - v_t^{Tr}\theta_n$, $t=1,\ldots,n$.

*Proof.* Trivial given Theorem 1, Remark 4, Lemma 2 and the fact that

$$\left|\frac{1}{n}\sum_{t=1}^n\left[\left(\xi_t u_t\right)^2-\left(y_t-v_t^{Tr}\theta_n\right)^2\right]\right|\leq\left\|\theta_n-\theta_0\right\|^2 O_p(1).$$

□

*Remark* 5. The result allows for matrix nomalization for the estimator when the regressors have diverging second moments. The second part implies that despite the non-standard rate of the OLSE, the classical self-normalized Wald test (for a relevant hypothesis structure) remains robust in the context of the present non-standard framework.

Now, an implication of Theorem 1 is that we can obtain the usual asymptotic results for the limit theory of the Gaussian QMLE in conditionally heteroskedastic models by replacing the assumption of finite fourth moments of the innovation process with the requirement that the truncated fourth moment is slowly varying at infinity without imposing any restrictive assumptions on the $u_i$ process. This is particularly useful in determining the limit theory of the QMLE in the EGARCH $(1,1)$ model. Remeber that the latter is defined by the system of recursions

$$X_t = \sigma_t Z_t,$$
$$\log \sigma_t^2 = a_0 + \beta_0 \log \sigma_{t-1}^2 + \gamma_0 Z_t + \delta_0 |Z_t|,$$

where the $(Z_t)_{t\in\mathbb{N}}$ are i.i.d. random variables such that $\mathbb{E}(Z_0) = 0$ and $\mathbb{E}(Z_0^2) = 1$.

Wintenberger and Cai [12] (see also Wintenberger [11]) derive the limit theory of the Gaussian QMLE by imposing, among others, restrictions on the parameter space along with ones on the distribution of $Z_0$ that permit existence and continous invertibility of ergodic solutions of stochastic recursions closely related to the above, conditions that imply the identifiability of $\theta_0 = (a_0, \beta_0, \gamma_0, \delta_0)^{Tr}$, the linear independence of derivatives, moment existence for several quantities e.t.c. In order to economize on space we do not describe those analytically. We simply note that if we allow for $\mathbb{E}(Z_0^4) = +\infty$ then an application of a result a la Hall and Yao [3] for the derivation of the limit theory of the QMLE for $\theta_0$, say $\theta_n$, would impose stricter than needed conditions on the parameter space in order to derive the asymptotic normality result.

The use of the theorem above avoids those strict conditions. In order to do so in this context, we specify $\xi_t$ as $Z_0^2 - 1$, $u_t$ as identically $1$ and $u_t$ as the unique ergodic solution of the gradient of the log-volatility recursion evaluated at $\theta_0$, whence $q = 4$. Then, using the notation of Wintenberger and Cai [12], Theorem 8 therein can be generalized as follows:

**Proposition 2.** *Suppose that assumptions of Theorem 7 in Wintenberger and Cai [12] are satisfied,* $H_\xi(x) := \mathbb{E}(Z_0^2 - 1)^2 1\{|Z_0^2 - 1| \leq x\}$ *is slowly varying at infinity and* $\mathbb{E}[\beta_0 - 2^{-1}(\gamma_0 Z_0 + \delta_0 |Z_0|)^2] < 1$. *Then*

$$\frac{\sqrt{n}}{\sqrt{r_n}}(\theta_n - \theta_0) \rightsquigarrow N(0, \mathbf{B}^{-1})$$

$\mathbf{B}^{-1}$ *is defined in Theorem 7 in Wintenberger and Cai [12].*

*Proof.* It follows exactly as in Wintenberger and Cai [12] where instead of using the CLT for square integrable martingale difference sequences we apply Theorem 1. □

*Remark* 6. (a) Wintenberger and Cai [12] show that the existence $\mathbf{B}$ is implied by the condition

$$\mathbb{E}[\beta_0 - 2^{-1}(\gamma_0 Z_0 + \delta_0 |Z_0|)^2] < 1.$$

This corresponds to the convergence of $K_n$ in our context. Hence the application of this special case of the present theorem generalizes the result of the aforementioned paper without the imposition of any further restrictions on the parameter space since it avoids the use of a condition of the form $\mathbb{E}\|v_1\|^{2+\epsilon} < \infty$. The generalization stems from the hypothesis that the fourth moment of $Z_0$ may diverge in a slowly varying fashion, in which case the limit theory will differ from the usual one, only due to the fact that the rate now becomes slower by a factor that represents this divergence. For example when $Z_0$ follows the $t_4$ distribution, then we have that $r_n = \frac{2}{3}\log n$. Obviously when $\mathbb{E}(Z_0^4) < +\infty$ we recover the standard result.

(b) A consequence of Proposition 2, and results implied by Proposition 5.2.12 of Straumann [8] is that

$$n(\theta_n - \theta_0)^{Tr} B_n (\theta_n - \theta_0) \rightsquigarrow \chi_2^q,$$

where $B_n$ is as in Remark 5.6.2 of Straumann (when the $\frac{1}{4}$ term is omitted). Analogously this implies that the standard Wald test (for the appropriate hypothesis structures) in this model is robust even if the requirement for finite fourth moments of the innovation process is replaced by Assumption 2.

## 4 Further Research

A possible route for further research would be the case where the partial sums of the scaling processes of the transform are only allowed, when appropriately normalized, to converge in distribution. This cannot be handled in the present framework and it could perhaps require a major generalization of the Principle of Conditioning of Jakubowski [5]. Such extensions would generalize the results of Wang [10] and be useful in a greater spectrum of econometric applications such as non linear co-integration e.t.c.

## Appendix

In what follows $q = \lambda = 1$.

**Lemma 1.** *Under Assumptions 1 and 2, for $q = 1$ and for any $\delta > 0$*

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n} v_i^2 - 1\right| \geq \delta\right) \to 0.$$

*Proof.* If $\kappa_n$ converges the result follows readily from the LLN of Birkhoff. If it diverges note that $\frac{1}{n\kappa_n}\sum_{i=1}^{n} v_{i,n}^2 - 1 = \frac{1}{n\kappa_n}\sum_{i=1}^{n}\left(v_{i,n}^2 - \mathbb{E}v_{i,n}^2\right)$ since $\mathbb{E}v_{i,n}^2 = H_v\left(\sqrt{n\kappa_n}\right)$. Then

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n}\left(v_i^2 - \mathbb{E}v_{i,n}^2\right)\right| > \delta\right) \leq \mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n}\left(v_{i,n}^2 - \mathbb{E}v_{i,n}^2\right)\right| > \delta\right) + \mathbb{P}\left(\sum_{i=1}^{n} v_i^2 \neq \sum_{i=1}^{n} v_{i,n}^2\right)$$

$$\leq \frac{\mathbb{E}v_{1,n}^4}{n\kappa_n^2\delta^2} - \frac{(\mathbb{E}v_{1,n}^2)^2}{n\kappa_n^2\delta^2} + \frac{2\sum_{i=1}^{n}(n-i+1)\,\mathbb{C}ov\left(v_{1,n}^2, v_{i+1,n}^2\right)}{n^2\kappa_n^2\delta^2} + n\mathbb{P}\left(|v_1| > \sqrt{n\kappa_n}\right),$$

where the last display follows from the inequality of Chebychev and stationarity. Due to the covariance summability condition in Assumption 2 the second term of the last display converges to zero. Also, the third term converges to $0$ due to Assumption 2 and by definition of $\kappa_n$. Regarding the first term we have that

$$\frac{\mathbb{E}v_{1,n}^4}{n\kappa_n^2} = -n\mathbb{P}\left(|v_1| > \sqrt{n\kappa_n}\right) + \frac{4}{n\kappa_n^2}\int_0^{\sqrt{n\kappa_n}} x^3\,\mathbb{P}\left(|v_1| > x\right)dx$$

where the first term on the right converges to $0$ as above while the second term is bounded from above by $\frac{4}{\sqrt{n\kappa_n}}\int_0^{\sqrt{n\kappa_n}} x^2\,\mathbb{P}\left(|v_1| > x\right)dx = \frac{4}{\sqrt{n\kappa_n}}H_v\left(\sqrt{n\kappa_n}\right)$ which again converges to $0$. □

**Lemma 2.** *Let Assumptions 1 and 2 for $q = 1$ and suppose that $(w_t)_{t\in\mathbb{N}}$ is stationary and ergodic sequence of random variables such that $\mathbb{E}|w_1| < +\infty$, and $w_t$ is independent of $v_t$ for all $t$. Then for any $\delta > 0$*

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n}\left[w_i - \mathbb{E}(w_1)\right]v_i^2\right| \geq \delta\right) \to 0.$$

*Proof.* Notice that

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n} w_i\left(v_i^2 - v_{i,n}^2\right)\right| + |\mathbb{E}(w_1)|\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n}\left(v_i^2 - v_{i,n}^2\right)\right| \geq \delta\right)$$

$$\leq n2\mathbb{P}\left(|v_0| > \sqrt{n\kappa_n}\right) = o(1).$$

We also have that for any $\epsilon > 0$

$$\frac{1}{n\kappa_n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left(\left(w_i v_{i,n}^2 1_{\{|w_i|v_{i,n}^2 > \epsilon n\kappa_n\}}\right)/\sigma(w_i)\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[w_i 1_{\{|w_i|>\epsilon\}}\frac{H_v\left(\sqrt{n\kappa_n}\right)-H_v\left(\sqrt{\frac{\epsilon n\kappa_n}{|w_i|}}\right)}{\kappa_n}\right]$$

and the latter converges to zero due to dominated convergence and Assumption 2. Hence there exists some $\epsilon_n \to 0$ for which

$$\lim_{n\to\infty}\frac{1}{n\kappa_n}\sum_{i=1}^{n}\mathbb{E}\left[w_i v_{i,n}^2 1_{\{|w_i|v_{i,n}^2>\epsilon_n n\kappa_n\}}\right]=0$$

and the result follows by Theorem 1 of Sung [9]. □

## References

[1] Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1989). *Regular variation* (Vol. 27). Cambridge university press.

[2] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, (Vol. 2). New York: Wiley.

[3] Hall, P., & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy–tailed errors. *Econometrica*, 71(1), 285-317.

[4] Ibragimov I.A. and Yu. V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*. WOLTERS-NOORDHOFF PUBLISHING GRONINGEN.

[5] Jakubowski, A. (2012). Principle of Conditioning revisited. Demonstratio Mathematica, 45(2), 325-36.

[6] Meerschaert, M. M. (1994). Norming operators for generalized domains of attraction. *Journal of Theoretical Probability*, 7(4), 793-798.

[7] Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences. MIMEO. url: https://hal.archives-ouvertes.fr/cel-00867106/document.

[8] Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models* (Vol. 181). New York: Springer.

[9] Sung, S. H. (1999). Weak law of large numbers for arrays of random variables. *Statistics & probability letters*, 42(3), 293-298.

[10] Wang, Q. (2014), ”Martingale Limit Theory Revisited and Non Linear Cointegrating Relations”. *Econometric Theory*, 30, 509–535.

[11] Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH (1, 1) model. *Scandinavian Journal of Statistics*, 40(4), 846-867.

[12] Wintenberger, O., & Cai, S. (2011). Parametric inference and forecasting in continuously invertible volatility models. arXiv preprint arXiv:1106.4983.