JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **45**: 298–319 (2024)
Published online 3 August 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12713

ORIGINAL ARTICLE

GOODNESS-OF-FIT TESTS FOR THE MULTIVARIATE STUDENT-*T* DISTRIBUTION BASED ON I.I.D. DATA, AND FOR GARCH OBSERVATIONS

SIMOS MEINTANIS,[a,b] BOJANA MILOŠEVIĆ,[c] MARKO OBRADOVIĆ[c] AND MIRJANA VELJOVIĆ[c]

[a]*Department of Economics, National and Kapodistrian University of Athens, Athens, Greece*
[b]*Pure and Applied Analytics, North-West University, Potchefstroom, South Africa*
[c]*Faculty of Mathematics, University of Belgrade, Belgrade, Serbia*

We consider goodness-of-fit tests for the multivariate Student's *t*-distribution with i.i.d. data and for the innovation distribution in a generalized autoregressive conditional heteroskedasticity model. The methods are based on the empirical characteristic function and are relatively easy to implement, invariant under linear transformations, and globally consistent. Asymptotic properties of the proposed procedures are investigated, while the finite-sample properties are illustrated by means of a Monte Carlo study. The procedures are also applied to real data from the financial markets.

Received 25 July 2022; Accepted 11 July 2023

Keywords: Goodness-of-fit test; heavy-tailed distribution; empirical characteristic function; CCC-GARCH.

JEL. 62G10; 62M10; 62M10.

1. INTRODUCTION

Student-*t* distributions (StDs) are extremely appealing for financial applications, the single major reason being that stylized facts of financial assets indicate that the corresponding empirical distribution is leptokurtic relative to the normal distribution. (Note that the tails of the StD range from those of the [heavy-tailed] Cauchy distribution and up to Gaussian tails as a limiting case). In this connection, applications of StDs on financial data date back to Praetz (1972), Blattberg and Gonedes (1974), and Kon (1984), but these are for univariate i.i.d. observations. More recently however, models with time-varying volatility have been employed in finance, with prime examples being several versions of generalized autoregressive conditional heteroskedasticity (GARCH) models with increasing flexibility, and in this regard applications of StDs with such models are numerous; for reviews of multivariate GARCH models the reader is referred to Bauwens *et al.* (2006), Tsay (2006), Silvennoinen and Teräsvirta (2009), and Boudt *et al.* (2019). In most of these reviews as well as in many other studies, the multivariate StD is consistently suggested as innovation distribution; see for instance, Harvey *et al.* (1992), Pesaran and Pesaran (2007), Santos *et al.* (2013), Rossi and Spazzini (2010), Diamantopoulos and Vrontos (2010), Creal *et al.* (2011), Wang and Tsay (2013), Asai and So (2015), Dube *et al.* (2016), Zheng *et al.* (2018), Chib and Zeng (2020), Chen and Gerlach (2021) and Hafner *et al.* (2020), among others.

In the aforementioned models and besides estimation of model parameters, there is also need for proper statistical validation of the model components, including the assumption of an StD for the law of unobserved innovations. Thus specification testing for the innovation distribution is an indispensable aspect of model validation, thereby propagating the need for a goodness-of-fit method for the StD. However, apart from a few informal diagnostic

* Correspondence to: Simos Meintanis, Department of Economics, National and Kapodistrian University of Athens, 1 Sophocleous Street, 10559 Athens, Greece.
Email: simosmei@econ.uoa.gr

recipes which we will review further down the article, performing a rigorous goodness-of-fit test is a non-trivial task in the multivariate setting, even for i.i.d. data. For instance, the classical Kolmogorov–Smirnov and Cramér–von Mises tests require substantial numerical work such as numerical integration in higher dimension. Certain alternative test procedures also involve sophisticated numerical techniques; see e.g. the tests of Hallin *et al.* (2021) based on the Wasserstein distance, of Meintanis *et al.* (2014), and Liang *et al.* (2019) based on the Rosenblatt transform, and that of Khmaladze (2016), based on the celebrated transformation bearing this author's name. Other tests such as the Mahalanobis test of McAssey (2013) and the smooth test of Ducharme and de Micheaux (2020) are relatively easier to apply but lack global consistency, while the general test of Ebner *et al.* (2018) based on nearest-neighbors is for simple hypotheses without estimated parameters.

In this article we propose a goodness-of-fit test that is tailored specifically for StDs. The new test is easy to implement, invariant with respect to affine transformations, and globally consistent. The remainder of the article unfolds as follows. In Section 2 we introduce the new test and show consistency, while in Section 3 we discuss estimation of parameters, all in the i.i.d. setting. In Section 4 we extend the method to a multivariate GARCH model and thereby test for StD innovations. In the same section we also present the corresponding asymptotics. The results of a Monte Carlo study illustrating the finite-sample properties of the method are presented in Section 5, followed by empirical applications in Section 6. Finally, we end in Section 7 with conclusions and discussion. Proofs are deferred to the Appendix, while the competitor test statistics and few extra Monte Carlo results are included in Appendix S1.

2. TEST STATISTIC

Let $(X_j,\ j = 1, \ldots, n)$ be independent copies of an arbitrary random vector $X \in \mathbb{R}^p$ of fixed dimension $p \geq 2$, and suppose we wish to test the null hypothesis

$$\mathcal{H}_0 : \text{The law of } X \in \mathcal{S}_{p,\nu}, \tag{2.1}$$

where $\mathcal{S}_{p,\nu} = \left\{S_{p,\nu}(\delta, V), (\delta, V) \in \mathbb{R}^p \times \mathbb{M}_+^p\right\}$ denotes the family of StDs with fixed degrees of freedom ν, arbitrary location vector δ, and scatter matrix V that belongs to the space $\mathbb{M}_+^p$ of $p \times p$ positive definite matrices. Here we focus attention on the 'purely' multivariate case ($p > 1$), but clearly the methods also apply to the univariate case.

Recall that if $X \sim S_{p,\nu}(\delta, V)$, then the corresponding density is given by

$$f_\nu(x) = \frac{\Gamma((\nu + p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)|V|^{1/2}}\left(1 + \frac{(x-\delta)^\top V^{-1}(x-\delta)}{\nu}\right)^{-(\nu+p)/2}; \tag{2.2}$$

see Kotz and Nadarajah (2004). Notice that for $\nu = 1$, (2.2) reduces to the density of the multivariate Cauchy distribution, while as $\nu \to \infty$, $f_\nu(\cdot)$ tends to the Gaussian density with mean δ and covariance matrix equal to V.

Also if $X \sim S_{p,\nu}(\delta, V)$, then the random vector $Y = (\nu V)^{-1/2}(X - \delta)$ satisfies $Y \sim S_{p,\nu}(0, \nu^{-1}\mathrm{I}_p)$, where I_p stands for the identity matrix in the indicated dimension. Moreover, the characteristic function (CF) of Y is given by (see Sutradhar, 1986)

$$\varphi_\mu(t) = \frac{2^{1-\mu}}{\Gamma(\mu)}\|t\|^\mu\ \mathrm{K}_\mu(\|t\|), \tag{2.3}$$

where $\nu = 2\mu$, and $\mathrm{K}_\mu(\cdot)$ stands for the MacDonald function of order $\mu > 0$, defined by

$$\mathrm{K}_\mu(x) = \left(\frac{2}{x}\right)^\mu \frac{\Gamma\left(\mu + \frac{1}{2}\right)}{\sqrt{\pi}} \int_0^\infty (1 + u^2)^{-(\mu+(1/2))} \cos(xu)\mathrm{d}u.$$

DOI: 10.1111/jtsa.12713 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Thus, and in view of the uniqueness of CFs, it is reasonable to base a test of the null hypothesis $\mathcal{H}_0$ figuring in (2.1) on the test statistic

$$T_{n,w} = n \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_\mu(t)|^2 w(t)\mathrm{d}t, \tag{2.4}$$

where $w(t) = w(-t) \geq 0$ is a weight function which will be shortly specified, and

$$\varphi_n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}t^\top \hat{Y}_j}, \ t \in \mathbb{R}^p, \tag{2.5}$$

$\left(\mathrm{i} = \sqrt{-1}\right)$, is the empirical CF computed from the standardized observations

$$\hat{Y}_j = \frac{\hat{V}_n^{-1/2}\left(X_j - \hat{\delta}_n\right)}{\sqrt{\nu}}, \quad j = 1, \ldots, n, \tag{2.6}$$

on the basis of estimators $\hat{\delta}_n$ and $\hat{V}_n$ of the parameters δ and V respectively. In (2.6), $\hat{V}_n$ is tacitly assumed to be non-singular, and $\hat{V}_n^{-1/2}$ denotes the unique symmetric square root of $\hat{V}_n^{-1}$. Rejection of the null hypothesis $\mathcal{H}_0$ is for large values of $T_{n,w}$.

In our test we will make use of the multivariate inversion theorem

$$f_X(x) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} \cos\left(t^\top x\right)\varphi_X(t)\mathrm{d}t, \tag{2.7}$$

whereby the density $f_X(\cdot)$ corresponding to a real and intergrable CF $\varphi_X(\cdot)$ may be obtained by inversion from this CF; see Ushakov (1999, Theorem 1.8.5).

By straightforward algebra it follows that if we set $w(t) = \varphi_\mu(t)/(2\pi)^p$ as weight function in (2.4) and use (2.7), then the test statistic, say $T_{n,\mu}$, may be written as

$$T_{n,\mu} = \frac{1}{n}\sum_{j,k=1}^{n} f_\mu\left(\hat{Y}_j - \hat{Y}_k\right) - 2\sum_{j=1}^{n} f_\mu^{(2)}\left(\hat{Y}_j\right) + n f_\mu^{(3)}(0), \tag{2.8}$$

where

$$f_\mu(x) = \frac{\Gamma\left(\mu + \frac{p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma(\mu)}\left(1 + \|x\|^2\right)^{-(\mu+\frac{p}{2})}, \quad x \in \mathbb{R}^p, \tag{2.9}$$

is the density corresponding to the CF given by (2.3), $f_\mu^{(2)}(\cdot)$ denotes the density of $Y_1 + Y_2$, and $f_\mu^{(3)}(\cdot)$ denotes the density of $Y_1 + Y_2 + Y_3$, with $(Y_k,\ k = 1, 2, 3)$, being independent random vectors with density $f_\mu(\cdot)$.

In this connection, the density of $Y_1 + Y_2$ has been obtained by Berg and Vignat (2010) as

$$f_\mu^{(2)}(x) = \sum_{k=0}^{\infty} c_{k,\mu}\zeta_{k,2\mu}(x), \tag{2.10}$$

where

$$c_{k,\mu} = \frac{1}{\mathrm{B}(\mu,\mu)}\frac{\Gamma\left(\frac{p}{2}+k\right)}{\Gamma\left(\frac{p}{2}\right)k!}\int_0^1 u^{\mu+\frac{p}{2}-1}(1-u)^{\mu+\frac{p}{2}-1}(1-u+u^2)^k\mathrm{d}u,$$

J. Time Ser. Anal. **45**: 298–319 (2024)
DOI: 10.1111/jtsa.12713

Table I. Values of $f_\mu^{(3)}(0)$ for different values of μ and p

	μ	1	1.5	2	2.5	3	5	7.5	15	25
$p = 2$	$f_\mu^{(3)}(0)$	0.057	0.104	0.153	0.204	0.256	0.465	0.728	1.522	2.583
$p = 5$	$f_\mu^{(3)}(0)$	0.002	0.006	0.014	0.026	0.043	0.171	0.497	2.990	11.004

and

$$\zeta_{k,\eta}(x) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}\mathrm{B}\left(\eta, k+\frac{p}{2}\right)} \frac{\|x\|^{2k}}{(1+\|x\|^2)^{k+\eta+\frac{p}{2}}}.$$

In practice, we approximated the infinite series in (2.10), by truncation of the corresponding sum at K terms, with $K = 50$ being chosen empirically.

Finally by using polar ($p = 2$), spherical ($p = 3$) or hyper-spherical ($p > 3$), coordinates, it follows that,

$$\begin{aligned}(2\pi)^p f_\mu^{(3)}(0) &= \int_{\mathbb{R}^p} \varphi_\mu^3(t)\mathrm{d}t \\ &= \left(\frac{2^{1-\mu}}{\Gamma(\mu)}\right)^3 2\pi^{\frac{p}{2}} \prod_{k=1}^{p-2} \frac{\Gamma\left(\frac{p-k}{2}\right)}{\Gamma\left(\frac{p-k+1}{2}\right)} \int_0^\infty u^{3\mu+p-1}\mathrm{K}_\mu^3(u)\mathrm{d}u, \end{aligned} \tag{2.11}$$

which involves just a single univariate integral, where we use the convention $\prod_{k=1}^0 \equiv 1$.

In Table I we give values of $f_\mu^{(3)}(0)$ for different values of μ and p.

Remark 2.1. The fact that the CF given by (2.3) is integrable, which is the condition for the inversion theorem in (2.7) to hold true, follows by direct calculation analogous to that leading to (2.11) that gives,

$$\int_{\mathbb{R}^p} \varphi_\mu(t)\mathrm{d}t = \pi^{\frac{p}{2}} \frac{2^p}{\Gamma(\mu)} \Gamma\left(\frac{p}{2}\right) \Gamma\left(\frac{p+2\mu}{2}\right) \prod_{k=1}^{p-2} \frac{\Gamma\left(\frac{p-k}{2}\right)}{\Gamma\left(\frac{p-k+1}{2}\right)}, \quad p \geq 2. \tag{2.12}$$

We now consider the limit behavior of the test statistics $T_{n,w}$ figuring in (2.4) under general conditions on the law of X and on the weight function $w(\cdot)$. Specifically we assume that the estimators $\widehat{\delta}_n$ and $\widehat{V}_n$ satisfy,

$$\widehat{\delta}_n \to \delta_X, \quad \widehat{V}_n \to V_X,$$

a.s. as $n \to \infty$, for some (finite) $\delta_X \in \mathbb{R}^p$, and some non-singular matrix V_X. Also suppose that the weight function is positive except for a set of measure zero and that $\int_{\mathbb{R}^p} w(t)\mathrm{d}t < \infty$, and write $\varphi_X(\cdot)$ for the CF of X. Then we have the following result:

Proposition 2.2. *Under the standing assumptions we have*

$$\frac{T_{n,w}}{n} \to \int_{\mathbb{R}^p} \left| \varphi_X\left(\frac{V^{-1/2}t}{\sqrt{\nu}}\right) - e^{\mathrm{i}t^\top \frac{V^{-1/2}}{\sqrt{\nu}}\delta} \varphi_\mu(t) \right|^2 w(t)\mathrm{d}t, \tag{2.13}$$

a.s. as $n \to \infty$.

Proof. Recall from (2.4) that

$$\frac{T_{n,w}}{n} = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_\mu(t)|^2 w(t)\mathrm{d}t, \tag{2.14}$$

while the strong uniform consistency of the empirical CF in bounded intervals (see Csörgő, 1981) entails

$$\varphi_n(t) \to \mathrm{e}^{-\mathrm{i}t^\top \frac{V^{-1/2}}{\sqrt{\nu}}\delta} \varphi_X\left(\frac{V^{-1/2}t}{\sqrt{\nu}}\right),$$

a.s. as $n \to \infty$. Consequently, since $|\varphi_n(t) - \varphi_\mu(t)|^2 \leq 4$, an application of Lebesgue's theorem of dominated convergence on (2.14) yields (2.13). ■

Since $w > 0$, the right-hand side of (2.13) is positive unless

$$\varphi_X\left(\frac{V^{-1/2}t}{\sqrt{\nu}}\right) = \mathrm{e}^{\mathrm{i}t^\top \frac{V^{-1/2}}{\sqrt{\nu}}\delta} \varphi_\mu(t),$$

identically in t, which is equivalent to

$$\varphi_X(t) = e^{\mathrm{i}t^\top \delta} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \|\sqrt{\nu V}t\|^{\frac{\nu}{2}} \mathrm{K}_{\frac{\nu}{2}}\left(\|\sqrt{\nu V}t\|\right). \tag{2.15}$$

The quantity in the right-hand side of (2.15) is the CF of the StD with density given by (2.2), and thus by the uniqueness of CFs, the test which rejects the null hypothesis $\mathcal{H}_0$ in (2.1) for large values of $T_{n,w}$ is consistent against such alternative distributions.

3. PARAMETER ESTIMATION

The family $\mathcal{S}_{p,\nu}$ of StDs is affine invariant (or simply invariant for short) meaning that if $X \sim S_{p,\nu}(\delta, V)$, then $AX + b \sim S_{p,\nu}(A\delta + b, AVA^\top)$, for each $p \times p$ non-singular matrix A and $b \in \mathbb{R}^p$. Therefore it is good statistical practice to require that any test statistic for $\mathcal{S}_{p,\nu}$, say T_{n,χ_n}, with $\chi_n = (X_1, \ldots, X_n)$ being the data vector, also be invariant, meaning that $T_{n,\chi_n} = T_{n,A\chi_n+b}$, for each (b, A). Such a property will guarantee that the decisions about rejection of the null hypothesis reached on the basis of this test statistic either by the use of the observations χ_n or by $A\chi_n + b$ are identical. For detailed discussions of the notion of invariant tests we refer to Henze (2002), Ebner and Henze (2020), and Ducharme and de Micheaux (2020).

With invariance in mind we now discuss the estimation of (δ, V). Specifically we assume that the estimators $\hat{\delta}_n := \hat{\delta}_{n,\chi_n}$ and $\hat{V}_n := \hat{V}_{n,\chi_n}$ satisfy,

$$\hat{\delta}_{n,A\chi_n+b} = A\hat{\delta}_{n,\chi_n} + b, \tag{3.1}$$

and

$$\hat{V}_{n,A\chi_n+b} = A\hat{V}_{n,\chi_n}A^\top; \tag{3.2}$$

see Bilodeau and Brenner (1999, chapter 13). In this connection, the moment estimators

$$\hat{\delta}_n = \overline{X}_n, \quad \hat{V}_n = \frac{\nu - 2}{\nu} S_n, \quad (\nu > 2), \tag{3.3}$$

with $\overline{X}_n = n^{-1}\sum_{j=1}^{n} X_j$ and $S_n = n^{-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)(X_j - \overline{X}_n)^\top$, typically satisfy (3.1) and (3.2). Maximum likelihood estimators also satisfy (3.1) and (3.2). (For estimation methods in the family of StDs see Kotz and Nadarajah, 2004, chapter 10, Liu, 1997, and Nadarajah and Kotz, 2008). With such invariant estimators and with a little reflection on (2.8), it follows that the test statistic $T_{n,\mu}$ depends on the observations solely via

$$\hat{Y}_j^\top \hat{Y}_k = \left(X_j - \hat{\delta}_n\right)^\top \hat{V}_n^{-1}\left(X_k - \hat{\delta}_n\right), \quad j, k = 1, \ldots, n, \tag{3.4}$$

which are in turn invariant given that (3.1) and (3.2) are satisfied, and consequently the test statistic $T_{n,\mu}$ is also rendered invariant. Also notice that, due to (3.4), the square root of the matrix $\hat{V}_n$ is not needed for the computation of $T_{n,\mu}$. A further advantage of using an invariant test such as $T_{n,\mu}$ is that the (finite-sample or asymptotic) distribution of $T_{n,\mu}$ does not depend on the actual values of (δ, V) which can be set equal to their standard values $(0, \mathrm{I}_p)$. We will study the asymptotic null distribution of the test statistic in the next section, whereby we consider the case of testing for the law of innovations in a GARCH model, a setting more general than the current setting, and for which the current i.i.d. setting, with fixed location at zero, may be viewed as a special case.

4. TEST FOR A GARCH MODEL WITH STUDENT-T INNOVATIONS

We propose an extension of the suggested method to test the validity of a Student-*t* innovation distribution in a given (multivariate) GARCH model, against general alternative distributions. In this connection, a wide variety of diagnostic procedures as well as other informal tests for certain modeling aspects within GARCH models have been applied, including likelihood criteria and information criteria (Rossi and Spazzini, 2010, Creal *et al.*, 2011), Ljung-Box, LM and portmanteau tests (Tsay, 2006; Bauwens *et al.*, 2006; Wang and Tsay, 2013), individual Q-Q plots and probability integral transform plots (Pesaran and Pesaran, 2007; Dube *et al.*, 2016) as well as sample autocorrelations (Zheng *et al.*, 2018). We also refer to the methods of Francq and Zakoïan (2022) for testing assumptions on specific characteristics of the innovation distribution, such as quantiles, moments, and asymmetry. These procedures however are either heuristic in nature, or indirectly test the model itself by targeting specific model-aspects, or test one fixed model against another fixed model, and thus they are not specifically aimed at the law of innovations. For rigorous GARCH-innovation tests we refer to Klar *et al.* (2012), Lee *et al.* (2015), Ghoudi and Rémillard (2014), and Dalla *et al.* (2017), all in the univariate setting. On the other hand and apart from the test of Henze *et al.* (2019) for the Gaussian distribution, there is relative scarcity of rigorous testing procedures tailored particularly to the law of innovations in multivariate GARCH models. In fact, and to the best of our knowledge, there exist only a couple of other existing tests, the first being that of Bai and Chen (2008), which involves the application of the Rosenblatt transform to reduce the data to p approximately independent uniform [0,1] variates and the Khmaladze transform that removes the effect of estimation. Thus this method is not so straightforward to apply, at least in the context of testing for GARCH innovations.

The other competing procedure can be found in the recent paper of Luo *et al.* (2023), and depends on a Stein-type characterization involving the score function. This method has the advantage of not being tailored, but applicable to any distribution under test provided of course that the underlying density (assumed to exist) is not too heavy-tailed, and that both density and score functions are sufficiently smooth, while for our test we only need the CF, which always exists, and moment assumptions are incidental in our case and due to the estimation step preceding the test. Moreover, and unlike our method, the test of Luo *et al.* (2023) requires data subsampling, and in the i.i.d. case is apparently not affine invariant, which implies potential dependence of the finite-sample distribution on the actual location vector and scatter matrix. These remarks notwithstanding, the test of Luo *et al.* (2023) remains a formidable competitor, and further comparison with our test will be discussed by means of Monte Carlo simulations in Section 5.

In developing our test we adopt as a reference GARCH model the CCC-GARCH of Bollerslev (1990) and Jeantheau (1998). This specification has been found to perform well in a few applications, see for instance Santos *et al.* (2013), but here it is adopted as a sort of 'threshold' model because of its simplicity. Otherwise our asymptotic

DOI: 10.1111/jtsa.12713 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

results do not make any specific use of the CCC-GARCH structure and thus hold true for the general MGARCH model defined in section 10.4 of Francq and Zakoïan (2019). In this connection extension of the test procedure to similar models seems to be in principle straightforward: Once the residual vector is obtained under whatever GARCH specification, the test is readily applied on this vector, and on the practical level the finite-sample behavior of the test is expected to be analogous. We will however come back to this issue in Section 6 where we discuss results for the diagonal CCC-GARCH, the non-diagonal ECCC-GARCH, and the dynamic conditional correlation model, the DCC-GARCH.

This remark notwithstanding, we point out that the asymptotic as well as the finite-sample behavior of our test is conditioned on the existence of a proper limit distribution of the estimators involved for the GARCH parameters; remark 10.6 of Francq and Zakoïan (2019) is relevant here. Thus caution should be exercised when applying our test to alternative GARCH specifications, such as the DCC-GARCH model, for which the domain of stationarity as well as the existence of a proper limit behavior of the corresponding estimators is still under scrutiny. In this regard it should be pointed out that alternative estimation methods could also have been employed for the GARCH parameters, but then the corresponding conditions for the limit null distribution of the test statistic should be analogously adjusted; by way of example we refer to the equation-by-equation estimator of Francq and Zakoïan (2016) where, for instance, the moment assumption for the law of GARCH errors is more restrictive than the one adopted herein (see condition (RA7) in Appendix A.1). For a more technical discussion we refer to Appendix A.1.

4.1. Setup

Assume that the observations $\left(X_j,\ j=1,\ \dots\ ,n\right)$, arise from a multivariate GARCH model defined by

$$X_j = \Sigma_j^{1/2}\varepsilon_j, \tag{4.1}$$

where $(\varepsilon_j,\ j=1,\ \dots\ ,n)$, are i.i.d. p-dimensional random vectors with mean zero and unit covariance matrix, and $\Sigma_j := \mathrm{Cov}(X_j|\mathcal{F}_{j-1}) = \Sigma\left(X_{j-1}, X_{j-2},\ \dots\ ;\theta\right)$, is a symmetric positive definite matrix of dimension $p\times p$ that depends on the past via a parameter vector θ to be specified below.

For the aforementioned CCC-GARCH model of Bollerslev (1990) and Jeantheau (1998), the matrix Σ_j is specified by

$$\Sigma_j(\theta, r, s) = D_j(\theta)R(\theta)D_j(\theta), \tag{4.2}$$

where $R(\cdot)$ is a $p\times p$ constant correlation matrix and $D_j(\cdot) = \mathrm{diag}(\sigma_{1,j},\ \dots\ ,\sigma_{p,j})$ is a diagonal matrix with elements satisfying

$$\sigma_j = \left(\sigma^2_{1,j},\ \dots\ ,\sigma^2_{p,j}\right)^\top = \omega + \sum_{k=1}^{r} A_k X^{(2)}_{j-k} + \sum_{k=1}^{s} B_k\sigma_{j-k}, \tag{4.3}$$

with $X_j^{(2)} = \left(X^2_{1,j},\ \dots\ ,X^2_{p,j}\right)^\top$. The parameter θ figuring in (4.2) comprises the elements of ω which is a $p\times 1$ vector of positive elements, the elements of R, and the elements of the $p\times p$ matrices A_k and B_k which are by definition non-negative.

Under the setting (4.1)–(4.3) we wish to test that for some $\theta=\theta_0$ the null hypothesis $\mathcal{H}_0$ stated in (2.1) holds true for the common law of the innovations $(\varepsilon_j, j=1,\ \dots\ ,n)$. To keep close to the standard GARCH setting we assume that the innovations have zero mean and unit covariance. (This corresponds to the specification $(\delta, V) = (0, ((\nu-2)/\nu)\mathrm{I}_p)$, in the original null hypothesis stated in (2.1)).

In this connection, any test aiming at the unobserved innovations should be actually applied on the corresponding residuals

$$\widetilde{\varepsilon}_j = \widetilde{\Sigma}_j^{-1/2}(\widehat{\theta}_n)X_j, \quad j=1,\ \dots\ ,n, \tag{4.4}$$

obtained on the basis of a consistent estimator $\hat{\theta}_n$ of θ that uses $\left(X_{j-1}^\top, X_{j-2}^\top, \ldots\right)^\top$ as input data. Note however that $\Sigma_j(\theta)$ depends on $\{X_k, -\infty < k \leq j-1\}$, whereas we only observe $X_1, \ldots, X_n$. Hence, for calculating the residuals we consider $\widetilde{\Sigma}_j(\hat{\theta}_n)$ defined as $\Sigma_j(\hat{\theta}_n)$, whereby we also employ initial values $\left(\widetilde{X}_0^\top, \ldots, \widetilde{X}_{1-r}^\top, \widetilde{\sigma}_0^\top, \ldots, \widetilde{\sigma}_{1-s}^\top\right)^\top$ to start the estimation process.

4.2. Asymptotic properties

In this part we obtain the limit null distribution of the test statistic. In doing so we also provide the missing asymptotics for the specification test of Klar *et al.* (2012) corresponding to univariate GARCH models. To keep close to the test statistic as defined by (2.4)–(2.6) we formulate our test as

$$\mathcal{T}_{n,w} = n \int_{\mathbb{R}^p} |\phi_n(t) - \varphi_\mu(t)|^2 w(t) \mathrm{d}t, \tag{4.5}$$

and use the same weight function, where

$$\phi_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}t^\top \widetilde{z}_j}, \quad t \in \mathbb{R}^p, \tag{4.6}$$

is the empirical CF of the standardized residuals $\widetilde{z}_j := \widetilde{\varepsilon}_j / \sqrt{\nu - 2}$, with $\widetilde{\varepsilon}_j$ defined in (4.4) and $\varphi_\mu(t)$ denotes the CF figuring in (2.3). Thus the test statistic is computed as $\mathcal{T}_{n,w} := T_{n,\mu}(\widetilde{z}_1, \ldots, \widetilde{z}_n)$ where $T_{n,\mu}(\widehat{Y}_1, \ldots, \widehat{Y}_n)$ denotes the 'i.i.d.' test statistic defined in (2.8).

We first state the result for the limit null distribution of the test statistic in the GARCH case.

Theorem 4.1. *Let the assumptions* RA1–RA7 *stated in the Appendix be satisfied. Moreover assume that the weight function satisfies* $\int_{\mathbb{R}^p} \|t\|^4 w(t) \mathrm{d}t < \infty$.[1] *Then under the null hypothesis there exists a zero mean Gaussian element* $\mathcal{G}$ *with covariance kernel* $K_G(s,t)$, *such that* $\mathcal{T}_{n,w} \to \|\mathcal{G}\|_{L^2}$.

The following corollary gives the asymptotic null distribution of the test statistic in the i.i.d. case.

Corollary 4.2. *Under the null hypothesis there exists a zero mean Gaussian element* $\mathcal{G}_0$ *with covariance kernel* $K_{G_0}(s,t) = \mathbb{E} U_n(t) U_n(s)$, *such that* $T_{n,w} \to \|\mathcal{G}_0\|_{L^2}$, *where* U_n *is defined in* (A2) *of the Appendix.*

The proofs of Theorem 4.1 and Corollary 4.2 are postponed to the Appendix. In the Appendix we also provide the expression for the covariance kernel of Theorem 4.1; see (A5).

As already stated, the test statistic itself as well as its asymptotics have been developed under the assumption that the degrees of freedom of the StD under test is fixed (known) in advance. An extension of the test formulated as in (4.5) to the case of estimated degrees of freedom is clear: We replace μ by its estimate $\hat{\mu}_n$, set the weight function w proportional to $\varphi_{\hat{\mu}_n}$, and carry out the computation of the test as in (2.8). The only problem is the extra randomness introduced in the weight function, which will be discussed in the next paragraph. Alternatively, and as far as practical applications are concerned, the test may be applied with a few fixed degrees of freedom, and the appropriate model can then be decided on the basis of likelihood or other model-choice criteria performed on non-rejected models; see for instance Hafner *et al.* (2020) and Wang and Tsay (2013). For a rigorous model selection procedure based on the CF we refer to Jiménez-Gamero *et al.* (2016).

In this connection and to avoid the problem with the random weight function when ν is unknown which could affect asymptotics, we point out that other weight functions, besides the CF $\varphi_\mu(\cdot)$ under the null hypothesis, may

[1] It may be shown that the Student-t CF $\varphi_\mu(\cdot)$ satisfies this condition.

DOI: 10.1111/jtsa.12713 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

be used in the test statistic. Specifically if the weight function for $T_{n,w}$ in (2.4) is set equal to the CF of any StD with $\nu' \neq \nu$ degrees of freedom, all computations follow through by using the distribution of two independent but not necessarily identically distributed Student-t random vectors obtained by Berg and Vignat (2010). Specifically let $\hat{\nu}_n$ be an estimate of ν, and set $\hat{\mu}_n = \hat{\nu}_n/2$. Then following analogous steps to those leading to (2.8) the test statistic is readily reduced to

$$\hat{T}_{n,\mu'} = \frac{1}{n}\sum_{j,k=1}^{n} f_{\mu'}\left(\hat{Y}_j - \hat{Y}_k\right) - 2\sum_{j=1}^{n} f^{(2)}_{\hat{\mu}_n,\mu'}\left(\hat{Y}_j\right) + nf^{(3)}_{\hat{\mu}_n,\mu'}(0), \tag{4.7}$$

where $f_{\mu'}(\cdot)$ is the density defined in (2.9) with μ replaced by μ' ($\nu' = 2\mu'$), $f^{(2)}_{\mu,\mu'}(\cdot)$ denotes the density of $Y_1 + Y_2$, with Y_1 (resp. Y_2) following a StD with ν (resp. ν') degrees of freedom, and independent, and $f^{(3)}_{\mu,\mu'}(\cdot)$ denotes the density of $Y_1 + Y_2 + Y_3$, with $(Y_k,\ k = 1, 2)$, as in $f^{(2)}_{\mu,\mu'}(\cdot)$, and Y_3 an independent copy of Y_1. This feature provides the test with a certain flexibility, a fact that is illustrated by means of the additional Monte Carlo results contained in Tables S5–S8.

The type of weight function $w(\cdot)$ adopted herein goes back to Epps (2005) who associates the choice of $w(\cdot)$ with the CF of the family being tested under the null. In fact it is precisely this choice, coupled with the inversion theorem for CFs and the results of Berg and Vignat (2010), that leads to the closed-form of the test statistics figuring in (2.8) and (4.7), and this computational expedience is no minor issue in the multivariate setting. Other aspects of the test that are safeguarded by choosing $\varphi_\mu(\cdot)$ as weight function are consistency and affine invariance. In this connection notice that the arguments leading to the consistency of the test require, among other things, that the weight function should be positive for all $t \in \mathbb{R}^p$; see the conditions preceding the statement of Prop. 2.2. Moreover invariance follows from the fact that by using $\varphi_\mu(\cdot)$ in $w(\cdot)$, results in a test statistic that depends solely on the so-called Mahalanobis distances defined in (3.4), thus rendering $T_{n,\mu}$ affine invariant. These considerations notwithstanding, there is reasonable concern as to how the choice of $w(\cdot)$ affects the power of the test. There exist some works in this line of research for testing univariate laws in the i.i.d. setting; see for instance Tenreiro (2019), Ebner and Henze (2021) and Tenreiro (2022). However, when it comes to multivariate distributions, the only available results for choosing $w(\cdot)$ are for testing normality and yet these results are based almost exclusively on Monte Carlo experiments; see Tenreiro (2009) and Tenreiro (2011).

To see where the problem lies we note that the limit null distribution of $T_{n,w}$ is that of $\sum_{j=1}^{\infty} \lambda_j \mathcal{N}_j^2$, where $(\mathcal{N}_j,\ j \geq 1)$ are i.i.d. standard normal random variables. In turn $\lambda_1 \leq \lambda_2 \leq \cdots$, are the eigenvalues of a complicated integral equation depending on the weight function $w(\cdot)$, the distribution being tested under the null as well as on the estimation of any parameters involved. Regarding power properties, the notion of Bahadur efficiency is often the preferred notion, but such efficiency requires calculation of the eigenvalues, and in particular the knowledge of λ_1, the largest eigenvalue, which however is rarely available. The corresponding technical analysis is highly non-trivial and in general it remains an open problem, but some results in this direction are available in Meintanis *et al.* (2022) and Móri *et al.* (2021), again for testing multivariate normality. Here instead we adopt the pragmatic approach of observing how the finite-sample power of the test varies with the weight function, and thus making choices. In this regard we refer to the Monte Carlo results in the Appendix S1 which indicate that the power varies to some extent with different degrees of freedom in the weight function $w(\cdot)$, sometimes rendering the test more powerful for smaller degrees while for other alternatives larger degrees of freedom in $w(\cdot)$ are preferable. In view of this behavior taking the weight function $w(\cdot)$ proportional to the CF of the particular StD being tested, i.e. with the same degrees of freedom, appears to be a good compromise. Nevertheless our results are not thorough and more work is needed in this direction.

5. MONTE CARLO STUDY

We present a study of the power performance of our test as compared to the following competitors: the test based on the Wasserstein metric proposed in Hallin *et al.* (2021) (denoted as W), the kernel Stein discrepancy- (KSD) based test proposed in Luo *et al.* (2023) (denoted as L), and the tests based on Rosenblatt transform for three different weight functions, denoted as (R_k, $k = 1, 2, 3$), which correspond to the statistics denoted as (T_k, $k = 1, 2, 3$), in the original paper of Meintanis *et al.* (2014). The test statistics are available in the Appendix S1. All tests were performed at 5% the level of significance.

The list of considered alternatives includes the following multivariate distributions: Student t, normal (N_p), Laplace (LA_p) (see Kotz *et al.*, 2012), generalized Gaussian (GN_p) (see Goodman and Kotz, 1973; Nardon and Pianca, 2009), skew-normal SN_p (see Arellano-Valle and Azzalini, 2008) and skew-t ST_p (see Bauwens *et al.*, 2006).

However, the limit null distributions derived in Theorem 4.1 and Corollary 4.2 are extremely complicated to use for actual test implementation; see for instance Meintanis and Swanepoel (2007). Thus we resort to resampling to approximate the actual distribution of the new test statistics. In the case of i.i.d. observations and due to the affine invariance property of the test discussed in Section 3, a simple Monte Carlo approximation is enough to compute critical points and carry out the test. On the other hand, in the cases such as for GARCH observations, where there is additional uncertainty involved in the test statistic due to parameter estimation a further bootstrap cycle of resampling is required. Here we use the warp-speed version of the parametric bootstrap method suggested by Giacomini *et al.* (2013). We note that the parametric bootstrap is specifically designed for test statistics with unknown parameters, and for dynamic models has been validated by Rémillard (2011). This bootstrap procedure is described in the Appendix S1.

5.1. Results in the i.i.d. case

Based on i.i.d. samples of sizes $n = 25$, 50 and $n = 100$, we test the null hypotheses that the sample comes from a p-variate Student distribution with 5 and 10 degrees of freedom.

All tests are applied to the standardized data, where the location and scatter parameters are estimated using the method of moments. The empirical sizes and powers are calculated using the Monte Carlo method with 10,000 replications. However since the tests based on the Rosenblatt transform are not distribution-free, the aforementioned warp-speed parametric bootstrap procedure with 5000 resamples is employed. The results are presented in Tables S1 and S2 ($p = 2$) and Tables S3 and S4 ($p = 5$).

5.2. Results in the GARCH case

Based on GARCH time series with 200 and 500 observations we test the null hypotheses that the innovations follow p-variate Student distribution with 5 and 10 degrees of freedom with unit covariance matrix.

The empirical sizes and powers are calculated using the warp-speed parametric bootstrap method with 5000 replicates. Here for dimension $p = 5$ we excluded the Rosenblatt transform tests due to computational complexity in higher dimensions. All tests are applied using the QMLE method for estimation of the GARCH parameters; see e.g. Francq and Zakoïan (2019).

The choice of considered models is the following: a bivariate GARCH(1,1) model with $\omega = (0.05, 0.1)$, $A = \text{diag}(0.06, 0.08)$, $B = \text{diag}(0.05, 0.4)$, and the 5-variate GARCH(1,1) model with $\omega = (0.03, 0.02, 0.02, 0.03, 0.01)$, $A = \text{diag}(0.1, 0.1, 0.1, 0.1, 0.1)$, $B = \text{diag}(0.1, 0.1, 0.2, 0.3, 0.3)$, with the matrix R in each case having all off-diagonal elements equal to 0.5. The results are presented in Table II. GARCH models with alternative parameter settings were also tried. The corresponding Monte Carlo results obtained are very similar and are shown in Table S9.

We now compare the performance of the W test, the (R_k, $k = 1, 2, 3$) tests and our test with the L test of Luo *et al.* (2023) by means of a Monte Carlo experiment employing precisely the GARCH models and parameter settings adopted in that paper. These models correspond to the non-diagonal version of the CCC-GARCH, i.e. the ECCC-GARCH, with dimensions $p = 2$ and $p = 5$, and the precise values of the GARCH parameters may be found in section 4.3 of Luo *et al.* (2023). The results are presented in Table III.

Table II. Empirical percentage of rejection for CCC-GARCH observations

	p = 2								p = 5				
n	$\mathcal{H}_0$	Alt.	R_1	R_2	R_3	W	L	$\mathcal{T}$	$\mathcal{H}_0$	Alt.	W	L	$\mathcal{T}$
200	$t_{2,10}$	$t_{2,10}$	0.04	0.05	0.05	0.05	0.05	0.04	$t_{5,10}$	$t_{5,10}$	0.05	0.05	0.05
200	$t_{2,10}$	$t_{2,5}$	0.07	0.05	0.05	0.30	0.35	0.35	$t_{5,10}$	$t_{5,5}$	0.54	0.31	0.75
200	$t_{2,10}$	LA_2	0.15	0.05	0.06	0.82	0.95	0.95	$t_{5,10}$	LA_5	1.00	0.95	1.00
200	$t_{2,10}$	GN_2	1.00	0.05	0.11	1.00	1.00	1.00	$t_{5,10}$	GN_5	1.00	1.00	1.00
200	$t_{2,10}$	N_2	0.06	0.06	0.05	0.07	0.10	0.12	$t_{5,10}$	N_5	0.04	0.06	0.41
200	$t_{2,10}$	SN_2	0.08	0.07	0.07	0.26	0.49	0.49	$t_{5,10}$	SN_5	0.18	0.26	0.85
200	$t_{2,10}$	ST_2	0.06	0.05	0.05	0.15	0.27	0.24	$t_{5,10}$	ST_5	0.43	0.51	0.72
200	$t_{2,5}$	$t_{2,5}$	0.05	0.05	0.05	0.06	0.06	0.05	$t_{5,5}$	$t_{5,5}$	0.05	0.05	0.06
200	$t_{2,5}$	$t_{2,10}$	0.06	0.05	0.06	0.07	0.09	0.25	$t_{5,5}$	$t_{5,10}$	0.18	0.04	0.70
200	$t_{2,5}$	LA_2	0.07	0.05	0.05	0.14	0.26	0.32	$t_{2,5}$	LA_2	0.51	0.10	1.00
200	$t_{2,5}$	GN_2	0.99	0.06	0.09	1.00	1.00	1.00	$t_{5,5}$	GN_5	1.00	0.85	1.00
200	$t_{2,5}$	N_2	0.06	0.05	0.05	0.34	0.42	0.78	$t_{5,5}$	N_5	0.79	0.08	1.00
200	$t_{2,5}$	SN_2	0.11	0.08	0.08	0.67	0.77	0.94	$t_{5,5}$	SN_5	0.97	0.35	1.00
200	$t_{2,5}$	ST_2	0.09	0.07	0.07	0.19	0.35	0.35	$t_{5,5}$	ST_5	1.00	1.00	1.00
500	$t_{2,10}$	$t_{2,10}$	0.05	0.05	0.05	0.06	0.05	0.05	$t_{5,10}$	$t_{5,10}$	0.05	0.05	0.05
500	$t_{2,10}$	$t_{2,5}$	0.11	0.05	0.06	0.68	0.70	0.73	$t_{5,10}$	$t_{5,5}$	0.93	0.75	0.99
500	$t_{2,10}$	LA_2	0.67	0.05	0.06	1.00	1.00	1.00	$t_{5,10}$	LA_5	1.00	1.00	1.00
500	$t_{2,10}$	GN_2	1.00	0.09	0.62	1.00	1.00	1.00	$t_{5,10}$	GN_5	1.00	1.00	1.00
500	$t_{2,10}$	N_2	0.06	0.05	0.05	0.11	0.24	0.31	$t_{5,10}$	N_5	0.09	0.13	0.92
500	$t_{2,10}$	SN_2	0.14	0.08	0.09	0.93	0.93	0.93	$t_{5,10}$	SN_5	0.87	0.77	1.00
500	$t_{2,10}$	ST_2	0.11	0.07	0.07	0.55	0.64	0.60	$t_{5,10}$	ST_5	0.97	0.94	0.99
500	$t_{2,5}$	$t_{2,5}$	0.05	0.05	0.05	0.05	0.05	0.05	$t_{5,5}$	$t_{5,5}$	0.05	0.05	0.05
500	$t_{2,5}$	$t_{2,10}$	0.07	0.08	0.08	0.36	0.39	0.69	$t_{5,5}$	$t_{5,10}$	0.91	0.10	0.99
500	$t_{2,5}$	LA_2	0.12	0.05	0.06	0.77	0.78	0.85	$t_{5,5}$	LA_5	1.00	0.54	1.00
500	$t_{2,5}$	GN_2	1.00	0.09	0.26	1.00	1.00	1.00	$t_{5,5}$	GN_5	1.00	1.00	1.00
500	$t_{2,5}$	N_2	0.12	0.06	0.05	0.98	0.98	1.00	$t_{5,5}$	N_5	1.00	0.68	1.00
500	$t_{2,5}$	SN_2	0.30	0.10	0.12	1.00	1.00	1.00	$t_{5,5}$	SN_5	1.00	0.99	1.00
500	$t_{2,5}$	ST_2	0.15	0.10	0.11	0.68	0.78	0.74	$t_{5,5}$	ST_5	1.00	1.00	1.00

5.3. Discussion

From the empirical sizes, we can see that all tests are well calibrated. In the i.i.d case our test is comparable to the Wasserstein distance test, the KSD based test, and the Rosenblatt transform test R_1, while the other tests based on the Rosenblatt transform are much less powerful. It is worth mentioning that our test is significantly better than competitors in the case of the standard normal alternative and the Student alternative with more degrees of freedom, while being slightly behind some of the competitors for the Laplace and generalized Gaussian distributions.

Both CCC-GARCH (Table II) and ECCC-GARCH (Table III) settings suggest that the new test is the most powerful nearly uniformly over sample size, dimension and type of alternative, and when $p = 5$ often by a wide margin, while in the cases that the $\mathcal{T}$ test is not the best test, the power differential is mostly quite small.

Moreover, the similarity of results in Tables II and III, and those in Table S9, suggests that the GARCH model parameters do not have a strong impact on the power performance of the considered tests. This however only applies to the above mentioned GARCH specifications, while analogous Monte Carlo trials showed that test performance is prone to numerical error under the DCC-GARCH model, thus rendering the corresponding results biased. For this reason these results are not reported.

DOI: 10.1111/jtsa.12713

Table III. Empirical percentage of rejection for ECCC-GARCH observations

	$p=2$								$p=5$				
n	$\mathcal{H}_0$	Alt.	R_1	R_2	R_3	W	L	$\mathcal{T}$	$\mathcal{H}_0$	Alt.	W	L	$\mathcal{T}$
200	$t_{2,10}$	$t_{2,10}$	0.05	0.05	0.05	0.05	0.05	0.05	$t_{5,10}$	$t_{5,10}$	0.05	0.05	0.05
200	$t_{2,10}$	$t_{2,5}$	0.06	0.05	0.05	0.22	0.28	0.26	$t_{5,10}$	$t_{5,5}$	0.36	0.20	0.49
200	$t_{2,10}$	LA_2	0.13	0.05	0.06	0.77	0.91	0.89	$t_{5,10}$	LA_5	1.00	0.81	1.00
200	$t_{2,10}$	GN_2	1.00	0.06	0.11	1.00	1.00	1.00	$t_{5,10}$	GN_5	1.00	1.00	1.00
200	$t_{2,10}$	N_2	0.05	0.05	0.05	0.07	0.10	0.14	$t_{5,10}$	N_5	0.05	0.06	0.55
200	$t_{2,10}$	SN_2	0.08	0.07	0.07	0.25	0.47	0.50	$t_{5,10}$	SN_5	0.26	0.27	0.91
200	$t_{2,10}$	ST_2	0.07	0.07	0.07	0.15	0.28	0.25	$t_{5,10}$	ST_5	0.35	0.49	0.66
200	$t_{2,5}$	$t_{2,5}$	0.05	0.05	0.05	0.05	0.04	0.04	$t_{5,5}$	$t_{5,5}$	0.05	0.05	0.05
200	$t_{2,5}$	$t_{2,10}$	0.05	0.05	0.05	0.10	0.11	0.29	$t_{5,5}$	$t_{5,10}$	0.23	0.05	0.78
200	$t_{2,5}$	LA_2	0.06	0.05	0.05	0.15	0.25	0.27	$t_{5,5}$	LA_5	0.27	0.06	0.97
200	$t_{2,5}$	GN_2	0.98	0.06	0.09	1.00	1.00	1.00	$t_{5,5}$	GN_5	1.00	0.69	1.00
200	$t_{2,5}$	N_2	0.06	0.05	0.05	0.38	0.46	0.83	$t_{5,5}$	N_5	0.80	0.11	0.99
200	$t_{2,5}$	SN_2	0.11	0.07	0.07	0.73	0.81	0.95	$t_{5,5}$	SN_5	0.97	0.39	1.00
200	$t_{2,5}$	ST_2	0.08	0.07	0.07	0.21	0.36	0.35	$t_{5,5}$	ST_5	1.00	1.00	1.00
500	$t_{2,10}$	$t_{2,10}$	0.05	0.05	0.05	0.05	0.05	0.05	$t_{5,10}$	$t_{5,10}$	0.05	0.05	0.05
500	$t_{2,10}$	$t_{2,5}$	0.09	0.05	0.05	0.69	0.68	0.69	$t_{5,10}$	$t_{5,5}$	0.85	0.60	0.96
500	$t_{2,10}$	LA_2	0.63	0.05	0.06	1.00	1.00	1.00	$t_{5,10}$	LA_5	1.00	1.00	1.00
500	$t_{2,10}$	GN_2	1.00	0.08	0.50	1.00	1.00	1.00	$t_{5,10}$	GN_5	1.00	1.00	1.00
500	$t_{2,10}$	N_2	0.05	0.04	0.05	0.10	0.23	0.33	$t_{5,10}$	N_5	0.16	0.14	0.95
500	$t_{2,10}$	SN_2	0.14	0.08	0.09	0.91	0.95	0.95	$t_{5,10}$	SN_5	0.89	0.81	1.00
500	$t_{2,10}$	ST_2	0.12	0.09	0.09	0.57	0.64	0.60	$t_{5,10}$	ST_5	0.97	0.95	0.99
500	$t_{2,5}$	$t_{2,5}$	0.05	0.05	0.05	0.05	0.05	0.05	$t_{5,5}$	$t_{5,5}$	0.05	0.05	0.05
500	$t_{2,5}$	$t_{2,10}$	0.08	0.06	0.06	0.41	0.43	0.75	$t_{5,5}$	$t_{5,10}$	0.92	0.14	0.99
500	$t_{2,5}$	LA_2	0.10	0.05	0.05	0.76	0.73	0.82	$t_{5,5}$	LA_5	0.99	0.38	1.00
500	$t_{2,5}$	GN_2	1.00	0.07	0.22	1.00	1.00	1.00	$t_{5,5}$	GN_5	1.00	1.00	1.00
500	$t_{2,5}$	N_2	0.12	0.04	0.06	0.99	0.98	1.00	$t_{5,5}$	N_5	1.00	1.00	1.00
500	$t_{2,5}$	SN_2	0.31	0.10	0.11	1.00	1.00	1.00	$t_{5,5}$	SN_5	1.00	1.00	1.00
500	$t_{2,5}$	ST_2	0.15	0.11	0.11	0.69	0.78	0.75	$t_{5,5}$	ST_5	1.00	1.00	1.00

6. REAL DATA EXAMPLES

We apply our test on a real data set.

We consider the monthly rates of stock return, defined as the first difference of the log index prices

$$R_{i,t} = 100(\ln P_{i,t} - \ln P_{i,t-1}),$$

of the IBM stock and the S&P 500 index from January 30, 1926 to December 31, 1999. This dataset consists of 888 observations and was analyzed in Tsay (2010) as well as in Henze and Jiménez-Gamero (2019). The data is available from R. Tsay's website.[2] Assuming constant trend, we subtracted mean values for considered period and proceeded with such 'centered' log return rates. The de-trended monthly log returns rates and their squares are presented on Figure 1, and suggest that a GARCH model may be appropriate. To model this time series, an extension of CCC-GARCH(1,1) where the matrices A_k and B_k in (4.3) are not diagonal, is employed. This model is also assumed in Henze and Jiménez-Gamero (2019).

[2] https://faculty.chicagobooth.edu/ruey-s-tsay/teaching

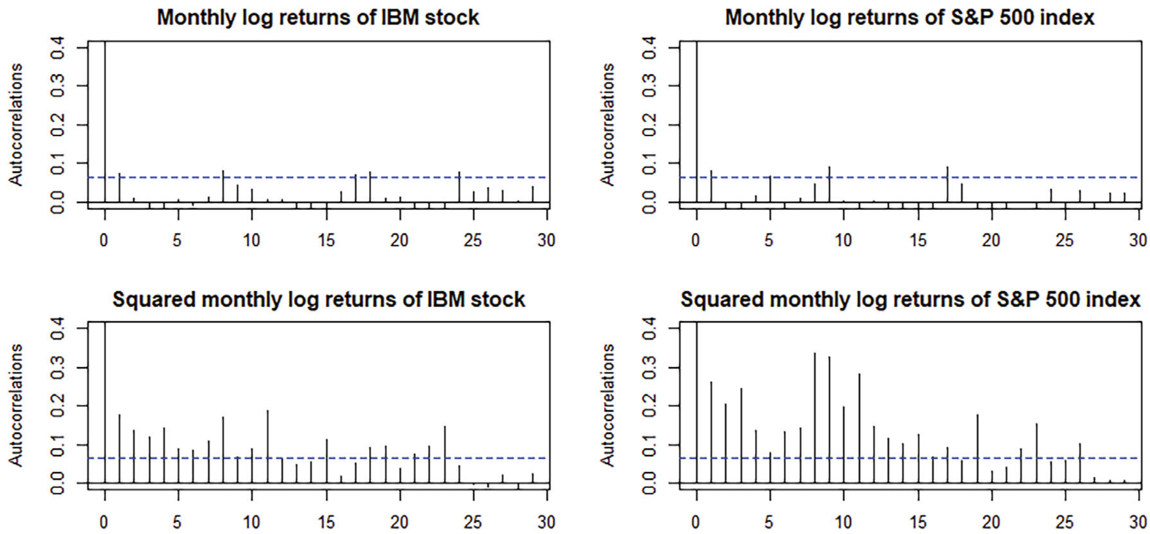

Figure 1. The monthly log returns and the square of monthly log returns

Table IV. Real data example: p-values (ECCC-GARCH)

		IBM stocks and S&P index				
$\mathcal{H}_0$	R_1	R_2	R_3	W	L	$\mathcal{T}$
$t_{2,5}$	0.70	0.80	0.86	0.21	0.44	0.50
$t_{2,7}$	0.67	0.74	0.80	0.20	0.39	0.47
$t_{2,10}$	0.51	0.74	0.281	0.11	0.17	0.20
$t_{2,11}$	0.55	0.77	0.77	0.09	0.12	0.16
$t_{2,12}$	0.51	0.79	0.79	0.07	0.10	0.13
$t_{2,13}$	0.48	0.78	0.75	0.07	0.08	0.11
$t_{2,14}$	0.48	0.78	0.77	0.05	0.06	0.08
$t_{2,15}$	0.45	0.77	0.74	0.06	0.07	0.08
$t_{2,20}$	0.38	0.75	0.72	0.03	0.02	0.03
$t_{2,30}$	0.32	0.72	0.66	0.04	0.01	0.03
$t_{2,50}$	0.29	0.73	0.64	0.03	0.02	0.02

For alternative degrees of freedom in the null hypothesis, the p-values of the proposed test, the Wasserstein test and of the Rosenblatt tests were calculated using the bootstrap procedure with 500 resamples and are reported in Table IV. From the table we can see that all tests do not reject the null hypotheses of a Student-t innovation distribution with $\nu \leq 15$. For higher degrees of freedom the results are mixed with a couple of tests leading to rejection (our test and the Wasserstein test), while the p-values of the Rosenblatt transform tests are clearly non-significant. In this connection, we mention that our results are in line with the conclusion of Henze and Jiménez-Gamero (2019), whereby a GARCH model with Gaussian innovations is rejected.

Following Luo *et al.* (2023), we now look whether the null distributions with highest p-values from Table IV give us better portfolio VaR forecasts in comparison with others. Let X_t be the bivariate time series of the aforementioned monthly log returns. We consider the equal-weight portfolio $w = (1/2, 1/2)^\top$ and forecast the VaR of $w^\top X_t$ at time t. Let $\Sigma_{t+1|t}$ and $VaR_{t+1|t}$ be the one step forecasts of Σ_{t+1} and VaR_{t+1}. Note that, since the log returns are centered, the forecast mean term $M_{t+1|t}$ which appears in (Luo *et al.*, 2023, section 5) is equal to zero. When the innovations

Table V. The empirical coverage rates at nominal level α and p-values of LR_{cc}

Position	α	ε_t: $t_{2,5}$	$t_{2,7}$	$t_{2,10}$	N_2
		Panel A: Empirical coverage rates			
Long	5%	0.055	0.05	0.045	0.042
	1%	0.012	0.012	0.012	0.015
Short	5%	0.043	0.042	0.04	0.04
	1%	0.0075	0.01	0.012	0.017
		Panel B: p-values of LR_{cc}			
Long	5%	0.89	1.00	0.90	0.78
	1%	0.89	0.89	0.89	0.64
Short	5%	0.79	0.79	0.64	0.64
	1%	0.87	1.00	0.89	0.39

ε_t have a multivariate StD with ν degrees of freedom we forecast for a long position at level α as

$$\mathrm{VaR}_{t+1|t} = \tilde{t}_{\nu;\alpha}\sqrt{w^\top \Sigma_{t+1|t} w},$$

where $\tilde{t}_{\nu;\alpha}$ is the size-α quantile of the univariate StD with unit variance and ν degrees of freedom. For a short position, the forecasting is done by replacing α with $1-\alpha$.

The forecasting is implemented mimicking the procedure from Luo *et al.* (2023): the first 400 observations were used to fit the model, and the fitted model is used to forecast $VaR_{t+1|t}$ for $t \geq 400$. We assess the forecast quality by empirical coverage rate and the conditional coverage test LR_{cc} from Christoffersen (1998). The innovation distributions used are bivariate StD with $\nu = 5, 7,$ and 10 degrees of freedom, and the normal distribution. The corresponding results presented in Table V suggest that the referenced GARCH model with innovations following a bivariate StD with 7 degrees of freedom renders well calibrated and relatively safe forecasts.

7. CONCLUSIONS

We suggest a novel approach of testing goodness-of-fit to a StD with i.i.d. data, including GARCH innovations. The tests are easy to implement, consistent, and attain a proper limit law under the null hypothesis. Since this law is complicated for actual use, we apply Monte Carlo approximation or bootstrap resampling to carry out the test in practice. The corresponding finite-sample results show that the new tests are competitive against alternative methods, and for GARCH innovations they are often more powerful by a wide margin. Moreover, the conclusions drawn from a couple of real-data applications are in line with the stylized fact of excess kurtosis of returns in the financial markets. It should finally be noted that while our method applies to any dimension p, the underlying setting is not that of high dimension ($p > n$). In this connection analogous methods for infinite dimensional functional data should be based on the characteristic functional rather than the CF and have been proposed by Jiang *et al.* (2019) and Henze and Jiménez-Gamero (2021), while those in Bugni *et al.* (2009) are based on functional analogues of the distribution function. For the standard setting of high dimension ($p > n$) we refer to Janková *et al.* (2020), Verzelen and Villers (2010), and Kock and Preinerstorfer (2019), but these methods are mostly for testing a high-dimensional parameter, rather than for goodness-of-fit testing for a family of distributions.

ACKNOWLEDGEMENTS

The work of B. Milošević and M. Veljović is supported by the Ministry of Science, Technological Developments and Innovations of the Republic of Serbia. This work is also supported by the COST action CA21163 – Text, functional and other high-dimensional data in econometrics: New models, methods, applications (HiTEc). The authors wish to sincerely thank Professor Ke Zhu (The University of Hong Kong) for providing the code of the test based on the Stein-type characterization. They are also indebted to two anonymous reviewers for the constructive criticism.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Arellano-Valle RB, Azzalini A. 2008. The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis* **99**(7):1362–1382.

Asai M, So MK. 2015. Long memory and asymmetry for matrix-exponential dynamic correlation processes. *Journal of Time Series Econometrics* **7**(1):69–94.

Bai J, Chen Z. 2008. Testing multivariate distributions in GARCH models. *Journal of Econometrics* **143**(1):19–36.

Bardet J-M, Wintenberger O. 2009. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* **37**(5B):2730–2759.

Bauwens L, Laurent S, Rombouts JV. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**(1):79–109.

Berg C, Vignat C. 2010. On the density of the sum of two independent Student t-random vectors. *Statistics & Probability Letters* **80**(13-14):1043–1055.

Billingsley P. 1968. *Convergence of Probability Measures* John Wiley & Sons, New York.

Bilodeau M, Brenner D. 1999. *Theory of Multivariate Statistics* Springer-Verlag, New York, NY.

Blattberg RC, Gonedes NJ. 1974. A comparison of the stable and Student distributions as statistical models for stock prices. *The Journal of Business* **47**(2):244–280.

Bollerslev T. 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics* **72**(3):498–505.

Bosq D. 2000. *Linear Processes in Function Spaces: Theory and Applications*, Vol. 149 Springer Science & Business Media, New York, NY.

Boudt K, Galanos A, Payseur S, Zivot E. 2019. *Multivariate GARCH models for large-scale applications: a survey*. In *Handbook of Statistics*, Vol. 41, Elsevier, Amsterdam, The Netherlands; 193–242.

Bugni F, Hall P, Horowitz J, Neumann G. 2009. Goodness–of–fit for functional data. *Econometrics Journal* **12**:S1–S18.

Chen WY, Gerlach RH. 2021. Semiparametric GARCH via Bayesian model averaging. *Journal of Business & Economic Statistics* **39**(2):437–452.

Chib S, Zeng X. 2020. Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics* **38**(4):771–783.

Christoffersen PF. 1998. Evaluating interval forecasts. *International Economic Review* **39**:841–862.

Comte F, Lieberman O. 2003. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**(1):61–84.

Creal D, Koopman SJ, Lucas A. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* **29**(4):552–563.

Csörgő S. 1981. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **55**(2):203–229.

Dalla V, Bassiakos Y, Meintanis SG. 2017. Characteristic function-based inference for GARCH models with heavy-tailed innovations. *Communications in Statistics-Simulation and Computation* **46**(4):2733–2755.

Diamantopoulos K, Vrontos ID. 2010. A Student-t full factor multivariate GARCH model. *Computational Economics* **35**(1):63–83.

Dube S et al. 2016. Multivariate t-distribution and GARCH modelling of volatility and conditional correlations on BRICS stock markets. *Journal of Applied Finance & Banking* **6**(2):53–82.

Ducharme GR, de Micheaux PL. 2020. A goodness-of-fit test for elliptical distributions with diagnostic capabilities. *Journal of Multivariate Analysis* **178**:104602.

Ebner B, Henze N. 2020. Tests for multivariate normality—a critical review with emphasis on weighted L^2-statistics. *TEST* **29**(4):845–892.

Ebner B, Henze N. 2021. Bahadur efficiency of the Epps–Pulley test for normality. *Zapiski Nauchnykh Seminarov POMI* **501**:302–314.

Ebner B, Henze N, Yukich JE. 2018. Multivariate goodness-of-fit on flat and curved spaces via nearest neighbor distances. *Journal of Multivariate Analysis* **165**:231–242.

Epps T. 2005. Tests for location-scale families based on the empirical characteristic function. *Metrika* **62**:99–114.

Francq C, Zakoïan J-M. 2016. Estimating multivariate volatility models equation by equation. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **78**:613–635.

Francq C, Zakoïan J-M. 2019. *GARCH Models: Structure, Statistical Inference and Financial Applications* John Wiley & Sons, Oxford, UK.

Francq C, Zakoïan J-M. 2022. Testing hypotheses on the innovations distribution in semi-parametric conditional volatility models. *Journal of Financial Econometrics*. https://doi.org/10.1093/jjfinec/nbac011.

Francq C, Zakoïan JM. 2012. QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* **28**(1):179–206.

Francq C, Jiménez-Gamero M, Meintanis S. 2017. Tests for conditional ellipticity in multivariate GARCH models. *Journal of Econometrics* **196**(2):305–319.

Ghoudi K, Rémillard B. 2014. Comparison of specification tests for GARCH models. *Computational Statistics & Data Analysis* **76**:291–300.

Giacomini R, Politis D, White H. 2013. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29**(3):567–589.

Goodman I, Kotz S. 1973. Multivariate θ-generalized normal distributions. *Journal of Multivariate Analysis* **3**:204–219.

Hafner CM, Herwartz H, Maxand S. 2020. Identification of structural multivariate GARCH models. *Journal of Econometrics* **227**(1):212–227.

Hallin M, Mordant G, Segers J. 2021. Multivariate goodness-of-fit tests based on Wasserstein distance. *Electronic Journal of Statistics* **15**(1):1328–1371.

Harvey A, Ruiz E, Sentana E. 1992. Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* **52**(1-2):129–157.

Henze N. 2002. Invariant tests for multivariate normality: a critical review. *Statistical Papers* **43**(4):467–506.

Henze N, Jiménez-Gamero MD. 2019. A new class of tests for multinormality with iid and GARCH data based on the empirical moment generating function. *TEST* **28**(2):499–521.

Henze N, Jiménez-Gamero M. 2021. A test for gaussianity in hilbert spaces via the empirical characteristic functional. *Scandinavian Journal of Statistics* **48**:406–428.

Henze N, Jiménez-Gamero MD, Meintanis SG. 2019. Characterizations of multinormality and corresponding tests of fit, including for GARCH models. *Econometric Theory* **35**(3):510–546.

Janková J, Shah R, Bühlmann P, Samworth R. 2020. Goodness-of-fit testing in high dimensional generalized linear models. *Journal of the Royal Statistical Society Series B* **82**:773–795.

Jeantheau T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**(1):70–86.

Jiang Q, Hušková M, Meintanis S, Zhu L. 2019. Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data. *Journal of Multivariate Analysis* **170**:202–220.

Jiménez-Gamero M, Batsidis A, Alba-Fernández M. 2016. Fourier methods for model selection. *Annals of the Institute of Statistical Mathematics* **68**(1):105–133.

Khmaladze E. 2016. Unitary transformations, empirical processes and distribution free testing. *Bernoulli* **22**(1):563–588.

Klar B, Lindner F, Meintanis SG. 2012. Specification tests for the error distribution in GARCH models. *Computational Statistics & Data Analysis* **56**(11):3587–3598.

Kock A, Preinerstorfer D. 2019. Power in high-dimensional testing problems. *Econometrica* **87**:1055–1069.

Kon SJ. 1984. Models of stock returns – a comparison. *The Journal of Finance* **39**(1):147–165.

Kotz S, Nadarajah S. 2004. *Multivariate t-distributions and Their Applications* Cambridge University Press, Cambridge, UK.

Kotz S, Kozubowski T, Podgorski K. 2012. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance* Springer Science & Business Media, New York, NY.

Lee J, Lee S, Park S. 2015. Maximum entropy test for GARCH models. *Statistical Methodology* **22**:8–16.

Liang J, Ng KW, Tian G. 2019. A class of uniform tests for goodness-of-fit of the multivariate L_p-norm spherical distributions and the L_p-norm symmetric distributions. *Annals of the Institute of Statistical Mathematics* **71**(1):137–162.

Liu C. 1997. ML estimation of the multivariate t distribution and the EM algorithm. *Journal of Multivariate Analysis* **63**(2):296–312.

Luo D, Zhu K, Gong H, Li D. 2023. Testing error distribution by kernelized Stein discrepancy in multivariate time series models. *Journal of Business & Economic Statistics* **41**(1): 111–125. https://doi.org/10.1080/07350015.2021.2002160.

McAssey MP. 2013. An empirical goodness-of-fit test for multivariate distributions. *Journal of Applied Statistics* **40**(5):1120–1131.

McLeish DL. 1974. Dependent central limit theorems and invariance principles. *Annals of Probability* **2**(4):620–628.

Meintanis S, Swanepoel J. 2007. Bootstrap goodness-of-fit tests with estimated parameters based on empirical transforms. *Statistics and Probability Letters* **77**:1004–1013.
Meintanis SG, Jiménez-Gamero MD, Alba-Fernández V. 2014. A class of goodness-of-fit tests based on transformation. *Communications in Statistics-Theory and Methods* **43**(8):1708–1735.
Meintanis S, Milošević B, Obradović M. 2022. Bahadur efficiency for certain goodness–of–fit tests based on the empirical characteristic function. *Metrika* 1–29. https://doi.org/10.1007/s00184-022-00891-0.
Móri T, Székely G, Rizzo M. 2021. On energy tests for normality. *Journal of Statistical Planning and Inference* **213**:1–15.
Nadarajah S, Kotz S. 2008. Estimation methods for the multivariate t distribution. *Acta Applicandae Mathematicae* **102**(1):99–118.
Nardon M, Pianca P. 2009. Simulation techniques for generalized gaussian densities. *Journal of Statistical Computation and Simulation* **79**(11):1317–1329.
Pesaran B, Pesaran MH. 2007. Modelling volatilities and conditional correlations in futures markets with a multivariate t distribution. *IZA Discussion Paper No 2906*; 1–35.
Praetz PD. 1972. The distribution of share price changes. *Journal of Business* **45**:49–55.
Rémillard B. 2011. Validity of the parametric bootstrap for goodness-of-fit testing in dynamic models. *SSRN*. doi:https://doi .org/10.2139/ssrn.1966476.
Rossi E, Spazzini F. 2010. Model and distribution uncertainty in multivariate GARCH estimation: a Monte Carlo analysis. *Computational Statistics & Data Analysis* **54**(11):2786–2800.
Santos AA, Nogales FJ, Ruiz E. 2013. Comparing univariate and multivariate models to forecast portfolio value-at-risk. *Journal of Financial Econometrics* **11**(2):400–441.
Silvennoinen A, Teräsvirta T. 2009. Multivariate GARCH models. In *Handbook of Financial Time Series*, Mikosch T, Kreiß J-P, Davis RA, Gustav Andersen T (eds.), Springer-Verlag, Berlin, Germany; 201–229.
Sutradhar BC. 1986. On the characteristic function of multivariate Student t-distribution. *The Canadian Journal of Statistics* **14**(4):329–337.
Tenreiro C. 2009. On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Computational Statistics & Data Analysis* **53**:1038–1053.
Tenreiro C. 2011. An affine invariant multiple test procedure for assessing multivariate normality. *Computational Statistics & Data Analysis* **55**:1980–1992.
Tenreiro C. 2019. On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation* **89**:1780–1797.
Tenreiro C. 2022. On automatic kernel density estimate-based tests for goodness-of-fit. *TEST* **31**:717–748.
Tsay RS. 2006. Multivariate volatility models. In *Time Series and Related Topics*, Ho H-C, Ing C-K, Leung Lai T (eds.), Institute of Mathematical Statistics, Beachwood, OH; 210–222.
Tsay RS. 2010. *Analysis of Financial Time Series* John Wiley & Sons, Hoboken, NJ.
Ushakov NG. 1999. *Selected Topics in Characteristic Functions* VSP Utrecht, The Netherlands.
Verzelen N, Villers F. 2010. Goodness-of-fit tests for high-dimensional Gaussian linear models. *The Annals of Statistics* **38**:704–752.
Wang Y, Tsay RS. 2013. On diagnostic checking of vector ARMA-GARCH models with Gaussian and Student-t innovations. *Econometrics* **1**(1):1–31.
Zheng Y, Li W, Li G. 2018. A robust goodness–of–fit test generalized autoregressive conditional heteroscedastic models. *Biometrika* **105**(1):73–89.

APPENDIX A

A.1. Regularity assumptions

(RA1) The estimator $\widehat{\theta}_n$ satisfies the Bahadur-type representation, $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} L_j + o_{\mathbb{P}}(1)$, where $L_j = H_j g_j$; $g_j = g(\theta_0; \varepsilon_j)$ is a vector of d^2 measurable functions such that $\mathbb{E}g_j = \mathbf{0}$ and $\mathbb{E}g_j^\top g_j < \infty$; and $H_j = H(\theta_0; \varepsilon_{j-1}, \varepsilon_{j-2,\dots})$ is an $m \times d^2$ matrix of measurable functions satisfying $\mathbb{E}\|H_j^\top H_j\|^2 < \infty$.[3]

(RA2) $\sup_{\theta\in\Theta}\left\|\widetilde{\Sigma}_j^{-1/2}(\theta)\right\| \le C, \sup_{\theta\in\Theta}\left\|\Sigma_j^{-1/2}(\theta)\right\| \le C$ a.s.

(RA3) $\sup_{\theta\in\Theta}\left\|\Sigma_j^{1/2}(\theta) - \widetilde{\Sigma}_j^{1/2}(\theta)\right\| \le C\rho^j$, where $\rho \in [0, 1)$ is a generic constant.

[3] The norm $\|A\|$ of matrix $A = \{a_{ij}\}$ is the L^1 norm defined as $\sum_{i,j} |a_{ij}|$.

(RA4) $\mathbb{E}\left\|X_j\right\|^{\varsigma} < \infty$ and $\mathbb{E}\left\|\Sigma_j^{1/2}\left(\theta_0\right)\right\|^{\varsigma} < \infty$ for some $\varsigma > 0$.

(RA5) For each sequence $x_1, x_2, \dots$ of vectors of $\mathbb{R}^p$, the function $\theta \mapsto \Sigma^{1/2}\left(x_1, x_2, \dots ; \theta\right)$ admits continuous second-order derivatives.

(RA6) For some neighborhood $V\left(\theta_0\right)$ 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

$$\mathbb{E} \sup_{\theta \in V\left(\theta_0\right)} \left\| \Sigma_j^{-1/2}(\theta) \frac{\partial^2 \Sigma_j^{1/2}(\theta)}{\partial \theta_k \partial \theta_\ell} \right\|^p < \infty$$

$$\mathbb{E} \sup_{\theta \in V\left(\theta_0\right)} \left\| \Sigma_j^{-1/2}(\theta) \frac{\partial \Sigma_j^{1/2}(\theta)}{\partial \theta_k} \right\|^q < \infty$$

$$\mathbb{E} \sup_{\theta \in V\left(\theta_0\right)} \left\| \Sigma_j^{1/2}\left(\theta_0\right) \Sigma_j^{-1/2}(\theta) \right\|^r < \infty, \quad 1 \le k, l \le v.$$

(RA7) $\mathbb{E}\|\varepsilon_j\|^4 < \infty$

Conditions (RA1)–(RA7) are equivalent to those typically adopted in the literature under similar settings; see e.g. Francq *et al.* (2017), Henze *et al.* (2019), and Francq and Zakoïan (2019). In this connection Comte and Lieberman (2003) and Bardet and Wintenberger (2009) showed that under certain mild regularity conditions the QMLE satisfies (RA1) for general MGARCH models. Moreover in Francq and Zakoïan (2019, p. 293) one can find the regularity conditions A1–A11, which are closely related to our conditions. In fact, these are the standard regularity conditions on which the consistency and asymptotic normality of the QMLE in the general MGARCH model are established; see Theorem 7 from Francq and Zakoïan (2019), where one can also find the expression for L_j in the Bahadur representation. On the other hand, in the case of the CCC-GARCH model the somewhat weaker and more explicit conditions CC1-CC7 of Francq and Zakoïan (2019, p. 298) (see also Francq and Zakoïan, 2012 and Francq *et al.*, 2017) are sufficient for the QMLE to satisfy (RA1). In addition (RA2)–(RA7) also follow from the aforementioned weaker conditions. For further discussion on the regularity conditions we refer the reader to Francq and Zakoïan (2019, section 10.4).

A.2. Proof of Theorem 4.1

For a symmetric weight function $w(t)$, the test statistic (4.5) can be expressed as

$$\mathcal{T}_{n,w} = \int_{\mathbb{R}^p} U_n^2(t, \widetilde{z}_n(\hat{\theta}_n))\, w(t)\, \mathrm{d}t, \tag{A1}$$

where

$$U_n(t, \widetilde{z}_n(\hat{\theta}_n)) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\sin\left(t^\top \widetilde{z}_j(\hat{\theta}_n)\right) + \cos\left(t^\top \widetilde{z}_j(\hat{\theta}_n)\right) - \varphi_\mu(t) \right), \quad t \in \mathbb{R}^p, \tag{A2}$$

and $\widetilde{z}_n(\theta) = (\widetilde{z}_1(\theta), \dots, \widetilde{z}_n(\theta))$.

We now prove that the empirical process $U_n(t, \widetilde{z}_n(\hat{\theta}_n))$ weakly converges to a certain Gaussian process. A convenient setting for asymptotics is the separable Hilbert space $\mathbb{H}$ of (equivalence classes of) measurable functions $f : \mathbb{R}^p \to \mathbb{R}$ satisfying $\int_{\mathbb{R}^p} f^2(t) w(t)\, \mathrm{d}t < \infty$. The inner product and the norm in $\mathbb{H}$ will be denoted by $\langle f, g \rangle_{\mathbb{H}} = \int_{\mathbb{R}^p} f(t) g(t) w(t)\, \mathrm{d}t$ and $\|f\|_{\mathbb{H}} = \langle f, f \rangle_{\mathbb{H}}^{1/2}$ respectively. With this notation, we have $\mathcal{T}_{n,w} = \|U_n\|_{\mathbb{H}}^2$, where the random element U_n of $\mathbb{H}$ is given in (A2).

  wileyonlinelibrary.com/journal/jtsa
DOI: 10.1111/jtsa.12713 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Set $z_n(\theta) := (z_1(\theta), \dots, z_n(\theta))$ and let

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

be the linear approximation of $z_j(\hat{\theta})$, where $\hat{\theta}_n = \left(\hat{\theta}_{n1}, \dots, \hat{\theta}_{nv}\right)^\top$ and $\theta_0 = \left(\theta_{01}, \dots, \theta_{0v}\right)^\top$.

We now proceed in two parts:

Part I: Find the limiting process of $U_n(t, \boldsymbol{W}_n)$, where $\boldsymbol{W}_n = (W_1, \dots, W_n)$;
Part II: show that

$$\sup_{t\in B_K} |U_n(t, \tilde{z}_n(\hat{\theta}_n)) - U_n(t, \boldsymbol{W}_n)| = o_{\mathbb{P}}(1), \tag{A3}$$

where $B_K = \{t \in \mathbb{R}^p : \|t\| \le K\}$, and K is such that $\int_{\mathbb{R}^p \setminus B_K} \mathbb{E}\|U_n(t, \mathbf{z}_n)\|^2 w(t)dt < \varepsilon$.

Part I. Notice that $\frac{\partial}{\partial \theta_k} z_j(\theta)\Big|_{\theta=\theta_0} = -\Sigma_j^{-1/2}(\theta)\frac{\partial}{\partial \theta_k}\Sigma_j^{1/2}(\theta) z_j$, and denote $A_{jk}(\theta) = \Sigma_j^{-1/2}(\theta)\frac{\partial}{\partial \theta_k}\Sigma_j^{1/2}(\theta)$ and $\alpha_k = \mathbb{E}\left[A_{jk}\left(\theta_0\right)\right]$. Then

$$W_j = z_j - \sum_{k=1}^{v} A_{jk}\left(\theta_0\right) z_j \left(\hat{\theta}_{nk} - \theta_{0k}\right),$$

and

$$U_n(t, \boldsymbol{W}_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\sin\left(t^\top W_j\right) + \cos\left(t^\top W_j\right) - \varphi_\mu(t)\right).$$

Denote for brevity $z_j = z_j(\theta_0)$. A Taylor expansion gives

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin\left(t^\top W_j\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin\left(t^\top z_j\right) + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \cos\left(t^\top z_j\right) \left(t^\top (W_j - z_j)\right) + R_s(t),$$

where

$$\begin{aligned} R_s(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{1}{2} \left(-\sin\left(t^\top \left(\gamma t^\top z_j + (1-\gamma) t^\top W_j\right)\right)\right) \left(t^\top \left(W_j - z_j\right)\right)^2 \\ &= \frac{1}{2\sqrt{n}} \sum_{j=1}^{n} \left(\sin\left(t^\top \left(\gamma t^\top z_j + (1-\gamma) t^\top W_j\right)\right)\right) \left(t^\top \sum_{k=1}^{v} A_{jk}\left(\theta_0\right) \left(\hat{\theta}_{nk} - \theta_{0k}\right) z_j\right)^2, \end{aligned}$$

for some $\gamma \in [0, 1]$. Applying the Cauchy-Schwarz inequality we get

$$\begin{aligned} |R_s(t)| &\le \frac{1}{2n^{\frac{3}{2}}} \sum_{j=1}^{n} \left(t^\top \sum_{k=1}^{v} A_{jk}\left(\theta_0\right) \sqrt{n} \left(\hat{\theta}_{nk} - \theta_{0k}\right) z_j\right)^2 \\ &\le \frac{1}{2n^{\frac{3}{2}}} \sum_{j=1}^{n} \|t\|^2 \left\| \sum_{k=1}^{v} A_{jk}\left(\theta_0\right) \sqrt{n} \left(\hat{\theta}_{nk} - \theta_{0k}\right) z_j \right\|^2 \\ &\le \|t\|^2 O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Similarly we obtain

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos\left(t^{\top}W_j\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos\left(t^{\top}z_j\right) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sin\left(t^{\top}z_j\right)\left(t^{\top}\left(W_j - z_j\right)\right) + R_c(t),$$

where $|R_c(t)| \leq \|t\|^2 O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ as above. Therefore,

$$\begin{aligned} U_n(t, \mathbf{W}_n) &= U_n(t, \boldsymbol{z}_n) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos t^{\top}z_j - \sin t^{\top}z_j\right) t^{\top}\left(W_j - z_j\right) + R_1(t) \\ &= U_n(t, \boldsymbol{z}_n) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos t^{\top}z_j - \sin t^{\top}z_j\right) t^{\top}\sum_{k=1}^{v}A_{jk}\left(\theta_0\right)\left(\hat{\theta}_{nk} - \theta_{0k}\right) z_j + R_1(t), \end{aligned}$$

where $R_1(t) = R_s(t) + R_c(t)$. Further,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos t^{\top}z_j - \sin t^{\top}z_j\right) t^{\top}\sum_{k=1}^{v}A_{jk}\left(\theta_0\right)\left(\hat{\theta}_{nk} - \theta_{0k}\right) z_j \\ &\quad = t^{\top}\sum_{k=1}^{v}\alpha_k\sqrt{n}\left(\hat{\theta}_{nk} - \theta_{0k}\right)\mathbf{\Psi}(t) + \sum_{k=1}^{v}\sqrt{n}\left(\hat{\theta}_{nk} - \theta_{0k}\right)\left(\frac{1}{n}\sum_{j=1}^{n}V_{kj}(t) + \frac{1}{n}\sum_{j=1}^{n}U_{kj}(t)\right), \end{aligned}$$

where

$$\begin{aligned} V_{kj}(t) &= t^{\top}A_{jk}\left(\theta_0\right)\left(\left(\cos t^{\top}z_j - \sin t^{\top}z_j\right) z_j - \mathbf{\Psi}(t)\right) \\ U_{kj}(t) &= t^{\top}\left(A_{jk}\left(\theta_0\right) - \alpha_k\right)\mathbf{\Psi}(t), \end{aligned}$$

and

$$\mathbf{\Psi}(t) - \mathbb{E}\left(z_j \sin t^{\top}z_j\right) = \nabla\mathbb{E}\cos\left(t^{\top}z_j\right) = \nabla\varphi_{\mu}(t).$$

After some calculation the components of the vector $\mathbf{\Psi}(t) = (\Psi(t_1), \ldots, \Psi(t_d))$ can be expressed as

$$\begin{aligned} \Psi(t_i) &= -\mathbb{E}z_j \sin t_i z_j \prod_{r\neq i}\mathbb{E}\cos t_r z_j \\ &= \frac{2^{d(1-\mu)}(2\mu)^{\frac{d\mu+1}{2}}}{(\Gamma(\mu))^d}\cdot(-t_i)\left|t_i\right|^{\mu-1}K_{1-\mu}\left(\sqrt{2\mu}\,|t_i|\right)\prod_{r\neq i}\left|t_j\right|^{\mu}K_{-\mu}\left(\sqrt{2\mu}\,|t_r|\right). \end{aligned}$$

Notice that $\mathbb{E}\left[V_{kj}\right] = 0, \mathbb{E}\left[\left\langle V_{kj}, V_{kr}\right\rangle\right] = 0, \forall j \neq r$, and $\mathbb{E}\left\|V_{kj}\right\|_{\mathrm{L2}}^2 < \infty$. Therefore, $\left\|\frac{1}{n}\sum_{j=1}^{n}V_{kj}(t)\right\|_{L_2} = o_{\mathbb{P}}(1)$. From the ergodic theorem $\frac{1}{n}\sum_{j=1}^{n}\left\{A_{jk}\left(\theta_0\right) - \alpha_k\right\} \to 0$ a.s., $1 \leq k \leq v$, so $\sup_{t\in B_k}\left\|\frac{1}{n}\sum_{j=1}^{n}U_{kj}(t)\right\|_{L_2} = o_{\mathbb{P}}(1)$. Therefore,

$$\sup_{t\in B_k}\left\|\sum_{k=1}^{v}\sqrt{n}\left(\hat{\theta}_{nk} - \theta_{0k}\right)\left(\frac{1}{n}\sum_{j=1}^{n}V_{kj}(t) + \frac{1}{n}\sum_{j=1}^{n}U_{kj}(t)\right)\right\|_{L_2} = o_{\mathbb{P}}(1).$$

DOI: 10.1111/jtsa.12713 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Finally, we obtain

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos t^{\top}z_j-\sin t^{\top}z_j\right)t^{\top}\sum_{k=1}^{v}A_{jk}\left(\theta_0\right)\left(\hat{\theta}_{nk}-\theta_{0k}\right)z_j$$
$$=t^{\top}\sum_{k=1}^{v}\alpha_k\sqrt{n}\left(\hat{\theta}_{nk}-\theta_{0k}\right)\boldsymbol{\Psi}(t)+o_{\mathbb{P}}(1),$$

and

$$U_n(t,\boldsymbol{W}_n)=U_n(t,z_n)-t^{\top}\sum_{k=1}^{m}\alpha_k\sqrt{n}\left(\hat{\theta}_{nk}-\theta_{0k}\right)\boldsymbol{\Psi}(t)+o_{\mathbb{P}}(1). \quad \text{(A4)}$$

From the central limit theorem in Hilbert spaces (see e.g. Bosq, 2000), the summand $U_n(t,z_n)$, as an i.i.d. sum of elements from a Hilbert space $\mathbb{H}$, converges to a Gaussian random element in this Hilbert space.

Since by Assumption RA1, $\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}L_j+o_{\mathbb{P}}(1)$, applying the central limit theorem for martingale differences (see e.g. McLeish, 1974) we get that $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}L_j$ converges weakly to a zero mean Gaussian random vector. Hence the second summand of (A4), being a product of a continuous function and a term which is $O_P(1)$ is tight, and converges in $C(B_K)$, the Banach space of real-valued continuous functions on B_K, to a Gaussian random element (see Billingsley, 1968).

Expressing (A4) as

$$U_n(t,\boldsymbol{W}_n)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Xi_j(t)+o_{\mathbb{P}}(1),$$

where

$$\Xi_j(t)=\sin(t^{\top}z_j)+\cos(t^{\top}z_j)-\varphi_{\mu}(t)+t^{\top}\langle\boldsymbol{\alpha},L_j\rangle\boldsymbol{\Psi}(t),$$

we get that $U_n(t,\boldsymbol{W}_n)$ converges to a zero mean random element G with covariance function

$$K_G(s,t)=\mathbb{E}\Xi_1(s)\Xi_1(t). \quad \text{(A5)}$$

Part II. Now we show (A3) in two steps:

(a) $\sup_{t\in B_K}\left|U_n\left(t,\tilde{z}_j(\hat{\theta}_n)\right)-U_n\left(t,z_j(\hat{\theta}_n)\right)\right|=o_{\mathbb{P}}(1)$

(b) $\sup_{t\in B_K}\left|U_n\left(t,z_j(\hat{\theta}_n)\right)-U_n\left(t,W_n\right)\right|=o_{\mathbb{P}}(1)$

The term under supremum in (a) is

$$U_n\left(t,\tilde{z}_j(\hat{\theta}_n)\right)-U_n\left(t,z_j(\hat{\theta}_n)\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\sin\left(t^{\top}\tilde{z}_j(\hat{\theta}_n)\right)+\cos\left(t^{\top}\tilde{z}_j(\hat{\theta}_n)\right)$$
$$-\sin\left(t^{\top}z_j(\hat{\theta}_n)\right)-\cos\left(t^{\top}z_j(\hat{\theta}_n)\right)\Big), \quad \text{(A6)}$$

for $t\in\mathbb{R}^p$. We now show that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\sin\left(t^{\top}\tilde{z}_j(\hat{\theta}_n)\right)-\sin\left(t^{\top}z_j(\hat{\theta}_n)\right)\right)=o_{\mathbb{P}}(1).$$

To this end, a Taylor expansion yields

$$\sin\left(t^\top \tilde{z}_j\left(\hat{\theta}_n\right)\right) - \sin\left(t^\top z_j\left(\hat{\theta}_n\right)\right) = t^\top \Lambda_{n,j} \cos\left(t^\top z_j\left(\hat{\theta}_n\right) + \alpha_{n,j} t^\top \Lambda_{n,j}\right)$$

where

$$\Lambda_{n,j} = \tilde{z}_j(\hat{\theta}_n) - z_j(\hat{\theta}_n) = \tilde{\Sigma}_j^{-1/2}\left(\hat{\theta}_n\right)\left(\Sigma_j^{1/2}\left(\hat{\theta}_n\right) - \tilde{\Sigma}_j^{1/2}\left(\hat{\theta}_n\right)\right)\Sigma_j^{-1/2}\left(\hat{\theta}_n\right) X_j,$$

for some $\alpha_{n,j} \in (0,1)$.

Then

$$\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\sin\left(t^\top \tilde{z}_j\left(\hat{\theta}_n\right)\right) - \sin\left(t^\top z_j\left(\hat{\theta}_n\right)\right)\right]\right| \le \frac{1}{\sqrt{n}}\sum_{j=1}^{n}|t^T \Lambda_{n,j}| \le \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\Lambda_{n,j}\|\|t\|,$$

and from conditions RA2 and RA3 we get that $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\Lambda_{n,j}\| = o_{\mathbb{P}}(1)$. The cosine part of (A6) is proven analogously and a) follows.

Turning to (b), consider the term under the supremum

$$\begin{aligned}
&|U_n(t, \boldsymbol{W}_n) - U_n(t, \boldsymbol{z}_n(\hat{\theta}_n))| \\
&= \left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\sin\left(t^\top W_j\right) + \cos\left(t^\top W_j\right) - \sin\left(t^\top z_j(\hat{\theta}_n)\right) - \cos\left(t^\top z_j(\hat{\theta}_n)\right)\right)\right| \\
&\le \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left|2\sin\frac{1}{2}\left(t^\top\left(W_j - z_j(\hat{\theta}_n)\right)\right)\cos\frac{1}{2}\left(t^\top\left(W_j + z_j(\hat{\theta}_n)\right)\right)\right. \\
&\quad \left. -\ 2\sin\frac{1}{2}\left(t^\top\left(W_j - z_j(\hat{\theta}_n)\right)\right)\sin\frac{1}{2}\left(t^\top\left(W_j + z_j(\hat{\theta}_n)\right)\right)\right| \\
&\le \frac{1}{n}\sum_{j=1}^{n} 4\sqrt{n}\left|\sin\frac{1}{2}\left(t^\top\left(W_j - z_j(\hat{\theta}_n)\right)\right)\right|,
\end{aligned}$$

and since

$$\sqrt{n}\left(W_j - z_j(\hat{\theta}_n)\right) = -\sum_{k,l=1}^{v}\frac{\partial^2}{\partial\theta_k\theta_l} z_j(\theta)\Bigg|_{\theta=\theta_*}\sqrt{n}\left(\hat{\theta}_{nk} - \theta_{0k}\right)\left(\hat{\theta}_{nl} - \theta_{0l}\right) = o_{\mathbb{P}}(1),$$

we get that $|U_n(t, \boldsymbol{W}_n) - U_n(t, \boldsymbol{z}_n(\hat{\theta}_n))| \le \|t\| \cdot o_{\mathbb{P}}(1)$, and (b) follows.

Applying the continuous mapping theorem to the representation (A1) ends the proof.