Characterizations of multinormality and corresponding tests of fit, including for GARCH models

Norbert Henze[1], M.D. Jiménez–Gamero[2], Simos G. Meintanis[3,4]

[1]*Institute of Stochastics, Karlsruhe Institute of Technology, Karlsruhe, Germany*
[2]*Department of Statistics and Operations Research, University of Seville, Seville, Spain*
[3]*Department of Economics, National and Kapodistrian University of Athens, Athens, Greece*
[4]*Unit for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa*[1]

Abstract. We provide novel characterizations of multivariate normality that incorporate both the characteristic function and the moment generating function, and we employ these results to construct a class of affine invariant, consistent and easy-to-use goodness-of-fit tests for normality. The test statistics are suitably weighted L^2-statistics, and we provide their asymptotic behavior both for i.i.d. observations as well as in the context of testing that the innovation distribution of a multivariate GARCH model is Gaussian. We also study the finite-sample behavior of the new tests and compare the new criteria with alternative existing tests.

Keywords. Characteristic function; Moment generating function; Goodness-of-fit test; multivariate normality; Gaussian GARCH model

AMS 2000 classification numbers: 62H15, 62G20.

---

[1]On sabbatical leave from the University of Athens

1 Introduction

Let X be a (univariate) random variable with characteristic function (CF) $\varphi_X(t) = \mathbb{E}[\exp(\mathrm{i}tX)]$, $t \in \mathbb{R}$. Assuming that the moment generating function (MGF) $M_X(t) = \mathbb{E}[\exp(tX)]$ of X exists for each $t \in \mathbb{R}$, [59] proved that the identity

$$(1.1) \quad \varphi_X(t)M_X(t) - 1 = 0 \quad \text{for each } t \in \mathbb{R}$$

characterizes the zero-mean Gaussian distribution. Notice that, by assuming the existence of the MGF in an interval around zero, the moments of any order are implicitly supposed to exist. A further point is that a CF satisfying (1.1) is necessarily real-valued. Hence, the symmetry of the distribution of X around zero is also implicit in (1.1), see [44], §3.1.

In the following we extend the characterization in (1.1) to non-centered random variables and to random vectors of arbitrary dimension. We then construct a goodness-of-fit test for multivariate normality based on a suitable weighted L^2-statistic in which both the CF and the MGF are estimated nonparametrically. Several variants of this test criterion are also suggested. Furthermore, we work out the asymptotics of our test statistics in the context of independent and identically distributed (i.i.d.) random vectors, thus obtaining a new test for multivariate normality, but also in the context of GARCH-type dependence. The latter test provides a novel method for assessing the celebrated question whether a Gaussian GARCH driven volatility process is adequate for explaining the heavy tails that are often observed with financial time series.

The rest of the paper unfolds as follows. In Section 2 we prove the above-mentioned extension of (1.1). On the basis of the resulting characterization, Section 3 suggests two new classes of affine invariant and easily computable statistics for testing for multivariate normality. Section 4 shows that a 'certain limit statistic' of one of these classes is a linear combination of two well-known measures of multivariate skewness, while the other is related to Mardia's multivariate sample kurtosis. In Section 5 we derive the limit null distribution of the new test statistics in the i.i.d. setting. Section 6 addresses the question of consistency of the new tests against general alternatives. Section 7 considers this criterion in the context of multivariate GARCH models in order to test for normality of innovations, and it provides the pertaining large sample theory. Sec-

tion 8 presents a Monte Carlo study that compares the new tests with competing ones. The article concludes with discussions in Section 9. The majority of proofs has been postponed to Section 10.

Throughout the paper, the letter d stands for dimension, and both random and nonrandom vectors are understood as column vectors; the transpose of a vector x will be denoted by $x^\top$; all random elements are supposed to be defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$; for any matrix $A = (a_{kj})$, we will use the norm defined by $\|A\| = \sum_{k,j} |a_{kj}|$; if A is a vector, $\|A\|$ denotes the Euclidean norm. If A is a square matrix, $\|A\|_2$ stands for the spectral norm of A. The unit matrix of order d will be denoted by I_d. Finally, $\mathrm{Re}(z) = a$ is the real part of a complex number $z = a + ib$.

2 Characterizations of multinormality

Let X be a d-variate non-degenerate random vector with CF $\varphi_X(t) = \mathbb{E}[\exp(it^\top X)]$ and MGF $M_X(t) = \mathbb{E}[\exp(t^\top X)] < \infty, t \in \mathbb{R}^d$. We then have the following characterization.

Proposition 2.1 *The identity*

$$(2.1) \quad \varphi_X(t)\, M_X(t) - e^{(i+1)t^\top \mu} = 0 \quad \text{for each } t \in \mathbb{R}^d$$

holds true for some $\mu \in \mathbb{R}^d$ if, and only if, X follows some normal distribution with mean μ.

PROOF. The "if" part is trivial. To prove the converse implication, suppose that (2.1) holds. Fix $a \in \mathbb{R}^d$ and put $Y_a = a^\top(X - \mu)$. We have $\varphi_{Y_a}(t) = e^{-ita^\top \mu}\, \varphi_X(ta)$, $M_{Y_a}(t) = e^{-ta^\top \mu}\, M_X(ta)$. Hence, (2.1) implies $\varphi_{Y_a}(t)\, M_{Y_a}(t) - 1 = 0$ for each $t \in \mathbb{R}$. In view of (1.1), it follows that Y_a has some zero-mean normal distribution. Moreover, the variance of Y_a is equal to $a^\top \Sigma a$, where Σ denotes the covariance matrix of X. Hence, $a^\top X = Y_a + a^\top \mu$ has a normal distribution with mean $a^\top \mu$ and variance $a^\top \Sigma a$. Since a was arbitrary, a well-known characterization of the multivariate normal distribution (see for instance [52], §8a.1–§8a.2) yields the assertion. ∎

The following result will be used in the construction of the test statistics and in order to prove consistency of one of our tests.

Corollary 2.2 *Assume that $\mathbb{E}X = 0$ and put $R_X(t) := Re\,\varphi_X(t)$. Then*

$$(2.2) \quad R_X(t)\,M_X(t) - 1 \;=\; 0 \quad \textit{for each } t \in \mathbb{R}^d$$

holds true if, and only if, X follows a zero-mean normal distribution.

PROOF. Since the "if" part is obviously true, suppose that (2.2) holds. Replace t with $-t$ to get

$$(2.3) \quad R_X(t)\,M_X(t) = 1 \;=\; R_X(-t)\,M_X(-t) \quad \text{for each } t \in \mathbb{R}^d.$$

However, $R_X(t) = R_X(-t)$, and since, by (2.2), $R_X(t) \neq 0$, (2.3) yields

$$(2.4) \quad M_X(t) = M_X(-t) \quad \text{for each } t \in \mathbb{R}^d.$$

Noticing that $M_X(-t) = M_{-X}(t)$, (2.4) gives $M_X = M_{-X}$ which, by uniqueness of the MGF, shows that the law of X is symmetric around zero, since X and $-X$ have the same distribution. Hence $R_X = \varphi_X$, and Proposition 2.1 completes the proof. ∎

3   New tests for multivariate normality

In this section, we assume that $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. copies of a d-dimensional random vector X, the distribution of which is assumed to be absolutely continuous with respect to d-dimensional Lebesgue measure. Writing $N_d(\mu, \Sigma)$ for the d-dimensional normal distribution with mean vector μ and non-degenerate covariance matrix Σ and $\mathcal{N}_d$ for the class of all non-degenerate d-variate normal distributions, a classical problem is to test the null hypothesis

$$H_0 : \text{ The law of } X \text{ belongs to } \mathcal{N}_d,$$

against general alternatives. Since the class $\mathcal{N}_d$ is closed with respect to full rank affine transformations, any genuine test statistic $T_n = T_n(X_1, \ldots, X_n)$ based on $X_1, \ldots, X_n$ should also be invariant with respect to such transformations, i.e., we should have

$$T_n(AX_1 + b, \ldots, AX_n + b) \;=\; T_n(X_1, \ldots, X_n)$$

for each nonsingular $d \times d$-matrix A and each $b \in \mathbb{R}^d$, see [31] for an account on the importance of affine invariance in connection with testing for multivariate normality. Writing $\overline{X}_n = n^{-1} \sum_{j=1}^n X_j$ for the sample mean and $S_n = n^{-1} \sum_{j=1}^n (X_j - \overline{X}_n)(X_j - \overline{X}_n)^\top$ for the sample covariance matrix of $X_1, \ldots, X_n$, a necessary and sufficient condition for a test statistic T_n to be affine invariant is that it is based on the Mahanalobis angles and distances

$$(3.1) \quad Y_{n,i}^\top Y_{n,j} = (X_i - \overline{X}_n)^\top S_n^{-1} (X_j - \overline{X}_n), \qquad i, j \in \{1, \ldots, n\},$$

where

$$Y_{n,j} = S_n^{-1/2}(X_j - \overline{X}_n), \qquad j = 1, \ldots, n,$$

are the so-called *scaled residuals* of $X_1, \ldots, X_n$, see [31]. Here, $S_n^{-1/2}$ denotes the unique symmetric square root of S_n which, due to the absolute continuity of the distribution of X, exists with probability one if $n \geq d + 1$, see [16]. The latter condition is tacitly assumed to hold in what follows.

Recall that in view of Corollary 2.2, we have $R_X(t) M_X(t) - 1 = 0$ for each $t \in \mathbb{R}^d$ if, and only if, the distribution of X is centered normal. Since the scaled residuals $Y_{n,1}, \ldots, Y_{n,n}$ provide an empirical standardization of $X_1, \ldots, X_n$, it seems tempting to introduce the *empirical cosine transform*

$$(3.2) \quad R_n(t) := \frac{1}{n} \sum_{j=1}^n \cos\left(t^\top Y_{n,j}\right), \quad t \in \mathbb{R}^d,$$

and the *empirical moment generating function*

$$(3.3) \quad M_n(t) := \frac{1}{n} \sum_{j=1}^n \exp\left(t^\top Y_{n,j}\right), \quad t \in \mathbb{R}^d,$$

of $Y_{n,1}, \ldots, Y_{n,n}$ and to base a test of H_0 on the weighted L^2-statistic

$$(3.4) \quad T_{n,\gamma} := \int_{\mathbb{R}^d} U_n^2(t) \, w_\gamma(t) \, \mathrm{d}t,$$

where

$$(3.5) \quad U_n(t) := \sqrt{n} \left(R_n(t) M_n(t) - 1\right), \qquad t \in \mathbb{R}^d,$$

5

$$(3.6) \quad w_\gamma(t) \;=\; \exp\left(-\gamma \|t\|^2\right),$$

and $\gamma > 0$ is some positive parameter. In principle, we could replace w_γ in (3.4) with a more general weight function satisfying some general conditions. The above special choice, however, leads to a test criterion with certain extremely appealing features. To this end, putting $Y_{jk}^{\pm} = Y_{n,j} \pm Y_{n,k}$, routine calculations give the representation

$$(3.7) \quad T_{n,\gamma} \;=\; \left(\frac{\pi}{\gamma}\right)^{d/2} \left\{ \frac{1}{2n^3} \sum_{j,k,\ell,m=1}^{n} \left[\exp\left(\frac{\|Y_{jk}^+\|^2 - \|Y_{\ell m}^-\|^2}{4\gamma}\right) \cos\left(\frac{Y_{jk}^{+\top} Y_{\ell m}^-}{2\gamma}\right) \right. \right.$$
$$\left. + \; \exp\left(\frac{\|Y_{jk}^+\|^2 - \|Y_{\ell m}^+\|^2}{4\gamma}\right) \cos\left(\frac{Y_{jk}^{+\top} Y_{\ell m}^+}{2\gamma}\right) \right]$$
$$\left. -\frac{2}{n} \sum_{j,k=1}^{n} \exp\left(\frac{\|Y_{n,j}\|^2 - \|Y_{n,k}\|^2}{4\gamma}\right) \cos\left(\frac{Y_{n,j}^\top Y_{n,k}}{2\gamma}\right) + n \right\},$$

which is amenable to computational purposes. Being a function of $Y_{n,i}^\top Y_{n,j}$ figuring in (3.1), the statistic $T_{n,\gamma}$ is affine invariant. Rejection of H_0 is for large values of $T_{n,\gamma}$.

In view of the fact that $T_{n,\gamma}$ contains a fourfold sum which implies $O(n^4)$ additions to compute $T_{n,\gamma}$, we also studied another statistic, which is

$$(3.8) \quad \widetilde{T}_{n,\gamma} \;:=\; \int_{\mathbb{R}^d} U_n(t)\, w_\gamma(t)\, \mathrm{d}t.$$

A simple calculation shows that $\widetilde{T}_{n,\gamma}$ takes the form

$$(3.9) \quad \widetilde{T}_{n,\gamma} \;=\; \left(\frac{\pi}{\gamma}\right)^{d/2} \sqrt{n} \left(\frac{1}{n^2} \sum_{j,k=1}^{n} \exp\left(\frac{\|Y_{n,j}\|^2 - \|Y_{n,k}\|^2}{4\gamma}\right) \cos\left(\frac{Y_{n,j}^\top Y_{n,k}}{2\gamma}\right) - 1 \right)$$

and hence is much faster to compute than $T_{n,\gamma}$.

We close this section by noting that tests for normality based on the empirical CF date back to [19]. For multivariate normality the first contribution is by [11], with later contributions by [4], [34], [33], [18], [51], [1], and [55]. For review material on the empirical CF we refer to [58]. On the other hand, the approach based on the empirical MGF is certainly less popular, and it appears to include only a few entries. In the multivariate case we refer to the test of [13] for bivariate exponentiality testing, the extension of [21] of the T_3-plot of [27] for testing normality and to [48] for testing skew-normality. In the univariate case we refer to [47] and [60], with the lacking theory

of the latter test provided by [32]. Finally, for a recent general account of weighted L^2-statistics such as ours we refer to [5].

4   The case $\gamma \to \infty$

In this section, we show that the statistic $T_{n,\gamma}$, after a suitable scaling, approaches a linear combination of two well-known measures of multivariate skewness as $\gamma \to \infty$. Likewise, $\widetilde{T}_{n,\gamma}$ is connected with a time-honored measure of multivariate kurtosis.

Theorem 4.1 *We have*

$$\lim_{\gamma \to \infty} \gamma^{3+d/2} \frac{96 T_{n,\gamma}}{n\pi^{d/2}} = 2b_{1,d} + 3\widetilde{b}_{1,d},$$

where

$$b_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} \left(Y_{n,j}^\top Y_{n,k} \right)^3, \qquad \widetilde{b}_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} Y_{n,j}^\top Y_{n,k} \, \|Y_{n,j}\|^2 \, \|Y_{n,k}\|^2,$$

are multivariate sample skewness in the sense of [46] and [49], respectively.

PROOF. From (3.7) and

$$\exp(y) = 1 + y + \frac{y^2}{2} + \frac{y^3}{6} + O(y^4), \qquad \cos(y) = 1 - \frac{y^2}{2} + O(y^4)$$

as $y \to 0$, the result follows by tedious but straightforward calculations, using the relations $\sum_{j=1}^{n} Y_{n,j} = 0$, $\sum_{j=1}^{n} \|Y_{n,j}\|^2 = nd$, $\sum_{j,k=1}^{n} \|Y_{n,j} + Y_{n,k}\|^2 = 2n^2 d$,

$$\sum_{j,k=1}^{n} \|Y_{n,j} + Y_{n,k}\|^4 = 2n^2 \left(\frac{1}{n} \sum_{j=1}^{n} \|Y_{n,j}\|^4 + d^2 + 2d \right),$$

$$\sum_{j,k=1}^{n} \|Y_{n,j} + Y_{n,k}\|^4 Y_{n,j}^\top Y_{n,k} = 8 \sum_{j,k=1}^{n} \left(Y_{n,j}^\top Y_{n,k} \right)^2 \|Y_{n,j}\|^2 + 4n^2 b_{1,d} + 2n^2 \widetilde{b}_{1,d}$$

as well as (writing $\mathrm{tr}(D)$ for the trace of a square matrix D)

$$
\begin{aligned}
\sum_{j,k=1}^{n} \left(Y_{n,j}^{\top} Y_{n,k}\right)^2 \|Y_{n,j}\|^2 &= \sum_{j,k=1}^{n} \mathrm{tr}\left(Y_{n,j}^{\top} Y_{n,k} Y_{n,k}^{\top} Y_{n,j} \|Y_{n,j}\|^2\right) \\
&= \mathrm{tr}\left(\sum_{k=1}^{n} Y_{n,k} Y_{n,k}^{\top} \sum_{j=1}^{n} Y_{n,j} Y_{n,j}^{\top} \|Y_{n,j}\|^2\right) \\
&= \mathrm{tr}\left(n I_d \sum_{j=1}^{n} Y_{n,j} Y_{n,j}^{\top} \|Y_{n,j}\|^2\right) \\
&= n \sum_{k=1}^{n} \mathrm{tr}\left(Y_{n,j}^{\top} Y_{n,j} \|Y_{n,j}\|^2\right) \\
&= n \sum_{j=1}^{n} \|Y_{n,j}\|^4. \quad\blacksquare
\end{aligned}
$$

Remark 4.2 *Theorem 4.1 parallels Theorem 2.1 of [30], who showed that the BHEP statistic for testing for multivariate normality, after suitable rescaling, approaches $2b_{1,d} + 3\widetilde{b}_{1,d}$ as a smoothing parameter (called β in that paper) tends to 0. Since β and γ are related by $\beta = \gamma^{-1/2}$, this corresponds to letting γ tend to infinity.*

Theorem 4.3 *We have*

$$
\lim_{\gamma \to \infty} \gamma^{2+d/2} \frac{16\widetilde{T}_{n,\gamma}}{\sqrt{n}\pi^{d/2}} = b_{2,d} - d(d+2),
$$

where

$$
b_{2,d} = \frac{1}{n} \sum_{j=1}^{n} \|Y_{n,j}\|^4
$$

is multivariate sample kurtosis in the sense of [46].

PROOF. From (3.9) we have

$$
\begin{aligned}
\gamma^{d/2} \frac{\widetilde{T}_{n,\gamma}}{\sqrt{n}\pi^{d/2}} = \frac{1}{n^2} \sum_{i,j=1}^{n} \Bigg\{ &\left(1 + \frac{\|Y_{n,i}\|^2 - \|Y_{n,j}\|^2}{4\gamma} + \frac{(\|Y_{n,i}\|^2 - \|Y_{n,j}\|^2)^2}{32\gamma^2}\right) \\
&\times \left(1 - \frac{(Y_{n,i}^{\top} Y_{n,j})^2}{8\gamma^2}\right) + O\left(\gamma^{-3}\right) \Bigg\} - 1
\end{aligned}
$$

8

as $\gamma \to \infty$. Since $\sum_{i,j}(\|Y_{n,i}\|^2 - \|Y_{n,j}\|^2) = 0$ and $\sum_{i,j}\left(Y_{n,i}^\top Y_{n,j}\right)^2 = n^2 d$, the result follows by straightforward calculations. ∎

Notice that $d(d+2)$ is the value of theoretical multivariate kurtosis in the sense of [46] attained by any non-degenerate d-variate normal distribution. In this sense the limit statistic of Theorem 4.3 maybe termed "excess kurtosis".

5 Asymptotic null distribution in the i.i.d. case

In this section we consider the case that $X_1, X_2, \ldots$ are i.i.d. d-dimensional random vectors with some non-degenerate normal distribution. Notice that the process U_n given in (3.5) is a random element of the separable Hilbert space $L^2 := L^2(\mathbb{R}^d, \mathcal{B}^d, w_\gamma(t)dt)$ of (equivalence classes of) measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ that are square integrable with respect to the finite measure on the σ-field $\mathcal{B}^d$ of Borel sets of $\mathbb{R}^d$ given by the weight function w_γ defined in (3.6). The inner product and the resulting norm in L^2 will be denoted by $\langle f, g \rangle = \int_{\mathbb{R}^d} f(t)\, g(t)\, w_\gamma(t)\, \mathrm{d}t$, $\|f\|_{L^2} = \sqrt{\langle f, f \rangle}$, respectively.

With this notation, the test statistics $T_{n,\gamma}$ and $\widetilde{T}_{n,\gamma}$ given in (3.4) and (3.8), respectively, take the form

$$(5.1) \quad T_{n,\gamma} = \|U_n\|_{L^2}^2$$

and

$$(5.2) \quad \widetilde{T}_{n,\gamma} = \langle U_n, 1 \rangle.$$

Writing "$\xrightarrow{\mathcal{D}}$" for convergence in distribution of random vectors and stochastic processes, the main result of this section is as follows.

Theorem 5.1 (*Convergence of U_n under H_0*)
Suppose that X has some non-degenerate d-variate normal distribution, and that $\gamma > 1$ in (3.6). Then there is a centered Gaussian random element W of L^2 having covariance kernel

$$C(s,t) = \mathrm{e}^{s^\top t} + \frac{1}{2}\left(\mathrm{e}^{s^\top t} + \mathrm{e}^{-s^\top t}\right) + 2\cos\left(s^\top t\right) - s^\top t - 4$$

so that $U_n \xrightarrow{\mathcal{D}} W$ as $n \to \infty$.

9

In view of (5.1) and (5.2), Theorem 5.1 and the Continuous Mapping Theorem immediately yield the following corollary.

Corollary 5.2 *Under the assumptions of Theorem 5.1, we have*

a) $T_{n,\gamma} \xrightarrow{\mathcal{D}} \|W\|_{L^2}^2$ *as* $n \to \infty$,

b) $\widetilde{T}_{n,\gamma} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$ *as* $n \to \infty$, *with* $\sigma^2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C(s,t) \mathrm{e}^{-\gamma(\|s\|^2 + \|t\|^2)} \, ds \, dt$.

Remark 5.3 *Routine calculations show that σ^2 in Corollary 5.2 (b) has the following expression:* $\sigma^2 = 2\pi^d(\gamma^2 - 0.25)^{-d/2} + 2\pi^d(\gamma^2 + 0.25)^{-d/2} - 4\pi^d \gamma^{-d}$.

Remark 5.4 *It is well-known that the distribution of $\|W\|_{L^2}^2$ is that of $\sum_{j=1}^{\infty} \lambda_j N_j^2$, where $\lambda_1, \lambda_2, \ldots$ are the positive eigenvalues of the integral operator $f \mapsto Af$ on L^2 associated with the kernel C given in Theorem 5.1, i.e., $(Af)(t) = \int C(s,t) f(s) \exp(-\gamma\|t\|^2) ds$, and $N_1, N_2, \ldots$ are i.i.d. standard normal random variables. We did not succeed in obtaining explicit solutions of this equation. However, since*

$$\mathbb{E}\|W\|_{L^2}^2 = \int_{\mathbb{R}^d} C(t,t) \exp\left(-\gamma\|t\|^2\right) \, dt$$

(see [53], p. 213), tedious but straighforward manipulations of integrals yield

$$
\begin{aligned}
\mathbb{E}\|W\|_{L^2}^2 \;=\; & \frac{3}{2}\left(\frac{\pi}{\gamma-1}\right)^{d/2} + \frac{1}{2}\left(\frac{\pi}{\gamma+1}\right)^{d/2} - \left(4 + \frac{d}{2\gamma}\right)\left(\frac{\pi}{\gamma}\right)^{d/2} \\
& +2 \left[\sqrt{\frac{\pi}{\gamma}} \frac{\cos\left(\frac{\tan^{-1}(1/\gamma)}{2}\right)}{(1 + (1/\gamma^2))^{1/4}}\right]^d \sum_{q=0}^{\lfloor \frac{d}{2} \rfloor} (-1)^q \binom{d}{2q} \left(\tan\left(\frac{\tan^{-1}(1/\gamma)}{2}\right)\right)^{2q},
\end{aligned}
$$

(5.3)

where $\lfloor \cdot \rfloor$ denotes the integer part of a given number.

6 Consistency

The next result shows that the test for multivariate normality based on $T_{n,\gamma}$ is consistent against general alternatives.

Theorem 6.1 *Suppose X has some absolutely continuous distribution, and that $M_X(t) = \mathbb{E}[\exp(t^\top X)] < \infty$, $t \in \mathbb{R}^d$. Then we have*

$$\liminf_{n \to \infty} \frac{T_{n,\gamma}}{n} \;\geq\; \int_{\mathbb{R}^d} \left(R_X(t) M_X(t) - 1\right)^2 w_\gamma(t) \, dt$$

almost surely.

10

PROOF. Because of affine invariance we may w.l.o.g. assume $\mathbb{E}X = 0$ and $\mathbb{E}XX^\top = \mathrm{I}_d$. With $\Delta_{n,j}$ given in (10.3), notice that

$$\xi_n := \max_{j=1,\dots,n} \|\Delta_{n,j}\| \leq \|S_n^{-1/2} - \mathrm{I}_d\|_2 \cdot \max_{j=1,\dots,n} \|X_j\| + \|S_n^{-1/2}\|_2 \|\overline{X}_n\|$$

which, by Theorem 5.2. of [6], implies

(6.1) $\lim_{n\to\infty} \xi_n = 0$ $\mathbb{P}$-a.s.

Fix $K > 0$, and recall $R_n(t)$, $M_n(t)$, from (3.2), (3.3), respectively. With $R_n^\circ(t)$, $M_n^\circ(t)$ given in (10.2), (10.4) yields

$$\max_{\|t\|\leq K} \left| M_n(t) - M_n^\circ(t) \right| \leq \max_{\|t\|\leq K} M_n(t) \left(K\xi_n + \frac{1}{2}K^2\xi_n^2 e^{K\xi_n} \right).$$

By (6.1) and the strong law of large numbers in the Banach space of continuous functions on $\{t \in \mathbb{R}^d : \|t\| \leq K\}$, we have $\lim_{n\to\infty} \max_{\|t\|\leq K} \left| M_n(t) - M_n^\circ(t) \right| = 0$ $\mathbb{P}$-a.s. Since $\max_{\|t\|\leq K} |R_n(t) - R_n^\circ(t)| \leq K\xi_n$, (6.1) implies $\lim_{n\to\infty} \max_{\|t\|\leq K} \left| R_n(t) - R_n^\circ(t) \right| = 0$ $\mathbb{P}$-a.s. In view of

$$
\begin{aligned}
R_n(t)M_n(t) - 1 &= (R_n(t) - R_n^\circ(t))(M_n(t) - M_n^\circ(t)) + M_n^\circ(t)(R_n(t) - R_n^\circ(t)) \\
&\quad + R_n^\circ(t)(M_n(t) - M_n^\circ(t)) + R_n^\circ(t)M_n^\circ(t) - 1,
\end{aligned}
$$

the strong law of large numbers in Banach spaces and Fatou's lemma yield

$$\liminf_{n\to\infty} \frac{T_{n,\gamma}}{n} \geq \int_{\{t:\|t\|\leq K\}} (R_X(t)M_X(t) - 1)^2 \, w_\gamma(t) \, \mathrm{d}t$$

almost surely. Since K is arbitrary, the assertion follows. ∎

In view of Corollary 2.2 and Theorem 6.1, the test that rejects the null hypothesis H_0 for large values of $T_{n,\gamma}$ is consistent against each alternative distribution the MGF of which exists on $\mathbb{R}^d$. We conjecture that this test is consistent against *any* alternative distribution. However, in view of the reasoning of [12], the behavior of $T_{n,\gamma}$ against heavy-tailed alternatives is a non-trivial problem.

7 Testing for normality in GARCH models

Consider the multivariate GARCH (MGARCH) model

(7.1) $X_j = \Sigma_j^{1/2}(\theta)\varepsilon_j, \ j \in \mathbb{Z},$

where $\theta \in \Theta \subseteq \mathbb{R}^v$, is an v-dimensional vector of unknown parameters. The unobservable random errors ε_j, $j \in \mathbb{Z}$, also referred to as *innovations*, are i.i.d. copies of a d-dimensional random vector ε, which is assumed to have mean zero and unit covariance matrix. Moreover, $\Sigma_j(\theta) = \Sigma(X_{j-1}, X_{j-2}, \ldots; \theta)$ is a $d \times d$ symmetric and positive definite matrix, which is the conditional variance matrix of X_j, given the past information. On the basis of the observations X_j, $j = 1, \ldots, n$, driven by equation (7.1), we wish to test the null hypothesis

$$H_{0,G} : \text{ The law of } \varepsilon \text{ is } N_d(0, I_d).$$

Notice that $H_{0,G}$ is equivalent to the hypothesis that, conditionally on $\{X_{j-1}, X_{j-2}, \ldots\}$, the law of X_j is $N_d(0, \Sigma_j(\theta))$, for some $\theta \in \Theta$. The difference with the i.i.d. setting is that, on the one hand, the innovations in (7.1) are already assumed to be centered at zero. On the other hand however, the covariance matrix $\Sigma_j(\theta)$ of X_j, unlike the i.i.d. case, here is allowed to be time-varying in a way that depends on the unknown parameter θ as well as on past observations.

For univariate GARCH models [39], [42], [26] and [14] suggested specification tests for the innovation distribution. However, with the exception of [2], corresponding tests are still scarce in the multivariate case. We now take some time to emphasize the importance of testing the null hypothesis of normality in GARCH models. First, notice that acceptance of the null hypothesis $H_{0,G}$ implies the validity of the classical Gaussian MGARCH model, which has been a benchmark for modelling certain economic and financial quantities. Although even today the normal distribution is most commonly used in applications, since the time of [45] and [20] there is empirical evidence that, for example, the distribution of financial variables is heavy-tailed, even after filtering the volatility clustering phenomenon produced by model (7.1). In order to capture this empirical so-called stylized fact, several authors suggested alternative innovation distributions, such as the (multivariate) Student-t distribution ([57], [2], and [15]), the stable distribution ([8], [50], and [23]), and the Laplace distribution ([56], [9]). In this connection, it is well-known that having erroneously accepted the normality assumption for the GARCH-residuals resulting from the estimation of model (7.1), leads to incorrect inferential procedures, such as assessment of standard risk measures like the value at risk (VaR); see for instance [54]. The preceding discussion provides

the ground on the basis of which the null hypothesis $H_{0,G}$ could be considered as highly relevant, particularly in statistical modelling with a view towards financial applications.

Notice that although $H_{0,G}$ is about the distribution of ε, the innovations themselves are unobservable in the context of model (7.1). Hence any decision regarding the null hypothesis $H_{0,G}$ should be based on residuals

$$(7.2) \quad \widetilde{\varepsilon}_j(\widehat{\theta}_n) = \widetilde{\Sigma}_j^{-1/2}(\widehat{\theta}_n)X_j, \ j = 1, \ldots, n.$$

Here, the matrix $\widetilde{\Sigma}_j(\cdot)$, apart from a suitable estimator $\widehat{\theta}_n$ that will be detailed later, also depends on specific initial values $\{\widetilde{X}_j, \ j \leq 0\}$ of X_j which, under certain conditions, are asymptotically irrelevant. Let U_n^G be defined as U_n in (3.5) by replacing $Y_{n,j}$ with $\widetilde{\varepsilon}_j(\widehat{\theta}_n), \ j = 1, \ldots, n$. The value of the test statistics $T_{n,\gamma}^G$ and $\widetilde{T}_{n,\gamma}^G$ are defined as in (3.4) and (3.8), respectively, with U_n changed for U_n^G.

In order to derive the asymptotic null distribution of U_n^G we will make the following assumptions (A.1)–(A.6), which will be commented on at the end of this section for specific instances of MGARCH models. In the sequel, $C > 0$ and ϱ, $0 < \varrho < 1$, denote generic constants, the values of which may vary across the text, and θ_0 stands for the true value of θ.

(A.1) The estimator $\widehat{\theta}_n$ satisfies $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_{\mathbb{P}}(1)$.

(A.2) $\sup_{\theta \in \Theta} \left\| \widetilde{\Sigma}_j^{-1/2}(\theta) \right\| \leq C, \quad \sup_{\theta \in \Theta} \left\| \Sigma_j^{-1/2}(\theta) \right\| \leq C$ a.s.

(A.3) $\sup_{\theta \in \Theta} \| \Sigma_j^{1/2}(\theta) - \widetilde{\Sigma}_j^{1/2}(\theta) \| \leq C\varrho^j$.

(A.4) $\mathbb{E} \left\| X_j \right\|^\varsigma < \infty$ and $\mathbb{E} \left\| \Sigma_j^{1/2}(\theta_0) \right\|^\varsigma < \infty$ for some $\varsigma > 0$.

(A.5) For each sequence $x_1, x_2, \ldots$ of vectors of $\mathbb{R}^d$, the function $\theta \mapsto \Sigma^{1/2}(x_1, x_2, \ldots; \theta)$ admits continuous second-order derivatives.

(A.6) For some neighborhood $V(\theta_0)$ of θ_0, there exist $p > 1$, $q > 2$ and $r > 1$ so that

$2p^{-1} + 2r^{-1} = 1$ and $4q^{-1} + 2r^{-1} = 1$, and

$$\mathbb{E} \sup_{\theta \in V(\Theta)} \left\| \sum_{k,\ell=1}^{v} \Sigma_j^{-1/2}(\theta) \frac{\partial^2 \Sigma_j^{1/2}(\theta)}{\partial \theta_k \partial \theta_\ell} \right\|^p < \infty,$$

$$\mathbb{E} \sup_{\theta \in V(\Theta)} \left\| \sum_{k=1}^{v} \Sigma_j^{-1/2}(\theta) \frac{\partial \Sigma_j^{1/2}(\theta)}{\partial \theta_k} \right\|^q < \infty,$$

$$\mathbb{E} \sup_{\theta \in V(\Theta)} \left\| \Sigma_j^{1/2}(\theta_0) \Sigma_j^{-1/2}(\theta) \right\|^r < \infty.$$

The next result gives the asymptotic null distribution of U_n^G.

Theorem 7.1 (*Convergence of U_n^G under $H_{0,G}$*)
Let $\{X_j\}$ be a strictly stationary process satisfying (7.1), with X_j being measurable with respect to the sigma-field generated by $\{\varepsilon_u, u \leq j\}$. Assume that (A.1)–(A.6) hold. Then under the null hypothesis $H_{0,G}$, there is a centered Gaussian random element W_G of L^2 having covariance kernel

$$C_G(s,t) = e^{s^\top t} + \frac{1}{2} \left(e^{s^\top t} + e^{-s^\top t} \right) + 2\cos\left(s^\top t \right) - 4$$

so that $U_n^G \xrightarrow{\mathcal{D}} W_G$ as $n \to \infty$.

From Theorem 7.1 and the Continuous Mapping Theorem we have the following corollary.

Corollary 7.2 *Under the assumptions of Theorem 7.1, and for $\gamma > 1$ in (3.6) we have*

a) $T_{n,\gamma}^G \xrightarrow{\mathcal{D}} \|W_G\|_{L^2}^2$ *as $n \to \infty$,*

b) $\widetilde{T}_{n,\gamma}^G \xrightarrow{\mathcal{D}} N(0, \sigma_G^2)$ *as $n \to \infty$, where $\sigma_G^2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C_G(s,t) e^{-\gamma(\|s\|^2 + \|t\|^2)} \, ds \, dt$.*

Remark 7.3 *Notice that the covariance kernels in Theorems 5.1 and 7.1 only differ by the term $s^\top t$. This difference is due to the fact that for the MGARCH case the innovations (and thus the data) are assumed to have zero mean. Hence, there is no centering involved. Second, since $\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} s^\top t e^{-\gamma(\|s\|^2 + \|t\|^2)} \, ds \, dt = 0$, it follows that $\sigma^2 = \sigma_G^2$, and thus $\widetilde{T}_{n,\gamma}$ and $\widetilde{T}_{n,\gamma}^G$ both have the same asymptotic null distribution. Perhaps the most surprising fact is that the limit null distribution of U_n^G does not depend on the*

estimation of the parameter θ, in contrast to the GARCH version of the normality test of [33], studied in [39], [37] and [38] for the usual (linear) univariate GARCH model. We underline that these observations refer to the asymptotic null distribution alone and that for finite sample sizes, the null distribution of both $T_{n,\gamma}^G$ and $\widetilde{T}_{n,\gamma}^G$ do depend on the estimation of the parameter θ as well as on the true value of this parameter.

Remark 7.4 *[10], [43], [3] and [25], among others, have shown that mild regularity conditions guarantee that the quasi maximum likelihood estimator (QMLE), defined by*

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta),$$

$$\mathcal{L}_n(\theta) = -\frac{1}{2}\sum_{j=1}^{n} \widetilde{\ell}_j(\theta), \quad \widetilde{\ell}_j(\theta) = X_j^\top \widetilde{\Sigma}_j^{-1}(\theta) X_j + \log\left|\widetilde{\Sigma}_j(\theta)\right|,$$

satisfies (A.1) for general MGARCH, or for particular specifications.

There are many MGARCH parametrizations for the matrix $\Sigma_j(\theta)$, see, e.g., [24]. One of the most widely used MGARCH models is the Constant Conditional Correlation (CCC) model proposed by [7] and extended by [36]. That model decomposes the conditional covariance matrix (7.1) into conditional standard deviations and a conditional correlation matrix, according to $\Sigma_j(\theta_0) = D_j(\theta_0)R_0 D_j(\theta_0)$. Here, $D_j(\theta_0)$ and R_0 are $d \times d$ matrices, with R_0 being a correlation matrix and $D_j(\theta_0)$ is a diagonal matrix so that $\sigma_j(\theta) = diag\left\{D_j^2(\theta)\right\}$ with

$$\sigma_j(\theta) = b + \sum_{k=1}^{p} B_k X_{j-k}^{(2)} + \sum_{k=1}^{q} \Gamma_k \sigma_{j-k}(\theta)$$

and $X_j^{(2)} = X_j \odot X_j$, where $\odot$ denotes the Hadamard product, that is, the element by element product. Moreover, b is a vector of dimension d and has positive elements, while $\{B_k\}_{k=1}^p$ and $\{\Gamma_k\}_{k=1}^q$ are $d \times d$ matrices with non-negative elements. Under certain weak assumptions, the QMLE for the parameters in this model satisfies (A.1), and (A.2)–(A.6) also hold (see [24] and [22]).

Remark 7.5 *Similar results for the consistency against alternatives to those stated in the i.i.d case can be given now. To save space we omit them.*

Table 1: Critical points for $(\gamma/\pi)^{d/2}T_{n,\gamma}$.

			γ						
d	n	α	1.2	1.3	1.4	1.5	2.0	2.5	3.0
2	20	0.05	7.95	5.62	3.94	2.71	0.80	0.31	0.15
		0.10	4.38	2.94	2.08	1.53	0.46	0.19	0.10
	50	0.05	12.62	7.19	4.82	3.74	1.06	0.43	0.21
		0.10	6.39	4.06	2.81	2.00	0.62	0.26	0.13
	100	0.05	14.37	8.11	5.22	3.60	1.04	0.44	0.22
		0.10	6.97	4.31	2.98	2.15	0.67	0.29	0.14
3	20	0.05	22.71	13.73	9.20	6.75	1.84	0.70	0.33
		0.10	13.37	8.44	5.80	4.29	1.22	0.48	0.23
	50	0.05	50.43	24.12	14.41	9.17	2.36	0.93	0.45
		0.10	23.51	12.45	7.92	5.63	1.54	0.63	0.31
	100	0.05	65.97	28.84	16.20	9.81	2.35	0.95	0.46
		0.10	28.58	15.06	9.07	6.02	1.61	0.66	0.33
5	20	0.05	124.67	63.48	37.98	25.23	5.99	2.18	0.98
		0.10	74.76	41.77	26.13	17.49	4.44	1.67	0.77
	50	0.05	670.14	222.95	100.01	54.12	8.46	3.05	1.41
		0.10	264.18	105.49	50.71	29.23	5.72	2.17	1.03
	100	0.05	1501.74	428.77	157.74	73.21	9.13	3.22	1.48
		0.10	514.14	170.50	71.73	36.88	5.98	2.27	1.09

8   Monte Carlo results

This section describes and summarizes the results of some simulation experiments. All computations have been performed using programs written in the R language.

8.1   Numerical experiments for i.i.d. data

Upper quantiles of the null distribution of $T_{n,\gamma}$ have been approximated by generating 10,000 samples from a law $N_d(0, I_d)$. Table 1 displays some critical values. Looking at this table we see that the speed of convergence to the asymptotic null distribution depends on the data dimension and the value of γ.

A natural competitor of the test based on $T_{n,\gamma}$ is the CF-based test studied in [33] (HW-test), since this test is simple to compute as well as affine invariant and has been shown to have good power performance vis–á-vis competitors. Since, according to [35], the global power function of any nonparametric test is flat on balls of alternatives except for alternatives coming from a finite-dimensional subspace, each test has a high power only against a specific set of alternatives. So, we carried out an extensive simulation study to compare their powers against a wide range of alternatives, with the aim of detecting those for which the test based on $T_{n,\gamma}$ is more powerful than the HW-test. As expected from Theorem 4.1 and Remark 4.2, for large γ and small β both tests behave very closely. For non-heavy-tailed distributions, we observed that the power of the proposed test is either similar or a bit less than that of the HW-test; for heavy-tailed distributions, however, the new test outperforms the HW-test. This observation can be appreciated by looking at Table 2, which displays the empirical power calculated by generating 1,000 samples (in each case), for the significance level $\alpha = 0.05$, from the following heavy-tailed alternatives: (LA) the multivariate symmetric Laplace distribution ([40]); (GN_θ) the multivariate β-generalized distribution ([29]), that coincides with the normal distribution for $\theta = 2$ and has heavy tails for $0 < \theta < 2$; (ASE_θ) the θ-stable and elliptically-contoured distribution; (T_θ) multivariate student-t with θ degrees of freedom.

As shown in Corollary 5.2, the test statistic $\widetilde{T}_{n,\gamma}$ is asymptotically normal with zero mean under the null hypothesis. We carried out some simulations to assess the normal approximation to the null distribution of $\widetilde{T}_{n,\gamma}$, and observed that it requires very large values of n, which depend on the value of γ and the data dimension. Table 3 exhibits the empirical power, for significance level $\alpha = 0.05$, against some heavy-tailed and light-tailed distributions. We write PII_a for the Pearson type II distribution with parameter a, and $U(0,1)^d$ for the uniform distribution on the d–dimensional cube. In Table 3 the one-sided test that rejects H_0 for large values of $\widetilde{T}_{n,\gamma}$ is codified as "one" while the two-sided test that rejects H_0 for large values of $|\widetilde{T}_{n,\gamma}|$ is codified as "two" with critical points calculated by simulation. Such results were calculated by generating 1,000 samples of size $n = 100$ in each case. For very heavy-tailed distributions, the new test is clearly more powerful than the HW-test. In these cases, the one-sided

17

Table 2: Percentage of rejection for nominal level $\alpha = 0.05$ and $n = 50$.

| | | Test based on $T_{n,\gamma}$ | | | | | HW-test | | |
| | | γ | | | | | β | | |
	d	1.3	1.4	1.5	2.0	2.5	0.1	0.5	1.0
LA	2	83.6	83.8	82.8	82.0	81.1	59.5	75.3	82.5
	3	91.9	92.8	93.8	93.7	92.8	77.0	89.5	94.2
	5	96.9	97.6	97.9	99.0	98.8	96.0	98.6	99.4
T_5	2	66.7	67.2	65.9	65.5	65.0	52.9	56.2	50.0
	3	75.5	76.6	78.3	78.7	78.0	67.5	69.9	59.6
	5	87.1	88.2	88.6	91.0	91.5	86.0	88.0	78.1
T_{10}	2	31.4	33.7	34.1	32.4	33.2	25.8	23.9	18.5
	3	38.4	40.0	40.0	41.3	41.3	33.7	31.7	22.3
	5	47.9	50.1	50.6	51.6	55.4	51.1	45.3	27.4
GN_1	2	87.7	87.8	85.7	83.5	82.1	59.6	74.5	87.5
	3	91.3	92.6	93.1	92.1	91.4	72.0	88.6	95.9
	5	92.2	93.4	94.7	96.5	96.3	86.4	95.2	97.9
$GN_{1.5}$	2	51.8	52.2	49.4	46.9	45.3	29.0	35.9	49.3
	3	56.2	58.5	60.7	59.0	56.3	38.9	48.5	66.6
	5	50.1	53.4	55.1	61.5	59.9	46.1	55.7	66.5
$ASE_{1.75}$	2	75.6	75.8	75.5	75.4	75.8	68.5	69.2	60.9
	3	82.6	82.6	83.0	83.2	83.6	79.2	78.5	68.2
	5	89.0	89.1	89.4	90.6	90.7	88.1	86.0	74.0
$ASE_{1.85}$	2	56.1	56.1	55.5	56.2	56.3	50.3	48.8	39.5
	3	63.2	63.7	64.4	64.1	63.6	57.5	55.5	42.8
	5	76.0	76.5	76.6	78.2	78.1	72.6	68.1	48.9

test gives slightly better results than the two-sided one. For distribution with very light tails, the one-sided test fails (same behaviour as the "quadratic" statistic), but the two-sided test is more powerful than the HW-test. These numerical results are in agreement with the statement of Theorem 4.3, which asserts that $\widetilde{T}_{n,\gamma}$ is close to a sample kurtosis measure.

8.2 Numerical experiments for GARCH models

Since usual practical applications of MGARCH models involve rather large sample sizes, this subsection studies the finite sample size behavior of the test based on $\widetilde{T}_{n,\gamma}^G$. With this aim, we must first specify a form for $\Sigma_j(\theta)$. In our simulations we considered a bivariate CCC–GARCH(1,1) model with

$$b = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, \ B_1 = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \ \Gamma_1 = \begin{pmatrix} 0.2 & 0.1 \\ 0.01 & 0.3 \end{pmatrix}, \ R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix},$$

Table 3: Percentage of rejection for nominal level $\alpha = 0.05$ and $n = 100$.

| | | | Test based on $\widetilde{T}_{n,\gamma}$ | | | | | HW-test | | |
| | | | γ | | | | | β | | |
	d		1.3	1.4	1.5	2.0	2.5	0.1	0.5	1.0
T_{10}	2	one	57.0	57.0	57.1	57.7	57.9	33.3	34.8	27.1
		two	46.1	46.1	46.2	46.4	46.4			
	3	one	74.9	74.7	74.6	74.4	74.1	44.7	47.4	38.9
		two	64.7	64.7	64.8	65.5	65.3			
	5	one	92.9	92.7	92.8	92.7	92.8	68.4	69.4	54.4
		two	88.0	88.3	88.3	87.9	87.8			
$ASE_{1.95}$	2	one	38.1	38.6	38.8	39.1	39.1	32.1	29.0	21.0
		two	33.9	34.1	34.2	34.4	34.5			
	3	one	45.8	46.1	46.3	46.6	46.5	39.9	35.9	24.3
		two	41.5	41.6	41.6	41.9	41.8			
	5	one	56.1	56.0	55.8	56.2	56.2	50.2	42.6	27.9
		two	52.4	52.4	52.3	52.1	518			
$U(0,1)^d$	2	one	0.0	0.0	0.0	0.0	0.0	2.0	28.3	97.4
		two	99.9	100.0	100.0	100.0	100.0			
	3	one	0.0	0.0	0.0	0.0	0.0	0.0	27.0	98.0
		two	100.0	100.0	100.0	100.0	100.0			
	5	one	0.0	0.0	0.0	0.0	0.0	0.0	22.4	95.8
		two	100.0	100.0	100.0	100.0	100.0			
PII_4	2	one	0.0	0.0	0.0	0.0	0.0	3.0	1.7	11.6
		two	47.8	48.0	48.2	48.1	48.6			
	3	one	0.0	0.0	0.0	0.0	0.0	0.2	1.7	17.9
		two	72.7	73.2	73.9	74.8	75.0			
	5	one	0.0	0.0	0.0	0.0	0.0	0.0	2.7	32.6
		two	93.9	94.5	94.8	95.3	95.6			

for $r = 0, 0.3$, and a trivariate bivariate CCC–GARCH(1,1) model with $b = (0.1, 0.1, 0.1)'$,

$$B_1 = \begin{pmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix}, \quad \Gamma_1 = \begin{pmatrix} 0.2 & 0.1 & 0.01 \\ 0.01 & 0.3 & 0.1 \\ 0.01 & 0.01 & 0.1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}$$

and r as before.

   As shown in Corollary 7.2, the test statistic $\widetilde{T}_{n,\gamma}^G$ is asymptotically normal with zero mean under the null hypothesis. We have computed the actual level for the one-sided test that rejects $H_{0,G}$ if $\widetilde{T}_{n,\gamma}^G/\sigma > u_{1-\alpha}$ and for the two-sided test rejecting $H_{0,G}$ when $|\widetilde{T}_{n,\gamma}^G/\sigma| > u_{1-\alpha/2}$ by generating 5,000 samples from the above bivariate

CCC–GARCH(1,1) model. The results are not reported in order to save space. We nevertheless have the following observations: The quality of the normal approximation strongly depends on the model parameter values, being conservative for $r = 0.0$ and a bit liberal for $r = 0.3$. As in the i.i.d.-case, rather large sample sizes are required for the normal approximation to work. In addition, the normal approximation gives actual levels closer to the nominal ones for the two-sided test. Nevertheless, as discussed before, there is empirical evidence that in typical applications the innovations have heavy-tailed distributions, and we have learnt from the i.i.d. setting that in such a case, the one-sided test is more powerful than the two-sided one. So, for moderate sample sizes, we should resort to another null distribution approximation.

In view of the remarks at the end of Remark 7.3, and in contrast to the i.i.d.-case, we cannot calculate critical points for each n by simulation since, in practice, the values of the true parameters involved in the specification of the conditional covariance matrix $\Sigma_j(\theta)$ are unknown. So, to approximate the null distribution of this test statistic we considered the following conditional resampling scheme, given the data $X_1, \ldots, X_n$:

(i) Calculate $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$, the residuals $\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n$ and the test statistic $\widetilde{T}_{n,\gamma}^G = \widehat{T}_{n,\gamma}^G(\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n)$.

(ii) Generate vectors $\varepsilon_1^*, \ldots, \varepsilon_n^*$ i.i.d. from a $N_d(0, I_d)$ distribution. Let $X_j^* = \Sigma_j^{1/2}(\widehat{\theta})\varepsilon_j^*$, $j = 1, \ldots, n$.

(iii) Calculate $\widehat{\theta}_n^* = \widehat{\theta}_n(X_1^*, \ldots, X_n^*)$, the residuals $\widetilde{\varepsilon}_1^*, \ldots, \widetilde{\varepsilon}_n^*$, and approximate the null distribution of $\widetilde{T}_{n,\gamma}^G$ by means of the conditional distribution, given the data, of $\widetilde{T}_{n,\gamma}^{G*} = \widetilde{T}_{n,\gamma}^G(\widetilde{\varepsilon}_1^*, \ldots, \widetilde{\varepsilon}_n^*)$.

In practice, the approximation in step (iii) is carried out by generating a large number of bootstrap replications of the test statistic $\widetilde{T}_{n,\gamma}^G$, for $b = 1, \ldots, M$, whose empirical distribution function is used to estimate the null distribution of $\widetilde{T}_{n,\gamma}^G$.

In our simulation study, for the distribution of the innovations, we took $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. from the distribution of ε with ε having a (N) multivariate normal distribution, in order to study the level of the resulting test; and to study the power we considered the following heavy-tailed distributions: T_θ and GN_θ (see §8.1), and the asymmetric exponential power distribution (AEP), whereby $(Z_1, \ldots, Z_d)^\top$, with $Z_1, \ldots, Z_d$ i.i.d.

from a univariate AEP distribution ([61]) with parameters $\alpha = 0.4$, $p_1 = 1.182$ and $p_2 = 1.820$; these settings gave useful results in practical applications for the errors in GARCH type models. As in the previous subsection, we also calculated the HW-test for several values of β with Y_j replaced by $\widetilde{\varepsilon}_j$.

The parameters in the CCC-GARCH models were estimated by QMLE using the package `ccgarch` of the language R. Table 4 reports the percentages of rejections for nominal significance level $\alpha = 0.05$ and sample size $n = 300$, for $r = 0$ and $r = 0.3$. In order to reduce the computational burden we adopted the warp-speed method of [28] for evaluating the above resampling scheme. With the warp-speed method, rather than computing critical points for each Monte Carlo sample, one resample is generated for each Monte Carlo sample, and the resampling test statistic is computed for that sample. Then the resampling critical values for $\widetilde{T}_{n,\gamma}^G$ are computed from the empirical distribution determined by the resampling replications of $\widetilde{T}_{n,\gamma}^{G*}$. In our simulations we took $10,000$ Monte Carlo samples for the level and $2,000$ for the power. Looking at Table 4, we conclude that the proposed test in most cases outperforms the HW-test by a wide margin.

9 Conclusion

We prove new characterizations of the multivariate normal distribution. Based on these characterizations we suggest tests for normality with i.i.d. data as well as tests for the null hypothesis of a classical Gaussian multivariate GARCH model. The new test statistics are simple to implement and in limiting cases yield well-known moment-based criteria of goodness-of-fit. Their asymptotic null distribution is studied and consistency against fixed alternatives is proved. In a series of Monte Carlo experiments the new tests are implemented either based on asymptotic theory or by means of resampling schemes showing that the resulting power is often higher than competitors. For the GARCH case in particular, a conditional resampling version of the procedure shows good power performance against certain heavy-tailed alternatives to the multivariate normal distribution.

Table 4: Percentage of rejection for nominal level $\alpha = 0.05$ and $n = 300$.

			Test based on $\widetilde{T}_{n,\gamma}^G$					HW-test				
			γ					β				
	d	r	1.2	1.3	1.4	1.5	2.0	0.5	1.0	1.5	2.0	2.5
N	2	0.0	4.86	4.86	4.90	4.82	4.68	5.33	5.27	5.04	5.02	4.88
		0.3	4.63	4.56	4.59	4.60	4.64	4.57	4.69	4.99	4.89	5.06
	3	0.0	3.81	3.77	3.75	3.81	3.88	4.68	4.68	5.01	4.98	5.03
		0.3	6.84	7.08	7.15	6.97	7.08	4.85	5.02	5.65	5.16	5.14
T_{10}	2	0.0	89.15	89.05	88.85	88.70	89.05	4.40	17.00	27.75	29.10	27.70
		0.3	81.30	81.15	81.10	81.40	81.95	5.10	18.75	26.50	27.25	26.00
	3	0.0	95.80	95.90	95.80	95.80	96.05	6.15	35.10	46.90	43.20	37.05
		0.3	16.10	13.15	12.90	13.25	15.35	9.10	37.05	45.85	43.75	37.25
$GN_{1.65}$	2	0.0	48.70	48.60	49.00	48.85	48.60	5.50	7.20	11.20	13.00	13.5
		0.3	48.15	47.80	47.30	47.20	46.25	4.75	7.40	12.10	13.80	14.25
	3	0.0	59.10	59.80	59.35	59.20	58.00	4.10	7.55	11.65	12.15	12.85
		0.3	54.55	54.00	54.25	54.45	53.80	4.35	8.15	11.85	12.70	11.95
AEP	2	0.0	84.25	83.90	83.60	83.55	83.40	6.90	31.10	47.40	48.95	46.10
		0.3	77.55	77.50	77.35	77.80	78.10	6.65	29.00	44.95	47.45	46.00
	3	0.0	86.45	86.15	86.40	86.30	86.15	6.00	29.90	44.35	44.45	37.3
		0.3	14.00	11.85	12.25	12.25	13.15	7.00	32.80	44.00	44.90	38.00

10   Appendix: Technical proofs

Subsection 10.1 sketches the proofs of theoretical results. Subsection 10.2 contains some auxiliary results used in the proofs of our main theorems.

10.1   Proofs

PROOF of Theorem 5.1. Because of affine invariance of T_n, we assume without loss of generality that $\mathbb{E}X_1 = 0$ and $\mathbb{E}[X_1 X_1^\top] = \mathrm{I}_d$. The main idea for showing convergence of the process $U_n(\cdot) = \sqrt{n}(M_n(\cdot)R_n(\cdot) - 1)$ is as follows: Putting

$$A_n(t) = \exp\left(-\|t\|^2/2\right)\sqrt{n}\left(M_n(t) - \exp\left(\|t\|^2/2\right)\right),$$
$$B_n(t) = \exp\left(\|t\|^2/2\right)\sqrt{n}\left(R_n(t) - \exp\left(-\|t\|^2/2\right)\right),$$

where M_n and R_n are given in (3.3) and (3.2), respectively, the crucial observation, obtained after straightforward computations, is the representation

(10.1) $\quad U_n(t) = \dfrac{1}{\sqrt{n}} A_n(t) B_n(t) + A_n(t) + B_n(t), \quad t \in \mathbb{R}^d.$

Hence, the program is to prove $A_n + B_n \xrightarrow{\mathcal{D}} W$ in L^2. Since the first term on the right-hand side of (10.1) will turn out to be asymptotically negligible, Slutzky's lemma then gives the convergence $U_n \xrightarrow{\mathcal{D}} W$.

The main problem of dealing with $A_n + B_n$ is that both A_n and B_n are sums of functions of the scaled residuals $Y_{n,1}, \ldots, Y_{n,n}$ and not of the i.i.d. random vectors $X_1, \ldots, X_n$. To make the reasoning transparent, put

$$
\begin{aligned}
A_n^\circ(t) &= \exp\left(-\|t\|^2/2\right) \sqrt{n}\left(M_n^\circ(t) - \exp\left(\|t\|^2/2\right)\right), \\
B_n^\circ(t) &= \exp\left(\|t\|^2/2\right) \sqrt{n}\left(R_n^\circ(t) - \exp\left(-\|t\|^2/2\right)\right),
\end{aligned}
$$

where, in contrast to R_n and M_n,

(10.2) $\quad R_n^\circ(t) := \dfrac{1}{n} \sum_{j=1}^{n} \cos\left(t^\top X_j\right), \qquad M_n^\circ(t) := \dfrac{1}{n} \sum_{j=1}^{n} \exp\left(t^\top X_j\right), \quad t \in \mathbb{R}^d,$

are based on $X_1, \ldots, X_n$. Notice that $A_n^\circ + B_n^\circ = n^{-1/2} \sum_{j=1}^{n} Z_j$, where $Z_j(t) = \mathrm{e}^{-\|t\|^2/2}\, \mathrm{e}^{t^\top X_j} - 1 + \mathrm{e}^{\|t\|^2/2} \cos\left(t^\top X_j\right) - 1, t \in \mathbb{R}^d$. Since $\mathbb{E}[Z_1(t)] = 0$, $t \in \mathbb{R}^d$, and (due to $\gamma > 1$) $\mathbb{E}\|Z_1\|_{\mathrm{L}^2}^2 = \int_{\mathbb{R}^d} \mathbb{E}\left[Z_1^2(t)\right] w_\gamma(t)\, \mathrm{d}t < \infty$, a Central Limit Theorem (CLT) in Hilbert spaces (see e.g., [41]) shows that there is a centered Gaussian random element W° (say) of L^2, so that $A_n^\circ + B_n^\circ \xrightarrow{\mathcal{D}} W^\circ$.

The idea now is to approximate each of the differences $A_n(t) - A_n^\circ(t)$ and $B_n(t) - B_n^\circ(t)$ by sums of i.i.d. random variables. To this end, put

(10.3) $\quad \Delta_{n,j} = Y_{n,j} - X_j = \left(S_n^{-1/2} - \mathrm{I}_d\right) X_j - S_n^{-1/2} \overline{X}_n, \qquad j = 1, \ldots, n.$

A Taylor expansion gives

(10.4) $\quad \mathrm{e}^{t^\top Y_{n,j}} - \mathrm{e}^{t^\top X_j} = \mathrm{e}^{t^\top X_j} \left(t^\top \Delta_{n,j} + \dfrac{1}{2}\left(t^\top \Delta_{n,j}\right)^2 \exp\left(\Theta_{n,j} t^\top \Delta_{n,j}\right)\right),$

where $|\Theta_{n,j}| \leq 1$. It follows that

(10.5) $\quad A_n(t) - A_n^\circ(t) = \exp\left(-\|t\|^2/2\right) \sqrt{n}\left(M_n(t) - M_n^\circ(t)\right) = V_{n,1}(t) + V_{n,2}(t),$

23

where

(10.6) $\quad V_{n,1}(t) \;=\; \exp\left(-\|t\|^2/2\right) \dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} e^{t^\top X_j} t^\top \Delta_{n,j},$

(10.7) $\quad V_{n,2}(t) \;=\; \exp\left(-\|t\|^2/2\right) \dfrac{1}{2}\dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} e^{t^\top X_j} \left(t^\top \Delta_{n,j}\right)^2 \exp\left(\Theta_{n,j} t^\top \Delta_{n,j}\right).$

By some tedious and delicate estimations, we have $\|V_{n,2}\|_{L^2}^2 = o_{\mathbb{P}}(1)$ (see Proposition 10.3 in the Appendix), and Proposition 10.4 yields

$$V_{n,1}(t) = -\frac{1}{2}\frac{1}{\sqrt{n}} \sum_{j=1}^{n} t^\top \left(X_j X_j^\top - I_d\right) t - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} t^\top X_j + o_{\mathbb{P}}(1),$$

where $o_{\mathbb{P}}(1)$ refers to convergence in L^2. Likewise, we have

$$B_n(t) - B_n^\circ(t) \;=\; \exp\left(\|t\|^2/2\right) \sqrt{n}\left(R_n(t) - R_n^\circ(t)\right) \;=\; W_{n,1}(t) + W_{n,2}(t),$$

where

(10.8) $\quad W_{n,1}(t) \;=\; -\exp\left(\|t\|^2/2\right) \dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} \sin\left(t^\top X_j\right) t^\top \Delta_{n,j},$

$\qquad\qquad W_{n,2}(t) \;=\; \exp\left(\|t\|^2/2\right) \dfrac{1}{2}\dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} \Psi_{n,j} \left(t^\top \Delta_{n,j}\right)^2$

and $|\Psi_{n,j}| \leq 1$. Since

$$|W_{n,2}(t)| \leq \exp\left(\|t\|^2/2\right) \frac{\|t\|^2}{2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\Delta_{n,j}\|^2$$

and $n^{-1/2} \sum_{j=1}^{n} \|\Delta_{n,j}\|^2 = o_{\mathbb{P}}(1)$ (see [33], p. 9), it follows that $\|W_{n,2}\|_{L^2}^2 = o_{\mathbb{P}}(1)$. Moreover, Proposition 10.5 gives

$$W_{n,1}(t) = \frac{1}{2}\frac{1}{\sqrt{n}} \sum_{j=1}^{n} t^\top \left(X_j X_j^\top - I_d\right) t + o_{\mathbb{P}}(1).$$

Summarizing, we have $A_n(t) + B_n(t) = n^{-1/2} \sum_{j=1}^{n} Z_j(t) + o_{\mathbb{P}}(1)$, where

$$Z_j(t) = \exp\left(-\|t\|^2/2\right) e^{t^\top X_1} - 1 + \exp\left(\|t\|^2/2\right) \cos\left(t^\top X_j\right) - 1 - t^\top X_j.$$

Since $\mathbb{E}Z_j(t) = 0$, $t \in \mathbb{R}^d$, and, in view of the condition $\gamma > 1$, $\mathbb{E}\|Z_1\|_{L^2}^2 < \infty$, a Hilbert space CLT (see [41]) yields $A_n + B_n \xrightarrow{\mathcal{D}} W$ for a centered Gaussian element W of L^2

24

having covariance kernel $C(s,t) = \mathbb{E}[Z_1(s)Z_1(t)]$. Using

$$\mathbb{E}\left[\cos\left(s^\top X\right) e^{t^\top X}\right] = e^{(\|t\|^2 - \|s\|^2)/2} \cos\left(s^\top t\right),$$

$$\mathbb{E}\left[s^\top X e^{t^\top X}\right] = s^\top t\, e^{\|t\|^2/2},$$

$$\mathbb{E}\left[\cos\left(s^\top X\right)\cos\left(t^\top X\right)\right] = \frac{1}{2} e^{-(\|s\|^2 + \|t\|^2)/2}\left(e^{s^\top t} + e^{-s^\top t}\right),$$

straightforward computations show that $C(s,t)$ takes the form given in Theorem 5.1. As a side-product of the derivations, we see that both A_n and B_n converge in distribution and thus are tight sequences in L^2. Hence, the first term in (10.1) is $o_\mathbb{P}(1)$, which completes the proof of Theorem 5.1. ■

PROOF OF (5.3): Putting $I_\gamma^{(0)} := I_\gamma(0,0)$, we have

$$\mathbb{E}\|W\|_{\mathrm{L}^2}^2 = \frac{3}{2}I_{\gamma-1}^{(0)} + \frac{1}{2}I_{\gamma+1}^{(0)} - 4I_\gamma^{(0)} - J_\gamma + 2K_\gamma,$$

where $J_\gamma = \int_{\mathbb{R}^d} \|t\|^2 \exp\left(-\gamma\|t\|^2\right) \mathrm{d}t$, $K_\gamma = \int_{\mathbb{R}^d} \cos(\|t\|^2) \exp\left(-\gamma\|t\|^2\right) \mathrm{d}t$. Straightforward algebra gives $J_\gamma = d\, j_\gamma^{(2)}\left(j_\gamma^{(0)}\right)^{d-1}$, where

$$j_\gamma^{(m)} = \int_{-\infty}^\infty t^m \exp\left(-\gamma t^2\right) \mathrm{d}t, \ m = 0, 2.$$

Invoking

$$\cos\left(\sum_{\ell=1}^d \theta_\ell\right) = \sum_{q=0}^{\lfloor\frac{d}{2}\rfloor}(-1)^q \sum_{1 \le j_1 < j_2 \ldots < j_{2q} \le d} \prod_{k=1}^{2q} \sin\theta_{j_k} \prod_{\ell \ne j_1, j_2, \ldots, j_{2q}}^d \cos\theta_\ell,$$

we obtain

$$K_\gamma = \sum_{q=0}^{\lfloor\frac{d}{2}\rfloor}(-1)^q \binom{d}{2q}\left(\kappa_\gamma^{(s)}\right)^{2q}\left(\kappa_\gamma^{(c)}\right)^{d-2q},$$

where $\kappa_\gamma^{(s)} = \int_{-\infty}^\infty \sin(t^2) \exp\left(-\gamma t^2\right) \mathrm{d}t$, $\kappa_\gamma^{(c)} = \int_{-\infty}^\infty \cos(t^2) \exp\left(-\gamma t^2\right) \mathrm{d}t$. Since

$$j_\gamma^{(0)} = \sqrt{\frac{\pi}{\gamma}}, \quad j_\gamma^{(2)} = \frac{1}{2\gamma}\sqrt{\frac{\pi}{\gamma}},$$

$$\kappa_\gamma^{(s)} = \sqrt{\frac{\pi}{\gamma}}\frac{\sin\left((1/2)\tan^{-1}(1/\gamma)\right)}{(1 + (1/\gamma^2))^{1/4}}, \quad \kappa_\gamma^{(c)} = \sqrt{\frac{\pi}{\gamma}}\frac{\cos\left((1/2)\tan^{-1}(1/\gamma)\right)}{(1 + (1/\gamma^2))^{1/4}},$$

the result follows by simple algebra. ■

PROOF of Theorem 7.1. Notice that U_n^G satisfies an equality similar to that in (10.1) with A_n and B_n replaced with A_n^G and B_n^G, respectively, where

$$A_n^G(t) = \exp(-\|t\|^2/2)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\exp\left(t^\top\widetilde{\varepsilon}_j(\widehat{\theta}_n)\right) - \exp(\|t\|^2/2)\right],$$

$$B_n^G(t) = \exp(\|t\|^2/2)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\cos\left(t^\top\widetilde{\varepsilon}_j(\widehat{\theta}_n)\right) - \exp(-\|t\|^2/2)\right].$$

To prove the result we will demonstrate that

$$\begin{aligned}(10.9)\qquad A_n^G(t) + B_n^G(t) \;&=\; \tfrac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\exp\left(-\tfrac{\|t\|^2}{2}\right)\exp(t^\top\varepsilon_j) + \exp\left(\tfrac{\|t\|^2}{2}\right)\cos(t^\top\varepsilon_j) - 2\right]\\ &\quad + r_n(t),\end{aligned}$$

with $\|r_n\|_{\mathrm{L}^2} = o_{\mathbb{P}}(1)$, and the result will follow from the CLT in Hilbert spaces. With this aim, we first introduce some notation. Let $\varepsilon_j(\theta) = \Sigma_j^{-1/2}(\theta)X_j$. Notice that $\varepsilon_j(\theta_0) = \varepsilon_j$. Let

$$W_j = \varepsilon_j + \sum_{k=1}^{v}\left.\frac{\partial}{\partial\theta_k}\varepsilon_j(\theta)\right|_{\theta=\theta_0}(\widehat{\theta}_{nk} - \theta_{0k}),$$

where $\widehat{\theta}_n = (\widehat{\theta}_{n1},\ldots,\widehat{\theta}_{nv})^\top$, $\theta_0 = (\theta_{01},\ldots,\theta_{0v})^\top$, and $\Delta_{n,j} = \varepsilon_j(\widehat{\theta}_n) - W_j$. Then

$$\widetilde{\varepsilon}_j(\widehat{\theta}_n) - W_j = \widetilde{\varepsilon}_j(\widehat{\theta}_n) - \varepsilon_j(\widehat{\theta}_n) + \Delta_{n,j}.$$

Let $A_{jk}(\theta) = \Sigma_j^{-1/2}(\theta)\frac{\partial}{\partial\theta_k}\Sigma_j^{1/2}(\theta)$ and $\mu_k = \mathbb{E}[A_{jk}(\theta)]$. To show (10.9) we prove that

(a.1) $\exp(-\|t\|^2/2)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\exp(t^\top W_j) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_j(t) + r_{n,1}(t)$, with

$$V_j(t) = \exp(-\|t\|^2/2 + t^\top\varepsilon_j) - \sum_{k=1}^{v}t^\top\mu_k t\sqrt{n}(\theta_{nk} - \theta_{0k}),\quad \|r_{n,1}\|_{\mathrm{L}^2} = o_{\mathbb{P}}(1),$$

(a.2) $\exp(\|t\|^2/2)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos(t^\top W_j) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}Z_j(t) + r_{n,2}(t)$, with

$$Z_j(t) = \exp(\|t\|^2/2)\cos(t^\top\varepsilon_j) + \sum_{k=1}^{v}t^\top\mu_k t\sqrt{n}(\theta_{nk} - \theta_{0k}),\quad \|r_{n,2}\|_{\mathrm{L}^2} = o_{\mathbb{P}}(1),$$

(b.1) $\|r_{n,3}\|_{\mathrm{L}^2} = o_{\mathbb{P}}(1)$, where $r_{n,3}(t) = \mathrm{e}^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\exp(t^\top W_j)\left(\exp(t^\top\Delta_{n,j}) - 1\right)$,

(b.2) $\|r_{n,4}\|_{\mathrm{L}^2} = o_{\mathbb{P}}(1)$, where $r_{n,4}(t) = \mathrm{e}^{\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos(t^\top W_j)\left(\exp(t^\top\Delta_{n,j}) - 1\right)$,

26

(c.1) $\|r_{n,5}\|_{L^2} = o_{\mathbb{P}}(1)$, where $r_{n,5}(t) = e^{-\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\exp(\widetilde{\varepsilon}_j(\widehat{\theta}_n)) - \exp(\varepsilon_j(\widehat{\theta}_n)) \right]$,

(c.2) $\|r_{n,6}\|_{L^2} = o_{\mathbb{P}}(1)$, where $r_{n,6}(t) = e^{\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\cos(\widetilde{\varepsilon}_j(\widehat{\theta}_n)) - \cos(\varepsilon_j(\widehat{\theta}_n)) \right]$.

<u>Proof of (a.1).</u> Notice that $\left. \frac{\partial}{\partial \theta_k} \varepsilon_j(\theta) \right|_{\theta=\theta_0} = -A_{jk}(\theta_0)\varepsilon_j$. A Taylor expansion gives

$$
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \exp(t^\top W_j) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \exp(t^\top \varepsilon_j)
$$

$$
- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \exp(t^\top \varepsilon_j) t^\top \sum_{k=1}^{v} A_{jk}(\theta_0)(\widehat{\theta}_{nk} - \theta_{0k})\varepsilon_j + \frac{R_{n,1}(t)}{2},
$$

with $|R_{n,1}| \leq n^{-1/2} R_{n,1,1} R_{n,1,2}$, where

$$
R_{n,1,1}(t) = \exp\left(\|t\| \max_j \| \sum_{k=1}^{v} A_{jk}(\theta_0) \| \|\widehat{\theta}_n - \theta_0\| \max_j \|\varepsilon_j\| \right),
$$

$$
R_{n,1,2}(t) = \sum_{j=1}^{n} \exp(t^\top \varepsilon_j) \left(t^\top \sum_{k=1}^{v} A_{jk}(\theta_0)(\widehat{\theta}_{nk} - \theta_{0k})\varepsilon_j \right)^2.
$$

We have

$$
\exp(-\frac{\|t\|^2}{2}) \frac{1}{\sqrt{n}} \sum_{j=1}^{n} e^{t^\top \varepsilon_j} t^\top \sum_{k=1}^{v} A_{jk}(\theta_0)(\widehat{\theta}_{nk} - \theta_{0k})\varepsilon_j = t^\top \sum_{k=1}^{v} \mu_k \sqrt{n}(\theta_{nk} - \theta_{0k})t
$$

$$
+ R_{n,1,3}(t),
$$

where

$$
R_{n,1,3}(t) = \sum_{k=1}^{v} \sqrt{n}(\widehat{\theta}_{nk} - \theta_{0k}) R_{n,1,3,k}(t), \quad R_{n,1,3,k}(t) = \frac{1}{n} \sum_{j=1}^{n} V_{kj}(t),
$$

$$
V_{kj}(t) = \exp(-\|t\|^2/2) \exp(t^\top \varepsilon_j) t^\top A_{jk}(\theta_0)\varepsilon_j - t^\top \mu_k t.
$$

By noting that $\mathbb{E}[V_{kj}] = 0$, $\mathbb{E}[\langle V_{kj}, V_{kr} \rangle] = 0$, $\forall j \neq r$, $\forall k$, and $\mathbb{E}\|V_{kj}\|_{L^2}^2 < \infty$, it follows that $\|R_{n,1,3,k}\|_{L^2} = o_{\mathbb{P}}(1)$, $1 \leq k \leq v$, and thus $\|R_{n,1,3}\|_{L^2} = o_{\mathbb{P}}(1)$. Similar calculations show $\| \exp(-\|t\|^2/2) R_{n,1,2} \|_{L^2} = O_{\mathbb{P}}(1)$. Let $\alpha > 0$ so that $1/q < \alpha < 1/2$, where q is as in (A.6). Then

$$
\mathbb{P}\left(\max_j \| \sum_{k=1}^{v} A_{jk}(\theta_0) \| > n^\alpha \right) \leq n \mathbb{P}\left(\| \sum_{k=1}^{v} A_{jk}(\theta_0) \| > n^\alpha \right)
$$

$$
\leq n \frac{\mathbb{E}(\| \sum_{k=1}^{v} A_{jk}(\theta_0) \|^r)}{n^{r\alpha}} \to 0 \quad \text{as} \quad n \to \infty.
$$

This convergence together with (A.1) and Proposition 10.1 imply

$$\max_j \left\| \sum_{k=1}^v A_{jk}(\theta_0) \right\| \|\widehat{\theta}_n - \theta_0\| \max_j \|\varepsilon_j\| = o_{\mathbb{P}} \left(\frac{\sqrt{\log(n)}}{n^{0.5-\alpha}} \right).$$

Therefore, Proposition 10.2 gives $\| \exp(-\|t\|^2/2) R_{n,1,1} \|_{\mathrm{L}^2} = O_{\mathbb{P}}(1)$. As a consequence, $\| \exp(-\|t\|^2/2) R_{n,1} \|_{\mathrm{L}^2} = o_{\mathbb{P}}(1)$, and (a.1) follows.

Proof of (a.2). The proof is similar to that of (a.1), so we omit it.

Proof of (b.1). Observe that $\exp(t^\top \Delta_{n,j}) - 1 = t^\top \Delta_{n,j} \exp(\alpha_{n,j} t^\top \Delta_{n,j})$ for some $\alpha_{n,j} \in (0,1)$, and

$$\Delta_{n,j} = \sum_{k,l=1}^v \frac{\partial^2}{\partial \theta_k \partial \theta_l} \varepsilon_j(\theta) \bigg|_{\theta = \widetilde{\theta}_n} (\widehat{\theta}_{nk} - \theta_{0k})(\widehat{\theta}_{nl} - \theta_{0l})$$

for some $\widetilde{\theta}_n$ between $\widehat{\theta}_n$ and θ_0. Now (A.1) and (A.6) yield $\Delta_{n,j}\| \leq D_j \|\varepsilon_j\| \|\widehat{\theta}_n - \theta_0\|^2$ for large enough n, where $\mathbb{E}D_j^2 < \infty$. The Cauchy–Schwarz inequality gives

$$|r_{n,3}(t)| \leq r_{n,3,1}(t)^{1/2} r_{n,3,2}(t)^{1/2},$$

where

$$r_{n,3,1}(t) = \mathrm{e}^{-\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^n \exp(t^\top W_j),$$

$$r_{n,3,2}(t) = \mathrm{e}^{-\|t\|^2/2} \|t\|^2 \|\sqrt{n}(\widehat{\theta}_n - \theta_0)\|^4 \frac{1}{\sqrt{n}n^2} \sum_{j=1}^n D_j^2 \|\varepsilon_j\|^2 \exp(2\|t\| \|D_j\| \|\varepsilon_j\| \|\widehat{\theta}_n - \theta_0\|^2).$$

Proceeding as in the proof of (a.1), it can be seen that $r_{n,3,1} = O_{\mathbb{P}}(1)$ (in L^2) and that $r_{n,3,2} = o_{\mathbb{P}}(1)$ (in L^2), showing that $r_{n,3} = o_{\mathbb{P}}(1)$ (in L^2).

Proof of (b.2). The proof is similar to that of (b.1) and is thus omitted.

Proof of (c.1). Let $\Lambda_{n,j} = \widetilde{\varepsilon}_j(\widehat{\theta}_n) - \varepsilon_j(\widehat{\theta}_n) = \widetilde{\Sigma}_j^{-1/2}(\widehat{\theta}_n) \left(\Sigma_j^{1/2}(\widehat{\theta}_n) - \widetilde{\Sigma}_j^{1/2}(\widehat{\theta}_n) \right) \Sigma_j^{-1/2}(\widehat{\theta}_n) X_j$. A Taylor expansion yields

$$(10.10) \quad \exp\left(t^\top \widetilde{\varepsilon}_j(\widehat{\theta}_n)\right) - \exp\left(t^\top \varepsilon_j(\widehat{\theta}_n)\right) = t^\top \Lambda_{n,j} \exp\left(t^\top \varepsilon_j(\widehat{\theta}_n) + \alpha_{n,j} t^\top \Lambda_{n,j}\right)$$

for some $\alpha_{n,j} \in (0,1)$. By the Cauchy–Schwarz inequality and (10.10),

$$\exp(-\|t\|^2/2) \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[\exp\left(t^\top \widetilde{\varepsilon}_j(\widehat{\theta}_n)\right) - \exp\left(t^\top \varepsilon_j(\widehat{\theta}_n)\right) \right] \right| \leq R_{n,1}(t)^{1/2} R_{n,2}(t)^{1/2},$$

where

$$R_{n,1}(t) = \exp(-\|t\|^2/2)\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \exp\left(2t^\top \varepsilon_j(\widehat{\theta}_n)\right),$$

$$R_{n,2}(t) = \exp(-\|t\|^2/2)\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|t\|^2 \|\Lambda_{n,j}\|^2 \exp(2\|t\|\|\Lambda_{n,j}\|).$$

Proceeding as in (a.1), we obtain $\|R_{n,1}\|_2 = O_\mathbb{P}(1)$. We next show $\|R_{n,2}\|_{\mathrm{L}^2} = o_\mathbb{P}(1)$, which yields (c.1). To this end, put $\Lambda_{n,j,k} = \|\Lambda_{n,j}\| + \|\Lambda_{n,k}\|$. From Proposition 10.2,

$$\|R_{n,2}\|_{\mathrm{L}^2}^2 \le \frac{1}{n} \sum_{j,k=1}^{n} \|\Lambda_{n,j}\|^2 \|\Lambda_{n,k}\|^2 \left(K_0 + K_1 \Lambda_{n,j,k}^{d+3}\right) \exp\left(\frac{\Lambda_{n,j,k}^2}{\gamma+1}\right).$$

Taking into account that for any $x, y \in \mathbb{R}$ and $r > 0$

$$(10.11) \quad |x+y|^r = c_r(|x|^r + |y|^r),$$

where $c_r = 1$ if $0 < r \le 1$ and $c_r = 2^{r-1}$, otherwise, we have

$$\|R_{n,2}\|_{\mathrm{L}^2}^2 \le C \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\Lambda_{n,j}\|^2 \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) \right\}^2$$

$$+ C \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\Lambda_{n,j}\|^2 \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) \right\} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\Lambda_{n,j}\|^{d+5} \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) \right\}.$$

We show that each expression within curly brackets is $o_\mathbb{P}(1)$. From (A.2) and (A.3),

$$(10.12) \quad \|\Lambda_{n,j}\| \le C\rho^j \|X_j\|.$$

As a consequence, (A.4) implies

$$(10.13) \quad \Lambda_{n,j} \to 0 \quad \text{a.s.}$$

as $j \to \infty$ for each n, because the upper bound in (10.12) does not depend on n (see Exercise 7.2 in [24]. From (10.13), it follows that for each $M > 1$ and each $\omega \in \Omega$, there is an integer $j_0 = j_0(\omega, M)$ so that $\exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) \le M$ for each n and each $j, k > j_0$. For $n > n_0$ and $r = 2$ or $r = d+5$,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\Lambda_{n,j}\|^r \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) \le T_1 + T_2,$$

29

where

$$T_1 = \frac{n_0}{\sqrt{n}} \max_{1 \le j \le n_0} \|\Lambda_{n,j}\|^r \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right), \quad T_2 = \frac{M}{\sqrt{n}} \sum_{j=n_0+1}^{n} \|\Lambda_{n,j}\|^r.$$

Let $0 < \alpha < 1/2$, and put $\log^+(x) = \max\{0, \log(x)\}$. Then

$$\mathbb{P}\left(\max_{1 \le j \le n_0} \|\Lambda_{n,j}\|^r \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) > n^\alpha\right) \le \sum_{j=1}^{n_0} \mathbb{P}\left(\|\Lambda_{n,j}\|^r \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) > n^\alpha\right)$$

$$\le C \sum_{j=1}^{n_0} \frac{\mathbb{E}[\log^+(\|\Lambda_{n,j}\|)^{\zeta/2}] + \mathbb{E}\|\Lambda_{n,j}\|^\zeta}{(\log(n))^{\zeta/2}} \le C \frac{\mathbb{E}\|X_1\|^\zeta \sum_{j=1}^{\infty} \rho^{j\zeta}}{(\log(n))^{\zeta/2}},$$

which implies $\max_{1 \le j \le n_0} \|\Lambda_{n,j}\|^r \exp\left(\frac{2\|\Lambda_{n,j}\|^2}{\gamma+1}\right) = o_{\mathbb{P}}(n^\alpha)$ and thus $T_1 = o_{\mathbb{P}}(1)$. From (A.4) and (10.11), for any $0 < \zeta \le \min\{1, \varsigma\}$, we have

$$\mathbb{E}\left(\sum_{j=n_0+1}^{n} \|\Lambda_{n,j}\|^r\right)^{\zeta/r} \le \sum_{j=n_0+1}^{n} \mathbb{E}\|\Lambda_{n,j}\|^\zeta \le \mathbb{E}\|X_1\|^\zeta \sum_{j=n_0+1}^{\infty} \rho^{j\zeta} < \infty.$$

Since $\sum_{j=n_0+1}^{n} \|\Lambda_{n,j}\|^r$ has finite moment of order ζ/r, it is finite almost surely. Thus $T_2 \to 0$ a.s. as $n \to \infty$, which completes the proof of (c.1).

Proof of (c.2). The proof is similar to that of (c.1) and is thus omitted. ■

10.2   Some auxiliary results

Proposition 10.1 *Let $X, X_1, X_2, \ldots$ be i.i.d. d-variate random vectors having the normal distribution $N_d(0, I_d)$, and put $F_n := \max_{1 \le j \le n} \|X_j\|$. Then*

$$\lim_{n \to \infty} \mathbb{P}\left(4\sqrt{2 \log n}\,(F_n - a_n) \le t\right) = \exp(-\exp(-t)), \qquad t \in \mathbb{R},$$

where

$$a_n = \sqrt{2 \log n} + \frac{(d-2) \log \log n}{2\sqrt{2 \log n}} - \frac{\log \Gamma(d/2)}{\sqrt{2 \log n}}.$$

PROOF. Since $\|X\|^2$ has a χ_d^2-distribution and thus a Gamma distribution, the distribution of F_n^2 is in the maximum domain of attraction of the Gumbel law. From p. 156 of [17] we therefore obtain $2(F_n^2 - d_n) \xrightarrow{\mathcal{D}} G$, where

$$d_n = 2\left(\log n + \left(\frac{d}{2} - 1\right) \log \log n - \log \Gamma(d/2)\right)$$

and G has a Gumbel distribution. Using $F_n^2 - d_n = (F_n - \sqrt{d_n})(F_n + \sqrt{d_n})$ together with $(F_n + \sqrt{d_n})/\sqrt{d_n} \xrightarrow{\mathbb{P}} 2$ and Sluzky's lemma, we have $4\sqrt{d_n}(F_n - \sqrt{d_n}) \xrightarrow{\mathcal{D}} Z$. Upon noting that $\sqrt{d_n} = a_n + o(1/\sqrt{2 \log n})$ as $n \to \infty$, the assertion follows. ∎

Proposition 10.2 *Let $k \in \mathbb{N} \cup \{0\}$ and $\alpha > 0$. Then*

$$\int_{\mathbb{R}^d} \|t\|^k \exp\left(-(1+\gamma)\|t\|^2 + 2\,\alpha\,\|t\|\right)\,dt \;\leq\; \left(K_0 + K_1\,\alpha^{d+k-1}\right) \exp\left(\frac{\alpha^2}{1+\gamma}\right)$$

for some constants K_0, K_1 that depend only on d, k and γ.

PROOF. Using spherical coordinates, the integral equals

$$C_d \exp\left(\frac{\alpha^2}{1+\gamma}\right) \int_0^\infty r^{k+d-1} \exp\left(-(1+\gamma)\left(r - \frac{\alpha}{1+\gamma}\right)^2\right)\,dr,$$

where C_d is a constant that depends only on d. This last integral, in turn, is equal to $\frac{1}{2}\sigma\sqrt{2\pi}\,\mathbb{E}\left[|N|^{k+d-1}\right]$, where $\sigma^2 = (2(1+\gamma))^{-1}$ and N has the normal distribution $N(\mu, \sigma^2)$, where $\mu = \alpha/(1+\gamma)$. From this, the result follows readily. ∎

Proposition 10.3 *We have $\|V_{n,2}\|_{L^2}^2 = o_{\mathbb{P}}(1)$, where $V_{n,2}$ is given in (10.7).*

PROOF. Putting $F_n := \max_{1 \leq j \leq n} \|X_j\|$, (10.3) gives

(10.14) $\quad |t^\top \Delta_{n,j}| \leq \|t\|\,\Lambda_n,$

where

(10.15) $\quad \Lambda_n = \|S_n^{-1/2} - I_d\|_2\,F_n + \|S_n^{-1/2}\|_2\,\|\overline{X}_n\|.$

On the other hand, we have $(t^\top \Delta_{n,j})^2 \leq 2(t^\top(S_n^{-1/2} - I_d)X_j)^2 + 2(t^\top S_n^{-1/2}\overline{X}_n)^2$, which implies $0 \leq V_{n,2}(t) \leq V_{n,2,1}(t) + V_{n,2,2}(t)$, where

$$V_{n,2,1}(t) = e^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^n e^{t^\top X_j}\left(t^\top\left(S_n^{-1/2} - I_d\right)X_j\right)^2 \exp\left(\Theta_{n,j}t^\top\Delta_{n,j}\right),$$

(10.16) $\quad V_{n,2,2}(t) = e^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^n e^{t^\top X_j}\left(t^\top S_n^{-1/2}\overline{X}_n\right)^2 \exp\left(\Theta_{n,j}t^\top\Delta_{n,j}\right).$

Since $\|V_{n,2}\|_{L^2}^2 \leq 2\|V_{n,2,1}\|_{L^2}^2 + 2\|V_{n,2,2}\|_{L^2}^2$, it suffices to prove that each of the last two summands is $o_{\mathbb{P}}(1)$. To tackle $V_{n,2,1}$, notice that

$$\left(t^\top\left(S_n^{-1/2} - I_d\right)X_j\right)^2 \leq \|t\|^2\,\|S_n^{-1/2} - I_d\|_2^2\,\|X_j\|^2$$

31

and hence, invoking (10.14),

$$V_{n,2,1}(t) \leq \|S_n^{-1/2} - \mathrm{I}_d\|_2^2 \, \|t\|^2 \, \mathrm{e}^{-\|t\|^2/2} \, \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathrm{e}^{t^\top X_j} \, \|X_j\|^2 \, \mathrm{e}^{\|t\|\Lambda_n}.$$

Consequently,

$$\|V_{n,2,1}\|_{\mathrm{L}^2}^2 \leq \|S_n^{-1/2} - \mathrm{I}_d\|_2^4 \frac{1}{n} \sum_{i,j=1}^n \|X_i\|^2 \, \|X_j\|^2 \int_{\mathbb{R}^d} \|t\|^4 \mathrm{e}^{-(1+\gamma)\|t\|^2} \, \mathrm{e}^{t^\top(X_i+X_j)+2\|t\|\Lambda_n} \, \mathrm{d}t.$$

Now, using $|t^\top(X_i + X_j)| \leq 2\|t\| \, F_n$ and Proposition 10.2 with $k = 4$ and $\alpha = F_n + \Lambda_n$, it follows that

$$(10.17) \quad \|V_{n,2,1}\|_{\mathrm{L}^2}^2 \leq \|S_n^{-1/2} - \mathrm{I}_d\|_2^4 \cdot n \cdot \frac{1}{n^2} \sum_{i,j=1}^n \|X_i\|^2 \, \|X_j\|^2 \, (K_0 + K_1 \Gamma_n^{d+3}) \, \exp\left(\frac{\Gamma_n^2}{1+\gamma}\right),$$

where $\Gamma_n = F_n + \Lambda_n$. From Proposition 10.1, we obtain

$$F_n = \sqrt{2 \log n} + \frac{(d-2) \log \log n}{2\sqrt{2 \log n}} + O_\mathbb{P}\left(\frac{1}{\sqrt{\log n}}\right).$$

In view of $\|S_n^{-1/2} - \mathrm{I}_d\|_2 = O_\mathbb{P}(n^{-1/2})$ and $\|\overline{X}_n\| = O_\mathbb{P}(n^{-1/2})$, we see that Λ_n figuring in (10.15) is of order $O_\mathbb{P}((\log n/n)^{1/2})$ and thus

$$\Gamma_n = \sqrt{2 \log n} + \frac{(d-2) \log \log n}{2\sqrt{2 \log n}} + O_\mathbb{P}\left(\frac{1}{\sqrt{\log n}}\right).$$

Hence, $\Gamma_n^2 = 2 \log n + (d-2) \log \log n + O_\mathbb{P}(1)$ and therefore

$$\frac{\Gamma_n^2}{1+\gamma} = \log\left(n^{2/(1+\gamma)}\right) + \log\left((\log n)^{(d-2)/(1+\gamma)}\right) + O_\mathbb{P}(1).$$

It follows that the rightmost factor figuring in (10.17) is

$$(10.18) \quad \exp\left(\frac{\Gamma_n^2}{1+\gamma}\right) = n^{2/(1+\gamma)} \, (\log n)^{(d-2)/(1+\gamma)} \, O_\mathbb{P}(1).$$

Since $\|S_n^{-1/2} - \mathrm{I}_d\|_2^4 = O_\mathbb{P}(n^{-2})$, $n^{-2} \sum_{i,j=1}^n \|X_i\|^2 \, \|X_j\|^2 = O_\mathbb{P}(1)$ and Γ_n^{d+3} is of order $O_\mathbb{P}((\log n)^{(d+3)/2})$, (10.18) yields

$$\|V_{n,2,1}\|_{\mathrm{L}^2}^2 = O_\mathbb{P}\left(n^{(1-\gamma)/(1+\gamma)}\right) \cdot (\log n)^{\frac{d-2}{1+\gamma} + \frac{d+3}{2}}$$

and thus $\|V_{n,2,1}\|_{\mathrm{L}^2}^2 = o_\mathbb{P}(1)$, since $\gamma > 1$.

To show that $\|V_{n,2,2}\|_{\mathrm{L}^2}^2 = o_\mathbb{P}(1)$, where $V_{n,2,2}$ is given in (10.16), a similar reasoning as above yields $\|V_{n,2,2}\|_{\mathrm{L}^2}^2 = O_\mathbb{P}(n^{-1}) \cdot \int_{\mathbb{R}^d} \|t\|^4 \mathrm{e}^{-(1+\gamma)\|t\|^2} \, \mathrm{e}^{2\|t\|(F_n+\Lambda_n)} \, \mathrm{d}t$. Hence, $\|V_{n,2,2}\|_{\mathrm{L}^2}^2$ is of the same order as $\|V_{n,2,1}\|_{\mathrm{L}^2}^2$, which completes the proof. ∎

Proposition 10.4 *We have*

$$V_{n,1}(t) = -\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}t^{\top}\left(X_jX_j^{\top}-I_d\right)t - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}t^{\top}X_j + o_{\mathbb{P}}(1),$$

where $V_{n,1}$ is given in (10.6) and $o_{\mathbb{P}}(1)$ is with respect to L^2.

PROOF. From the definition of $\Delta_{n,j}$, we have $V_{n,1}(t) = V_{n,1,1}(t) - V_{n,1,2}(t) - V_{n,1,3}(t)$, where

$$V_{n,1,1}(t) = e^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}e^{t^{\top}X_j}t^{\top}\left(S_n^{-1/2}-I_d\right)X_j,$$

$$V_{n,1,2}(t) = e^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}e^{t^{\top}X_j}t^{\top}\left(S_n^{-1/2}-I_d\right)\overline{X}_n,$$

$$V_{n,1,3}(t) = e^{-\|t\|^2/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}e^{t^{\top}X_j}t^{\top}\overline{X}_n.$$

Since $|t^{\top}\left(S_n^{-1/2}-I_d\right)\overline{X}_n| \le \|t\|\,\|S_n^{-1/2}-I_d\|_2\,\|\overline{X}_n\| = \|t\|\,O_{\mathbb{P}}(n^{-1})$, it is readily seen that $\|V_{n,1,2}\|_{L^2}^2 = o_{\mathbb{P}}(1)$. Now,

$$V_{n,1,3}(t) = e^{-\|t\|^2/2}\left(M_n^{\circ}(t)-m(t)\right)\frac{1}{\sqrt{n}}\sum_{k=1}^{n}t^{\top}X_k + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}t^{\top}X_j.$$

with $M_n^{\circ}(t)$ given in (10.2) and $m(t) = \exp(\|t\|^2/2)$. Taking expectations and using Fubini's Theorem, it follows that $V_{n,1,3}(t) = n^{-1/2}\sum_{j=1}^{n}t^{\top}X_j + o_{\mathbb{P}}(1)$, where $o_{\mathbb{P}}(1)$ is understood with respect to L^2. Finally, we use the relation

$$(10.19)\quad \sqrt{n}\left(S_n^{-1/2}-I_d\right) = -\frac{1}{2\sqrt{n}}\sum_{k=1}^{n}\left(X_kX_k^{\top}-I_d\right) + O_{\mathbb{P}}\left(n^{-1/2}\right)$$

(see display (2.13) of [33]). Replacing $\sqrt{n}(S_n^{-1/2}-I_d)$ in the expression of $V_{n,1,1}(t)$ with the right-hand side of (10.19), we obtain

$$V_{n,1,1}(t) = -\frac{1}{2}e^{-\|t\|^2/2}\left(\frac{1}{n}\sum_{j=1}^{n}e^{t^{\top}X_j}X_j\right)^{\top}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(X_jX_j^{\top}-I_d\right)t + o_{\mathbb{P}}(1).$$

Replacing $n^{-1}\sum_{j=1}^{n}\exp(t^{\top}X_j)X_j$ with its expectation $\mathbb{E}[\exp(t^{\top}X_1)X_1] = \exp(\|t\|^2/2)t$ means adding a term that is $o_{\mathbb{P}}(1)$ in L^2 which, upon combining with the non-negligible term of $V_{n,1,3}$, yields the assertion. ∎

Proposition 10.5 *We have*

$$W_{n,1}(t) = \frac{1}{2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} t^\top \left(X_j X_j^\top - I_d \right) t + o_{\mathbb{P}}(1),$$

where $W_{n,1}$ is given in (10.8) and $o_{\mathbb{P}}(1)$ is with respect to L^2.

PROOF. We have $W_{n,1}(t) = -W_{n,1,1}(t) + W_{n,1,2}(t) + W_{n,1,3}(t)$, where

$$W_{n,1,1}(t) = e^{\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin\left(t^\top X_j\right) t^\top \left(S_n^{-1/2} - I_d\right) X_j,$$

$$W_{n,1,2}(t) = e^{\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin\left(t^\top X_j\right) t^\top \left(S_n^{-1/2} - I_d\right) \overline{X}_n,$$

$$W_{n,1,3}(t) = e^{\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin\left(t^\top X_j\right) t^\top \overline{X}_n.$$

By complete analogy with the reasoning given in the proof of Proposition 10.4, we have $\|W_{n,1,2}\|_{L^2}^2 = o_{\mathbb{P}}(1)$, and it is readily seen that also $\|W_{n,1,3}\|_{L^2}^2 = o_{\mathbb{P}}(1)$. Finally, by making use of (10.19) and proceeding as in the proof of Proposition 10.4 (notice that $\mathbb{E}[\sin(t^\top X_1)X_1] = \exp(-\|t\|^2/2)t)$, the assertion follows. ∎

Acknowledgements

References

[1] Arcones, M. (2007). Two tests for multivariate normality based on the characteristic function. *Math. Methods Statist.*, **16**, 177–201.

[2] Bai, J., and Chen, Z. (2008). Testing multivariate distributions in GARCH models. *J. Econometrics*, **143**, 19–36.

[3] Bardet, J. M. and Wintenberger, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.*, **37**, 2730–2759.

[4] Baringhaus, L., and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, **35**, 339–348.

[5] Baringhaus, L., Ebner, B., and Henze, N. (2016). The limit distribution of weighted L^2-goodness-of fit statistics under fixed alternatives, with applications. *Ann. Inst. Stat. Math.*, doi:10.1007/s10463-016-0567-8

[6] Barndorff-Nielsen, O. (1963). On the limit behaviour of extreme order statistics. *Ann. Math. Statist.*, **34**, 992–1012.

[7] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Rev. Econ. Stat.*, **72**, 498–505.

[8] Bonato, M. (2012). Modeling fat tails in stock returns: a multivariate stable-GARCH approach. *Comput. Statist.*, **27**, 499–521.

[9] Chen, Q., Gerlach, R., and Lu, Z. (2012) Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Comput. Statist. Data Anal.*, **56**, 3498–3516.

[10] Comte, F. and Lieberman, O. (2003) Asymptotic theory for multivariate GARCH processes. *J. Multivariate Anal.*, **84**, 61–84.

[11] Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.*, **14**, 708–723.

[12] Csörgő, S. (1989), Consistency of some tests for multivariate normality. *Metrika*, **36**, 107–116.

[13] Csörgő, S., and Welsh, A.H. (1989). Testing for exponential and Marshall–Olkin distributions. *J. Statist. Plann. Inference*, **23**, 287–300.

[14] Dalla, V., Meintanis, S.G. and Bassiakos, Y. (2017). Characteristic function-based inference for GARCH models with heavy-tailed innovations. *Comm. Statist. Simulation Comput.*, **46**, 2733–2755.

[15] Diamantopoulos, K., and Vrontos, I.D. (2010). A Student-t full factor multivariate GARCH model. *Comput. Econom.* , **35**, 63–83.

[16] Eaton, M.L., and Perlman, M.D. (1973). The non-singularity of generalized sample covariance matrices. *Ann. Statist.*, **1**, 710–717.

[17] Embrechts, P., Klüppelberg, C., and Mikosch, Th. (1997). *Modelling extremal events.* Springer, New York.

[18] Epps, T.W. (1999). Limiting behavior of the ICF test for normality under Gram–Charlier alternatives. *Statist. Probab. Lett.*, **42**, 175–184.

[19] Epps, T.W., and Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, **70**, 723–726.

[20] Fama, E. (1965). The behavior of stock market prices. *J. Bus.*, **38**, 34–105.

[21] Fang, K.–T., Li, R.–Z. and Liang, J.–J. (1998). A multivariate version of Ghosh's T_3-plot to detect non-multinormality. *Comput. Statist. Data Anal.*, **28**, 371–386.

[22] Francq, C., Jiménez-Gamero, M.D., and Meintanis, S.G. (2017). Tests for sphericity in multivariate GARCH models. *J. Econometrics*, **196**, 305–319.

[23] Francq, C., and Meintanis, S.G. (2016). Fourier-type estimation of the power GARCH model with stable Paretian innovations. *Metrika*, **79**, 389–424.

[24] Francq, C. and Zakoïan, J.M. (2010). *GARCH Models: Structure, Statistical Inference and Applications*. Wiley, London.

[25] Francq, C. and Zakoïan, J.M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* , **28**, 179–206.

[26] Ghoudi K., and Rémillard, B. (2014). Comparison of specification tests for GARCH models. *Computat. Statist. Data Anal.*, **76**, 291–300.

[27] Ghosh, S. (1996). A new graphical tool to detect non-normality. *J. Roy. Statist. Soc. Ser. B*, **58**, 691–702.

[28] Giacomini, R., Politis, D.N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**, 567–589.

[29] Goodman, I.R., Kotz, S. (1973). Multivariate θ-generalized normal distributions. *J. Multivariate Anal.*, **3**, 204–219.

[30] Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statist. Probab. Lett.*, **35**, 203–213.

[31] Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statist. Papers*, **43**, 467–506.

[32] Henze, N., and Koch, S. (2016) On a test of normality based on the empirical moment generating function. *arXiv:1612.08586*.

[33] Henze, N., and Wagner, T. (1997) A new approach to the BHEP tests for multivariate normality. *J. Multivariate Anal.*, **62**, 1–23.

[34] Henze, N., and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Comm. Statist. Theory Methods*, **19**, 3595–3617.

[35] Janssen, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.*, **28**, 239–253.

[36] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, **14**, 70–86.

[37] Jiménez-Gamero, M.D. (2014). On the empirical characteristic function process of the residuals in GARCH models and applications. *Test*, **23**, 409–432.

[38] Jiménez-Gamero, M.D., and Pardo-Fernández, J.C. (2017). Empirical characteristic function tests for GARCH innovation distribution using multipliers. *J. Stat. Comput. Simul.*, **87**, 2069–2093.

[39] Klar, B., Lindner, F. and Meintanis, S.G. (2012). Specification tests for the error distribution in GARCH models. *Comput. Statist. Data Anal.*, **56**, 3587–3598.

[40] Kotz, S., Kozubowski, T.J. and Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance*. Birkhäuser, Boston.

[41] Kundu, D., Majumdar, S., and Mukherjee, K. (2000). Central Limit theorems revisited. *Statist. Probab. Lett.*, **47**, 265–275.

[42] Lee, J., Lee, S. and Park, S. (2014). Maximum entropy test for GARCH models. *Stat. Methodol.*, **22**, 8–16.

[43] Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, **19**, 280–310.

[44] Lukacs, E. (1970). *Characteristic Functions*. Griffin, London.

[45] Mandelbrot, B. (1963). The variation of certain speculative prices. *J. Bus.*, **36**, 394–419.

[46] Mardia, K.V. (1070). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

[47] Meintanis, S.G. (2007). A Kolmogorov-Smirnov type test for Skew Normal distributions based on the empirical moment generating function. *J. Statist. Plann. Inference*, **137**, 2681–2688.

[48] Meintanis, S.G, and Hlávka. Z. (2010). Goodness-of-fit test for bivariate and multivariate skew-normal distributions. *Scand. J. Statist.*, **37**, 701–714.

[49] Móri, T.F., Rohatgi, V.K., and Székely, G.J. (1993). On multivariate skewness and kurtosis. *Theory Probab. Appl.* **38**, 547–551.

[50] Ogata, H. (2013). Estimation of multivariate stable distributions with generalized empirical likelihood. *J. Econometrics*, **172**, 248–254.

[51] Pudelko, J. (2005). On a new affine invariant and consistent test for multivariate normality. *Probab. Math. Statist.*, **25**, 43–54.

[52] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

[53] Shorack, G. R., and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[54] Spierdijk, L. (2016). Confidence intervals for ARMA-GARCH Value-at-Risk: The case of heavy tails and skewness. *Computat. Statist. Data Anal.*, **100**, 545–559.

[55] Tenreiro, C. (2011). An affine invariant multiple test procedure for assessing multivariate normality. *Comput. Statist. Data Anal.*, **55**, 1980–1992.

[56] Trindade, A.A., and Zhu, Y. (2007). Approximating the distributions of estimators of financial risk under an asymmetric Laplace law. *Comput. Statist. Data Anal.*, **51**, 3433–3447.

[57] Tsay, R.S. (2006). Multivariate volatility models. In: *IMS Lecture Notes–Monograph Series, Vol. 52: Time series and Related Topics*. Institute of Mathematic Statistics, 210–222.

[58] Ushakov, N.G. (1999). *Selected Topics in Characteristic Functions*. VSP, Utrecht.

[59] Volkmer, H. (2014) A characterization of the normal distribution. *J. Statist. Theor. Applic.*, **13**, 83–85.

[60] Zghoul, A.A. (2010) A goodness-of-fit test for normality based on the empirical moment generating function. *Comm. Statist. Simulation Comput.*, **39**, 1292–1304.

[61] Zhu, D., and Zinde-Walsh, V. (2009) Properties and estimation of asymmetric exponential power distribution. *J. Econometrics*, **148**, 86–99.