Tests for sphericity in multivariate garch models

Francq, Christian and Jiménez Gamero, Maria Dolores and Meintanis, Simos

September 2015

Tests for sphericity in multivariate garch models

C. Francq[a], M.D. Jiménez–Gamero[d], S.G. Meintanis[b,c,1]

[a] *CREST and University Lille 3, BP 60149, 59653 Villeneuve d'Ascq cedex, France*

[b] *Department of Economics, National and Kapodistrian University of Athens Athens, Greece*

[c] *Unit for Business Mathematics and Informatics, North–West University Potchefstroom, South Africa*

[d] *Department of Statistics and Operations Research, University of Sevilla Sevilla, Spain*

Abstract. Tests for spherical symmetry of the innovation distribution are proposed in multivariate GARCH models. The new tests are of Kolmogorov–Smirnov and Cramér–von Mises–type and make use of the common geometry underlying the characteristic function of any spherically symmetric distribution. The asymptotic null distribution of the test statistics as well as the consistency of the tests is investigated under general conditions. It is shown that both the finite sample and the asymptotic null distribution depend on the unknown distribution of the Euclidean norm of the innovations. Therefore a conditional Monte Carlo procedure is used to actually carry out the tests. The validity of this resampling scheme is formally justified. Results on the behavior of the test in finite–samples are included, as well as an application on financial data.

Keywords. Extended CCC-GARCH; Spherical symmetry; Empirical characteristic function; Conditional Monte Carlo test.

JEL classification : C12, C15, C32, C58

[1] On sabbatical leave from the University of Athens

1 Introduction

For $d \geq 1$, consider the multivariate observation vector $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{dt})'$, from the model

$$\text{(1.1)} \quad \boldsymbol{y}_t = \mathbf{C}_t^{1/2} \boldsymbol{\varepsilon}_t,$$

where the (unobserved) random errors $\{\boldsymbol{\varepsilon}_t\}_t$ (also referred to as innovations), are independent and follow an unspecified distribution which remains invariant with respect to time t, and has mean zero and identity covariance matrix. We assume that given the information set available at time t, the conditional covariance matrix of $\boldsymbol{y}_t$ equals $\mathbf{C}_t$, $\mathbf{C}_t$ being a $(d \times d)$ symmetric and positive definite matrix. This is the setting of the multivariate GARCH (MGARCH) model, and under this model we are interested in testing the null hypothesis of spherical symmetry for the distribution of the innovations. Specifically, and on the basis of observations $\{\boldsymbol{y}_t, \ t = 1, \ldots, T\}$ driven by the equation (1.1), we wish to test the null hypothesis

$$\mathcal{H}_0 : \text{ the law of } \{\boldsymbol{\varepsilon}_t\}_t \text{ belongs to the family of spherically symmetric laws} \in \mathbb{R}^d,$$

against general alternatives. Note that the hypothesis that $\{\boldsymbol{\varepsilon}_t\}_t$ belongs to the class of spherically symmetric distributions (SSD) is equivalent to assuming that the corresponding distribution is invariant under the group of transformations $\boldsymbol{\varepsilon}_t \mapsto \mathbf{H}\boldsymbol{\varepsilon}_t$, where $\mathbf{H}$ is any orthogonal $(d \times d)$–matrix.

The null hypothesis $\mathcal{H}_0$ implies a model that lies somewhere between a fully parametric MGARCH, and an MGARCH model with a completely unspecified innovation distribution. Of course in the i.i.d. setting, the importance of the class of SSD is well known: Several notions and procedures extend nicely from the classical Gaussian context to spherical symmetry; see, for instance, Jones (2008), Cacoullos (2014), Zuo and Serfling (2000), Hallin and Paindaveine (2002), and Hallin and Werker (2003). On the other hand, and in the context of dynamic models, it may been shown, see e.g. Embrechts et al. (2002) and Berk (1997), that an innovation distribution belonging to the SSD class renders model (1.1) conveniently amenable to standard approaches of risk management such as Value–at–Risk and the mean–variance approach to risk management and portfolio optimization. Hence spherical symmetry has often been a point of departure for financial data. In fact, many fully parametric versions make use of innovation distributions belonging to the family of SSD. Examples are the Gaussian (M)GARCH specification of Bai and Chen (2008), Lee et al. (2010), and Lee et al. (2014), or its Student–t counterpart. For further families and for statistical procedures within dynamic models involving SSDs see Amengual and Sentana (2011), and Liu et al. (2011). For more general specification tests in conditional models the reader is referred to Delgado and Stute (2008) and Koul and Stute (1999). As already mentioned, an MGARCH model with a completely unspecified innovation distribution may also be entertained; see Hafner and Rombouts (2007) for instance. However even in this case, Hafner and Rombouts (2007) assume an innovation distribution in the SSD class for their nonparametric estimator of the innovation distribution to avoid the 'curse of dimensionality' and capture the univariate convergence rate. Nevertheless and despite the popularity of the SSD class, there is recently a strong tendency to allow for skewness in GARCH models for financial returns, and one way to do so is via the conditional distribution of the observations; see Mittnik and Paolella (2000), Bauwens and Laurent (2005), De Luca et al. (2006), Trindade and Zhu (2007), Haas et al. (2009), and Chen et al. (2012). This recent tendency in conjunction with the earlier bias towards a SSD for the innovations provides the ground on the basis of which the null hypothesis $\mathcal{H}_0$ could be considered as highly relevant, particularly in statistical modelling with a view towards financial applications.

For i.i.d. data, there exist several works on testing spherical symmetry; see for instance Koltchinskii and Li (1998), Baringhaus (1991), Kariya and Eaton (1977) and the review article by Meintanis and Ngatchou–Wandji (2012). Tests for conditional symmetry maybe found in Bai and Ng (2001) and Delgado and Escanciano (2007). The method presented here however is related more with the approaches suggested by Ghosh and Ruymgaart (1992), Diks and Tong (1999), Zhu and Neuhaus (2000), Zhu (2005) and Henze et al. (2014). The common theme in all these works is that the authors use specific properties of the characteristic function of SSDs in their test statistics.

The purpose of this paper is to extend the test procedure of Henze et al. (2014) from the i.i.d. context to models involving dependence, with special emphasis on

MGARCH models. In doing so we derive the limit properties of the procedure under GARCH-type dependence. In addition we suggest and show the consistency of a modified version of the resampling counterpart of the test statistic employed in Henze et al. (2014). Although in the proofs we make use of constant correlations, our simulations also include time–dependent correlations.

In order to introduce the proposed procedure, let $\boldsymbol{X} \in \mathbb{R}^d$ be an arbitrary random variable with corresponding characteristic function (CF) $\varphi(\mathbf{u}) = E[\exp(\mathrm{i}\mathbf{u}'\boldsymbol{X})]$, $\mathbf{u} \in \mathbb{R}^d$. We will make use of the following characterization of SSD: The CF $\varphi(\mathbf{u})$ is the CF of a SSD if and only if there exists some function $\phi : \mathbb{R} \to \mathbb{R}$ such that

$$\varphi(\mathbf{u}) = \phi(\|\mathbf{u}\|^2), \tag{1.2}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. Characterization (1.2) may be found in Fang et al. (1990), together with a wealth of material on SSD.

Along the lines proposed by Henze et al. (2014), we suggest to use the process

$$\Delta_T(\mathbf{u},\mathbf{v}) = \varphi_T(\mathbf{u}) - \varphi_T(\mathbf{v}),\ \ \mathbf{u},\mathbf{v} \in \mathbb{R}^d,$$

where

$$\varphi_T(\mathbf{u}) = \frac{1}{T}\sum_{t=1}^{T} e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t},$$

is the empirical CF of the innovations $\boldsymbol{\varepsilon}_t = \mathbf{C}_t^{-1/2}\boldsymbol{y}_t,\ t = 1,\ldots,T$. Then, in view of characterization (1.2) and the consistency of the empirical CF, we expect that for large T, the value of $\Delta_T(\mathbf{u},\mathbf{v})$ should be close to zero under the null hypothesis $\mathcal{H}_0$ provided that this value is computed over pairs of points $\mathbf{u},\mathbf{v}$ such that $\|\mathbf{u}\| = \|\mathbf{v}\|$.

Since $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_T$, are unobserved, any decision regarding the innovation–distribution should naturally be based on the residuals

$$\widetilde{\boldsymbol{\varepsilon}}_t = \widetilde{\mathbf{C}}_t^{-1/2}\boldsymbol{y}_t,\ t = 1,\ldots,T,$$

where $\widetilde{\mathbf{C}}_t$ denotes an appropriate estimator of the covariance matrix $\mathbf{C}_t$ that will be detailed later. Specifically, we consider test statistics involving the process

$$D_T(\mathbf{u},\mathbf{v}) = \widetilde{\varphi}_T(\mathbf{u}) - \widetilde{\varphi}_T(\mathbf{v}), \tag{1.3}$$

where

$$\widetilde{\varphi}_T(\mathbf{u}) = \frac{1}{T}\sum_{t=1}^{T} e^{\mathrm{i}\mathbf{u}'\widetilde{\boldsymbol{\varepsilon}}_t},$$

is the empirical CF of the residuals $\widetilde{\boldsymbol{\varepsilon}}_t, \ t = 1, \ldots, T$.

The remainder of this paper is outlined as follows. In Section 2, the test statistics are defined, while Section 3 we discuss procedures of estimating the covariance matrix $\mathbf{C}_t$ under specific versions of MGARCH models. In Section 4 large–sample properties of the proposed tests are studied, while in Section 5 we introduce and prove the validity of a resampling scheme that removes the drawbacks encountered when one relies entirely on asymptotics in order to actually carry out the tests. Simulations and a real data application are presented in Section 6, while in the last part of the paper in Section 7 we draw some conclusions and consider possible extensions. All proofs, as well as some intermediate results, are sketched in Section 8.

2 Test statistics

We consider Kolmogorov-Smirnov (KS) and Cramér–von–Mises type (CM) test statistics involving the process $D_T(\mathbf{u}, \mathbf{v})$. Specifically and since $D_T(\cdot, \cdot)$ is a complex function, for the purpose of testing the null hypothesis $\mathcal{H}_0$ we shall monitor the values of the function $|D_T(\mathbf{u}, \mathbf{v})|$, over pairs of points $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^d$ which are equidistant from the origin. Intuitively and in view of characterization (1.2), we expect these values to be 'small' under the null hypothesis and as $T \to \infty$, and consequently large values of this function should lead to rejection of $\mathcal{H}_0$. However, any test statistic conducted on the basis of this characterization should, at least in principle, take into account the full variation of this function over all possible such pairs $(\mathbf{u}, \mathbf{v})$. As a compromise, we choose for a fixed integer $K \geq 1$, a finite collection

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_K\} \in \mathbb{S}_\circ,$$

of points lying in the unit sphere $\mathbb{S}_\circ := \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$, and which are scattered as uniformly as possible over $\mathbb{S}_\circ$. We shall base our test statistics on the variation of $|D_T(\mathbf{u}, \mathbf{v})|$ realized over pairs $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^d$ lying in directions away from the origin specified by this collection. For the CM test statistic we do not limit the extend to which the points $(\mathbf{u}, \mathbf{v})$ will stretch away from the origin. For the KS statistic however we restrict this range by defining, for a fixed integer $L \geq 1$, another finite collection of points

$$0 < \rho_1 < \rho_2 < \ldots \rho_L < \infty,$$

and by considering the variation of $|D_T(\mathbf{u}, \mathbf{v})|$ over points $(\mathbf{u}, \mathbf{v})$ such that, $\|\mathbf{u}\| = \|\mathbf{v}\| = \rho_\ell$, $\forall \ell \in \{1, \ldots, L\}$.

Based on the above notation and reasoning we suggest to reject the null hypothesis $\mathcal{H}_0$ for large values of the KS test statistic

$$\mathrm{KS}_T = \sqrt{T} \max_{l=1,\ldots,L} \max_{j,m=1,\ldots,K} |D_T(\rho_l \mathbf{u}_j, \rho_l \mathbf{u}_m)| \, .$$

Likewise the proposed CM test statistic is defined as,

$$\mathrm{CM}_T = T \int_0^\infty \left(\sum_{j,m=1}^{K} |D_T(\rho \mathbf{u}_j, \rho \mathbf{u}_m)|^2 \right) \omega(\rho) d\rho,$$

where $\omega(\cdot)$ denotes a nonnegative weight function satisfying

$$\int_0^\infty \omega(\rho) d\rho < \infty, \qquad \int_0^\infty \rho^2 \omega(\rho) d\rho < \infty. \tag{2.1}$$

In fact, if we let $\widetilde{\boldsymbol{\varepsilon}}_{st} = \widetilde{\boldsymbol{\varepsilon}}_s - \widetilde{\boldsymbol{\varepsilon}}_t$ and

$$I_\omega(\mathbf{z}) := \int_0^\infty \cos(\rho \mathbf{z}) \omega(\rho) d\rho,$$

then direct computation yields

$$\mathrm{CM}_T = \frac{1}{T} \sum_{j,m=1}^{K} \sum_{s,t=1}^{T} \left[I_\omega(\mathbf{u}_j' \widetilde{\boldsymbol{\varepsilon}}_{st}) + I_\omega(\mathbf{u}_m' \widetilde{\boldsymbol{\varepsilon}}_{st}) - 2 I_\omega(\mathbf{u}_m' \widetilde{\boldsymbol{\varepsilon}}_s - \mathbf{u}_j' \widetilde{\boldsymbol{\varepsilon}}_t) \right],$$

which shows that a suitable choice of the weight function, such as $\omega(\rho) = e^{-a\rho^b}$, with $a > 0$, and $b = 1$ or $b = 2$, renders the CM test statistic in a closed form convenient for computations. Note that both test statistics are computed on the basis of the residuals obtained and that therefore we should also consider the problem of computing an estimate $\widetilde{\mathbf{C}}_t$ of the covariance matrix $\mathbf{C}_t$, as this estimate is used in (1.3) in order to obtain the residuals. Estimation of $\mathbf{C}_t$ will be carried out next in the context of specific MGARCH structures.

3 Estimation under GARCH models

There exist several versions of MGARCH models. The reader is referred to Tsay (2014), Francq and Zakoïan (2010) and Silvennoinen and Teräsvirta (2009) for some recent accounts. To introduce MGARCH consider the covariance matrix in (1.1) and write $\mathbf{C}_t := \mathbf{C}_t(\boldsymbol{\vartheta})$ to indicate that this matrix depends on a parameter vector $\boldsymbol{\vartheta}$. Different versions of MGARCH deviate in the specification of the dependence structure of $\mathbf{C}_t$ with respect to the past, one of the main issues being the dimensionality of the parameter $\boldsymbol{\vartheta}$ with increasing dimension d. A specific instance of MGARCH which is both intuitively and computationally attractive is the so–called extended constant conditional correlation (E)CCC–GARCH(p, q) model. This specification is defined by

$$(3.1) \quad \mathbf{C}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t,$$

where $\mathbf{D}_t$ and $\mathbf{R}$ are $(d \times d)$ matrices with $\mathbf{D}_t$ diagonal and $\mathbf{R}$ being a correlation matrix. If $\mathbf{A}$ is a square matrix, then $\text{diag}(\mathbf{A})$ denotes the vector of the diagonal elements of $\mathbf{A}$. If $\mathbf{a}$ is a vector, then $\text{diag}(\mathbf{a})$ denotes the diagonal matrix whose diagonal is $\mathbf{a}$. The matrix $\mathbf{D}_t$ is related to a volatility vector $\boldsymbol{\sigma}_t = \text{diag}(\mathbf{D}_t^2)$ by

$$(3.2) \quad \boldsymbol{\sigma}_t = \boldsymbol{b} + \sum_{j=1}^{p} \mathbf{B}_j \boldsymbol{y}_{t-j}^{(2)} + \sum_{j=1}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\sigma}_{t-j},$$

$\boldsymbol{y}_t^{(2)} = \boldsymbol{y}_t \odot \boldsymbol{y}_t$, $\odot$ denoting the Hadamard product, that is, the element by element product. In (3.2), the vector $\boldsymbol{b}$ is of dimension d and has positive elements, while $\{\mathbf{B}_j\}_{j=1}^{p}$ and $\{\boldsymbol{\Gamma}_j\}_{j=1}^{q}$ are $(d \times d)$ matrices with non–negative elements. The CCC-GARCH model has been introduced by Bollerslev (1990) with diagonal matrices $\mathbf{B}_j$ and $\boldsymbol{\Gamma}_j$. We consider here the extended version of Jeantheau (1998), in which the matrices of (3.2) are allowed to be nondiagonal.

The model (3.2) could be extended by introducing asymmetries, as in Francq and Zakoïan (2012). This would not change the resampling scheme that we propose in Section 5 below, but would entail heavier notation and additional technical difficulties. We therefore concentrate on the formulation (3.2) for the theoretical results, but we will consider alternative GARCH formations in the applications.

As observed before, an estimator of the covariance matrices $\mathbf{C}_t$, $t = 1, \ldots, T$, is required in order to calculate the residuals. Note that $\mathbf{C}_t$ depends on $\{\boldsymbol{y}_k, t - p \leq$

$k \leq t-1\}$ and $\{\boldsymbol{\sigma}_k,\ t-q \leq k \leq t-1\}$, whereas we only observe $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$. Because of this reason initial values $(\boldsymbol{y}_{1-p}, \ldots, \boldsymbol{y}_0)$ and $(\widetilde{\boldsymbol{\sigma}}_{1-q}, \ldots, \widetilde{\boldsymbol{\sigma}}_0)$ are necessary in order to start the recursion implied by (3.1) and (3.2), and we shall denote by $\widetilde{\mathbf{C}}_t$ the covariance matrix computed recursively on the basis of the aforementioned initial values.

The standard estimation method for the Gaussian MGARCH model is maximum likelihood. However, it has been shown that even with non–Gaussian innovations, under quite general conditions, maximizing the Gaussian likelihood leads to a consistent and asymptotically normal estimator (see, e.g., Francq and Zakoïan, 2010). This estimator is called the quasi–MLE (QMLE), and is formally defined as

$$\widehat{\boldsymbol{\vartheta}}_T = \arg\max_{\boldsymbol{\vartheta} \in \Theta} \mathcal{L}_T(\boldsymbol{\vartheta}),$$

where Θ denotes the parameter space,

$$\mathcal{L}_T(\boldsymbol{\vartheta}) = -\frac{1}{2} \sum_{t=1}^{T} \widetilde{\boldsymbol{\ell}}_t,$$

and

$$\widetilde{\boldsymbol{\ell}}_t := \widetilde{\boldsymbol{\ell}}_t(\boldsymbol{\vartheta}) = \boldsymbol{y}_t' \widetilde{\mathbf{C}}_t^{-1} \boldsymbol{y}_t + \log \left| \widetilde{\mathbf{C}}_t \right|.$$

Note that, as it is well known, the initial values $(\boldsymbol{y}_{1-p}, \ldots, \boldsymbol{y}_0)$ and $(\widetilde{\boldsymbol{\sigma}}_{1-q}, \ldots, \widetilde{\boldsymbol{\sigma}}_0)$ have no influence on the asymptotic properties of the QMLE.

4 Asymptotic properties

Here and in what follows, the notation $\xrightarrow{\mathcal{D}}$ means convergence in distribution of random elements and random variables, $\mathbf{o}_{\mathrm{P}}(1)$ stands for a vector consisting of $o_{\mathrm{P}}(1)$ elements and all limits are taken when $T \to \infty$. We now distinguish the true value $\boldsymbol{\vartheta}_0$ of the parameter and a generic element $\boldsymbol{\vartheta}$ of the parameter space Θ. Denoting by $r_{\ell j}$ the element of the row ℓ and column j of the matrix $\mathbf{R}$, we can write

$$\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{s_0})' \quad = \quad (\boldsymbol{b}', \mathrm{vec}'(\mathbf{B}_1), \ldots, \mathrm{vec}'(\boldsymbol{\Gamma}_q), \mathbf{r}')',$$

where $\mathbf{r}' = (r_{21}, \ldots, r_{d1}, r_{32}, \ldots, r_{d,d-1}) \in \mathbb{R}^{s_2}$, $s_0 = s_1 + s_2$ with $s_1 = d + (p+q)d^2$ and $s_2 = d(d-1)/2$. If necessary, we write $\widetilde{\mathbf{C}}_t(\boldsymbol{\vartheta})$ or $\widetilde{\mathbf{C}}_t(\boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1; \boldsymbol{\vartheta})$ instead of $\widetilde{\mathbf{C}}_t$, but we keep the simplest notation when there is no confusion. A similar convention is used for other terms, such as $\mathbf{D}_t(\boldsymbol{\vartheta})$ or $\widetilde{\mathbf{D}}_t(\boldsymbol{\vartheta})$. For any matrix $\mathbf{A} = (a_{\ell j})$, we will use the norm defined by $\|\mathbf{A}\| = \sum_{\ell,j} |a_{\ell j}|$; if $\mathbf{A}$ is a vector, $\|\mathbf{A}\|$ denotes the Euclidean norm.

We obtain the asymptotic null distribution of the test statistics under an arbitrary estimator $\widehat{\boldsymbol{\vartheta}}_T$ of the parameter $\boldsymbol{\vartheta}_0$. In doing so we assume an asymptotic representation for $\widehat{\boldsymbol{\vartheta}}_T$ which is relatively general and applies to most estimators of interest, such as the QMLE (see Lemma 4.2). Also in order to obtain consistency of the proposed tests we impose a weak condition on the CF under innovation distributions not belonging to the family of SSD. In particular:

(A.1) Assume that the estimator is strongly consistent and admits the following representation,

$$\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\psi}_{0,t-1} \boldsymbol{g}_{0t} + \mathbf{o}_{\mathrm{P}}(1),$$

where $\boldsymbol{g}_{0t} := \boldsymbol{g}(\boldsymbol{\vartheta}_0, \boldsymbol{\varepsilon}_t)$ is a vector of d^2 measurable functions such that $E(\boldsymbol{g}_{0t}) = \mathbf{0}$ and $E(\boldsymbol{g}'_{0t}\boldsymbol{g}_{0t})^2 < \infty$, and $\boldsymbol{\psi}_{0t} := \boldsymbol{\psi}(\boldsymbol{\vartheta}_0; \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$ is a $s_0 \times d^2$ matrix of measurable functions such that $E\|\boldsymbol{\psi}_{0t}\boldsymbol{\psi}'_{0t}\|^2 < \infty$.

(A.2) Assume that under a fixed alternative distribution, the CF of the innovation distribution satisfies

$$\varphi(\rho_0 \mathbf{u}_0) \neq \varphi(\rho_0 \mathbf{v}_0),$$

for some $\mathbf{u}_0, \mathbf{v}_0 \in \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ and some $\rho_0 \in \{\rho_1, \ldots, \rho_L\}$.

The necessary and sufficient condition for the existence of a (unique) strictly stationary (and non anticipative) solution to the CCC-GARCH model defined by (1.1) and (3.1)-(3.2) is $\gamma_0 < 0$, where γ_0 is the top-Lyapounov exponent of the model (as defined by (2.23) and (11.36) in Francq and Zakoïan, 2010). The number γ_0 depends, in a non explicit way, on $\boldsymbol{\vartheta}_0$ and on the distribution of $\boldsymbol{\varepsilon}_t$ and its value can be evaluated by Monte Carlo simulations. We will also assume that the parameter is identifiable (uniqueness of the parametrization). Several types of conditions can be used to ensure it. Here we will assume that Assumption **A4** below holds. Although a bit restrictive, it is quite simple; for weaker alternative conditions ensuring the identifiability see for instance, Reinsel, 1997, p. 37–40. Denote by $\mathbf{I}_d$ the $d \times d$ identity matrix, and by $\mathbf{e}_j$ the j-th column of $\mathbf{I}_d$. Let $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = \sum_{j=1}^{p} \mathbf{B}_j z^j$ when $p > 0$ and $\mathcal{G}_{\boldsymbol{\vartheta}}(z) = \mathbf{I}_d - \sum_{j=1}^{q} \boldsymbol{\Gamma}_j z^j$ when $q > 0$.

The following assumptions will be assumed to derive all results in this section.

A1: $\boldsymbol{\vartheta}_0 \in \Theta$ and Θ is a compact subset of $(0, +\infty)^d \times [0, +\infty)^{d^2(p+q)} \times (-1, 1)^{d(d-1)/2}$.

A2: $\gamma_0 < 0$ and $\forall \boldsymbol{\vartheta} \in \Theta$, $|\mathcal{G}_{\boldsymbol{\vartheta}}(z)| = 0 \Rightarrow |z| > 1$.

A3: For $j = 1, \ldots, d$ the distribution of $\mathbf{e}_j' \boldsymbol{\varepsilon}_1$ is not concentrated on two points and $P(\mathbf{e}_j' \boldsymbol{\varepsilon}_1 > 0) \in (0, 1)$.

A4: If $q > 0$ then $\mathcal{B}_{\boldsymbol{\vartheta}_0}(1) \neq 0$, the polynomials $\mathcal{B}_{\boldsymbol{\vartheta}_0}(z)$ and $\mathcal{G}_{\boldsymbol{\vartheta}_0}(z)$ are left coprime and the matrix $[\mathbf{B}_{0p}\, \boldsymbol{\Gamma}_{0q}]$ has full rank d.

A5: $\mathbf{R}$ is a positive-definite correlation matrix for all $\boldsymbol{\vartheta} \in \Theta$.

A6: $\boldsymbol{\vartheta}_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ is the interior of Θ.

A7: $E\|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\|^2 < \infty$.

The following result gives the asymptotic null distribution of the test statistics KS_T and CM_T.

Theorem 4.1 *Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$, follow a MGARCH model as specified by (1.1), (3.1) and (3.2), and assume that (2.1) and* **A1**-**A7** *hold. Assume that $\widehat{\boldsymbol{\vartheta}}_T$ satisfies (A.1). Then under the null hypothesis $\mathcal{H}_0$,*

$$\text{(4.1)} \quad \mathrm{KS}_T \xrightarrow{\mathcal{D}} \max_{l=1,\ldots,L} \max_{j,m=1,\ldots,K} |\mathcal{W}(\rho_l \mathbf{u}_j, \rho_l \mathbf{u}_m)|$$

and

$$\text{(4.2)} \quad \mathrm{CM}_T \xrightarrow{\mathcal{D}} \int_0^\infty \left(\sum_{j,m=1}^{K} |\mathcal{W}(\rho \mathbf{u}_j, \rho \mathbf{u}_m)|^2 \right) \omega(\rho) d\rho,$$

where $\mathcal{W} = \{\mathcal{W}(\mathbf{u}, \mathbf{v});\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^d\}$ is a complex valued zero–mean Gaussian random field with covariance kernel equal to that of

$$\text{(4.3)} \quad g_1(\mathbf{u}) - g_1(\mathbf{v}),$$

where $g_t(\mathbf{u}) = e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t} + \boldsymbol{g}'_{0t}\boldsymbol{\psi}'_{0,t-1}\dot{\boldsymbol{\varphi}}(\mathbf{u})$ *and* $\dot{\boldsymbol{\varphi}}(\cdot)$ *is a real vector defined in the proof.*

Because of its convenient properties, a commonly used estimator of $\boldsymbol{\vartheta}$ is the QMLE. The next Lemma shows that it satisfies (A.1).

Lemma 4.2 *Under Assumptions* **A1**-**A7**, *the QMLE* $\widehat{\boldsymbol{\vartheta}}_T$ *satisfies (A.1).*

The last result in this section gives the asymptotic behavior of the test statistics KS_T and CM_T under alternatives.

Theorem 4.3 *Suppose that the assumptions of Theorem 4.1 are satisfied, but instead of the null hypothesis* $\mathcal{H}_0$, *consider any fixed alternative innovation distribution satisfying (A.2). Then we have*

$$\liminf \frac{\mathrm{KS}_T}{\sqrt{T}} \geq |\varphi(\rho_0\mathbf{u}_0) - \varphi(\rho_0\mathbf{v}_0)|\,, \tag{4.4}$$

and

$$\liminf \frac{\mathrm{CM}_T}{T} \geq \int_0^\infty |\varphi(\rho\mathbf{u}_0) - \varphi(\rho\mathbf{v}_0)|^2\,\omega(\rho)d\rho, \tag{4.5}$$

almost surely.

Remark 4.4 *As a result of Theorems 4.1 and 4.3, the test which rejects the null hypothesis* $\mathcal{H}_0$ *for large values of the test statistic* KS_T *(resp.* CM_T*) is consistent against each non–spherically symmetric alternative innovation distribution satisfying (A.2).*

5 A conditional resampling scheme

Both the finite–sample and the asymptotic distribution of the test statistics under the null hypothesis $\mathcal{H}_0$ of spherical symmetry depend on the unknown distribution of the Euclidean norm of the underlying random vector $\boldsymbol{\varepsilon}_t$. From Theorem 4.1 it is also clear that these null distributions also depend on the estimator of $\boldsymbol{\vartheta}$ employed. A well known result which will be used below is that $\boldsymbol{\varepsilon}_t = \|\boldsymbol{\varepsilon}_t\|(\boldsymbol{\varepsilon}_t/\|\boldsymbol{\varepsilon}_t\|)$ and that under $\mathcal{H}_0$, $\|\boldsymbol{\varepsilon}_t\|$ and $\boldsymbol{\varepsilon}_t/\|\boldsymbol{\varepsilon}_t\|$ are independent, and the latter random variable follows a uniform distribution over the unit sphere $\mathbb{S}_\circ$. In view of these observations, we consider the following conditional resampling scheme, given the data $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$, where for simplicity we write $\mathcal{T}$ for the test statistic:

(i) Calculate $\widehat{\boldsymbol{\vartheta}}_T = \widehat{\boldsymbol{\vartheta}}_T(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$, the residuals $\widetilde{\boldsymbol{\varepsilon}}_1, \ldots, \widetilde{\boldsymbol{\varepsilon}}_T$ and the test statistic $\mathcal{T} := \mathcal{T}(\widetilde{\boldsymbol{\varepsilon}}_1, \ldots, \widetilde{\boldsymbol{\varepsilon}}_T)$.

(ii) Generate vectors $\boldsymbol{s}_t^*$, $t = 1, \ldots, T$, that are independent and uniformly distributed on $\mathbb{S}_\circ$, independently generate vectors $\underline{\boldsymbol{\varepsilon}}_t^*$, $t = 1, \ldots, T$, that are independent and uniformly distributed on $\{\widetilde{\boldsymbol{\varepsilon}}_{01}, \ldots, \widetilde{\boldsymbol{\varepsilon}}_{0T}\}$, where $\widetilde{\boldsymbol{\varepsilon}}_{0j} = S_T^{-1/2}(\widetilde{\boldsymbol{\varepsilon}}_j - \widetilde{\boldsymbol{\varepsilon}}_.)$, $\widetilde{\boldsymbol{\varepsilon}}_.$ is the sample mean of the residuals and S_T is the sample covariance matrix of the residuals, compute $\boldsymbol{\varepsilon}_t^* = \|\underline{\boldsymbol{\varepsilon}}_t^*\| \boldsymbol{s}_t^*$ and let $\boldsymbol{y}_t^* = \widetilde{\mathbf{C}}_t^{1/2}(\boldsymbol{y}_{t-1}^*, \ldots, \boldsymbol{y}_1^*; \widehat{\boldsymbol{\vartheta}}_T)\boldsymbol{\varepsilon}_t^*$, $t = 1, \ldots, T$.

(iii) Calculate $\widehat{\boldsymbol{\vartheta}}_T^* = \widehat{\boldsymbol{\vartheta}}_T(\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_T^*)$, the resampling residuals

$$\widetilde{\boldsymbol{\varepsilon}}_t^* = \widetilde{\mathbf{C}}_t^{-1/2}(\boldsymbol{y}_{t-1}^*, \ldots, \boldsymbol{y}_1^*; \widehat{\boldsymbol{\vartheta}}_T^*)\boldsymbol{y}_t^*, \qquad t = 1, \ldots, T$$

and the test statistic $\mathcal{T}^* = \mathcal{T}(\widetilde{\boldsymbol{\varepsilon}}_1^*, \ldots, \widetilde{\boldsymbol{\varepsilon}}_T^*)$.

(iv) Repeat steps (ii) and (iii) a number of times B and calculate the corresponding test statistic values $\mathcal{T}_1^*, \ldots, \mathcal{T}_B^*$.

(v) Reject the null hypothesis if $\mathcal{T} > \mathcal{T}_{(B-[\alpha B])}^*$, where $\mathcal{T}_{(1)}^* \leq \ldots \leq \mathcal{T}_{(B)}^*$ denote the corresponding order statistics and α denotes the prescribed level of significance.

Let us write $\boldsymbol{\varepsilon}_{t,T}^*$ and $\boldsymbol{y}_{t,T}^*$ instead of $\boldsymbol{\varepsilon}_t^*$ and $\boldsymbol{y}_t^*$ when it is useful to emphasize that the distributions of these random vectors depend on T. Observe that $E\left(\boldsymbol{\varepsilon}_{1,T}^* \mid \boldsymbol{y}\right) = E(\boldsymbol{\varepsilon}_1) = \mathbf{0}$ and $E\left(\boldsymbol{\varepsilon}_{1,T}^* \boldsymbol{\varepsilon}_{1,T}^{*'} \mid \boldsymbol{y}\right) = E(\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1') = \mathbf{I}_d$.

Note that, at least in the case where $\widehat{\boldsymbol{\vartheta}}_T$ is the QMLE, one can avoid the costly optimizations required in step (iii). Indeed the QMLE is obtained by iterating a Newton-Raphson equation. A standard solution for avoiding the optimization (see *e.g.* Kreiss et al. (2011), Shimizu (2013), Francq et al. (2014), and the references therein) consists in bootstrapping a single Newton-Raphson iteration. In the present framework, using the notations in the proof of Lemma 4.2, this leads to set

$$(5.1) \quad \widehat{\boldsymbol{\vartheta}}_T^* = \widehat{\boldsymbol{\vartheta}}_T + \frac{1}{T}\sum_{t=1}^{T} \widetilde{\boldsymbol{\psi}}_{t-1}\widetilde{\boldsymbol{g}}_t^*, \qquad \widetilde{\boldsymbol{\psi}}_{t-1} = -\mathbf{J}_T^{-1}\widetilde{\boldsymbol{\Delta}}_{t-1},$$

where

$$\mathbf{J}_T = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_T)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}, \quad \widetilde{\boldsymbol{g}}_t^* = \mathrm{vec}\left(\mathbf{I}_d - \widehat{\mathbf{R}}^{1/2}\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}\widehat{\mathbf{R}}^{-1/2}\right),$$

$\mathbf{e}_j'\widetilde{\boldsymbol{\Delta}}_{t-1} = 2\mathrm{vec}\left(\widehat{\mathbf{D}}_t^{-1}\widehat{\mathbf{D}}_t^{(j)}\right)$ for $j = 1,\ldots,s_1$, and $\mathbf{e}_j'\widetilde{\boldsymbol{\Delta}}_{t-1} = \mathrm{vec}\left(\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{R}}^{(j)}\right)$ for $j = s_1+1,\ldots,s_0$, with $\widehat{\mathbf{R}} = \widetilde{\mathbf{R}}(\widehat{\boldsymbol{\vartheta}}_T)$, $\widehat{\mathbf{D}}_t = \widetilde{\mathbf{D}}_t(\widehat{\boldsymbol{\vartheta}}_T)$ and similar notations for the derivatives.

The next result gives the asymptotic behavior of the proposed test statistics when evaluated on the bootstrap residuals, say KS_T^* and CM_T^*. Let $\boldsymbol{\varepsilon}_{01}$ be distributed as $\|\boldsymbol{\varepsilon}_1\|\boldsymbol{s}$, with $\boldsymbol{s}$ uniformly distributed on $\mathbb{S}_\circ$ and independent of $\|\boldsymbol{\varepsilon}_1\|$. Observe that if the null hypothesis holds then $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_{01}$ both have the same distribution. Let $\varphi_0(\mathbf{u})$ denote the CF of $\boldsymbol{\varepsilon}_{01}$.

Theorem 5.1 *Let $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T$ follow a MGARCH model as specified by (1.1), (3.1) and (3.2), and assume that (2.1) and* **A1**-**A7** *hold. Assume that the distribution of $\|\boldsymbol{\varepsilon}_1\|$ admits a bounded density f with respect to the Lebesgue measure. Assume that $\widehat{\boldsymbol{\vartheta}}_T$ is the QMLE, and that the condition $\widehat{\boldsymbol{\vartheta}}_T^* = \widehat{\boldsymbol{\vartheta}}_T(\boldsymbol{y}_1^*,\ldots,\boldsymbol{y}_T^*)$ in (iii) is replaced by* (5.1). *Then for almost all sequences $\boldsymbol{y} = \{\boldsymbol{y}_t\}$ satisfying (1.1), (3.1), (3.2), and conditionally on $\boldsymbol{y}$,*

$$\text{(5.2)} \quad \mathrm{KS}_T^* \xrightarrow{\mathcal{D}} \max_{l=1,\ldots,L}\ \max_{j,m=1,\ldots,K} |\mathcal{W}_0(\rho_l\mathbf{u}_j, \rho_l\mathbf{u}_m)|,$$

and

$$\text{(5.3)} \quad \mathrm{CM}_T^* \xrightarrow{\mathcal{D}} \int_0^\infty \left(\sum_{j,m=1}^{K} |\mathcal{W}_0(\rho\mathbf{u}_j, \rho\mathbf{u}_m)|^2\right)\omega(\rho)d\rho,$$

where $\mathcal{W}_0$ is as defined in Theorem 4.1 when the innovations are distributed as $\boldsymbol{\varepsilon}_{01}$.

The result in Theorem 5.1 is valid whether the null hypothesis is true or not. When the null hypothesis is true, the limits in Theorems 4.1 and 5.1 coincide, and thus the proposed bootstrap provides a consistent approximation to the null distribution of the test statistics. On the other hand, if the null hypothesis is not true, we still have that $\mathrm{KS}_T^* = O_P(1)$ and $\mathrm{CM}_T^* = O_P(1)$. In view of the result in Theorem 4.3, it is concluded that the bootstrap test is consistent against each non–spherically symmetric alternative innovation distribution satisfying (A.2).

6 Monte Carlo results and application

This section is devoted to the study of the finite–sample performance of the proposed tests in terms of level approximation and power. With this aim a Monte Carlo simulation study was conducted. We first considered a bivariate CCC–GARCH(1,1) model with

$$\boldsymbol{b} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, \ \mathbf{B}_1 = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \ \boldsymbol{\Gamma}_1 = \begin{pmatrix} 0.2 & 0.1 \\ 0.01 & 0.3 \end{pmatrix}, \ \mathbf{R} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix},$$

for $r = 0, 0.3$. For the distribution of the innovations we took $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T$, i.i.d. from the distribution of $\boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon}$ as in the following cases:

(i) $\boldsymbol{\varepsilon} \sim N_2(0, \mathbf{I}_2)$,

(ii)-(iv) $\boldsymbol{\varepsilon} = |t_\nu| R$, where $R = (R_1, R_2)'$ is uniformly distributed on the unit circle and the random variable t_ν has t-distribution with ν d.f., $\nu = 5, 6, 7$,

(v) $\boldsymbol{\varepsilon} = (Z_1, Z_2)'$, with Z_1, Z_2 i.i.d. from an asymmetric exponential power distribution (Zhu and Zinde-Walsh, 2009) with parameters $\alpha = 0.4$, $p_1 = 1.182$ and $p_2 = 1.820$ ($\mu = 0$, $\sigma = 1$).

(vi) $\boldsymbol{\varepsilon} = (Z_1, Z_2)'$, with Z_1, Z_2 independent, $Z_1 \sim N(0, 1)$, $Z_2 \sim t_5$.

(vii) the distribution of $\boldsymbol{\varepsilon}$ is an equal mixture of two bivariate normal distributions with unit covariance matrices and means $(-1.5, 0)'$ and $(1.5, 0)'$,

(viii) $\boldsymbol{\varepsilon}$ has a skew-normal distribution with direct parameters (Arellano-Valle and Azzalini, 2008) $\mu = (0, 0)$, $\Sigma = \mathbf{I}_2$ and $\alpha = (0, 0.25)$.

Each distribution was suitably modified so that $E(\boldsymbol{\varepsilon}_t) = 0$ and $Var(\boldsymbol{\varepsilon}_t) = \mathbf{I}_2$. The cases (i)–(iv) obey $\mathcal{H}_0$, while cases (v)–(viii) correspond to the alternative hypothesis.

There is a number of parameters that should be specified for the application of the test statistics. These parameters affect the performance of the tests. Here however we do not investigate this aspect of the methods and remain within the suggestions made by Henze et al. (2014) for the values of these user parameters. In particular for the

KS test statistic we took $K = 9$ grid points and $L = 8$ with $\rho_l = l/L$ (denoted in the tables as $\text{KS}^{(1)}$) and $\rho_l = 2l/L$ (denoted in the tables as $\text{KS}^{(2)}$), $l = 1, \ldots, L$; likewise, for the CM test statistic we took $K = 17$ grid points. Also for the CM statistic we use the weight function $\omega(\rho) = \exp(-a\rho^b)$ for ease of computation. The values of b employed are $b = 1$ (denoted in the tables as $\text{CM}^{(1)}$) and $b = 2$ (denoted in the tables as $\text{CM}^{(2)}$), while for the values of a we took $a = 0.4$ for $b = 1$ and $a = 0.1$ for $b = 2$.

The parameters in the CCC-GARCH models were estimated by QMLE using the package `ccgarch` of the language R. Tables 1 and 2 report the percentages of rejections for nominal significance levels $\alpha = 0.01, 0.05, 0.10$ and sample sizes $T = 300, 400$, for $r = 0$ and $r = 0.3$, respectively. In order to reduce the computational burden we adopted the 'Warp–Speed' method of Giacomini et al. (2013) for evaluating the resampling scheme proposed in Section 5. With the Warp–Speed method, rather than computing critical points for each Monte Carlo sample, one resample is generated for each Monte Carlo sample and the resampling test statistic $\mathcal{T}^*$ is computed for that sample. Then the resampling critical values for $\mathcal{T}$ are computed from the empirical distribution determined by the resampling repetitions of $\mathcal{T}^*$. In our simulations we took $10,000$ Monte Carlo samples.

We repeated the above experiment for a trivariate CCC–GARCH(1,1) model with $\boldsymbol{b} = (0.1, 0.1, 0.1)'$,

$$\mathbf{B}_1 = \begin{pmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix}, \ \mathbf{\Gamma}_1 = \begin{pmatrix} 0.2 & 0.1 & 0.01 \\ 0.01 & 0.3 & 0.1 \\ 0.01 & 0.01 & 0.1 \end{pmatrix}, \ \mathbf{R} = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix},$$

with r as before. For the distribution of the innovations we considered the following generalizations of the cases considered before:

(i) $\boldsymbol{\varepsilon} \sim N_3(0, \mathbf{I}_3)$,

(ii)-(iv) $\boldsymbol{\varepsilon} = |t_\nu| R$, where $R = (R_1, R_2, R_3)'$ is uniformly distributed on the unit sphere and the random variable t_ν has t-distribution with ν d.f., $\nu = 5, 6, 7$,

(v) $\boldsymbol{\varepsilon} = (Z_1, Z_2, Z_3)'$, with Z_1, Z_2, Z_3 i.i.d. from an asymmetric exponential power distribution (Zhu and Zinde-Walsh, 2009) with parameters $\alpha = 0.4$, $p_1 = 1.182$ and $p_2 = 1.820$ ($\mu = 0$, $\sigma = 1$).

		$T = 300$				$T = 400$			
case	α	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$
(i)	0.01	0.64	1.08	1.30	1.28	1.20	1.00	1.00	0.96
	0.05	4.96	5.08	4.98	4.86	5.74	4.82	5.08	5.06
	0.10	9.52	10.74	10.34	10.28	10.74	10.08	9.98	9.70
(ii)	0.01	1.00	1.08	1.08	1.14	0.88	1.28	1.24	1.20
	0.05	5.10	5.36	5.62	5.72	5.61	5.56	6.42	6.20
	0.10	9.84	9.90	11.02	11.02	10.66	10.67	11.93	11.83
(iii)	0.01	1.09	0.91	0.94	1.00	1.07	1.02	1.23	1.15
	0.05	5.23	4.88	5.62	5.50	5.23	4.92	5.37	5.36
	0.10	10.20	9.37	10.30	10.51	10.02	10.59	11.16	11.16
(iv)	0.01	1.14	1.17	1.21	1.18	0.92	1.24	1.09	1.03
	0.05	5.19	5.20	5.19	5.13	4.96	5.30	5.84	5.49
	0.10	10.56	9.81	10.50	10.52	9.83	10.64	10.67	10.82
(v)	0.01	1.80	1.36	1.86	1.82	1.78	1.36	1.88	1.76
	0.05	7.44	6.60	7.92	8.12	6.66	7.40	8.46	8.28
	0.10	12.94	12.60	15.90	15.50	13.32	14.10	16.02	15.56
(vi)	0.01	1.40	5.64	5.24	5.6	1.16	10.26	7.56	8.16
	0.05	5.92	20.98	18.26	19.2	5.74	30.80	25.92	26.94
	0.10	10.80	35.24	29.38	30.8	11.02	46.64	38.92	41.44
(vii)	0.01	4.94	19.60	22.34	23.16	5.60	49.60	46.24	47.76
	0.05	14.92	61.20	58.20	61.30	17.94	85.88	81.30	82.62
	0.10	24.90	81.04	77.24	78.26	29.14	95.20	92.32	93.40
(viii)	0.01	1.04	8.56	47.60	46.68	0.82	12.90	66.30	65.00
	0.05	4.50	22.42	70.10	68.98	4.90	30.62	84.86	84.02
	0.10	8.96	33.48	79.74	79.40	10.20	42.56	90.62	89.84

Table 1: Percentages of rejections of KS and CM tests in dimension $d = 2$ and $r = 0$.

		$T = 300$				$T = 400$			
case	α	$\text{KS}^{(1)}$	$\text{KS}^{(2)}$	$\text{CM}^{(1)}$	$\text{CM}^{(2)}$	$\text{KS}^{(1)}$	$\text{KS}^{(2)}$	$\text{CM}^{(1)}$	$\text{CM}^{(2)}$
(i)	0.01	1.00	0.66	1.26	1.06	1.18	1.10	1.12	1.24
	0.05	4.40	5.44	5.42	5.28	5.70	5.82	5.82	5.86
	0.10	9.74	11.00	10.42	10.36	10.52	9.98	10.74	10.80
(ii)	0.01	1.24	1.0	1.04	0.98	1.09	1.01	0.87	0.95
	0.05	5.18	5.3	5.64	5.60	4.92	4.96	5.24	5.24
	0.10	10.52	10.7	10.78	10.42	10.16	10.18	10.63	10.76
(iii)	0.01	1.06	1.03	0.90	0.94	1.07	1.24	1.36	1.50
	0.05	4.57	4.90	5.30	5.21	4.97	5.57	5.84	5.94
	0.10	9.78	10.07	10.64	10.52	10.02	11.24	10.90	10.94
(iv)	0.01	1.00	1.04	1.27	1.30	0.91	0.99	1.20	1.22
	0.05	4.98	5.02	5.21	5.27	5.21	5.00	5.26	5.03
	0.10	9.88	10.39	10.17	10.33	10.19	10.07	10.31	10.22
(v)	0.01	1.50	0.88	1.82	1.62	1.72	1.40	1.76	1.90
	0.05	6.28	6.62	7.88	7.86	7.58	6.08	7.90	7.98
	0.10	12.24	12.62	14.24	14.08	13.72	12.48	15.66	15.88
(vi)	0.01	1.14	4.38	4.66	5.40	1.14	8.68	8.10	8.76
	0.05	6.04	18.78	17.90	18.56	6.16	30.08	25.48	26.68
	0.10	10.90	33.38	30.40	31.24	10.68	48.24	38.04	39.64
(vii)	0.01	4.54	19.82	18.76	20.36	4.94	39.64	34.86	35.82
	0.05	13.28	60.82	55.88	58.30	18.70	80.64	75.68	78.78
	0.10	23.38	79.12	76.82	78.18	29.86	92.70	90.38	91.34
(viii)	0.01	1.24	6.96	44.42	42.18	0.76	12.70	62.98	61.56
	0.05	4.68	21.40	70.62	69.38	4.62	28.28	83.16	82.72
	0.10	8.66	33.38	80.52	80.08	10.02	40.58	89.94	89.40

Table 2: Percentages of rejections of KS and CM tests in dimension $d = 2$ and $r = 0.3$.

(vi) $\boldsymbol{\varepsilon} = (Z_1, Z_2, Z_3)'$, with Z_1, Z_2, Z_3 independent, $Z_1, Z_2 \sim N(0,1)$, $Z_3 \sim t_5$.

(vii) the distribution of $\boldsymbol{\varepsilon}$ is an equal mixture of two trivariate normal distributions with unit covariance matrices and means $(-1.5, 0, 0)'$ and $(1.5, 0, 0)'$,

(viii) $\boldsymbol{\varepsilon}$ has a skew-normal distribution with direct parameters (Arellano-Valle and Azzalini, 2008) $\mu = (0, 0)$, $\Sigma = \mathbf{I}_3$ and $\alpha = (0, 0, 0.25)$.

In light of the results in Tables 1–4, the following comments can be made: the level of all the tests is quite close to the nominal value, specially for lighter tail distributions. Also it is observed that the true level becomes closer to the nominal one as ν increases in cases (ii) to (iv) and (i) (which corresponds to $\nu = \infty$); as for the power, we see that higher powers are obtained for distributions "less spherically symmetric"; as expected, the power increases with the sample size; the powers for $r = 0$ are a bit larger than for $r = 0.3$; overall, the CM tests are more powerful than the KS tests, while the power of the two CM tests considered is quite close.

In addition to the above models we also tried another one for $d = 2$ with more persistence, that is, with larger diagonal elements of Γ_1. Specifically, we took

$$\mathbf{B}_1 = \begin{pmatrix} 0.01 & 0.01 \\ 0.01 & 0.01 \end{pmatrix}, \ \boldsymbol{\Gamma}_1 = \begin{pmatrix} 0.85 & 0.01 \\ 0.01 & 0.85 \end{pmatrix},$$

$\boldsymbol{b}$ and $\mathbf{R}$ as before. Tables 5–6 display the obtained results. Similar comments can be made.

As observed in Section 3, there are many ways to generalize the univariate GARCH model to the multivariate setting. In this paper we consider a particular model, the constant conditional correlation model. Nevertheless, under suitable conditions, the proposed tests could be applied for other specifications of the conditional covariance matrix. Moreover, different conclusions could be reached when applying the proposed tests for different fitted models. 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 (see Example 9.2 in Tsay (2002), the data is available from the website of the author `http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/`). This data set was also analyzed in Bai and Chen (2008), where the authors tested normality

case	α	$T=300$				$T=400$			
		$\mathrm{KS}^{(1)}$	$\mathrm{KS}^{(2)}$	$\mathrm{CM}^{(1)}$	$\mathrm{CM}^{(2)}$	$\mathrm{KS}^{(1)}$	$\mathrm{KS}^{(2)}$	$\mathrm{CM}^{(1)}$	$\mathrm{CM}^{(2)}$
(i)	0.01	0.85	0.86	1.09	1.10	0.66	0.87	1.01	1.10
	0.05	4.83	5.02	5.68	5.40	4.68	5.19	4.94	4.90
	0.10	9.81	10.03	11.10	10.94	9.66	10.30	9.66	9.65
(ii)	0.01	1.09	0.94	1.16	1.13	1.07	1.16	1.71	1.77
	0.05	5.64	5.31	6.07	6.38	4.90	5.47	6.50	6.23
	0.10	10.67	11.04	12.40	12.16	9.86	11.32	12.40	12.37
(iii)	0.01	1.31	1.08	1.54	1.44	1.04	1.16	1.48	1.41
	0.05	5.33	5.56	5.64	5.48	5.44	5.38	5.91	6.09
	0.10	10.86	10.77	11.20	11.23	11.01	10.10	11.35	11.69
(iv)	0.01	0.87	0.99	1.17	1.23	1.05	1.21	1.10	1.12
	0.05	5.08	4.86	5.64	5.58	5.11	5.33	5.36	5.20
	0.10	9.71	9.81	11.34	11.17	10.24	10.34	10.32	10.18
(v)	0.01	0.80	1.88	2.00	1.92	1.52	1.84	2.54	2.60
	0.05	5.14	7.00	7.46	7.22	6.12	7.94	9.52	9.44
	0.10	10.40	13.84	13.62	13.78	11.32	15.60	15.90	16.30
(vi)	0.01	0.90	4.70	3.18	3.58	0.92	6.92	4.16	4.62
	0.05	5.36	18.58	12.94	13.34	4.42	27.46	16.68	17.82
	0.10	10.18	31.52	22.48	23.58	9.64	41.38	28.68	30.80
(vii)	0.01	1.25	3.14	6.50	6.51	1.40	6.80	12.57	13.72
	0.05	6.32	21.49	25.26	26.37	7.76	35.94	40.23	42.59
	0.10	13.27	38.44	41.38	43.12	14.70	55.45	59.85	62.58
(viii)	0.01	0.98	6.44	45.62	44.66	0.94	10.50	67.48	66.28
	0.05	4.86	20.82	68.20	67.04	4.36	26.22	83.86	83.02
	0.10	9.60	32.60	79.04	78.24	9.44	38.96	90.04	89.40

Table 3: Percentages of rejections of KS and CM tests in dimension $d = 3$ and $r = 0$.

case	α	$T = 300$				$T = 400$			
		$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$
(i)	0.01	1.05	1.28	1.08	1.14	0.87	1.42	1.34	1.38
	0.05	4.72	5.25	4.84	5.10	4.70	5.59	5.29	5.46
	0.10	10.13	9.63	9.43	9.29	9.20	10.58	10.28	10.10
(ii)	0.01	0.85	1.04	1.43	1.36	1.19	1.28	1.42	1.35
	0.05	4.90	5.29	6.02	6.09	5.13	5.69	6.52	6.39
	0.10	10.41	11.12	11.62	11.97	10.09	10.99	12.37	12.40
(iii)	0.01	1.20	1.22	1.47	1.47	1.02	1.56	1.51	1.42
	0.05	5.18	5.24	5.28	5.28	5.28	5.61	6.23	6.38
	0.10	10.06	10.47	10.53	10.70	10.33	10.93	12.26	12.08
(iv)	0.01	0.97	1.48	1.31	1.28	0.89	1.12	1.50	1.58
	0.05	4.75	5.45	5.80	6.08	4.79	5.37	6.06	6.01
	0.10	9.36	10.98	11.56	11.90	10.13	10.38	11.15	11.59
(v)	0.01	1.22	1.48	1.20	1.20	1.34	1.86	2.20	2.18
	0.05	5.66	6.54	7.06	7.24	6.18	7.92	8.24	8.26
	0.10	11.22	14.14	1 3.72	13.84	11.10	14.90	15.84	16.14
(vi)	0.01	1.16	3.32	2.08	2.32	0.80	5.52	2.80	3.10
	0.05	4.74	16.42	10.54	10.84	5.12	23.48	14.20	14.86
	0.10	9.30	27.58	19.78	20.76	10.36	39.46	26.34	26.96
(vii)	0.01	2.11	2.96	6.59	6.77	2.26	7.24	11.19	12.84
	0.05	7.49	16.19	26.32	27.81	7.86	33.22	39.32	41.39
	0.10	13.71	34.05	43.16	45.17	13.24	53.28	59.47	60.99
(viii)	0.01	0.88	6.26	43.50	42.00	0.74	9.56	63.02	62.24
	0.05	4.42	20.60	68.18	67.20	4.64	28.48	82.16	81.20
	0.10	8.96	31.22	76.82	76.62	9.36	40.40	89.60	88.98

Table 4: Percentages of rejections of KS and CM tests in dimension $d = 3$ and $r = 0.3$.

		$T = 300$				$T = 400$			
case	α	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$
(i)	0.01	0.97	0.97	0.96	0.92	0.71	1.05	0.96	1.06
	0.05	4.56	4.58	4.54	4.62	4.62	5.03	4.89	5.12
	0.10	9.84	9.57	9.29	9.47	9.74	10.15	9.61	9.78
(ii)	0.01	1.12	0.89	1.23	1.37	0.96	1.27	1.23	1.16
	0.05	5.23	5.83	5.75	5.95	5.03	5.62	5.87	5.86
	0.10	10.17	11.44	11.83	11.76	9.85	10.76	11.16	11.35
(v)	0.01	1.74	1.44	1.48	1.42	1.66	1.22	2.18	1.90
	0.05	7.72	7.02	7.26	6.94	6.66	6.98	8.62	8.60
	0.10	13.54	13.30	14.56	14.84	12.46	14.04	15.54	15.52
(vi)	0.01	1.30	4.90	6.18	6.66	0.92	8.48	8.20	8.84
	0.05	6.12	22.76	20.30	20.64	5.08	31.14	25.32	27.90
	0.10	11.84	35.68	32.42	33.86	10.30	47.56	39.62	41.70
(vii)	0.01	3.36	20.92	22.32	24.78	5.92	41.68	41.70	44.42
	0.05	13.58	62.98	58.92	60.34	17.48	84.24	78.72	81.00
	0.10	22.20	79.46	76.58	78.00	30.32	94.94	91.32	92.30
(viii)	0.01	0.72	7.76	48.62	48.02	0.70	13.22	66.04	63.54
	0.05	5.20	21.96	71.62	70.98	4.96	30.58	85.64	84.54
	0.10	9.46	33.32	82.60	82.00	10.24	42.54	91.20	91.00

Table 5: Percentages of rejections of KS and CM tests in dimension $d = 2$ and $r = 0$.

case	α	$T=300$				$T=400$			
		$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$	$KS^{(1)}$	$KS^{(2)}$	$CM^{(1)}$	$CM^{(2)}$
(i)	0.01	0.93	0.78	0.80	0.79	0.97	0.66	0.80	0.76
	0.05	5.14	4.93	4.94	4.70	4.92	4.90	4.75	4.64
	0.10	9.85	9.56	9.30	9.78	10.05	9.75	9.59	9.35
(ii)	0.01	0.85	1.05	1.34	1.27	0.96	1.11	1.34	1.25
	0.05	5.01	5.38	6.04	6.50	5.89	5.06	5.77	5.89
	0.10	9.93	10.84	12.08	11.86	10.67	10.44	11.26	11.25
(v)	0.01	1.00	1.20	1.60	1.66	1.32	1.66	1.96	2.04
	0.05	6.58	5.94	6.68	6.72	6.84	6.96	8.44	8.46
	0.10	12.34	11.54	14.10	14.44	13.46	14.32	15.64	15.62
(vi)	0.01	1.00	5.18	4.26	4.62	1.40	9.58	7.98	8.86
	0.05	5.74	19.48	18.98	19.60	5.58	31.70	24.52	25.92
	0.10	11.54	33.36	31.30	33.50	11.32	47.78	39.76	42.06
(vii)	0.01	3.92	22.60	24.76	26.14	5.96	42.94	40.96	44.66
	0.05	13.68	61.18	57.14	59.52	19.02	84.08	78.98	80.40
	0.10	22.96	79.88	76.74	78.20	29.90	94.40	91.24	92.44
(viii)	0.01	1.04	8.16	46.60	45.96	0.92	12.42	67.48	65.30
	0.05	5.02	23.74	72.30	71.74	4.94	29.78	85.20	84.68
	0.10	10.42	35.46	82.74	82.26	10.28	43.36	91.28	90.92

Table 6: Percentages of rejections of KS and CM tests in dimension $d=2$ and $r=0.3$.

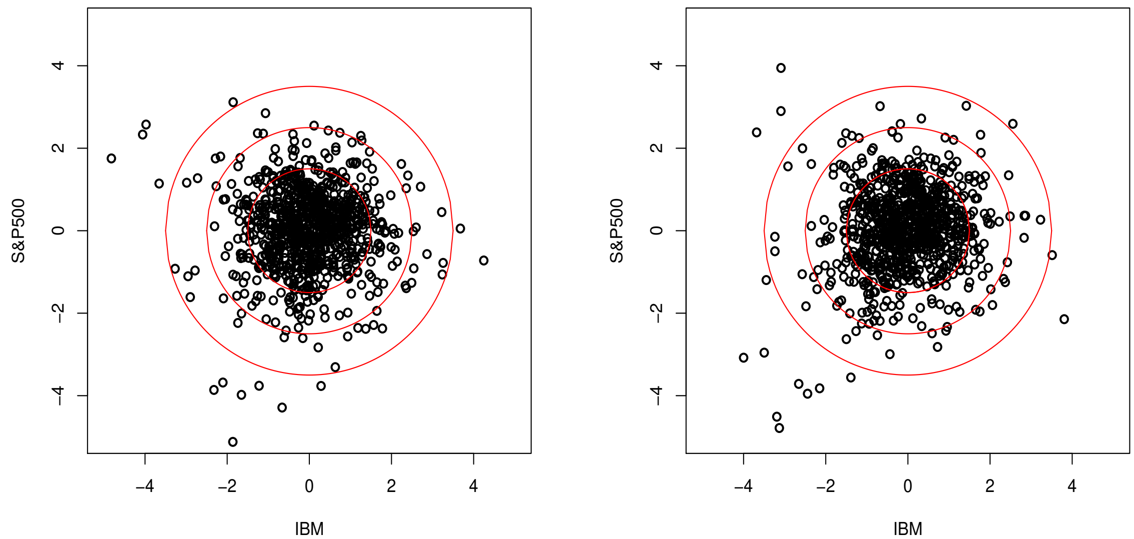

Figure 1: Scatter plot of the residuals when a CCC-GARCH(1,1) is fitted (left), the bivariate GARCH model in Example 9.2 of Tsay (2002) is fitted (right).

and a t-distribution for the bivariate error distribution in a bivariate GARCH(1,1) model. The normality hypothesis was rejected, but not the t-distribution.

We first fitted a CCC–GARCH(1,1) to the centred data $\{\boldsymbol{y}_t - \bar{\boldsymbol{y}},\ t = 1, \ldots, 888\}$ and applied the two CM–type tests described above to the residuals. The p–values obtained by generating $B = 1,000$ bootstrap samples are 0.031 and 0.035. Therefore the hypothesis of spherical symmetry is rejected. Then, we fitted the time–varying bivariate GARCH model considered in Bai and Chen (2008) (also in Example 9.2 of Tsay, 2002) and applied the two CM–type tests described above to the residuals. To get the p-values we applied the bootstrap algorithm described in Section 5, where it is understood that to get $\boldsymbol{\varepsilon}_t^*$ and $\widehat{\boldsymbol{\vartheta}}_T^*$ we considered the same model that was fitted to the original data. The two p-values coincide and they are equal to 0.989. Therefore the hypothesis of spherical symmetry is accepted, as expected from the results in Bai and Chen (2008). Looking at Figure 1 one may appreciate the difference between the scatter plot of the residuals between the two fitted models.

7 Conclusion

We propose Kolmogorov–Smirnov and Cramér–von Mises tests for the null hypothesis that the innovations of a given multivariate GARCH model follow a spherically symmetric distribution. The tests are based on the characteristic function, and specifically use the fact that the level curves of each spherically symmetric characteristic function are defined via equidistant from the origin arguments of this function. The consistency of both tests is shown and the limit null distribution of the test statistics is derived. As this distribution is complicated and not practical to apply we suggest a resampling procedure in order to actually carry out the tests, and prove its validity. Simulation results indicate that the tests behave reasonably well even under conditions not covered by our theory. Hence while our theoretical results were model–specific, these simulations suggest that the methods may be extended to alternative GARCH models, as well as to other models of conditional heteroscedasticity under which the hypothesis of sphericity is also relevant; e.g. the multivariate stochastic volatility models.

8 Proofs

We first collect some known results concerning the CCC-GARCH models and their estimation. First recall that the strict stationarity condition **A2** entails $E\left\|\boldsymbol{y}_t\right\|^s < \infty$ for some small $s > 0$ (Corollary 11.2 in Francq and Zakoïan, 2010). In the sequel C and ϱ denote generic constants or random variables, whose values are unimportant and may vary across the text, such that $C > 0$ and $0 < \varrho < 1$. By (11.55) in Francq and Zakoïan (2010), under **A1**-**A2**, we have almost surely

$$(8.1) \quad \sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\mathbf{D}_t(\boldsymbol{\vartheta}) - \widetilde{\mathbf{D}}_t(\boldsymbol{\vartheta})\right\| \le C\varrho^t, \quad \forall t.$$

By (11.71), (11.72) and (11.81) of the previous reference, under **A2** and **A6**, for any $r_0 \ge 0$ there exists a neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that, for all $k, \ell \in \{1, \ldots, s_0\}$

$$(8.2) \quad E \sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)} \left\|\frac{\partial \mathbf{D}_t(\boldsymbol{\vartheta})}{\partial \vartheta_k}\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\right\|^{r_0} < \infty, \quad E \sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)} \left\|\frac{\partial^2 \mathbf{D}_t(\boldsymbol{\vartheta})}{\partial \vartheta_\ell \partial \vartheta_k}\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\right\|^{r_0} < \infty$$

and

$$(8.3) \quad E \sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)} \left\|\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\mathbf{D}_t(\boldsymbol{\vartheta}_0)\right\|^{r_0} < \infty.$$

It is also known that, under **A1**-**A7**, the QMLE satisfies

$$\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}_0\right)=-\mathbf{J}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}+o_P(1), \tag{8.4}$$

where

$$\ell_t(\boldsymbol{\vartheta})=\boldsymbol{y}_t'\mathbf{C}_t^{-1}\boldsymbol{y}_t+\log|\mathbf{C}_t| \quad \text{and} \quad \mathbf{J}=E\left(\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right).$$

Proof of Lemma 4.2. For simplicity, from now on we take the square root $\mathbf{C}_t^{1/2}=\mathbf{D}_t\mathbf{R}^{1/2}$ and $\mathbf{C}_t^{-1/2}=\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}$. Note that these matrices are well defined because $\mathbf{R}$ is invertible and $\boldsymbol{b}>0$, for all $\boldsymbol{\vartheta}\in\Theta$ (by **A1** and **A5**). Letting $\mathbf{D}_{0t}=\mathbf{D}_t(\boldsymbol{\vartheta}_0)$, $\mathbf{R}_0=\mathbf{R}(\boldsymbol{\vartheta}_0)$,

$$\mathbf{D}_{0t}^{(j)}=\frac{\partial\mathbf{D}_t}{\partial\vartheta_j}(\boldsymbol{\vartheta}_0), \qquad \mathbf{R}_0^{(j)}=\frac{\partial\mathbf{R}}{\partial\vartheta_j}(\boldsymbol{\vartheta}_0),$$

we have that

$$\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\vartheta_j}=2\mathrm{tr}\left\{\left(\mathbf{I}_d-\mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{R}_0^{-1/2}\right)\mathbf{D}_{0t}^{-1}\mathbf{D}_{0t}^{(j)}\right\},$$

for $j=1,\ldots,s_1$, and

$$\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\vartheta_j}=\mathrm{tr}\left\{\left(\mathbf{I}_d-\mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{R}_0^{-1/2}\right)\mathbf{R}_0^{-1}\mathbf{R}_0^{(j)}\right\},$$

for $j=s_1+1,\ldots,s_0$. Using the elementary relation $\mathrm{tr}(\mathbf{A}'\mathbf{B})=(\mathrm{vec}\mathbf{B})'\mathrm{vec}\mathbf{A}$, we finally obtain (A.1) from (8.4), with

$$\boldsymbol{g}_{0t}=\mathrm{vec}\left(\mathbf{I}_d-\mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{R}_0^{-1/2}\right).$$

The existence of second-order moments for $\boldsymbol{\psi}_{0t}$ and $\boldsymbol{g}_{0t}$ come from **A7** and the first part of (8.2). □

Before proving Theorem 4.1, we will give some preliminary results. With this aim, we first introduce some further notation. Let

$$\varphi_T(\mathbf{u},\boldsymbol{\vartheta})=\frac{1}{T}\sum_{t=1}^{T}e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}, \qquad \widetilde{\varphi}_T(\mathbf{u},\boldsymbol{\vartheta})=\frac{1}{T}\sum_{t=1}^{T}e^{\mathrm{i}\mathbf{u}'\widetilde{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}$$

with $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})=\mathbf{C}_t^{-1/2}(\boldsymbol{\vartheta})\boldsymbol{y}_t$ and $\widetilde{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})=\widetilde{\mathbf{C}}_t^{-1/2}(\boldsymbol{\vartheta})\boldsymbol{y}_t$, so that $\widetilde{\varphi}_T(\mathbf{u})=\widetilde{\varphi}_T(\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T)$.

The next result shows that the initial values have no influence on the asymptotic behavior of the empirical CF.

Lemma 8.1 *Under Assumptions* **A1**, **A2** *and* **A5**, *for all* $\mathbf{u}$, *almost surely*

$$\sqrt{T}\sup_{\boldsymbol{\vartheta}\in\Theta}|\varphi_T(\mathbf{u},\boldsymbol{\vartheta})-\widetilde{\varphi}_T(\mathbf{u},\boldsymbol{\vartheta})|=o(1).$$

Proof of Lemma 8.1. From now on, for any complex number z, $\mathrm{Re}(z)$ ($\mathrm{Im}(z)$) stands for its real part (imaginary) part, that is, $z=\mathrm{Re}(z)+\mathrm{i}\,\mathrm{Im}(z)$ and $\overline{z}=\mathrm{Re}(z)-\mathrm{i}\,\mathrm{Im}(z)$.

Using (8.1), the elementary relation $|\cos x-\cos y|\leq|x-y|$, and the fact that

$$\sup_{\boldsymbol{\vartheta}\in\Theta}\max\left\{\left\|\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\right\|,\left\|\widetilde{\mathbf{D}}_t^{-1}(\boldsymbol{\vartheta})\right\|\right\}\leq d\,\frac{1}{\inf_{\boldsymbol{\vartheta}\in\Theta}\boldsymbol{b}_0}\leq C, \tag{8.5}$$

we obtain

$$\begin{aligned}
\sqrt{T}\left|\mathrm{Re}\left\{\varphi_T(\mathbf{u},\boldsymbol{\vartheta})\right\}-\mathrm{Re}\left\{\widetilde{\varphi}_T(\mathbf{u},\boldsymbol{\vartheta})\right\}\right| &\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\mathbf{u}'\widetilde{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})\right| \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|\mathbf{u}'\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}\left(\widetilde{\mathbf{D}}_t-\mathbf{D}_t\right)\widetilde{\mathbf{D}}_t^{-1}\boldsymbol{y}_t\right| \\
&\leq \frac{C}{\sqrt{T}}\left\|\mathbf{u}\right\|\sum_{t=1}^{T}\left\|\boldsymbol{y}_t\right\|\varrho^t.
\end{aligned}\tag{8.6}$$

Since $E\left\|\boldsymbol{y}_t\right\|^s<\infty$ for some $s>0$ and $\left(\sum_{t=1}^{\infty}\left\|\boldsymbol{y}_t\right\|\varrho^t\right)^\delta\leq\sum_{t=1}^{\infty}\left\|\boldsymbol{y}_t\right\|^\delta\varrho^{\delta t}$, for any $\delta\in(0,1)$, it follows that $\sum_{t=1}^{\infty}\left\|\boldsymbol{y}_t\right\|\varrho^t$ has a finite moment of order s, and thus it is almost surely finite. Thus, the right-hand side of the inequality (8.6) tends to 0. The same convergence holds for the imaginary part, and the conclusion follows. □

Lemma 8.2 *Under the assumptions of Theorem 4.1, and for any sequence* $\boldsymbol{\vartheta}_T$ *tending to* $\boldsymbol{\vartheta}_0$ *as* $T\to\infty$, *for some real vector* $\dot{\boldsymbol{\varphi}}(\mathbf{u})$ *we have almost surely*

$$\frac{\partial}{\partial\boldsymbol{\vartheta}}\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_T)\to\dot{\boldsymbol{\varphi}}(\mathbf{u}).$$

Proof of Lemma 8.2. Elementary multivariate differentiation rules yield

$$\begin{aligned}
\frac{\partial}{\partial\vartheta_j}\mathrm{Re}\,e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} &= \sin\left\{\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\}\mathrm{tr}\left\{\mathbf{D}_t^{-1}\boldsymbol{y}_t\mathbf{u}'\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}\frac{\partial\mathbf{D}_t}{\partial\vartheta_j}\right\} \\
&= \sin\left\{\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\}\mathbf{u}'\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}\frac{\partial\mathbf{D}_t}{\partial\vartheta_j}\mathbf{D}_t^{-1}\boldsymbol{y}_t,
\end{aligned}\tag{8.7}$$

for $j=1,\ldots,s_1$ and

$$\frac{\partial}{\partial\vartheta_j}\mathrm{Re}\,e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} = \sin\left\{\mathbf{u}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\}\mathbf{u}'\mathbf{R}^{-1/2}\frac{\partial\mathbf{R}^{1/2}}{\partial\vartheta_j}\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}\boldsymbol{y}_t, \tag{8.8}$$

for $j = s_1+1,\dots,s_0$. Similar expressions hold for the imaginary part. Using the notation introduced in the proof of Lemma 4.2 and the extra notation

$$\mathbf{R}_0^{\frac{1}{2}(j)} = \frac{\partial \mathbf{R}^{1/2}}{\partial \vartheta_j}(\boldsymbol{\vartheta}_0),$$

we then obtain

$$\frac{\partial}{\partial \boldsymbol{\vartheta}}\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_0) \;\; = \;\; \frac{1}{T}\sum_{t=1}^{T} -\mathrm{i}e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t}\boldsymbol{S}_t,$$

with

$$\boldsymbol{S}_t' = \mathbf{u}'\boldsymbol{M}_{t-1}(\mathbf{I}_{s_0}\otimes\boldsymbol{\varepsilon}_t), \quad \boldsymbol{M}_{t-1} = \left(\mathbf{R}_0^{-1/2}\mathbf{D}_{0t}^{-1}\mathbf{D}_{0t}^{(1)}\mathbf{R}_0^{1/2} \;\cdots\; \mathbf{R}_0^{-1/2}\mathbf{R}_0^{\frac{1}{2}(s_0)}\right).$$

By the ergodic theorem,

$$\frac{\partial}{\partial \boldsymbol{\vartheta}}\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_0) \;\to\; \dot{\boldsymbol{\varphi}}(\mathbf{u}) := -E\left\{\mathrm{i}e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_1}(\mathbf{I}_{s_0}\otimes\boldsymbol{\varepsilon}_1')\right\}\boldsymbol{M}'\mathbf{u}, \quad \boldsymbol{M} = E\boldsymbol{M}_1.$$

Note that under the null hypothesis $\mathcal{H}_0$,

$$E\{\cos(\mathbf{u}'\boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1'\} = \mathbf{0}_d,$$

and thus the vector $\dot{\varphi}(\mathbf{u})$ is real, specifically,

$$\dot{\boldsymbol{\varphi}}(\mathbf{u}) = E\left\{\sin(\mathbf{u}'\boldsymbol{\varepsilon}_1)(\mathbf{I}_{s_0}\otimes\boldsymbol{\varepsilon}_1')\right\}\boldsymbol{M}'\mathbf{u}.$$

Doing a Taylor expansion

$$\frac{\partial}{\partial \vartheta_j}\mathrm{Re}\,\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_T) \;\; = \;\; \frac{\partial}{\partial \vartheta_j}\mathrm{Re}\,\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_0) + (\boldsymbol{\vartheta}_T-\boldsymbol{\vartheta}_0)'\frac{\partial^2}{\partial\vartheta_j\partial\boldsymbol{\vartheta}}\mathrm{Re}\,\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_T^\circ)$$

where $\boldsymbol{\vartheta}_T^\circ$ is between $\boldsymbol{\vartheta}_T$ and $\boldsymbol{\vartheta}_0$. A similar expression holds for the imaginary part. Therefore the result will follow from the ergodic theorem, by showing that for all $\ell, j \in \{1,\dots,s_0\}$ there exists a neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$(8.9) \quad E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\max\left\{\left|\frac{\partial^2}{\partial\vartheta_\ell\partial\vartheta_j}\cos\{\mathbf{u}'\boldsymbol{\varepsilon}_1(\boldsymbol{\vartheta})\}\right|, \left|\frac{\partial^2}{\partial\vartheta_\ell\partial\vartheta_j}\sin\{\mathbf{u}'\boldsymbol{\varepsilon}_1(\boldsymbol{\vartheta})\}\right|\right\} < \infty.$$

Differentiating (8.7) and (8.8), one can see that

$$\left|\frac{\partial^2}{\partial\vartheta_\ell\partial\vartheta_j}\cos\{\mathbf{u}'\boldsymbol{\varepsilon}_1(\boldsymbol{\vartheta})\}\right| + \left|\frac{\partial^2}{\partial\vartheta_\ell\partial\vartheta_j}\sin\{\mathbf{u}'\boldsymbol{\varepsilon}_1(\boldsymbol{\vartheta})\}\right|$$

is bounded by a sum of products of the terms

$$\|\mathbf{u}\|, \left\|\mathbf{R}^{-1/2}\right\|, \left\|\frac{\partial \mathbf{R}^{1/2}}{\partial \vartheta_k}\right\|, \left\|\frac{\partial^2 \mathbf{R}^{1/2}}{\partial \vartheta_\ell \partial \vartheta_j}\right\|, \left\|\frac{\partial \mathbf{D}_t}{\partial \vartheta_k}\mathbf{D}_t^{-1}\right\|, \left\|\frac{\partial^2 \mathbf{D}_t}{\partial \vartheta_\ell \partial \vartheta_j}\mathbf{D}_t^{-1}\right\| \tag{8.10}$$

for $k \in \{\ell, j\}$, and of

$$\left\|\mathbf{D}_t^{-1}\boldsymbol{y}_t\right\| = \left\|\mathbf{D}_t^{-1}\mathbf{D}_{0t}\mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_t\right\|. \tag{8.11}$$

By the compactness of Θ and by **A5**, the second, third and fourth terms of (8.10) are bounded uniformly in $\boldsymbol{\vartheta}$. By (8.2), the suprema over some neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of the last two terms of (8.10) admit moments of any fixed order r_0. By (8.3) and **A7**, for some neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$, the supremum of (8.11) over $\boldsymbol{\vartheta} \in \mathcal{V}(\boldsymbol{\vartheta}_0)$ admits a moment of order 2. Therefore, (8.9) follows from the Hölder inequality, which completes the proof. □

Lemma 8.3 *Under the assumptions of Theorem 4.1,* $|\mathrm{CM}_T - \mathrm{CM}_T^{\circ}| = o(1)$ *almost surely, where*

$$\mathrm{CM}_T^{\circ} = T\int_0^{\infty}\left(\sum_{j,m=1}^{K}|D_T^{\circ}(\rho\mathbf{u}_j, \rho\mathbf{u}_m)|^2\right)\omega(\rho)d\rho,$$

with $D_T^{\circ}(\mathbf{u}, \mathbf{v}) = \varphi_T(\mathbf{u}, \widehat{\boldsymbol{\vartheta}}_T) - \varphi_T(\mathbf{v}, \widehat{\boldsymbol{\vartheta}}_T)$.

Proof of Lemma 8.3. In view of (8.6) and arguments of the proof of Lemma 8.1,

$$|D_T^{\circ}(\mathbf{u}, \mathbf{v}) - D_T(\mathbf{u}, \mathbf{v})| \leq \frac{C}{T}\left(\|\mathbf{u}\| + \|\mathbf{v}\|\right).$$

We then obtain

$$\begin{aligned}
&|\mathrm{CM}_T^{\circ} - \mathrm{CM}_T| \\
= &\left|T\int_0^{\infty}\sum_{j,m=1}^{K}\left[\{D_T^{\circ}(\rho\mathbf{u}_j, \rho\mathbf{v}_m) - D_T(\rho\mathbf{u}_j, \rho\mathbf{v}_m)\}\overline{D}_T^{\circ}(\rho\mathbf{u}_j, \rho\mathbf{v}_m)\right.\right. \\
&\left.\left.+D_T(\rho\mathbf{u}_j, \rho\mathbf{v}_m)\left\{\overline{D}_T^{\circ}(\rho\mathbf{u}_j, \rho\mathbf{v}_m) - \overline{D}_T(\rho\mathbf{u}_j, \rho\mathbf{v}_m)\right\}\right]\omega(\rho)d\rho\right| \\
\leq\ & C\int_0^{\infty}\rho\sum_{j,m=1}^{K}\left(\|\mathbf{u}_j\| + \|\mathbf{v}_m\|\right)\left(\overline{D}_T^{\circ}(\rho\mathbf{u}_j, \rho\mathbf{v}_m) + D_T(\rho\mathbf{u}_j, \rho\mathbf{v}_m)\right)\omega(\rho)d\rho.
\end{aligned}$$

Using Lebesgue's dominated convergence theorem, with (2.1) and the fact that $\overline{D}_T^{\circ}(\cdot, \cdot)$ and $D_T(\cdot, \cdot)$ tend to zero and are bounded uniformly in T, the right-hand side of the inequality tends to zero a.s.. □

Lemma 8.4 *Under the assumptions of Theorem 4.1,* $|\mathrm{CM}_T^{\circ} - \mathrm{CM}_T^{\circ\circ}| = o_P(1)$*, where*

$$\mathrm{CM}_T^{\circ\circ} = T \int_0^\infty \left(\sum_{j,m=1}^{K} |D_T^{\circ\circ}(\rho \mathbf{u}_j, \rho \mathbf{u}_m)|^2 \right) \omega(\rho) d\rho,$$

with $D_T^{\circ\circ}(\mathbf{u}, \mathbf{v}) = T^{-1} \sum_{t=1}^{T} \{g_t(\mathbf{u}) - g_t(\mathbf{v})\}$.

Proof of Lemma 8.4. We have

$$\sqrt{T} \{D_T^{\circ}(\mathbf{u}, \mathbf{v}) - D_T^{\circ\circ}(\mathbf{u}, \mathbf{v})\} = a_T(\mathbf{u}) - a_T(\mathbf{v}),$$

where, for some $\boldsymbol{\vartheta}_T$ between $\widehat{\boldsymbol{\vartheta}}_T$ and $\boldsymbol{\vartheta}_0$,

$$\begin{aligned}
\mathrm{Re}\{a_T(\mathbf{u})\} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \cos(\mathbf{u}'\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\vartheta}}_T)) - \cos(\mathbf{u}'\boldsymbol{\varepsilon}_t) - \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}'\dot{\boldsymbol{\varphi}}(\mathbf{u}) \right\} \\
&= \sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)' \frac{\partial}{\partial \boldsymbol{\vartheta}} \mathrm{Re}\{\varphi_T(\mathbf{u}, \boldsymbol{\vartheta}_T)\} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}'\dot{\boldsymbol{\varphi}}(\mathbf{u}) \\
&= \left\{ \sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)' - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}' \right\} \frac{\partial}{\partial \boldsymbol{\vartheta}} \mathrm{Re}\{\varphi_T(\mathbf{u}, \boldsymbol{\vartheta}_T)\} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}' \left\{ \frac{\partial}{\partial \boldsymbol{\vartheta}} \mathrm{Re}\{\varphi_T(\mathbf{u}, \boldsymbol{\vartheta}_T)\} - \dot{\boldsymbol{\varphi}}(\mathbf{u}) \right\}.
\end{aligned}$$

By the CLT for stationary martingale differences $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}' = O_P(1)$. Using also Lemma 8.2 and (A.1) it follows that $\mathrm{Re}\{a_T(\mathbf{u})\} = o_P(1)$. Proceeding analogously, we get $\mathrm{Im}\{a_T(\mathbf{u})\} = o_P(1)$. More precisely, $|a_T(\mathbf{u})| \leq C_T \|\mathbf{u}\|$, with $C_T = O_P(1)$. Therefore the conclusion follows from (2.1), as in the proof of Lemma 8.3. □

Lemma 8.5 *Under the assumptions of Theorem 4.1, for any* $\overline{\rho} > 0$ *and any* $\mathbf{u} \in \mathbb{R}^d$*, the sequences*

$$\left\{ T^{-1/2} \sum_{t=1}^{T} Re\, g_t(\rho \mathbf{u}),\ \rho \in [0, \overline{\rho}] \right\}_{T \geq 1} \quad \text{and} \quad \left\{ T^{-1/2} \sum_{t=1}^{T} Im\, g_t(\rho \mathbf{u}),\ \rho \in [0, \overline{\rho}] \right\}_{T \geq 1}$$

are tight in $\mathcal{C}([0, \overline{\rho}])$*, the Banach space of the real-valued continuous functions on* $[0, a]$*, endowed with the supremum norm.*

Proof of Lemma 8.5. We have

$$Y_T(\rho) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{Re}\, g_t(\rho \mathbf{u}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \cos(\rho \mathbf{u}'\boldsymbol{\varepsilon}_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}'\dot{\boldsymbol{\varphi}}(\rho \mathbf{u}).$$

From **A7**, it follows that the first term in the right side of the above equality is tight in every finite interval (see Csörgő, 1981). Because by the CLT for stationary martingale differences $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}' = O_P(1)$ and by **A7** $\dot{\boldsymbol{\varphi}}(\rho\mathbf{u})$ is a continuous function of ρ, we get that the second term in the right side of the above equality is also tight in every finite interval. Thus, $Y_T(\rho)$ is tight in every finite interval, specifically it is tight in $[0,\overline{\rho}]$, for any $\overline{\rho} > 0$. The second sequence is treated more easily since it does not depend on the real vector $\dot{\boldsymbol{\varphi}}(\rho\mathbf{u})$. □

Proof of Theorem 4.1. We now employ the notation $a \stackrel{c}{=} b$ when $a = b + c$. Using Lemma 8.1, a Taylor expansion with Lemma 8.2, and (A.1), for any fixed $\mathbf{u}$ we have

$$\begin{aligned}\sqrt{T}\left\{\widetilde{\varphi}_T(\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T) - \varphi(\mathbf{u})\right\} &\stackrel{o_P(1)}{=} \sqrt{T}\left\{\varphi_T(\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T) - \varphi(\mathbf{u})\right\} \\ &\stackrel{o_P(1)}{=} \sqrt{T}\left\{\varphi_T(\mathbf{u},\boldsymbol{\vartheta}_0) - \varphi(\mathbf{u})\right\} + \sqrt{T}\left(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right)'\dot{\boldsymbol{\varphi}}(\mathbf{u}) \\ &\stackrel{o_P(1)}{=} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t} - \varphi(\mathbf{u}) + \boldsymbol{g}_{0t}'\boldsymbol{\psi}_{0,t-1}'\dot{\boldsymbol{\varphi}}(\mathbf{u})\right\}.\end{aligned}$$

Now, note if $\mathbf{u} = \rho\mathbf{u}_j$ and $\mathbf{v} = \rho\mathbf{u}_m$ with $(\mathbf{u}_j,\mathbf{u}_m) \in \mathbb{S}_\circ^2$ then, under $\mathcal{H}_0$, $\varphi(\mathbf{u}) = \varphi(\mathbf{v})$. We thus have

$$\sqrt{T}D_T(\mathbf{u},\mathbf{v}) \stackrel{o_P(1)}{=} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\mathbf{u}) - g_t(\mathbf{v}).$$

The complex-valued random field of the right-hand side, indexed by $(\mathbf{u},\mathbf{v})$, has the same covariance kernel as (4.3). The central limit theorem (CLT) for squared integrable martingale differences then entails that

$$\begin{aligned}&\left\{\sqrt{T}D_T(\rho_l\mathbf{u}_j,\rho_l\mathbf{v}_m); l = 1,\ldots,L; j,m = 1,\ldots,K\right\} \\ \xrightarrow{\mathcal{D}}\;& \left\{\mathcal{W}(\rho_l\mathbf{u}_j,\rho_l\mathbf{v}_m); l = 1,\ldots,L; j,m = 1,\ldots,K\right\}.\end{aligned}$$

By the continuous mapping theorem, (4.1) follows.

To show (5.3) it is necessary to rely on a functional CLT. Lemmas 8.3 and 8.4 show that CM_T and $\mathrm{CM}_T^{\circ\circ}$ have the same asymptotic distribution. Note that $\mathrm{CM}_T^{\circ\circ}$ is a continuous functional of the process

$$\rho \mapsto d_T(\rho) := \sum_{j,m=1}^{K}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{g_t(\rho\mathbf{u}_j) - g_t(\rho\mathbf{u}_m)\right\}\right|^2.$$

Alternatively, $\mathrm{CM}_T^{\circ\circ}$ can be considered as a continuous functional of the multivariate process $\rho \mapsto T^{-1/2}\sum_{t=1}^{T}\mathbf{G}_t(\rho)$, where $\mathbf{G}'_t(\rho) = \{g_t(\rho\mathbf{u}_1),\dots,g_t(\rho\mathbf{u}_M)\}$. Let $\overline{\rho} > 0$. We have already shown the convergence of the finite-dimensional distributions of $\mathcal{D}_T := \{d_T(\rho), \rho \in [0,\overline{\rho}]\}$ to those of $\mathcal{D} = \left\{\sum_{j,m=1}^{K}|\mathcal{W}(\rho\mathbf{u}_j,\rho\mathbf{u}_m)|^2, \rho \in [0,\overline{\rho}]\right\}$ as $T \to \infty$. Because a vectorial sequence of random elements is tight when the sequences of each of its components are tight, Lemma 8.5 shows that the sequence $\left\{T^{-1/2}\sum_{t=1}^{T}\mathbf{G}_t(\rho), \rho \in [0,\overline{\rho}]\right\}$ is tight. Since a continuous transformation of a tight sequence is tight, we have shown the tightness of the sequence $(\mathcal{D}_T)_{T\geq 1}$, and thus the weak convergence of $\mathcal{D}_T$ to $\mathcal{D}$ in the Banach space $\mathcal{C}([0,\overline{\rho}])$ of the complex-valued continuous functions on $[0,\overline{\rho}]$, endowed with the supremum norm.

The continuous mapping theorem then entails that

$$T\int_0^{\overline{\rho}}\left(\sum_{j,m=1}^{K}|D_T(\rho\mathbf{u}_j,\rho\mathbf{u}_m)|^2\right)\omega(\rho)d\rho \xrightarrow{\mathcal{D}} \int_0^{\overline{\rho}}\left(\sum_{j,m=1}^{K}|\mathcal{W}(\rho\mathbf{u}_j,\rho\mathbf{u}_m)|^2\right)\omega(\rho)d\rho.$$

Since $\overline{\rho}$ can be chosen arbitrarily large, we conclude as in the proof of Corollary 3.2 of Henze et al. (2014). □

Proof of Theorem 4.3. We have already shown that, almost surely, $\widehat{\varphi}_T(\mathbf{u}) \to \varphi(\mathbf{u})$, therefore

$$\lim D_T(\rho_l\mathbf{u}_j,\rho_l\mathbf{v}_m) = \varphi(\rho_l\mathbf{u}_j) - \varphi(\rho_l\mathbf{v}_m)$$

almost surely, and (4.4) follows. To show (4.5) we note that

$$\lim\int_0^{\infty}|D_T(\rho\mathbf{u}_0,\rho\mathbf{v}_0)|^2\,\omega(\rho)d\rho = \int_0^{\infty}\lim|D_T(\rho\mathbf{u}_0,\rho\mathbf{v}_0)|^2\,\omega(\rho)d\rho$$

by Legesgue's dominated convergence theorem, using $|D_T(\rho\mathbf{u}_0,\rho_l\mathbf{v}_0)| \leq 2$ and (2.1). □

To establish the asymptotic validity of the resampling scheme, we first show that the conditional distribution of $\boldsymbol{\varepsilon}^*_{t,T}$ tends to a well-defined distribution which coincides with that of $\boldsymbol{\varepsilon}_t$ when the null hypothesis is true.

Lemma 8.6 *Suppose that the assumptions of Theorem 5.1 are satisfied, except that we do not have to assume that $\widehat{\boldsymbol{\vartheta}}_T$ is the QMLE. Assume that $\widehat{\boldsymbol{\vartheta}}_T$ is any estimator satisfying (A.1). For almost all sequence $\boldsymbol{y} = \{\boldsymbol{y}_t\}$ satisfying (1.1), (3.1) and (3.2), the distribution of $\boldsymbol{\varepsilon}^*_{t,T}$ conditionally on $\boldsymbol{y}$ tends to the unconditional distribution of $\boldsymbol{\varepsilon}_{01}$.*

Proof of Lemma 8.6. To show that the conditional (on $\{\boldsymbol{y}_t\}$) distribution of $\boldsymbol{\varepsilon}^*_{t,T}$ converges to the (non conditional) distribution of $\boldsymbol{\varepsilon}_{01}$, it suffices to show that the conditional distribution of $\|\boldsymbol{\varepsilon}^*_{t,T}\|$ converges to the distribution of $\|\boldsymbol{\varepsilon}_t\|$.

Recall that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \mathbf{C}_t^{-1/2}(\boldsymbol{\vartheta})\boldsymbol{y}_t$ and $\widetilde{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) = \widetilde{\mathbf{C}}_t^{-1/2}(\boldsymbol{\vartheta})\boldsymbol{y}_t$. We have

$$\mathbf{C}_t^{-1/2}(\boldsymbol{\vartheta}) - \widetilde{\mathbf{C}}_t^{-1/2}(\boldsymbol{\vartheta}) = \mathbf{R}^{-1/2}\mathbf{D}_t^{-1}\left(\widetilde{\mathbf{D}}_t - \mathbf{D}_t\right)\widetilde{\mathbf{D}}_t^{-1}(\boldsymbol{\vartheta}).$$

Moreover, by (8.3), there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \|\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\mathbf{C}_t^{1/2}(\boldsymbol{\vartheta}_0)\|^2 < \infty.$$

Using (8.1) and (8.5), it follows that for T large enough

$$\left\|\widetilde{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\vartheta}}_T)\right\| \le C\varrho^t u_t\|\boldsymbol{\varepsilon}_t\|,$$

where $\{u_t\}$ is a stationary sequence of positive random variables, measurable with respect to the sigma-field $\mathcal{F}_{t-1}$ generated by $\{\boldsymbol{\varepsilon}_u, u < t\}$, and such that $E(u_t^2) < \infty$. We also have

$$\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\vartheta}}_T) = \boldsymbol{\varepsilon}_t + \frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_T)}{\partial \boldsymbol{\vartheta}'}\left(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right),$$

with $\boldsymbol{\vartheta}_T$ between $\widehat{\boldsymbol{\vartheta}}_T$ and $\boldsymbol{\vartheta}_0$. Note that, for $\ell = 1, \dots, s_1$ and $j = 1, \dots, d$, we have

$$\frac{\partial \mathbf{e}_j\mathbf{C}_t^{-1/2}(\boldsymbol{\vartheta})\boldsymbol{y}_t}{\partial \vartheta_\ell} = -\mathrm{tr}\left\{\mathbf{D}_t^{-1}\boldsymbol{y}_t\mathbf{e}_j\mathbf{R}^{-1/2}\mathbf{D}_t^{-1}(\boldsymbol{\vartheta})\frac{\partial \mathbf{D}_t(\boldsymbol{\vartheta})}{\partial \vartheta_\ell}\right\}.$$

Similar expressions hold for $\ell = s_1 + 1, \dots, s_0$. By the previous arguments, it follows that for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ we have

$$\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'}\right\| \le u_t\|\boldsymbol{\varepsilon}_t\|.$$

We thus have

$$\|\widetilde{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t\| \le C\left(\varrho^t + \left\|\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right\|\right)u_t\|\boldsymbol{\varepsilon}_t\|, \tag{8.12}$$

for T large enough. We also have that

$$\widetilde{\boldsymbol{\varepsilon}}_{0t} - \widetilde{\boldsymbol{\varepsilon}}_t = (S_T^{-1/2} - \mathbf{I}_d)\widetilde{\boldsymbol{\varepsilon}}_t - S_T^{-1/2}\widetilde{\boldsymbol{\varepsilon}}_..$$

From (8.12) it follows that

$$\widetilde{\boldsymbol{\varepsilon}}_. \to 0, \quad S_T \to \mathbf{I}_d,$$

for almost all sequences $\{\boldsymbol{y}_t\}$. Therefore,

$$|\|\widetilde{\boldsymbol{\varepsilon}}_{0t}\| - \|\boldsymbol{\varepsilon}_t\|| \leq a_{1T} + a_{2T}\|\boldsymbol{\varepsilon}_t\| + C\left(\varrho^t + a_{3T}\right)u_t\|\boldsymbol{\varepsilon}_t\|,$$

for T large enough, a_{jT} being a positive random variable tending almost surely to 0, $j = 1, 2, 3$. Denote by 1_A the indicator function of an event A. For all $x \in \mathbb{R}$, all $\epsilon > 0$ and all $M > 0$, we then have

$$\begin{aligned}
&\left|1_{\{\|\widetilde{\boldsymbol{\varepsilon}}_{0t}\| \leq x\}} - 1_{\{\|\boldsymbol{\varepsilon}_t\| \leq x\}}\right| \\
&\quad \leq 1_{\{x - a_{1T} - a_{2T}\|\boldsymbol{\varepsilon}_t\| - C(\varrho^t + a_{3T})u_t\|\boldsymbol{\varepsilon}_t\| \leq \|\boldsymbol{\varepsilon}_t\| \leq x + a_{1T} + a_{2T}\|\boldsymbol{\varepsilon}_t\| + C(\varrho^t + a_{3T})u_t\|\boldsymbol{\varepsilon}_t\|\}} \\
&\quad \leq 1_{A_{t,\epsilon,M}} + 1_{B_{1,\epsilon}} + 1_{B_{2,\epsilon}} + 1_{B_{3,\epsilon}} + 1_{C_{t,M}},
\end{aligned}$$

with the events being

$$A_{t,\epsilon,M} = \left\{x - \epsilon(1+M) - C\left(\varrho^t + \epsilon\right)u_t M \leq \|\boldsymbol{\varepsilon}_t\| \leq x + \epsilon(1+M) + C\left(\varrho^t + \epsilon\right)u_t M\right\},$$

$B_{j,\epsilon} = \{a_{jT} > \epsilon\}$, $j = 1, 2, 3$, and $C_{t,M} = \{\|\boldsymbol{\varepsilon}_t\| > M\}$. The almost sure convergence to 0 of a_{jT} entails that $E1_{B_{j,\epsilon}} \to 0$, $j = 1, 2, 3$. We also have $E1_{C_{t,M}} \to 0$ as $M \to \infty$. Conditioning on $\mathcal{F}_{t-1}$, we have

$$EA_{t,\epsilon,M} = E\int_{x-(1+M)\epsilon - C(\varrho^t+\epsilon)u_t M}^{x+(1+M)\epsilon + C\left(\varrho^t+\epsilon\right)u_t M} f(y)dy \leq 2\max_{y\in\mathbb{R}} f(y)\left\{(1+M)\epsilon + C\left(\varrho^t + \epsilon\right)MEu_t\right\}.$$

For all $\kappa > 0$, we thus have a small $\varepsilon > 0$ and a large $M > 0$ such that

$$\lim_{t\to\infty} E\left\{1_{A_{t,\epsilon,M}} + 1_{B_{1,\epsilon}} + 1_{B_{2,\epsilon}} + 1_{B_{3,\epsilon}} + 1_{C_{t,M}}\right\} \leq \kappa.$$

It follows that, for almost all sequences $\{\boldsymbol{y}_t\}$, we have

$$\lim \frac{1}{T}\sum_{t=1}^{T} 1_{\{\|\widetilde{\boldsymbol{\varepsilon}}_{0t}\| \leq x\}} = \lim \frac{1}{T}\sum_{t=1}^{T} 1_{\{\|\boldsymbol{\varepsilon}_t\| \leq x\}} = P\left(\|\boldsymbol{\varepsilon}_t\| \leq x\right), \quad \forall x \in \mathbb{R},$$

which allows to conclude. □

The following result shows that, in addition to the convergence in distribution of Lemma 8.6, we have convergence of the conditional moments of $\boldsymbol{\varepsilon}^*_{t,T}$ to the (unconditional) moments of $\boldsymbol{\varepsilon}_1$. Recall that by construction, $E\left(\boldsymbol{\varepsilon}^*_{1,T} \mid \boldsymbol{y}\right) = E(\boldsymbol{\varepsilon}_1) = \mathbf{0}$ and $E\left(\boldsymbol{\varepsilon}^*_{1,T}\boldsymbol{\varepsilon}^{*\prime}_{1,T} \mid \boldsymbol{y}\right) = E(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1') = \mathbf{I}_d$.

Lemma 8.7 *Suppose that the assumptions of Lemma 8.6 are satisfied. Then for almost all sequences* $\boldsymbol{y} = \{\boldsymbol{y}_t\}$ *satisfying (1.1), (3.1), (3.2), we have*

$$(8.13)\quad E\left(\left\|\boldsymbol{\varepsilon}_{1,T}^*\right\|^4 \mid \boldsymbol{y}\right) \to E\left\|\boldsymbol{\varepsilon}_{01}\right\|^4,$$

and for any continuous function $\mathbf{h}$ *such that* $\|\mathbf{h}(\boldsymbol{x})\| \leq a\|\boldsymbol{x}\|^4 + b$, *with* $a, b > 0$, *it follows*

$$(8.14)\quad E\left\{\mathbf{h}\left(\boldsymbol{\varepsilon}_{1,T}^*\right) \mid \boldsymbol{y}\right\} \to E\mathbf{h}\left(\boldsymbol{\varepsilon}_{01}\right).$$

Proof of Lemma 8.7. (8.13) can be shown by employing the arguments used in the proof of Lemma 8.6. To show (8.14), first note that Lemma 8.6 and the continuity of $\mathbf{h}$ entail that the distribution of $\mathbf{h}(\boldsymbol{\varepsilon}_{t,T}^*)$ conditional on $\boldsymbol{y}$ tends to the law of $\mathbf{h}(\boldsymbol{\varepsilon}_{0t})$. By Theorem 5.4 in Billingsley (1968), (8.13) entails that, conditional on $\boldsymbol{y}$, the sequence $\|\boldsymbol{\varepsilon}_{t,T}^*\|^4$ is uniformly integrable in T. The same theorem also shows that (8.14) is obtained by showing that the sequence $\mathbf{h}(\boldsymbol{\varepsilon}_{t,T}^*)$ is uniformly integrable in T, which is obvious from $E\left\|\mathbf{h}(\boldsymbol{\varepsilon}_{t,T}^*)\right\| \leq aE\left\|\boldsymbol{\varepsilon}_{t,T}^*\right\|^4 + b$. □

We now show that one can obtain results similar to those of Lemmas 8.6 and 8.7 for more complex functions of the $\boldsymbol{\varepsilon}_t^*$'s, such as $\widehat{\mathbf{C}}_t^* := \widetilde{\mathbf{C}}_t(\boldsymbol{y}_{t-1}^*, \ldots, \boldsymbol{y}_1^*; \widehat{\boldsymbol{\vartheta}}_T)$. With this aim, we will assume that the initial values are taken $\boldsymbol{y}_{1-p}^* = \ldots = \boldsymbol{y}_0^* = \widetilde{\boldsymbol{\sigma}}_{1-q}^* = \ldots = \widetilde{\boldsymbol{\sigma}}_0^* = \mathbf{0}$.

Lemma 8.8 *Under the assumptions of Lemma 8.6 and for almost all sequence* $\boldsymbol{y} = \{\boldsymbol{y}_t\}$ *satisfying the assumptions of this lemma, the distribution of* $\widehat{\mathbf{C}}_t^*$ *conditionally on* $\boldsymbol{y}$ *tends to the unconditional distribution of*

$$(8.15)\quad \widetilde{\mathbf{C}}_t(\widetilde{\boldsymbol{y}}_{t-1}, \ldots, \widetilde{\boldsymbol{y}}_1; \boldsymbol{\vartheta}_0), \qquad \text{as} \quad T \to \infty,$$

where the sequence $(\widetilde{\boldsymbol{y}}_t)_{t\geq 1}$ *satisfies the same recursive equation as* $(\boldsymbol{y}_t)_{t\geq 1}$, *with the initial values* $\boldsymbol{y}_{1-p} = \ldots = \boldsymbol{y}_0 = \widetilde{\boldsymbol{\sigma}}_{1-q} = \ldots = \widetilde{\boldsymbol{\sigma}}_0 = \mathbf{0}$.

Proof of Lemma 8.8 For simplicity, assume for the proof that $p = q = 1$ in (3.2). Note that $\boldsymbol{y}_t^{(2)} = \boldsymbol{\Upsilon}_t\boldsymbol{\sigma}_t$, where $\boldsymbol{\Upsilon}_t = \mathrm{diag}(\eta_{1t}^2, \ldots, \eta_{dt}^2)$ and $(\eta_{1t}, \ldots, \eta_{dt})' = \mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_t$. Similarly $\boldsymbol{y}_t^{*(2)} := \boldsymbol{y}_t^* \odot \boldsymbol{y}_t^* = \boldsymbol{\Upsilon}_t^*\boldsymbol{\sigma}_t^*$, where $\boldsymbol{\Upsilon}_t^* = \mathrm{diag}(\eta_{1t}^{*2}, \ldots, \eta_{dt}^{*2})$ with $(\eta_{1t}^*, \ldots, \eta_{dt}^*)' = \widehat{\mathbf{R}}^{1/2}\boldsymbol{\varepsilon}_t^*$ and $\boldsymbol{\sigma}_t^* = \widehat{\boldsymbol{b}} + \widehat{\mathbf{B}}\boldsymbol{y}_{t-1}^{*(2)} + \widehat{\boldsymbol{\Gamma}}\boldsymbol{\sigma}_{t-1}^*$, for $t = 1, 2, \ldots$. Because the initial values are

$\boldsymbol{y}_0^* = \widetilde{\boldsymbol{\sigma}}_0^* = \mathbf{0}$, we have

$$\boldsymbol{\sigma}_t^* = \widehat{\boldsymbol{b}} + \sum_{j=1}^{t-1} \prod_{k=1}^{j} \left(\widehat{\mathbf{B}} \boldsymbol{\Upsilon}_{t-k}^* + \widehat{\boldsymbol{\Gamma}} \right) \widehat{\boldsymbol{b}}.$$

By the consistency of $\widehat{\boldsymbol{\vartheta}}_T$, Lemma 8.6 and Slutsky's lemma, the conditional distribution of $\boldsymbol{\Upsilon}_t^*$, given $\boldsymbol{y}$, tends to the law of $\boldsymbol{\Upsilon}_t$. By the same arguments, the conditional distribution of $\boldsymbol{\sigma}_t^*$, given $\boldsymbol{y}$, tends to the law of

$$\widetilde{\boldsymbol{\sigma}}_t = \boldsymbol{b}_0 + \sum_{j=1}^{t-1} \prod_{k=1}^{j} \left(\mathbf{B}_0 \boldsymbol{\Upsilon}_{t-k} + \boldsymbol{\Gamma}_0 \right) \boldsymbol{b}_0$$

and the conclusion follows. □

Let $\varphi_T^*(\mathbf{u}) = E\left(e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_{1,T}^*} \mid \boldsymbol{y}\right) = E\left\{\cos\left(\mathbf{u}'\boldsymbol{\varepsilon}_{1,T}^*\right) \mid \boldsymbol{y}\right\}$ and let $\widetilde{\varphi}_{0T}(\mathbf{u}, \widehat{\boldsymbol{\vartheta}}_T)$ and $\varphi_0(\mathbf{u})$ denote the analogues of $\widetilde{\varphi}_T(\mathbf{u}, \widehat{\boldsymbol{\vartheta}}_T)$ and $\varphi(\mathbf{u})$, respectively, when the innovations are distributed as $\boldsymbol{\varepsilon}_{01}$.

Lemma 8.9 *Under the assumptions of Theorem 5.1, conditional to almost all sequence* $\boldsymbol{y} = \{\boldsymbol{y}_t\}$*, the finite-dimensional distributions of*

$$\left(\sqrt{T}\left\{\widetilde{\varphi}_T^*(\mathbf{u}, \widehat{\boldsymbol{\vartheta}}_T^*) - \varphi_T^*(\mathbf{u})\right\}; \mathbf{u} \in \mathbb{R}^d\right)$$

converge to those of $\left(\sqrt{T}\left\{\widetilde{\varphi}_{0T}(\mathbf{u}, \widehat{\boldsymbol{\vartheta}}_T) - \varphi_0(\mathbf{u})\right\}; \mathbf{u} \in \mathbb{R}^d\right)$.

Proof of Lemma 8.9 Letting $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}) = \widetilde{\mathbf{C}}_t^{-1/2}(\boldsymbol{y}_{t-1}^*, \ldots, \boldsymbol{y}_1^*; \boldsymbol{\vartheta})\boldsymbol{y}_t^*$, we have the resampling innovations $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t^*(\widehat{\boldsymbol{\vartheta}}_T)$ and the resampling residuals $\widetilde{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t^*(\widehat{\boldsymbol{\vartheta}}_T^*)$. Lemma 8.6 shows that, asymptotically, the resampling innovations satisfactorily mimic the GARCH innovations $\boldsymbol{\varepsilon}_t$. Informally, it remains to show that the resampling residuals properly mimic the behavior of the GARCH residuals $\widetilde{\boldsymbol{\varepsilon}}_t$.

For the sake of simplicity, we only show the convergence of the marginal distributions of the real part. The convergence of the imaginary part and of the multidimensional distributions can be obtained by similar arguments.

Let

$$\widehat{\varphi}_T^*(\mathbf{u}, \boldsymbol{\vartheta}) = \frac{1}{T} \sum_{t=1}^{T} e^{\mathrm{i}\mathbf{u}'\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})}.$$

A Taylor expansion entails that, for any fixed $\mathbf{u}$, we have

$$\begin{aligned}\sqrt{T}\left\{\mathrm{Re}\,\widehat{\varphi}^*_T(\mathbf{u},\widehat{\boldsymbol{\vartheta}}^*_T)-\varphi^*_T(\mathbf{u})\right\} &= \frac{1}{\sqrt{T}}\sum_{t=1}^T\{\cos(\mathbf{u}'\boldsymbol{\varepsilon}^*_t)-\varphi^*_T(\mathbf{u})\}\\ &+\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}}^*_T-\widehat{\boldsymbol{\vartheta}}_T\right)'\frac{1}{T}\sum_{t=1}^T\frac{\partial}{\partial\boldsymbol{\vartheta}}\cos\{\mathbf{u}'\boldsymbol{\varepsilon}^*_t(\boldsymbol{\vartheta}_T)\}.\end{aligned}$$

for some $\boldsymbol{\vartheta}_T$ between $\widehat{\boldsymbol{\vartheta}}^*_T$ and $\widehat{\boldsymbol{\vartheta}}_T$. Note that $E(\widetilde{\boldsymbol{g}}^*_t \mid \boldsymbol{y})=0$. Moreover (8.13) entails that $\sup_{t,T}E\left(\|\widetilde{\boldsymbol{g}}^*_t\widetilde{\boldsymbol{g}}^{*'}_t\|\mid\boldsymbol{y}\right)<\infty$. It follows that any sequence satisfying (5.1) is such that $\widehat{\boldsymbol{\vartheta}}^*_T\to\boldsymbol{\vartheta}_0$ in probability. Using the notation introduced in the proof of Lemma 8.2, it can also be shown that

$$(8.16)\quad \frac{1}{T}\sum_{t=1}^T\frac{\partial}{\partial\boldsymbol{\vartheta}}\cos\{\mathbf{u}'\boldsymbol{\varepsilon}^*_t(\boldsymbol{\vartheta}_T)\}\to\dot{\boldsymbol{\varphi}}_0(\mathbf{u})\quad\text{in probability}$$

for any sequence $\boldsymbol{\vartheta}_T$ tending to $\boldsymbol{\vartheta}_0$ in probability, where $\dot{\boldsymbol{\varphi}}_0(\mathbf{u})$ denotes the analogue of $\dot{\boldsymbol{\varphi}}(\mathbf{u})$ when the innovations are distributed as $\boldsymbol{\varepsilon}_{01}$. Indeed, in view of (8.7) we have, for $j=1,\dots,s_1$,

$$(8.17)\quad \frac{\partial}{\partial\vartheta_j}\cos\left\{\mathbf{u}'\boldsymbol{\varepsilon}^*_t(\widehat{\boldsymbol{\vartheta}}_T)\right\} = \sin(\mathbf{u}'\boldsymbol{\varepsilon}^*_t)\,\mathbf{u}'\widehat{\mathbf{R}}^{-1/2}\widehat{\mathbf{D}}^{*-1}_t\widehat{\mathbf{D}}^{*(j)}_t\widehat{\mathbf{R}}^{1/2}\boldsymbol{\varepsilon}^*_t,$$

where $\widehat{\mathbf{D}}^*_t=\widetilde{\mathbf{D}}_t(\boldsymbol{y}^*_{t-1},\dots,\boldsymbol{y}^*_1;\widehat{\boldsymbol{\vartheta}}_T)$ and $\widehat{\mathbf{D}}^{*(j)}_t=\partial\widehat{\mathbf{D}}^*_t/\partial\vartheta_j$. As in Lemma 8.8, one can show that the distribution of the random variable defined by (8.17) conditional on $\boldsymbol{y}$ tends to that of

$$\sin(\mathbf{u}'\boldsymbol{\varepsilon}_t)\,\mathbf{u}'\mathbf{R}_0^{-1/2}\mathbf{D}_{0t}^{-1}\mathbf{D}_{0t}^{(j)}\mathbf{R}_0^{1/2}\boldsymbol{\varepsilon}_{0t}.$$

Note that the expectation of the previous variable is equal to the j-th element of $\dot{\boldsymbol{\varphi}}_0(\mathbf{u})$. One can then obtain (8.16) by the arguments used in the proof of Lemma 8.2. Using (5.1), we thus have

$$\sqrt{T}\left\{\mathrm{Re}\,\widehat{\varphi}^*_T(\mathbf{u},\widehat{\boldsymbol{\vartheta}}^*_T)-\varphi^*_T(\mathbf{u})\right\}\overset{o_P(1)}{=}\frac{1}{\sqrt{T}}\sum_{t=1}^T x^*_{t,T}$$

with

$$x^*_{t,T}=\cos(\mathbf{u}'\boldsymbol{\varepsilon}^*_t)-\varphi^*_T(\mathbf{u})+\widetilde{\boldsymbol{g}}^{*'}_t\widetilde{\boldsymbol{\psi}}'_{t-1}\dot{\boldsymbol{\varphi}}_0(\mathbf{u}).$$

Note that the asymptotic distribution of $\sqrt{T}\text{Re}\left\{\widehat{\varphi}_{0T}(\mathbf{u},\widehat{\boldsymbol{\vartheta}}_{\mathbf{T}})-\varphi_0(\mathbf{u})\right\}$ is $\mathcal{N}(0,\sigma^2)$, where

$$\sigma^2=\text{Var}\left\{\text{Re}\, g_{01}(\mathbf{u})\right\},$$

and $g_{01}(\mathbf{u})$ stands for the analogue of $g_1(\mathbf{u})$ when the innovations are distributed as $\boldsymbol{\varepsilon}_{01}$ (see the proof of Theorem 4.1). It now suffices to show that, conditional on $\boldsymbol{y}$, for any fixed $\mathbf{u}$,

$$\text{(8.18)}\quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{t,T}^{*}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2\right).$$

Note that, conditional on $\boldsymbol{y}$, for each T the random variables $x_{1,T}^{*},x_{2,T}^{*},\dots$ are independent and centered, with finite second-order moments. From the Lindeberg CLT for triangular arrays, to prove (8.18) it suffices to show that

$$\text{(8.19)}\quad \frac{1}{T}\sum_{t=1}^{T}\text{Var}\left(x_{t,T}^{*}\mid\boldsymbol{y}\right)\to\sigma^2,$$

and that for all $\epsilon>0$

$$\text{(8.20)}\quad \frac{1}{T}\sum_{t=1}^{T}E\left\{x_{t,T}^{*2}1_{\left\{|x_{t,T}^{*}|\geq\sqrt{T}\epsilon\right\}}\mid\boldsymbol{y}\right\}\to 0.$$

Let

$$\widetilde{\boldsymbol{g}}_t(\boldsymbol{\varepsilon})=\text{vec}\left(\mathbf{I}_d-\widehat{\mathbf{R}}^{1/2}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\widehat{\mathbf{R}}^{-1/2}\right).$$

Note that $\widetilde{\boldsymbol{g}}_t^{*}=\widetilde{\boldsymbol{g}}_t(\boldsymbol{\varepsilon}_t^{*})$. Define similarly $x_{t,T}(\boldsymbol{\varepsilon})$ such that $x_{t,T}=x_{t,T}(\boldsymbol{\varepsilon}_t^{*})$. By (8.14) of Lemma 8.7, for almost all $\boldsymbol{y}$ as $T\to\infty$, we then have

$$\text{Var}\left(x_{t,T}^{*}\mid\boldsymbol{y}\right)\to\lim_{T\to\infty}Ex_{t,T}^{2}(\boldsymbol{\varepsilon}_{0t})=\sigma^2,$$

and (8.19) follows from Cesàro's lemma. Because $E\left\{x_{t,T}^{*2}\mid\boldsymbol{y}\right\}<\infty$ and the event $\left\{|x_{t,T}^{*}|\geq\sqrt{T}\epsilon\right\}$ tends to $\emptyset$ as $T\to\infty$, the dominated convergence theorem shows (8.20), which completes the proof. □

Lemma 8.10 *Under the assumptions of Theorem 5.1, conditional on almost all sequences $\boldsymbol{y}=\{\boldsymbol{y}_t\}$, for any $\overline{\rho}>0$ and any $\mathbf{u}\in\mathbb{R}^d$, the sequences*

$$\left\{T^{-1/2}\sum_{t=1}^{T}Re\left\{\widetilde{\varphi}_T^{*}(\rho\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T^{*})-\varphi_T^{*}(\rho\mathbf{u})\right\},\ \rho\in[0,\overline{\rho}]\right\}_{T\geq 1}$$

and

$$\left\{T^{-1/2}\sum_{t=1}^{T} Im\left\{\widetilde{\varphi}_T^*(\rho\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T^*)-\varphi_T^*(\rho\mathbf{u})\right\},\ \rho\in[0,\overline{\rho}]\right\}_{T\geq 1}$$

are tight in $\mathcal{C}([0,\overline{\rho}])$.

Proof of Lemma 8.10 From the proof of Lemma 8.9,

$$\sqrt{T}\mathrm{Re}\left\{\widetilde{\varphi}_T^*(\rho\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T^*)-\varphi_T^*(\rho\mathbf{u})\right\}\overset{o_P(1)}{=}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\cos\left(\rho\mathbf{u}'\boldsymbol{\varepsilon}_t^*\right)-\varphi_T^*(\rho\mathbf{u})+\widetilde{\boldsymbol{g}}_t^{*\prime}\widetilde{\boldsymbol{\psi}}_{t-1}'\dot{\boldsymbol{\varphi}}_0(\mathbf{u})\right\},$$

uniformly in $\rho\in[0,\overline{\rho}]$. Thus, to show the tightness of the left hand side part of the above expression it suffices to show the tightness of the process on the right hand side. The tightness of $\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\cos(\rho\mathbf{u}'\boldsymbol{\varepsilon}_t^*)-\varphi_T^*(\rho\mathbf{u}),\ \rho\in[0,\overline{\rho}]\}_{T\geq 1}$ comes from the inequality $|\cos(x)-\cos(y)|\leq|x-y|$, **A7** and Theorem 12.3 in Billingsley (1968) The rest of the proof follows the lines of that of Lemma 8.5. □

Proof of Theorem 5.1. The statistic KS_T^* is based on the empirical CF $\widehat{\varphi}_T^*(\mathbf{u},\widehat{\boldsymbol{\vartheta}}_T^*)$ at a finite number of points $\mathbf{u}$. Thus the result for this test statistic is a direct consequence of Lemma 8.9. For the statistic CM_T^* the result is obtained by using Lemmas 8.9 and 8.10 and proceeding as in the last part of the proof of Theorem 4.1. □

Acknowledgement The work of S. Meintanis was supported by research grant no. 11699 of the Special Account for Research Grants (ELKE) of the National and Kapodistrian University of Athens. M.D. Jiménez-Gamero acknowledges financial support from grant MTM2014-55966-P (FEDER support included) of the Spanish Ministry of Economy and Competitiveness.

References

Amengual, D. and Sentana, E. (2011). Inference in multivariate dynamic models with elliptical innovations. *Econometrics Seminar*, *TSE*, Toulouse.

Arellano-Valle, R.B., Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distrinution. *J. Multivariate Anal.*, *99*, 1362–1382.

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

Bai, J. and Ng, S. (2001). A test for conditional symmetry in time series models. *J. Econometr.*, *103*, 225–258.

Baringhaus, L. (1991). Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.*, *19*, 899–917.

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to GARCH models. *J. Bus. Econom. Statist.*, *23*, 346–354.

Berk, J. (1997). Necessary conditions for the CAPM. *J. Econom. Theor.*, *73*, 245–257.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

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

Cacoullos, T. (2014). Polar angle tangent vectors follow Cauchy distributions under spherical symmetry. *J. Multivar. Anal.*, *128*, 147–153.

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

Csörgő, S. (1981). Limit behavior of the empirical characteristic function. *Ann. Probab.*, *9*, 130–144.

Delgado, M., and Escanciano, J.C. (2007). Nonparametric tests for conditional symmetry in dynamic models. *J. Econometr.*, *141*, 652–682.

Delgado, M., and Stute, W. (2008). Distribution–free specification tests of conditional models. *J. Econometr.*, *143*, 37–55.

De Luca, G., Genton, M.G., and Loperfido, N. (2006). A multivariate skew–GARCH model. In D. Terrell (Ed.) *Advances in Econometrics: Econometric Analysis of Economic and Financial Timee Series, Part A*, Elsevier, pp. 33–57.

Diks, C. and Tong, H. (1999). A test for symmetries of multivariate probability distributions. *Biometrika*, *86*, 605–614.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In M.A.H. Dempster (Ed) *Risk Management: Value at Risk and Beyond*, Cambridge University Press, London, pp. 176-223.

Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

Francq, C., Horváth, L., and Zakoïan, J.M. (2015). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, published online, DOI: 10.1093/jjfinec/nbu030.

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. *Econom. Theory* , *28*, 179–206.

Ghosh, S. and Ruymgaart, F.H. (1992). Applications of empirical characteristic functions in some multivariate problems. *Canad. J. Statist.*, *20*, 429–440.

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

Haas, M., Mittnik, S, and Paolella, M.S. (2009). Asymmetric multivariate normal mixture GARCH. *Comput. Statist. Dat. Anal.*, *53*, 2129–2154.

Hafner, C.M. and Rombouts, J.V.K. (2007). Semiparametric multivariate volatility models. *Econometr. Theor.*, *23*, 251–280.

Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo–Mahalanobis ranks. *Ann. Statist.*, *30*, 1103–1133.

Hallin, M. and Werker, B.J.M. (2003). Semi–parametric efficiency, distribution–freeness and invariance. *Bernoulli*, *9*, 137–165.

Henze, N., Hlávka, Z., and Meintanis, S.G. (2014). Testing for spherical symmetry via the empirical characteristic function. *Statisics*, *48*, 1282–1296.

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

Jones, M. C. (2008). The distribution of the ratio X/Y for all centered elliptically symmetric distributions. *J. Multivar. Anal.*, *99*, 572–573.

Kariya, T. and Eaton, M.L. (1977). Robust tests for spherical symmetry. *Ann. Statist.*, *5*, 206–215.

Koltchinskii, V.I. and Li, L. (1998). Testing for spherical symmetry of a multivariate distribution. *J. Multivar. Anal.*, *65*, 228–244.

Kreiss, J.P., Paparoditis, E. and Politis, D.N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.*, *39*, 2103–2130.

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

Lee, S., Park, S. and Lee, T. (2010). A note on the Jarque–Bera normality test for GARCH innovations. *J. Kor. Statist. Soc.*, *39*, 93–102.

Liu, S., Heyde, C.C. and Wong, W.K. (2011). Moment matrices in conditional heteroskedastic models under elliptical distributions with applications in AR-ARCH models. *Statist. Pap.*, *52*, 621–632.

Meintanis, S.G. and Ngatchou-Wandji, J. (2012). Recent tests for symmetry with multivariate and structured data: A review. In *Nonparametric Statistical Methods and Related Topics* 35–73, World Scientific Publishing Company, London.

Mittnik, S. and Paolella, M.S. (2000). Conditional density and value–at–risk prediction of Asian currency exchange rates. *J. Forecasting*, *19*, 313–333.

Reinsel, G.C. (1997). Elements of Multivariate Time Series Analysis. Springer-Verlag, New York.

Shimizu, K. (2013). The bootstrap does not always work for heteroscedastic models. *Statistics & Risk Modeling*, *30*, 189–204.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In T.G. Andersen et al. (eds.): *Handbook of Financial Time Series* 201–229, Springer, Berlin.

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

Tsay, R.S. (2002). Analysis of Financial Time Series. Wiley, New York.

Tsay, R.S. (2014). Multivariate Time Series Analysis: with R and Financial Applications. Wiley.

Zhu, L.X. (2005). Asymptotics of goodness–of–fit tests for symmetry. In: Lecture Notes in Statistics *20*, *NonParametric Monte Carlo Tests and Their Applications*, 27–43, Springer, New York.

Zhu, L.X. and Neuhaus, G. (2000). Nonparametric Monte Carlo tests for multivariate distributions. *Biometrika*, *87*, 919–928.

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

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, *28*, 461–482.