A CLT For Martingale Transforms With Slowly Varying Second Moments and the Limit Theory of the QMLE for Conditionally Heteroskedastic Models

Stelios Arvanitis and Alexandros Louka

September 30, 2013

Abstract

We provide a limit theorem to a normal limit for the standardized sum of a martingale transform that holds even in cases where the second moments diverge at an appropriately slow rate. This extends relevant results with stable but non normal limits to the case of asymptotic normality, as well as results of asymptotic normality by allowing domains of non-normal attraction. In those cases the rate is slower than $\sqrt{n}$ and it contains information for the rate of divergence of the truncated second moments. A major application concerns the characterization of the rate and the limiting distribution of the Gaussian QMLE in the case of GARCH type models. By extending the relevant framework we accommodate for the case of slowly varying and potentially diverging fourth moments for the innovation process as well as the possibility that the parameter lies on the boundary. The results are of potential interest to financial econometrics in view of the conditional leptokurtosis of the empirical distributions of asset returns.

KEYWORDS: CLT, Domain of non normal Attraction, Martingale Transform, Slowly Varying Second Moment, Stationarity, Ergodicity, Conditional heteroskedasticity, Gaussian quasi likelihood, QMLE, boundary point, infinite fourth moments, leptokurtosis.

JEL: C10, C13.

1 Introduction-Motivation

It is empirically known that distributions of financial asset returns exhibit fat tail behavior. Modelling the conditional moments of such processes using GARCH-type models has only partly explained this behavior (the *standardized* returns exhibit among others, significant skewness and leptokurtosis-see for example Diebold [11]) and therefore considering heavy-tailed distributions for the innovation process is of particular interest for applications in finance. The use of the Gaussian QMLE for the parameter estimation of such models is very convenient as it has been shown to be consistent and asymptotically normal under mild conditions and thus reducing the risk of model misspecification. However, asymptotic normality with the usual $\sqrt{n}$ rate breaks down when the fourth moment of the error process is infinite. It is worth noting that, contrary to the QMLE, the MLE estimator can be $\sqrt{n}$-consistent in this case, which has led to the consideration of the MLE and the QMLE using non-Gaussian densities that allow fat tail behavior such as the Student' t distribution with unknown degrees of freedom or the generalized error distribution (see among others Engle and Bollerslev [12], Calzolari et al. [7], Hansen [14], Hsieh [15], Nelson [26]). However, as the error distribution

is generally unknown, this could result to the inconsistency of such estimators as Straumann [28] showed.

In this respect, the relative superiority of the QMLE has led to the construction and use of limit theorems for sequences with infinite second moments that imply distributional convergence to a-stable distributions with rate of convergence slower than $\sqrt{n}$. Note that for an a-stable distribution the stability parameter a which takes values in the interval $(0,2]$ controls the tail behavior of the distribution and $a=2$ corresponds to the Gaussian distribution. Mikosch and Straumann [23], extending the work of Hall and Yao [13], derived a limit theorem to stable laws with $a \in (0,2)$ for martingale transforms of the form $\sum_{t=1}^{n} \xi_t V_t$ where $(\xi_t)_{t\in\mathbb{Z}}$ is an iid sequence, $(V_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence with sufficiently high moments, that also satisfies a certain mixing condition. Note that the rate is partially characterized but not analytically obtained. Furthermore, several properties of the limit remain incompletely specified (with the exception of a itself-see Remark 2.3 of Mikosch and Straumann [23]). Then applying this limit theorem they derive the convergence in distribution of the QMLE for the $\mathrm{GARCH}(p,q)$ model to an a-stable distribution. Surgailis [30] also derived a limit theorem for analogous zero mean sums for $a \in (1,2)$ when the distribution of ξ_0 lies in the domain of *normal*[1] attraction of an a-stable distribution, without the need of the mixing condition. Jakubowski [17] improved the result by requiring the existence of lower moments for V_0. Here, the rate is completely specified, and the characteristics of the limit are specified up to linear transformations. Arvanitis and Louka [4] applied Surgailis' limit theorem to derive the asymptotic distribution of the QMLE for the $\mathrm{GQARCH}(1,1)$ model with rate $n^{1-1/a}$. A detailed comparison with Mikosch and Straumann's approach can be found in the latter.

Our motivation lies in the fact that, although not evident in the literature, *it is possible to derive asymptotic normality of the* QMLE *even when the error process has infinite fourth moment*, under conditions regarding the rate of divergence of the truncated second moment of the squared error. We derive a CLT for sums of (vector) martingale transforms which essentially extends that of Surgailis for $a=2$ allowing for *domains of non-normal attraction* to the normal distribution. The latter is used to show the asymptotic normality of the QMLE estimator with a rate of convergence that is slower than $\sqrt{n}$ but faster than the rate that is implied in the relative literature. Thus, the main purpose of this paper is to enhance the limit theory of the QMLE by including this intermediate case concerning the distribution of the errors. The framework we use to examine the asymptotic behavior of the Gaussian QMLE builds on the application of the Lipschitz stochastic recurrence equation (SRE) theory by Straumann [28] as well as Wintenberger and Cai [33] and is also closely related to that of Arvanitis and Louka [4] which allows the parameter of interest to lie on the boundary of the parameter space. They partially use the methodology of Andrews [3] when it comes to the asymptotic approximation of the likelihood function by a second order polynomial, but they depart from the latter by characterizing the asymptotic parameter space using upper Painleve-Kuratowski limits (see for example van der Vaart [32] Lemma 7.13.2-3). The main departure with respect to the Arvanitis and Louka [4] paper, is that we extend the results to allow for a non normal domain attraction, that we restrict our attention to the normal law, that is when $a=2$ and that by using the theory of Wintenberger and Cai [33] we derive the results by utilizing weaker moment conditions for the derivatives of the ergodic version of the quasi likelihood function.

The remaining structure of the paper is organized as follows. First, the main result is

[1] Remember that a distribution lies in the domain of normal attraction of an a-stable distribution ($a \in (0,2]$) iff the slowly varying function appearing in the local to zero representation of its characteristic function has a limit. In this case the standardization is of the form $\frac{1}{n^{1/a}}$.

presented regarding the CLT involving martingale transforms with slowly varying truncated second moments that could diverge slowly enough. Notice that among others the CLT extends the results of Abadir and Magnus [1] when ξ_0 follows a Student's t distribution with four degrees of freedom. Secondly, we provide the framework and the required assumptions regarding the consistency and asymptotic normality of the QMLE. We also provide an extension of the limit theorem that incorporates simple cases that do not conform to the aforementioned framework. We then perform a Monte Carlo study regarding the QMLE for the $\mathrm{GARCH}(1,1)$ model which indicates that the asymptotic results approximately carry over to (sufficiently large) samples. Finally we conclude.

Notation Let us initially fix a general framework along with some notation. To this end $(\Omega, \mathcal{F}, \mathbb{P})$ denotes an underlying complete probability space and thereby $\mathbb{P}$ a.s. denotes that a property holds for any element in a measurable subset of Ω of unit $\mathbb{P}$ probability, $\underset{p}{\rightarrow}$, $\rightsquigarrow$ denote convergence in probability and distribution respectively while $\backsim$ denotes asymptotic equivalence as $n \rightarrow \infty$. $\overset{e.a.s.}{\rightarrow}$ denotes exponentially fast almost sure convergence under $\mathbb{P}$. Remember that $x_n \overset{e.a.s.}{\rightarrow} 0$ iff $\gamma^n x_n \rightarrow 0$ $\mathbb{P}$ a.s. for some $\gamma > 1$ (see Paragraph 2.5 of Straumann [28]). $\|\cdot\|$ denotes the usual Euclidean norm or the Frobenius matrix norm, the distinction will be obvious by the context, and $B(\theta, \varepsilon)$, $\bar{B}(\theta, \varepsilon)$ denote the open and the closed ball respectively, centered at θ and having radii $\varepsilon > 0$ in some relevant metric space. For a self function (say Φ) on an appropriate space, $\cdot^{(m)}$ denotes m-fold self composition, with the convention that for $m = 0$ we obtain the identity function, while if the function is Lipschitz then $\Lambda^{(m)}(\Phi)$ denotes the Lipschitz coefficient of $\Phi^{(m)}$, i.e. $(\Lambda(\Phi))^m$. $\cdot^T$ denotes transposition and $\mathrm{diag}(x)$ denotes the diagonalization of a vector x to a square matrix. Finally, C denotes a *generic* positive constant, while t_v denotes the Student's t distribution with v degrees of freedom and $\ln^+(x) = 1_{[1,+\infty)}(x) \ln x$.

2 A CLT Involving Martingale Transforms With Slowly Varying Truncated Second Moments

In this section we present the main probabilistic result which is a direct consequence of the *principle of conditioning*-see Jakubowski [16] and [17], Kwapien and Woyczynski [21]-Theorem 5.8.3 and Surgailis [30]-Theorem B.2. It concerns the convergence in distribution of a properly standardized sum of (vector) martingale transforms to a normally distributed random vector. The result nests a special case of the CLT for square integrable stationary and ergodic martingale differences but it also allows for cases in which second conditional moments do not exist, as long as their truncated versions diverge slowly enough. In such cases the rate ceases to be of asymptotic order $\sqrt{n}$ but it also contains information on the speed of divergence of the aforementioned moments.

Remember that if $\wp$ is a (Lebesgue measurable) real function defined on a neighborhood of infinity (i.e. on an interval of the form $[x_0, +\infty)$ for some $x_0 \in \mathbb{R}$) we say that it varies at infinity with index $a \geq 0$, iff for any $t > 0$, $\lim_{x\rightarrow\infty} \frac{\wp(tx)}{\wp(x)} = t^a$. The variation is termed *regular* if $a \neq 0$ otherwise it is termed *slow* (see among others the Definition and the discussion in paragraph 1.8 of Bingham et. al. [6]). The Karamata representation (see Theorem 1.3.1 of Bingham et. al. [6]) of a slowly varying function at infinity (say $\wp$) has the form

$$\wp(x) = \exp\left(c(x) + \int_{x_0}^{x} \frac{\epsilon(t)}{t} dt\right)$$

where $c(\cdot)$, $\epsilon(\cdot)$ are Lebesgue measurable, while $c(x) \to C$, $\epsilon(x) \to 0$ as $x \to +\infty$. Considerations of asymptotic equivalence may imply that it is of no loss of generality to restrict our focus to functions that admit the representation

$$C \exp\left(\int_{x_0}^{x} \frac{\epsilon(t)}{t} dt\right).$$

The latter is said to belong to the *Zygmund* class of slowly varying functions (see Theorem 1.5.5 of Bingham et. al. [6]) and this along with the measurability of $\epsilon(\cdot)$ imply that $\epsilon(x)\wp(x) = x\wp'(x)$, Lebesgue almost everywhere. We are now ready to state the result.

Theorem 2.1 *Let the following hold:*

1. $(\xi_t)_{t\in\mathbb{N}}$ *is an* iid *sequence of random variables and* $(V_t)_{t\in\mathbb{N}}$ *is an* $\mathbb{R}^d$*-valued sequence of random elements that is point-wise stationary and ergodic. For the filtration* $(\mathcal{F}_t)_{t\in\mathbb{N}}$ *with* $\mathcal{F}_t \triangleq \sigma\left(\xi_t, V_t, \xi_{t-1}, V_{t-1}, \xi_{t-2}, V_{t-2}, \ldots\right)$, ξ_t *is independent of* $\mathcal{F}_{t-1}$ *and* V_t *is measurable w.r.t.* $\mathcal{F}_{t-1}$.

2. *For the second truncated moment of the distribution of* ξ_0*, say* F_{ξ_0}

$$\wp(x) \backsim \int_{-x}^{x} \xi_0^2 dF_{\xi_0},$$

   where $\wp$ *is a continuously differentiable, slowly varying function of the Zygmund class such that* $\epsilon(x)\wp(x)$ *is bounded.*

3. *If* $\wp$ *has a limit then* $\mathbb{E}\left\|V_0\right\|^2 < +\infty$*, otherwise* $\mathbb{E}\left\|\mathrm{diag}^2(V_0)\left(\ln|V_{0_i}|\right)_{i=1,\cdots,d}\right\| < +\infty$.

Then as $n \to \infty$

$$\frac{1}{\sqrt{n\wp\left(\sqrt{n}\right)}} \sum_{t=1}^{n} \xi_t V_t \rightsquigarrow N\left(\mathbf{0}_d, \mathbb{E}\left(V_0 V_0^T\right)\right).$$

The first condition describes the martingale transform emerging from a multiplicative structure between an iid and a stationary and ergodic sequence. This structure is typical in conditionally heteroskedastic models either for the process itself or for the score process emerging from the relevant Gaussian quasi likelihood function. From this we derive our initial motivation. The third condition concerns further restrictions on the distribution of the $(V_n)_{n\in\mathbb{N}}$ in the form of moment conditions. The first case is essentially a special case of the martingale difference CLT. The stricter moment condition in the second one, would be satisfied if $\mathbb{E}\left\|V_0\right\|^{2\lambda} < +\infty$ for some $\lambda > 1$. These conditions can be easily verified in the background of the Gaussian QML estimation for the score process of several GARCH-type models.

The second condition imposes a special structure on the form of the slowly varying function that is asymptotically equivalent to the second truncated moment of ξ_0. First notice that if $\wp$ has a limit then it is always asymptotically equivalent to a function that admits this structure in a trivial manner. Second, it is easy to see that the condition is satisfied by functions of the form $\wp(x) = C\ln x$, an observation that generates the concluding remarks of the present section. Given this, a non trivial class of slowly varying functions can be built that satisfy the restrictions of condition 2, as non exhaustively exhibited in the following remark.

Remark R.1 *It is a matter of routine calculations to see that functions of the form* $C\frac{(\ln^{(m_1)}(x))^{\delta_1}}{(\ln^{(m_2)}(x))^{\delta_2}}$ *for* $\delta_1 \in [0,1], : \delta_2 \geq 0$ *or of the form* $C(\ln^{(m)}(\exp(1)+x))^{\delta}$ *for* $\delta \in [0,1]$*, with appropriate domain, satisfy 2.1.2. Furthermore, if* $\wp_i, : i = 1, \cdots, k$ *satisfies the aforementioned condition then the (appropriately restricted) function* $\wp(x) = \prod_{i=1}^{k}(\wp_i(x))^{\delta_i}$ *for* $\delta_i \in [0,1]$ *for all* i *does too. Notice that counterexamples can be constructed from the previous considerations by exponentiation. For example a function of the form* $\exp((\ln^{(m_1)}(x))^{\delta}(\ln^{(m_2)}(x))^{-1})$ *for* $\delta \in (0,1]$ *would also be in the Zygmund class and satisfy the differentiability restriction but it would fail to satisfy the boundedness part of the aforementioned condition if* $m_1 < m_2$*. The same restriction generally precludes functions of "infinite oscillation" such as those of the form* $C\exp(\ln^{1/3}(x)\cos^{1/3}(x))$ *that also meet the other requirements.*

Hence the theorem allows for slower than logarithmic rates of divergence for the truncated second moment.

Remark R.2 *Using the direct and the dual versions of Potter's Theorem (see Theorem 1.5.6 in Bingham et. al. [6]) it is easy to see that (suppose without loss of generality that* $d = 1$*) for any increasing slowly varying* $\wp$ *if* $\mathbb{E}\left|V_0\right|^{2\lambda} < +\infty$ *for some* $\lambda > 1$ *then*

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}V_i^2\frac{\wp\left(\frac{\sqrt{m\wp(\sqrt{m})}}{|t||V_i|}\right)}{\wp\left(\sqrt{m}\right)} \leq A\mathbb{E}\left(\left|V_0\right|^2\max\left\{V_0^{-\delta},V_0^{\delta}\right\}\right)\right) = 1$$

and

$$\lim_{n\to\infty}\mathbb{P}\left(\inf_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}V_i^2\frac{\wp\left(\frac{\sqrt{m\wp(\sqrt{m})}}{|t||V_i|}\right)}{\wp\left(\sqrt{m}\right)} \geq \frac{1}{A}\mathbb{E}\left(\left|V_0\right|^2\max\left\{V_0^{-\delta},V_0^{\delta}\right\}\right)\right) = 1$$

for any $A > 1$ *and any* $0 < \delta \leq 2(\lambda - 1)$ *(see the previous remark). This is due to the fact that* $\frac{\sqrt{n\wp(\sqrt{n})}}{|t||V_i|} \geq \frac{\sqrt{\wp(\sqrt{n})}}{t_0}$ *with* $\mathbb{P}$ *probability converging to* 1 *(see the set* $C_{n,K}$ *defined in the proof of theorem 2.1) and since* $\wp(\sqrt{n}) \to \infty$ *Potter's Theorem is applicable. In order to obtain results such as the one in theorem 2.1 further structure is imposed on* $\wp$ *that allows that limits can simultaneously be considered as* $A \downarrow 1$ *and* $\delta \downarrow 0$*.*

Remark R.3 *The convergence rate* $\frac{\sqrt{n}}{\sqrt{\wp(\sqrt{n})}}$ *of the arithmetic mean* $\frac{1}{n}\sum_{t=1}^{n}\xi_t V_t$ *can for example be replaced by* $\frac{\sqrt{n}}{\sqrt{\wp\left(\sqrt{n\wp(\sqrt{n})}\right)}}$ *since from the proof of theorem 2.1 we have that any function satisfying 2.1.2 also satisfies* $\frac{\wp(x\sqrt{\wp(x)})}{\wp(x)} \to 1$ *as* $x \to +\infty$*. More generally, given a sequence* $(\eta_n)_{n\in\mathbb{N}}$ *such that* $\eta_n \downarrow 0$ *the former is asymptotically equivalent to* $\frac{\sqrt{n}}{q_n}$ *where* $q_n = \max_z\left\{\frac{z^2}{\wp(\sqrt{n}z)} \leq 1+\eta_n\right\}$ *(see for example the proof of Theorem 2.6.2 of Ibragimov and Linnik [19]). Given its definition it can be seen that* q_n *must be of the form* $\wp^{\star}(n)$ *where again* $\wp^{\star}$ *is a slowly varying function (see paragraph 2.2 of Ibragimov and Linnik [19]). When and only when* $\wp$ *has a limit, then the rate is of the form* $C\sqrt{n}$ *whereupon we obtain the notion of the domain of normal attraction to the normal distribution.*

Some special applications of theorem 2.1 are the following.

Remark R.4 *The theorem nests as a special case the results of Abadir and Magnus [1]. More specifically, when* $d=1$, $V_n=1$ $\mathbb{P}$ *a.s. for all* n *and* $\xi_0 \sim t_2$ *then* $\wp(x)=2\ln x$ *and therefore* $\frac{1}{\sqrt{n\ln n}}\sum_{t=1}^{n}\xi_k \rightsquigarrow N(0,1)$. *A more complex case is (among many similar others) the one where* $(V_n^2)_{n\in\mathbb{N}}$ *is a stochastic volatility sequence, e.g. it satisfies the* $\mathrm{SV}(1,1)$ *recursion*

$$V_t^2=\exp\left(\omega+\alpha u_{t-1}+\beta\ln V_{t-1}^2\right)$$

where $(u_n)_{n\in\mathbb{Z}}$ *is* iid *with* $\mathbb{E}(u_0)=0$, *the distribution of* u_0 *has a moment generating function (say* $\mathbb{M}$*),* u_t *is independent of* u_s *for any* t,s. *If* $|\beta|<1$ *then Theorem 2.6.1 of Straumann [28] and the fact that continuous transformations preserve stationarity and ergodicity imply the validity of condition 1. Furthermore, when* $|\beta|<1$ *and* λa, *for some* $\lambda>1$, *lies in the interval of absolute convergence of* $\ln\mathbb{M}$ *(i.e. the cumulant generating function) then 2.1.3 is valid and thereby some calculations show that*

$$\frac{1}{\sqrt{n\ln n}}\sum_{t=1}^{n}\xi_t V_t \rightsquigarrow N\left(0,\exp\left(\frac{\omega}{1-\beta}+\sum_{i=0}^{\infty}\ln\mathbb{M}\left(\alpha\beta^i\right)\right)\right).$$

In the case where u_0 *follows the standard normal distribution we have that* $\ln\mathbb{M}\left(\alpha\beta^i\right)=\frac{a^2\beta^{2i}}{2}$ *and the asymptotic variance becomes* $\exp\left(\frac{\omega}{1-\beta}+\frac{a^2}{2\left(1-\beta^2\right)}\right)$.

We distinguish the following result in a separate corollary since it constitutes a special case of the form of the score for the Gaussian (quasi-) likelihood function in the context of conditionally heteroskedastic models.

Corollary 2.2 *Let* $\xi_n=z_n^2-1$ *where* $\sqrt{2}z_0\sim t_4$, *and suppose that conditions 2.1.1,3 hold. Then*

$$\frac{1}{\sqrt{n\ln n}}\sum_{t=1}^{n}\xi_t V_t \rightsquigarrow N\left(\mathbf{0}_d,\frac{3}{2}\mathbb{E}\left(V_0V_0^T\right)\right).$$

When $\xi_n=z_n^2-1$ the aforementioned score has the form of the sum in theorem 2.1. This is essentially motivation for the considerations in the following section.

3 Limit Theory of the QMLE for GARCH Type Models With Slowly Varying Fourth Conditional Moments

A major application of the theorem presented in the previous section concerns the characterization of the rate and the asymptotic distribution of the Gaussian QMLE in GARCH type models. In what follows we briefly describe the framework and derive the results. The derivations draw heavily on the theory developed by Straumann [28] as well as Wintenberger and Cai [33]. The differences correspond *first* to the fact that we allow for *the centralized squares* of the elements of the *structuring sequence* to lie in the domain of *non normal attraction to the normal distribution* and *second* to the parameter of interest to be on the boundary of the relevant parameter space.

The framework is structured as follows: first, we define the process as the unique stationary and ergodic solution of a stochastic recurrence system of equations, second we are occupied with the issue of existence, uniqueness, stationarity and ergodicity of the solution

of a transformation of the aforementioned recurrence, that essentially enables the invertibility of the volatility process for any parameter value. This allows the approximation of the latter process, which is *latent*, by filters that are measurable functions of the *observed* heteroskedastic process (this is related to the notion of *observable invertibility* essentially appearing in Straumann [28]-see definition 2 of Wintenberger and Cai [33]). Third, we define the QMLE and given the previous, we describe sufficient conditions (e.g. existence of logarithmic moments and of universal lower bounds for the filtered processes) that establish its strong consistency. Finally, we are occupied with the issue of existence, uniqueness, stationarity and ergodicity of the solutions of recurrence equations that emerge by differentiating the previous equations, along with analogous (moment existence, linear independence etc.) conditions for those solutions that permit among others the application of the CLT of the previous section, and are in any case helpful for the establishment of the rate and the weak limit of the QMLE via the results in the last part of the Appendix.

The process Suppose that Θ is a compact subset of $\mathbb{R}^d$ and let θ_0 be an *arbitrary* member of Θ. Consider the conditionally heteroskedastic process (w.r.t. θ_0) defined by

$$\left\{ \begin{array}{c} y_t = \sigma_t z_t \\ \sigma_t^2 = g_{\theta_0}\left(z_{t-1}, \ldots, z_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-l}^2\right) \end{array} \right. , t \in \mathbb{Z} \tag{1}$$

where the *structuring sequence* $(z_t)_{t\in\mathbb{Z}}$ is a process of iid random variables with $\mathbb{E}z_0 = 0$ and $\mathbb{E}z_0^2 = 1$, $g. \in \mathbb{C}\left(\Theta \times \mathbb{R}^p \times (\mathbb{R}^+)^q, \mathbb{R}^{++}\right)$ for any $\theta \in \Theta$ and $l = \max(p, q)$. Let

$$\Phi_{t,\theta_0}(x) \doteqdot \left(g_{\theta_0}\left(z_t, \ldots, z_{t-p+1}, x_1, \ldots, x_l\right), x_1, \ldots, x_{l-1}\right)'.$$

Given the definition of $(z_t)_{t\in\mathbb{Z}}$ and the properties of g_{θ_0}, the sequence $(\Phi_{t,\theta_0}(x))_{t\in\mathbb{Z}}$ is stationary and ergodic for any x due to Proposition 2.1.1 of Straumann [28].

Assumption A.1 *Suppose that*

$$\mathbb{E}\ln^+ \left|g_{\theta_0}\left(z_0, \ldots, z_{-p+1}, y_1, \ldots, y_l\right)\right| < +\infty,$$

for some $y_1, \ldots, y_l \in \mathbb{R}^{++}$, Φ_{t,θ_0} *is* $\mathbb{P}$ *a.s. Lipschitz w.r.t.* x *with coefficient* $\Lambda(\Phi_{t,\theta_0})$ *that satisfies*

$$\mathbb{E}\ln^+ \Lambda(\Phi_{0,\theta_0}) < +\infty \text{ and for some } m \in \mathbb{N}^*, \mathbb{E}\ln\Lambda\left(\Phi_{0,\theta_0}^{(m)}\right) < 0.$$

The previous assumption along with Theorem 2.6.1 of Straumann [28], imply that the stochastic recurrence equation (SRE) in (1) admits a unique (up to indistinguishability) stationary and ergodic solution $(\sigma_t^2)_{t\in\mathbb{Z}}$ and furthermore any other solution converges exponentially almost surely to this one as $t \to \infty$. Due to continuity those properties extend to the heteroskedastic process itself.

Continuous Invertibility and the $(h_t)_{t\in\mathbb{Z}}$ Process Given the described process, the next part of the framework concerns the issue of *continuous invertibility* (see Definition 4 of Wintenberger and Cai [33]). This is closely connected to the properties of the filtering of the latent volatility process and thereby to the optimization procedure on the relevant likelihood function. Consider g_θ from before along with the first equation of (1). Given the process $(y_t)_{t\in\mathbb{Z}}$ consider the following stochastic recursion

$$h_t(\theta) = g_\theta\left(\frac{y_{t-1}}{\sqrt{h_{t-1}(\theta)}}, \ldots, \frac{y_{t-p-1}}{\sqrt{h_{t-p-1}(\theta)}}, h_{t-1}(\theta), \ldots, h_{t-q-1}(\theta)\right), \tag{2}$$

where $t \in \mathbb{Z}$ and $\theta \in \Theta$. Likewise to the previous section consider

$$\Psi_{t,\theta}(x) \doteqdot \left(g_\theta \left(\frac{y_{t-1}}{\sqrt{x_1}}, \ldots, \frac{y_{t-p-1}}{\sqrt{x_p}}, x_1, \ldots, x_l \right), x_1, \ldots, x_{l-1} \right).$$

Analogously, the sequence $(\Psi_{t,\theta}(x))_{t\in\mathbb{Z}}$ is stationary and ergodic for any x, θ. The following assumption is essentially condition (**CI**) of Wintenberger and Cai [33].

Assumption A.2 *Suppose that*

$$\mathbb{E} \ln^+ \left(\sup_{\theta\in\Theta} \left| g_\theta \left(\frac{y_{t-1}}{\sqrt{x_1}}, \ldots, \frac{y_{t-p-1}}{\sqrt{x_p}}, x_1, \ldots, x_l \right) \right| \right) < +\infty,$$

for some $x_1, \ldots, x_l \in \mathbb{R}^{++}$. $\Psi_{t,\theta}$ *is* $\mathbb{P}$ *a.s. Lipschitz w.r.t.* x *with coefficient* $\Lambda(\Psi_{t,\theta})$ *that is* $\mathbb{P}$ *a.s. continuous w.r.t.* θ *and satisfies*

$$\mathbb{E} \ln^+ \sup_{\theta\in\Theta} \Lambda(\Psi_{0,\theta}) < +\infty \text{ and for some } m \in \mathbb{N}^*, \ \mathbb{E} \ln \Lambda\left(\Psi_{0,\theta}^{(m)}\right) < 0 \text{ for all } \theta \in \Theta.$$

The following lemma summarizes some of the implications of the first pair of assumptions. It is essentially Theorem 3 of Wintenberger and Cai [33].

Lemma 3.1 *Under assumptions A.1 and A.2 for any* $\theta \in \Theta$ *there exists a unique stationary and ergodic solution* $(h_t(\theta))_{t\in\mathbb{Z}}$ *to (2). Moreover* $h_t(\theta)$ *is continuous w.r.t.* θ. *Furthermore for any* $\theta \in \Theta$ *and any other solution to (2), say* $\left(\hat{h}_t(\theta)\right)_{t\in\mathbb{Z}}$, *there exists* $\varepsilon > 0$ *such that* $\sup_{\theta'\in B(\theta,\varepsilon)\cap\Theta} \left| h_t(\theta') - \hat{h}_t(\theta') \right| \overset{e.a.s.}{\to} 0$.

This is extremely helpful since the actual evaluation at each parameter value, and thereby the computability of the optimization of the likelihood function, depends on solutions of (2) based on initial conditions. It implies that any such solution (that is in general non stationary due to its dependence on initial conditions) will converge to the stationary and ergodic solution fast enough as $t \to \infty$. The local uniformity of the approximation, the stationarity and ergodicity of the solution, along with some moment existence could imply the convergence of arithmetic means of the $\left(\hat{h}_t(\theta)\right)_{t\in\mathbb{Z}}$ process evaluated at a convergent sequence to the expectation of the ergodic solution evaluated at the limit of the aforementioned sequence. All these will be convenient for the establishment of the asymptotic properties of the estimator.

The QMLE-Definition and Existence Given a finite realization $(y_t)_{t=1,\ldots,n}$ from the heteroskedastic process, the following defines the Gaussian quasi likelihood function $\hat{c}_n$. The term is used in an abusive manner since the original function would be constructed as $-\frac{1}{2} * \hat{c}_n(\theta) + \mathrm{const}$. This form enables the characterization of the QMLE as an approximate minimizer.

Assumption A.3 *Suppose that* $\varsigma_{k,\theta} : \Omega \to \mathbb{R}^{++}$ *is measurable for any* $\theta \in \Theta$ *and* $\mathbb{P}$ *almost surely continuous w.r.t.* θ *for all* $k = 0, \ldots, l-1$ *and,* $\zeta_{k,\theta} : \Omega \to \mathbb{R}$ *is measurable for any* $\theta \in \Theta$ *and* $\mathbb{P}$ *a.s. continuous w.r.t.* θ *for all* $k = 0, \ldots, p-1$.

Definition D.1 *Define the filter* $\left(\hat{h}_t(\theta)\right)_{t=1,\ldots,n}$ *for* $\theta \in \Theta$ *by*

$$\begin{aligned}\hat{h}_k(\theta) &= \varsigma_{k,\theta} \text{ when } k=0,\ldots,l-1 \text{ and } y_k=\zeta_{k,\theta} \text{ when } k=0,\ldots,p-1 \text{ and}\\ \hat{h}_t(\theta) &= g_\theta\left(\frac{y_{t-1}}{\sqrt{\hat{h}_{t-1}(\theta)}},\ldots,\frac{y_{t-p-1}}{\sqrt{\hat{h}_{t-p-1}(\theta)}},\hat{h}_{t-1}(\theta),\ldots,\hat{h}_{t-q-1}(\theta)\right).\end{aligned}$$

We can now define the Gaussian quasi likelihood function and the subsequent estimator, as a (possibly measurable selection) of its approximate $\arg\min$.

Definition D.2 *The Gaussian quasi likelihood function is*

$$\hat{c}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\hat{\ell}_t(\theta)$$

where

$$\hat{\ell}_t(\theta) = \ln\hat{h}_t(\theta) + \frac{y_t^2}{\hat{h}_t(\theta)}.$$

For ε_n *an* $\mathbb{P}$ *almost surely non negative random variable the* QMLE θ_n *is defined by*

$$\hat{c}_n(\theta_n) \leq \inf_{\theta\in\Theta}\hat{c}_n(\theta) + \varepsilon_n.$$

ε_n can be perceived as an *optimization error*, and thereby the definition is wide enough to include the estimator obtained (as is usually the case) by numerical optimization of $\hat{c}_n$. The $\mathbb{P}$ almost sure continuity (w.r.t. θ) of the filter, inherited by the definition of g_θ and assumption A.3 along with the compactness and the separability of Θ imply the existence of θ_n even when $\varepsilon_n = 0$ $\mathbb{P}$ a.s. This is rigorously established in the proof of the following proposition.

Proposition 3.2 *Suppose that assumption A.3 holds, then the* QMLE *exists.*

Consistency We turn to the limit theory for the estimator. The aforementioned *exponentially fast* approximation of the filter by the stationary and ergodic inverted process $(h_t)_{t\in\mathbb{Z}}$ (locally uniformly) along with the consequences of assumption A.1 enable the asymptotic approximation of $\hat{c}_n$ by an average of ergodic contributions obtained as

$$c_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\ell_t(\theta)$$

with

$$\ell_t(\theta) = \ln h_t(\theta) + \frac{y_t^2}{h_t(\theta)}$$

We can address c_n as the "ergodic likelihood". Several of its properties are appropriate approximations of analogous properties of $\hat{c}_n$ and thereby they will be used for the establishment of the limit theory. In this respect, given the previous, the following assumption provides with sufficient conditions for strong consistency.

Assumption A.4 *Suppose that:*

1. $\varepsilon_n \to 0$ $\mathbb{P}$ *a.s.*

2. $\mathbb{E} \ln^+ \sigma_0^2 < +\infty$.

3. $\inf_\Theta h_0(\theta) \geq C \quad \mathbb{P}$ *a.s.*

4. *For any* $\theta \in \Theta$*:*

$$h_0(\theta) = \sigma_0^2 \Leftrightarrow \theta = \theta_0.$$

Condition A.4.1 implies that the optimization error vanishes asymptotically. A.4.2 requires the existence of logarithmic moments for the volatility process and due to the properties of z_0, it also implies that $\mathbb{E} \ln^+ y_0^2 < \infty$. By Theorem 2 of Wintenberger and Cai [33] it follows from assumption A.1 and a condition of the form $\mathbb{E}\left(\ln^+ \left|g_{\theta_0}(z_0, \ldots, z_{-p+1}, y_1, \ldots, y_l)\right|\right)^2 < +\infty$ for some $y \in \mathbb{R}^{++}$. A.4.3 requires the existence of a universal deterministic lower bound for the volatility processes that is naturally obtained in several GARCH-type models again due to the form of the recursion, the positivity constraints and the inclusion of a strictly positive constant. In more complex cases (e.g. the EGARCH model), it could be obtained by placing further restrictions on the parameter space. A.4.4 is an identification condition that can be obtained by requiring more structure on the support of the distribution of z_0 as well as on the form of the defining recursion. The result is presented in the following theorem.

Theorem 3.3 *Suppose that assumptions A.1, A.2, A.3 and A.4 hold, then the* QMLE *is strongly consistent.*

Notice that assumptions A.1, A.2, A.3 along with conditions A.4.2-4 are identical to the conditions C.1-C.4 of the relevant theorem 5.3.1 of Straumann [28] (see the proof of the second part) or Theorem 4 of Wintenberger and Cai [33]. Hence theorem 3.3 is essentially an *extension* by allowing the existence of an asymptotically negligible optimization error, and thereby by providing sufficient conditions for the *consistency of approximate optimizers* of the likelihood function.

Rate and Asymptotic Distribution The remaining elements of the limit theory, i.e. the rate and the limiting distribution can be established by conditions that are local in nature. The results depend crucially on the asymptotic existence of a local to θ_0 quadratic approximation of c_n^*, as required by theorem 5.2. In accordance with the differentiability properties of $\hat{h}_t$ for a variety of heteroskedastic models, we will assume that the approximation has the form of a second order Taylor expansion. Hence due to the possibility of θ_0 being on the boundary of Θ we will need a form of differentiability for the filter (and the subsequent stationary and ergodic approximation) that is consistent with this. We will use the notion of *left/right* (*l/r*) partial derivatives as in paragraph 3.3. of Andrews [3]. This requires some further structure on the set on which θ_n at least asymptotically attains its values. The following assumption takes care of those concepts.

Assumption A.5 *Suppose that:*

1. For some $\eta \leq \frac{\varepsilon}{m}$ for some $1 < m \in \mathbb{N}$ and the $\varepsilon > 0$ that corresponds to θ_0 in lemma 3.1, $\Theta \cap \bar{B}(\theta_0, \eta)$ coincides with the closure of its interior. Furthermore, $\Theta \cap \bar{B}(\theta_0, \eta) - \theta_0$ equals the intersection of a union of orthants and an open cube.

2. The function

$$\left(\theta^T, x_1, \ldots, x_l\right) \to g_\theta\left(\frac{y_1}{\sqrt{x_1}}, \ldots, \frac{y_p}{\sqrt{x_p}}, x_1, \ldots, x_l\right)$$

has continuous second order (l/r) partial derivatives differentiable on $\Theta \cap \bar{B}(\theta_0, \eta) \times (\mathbb{R}^{++})^l$ for every fixed $(y_1, \ldots, y_p) \in \mathbb{R}^p$.

3. The functions $\varsigma_{k,\theta}$ and $\zeta_{k,\theta}$ have continuous second order (l/r) partial derivatives on $\Theta \cap \bar{B}(\theta_0, \eta)$, $\mathbb{P}$ a.s., for all $k = 0, \ldots, l-1$ and $k = 0, \ldots, p-1$.

A.5.1 ensures that at any point of $\Theta \cap \bar{B}(\theta_0, \eta)$, there exists enough space around each of its elements so that a left and/or right perturbation can be defined, and its second part is essentially Assumption 2^{2*}.(a) of Andrews [3]. This implies that at any such point a left and/or right partial derivative could be in principle defined. A.5.2 and A.5.3 ensure that both g_θ and the initial conditions have well defined and continuous left and/or right second order partial derivatives. Given those, the Taylor approximation is valid on any K that is a non empty compact subset of $\Theta \cap \bar{B}(\theta_0, \eta)$ even if the coefficients of the relevant polynomials may depend on random elements that can take values outside K with positive $\mathbb{P}$ probability. Furthermore, since the vector $(x_1, \ldots, x_l)$ belongs to $(\mathbb{R}^{++})^l$ the relevant derivatives w.r.t. to the elements of this vector are by construction left and right. Due to the chain rule (see Appendix A. of Andrews [3]), they imply that the analogous derivatives of the filter (w.r.t. θ) are also well defined. In what follows we denote the matrices of first and second order (l/r) partial derivatives with $\cdot'$ and $\cdot''$ respectively. Their existence along with the form of $\hat{c}_n$ and Theorem 6 of Andrews [3] imply the $\mathbb{P}$ a.s. existence of a second order Taylor expansion of the likelihood function around θ_0. This does not suffice for the second part of Assumption A.8 to hold, and thereby theorem 5.2 cannot be directly used. The possibility of the existence, stationarity and ergodicity of h_t' and h_t'' along with the possibility that they provide geometric approximations of $\hat{h}_t'$ and $\hat{h}_t''$ respectively could enable the verification of the aforementioned conditions. The following assumption and the subsequent proposition takes care of this after the establishment of some notation.

Let k_i be the i-th element of the vector $\left(\theta^T, x_1, \ldots, x_l\right)$. Then for $i, j = 1, \ldots, d, \ldots, d+l$ define

$$\partial^i \psi_t\left(\theta^T, x_1, \ldots, x_l\right) = \frac{\partial}{\partial k_i} g_\theta\left(\frac{y_t}{\sqrt{x_1}}, \ldots, \frac{y_{t-p+1}}{\sqrt{x_p}}, x_1, \ldots, x_l\right)$$

and

$$\partial^{i,j} \psi_t\left(\theta^T, x_1, \ldots, x_l\right) = \frac{\partial^2}{\partial k_i \partial k_j} g_\theta\left(\frac{y_t}{\sqrt{x_1}}, \ldots, \frac{y_{t-p+1}}{\sqrt{x_p}}, x_1, \ldots, x_l\right).$$

Assumption A.6 *Suppose that:*

1. for $i = 1, \ldots, d, \ldots, d+l$

$$\mathbb{E}\left[\ln^+\left(\sup_{\theta \in \Theta \cap \bar{B}(\theta_0, \eta)} \left|\partial^i \psi_0\left(\theta, h_0(\theta), \ldots, h_{-l}(\theta)\right)\right|\right)\right] < +\infty$$

Furthermore for every $i = 1, \ldots, d, \ldots, d+l$, there exist a stationary sequence $\left(\bar{C}_{i,1}(t)\right)$

with $\mathbb{E}\left[\ln^{+}\bar{C}_{i,1}(0)\right]<\infty$ and some function $r_1:\mathbb{R}\to\mathbb{R}^{+}$ that is continuously differentiable in a compact neighborhood of zero and $r_1(0)=0$ such that

$$\sup_{\theta\in\Theta\cap\bar{B}(\theta_0,\eta)}\left|\partial^{i}\psi_t\left(\theta^T,x_1,\ldots,x_l\right)-\partial^{i}\psi_t\left(\theta^T,x_1',\ldots,x_l'\right)\right|\leq\bar{C}_{i,1}(t)\,r_1\left(|x-x'|\right), \quad (3)$$

where $x=(x_1,\ldots,x_l)$ and $x'=(x_1',\ldots,x_l')$ in $\left(\mathbb{R}^{l}\right)^{++}$.

2. For $i,j=1,\ldots,p,\ldots,p+q$

$$\mathbb{E}\left[\ln^{+}\left(\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left|\partial^{i,j}\psi_0\left(\theta^T,h_0(\theta),\ldots,h_{-l}(\theta)\right)\right|\right)\right] < +\infty,$$

$$\text{and } \mathbb{E}\ln^{+}\left(\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left|h_0'(\theta)\right|\right) < +\infty$$

Furthermore for every $i,j=1,\ldots,d,\ldots,d+l$, there exists a stationary sequence $\left(\bar{C}_{i,j,2}(t)\right)$ with $\mathbb{E}\left[\ln^{+}\bar{C}_{i,j,2}(0)\right]<\infty$ and some function $r_2:\mathbb{R}\to\mathbb{R}^{+}$ that is continuously differentiable in a compact neighborhood of zero and $r_2(0)=0$ such that

$$\sup_{\theta\in\Theta\cap\bar{B}(\theta_0,\eta)}\left|\partial^{i,j}\psi_t\left(\theta^T,x_1,\ldots,x_l\right)-\partial^{i,j}\psi_t\left(\theta^T,x_1',\ldots,x_l'\right)\right|\leq\bar{C}_{i,j,2}(t)\,r_2\left(|x-x'|\right). \quad (4)$$

This assumption essentially implies the existence and uniqueness of stationary and ergodic solutions to the SRE's obtained by (*l*/*r*) first and second order differentiation of the second equation in (1) w.r.t. θ. Furthermore, first those solutions are identified with h_t' and h_t'' which are continuous w.r.t the parameter and $\hat{h}_t'$, $\hat{h}_t''$ rapidly converge to their ergodic version uniformly in a neighborhood of θ_0 which without any damage to generality and for notational simplicity we assume that it coincides with $\Theta\cap\bar{B}(\theta_0,\eta)$. The derivation of the previous along with their implications on the asymptotic relation between the Taylor expansions of $\hat{c}_n$ and c_n are obtained in the proof of the following lemma.

Lemma 3.4 *Suppose that assumptions A.1, A.2, A.3, A.5 and A.6 hold. Then*

1. *h_t' and h_t'' are continuous w.r.t. θ, for all $t\in\mathbb{Z}$,*

$$\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\left\|h_t'(\theta)-\hat{h}_t'(\theta)\right\|\overset{e.a.s.}{\to}0 \text{ and } \sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\left\|h_t''(\theta)-\hat{h}_t''(\theta)\right\|\overset{e.a.s.}{\to}0, \text{ and} \quad (5)$$

$$n\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\|c_n'(\theta)-\hat{c}_n'(\theta)\| \text{ and } n\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\|c_n''(\theta)-\hat{c}_n''(\theta)\| \text{ converge } \mathbb{P} \text{ a.s.} \quad (6)$$

2. *If for some $(r_n)_{n\in\mathbb{N}}$ such that $r_n\to+\infty$, with $r_n=o(n)$ and $r_nc_n'(\theta_0)\rightsquigarrow z_{\theta_0}$ for z_{θ_0} some well defined random vector, then $r_n\hat{c}_n'(\theta_0)\rightsquigarrow z_{\theta_0}$, and*

3. *if $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\|\ell_0''(\theta)\|^2<+\infty$ then for any sequence $\vartheta_n\to\theta_0$ $\mathbb{P}$ a.s., $\hat{c}_n''(\vartheta_n)\rightsquigarrow$ $\mathbf{J}_{\theta_0}=\mathbb{E}\ell_0''(\theta_0)=\mathbb{E}\left(\frac{h_t'(\theta_0)[h_t'(\theta_0)]^T}{\sigma_t^4}\right)$.*

In order to be able to use the results in 2.1, 3.4 and 5.2 for the characterization of the rate and the limit distribution we need a final assumption that takes care of the asymptotic behavior of c_n' and c_n''. In what follows K denotes a compact non empty subset of Θ of possibly small enough diameter that *contains* θ_0 and is a subset of $\Theta \cap \bar{B}(\theta_0, \eta)$, such that $\theta_n \in K$ with $\mathbb{P}$-probability that converges to one as $n \to \infty$. Given theorem 3.3, K could for example be chosen as $\Theta \cap \bar{B}(\theta_0, \eta)$ itself. Furthermore, let $\mathcal{H}_n = \sqrt{n}\left(\sqrt{\wp(\sqrt{n})}\right)^{-1}(K - \theta_0)$ where the function $\wp$ is as in Theorem 2.1 and is specified next. The asymptotic parameter space in defined next as an appropriate limit of $\mathcal{H}_n$.

Definition D.3 $\mathcal{H} = \limsup_{n\to\infty} \mathcal{H}_n$ *i.e. it is the set containing any* $x \in \mathbb{R}^d$ *such that* x *is a cluster point of some* $(x_n)_{n\in\mathbb{N}}$ *with* $x_n \in \mathcal{H}_n$.

$\mathcal{H}$ is essentially the upper limit in the Painleve-Kuratowski sense of $(\mathcal{H}_n)_{n\in\mathbb{N}}$ (see for example Appendix B of Molchanov [24]). The definition is equivalent to that $x \in \mathcal{H}$ iff there exists an infinite subset of $\mathbb{N}$ (say $\mathcal{N}$) such that for any $\varepsilon > 0$, $\mathcal{H} \cap B(x, \varepsilon) \neq \varnothing$ for all $n \in \mathcal{N}$. Notice that $\mathcal{H}$ always exists and it is a closed subset of $\mathbb{R}^d$ (see Proposition 4.4 of Rockafellar and Wets [27]). In our case it is always different from $\varnothing$ since it contains 0. When θ_0 is an interior point then $\mathcal{H} = \mathbb{R}^d$. This definition is not less general compared to Assumption 5 of Andrews [3] as Lemma 3.8 of Arvanitis and Louka [4] implies.

Assumption A.7 *Suppose that:*

1. *For the second moment of the distribution of* $(z_0^2 - 1)$, *we have that*

$$\wp(x) \backsim \int_{-1}^{x} \left(z_0^2 - 1\right)^2 dF_{z_0^2-1}, \text{ as } x \to +\infty$$

   where $\wp$ *conforms 2.1.2.*

2. $\mathbb{E}\left\|\frac{h_0'(\theta_0)}{h_0(\theta_0)}\right\|^{2\lambda} < +\infty$ *with* $\lambda \geq 1$ *if* $\wp$ *has a limit as* $x \to \infty$, *otherwise* $\lambda > 1$. *Furthermore,* $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0'}{h_0}\right\|^2 < +\infty$, $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0''}{h_0}\right\|^{\lambda'} < +\infty$ *with* $\lambda' > 1$ *and* $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\frac{\sigma_0^{2\lambda^*}}{h_0^{\lambda^*}} < +\infty$ *for some* $\lambda^* \geq \max\left(2, \frac{1}{1-\frac{1}{\lambda'}}\right)$ *such that* $\mathbb{E}\left|z_0\right|^{2\lambda^*} < +\infty$.

3. *The components of the vector* $\frac{\partial}{\partial\theta}g_\theta\left(\frac{y_{t-1}}{\sqrt{h_{t-1}(\theta)}}, \ldots, \frac{y_{t-p-1}}{\sqrt{h_{t-p-1}(\theta)}}, h_{t-1}(\theta), \ldots, h_{t-q-1}(\theta)\right)\Big|_{\theta=\theta_0}$ *are linearly independent random variables.*

4. $\mathcal{H}$ *is convex.*

A.7.1 and the first part of A.7.2 enable the use of theorem 2.1. Notice that when $\wp(x)$ has a limit then this is $\mathbb{E}z_0^4 - 1$, while when it diverges as $x \to +\infty$ then equivalently $\wp(x) \backsim \int_{-\sqrt{2(x+1)}}^{\sqrt{2(x+1)}} z_0^4 dF_{z_0}$ as $x \to +\infty$, due to the fact that by definition $\mathbb{E}z_0^2 = 1$. The second part of A.7.2 also along with lemma 3.4, theorem 3.3 and the ULLN for stationary and ergodic sequences imply the convergence in probability of $\hat{c}_n''(\theta_n)$ to $\mathbb{E}\ell_0''(\theta_0)$. Notice also that in a variety of heteroskedastic models (see for example paragraph 5.7.1 of Straumann [28] concerning the $\mathrm{AGARCH}(p, q)$ model) under the appropriate conditions, the

$\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0'}{h_0}\right\|^{\lambda}, \mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0''}{h_0}\right\|^{\lambda'} < +\infty$ holds for arbitrarily large λ, λ'. This implies that under additional conditions concerning the behavior of the distribution of z_0 in shrinking neighborhoods of zero (see for example condition 4.5 in Theorem 4.1 of Berkes et al. [5]) $\mathbb{E}\sup_K \frac{\sigma_0^{2\lambda^*}}{h_0^{\lambda^*}} < +\infty$ holds for λ^*arbitrarily close to 1 and this is also conforming to $\mathbb{E}\left|z_0\right|^{2\lambda^*} < +\infty$ even when $\wp$ diverges due to Theorem 2.6.4 of Ibragimov and Linnik [19]. A.7.3 implies that $\mathbb{E}\ell_0''(\theta_0)$ is positive definite. A.7.4 implies the uniqueness of the limit established in the final theorem and it is analogous to Assumption 6 of Andrews [3]. The following counterexample implies that condition A.7.4 is not trivial by considering a K with empty interior.

Example: K is comprised by the elements and the limit of a converging sequence. Let $(\gamma_m)_{m\in\mathbb{Z}}$ denote a real sequence that converges to zero and suppose without loss of generality that $\theta_0 = 0$. For some x in $\mathbb{R}^d$ let $K = K - \theta_0 = \{\gamma_m x, m \geq 1\} \cup \{0\}$. If $c = \lim_{m\to\infty}\gamma_m\frac{\sqrt{m}}{\sqrt{\wp(\sqrt{m})}}$ then due to the defining properties of $\wp$, $\mathcal{H} = \left\{\frac{c}{\sqrt{k}}x, k = 1, 2, \ldots\right\}\cup\{0\}$. Obviously A.7.4 fails if $x \neq 0$.

If K itself contains a set of the form $\Theta\cap\bar{B}(\theta_0,\eta^*)$ with $0 < \eta^* \leq \eta$ then condition A.7.4 implies that $\mathcal{H}$ coincides with the closure of its interior. This is due to the fact that $K - \theta_0$ must contain a neighborhood of zero of the form $\prod_{i=1}^k [l_i, u_i]$ where some of the lower or upper bounds could be zero but not simultaneously for the same i. Choose an arbitrary non zero point in the previous set. It is easy to see that this belongs to $\mathcal{H}_n$ for all n and thereby to $\mathcal{H}$ which is by construction convex.

The following theorem is essentially the second main result of the paper.

Theorem 3.5 *Suppose that assumptions A.1, A.2, A.3, A.4, A.5, A.6 and A.7.1-3 hold. If $\varepsilon_n = O_p\left(\frac{\wp(\sqrt{n})}{n}\right)$ then*

$$\sqrt{\frac{n}{\wp(\sqrt{n})}}(\theta_n - \theta_0) = O_p(1).$$

If furthermore A.7.4 holds and $\varepsilon_n = o_p\left(\frac{\wp(\sqrt{n})}{n}\right)$, then

$$\sqrt{\frac{n}{\wp(\sqrt{n})}}(\theta_n - \theta_0) \rightsquigarrow \tilde{h}_{\theta_0}$$

where $\tilde{h}_{\theta_0}$ is uniquely defined by $q\left(\tilde{h}_{\theta_0}\right) = \inf_{h\in\mathcal{H}} q(h)$ and $q(h) := \left(h - \mathbf{J}_{\theta_0}^{-1}z_{\theta_0}\right)'\mathbf{J}_{\theta_0}\left(h - \mathbf{J}_{\theta_0}^{-1}z_{\theta_0}\right)$ for $\mathbf{J}_{\theta_0}$=$\mathbb{E}\ell_0''(\theta_0) = \mathbb{E}\left(\frac{h_t'(\theta_0)[h_t'(\theta_0)]^T}{\sigma_t^4}\right)$ which is positive definite and $z_{\theta_0} \sim N(0, \mathbf{J}_{\theta_0})$.

Notice from the proof of theorem 5.2 that in the case that condition A.7.4 fails theorem 3.5 could retain some information for the limit distribution of the estimator, in the sense that if $\mathcal{H}$ is such that $\arg\min_{\mathcal{H}} q(h)$ is non empty, then the limit in distribution of the latter is "hiding" inside this set of minimizers. Furthermore the issue of the non unique choice of K is unimportant w.r.t. to the characterization of the limit since the latter is unique. Moreover if K can be chosen so that it contains a set of the form $\Theta\cap\bar{B}(\theta_0,\eta^*)$ with $0 < \eta^* \leq \eta$, then $\mathcal{H}$ is essentially independent of the choice of K and thereby this is also true for the *representation* of the limit $\tilde{h}_{\theta_0}$ as a minimizer. This is due to the facts that $\mathbf{J}_{\theta_0}^{-1}z_{\theta_0}$ follows

a non degenerate Normal distribution and that any two convex subsets of $\mathbb{R}^d$ that coincide with the closure of their interior if they are non equal then they must differ at interior points. Those imply that the function $\inf_{h\in\cdot} q(h)$ must be bijective when defined on the collection of closed non empty convex subsets of $\mathbb{R}^d$. Hence any other compatible choice of K would result to the same asymptotic parameter space.

Now, theorem 3.5 encompasses the results of Theorem 5.6.1 of Straumann [28] when θ_0 is an interior point and $\mathbb{E}z_0^4 < +\infty$. In this case the rate is $\sqrt{n}$ and the limit distribution is $N\left(0, \left(\mathbb{E}z_0^4 - 1\right)\mathbf{J}_{\theta_0}^{-1}\right)$. In the same case when θ_0 is a boundary point then the rate is again $\sqrt{n}$ and $\sqrt{n}\left(\theta_n - \theta_0\right) \rightsquigarrow \sqrt{\mathbb{E}z_0^4 - 1}\tilde{h}_{\theta_0}$. When $\wp$ is diverging and θ_0 is an interior point, then $\sqrt{\frac{n}{\wp(\sqrt{n})}}\left(\theta_n - \theta_0\right) \rightsquigarrow N\left(0, \mathbf{J}_{\theta_0}^{-1}\right)$ a novel result in this framework, that implies that we can obtain asymptotic normality even in cases where the fourth moments do not exist albeit at a *slower rate*. The following corollary, linked with corollary 2.2, handles the case where z_0 follows a normalized Student' s t distribution with 4 degrees of freedom as a prominent example of a diverging truncated fourth moment.

Corollary 3.6 *Suppose that $\sqrt{2}z_0 \sim t_4$ and assumptions A.1, A.2, A.3, A.4, A.5, A.6 and A.7.2-3 hold with $\delta = 2$. If $\varepsilon_n = O_p\left(\frac{\ln n}{n}\right)$ then*

$$\sqrt{\frac{n}{\ln n}}\left(\theta_n - \theta_0\right) = O_p\left(1\right).$$

If furthermore A.7.4 holds and $\varepsilon_n = o_p\left(\frac{\ln n}{n}\right)$, then

$$\sqrt{\frac{n}{\ln n}}\left(\theta_n - \theta_0\right) \rightsquigarrow \sqrt{\frac{3}{2}}\tilde{h}_{\theta_0}$$

where $\tilde{h}_{\theta_0}$ is as in theorem 3.5.

The following example concerns the verification of the assumption framework above (given the a priori validity of the assumption A.3 and of the conditions A.4.1, A.5.3 and A.7.3) for the $\mathrm{GQARCH}\left(1,1\right)$ model introduced by Sentana [29].

GQARCH(1,1) Let $p = q = 1$, $g_\theta\left(z_{t-1}, x\right) = \omega + \alpha\left(z_{t-1}\sqrt{x} + \frac{\gamma}{2\alpha}\right)^2 + \beta x$ and Θ is a compact subset of $\mathbb{R}^{++} \times \mathbb{R}^{++} \times \mathbb{R}^{-} \times [0,1)$. The results referenced below are established in Arvanitis and Louka [4]. Remark R.1 implies that there exist $\theta_0 \equiv (\omega_0, \alpha_0, \gamma_0, \beta_0) \in \Theta$ such that Assumption A.1 is satisfied. This is permitted even in cases where $\alpha_0 + \beta_0 > 1$ implying the existence of solutions with the required properties that are not covariance stationary. Assumption A.2 is satisfied since $\beta < 1$. For assumption A.4.2 and A.4.4 see lemma 2.2 and lemma 3.3. For assumption A.4.3 see lemma 2.1. The latter holds of the distribution of z_0 is not concentrated in two points. For Assumption A.5.1 see lemma 3.5 and for the rest of Assumption A.5 and Assumption A.6 see the proofs of lemmata 4.7 and 4.8. If $\mathbb{P}\left(z_0^2 \leq t\right) = o\left(t^\mu\right)$ for $t \downarrow 0$, then Assumption A.7.2-3 follows from lemmata 4.2, 4.3 and 4.6 which hold even if $\wp$ is diverging since $\mathbb{E}\left|z_0\right|^\lambda < +\infty$ for any $0 < \lambda < 4$ due to Theorem 2.6.4 of Ibragimov and Linnik [19]. Examples of Θ's that satisfy condition A.7.4 of Assumption A.7 can also be found in section 3.3. of Arvanitis and Louka [4].

Several other examples can be constructed using the work of Straumann [28]. For instance the verification of assumptions A.1, A.2, A.4, A.5, A.6 and A.7.2-3, for the $\mathrm{AGARCH}\left(p,q\right)$

model when the parameter space is appropriately restricted and under the condition $\mathbb{P}\left(z_0^2 \leq t\right) = o\left(t^{\mu}\right)$ as $t \downarrow 0$, would follow from the results appearing in examples 5.2.5 and 5.2.11 and paragraphs 5.4.2 and 5.7.1 of Straumann [28] by noticing first that for the set $[\omega_{\min}, +\infty) \times [0, +\infty)^p \times B \times [-1, 1]$ for $\omega_{\max} > 0$ and $B = \{x \in (\mathbb{R}^+)^q : \sum_{i=1}^q x_i < 1\}$ assumption A.5.1 follows, second that Lemma 5.7.3 of Straumann [28] can be seen to hold even when θ_0 lies on the boundary of Θ which by construction is a compact subset of the previous set with non empty interior (as long as the conditions $a_{i_0} \neq 0$ for some $i = 1, 2, \ldots, p$ and $\left(a_{p_0}, \beta_{q_0}\right) \neq (0, 0)$-simply express the linear combination w.r.t. a_{i_0} instead of a_{1_0}) and third that this is also true for lemmata 5.1 and 3.2 of Berkes et al. [5] as well as for their extensions concerning the $\mathrm{AGARCH}\left(p, q\right)$ case, i.e. lemmata 5.7.4 and 5.7.5 of Straumann [28]. Notice that it is easy to construct examples of Θ for which both the previous assumptions and condition A.7.4 of Assumption A.7 hold. Consider for example the case where $p = 2$, $q = 1$ and $\Theta = [\omega_{\min}, \omega_{\max}] \times [0, a_{1_{\max}}] \times [a_{2_{\min}}, a_{2_{\max}}] \times [\beta_{\min}, \beta_{\max}] \times [-1, 1]$ with the obvious notation, where $a_{2_0} = a_{2_{\max}}$, $\gamma_0 = -1$, and the other elements of θ_0 lie in the interior of their defining intervals, in which case $\mathcal{H} = \mathbb{R}^2 \times (-\infty, 0] \times \mathbb{R} \times [0, +\infty)$. In such a case it is easy to see that $\tilde{h}_{\theta_0} = \left(L'\right)^{-1}\left(z_1, z_2, \min\left\{0, z_3\right\}, z_4, \max\left\{0, z_5\right\}\right)'$ where $\left(z_1, z_2, z_3, z_4, z_5\right)' \sim N\left(0, \mathrm{Id}_5\right)$ and $\mathbf{J}_{\theta_0} = LL'$.

It is also easy to extend the assumption framework so that the recursions in (1) and 3.2 define not the volatility processes per se but their composition with some common *bijective* transformation. If assumptions A.1 and A.2 hold w.r.t the transformed processes, assumption A.3 incorporates the condition that its inverse (the *link* function as termed by Wintenberger and Cai 3.1) is continuous, and assumption A.6 is augmented by the condition that the inverse has first and second derivatives that are Lipschitz continuous on the bounded away from zero-due to condition A.4.3- domain of the volatilities, then theorem 3.5 would also hold.

This however would not suffice for the full scope of theorem 3.5 to incorporate a model such as the EGARCH one, at least using the approach adopted in paragraph 5.7.2 of Straumann [28], or in Demos and Kyriakopoulou [10]. This is due to the fact that given the aforementioned results, the second part of assumption A.7.2 would not hold if $\mathbb{E}\left(z_0^4\right) = +\infty$, even though the remaining conditions would be easily validated. The results in paragraph 4 of Wintenberger and Cai [33] or in lemma 3 of Wintenberger [34] could indicate a possible alternative route via the substitution of this condition by a much weaker one. We are not pursuing this line of reasoning any further since it is out of the scope of the present paper.

Dropping the Stationarity Assumption: A Simple ARCH(1) Example The conditions imposed for the validity of theorem 2.1 incorporate first the restriction that the stochastic sequence $\left(V_t\right)_{t \in \mathbb{N}}$ is stationary and ergodic and second that it is also appropriately integrable. It is easy to see from the proof of the theorem that these restrictions essentially facilitate the use of a local representation of an appropriate conditional characteristic function and the convergence of some random series. If these results can be established without those conditions, then analogues of the theorem can be obtained even in cases of *non stationarity*. The following theorem provides with such an example that essentially exhibits the analytical strength of the principle of conditioning. Remember that the filtration $\left(\mathcal{F}_t\right)_{t \in \mathbb{N}}$ is defined by $\mathcal{F}_t \triangleq \sigma\left(\xi_t, V_t, \xi_{t-1}, V_{t-1}, \xi_{t-2}, V_{t-2}, \ldots\right)$.

Theorem 3.7 *Let the following hold:*

1. $\left(\xi_t\right)_{t \in \mathbb{N}}$ *is an* iid *sequence of random variables and* $\left(V_t\right)_{t \in \mathbb{N}}$ *is an real valued sequence of random elements such that* ξ_t *is independent of* $\mathcal{F}_{t-1}$ *and* V_t *is measurable w.r.t.* $\mathcal{F}_{t-1}$.

2. *For the second truncated moment of the distribution of* ξ_0,

$$\wp(x) \backsim \int_{-x}^{x} \xi_0^2 dF_{\xi_0},$$

where $\wp$ *is slowly varying.*

3. $V_n^2 \to v > 0$, $\mathbb{P}$ *a.s. as* $n \to \infty$.

Then as $n \to \infty$

$$\frac{1}{\sqrt{n}\wp^\star(n)} \sum_{t=1}^{n} \xi_t V_t \rightsquigarrow N(0, v).$$

where $\wp^\star$ *is a slowing varying function satisfying* $\wp(\sqrt{x}\wp^\star(x)) \backsim (\wp^\star(x))^2$.

Remark R.5 *For the definition of* $\wp^\star$*see remark R.3. Furthermore, notice from the same remark that if* $\wp$ *satisfies* $\frac{\wp(x\sqrt{\wp(x)})}{\wp(x)} \to 1$ *then* $\wp^\star(x)$ *can be chosen to be* $\sqrt{\wp(\sqrt{x})}$.

Comparing theorem 3.7 with 2.1 we have that the existence of a $\mathbb{P}$ a.s. (strictly positive) limit for the $(V_t^2)_{t\in\mathbb{N}}$ substitutes the stationarity and ergodicity condition, the existence of appropriate moments for $(V_t)_{t\in\mathbb{N}}$ as well as the restrictions on the asymptotic behavior of $\wp$. The result holds for *any* slowly varying function. A simple application of this concerns the limit theory of the QMLE in the case of a simple non stationary $\mathrm{ARCH}(1)$ model. Notice that under the assumption of the existence of a limit for $\wp$, this theory was established by Jensen and Rahbek [18].

To this end, let $p = q = 1$, $g_\theta(z_{t-1}, x) = 1 + \alpha z_{t-1}\sqrt{x}$ and $\Theta = [a_l, a_u]$ where $a_l \geq \exp(-2\mathbb{E}\ln|z_0|) > 0$. Notice that due to the results in Nelson [25] any solution to the previous recursion is *non stationary*. Suppose *without loss of generality* that $\theta_0 \in (a_l, a_u)$ and consider the QMLE (with zero optimization error) for the θ_0. We obtain the following proposition that is a direct consequence of theorem 3.7 in conjunction with lemmata 1 and 2 in Jensen and Rahbek [18].

Proposition 3.8 *Suppose that for the second moment of the distribution of* $(z_0^2 - 1)$, *we have*

$$\wp(x) \backsim \int_{-1}^{x} \left(z_0^2 - 1\right)^2 dF_{z_0^2-1}, \text{ as } x \to +\infty$$

where $\wp$ *conforms to condition 3.7.2, that the conditions described in the previous paragraph hold, as well as that* $\varepsilon_n = 0$ $\mathbb{P}$ *a.s. Then*

$$\frac{\sqrt{n}}{\wp^\star(n)}(\theta_n - \theta_0) \rightsquigarrow N\left(0, \theta_0^2\right).$$

Obviously this is an extension of Theorem 1 of Jensen and Rahbek [18] that establishes the asymptotic normality (in the interior case) of the QMLE in this simple model even when stationarity fails. It could be interesting to explore analogous results in more complex models.

4 Monte Carlo Evidence

In this section we perform two Monte Carlo experiments to assess the quality of the approximation of the limit distribution of the QMLE in finite samples. Using the framework described in section 3 we examine the simple $\mathrm{GARCH}\,(1,1)$ model $\sqrt{2}z_0 \sim t_4$, thus letting $\sigma_t^2 = g_{\theta_0}\left(z_{t-1}, \sigma_{t-1}^2\right) = \omega_0 + \left(\alpha_0 z_{t-1}^2 + \beta_0\right)\sigma_{t-1}^2$ in (1). In the first experiment we let θ_0 be an interior point and choose $\omega_0 = 0.5$, $\alpha_0 = 0.05$, $\beta_0 = 0.90$, while in the second case we examine the case where θ_0 is a boundary point choosing $\omega_0 = 0.5$, $\alpha_0 = 1$, $\beta_0 = 0$ pertaining to the $\mathrm{ARCH}(1)$ case. In each experiment we consider $n = 100$, 1000, 10000 and 100000 and the number of Monte Carlo simulations is 1000. For every sample we plot the Kaplan-Meier estimate of the cumulative distribution function over the Monte Carlo simulations separately for each element of the $\theta_n - \theta_0$ multiplied by $\sqrt{\frac{n}{\ln n}}$. By Corollary 3.6 in the first case $\tilde{h}_{\theta_0} \sim N\left(0, \mathbf{J}_{\theta_0}^{-1}\right)$ and in the second case calculations show that $\tilde{h}_{\theta_0} = (L')^{-1}\left(z_1, z_2, \max\{0, z_3\}\right)'$ where $(z_1, z_2, z_3)' \sim N\left(0, 3/2\,\mathrm{Id}\right)$ and $\mathbf{J}_{\theta_0} = LL'$. Furthermore, in the first case $\mathbf{J}_{\theta_0}$ is approximated by computing the simulated average of the outer product of the gradient of $\ln h_T(\theta)$ at θ_0 via numerical procedures over samples of size $T = 1000$. In the second case $\frac{h_t'(\theta_0)[h_t'(\theta_0)]^T}{\sigma_t^4}$ has a simple analytical expression which depends on y_{t-1} and σ_{t-1}. We approximate its expectation by using a double average across samples of size 10000000 and over 1000 simulations for increased precision. Then the limit distributions are constructed for each element of θ by plotting the Kaplan-Meier estimate of the cumulative distribution function of independent draws of the latter distributions. The relevant figures are presented in the Appendix. For economy of space, in the first case we provide only the figures concerning β, as the rest are similar, and in the second case we present the figures concerning α and β due to the analogous similarity of the results concerning a and ω. We observe that although the empirical distributions in question seem to show some heavy tail behavior, they resemble those of the theoretical limits as the sample size increases.

5 Conclusions

We have provided a CLT to a normal limit for the standardized sum of a martingale transform, constructed by point-wise multiplication between an iid and a stationary and ergodic sequence that holds even in cases where the relevant second moments diverge at an appropriately slow rate. This extends relevant results with stable but non normal limits to the case of asymptotic normality, as well as the results concerning asymptotic normality but allowing domains of non-normal attraction. In those cases the rate is slower that $\sqrt{n}$ and it contains information for the rate of divergence of the truncated second moments of the relevant sequence.

A major application concerns the characterization of the rate and the limiting distribution of the score vector of the Gaussian quasi likelihood function in the case of GARCH type models, and the subsequent characterization of the limit theory for the QMLE. Given this we have also reviewed the analogous framework for the establishment of the relevant limit theory and extended it, so that it can accommodate the case of slowly varying fourth moments for the innovation process by an application of the aforementioned theorem and the possibility that the parameter of interest lies on the boundary of the parameter space.

Possible probabilistic extensions concern, first, the establishment of analogous results if we allow the rate of divergence of the second moments to be described by larger classes of slowly varying functions. Second, the extension of the results when the truncated moment of order a is slowly varying, for $a \in (1, 2)$ whereby the limit would be a multivariate a-stable

distribution and the complete characterization of which could be achieved by considering the set of non trivial linear transformations of the relevant random vector. This could provide with special cases of the results of Mikosch and Straumann [23], without the need for verification of the mixing condition that they are using, in which both the rate and the characteristics of the limit are completely described. In such cases we also expect that the rate would be slower than the one obtained in the normal case and to be of the form $\frac{n^{\frac{a-1}{a}}}{\wp^{1/a}\left(n^{1/a}\right)}$.

Finally, a possible statistical application would concern the issue of the existence of appropriate studentizations of the Wald statistic that is based on the QMLE considered before, under the maintained hypothesis of the domain of (normal or not) attraction to the normal distribution. This could be useful in the cases of the non normal domain for the obvious asymptotic size corrections, while retaining consistency in the normal domain case.

References

[1] Abadir K. and J. Magnus (2004), "The Central Limit Theorem for Student's Distribution", *Econometric Theory*, Vol. 20, No. 6, pp. 1261-1263.

[2] Aliprantis C.D. and K.C. Border (1999), "*Infinite Dimensional Analysis. A Hitchhiker's Guide.*", Springer.

[3] Andrews, D. W. K., "Estimation when a parameter is on boundary", *Econometrica, 67:1341-1383, 1999.*

[4] Arvanitis S. and A. Louka (2013), "Limit Theory for the QMLE of the GQARCH(1,1) model", *Communications in Statistics-Theory and Methods* (to appear).

[5] Berkes, I. , L. Horvath, and P. Kokoszka, "GARCH processes: structure and estimation", *Bernoulli,* 9:201-228, 2003.

[6] Bingham N.H., Goldie C.M. and J.L. Teugels (1992), "*Regular Variation*", Encyclopedia of Mathematics and Its Applications, Vol. 27, C.U.P.

[7] Calzolari G., Fiorentini G. and E. Sentana (2003), "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models with Student t Innovations", *Journal of Business and Economic Statistics*, 21, 4, 532-546.

[8] Choirat C., Hess C., and R. Seri (2003), "A Functional Version of the Birkhoff Ergodic Theorem for a Normal Integrand: A Variational Approach.", *The Annals of Probability*, Vol. 31, No 1, pp. 63-92.

[9] Davidson J. (1995), "*Stochastic Limit Theory.*", Advanced Texts in Econometrics, Oxford University Press.

[10] Demos A. and D. Kyriakopoulou (2009), "Asymptotic normality of the QMLE in the EGARCH (1, 1) model". working paper, 2013.

[11] Diebold, F. (1988). "*Empirical modeling of exchange rate dynamics*". Springer-Verlag.

[12] Engle, R. F. and Bollerslev, T. (1986). "Modelling the persistence of conditional variances". *Econometric Reviews*, 5 1-50.

[13] Hall P. and Yao Q., "Inference in ARCH and GARCH Models with Heavy Tailed Errors", *Econometrica*, 71, 285-317, 2003.

[14] Hansen, B. E. (1994). Autoregressive conditional density estimation. International Economic Review, 705-730.

[15] Hsieh, D. A. (1989). "Modeling heteroscedasticity in daily foreign-exchange rates". *Journal of Business and Economic Statistics*, 7, 307-317.

[16] Jacubowski A. (1986), "Principle of Conditioning In Limit Theorems For Sums of Random Variables", *The Annals of Probability*, Vol. 14, No 2, pp. 902-915.

[17] Jacubowski A. (2012), "Principle of Conditioning Revisited", *Demonstratio Mathematica*, Vol. XLV, No 2.

[18] Jensen, S. T., and Rahbek, A. (2004), "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case". *Econometrica*, 72(2), 641-646.

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

[20] Klein Er., and Thompson A.C. (1984), "*Theory of Correspondences.*", CMS Series of Monographs, Wiley Interscience.

[21] Kwapien S., W.A. Woyczynski (1992), "*Random Series and Stochastic Integrals: Single and Multiple*", Probability and Its Applications, Birkhauser.

[22] Lachout P., E. Liebsher, and S. Vogel, (2005), "Strong Convergence of Estimators as ε_n-Minimisers of Optimization Problems.", *Ann. Inst. Stat. Math.*, Vol. 57-2, pp. 291-313.

[23] Mikosch T. and D. Straumann, "Stable Limits of Martingale Transforms With Application to the Estimation of GARCH Parameters", *The Annals of Statistics*, Vol.34, No 1, 439-522, 2006.

[24] Molchanov, Ilya (2004), "*Theory of Random Sets.*", Probability and Its Applications, Springer.

[25] Nelson, D. B. (1990). "Stationarity and persistence in the GARCH(1,1) model". *Econometric theory*, 6(03), 318-334.

[26] Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach". *Econometrica*, 59, 347-370.

[27] Rockafellar T.R. and Wetts J-B (1997), "*Variational Analysis.*", Springer-Verlag.

[28] Straumann, D. (2004), "*Estimation in Conditionally Heteroscedastic Time Series Models.*", Springer.

[29] Sentana, E., "Quadratic ARCH Models", *The Review of Economic Studies*, 62:639-661, 1995.

[30] Surgailis, D., "A Quadratic ARCH(∞) Model with Long Memory and Levy Stable Behavior of Squares", *Adv. Appl. Prob.*, 40:1198-1222, 2008.

[31] van der Vaart, A.W., "*Asymptotic Statistics*" CUP, 1997.

[32] van der Vaart, A.W. and J.A. Wellner, "*Weak Convergence and Empirical processes*" Springer, 2000.

[33] Wintenberger, O., and S. Cai, "Parametric Inference and Forecasting for Continuously Invertible Volatility Models", http://arxiv.org/pdf/1106.4983.pdf.

[34] Wintenberger, O. (2012), "QML estimation of the continuously invertible EGARCH(1,1) model". *arXiv preprint* arXiv:1211.3292.

Appendix

Proofs

Proof of Theorem 2.1. When $\wp$ has a limit the result follows directly form the CLT for squared integrable stationary and ergodic martingale difference sequences (see for example the more general Theorem 24.3 of Davidson [9]) along with the Cramer-Wold device (see Theorem 25.5 of Davidson [9]). When $\wp$ does not have a limit, consider first the case where $d = 1$. Consider $(\xi_k^*)_{k\in\mathbb{Z}}$ to be an independent copy of $(\xi_k)_{k\in\mathbb{Z}}$ that is also independent of $(V_k)_{k\in\mathbb{Z}}$ and without loss of generality adopted to $(\mathcal{F}_k)_{k\in\mathbb{Z}}$, by possibly enlarging $\mathcal{F}_k$ since even in this case the original sequences would satisfy the same properties w.r.t. the enlarged filtration. Using the notation of Kwapien and Woyczynski [21] define $Y_{n,k} = \frac{1}{\sqrt{n\wp(\sqrt{n})}}\xi_k^* V_k$, $X_{n,k} = \frac{1}{\sqrt{n\wp(\sqrt{n})}}\xi_k V_k$, $\mathcal{F}_{n,k} \triangleq \mathcal{F}_k$, $\mathcal{G}_n = \sigma\{V_k, k \le n\}$ and notice that the tangency condition w.r.t. $(\mathcal{F}_{n,k})_{n\in\mathbb{N},k\in\mathbb{Z}}$ and the conditional independence condition w.r.t. $\mathcal{G}_n$ of Kwapien and Woyczynski [21] (see their Definitions 4.3.1 and 4.3.2. respectively) are valid for the sequences $(Y_{n,k})_{n\in\mathbb{N},k\in\mathbb{Z}}$, $(X_{n,k})_{n\in\mathbb{N},k\in\mathbb{Z}}$, $(\mathcal{F}_{n,k})_{n\in\mathbb{N},k\in\mathbb{Z}}$ and $(\mathcal{G}_n)_{n\in\mathbb{N}}$. Hence by Theorem 5.8.3 of Kwapien and Woyczynski [21] the result would follow if we prove that for all $t \in \mathbb{R}$

$$\mathbb{E}\left(\exp\left(it\frac{1}{\sqrt{n\wp(\sqrt{n})}}\sum_{k=1}^n \xi_k^* V_k\right)/\mathcal{G}_n\right) \underset{p}{\to} \exp\left(-\frac{t^2\mathbb{E}(V_0^2)}{2}\right).$$

Then notice that due to the assertion 2.6.21 and Theorem 2.6.1 of Ibragimov and Linnik [19] we have that the characteristic function $f(t)$ of ξ_0 and thereby of ξ_0^* has the following representation for $t \in (-t_0, t_0)$, and some $t_0 > 0$

$$\ln f(t) = -\frac{6}{2}|t|^2\wp(|t|^{-1}).$$

Now, fix $t \neq 0$ and define the event

$$C_{n,K} := \{\omega \in \Omega : |V_i| \le K_t\sqrt{n\wp(\sqrt{n})} \quad ,\forall i = 1,\dots,n\}$$

where $K_t < \frac{t_0}{|t|}$. Then

$$\begin{aligned}\mathbb{P}(C_{n,K}^c) &\le \sum_{i=1}^n \mathbb{P}[|V_i| > K_t\sqrt{n\wp(\sqrt{n})}] \\ &\le \sum_{i=1}^n \frac{\mathbb{E}[V_0^2]}{K_t^2 n\wp(\sqrt{n})} = \frac{\mathbb{E}{V_0}^2]}{K_t^2\wp(\sqrt{n})} \to 0,\end{aligned}$$

due to the third condition and Markov's inequality. Combining the previous, if $\omega \in C_{n,K}$ then $\ln E\left(\exp\left(it\frac{1}{\sqrt{n\wp(\sqrt{n})}}\sum_{k=1}^{n}\xi_k V_k\right)/\mathcal{G}_n\right)$ equals

$$
\begin{aligned}
&-\frac{t^2}{2n\wp\left(\sqrt{n}\right)}\sum_{i=1}^{n}V_i^2\wp\left(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t||V_i|}\right)\\
&=-\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t||V_i|}\right)}{\wp\left(\sqrt{n}\right)}\\
&=-\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2\left[\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t||V_i|}\right)}{\wp\left(\sqrt{n}\right)}-1\right]-\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2.
\end{aligned}
$$

It suffices to prove that the last term in the previous display converges in probability to $-\frac{t^2}{2}\mathbb{E}V_0^2$. Due to the first condition $(V_n^2)_{n\in\mathbb{N}}$ is stationary and ergodic and $\mathbb{E}V_0^2$ exists. Hence from the continuous mapping theorem and Birkhoff's LLN $-\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2 \underset{p}{\rightarrow} -\frac{t^2}{2}\mathbb{E}V_0^2$. It suffices that the remaining term converges in probability to zero. From the properties if $\wp\left(x\right)$, its Karamata representation and the mean value theorem we have that for any $x\in\mathbb{R}$

$$
\begin{aligned}
&\wp\left(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\exp\left(x\right)\right)\\
=\;&\wp\left(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\right)+\wp'\left(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\exp\left(x^*\right)\right)\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\exp\left(x^*\right)x\\
=\;&\wp\left(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\right)+\epsilon(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\exp\left(x^*\right))\wp(\frac{\sqrt{n\wp\left(\sqrt{n}\right)}}{|t|}\exp\left(x^*\right))x
\end{aligned}
$$

where x^* lies between x and zero. Thereby letting $x=-\ln|V_i|$ we obtain for any $\omega\in C_{n,K}$ and any i

$$
\begin{aligned}
&\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t||V_i|}\right)}{\wp\left(\sqrt{n}\right)}-1\right|\\
\leq\;&\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\right)}{\wp\left(\sqrt{n}\right)}-1\right|+\frac{\left|\epsilon(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\exp\left(x^*\right))\wp(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\exp\left(x^*\right)\right|}{\wp\left(\sqrt{n}\right)}\left|\ln|V_i|\right|\\
\leq\;&\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\right)}{\wp\left(\sqrt{n}\right)}-1\right|+\frac{C}{\wp\left(\sqrt{n}\right)}\left|\ln|V_i|\right|
\end{aligned}
$$

due to the fact that $\epsilon(x)\wp(x)$ is by assumption bounded. Hence

$$\mathbb{P}(\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t||V_i|}\right)}{\wp(\sqrt{n})}-1\right| > \varepsilon) \leq$$

$$\mathbb{P}(\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\right)}{\wp(\sqrt{n})}-1\right|\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2 > \frac{\varepsilon}{2}) + \mathbb{P}(\frac{C}{\wp(\sqrt{n})}\frac{t^2}{2n}\sum_{i=1}^{n}V_i^2\left|\ln|V_i|\right| > \frac{\varepsilon}{2}).$$

Due to stationarity and ergodicity and the assumed existence of moments the $\frac{1}{n}\sum_{i=1}^{n}V_i^2$ and $\frac{1}{n}\sum_{i=1}^{n}V_i^2\left|\ln|V_i|\right|$ are bounded in $\mathbb{P}$-probability. Using again a mean value expansion we have that for any y large enough

$$\wp\left(\frac{xy}{|t|}\right) = \wp\left(\frac{x}{|t|}\right) + \epsilon\left(\frac{xy^*}{|t|}\right)\wp\left(\frac{xy^*}{|t|}\right)\left(\frac{y}{y^*}-\frac{1}{y^*}\right)$$

where y^* is similarly between y and 1 and thereby letting $x=\sqrt{n}$ and $y=\sqrt{\wp(\sqrt{n})}$ using again the boundedness of $\epsilon(x)\wp(x)$ we have that

$$\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\right)}{\wp(\sqrt{n})}-1\right| \leq \left|\frac{\wp\left(\frac{\sqrt{n}}{|t|}\right)}{\wp(\sqrt{n})}-1\right| + C\left(\frac{1}{y^*\sqrt{\wp(\sqrt{n})}}+\frac{1}{y^*\wp(\sqrt{n})}\right)$$

and the last term in the previous display converges to zero as $n\to+\infty$ due to the fact that $\wp$ is diverging and that it is slowly varying at infinity. Hence $\left|\frac{\wp\left(\frac{\sqrt{n\wp(\sqrt{n})}}{|t|}\right)}{\wp(\sqrt{n})}-1\right|$ converges to zero and and so does $\frac{C}{\wp(\sqrt{n})}$ establishing the result. For the case where $d>1$ repeat the previous proof by replacing V_i with $\lambda' V_i$ for λ an arbitrary non zero d dimensional vector. The result would then follow from the Cramer-Wold device and the identification of the multivariate normal distribution via the use of linear transformations. ■

Proof of Corollary 2.2. Notice first that for $z>-1$,

$$\begin{aligned}
&\mathbb{E}\left(z_0^2-1\right)^2\mathbf{1}_{(z_0^2-1\leq z)}\\
=\;& 12\int_{-\sqrt{2(z+1)}}^{\sqrt{2(z+1)}}\left(\frac{x^2}{2}-1\right)^2\left(4+x^2\right)^{-5/2}dx\\
=\;& -12\frac{(1.5+0.75z)\sqrt{(1+z)(3+z)}}{(3+z)^2}+6\,\mathrm{ArcSinh}\left(\sqrt{\frac{z+1}{2}}\right),
\end{aligned}$$

where the expression of the right hand-side is asymptotically equivalent to $3\ln z$. Then the result follows from theorem 2.1 for $\wp(z)=3\ln z$. ■

Proof of Proposition 3.2. Notice first that $\hat{c}_n$ is a Caratheodory function, i.e. continuous w.r.t. θ (due to the continuity of the filter $\hat{h}_t(\theta)$) and point-wise measurable. Then the separability of K and lemma 4.51 of Aliprantis and Border [2] imply that $\hat{c}_n$ is jointly measurable. Furthermore it is proper (i.e. it does not attain the value $-\infty$ and there exists at least one $\theta \in K$ such that $c_n(\theta) \in \mathbb{R}$) since by $\hat{c}_n$ being a Gaussian quasi likelihood function it $\mathbb{P}$ a.s. does not attain the values $\pm\infty$ $\mathbb{P}$ a.s. This implies that it is a *proper normal integrand* in the sense of definition 3.5 (Ch. 5) of Molchanov [24] due to Proposition 3.6 (Ch. 5) in the same reference. The result now follows by the Theorem of Measurable Projections in van der Vaart and Wellner [32], example 1.7.5 p. 47, Proposition 3.10.i (Ch. 5-by setting $a = \inf_K \hat{c}_n + \varepsilon_n$) and the fundamental selection theorem (Theorem 2.13-Ch. 1) of Molchanov [24] (see also the proof of Theorem 3.24.(**i**)-Ch. 5 in the same reference). ■

Proof of Theorem 3.3. Due to assumptions A.1, A.2 and A.2.C.2, 3.1 and lemma Proposition 5.2.12 of Straumann [28] imply that for any $\theta \in \Theta$ there exists an $\varepsilon > 0$ such that

$$\sup_{\Theta \cap B(\theta,\varepsilon)} |c_n - \hat{c}_n| \to 0 \ \mathbb{P} \text{ a.s.}$$

due to Part 1.(i) of the proof of Theorem 5.3.1 of Straumann [28]. This locally uniform asymptotic approximation implies the analogous asymptotic approximation w.r.t. the topology of epi-convergence by the sequential characterization of the latter (see Definitions 2.1 and 2.2 of Lachout et al. [22]). This in turns implies that if $(c_n)_{n\in\mathbb{N}}$ epi-converges to a limit function, then so does $(\hat{c}_n)_{n\in\mathbb{N}}$ to the same limit. To this end, let $\rho_0 \doteqdot \inf_{\theta\in K}\left(\ln h_0(\theta) + \frac{z_0^2\sigma_0^2(\theta_0)}{h_0(\theta)}\right)$ and notice that

$$\begin{aligned}
& \mathbb{E}\left|\inf_{\theta\in K}\left(\ln h_0(\theta) + \frac{z_0^2\sigma_0^2(\theta_0)}{h_0(\theta)}\right)\right| \\
= \ & -\mathbb{E}\rho_0 1_{\rho_0\le 0} + \mathbb{E}\rho_0 1_{\rho_0>0} \\
\le \ & -\mathbb{E}\inf_{\theta\in K}\ln h_0(\theta)\, 1_{\rho_0\le 0} + \mathbb{E}\ln\sigma_0^2 1_{\rho_0>0} + \mathbb{E}\frac{\sigma_0^2(\theta_0)}{h_0(\theta_0)} 1_{\rho_0>0} \\
\le \ & C + \mathbb{E}\ln\sigma_0^2 1_{\rho_0>0}
\end{aligned}$$

for some $C > 0$ that exists due to assumption A.4.3-4. Similarly since σ_0^2 is bounded away from zero and due to A.4.3, $\mathbb{E}\ln\sigma_0^2 1_{\rho_0>0} < +\infty$. Then due to Part 1.(iii) of the proof of Theorem 5.3.1 of Straumann [28] implies that $\theta_0 = \arg\min_\Theta \mathbb{E}\left(\ln h_0(\theta) + \frac{y_0^2}{h_0(\theta)}\right)$. Hence taking also into account A.4 A.4.1 we have that lemma 5.1 is applicable. ■

Proof of Lemma 3.4. **1**. The implications in (5) follow in an essentially similar manner to the proofs of Propositions 5.5.1 and 5.5.2 of Straumann [28] (with the analogous use of the conventions formulated there in order to describe the SRE's that are constructed by differentiations). The differences to those proofs are the following. First Theorem 3 of Wintenberger and Cai [33] is used in place of Proposition 5.2.12 of Straumann [28]. Second, (3), (4) are generalizations of the Holder type continuity conditions imposed in the relevant results by Straumann. The continuous differentiability around zero also imply the implications of the conditions of Straumann by an application of the mean value theorem around zero. Third, the identification of the solutions of the SRE's obtained by differentiation with h_t' and h_t'' respectively is obtained by a lemma that prescribes that under uniform convergence and the existence of a uniform limit of the first derivatives the limit function is differentiable and the limit of the derivatives is the derivative of the limit. Via the results of Appendix A of Andrews [3], this can be also seen to hold for (l/r) derivatives. Given those results the first

implication in (6) is obtained by an application of the mean value theorem to the function $f(a,b)=\frac{a}{b}\left(1-\frac{y_t^2}{b}\right),\ a\in\mathbb{R},\ b>0$, that in turn implies

$$\sup_K\left\|\hat{\ell}_t'(\theta)-\ell_t'(\theta)\right\|\leq c\left(1+y_t^2\right)\left[\sup_K\left|h_t-\hat{h}_t\right|+\sup_K\left\|h_t'-\hat{h}_t'\right\|\right]$$

for some $c>0$. The previous along with $\mathbb{E}\ln^+ y_t^2<+\infty$, Proposition 2.5.1 of Straumann [28] and

$$n\sup_K\|\hat{c}_n'(\theta)-c_n'(\theta)\|\leq\sum_{t=1}^{\infty}\sup_K\left\|\hat{\ell}_t'(\theta)-\ell_t'(\theta)\right\|<+\infty$$

imply the first result. For the second we have that the triangle inequality and the mean value theorem for the functions $f(a,b)=\frac{a}{b}\left(1-\frac{y_t^2}{b}\right)$ and $g(a,b)=\left(\frac{2y_t^2}{a}-1\right)\frac{b}{a^2}$ imply

$$\begin{aligned}\sup_K\left\|\hat{\ell}_t''(\theta)-\ell_t''(\theta)\right\|&\leq c_1\left(1+y_t^2\right)\left[\sup_K\left|h_t-\hat{h}_t\right|+\sup_K\left\|h_t''-\hat{h}_t''\right\|\right]\\&+c_2\left(1+y_t^2\right)\left[\sup_K\left|h_t-\hat{h}_t\right|+\sup_K\left\|\hat{h}_t'\left(\hat{h}_t'\right)^T-h_t'\left(h_t'\right)^T\right\|\right].\end{aligned}$$

for some $c_1,c_2>0$ which exist due to compactness of K^* and the uniform boundedness of the volatility filters away from zero. Analogously to the previous and due to the fact

$$n\sup_K\|c_n''(\theta)-\hat{c}_n''(\theta)\|\leq\sum_{t=1}^{\infty}\sup_K\left\|\ell_t''(\theta)-\hat{\ell}_t''(\theta)\right\|_{K^*}<\infty$$

we obtain the needed result. **2**. It is obtained by the first implication in (6), the convergence in distribution of $r_nc_n'(\theta_0)$, the assumption that $\frac{r_n}{n}\to 0$ and the triangle inequality. **3**. Follows directly form the triangle inequality and the ergodic ULLN. ■

Proof of Theorem 3.5. The theorem 3.3 and lemma 3.4 imply that the result would hold via theorem 5.2 if the following hold. First $r_nc_n'(\theta_0)=\frac{1}{\sqrt{n\wp(\sqrt{n})}}\sum_{t=1}^n\left(z_t^2-1\right)\frac{h_t'(\theta_0)}{\sigma_t^2}$ converges in distribution for $r_n=\sqrt{\frac{n}{\wp(\sqrt{n})}}$. Conditions A.7.1-2 and theorem 2.1 imply that this holds with limit $z_{\theta_0}\sim N\left(0,\mathbf{J}_{\theta_0}\right)$ since it is easy to see that $\wp(x)\backsim\int_{-x}^{x}\left(z_0^2-1\right)^2dF_{z_0}$. Second A.7.2 implies the validity of the result in the third part of 3.4. Finally, the last condition of the second part of assumption A.8 follows from condition A.7.3 along with lemma 5.6.3 of Straumann [28] while the third part of assumption A.8 is essentially A.7.4. ■

Proof of Corollary 3.6. Combine theorem 3.5 with the proof of corollary 2.2 and the continuous mapping theorem. ■

Proof of Theorem 3.7. The result follows in the same lines with the proof of theorem 2.1 with the following modifications. First due to third condition of the theorem and Egoroff's Theorem we have that for some $\varepsilon>0$, there exists a N_ε that is $\mathbb{P}$ a.s. independent of ω for which $|V_n|\in\left(\sqrt{v}-\varepsilon,\sqrt{v}+\varepsilon\right)$ $\mathbb{P}$ a.s. for any $n\geq N_\varepsilon$. Hence we have that for any $i\geq\max\left(\sqrt{v}+\varepsilon,N_\varepsilon\right)$, $|V_i|\leq K_t\sqrt{n}\wp^\star(n)$ $\mathbb{P}$ a.s. Hence we can without loss of generality assume that the condition appearing in theorem 2.1 $\mathbb{P}(C_{n,K}^c)\to 0$ holds, due to the $\mathbb{P}$ a.s. asymptotic negligibility of the terms $\frac{|V_i|}{n}$ in the series under examination for all $i<\max\left(\sqrt{v}+\varepsilon,N_\varepsilon\right)$. Then due to condition 3 and the Cezaro sum theorem we have that $\frac{1}{n}\sum_{i=1}^nV_i^2\to v$ $\mathbb{P}$ a.s. Finally, using the argument above for the $\mathbb{P}$ a.s. confinement of

the $(|V_i|)_{i\in\mathbb{N}}$ sequence in a compact interval for large enough i, we have that as $n \to \infty$, $\left[\frac{\wp\left(\frac{\sqrt{n}\wp^\star(n)}{|t||V_i|}\right)}{(\wp^\star(n))^2} - 1\right] = \left[\frac{\wp\left(\frac{\sqrt{n}\wp^\star(n)}{|t||V_i|}\right)}{\wp\left(\sqrt{n}\wp^\star(n)\right)} \frac{\wp\left(\sqrt{n}\wp^\star(n)\right)}{(\wp^\star(n))^2} - 1\right] \to 0$ $\mathbb{P}$ a.s. due to the Uniform Convergence Theorem (see Theorem 1.2.1 of Bingham et al. [6]) and the definition of $\wp^\star$. Using again the confinement argument along with the previous and the asymptotic negligibility of the terms $\frac{|V_i|}{n}\left[\frac{\wp\left(\frac{\sqrt{n}\wp^\star(n)}{|t||V_i|}\right)}{(\wp^\star(n))^2} - 1\right]$ we obtain that the series $\frac{1}{n}\sum_{i=1}^n V_i^2 \left[\frac{\wp\left(\frac{\sqrt{n}\wp^\star(n)}{|t||V_i|}\right)}{(\wp^\star(n))^2} - 1\right] \to 0$ $\mathbb{P}$ a.s. ■

Proof of Proposition 3.8. The result follows exactly as in the proof of Theorem 1 of Jensen and Rahbek [18] except for the replacement of their Lemma 3 by a use of Theorem 3.7. To this end consider equation (4) in Jensen and Rahbek [18] and evaluate the score at θ_0. Due to Lemma 2 of Jensen and Rahbek [18], Theorem 3.7 is applicable with $\xi_t = (z_t^2 - 1)$, $V_t = \frac{1}{2}\frac{y_{t-1}^2}{1+\theta_0 y_{t-1}^2}$ and $v = \frac{1}{4\theta_0^2}$. This implies that $\frac{1}{2\sqrt{n}\wp^\star(n)}\sum_{t=1}^n (z_t^2 - 1)\frac{y_{t-1}^2}{1+\theta_0 y_{t-1}^2} \rightsquigarrow N\left(0, \frac{1}{4\theta_0^2}\right)$ and the rest follow via the use of Lemmata 4 and 5 of Jensen and Rahbek [18]. ■

Helpful: Strong Consistency, Rate of Convergence and Asymptotic Distribution

Suppose that Θ is a compact subset of $\mathbb{R}^d$ equipped with the relevant Euclidean topology. Let $c_n : \Omega \times \Theta \to \mathbb{R}$ be jointly measurable, θ_n be defined as a $\mathbb{P}$ a.s. approximate minimizer of c_n with optimization error ε_n a $\mathbb{P}$ a.s. non negative random variable. The following result provides with sufficient conditions that characterize the rate of convergence and the asymptotic distribution of θ_n given *consistency*. Let $\theta_0 \in \Theta$. For reasons of notational economy we suppress the dependence on ω. The following lemma provides with sufficient conditions for strong consistency when c_n has the form of an ergodic mean, allowing for cases where the analogous expectation does not exist.

Lemma 5.1 *Suppose that $c_n(\theta) = \frac{1}{n}\sum_{i=1}^n m_i(\theta)$, $(m_i(\theta))_{i\in\mathbb{Z}}$ is ergodic for any θ, c_n is jointly continuous $\mathbb{P}$ a.s., there exists a finite open cover of Θ, such that $\mathbb{E}\left|\inf_{\theta\in A} m_0(\theta)\right| < +\infty$, for any A in the cover, $\mathbb{E}m_0(\theta)$ assumes values in $\overline{\mathbb{R}}$ for any θ in a countable dense subset of Θ. Suppose furthermore that $\theta_0 = \arg\min_\Theta \mathbb{E}m_0(\theta)$ and that $\varepsilon_n \to 0$, $\mathbb{P}$ a.s. Then $\theta_n \to \theta_0$ $\mathbb{P}$ a.s.*

Proof. The first part of the assumption framework of the lemma implies condition $\mathcal{C}_0$ and thereby Theorem 2.3 of Choirat , Hess, and Seri, [8], which implies the joint $\mathbb{P}$ a.s. epi-convergence of c_n to $\mathbb{E}m_0$. Let $\mathbf{epi}$ denote the epigraph of a given function (see e.g. Paragraph 3.1-Ch.5 of Molchanov [24]). Then the assumed properties of c_n Proposition 3.6 and Definition 3.5 (Ch. 5) of Molchanov [24] imply that $\mathbf{epi}_n \doteqdot \mathbf{epi}(c_n)$ is a jointly measurable closed valued correspondence. Conditions 1. and 2. are essentially the sequential characterization of P a.s. *epi-convergence* of c_n to $\mathbb{E}m_0$ (see Definitions 2.1 and 2.2 of Lachout et al. [22]). It follows that $\mathbb{E}m_0$ is an lsc function (see proposition 7.4.a of Rockafellar and Wets [27]). Hence $\mathbf{epi}(\mathbb{E}m_0)$ is a closed valued correspondence. Due to Molchanov [24], paragraph 1.1, and Klein and Thompson [20], Definition 4.5.1 this $\mathbb{P}$ a.s. epi-convergence is equivalent to the following (**i**)-(**ii**) conditions. (**i**) for large enough n, and for all ω in a measurable subset of Ω of unit $\mathbb{P}-$probability, $\mathbf{epi}_n \cap \Theta \times (\mathbb{E}m_0(\theta_0), +\infty) \neq \varnothing$ since $\Theta \times (\mathbb{E}m_0(\theta_0), +\infty)$ is open in the relevant product topology and $\mathbf{epi}(\mathbb{E}m_0) \cap \Theta \times (\mathbb{E}m_0(\theta_0), +\infty) \neq \varnothing$. Hence $\inf_\Theta c_n(\theta) \geq \mathbb{E}m_0(\theta_0)$ for all ω described previously which implies that $\liminf_n \inf_\Theta c_n \geq \mathbb{E}m_0(\theta_0)$ $\mathbb{P}$ a.s. Furthermore (**ii**) for any $\varepsilon > 0$, we have

that for large n, and for all ω in a (possibly different than the previous) measurable subset of Ω of unit $\mathbb{P}-$probability, $\mathbf{epi}_n \cap \Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon] = \varnothing$ P a.s. since $\Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon]$ is compact in the relevant product topology and $\mathbf{epi}(\mathbb{E}m_0) \cap \Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon] = \varnothing$. This implies that $\limsup_n \inf_\Theta c_n(\theta) \leq \mathbb{E}m_0(\theta_0)$ $\mathbb{P}$ a.s. Now let x_n be a measurable selection from the random compact set

$$\left\{\theta \in \Theta : c_n(\theta) \leq \inf_\Theta c_n + \varepsilon_n\right\}$$

such that for some subsequence (x_{n_k}), $x_{n_k} \to x$ $\mathbb{P}$ a.s. Its existence is guaranteed by the fundamental selection theorem (Theorem 2.13-Ch. 1 of Molchanov [24]). Then

$$\begin{aligned}
\mathbb{E}m_0(x) &\leq \liminf_{n_k} c_{n_k}(x_{n_k}) \ \mathbb{P} \text{ a.s.} \\
&\leq \limsup_{n_k} c_{n_k}(x_{n_k}) \ \mathbb{P} \text{ a.s.} \\
&= \limsup_{n_k} \left(\inf_\Theta c^*_{n_k} + \varepsilon_{n_k}\right) \ \mathbb{P} \text{ a.s.} \\
&\leq \mathbb{E}m_0(\theta_0) \ \mathbb{P} \text{ a.s.}
\end{aligned}$$

establishing that any $\mathbb{P}$ a.s. cluster point of such a measurable selection coincides with θ_0. The result now follows from the fact that Θ is compact. ■

For $r_n \to +\infty$, we denote with $\mathcal{H}_n$ the $r_n(\Theta - \theta_0) = \{r_n(x - \theta_0), x \in \Theta\}$ and notice that $\mathcal{H}_n$ is compact and contains 0. Furthermore we denote with $\mathcal{H} = \limsup_{n\to\infty} \mathcal{H}_n$ in the sense of the obvious generalization of definition D.3.

Consider the following assumption that provides more structure for the asymptotic properties of c_n.

Assumption A.8 *Assume that the following hold:*

1. *For any sequence (ϑ_n) with values in Θ such that $\vartheta_n \xrightarrow{p} \theta_0$, $c_n(\vartheta_n) - c_n(\theta_0) = (\vartheta_n - \theta_0)' q_n + (\vartheta_n - \theta_0)' g_n (\vartheta_n - \theta_0)$, with $\mathbb{P}$ probability that converges to 1. g_n is a random $q \times q$ matrix that can be defined in any point of the aforementioned line $\mathbb{P}$ a.s. q_n is a random $q \times 1$ matrix.*

2. *For some positive real sequence $r_n \to +\infty$, $r_n q_n \rightsquigarrow z_{\theta_0}$ which is a random vector whose distribution can depend on θ_0 and $g_n \xrightarrow{p} \mathbf{J}_{\theta_0}$ a non singular matrix independent of ω that may depend on θ_0.*

3. *$\mathcal{H}$ is convex.*

The next theorem is the final result of this section.

Theorem 5.2 *Assume that $\theta_n \underset{p}{\to} \theta_0$. If A.8.1,2 hold and $\varepsilon_n = O_p(r_n^{-2})$ then*

$$r_n(\theta_n - \theta_0) = O_p(1). \tag{7}$$

If moreover A.8.3 holds and $\varepsilon_n = o_p(r_n^{-2})$ then

$$r_n(\theta_n - \theta_0) \rightsquigarrow \tilde{h}_{\theta_0} \tag{8}$$

with $\tilde{h}_{\theta_0}$ defined uniquely by $q\left(\tilde{h}_{\theta_0}\right) = \inf_{h\in\mathcal{H}} q(h)$ and $q(h) := \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)$.

Proof. Notice that due to the definition of θ_n we have

$$c_n(\theta_n) - c_n(\theta_0) \leq O_p\left(r_n^{-2}\right).$$

From $\theta_n \underset{p}{\to} \theta_0$ and employing assumption A.8.1,2

$$\nu_n' r_n q_n + \nu_n' g_n(b_n^*)\nu_n \leq O_p(1)$$

where $\nu_n = r_n(\theta_n - \theta_0)$ and b_n^* as in A.8.1. Hence due to consistency

$$\nu_n' r_n q_n + \nu_n'(\mathbf{J}_{\theta_0} + o_p(1))\nu_n \leq O_p(1).$$

Assumption A.8.2 then implies that there exists some positive $c > 0$ such that

$$\|\nu_n\| O_p(1) + c\|\nu_n\|^2 + \|\nu_n\|^2 o_p(1) \leq O_p(1)$$

which implies that

$$\|\nu_n\|^2(1 + o_p(1)) + 2\|\nu_n\| O_p(1)(1 + o_p(1)) + O_p(1) \leq O_p(1)$$

Hence

$$\|\nu_n\|(1 + o_p(1)) \leq O_p(1)$$

establishing (7). Now given the definition of $\mathcal{H}$ consider the following. From consistency and assumption A.8 we can define $\varpi_n : \mathbb{R}^q \to \mathbb{R}$ as

$$\begin{aligned}\varpi_n(h) &\equiv r_n^2\left(c_n\left(\theta_0 + \frac{h}{r_n}\right) - c_n(\theta_0)\right)\\ &= h' r_n q_n + h' g_n(b_n^*) h\end{aligned}$$

From the first part of the present proof we have that for U an arbitrary compact subset of $\mathbb{R}^q$

$$\varpi_n(h) \rightsquigarrow h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \text{ in } C(U, \mathbb{R}).$$

Hence for any A compact subset of $\mathbb{R}^q$,

$$\inf_{h \in A} \varpi_n(h) \rightsquigarrow \inf_{h \in A}\left(h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h\right). \tag{9}$$

Due to (7) $h_n \doteqdot r_n(\beta_n - b(\theta_n)) \in \mathcal{H}_n \cap B(0, r_n\varepsilon) \doteqdot M_n$ with $\mathbb{P}$-probability tending to 1 for some $\varepsilon > 0$. If F is a closed non empty subset of $\mathbb{R}^q$, and $h_n \in F$, then for large enough n, either $M_n \subset F$, or $M_n \not\subseteq F$ but $M_n \cap F \neq \varnothing$. In either case due to the definitions of θ_n, β_n ϖ_n and the fact that $\varepsilon_n = o_p(r_n^{-2})$

$$\inf_{h \in M_n \cap F} \varpi_n(h) \leq \inf_{h \in M_n} \varpi_n(h) + o_p(1)$$

and therefore due to Slutsky's lemma

$$\begin{aligned}&\mathbb{P}(h_n \in F) \leq \mathbb{P}\left(\inf_{h \in M_n \cap F} \varpi_n(h) \leq \inf_{h \in M_n} \varpi_n(h) + o_p(1)\right)\\ \leq\ &\mathbb{P}\left(\inf_{h \in M_n \cap F} \varpi_n(h) \leq \inf_{h \in M_n} \varpi_n(h)\right) + o(1)\end{aligned}$$

Now notice that $M_n = M_n \cap \mathbb{R}^q$ and $\mathbb{R}^q$ is open, $\limsup_{n\to\infty} M_n = \mathcal{H}$, since $\limsup_{n\to\infty} \mathcal{H}_n = \mathcal{H}$ and $r_n \to \infty$. Furthermore equation (9) and the continuous mapping theorem imply that Lemma 7.13.2-3 of van der Vaart [31] is applicable, so that the last probability is less than or equal to

$$\mathbb{P}\left(\inf_{h\in\mathcal{H}\cap F} \varpi_n(h) \leq \inf_{h\in\mathcal{H}} \varpi_n(h) + o_p(1)\right) \leq \mathbb{P}\left(\inf_{h\in\mathcal{H}\cap F} \varpi_n(h) \leq \inf_{h\in\mathcal{H}} \varpi_n(h)\right) + o(1)$$

due to Slutsky's Lemma. Now from equation (9), the continuous mapping theorem and Portmanteau Lemma we have that the $\limsup$ of the probability in the right hand side of the last display is less than or equal to

$$\mathbb{P}\left(\inf_{h\in\mathcal{H}\cap F} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \leq \inf_{h\in\mathcal{H}} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h\right)$$

which equals

$$\begin{aligned} & \mathbb{P}\left(\inf_{h\in\mathcal{H}\cap F} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \pm \frac{1}{2} z'_{\theta_0} \mathbf{J}_{\theta_0}^{-1} z_{\theta_0} \leq \inf_{h\in\mathcal{H}} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \pm \frac{1}{2} z'_{\theta_0} \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right) \\ = \; & \mathbb{P}\left(\inf_{h\in\mathcal{H}\cap F} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right) \leq \inf_{h\in\mathcal{H}} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)\right) \end{aligned}$$

Since H^* is closed and convex and $\mathbf{J}_{\theta_0}$ is positive definite $\tilde{h}_{\theta_0}$ is unique, and thereby when

$$\inf_{h\in\mathcal{H}\cap F} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right) \leq \inf_{h\in\mathcal{H}} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)$$

holds then

$$\tilde{h}_{\theta_0} \in \mathcal{H} \cap F$$

and therefore the last probability is less than or equal to

$$\mathbb{P}\left(\tilde{h}_{\theta_0} \in \mathcal{H} \cap F\right) \leq \mathbb{P}\left(\tilde{h}_{\theta_0} \in F\right)$$

hence we have proven that

$$\limsup_{n\to\infty} \mathbb{P}(h_n \in F) \leq \mathbb{P}\left(\tilde{h}_{\theta_0} \in F\right)$$

and (8) follows from the Portmanteau theorem due to the fact that F is chosen arbitrarily. ■

Appendix 2: Figures

Figure 1: First Experiment - Empirical CDF for the QMLE of β

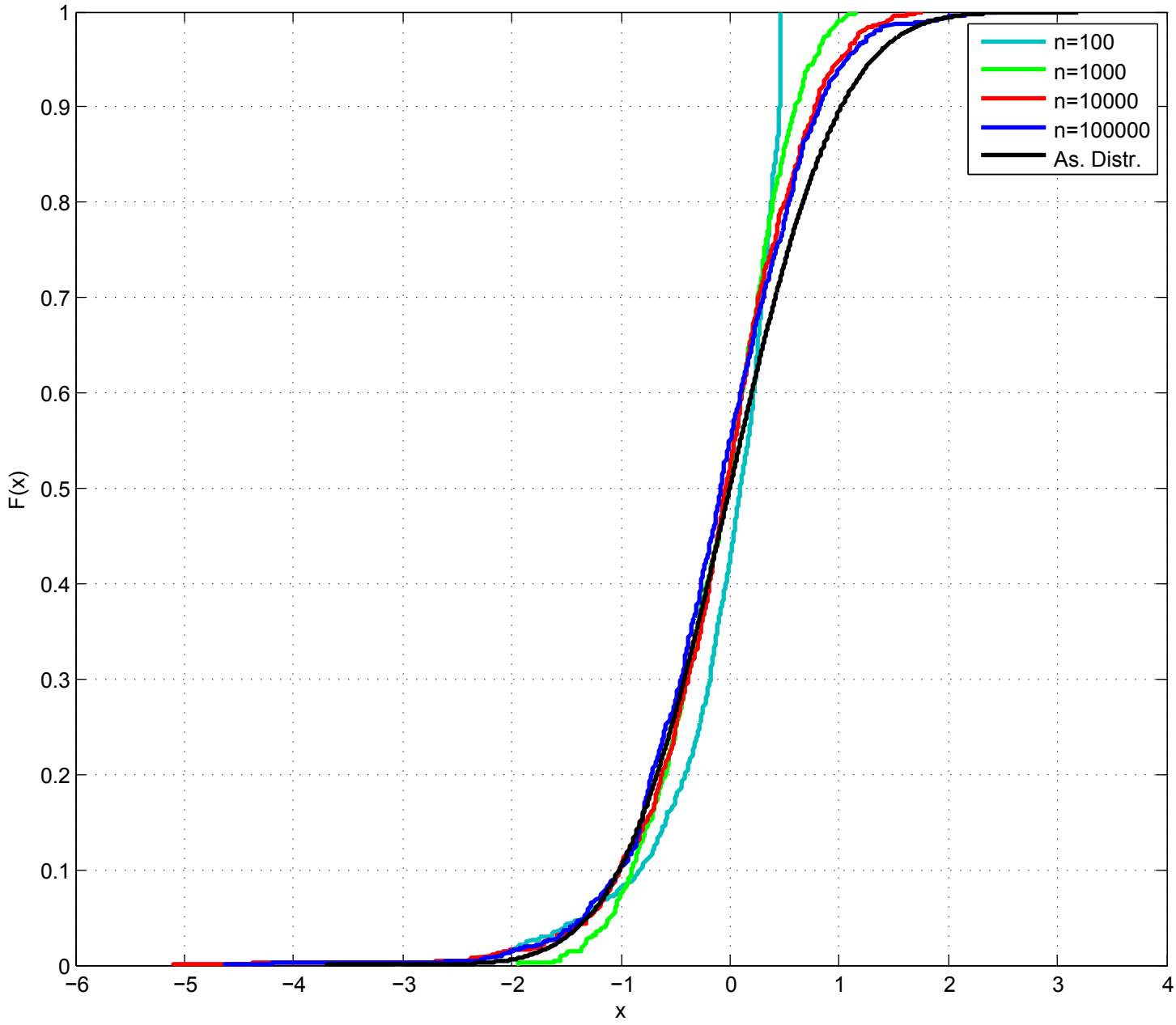

Figure 2: Second Experiment - Empirical CDF for the QMLE of α

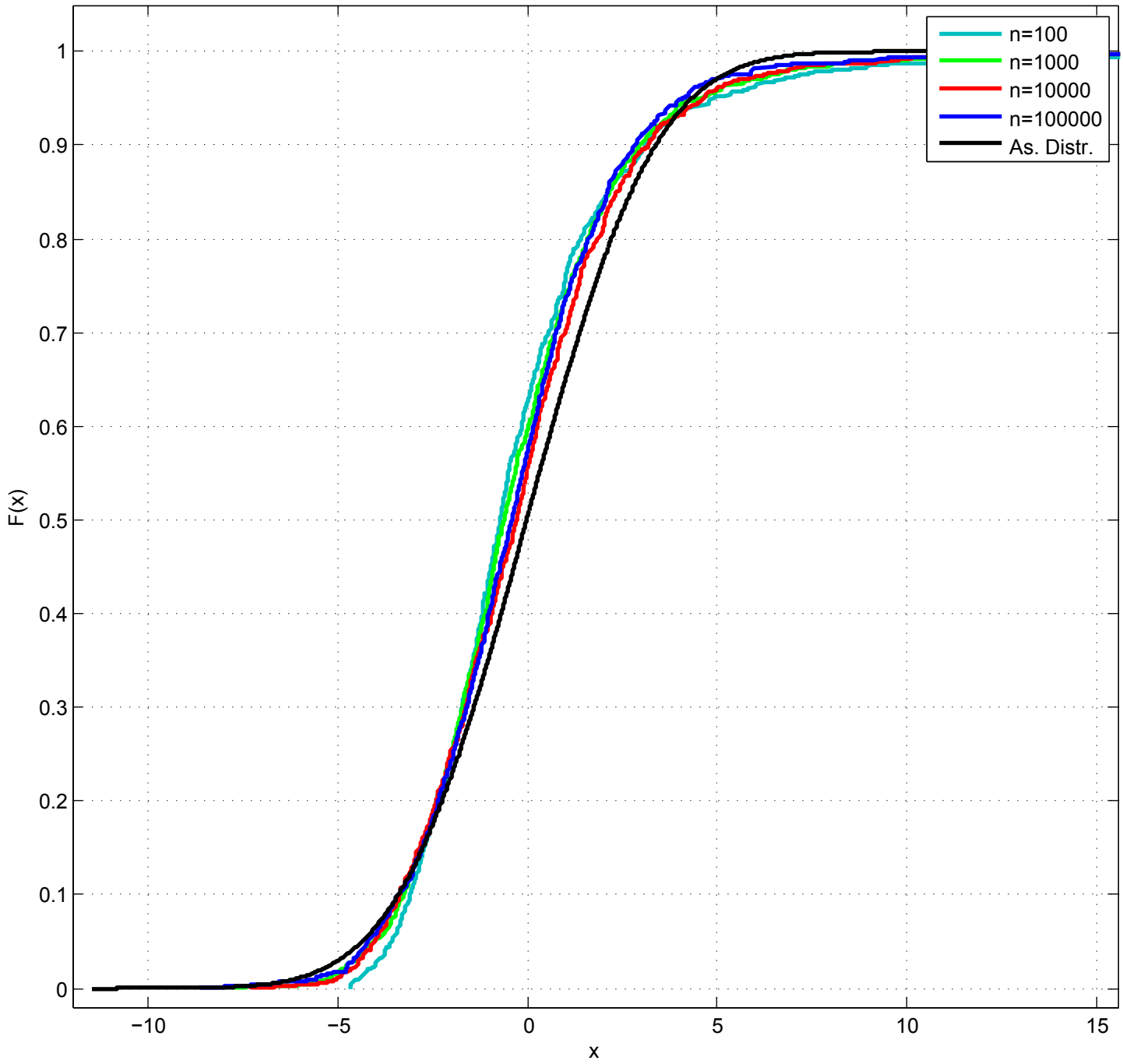

Figure 3: Second Experiment - Empirical CDF for the QMLE of β