



**HAL**  
open science

# A Goodness-of-Fit Test for Elliptical Distributions with Diagnostic Capabilities

Gilles R. Ducharme, Pierre Lafaye de Micheaux

► **To cite this version:**

Gilles R. Ducharme, Pierre Lafaye de Micheaux. A Goodness-of-Fit Test for Elliptical Distributions with Diagnostic Capabilities. 2019. hal-02070401

**HAL Id: hal-02070401**

**<https://hal.science/hal-02070401>**

Preprint submitted on 17 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A GOODNESS-OF-FIT TEST FOR ELLIPTICAL DISTRIBUTIONS WITH DIAGNOSTIC CAPABILITIES

BY GILLES R. DUCHARME<sup>†</sup>, AND PIERRE LAFAYE DE MICHEAUX<sup>‡</sup>

*Université de Montpellier<sup>†</sup> and UNSW Sydney<sup>‡</sup>*

This paper develops a smooth test of goodness-of-fit for elliptical distributions. The test is adaptively omnibus, invariant to affine-linear transformations and has a convenient expression that can be broken into components. These components have diagnostic capabilities and can be used to identify specific departures. This helps in correcting the null model when the test rejects. As an example, the results are applied to the multivariate normal distribution for which the R package `ECGofTestDx` is available. It is shown that the proposed test strategy encompasses and generalizes a number of existing approaches. Some other cases are studied, such as the bivariate Laplace, logistic and Pearson type II distribution. A simulation experiment shows the usefulness of the diagnostic tools.

**1. Introduction.** Elliptically contoured (EC), or elliptical for short, distributions have become important tools in the analysis of multivariate data. They retain from the multivariate normal (MVN) distribution the feature of elliptical symmetry about a location  $\boldsymbol{\mu}$ . They extend the MVN to allow modelling data with short or large tails. Tools of multivariate analysis such as regression, correlation, PCA, discrimination, are easily ported to them. Many inferential procedures from MVN theory remain valid, after slight modifications. They pave the way toward more flexible models, such as skew-elliptic distributions, or more specialized, such as elliptical copulas, elliptical complex and elliptical matrix distributions. Consequently, they are used in many applications, such as portfolio theory, risk management, radioimmunoassay, astronomy, physics, signal processing, etc. (see [8] for a bibliography).

In this paper we focus on  $m$ -dimensional EC distributions with density of the form

---

\*The authors would like to thank Bernard Boulerice for his contribution to a preliminary version of this paper.

*MSC 2010 subject classifications:* Primary 62F03, 62H05; secondary 62E10

*Keywords and phrases:* Diagnostic Information, Elliptical Distribution, Goodness-of-fit test, Multivariate Laplace distribution, Multivariate logistic distribution, Multivariate normal distribution, Multivariate Pearson type II distribution, Smooth tests

$$(1.1) \quad f(\mathbf{x}; \eta) = c_m \sqrt{\det(\mathbf{V}^{-1})} \phi_m((\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})),$$

where  $c_m = \Gamma(m/2) / (\pi^{m/2} \int_0^\infty y^{m/2-1} \phi_m(y) dy)$  is a normalizing constant, the density generator  $\phi_m(\cdot)$  is defined over  $\mathbb{R}_{\geq 0}$  and the parameter  $\eta = (\boldsymbol{\mu}, \mathbf{V}^{-1}) \in \Xi = \mathbb{R}^m \times \mathbb{M}_+^m$ , the space of  $m \times m$  positive definite matrices, is unknown. To give a few examples taken from [4, Chap. 13], the MVN has  $\phi_m(y) = \exp\{-y/2\}$ ; the multiStudent with  $\nu$  degrees of freedom has  $\phi_m(y) = (1 - y/\nu)^{-(m+\nu)/2}$ ; the multivariate  $\alpha$ -th power exponential distribution has  $\phi_m(y) = \exp\{-y^\alpha\}$ . The multivariate slash-elliptical distribution with shape parameter  $\alpha$  [50] has  $\phi_m(y) = y^{-(m+\alpha)/2} \int_0^{y/2} t^{(m+\alpha)/2} e^{-t} dt$ . A more elaborate version, the multivariate extended slash-elliptical distribution has been introduced recently [45] to fit heavy-tailed data. Additions to the catalogue of EC distributions are periodically made in response to challenging new data sets.

To exploit these statistical models in applications, a user needs tools to help in selecting a distribution from this growing catalogue. One such tool is a Goodness-of-Fit (GoF) test of the null hypothesis that some data arise from a postulated EC density of the form (1.1). The goal of the present work is to develop such a GoF test having attractive features.

An important feature of a GoF test is its power in detecting departures from the null density. GoF tests can be roughly divided into omnibus and directional. An omnibus test will have power converging to one (i.e. be consistent) for any departure from the null. A directional GoF test will be consistent for only some departures. At first glance, the omnibus property appears essential. But equal power cannot be allocated to each departure and, for finite samples, only some will have significant power (see [47, Chap. 5]). Thus the interest for compromising approaches offering some control over the power allocation, such as the smooth test paradigm introduced by [40]. The smooth test is directional along a set of  $K$  departures, but the value of  $K$  can be selected. This creates a slider between directional and omnibus GoF tests that yields an adaptive form of the omnibus property which has been found, in many simulations, to be very effective. This has brought [44, p. 9] to recommend : “*Don’t use those other methods—use a smooth test!*”. In its original form, its main weakness is the lack of criteria to position  $K$  on the slider. To correct this, [31] has introduced a version that selects  $K$  in a data-driven fashion. The resulting data-driven smooth test constitutes, power-wise, a significant improvement. This has prompted [23] to sharpen [44]’s recommendation into : “*use a data-driven smooth test!*”.

Another desirable feature of a GoF test is the ability to extract, when

the test rejects, some information regarding the aspects of the null density contradicted by the data. As stated in [15, p. 800] regarding the process of model checking: “*the purpose of the checking is not simply to “accept” or “reject”, but rather to discover what aspects of the data are not being fit well by the model*”. A GoF test that can provide such information is said to have diagnostic (Dx) capabilities. The smooth test has Dx capabilities, see [11].

Encouraged by these features, much work ([21, 49, 29, 23, 43, 48] among others) has been done to extend the smooth test paradigm. However, for multivariate densities, few extensions have appeared. One hurdle comes from the fact that EC distributions are closed under the group of affine-linear transformations of the data. The invariance principle states that if a statistical problem is left invariant after a transformation, then its solution should also be invariant under this transformation, otherwise interpretability of the inference can be compromised. [19, p. 469] nicely summarizes the importance of this principle : “... *any proposal to use a non-invariant test... must come together with a special justification*”. For the case of the bivariate normal, [5] has developed a data-driven smooth test that cleverly combines the smooth test statistics for univariate normality. But it is not invariant to rotations, so that an unscrupulous user could simply rotate the data to reach a pre-specified conclusion. Also her approach breaks down for general EC distributions because a multivariate EC density does not always have the same density generator as its marginals (the so-called inconsistency property, see [4, p. 209]). [6] develops a smooth GoF test invariant to rotations, but their method is confined to data lying on an hypersphere.

Because of this scarcity, users facing the problem of assessing the fit of a multivariate EC density may feel abandoned by statistical methodology. When the null density is the MVN, numerous GoF tests exist (see [28, 46, 19, 41] and references therein). However, GoF tests tailored to other EC distributions are almost non-existent (with the noticeable exceptions of [14] and [12]) and a user must resort to a general-purpose GoF test, such as the multivariate version of the Cramer von Mises approach (see [36] for a short review), or ad hoc methods.

In this paper, we develop a smooth GoF test for EC densities of the form (1.1) with unknown parameter  $\eta$ , that is adaptively omnibus, has Dx capabilities and is affine-linear invariant. As a bonus, the asymptotic reference distribution is a standard  $\chi^2$  and its power function can be approximated by a sum of independent non-central  $\chi^2$ . The approach is based on the ideas in [6] but adapted to the case of affine-linear transformations. Section 2 adapts the smooth test paradigm to the case of a general EC density and sets the stage for the several levels of invariance required in our approach:

problem invariance, test statistic invariance and  $D_x$  invariance. Section 3 makes explicit the orthonormal basis on which our smooth test is based and derives the associated test strategy by exploiting a variant of Rao's score test. Section 4 shows that the smooth test statistic possesses a convenient explicit expression from which emerges a decomposition into a sum of three invariant and asymptotically independent  $\chi^2$  distributed components, which we refer to as the  $\mathcal{Q} = \mathcal{UIR}$  decomposition. It is explained that each component can provide interpretable  $D_x$  information about the aspects of the null EC density contradicted by the data. It is also explained how the omnibus/directional slider must be set up to preserve invariance in all components of this decomposition. Section 5 computes the  $\mathcal{Q} = \mathcal{UIR}$  decomposition for the multivariate normal (MVN) distribution and discusses how our smooth test extends and relates to a number of proposals for this problem. The R package `ECGofTestDx` can perform the necessary calculations. The case of the bivariate Laplace is also considered. Section 6 reports on an experiment that was conducted to see how the elements of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition behave under a number of alternative densities somewhat representative of what can be encountered in actual situations. It is seen that the decomposition can indeed help in providing illuminating  $D_x$  information. Some points requiring further research are collected in Section 7. Proofs are confined to Appendix A and Appendix B gives the details for the computation of the smooth test for the MVN when  $m = 2, 3$  as well as MATHEMATICA commands to extend in higher dimensions. Appendix C gives the smooth test for two other EC distribution, the bivariate logistic and Pearson type II distributions.

**2. The Smooth Test Paradigm for Elliptical Distributions.** Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent and identically distributed  $m$ -dimensional observations with density  $\psi(\cdot)$ . Consider the problem of testing :

$$(2.1) \quad H_0 : \psi(\cdot) \in \mathcal{F}_0 = \{f(\cdot; \eta), \eta \in \Xi\} \quad vs \quad H_1 : \psi(\cdot) \notin \mathcal{F}_0,$$

where  $f(\cdot; \eta)$  has the form (1.1) with a given  $\phi_m(\cdot)$ . Here we consider the usual case where  $\eta = (\boldsymbol{\mu}, \mathbf{V}^{-1})$  is unknown. We set the following :

*Assumption A* : The support of  $f(\cdot; \eta)$  does not depend on  $\eta$  and the mapping from  $\Xi$  to  $\mathcal{F}_0$  is one-to-one. Moreover  $f(\cdot; \eta)$  has a moment generating function so that all moments of  $\mathbf{X}_1$  exist. Also  $\mathbb{P}[\mathbf{X}_1 = \mathbf{0}] = 0$ .

*Assumption B* :  $\hat{\eta}$  is an affine-equivariant estimator (in the sense of Definition 13.1 in [4]) of  $\eta$  such that  $\sqrt{n}(\hat{\eta} - \eta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\ell}(\mathbf{X}_i, \eta) + o_p(1)$ , where  $\mathbb{E}_0(\boldsymbol{\ell}(\mathbf{X}_1, \eta)) = \mathbf{0}$  and the covariance matrix  $\mathbb{V}_0(\boldsymbol{\ell}(\mathbf{X}_1, \eta))$  is finite.

Now confine  $\psi(\cdot)$  to the embedding family of densities

$$(2.2) \quad \mathcal{G} = \{g(\cdot; \eta, h(\cdot)) \mid \eta \in \Xi, h(\cdot) \in \mathbb{H} \text{ and } g(\cdot; \eta, \mathbf{0}(\cdot)) = f(\cdot; \eta)\}$$

where  $\mathbf{0}(\cdot)$  is the zero function assumed to belong to a suitable space of functions  $\mathbb{H}$ . This allows to restate problem (2.1) as that of testing

$$(2.3) \quad H_0 : h(\cdot) = \mathbf{0}(\cdot) \quad vs \quad H_1 : h(\cdot) \neq \mathbf{0}(\cdot).$$

Here we adopt from [10] embedding functions of the form

$$(2.4) \quad g(\cdot; \eta, h(\cdot)) = f(\cdot; \eta) \times \frac{(1 + h(\cdot) - \mathbb{E}_0(h(\mathbf{X})))^2}{1 + \|h(\cdot) - \mathbb{E}_0(h(\mathbf{X}))\|_{(f, \eta)}^2},$$

where the subscript “0” refers to a statistical operator evaluated under  $H_0$  and  $\|h(\cdot)\|_{(f, \eta)}^2 = \int h^2(\mathbf{x})f(\mathbf{x}; \eta)d\mathbf{x}$ . These are related to Hellinger’s metric.

The confinement to  $\mathcal{G}$ , and thus the choice of  $\mathbb{H}$ , affects the properties of our test. Regarding invariance, let  $AL(m)$  be the group of affine-linear transformations on  $\mathbb{R}^m$  with generic element  $\gamma = \gamma(\mathbf{x}) = \mathbf{A}(\mathbf{x} + \mathbf{b})$ , represented by the pair  $(\mathbf{A}, \mathbf{b})$  where  $\mathbf{A}$  is a  $m \times m$  non-singular real matrix and  $\mathbf{b} \in \mathbb{R}^m$ . EC distributions are closed under these transformations: the density of  $\gamma(\mathbf{X})$  has the form (1.1) with its parameters transformed from  $\eta = (\boldsymbol{\mu}, \mathbf{V}^{-1})$  to  $\gamma^*(\eta) = (\mathbf{A}(\boldsymbol{\mu} + \mathbf{b}), (\mathbf{A}^{-1})^T \mathbf{V}^{-1} \mathbf{A}^{-1})$  [4, p. 207]. Our GoF test, and its Dx information, must be invariant to such transformations. To this end, let  $L^2(f, \eta) = \{h(\cdot) \mid \|h(\cdot)\|_{(f, \eta)}^2 < \infty\}$  and set  $\mathbb{H} = \{h \in L^2(f, \eta) \mid \|h \circ \gamma(\cdot)\|_{(f, \eta)} < \infty, \forall \gamma \in AL(m)\}$ . With this choice, embedding family (2.2) does not depend on the value of  $\eta$  and is obviously closed under affine-linear transformations. This will be important to ensure invariance. Note that  $\mathbb{H}$  is a slight restriction of the Banach space  $L^2(f, \eta)$ , so that almost any reasonable  $\psi(\cdot)$  can be written as (2.4). This is important for the omnibus property.

We now recall the following terminology. A subspace  $\mathcal{H}$  of  $\mathbb{H}$  is said *AL(m)-invariant* if  $h(\cdot) \in \mathcal{H}$  implies  $h \circ \gamma(\cdot) \in \mathcal{H}$  for all  $\gamma \in AL(m)$ . An *AL(m)-invariant* subspace is further said to be *irreducible* if it contains no nontrivial *AL(m)-invariant* subspaces. For any  $\eta \in \Xi$ , let  $\langle \cdot, \cdot \rangle_{(f, \eta)}$  be the

scalar product associated with  $\|\cdot\|_{(f,\eta)}$ .  $\mathbb{H}$  equipped with this scalar product is a Hilbert space denoted  $\mathbb{H}_{(f,\eta)}$ . With these definitions at hand, a variant of the argument in [6] shows that  $\mathbb{H}_{(f,\eta)}$  can be decomposed into a sequence of disjoint irreducible  $AL(m)$ -invariant subspaces  $\{\Pi_k, k \geq 0\}$ , each of a finite dimension  $d_m(k)$  that will be made explicit in Theorem 3.1 below. Hence,

$$(2.5) \quad \mathbb{H}_{(f,\eta)} = \bigoplus_{k=0}^{\infty} \Pi_k,$$

where  $\bigoplus$  is the direct sum operator with respect to  $\langle \cdot, \cdot \rangle_{(f,\eta)}$  and  $\Pi_0$  is the set of constant functions. For each  $k \geq 0$ , let  $\{\pi_{k,j,\ell}(\cdot; \eta), (j, \ell) \in B_k\}$  be a complete orthonormal basis (CONB) for  $\Pi_k$ . Here  $B_k$  is a set of  $d_m(k)$  pairs of integers that will be explicated, along with a choice  $\pi_{k,j,\ell}(\cdot; \eta)$  in Theorem 3.1. Thus  $\text{Span}\{\pi_{k,j,\ell}(\cdot; \eta), (j, \ell) \in B_k\} = \Pi_k$  and for any  $(j, \ell), (j', \ell') \in B_k$ ,  $\langle \pi_{k,j,\ell}(\cdot; \eta), \pi_{k,j',\ell'}(\cdot; \eta) \rangle_{(f,\eta)} = \delta_{jj'} \delta_{\ell\ell'}$  where  $\delta$  is Kronecker's delta. It follows from (2.5) that the whole set of  $\pi_{k,j,\ell}(\cdot; \eta)$  forms a CONB for  $\mathbb{H}_{(f,\eta)}$  and that any  $h(\cdot) \in \mathbb{H}_{(f,\eta)}$  can be written as

$$(2.6) \quad h(\cdot) = \sum_{k=0}^{\infty} \sum_{(j,\ell) \in B_k} \theta_{k,j,\ell}(\eta) \pi_{k,j,\ell}(\cdot; \eta),$$

where  $\theta_{k,j,\ell}(\eta) = \langle h(\cdot), \pi_{k,j,\ell}(\cdot; \eta) \rangle_{(f,\eta)}$ . With (2.6) the problem of testing (2.3) reduces to that of testing the nullity of all (for an omnibus test within  $\mathcal{G}$ ) or part (for a directional version) of the  $\theta_{k,j,\ell}(\eta)$ . Focusing on the latter, fix  $K$  and consider the subset of  $\mathcal{G}$  that contains the functions of the form

$$(2.7) \quad g_K(\cdot; \eta, \boldsymbol{\theta}(\eta)) = f(\cdot; \eta) \times \frac{\left(1 + \sum_{k=1}^K \sum_{(j,\ell) \in B_k} \theta_{k,j,\ell}(\eta) \pi_{k,j,\ell}(\cdot; \eta)\right)^2}{1 + \sum_{k=1}^K \sum_{(j,\ell) \in B_k} \theta_{k,j,\ell}^2(\eta)}.$$

Further confining  $\psi(\cdot)$  to this subset reduces the problem of testing (2.1) to that of testing

$$(2.8) \quad H_0 : \boldsymbol{\theta}(\eta) = \mathbf{0} \quad vs \quad H_1 : \boldsymbol{\theta}(\eta) \neq \mathbf{0},$$

where  $\boldsymbol{\theta}(\eta) = \{\theta_{k,j,\ell}(\eta), 1 \leq k \leq K, (j, \ell) \in B_k\}$ . Under (2.7), it is easy to see that if  $\boldsymbol{\theta}(\eta)$  is small, the dominant term in the maximum likelihood estimator (mle) of  $\theta_{k,j,\ell}(\eta)$  is  $n^{-1} \sum_{i=1}^n \pi_{k,j,\ell}(\mathbf{X}_i; \eta)$ . Hence, it makes statistical sense (for another justification, see [27]) to base a test of (2.8) on

$$(2.9) \quad \bar{\boldsymbol{\pi}}(\hat{\eta}) = \left( n^{-1} \sum_{i=1}^n \pi_{k,j,\ell}(\mathbf{X}_i; \hat{\eta}), 1 \leq k \leq K, (j, \ell) \in B_k \right)^T.$$

If the  $\pi_{k,j,\ell}(\cdot; \cdot)$  are continuously differentiable in  $\eta$ , then Theorem 2.1 a) of [27] ensures the asymptotic normality (with asymptotic expectation  $\mathbf{0}$ ) of  $\sqrt{n}\bar{\boldsymbol{\pi}}(\hat{\eta})$  under  $H_0$ . The asymptotic covariance matrix depends on the choice of  $\hat{\eta}$ . In particular, if  $f(\cdot; \eta)$  satisfies the conditions in [13, p. 121], then  $\hat{\eta}$  can be taken as the mle of  $\eta$ , which satisfies *Assumption B* with  $\mathbb{V}_0(\ell(\mathbf{X}_i, \eta)) = \mathcal{J}_\eta^{-1}$ , where  $\mathcal{J}_\eta$  is Fisher's information for  $\eta$  under  $H_0$ . If  $\mathbf{J}_\eta$  denotes the matrix with elements  $\text{Cov}_0(\pi_{k,j,\ell}(\mathbf{X}_i; \eta), \partial \log f(\mathbf{X}_i; \eta) / \partial \eta_a)$  where  $\eta_a$  is a  $a$ -th component of  $\eta$ , then from Theorem 2.1 c) of [27], this covariance matrix is  $\mathbf{I}_{\nu_Q} - \mathbf{J}_\eta \mathcal{J}_\eta^{-1} \mathbf{J}_\eta^T$  where  $\mathbf{I}_\nu$  is the identity matrix of order  $\nu$  and  $\nu_Q = \sum_{k=1}^K d_m(k)$ . Assume this matrix invertible. Then test statistic

$$(2.10) \quad \mathcal{Q}_K = \mathcal{Q}_K(\mathbf{X}, \hat{\eta}) = n \bar{\boldsymbol{\pi}}^T(\hat{\eta}) (\mathbf{I}_{\nu_Q} - \mathbf{J}_{\hat{\eta}} \mathcal{J}_{\hat{\eta}}^{-1} \mathbf{J}_{\hat{\eta}}^T)^{-1} \bar{\boldsymbol{\pi}}(\hat{\eta})$$

is under  $H_0$  asymptotically  $\chi_{\nu_Q}^2$ . We refer to  $\mathcal{Q}_K$  as a *global* test statistic for  $H_0$ . As a by-product of the CONB introduced in the next section, an explicit expression for this test statistic is available (see (4.1) and (4.2)).

For this to make statistical sense requires  $AL(m)$ -invariance of version (2.8) of problem (2.1) and  $AL(m)$ -invariance of test statistic  $\mathcal{Q}_K(\mathbf{X}, \hat{\eta})$ . The latter will be tackled in Sections 3 and 4. The former requires that, after transforming into  $\gamma(X)$ , the new null  $\boldsymbol{\theta}(\gamma^*(\eta)) = \mathbf{0}$  holds if and only if  $\boldsymbol{\theta}(\eta) = \mathbf{0}$ . Because of the properties of the  $\Pi_k$  in (2.5), a condition for this is that all  $\pi_{k,j,\ell}(\cdot; \eta)$ ,  $(j, \ell) \in B_k$  must appear in (2.7) with no  $\theta_{k,j,\ell}(\eta)$  structurally set to zero (i.e. no term  $\pi_{k,j,\ell}(\cdot; \eta)$  systematically excluded). As will be seen in Remark 5, some components of  $\bar{\boldsymbol{\pi}}(\hat{\eta})$  may systematically vanish, i.e.  $\pi_{k,j,\ell}(\mathbf{X}_i; \hat{\eta}) \equiv 0$ . Then all other terms in  $\{\pi_{k,j,\ell}(\cdot; \eta), (j, \ell) \in B_k\}$  must be dropped from (2.9) to preserve problem invariance. As will be seen in Section 4, the same condition, coupled with our choice for the  $\pi_{k,j,\ell}(\cdot; \eta)$  also ensures  $AL(m)$ -invariance of  $\mathcal{Q}_K(\mathbf{X}, \hat{\eta})$ .

REMARK 1. *Proving the unicity of the mle  $\hat{\eta}$  can be difficult. Some conditions are given in [26]. The verification of regularity conditions ensuring Assumption B (e.g. [13]) is also tedious. Some work has been done in [16] for particular cases of the  $\alpha$ -th power exponential distribution. Other estimators could in principle be used but the matrix in (2.10) becomes more complicated. Theorem 2.1.a) in [27] gives its general expression without exploiting the particular structure of elliptical distributions.*

REMARK 2. *In view of (2.10), the power function for a fixed  $\psi(\cdot)$  can,*



under mild assumptions (see [20]), be approximated by

$$(2.11) \quad \mathbb{P}_\psi[\mathcal{Q}_K > c] = P \left[ \sum_{k=1}^K \sum_{(j,\ell) \in B_k} \lambda_{k,j,\ell} \chi_1^2(n\nu_{k,j,\ell}^2) > c \right] + O(n^{-1/2}),$$

where  $\nu_{k,j,\ell}$  are functions of  $\int \pi_{k,j,\ell}(\mathbf{x}; \eta) \psi(\mathbf{x}) d\mathbf{x}$  and  $\lambda_{k,j,\ell}$  are the eigenvalues of a complicated matrix (see [11, Sec. 3.2] for details in the case where  $\mathbf{J}_\eta = \mathbf{0}$ , as in Section 5.1). When  $\psi(\cdot)$  is of the form (2.10),  $\nu_{k,j,\ell} \simeq O(\theta_{k,j,\ell}(\eta))$ . If one component of  $\boldsymbol{\theta}(\eta)$  in (2.8) differs from 0, the test will be consistent. This offers a handle on balancing the directional/omnibus slider. On the one hand, power is lost when, for  $k > K$ , one  $\theta_{k,j,\ell}(\eta)$  is large. On the other hand, because  $\theta_{k,j,\ell}(\eta) \rightarrow 0$  with  $k$ , a large  $K$  could add terms close to a  $\chi_1^2(0)$  in (2.11), leading to power dilution. To properly set a balance, some knowledge about  $\psi(\cdot)$  is needed; otherwise see Remark 6 for choosing  $K$  in a data-driven fashion. But these remarks pertain to the  $\theta_{k,j,\ell}(\eta)$ : a compounding difficulty is that some  $\lambda_{k,j,\ell}$  may also be small, with the corresponding  $\lambda_{k,j,\ell} \chi_1^2(n\nu_{k,j,\ell}^2)$  stochastically different from a  $\chi_1^2(0)$ , which again affects power, in some cases rendering the test biased for small samples; see the  $\alpha$ -th power exponential case in Section 6.

REMARK 3. The above test can in principle be extended to the case where the density generator  $\phi_m(\cdot)$  in (1.1) involves a shape parameter, such as with the  $\alpha$ -th power exponential distribution. We do not pursue this further here as the necessary developments are beyond the scope of the paper.

REMARK 4. Convergence toward the asymptotic  $\chi^2$  distribution can be slow. A benefit of invariance is that better approximations can be obtained by Monte Carlo resampling from any convenient  $f(\cdot; \eta)$  distribution in  $\mathcal{F}_0$ .

### 3. The smooth test statistic.

3.1. *Construction of the basis.* Suppose that  $\mathbf{X}$  has a density of the form (1.1). The representation in [7] states that the random vector  $\mathbf{Y} = \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$  has a spherical distribution (i.e. density (1.1) with  $\eta = (\mathbf{0}, \mathbf{I}_m)$ ) with stochastic representation

$$(3.1) \quad \mathbf{Y} = R\mathbf{U},$$

where  $R = R(\mathbf{X}) = \|\mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|$  is referred to as the radius, independent of  $\mathbf{U} = \mathbf{U}(\mathbf{X}) = \mathbf{Y}/\|\mathbf{Y}\| \sim U(\Omega_m)$ , the uniform distribution on the unit

sphere  $\Omega_m$ . Here  $\|\cdot\|$  is the Euclidean norm. The density of  $\mathbf{Y}$  can thus be parametrized in polar coordinates as

$$(3.2) \quad f_{R,\mathbf{U}}(r, \mathbf{u}) = c_m \phi_m(r^2) r^{m-1} dr d\omega_m(\mathbf{u}),$$

where  $d\omega_m(\mathbf{u})$  denotes the area element on  $\Omega_m$  with  $\omega_m(\Omega_m) = 2\pi^{m/2}/\Gamma(m/2)$ . This mapping from  $\mathbf{X} \in \mathbb{R}^m$  to  $(R, \mathbf{U}) \in \mathbb{R}^+ \times \Omega_m$  depends on  $\eta$  but  $f_{R,\mathbf{U}}(r, \mathbf{u})$  is invariant to the choice of  $\mathbf{V}^{-1/2}$ .

A CONB for  $\mathbb{H}_{(f,\eta)}$  will be obtained by combining elements of CONBs associated with the distributions of  $R$  and  $\mathbf{U}$ , so that  $\pi_{k,j,\ell}(\mathbf{X}; \eta)$  can be written as  $\pi_{k,j,\ell}(R(\mathbf{X}), \mathbf{U}(\mathbf{X}))$ .

First consider  $\mathbf{U}$ . Let  $\mathcal{O}(m)$  be the group of rotations on  $\mathbb{R}^m$ . From [17, p. 17], the space  $E_m(k)$  of hyperspherical harmonics of degree  $k$  in  $m$  dimensions is an irreducible  $\mathcal{O}(m)$ -invariant subspace of the space of homogeneous polynomials of degree  $k$  on  $\Omega_m$  and, from [1, Thm. 5.12, p. 81], the space of square integrable functions on  $\Omega_m$  can be decomposed as :

$$L^2(\Omega_m) = \bigoplus_{k=0}^{\infty} E_m(k).$$

Again from [1, Prop. 5.8, p. 78],

$$(3.3) \quad \dim(E_m(k)) = e_m(k) = \begin{cases} 1 & \text{if } k = 0 \\ m & \text{if } k = 1 \\ C_{m-1}^{m+k-1} - C_{m-1}^{m+k-3} & \text{if } k \geq 2 \end{cases}$$

Let  $\{\Psi_{k,\ell}(\mathbf{u}), \ell = 1, \dots, e_m(k)\}$  be a CONB for  $E_m(k)$  with respect to the scalar product  $\langle h_1, h_2 \rangle_{d\omega_m} = (\omega_m(\Omega_m))^{-1} \int_{\Omega_m} h_1(\mathbf{u}) h_2(\mathbf{u}) d\omega_m(\mathbf{u})$ . When  $m = 2$ , writing  $\mathbf{u}^T = (\mu_1, \mu_2) = (\cos \theta, \sin \theta)$  we have  $\Psi_{k,1}(\mu_1, \mu_2) = \sqrt{2} \cos(k\theta)$ ,  $\Psi_{k,2}(\mu_1, \mu_2) = \sqrt{2} \sin(k\theta)$ . When  $m = 3$ ,  $\Psi_{k,\ell}(\mu_1, \mu_2, \mu_3) = P_{k-j}^j(\mu_1) \Psi_{k,j'}(\mu_2, \mu_3)$  where  $P_{k-j}^j(\cdot)$  is the associated Legendre function of the first kind of order  $j$  and degree  $k-j$ , for  $j = 0, \dots, k$ ,  $j' = 1, \min\{j+1, 2\}$ . This is a particular case of a recurrence formula, explained in [6], that generates the hyperspherical harmonics in dimension  $m$  from those in lower dimensions. Such calculations require symbolic manipulations. In particular, [1, Appendix B] give details about the MATHEMATICA program HFT10.m that computes  $\{\Psi_{k,\ell}(\mathbf{u}), \ell = 1, \dots, e_m(k)\}$  for any  $m$ . To facilitate the application of the methods of the paper, the instructions to generate  $\Psi_{k,\ell}(\mathbf{u})$  with this program are detailed in our Appendix B.

Next, we look at the radius  $R$ . For each  $i \geq 0$ , consider the set of functions  $\{s_{j,i}(r), j \geq 0\}$  where  $s_{j,i}(\cdot)$  is a polynomial of degree  $j$  in  $r^2$  satisfying

$$(3.4) \quad \omega_m(\Omega_m)c_m \int_0^\infty s_{j,i}(r)s_{j',i}(r)\phi_m(r^2)r^{m-1+2i}dr = \delta_{j,j'}.$$

To compute these polynomials, we apply the method described in Section 3 in [6], but using the scalar product (3.4). For  $i \geq 0$ , let  $\mu_{j,i} = \mathbb{E}_0(R^{2(j+i)})$ . Write  $\boldsymbol{\mu}_{j,i} = (\mu_{j,i}, \mu_{j+1,i}, \dots, \mu_{2j-1,i})^T$  and

$$\mathbf{M}_{j,i} = \begin{pmatrix} \mu_{0,i} & \cdots & \mu_{j-1,i} \\ \vdots & \ddots & \vdots \\ \mu_{j-1,i} & \cdots & \mu_{2j-2,i} \end{pmatrix}.$$

Set these quantities to 0 when  $j = 0$ . Using the argument leading to (4.4) of [6], we get

$$(3.5) \quad s_{j,i}(r) = \frac{r^{2j} - (1, r^2, \dots, r^{2j-2})\mathbf{M}_{j,i}^{-1}\boldsymbol{\mu}_{j,i}}{\sqrt{\mu_{2j,i} - \boldsymbol{\mu}_{j,i}^T\mathbf{M}_{j,i}^{-1}\boldsymbol{\mu}_{j,i}}}.$$

The MATHEMATICA commands for these functions in the important case of the MVN distribution are in Appendix B. The following theorem, whose proof is given in Appendix A.1, explains how to construct a CONB (i.e. the  $\pi_{k,j,\ell}(\cdot; \cdot)$  of the previous section) associated with an  $f(\cdot; \eta)$  of the form (1.1).

**THEOREM 3.1.** *Fix  $\eta \in \Xi$ . Let  $\mathcal{P}_k$  be the space of polynomials of degree  $k$  in  $\mathbf{x} \in \mathbb{R}^m$ . Let  $\Pi_0 = \mathcal{P}_0$  and for  $k \geq 1$ , let  $\Pi_k$  be the subspace of  $\mathcal{P}_k$  such that  $p(\cdot) \in \Pi_k$  if*

$$(3.6) \quad \int_{\mathbb{R}^m} p(\mathbf{x})p_{k'}(\gamma(\mathbf{x}))f(\mathbf{x}; \eta)d\mathbf{x} = 0$$

for all  $p_{k'}(\cdot) \in \Pi_{k'}$ ,  $0 \leq k' < k$  and  $\gamma \in AL(m)$ . Then,

i) *The subspaces  $\Pi_k$  are orthogonal with respect to  $\langle \cdot, \cdot \rangle_{(f,\eta)}$ . Moreover,*

$$\mathcal{P}_k = \Pi_0 \oplus \Pi_1 \oplus \cdots \oplus \Pi_k$$

and

$$(3.7) \quad \mathbb{H}_{(f,\eta)} = \bigoplus_{k=0}^{\infty} \Pi_k.$$

ii) The subspace  $\Pi_k$  is an irreducible  $AL(m)$ -invariant subspace of the space  $\mathcal{P}_k$ . The polynomials

$$(3.8) \quad \pi_{k,j,\ell}(R(\mathbf{x}), \mathbf{U}(\mathbf{x})) = R(\mathbf{x})^{k-2j} s_{j,k-2j}(R(\mathbf{x})) \Psi_{k-2j,\ell}(\mathbf{U}(\mathbf{x})),$$

with  $(j, \ell) \in B_k = \{j = 0, \dots, [k/2], \ell = 1, \dots, e_m(k-2j)\}$  form a CONB for  $\Pi_k$  with respect to  $\langle \cdot, \cdot \rangle_{(f,\eta)}$ . Here  $[z]$  denotes the integer part of  $z$ . Moreover  $\Pi_k$  is of dimension  $d_m(k) = \sum_{j=0}^{[k/2]} e_m(k-2j) = C_k^{m+k-1}$ . Finally each  $\Pi_k$  can be further decomposed as

$$(3.9) \quad \Pi_k = \bigoplus_{j=0}^{[k/2]} \Pi_{k,k-2j},$$

where  $\Pi_{k,j}$  is spanned by  $\{\pi_{k,j,\ell}(R(\mathbf{x}), \mathbf{U}(\mathbf{x})), \ell = 1, \dots, e_m(k-2j)\}$  and is an irreducible  $\mathcal{O}(m)$ -invariant subspace of  $\Pi_k$  whose dimension is  $e_m(k-2j)$ .

Beside giving the expression (3.8) for the  $\pi_{k,j,\ell}(\cdot; \eta)$ , this theorem is crucial for at least two other points. First, (3.7) ensures that an embedding family of the form (2.7) with  $K$  properly chosen and with (3.8) as the elements in the basis, can approximate almost any alternatives. Thus the test can be made adaptively omnibus; see Remark 6. Second, as will be seen in Section 4, (3.9) provides conditions ensuring the invariance of Dx tools. Note also that the differentiability of the  $\pi_{k,j,\ell}(\cdot; \eta)$  required to apply [27]'s results, follows from the structure in (3.8).

3.2. *The test statistic.* The smooth test for null hypothesis (2.1) can now be explicited. Estimate  $\eta$  by the mle  $\hat{\eta} = (\hat{\boldsymbol{\mu}}, \hat{\mathbf{V}}^{-1})$  with estimated information matrix  $\mathcal{J}_{\hat{\eta}}$ . Set  $\hat{\mathbf{Y}}_i = \hat{\mathbf{V}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})$  and compute  $\hat{R}_i = \|\hat{\mathbf{Y}}_i\|$ ,  $\hat{\mathbf{U}}_i = \hat{\mathbf{Y}}_i / \|\hat{\mathbf{Y}}_i\|$ . Inject these into the  $\pi_{k,j,\ell}(\hat{R}_i, \hat{\mathbf{U}}_i)$  to compute  $\bar{\pi}_{k,j,\ell} = n^{-1} \sum_{i=1}^n \pi_{k,j,\ell}(\hat{R}_i, \hat{\mathbf{U}}_i)$ . Group these into vector  $\bar{\boldsymbol{\pi}}(\hat{\eta})$  of (2.9) and compute  $\mathbf{J}_{\hat{\eta}}$ . Test statistic (2.10) can be obtained. Null hypothesis (2.1) is rejected at approximate level  $\alpha$  if  $\mathcal{Q}_K$  is greater than the  $1 - \alpha$ -th quantile of the reference distribution, e.g. the Monte Carlo approximation or the  $\chi_{\nu_{\mathcal{Q}}}^2$  distribution with  $\nu_{\mathcal{Q}} = \sum_{k=k_{min}}^K C_k^{m+k-1}$ .

REMARK 5. As stated in Section 2, to ensure invariance care must be taken in selecting the elements of  $\bar{\boldsymbol{\pi}}(\hat{\eta})$ . First, any  $\pi_{k,j,\ell}(\cdot, \cdot)$  that is a linear combination of some of the  $\partial \log f(\mathbf{X}; \eta) / \partial \eta_a$  must be excluded because they will contribute nothing to the test statistic. Next, irreducibility of  $\Pi_k$  dictates that to retain invariance, when  $\pi_{k,j,\ell}(\cdot, \cdot)$  has been excluded, then all other  $\{\pi_{k,j',\ell'}(\cdot, \cdot) \mid (j', \ell') \in B_k\}$  must also be excluded.

REMARK 6. *One difficulty in applying any smooth test resides in selecting the value of  $K$ . For univariate data, [31]’s data-driven smooth test first chooses an integer  $d(n)$ , performs a selection procedure to get a “good”  $\hat{K} \in [1, \dots, d(n)]$ , and finally computes the associated test statistic. The framework where  $d(n) \rightarrow \infty$  leads to an adaptively omnibus test and is referred to as the infinite horizon case [48, p. 102]. The available theory about the rate of divergence of  $d(n)$  is impressive but hardly translates into a precise value, as it is expressed in terms of  $o(\cdot)$ . Thus  $d(n)$  is in practice fixed by external considerations, the finite horizon framework. Fortunately, simulations show that the power of the data-driven smooth test stabilizes rapidly as  $d(n)$  increases. Hence in practice both frameworks lead to the same modus operandi as long as  $d(n)$  is not too small.*

*The methodology that derives from Theorem 3.1 allows to compute the smooth test up to any desired  $d(n)$ , so it is in principle possible to develop both infinite and finite horizon data-driven tests in our context. Here we only sketch a simple invariant adaptation of [31]’s approach and focus on the finite horizon case. Suppose  $d(n)$  is given. Define*

$$(3.10) \quad \hat{K} = \min\left\{k : 1 \leq k \leq d(n), \mathcal{Q}_k - \left[ \sum_{j=1}^k \text{Card}(B_j) \right] \log(n) \geq \mathcal{Q}_\ell - \left[ \sum_{j=1}^\ell \text{Card}(B_j) \right] \log(n), \ell = 1, \dots, d(n)\right\}.$$

*The asymptotic reference distribution of test statistic  $\mathcal{Q}_{\hat{K}}$  is a  $\chi^2$  with  $C_1^m$  degrees of freedom. As  $d(n)$  is allowed to increase,  $\mathcal{Q}_{\hat{K}}$  spreads its power in a data-driven fashion over an increasing number of directions. The exploration of this and other scenarios is left for future work ; see Section 7.*

**4. Invariance and the  $\mathcal{Q} = \mathcal{UIR}$  decomposition.** We now take a closer look at matrix  $(\mathbf{I}_{v_{\mathcal{Q}}} - \mathbf{J}_\eta \mathcal{J}_\eta^{-1} \mathbf{J}_\eta^T)^{-1}$  in (2.10). It is explained in Section A.2 that this is a matrix of constants and that after permuting the  $\pi_{k,j,\ell}(\cdot, \cdot)$  in (2.9) according to the values of  $k - 2j$ , test statistic (2.10) can be conveniently written as (A.17), which leads to

$$(4.1) \quad \mathcal{Q}_K = \mathcal{Q}_K(\mathbf{X}, \hat{\eta}) = \mathcal{U}_K + \mathcal{I}_K + \mathcal{R}_K,$$

which we refer to as the  $\mathcal{Q} = \mathcal{UIR}$  decomposition, where

$$\begin{aligned}
 \mathcal{U}_K &= n \|\bar{\boldsymbol{\pi}}_{\mathcal{U}}\|^2, \\
 \mathcal{I}_K &= n \left\{ \|\bar{\boldsymbol{\pi}}_{\mathcal{I},1}\|^2 + d_1 \text{tr}(\mathbf{c}_1 \mathbf{c}_1^T \bar{\mathcal{J}}_1 \bar{\mathcal{J}}_1^T) + \|\bar{\boldsymbol{\pi}}_{\mathcal{I},2}\|^2 + d_2 \text{tr}(\mathbf{c}_2 \mathbf{c}_2^T \bar{\mathcal{J}}_2 \bar{\mathcal{J}}_2^T) + \|\bar{\boldsymbol{\pi}}_{\mathcal{I},3}\|^2 \right\}, \\
 \mathcal{R}_K &= n \left\{ \|\bar{\boldsymbol{\pi}}_{\mathcal{R}}\|^2 + d_0 \text{tr}(\mathbf{c}_0 \mathbf{c}_0^T \bar{\boldsymbol{\pi}}_{\mathcal{R}} \bar{\boldsymbol{\pi}}_{\mathcal{R}}^T) \right\},
 \end{aligned}
 \tag{4.2}$$

are asymptotically independent  $\chi^2$  with degrees of freedom  $\nu_{\mathcal{U}}, \nu_{\mathcal{I}}, \nu_{\mathcal{R}}$  given in (A.2) under  $H_0$ . Note that, beyond the moments of  $R^2$  required for the  $s_{j,i}(\cdot)$ , the computation of this  $\mathcal{Q} = \mathcal{UIR}$  decomposition conveniently requires only the  $2 + 2[K/2] + [(K + 1)/2]$  quantities described in Section A.2 to get the vectors  $\sqrt{d_0} \mathbf{c}_0, \sqrt{d_1} \mathbf{c}_1$  and  $\sqrt{d_2} \mathbf{c}_2$ .

The elements  $\pi_{k,0,\ell}(r, \mathbf{u})$  in  $\mathcal{U}_K$  are proportional to  $r^k \Psi_{k,\ell}(\mathbf{u})$ . The components of  $\bar{\boldsymbol{\pi}}_{\mathcal{U}}$  are thus weighted averages of polynomials in  $\hat{\mathbf{U}}_i$ . Note that up to a constant, the  $\pi_{k,0,\ell}(r, \mathbf{u})$  are identical for any EC distribution and thus cannot be related to the distribution of  $R^2$ . Hence  $\mathcal{U}_K$  serves to detect departures from the uniformity of  $\mathbf{U}$ . When such departures are detected, the true density of the data may not be constant on ellipses centered at  $\boldsymbol{\mu}$  and a model incorporating this feature should be sought, e.g. a distribution where  $\mathbf{U}$  possesses a more complex density on  $\Omega_m$ .

Similarly, because  $\Psi_{0,1}(\mathbf{u}) \equiv 1$ , the elements of  $\mathcal{R}_K$  are  $\pi_{k,k/2,1}(r, \mathbf{u}) = s_{k,0}(r)$  with  $k$  even. Hence this component, which under  $H_0$  is asymptotically  $\chi_{\nu_{\mathcal{R}}}^2$ , serves to detect departures from the distribution of the “radius”  $R$ . These departures may then be identified, visually or otherwise, and the null model corrected accordingly.  $\mathcal{R}_K$  can also be used when, as in [24], it is desired to have a test whose power is directed toward alternatives that are also elliptic about the unknown  $\boldsymbol{\mu}$ .

Finally, consider the elements in  $\mathcal{I}_K$ . These are products of polynomials in  $R$  with hyperspherical harmonics in  $\mathbf{U}$ . Thus  $\mathcal{I}_K$ , which is approximately  $\chi_{\nu_{\mathcal{I}}}^2$ , detects correlations between  $\mathbf{U}$  and  $R$ . When this occurs, the structure of the true density is complicated and a user could look at more involved densities, e.g. skew-densities of some sort.

Summing up, each component of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition can detect a meaningful, in terms of stochastic representation (3.1), type of departure from the null density.

$AL(m)$ -invariance of  $\mathcal{Q}_K$  and each of its components follows from the following considerations that hold provided all  $\pi_{k,j,\ell}(\cdot, \cdot)$  spanning each  $\Pi_k$ , and thus each  $\Pi_{k,k-2j}$  in (3.9), appear in (2.9). For any  $\gamma = (\mathbf{A}, \mathbf{b}) \in AL(m)$ , write  $\mathbf{X}_i^* = \gamma(\mathbf{X}_i)$ ,  $\hat{\boldsymbol{\eta}}^* = \gamma^*(\hat{\boldsymbol{\eta}}) = (\hat{\boldsymbol{\mu}}_*, \hat{\mathbf{V}}_*^{-1})$  defined in Section 2. Also

define  $\hat{\mathbf{Y}}_i^* = \hat{\mathbf{V}}_*^{-1/2}(\mathbf{X}_i^* - \hat{\boldsymbol{\mu}}_*)$  and similarly for  $\hat{R}_i^*$  and  $\hat{\mathbf{U}}_i^*$ . Observe first that  $\hat{R}_i = \hat{R}_i^*$  so that  $s_{k,0}(\hat{R}_i) = s_{k,0}(\hat{R}_i^*)$  and component  $\mathcal{R}_K$ , whose elements span the  $\Pi_{k,0}$  of Theorem 3.1, is trivially  $AL(m)$ -invariant. For component  $\mathcal{U}_K$ , whose elements span the  $\mathcal{O}(m)$ -invariant  $\Pi_{k,[k/2]}$ , notice that  $\pi_{k,0,\ell}(\hat{R}_i^*, \hat{\mathbf{U}}_i^*) = \pi_{k,0,\ell}(\hat{R}_i, \mathbf{O}^* \hat{\mathbf{U}}_i)$ , where  $\mathbf{O}^* = (\mathbf{A} \hat{\mathbf{V}}^{-1} \mathbf{A}^T)^{-1/2} \mathbf{A} \hat{\mathbf{V}}^{-1/2} \in \mathcal{O}(m)$ . It follows that

$$(\pi_{k,0,\ell}(\hat{R}_i, \mathbf{O}^* \hat{\mathbf{U}}_i), \ell = 1, \dots, e_m([k/2]))^T = \mathbf{O}^{**} (\pi_{k,0,\ell}(\hat{R}_i, \hat{\mathbf{U}}_i), \ell = 1, \dots, e_m([k/2]))^T,$$

for some  $\mathbf{O}^{**} \in \mathcal{O}(e_m([k/2]))$  by standard properties of Wigner  $d$ -matrices. Hence each  $(\bar{\pi}_{k,0,\ell}, \ell = 1, \dots, e_m([k/2]))$ , and thus  $\mathcal{U}_K$ , is invariant to  $AL(m)$ -transformations. A similar argument applies to the parts  $\|\bar{\pi}_{\mathcal{I},j}\|^2, j = 1, 2, 3$  of component  $\mathcal{I}_K$  which span the intermediate  $\mathcal{O}(m)$ -invariant  $\Pi_{k,k-2j}$ . Finally, consider the term  $\text{tr}(\mathbf{c}_a \mathbf{c}_a^T \bar{\mathcal{J}}_a \bar{\mathcal{J}}_a^T), a = 1, 2$ . Because  $\bar{\mathcal{J}}_a$  averaged over all  $(\hat{R}_i^*, \hat{\mathbf{U}}_i^*) = \bar{\mathcal{J}}_a \mathbf{O}_a^{**}$ , it follows that  $\mathcal{I}_K$  is  $AL(m)$ -invariant. Then so is global statistic  $\mathcal{Q}_K$  as  $\mathcal{Q}_K(\gamma(\mathbf{X}), \gamma^*(\hat{\eta})) = \mathcal{Q}_K(\mathbf{X}, \hat{\eta})$ .

To be diagnostic, the components in (4.1) must be further processed and, in particular, [18] shows that they must be scaled. This operation must be done with some care in order for the scaled statistics to retain both  $AL(m)$ -invariance and their meaningful interpretations. One possibility is the following; let  $\hat{\Sigma}_{\mathcal{U}\mathcal{U}}$  be a block diagonal matrix where each block is the empirical covariance matrix of the  $\pi_{k,0,\ell}(\hat{R}_i, \hat{\mathbf{U}}_i) \in \Pi_{k,[k/2]}$  appearing in  $\boldsymbol{\pi}_{\mathcal{U}}$ . Define similarly  $\hat{\Sigma}_{\mathcal{I}\mathcal{I}}, \hat{\Sigma}_{\mathcal{R}\mathcal{R}}$  for  $\boldsymbol{\pi}_{\mathcal{I}}, \boldsymbol{\pi}_{\mathcal{R}}$ . The scaled components  $\mathcal{U}_K^{(s)} = n \bar{\boldsymbol{\pi}}_{\mathcal{U}}^T \hat{\Sigma}_{\mathcal{U}\mathcal{U}}^{-1} \bar{\boldsymbol{\pi}}_{\mathcal{U}}$ ,  $\mathcal{I}_K^{(s)} = n \bar{\boldsymbol{\pi}}_{\mathcal{I}}^T \hat{\Sigma}_{\mathcal{I}\mathcal{I}}^{-1} \bar{\boldsymbol{\pi}}_{\mathcal{I}}$  and  $\mathcal{R}_K^{(s)} = n \bar{\boldsymbol{\pi}}_{\mathcal{R}}^T \hat{\Sigma}_{\mathcal{R}\mathcal{R}}^{-1} \bar{\boldsymbol{\pi}}_{\mathcal{R}}$  are  $AL(m)$ -invariant and diagnostic. See Section 6 for some examples of their usefulness.

REMARK 7. *Statistic  $\mathcal{Q}_1$  is a function of  $\pi_{1,0,\ell}(r, u) \propto ru_\ell$  and is thus basically a distance between  $\bar{\mathbf{X}}$  and the mle  $\hat{\boldsymbol{\mu}}$ . As such, it can be useful in discriminating  $H_0$  but provides little Dx information in any senses associated with the  $\mathcal{Q} = \mathcal{U}\mathcal{I}\mathcal{R}$  decomposition. In the above, this term is bundled into component  $\mathcal{I}_K$ . Similarly  $\mathcal{Q}_2 - \mathcal{Q}_1$ , which is spread over  $\mathcal{I}_K$  and  $\mathcal{R}_K$ , is related to  $\text{tr}(\hat{\mathbf{V}}^{-1} \mathbf{S})$  ( $\mathbf{S}$  is the empirical covariance matrix of the  $\mathbf{X}_i$ ) and again can be useful in discriminating  $H_0$  but otherwise provides little Dx insights. [9] advises that one should use a powerful test for the null hypothesis, followed by less formal procedures when the null is rejected. Thus it can be a reasonable strategy to consider only the elements  $\mathcal{Q}_k - \mathcal{Q}_2$  at the Dx stage.*

**5. Applications of the smooth test methodology to some EC distributions.** The above smooth GoF test has the desirable features listed in Section 1: with  $K$  chosen appropriately, it can be made adaptively omnibus; it possesses Dx capabilities; and it is  $AL(m)$ -invariant. In addition,

its behavior under both  $H_0$  and  $H_1$  can be conveniently approximated. In this section, we develop the test strategy for two popular EC distributions. Two others are treated in Appendix C.

5.1. *The multivariate normal (MVN) distribution.* Consider the important problem of testing the null hypothesis  $H_0$  that a sample  $\mathbf{X}_1, \dots, \mathbf{X}_n$  arises from an  $m$ -dimensional MVN distribution with density

$$f(\mathbf{x}; \eta) = (2\pi)^{-m/2} \det(\mathbf{V}^{-1})^{1/2} \exp \left\{ -(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2 \right\}.$$

Under  $H_0$ ,  $R^2$  has a  $\chi_m^2$  distribution with  $j$ -th moment  $2^j \Gamma(m/2 + j) / \Gamma(m/2)$ . The elements of an orthonormal basis satisfying (3.4) are, via (3.5),

$$(5.1) \quad s_{j,i}(r) = (-1)^j \sqrt{\frac{j! \Gamma(m/2)}{2^i \Gamma(m/2 + j + i)}} L_j^{m/2+i-1}(r^2/2),$$

where  $L_j^\alpha(\cdot)$  is the  $j$ -th generalized Laguerre polynomial of order  $\alpha$ ; see Appendix B for MATHEMATICA commands to generate these quantities along with those for  $\{\Psi_{k,j}(\cdot), j = 1, \dots, e_m(k)\}$ . Tables 3 and 4 list all  $\pi_{k,j,\ell}(r, \mathbf{u})$  for  $k = 3, 4, 5$  and  $m = 2, 3$ .

Next, it is easy to see that for the MVN,  $\zeta(\cdot) = 1$  (defined in Section A.2) and  $\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2$  vanish so that the  $\mathcal{Q} = \mathcal{U}\mathcal{I}\mathcal{R}$  decomposition (4.2) takes a particularly simple form. To get its explicit expression, first compute  $\hat{\mathbf{Y}}_i, \hat{\mathbf{R}}_i$  and  $\hat{\mathbf{U}}_i$  with  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$  and  $\hat{\mathbf{V}}^{-1} = \mathbf{S}^{-1}$ , where  $(\bar{\mathbf{X}}, \mathbf{S})$  is the mle of  $(\boldsymbol{\mu}, \mathbf{V})$ . Then all  $\bar{\pi}_{k,j,\ell} = 0$  when  $k = 1, 2$  and in view of Remark 5, this leads to  $\mathcal{Q}_1 = \mathcal{Q}_2 = 0$ . Upon writing  $\mathcal{C}_{k,j}^2 = n \sum_{\ell=1}^{e_m(k-2j)} (\bar{\pi}_{k,j,\ell})^2$ , (2.10) becomes:

$$(5.2) \quad \mathcal{Q}_K = \sum_{k=3}^K \sum_{j=0}^{\lfloor k/2 \rfloor} \mathcal{C}_{k,j}^2,$$

which is asymptotically  $\chi_{\nu_{\mathcal{Q}}}^2$  under  $H_0$ , where  $\nu_{\mathcal{Q}} = \sum_{k=3}^K C_k^{m+k-1}$ . Also

$$(5.3) \quad \mathcal{R}_K = \sum_{\substack{k=3 \\ k:\text{even}}}^K \mathcal{C}_{k,k/2}^2$$

is, under  $H_0$ , asymptotically  $\chi_{\nu_{\mathcal{R}}}^2$  where  $\nu_{\mathcal{R}} = \lfloor (K - 2)/2 \rfloor$ . Moreover,

$$(5.4) \quad \mathcal{U}_K = \sum_{k=3}^K \mathcal{C}_{k,0}^2$$



has for reference distribution a  $\chi_{\nu_{\mathcal{U}}}^2$ , where  $\nu_{\mathcal{U}} = \sum_{k=3}^K e_m(k)$ . Finally,

$$(5.5) \quad \mathcal{I}_K = \sum_{k=3}^K \sum_{j=1}^{\lfloor (k-1)/2 \rfloor} \mathcal{C}_{k,j}^2.$$

is approximately  $\chi_{\nu_{\mathcal{I}}}^2$  with  $\nu_{\mathcal{I}} = \nu_{\mathcal{Q}} - \nu_{\mathcal{U}} - \nu_{\mathcal{R}}$ .

To scale the components in the  $\mathcal{Q} = \mathcal{U}\mathcal{I}\mathcal{R}$  decomposition, let  $\hat{\Sigma}_{k,j}$  denote the empirical covariance matrix of the  $n$  random vectors  $(\pi_{k,j,\ell}(\hat{R}_i, \hat{U}_i), \ell = 1, \dots, e_m(k-2j))$  with empirical mean  $\bar{\pi}_{k,j} = (\bar{\pi}_{k,j,\ell}, \ell = 1, \dots, e_m(k-2j))$ . The scaled components are  $(\mathcal{C}_{k,j}^{(s)})^2 = n\bar{\pi}_{k,j}^T \hat{\Sigma}_{k,j}^{-1} \bar{\pi}_{k,j}$ . These  $(\mathcal{C}_{k,j}^{(s)})^2$ , bundled into  $\mathcal{U}^{(s)}, \mathcal{I}^{(s)}$  or  $\mathcal{R}^{(s)}$  are diagnostic to identify the aspects of the MVN not supported by the data. The R package ECGoFTTestDx computing the above smooth test for the MVN has been deposited on CRAN. An application to a data set is shown in Appendix D.

When  $m = 1$ ,  $\Pi_k$  is spanned by  $\pi_{k,\lfloor k/2 \rfloor, 1}(r, u)$ , with  $u = \pm 1$ . These are the Hermite polynomials of [43]'s smooth test of univariate normality. Thus our approach generalizes their test.

For univariate distributions, the components of smooth test statistics are often related to GoF tests that have been introduced from other principles. The same occurs here and some components of  $\mathcal{Q}_K$  turn out to be well-known GoF test statistics for the MVN. In particular,

$$(5.6) \quad \begin{aligned} \mathcal{Q}_4 &= \mathcal{C}_{3,0}^2 + \mathcal{C}_{1,1}^2 \quad (= \mathcal{Q}_3) \\ &+ \mathcal{C}_{4,0}^2 + \mathcal{C}_{4,1}^2 + n(\bar{\pi}_{4,2,1})^2. \end{aligned}$$

Inspection of these terms shows that  $\mathcal{R}_4 = n(\bar{\pi}_{4,2,1})^2 = nb_{1,m}/6$  where  $b_{1,m}$  is [33]'s multivariate measure of skewness. Moreover,  $n(\bar{\pi}_{4,2,1})^2 = n(b_{2,m} - m(m+2))^2 / (8m(m+2))$ , where  $b_{2,m}$  is [33]'s multivariate measure of kurtosis. These are popular measures in multivariate analysis, whose usefulness in providing Dx information arises by extending the meaning of their univariate counterparts: in particular,  $b_{1,m}$  vanishes under elliptical symmetry and  $b_{2,m}$  quantifies the tails of the distribution. Consequently, they have been used as GoF tests for the MVN and, in particular, [46] have concluded that tests based on them rank among the bests in terms of power. [25] have combined them (i.e.  $\mathcal{Q}_3 + n(\bar{\pi}_{4,2,1})^2$ ) to get a multivariate version of the popular Jarque-Bera test of univariate normality.

[35] have developed a test of multivariate normality as an example of some simplifications that occur with Rao's score method when a suitable group structure exists. By brute force calculations, they were able to obtain  $\mathcal{Q}_4$  and state to have been unable to generalize to  $K > 4$ . Here, by making

use of recent advances in computational tools for harmonic analysis, we can easily go beyond this limitation. Moreover, as a by-product, we get for any  $K$  the informative  $\mathcal{Q}=\mathcal{UITR}$  decomposition.

The idea of using the elements of stochastic representation (3.1) for testing MVN is not new. Many authors have proposed testing the distribution of  $R$  and/or  $\mathbf{U}$  individually (see [28, 19] for reviews), thus in effect crafting in an *ad hoc* fashion fragments of the  $\mathcal{Q}=\mathcal{UITR}$  decomposition. [29], followed by [42] and [3], have taken a different route and proposed a smooth test based on the fact that the multivariate normal density is the product of  $m$  univariate normal densities when  $\mathbf{V} = \mathbf{I}_m$ . For this particular case, a CONB can be obtained by the tensor product of elements of CONB associated with each univariate distribution [30, p. 51, Theorem 4.3]. However, this approach breaks down for general elliptical distributions, as in the following application, because  $\mathbf{V} = \mathbf{I}$  is not associated with independence [38] and in view of the inconsistency property of many EC distributions.

5.2. *The bivariate Laplace distribution.* There are several bivariate extensions of the univariate Laplace ; we take here the variant discussed in [39] obtained from the power exponential distribution in Section 1 by setting  $\alpha = 1/2$ . This distribution has larger tails than the MVN. The  $j$ -th moment of  $R^2$  is  $\Gamma(2(j+1))$ . We take  $K=5$ , a reasonable value, and from (3.5), the  $s_{j,k-2j}(\cdot)$  required for  $\mathcal{Q}_5$  are :  $s_{0,1}(r) = (\sqrt{6})^{-1}$ ;  $s_{0,2}(r) = (2\sqrt{30})^{-1}$ ;  $s_{0,3}(r) = (12\sqrt{35})^{-1}$ ;  $s_{0,4}(r) = (72\sqrt{70})^{-1}$ ;  $s_{0,5}(r) = (720\sqrt{77})^{-1}$ ;  $s_{1,0}(r) = (r^2 - 6)/(2\sqrt{21})$ ;  $s_{1,1}(r) = (r^2 - 20)/(4\sqrt{165})$ ;  $s_{1,2}(r) = (r^2 - 42)/(60\sqrt{42})$ ;  $s_{1,3}(r) = (r^2 - 72)/(144\sqrt{665})$ ;  $s_{2,0}(r) = (7r^4 - 360r^2 + 1320)/(24\sqrt{10745})$  and  $s_{2,1}(r) = (11r^4 - 10926r^2 + 12600)/(72\sqrt{225610})$ . Along with the expression for  $\Psi_{k,\ell}(\cdot)$  in Section 3.1 when  $m = 2$ , it is an easy exercise to compute the  $\pi_{k,j,\ell}(\cdot, \cdot)$ . Next, because  $\zeta(r^2) = r^2 \times \sqrt{\frac{1}{r^2}}$ , one finds  $\sigma_1 = 1$ ,  $\sigma_2 = 6$  and

$$\begin{aligned} \sqrt{d_0}\mathbf{c}_0 &= (7)^{-1}\{-\sqrt{614}, 4\sqrt{\frac{10}{3}}\}, \\ \sqrt{d_1}\mathbf{c}_1 &= (\sqrt{1203})^{-1}\{\sqrt{4102}, -4\sqrt{\frac{2051}{55}}, 48\sqrt{\frac{6}{55}}\}, \\ \sqrt{d_2}\mathbf{c}_2 &= (\sqrt{19})^{-1}\{-2\sqrt{35}, 4\}. \end{aligned}$$

After computing the mle  $\hat{\eta} = (\hat{\boldsymbol{\mu}}, \hat{\mathbf{V}}^{-1})$  and in turn,  $\hat{\mathbf{Y}}_i$ ,  $\hat{R}_i$  and  $\hat{\mathbf{U}}_i$ , plug these into the  $\pi_{k,j,\ell}(\hat{R}_i, \hat{\mathbf{U}}_i)$  to get the  $\bar{\pi}_{k,j,\ell}$ . Setting  $\bar{\pi}_{\mathcal{U}} = (\bar{\pi}_{3,0,1}, \bar{\pi}_{3,0,2}, \bar{\pi}_{4,0,1}, \bar{\pi}_{4,0,2}, \bar{\pi}_{5,0,1}, \bar{\pi}_{5,0,2})$ ,  $\bar{\pi}_{\mathcal{R}} = (\bar{\pi}_{2,1,1}, \bar{\pi}_{4,2,1})$ ,  $\bar{\pi}_{\mathcal{I},1} = (\bar{\pi}_{1,0,1}, \bar{\pi}_{3,1,1}, \bar{\pi}_{5,1,1}, \bar{\pi}_{5,2,1}, \bar{\pi}_{1,0,2}, \bar{\pi}_{3,1,2})$ ,  $\bar{\pi}_{\mathcal{I},2} = (\bar{\pi}_{2,0,1}, \bar{\pi}_{4,1,1}, \bar{\pi}_{2,0,2}, \bar{\pi}_{4,1,2})$  and  $\bar{\pi}_{\mathcal{I},3} = (\bar{\pi}_{5,1,2}, \bar{\pi}_{5,2,2})$  yields all the elements required to compute the various test components. Their degrees of freedom are those in Section 4. For a GoF test for another bivariate Laplace distribution, see [14].

**6. Performance of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition in deriving Dx information.** There can be many directions in which an alternative may depart from a null model. The present Dx tool, based on the  $\mathcal{Q} = \mathcal{UIR}$  decomposition, pertains to departures from the easily interpretable representation (3.1). To evaluate its performance, we need to determine how well it detects those departures, ideally in a context where other departures coexist to complicate matters. Performance here refers to the confidence allocated to the Dx information extracted from a sample, i.e. sensibility and specificity. As our tests are performed at a given level (here 5%), specificity is fixed and we concentrate on sensitivity, the probability that a departure from representation (3.1) will be detected, which is the power of the tests based on the components of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition.

To start building this confidence, a small experiment was performed. We adopt the context of Section 5.1 where the interest lies in assessing the null hypothesis of bivariate normality. Because representation (3.1) is common to all EC distributions, the results below could be representative of what could be obtained with other EC null models and higher dimensions. Samples were generated from thirty distributions taken from [37, 5, 22]. Statistics  $\mathcal{Q}_K, \mathcal{U}_K^{(s)}, \mathcal{I}_K^{(s)}$  and  $\mathcal{R}_K^{(s)}$ , ( $K = 3, \dots, 12$ ) were computed and compared to their null Monte Carlo (based on 20,000 replications) approximation, as explained in Remark 4. To yield interesting power values, this was replicated 5,000 times and the sample size  $n$  was adapted to each alternative.

We stress that our goal is to appreciate the usefulness of the Dx procedure based of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition and how these components relate to the global  $\mathcal{Q}_K$ . It is *not* to thoroughly compare the power of  $\mathcal{Q}_K$  to its competitors, as we have made no attempt to optimize the choice of  $K$ . Nevertheless, to offer some perspectives we did compute the power of the BHEP test [2] with the tuning parameter set at 1.41; this test has been recommended in some simulation studies [37] but offers no Dx information. We also computed the power of the multivariate Jarque & Bera test [25] that does offer some Dx information via its skewness-kurtosis components.

The alternatives were chosen to somewhat resemble the bivariate normal and to depart along one or two directions identifiable by the  $\mathcal{Q} = \mathcal{UIR}$  decomposition, in addition to other departures. To screen the alternatives along these criteria, we generated 100 000 observations  $(X_1, X_2)$  from each, transformed them into  $(Y_1, Y_2)$  and then into  $(R, \mathbf{U})$ , and then again  $\mathbf{U}$  into  $\theta = \arctan(U_2/U_1)$ . Plots of the various univariate and bivariate densities were made, and various statistics and characteristics were computed. We then excluded the densities where all  $\mathcal{UIR}$  components are large, because the resulting knowledge (i.e. all is wrong) is not easy to use in the iterative

process of correcting a null model after Dx analysis.

From the results from these thirty alternatives, we report here on five that cover most of the behavior we have observed. Also, we report here on the case  $K = 5$ .

The first alternative is a member of the Khintchine family of distributions [22, Chap. 8]. Synthetic data from this distribution are obtained by generating  $Z \sim \Gamma(1.5, 1)$  and then setting

$$(X_1, X_2) = \sqrt{\frac{3\Gamma(1.5)}{\Gamma(1.5 + 2 \times 0.3998935)}} Z^{0.398935} \times 2(V_1 - 0.5, V_2 - 0.5),$$

where  $V_1, V_2$  are independent  $U(0, 1)$ . In the above expression, the various constants are such that the marginals of  $(Y_1, Y_2)$  are nearly independent  $N(0, 1)$  and the marginal density of  $R$  is approximately a  $\sqrt{\chi_2^2}$ . However  $\mathbf{U}$  is non-uniform, as the density of  $\theta$  oscillates as  $\sin(4\theta)$  with modes at  $\pm\pi/4, \pm3\pi/4$ . The density of  $(Y_1, Y_2)$  has squarish contours but otherwise resembles the bivariate normal. These oscillations induce a complicated relationship between  $R$  and  $\mathbf{U}$  which, at the first order, can be assimilated to near independence, the regression of  $R$  on  $\theta$  involving  $\sin(k\theta), \cos(k\theta)$ ,  $k = 1, \dots, 3$  being nearly constant with a Spearman's correlation coefficient ( $\rho_S = -0.002$ )<sup>1</sup>. Because of these, we expect :  $\text{power}(\mathcal{R}_5^{(s)}) \approx \text{power}(\mathcal{I}_5^{(s)}) \approx 5\% < \text{power}(\mathcal{U}_5^{(s)})$ . Note that this, and further, expectation derives from the 100 000 samples generated when screening the alternatives. This was blinded in the following simulation study. The sample size is  $n = 400$ .

The second distribution is the generalized Burr-Pareto-Logistic (with  $\alpha = 1, \beta = 0$ ) in [22, Chap. 9]; see his Figure 9.8 for a sketch of the density and p. 167 for an algorithm to generate synthetic data. The distribution of  $(X_1, X_2)$  has dependent  $N(0, 1)$  marginals with contours of triangular shape. Here the distribution of  $R$  is almost a  $\sqrt{\chi_2^2}$ ,  $\mathbf{U}$  is clearly non-uniform with a trimodal distribution, while  $R$  and  $\mathbf{U}$  are slightly correlated (with  $\rho_S \approx 0.04$ ). Hence we should find :  $\text{power}(\mathcal{R}_5^{(s)}) \approx 5\% < \text{power}(\mathcal{I}_5^{(s)}) \leq \text{power}(\mathcal{U}_5^{(s)})$ . The sample size is  $n = 250$ .

The third distribution is the contaminated binormal :  $0.8 \times MVN_2(0, \mathbf{I}_2) + 0.2 \times MVN_2((1, 1), \text{Diag}\{1, 2\})$ , see [22, Chap. 4].  $R$  is again very close to

---

<sup>1</sup>Pearson's correlation coefficient is not a good measure of the dependency between these quantities A normalized version of mutual information seems better adapted but we have found its scale difficult to appreciate from one problem to another. We have settled on Spearman's correlation coefficient mainly because it remains the same between  $\theta$  and  $R$  or  $R^2$  and because a sophisticated user can in general make some intuitive sense of differences between its values

a  $\sqrt{\chi_2^2}$ ,  $\mathbf{U}$  has a distribution that slightly differs from uniformity while  $R$  and  $\mathbf{U}$  are slightly correlated ( $\rho_S = -0.03$ ). Again  $n = 400$  and we expect to find :  $\text{power}(\mathcal{R}_5^{(s)}) \approx 5\% \leq \text{power}(\mathcal{U}_5^{(s)}), \text{power}(\mathcal{I}_5^{(s)})$ , with little insight about the comparative power of  $\mathcal{U}_5^{(s)}, \mathcal{I}_5^{(s)}$ .

The fourth density is a Laplace-type (because of its marginals) distribution generated by the following scheme : take  $W_0, W_1, W_2 \sim \text{Exp}(1)$  and form  $(X_1, X_2) = (W_1 - W_0, W_2 - W_0)$ . Here,  $\mathbf{U}$  is non-uniform,  $R$  departs from the  $\sqrt{\chi_2^2}$  but  $R$  and  $\theta$  are approximately independent ( $\rho_S = -0.004$ ), so one can expect:  $\text{power}(\mathcal{I}_5^{(s)}) \approx 5\% \leq \text{power}(\mathcal{U}_5^{(s)}), \text{power}(\mathcal{R}_5^{(s)})$ . We have taken  $n = 75$  to mitigate the high power of  $\mathcal{Q}_5$  with that of its components.

The last distribution is a 2-th power exponential distribution generated by taking  $R \sim \Gamma(2, 2)$ ,  $\theta \sim U(0, 2\pi)$  and forming  $(X_1, X_2) = R \times (\cos \theta, \sin \theta)$ . This yield an EC distribution where  $R$  is markedly different from a  $\sqrt{\chi_2^2}$ . Thus here we should find :  $5\% \approx \text{power}(\mathcal{U}_5^{(s)}) \approx \text{power}(\mathcal{I}_5^{(s)}) \leq \text{power}(\mathcal{R}_5^{(s)})$ . We have taken  $n = 100$ .

Table 1 shows the power (in %) of the various tests. The shaded cells are located where the power of the components of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition are expected to be greater than nominal (5%). A first general conclusion is that no global test dominates. An interesting comparison is between J-B and  $\mathcal{Q}_5$  where, as stated in Section 5.1, the first is a fragment of the second. For some alternatives (e.g. Burr-Pareto-Logistic), the extra components have little power and lead to power dilution of  $\mathcal{Q}_5$  with respect to J-B. For other alternatives (e.g. Khintchine), power of J-B test is much lower because the extra components in  $\mathcal{Q}_5$  detect departures that hardly translate into skewness or kurtosis. A data-driven approach to select the value of  $K$  could be one way to further increase the power of  $\mathcal{Q}_K$ , but see Section 7. One atypical case is the 2-power exponential. This will be discussed below.

Regarding the power (i.e. sensitivity) of the components, the Dx information derived from the  $\mathcal{Q} = \mathcal{UIR}$  decomposition is consonant with the expectations related to each alternative. Furthermore, the magnitudes of the power in the shaded cells somewhat reflect the severeness of the departures (when grossly quantifiable). For the Khintchine, both  $\mathcal{R}_5^{(s)}, \mathcal{I}_5^{(s)}$  have power approximately equal to level and the severe non-uniformity of  $\mathbf{U}$  is detected with good power. For the Burr-Pareto-Logistic, the expectation :  $\text{power}(\mathcal{R}_5^{(s)}) \approx 5\% < \text{power}(\mathcal{I}_5^{(s)}) \leq \text{power}(\mathcal{U}_5^{(s)})$  is realized while the small power of  $\mathcal{I}_5^{(s)}$  (21.8%) reflects the slight correlation between  $R$  and  $\mathbf{U}$ . Again, the severe non-uniformity of  $\mathbf{U}$  is almost always detected. For the contaminated MVN,  $\mathcal{R}_5^{(s)}$  has, as expected, trivial power,  $\mathcal{I}_5^{(s)}$  picks up rather well the dependency between  $R$  and  $\mathbf{U}$  while the slight non-uniformity

Alt. / Test	$n$	BHEP	J-B	$\beta_1 = \mathcal{Q}_3$	$\beta_2$	$\mathcal{Q}_5$	$\mathcal{U}_5^{(s)}$	$\mathcal{I}_5^{(s)}$	$\mathcal{R}_5^{(s)}$
K	400	47.9	6.9	6.8	6.23	63.5	97.6	4.6	5.4
BPL	250	52.2	95.6	96.7	21.5	68.9	94.9	21.8	1.7
CMVN	400	50.6	76.3	77.3	40.8	54.5	19.5	41.9	4.6
Lt	75	96.6	99.0	98.1	95.0	94.9	39.5	7.6	21.9
2PE	100	45.0	2.1	0.0	0.0	0.0	6.5	6.4	92.2

TABLE 1

Power in % (based on 5,000 replications at level  $\alpha = 5\%$ ) of the test of bivariate normality based on  $\mathcal{Q}_5$  and its  $ULR$  components. The reference distribution has been approximated by 20,000 Monte Carlo samples of size  $n$  from the  $MVN(\mathbf{0}, \mathbf{I}_2)$ .

Alternatives are  $K$ =Khintchine,  $BPL$ =Burr-Pareto-Logistic,  $CMVN$ =Contaminated  $MVN$ ,  $Lt$ =Laplace -type and  $2PE$ = $\alpha = 2$ -power exponential. Also shown is the BHEP (Baringhauss-Henze-Epps-Pulley) test and the J-B (multivariate Jarque-Bera) test along with its components  $\beta_1, \beta_2$  (both using Monte Carlo quantiles). The shaded cells correspond to those where we expect power  $> 5\%$ .

of  $\mathbf{U}$  is detected with moderate power. For the Laplace-type case, the departures on the marginal distributions of  $R$  and  $\mathbf{U}$  are correctly detected. For the 2-power exponential,  $\text{power}(\mathcal{R}_5^{(s)}) \approx 91.6$ , showing that the departures along the distribution of  $R$  are correctly captured. The two other components have power close to nominal, as they should.

To summarize, one can have some confidence that our Dx procedure can be a useful tool in the iterative process of modeling a data set.

In some cases, at least one component yield better power than the global test statistic from which it is extracted. Thus the natural two-stage strategy where one first test the global null hypothesis using  $\mathcal{Q}_K$  and, if significant, proceed at the Dx level, would lose valuable information. This can be explained by a number of factors. First, when only one component ought to be significant, as with the Khintchine, the degrees of freedom associated with this component (6 in this case) is much smaller than that of the global test (here 15), thus diluting power. This effect is further confounded by the scaling of the components, which also affects power, and suggests looking at a scaled global statistic. Here  $\text{power}(\mathcal{Q}_5^{(s)} = \mathcal{U}_5^{(s)} + \mathcal{I}_5^{(s)} + \mathcal{R}_5^{(s)}) = 83.5$ , greater than the other global tests, which suggest replacing  $\mathcal{Q}_5$  by its scaled version. However the benefit of this replacement is not universal : for the contaminated normal,  $\text{power}(\mathcal{Q}_5) = 54.5$  while  $\text{power}(\mathcal{Q}_5^{(s)}) = 45.6$ . Thus it is not clear which of the above factors (degrees of freedom vs scaling) has more impact on power. An extreme case of this is the 2-th power exponential where the power of  $\mathcal{Q}_5$  is almost zero and the J-B test is not much better. Here scaling has a large effect on power as  $\text{power}(\mathcal{Q}_5^{(s)}) = 47.5$ . This is an instance where the values of the  $\lambda_{k,j,\ell}$  in Remark 2 renders the global test biased. Scaling partly corrects the problem, as explained in [11]. A pro-

cedure that would select in a data-driven fashion between  $\mathcal{Q}_K$  and  $\mathcal{Q}_K^{(s)}$ , in addition to a proper value of  $K$  globally or at the component level, would perhaps reduce such variations. See Section 7.

The J-B test also offers some Dx information via its  $b_{1,m}, b_{2,m}$  components. But they are not scaled and, in particular,  $\mathcal{R}_5^{(s)}$  is the scaled version of  $b_{2,m}$ . Comparing their powers shows the pitfalls in attempting to interpret unscaled Dx statistics. However, being fragments of our components, they can be rescaled and exploited to further refine the Dx information, a problem we now look into from another angle.

In the above, we have taken  $K = 5$  and shown that our procedure can provide useful Dx information supplementing the global test. Now we go further and vary  $K$  to see if more precise Dx information can be extracted from a set of components. Table 2 presents further results from our experiment, namely the values of  $(\mathcal{U}_k^{(s)}, k = 3, \dots, 7)$  and  $(\mathcal{I}_k^{(s)}, k = 3, \dots, 7)$  in the case of the Khintchine distribution ( $n = 400$ ). Notice the sharp increase in power between  $\mathcal{U}_3^{(s)}$  and  $\mathcal{U}_4^{(s)}$ . This indicates that the hypersphericals in the added components of  $\mathcal{U}_4^{(s)}$  pick up some important departure from uniformity. To see what has been detected, we refer to the definition of projection-based Dx information in [11]. Recall that the elements in  $\mathcal{U}_3^{(s)}$  have the form  $\sin(k\theta), \cos(k\theta), k = 1, 2, 3$  and thus have at most 3 maximums in  $[0, 2\pi)$ . Those added in  $\mathcal{U}_4^{(s)}$  have the form  $\sin(4\theta), \cos(4\theta)$  and the increased power indicates that they capture higher oscillations, suggesting that the density of  $\mathbf{U}$  (or  $\theta$ ) could have at least four maximums. This is consonant with the fact that  $\theta$  has a distribution that oscillates as  $\sin(4\theta)$ . Also, recall that the scatter plot of  $(R, \theta)$  reveals that these quantities have a complicated relationship. A similar increase of power from  $\mathcal{I}_5^{(s)}$  to  $\mathcal{I}_6^{(s)}$  and afterwards indicates again that some correlation has been detected at the order of functions of the form  $\cos(6\theta), \sin(6\theta)$ . Indeed a regression of  $R$  on  $\theta$  with basis functions  $\sin(k\theta), \cos(k\theta), k = 1, \dots, 7$  retains the functions  $\cos(4\theta)$  and  $\sin(6\theta)$  as significant. Hence the results in Table 2 indicate that beyond the basic analysis of the  $\mathcal{Q} = \mathcal{UIR}$  components, a finer analysis of the subcomponents could be exploited to pinpoint with better precision the nature of the departures.

**7. Conclusions.** This paper develops a smooth test of goodness-of-fit for EC distributions that, through the invariant  $\mathcal{Q} = \mathcal{UIR}$  decomposition, allows to extract illuminating Dx information. Some applications are worked out, in particular the important case of the MVN, where the usefulness of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition is demonstrated. For most EC distributions, there exists no competitor to our approach. For the MVN, a small experiment

Component	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$
$\mathcal{U}_K^{(s)}$	5.2	98.7	97.6	96.3	94.5
$\mathcal{I}_K^{(s)}$	4.8	4.0	4.6	56.1	47.1

TABLE 2

Power in % (based on 5000 replications) of some components in the  $UIR$  decomposition for the Khintchine alternative. The level is  $\alpha = 5\%$ . The reference distribution has been approximated by 20 000 Monte Carlo samples of size  $n = 400$  from the  $MVN(\mathbf{0}, \mathbf{I}_2)$ .

suggests that our smooth test can be competitive powerwise.

In the present work, no attempt is made to optimize the selection of the hyperparameter  $K$ . This topic deserves a separate analysis because the present context opens up many new possibilities beyond the obvious extension to select among  $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_K$  in a data-driven fashion via (3.10). One of these is to increase sensitivity of the  $D_x$  by selecting among the components, say  $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_K$ . In addition, it was shown in Table 2 that this could also prove useful in better understanding the nature of the detected departure. Another is to select among  $\{\mathcal{U}_K, \mathcal{U}_K + \mathcal{I}_K, \mathcal{U}_K + \mathcal{I}_K + \mathcal{R}_K = \mathcal{Q}_K\}$  or some other (data-driven) permutation of the components. Yet another is to select between the raw components and their rescaled versions.

In multivariate contexts, the computation of the mle may be difficult or the information  $\mathcal{J}_\eta$  may not exist. This is the case for example in the Laplace discussed by [14], thus forcing the use of moment estimators. Deriving the  $\mathcal{Q} = UIR$  decomposition in such cases remains to be investigated.

Finally, EC distributions often involve (e.g. the  $\alpha$ -th power exponential), in addition to  $\eta$ , a shape parameter that offers more flexibility in adjusting the data. The shape parameter must in general be estimated and this does not, in principle, affect the use of a test statistic similar to (2.9). However, an open question is how does this estimation impact the  $\mathcal{Q} = UIR$  decomposition.

## References.

- [1] AXLER, S., BOURDON, P. and RAMEY, W. (2001). *Harmonic function theory*, 2<sup>nd</sup> edition. Springer-Verlag, New York.
- [2] BARINGHAUS, L. and HENZE, N. (1988) : A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, **35**, 339–348.
- [3] BEST D.J. and RAYNER, J.C.W. (1988). A test for bivariate normality. *Stat. Probab. Lett.*, **6**, 407–12.
- [4] BILODEAU, M. and BRENNER, D. (1999). *Theory of multivariate statistics*. Springer, New York.
- [5] BOGDAN, M. (1999). Data driven smooth tests for bivariate normality. *J. Multivar. Anal.*, **68**, 26–53.
- [6] BOULERICE, B. and DUCHARME, G.R. (1997). Smooth tests of goodness-of-fit for directional and axial data. *J. Multivar. Anal.*, **60**, 154–175.



- [7] CAMBANIS, S., HUANG, S., SIMONS, G. (1981) : On the Theory of Elliptically Contoured Distributions. *J. Multivar. Anal.*, **11**, 368–338.
- [8] CHMIELEWSKI, M.A. (1981). Elliptically Symmetric Distributions: A Review and Bibliography. *Int. Stat. Rev.*, **49**, 67–74.
- [9] CSÖRGÖ, S., (1989) : Consistency of some tests for multivariate normality. *Metrika*, **36**, 107–116.
- [10] DUCHARME, G.R. (2001) : Goodness-of-fit tests for the inverse Gaussian and related distributions. *Test*, **10**, 271–290.
- [11] DUCHARME, G.R., and AL AKHRAS, W. (2016) : Tree based Diagnostic Procedures Following a Smooth Test of goodness-of-Fit. *Metrika*, **79**, 971–989.
- [12] EBNER, B. (2012) : Asymptotic theory for the test for multivariate normality by Cox and Small. *J. Multivar. Anal.* **111**, 368–379.
- [13] FERGUSON, T.S. (1996): *A course in large sample theory*. Chapman and Hall, London.
- [14] FRAGIADAKIS, K. and MEINTANIS, S. G. (2011) : Goodness-of-fit tests for multivariate Laplace distributions. *Math. Comput. Model.*, **53**, 769–779.
- [15] GELMAN, A., MENG, X.L. and STERN, H. (1996) : Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.*, **6**, 733–807.
- [16] HASSAN, M. (2005) : *Estimation dans les modèles elliptiques avec une application à la biostatistique*. Thèse Université Montpellier 2 (in French).
- [17] HELGASON, S. (1984). *Groups and geometric analysis: Integral geometry, invariant differential operators and spherical functions*. Academic Press, New York.
- [18] HENZE, N. (1997). Do components of smooth tests of fit have diagnostic properties ? *Metrika*, **45**, 121–130.
- [19] HENZE, N. (2002) : Invariant tests for multivariate normality : a critical review. *Stat. Papers*, **43**, 467-506.
- [20] INGLOT T., KALLENBERG, W.C.M. and LEDWINA, T. (1994). Power approximations to and power comparison of smooth goodness-of-fit tests. *Scand. Stat. Theory. Appl.*, **21**, 131–145.
- [21] JAVITZ, H. (1975): *Generalized smooth tests of goodness-of-fit, independence and equality of distributions*. Ph.D. Thesis, University of California, Berkeley.
- [22] JOHNSON, M.E. (1987) : *Multivariate Statistical Simulation*. Wiley, New York.
- [23] KALLENBERG, W.C.M. and LEDWINA, T. (1997). Data driven smooth tests when the hypothesis is composite. *J. Am. Stat. Assoc.*, **92**, 1094–1104.
- [24] KARIYA, T. and GEORGE, E.I. (1995). LBI tests for multivariate normality in curved families and Mardia’s test. *Sankhyā: The Indian Journal of Statistics, Ser A*, **57**, 440–451.
- [25] KOIZUMI, K., OKAMOTO, N., and SEO, T. (2009) : On Jarque-Bera tests for assessing multivariate normality. *Journal of Statistics: Advances in Theory and Applications*, **1**, 207–220.
- [26] KENT, J.T. and TYLER, D.E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist* **19**, 2102–2119.
- [27] KLAR, B. (2000). Diagnostic smooth tests of fit. *Metrika*, **52**, 237–252.
- [28] KOZIOL, J.A. (1986). Assessing multivariate normality: A compendium. *Comm. Statist. Theory Methods*, **15**, 2763–2783.
- [29] KOZIOL, J.A. (1987). An alternative formulation of Neyman’s smooth goodness-of-fit tests under composite alternatives. *Metrika*, **34**, 17–24.
- [30] LANCASTER, H.O. (1969). *The chi-square distribution*. Wiley, New York.
- [31] LEDWINA, T. (1994). Data-Driven Version of Neyman’s Smooth Test of Fit, *J. Am. Stat. Assoc.*, **89**, 1000–1005.
- [32] LEMONTE A.J., PATRIOTA, A. (2011) : Multivariate elliptical models with general

- parametrization. *Stat. Methodol.*, **8**, 389–400.
- [33] MARDIA, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–30.
- [34] MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979). *Multivariate Analysis*. Academic Press, London.
- [35] MARDIA, K.V. and KENT, J.T. (1991). Rao score tests for goodness-of-fit and independence. *Biometrika*, **78**, 355–363.
- [36] MCASSEY, M.P. (2013) . An empirical goodness-of-fit test for multivariate distributions. *J. Appl. Stat.*, **40**, 1120–1131.
- [37] MECKLIN, C.J. and MUNDFROM, D.J. (2005): A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *J. Stat. Comput. Simul.*, **75**, 93–107.
- [38] MUIRHEAD, R.B. (1982): *Aspects of multivariate statistical theory*. Wiley, New York.
- [39] NAIK, D.N., PLUNGPOGPUN, K. (2006) : *A Kotz-type distribution for multivariate statistical inference*. In Advances in distribution theory, order statistics, and inference, N. Balakrishnan, J.M. Sarabia, E. Castillo eds. Statistics for Industry and Technology Part II, Birkhäuser, Boston, 111–124
- [40] NEYMAN, J. (1937). Smooth tests for goodness-of-fit. *Skand. Aktuarietidskr*, **20**, 150–199.
- [41] OZLEM A. and DEMET Y., (2016) : Comparison of some multivariate normality tests: A simulation study. *Int. J. Adv. Applied Sciences*, **3**, 73–85.
- [42] RAYNER, J.C.W. and BEST D.J. (1988). Smooth tests of goodness-of-fit for regular distributions. *Commun. Stat. Theory. Methods.*, **17**, 3235–67.
- [43] RAYNER, J.C.W., THAS, O. and BEST, D.J. (2009). *Smooth tests of goodness-of-fit using R, Second edition*. John Wiley & Sons, Singapore.
- [44] RAYNER, J.C.W. and BEST, D.J. (1990). Smooth Tests of Goodness of Fit: An Overview, *Int. Stat. Rev.*, **58**, 9–17.
- [45] ROJAS, M.A., BOLFARINE, H., and GOMEZ, H.W. (2014) : An extension of the slash-elliptical distribution. *Stat. Oper. Res.*, **38**, 215–230.
- [46] ROMEU J.L. and **Ozturk, A.** (1993). A comparative study of goodness-of-fit tests for multivariate normality. *J. Multivar. Anal.*, **46**, 309–34.
- [47] SHORACK, G.R. and WELLNER, J.A., (1986) : *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [48] THAS, O. (2010) : *Comparing Distributions*, Springer, New York.
- [49] THOMAS, D.R. and PIERCE, D.A. (1979). Neyman’s smooth goodness-of-fit test when the hypothesis is composite. *J. Am. Stat. Assoc.*, **74**, 441–5.
- [50] WANG, J., and GENTON, M.G. (2006). The multivariate skew-slash distribution. *J. Stat. Plan. Inference*, **136**, 209–220.

## APPENDIX A: PROOFS AND AUXILIARY RESULTS

**A.1. Proof of Theorem 3.1.** Let  $k \geq 0$  and  $\mathbf{x} \in \mathbb{R}^m$ . For any fixed  $\gamma = (\mathbf{A}, \mathbf{b}) \in AL(m)$ , define the functions

$$\tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma) = \Psi_{i,\ell}(\gamma(\mathbf{x})/\|\gamma(\mathbf{x})\|)\|\gamma(\mathbf{x})\|^i,$$

where the  $\Psi_{i,\ell}(\cdot)$  are the spherical harmonics evoked in Section 3.1. To avoid trivialities, set  $\tilde{\Psi}_{i,\ell}(-\mathbf{b}; \gamma) = 0$  for  $i > 0$  and  $\tilde{\Psi}_{0,1}(-\mathbf{b}; \gamma) = 1$ . It follows that  $\tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma)$  is an homogeneous polynomial of degree  $i$  in the components of  $\gamma(\mathbf{x}) = \mathbf{A}(\mathbf{x} + \mathbf{b})$  and thus a polynomial of degree  $i$  in  $\mathbf{x} \in \mathbb{R}^m$ . We need the following lemma which states that for any given  $\gamma$ , any polynomial in  $\mathbb{R}^m$  can be expressed as a linear combination of the  $\pi_{k,j,\ell}(r, \mathbf{u})$  of (3.8) and explains how the coefficients of this linear combination are affected by the choice of  $\gamma$ .

LEMMA. *Let  $p(\cdot) \in \mathcal{P}_k$ . Then, for any  $\gamma = (\mathbf{A}, \mathbf{b}) \in AL(m)$ , we have*

$$(A.1) \quad p(\mathbf{x}) = \sum_{i=0}^k \sum_{j=0}^{\lfloor (k-i)/2 \rfloor} \sum_{\ell=1}^{e_m(i)} a_{j,i,\ell}(\gamma) s_{j,i}(\|\gamma^{-1}(\mathbf{x})\|) \tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma^{-1}),$$

for all  $\mathbf{x} \in \mathbb{R}^m$ , where  $\gamma^{-1} = (\mathbf{A}^{-1}, -\mathbf{A}\mathbf{b})$  and

$$(A.2) \quad a_{j,i,\ell}(\gamma) = c_m \det(\mathbf{A}) \int_{\mathbb{R}^m} p(\mathbf{x}) s_{j,i}(\|\gamma^{-1}(\mathbf{x})\|) \tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma^{-1}) \phi_m(\|\gamma^{-1}(\mathbf{x})\|^2) d\mathbf{x}.$$

PROOF. Let  $\mathbf{x} = r\mathbf{u}$  where  $r = \|\mathbf{x}\|$  and  $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$ . To avoid trivialities again, when  $\mathbf{x} = 0$ , set  $r = 0$ , define  $\mathbf{u}$  arbitrarily and agree to take  $r^i = 1$ . Consider the polynomial  $p(\gamma(r\mathbf{u}))$  on  $\mathbb{R}^+ \times \Omega_m$ . For fixed  $r$ , this is a polynomial of degree at most  $k$  in the components of  $\mathbf{u}$ . Now, because the spherical harmonics  $\{\Psi_{i,\ell}(\cdot), \ell = 1, \dots, e_m(i), i \geq 0\}$  form a CONB with respect to the scalar product  $\langle \cdot, \cdot \rangle_{d\omega_m}$  for  $L^2(\Omega_m)$ , and thus for the restriction to  $\Omega_m$  of the space of polynomials of degree  $k$  on  $\mathbb{R}^m$ , it follows that

$$p(\gamma(r\mathbf{u})) = \sum_{i=0}^k \sum_{\ell=1}^{e_m(i)} q_{i,\ell}(r; \gamma) \Psi_{i,\ell}(\mathbf{u}),$$

where

$$q_{i,\ell}(r; \gamma) = \frac{1}{\omega_m(\Omega_m)} \int_{\Omega_m} p(\gamma(r\mathbf{u})) \Psi_{i,\ell}(\mathbf{u}) d\omega_m(\mathbf{u}).$$

Setting  $e = (\mathbf{I}_m, \mathbf{0})$  as the identity element of  $AL(m)$ , this shows that

$$\begin{aligned} p(\gamma(\mathbf{x})) &= \sum_{i=0}^k \sum_{\ell=1}^{e_m(i)} w_{i,\ell}(\|\mathbf{x}\|; \gamma) \tilde{\Psi}_{i,\ell}(\mathbf{x}; e) \\ (A.3) \quad &= \sum_{i=0}^k \sum_{\ell=1}^{e_m(i)} w_{i,\ell}(\|\mathbf{x}\|; \gamma) \tilde{\Psi}_{i,\ell}(\gamma(\mathbf{x}); \gamma^{-1}), \end{aligned}$$

where  $w_{i,\ell}(\|\mathbf{x}\|; \gamma) = \|\mathbf{x}\|^{-i} q_{i,\ell}(\|\mathbf{x}\|; \gamma)$ . Thus

$$(A.4) \quad p(\mathbf{x}) = \sum_{i=0}^k \sum_{\ell=1}^{e_m(i)} w_{i,\ell}(\|\gamma^{-1}(\mathbf{x})\|; \gamma) \tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma^{-1}).$$

Now, because  $\tilde{\Psi}_{i,\ell}(\mathbf{x}; \gamma^{-1})$  is a polynomial of degree  $i$  in  $\mathbf{x}$  and  $p(\gamma(\mathbf{x}))$  is a polynomial of degree at most  $k$ , it follows that each  $w_{i,\ell}(\|\mathbf{x}\|; \gamma)$  is a polynomial of degree at most  $[(k-i)/2]$  in  $\|\mathbf{x}\|^2$ . Hence in turn, it can be written as a linear combination of the polynomials  $s_{j,i}(\|\mathbf{x}\|)$ ,  $i = 0, \dots, [(k-i)/2]$  of (3.4), namely

$$(A.5) \quad w_{i,\ell}(r; \gamma) = \sum_{j=0}^{[(k-i)/2]} a_{j,i,\ell}(\gamma) s_{j,i}(r).$$

Combining (A.4) and (A.5) yields (A.1). Now using (3.2) and (3.4) with (A.5), we have

$$(A.6) \quad a_{j,i,\ell}(\gamma) = \omega_m(\Omega_m) c_m \int_0^\infty w_{i,\ell}(r; \gamma) s_{j,i}(r) \phi_m(r^2) r^{m-1+2j} dr,$$

$$(A.7) \quad = c_m \int_{\mathbb{R}^m} w_{i,\ell}(\|\mathbf{x}\|; \gamma) s_{j,i}(\|\mathbf{x}\|) \phi_m(\|\mathbf{x}\|^2) \|\mathbf{x}\|^{2j} d\mathbf{x},$$

$$(A.8) \quad = c_m \det(\mathbf{A}) \int_{\mathbb{R}^m} w_{i,\ell}(\|\gamma^{-1}(\mathbf{x})\|; \gamma) s_{j,i}(\|\gamma^{-1}(\mathbf{x})\|) \phi_m(\|\gamma^{-1}(\mathbf{x})\|^2) \|\gamma^{-1}(\mathbf{x})\|^{2j} d\mathbf{x},$$

upon transforming from polar to cartesian coordinates. Finally, substitute (A.4) into the right-hand side of (A.2). Simple algebra using the orthogonality of the  $\Psi_{i,\ell}(\cdot)$  yields the right-hand side of (A.8). This concludes the proof of the lemma.  $\square$

Proof of *i*). A change of variable in 3.6 shows that, without loss of generality, we can set  $\eta = (\mathbf{0}, \mathbf{I}_m)$ . By (3.6), when  $k \neq k'$ , the spaces  $\Pi_k$  and  $\Pi_{k'}$  are orthogonal with respect to  $\langle \cdot, \cdot \rangle_{(f, \eta)}$ . By the lemma, all polynomials  $p(\cdot)$  in  $\mathcal{P}_k$  have an expansion of the form (A.1) where from (A.8) we can take  $\gamma = e$  to simplify the derivations, while all  $p_{k'}(\gamma(\cdot)) \in \Pi_{k'}$  have an expansion of the form A.3. Substituting these expansions into (3.6) shows, after careful identification and using the orthonormality of the  $s_{j,i}(\cdot)$ ,  $\Psi_{i,\ell}(\cdot)$ , that  $p(\cdot) \in \Pi_k$  if and only if its coefficient in (A.1) satisfies  $a_{j,i,\ell}(e) = 0$  for every  $i, j$  such that  $i \neq k - 2j$ . Thus, after rearranging the indices and extending to any  $\gamma \in AL(m)$ , each  $p(\cdot) \in \Pi_k$  has an expansion of the form

$$(A.9) \quad p(\mathbf{x}) = \sum_{j=0}^{\lfloor k/2 \rfloor} \sum_{\ell=1}^{e_m(j)} a_{j,k-2j,\ell}(\gamma) s_{j,k-2j}(\|\gamma^{-1}(\mathbf{x})\|) \tilde{\Psi}_{k-2j,\ell}(\mathbf{x}; \gamma^{-1}).$$

Because the set of polynomials  $s_{j,k-2j}(\|x\|) \tilde{\Psi}_{k-2j,\ell}(\mathbf{x})$  are orthonormal with respect to  $\langle \cdot, \cdot \rangle_{(f, \eta)}$  and span  $\Pi_k$ , it follows from (A.1) and (A.9) that  $\mathcal{P}_k = \Pi_k \oplus \mathcal{P}_{k-1}$ . This proves the first part of *i*). The second part comes from the fact that the polynomials are dense in  $L^2_{AL(m)}(f, \eta)$ .

Proof of *ii*) By construction,  $\Pi_k$  is an  $AL(m)$ -invariant subspace. From (A.9), for any  $\gamma \in AL(m)$ , there is only one element of  $\Pi_k$ , up to a multiplicative constant, that is invariant over the subgroup of rotations. It follows from Schur's lemma that  $\Pi_k$  is irreducible. Indeed, Schur's lemma [17, p. 390] asserts that if two representations of a group are irreducible then every commuting non-zero linear map between them is an isomorphism and that a representation is irreducible if and only if every such linear map is a scalar multiple of the identity element. In the present context, showing that the space  $\Pi_k$  is irreducible amounts to show that the representation of  $AL(m)$  induced by the coefficients  $a_{k,j,\ell}(\gamma)$  of its basis in (A.9) is irreducible. The identity element is then the element that is invariant over rotations which are identified with the commuting maps.

Setting  $g^{-1} = (V^{-1/2}, -V^{-1/2}\mu)$ , the

$$\pi_{k,j,\ell}(r, u) = s_{j,k-2j}(\|g^{-1}(x)\|) \tilde{\Psi}_{k-2j,\ell}(x, g^{-1})$$

are orthonormal with respect to  $\langle \cdot, \cdot \rangle_{(f, \eta)}$  and, from (A.9), form a complete basis for  $\Pi_k$ . Let  $\Pi_{k,j}(g^{-1})$  be the space generated by the span of  $\{s_{j,k-2j}(\|g^{-1}(x)\|) \tilde{\Psi}_{k-2j,\ell}(x; g^{-1}), \ell = 1, \dots, e_m(k - 2j)\}$ . Because, for fixed  $(k, j)$ ,  $\Pi_{k,j}(g^{-1})$  is closed under rotations, the functions that span  $\Pi_{k,j}(g^{-1})$  are, up to multiplicative constants, the same as those that span  $E_m(k - 2j)$ . Thus  $\Pi_{k,j}(g^{-1})$  is an irreducible rotation-invariant subspace of  $\Pi_k$  of dimension  $e_m(k - 2j)$ . Combining these functions shows that for  $g^{-1} = (V^{-1/2}, -\mu)$ ,

$\Pi_k = \Pi_{k,0}(g^{-1}) \oplus \dots \oplus \Pi_{k,[k/2]}(g^{-1})$ , with respect to  $\langle \cdot, \cdot \rangle_{(f,\eta)}$ . The dimension of  $\Pi_k$  is computed from this direct sum as a combinatorial exercise.

**A.2. Derivation of the  $\mathcal{Q} = \mathcal{UIR}$  decomposition.** We derive (4.1) and (4.2). First reparametrize  $\boldsymbol{\eta} = (\boldsymbol{\mu}, \mathbf{V}^{-1})$  into the  $m + m(m+1)/2$  vector  $\boldsymbol{\eta} = (\boldsymbol{\mu}, \text{lvec}(\mathbf{V}^{-1}))^T$ , where  $\text{lvec}(\cdot)$  is the ‘‘lower vec’’ operator. Introduce the  $m^2 \times m(m+1)/2$  matrix  $\mathbf{B}$  with element :

$$b_{ij,k\ell} = \begin{cases} 1 & \text{if } (i = k \text{ and } j = \ell) \text{ or } (i = \ell \text{ and } j = k) \\ 0 & \text{otherwise} \end{cases},$$

where  $i$  first goes from 1 to  $m$  and then  $j$  goes from 1 to  $m$  while  $k$  varies from  $\ell$  to  $m$  followed by  $\ell$  varying from 1 to  $m$ . This matrix, which has rank  $m(m+1)/2$ , is such that  $\text{vec}(\mathbf{V}^{-1}) = \mathbf{B} \cdot \text{lvec}(\mathbf{V}^{-1})$ . Now some algebra yields

$$\frac{\partial \log f(\mathbf{x}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbf{N}_\eta \mathbf{Z},$$

where ( $\otimes$  denotes Kronecker’s product)

$$\mathbf{N}_\eta = \begin{pmatrix} \mathbf{V}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{B}^T (\mathbf{V}^{-1/2} \otimes \mathbf{V}^{-1/2}) \mathbf{B} \end{pmatrix},$$

$$(A.10) \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} = \begin{pmatrix} R \times g(R^2) \times \mathbf{U} \\ \text{lvec}(\mathbf{I}_m - \zeta(R^2) \mathbf{U} \mathbf{U}^T) \end{pmatrix},$$

with  $g(r^2) = -2\phi'_m(s)/\phi_m(s)|_{s=r^2}$  and  $\zeta(r^2) = r^2 \times g(r^2)$ . Note that  $\mathbf{N}_\eta$  is invertible and the distribution of  $\mathbf{Z}$  does not depend on  $\boldsymbol{\eta}$ ; it has expectation 0 and, in view of results about the  $U(\Omega_m)$  distribution (see [4, p. 239, Ex. 4], its variance is block-diagonal with the inverse of these blocks being :

$$(A.11) \quad \mathbb{V}^{-1}(\mathbf{Z}_1) = m \mathbf{I}_m / \sigma_1,$$

$$\mathbb{V}^{-1}(\mathbf{Z}_2) = \sigma_2 [\mathbf{I}_{m(m+1)/2} - \frac{1}{2} \text{Diag}(\text{lvec}(\mathbf{I}_m))] - \frac{\sigma_2(1 - \sigma_2)}{4 + 2m(1 - \sigma_2)} \text{lvec}(\mathbf{I}_m) \text{lvec}^T(\mathbf{I}_m),$$

where  $\sigma_1 = \mathbb{E}_0(R^2 \times g^2(R^2))$  and  $\sigma_2 = m(m+2)/\mathbb{E}_0(\zeta^2(R^2))$ . Then  $\mathcal{J}_\eta = \mathbf{N}_\eta \mathbb{V}(\mathbf{Z}) \mathbf{N}_\eta^T$  is also block-diagonal. Moreover, the lines of  $\mathbf{J}_\eta$  are

$$\text{Cov}_0(\pi_{k,j,\ell}(R, \mathbf{U}), \partial \log f(\mathbf{X}; \boldsymbol{\eta}) / \partial \boldsymbol{\eta}^T) = \mathbf{N}_\eta \mathbb{E}_0(R^{k-2j} s_{j,k-2j}(R^2) g(R^2) \Psi_{k-2j,\ell}(\mathbf{U}) \mathbf{Z}^T).$$

Hence the elements of  $\mathbf{J}_\eta \mathcal{J}_\eta^{-1} \mathbf{J}_\eta^T$  do not involve  $\boldsymbol{\eta}$  and  $(\mathbf{I}_{v_\mathcal{Q}} - \mathbf{J}_\eta \mathcal{J}_\eta^{-1} \mathbf{J}_\eta^T)$  depends only on the distribution of  $R$ . Defining  $\boldsymbol{\Psi}_k = (\Psi_{k,\ell}(\mathbf{U}), \ell = 1, \dots, e_m(k))^T$ ,

it follows that  $\mathbf{U} = \Psi_1/\sqrt{m}$  while  $\text{lvec}(\mathbf{U}\mathbf{U}^T) = \mathbf{A}_m(\Psi_0^T, \Psi_2^T)^T$ , where  $\mathbf{A}_m$  is a square matrix of order  $m(m+1)/2$  partitioned as  $(\mathbf{A}_{m(1)} \mid \mathbf{A}_{m(2)})$  with  $\mathbf{A}_{m(1)} = \text{lvec}(\mathbf{I}_p)/m$  being its first column. The expression of  $\mathbf{A}_{m(2)}$  is not required beyond the relationships  $\text{lvec}(\mathbf{I}_p)^T \mathbf{A}_{m(2)} = \mathbf{0}$  and  $\mathbf{A}_{m(2)}^T \mathbf{A}_{m(2)} = \frac{1}{m(m+2)} \mathbf{I}_{\frac{m(m+1)}{2}-1}$ , from the expression for  $\mathbb{V}(\text{lvec}(\mathbf{U}\mathbf{U}^T))$  in [4, p. 239, Ex. 4]. It follows from the orthogonality of the  $\Psi_{k-2j,\ell}(\cdot)$  that

$$(A.12) \quad \mathbb{E}_0 \left( R^{k-2j} s_{j,k-2j}(R^2) \Psi_{k-2j} \mathbf{Z}_1^T \right) = \begin{cases} \mathbb{E}_0 (s_{j,1}(R^2) \zeta(R^2)) \mathbf{I}_m / \sqrt{m} & \text{if } k-2j=1 \\ \mathbf{0} & \text{otherwise} \end{cases},$$

while adding the orthogonality of the  $s_{j,i}(\cdot)$ ,

$$(A.13) \quad \mathbb{E}_0 \left( R^{k-2j} s_{j,k-2j}(R^2) \Psi_{k-2j} \mathbf{Z}_2^T \right) = \begin{cases} -\mathbb{E}_0 (s_{j,0}(R^2) \zeta(R^2)) \mathbf{A}_{m(1)}^T & \text{if } k-2j=0 \\ -\mathbb{E}_0 (R^2 s_{j,2}(R^2) \zeta(R^2)) \mathbf{A}_{m(2)}^T & \text{if } k-2j=2 \\ \mathbf{0} & \text{otherwise} \end{cases}.$$

Collect the constants in these expressions into the following vectors : for  $k-2j=0$ ,  $\mathbf{c}_0^T = (-\mathbb{E}_0 (s_{j,0}(R^2) \zeta(R^2)), j=1, \dots, [K/2])$ , for  $k-2j=1$ ,  $\mathbf{c}_1^T = (\mathbb{E}_0 (s_{j,1}(R^2) \zeta(R^2)) / \sqrt{m}, j=1, \dots, [(K+1)/2])$  and for  $k-2j=2$ ,  $\mathbf{c}_2^T = (-\mathbb{E}_0 (R^2 s_{j,2}(R^2) \zeta(R^2)), j=1, \dots, [K/2])$ . Permute the components of  $(\pi_{k,j,\ell}(\cdot, \cdot), k=1, \dots, K, (j, \ell) \in B_k)^T$  into  $(\pi_{\mathcal{U}}^T, \pi_{\mathcal{I}}^T, \pi_{\mathcal{R}}^T)^T$  so that all  $\pi_{k,0,\ell}(\cdot, \cdot)$  with  $k \geq 3$  are in  $\pi_{\mathcal{U}}$  in increasing order of  $\ell$  and then  $k$ , all  $\pi_{k,[k/2],1}(\cdot, \cdot)$  with  $k \geq 2$  are in  $\pi_{\mathcal{R}}$  while all other terms are regrouped in  $\pi_{\mathcal{I}}$  in the following way : let  $\mathfrak{J}_1$  be the matrix with rows  $(\pi_{k,j,\ell}(r, \mathbf{u}) \mid k-2j=1, \ell=1, \dots, e_m(k-2j))$  ordered from row to row according to  $k$ . Similarly define  $\mathfrak{J}_2$  with  $k-2j=2$  and regroup all other terms (those with  $k-2j > 2$ ) into  $\pi_{\mathcal{I},3}$  ordered lexicographically. Then  $\pi_{\mathcal{I}}^T = (\text{vec}(\mathfrak{J}_1)^T, \text{vec}(\mathfrak{J}_2)^T, \pi_{\mathcal{I},3}^T)^T = (\pi_{\mathcal{I},1}^T, \pi_{\mathcal{I},2}^T, \pi_{\mathcal{I},3}^T)^T$ . Form  $(\bar{\pi}_{\mathcal{U}}^T, \bar{\pi}_{\mathcal{I}}^T, \bar{\pi}_{\mathcal{R}}^T)^T$ , the permuted version of (2.9) ordered in the same way with  $\bar{\pi}_{\mathcal{I}}^T = (\bar{\pi}_{\mathcal{I},1}^T = \text{vec}(\tilde{\mathfrak{J}}_1)^T, \bar{\pi}_{\mathcal{I},2}^T = \text{vec}(\tilde{\mathfrak{J}}_2)^T, \bar{\pi}_{\mathcal{I},3}^T)^T$ . Easy calculations show that  $\mathbf{J}_\eta$  with rows permuted accordingly has the form

$$(A.14) \quad \mathbf{J}_\eta = \mathbf{N}_\eta \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \nu_{\mathcal{U}} \times m & \nu_{\mathcal{U}} \times m(m+1)/2 \\ \tau_{11} & \tau_{12} \\ \nu_{\mathcal{I}} \times m & \nu_{\mathcal{I}} \times m(m+1)/2 \\ \mathbf{0} & \tau_{22} \\ \nu_{\mathcal{R}} \times m & \nu_{\mathcal{R}} \times m(m+1)/2 \end{pmatrix},$$

where  $\nu_{\mathcal{R}} = [K/2]$ ,  $\nu_{\mathcal{U}} = \sum_{k=1}^K e_m(k)$ ,  $\nu_{\mathcal{I}} = \nu_{\mathcal{Q}} - \nu_{\mathcal{R}} - \nu_{\mathcal{U}}$  where  $\nu_{\mathcal{Q}} = \sum_{k=1}^K C_k^{m+k-1}$ . Recall that the  $\tau_{ij}$  depend only on the distribution of  $R$

under  $H_0$ . Thus

$$(A.15) \quad (\mathbf{I}_{\nu_{\mathcal{Q}}} - \mathbf{J}_{\eta} \mathcal{J}_{\eta}^{-1} \mathbf{J}_{\eta}^T) = \begin{pmatrix} \mathbf{I}_{\nu_{\mathcal{U}}} & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{R}} \\ 0 & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{R}}^T & \boldsymbol{\Sigma}_{\mathcal{R}\mathcal{R}} \end{pmatrix}.$$

We focus on  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{R}} = -\boldsymbol{\tau}_{12} \mathbb{V}^{-1}(\mathbf{Z}_2) \boldsymbol{\tau}_{22}^T$ . The lines of matrix  $\boldsymbol{\tau}_{22}$  have the form  $-\mathbb{E}_0(s_{j,0}(R^2) \zeta(R^2)) (\text{lvec}(\mathbf{I}_p))^T$ . Combining this with the expression of  $\mathbb{V}^{-1}(\mathbf{Z}_2)$  and the fact that the lines of  $\boldsymbol{\tau}_{12}$  either vanish or, from (A.13) have the form  $-\mathbb{E}_0(R^2 s_{j,2}(R^2) \zeta(R^2)) \mathbf{A}_{m(2)}^T$ , it follows that  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{R}} = 0$ . Also  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I}} = \mathbf{I}_{\nu_{\mathcal{I}}} - \boldsymbol{\tau}_{11} \mathbb{V}^{-1}(\mathbf{Z}_1) \boldsymbol{\tau}_{11}^T - \boldsymbol{\tau}_{12} \mathbb{V}^{-1}(\mathbf{Z}_2) \boldsymbol{\tau}_{12}^T$  and  $\boldsymbol{\Sigma}_{\mathcal{R}\mathcal{R}} = \mathbf{I}_{\nu_{\mathcal{R}}} - \boldsymbol{\tau}_{22} \mathbb{V}^{-1}(\mathbf{Z}_2) \boldsymbol{\tau}_{22}^T$ . It is easy to see that  $\boldsymbol{\Sigma}_{\mathcal{R}\mathcal{R}}^{-1} = \mathbf{I}_{[K/2]} + d_0 \mathbf{c}_0 \mathbf{c}_0^T$ , where  $d_0 = \frac{\sigma_2}{m(2+m(1-\sigma_2)-\sigma_2\|\mathbf{c}_0\|^2)}$ . It is again easy to see that  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I}}^{-1}$  is block diagonal, with block  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I},1}^{-1} = \mathbf{I}_{e_m(1) \times [(K+1)/2]} + \mathbf{I}_{e_m(1)} \otimes d_1 \mathbf{c}_1 \mathbf{c}_1^T$ , where  $d_1 = \frac{m}{\sigma_1 - m\|\mathbf{c}_1\|^2}$ , block  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I},2}^{-1} = \mathbf{I}_{e_m(2) \times [K/2]} + \mathbf{I}_{e_m(2)} \otimes d_2 \mathbf{c}_2 \mathbf{c}_2^T$  with  $d_2 = \frac{\sigma_2}{m(2+m)-\sigma_2\|\mathbf{c}_2\|^2}$  and block  $\boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I},3}^{-1} = \mathbf{I}_{\nu_{\mathcal{I}} - e_m(1) \times ((K+1)/2) + [K/2]}$ . Using standard properties of the  $\text{vec}$  and  $\otimes$  operators, we get  $\boldsymbol{\pi}_{\mathcal{I},1}^T \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{I},1}^{-1} \boldsymbol{\pi}_{\mathcal{I},1} = d_1 \text{tr}(\mathbf{c}_1 \mathbf{c}_1^T \tilde{\mathcal{J}}_1 \tilde{\mathcal{J}}_1^T)$  and similarly for the other term. Collecting these, we finally get :

$$(A.16) \quad \mathcal{Q}_K = n \left[ \|\bar{\boldsymbol{\pi}}_{\mathcal{U}}\|^2 + \|\bar{\boldsymbol{\pi}}_{\mathcal{I},1}\|^2 + d_1 \text{tr}(\mathbf{c}_1 \mathbf{c}_1^T \tilde{\mathcal{J}}_1 \tilde{\mathcal{J}}_1^T) + \|\bar{\boldsymbol{\pi}}_{\mathcal{I},2}\|^2 \right. \\ (A.17) \quad \left. + d_2 \text{tr}(\mathbf{c}_2 \mathbf{c}_2^T \tilde{\mathcal{J}}_2 \tilde{\mathcal{J}}_2^T) + \|\bar{\boldsymbol{\pi}}_{\mathcal{I},3}\|^2 + \|\bar{\boldsymbol{\pi}}_{\mathcal{R}}\|^2 + d_0 \text{tr}(\mathbf{c}_0 \mathbf{c}_0^T \bar{\boldsymbol{\pi}}_{\mathcal{R}} \bar{\boldsymbol{\pi}}_{\mathcal{R}}^T) \right].$$

## APPENDIX B: MATHEMATICA COMMANDS TO GENERATE THE $\pi_{K,J,\ell}(R, U)$ OF SECTION 3.1

We give the MATHEMATICA commands to generate the  $\{\Psi_{k,j}(u) \mid j = 1, \dots, e_m(k)\}$  of Section 3.1. The package `HFT10.m` must first be downloaded from the site given in Appendix B of [1] and loaded into MATHEMATICA via the command `<<'.../.../.../HFT10.m'>>`, where `.../.../.../` is the path leading to where the package `HFT10.m` is stored on the computer; after typing `<<`, one can use the “File Path...” command in the “Insert” menu to automatically generate this path.

Once the package is loaded (the text “\* You can now use the functions in this package.” will appear on the MATHEMATICA notebook), the user needs only to set the dimension  $m$  via the command `setDimension[u, m]`. Here  $u = (u_1, \dots, u_m)$  is the vector of variables in which the spherical harmonics will be expressed. Then the command

$$\Psi[\mathbf{k}_-, \mathbf{u}_-] := \text{basisH}[\mathbf{k}, \mathbf{u}, \text{Sphere}]/\|\mathbf{u}\| \rightarrow 1;$$



$\pi_{3,0,1} = \frac{r^3 u_1 (3-4u_1^2)}{2\sqrt{6}}$	$\pi_{3,0,2} = \frac{r^3 u_2 (1-4u_1^2)}{2\sqrt{6}}$	$\pi_{3,1,1} = \frac{r(r^2-4)u_1}{2\sqrt{2}}$
$\pi_{3,1,2} = \frac{r(r^2-4)u_2}{2\sqrt{2}}$	$\pi_{4,0,1} = \frac{r^4(1-8u_1^2+8u_1^4)}{8\sqrt{3}}$	$\pi_{4,0,2} = \frac{r^4 u_2 (u_1-2u_1^3)}{2\sqrt{3}}$
$\pi_{4,1,1} = \frac{r^2(r^2-6)(1-2u_1^2)}{4\sqrt{3}}$	$\pi_{4,1,2} = \frac{r^2(r^2-6)u_1 u_2}{2\sqrt{3}}$	$\pi_{4,2,1} = \frac{r^4-8r^2+8}{8}$
$\pi_{5,0,1} = \frac{r^5 u_1 (5-20u_1^2+16u_1^4)}{8\sqrt{30}}$	$\pi_{5,0,2} = \frac{r^5 u_2 (1-12u_1^2+16u_1^4)}{8\sqrt{30}}$	$\pi_{5,1,1} = \frac{r^3(r^2-8)u_1(3-4u_1^2)}{8\sqrt{6}}$
$\pi_{5,1,2} = \frac{r^3(r^2-8)u_2(1-4u_1^2)}{8\sqrt{6}}$	$\pi_{5,2,1} = \frac{r(r^4-12r^2+24)u_1}{8\sqrt{3}}$	$\pi_{5,2,2} = \frac{r(r^4-12r^2+24)u_2}{8\sqrt{3}}$

TABLE 3

Table of the polynomials  $\pi_{k,j,\ell}(r, u)$  of Theorem 3.1 in the case of a MVN null hypothesis, for  $m = 2$  and  $k = 3, 4$  and 5.

generates the basis of dimension  $e_m(k)$  for  $E_m(K)$ . For example, typing

```
setDimension[u, 5];
Do[Print[TableForm[Ψ(k, u)]], {k, 1, 4}];
```

prints the  $\{\Psi_{k,j}(u) \mid j = 1, \dots, e_5(k)\}$ , for  $k = 1, \dots, 4$ . To generate the  $s_{j,i}(r)$  for the case of the null MVN distribution, the command is

```
s[j_, i_, r_] := (-1)^j*Sqrt[j!*Gamma[m/2]/(2^i*Gamma[m/2 + j + i])] * LaguerreL[j, m/2 + i - 1, r^2/2];
```

Finally, to generate the  $\pi_{k,j,\ell}(r, u)$ , the commands are

```
e[m_, k_] := Which[k == 0, 1, k == 1, m, k >= 2, Binomial[m + k - 1, m - 1] - Binomial[m + k - 3, m - 1]];
πfunct[k_, j_, ℓ_, r_] := r^(k - 2*j)*s[j, k - 2*j, r]*Ψ[k - 2*j, u][[ℓ]];
CONBII[k_] := Flatten[Table[πfunct[k, j, ℓ, r], {j, 0, Floor[k/2]}, {1, 1, e[m, k - 2*j]}]]];
```

Table 3 lists the polynomials  $\pi_{k,j,\ell}(r, u)$  for the case  $m = 2, k = 3, \dots, 5$  that are used in the experiment of Section 6.

Table 4 lists the  $\pi_{k,j,\ell}(r, u)$  required for the Open/Closed book example of Section 6 ( $m = 3$  and  $k = 3, 4$  and 5)

## APPENDIX C: SMOOTH TEST FOR THE BIVARIATE LOGISTIC AND PEARSON TYPE II DISTRIBUTIONS

**C.1. The bivariate logistic distribution.** We consider the bivariate logistic distribution in [32] with density generator  $\phi_m(y) = e^{-y}/(1 + e^{-y})^2$ . This is again a competitor to the MVN but with shorter tails. The moments of  $R^2$  have explicit but complicated expressions and it is shorter to use numerical approximations. Consequently, the  $s_{j,k-2j}(\cdot)$  required for  $\mathcal{Q}_5$  are :

$\pi_{3,0,1} = \frac{r^3 u_1 (1-5u_2^2)}{2\sqrt{10}}$	$\pi_{3,0,2} = \frac{r^3 u_2 (3-5u_2^2)}{2\sqrt{15}}$	$\pi_{3,0,3} = \frac{r^3 u_3 (1-5u_2^2)}{2\sqrt{10}}$
$\pi_{3,0,4} = \frac{r^3 u_1 (u_1^2 - 3u_3^2)}{2\sqrt{6}}$	$\pi_{3,0,5} = \frac{r^3 u_2 (u_1^2 - u_3^2)}{2}$	$\pi_{3,0,6} = \frac{r^3 u_3 (3u_1^2 - u_3^2)}{2\sqrt{6}}$
$\pi_{3,0,7} = r^3 u_1 u_2 u_3$	$\pi_{3,1,1} = \frac{r(r^2-5)u_1}{\sqrt{10}}$	$\pi_{3,1,2} = \frac{r(r^2-5)u_2}{\sqrt{10}}$
$\pi_{3,1,3} = \frac{r(r^2-5)u_3}{\sqrt{10}}$	$\pi_{4,0,1} = \frac{r^4(3-30u_2^2+35u_2^4)}{8\sqrt{105}}$	$\pi_{4,0,2} = \frac{r^4 u_2 u_3 (3-7u_2^2)}{2\sqrt{42}}$
$\pi_{4,0,3} = \frac{r^4(7u_2^2-1)(u_3^2-u_1^2)}{4\sqrt{21}}$	$\pi_{4,0,4} = \frac{r^4 u_2 u_3 (3u_1^2-u_3^2)}{2\sqrt{6}}$	$\pi_{4,0,5} = \frac{r^4(u_1^4-6u_1^2 u_3^2+u_3^4)}{8\sqrt{3}}$
$\pi_{4,0,6} = \frac{r^4 u_1 u_2 (3-7u_2^2)}{2\sqrt{42}}$	$\pi_{4,0,7} = \frac{r^4 u_1 u_3 (1-7u_2^2)}{4\sqrt{21}}$	$\pi_{4,0,8} = \frac{r^4 u_1 u_2 (u_1^2-3u_3^2)}{2\sqrt{6}}$
$\pi_{4,0,9} = \frac{r^4 u_1 u_3 (u_1^2-u_3^2)}{2\sqrt{3}}$	$\pi_{4,1,1} = \frac{r^2(r^2-7)(1-3u_2^2)}{2\sqrt{42}}$	$\pi_{4,1,2} = \frac{r^2(r^2-7)u_2 u_3}{\sqrt{14}}$
$\pi_{4,1,3} = \frac{r^2(r^2-7)(u_1^2-u_3^2)}{2\sqrt{14}}$	$\pi_{4,1,4} = \frac{r^2(r^2-7)u_1 u_2}{\sqrt{14}}$	$\pi_{4,1,5} = \frac{r^2(r^2-7)u_1 u_3}{\sqrt{14}}$
$\pi_{4,2,1} = \frac{(r^4-10r^2+15)}{2\sqrt{30}}$	$\pi_{5,0,1} = \frac{r^5 u_2 (15-70u_2^2+63u_2^4)}{24\sqrt{105}}$	$\pi_{5,0,2} = \frac{r^5 u_3 (1-14u_2^2+21u_2^4)}{24\sqrt{7}}$
$\pi_{5,0,3} = \frac{r^5 u_2 (1-3u_2^2)(u_1^2-u_3^2)}{12}$	$\pi_{5,0,4} = \frac{r^5 u_3 (1-9u_2^2)(3u_1^2-u_3^2)}{24\sqrt{6}}$	$\pi_{5,0,5} = \frac{r^5 u_2 (9u_1^4-6u_1^2 u_3^2-7u_3^4)}{8\sqrt{3}}$
$\pi_{5,0,6} = \frac{r^5 u_3 (5u_1^4-10u_1^2 u_3^2+u_3^4)}{8\sqrt{30}}$	$\pi_{5,0,7} = \frac{r^5 u_1 (1-14u_2^2+21u_2^4)}{24\sqrt{7}}$	$\pi_{5,0,8} = \frac{r^5 u_1 u_2 u_3 (1-3u_2^2)}{6}$
$\pi_{5,0,9} = \frac{r^5 u_1 (1-9u_2^2)(u_1^2-3u_3^2)}{24\sqrt{6}}$	$\pi_{5,0,10} = \frac{r^5 u_1 u_2 u_3 (u_1^2-u_3^2)}{2\sqrt{3}}$	$\pi_{5,0,11} = \frac{r^5 u_1 (u_1^4-10u_1^2 u_3^2+5u_3^4)}{8\sqrt{30}}$
$\pi_{5,1,1} = \frac{r^3(r^2-9)u_2(3-5u_2^2)}{6\sqrt{30}}$	$\pi_{5,1,2} = \frac{r^3(r^2-9)u_3(1-5u_2^2)}{12\sqrt{5}}$	$\pi_{5,1,3} = \frac{r^3(r^2-9)u_2(u_1^2-u_3^2)}{6\sqrt{2}}$
$\pi_{5,1,4} = \frac{r^3(r^2-9)u_3(3u_1^2-u_3^2)}{12\sqrt{3}}$	$\pi_{5,1,5} = \frac{r^3(r^2-9)u_1(1-5u_2^2)}{12\sqrt{5}}$	$\pi_{5,1,6} = \frac{r^3(r^2-9)u_1 u_2 u_3}{3\sqrt{2}}$
$\pi_{5,1,7} = \frac{r^3(r^2-9)u_1(u_1^2-3u_3^2)}{12\sqrt{3}}$	$\pi_{5,2,1} = \frac{r(r^4-14r^2+35)u_1}{2\sqrt{70}}$	$\pi_{5,2,2} = \frac{r(r^4-14r^2+35)u_2}{2\sqrt{70}}$
$\pi_{5,2,3} = \frac{r(r^4-14r^2+35)u_3}{2\sqrt{70}}$		

TABLE 4

Table of the polynomials  $\pi_{k,j,\ell}(r, u)$  of Theorem 3.1 in the case of a MVN null hypothesis, for  $m = 3$  and  $k = 3, 4$  and  $5$ .

$s_{0,1}(r) = 0.849322$ ;  $s_{0,2}(r) = 0.551329$ ;  $s_{0,3}(r) = 0.30403$ ;  $s_{0,4}(r) = 0.148319$ ;  
 $s_{0,5}(r) = 0.0654688$ ;  $s_{1,0}(r) = -1.18523 + 0.854964 r^2$ ;  $s_{1,1}(r) = -1.36758 +$   
 $0.576276 r^2$ ;  $s_{1,2}(r) = -1.0461 + 0.318116 r^2$ ;  $s_{1,3}(r) = -0.646026 + 0.153749 r^2$ ;  
 $s_{2,0}(r) = 1.24468 - 1.86588 r^2 + 0.407913 r^4$  and  $s_{2,1}(r) = 1.79127 - 1.51119 r^2 +$   
 $0.230011 r^4$ . Next,  $\zeta(r^2) = 2r^2 \tanh(r^2/2)$ ,  $\sigma_1 = 3.18173$ ,  $\sigma_2 = 0.82306$  and

$$\begin{aligned}\sqrt{d_0}\mathbf{c}_0 &= \{-9.45511, -0.77618\}, \\ \sqrt{d_1}\mathbf{c}_1 &= \{12.55, 3.40145, -1.89893\}, \\ \sqrt{d_2}\mathbf{c}_2 &= \{-8.04922, -1.2599\}.\end{aligned}$$

Continue as in the bivariate Laplace case.

**C.2. The bivariate Pearson type II distribution.** We consider the bivariate Pearson type II distribution described in [22, Sec 6.2] with density generator  $\phi_m(y) = (1-y)^\alpha$  with  $y \in [0, 1]$ . For  $\alpha > 0$ , this is another EC distribution that somewhat resembles the MVN. The  $j$ -th moment of  $R^2$  is  $(2j + \alpha)B(2 + \alpha, 1 + j)$  and from (3.5), the  $s_{j,k-2j}(\cdot)$  required for  $\mathcal{Q}_5$  are :  $s_{0,j}(r) = (\sqrt{(j+1)B(2+\alpha, j+1)})^{-1}$  for  $j = 1, \dots, 5$ ;  $s_{1,j}(r) = \frac{(2+j+\alpha)r^2 - (1+j)}{\sqrt{B(2+\alpha, j+2)(1+\alpha)(2+j+\alpha)}}$  for  $j = 0, \dots, 3$ ;  $s_{2,0}(r) = \frac{(2+r^2(3+\alpha)(r^2(4+\alpha)-4)\sqrt{5+\alpha}}{2\sqrt{1+\alpha}}$  and  $s_{2,1}(r) = \frac{(6+r^2(4+\alpha)(r^2(5+\alpha)-6)\sqrt{(3+\alpha)(6+\alpha)}}{2\sqrt{3}\sqrt{1+\alpha}}$ . Next, because  $\zeta(r^2) = \frac{2\alpha r^2}{1-r^2}$ , one finds  $\sigma_1 = \frac{4\alpha(\alpha+1)}{\alpha-1}$ ,  $\sigma_2 = 1 - \alpha^{-1}$  (hence  $\alpha > 1$  in the sequel) and

$$\begin{aligned}\sqrt{d_0}\mathbf{c}_0 &= -\frac{\sqrt{\alpha-1}}{3} \left\{ \frac{(\alpha+2)}{2} \sqrt{(\alpha+3)}, \sqrt{\alpha+5} \right\}, \\ \sqrt{d_1}\mathbf{c}_1 &= \frac{\sqrt{\alpha-1}}{2} \left\{ \frac{\sqrt[3]{\alpha+2}\sqrt{(\alpha+1)(\alpha+3)}}{6}, \frac{\sqrt{(\alpha+2)(\alpha+3)(\alpha+4)}}{3\sqrt{2}}, \frac{\sqrt{\alpha+6}}{\sqrt{3}} \right\}, \\ \sqrt{d_2}\mathbf{c}_2 &= -\frac{\sqrt{\alpha-1}}{2\sqrt{2}} \left\{ \frac{(3+\alpha)\sqrt{\alpha+1}}{\sqrt{3}}, \sqrt{5+\alpha} \right\}.\end{aligned}$$

Continue as in the previous cases. Note that here  $\alpha$  is supposed known, but the above can serve to prefigure the difficulties to be encountered with an unknown shape parameter.

#### APPENDIX D: AN APPLICATION

As an application of the methods of the paper, consider the ‘‘Open-book closed-book examination’’ data set [34, p. 3-4] corresponding to examination marks in Mechanics, Vectors, Algebra, Analysis and Statistic for a sample of 88 students. Here we consider only the marks in Vectors, Algebra and Statistics, so that  $m = 3$  and we wish to test a trivariate MVN. We apply

Test Statistics	Value	d.f.	p-value $\chi^2$	p-value MC
$\mathcal{Q}_5$	98.62	46	0.004	0.027
$\mathcal{U}_5^{(s)}$	37.38	27	0.088	0.163
$\mathcal{I}_5^{(s)}$	33.14	18	0.016	0.028
$\mathcal{R}_5^{(s)}$	0.77	1	0.381	0.466

TABLE 5

*Examination marks ( $n = 88$ ) in Vectors, Algebra and Statistics from the “Open book-Closed book examination” data set [34, p. 3-4]. The null hypothesis is a trivariate normal distribution; d.f. refers to degrees of freedom of the asymptotic  $\chi^2$  approximation, p-value  $\chi^2$  are computed from these reference distributions, while p-value MC refers to the p-value computed from a Monte Carlo approximation (20 000 replications)*

our test strategy with  $K = 5$ . All polynomials required for these computations appear in Table 4. The results are shown in Table 5 which lists test statistic  $\mathcal{Q}_5$  and its scaled  $UIR$  components, as well as the  $p$ -values obtained from the reference asymptotic  $\chi^2$  distribution and from Monte Carlo (20 000 replications) approximations. The null hypothesis of trivariate MVN is rejected at the 5% level. Inspection of the scaled components shows that both the distributions of  $R$  and  $\mathbf{U}$  appear consonant with the MVN and that rejection comes from correlations between  $R$  and  $\mathbf{U}$ . If the confidence build in this section for  $m = 2$  can be transferred to the case  $m = 3$ , a refined model should try to take into account the dependencies between these random quantities.

IMAG, UNIV. MONTPELLIER, CNRS, MONTPELLIER, FRANCE  
 SCHOOL OF MATHEMATICS AND STATISTICS, UNSW SYDNEY, NSW 2052 AUSTRALIA  
 E-MAIL: [gilles.ducharme@umontpellier.fr](mailto:gilles.ducharme@umontpellier.fr)  
[lafaye@unsw.edu.au](mailto:lafaye@unsw.edu.au)