CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

Local Asymptotic Normality of General Conditionally Heteroskedastic and Score-Driven Time-Series Models

Christian Francq, Jean-Michel Zakoïan

N°06 / March 2022

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Local Asymptotic Normality of General Conditionally Heteroskedastic and Score-Driven Time-Series Models

Christian Francq* and Jean-Michel Zakoian†‡

Abstract

The paper establishes the Local Asymptotic Normality (LAN) property for general conditionally heteroskedastic time series models of multiplicative form, $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$, where the volatility $\sigma_t(\boldsymbol{\theta}_0)$ is a parametric function of $\{\epsilon_s, s < t\}$, and (η_t) is a sequence of i.i.d. random variables with common density $f_{\boldsymbol{\theta}_0}$. In contrast with earlier results, the finite dimensional parameter $\boldsymbol{\theta}_0$ enters in both the volatility and the density specifications. To deal with non-differentiable functions, we introduce a conditional notion of the familiar quadratic mean differentiability condition which takes into account parameter variation in both the volatility and the errors density. Our results are illustrated on two particular models: the APARCH with Asymmetric Student-t distribution, and the Beta-t-GARCH model, and are extended to handle a conditional mean.

Keywords: APARCH, Asymmetric Student-t distribution, Beta-t-GARCH, Conditional heteroskedasticity, LAN in time series, Quadratic mean differentiability.

*CREST and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-Mail: christian.francq@univ-lille3.fr

†Corresponding author: Jean-Michel Zakoïan, University of Lille and CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. E-mail: zakoian@ensae.fr.

‡The authors are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MLforRisk (ANR-21-CE26-0007) and to the Labex ECODEC.

1 Introduction

Local Asymptotic Normality (LAN) is a crucial property for comparing the asymptotic performance of statistical procedures in parametric or semi-parametric models (parameterized by finite-dimensional and infinite-dimensional nuisance parameters). For independent and identically distributed (iid) data, a comprehensive account on the LAN theory can be found in the books by van der Vaart (1998), and Lehmann and Romano (2006). Swensen (1985) established the LAN property for finite-order AR models with a regression trend. The proof of the LAN property for ARMA models is due to Kreiss (1987), while Koul and Schick (1996) considered random coefficients AR models. LAN results for a large class of time series models, in particular models with time-varying location and scale, were obtained by Drost, Klaassen and Werker (1997). The LAN property was also established for long-memory time series models, see Hallin, Taniguchi, Serroukh and Choy (1999).

In GARCH models $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$, where the volatility $\sigma_t(\boldsymbol{\theta}_0)$ belongs to the σ-field generated by the past of ϵ_t and (η_t) is an iid sequence with density f, the most popular estimation method for the parameter $\boldsymbol{\theta}_0$ is the QMLE (Quasi-Maximum Likelihood Estimation) which uses a criterion based on a Gaussian density for η_t. For standard GARCH, the asymptotic properties of the QMLE were derived under mild regularity conditions by Berkes, Horváth and Kokoszka (2003), and by Francq and Zakoian (2004). When the distribution of η_t is not normal, the QMLE may not be efficient (in particular in the minimax sense; see van der Vaart (1998)). Efficient estimators of (some components of) $\boldsymbol{\theta}_0$ can be obtained, when f is unknown, via an adaptive estimation procedure. This problem was studied, among others, by Linton (1993), Jeganatan (1995), Drost and Klaassen (1997) who proved the LAN property for ARCH models, and Lee and Taniguchi (2005) who considered the inclusion of a stochastic mean and dealt with initial values in the DGP.

The results established in the aforementioned articles hold under the assumption that the errors density f is a nuisance parameter. Recent references on GARCH-type and score-driven volatility models underlined the interest of parametrizing the errors density. This can be done by letting this density depend on a finite-dimensional parameter $\boldsymbol{\nu}$, hence $f(\cdot) = f(\cdot;\boldsymbol{\nu}_0)$, which is independent of the volatility parameter $\boldsymbol{\theta}_0$. The LAN property was established in this context, for ARMA-GARCH models, by Ling and McAleer (2003). In other formulations, the density parameter enters directly as a parameter of the volatility dynamics. This is the case of the score-driven volatility models introduced by Creal, Koopman and Lucas (2008) and Harvey and Chakravarty (2008). To our knowledge, no LAN result exists for handling such volatility models.

The aim of the present contribution is to establish the LAN property under mild conditions in a fully parametric framework of general GARCH time series models, where the finite dimensional parameter $\boldsymbol{\theta}_0$ enters in both the volatility and the density specifications. We first consider the case where both the volatility and the errors density are smooth functions. In the usual setting, it is known that such smoothness assumptions can be replaced by the concept of Quadratic Mean Differentiability (see e.g. van der Vaart (1998)). However, because the lack of differentiability may concern both the volatility and the density functions, QMD is not sufficient in our framework and the main challenge is to extend this concept. We introduce a related concept, called *Conditional Quadratic Mean Differentiability (CQMD)*, which expands, around the true parameter value, the *conditional density* rather than the density of the observations.

Without the assumption of zero-mean innovatons, GARCH models allow for a time-varying mean, but the conditional mean is proportional to $\sigma_t(\boldsymbol{\theta}_0)$. We will extend the analysis to cover more general conditional means of returns, with models of the form $y_t = m_t(\boldsymbol{\theta}_0) + \sigma_t(\boldsymbol{\theta}_0)\eta_t$. However, the assumptions being more demanding and the LAN result more complex, we prefer to start by studying the pure GARCH model.

The plan of the paper is as follows. In Section 2, we present our assumptions on the GARCH-type model and provide our main results on the LAN property. In Section 3, we use the LAN property to derive local asymptotic powers of tests. Examples are developed in Section 4. For completeness we also consider in Section 5 the case where a conditional mean is included in the model. Concluding remarks are displayed in Section 6. Most proofs can be found in the appendix.

2 General GARCH model and LAN result

We consider a general volatility model $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$ where $\sigma_t(\boldsymbol{\theta}_0) = \sigma_{\boldsymbol{\theta}_0}(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$, the sequence (η_t) is iid, [1] and $\boldsymbol{\theta}_0$ belongs to a convex subset Θ of $\mathbb{R}^d$. Since we are going to consider local properties of the model around $\boldsymbol{\theta}_0$, we will assume, without loss of generality, that Θ is bounded. Denote by $\boldsymbol{\theta}$ a generic element of Θ. Let $\mathcal{F}_t$ be the sigma-field generated by $\{\eta_u, u \leq t\}$. Our assumptions on

[1] A usual assumption is that $E\eta_t = 0$ and $E\eta_t^2 = 1$ but, in this fully parametric framework, we do not require such moment assumptions.

the model are summarized in

A1($\boldsymbol{\theta}_0$): (ϵ_t) satisfies $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$ where η_t has density $f_{\boldsymbol{\theta}_0}$ with respect to a sigma-finite measure μ and, for all $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, $\{\sigma_t(\boldsymbol{\theta})\}$ is a stationary sequence with $\sigma_t(\boldsymbol{\theta}) \in \mathcal{F}_{t-1}$ and $\sigma_t(\boldsymbol{\theta}) > 0$.

For $\boldsymbol{\tau} \in \mathbb{R}^d$, let the sequence of local parameters $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}$ such that $\boldsymbol{\theta}_n \in \Theta$ for n large enough. We denote by P_0 (resp. $P_{n,\boldsymbol{\tau}}$) the stationary distribution of the process (ϵ_t) when the parameter is $\boldsymbol{\theta}_0$ (resp. $\boldsymbol{\theta}_n$), *i.e.* under **A1**($\boldsymbol{\theta}_0$) (resp. **A1**($\boldsymbol{\theta}_n$)). Under **A1**($\boldsymbol{\theta}_n$), the process could be denoted $(\epsilon_{t,n})_{t\in\mathbb{Z}}$ but it is standard to avoid this heavy notation. Because the η_t's are iid with density $f_{\boldsymbol{\theta}}$, the likelihood of $\epsilon_1, \dots, \epsilon_n$ conditional on $\mathcal{F}_0$ is

$$L_n(\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{1}{\sigma_t(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right), \quad \eta_t(\boldsymbol{\theta}) = \frac{\epsilon_t}{\sigma_t(\boldsymbol{\theta})}.$$

We will study the conditional log-likelihood ratio

$$\Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \log \frac{L_n(\boldsymbol{\theta}_n)}{L_n(\boldsymbol{\theta}_0)}.$$

Note that $\sigma_t(\boldsymbol{\theta})$ generally involves the infinite past of the process (ϵ_t) (and thus of (η_t)) and that no initial conditions are introduced here[2]. In many models, both the density and the volatility are smooth functions. We start by deriving LAN results in this situation, for which more explicit conditions can be provided.

2.1 LAN property under differentiability

Assume the following regularity conditions.

A2: For all $\boldsymbol{\theta} \in \Theta$, $y \mapsto f_{\boldsymbol{\theta}}(y)$ admits continuous second-order derivatives. For all $t \geq 1$, $\boldsymbol{\theta} \mapsto \sigma_t(\boldsymbol{\theta})$ admits continuous second-order derivatives. For all $y \in \mathbb{R}$, $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}(y)$ admits continuous second-order derivatives.

We also need to introduce the notations

$$g_{\boldsymbol{\theta}}(y) = 1 + y\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}(y), \quad \boldsymbol{f}_{\boldsymbol{\theta}}(y) = \frac{\partial \log f_{\boldsymbol{\theta}}(y)}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{g}_{\boldsymbol{\theta}}(y) = \frac{\partial g_{\boldsymbol{\theta}}(y)}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{F}_{\boldsymbol{\theta}}(y) = \frac{\partial^2 \log f_{\boldsymbol{\theta}}(y)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

where prime denotes derivative with respect to y. Assuming

[2]A different approach was adopted by Drost, Klaassen and Werker (1997) who assumed that the DGP includes initial conditions. On the other hand, Ling and McAleer (2003) considered the likelihood of the observations and an initial value.

A3: $Eg^2_{\boldsymbol{\theta}_0}(\eta_t) < \infty$, $E\|\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)\|^2 < \infty$ and $E\|\frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\|^2 < \infty$,

let

$$\mathfrak{J} = \iota_f \boldsymbol{J} - \boldsymbol{\Omega} \boldsymbol{f}^\top - \boldsymbol{f} \boldsymbol{\Omega}^\top + \boldsymbol{F}, \tag{2.1}$$

with $\iota_f = Eg^2_{\boldsymbol{\theta}_0}(\eta_t)$, $\boldsymbol{J} = E\frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}$, $\boldsymbol{\Omega} = E\frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$, $\boldsymbol{F} = E\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{f}^\top_{\boldsymbol{\theta}_0}(\eta_t)$, and $\boldsymbol{f} = Eg_{\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)$.

Finally, we assume that

A4: there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\boldsymbol{F}_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))\| < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\boldsymbol{f}_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))\|^2 < \infty,$$

and three pairs of conjugate numbers $p_i > 1$, $q_i > 1$, $1/p_i + 1/q_i = 1$, for $i = 1, 2, 3$, such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |g_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))|^{p_1} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\|^{q_1} < \infty,$$

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| g'_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))\, \eta_t(\boldsymbol{\theta}) \right|^{p_2} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{2q_2} < \infty,$$

and

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\boldsymbol{g}_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))\|^{p_3} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{q_3} < \infty.$$

Let the central sequence

$$\boldsymbol{\Delta}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) - g_{\boldsymbol{\theta}_0}(\eta_t) \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\}.$$

Note that the term $\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)$ vanishes when, as in Drost and Klaassen (1997), Drost, Klaassen and Werker (1997) or Lee and Taniguchi (2005), the density f of η_t does not depend on $\boldsymbol{\theta}$. Note also that our central sequence is not measurable with respect to the observations. For most volatility models the effect of deterministic initial values is negligible asymptotically. This issue will be considered below.

Our first result is the following.

Proposition 2.1. *Let Θ be a bounded convex subset of $\mathbb{R}^d$ such that $\boldsymbol{\theta}_0 \in \Theta$. Assume* **A1**$(\boldsymbol{\theta}_0)$ *and* **A2**-**A4**. *When $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n} \in \Theta$ for n large enough, we have the LAN property*

$$\Lambda_n(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) = \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n - \frac{1}{2}\boldsymbol{\tau}^\top \mathfrak{J} \boldsymbol{\tau} + o_{P_0}(1) \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2}\boldsymbol{\tau}^\top \mathfrak{J} \boldsymbol{\tau}, \boldsymbol{\tau}^\top \mathfrak{J} \boldsymbol{\tau}\right) \quad \textit{under } P_0.$$

Note that in the particular case where the density f is a nuisance parameter (i.e. independent of $\boldsymbol{\theta}_0$), we retrieve the usual expansion with $\mathfrak{J} = \iota_f \boldsymbol{J}$.

In Proposition 2.1 the asymptotic distribution of the likelihood ratio is obtained without considering initial values. As in Lee and Taniguchi (2005), we now introduce a version of the central sequence that takes into account initial values for $\{\epsilon_j, j \leq 0\}$. Let for $t > 0$, $\widetilde{\sigma}_t(\boldsymbol{\theta}) = \sigma_{\boldsymbol{\theta}}(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \widetilde{\epsilon}_0, \widetilde{\epsilon}_{-1}, \dots)$, where the $\widetilde{\epsilon}_j$'s are fixed initial values. Let the observation-measurable version of the central sequence

$$\widetilde{\boldsymbol{\Delta}}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \boldsymbol{f}_{\boldsymbol{\theta}_0}(\widetilde{\eta}_t) - g_{\boldsymbol{\theta}_0}(\widetilde{\eta}_t) \frac{\partial \log \widetilde{\sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\}, \quad \widetilde{\eta}_t = \widetilde{\eta}_t(\boldsymbol{\theta}_0), \quad \widetilde{\eta}_t(\boldsymbol{\theta}) = \frac{\epsilon_t}{\widetilde{\sigma}_t(\boldsymbol{\theta})}.$$

For many volatility models, such as those considered in Section 4 below, the following assumptions are satisfied. In particular, the moment condition in the next assumption holds true when the volatility is bounded below.

A5: We have $E\sigma_t^{-s}(\boldsymbol{\theta}_0) < \infty$ for some $s > 0$. Moreover, there exist $K > 0$ and $\rho \in [0, 1)$ such that

$$|\sigma_t(\boldsymbol{\theta}_0) - \widetilde{\sigma}_t(\boldsymbol{\theta}_0)| + \left\| \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \leq K\rho^t \quad \text{a.s.}$$

A6: the functions $y \mapsto \boldsymbol{f}_{\boldsymbol{\theta}_0}(y)$ and $y \mapsto g_{\boldsymbol{\theta}_0}(y)$ have (componentwise) bounded derivatives.

The following result shows that the initial values are generally irrelevant for the asymptotic distribution of the central sequence.

Proposition 2.2. *The LAN property of Proposition 2.2 remains valid when* $\boldsymbol{\Delta}_n$ *is replaced by* $\widetilde{\boldsymbol{\Delta}}_n$*, under the additional assumptions* **A5**-**A6**.

2.2 LAN property under CQMD

Assumption **A2** is standard and is sufficient for most applications, but it can be replaced by the following CQMD condition.

A2*: For all $t \in \mathbb{Z}$, there exists a vector $\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) := \boldsymbol{s}_{\boldsymbol{\theta}_0}(y, \eta_{t-1}, \eta_{t-2}, \dots) \in \mathbb{R}^d$ where $\boldsymbol{s}_{\boldsymbol{\theta}_0}$ is a measurable function, such that

$$\sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{h})} f_{\boldsymbol{\theta}_0 + \boldsymbol{h}} \left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{h})} y \right)} = \sqrt{f_{\boldsymbol{\theta}_0}(y)} + \frac{1}{2} \boldsymbol{h}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) \sqrt{f_{\boldsymbol{\theta}_0}(y)} + r_{t,\boldsymbol{h}}(y), \tag{2.2}$$

with

$$\|r_{t,\boldsymbol{h}}(\cdot)\|_{L^2(\mu)}^2 := \int r_{t,\boldsymbol{h}}^2(y) d\mu(y) = o_{P_0}(\|\boldsymbol{h}\|^2) \quad \text{as} \quad \boldsymbol{h} \to 0.$$

Note that when f is not parametrized by $\boldsymbol{\theta}_0$, it is enough to suppose QMD for $\sqrt{f}$ as in Drost, Klaassen and Werker (1997). Note also that under **A2-A4**, a Taylor expansion and tedious computations show that (2.2) holds with

$$\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) = \frac{\partial}{\partial \boldsymbol{h}} \log \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})} f_{\boldsymbol{\theta}_0+\boldsymbol{h}}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})} y\right)\Bigg|_{\boldsymbol{h}=0} = \boldsymbol{f}_{\boldsymbol{\theta}_0}(y) - g_{\boldsymbol{\theta}_0}(y) \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \tag{2.3}$$

In the sequel we no longer assume **A2** but, instead, assume the CQMD condition **A2***. We have the following lemma.

Lemma 2.1. *Under* $\mathbf{A1}(\boldsymbol{\theta}_0)$ *and* $\mathbf{A2}^*$

$$E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|\mathcal{F}_{t-1}) = \mathbf{0} \quad \textit{and} \quad \mathfrak{J}_t := E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{s}^\top_{t,\boldsymbol{\theta}_0}(\eta_t)|\mathcal{F}_{t-1}) \;\; \textit{exists}, \quad \textit{a.s.} \tag{2.4}$$

Note that **A2*** entails that

$$\|r_{t,\boldsymbol{h}}(\cdot)\|_{L^2(\mu)} \leq 2 + \frac{1}{2}\{\boldsymbol{h}^\top \boldsymbol{\mathfrak{J}_t} \boldsymbol{h}\}^{1/2}. \tag{2.5}$$

Let the assumption

A3***:** The following matrix exists

$$\mathfrak{J} := E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{s}^\top_{t,\boldsymbol{\theta}_0}(\eta_t)).$$

Note that under (2.3), $\mathfrak{J}$ coincides with the matrix in (2.1). It follows from (2.5) and **A3*** that for any bounded sequence $(\boldsymbol{h}_n)$, we have uniform integrability of the sequence $(\|r_{t,\boldsymbol{h}_n}(\cdot)\|_{L^2(\mu)})_n$. Therefore, using Theorem 3.5 of Billingsley (1999), we have

$$E\int r^2_{t,\boldsymbol{h}}(y)d\mu(y) = o(\|\boldsymbol{h}\|^2) \quad \text{as} \quad \boldsymbol{h} \to 0. \tag{2.6}$$

Our main result is the following.

Proposition 2.3. *Proposition 2.1 remains valid when* **A2-A4** *is replaced by* **A2*****-A3*** *and the central sequence is defined by* $\boldsymbol{\Delta}_n = n^{-1/2}\sum_{t=1}^n \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)$.

Extending Proposition 2.2 by introducing initial values in the central sequence of Proposition 2.3 seems only possible on a case-by-case basis.

3 Testing linear hypotheses

In this section, we study how our LAN properties can be used to derive the local asymptotic powers of tests. Consider testing an assumption of the form $H_0 : \boldsymbol{R}\boldsymbol{\theta}_0 = \boldsymbol{r}$ where $\boldsymbol{R}$ is a full row rank $p \times d$ matrix and $\boldsymbol{r} \in \mathbb{R}^p$. Assume that $\boldsymbol{\theta}_0$ belongs to the interior $\overset{\circ}{\Theta}$ of Θ and that, for an estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$, the following Bahadur expansion holds

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Psi}_{t-1} \boldsymbol{V}(\eta_t) + o_{P_0}(1),$$

where $\boldsymbol{V}(\cdot)$ is a measurable function, $\boldsymbol{V} : \mathbb{R} \mapsto \mathbb{R}^k$ for some positive integer k, and $\boldsymbol{\Psi}_{t-1}$ is a $\mathcal{F}_{t-1}$-measurable $d \times k$ matrix, $(\boldsymbol{\Psi}_t)$ being stationary. We assume the variables $\boldsymbol{\Psi}_t$ and $\boldsymbol{V}(\eta_t)$ belong to L^2, $E\boldsymbol{V}(\eta_t) = 0$, $\operatorname{var}\{\boldsymbol{V}(\eta_t)\} = \boldsymbol{\Upsilon}$ is nonsingular and, for any $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{x}'\boldsymbol{\Psi}_t = 0$ a.s. entails $\boldsymbol{x} = 0$.

When $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n^{ML}$ is the Maximum Likelihood Estimator (MLE), the Bahadur expansion holds under some regularity conditions, and we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{ML} - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathfrak{J}^{-1} \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t) + o_{P_0}(1). \tag{3.1}$$

When $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n^{QML}$ is the QMLE, the Bahadur expansion also holds under some regularity conditions, with

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{QML} - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{2} \boldsymbol{J}^{-1} \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} (\eta_t^2 - 1) + o_{P_0}(1). \tag{3.2}$$

It should be noted that initial values may (and generally have to) be introduced in the definition of the (Q)MLE. However, the log-likelihood ratio remains throughout defined using the infinite past of the process, that is, without initial values.

We wish to test H_0 against the sequence of local alternatives $H_n : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}$, $\boldsymbol{\tau} \in \mathbb{R}^d$, where $\boldsymbol{R}\boldsymbol{\theta}_0 = \boldsymbol{r}$ and $\boldsymbol{R}\boldsymbol{\tau} \neq \boldsymbol{0}$.[3]

Assuming that the LAN property holds, under the conditions of either Propositions 2.1 or 2.3, we have, under H_0,

$$\begin{pmatrix} \sqrt{n}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{r}\right) \\ \Lambda_n(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{R}\boldsymbol{\Psi}_{t-1}\boldsymbol{V}(\eta_t) \\ \boldsymbol{\tau}^\top \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t) - \frac{1}{2}\boldsymbol{\tau}^\top \mathfrak{J} \boldsymbol{\tau} \end{pmatrix} + o_{P_0}(1).$$

[3] In other words, under H_n the true parameter value is $\boldsymbol{\theta}_n$ instead of $\boldsymbol{\theta}_0$ and the null hypothesis is not satisfied under H_n ($\boldsymbol{R}\boldsymbol{\theta}_n \neq \boldsymbol{r}$).

Consequently,

$$\begin{pmatrix} \sqrt{n}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{r}\right) \\ \Lambda_n(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{ \begin{pmatrix} 0 \\ -\frac{1}{2}\boldsymbol{\tau}^\top \boldsymbol{\mathfrak{J}}\boldsymbol{\tau} \end{pmatrix}, \begin{pmatrix} \boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}^\top & \boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) \\ \boldsymbol{c}^\top_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) & \boldsymbol{\tau}^\top \boldsymbol{\mathfrak{J}}\boldsymbol{\tau} \end{pmatrix} \right\}, \quad \text{under } P_0,$$

where $\boldsymbol{\Sigma} = E(\boldsymbol{\Psi}_t \boldsymbol{\Upsilon} \boldsymbol{\Psi}_t^\top)$, $\boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) = \boldsymbol{R}E[\boldsymbol{\Psi}_{t-1}E_{t-1}\{\boldsymbol{V}(\eta_t)\boldsymbol{s}^\top_{t,\boldsymbol{\theta}_0}(\eta_t)\}]\boldsymbol{\tau}$.

In the particular case where (2.3) holds, we thus have

$$\boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) = \boldsymbol{R}E(\boldsymbol{\Psi}_{t-1})E\{\boldsymbol{V}(\eta_t)\boldsymbol{f}^\top_{\boldsymbol{\theta}_0}(\eta_t)\}\boldsymbol{\tau} - \boldsymbol{R}E\left[\boldsymbol{\Psi}_{t-1}E\{g_{\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{V}(\eta_t)\}\frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right]\boldsymbol{\tau}.$$

Le Cam's third lemma and the contiguity of the probabilities P_0 and $P_{n,\boldsymbol{\tau}}$ (a consequence of the LAN property) entail that

$$\sqrt{n}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{r}\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}), \boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}^\top\right) \quad \text{under } H_n. \tag{3.3}$$

The Wald test, at asymptotic level $\underline{\alpha} \in (0,1)$, is defined by the rejection region $\{W_{n,f} > \chi^2_p(1-\underline{\alpha})\}$ where $\chi^2_p(1-\underline{\alpha})$ is the $(1-\underline{\alpha})$-quantile of the chi-square distribution with p degrees of freedom and

$$W_{n,f} = n\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{r}\right)^\top \{\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}\boldsymbol{R}^\top\}^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{r}\right),$$

where $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$. Under H_n, in view of (3.3), $W_{n,f}$ follows asymptotically a non-central chi-square distribution with p degrees of freedom and non-centrality parameter

$$\boldsymbol{c}^\top_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau})\{\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}^\top\}^{-1}\boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}).$$

Denoting by $\Phi_{\boldsymbol{\tau}}$ the cdf of this distribution, the Wald test has Local Asymptotic Power (LAP) $1 - \Phi_{\boldsymbol{\tau}}\{\chi^2_p(1-\underline{\alpha})\}$.

The following proposition can be used to quantify the local asymptotic efficiency loss of the QMLE with respect to the MLE for testing linear restrictions on parameters involved in the volatility or/and the density of the innovations.

Proposition 3.1. *Assume* **A1**$(\boldsymbol{\theta}_0)$, *either* **A2**-**A4** *or* **A2***-**A3***, **A5**-**A6** *and* (2.3). *For the MLE satisfying* (3.1) *and the QMLE satisfying* (3.2), *we have* $\boldsymbol{c}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) = \boldsymbol{R}\boldsymbol{\tau}$.

4 Examples

In this section we present two examples of popular GARCH specifications for which our LAN result can be derived, under more explicit assumptions than in the general model. The first example deals with a class of nonlinear GARCH models for which the smoothness assumptions required in Proposition 2.1 are not satisfied. We will therefore rely on Proposition 2.3. The second example illustrates a situation where the volatility and density have common parameters.

4.1 Application to APARCH(1,1) models with Student errors

The following generalized asymmetric Student-t distribution was proposed by Zhu and Galbraith (2010)

$$f_{\boldsymbol{\theta}}(y) = \begin{cases} \frac{\alpha}{\alpha^*} K(\nu_1) \left[1 + \frac{1}{\nu_1}\left(\frac{y}{2\alpha^*}\right)^2\right]^{-\frac{\nu_1+1}{2}}, & y \leq 0, \\ \frac{1-\alpha}{1-\alpha^*} K(\nu_2) \left[1 + \frac{1}{\nu_2}\left(\frac{y}{2(1-\alpha^*)}\right)^2\right]^{-\frac{\nu_2+1}{2}}, & y > 0, \end{cases} \tag{4.1}$$

where $K(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}$ (where $\Gamma(\cdot)$ is the Gamma function), $\alpha \in (0,1)$ is the skewness parameter, $\nu_1, \nu_2 > 0$ are respectively the left and right tail parameters, and α^* is defined as $\alpha^* = \alpha K(\nu_1)/[\alpha K(\nu_1) + (1-\alpha)K(\nu_2)]$. This density is continuous (in y) and admits a finite variance provided $\nu_1 \wedge \nu_2 > 2$. See Zhu and Galbraith (2010) for a detailed study of this distribution, including the asymptotic properties of the ML estimator for iid observations.

Consider the class of APARCH (Asymmetric Power ARCH) models introduced by Ding, Granger and Engle (1993), defined as

$$\begin{cases} \epsilon_t & = & \sigma_t(\boldsymbol{\theta}_0)\eta_t, \\ \sigma_t^\delta(\boldsymbol{\theta}) & = & \omega + \alpha_+|\epsilon_{t-1}|^\delta \mathbb{1}_{\epsilon_{t-1}>0} + \alpha_-|\epsilon_{t-1}|^\delta \mathbb{1}_{\epsilon_{t-1}<0} + \beta\sigma_{t-1}^\delta, \end{cases} \tag{4.2}$$

and assume that the density of η_t is given by (4.1) with parameters indexed by 0. Let

$$\boldsymbol{\theta} = (\omega, \alpha_+, \alpha_-, \beta, \delta, \alpha, \nu_1, \nu_2)' \in \Theta \subset [\underline{\omega}, \infty) \times [0,\infty)^2 \times [0,1) \times (0,\infty) \times (0,1) \times (0,\infty)^2. \tag{4.3}$$

Corollary 4.1 (**APARCH with asymmetric Student innovation**). *The LAN property holds for Model* (4.1)-(4.2) *if* Θ *satisfies* (4.3) *and*

$$E \log a_{\boldsymbol{\theta}_0}(\eta_1) < 0, \quad \textit{where} \quad a_{\boldsymbol{\theta}}(z) = \alpha_+ z^\delta \mathbb{1}_{z>0} + \alpha_-|z|^\delta \mathbb{1}_{z<0} + \beta.$$

For this model, despite the lack of differentiability of the density function, the LAN property holds under the strict stationarity condition. The following example shows that the strict stationarity condition may not suffice for the LAN property to hold. A similar situation occurs for ARMA models where the LAN property is satisfied if the parameter space is chosen in such a way that both the AR and MA polynomials have no zeros with magnitude less or equal to one (see Kreiss, 1987). A unit root in the AR part can also be handled (see Ling and McAleer, 2003).

4.2 Application to the Beta-t-GARCH(1,1)

The class of the Beta-t-GARCH was studied by Harvey (2013) and Creal, Koopman and Lucas (2013). Assume that the errors of the GARCH model follow a Student's t distribution with ν degrees of freedom, that is

$$f_{\boldsymbol{\theta}}(y) = \frac{1}{\sqrt{(\nu-2)\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \tag{4.4}$$

with $\nu > 2$, and assume that

$$\sigma_t^2(\boldsymbol{\theta}) = \omega + \beta\sigma_{t-1}^2(\boldsymbol{\theta}) + \alpha\frac{(\nu+1)\epsilon_{t-1}^2}{(\nu-2) + \epsilon_{t-1}^2/\sigma_{t-1}^2(\boldsymbol{\theta})}, \tag{4.5}$$

where $\boldsymbol{\theta} = (\omega, \alpha, \beta, \nu)'$ belongs to the parameter space Θ, a subset of $(\underline{\omega}, \infty)^2 \times [0,1) \times (2, \infty)$ for some $\underline{\omega} > 0$. Note that the parameter ν is involved in both the density and the volatility.

By the Cauchy root test, it can be easily seen that, at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, there exists a stationary and ergodic solution to this model, explicitly given by $\epsilon_t = \sigma_t\eta_t$ with

$$\sigma_t^2 = \sigma_t^2(\boldsymbol{\theta}_0) = \omega_0\left\{1 + \sum_{i=1}^{\infty} a_{\boldsymbol{\theta}_0}(\eta_{t-1})\cdots a_{\boldsymbol{\theta}_0}(\eta_{t-i})\right\}, \quad a_{\boldsymbol{\theta}}(z) = \alpha\frac{(\nu+1)z^2}{\nu - 2 + z^2} + \beta,$$

when $\boldsymbol{\theta}_0$ is such that

$$E\log a_{\boldsymbol{\theta}_0}(\eta_1) < 0. \tag{4.6}$$

The arguments of the proof of Lemma 2.3 in Berkes, Horváth and Kokoszka (2003) entail that under (4.6) there exists $s > 0$, such that

$$E|\epsilon_t|^s < \infty, \qquad E\sigma_t^s < \infty. \tag{4.7}$$

Assumption $\mathbf{A1}(\boldsymbol{\theta}_0)$ also requires stationarity of the sequence $\{\sigma_t(\boldsymbol{\theta})\}$ together with $\sigma_t(\boldsymbol{\theta}) \in \mathcal{F}_{t-1}$ for any value $\boldsymbol{\theta}$ of the parameter space. This property requires additional conditions contrary to the previous example where it was trivially satisfied under the condition $|\beta| < 1$. Note that $\sigma_t^2(\boldsymbol{\theta})$ is a solution of a Stochastic Recurrence Equation (SRE) of the form

$$\sigma_t^2(\boldsymbol{\theta}) = \varphi(\epsilon_{t-1}^2, \sigma_{t-1}^2(\boldsymbol{\theta})), \qquad \varphi(\epsilon^2, \sigma^2) = \alpha\frac{(\nu+1)\epsilon^2}{\nu - 2 + \epsilon^2/\sigma^2} + \beta\sigma^2.$$

According to the SRE theory (see Straumann and Mikosch, 2006) the model is invertible at $\boldsymbol{\theta}$, *i.e.* $\sigma_t^2(\boldsymbol{\theta})$ can be written as a measurable function of $\{\epsilon_u, u < t\}$, if

$$i)\ E \log \sup_{\sigma^2} \left| \frac{\partial \varphi(\epsilon_t^2, \sigma^2)}{\partial \sigma^2} \right| < 0, \qquad ii)\ E \log^+ \left| \varphi(\epsilon_t^2, \sigma_0^2) \right| < \infty$$

for some $\sigma_0^2 > 0$. Condition $ii)$ is always satisfied and, since $\sigma_t^2 \geq \omega/(1-\beta)$ condition $i)$ holds if

$$E \log \left(\alpha \frac{(\nu+1)\epsilon_1^4}{\left\{(\nu-2)\omega/(1-\beta)+\epsilon_1^2\right\}^2} + \beta \right) < 0. \tag{4.8}$$

Note that the constraint (4.8), which depends on $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, can be tested using Monte Carlo simulations. We thus have seen that $\mathbf{A1}(\boldsymbol{\theta}_0)$ is satisfied under (4.6) and (4.8). Assumption **A2** holds true without additional conditions. Now, note that

$$g_{\boldsymbol{\theta}}(y) = 1 - \frac{(\nu+1)y^2}{\nu-2+y^2}, \quad \boldsymbol{f}_{\boldsymbol{\theta}}(y) = \begin{pmatrix} \mathbf{0}_3 \\ \frac{1}{2}\left\{ \frac{\nu}{\nu-2} + \psi_0\left(\frac{\nu+1}{2}\right) - \psi_0\left(\frac{\nu}{2}\right) - \log\left(1+\frac{y^2}{\nu-2}\right) - \frac{\nu+1}{\nu-2+y^2} \right\} \end{pmatrix},$$

where $\psi_0(x) = \log'\{\Gamma(x)\}$ is the digamma function. The first two moment conditions of **A3** are thus satisfied. The last condition is implied by Lemma G.1 in the appendix.

Now we turn to **A4**. We have

$$\frac{\partial^2 \log f_{\boldsymbol{\theta}}(y)}{\partial \nu^2} = \frac{1}{4}\left\{ \frac{-1}{(\nu-2)^2} + \psi_1\left(\frac{\nu+1}{2}\right) - \psi_1\left(\frac{\nu}{2}\right) + \frac{y^2}{(\nu-2+y^2)(\nu-2)} - \frac{y^2-3}{(\nu-2+y^2)^2} \right\},$$

where ψ_1 is the trigamma function. Note that this function is bounded. Thus the first moment condition in **A4** is satisfied. The second inequality is also satisfied using (4.7), the elementary inequality $\log(1+y) \leq K(1+y^s)$ for $y > 0$ and the lower bound for $\sigma_t(\boldsymbol{\theta})$. Moreover the function $y g'_{\boldsymbol{\theta}}(y)$ being bounded, the third condition is satisfied for any p_1. Similarly, the fifth and seventh inequalities hold for any p_2, p_3. Thus **A4** is satisfied provided, for some $r > 0$,

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{1+r} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\|^{1+r} < \infty. \tag{4.9}$$

These moment conditions require an extension of Lemma G.1 which is discussed in Blasques, Koopman and Lucas (2014) through the notion of moment preserving maps. We have shown the following result.

Corollary 4.2 (Beta-t-GARCH). *The LAN property holds for Model* (4.4)-(4.5) *with* $\beta_0 \neq 0$ *if* (4.6), (4.8) *and* (4.9) *are satisfied.*

For the sake of illustration we consider testing the assumption $H_0 : \nu = \nu_0$ against $H_n : \nu = \nu_0 + \tau/\sqrt{n}$ in Model (4.4)-(4.5) with $\omega_0 = 0.5, \alpha_0 = 0.1, \beta_0 = 0.88$. The LAPs of the tests based on the QMLE and MLE are displayed in Figure 1. By Proposition 3.1, these LAPs only differ by the asymptotic variances $\boldsymbol{\Sigma}$ of the estimators, which were numerically obtained from simulations of size $n = 100,000$. As expected the discrepancy is large for small values of ν_0 and reduces as ν_0 increases, with a degeneracy of the two powers at $\nu = \infty$ since the parameter is no longer identifiable. Next, we consider testing the assumption $H_0 : \alpha = \alpha_0$ against $H_n : \alpha = \alpha_0 + \tau/\sqrt{n}$ for the same model. The LAPs of the tests based on the QMLE and MLE are displayed in Figures 2 (when ν_0 varies) and 3 (when β_0 varies). The efficiency loss when going from ML to QML tends to zero as ν_0 increases. On the contrary when β_0 varies for a given value of ν_0, the efficiency loss is not much affected. Note that the strict stationarity condition (4.6) is satisfied also for the bottom panels with $\alpha_0 + \beta_0 > 1$. Contrary to the test of ν_0, the powers of the test of α_0 do not diminish when ν_0 increases (compare the range of values of τ in Figures 1 and 2-3). Surprisingly, the LAP of the test of α_0 improves when β_0 approaches 1.

5 Including a conditional mean

In this section, we extend our LAN results to the conditional location-scale model

$$y_t = m_t(\boldsymbol{\theta}_0) + \epsilon_t(\boldsymbol{\theta}_0), \quad \epsilon_t(\boldsymbol{\theta}_0) = \sigma_t(\boldsymbol{\theta}_0)\eta_t \tag{5.1}$$

under the same assumptions on (η_t) and $\boldsymbol{\theta}$ as in the previous sections, with $m_t(\boldsymbol{\theta}_0) \in \mathcal{F}_{t-1}$ for all $\boldsymbol{\theta} \in \Theta$. The conditional log-likelihood ratio has the same expression as before with

$$\eta_t(\boldsymbol{\theta}) = \frac{\epsilon_t(\boldsymbol{\theta})}{\sigma_t(\boldsymbol{\theta})} = \frac{y_t - m_t(\boldsymbol{\theta})}{\sigma_t(\boldsymbol{\theta})}.$$

We start by studying the LAN property under differentiability. We introduce the following assumptions.

B1($\boldsymbol{\theta}_0$): (y_t) satisfies (5.1) where η_t has density $f_{\boldsymbol{\theta}_0}$ and, for all $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, $\{m_t(\boldsymbol{\theta}), \sigma_t(\boldsymbol{\theta})\}$ is a stationary sequence with $m_t(\boldsymbol{\theta}), \sigma_t(\boldsymbol{\theta}) \in \mathcal{F}_{t-1}$ and $\sigma_t(\boldsymbol{\theta}) > 0$.

B2: For all $t \geq 1$, $\boldsymbol{\theta} \mapsto m_t(\boldsymbol{\theta})$ has continuous second-order derivatives and $E\|\frac{1}{\sigma_t(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\|^2 < \infty$.

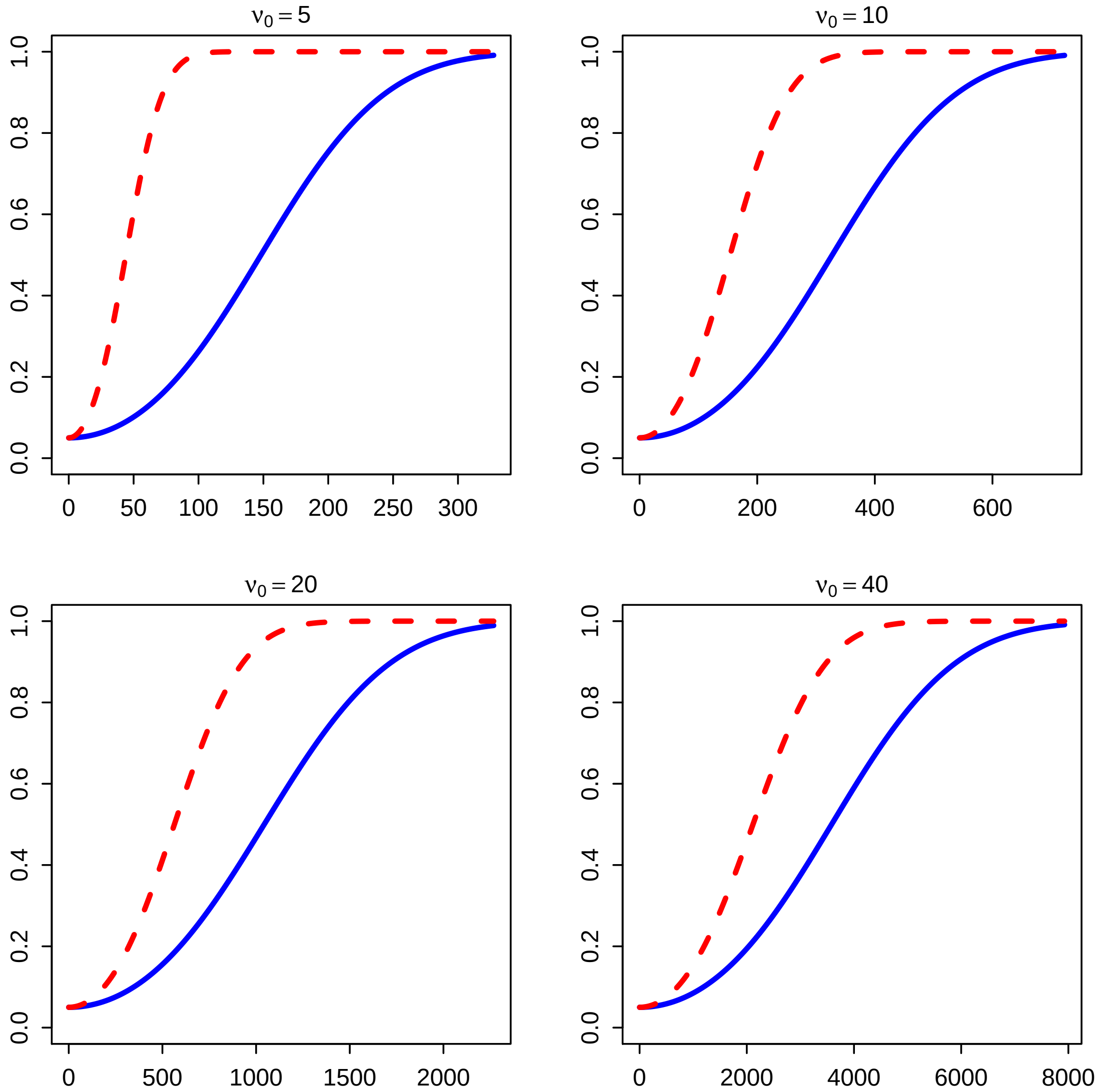

Figure 1: LAPs of the tests of $H_0 : \nu = \nu_0$ based on QML (blue line) and ML (dotted red line), as functions of τ, for the Beta-t-GARCH.

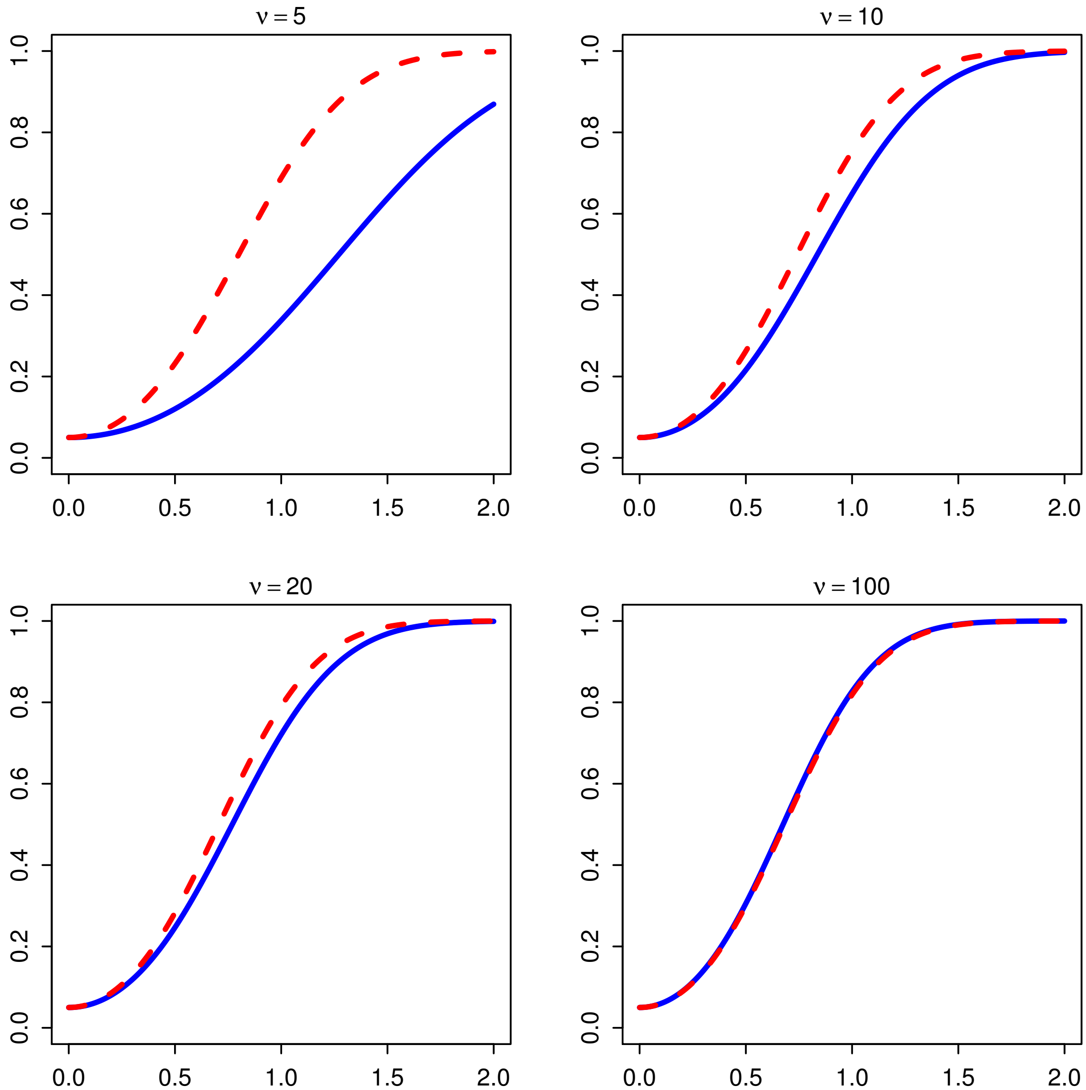

Figure 2: LAPs of the tests of $H_0 : \alpha = 0.1$ based on QML (blue line) and ML (dotted red line), as functions of τ, for the Beta-t-GARCH with different values of ν (and $\beta_0 = 0.88$).

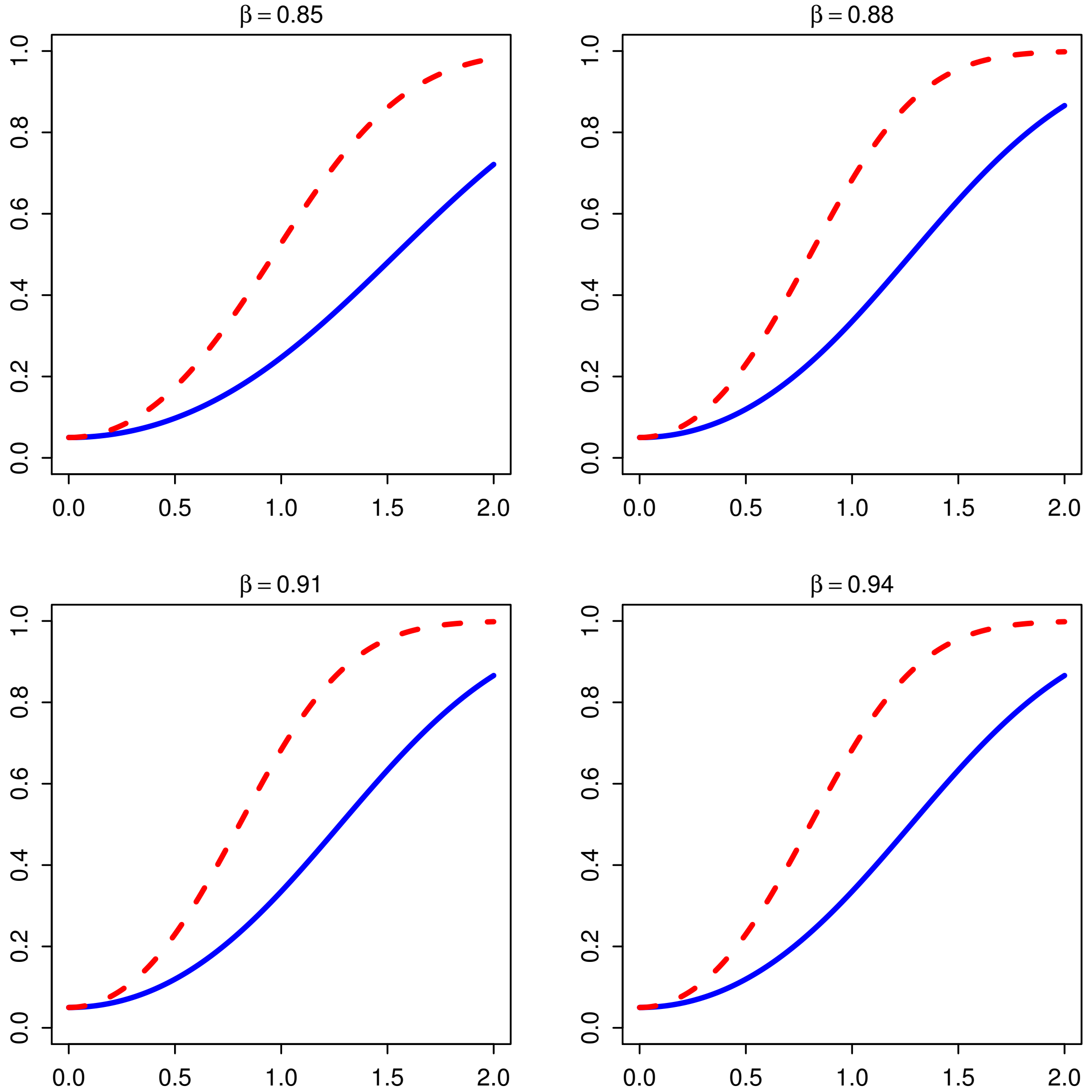

Figure 3: LAPs of the tests of $H_0 : \alpha = 0.1$ based on QML (blue line) and ML (dotted red line), as functions of τ, for the Beta-t-GARCH with different values of β (and $\nu_0 = 5$).

B3: We have $E\|\frac{f'_{\boldsymbol{\theta}_0}}{f_{\boldsymbol{\theta}_0}}(\eta_t)\|^2 < \infty$. Moreover, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and four pairs of conjugate numbers $p_i > 1$, $q_i > 1$, $1/p_i + 1/q_i = 1$, for $i = 4, 5, 6, 7$, such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \left(\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right)' (\eta_t(\boldsymbol{\theta})) \right|^{p_4} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{2q_4} < \infty,$$

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \left(\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right) (\eta_t(\boldsymbol{\theta})) \right|^{p_5} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\|^{q_5} < \infty,$$

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| g'_{\boldsymbol{\theta}} (\eta_t(\boldsymbol{\theta})) \right|^{p_6} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{q_6} < \infty,$$

and

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \boldsymbol{f}'_{\boldsymbol{\theta}} (\eta_t(\boldsymbol{\theta})) \right\|^{p_7} < \infty, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{q_7} < \infty.$$

Let $\boldsymbol{D}_t = \frac{1}{\sigma_t} \left(\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}, \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} \right)^\top$ and

$$\mathfrak{I} = \iota_f \boldsymbol{J}_{\sigma\sigma} - \upsilon_f (\boldsymbol{J}_{m\sigma} + \boldsymbol{J}_{\sigma m}) + \gamma_f \boldsymbol{J}_{mm} - \boldsymbol{\Omega}_\sigma \boldsymbol{f}^\top - \boldsymbol{f} \boldsymbol{\Omega}_\sigma^\top - \boldsymbol{h} \boldsymbol{\Omega}_m^\top - \boldsymbol{\Omega}_m \boldsymbol{h}^\top + \boldsymbol{F}, \tag{5.2}$$

with (recalling some notations) $\iota_f = E g^2_{\boldsymbol{\theta}_0}(\eta_t)$, $\upsilon_f = E g'_{\boldsymbol{\theta}_0}(\eta_t)$, $\gamma_f = E\left[\left(\frac{f'_{\boldsymbol{\theta}_0}}{f_{\boldsymbol{\theta}_0}} \right)' (\eta_t) \right]$, $\boldsymbol{J} = E \boldsymbol{D}_t \boldsymbol{D}_t^\top = \begin{pmatrix} \boldsymbol{J}_{mm} & \boldsymbol{J}_{m\sigma} \\ \boldsymbol{J}_{\sigma m} & \boldsymbol{J}_{\sigma\sigma} \end{pmatrix}$, $\boldsymbol{\Omega} = E \boldsymbol{D}_t = \begin{pmatrix} \boldsymbol{\Omega}_m \\ \boldsymbol{\Omega}_\sigma \end{pmatrix}$, $\boldsymbol{F} = E \boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) \boldsymbol{f}^\top_{\boldsymbol{\theta}_0}(\eta_t)$, $\boldsymbol{h} = E \boldsymbol{f}'_{\boldsymbol{\theta}_0}(\eta_t)$, and $\boldsymbol{f} = E g_{\boldsymbol{\theta}_0}(\eta_t) \boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)$.

The central sequence is now given by

$$\boldsymbol{\Delta}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) - g_{\boldsymbol{\theta}_0}(\eta_t) \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{f'_{\boldsymbol{\theta}_0}}{f_{\boldsymbol{\theta}_0}}(\eta_t) \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\}.$$

Proposition 5.1. *Let Θ be a bounded convex subset of $\mathbb{R}^d$ such that $\boldsymbol{\theta}_0 \in \Theta$. Assume* **B1**$(\boldsymbol{\theta}_0)$, **A2**-**A4** *and* **B2**-**B3**. *When $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n} \in \Theta$ for n large enough, we have the LAN property*

$$\Lambda_n(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) = \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n - \frac{1}{2} \boldsymbol{\tau}^\top \mathfrak{I} \boldsymbol{\tau} + o_{P_0}(1) \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2} \boldsymbol{\tau}^\top \mathfrak{I} \boldsymbol{\tau}, \boldsymbol{\tau}^\top \mathfrak{I} \boldsymbol{\tau} \right) \quad \text{under } P_0.$$

When differentiability does not hold, the previous assumptions can be replaced by the following conditions.

B2*: For all $t \in \mathbb{Z}$, there exists a vector $\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) := \boldsymbol{s}_{\boldsymbol{\theta}_0}(y, \eta_{t-1}, \eta_{t-2}, \ldots) \in \mathbb{R}^d$ where $\boldsymbol{s}_{\boldsymbol{\theta}_0}$ is a measurable function, such that

$$\sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{h})} f_{\boldsymbol{\theta}_0 + \boldsymbol{h}} \left(\frac{m_t(\boldsymbol{\theta}_0) - m_t(\boldsymbol{\theta}_0 + \boldsymbol{h})}{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{h})} + \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{h})} y \right)}$$
$$= \sqrt{f_{\boldsymbol{\theta}_0}(y)} + \frac{1}{2} \boldsymbol{h}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) \sqrt{f_{\boldsymbol{\theta}_0}(y)} + r_{t,\boldsymbol{h}}(y), \qquad \|r_{t,\boldsymbol{h}}(\cdot)\|^2_{L^2(\mu)} = o_{P_0}(\|\boldsymbol{h}\|^2). \tag{5.3}$$

B3*: The following matrix exists

$$\mathfrak{I} := E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{s}_{t,\boldsymbol{\theta}_0}^\top(\eta_t)).$$

Proposition 5.2. *Proposition 5.1 remains valid when* **A2**-**A4** *and* **B2-B3** *are replaced by* **B2***-**B3*** *and the central sequence is defined by* $\boldsymbol{\Delta}_n = n^{-1/2}\sum_{t=1}^n \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)$.

6 Conclusion

In this paper, we proved the LAN property for general conditional location-scale models where the parameter of the errors density has common components with that of the mean and volatility. A typical example where this situation occurs is the case of some score-driven volatility models. Our assumptions on the volatility model are rather weak, in particular they are compatible with high persistence introduced through ARCH(∞) models (see e.g. Robinson and Zaffaroni (2006), Royer (2022)). The introduction of the notion of CQMD allows to handle situations where some regularity assumptions on the volatility and/or the density functions are in failure. As examples of application of the LAN property, we consider tests of linear restrictions. Using the LAN property, we are able to quantify the asymptotic discrepancy in local power between the QML and ML estimators. Interesting future areas of research are the extension of the framework of this article to more general score-driven specifications, or to multivariate models.

APPENDIX

A Proof of Proposition 2.1

Note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log \left\{ \frac{1}{\sigma_t(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \right\} = -g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \boldsymbol{f}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)$$

and

$$\begin{aligned}\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log \left\{ \frac{1}{\sigma_t(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \right\} =& - g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{\boldsymbol{\theta}}^\top\left(\eta_t(\boldsymbol{\theta})\right) \\ &+ g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\ &- \eta_t(\boldsymbol{\theta}) \boldsymbol{f}'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{F}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)\end{aligned}$$

where $\boldsymbol{f}'_{\boldsymbol{\theta}}(y)$ denotes the vector of the derivatives of the elements of $\boldsymbol{f}_{\boldsymbol{\theta}}(y)$. Note that $y\boldsymbol{f}'_{\boldsymbol{\theta}}(y) = \boldsymbol{g}_{\boldsymbol{\theta}}(y)$. A Taylor expansion of $\boldsymbol{\theta}_n \mapsto \Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ thus yields

$$\Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n - \frac{1}{2} \boldsymbol{\tau}^\top \mathfrak{J}_n(\boldsymbol{\theta}_n^*) \boldsymbol{\tau}, \tag{A.1}$$

where $\boldsymbol{\theta}_n^*$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_n$, and

$$\begin{aligned}\mathfrak{J}_n(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{t=1}^n g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{1}{n} \sum_{t=1}^n g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\ &+ \frac{1}{n} \sum_{t=1}^n \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{\boldsymbol{\theta}}^\top\left(\eta_t(\boldsymbol{\theta})\right) + \frac{1}{n} \sum_{t=1}^n \boldsymbol{g}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - \frac{1}{n} \sum_{t=1}^n \boldsymbol{F}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right).\end{aligned}$$

Note that under $\mathbf{A1}(\boldsymbol{\theta}_0)$ and $\mathbf{A3}$, $\left\{(g_{\boldsymbol{\theta}_0}(\eta_t) \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}, \boldsymbol{f}_{\boldsymbol{\theta}_0}^\top(\eta_t))^\top, \mathcal{F}_t\right\}$ is a square integrable martingale difference. By the central limit theorem of Billingsley (1961) we have $\boldsymbol{\Delta}_n \xrightarrow{d} \mathcal{N}\left\{\mathbf{0}, \mathfrak{J}\right\}$ under P_0 as $n \to \infty$. Moreover, integrations by parts show that

$$\iota_f = -E g'_{\boldsymbol{\theta}_0}\left(\eta_t\right) \eta_t = -1 + \int y^2 \frac{(f'_{\boldsymbol{\theta}_0}(y))^2}{f_{\boldsymbol{\theta}_0}(y)} dy, \quad E\boldsymbol{g}_{\boldsymbol{\theta}_0}(\eta_t) = -\boldsymbol{f}.$$

For the last equality, we use the fact that $\partial \int f_{\boldsymbol{\theta}}(y) g_{\boldsymbol{\theta}}(y) dy / \partial \boldsymbol{\theta} = \mathbf{0}$ because $\int f_{\boldsymbol{\theta}}(y) g_{\boldsymbol{\theta}}(y) dy = 0$ for all $\boldsymbol{\theta}$. Note also that $\boldsymbol{F} = -E\boldsymbol{F}_{\boldsymbol{\theta}_0}(\eta_t)$. The ergodic theorem then entails that $\mathfrak{J}_n(\boldsymbol{\theta}_0) \to \mathfrak{J}$ a.s. as $n \to \infty$.

It remains to establish that, as $n \to \infty$,

$$\|\mathfrak{J}_n(\boldsymbol{\theta}_n^*) - \mathfrak{J}_n(\boldsymbol{\theta}_0)\| \to 0 \quad \text{in probability.} \tag{A.2}$$

We only give the proof of

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{F}_{\boldsymbol{\theta}_n^*} \{\eta_t(\boldsymbol{\theta}_n^*)\} - \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{F}_{\boldsymbol{\theta}_0} (\eta_t) \right\| \to 0 \quad \text{a.s.} \tag{A.3}$$

The other convergences showing (A.2) are obtained similarly. By the ergodic theorem, (A.3) is obtained by showing that for all $\varepsilon > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\boldsymbol{F}_{\boldsymbol{\theta}} \{\eta_t(\boldsymbol{\theta})\} - \boldsymbol{F}_{\boldsymbol{\theta}_0} (\eta_t)\| \leq \varepsilon.$$

By the dominated convergence theorem, **A2** and the first moment condition of **A4**, the left-hand side of the previous inequality tends to 0 when the neighborhood $V(\boldsymbol{\theta}_0)$ shrinks to the singleton $\{\boldsymbol{\theta}_0\}$, and (A.3) follows. The rest of the proof follows by the same arguments. □

B Proof of Proposition 2.2

Let $\boldsymbol{f}_i$ the i-th component of $\boldsymbol{f}_{\boldsymbol{\theta}_0}$ and $K = \sup_y \sup_{1 \leq i \leq d} |\boldsymbol{f}_i'(y)|$. We have, from **A5-A6**

$$\begin{aligned} \left\| \boldsymbol{\Delta}_n - \widetilde{\boldsymbol{\Delta}}_n \right\| &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} K\, |\eta_t - \widetilde{\eta}_t| \left(1 + \left\| \frac{\partial \log \tilde{\sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \right) + |g_{\boldsymbol{\theta}_0}(\eta_t)| \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{1}{\tilde{\sigma}_t(\boldsymbol{\theta}_0)} \frac{\partial \tilde{\sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \\ &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{\infty} K \rho^t (|\eta_t| + |g_{\boldsymbol{\theta}_0}(\eta_t)) \left(1 + \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \left\| \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \right). \end{aligned}$$

By **A3** and the first part of **A5**, the infinite sum is finite a.s. It follows that $\left\| \boldsymbol{\Delta}_n - \widetilde{\boldsymbol{\Delta}}_n \right\| = o_P(1)$. The conclusion follows. □

C Proof of Lemma 2.1

The proof is adapted from the iid case (see for instance Lehmann and Romano (2006), Lemma 12.2.1). We start by showing the second result. Taking $\boldsymbol{h} = h\boldsymbol{\tau}$ where $h > 0$, we get from **A2***

$$\|g_h - g\|_{L^2(\mu)} \to 0 \quad \text{when } h \to 0$$

where $g(y) = \frac{1}{2} \boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y) \sqrt{f_{\boldsymbol{\theta}_0}(y)}$ and

$$g_h(y) = \frac{1}{h} \left\{ \sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + h\boldsymbol{\tau})} f_{\boldsymbol{\theta}_0 + h\boldsymbol{\tau}} \left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0 + h\boldsymbol{\tau})} y \right)} - \sqrt{f_{\boldsymbol{\theta}_0}(y)} \right\}.$$

Since $\|g_h\|_{L^2(\mu)} < \infty$, it follows that $\|g\|^2_{L^2(\mu)} = \frac{1}{4}\boldsymbol{\tau}^\top \mathfrak{J}_t \boldsymbol{\tau} < \infty$.

Now taking, conditionally on $\mathcal{F}_{t-1}$, the squared $L^2(\mu)$-norm of both sides of the equality (2.2), we obtain

$$\begin{aligned} 0 &= \frac{1}{4}\boldsymbol{h}^\top \mathfrak{J}_t \boldsymbol{h} + \int r^2_{t,\boldsymbol{h}}(y)d\mu(y) + \boldsymbol{h}^\top E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|\mathcal{F}_{t-1}) \\ &\quad +2\int r_{t,\boldsymbol{h}}(y)\sqrt{f_{\boldsymbol{\theta}}(y)}d\mu(y) + \int \boldsymbol{h}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\sqrt{f_{\boldsymbol{\theta}}(y)}r_{t,\boldsymbol{h}}(y)d\mu(y) \quad a.s. \end{aligned}$$

Noting that, by the Cauchy-Schwarz inequality, $\int r_{t,\boldsymbol{h}}(y)\sqrt{f_{\boldsymbol{\theta}}(y)}d\mu(y) = o_{P_0}(\|\boldsymbol{h}\|)$ and $\int \boldsymbol{h}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\sqrt{f_{\boldsymbol{\theta}}(y)}r_{t,\boldsymbol{h}}(y)d\mu(y) = o_{P_0}(\|\boldsymbol{h}\|^2)$, and comparing the orders as $\boldsymbol{h} \to 0$, we deduce the first equality in (2.4) (a well known result when **A2** holds). □

D Proof of Proposition 2.3

Letting

$$W_{t,n} = \sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_n)} \frac{f_{\boldsymbol{\theta}_n}(\eta_t(\boldsymbol{\theta}_n))}{f_{\boldsymbol{\theta}_0}(\eta_t)}} - 1$$

and using $\log(y+1) = y - y^2/2 + y^2\xi(y)$ with $\xi(y) \to 0$ as $y \to 0$, we have

$$\Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = 2\sum_{t=1}^n \log(W_{t,n}+1) = 2\sum_{t=1}^n W_{t,n} - \sum_{t=1}^n W^2_{t,n} + 2\sum_{t=1}^n W^2_{t,n}\xi(W_{t,n}).$$

We will show that

$$2\sum_{t=1}^n \{W_{t,n} - E(W_{t,n} \mid \mathcal{F}_{t-1})\} = \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n + o_{P_0}(1), \tag{D.1}$$

$$2\sum_{t=1}^n E(W_{t,n} \mid \mathcal{F}_{t-1}) = -\frac{1}{4}\boldsymbol{\tau}^\top \mathfrak{J}\boldsymbol{\tau} + o_{P_0}(1), \tag{D.2}$$

$$\sum_{t=1}^n W^2_{t,n} = \frac{1}{4}\boldsymbol{\tau}^\top \mathfrak{J}\boldsymbol{\tau} + o_{P_0}(1), \tag{D.3}$$

$$\sum_{t=1}^n W^2_{t,n}\xi(W_{t,n}) = o_{P_0}(1). \tag{D.4}$$

Under **A1**$(\boldsymbol{\theta}_0)$ and the CQMD condition, it can be seen that $(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t))$ is a stationary and ergodic sequence. The conclusion will follow by noting that $\{\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t), \mathcal{F}_t\}$ is a square integrable martingale difference by (2.4) and **A3***.

By **A2***, we have

$$W_{t,n} - E\left(W_{t,n} \mid \mathcal{F}_{t-1}\right) = \frac{1}{2\sqrt{n}} \boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t) + R_{t,n}, \quad R_{t,n} = \frac{r_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{\sqrt{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)}} - E\left(\frac{r_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{\sqrt{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)}} \mid \mathcal{F}_{t-1}\right).$$

Noting that $(R_{t,n})$ is a stationary martingale difference, we have

$$\mathrm{Var}\left(\sum_{t=1}^n R_{t,n}\right) = n\mathrm{Var}\left(R_{t,n}\right) \leq nEE\left(\left\{\frac{r_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{\sqrt{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)}}\right\}^2 \mid \mathcal{F}_{t-1}\right) = nE\int r^2_{t,n^{-1/2}\boldsymbol{\tau}}(y)d\mu(y) = o(1)$$

where the last equality follows from (2.6). Thus (D.1) follows.

By **A2*** again, we have

$$\begin{aligned}
\sum_{t=1}^n E\left(W_{t,n} \mid \mathcal{F}_{t-1}\right) &= \sum_{t=1}^n \int \left\{\frac{\sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_n)} f_{\boldsymbol{\theta}_n}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_n)} y\right)}}{\sqrt{f_{\boldsymbol{\theta}_0}\left(y\right)}} - 1\right\} f_{\boldsymbol{\theta}_0}\left(y\right) d\mu(y) \\
&= -\frac{1}{2}\sum_{t=1}^n \int \left\{\sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_n)} f_{\boldsymbol{\theta}_n}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_n)} y\right)} - \sqrt{f_{\boldsymbol{\theta}_0}\left(y\right)}\right\}^2 d\mu(y) \\
&= -\frac{1}{2}\sum_{t=1}^n \int \left\{\frac{1}{2\sqrt{n}} \boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\sqrt{f_{\boldsymbol{\theta}_0}(y)} + r_{t,\boldsymbol{\tau}/\sqrt{n}}(y)\right\}^2 d\mu(y) \\
&= -\frac{1}{8n}\sum_{t=1}^n \int \left\{\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\right\}^2 f_{\boldsymbol{\theta}_0}(y)d\mu(y) + n\, o_{P_0}(\|\boldsymbol{\tau}/\sqrt{n}\|^2)
\end{aligned}$$

and (D.2) follows from the ergodic theorem and **A3***.

We also have

$$\sum_{t=1}^n W^2_{t,n} = \frac{1}{4n}\sum_{t=1}^n \left(\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\right)^2 + \sum_{t=1}^n \frac{r^2_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)} + \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\frac{r_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{\sqrt{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)}}.$$

By the ergodic theorem, the first term of the right-hand side of the equality tends almost surely to $\frac{1}{4}\boldsymbol{\tau}^\top \mathfrak{J}\boldsymbol{\tau}$. The expectation of the second term term is equal to $nE\int r^2_{t,n^{-1/2}\boldsymbol{\tau}}(y)d\mu(y) = o(1)$, and thus this positive term tends to zero in probability. The third term also tends to zero in probability, by the Cauchy-Schwarz inequality and the two previous convergence results. Therefore (D.3) is shown.

For all $\varepsilon > 0$ there exists $\delta > 0$ such that $|\xi(y)| \leq \varepsilon$ if $|y| \leq \delta$. Therefore we have

$$\begin{aligned}
&\sum_{t=1}^n W^2_{t,n}\xi(W_{t,n}) \leq \varepsilon \sum_{t=1}^n W^2_{t,n} + \sum_{t=1}^n W^2_{t,n} 1_{|W_{t,n}|>\delta} \\
&\leq \varepsilon\, O_{P_0}(1) + \frac{1}{n}\sum_{t=1}^n \left(\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\right)^2 1_{|\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|>n^{1/2}\delta} + 4\sum_{t=1}^n \frac{r^2_{t,n^{-1/2}\boldsymbol{\tau}}(\eta_t)}{f_{\boldsymbol{\theta}_0}\left(\eta_t\right)}
\end{aligned}$$

using (D.3) and the elementary inequality $(a+b)^2 1_{|a+b|>\delta} \leq 4a^2 1_{|a|>\delta/2} + 4b^2$. We have already seen that the last sum is an $o_P(1)$. Now, for all $M > 0$, when n is sufficiently large we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\right)^2 1_{|\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|>n^{1/2}\delta} \leq \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\right)^2 1_{|\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|>M}$$

and, by the ergodic theorem, $\mathbf{A1}(\boldsymbol{\theta}_0)$ and $\mathbf{A3}^*$, the right-hand side converges almost surely to $E\left(\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\right)^2 1_{|\boldsymbol{\tau}^\top \boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|>M}$, which is arbitrarily small when M is large. The conclusion follows.

□

E Proof of Proposition 3.1

For the MLE, by (3.1) we find

$$\boldsymbol{c}^{ML}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) = \text{Cov}_{as}\left(\boldsymbol{R}\boldsymbol{\mathfrak{J}}^{-1}\boldsymbol{\Delta}_n, \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n\right) = \boldsymbol{R}\boldsymbol{\tau}$$

and for the QMLE, by (3.2),

$$\begin{aligned}\boldsymbol{c}^{QML}_{\boldsymbol{\theta}_0,f}(\boldsymbol{\tau}) =& \text{Cov}\left(\frac{1}{2}\boldsymbol{R}\boldsymbol{J}^{-1}(\eta_t^2-1)\frac{\partial \log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}, \boldsymbol{\tau}^\top \boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) - g_{\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{\tau}^\top \frac{\partial \log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)\\ =& \frac{1}{2}\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{\Omega}\boldsymbol{\tau}^\top E(\eta_t^2-1)\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) + \frac{1}{2}E\left[(1-\eta_t^2)g_{\boldsymbol{\theta}_0}(\eta_t)\right]\boldsymbol{R}\boldsymbol{\tau}.\end{aligned}$$

Now we have

$$\begin{aligned}E(\eta_t^2-1)\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t) =& E\eta_t^2\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\\ =& E\int x^2\frac{\partial}{\partial\boldsymbol{h}}\log\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}f_{\boldsymbol{\theta}_0+\boldsymbol{h}}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}x\right)\Bigg|_{\boldsymbol{h}=0} f_{\boldsymbol{\theta}_0}(x)dx\\ =& E\int x^2\frac{\partial}{\partial\boldsymbol{h}}\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}f_{\boldsymbol{\theta}_0+\boldsymbol{h}}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}x\right)\Bigg|_{\boldsymbol{h}=0} dx\\ =& E\frac{\partial}{\partial\boldsymbol{h}}\int x^2\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}f_{\boldsymbol{\theta}_0+\boldsymbol{h}}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+\boldsymbol{h})}x\right)dx\Bigg|_{\boldsymbol{h}=0}\\ =& E\frac{\partial}{\partial\boldsymbol{h}}\frac{\sigma_t^2(\boldsymbol{\theta}_0+\boldsymbol{h})}{\sigma_t^2(\boldsymbol{\theta}_0)}\int y^2 f_{\boldsymbol{\theta}_0+\boldsymbol{h}}(y)dx\Bigg|_{\boldsymbol{h}=0}\\ =& E\frac{\partial}{\partial\boldsymbol{h}}\frac{\sigma_t^2(\boldsymbol{\theta}_0+\boldsymbol{h})}{\sigma_t^2(\boldsymbol{\theta}_0)}\Bigg|_{\boldsymbol{h}=0} = E\,\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 2\boldsymbol{\Omega}.\end{aligned}$$

Moreover,

$$E(\eta_t^2-1)\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t) = E(\eta_t^2-1)\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t) - E(\eta_t^2-1)g_{\boldsymbol{\theta}_0}(\eta_t)E\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}},$$

with

$$E(\eta_t^2-1)g_{\boldsymbol{\theta}_0}(\eta_t)=\int(x^2-1)\left(1+x\frac{f'_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}}(x)}\right)f_{\boldsymbol{\theta}}(x)dx=1+\int x^3f'_{\boldsymbol{\theta}}(x)dx=-2.$$

It follows that

$$E(\eta_t^2-1)\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)=2\boldsymbol{\Omega}+E(\eta_t^2-1)g_{\boldsymbol{\theta}_0}(\eta_t)E\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}=2\boldsymbol{\Omega}-2\boldsymbol{\Omega}=0.$$

□

F Proof of Corollary 4.1

Note that $E\log^+ a_{\boldsymbol{\theta}_0}(\eta_1)<\infty$ because $E\log^+|\eta_t|<\infty$. It follows that, by the Cauchy rule

$$\sigma_t^{\delta_0}(\boldsymbol{\theta}_0)=\omega_0+a_{\boldsymbol{\theta}_0}(\eta_{t-1})\sigma_{t-1}^{\delta_0}=\omega_0\left(1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}a_{\boldsymbol{\theta}_0}(\eta_{t-j})\right).$$

Therefore $\mathbf{A1}(\boldsymbol{\theta}_0)$ reduces to $E\log a_{\boldsymbol{\theta}_0}(\eta_1)<0$ and $\sup_\Theta\beta<1$. For some $\boldsymbol{\theta}$, the function $y\mapsto f_{\boldsymbol{\theta}}(y)$ is differentiable only once at $y=0$. Therefore **A2** is not satisfied and the result cannot be obtained from Proposition 2.1. We will show the CQMD of Proposition 2.3.

By Lemma 2.1 of Garel and Hallin (1995) (see also Lind and Roussas (1972)) multivariate QMD is equivalent to partial QMD component by component. Note that a similar property does not hold for the classical differentiability. Reasoning conditional to $\mathcal{F}_{t-1}$, establishing **A2*** is thus equivalent to showing, for $i=1,\dots,d$,

$$\frac{1}{h^2}\int\left\{\sqrt{\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+h\boldsymbol{e}_i)}f_{\boldsymbol{\theta}_0+h\boldsymbol{e}_i}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+h\boldsymbol{e}_i)}y\right)}-\sqrt{f_{\boldsymbol{\theta}_0}(y)}-\frac{1}{2}h\boldsymbol{e}_i^\top\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\sqrt{f_{\boldsymbol{\theta}_0}(y)}\right\}^2dy=o_P(1)$$

as $h\to 0$, where $\boldsymbol{e}_i$ is the i-th element of the canonical basis of $\mathbb{R}^d$ and $\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)\in\mathcal{F}_{t-1}$. We will show the result with

$$\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(y)=\boldsymbol{f}_{\boldsymbol{\theta}_0}(y)-g_{\boldsymbol{\theta}_0}(y)\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}.$$

By Proposition 2 in Zhu and Galbraith (2010), the information matrix $\boldsymbol{F}=E\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_1)\boldsymbol{f}_{\boldsymbol{\theta}_0}^\top(\eta_1)$ exists and is continuous. Noting that $g_{\boldsymbol{\theta}}(\cdot)$ is bounded, $\iota_f=Eg^2_{\boldsymbol{\theta}}(\eta_t)$ and $\boldsymbol{f}=Eg_{\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{f}_{\boldsymbol{\theta}_0}(\eta_t)$ exist. Moreover, they are continuous at $\boldsymbol{\theta}_0$. It follows that

$$\mathfrak{I}_t=E\left(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{s}_{t,\boldsymbol{\theta}_0}^\top(\eta_t)\mid\mathcal{F}_{t-1}\right)=\iota_f\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}-\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}\boldsymbol{f}^\top-\boldsymbol{f}\frac{\partial\log\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}+\boldsymbol{F}$$

exists and is continuous at $\boldsymbol{\theta}_0$. Given $\mathcal{F}_{t-1}$, the application $h \mapsto \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+h\boldsymbol{e}_i)} f_{\boldsymbol{\theta}_0+h\boldsymbol{e}_i}\left(\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0+h\boldsymbol{e}_i)} y\right)$ is continuously differentiable, and thus absolutely continuous in a neighborhood of 0. By Theorem 12.2.1 in Lehmann and Romano (2006) (see also Theorem 1.117 in Liese and Miescke (2008)) the result follows by the fact that $\boldsymbol{e}_i^\top \mathfrak{J}_t \boldsymbol{e}_i$ exists and is continuous. Hamadeh and Zakoian (2011) showed that $\partial \log \sigma_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ admits moments of any order (see their Equation 5.20). It follows that $\mathfrak{J} = E\mathfrak{J}_t$ exists, which shows **A3*** and completes the proof. □

G Complement to the proof of Corollary 4.2

Lemma G.1. *Under* (4.6)*, when* $\beta_0 \neq 0$*, the Beta-t-GARCH(1,1) satisfies*

$$E\left\|\frac{\partial \log \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\|^r < \infty, \quad \text{for all } r > 0.$$

Proof. Letting $a_t(\boldsymbol{\theta}) = a_{\boldsymbol{\theta}}(\eta_t(\boldsymbol{\theta}))$, for all $i \geq 1$ we have

$$\sigma_t^2(\boldsymbol{\theta}) = \omega\left\{1 + \sum_{k=1}^{i-1}\prod_{j=1}^{k} a_{t-j}(\boldsymbol{\theta})\right\} + \sigma_{t-i}^2(\boldsymbol{\theta})\prod_{j=1}^{i} a_{t-j}(\boldsymbol{\theta}).$$

Therefore

$$\frac{\sigma_{t-i}^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})} \leq \frac{1}{\prod_{j=1}^{i} a_{t-j}(\boldsymbol{\theta})}.$$

We also have

$$\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} 1 \\ \frac{(\nu+1)\epsilon_{t-1}^2}{(\nu-2)+\frac{\epsilon_{t-1}^2}{\sigma_{t-1}^2(\boldsymbol{\theta})}} \\ \sigma_{t-1}^2(\boldsymbol{\theta}) \\ \frac{\alpha\epsilon_{t-1}^2}{(\nu-2)+\frac{\epsilon_{t-1}^2}{\sigma_{t-1}^2(\boldsymbol{\theta})}} - \frac{\alpha(\nu+1)\epsilon_{t-1}^2}{\left\{(\nu-2)+\frac{\epsilon_{t-1}^2}{\sigma_{t-1}^2(\boldsymbol{\theta})}\right\}^2} \end{pmatrix} + b_{t-1}(\boldsymbol{\theta})\frac{\partial \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

with

$$b_t(\boldsymbol{\theta}) = \beta + \frac{\alpha(\nu+1)\epsilon_t^4}{\left\{(\nu-2)\sigma_t^2(\boldsymbol{\theta}) + \epsilon_t^2\right\}^2} = \beta + \frac{\alpha(\nu+1)\eta_t^4(\boldsymbol{\theta})}{\left\{\nu-2+\eta_t^2(\boldsymbol{\theta})\right\}^2} < a_t(\boldsymbol{\theta}) \quad \text{a.s.}$$

In particular, we have

$$\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \beta} \leq \sum_{i=0}^{\infty}\frac{1}{a_{t-i-1}(\boldsymbol{\theta})}\prod_{j=1}^{i}\frac{b_{t-j}(\boldsymbol{\theta})}{a_{t-j}(\boldsymbol{\theta})}.$$

Let $a_t = a_t(\boldsymbol{\theta}_0)$ and $b_t = b_t(\boldsymbol{\theta}_0)$. Note that there exist $0 < \underline{\eta} < \overline{\eta}$ and $\rho < 1$ such that

$$\frac{b_t}{a_t} \leq \rho 1_{\eta_t^2 \in [\underline{\eta}, \overline{\eta}]} + 1_{\eta_t^2 \notin [\underline{\eta}, \overline{\eta}]}.$$

Therefore, letting $\pi = P(\eta_t^2 \in [\underline{\eta}, \overline{\eta}]) \in (0,1)$, we have

$$E\left(\frac{b_t}{a_t}\right)^r \leq \rho^r \pi + 1 - \pi < 1.$$

Moreover $a_t^{-1} < \beta_0^{-1}$. Thus $\frac{\partial \log \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \beta}$ admits moments at any order. The other derivatives can be handled similarly. □

H Proof of Proposition 5.1

We have

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log\left\{\frac{1}{\sigma_t(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)\right\} = -g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \boldsymbol{f}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)$$

and

$$\begin{aligned}
&\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log\left\{\frac{1}{\sigma_t(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)\right\} \\
=& - g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \boldsymbol{g}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&+ g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \left\{\eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&- \frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \left(-\frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right) \\
&+ \left(\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right)'\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \left\{\eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&- \frac{\partial}{\partial \boldsymbol{\theta}}\left\{\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right\}\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&- \left\{\eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} (\boldsymbol{f}'_{\boldsymbol{\theta}})^\top \left(\eta_t(\boldsymbol{\theta})\right) + \boldsymbol{F}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \\
=& - g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \boldsymbol{g}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{\boldsymbol{\theta}}^\top \left(\eta_t(\boldsymbol{\theta})\right) \\
&+ g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \left\{\eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{1}{\sigma_t(\boldsymbol{\theta})} \left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right)\right\} \\
&- \frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \left(\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right) + \left(\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right)'\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&- \frac{1}{\sigma_t(\boldsymbol{\theta})} \left\{\boldsymbol{f}'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (\boldsymbol{f}'_{\boldsymbol{\theta}})^\top \left(\eta_t(\boldsymbol{\theta})\right)\right\} + \boldsymbol{F}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right)
\end{aligned}$$

recalling that $\boldsymbol{f}'_{\boldsymbol{\theta}}(y)$ denotes the vector of the derivatives of the elements of $\boldsymbol{f}_{\boldsymbol{\theta}}(y)$ and that $y\boldsymbol{f}'_{\boldsymbol{\theta}}(y) = \boldsymbol{g}_{\boldsymbol{\theta}}(y)$. A Taylor expansion of $\boldsymbol{\theta}_n \mapsto \Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ thus yields $\Lambda_n(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \boldsymbol{\tau}^\top \boldsymbol{\Delta}_n - \frac{1}{2}\boldsymbol{\tau}^\top \mathfrak{I}_n(\boldsymbol{\theta}_n^*)\boldsymbol{\tau}$, where $\boldsymbol{\theta}_n^*$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_n$, and

$$\begin{aligned}
\mathfrak{I}_n(\boldsymbol{\theta}) =& \frac{1}{n}\sum_{t=1}^n g_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial^2 \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{1}{n}\sum_{t=1}^n g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \eta_t(\boldsymbol{\theta}) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&+ \frac{1}{n}\sum_{t=1}^n \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{\boldsymbol{\theta}}^\top\left(\eta_t(\boldsymbol{\theta})\right) + \frac{1}{n}\sum_{t=1}^n \boldsymbol{g}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - \frac{1}{n}\sum_{t=1}^n \boldsymbol{F}_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \\
&- \frac{1}{n}\sum_{t=1}^n g'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{1}{\sigma_t(\boldsymbol{\theta})} \left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \log \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right) \\
&+ \frac{1}{n}\sum_{t=1}^n \frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t(\boldsymbol{\theta})} \left(\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) - \left(\frac{f'_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right)' \left\{\eta_t(\boldsymbol{\theta})\right\} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\
&+ \frac{1}{\sigma_t(\boldsymbol{\theta})} \left\{ \boldsymbol{f}'_{\boldsymbol{\theta}}\left(\eta_t(\boldsymbol{\theta})\right) \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (\boldsymbol{f}'_{\boldsymbol{\theta}})^\top \left(\eta_t(\boldsymbol{\theta})\right) \right\}.
\end{aligned}$$

Under $\mathbf{B1}(\boldsymbol{\theta}_0)$, **A3** and **B2**-**B3**, $\left\{ (g_{\boldsymbol{\theta}_0}(\eta_t) \frac{\partial \log \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}, \frac{f'_{\boldsymbol{\theta}_0}}{f_{\boldsymbol{\theta}_0}}(\eta_t) \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}, \boldsymbol{f}_{\boldsymbol{\theta}_0}^\top(\eta_t))^\top, \mathcal{F}_t \right\}$ is a square integrable martingale difference. By the central limit theorem of Billingsley (1961) we have $\boldsymbol{\Delta}_n \xrightarrow{d} \mathcal{N}\left\{\mathbf{0}, \mathfrak{I}\right\}$ under P_0 as $n \to \infty$. The ergodic theorem entails that $\mathfrak{I}_n(\boldsymbol{\theta}_0) \to \mathfrak{I}$ a.s. as $n \to \infty$. The rest of the proof follows by the arguments given to establish Proposition 2.1. □

I Proof of Proposition 5.2

The proof of Lemma 2.1 can be transposed directly when $\mathbf{B1}(\boldsymbol{\theta}_0)$ and $\mathbf{B2}^*$ hold (instead of $\mathbf{A1}(\boldsymbol{\theta}_0)$ and $\mathbf{A2}^*$). We thus have that

$$E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)|\mathcal{F}_{t-1}) = \mathbf{0} \quad \text{and} \quad \mathfrak{I}_t := E(\boldsymbol{s}_{t,\boldsymbol{\theta}_0}(\eta_t)\boldsymbol{s}_{t,\boldsymbol{\theta}_0}^\top(\eta_t)|\mathcal{F}_{t-1}) \ \text{ exists}, \quad a.s. \tag{I.1}$$

The proof of Proposition 2.3 also applies without much difference: defining $W_{t,n}$ as before (but with now $\eta_t(\boldsymbol{\theta}_n) = (y_t - m_t(\boldsymbol{\theta}_n))/\sigma_t(\boldsymbol{\theta}_n))$, the proof relies on establishing (D.1)-(D.4). The proof of (D.1), (D.3) and (D.4) is unchanged, while the proof of (D.2) is straightforwardly adapted using $\mathbf{B2}^*$ instead of $\mathbf{A2}^*$. The conclusion follows. □

References

Berkes, I., Horváth, L. and P.S. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P.(1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Billingsley P. (1999) *Convergence of Probability Measures* 2nd ed. John Wiley, New York.

Blasques, F., Koopman, S.J. and A. Lucas (2014) Maximum likelihood estimation for score-driven models. TI 2014-029/III Tinbergen Institute Discussion Paper.

Creal, D., Koopman, S. J., and A. Lucas (2008) A general framework for observation driven time-varying parameter models. TI 08-108/4 Tinbergen Institute Discussion Paper.

Creal, D., Koopman, S. J., and A. Lucas (2013) Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795.

Drost, F. C. and C. A. J. Klaassen (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193–221.

Drost, F.C., Klaassen, C.A.J. and B.J.M. Werker (1997) Adaptive estimation in time-series models. *Annals of Statistics* 25, 786–817.

Francq, C. and J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Garel, B. and M. Hallin (1995) Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics* 47, 551–579.

Hallin, M., Taniguchi, M., Serroukh, A., and K. Choy (1999) Local asymptotic normality for regression models with long-memory disturbance. *The Annals of Statistics* 27, 2054–2080.

Hamadeh, T., and J.M. Zakoïan (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *Journal of Statistical Planning and Inference*, 141, 488–507.

Harvey, A. (2013) *Dynamic models for volatility and heavy tails.* Cambridge University Press.

Harvey, A. C., and T. Chakravarty (2008) Beta-t-(E)GARCH. University of Cambridge, Faculty of Economics, Working Paper CWPE 08340.

Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

Koul, H. L. and A. Schick (1996) Adaptive estimation in a random coefficient autoregressive model. *Annals of Statistics* 24, 1025–1052.

Kreiss, J.P. (1987) On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15, 112–133.

Lee, S. and M. Taniguchi (2005) Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. *Statistica Sinica* 15, 215–234.

Lehmann, E.L. and J.P. Romano (2006) *Testing statistical hypotheses*, 3rd edition, Springer-Verlag: New-York.

Lehmann, E. L. and J. P. Romano (2006) *Testing statistical hypotheses.* Third Edition, Springer.

Liese, F. and K. J. Miescke (2008) *Statistical decision theory.* Springer, New York.

Lind, B. and G. Roussas (1972) A remark on quadratic mean differentiability. *The Annals of Mathematical Statistics* 43, 1030–1034.

Ling, S. and M. McAleer (2003) On adaptative estimation in nonstationary ARMA models with GARCH errors. *The Annals of Statistics* 31, 642–674.

Linton, O. (1993) Adaptive estimation in ARCH models. *Econometric Theory* 9, 539–569.

Robinson, P.M., and P. Zaffaroni (2006) Pseudo-maximum likelihood estimation of ARCH(1) models. *The Annals of Statistic* 34, 1049–1074.

Royer, J. (2022) Conditional asymmetry in Power ARCH(∞) models. *Journal of Econometrics*, in press, https://doi.org/10.1016/j.jeconom.2021.10.013

Straumann, D., and T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449–2495.

Swensen, A. R. (1985) The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* 16, 54–70.

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

Zhu, D., and J.W. Galbraith (2010) A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics* 157, 297–305.

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

The Center for Research in Economics and Statistics (CREST) is a leading French scientific institution for advanced research on quantitative methods applied to the social sciences.

CREST is a joint interdisciplinary unit of research and faculty members of CNRS, ENSAE Paris, ENSAI and the Economics Department of Ecole Polytechnique. Its activities are located physically in the ENSAE Paris building on the Palaiseau campus of Institut Polytechnique de Paris and secondarily on the Ker-Lann campus of ENSAI Rennes.