A NEW CLASS OF TESTS FOR MULTINORMALITY WITH I.I.D. AND GARCH DATA BASED ON THE EMPIRICAL MOMENT GENERATING FUNCTION

NORBERT HENZE[1], MARÍA DOLORES JIMÉNEZ–GAMERO[2],

[1]*Institute of Stochastics, Karlsruhe Institute of Technology, Karlsruhe, Germany*
[2]*Department of Statistics and Operations Research, University of Seville, Seville, Spain*

Abstract. We generalize a recent class of tests for univariate normality that are based on the empirical moment generating function to the multivariate setting, thus obtaining a class of affine invariant, consistent and easy-to-use goodness-of-fit tests for multinormality. 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 study the finite-sample behavior of the new tests, compare the criteria with alternative existing procedures, and apply the new procedure to a data set of monthly log returns.

1. Introduction

As evidenced by the papers Arcones (2007), Batsidis et al. (2013), Cardoso de Oliveira and Ferreira (2010), Ebner (2012), Enomoto et al. (2012), Farrel et al. (2007), Hanusz and Tarasińska (2008, 2012), Henze et al. (2017), Joenssen and Vogel (2014), Jönsson (2011), Kim (2016), Koizumi et al. (2014), Mecklin and Mundfrom (2005), Pudelko (2005), Székely and Rizzo (2005), Tenreiro (2011, 2017), Thulin (2014), Villa-

señor-Alva and Estrada (2009), Voinov et al. (2016), Yanada et al. (2015), and Zhou and Shao (2014), there is an ongoing interest in the problem of testing for multivariate normality. Without claiming to be exhaustive, the above list probably covers most of the publications in this field since the review paper Henze (2002).

Recently, Henze and Koch (2017) provided the lacking theory for a test for *univariate* normality suggested by Zghoul (2010). The purpose of this paper is twofold. First, we generalize the results of Henze and Koch (2017) to the multivariate case, thus obtaining a class of affine invariant and consistent tests for multivariate normality. Second, in contrast to that paper (and most of the other publications), which considered only independent and identically distributed (i.i.d.) observations, we also provide the asymptotics of our test statistics in the context of GARCH-type dependence.

To be more specific, let (for the time being) $X, X_1, X_2, \ldots$ be a sequence of i.i.d. d-variate random column vectors that are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that the distribution $\mathbb{P}^X$ is absolutely continuous with respect to Lebesgue measure. Let $N_d(\mu, \Sigma)$ denote the d-variate normal distribution with mean vector μ and non-degenerate covariance matrix Σ, and write $\mathcal{N}_d$ for the class of all non-degenerate d-dimensional normal distributions. A test for multivariate normality is a test of the null hypothesis

$$H_0: \ \mathbb{P}^X \in \mathcal{N}_d,$$

and usually such a test should be consistent against any fixed non-normal alternative distribution. 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 affine invariant, 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 Henze (2002) for a critical account on affine invariant tests for multivariate normality.

In what follows, let $\overline{X}_n = n^{-1} \sum_{j=1}^n X_j$, $S_n = n^{-1} \sum_{j=1}^n (X_j - \overline{X}_n)(X_j - \overline{X}_n)^\top$ denote the sample mean and the sample covariance matrix of $X_1, \ldots, X_n$, respectively, where $\top$ means transposition of vectors and matrices. Furthermore, let

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

be the so-called *scaled residuals* of $X_1, \ldots, X_n$, which provide an empirical standard-

ization of $X_1, \ldots, X_n$. Here, $S_n^{-1/2}$ denotes the unique symmetric square root of S_n. Notice that S_n is invertible with probability one provided that $n \geq d + 1$, see Eaton and Perlman (1973). The latter condition is tacitly assumed to hold in what follows. Letting

$$M_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp\left(t^\top Y_{n,j}\right), \quad t \in \mathbb{R}^d, \tag{1.1}$$

denote the empirical moment generating function of $Y_{n,1}, \ldots, Y_{n,n}$, $M_n(t)$ should be close to

$$m(t) = \exp(\|t\|^2 / 2),$$

which is the moment generating function of the standard normal distribution $N_d(0, I_d)$. Here and in the sequel, $\|\cdot\|$ stands for the Euclidean norm on $\mathbb{R}^d$, and I_d is the unit matrix of order d.

The statistic proposed in this paper is the weighted L^2-statistic

$$T_{n,\beta} = n \int_{\mathbb{R}^d} \left(M_n(t) - m(t)\right)^2 w_\beta(t) \, \mathrm{d}t, \tag{1.2}$$

where

$$w_\beta(t) = \exp\left(-\beta \|t\|^2\right), \tag{1.3}$$

and $\beta > 1$ is some fixed parameter, the role of which will be discussed later. Notice that $T_{n,\beta}$ is the 'moment generating function analogue' to the BHEP-statistics for testing for multivariate normality (see, e.g., Baringhaus and Henze (1988), Henze and Zirkler (1990), Henze and Wagner (1997)). The latter statistics originate if one replaces $M_n(t)$ with the empirical *characteristic* function of the scaled residuals and $m(t)$ with the characteristic function $\exp(-\|t\|^2 / 2)$ of the standard normal distribution $N_d(0, I_d)$. For a general account on weighted L^2-statistics see, e.g., Baringhaus et al. (2017).

In principle, one could replace w_β in (1.3) 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, since straightforward calculations yield the representation

$$T_{n,\beta} = \pi^{d/2} \left(\frac{1}{n} \sum_{i,j=1}^{n} \frac{1}{\beta^{d/2}} \exp\left(\frac{\|Y_{n,i} + Y_{n,j}\|^2}{4\beta}\right) + \frac{n}{(\beta - 1)^{d/2}}\right. \tag{1.4}$$
$$\left. -2 \sum_{j=1}^{n} \frac{1}{(\beta - 1/2)^{d/2}} \exp\left(\frac{\|Y_{n,j}\|^2}{4\beta - 2}\right)\right),$$

3

which is amenable to computational purposes. Notice that the condition $\beta > 1$ is necessary for the integral in (1.2) to be finite. Later, we have to impose the further restriction $\beta > 2$ to prove that $T_{n\beta}$ has a non-degenerate limit null distribution as $n \to \infty$. We remark that $T_{n,\beta}$ is affine invariant since it only depends on the Mahanalobis angles and distances $Y_{n,i}^\top Y_{n,j}$, $1 \le i, j \le n$. Rejection of H_0 is for large values of $T_{n,\beta}$.

The rest of the paper unfolds as follows. The next section shows that letting β tend to infinity in (1.2) yields a linear combination of two well-known measures of multivariate skewness. In Section 3 we derive the limit null distribution of $T_{n,\beta}$ in the i.i.d. setting. Section 4 addresses the question of consistency of the new tests against general alternatives, while Section 5 considers the new criterion in the context of multivariate GARCH models in order to test for normality of innovations, and it provides the pertaining large-sample theory. Section 6 presents a Monte Carlo study that compares the new tests with competing ones, and it considers a real data set from the financial market. The article concludes with discussions in Section 7.

2. The case $\beta \to \infty$

In this section, we show that the statistic $T_{n,\beta}$, after a suitable scaling, approaches a linear combination of two well-known measures of multivariate skewness as $\beta \to \infty$.

Theorem 2.1 *We have*

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

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 Mardia (1970) and Móri, Rohatgi and Székely (1993), respectively.

PROOF. Let $b_{2,d} = n^{-1} \sum_{j=1}^{n} \|Y_{n,j}\|^4$ denote multivariate sample kurtosis in the sense of Mardia (1970). From (1.4) and

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

4

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

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

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

$$\sum_{j,k=1}^{n} \|Y_{n,j} + Y_{n,k}\|^6 = 2n \sum_{j=1}^{n} \|Y_{n,j}\|^6 + 6(d+4)n^2 b_{2,d} + 8n^2 b_{1,d} + 12n^2 \widetilde{b}_{1,d}.$$

For the derivation of the second but last expression, see the proof of Theorem 4.1 of Henze et al. (2017). We stress that although $b_{2,d}$ and $\sum_{j=1}^{n} \|Y_{n,j}\|^6$ show up in some of the equations above, these terms cancel out in the derivation of the final result. ∎

Remark 2.2 Interestingly, $T_{n,\beta}$ exhibits the same limit behavior as $\beta \to \infty$ as both the statistic studied by Henze et al. (2017), which is based on a weighted L^2-distance involving both the empirical characteristic function and the empirical moment generating function, and the BHEP-statistic for testing for multivariate normality, which is based on the empirical characteristic function, see Theorem 2.1 of Henze (1997). At first sight, Theorem 2.1 seems to differ from Theorem 4 of Henze and Koch (2017) which covers the special case $d = 1$, but a careful analysis shows that – with the notation $\tau(\beta)$ in that paper – we have $\lim_{\beta \to \infty} \beta^{7/2} \tau(\beta) = 0$.

3. 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. The key observation for deriving the limit distribution of $T_{n,\beta}$ is the fact that

$$T_{n,\beta} = \int_{\mathbb{R}^d} W_n(t)^2 \, w_\beta(t) \, \mathrm{d}t,$$

where

$$W_n(t) = \sqrt{n} \left(M_n(t) - m(t) \right), \quad t \in \mathbb{R}^d, \tag{3.1}$$

with $M_n(t)$ given in (1.1). Notice that W_n is a random element of the Hilbert space

$$\mathrm{L}_\beta^2 := \mathrm{L}^2(\mathbb{R}^d, \mathcal{B}^d, w_\beta(t)\mathrm{d}t) \tag{3.2}$$

5

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_β defined in (1.3). The resulting norm in L_β^2 will be denoted by $\|f\|_{\mathrm{L}_\beta^2} = \sqrt{\langle f, f \rangle}$. With this notation, $T_{n,\beta}$ takes the form

$$T_{n,\beta} = \|W_n\|_{\mathrm{L}_\beta^2}^2. \tag{3.3}$$

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

Theorem 3.1 (*Convergence of W_n under H_0*)
Suppose that X has some non-degenerate d-variate normal distribution, and that $\beta > 2$ in (1.3). Then there is a centred Gaussian random element W of L^2 having covariance kernel

$$C(s, t) = \exp\left(\frac{\|s\|^2 + \|t\|^2}{2}\right)\left(e^{s^\top t} - 1 - s^\top t - \frac{\left(s^\top t\right)^2}{2}\right), \quad s, t \in \mathbb{R}^d,$$

so that $W_n \overset{\mathcal{D}}{\longrightarrow} W$ as $n \to \infty$.

In view of (3.3), the Continuous Mapping Theorem yields the following result.

Corollary 3.2 If $\beta > 2$, then, under the null hypothesis H_0,

$$T_{n,\beta} \overset{\mathcal{D}}{\longrightarrow} \|W\|_{\mathrm{L}_\beta^2}^2 \quad \text{as } n \to \infty.$$

Remark 3.3 The distribution of $T_{\infty,\beta} := \|W\|_{\mathrm{L}_\beta^2}^2$ (say) 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_β^2 associated with the kernel C given in Theorem 3.1, i.e., $(Af)(t) = \int C(s, t) f(s) \exp(-\beta\|s\|^2) \mathrm{d}s$, 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

$$\mathbb{E}(T_{\infty,\beta}) = \int_{\mathbb{R}^d} C(t, t) \, w_\beta(t) \, \mathrm{d}t,$$

$$\mathbb{V}(T_{\infty,\beta}) = 2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C^2(s, t) w_\beta(s) w_\beta(t) \, \mathrm{d}s \mathrm{d}t$$

(see Shorack and Wellner, 1986, p. 213), tedious but straighforward manipulations of integrals yield the following result, which generalizes Theorem 2 of Henze and Koch (2017).

Theorem 3.4 *If $\beta > 2$ we have*

a)

$$\mathbb{E}(T_{\infty,\beta}) = \pi^{d/2} \left(\frac{1}{(\beta-2)^{d/2}} - \frac{1}{(\beta-1)^{d/2}} - \frac{d}{2(\beta-1)^{d/2+1}} - \frac{d(d+2)}{8(\beta-1)^{d/2+2}} \right),$$

b)

$$
\mathbb{V}(T_{\infty,\beta}) = 2\pi^d \left(\frac{1}{(\beta(\beta-2))^{d/2}} - \frac{2^{d+1}}{\eta^{d/2}} - \frac{(1+2d)2^d}{\eta^{d/2+1}} - \frac{d(d+2)2^d}{\eta^{d/2+2}} \right.
$$
$$
\left. + \frac{1}{(\beta-1)^d} + \frac{d}{2(\beta-1)^{d+2}} + \frac{3d(d+2)}{64(\beta-1)^{d+4}} \right),
$$

where $\eta = 4(\beta-1)^2 - 1$.

PROOF of Theorem 3.1. In view of affine invariance, we assume w.l.o.g. that the distribution of X is $\mathrm{N}_d(0, \mathrm{I}_d)$. In Henze et al. (2017), the authors considered the "exponentially down-weighted empirical moment generating function process"

$$A_n(t) = \exp\left(-\frac{\|t\|^2}{2}\right) M_n(t), \quad t \in \mathbb{R}^d. \tag{3.4}$$

Notice that, with the notation given in (3.2), we have

$$\|A_n\|_{\mathrm{L}_\beta^2}^2 = \|M_n\|_{\mathrm{L}_\gamma^2}^2,$$

where $\gamma = \beta - 1$ From display (10.5) and Propositions 10.3 and 10.4 of Henze et al. (2017) we have

$$A_n(t) = \exp\left(-\frac{\|t\|^2}{2}\right) \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \mathrm{e}^{t^\top X_j} - m(t) \right) + V_n(t) + R_n(t),$$

where $\int_{\mathbb{R}^d} R_n^2(t) w_\gamma(t) \mathrm{d}t = o_{\mathbb{P}}(1)$, and

$$V_n(t) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^{n} \left((t^\top X_j)^2 - \|t\|^2 \right) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} t^\top X_j.$$

Display (3.4) and the representation of A_n as a sum yield

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j(t) + m(t) R_n(t),$$

7

where

$$Z_j(t) = e^{t^\top X_j} - m(t) - \frac{m(t)}{2} \left((t^\top X_j)^2 - \|t\|^2 \right) - m(t) t^\top X_j.$$

Notice that $Z_1, Z_2, \ldots$ are i.i.d. centred random elements of L_β^2. Since

$$\int_{\mathbb{R}^d} (m(t) R_n(t))^2 w_\beta(t) \, dt = \int_{\mathbb{R}^d} R_n^2(t) w_\gamma(t) \, dt = o_{\mathbb{P}}(1),$$

a Central Limit Theorem in Hilbert spaces (see e.g., Bosq, 2000) shows that there is a centered Gaussian random element W of L_β^2, so that $W_n \xrightarrow{\mathcal{D}} W$. Using the fact that $t^\top X$ has the normal distribution $N(0, \|t\|^2)$ and the relations

$$
\begin{aligned}
\mathbb{E}\left[e^{s^\top X} (t^\top X)^2 \right] &= m(s) \left((s^\top t)^2 + \|t\|^2 \right), \\
\mathbb{E}\left[e^{s^\top X} t^\top X \right] &= m(s) s^\top t, \\
\mathbb{E}\left[(s^\top X)^2 (t^\top X)^2 \right] &= 2(s^\top t)^2 + \|s\|^2 \|t\|^2,
\end{aligned}
$$

some straightforward algebra shows that the covariance kernel $C(s,t)$ figuring in the statement of Theorem 3.1 equals $\mathbb{E} Z_1(s) Z_1(t)$. ∎

4. Consistency

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

Theorem 4.1 *Suppose X has some absolutely continuous distribution, and that $M_X(t) = \mathbb{E}[\exp(t^\top X)] < \infty$, $t \in \mathbb{R}^d$. Furthermore, let $\widetilde{X} = \Sigma^{-1/2}(X - \mu)$, where $\mu = \mathbb{E}(X)$ and $\Sigma^{-1/2}$ is the symmetric square root of the inverse of the covariance matrix Σ of X. Letting $M_{\widetilde{X}}(t) = \mathbb{E}[\exp(t^\top \widetilde{X})]$, we have*

$$\liminf_{n \to \infty} \frac{T_{n,\beta}}{n} \geq \int_{\mathbb{R}^d} \left(M_{\widetilde{X}}(t) - m(t) \right)^2 w_\beta(t) \, dt$$

almost surely.

PROOF. Because of affine invariance we may w.l.o.g. assume $\mathbb{E} X = 0$ and $\Sigma = I_d$. Fix $K > 0$ and put $M_n^\circ(t) = n^{-1} \sum_{j=1}^n \exp(t^\top X_j)$. From the proof of Theorem 6.1 of Henze et al. (2017) we have

$$\lim_{n \to \infty} \max_{\|t\| \leq K} \left| M_n(t) - M_n^\circ(t) \right| = 0$$

$\mathbb{P}$-almost surely. Now, the strong law of large numbers in the Banach space of continuous functions on $B(K) := \{t \in \mathbb{R}^d : \|t\| \leq K\}$ and Fatou's lemma yield

$$
\begin{aligned}
\liminf_{n \to \infty} \frac{T_{n,\beta}}{n} &\geq \liminf_{n \to \infty} \int_{B(K)} (M_n(t) - m(t))^2 \, w_\beta(t) \, \mathrm{d}t \\
&\geq \int_{B(K)} \left(\mathbb{E}\mathrm{e}^{t^\top X} - m(t) \right)^2 w_\beta(t) \, \mathrm{d}t
\end{aligned}
$$

$\mathbb{P}$-almost surely. Since K is arbitrary, the assertion follows. ∎

Now, suppose that X has an alternative distribution (which is assumed to be standardized) satisfying the conditions of Theorem 4.1. Since $\mathbb{E}\exp(t^\top X) - m(t) \neq 0$ for at least one t, Theorem 4.1 shows that $\lim_{n \to \infty} T_{n,\beta} = \infty$ $\mathbb{P}$-almost surely. Since, for any given nominal level $\alpha \in (0, 1)$, the sequence of critical values of a level-α-test based on $T_{n\beta}$ that rejects H_0 for large values of $T_{n,\beta}$ converges according to Theorem 3.1, this test is consistent against such an alternative. It should be 'all the more consistent' against any distribution not satisfying the conditions of Theorem 4.1 but, in view of the reasoning given in Csörgő (1986), the behavior of $T_{n,\beta}$ against such alternatives is a difficult problem.

5. Testing for normality in GARCH models

In this section we consider the multivariate GARCH (MGARCH) model

$$
X_j = \Sigma_j^{1/2}(\theta)\varepsilon_j, \quad j \in \mathbb{Z}, \tag{5.1}
$$

where $\theta \in \Theta \subseteq \mathbb{R}^v$ is a v-dimensional vector of unknown parameters. The unobservable random errors or innovations $\{\varepsilon_j, \ j \in \mathbb{Z}\}$ are i.i.d. copies of a d-dimensional random vector ε, which is assumed to have mean zero and unit covariance matrix. Hence

$$
\Sigma_j(\theta) = \Sigma(\theta; X_{j-1}, X_{j-2}, \ldots)
$$

is the conditional variance of X_j, given $X_{j-1}, X_{j-2}, \ldots$. The explicit expression of $\Sigma_j(\theta)$ depends on the assumed MGARCH model (see, e.g., Francq and Zakoïan, 2010, for a detailed description of several relevant models). The interest in testing for normality of the innovations stems from the fact that this distributional assumption is made in

9

some applications, and that, if erroneously accepted, some inferential procedures can lead to wrong conclusions (see, e.g., Spierdijk, 2016, for the effect on the assessment of standard risk measures such as the value at risk).

Therefore, an important step in the analysis of GARCH models is to check whether the data support the distributional hypotheses made on the innovations. Because of this reason, a number of goodness-of-fit tests have been proposed for the innovation distribution. The papers by Klar et al. (2012) and Ghoudi and Rémillard (2014) contain an extensive review of such tests as well as some numerical comparisons between them for the special case of testing for univariate normality. The proposals for testing goodness-of-fit in the multivariate case are rather scarce.

The class of GARCH models has been proved to be particularly valuable in modeling financial data. As discussed, among others in Rydberg (2000), one of the stylized features of financial data is that they are heavy-tailed. From an extensive simulation study (a summary is reported in Section 6), we learnt that, for i.i.d. data, the test of normality based on $T_{n,\beta}$ exhibits a high power against heavy-tailed distributions. Because of these reasons, this section is devoted to adapt that procedure to testing for normality of the innovations based on data $X_1, \ldots, X_n$ that are driven by equation (5.1). Therefore, on the basis of the observations, we wish to test the null hypothesis

$$H_{0,G} : \text{ The law of } \varepsilon \text{ is } \mathrm{N}_d(0, \mathrm{I}_d).$$

against general alternatives. 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 $\mathrm{N}_d(0, \Sigma_j(\theta))$, for some $\theta \in \Theta$. Two main differences with respect to the i.i.d. case are: (a) the innovations in (5.1) are assumed to be centered at zero with unit covariance matrix; and (b) the conditional covariance matrix $\Sigma_j(\theta)$ of X_j is time-varying in a way that depends on the unknown parameter θ and on past observations.

Notice that although $H_{0,G}$ is about the distribution of ε, the innovations are unobservable in the context of model (5.1). Hence any inference on the distribution of the innovations should be based on residuals

$$\widetilde{\varepsilon}_j(\widehat{\theta}_n) = \widetilde{\Sigma}_j^{-1/2}(\widehat{\theta}_n)X_j, \quad 1 \le j \le n. \tag{5.2}$$

Recall that $\Sigma_j(\theta) = \Sigma(\theta; X_{j-1}, X_{j-2}, \ldots)$, but we only observe $X_1, \ldots, X_n$. Therefore,

to estimate $\Sigma_j(\theta)$, apart from a suitable estimator $\widehat{\theta}_n$ of θ, we also need to specify values for $\{X_j, \ j \leq 0\}$, say $\{\widetilde{X}_j, \ j \leq 0\}$. So we write $\widetilde{\Sigma}_j(\theta)$ for $\Sigma(\theta; X_{j-1}, \ldots, X_1, \widetilde{X}_0, \widetilde{X}_{-1} \ldots)$. Under certain conditions, these arbitrarily fixed initial values are asymptotically irrelevant.

Taking into account that the innovations have mean zero and unit covvariance matrix, we will work directly with the residuals, without standardizing them. Let M_n^G be defined as M_n in (1.1) by replacing $Y_{n,j}$ with $\widetilde{\varepsilon}_j(\widehat{\theta}_n)$, $1 \leq j \leq n$, and define $T_{n,\beta}^G$ as $T_{n,\beta}$ in (3.3) with W_n changed for W_n^G, where W_n^G is defined as W_n in (3.1) with M_n replaced by M_n^G. In order to derive the asymptotic null distribution of W_n^G we will make the assumptions (A.1)–(A.6) below. In the sequel, $C > 0$ and ϱ, $0 < \varrho < 1$, denote generic constants, the values of which may vary across the text, θ_0 stands for the true value of θ, and for any matrix $A = (a_{kj})$, $\|A\| = \sum_{k,j} |a_{kj}|$ denotes the l_1-norm (we use the same notation as for the Euclidean norm of vectors).

(A.1) The estimator $\widehat{\theta}_n$ satisfies $\sqrt{n}(\widehat{\theta}_n - \theta_0) = n^{-1/2} \sum_{j=1}^n L_j + o_{\mathbb{P}}(1)$, where $L_j = h_j g_j$, $g_j = g(\theta_0; \varepsilon_j)$ is a vector of d^2 measurable functions such that $\mathbb{E}(g_j) = 0$ and $\mathbb{E}(g_j^\top g_j) < \infty$, and $h_j = h(\theta_0; \varepsilon_{j-1}, \varepsilon_{j-2} \ldots)$ is a $v \times d^2$-matrix of measurable functions satisfying $\mathbb{E}(\|h_j h_j^\top\|^2) < \infty$,

(A.2) $\sup_{\theta \in \Theta} \left\| \widetilde{\Sigma}_j^{-1/2}(\theta) \right\| \leq C$, $\quad \sup_{\theta \in \Theta} \left\| \Sigma_j^{-1/2}(\theta) \right\| \leq C$ $\quad \mathbb{P}$-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) $\mathbb{E} \|X_j\|^\varsigma < \infty$ and $\mathbb{E} \left\| \Sigma_j^{1/2}(\theta_0) \right\|^\varsigma < \infty$ for some $\varsigma > 0$,

(A.5) for any sequence $x_1, x_2, \ldots$ of vectors of $\mathbb{R}^d$, the function $\theta \mapsto \Sigma^{1/2}(\theta; x_1, x_2, \ldots)$ 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 W_n^G.

Theorem 5.1 (*Convergence of W_n^G under $H_{0,G}$*)
Let $\{X_j\}$ be a strictly stationary process satisfying (5.1), 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 and that $\beta > 2$. Then under the null hypothesis $H_{0,G}$, there is a centered Gaussian random element W_G of L_β^2, having covariance kernel $C_G(s,t) = cov(U(t), U(s))$, so that $W_n^G \xrightarrow{\mathcal{D}} W_G$ as $n \to \infty$, where

$$U(t) = \exp(t^\top \varepsilon_1) - m(t) - m(t) a(t)^\top L_1,$$

$a(t)^\top = (t^\top \mu_1 t, \dots, t^\top \mu_v t)$, $\mu_k = \mathbb{E}[A_{1k}(\theta_0)]$, $A_{1k}(\theta) = \Sigma_1^{-1/2}(\theta) \frac{\partial}{\partial \theta_k} \Sigma_1^{1/2}(\theta)$, $1 \leq k \leq v$.

From Theorem 5.1 and the Continuous Mapping Theorem we have the following corollary.

Corollary 5.2 Under the assumptions of Theorem 5.1, we have

$$T_{n,\beta}^G \xrightarrow{\mathcal{D}} \|W_G\|_{L_\beta^2}^2 \quad \text{as } n \to \infty.$$

The standard estimation method for the parameter θ in GARCH models is the quasi maximum likelihood estimator (QMLE), defined as

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

where

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

Comte and Leiberman (2003) and Bardet and Wintenberger (2009), among others, have shown that under certain mild regularity conditions the QMLE satisfies (A.1) for general MGARCH models.

As observed before, there are many MGARCH parametrizations for the matrix $\Sigma_j(\theta)$. Nevertheless, there exist only partial theoretical results for such models. The Constant Conditional Correlation model, proposed by Bollerslev (1990) and extended by Jeantheau (1998), is an exception, since its properties have been thoroughly studied. This model decomposes the conditional covariance matrix figuring in (5.1) into conditional standard deviations and a conditional correlation matrix, according to $\Sigma_j(\theta_0) = D_j(\theta_0)R_0D_j(\theta_0)$, where $D_j(\theta_0)$ and R_0 are $d \times d$-matrices, R_0 is a correlation matrix, and $D_j(\theta_0)$ is a diagonal matrix so that $\sigma_j^2(\theta) = \text{diag}\left\{D_j^2(\theta)\right\}$ with

$$\sigma_j^2(\theta) = b + \sum_{k=1}^{p} B_k X_{j-k}^{(2)} + \sum_{k=1}^{q} \Gamma_k \sigma_{j-k}^2(\theta). \tag{5.3}$$

Here, $X_j^{(2)} = X_j \odot X_j$, where $\odot$ denotes the Hadamard product, that is, the element by element product, b is a vector of dimension d with positive elements, and $\{B_k\}_{k=1}^{p}$ and $\{\Gamma_k\}_{k=1}^{q}$ are $d \times d$ matrices with non-negative elements. This model will be referred to as CCC-GARCH(p,q). Under certain weak assumptions, the QMLE for the parameters in this model satisfies (A.1), and (A.2)–(A.6) also hold, see Francq and Zakoïan (2010) and Francq et al. (2017).

The asymptotic null distribution of $T_{n,\beta}^G$ depends on the equation defining the GARCH model and on θ_0 through the quantities $\mu_1, \ldots, \mu_v$, as well as on which estimator of θ has been employed. Therefore, the asymptotic null distribution cannot be used to approximate the null distribution of $T_{n,\beta}^G$. Following Klar et al. (2012), we will estimate the null distribution of $T_{n,\beta}^G$ by using the following parametric bootstrap algorithm:

(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 $T_{n,\beta}^G = T_{n,\beta}^G(\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n)$.

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

(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 $T_{n,\beta}^G$ by means of the conditional distribution, given the data, of $T_{n,\beta}^{G*} = T_{n,\beta}^G(\widetilde{\varepsilon}_1^*, \ldots, \widetilde{\varepsilon}_n^*)$.

In practice, the approximation in step (iii) is carried out by generating a large number of bootstrap replications of the test statistic $T_{n,\beta}^G$, whose empirical distribution function is used to estimate the null distribution of $T_{n,\beta}^G$. Similar steps to those given in the proof of Theorem 5.1 show that if one assumes that (A.1)–(A.6) continue to hold when θ_0 is replaced by θ_n, with $\theta_n \to \theta_0$ as $n \to \infty$, and $\varepsilon \sim \mathrm{N}_d(0, \mathrm{I}_d)$, then the conditional distribution of $T_{n,\beta}^{G*}$, given the data, converges in law to $\|W_G\|_{\mathrm{L}_\beta^2}^2$, with W_G as defined in Theorem 5.1. Therefore, the above bootstrap procedure provides a consistent null distribution estimator.

Remark 5.3 The practical application of the above bootstrap null distribution estimator entails that the parameter estimator of θ and the residuals must be calculated for each bootstrap resample, which results in a time-consuming procedure. Following the approaches in Ghoudi and Rémillard (2014) and Jiménez-Gamero and Pardo-Fernández (2017) for other goodness-of-fit tests for univariate GARCH models, we could use a weighted bootstrap null distribution estimator in the sense of Burke (2000). From a computational point of view, it provides a more efficient estimator. Nevertheless, it can be verified that the consistency of the weighted bootstrap null distribution estimator of $T_{n,\beta}^G$ requires the existence of the moment generating function of the true distribution generating the innovations, which is a rather strong condition, specially taking into account that the alternatives of interest are heavy-tailed.

As in the i.i.d. case, the next result shows that the test for multivariate normality based on $T_{n,\beta}^G$ is consistent against general alternatives.

Theorem 5.4 *Let $\{X_j\}$ be a strictly stationary process satisfying (5.1), 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, that ε has some absolutely continuous distribution, and that $M_\varepsilon(t) = \mathbb{E}[\exp(t^\top \varepsilon)] < \infty$, $t \in \mathbb{R}^d$. We then have*

$$\liminf_{n \to \infty} \frac{T_{n,\beta}^G}{n} \geq \int_{\mathbb{R}^d} (M_\varepsilon(t) - m(t))^2\, w_\beta(t)\, \mathrm{d}t$$

in probability.

Similar comments to those made after Theorem 4.1 for the i.i.d. case can be done in this setting.

PROOF of Theorem 5.1. From the proof of Theorem 7.1 in Henze et al. (2017), it follows that $W_n^G(t) = W_{1,n}^G(t) + r_{n,1}(t)$, with $W_{1,n}^G(t) = n^{-1/2} \sum_{j=1}^n V_j(t)$,

$$V_j(t) = \exp(t^\top \varepsilon_j) - m(t) a(t)^\top \sqrt{n}(\widehat{\theta}_n - \theta_0) - m(t),$$

and $\|r_{n,1}\|_{L_\beta^2} = o_\mathbb{P}(1)$. By Assumption A.1, $W_{1,n}^G(t) = W_{2,n}^G(t) + r_{n,2}(t)$, with $W_{2,n}^G(t) = n^{-1/2} \sum_{j=1}^n U_j(t)$,

$$U_j(t) = \exp(t^\top \varepsilon_j) - \exp(-\|t\|^2/2) a(t)^\top L_j - \exp(-\|t\|^2/2),$$

and $\|r_{n,2}\|_{L_\beta^2} = o_\mathbb{P}(1)$.

To prove the result we will apply Theorem 4.2 in Billingsley (1968) to $\{W_{2,n}^G(t), t \in \mathbb{R}^d\}$ by showing that (a) for each positive M, $\{W_{2,n}^G(t), t \in B(K)\}$ converges in law to $\{W^G(t), t \in B(K)\}$ in $C(B(K))$, the Banach space of real-valued continuous functions on $B(K) := \{t \in \mathbb{R}^d : \|t\| \leq K\}$, endowed with the supremum norm; (b) for each $\varepsilon > 0$, there is a positive K so that

$$\int_{\mathbb{R}^d \setminus B(K)} \mathbb{E}\left[W_{2,n}^G(t)^2 \right] w_\beta(t)\, \mathrm{d}t < \varepsilon, \tag{5.4}$$

$$\int_{\mathbb{R}^d \setminus B(K)} \mathbb{E}\left[W^G(t)^2 \right] w_\beta(t)\, \mathrm{d}t < \varepsilon. \tag{5.5}$$

Proof of (a): By applying the central limit theorem for martingale differences, the finite-dimensional distributions of $\{W_{2,n}^G(t), t \in \mathbb{R}^d\}$ converge to those of $\{W_G(t), t \in \mathbb{R}^d\}$. Hence, to prove (a) we must show that $\{W_{2,n}^G(t), t \in B(K)\}$ is tight. With this aim we write $W_{2,n}^G(t) = W_{3,n}^G(t) - W_{4,n}^G(t)$, with $W_{3,n}^G(t) = n^{-1/2} \sum_{j=1}^n \{\exp(t^\top \varepsilon_j) - m(t)\}$ and $W_{4,n}^G(t) = m(t) a(t)^\top n^{-1/2} \sum_{j=1}^n L_j$. The mean value theorem gives

$$\mathbb{E}\left[\{\exp(t^\top \varepsilon) - \exp(s^\top \varepsilon)\}^2 \right] \leq \kappa \|t - s\|^2, \quad s, t \in B(K),$$

for some positive κ. From Theorem 12.3 in Billingsley (1968), the process $\{W_{3,n}^G(t), t \in B(K)\}$ is tight. By the central limit theorem for martingale differences, $n^{-1/2} \sum_{j=1}^n L_j$ converges in law to a v-variate zero mean normal random vector. Hence $\{W_{4,n}^G(t), t \in B(K)\}$, being a product of a continuous function and a term which is $O_\mathbb{P}(1)$, is tight, and the same property holds for $\{W_{2,n}^G(t), t \in B(K)\}$.

Proof of (b): In view of $\mathbb{E}\left[W_{2,n}^G(t)^2\right] = \mathbb{E}\left[U_1(t)^2\right] < \infty$, for each $\varepsilon > 0$ there is some positive constant K so that (5.4) holds. Likewise, (5.5) holds, which completes the proof. $\blacksquare$

PROOF of Theorem 5.4. Let $\varepsilon_j(\theta) = \Sigma_j^{-1/2}(\theta)X_j$. Notice that $\varepsilon_j(\theta_0) = \varepsilon_j$. Let $\widetilde{M}_n^G(t) = n^{-1}\sum_{j=1}^n \exp\{t^\top \widetilde{\varepsilon}_j(\hat{\theta}_n)\}$, $\widehat{M}_n^G(t) = n^{-1}\sum_{j=1}^n \exp\{t^\top \varepsilon_j(\hat{\theta}_n)\}$, $M_n^\circ(t) = n^{-1}\sum_{j=1}^n \exp\{t^\top \varepsilon_j\}$ and $B(K) := \{t \in \mathbb{R}^d : \|t\| \le K\}$. To show the result we will prove

(a) $\sup_{t\in B(K)} |\widehat{M}_n^G(t) - M_n^\circ(t)| = o_\mathbb{P}(1)$,

(b) $\sup_{t\in B(K)} |\widetilde{M}_n^G(t) - \widehat{M}_n^G(t)| = o_\mathbb{P}(1)$,

and the result will follow by using the same proof as in the i.i.d. case.

Proof of (a): Let $\hat{\theta}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nv})^\top$, $\theta_0 = (\theta_{01}, \ldots, \theta_{0v})^\top$ and $A_{jk}(\theta) = \Sigma_j^{-1/2}(\theta)\frac{\partial}{\partial\theta_k}\Sigma_j^{1/2}(\theta)$. We have $\varepsilon_j(\hat{\theta}_n) = \varepsilon_j + \Delta_{n,j}$, with $\Delta_{n,j} = -\sum_{k=1}^v A_{jk}(\widetilde{\theta}_{n,j})\varepsilon_j(\hat{\theta}_{nk} - \theta_{0k})$, for some $\widetilde{\theta}_{n,j}$ between $\hat{\theta}_n$ and θ_0. 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 $\alpha_{n,j} \in (0,1)$. Now (A.1) and (A.6) yield $\|\Delta_{n,j}\| \le D_j\|\varepsilon_j\|\|\hat{\theta}_n - \theta_0\|$ for large enough n, where $\mathbb{E}(D_j^2) < \infty$. The Cauchy–Schwarz inequality gives

$$|\widehat{M}_n^G(t) - M_n^\circ(t)| = \left|\frac{1}{n}\sum_{j=1}^n \exp(t^\top \varepsilon_j)\left\{\exp(t^\top \Delta_{n,j}) - 1\right\}\right| \le r_{1,n}(t)^{1/2}r_{2,n}(t)^{1/2},$$

where $r_{1,n}(t) = M_n(2t)$, and

$$r_{2,n}(t) = \|t\|^2\|\hat{\theta}_n - \theta_0\|^2\exp\left\{2\|t\|\|\hat{\theta}_n - \theta_0\|\max_{1\le j\le n}D_j\|\varepsilon_j\|\right\}\frac{1}{n}\sum_{j=1}^n D_j^2\|\varepsilon_j\|^2.$$

From the strong law of large numbers in the Banach space of continuous functions on $B(K)$, we have

$$\sup_{t\in B(K)} r_{1,n}(t) \le \sup_{t\in B(K)} M_\varepsilon(2t) + \sup_{t\in B(K)} |M_n^G(2t) + M_\varepsilon(2t)| < K_1 \quad \mathbb{P}\text{-a.s.}$$

for some positive constant K_1. From the ergodic theorem, $n^{-1}\sum_{j=1}^n D_j^2\|\varepsilon_j\|^2 < K_2$ $\mathbb{P}$-almost surely for some positive constant K_2. Using stationarity and finite second moments, if follows that $\max_{1\le j\le n} D_j\|\varepsilon_j\|/\sqrt{n} \to 0$, $\mathbb{P}$-almost surely. Hence (A.1) yields $\sup_{t\in B(K)} r_{2,n}(t) \to 0$, in probability. This concludes the proof of (a).

Proof of (b): The reasoning follows similar steps as the proof of fact (c.1) in the proof of Theorem 7.1 in Henze et al. (2017) and is thus omitted. $\blacksquare$

6. Monte Carlo results

This section describes and summarizes the results of an extensive simulation experiment carried out to study the finite-sample performance of the proposed tests. Moreover, we consider a real data set of monthly log returns. All computations have been performed using programs written in the R language.

6.1. Numerical experiments for i.i.d. data

Upper quantiles of the null distribution of $T_{n,\beta}$ have been approximated by generating 100,000 samples from a law $N_d(0, I_d)$. Table 1 displays some critical values with the convention that an entry like $^{-4}1.17$ stands for 1.17×10^{-4}. The results show that large sample sizes are required to approximate the critical values by their corresponding asymptotic values.

A natural competitor of the test based on $T_{n,\beta}$ is the CF-based test studied in Henze and Wagner (1997) (HW-test). The latter procedure is simple to compute as well as affine invariant, and it has revealed good power performance with regard to competitors. The behaviour of the test based on $T_{n,\beta}$ in relation to the HW-test depends on whether the distribution is heavy-tailed or not. We tried a number of non-heavy-tailed distributions (specifically, the multivariate Laplace distribution, finite mixtures of normal distributions, the skew-normal distribution, the multivariate χ^2-distribution, the Khintchine distribution, the uniform distribution on $[0,1]^d$ and the Pearson type II family). For these distributions we observed that the power of the proposed test is either similar or smaller than that of the HW-test; for very heavy-tailed distributions, 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 10,000 samples (in each case), for the significance level $\alpha = 0.05$, from the following heavy-tailed alternatives: (ASE_θ) the θ-stable and elliptically-contoured distribution and the (T_θ) multivariate Student's t with θ degrees of freedom. The same fact was also observed in Zghoul (2010), who numerically studied the test based on $T_{n,\beta}$ for univariate data.

In our simulations we tried a large number of values for β for the proposed test as well as for the HW-test. The tables display the results for those values of β giving the highest power in most of the cases considered. The same can be said for the simulations

17

Table 1: Critical points for $\pi^{-d/2}T_{n,\beta}$.

			β						
d	n	α	2.5	3.0	3.5	4.0	5.0	6.0	10.0
2	20	0.05	0.213	$^{-1}0.751$	$^{-2}3.269$	$^{-2}1.639$	$^{-3}5.408$	$^{-3}2.266$	$^{-4}2.241$
		0.10	0.339	$^{-1}1.147$	$^{-2}4.857$	$^{-2}2.380$	$^{-3}7.638$	$^{-3}3.150$	$^{-4}3.025$
	50	0.05	0.391	$^{-1}1.246$	$^{-2}5.098$	$^{-2}2.436$	$^{-3}7.594$	$^{-3}3.078$	$^{-4}2.875$
		0.10	0.661	$^{-1}1.997$	$^{-2}7.802$	$^{-2}3.624$	$^{-3}10.917$	$^{-3}4.330$	$^{-4}3.897$
	100	0.05	0.511	$^{-1}1.539$	$^{-2}6.073$	$^{-2}2.838$	$^{-3}8.620$	$^{-3}3.429$	$^{-4}3.111$
		0.10	0.868	$^{-1}2.432$	$^{-2}9.168$	$^{-2}4.153$	$^{-3}12.094$	$^{-3}4.724$	$^{-4}4.143$
	200	0.05	0.612	$^{-1}1.757$	$^{-2}6.719$	$^{-2}3.085$	$^{-3}9.181$	$^{-3}3.616$	$^{-4}3.232$
		0.10	1.028	$^{-1}2.726$	$^{-2}9.908$	$^{-2}4.382$	$^{-3}12.528$	$^{-3}4.845$	$^{-4}4.221$
	300	0.05	0.679	$^{-1}1.894$	$^{-2}7.114$	$^{-2}3.223$	$^{-3}9.466$	$^{-3}3.719$	$^{-4}3.296$
		0.10	1.132	$^{-1}2.878$	$^{-2}10.259$	$^{-2}4.518$	$^{-3}12.748$	$^{-3}4.905$	$^{-4}4.232$
	400	0.05	0.701	$^{-1}1.925$	$^{-2}7.165$	$^{-2}3.248$	$^{-3}9.502$	$^{-3}3.721$	$^{-4}3.283$
		0.10	1.148	$^{-1}2.868$	$^{-2}10.084$	$^{-2}4.417$	$^{-3}12.521$	$^{-3}4.843$	$^{-4}4.187$
3	20	0.05	0.356	$^{-1}1.066$	$^{-2}4.095$	$^{-2}1.851$	$^{-3}5.218$	$^{-3}1.942$	$^{-4}1.413$
		0.10	0.520	$^{-1}1.504$	$^{-2}5.629$	$^{-2}2.503$	$^{-3}6.886$	$^{-3}2.518$	$^{-4}1.773$
	50	0.05	0.719	$^{-1}1.906$	$^{-2}6.760$	$^{-2}2.894$	$^{-3}7.598$	$^{-3}2.709$	$^{-4}1.828$
		0.10	1.153	$^{-1}2.879$	$^{-2}9.789$	$^{-2}4.073$	$^{-3}10.317$	$^{-3}3.593$	$^{-4}2.334$
	100	0.05	0.988	$^{-1}2.433$	$^{-2}8.258$	$^{-2}3.426$	$^{-3}8.696$	$^{-3}3.043$	$^{-4}1.992$
		0.10	1.646	$^{-1}3.732$	$^{-2}11.943$	$^{-2}4.788$	$^{-3}11.572$	$^{-3}3.945$	$^{-4}2.489$
	200	0.05	1.232	$^{-1}2.851$	$^{-2}9.322$	$^{-2}3.781$	$^{-3}9.365$	$^{-3}3.231$	$^{-4}2.078$
		0.10	2.046	$^{-1}4.319$	$^{-2}13.243$	$^{-2}5.167$	$^{-3}12.210$	$^{-3}4.123$	$^{-4}2.567$
	300	0.05	1.332	$^{-1}2.979$	$^{-2}9.555$	$^{-2}3.849$	$^{-3}9.431$	$^{-3}3.242$	$^{-4}2.072$
		0.10	2.187	$^{-1}4.441$	$^{-2}13.364$	$^{-2}5.156$	$^{-3}12.105$	$^{-3}4.073$	$^{-4}2.527$
	400	0.05	1.397	$^{-1}3.061$	$^{-2}9.725$	$^{-2}3.893$	$^{-3}9.509$	$^{-3}3.260$	$^{-4}2.084$
		0.10	2.245	$^{-1}4.481$	$^{-2}13.341$	$^{-2}5.122$	$^{-3}12.010$	$^{-3}4.046$	$^{-4}2.519$
5	20	0.05	0.597	$^{-1}1.347$	$^{-2}4.130$	$^{-2}1.554$	$^{-3}3.275$	$^{-3}0.971$	$^{-5}3.884$
		0.10	0.774	$^{-1}1.691$	$^{-2}5.089$	$^{-2}1.886$	$^{-3}3.900$	$^{-3}1.142$	$^{-5}4.474$
	50	0.05	1.519	$^{-1}2.868$	$^{-2}7.862$	$^{-2}2.731$	$^{-3}5.215$	$^{-3}1.460$	$^{-5}5.283$
		0.10	2.332	$^{-1}4.130$	$^{-2}10.801$	$^{-2}3.633$	$^{-3}6.633$	$^{-3}1.809$	$^{-5}6.260$
	100	0.05	2.315	$^{-1}3.947$	$^{-2}10.132$	$^{-2}3.381$	$^{-3}6.134$	$^{-3}1.667$	$^{-5}5.768$
		0.10	3.782	$^{-1}5.884$	$^{-2}14.199$	$^{-2}4.530$	$^{-3}7.779$	$^{-3}2.051$	$^{-5}6.782$
	200	0.05	3.047	$^{-1}4.744$	$^{-2}11.541$	$^{-2}3.736$	$^{-3}6.565$	$^{-3}1.755$	$^{-5}5.962$
		0.10	4.969	$^{-1}6.964$	$^{-2}15.880$	$^{-2}4.896$	$^{-3}8.131$	$^{-3}2.112$	$^{-5}6.875$
	300	0.05	3.346	$^{-1}5.016$	$^{-2}11.974$	$^{-2}3.829$	$^{-3}6.636$	$^{-3}1.769$	$^{-5}5.985$
		0.10	5.445	$^{-1}7.343$	$^{-2}16.307$	$^{-2}4.960$	$^{-3}8.119$	$^{-3}2.100$	$^{-5}6.832$
	400	0.05	3.608	$^{-1}5.234$	$^{-2}12.292$	$^{-2}3.889$	$^{-3}6.679$	$^{-3}1.776$	$^{-5}5.997$
		0.10	5.838	$^{-1}7.586$	$^{-2}16.477$	$^{-2}4.958$	$^{-3}8.085$	$^{-3}2.085$	$^{-5}6.821$

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

| | d | \multicolumn{6}{c}{Test based on $T_{n,\beta}$} | \multicolumn{3}{c}{HW-test} |
| | | \multicolumn{6}{c}{β} | \multicolumn{3}{c}{β} |
		3	3.5	4.0	5.0	6.0	10.0	0.1	0.5	1.0
$ASE_{1.75}$	2	72.47	72.62	72.43	72.08	71.59	70.34	67.29	67.75	59.91
	3	82.70	82.78	82.76	82.69	82.52	81.92	79.07	78.16	68.60
	5	90.51	90.86	90.95	91.25	91.32	91.05	88.89	87.46	75.71
$ASE_{1.85}$	2	54.43	54.39	54.35	53.91	53.55	52.59	50.00	48.17	39.35
	3	62.72	62.67	62.61	62.46	62.39	61.58	57.95	54.67	42.44
	5	75.31	75.52	75.65	75.96	76.03	75.63	71.81	66.66	47.82
$ASE_{1.95}$	2	24.67	24.62	24.52	24.22	24.11	23.56	22.44	20.78	15.38
	3	29.31	29.37	29.47	29.12	28.91	28.45	26.37	24.04	16.79
	5	38.28	38.39	38.37	38.27	38.00	37.55	33.99	29.35	17.39
T_5	2	58.77	58.82	58.74	58.26	57.82	56.21	51.44	54.20	47.58
	3	40.59	40.76	40.98	41.11	41.34	41.31	39.69	37.79	28.99
	5	87.14	87.80	88.43	89.17	89.59	89.76	86.36	87.21	77.92
T_7	2	42.33	42.21	42.16	41.86	41.43	39.68	36.19	36.30	28.97
	3	55.23	55.49	55.51	55.62	55.30	54.38	49.20	48.82	37.89
	5	71.51	72.50	73.17	74.07	74.47	74.30	69.35	67.75	51.33
T_{10}	2	28.80	28.82	28.73	28.29	27.99	27.12	24.48	22.94	16.23
	3	38.34	38.56	38.58	38.51	38.33	37.00	32.97	30.55	20.84
	5	51.64	52.38	53.01	53.91	54.25	54.33	48.85	45.36	28.41

in the next subsection.

6.2. Numerical experiments for GARCH data

In our simulations we considered a bivariate CCC–GARCH(1,1) model with

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

for $\gamma = 0.3, 0.4, 0.5$ and $r = 0, 0.3$, and a trivariate CCC–GARCH(1,1) model with $b = (0.1, 0.1, 0.1)'$,

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

and γ and r as before. The parameters in the CCC-GARCH models were estimated by their QMLE using the package `ccgarch` of the language R. 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 bootstrap

test. To assess the power we considered the following heavy-tailed distributions: T_θ, the multivariate β-generalized distribution (GN_θ), that coincides with the normal distribution for $\theta = 2$ and has heavy tails for $0 < \theta < 2$ (Goodman and Kotz, 1973), and the asymmetric exponential power distribution (AEP), whereby $(X_1, \ldots, X_d)^\top$, with $X_1, \ldots, X_d$ i.i.d. from a univariate AEP distribution (Zhu and Zinde-Walsh, 2009) 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.

Table 3 reports the percentages of rejections for nominal significance level $\alpha = 0.05$ and sample size $n = 300$, for $r = 0$, 0.3 and $\gamma = 0.4$. The resulting pictures for $\gamma = 0.3$, 0.5 are quite similar so, to save space, we omit the results for these values of γ. In order to reduce the computational burden we adopted the warp-speed method of Giacomini et al. (2013), which works as follows: 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 $T_{n,\beta}^G$ are computed from the empirical distribution determined by the resampling replications of $T_{n,\beta}^{G*}$. In our simulations we generated $10,000$ Monte Carlo samples for the level and $2,000$ for the power. Looking at Table 3, we conclude that: the actual level of the proposed bootstrap test is very close to the nominal level, and this is also true for the HW-test (although to the best of our knowledge, the consistency of the bootstrap null distribution estimator of the HW-test statistic has been proved only for the univariate case in Jiménez-Gamero, 2014); and with respect to the power, the proposed test in most cases outperforms the HW-test.

6.3. A real data set application

As an illustration, we consider the monthly log returns of IBM stock and the S&P 500 index from January 1926 to December 2008 with 888 observations. This data set was analyzed in Example 10.5 of Tsay(2010), where it is showed that a CCC-GARCH(1,1) model provides a adequate description of the data, which is available from the website `http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/`) of the author. We applied the proposed test and the HW test for testing $H_{0,G}$. The p-values were obtained by generating 1000 bootstrap samples. For all values of β in

Table 3: Percentage of rejections for nominal level $\alpha = 0.05$, $\gamma = 0.4$ and $n = 300$.

			Test based on $T_{n,\beta}^G$					HW-test			
			β					β			
	d	r	2.1	2.2	2.3	2.4	2.5	1.0	1.5	2.0	2.5
N	2	0.0	4.96	4.85	4.81	4.79	4.73	5.06	4.80	4.97	4.82
		0.3	4.14	4.33	4.38	4.40	4.27	4.95	5.45	5.36	5.29
	3	0.0	4.54	4.71	4.73	4.74	4.73	4.64	4.64	4.88	4.51
		0.3	4.96	4.85	4.81	4.79	4.73	5.06	4.80	4.97	4.82
T_{10}	2	0.0	61.85	61.20	59.25	57.55	55.50	26.70	36.70	37.20	34.85
		0.3	66.95	66.80	65.85	64.15	61.35	20.50	31.70	32.10	30.60
	3	0.0	81.45	80.95	80.15	79.65	78.20	45.75	55.40	50.95	43.80
		0.3	78.30	78.05	78.20	77.20	77.15	42.40	55.70	52.85	44.00
$GN_{1.65}$	2	0.0	22.40	21.05	20.10	18.95	17.85	8.65	15.20	16.45	16.75
		0.3	18.30	17.80	17.70	16.80	16.10	8.00	14.00	16.00	14.30
	3	0.0	17.55	18.40	18.10	17.80	16.90	9.10	14.85	15.35	15.60
		0.3	20.00	19.65	19.85	19.80	18.90	9.70	13.95	15.55	15.15
AEP	2	0.0	56.75	55.50	53.35	51.10	49.00	29.55	49.85	52.85	51.45
		0.3	52.70	51.20	49.65	47.90	45.75	26.35	45.20	50.00	49.20
	3	0.0	55.40	55.85	54.85	53.75	51.65	38.25	54.25	55.45	49.25
		0.3	59.55	59.30	58.75	57.15	57.00	33.15	53.65	53.90	49.70

Table 3 we get the same p-value, 0.000, which leads us to reject $H_{0,G}$, as expected by looking at Figure 1, which displays the scatter plot of the residuals after fitting a CCC-GARCH(1,1) model to the log returns, and Figure 2, that represents the histograms of the marginal residuals with the probability density function of a standard normal law superimposed.

7. Conclusions

We have studied a class of affine invariant tests for multivariate normality both in an i.i.d. setting and in the context of testing that the innovation distribution of a multivariate GARCH model is Gaussian, thus generalizing results of Henze and Koch (2017) in two ways. The test statistics are suitably weighted L^2-statistics based on the difference between the empirical moment generating function of scaled residuals of the data and the moment generating function of the standard normal distribution in $\mathbb{R}^d$. As such, they can be considered as 'moment generating function analogues' to
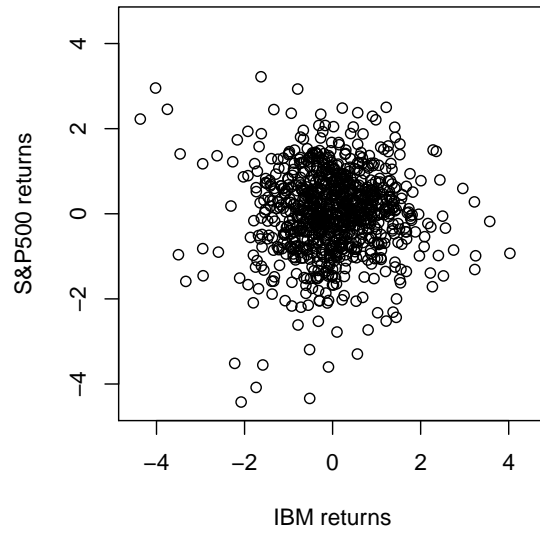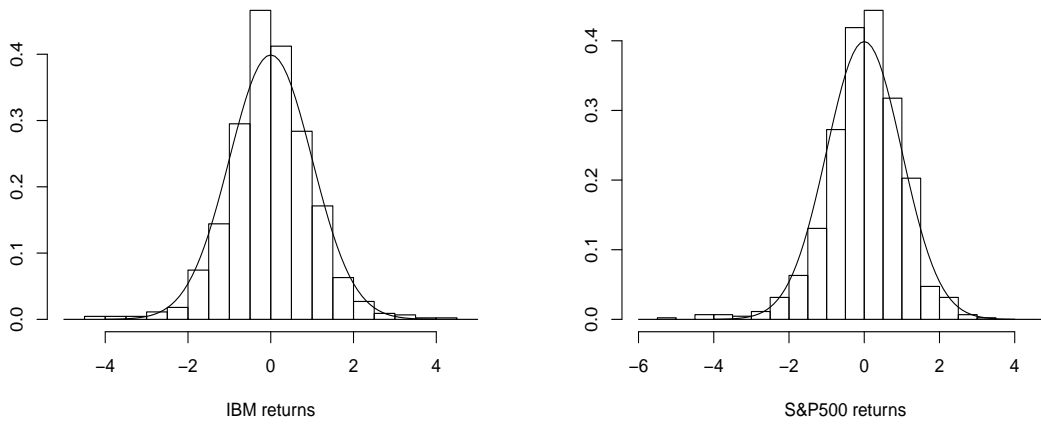
Figure 1: Scatter plot of the residuals.

Figure 2: Histograms of the residuals.

22

the time-honored class of BHEP tests that use the empirical *characteristic* function. As the decay of a weight function figuring in the test statistic tends to infinity, the test statistic approaches a certain linear combination of two well-known measures of multivariate skewness. The tests are easy to implement, and they turn out to be consistent against a wide range of alternatives. In contrast to a recently studied L^2-statistic of Henze et al. (2017) that uses both the empirical moment generating and the empirical characteristic function, our test is also feasible for larger sample sizes since the computational complexity is of order $O(n^2)$. Regarding power, the new tests outperform the BHEP-tests against heavy-tailed distributions.

Acknowledgements

References

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

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

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

Baringhaus, L., Ebner, B., and Henze, N. (2017), The limit distribution of weighted L^2-goodness-of fit statistics under fixed alternatives, with applications, *Ann. Inst. Stat. Math.*, **69**, 969–995.

Batsidis, A., Martin, N., Pardo, L., and Zografos, K. (2013), A necessary power divergence type family of tests for multivariate normality, *Comm. Statist. – Simul. Comput.*, **42**, 2253–2271.

Billingsley, P. (1968), *Convergence of Probability Measures.* John Wiley & Sons, New York-London-Sidney.

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

Bosq, D. (2000), *Linear Processes in Function Spaces*. Springer, New York.

Burke M.D. (2000), Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap, *Statist. Probab. Lett.*, **46**, 13–20.

Cardoso de Oliveira, I.R., and Ferreira, D.F. (2010), Multivariate extension of chi-squared univariate normality test, *J. Statist. Comput. Simul.*, **80**, 513–526.

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

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

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

Ebner, B. (2012), Asymptotic theory for the tests of multivariate normality by Cox and Small, *J. Multiv. Anal.*, **111**, 368–379.

Enomoto, R., Okamoto, N., and Seo, T. (2012) Multivariate normality test using Srivastava's skewness and kurtosis, *SUT J. Math.*, **48**, 103–115.

Farrel, P.J., Salibian–Barrera, M., and Naczk, K. (2007) On tests for multivariate normality and associated simulation studies, *J. Statist. Comput. Simul.*, **77**, 1053–1068.

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

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

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

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

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.

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

Hanusz, Z., and Tarasińska, J. (2008), A note on Srivastava and Hui's test of multivariate normality, *J. Multiv. Anal.*, **99**, 2364–2367.

Hanusz, Z., and Tarasińska, J. (2012). New test for multivariate normality based on Small's and Srivastava's graphical methods, *J. Statist. Comput. Simul.*, **82**, 1743–1752.

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

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

Henze, N., Jiménez-Gamero, M.D., and Meintanis, S.G. (2017), Characterizations of multinormality and and corresponding tests of fit, including for Garch models, *arXiv:1706.03029*.

Henze, N., and Koch, S. (2017), On a test of normality based on the empirical moment generating function, *Statist. Papers*, doi:10.1007/s00362-017-0923-7.

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

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

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

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

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.

Joenssen, D.W., and Vogel, J. (2014), A power study of goodness-of-fit tests for multivariate normality implemented in R, *J. Stat. Comput. Simul.*, **84**, 1055–1078.

Jönsson, K., (2011), A robust test for multivariate normality, *Econom. Lett.*, **113**, 199–201.

Kim, N. (2016), A robustified Jarque–Bera test for multivariate normality, *Econom. Lett.*, **140**, 48–52,

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.

Koizumi, K., Hyodo, M., and Pavlenko, T. (2014), Modified Jarque–Bera tests for multivariate normality in a high-dimensional framework, *J. Statist. Th. Pract.*, **8**, 382–399.

Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519–530.

Mecklin, Ch. J., and Mundfrom, D.J. (2005), A Monte Carlo comparison of the type I and type II error rates of tests of multivariate normality, *J. Statist. Comput. Simul.*, **75**, 93–107.

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

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

Rydberg, T.H. (2000), Realistic statistical modelling of financial data, *Int. Stat. Rev.*, **68**, 233–258.

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, 1986.

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.

Székeley, G.J., and Rizzo, M.L. (2005), A new test for multivariate normality, *J. Multiv. Anal.*, **93**, 58–80.

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

Tenreiro, C. (2017), A new test for multivariate normality by combining extreme and nonextreme BHEP tests, *Commun. Statist. –Th. Meth.*, **46**, 1746–1759.

Thulin, M. (2014), Tests for multivariate normality based on canonical correlations, *Stat. Meth. Appl.*, **23**, 189–208.

Tsay, R.S. (2010). Analysis of Financial Time Series. Wiley, Hoboken, New Jersey.

Villaseñor-Alva, J.A., and Estrada, E.G. (2009), A generalization of Shapiro–Wilk's test for multivariate normality, *Commun. Statist. – Th. Meth.*, **38**, 1870–1883.

Voinov, V., Pya, N., Makarov, R. and Voinov, Y. (2016), New invariant and consistent chi-squared type goodness-of-fit tests for multivariate normality and a related comparative simulation study, *Commun. Statist. – Th. Meth.* **45**, 3249–3263.

Yanada, T., Romer, M.M., and Richards, D. St. P. (2015), Kurtosis tests for multivariate normality with monotone incomplete data, *TEST* **24**, 532–557.

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

Zhou, S., and Shao, Y. (2014), A powerful test for multivariate normality, *J. Appl. Statist.*, **41**, 351–363.

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