cemfi

working paper
1709

Testing Distributional Assumptions Using a Continuum of Moments

Dante Amengual
Marine Carrasco
Enrique Sentana

March 2017

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 1709
March 2017

Testing Distributional Assumptions Using a Continuum of Moments

Abstract

We propose specification tests for parametric distributions that compare theoretical and empirical characteristic functions. Our proposal is the continuum of moment conditions analogue to the usual overidentifying restrictions test, which takes into account the correlation between influence functions for different argument values. We derive its asymptotic distribution for fixed regularization parameter and when this vanishes with the sample size. We show its consistency against any deviation from the null, study its local power and compare it with existing tests. An extensive Monte Carlo exercise confirms that our proposed tests display good power in finite samples against a variety of alternatives.

JEL Codes: C01, C12, C52.

Keywords: Consistent tests, characteristic function, GMM, continuum of moment conditions, goodness-of-fit, Tikhonov regularization.

Dante Amengual
CEMFI
amengual@cemfi.es

Marine Carrasco
Université de Montréal
marine.carrasco@umontreal.ca

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

We are grateful to the participants at CIREQ conference in honor of Jean-Marie Dufour (Montréal) for helpful comments, discussions and suggestions. We also thank David Benatia, Alex Heinemann, Ran Liu and Siqi Wei for their able research assistance at various stages of this project. We gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness through grant ECO 2014-59262 (Amengual and Sentana) as well as from the Canadian Social Sciences and Humanities Research Council (Carrasco). Of course, the usual caveat applies.

1 Introduction

Goodness-of-fit tests are important to assess whether a parametric distribution provides an appropriate representation of the data. These tests can be divided in two main categories: (i) directional tests, which are designed to have power against specific alternatives, such as Neyman smooth test (Neyman, 1937 and Rayner and Best, 1989), Jarque and Bera's (1980) test of normality, as well as those proposed by Sefton (1992), Fiorentini, Sentana and Calzolari (2003), Bontemps and Meddahi (2005, 2012), Mencía and Sentana (2012) and Tuvaandorj and Zinde-Walsh (2014) among many others; (ii) omnibus tests, which are consistent against any alternative to the null hypothesis, for instance the integrated conditional moment test of Bierens (1982) and Bierens and Ploberger (1997), the conditional Kolmogorov test of Andrews (1997), and the copula goodness-of-fit test of Genest, Huang and Dufour (2013). Our proposed tests fall in this second category.

In particular, our testing procedure is based on the difference between the empirical and theoretical characteristic functions (CF) for all possible values of their argument. This gives rise to a continuum of moments in a L^2 space. Our aim is to construct a J test for overidentifying restrictions based on these moments, as in Hansen (1982). However, what plays the role of the covariance matrix in his test becomes now a covariance operator, whose inverse is unbounded. Therefore, some regularization is needed to stabilize the inverse. We propose to use Tikhonov regularization (see Kress, 1999) and consider two types of tests. The first one uses a fixed value of the regularization parameter α. Given that α can be regarded as a bandwidth, this approach is analogous to the fixed b asymptotics used in Kiefer and Vogelsang (2002). The second type of tests allows α to converge to zero at an appropriate rate, in which case our proposed test is closer in spirit to Hansen (1982)'s J test. In this second instance, however, the statistics would tend to a diverging χ^2 with infinite degrees of freedom. For that reason, we center and rescale it following the procedure put forward by Carrasco and Florens (2000), who presented this type of test for the first time.

We will consider various versions of our proposed tests depending on whether the parameter vector θ is known in advance or replaced by a consistent estimator, and whether we make use of the analytical expression for the covariance operator or estimate it. We will derive the asymptotic distribution of our tests under the null hypothesis and under local alternatives. We will also characterize the alternatives for which our tests have maximum power.

The advantages of using the CF are multiple: (a) in some important examples, the distribution function is only known in integral form whereas the CF has a closed form expression, as in the cases of stable distributions and affine diffusions (see Singleton (2001) and Carrasco, Chernov, Florens, and Ghysels (2007)); (b) handling multivariate random variables can be done just as easily as the scalar case; (c) our tests have the same form and are computed in the same manner for any CF tested; (d) our tests are consistent against any alternative to the null hypothesis.

Various tests based on the empirical CF have been previously proposed: Feuerverger and Mureika (1977), Epps and Pulley (1983), Hall and Welsh (1983), Baringhaus and Henze (1988), Ghosh and Ruymgaart (1992), Fan (1997), Hong (1999), Su and White (2007), Chen and Hong (2010), Bierens and Wang (2012) and Leucht (2012) among others. The main difference with ours is that we not only consider a continuum of moments, but we also explicitly take into account the correlation between the empirical CF for different values of its argument. Our work is also related to Dufour and Valery (2016), who propose a regularized Wald test to deal with the singularity of the covariance matrix.

The remainder of the paper is organized as follows. We introduce the tests in Section 2 and derive the asymptotic properties of the J test with fixed regularization parameter α and known (unknown) θ in Section 3 (4). Next, we study the J test with vanishing α in Section 5. Finally, Section 6 presents the results of our Monte Carlo simulations while Section 7 concludes. All the proofs are collected in Appendix A and computational aspects are discussed in Appendix B.

2 Presentation of the tests and overview

Assume we observe a sample of random variables $X_1, X_2, ..., X_n$ independent and identically distributed (*iid*) taking their values on $\mathbb{R}^q$ with $q \geq 1$. The X_j have probability density function (pdf) $f(x;\theta)$ indexed by a finite dimensional parameter θ, which may be known or unknown, and CF $\psi(t;\theta) = E[e^{itX}]$, where $t \in \mathbb{R}^q$ is its argument. As is well known, $f(x;\theta)$ and $\psi(t;\theta)$ are intimately related because the former is the Fourier transform of the latter, i.e.

$$\psi(t;\theta) = \int e^{itx} f(x;\theta)\, dx. \tag{1}$$

Figure 1 presents the CFs for the univariate distributions that we consider in our Monte Carlo study, namely, a standard normal, as well as standardized (zero mean - unit variance) versions of the uniform and $\chi^2(2)$ distributions. Given that the first two examples are symmetrically distributed around 0, the CF is real and symmetric around 0 too. In contrast, it contains an (odd) imaginary component in the case of the asymmetric chi-square.

We are interested in testing $H_0 : \psi = \psi_0(.;\theta_0)$, where ψ_0 is a known CF and θ_0 is some element of $\Theta \subset \mathbb{R}^p$. Our testing procedures are based on the difference between the empirical and theoretical CFs. Specifically, the relevant influence functions are

$$\hat{h}(t;\theta) = \frac{1}{n}\sum_{j=1}^{n} h_j(t;\theta), \tag{2}$$

$$h_j(t;\theta) = e^{itX_j} - \psi_0(t;\theta). \tag{3}$$

This gives rise to a continuum of moments since under the null $E[h_j(t;\theta_0)] = 0$ for all $t \in \mathbb{R}^q$.

Let π be an arbitrary probability density function on $\mathbb{R}^q$. Then, the function $h_j(t;\theta)$ is a random element of $L^2(\pi)$, the space of complex-valued functions which are square integrable with respect to the density π. The inner product on this space is defined for any functions f and g of $L^2(\pi)$ as $\langle f, g\rangle = \int f(t)\overline{g(t)}\pi(t)\,dt$, where the bar denotes the complex conjugate. $L^2(\pi)$ is a Hilbert space and we will work on this space to derive the asymptotic distribution of our test statistics.

By the central limit theorem of *iid* random elements of a separable Hilbert space (see e.g. Example 1.8.5 of van der Vaart and Wellner (1996)), we have that under H_0, as n goes to infinity

$$\sqrt{n}\hat{h}(t;\theta_0) \Rightarrow \mathcal{N}(0,K)$$

in $L^2(\pi)$, where $\mathcal{N}(0,K)$ denotes a Gaussian process of $L^2(\pi)$. K is an integral operator from $L^2(\pi)$ to $L^2(\pi)$ such that

$$(Kf)(s) = \int k(s,t)\, f(t)\, \pi(t)\, dt, \tag{4}$$

where

$$k_0(s,t) = E[h_j(s;\theta_0)\overline{h_j(s;\theta_0)}] = \psi_0(s-t;\theta_0) - \psi_0(s;\theta_0)\,\psi_0(-t;\theta_0). \tag{5}$$

In the sequel, we denote by λ_j and ϕ_j the eigenvalues and eigenfunctions of K, respectively, which are solutions to the functional equation $(K\phi_j)(t) = \lambda_j\phi_j(t)$. Figures 2a and 2c present the eigenfunctions associated with the largest two eigenvalues for the covariance operator K for the standard normal when the weighting function π is itself a normal with zero mean and scale parameter ω for two values of ω. In turn, Figures 2b and 2d do the same but for the corresponding operator of the standardised uniform distribution on $(-\sqrt{3}, \sqrt{3})$. As can be seen in these figures, if we arrange the eigenvalues in decreasing order, the eigenfunctions associated to even (odd) eigenvalues are even (odd) functions in these two examples. We also report in Figures 2e and 2f the largest five eigenvalues for those distributions. As we shall see below, the main effect of changing ω will be to change the relative weights given to small and large values of the CF argument t.

We are interested in applying Hansen (1982)'s J test of overidentifying restrictions to our continuum of moments. To illustrate the difficulties that may arise, assume for a moment that $\hat{\underline{h}}(\theta)$ is a finite dimensional M-vector obtained from a rough discretization of $\mathbb{R}^q$, so that $\sqrt{n}\hat{\underline{h}}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{K})$ and $\mathcal{K}$ is a nonsingular $M \times M$ matrix. Assuming for simplicity that both $\mathcal{K}$ and θ are known, the usual J test for overidentifying restrictions is

$$J = n\hat{\underline{h}}^{\star}(\theta)\mathcal{K}^{-1}\hat{\underline{h}}(\theta), \tag{6}$$

where * denotes the complex conjugate transpose of a vector/matrix. Now if we let M grow by taking a denser and denser grid, then the matrix $\mathcal{K}$ becomes increasingly ill-conditioned, in the sense that the ratio of its largest eigenvalue to its smallest one increases dramatically, so $\mathcal{K}^{-1}$ may be numerically unreliable for large M.

In our setting, the covariance matrix $\mathcal{K}$ is replaced by the aforementioned covariance operator K (see Appendix B.1), which has a countable infinite number of eigenvalues λ_j, $j = 1, 2, \ldots$ (arranged in decreasing order) and associated eigenfunctions ϕ_j. As we will see later on, this operator is compact, meaning that its inverse is not bounded. Consequently, its smallest eigenvalues will converge to zero as j goes to infinity, so taking the inverse of K is problematic. In terms of the spectral decomposition of K, the direct analogue to the J test statistic in (6) would be written as

$$\left\langle \sqrt{n}\hat{h}, K^{-1}\sqrt{n}\hat{h} \right\rangle = \sum_j \frac{1}{\lambda_j} \left\langle \sqrt{n}\hat{h}, \phi_j \right\rangle^2 \tag{7}$$

where the dependence on θ is omitted for simplicity. This expression will blow up because of the division by the small eigenvalues λ_j for large j. This is related to the problem of solving an integral equation $Kf = g$ where g is known and f is the object of interest. This problem is said to be ill-posed because f is not continuous in g. Indeed, a small perturbation in g will result in a large change in f. To stabilize the solution, one needs to use some regularization scheme (see Kress (1999) and Carrasco, Florens, and Renault (2007) for various possibilities). As in Carrasco and Florens (2000), we use Tikhonov regularization, which consists in replacing $K^{-1}g$ by the regularized solution $\left(K^2 + \alpha I\right)^{-1} Kg$ where $\alpha \geq 0$ is a regularization parameter. We use the notation $(K^\alpha)^{-1}$ for $\left(K^2 + \alpha I\right)^{-1} K$, which is the operator with eigenvalues $\frac{\lambda_j}{\lambda_j^2+\alpha}$ and corresponding eigenfunctions ϕ_j, and $(K^\alpha)^{-1/2}$ for the operator with eigenvalues $\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2+\alpha}}$ and the same eigenfunctions.

Thus, the regularized version of the J test is

$$\left\| (K^\alpha)^{-1/2} \sqrt{n}\hat{h} \right\|^2 = \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha} \left\langle \sqrt{n}\hat{h}, \phi_j \right\rangle^2. \tag{8}$$

Comparing the expressions (7) and (8), we observe that $\frac{1}{\lambda_j}$ has been replaced by $\frac{\lambda_j}{\lambda_j^2+\alpha}$, which is bounded.

We will consider various versions of this test depending on whether:

- θ is known or estimated,
- K is known or estimated,
- α is fixed or goes to zero.

Consider the case where α is fixed; if we are willing to assume that θ is known, so that the distribution under the null hypothesis is completely specified and the operator K is known, then the first test we should consider is

$$J(\theta_0, K) = \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha} \left\langle \sqrt{n}\hat{h}, \phi_j \right\rangle^2 . \tag{9}$$

As we explain in Appendix B.1, the test statistic (9) can be arbitrarily approximated from a numerical point of view by a modified version of the matrix expression (6). Specifically, if we evaluated the CF at a very fine but discrete grid of M points over a finite range of values of the argument t, then

$$J(\theta_0, K) = n\underline{\hat{h}}(\theta)^{\star} \left(\frac{\mathcal{K}}{M}\right)^{1/2} \left[\left(\frac{\mathcal{K}}{M}\right)^2 + \alpha I\right]^{-1} \left(\frac{\mathcal{K}}{M}\right)^{1/2} \underline{\hat{h}}(\theta) . \tag{10}$$

Several issues related to the practical implementation of this test (in particular the computation of the eigenelements of K) are discussed in Appendix B.1.

When θ is unknown, however, the operator K is only known up to θ. Let $\tilde{\theta}$ be a consistent estimator of θ obtained for instance from

$$\tilde{\theta} = \arg\min_{\theta\in\Theta} \left\| \hat{h}(.;\theta) \right\|^2 .$$

In this context, the integral operator $K_{\tilde{\theta}}$ can be defined as in (4) but with kernel

$$k(s,t) = \psi_0(s-t;\tilde{\theta}) - \psi_0(s;\tilde{\theta})\psi_0(-t;\tilde{\theta}).$$

Let $\{\lambda_{j\tilde{\theta}}, \phi_{j\tilde{\theta}}\}$ $j = 1, ..., M$ be the eigenvalues and eigenfunctions of the operator $K_{\tilde{\theta}}$. Then the second test we consider is

$$J(\hat{\theta}, K_{\hat{\theta}}) = \sum_j \frac{\lambda_{j\tilde{\theta}}}{\lambda_{j\tilde{\theta}}^2 + \alpha} \left\langle \sqrt{n}\hat{h}(.;\hat{\theta}), \phi_{j\tilde{\theta}} \right\rangle^2 = \arg\min_{\theta\in\Theta} \sum_j \frac{\lambda_{j\tilde{\theta}}}{\lambda_{j\tilde{\theta}}^2 + \alpha} \left\langle \sqrt{n}\hat{h}(.;\theta), \phi_{j\tilde{\theta}} \right\rangle^2 .$$

Alternatively, we may prefer to estimate K using a sample covariance operator. In fact, there are two obvious possibilities. The first one is to use the integral estimator $\hat{K}$ with uncentred kernel

$$\hat{k}(s,t) = \frac{1}{n}\sum_{i=1}^{n} h_i(s;\tilde{\theta})h_i(-t;\tilde{\theta}),$$

where $\tilde{\theta}$ is a consistent first step estimator of θ. On the other hand, the second possibility is the integral operator $\mathring{K}$ with centred kernel

$$\mathring{k}\,(s,t) = \frac{1}{n}\sum_{i=1}^{n} h_i\,(s)\,h_i\,(-t)\,,$$

where

$$h_i\,(s) = h_i\,(s;\theta) - \hat{h}\,(t;\theta) = e^{itX_i} - \frac{1}{n}\sum_{l=1}^{n} e^{itX_l}.$$

The advantage of the second estimator is that it does not require a first step estimator of θ and thereby it may be more robust to misspecification. Either way, given that $\hat{K}$ and $\mathring{K}$ have finite range, they will have at most n nonzero eigenvalues, which (under some conditions) will be consistent estimators of the largest eigenvalues of K.

For computational reasons, it is convenient to rewrite the test statistics (8), which use as eigenvalues and eigenfunctions those of $\hat{K}$ and $\mathring{K}$, in terms of certain matrices and vectors (see Carrasco et al (2007) for analogous expressions for $\hat{K}$ under time series dependence). Specifically, we obtain the following two expressions:

i) The test based on $\hat{K}$, which can be computed as

$$J(\theta,\hat{K}_{\tilde{\theta}}) = \arg\min_{\theta\in\Theta} \underline{v}\,(\theta)^{\star}\left[\alpha I + C^2\right]^{-1}\underline{v}\,(\theta) \tag{11}$$

where $\underline{v}\,(\theta)$ is a $n\times 1$ vector with l-th element $v_l\,(\theta) = \int \overline{h_l(t;\tilde{\theta})}\hat{h}\,(t;\theta)\,\pi\,(t)\,dt$, C is an $n\times n$ matrix with (i,l) element c_{il}/n with $c_{il} = \left\langle h_l(t;\tilde{\theta}), h_i(t;\tilde{\theta})\right\rangle$ (see Appendix B.2 for analytical expressions for these integrals).

ii) The test based on $\mathring{K}$, whose matrix expression is

$$J(\theta,\mathring{K}) = \arg\min_{\theta\in\Theta} \underline{\mathring{v}}\,(\theta)^{\star}\,[\alpha I + \mathring{C}^2]^{-1}\underline{\mathring{v}}\,(\theta) \tag{12}$$

where $\underline{\mathring{v}}\,(\theta)$ is a $n\times 1$ vector with l-th element $\mathring{v}_l\,(\theta) = \int \overline{h_l\,(t)}\hat{h}\,(t;\theta)\,\pi\,(t)\,dt$, $\mathring{C}$ is an $n\times n$ matrix with (i,l) element $\mathring{c}_{il}/n$ with $\mathring{c}_{il} = \langle h_l\,(t)\,, h_i\,(t)\rangle$. In this regard, note that $\mathring{C} = (I-\ell\ell'/n)C(I-\ell\ell'/n)$, where ℓ is a vector of n ones.

In Sections 3 and 4, we will study the asymptotic distribution of the test statistics $J(\theta_0,K)$, $J(\hat{\theta},K_{\hat{\theta}})$, $J(\theta,\hat{K}_{\tilde{\theta}})$ and $J(\theta,\mathring{K})$ and show that they converge under H_0 to a weighted sum of χ^2's whose weights depend on θ. Given the eigenvalues, those weights and hence their asymptotic distributions are known, so we can compute the p-value of these quadratic forms in normal variables using Imhof (1961). Nevertheless, we rely on the parametric bootstrap in the simulations to improve the small sample properties of our proposed procedures.

For all the tests presented so far, α is fixed, so that our regularized inverse $(K^{\alpha})^{-1}$ is a biased approximation of K^{-1}. It is possible to approach K^{-1} by letting α go to zero at a suitable rate. However, a test based on (8) with α going to zero would tend to a chi-square with infinite degrees of freedom, and hence diverge. For that reason, we explain next how to center and rescale it following Carrasco and Florens (2000). Let $h_j(t;\hat{\theta})$ denote the influence function (3) evaluated at a consistent estimator of θ. Similarly, let $\hat{\lambda}_j$ denote the eigenvalues of $\hat{K}$,

$$\hat{a}_j = \frac{\hat{\lambda}_j^2}{\hat{\lambda}_j^2 + \alpha}, \quad \hat{p}_n = \sum_{j=1}^{n} \hat{a}_j \quad \text{and} \quad \hat{q}_n = 2\sum_{j=1}^{n} \hat{a}_j^2. \tag{13}$$

After appropriate centering and rescaling, we obtain:

$$J_{\alpha_n}(\hat{\theta}, K_{\hat{\theta}}) = \frac{\left\| (\hat{K}^{\alpha_n})^{-1/2} \sqrt{n} \hat{h}(.;\hat{\theta}) \right\|^2 - \hat{p}_n}{\sqrt{\hat{q}_n}}. \tag{14}$$

In Section 5, we show that J_{α_n} converges to a standard normal distribution under the null.

3 J test when α is fixed and the parameter is known

3.1 Distribution under local alternatives

The $J(\theta_0, K)$ statistic in (9) with α fixed is part of a larger class of tests based on weighted L^2 statistics that we will denote by T_B in the sequel. Let B be a nonrandom bounded linear operator from $L^2(\pi)$ to $L^2(\pi)$ and B_n be a sequence of random bounded linear operators from $L^2(\pi)$ to $L^2(\pi)$ such that $\|B_n - B\| \xrightarrow{P} 0$ as n goes to infinity, where $\|.\|$ is the sup-norm. Assume moreover that the null space of B equals $\{0\}$; otherwise the test would lack power against certain alternatives. Popular choices of B satisfying our assumptions include $B = I$ as in Epps and Pulley (1983), Bierens and Wang (2012) and Leucht (2012), as well as $B = (K^{\alpha})^{-1/2}$ with $\alpha > 0$ fixed.

In this section and the next one we focus on tests based on weighted L^2 statistics

$$T_B = \left\| B_n \sqrt{n} \hat{h} \right\|^2 = \int (B_n \sqrt{n} \hat{h})^2(t)\, \pi(t)\, dt, \tag{15}$$

where $\hat{h}(t) = \sum_{j=1}^{n} [e^{itX_j} - \psi_0(t)]$ and $\psi_0(t) = \psi_0(t;\theta_0)$.

We look at local alternatives of the form

$$H_{1n} : \psi_n = \psi_0 + \frac{c\eta}{\sqrt{n}}, \tag{16}$$

where c is a scalar. In this context, η represents the direction of the alternative, while c represents the distance from the null. To guarantee the uniqueness of the representation, η needs to be normalized. There are many possibilities. As the results of this subsection are not affected by the choice of the normalization, we will not specify a normalization at this stage. Nevertheless, we impose $\eta(0) = 0$ to preserve the property that $\psi_n(0) = 1$ which every CF needs to satisfy. Similarly, we also need $\overline{\eta(t)} = \eta(-t)$ to preserve the property that $\overline{\psi_n(t)} = \psi_n(-t)$ for any CF. Assume moreover that there is a constant $\mathfrak{C} > 0$ such that $|\eta(t)| \leq \mathfrak{C}$ for all t, and $0 < \|B\eta\| < \infty$. CFs need to satisfy the condition $|\psi_n| \leq 1$, which hopefully will be satisfied by ψ_n under H_{1n} for n sufficiently large.

First, we establish some results on the operator K of form (4) with kernel (5), suppressing the dependence on θ_0 for simplicity.

Lemma 1 *K is a self-adjoint positive definite Hilbert-Schmidt operator from $L^2(\pi)$ to $L^2(\pi)$ and the sum of its eigenvalues is bounded by 1.*

Lemma 1 implies two things: that K has a countable spectrum, and that the sum of its eigenvalues is less than 1.

Example. Consider the CF of a univariate normal with mean μ and variance σ^2; it turns out that when using a normal weighting function with zero mean and scale parameter ω, we can obtain analytical solutions for the sums of both λ's and λ^2's. Specifically, the expressions are

$$\sum_j \lambda_j = 1 - \frac{1}{\sqrt{1 + 2\sigma^2\omega^2}}$$

and

$$\sum_j \lambda_j^2 = \frac{1}{\omega + 2\sigma^2\omega^3} \left(\frac{e^{-\frac{4\mu^2\omega^2}{1+2\sigma^2\omega^2}}}{\sqrt{1+2\sigma^2\omega^2}} + \frac{e^{-\frac{4\mu^2\omega^2}{1+4\sigma^2\omega^2}}}{\sqrt{1+4\sigma^2\omega^2}} - \frac{2}{\sqrt{1+4\sigma^2\omega^2+3\sigma^4\omega^4}} \right).$$

As can be seen from the above expressions, the sums of both λ's and λ^2's depend on the scale ω of the weighting function.

Proposition 2 *Under H_{1n}, as n goes to infinity*

$$\sqrt{n}\widehat{h} \Rightarrow \mathcal{N}(c\eta, K)$$

in $L^2(\pi)$.

Let a_j, ϕ_j, $j = 1, 2..., J$ be the eigenvalues (arranged in decreasing order) and eigenvectors of BKB^*. Further, let $\delta_j = c^2 \left\langle B\eta, \phi_j \right\rangle^2 / a_j$.

Proposition 3 *Under H_{1n}, we have*

$$T_B \xrightarrow{d} \sum_{j=1}^{\infty} a_j \chi_j^2 (1, \delta_j) = \sum_{j=1}^{\infty} a_j \left(e_j + \frac{c \left\langle B\eta, \phi_j \right\rangle}{\sqrt{a_j}} \right)^2$$

where $\chi_j^2 (1, \delta_j)$, $j = 1, 2, ...$ denote independent noncentral chi-square random variables with 1 degree of freedom and non centrality parameter δ_j while e_j, $j = 1, 2, ...$ are the underlying independent standard normal variables.

Remark 1. The previous proposition will not apply if $B = K^{-1/2}$. In that case B is not bounded violating one of the assumptions. Moreover, $\mathcal{N}(0, I)$ is not a Gaussian process because the trace of its covariance operator (the identity operator) is infinite. We will discuss the case $B = (K^\alpha)^{-1/2}$ when α goes to zero in Section 5.

Remark 2. We see that as soon as $\left\langle B\eta, \phi_j \right\rangle \neq 0$ for some j, the test statistic T_B will have non trivial power. But because $\{\phi_j\}$ forms an orthonormal basis of $L^2(\pi)$, then $B\eta = \sum_j \left\langle B\eta, \phi_j \right\rangle \phi_j$ and by Parseval's identity, $\|B\eta\|^2 = \sum_j \left\langle B\eta, \phi_j \right\rangle^2 > 0$. It follows that $\left\langle B\eta, \phi_j \right\rangle$ cannot all be zero simultaneously. Therefore, T_B has indeed non trivial power against all local alternatives of the form H_{1n}, and against all fixed alternatives *a fortiori*. However, if $\left\langle B\eta, \phi_j \right\rangle^2$ is small (as will be the case for most j since the sequence $\left\langle B\eta, \phi_j \right\rangle^2$ is summable), the power against local alternatives in the jth direction may be poor. In the next subsection, we will study the power properties of these tests in more detail.

3.2 Alternatives with maximum power

It is well-known that there is no uniformly most powerful test for assessing H_0 and that goodness-of-fit tests have good power only against certain local alternatives (see Neuhaus (1976), Janssen (2000), Escanciano (2009), and Lehmann and Romano (2005, Section 14.6)). In this subsection, we will characterize the alternative with maximum power.

Given that there is a one-to-one mapping between the density and the CF through the Fourier inversion theorem (see (1)), we can reformulate H_0 and H_{1n} in terms of the density instead. Thus, we obtain

$\tilde{H}_0 : f(x) = f_0(x)$,

$\tilde{H}_{un}(c) : f_n(x) = f_0(x) \left[1 + \frac{cu(x)}{\sqrt{n}} \right]$.

Let $L^2(f_0) < \infty$ denote the L^2 space of real functions $\varphi(X)$ such that we can define $\|\varphi\|^2_{L^2(f_0)} = \int \varphi^2(x) f_0(x) dx$. Note that c in $\tilde{H}_{un}(c)$ is the same as c in H_{1n} and $u(x)$ defined in $\tilde{H}_{un}(c)$ is related to η defined in H_{1n} through the relations

$$\begin{aligned} u(x) &= \frac{\frac{1}{2\pi}\int e^{-itx}\eta(t)\,dt}{f_0(x)}, \\ \eta(t) &= \int e^{itx}u(x)\,f_0(x)\,dx. \end{aligned}$$

Moreover, the condition $\eta(0)=0$ implies $\int u(x)\,f_0(x)\,dx=0$. Still, u needs to be normalized. Many normalizations could be used. For convenience, we impose the normalization condition $\|u\|_{L^2(f_0)} = E[u^2(X)] = 1$, which corresponds to the following condition on η:

$$\int \left| \int e^{-itx}\eta(t)\,dt \right|^2 \frac{1}{f_0(x)}dx = 1.$$

In this set up, we define the asymptotic local power function $\Pi_B(a,c,u)$ as

$$\Pi_B(a,c,u) = \lim_{n\to\infty} P[T_B \geq c_a | \tilde{H}_{un}(c)],$$

where c_a is the critical value such that T_B achieves a level a, i.e. $\lim_{n\to\infty} P(T_B \geq c_a | H_0) = a$.

To analyze the power of these statistics, it is useful to rewrite K as T^*T, where T is an operator from $L^2(\pi)$ to $L^2(f_0)$ and T^* is the adjoint operator from $L^2(f_0)$ to $L^2(\pi)$. Such a decomposition has been used to study the power of Cramer von Mises type tests by Neuhaus (1976, equation (1.9)) and Escanciano (2009, p.168).

The operators T and T^* are as follows:

$T: L^2(\pi) \to L^2(f_0)$,

$(T\varphi)(X) = \int \overline{h(X;t)}\varphi(t)\,\pi(t)\,dt$,

$T^*: L^2(f_0) \to L^2(\pi)$, and

$(T^*\phi)(t) = \int h(x;t)\,\phi(x)\,f_0(x)\,dx$.

Moreover, TB^* is compact and admits a singular system $\{\sqrt{a_j},\phi_j,\varphi_j\}$, where $TB^*\phi_j = \sqrt{a_j}\varphi_j$ and $BT^*\varphi_j = \sqrt{a_j}\phi_j$. Therefore, ϕ_j are the eigenfunctions of $BT^*TB^* = BKB^*$ and φ_j those of B^*TT^*B. Thus, φ_j can be interpreted as principal components weights.

Observe that $\eta = T^*u$. Indeed, if we use the property of Fourier transforms and the fact that $\int u(x)\,f_0(x)\,dx = 0$, we will have that

$$\begin{aligned} (T^*u)(t) &= \int [e^{itx} - \psi_0(t)]u(x)\,f_0(x)\,dx \\ &= \int e^{itx}u(x)\,f_0(x)\,dx - \psi_0(t)\int u(x)\,f_0(x)\,dx = \eta(t), \end{aligned}$$

Hence, the relation $\eta = T^*u$ implies that

$$\frac{\langle B\eta, \phi_j \rangle}{\sqrt{a_j}} = \frac{\langle BT^* u, \phi_j \rangle}{\sqrt{a_j}} = \frac{\langle u, TB^* \phi_j \rangle}{\sqrt{a_j}} = \langle u, \varphi_j \rangle . \tag{17}$$

From (17) and Proposition 2, it follows that under $\tilde{H}_{un}(c)$,

$$T_B \xrightarrow{d} \sum_{j=1}^{\infty} a_j (e_j + c \langle u, \varphi_j \rangle)^2. \tag{18}$$

Note that the sequence φ_j $j = 1, 2, ...$ forms a complete orthonormal basis of $\overline{\mathcal{R}(TB^*)} = L^2(f_0) \cap \{u : E(u) = 0\}$. Hence, the alternatives of interest are linear combinations of the eigenfunctions φ_j. In this context, the analysis of the limiting distribution in (18) and the orthogonality of the $\varphi_j' s$ allow us to establish the following results:

Proposition 4 *The limiting power of $\Pi_B(\alpha, c, u)$ has the following properties.*

(a) $\{\Pi_B(a, c, u) : u \in L^2(f_0), E(u) = 0, \|u\|_{L^2(f_0)} = 1\} = \Pi_B(a, c, \varphi_1)$,

(b) $\Pi_B(a, c, \varphi_j) \leq \Pi_B(a, c, \varphi_i)$ *for* $1 \leq j \leq i$,

(c) $lim_{j\to\infty}\Pi_B(a, c, \varphi_j) = a$.

Proposition 4 says that (a) the maximum power is achieved for the local alternative $u = \varphi_1$ corresponding to the first principal component, (b) the power decreases when considering higher-order principal components, (c) finally, the power goes down to the level of the test, a, for the highest frequency (case $j \to \infty$).

As we saw before, in general φ_j depends on B, so that the alternative with maximum power will be different for different tests T_B.

But if we consider more specifically the cases $B = I$ and $B = (K^\alpha)^{-1/2}$, the φ_j are the same because they correspond to the eigenfunctions of TT^*. Hence, the alternative for which the tests T_B for $B = I$ and $B = (K^\alpha)^{-1/2}$ are the most powerful coincides, and corresponds to $\eta = \varphi_1$.

When $B = I$, then $a_j = \lambda_j$, i.e. the eigenvalues of K, which decline quickly towards 0. So the test T_B with $B = I$ concentrates its power on the first principal component. On the other hand, when $B = (K^\alpha)^{-1/2}$, $a_j = \frac{\lambda_j^2}{\lambda_j^2+\alpha}$ instead, which will decline slower towards 0 if α is relatively small. Consequently, power will be more balancedly spread among the first few directions when $B = (K^\alpha)^{-1/2}$ than when $B = I$. Figure 3 illustrates the decline of λ_j and a_j in the case of a uniform distribution. In the extreme case where $\alpha = 0$, we would have $a_j = 1$, which means that power would be evenly spread among all alternatives. However, in this case the null distribution is a Chi-square with infinite degrees of freedom and the resulting test has power equal to size for any local alternative; see Lemma 14.3.1 of Lehmann and Romano (2005). We will consider the case where $\alpha \to 0$ in greater detail in Section 5.

4 J test when α is fixed and the parameter is unknown

4.1 Distribution under local alternatives

Consider again local alternatives of the form (16), where c is a constant, $\eta \in L^2(\pi)$, $\eta(0) = 0$, $\overline{\eta(t)} = \eta(-t)$ and $|\eta(t)| < \mathfrak{C}$ for some constant $\mathfrak{C}$.

Assumption 1. X_i, $i = 1, 2, ...$ are independent, identically distributed.

Assumption 2. Under H_{1n}, $\|B_n - B\| \overset{P}{\to} 0$. Under H_1, $\|B_n - B_1\| \overset{P}{\to} 0$ where both B and B_1 are bounded linear operators and B_1 may differ from B. The null spaces of B and B_1 equal $\{0\}$.

In the sequel, we denote by P_0 the law of X_i under H_0, P_n the law of X_i under H_{1n}, and P_1 the law of X_i under H_1.

Assumption 3. P_n is contiguous to P_0.

This condition is standard in the goodness-of-fit literature and imposes some mild restriction on the density. Sufficient conditions for this assumption to be true are given in Lehmann and Romano (2005). They also provide a variety of examples.

Assumption 4. The parameter space Θ is a compact subset of $\mathbb{R}^p$. The true parameter θ_0 is contained in the interior of Θ. $\psi_0(\tau; \theta)$ is continuously differentiable with respect to θ.

Assumption 5 (identification). $\psi_0(\tau; \theta) = \psi_0(\tau; \theta_0)$ for all $\tau \Leftrightarrow \theta = \theta_0$.

Let

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\| B_n \hat{h}(.; \theta) \right\|$$

and define

$$D_0 = \left. \frac{\partial \psi_0(.; \theta)}{\partial \theta} \right|_{\theta = \theta_0}.$$

Note the result in Proposition 2 remains valid here. Namely, $\sqrt{n}\hat{h}(\theta_0) \Rightarrow \mathcal{N}(c\eta, K)$ under H_{1n}, where K is an integral operator with kernel $k(s,t) = \psi(s - t; \theta_0) - \psi(s; \theta_0)\psi(-t; \theta_0)$.

Then

Proposition 5 *Suppose Assumptions 1-5 hold. Under H_0, $\hat{\theta}$ is a consistent estimator of θ_0 and*

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} \mathcal{N}(0, \langle BD_0, BD_0\rangle^{-1} \langle BD_0, (BKB^*) BD_0\rangle \langle BD_0, BD_0\rangle^{-1}).$$

Moreover, under H_1,

$$\hat{\theta} \overset{P_1}{\to} \theta_1 = \arg\min_{\theta \in \Theta} \left\| B_1 E^{P_1}(h_j(.; \theta)) \right\|.$$

Let L be the operator from $L^2(\pi)$ to $L^2(\pi)$ such that for all $\varphi \in L^2(\pi)$

$$(L\varphi)(\tau) = \varphi(\tau) - D_0(\tau) \langle BD_0, BD_0\rangle^{-1} \langle B^* BD_0, \varphi\rangle.$$

Let $\tilde{K}$ be the integral operator from $L^2(\pi)$ to $L^2(\pi)$ with kernel

$$\begin{aligned}\tilde{k}(s,t) &= k(s,t) - D_0(s)\langle BD_0, BD_0\rangle^{-1}(KB^*BD_0)(t) \\ &\quad -\overline{D_0(t)}\langle BD_0, BD_0\rangle^{-1}\overline{(KB^*BD_0)}(s) \\ &\quad +D_0(s)\langle BD_0, BD_0\rangle^{-1}\langle BD_0, (BKB^*)BD_0\rangle\langle BD_0, BD_0\rangle^{-1}\overline{D_0(t)}'\end{aligned}$$

In addition, let $\tilde{a}_j$, $\tilde{\phi}_j$, $j = 1, 2..., J$ denote the eigenvalues (arranged in decreasing order) and eigenvectors of $B\tilde{K}B^*$. Finally, define $\tilde{\delta}_j = c^2\left\langle BL\eta, \tilde{\phi}_j\right\rangle^2/\tilde{a}_j$.

Proposition 6 *Suppose Assumptions 1-5 hold. Under H_{1n}, we have*

(i) $\sqrt{n}\hat{h}(\hat{\theta}) \Rightarrow \mathcal{N}(cL\eta, \tilde{K})$ *in* $L^2(\pi)$.

(ii)

$$T_B \xrightarrow{d} \sum_{j=1}^{\infty}\tilde{a}_j\chi_j^2(1,\tilde{\delta}_j) = \sum_{j=1}^{\infty}\tilde{a}_j\left(e_j + \frac{c\left\langle BL\eta, \tilde{\phi}_j\right\rangle}{\sqrt{\tilde{a}_j}}\right)^2$$

where $\chi_j^2(1,\tilde{\delta}_j)$, $j = 1, 2, ..., L$ denote independent noncentral chi-square r.v. with 1 degree of freedom and non centrality parameter $\tilde{\delta}_j$ and e_j, $j = 1, 2, ...$ are the underlying independent standard normal variables.

Proposition 6 implies that T_B has non trivial power against all local alternatives η for which $L\eta \neq 0$, i.e. those η such that $\eta \neq v'D_0$, where v is some $p \times 1$ vector of constants. The following example illustrates this condition:

Lemma 7 *Assume $H_0 : \psi = \psi_0$ where ψ_0 is the characteristic function of the $\mathcal{N}(\mu, \sigma^2)$. Let f_0 be the pdf of the $\mathcal{N}(\mu, \sigma^2)$. The test T_B has only trivial power against local alternatives of the form*

$$H_{1n} : \psi_n(\tau) = \left(1 + \frac{ai\tau}{\sqrt{n}} - \frac{b\tau^2}{2\sqrt{n}}\right)\psi_0(\tau)$$

for some constants a and b. Moreover the density corresponding to ψ_n is

$$f_n(x) = \left\{1 + \frac{a}{\sqrt{n}}\frac{(x-\mu)}{\sigma^2} + \frac{b}{2\sqrt{n}}\left[\frac{(x-\mu)^2 - \sigma^2}{2\sigma^4}\right]\right\}f_0(x). \tag{19}$$

It follows from Lemma 7 that when μ and σ^2 are estimated, the test T_B has trivial power against alternatives of the form (19), which correspond to a second order Hermite expansion of the Gaussian density. The two additive terms in (19) contain the first two Hermite polynomials, which will be close to zero once μ and σ^2 are estimated. This is similar to what is found in other tests. For example, Bontemps and Meddahi (2005)'s moment test of normality cannot make use of the first two Hermite polynomials evaluated at the estimated parameters because their sample means will converge to 0 in probability even after scaling them by $\sqrt{n}$.

The following result establishes that T_B has power against all fixed alternatives (including those such that $L\eta = 0$).

Proposition 8 *Suppose Assumptions 1-5 hold. The test T_B is consistent.*

In the next subsection, we analyze the power of our test in more detail.

4.2 Alternative with maximum power

We can follow the same steps as in Section 3.2 to characterize the alternatives for which the tests T_B have maximum power.

As mentioned earlier, the null and alternatives can be equivalently expressed in terms of either the characteristic function or the density of X_j. Consider the hypotheses $\tilde{H}_0$ and $\tilde{H}_{un}(c)$ as defined in Section 3.2., where f_0 denotes now $f_0(x;\theta_0)$. Let $L^2(f_0)$ and assume the same normalization of u and the same power function $\Pi_B(a,c,u)$ as before. Following Neuhaus (1976) and Escanciano (2009), we can determine for which local alternative the test T_B has maximum power.

Let $h(x;t) = e^{itx} - \psi_0(.;\theta_0)$. To analyze power, it is useful to rewrite the covariance operator $\tilde{K}$ as $\tilde{T}^*\tilde{T}$, where $\tilde{T}$ is an operator from $L^2(\pi)$ to $L^2(f_0)$ and $\tilde{T}^*$ is the operator from $L^2(f_0)$ to $L^2(\pi)$, $\tilde{T}^*$ being the adjoint of $\tilde{T}$. The operators $\tilde{T}$ and $\tilde{T}^*$ are as follows;

$\tilde{T} : L^2(\pi) \to L^2(f_0)$,

$\left(\tilde{T}\varphi\right)(X) = \int [\overline{h(X;t)} - \overline{D_0(t)} \langle BD_0, BD_0\rangle^{-1} \overline{\langle B^*BD_0, h(X;.)\rangle}]\varphi(t)\pi(t)\,dt$,

$\tilde{T}^* : L^2(f_0) \to L^2(\pi)$, and

$(T^*\phi)(t) = \int [h(x;t) - D_0(t)\langle BD_0, BD_0\rangle^{-1} \langle B^*BD_0, h(x;.)\rangle]\phi(x) f_0(x)\,dx$.

Moreover, $B\tilde{T}^*\tilde{T}B^* = B\tilde{K}B^*$ is compact and admits a singular system $\{\tilde{a}_j, \tilde{\phi}_j, \tilde{\varphi}_j\}$, where $\tilde{T}B_j^*\tilde{\phi} = \sqrt{\tilde{a}_j}\tilde{\varphi}_j$ and $\tilde{B}\tilde{T}^*\tilde{\varphi}_j = \sqrt{\tilde{a}_j}\tilde{\phi}_j$. $\tilde{\phi}_j$ are the eigenfunctions of $B\tilde{T}^*\tilde{T}B$ and $\tilde{\varphi}_j$ are the eigenfunctions of $B^*\tilde{T}\tilde{T}^*B$. $\tilde{\varphi}_j$, which can again be interpreted as principal components weights.

Observe that $L\eta = \tilde{T}^*u$. Hence,

$$\frac{\left\langle BL\eta, \tilde{\phi}_j\right\rangle}{\sqrt{\tilde{a}_j}} = \frac{\left\langle B\tilde{T}^*u, \tilde{\phi}_j\right\rangle}{\sqrt{\tilde{a}_j}} = \frac{\left\langle u, \tilde{T}B^*\tilde{\phi}_j\right\rangle}{\sqrt{\tilde{a}_j}} = \left\langle u, \tilde{\varphi}_j\right\rangle. \tag{20}$$

From (17) and Proposition 5, it follows that under $\tilde{H}_{un}(c)$,

$$T_B \xrightarrow{d} \sum_{j=1}^{\infty} \tilde{a}_j \left(e_j + c\left\langle u, \tilde{\varphi}_j\right\rangle\right)^2. \tag{21}$$

For those u such that $T^*u = 0$, the tests T_B have power equal to size. Therefore, we will focus on alternatives such that $T^*u \neq 0$, alternatives which belong to the orthogonal space to the null space of T^* (denoted $\mathcal{N}(T^*)$) – these are the alternatives corresponding to η such that $L\eta \neq 0$. For any compact operator T we have the relation, $\mathcal{N}(T^*)^\perp = \overline{\mathcal{R}(T)}$, where $\overline{\mathcal{R}(T)}$ is the closure of the range of T. Note that the sequence $\tilde{\varphi}_j$, for $j = 1, 2, ...$ form a complete orthonormal basis of $\overline{\mathcal{R}(T)}$. Hence, the alternatives of interest are linear combinations of the $\tilde{\varphi}_j$. The analysis of the limiting distribution in (21) and the orthogonality of the $\tilde{\varphi}_j$ allow us to establish an analogous result to Proposition 4:

Proposition 9 *Suppose Assumptions 1-5 hold. The limiting power of* $\Pi_B(\alpha, c, u)$ *has the following properties.*

(a) $\max_u\{\Pi_B(a, c, u) : u \in \overline{\mathcal{R}(T)},\ \|u\|_{L^2(f_0)} = 1\} = \Pi_B(a, c, \tilde{\varphi}_1)$,

(b) $\Pi_B(a, c, \tilde{\varphi}_j) \leq \Pi_B(a, c, \tilde{\varphi}_i)$ *for* $1 \leq j \leq i$,

(c) $\lim_{j\to\infty}\Pi_B(a, c, \tilde{\varphi}_j) = a$.

As before, we observe that the maximum power is reached for the first principal component, and that power declines toward size a for subsequent directions.

5 J test when α goes to zero

5.1 Distribution under local alternatives

As we discussed at the end of Section 2, the continuum of moments analogue to the overidentification restrictions test diverges when α goes to zero, so we need to center and re-scale this statistic appropriately as in (14). But because $\hat{q}_n$ in the denominator of this expression diverges as n goes to infinity, the rescaled test does not have power against contiguous alternatives. Therefore, we need to consider alternatives that converges to H_0 slower than the usual $n^{-1/2}$ rate. For that reason, in what follows we study the properties of $J_{\alpha_n}(\hat{\theta}, \hat{K})$ under local alternatives of the form

$$H_{2n} : \psi_n(t) = \psi_0(t; \theta_0) + b_n\eta(t)$$

where $\eta \in L^2(\pi)$, $\eta(0) = 0$, $\overline{\eta(t)} = \eta(-t)$, $|\eta(t)| < \mathfrak{C}$ for some constant $\mathfrak{C}$, and b_n is a sequence of numbers going to zero at a rate slower than $\sqrt{n}$. The precise rate will be specified later on. In the sequel, P_{2n} denotes the law of X_i under H_{2n}.

Assumption 6. Under H_{2n},

$$\sqrt{n}(\hat{\theta} - \theta_0) = H^{-1}\left\langle G(.), \sqrt{n}\hat{h}(.; \theta_0)\right\rangle + o_{P_{2n}}(1),$$

where H is a positive definite $p \times p$ matrix and G is a $p \times 1$ vector of $L^2(\pi)$.

The estimators mentioned in the previous sections satisfy this condition under H_0. In particular, in the case of the GMM estimator:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\| (\hat{K}^{\alpha})^{-1/2} \sqrt{n} \hat{h}(.; \hat{\theta}) \right\|^2,$$

the condition is satisfied under H_0 for $H = \langle K^{-1/2} D_0, K^{-1/2} D_0 \rangle = \mathcal{I}$ and $G = K^{-1} D_0$, where $D_0(t) = \partial \psi_0(t; \theta_0) / \partial \theta$ and $\mathcal{I}$ is the information matrix provided α goes to zero at a certain rate (see Carrasco and Florens (2000)). Similarly, the MLE satisfies also this assumption under H_0 with $H = \mathcal{I}$ and

$$G(t) = \frac{1}{\pi(t)} \frac{1}{2\pi} \int e^{itx} \frac{\partial \ln f_0(x)}{\partial \theta} dx.$$

To see this, check that

$$\langle e^{itx}, G(t) \rangle = \int e^{itx} \overline{G(t)} \pi(t) \, dt = \frac{\partial \ln f_0(x)}{\partial \theta}$$

and

$$\langle \psi_0(t; \theta_0), G(t) \rangle = 0.$$

Under H_{2n}, we have

$$\sqrt{n} \{ \hat{h}(.; \hat{\theta}) - E^{P_{2n}} [\hat{h}(.; \hat{\theta})] \} \Rightarrow \mathcal{N}(0, K_\omega)$$

in $L^2(\pi)$ where $K_\omega : L^2(\pi) \to L^2(\pi)$ such that

$$(K_\omega \varphi)(s) = \int k_\omega(s, t) \pi(t) \, dt$$

with

$$k_\omega(s,t) = E^{P_0}[(h_j(s; \theta_0) - D_0(s) H^{-1} \langle G, h_j(.; \theta_0) \rangle) (h_j(t; \theta_0) - D_0(t) H^{-1} \langle G, h_j(.; \theta_0) \rangle)^*].$$

Assumption 7. Under H_{2n}, $\left\| \hat{K} - K_\omega \right\| \to 0$ and $\left\| (\hat{K}^{\alpha})^{-1/2} - (K_\omega^{\alpha})^{-1/2} \right\| = O_{P_n}\left(\frac{1}{\sqrt{n} \alpha^{3/4}} \right)$.

The results stated in Assumption 7 are easy to establish under H_0 (see Carrasco and Florens (2000) and Carrasco et al (2007)).

Let $\{\lambda_{j\omega}, \phi_{j\omega}\}$ $i = 1, 2, ...$ be the eigenvalues and eigenfunctions of K_ω and $a_{j\omega} = \frac{\lambda_{j\omega}^2}{\lambda_{j\omega}^2 + \alpha}$. Let $p_n = \sum_{j=1}^n a_{j\omega}$, $q_n = \sum_{j=1}^n a_{j\omega}^2$, and

$$(L_\omega \eta)(s) = \eta(s) - D_0(s) H^{-1} \langle G, \eta \rangle.$$

Moreover, let $\mathcal{H}_K$ be the reproducing kernel Hilbert space associated with K_ω, defined as

$$\mathcal{H}_K = \left\{ f \in L^2(\pi) : \|f\|_K^2 = \sum \frac{\langle f, \phi_{j\omega} \rangle^2}{\lambda_{j\omega}} < \infty \right\}.$$

Assumption 8.

$$E^{P_n}[\sqrt{n}\hat{h}(s;\hat{\theta})] - \sqrt{n} b_n (L_\omega \eta)(s) = \nu_n(s)$$

where $\nu_n \in \mathcal{H}_K$ and $\nu_n \to 0$ as n goes to infinity.

Assumption 8 slightly strengthens Assumption 6. Indeed by the mean value theorem, we have

$$\begin{aligned} \sqrt{n}\hat{h}(s;\hat{\theta}) &= \sqrt{n}\hat{h}(s;\theta_0) - \frac{\partial \psi(\bar{\theta})}{\partial t}(\hat{\theta} - \theta_0) \\ &= \sqrt{n}\hat{h}(s;\theta_0) - D_0(s) H^{-1} \left\langle G, \sqrt{n}\hat{h}(.;\theta_0) \right\rangle + o_{P_{2n}}(1) \end{aligned} \tag{22}$$

Moreover, $E^{P_n}[\sqrt{n}\hat{h}(s;\theta_0)] = \sqrt{n} b_n \eta(s)$. Thus, Assumption 8 says that the expectation of the term $o_{P_{2n}}(1)$ in Equation (22) belongs to the space $\mathcal{H}_K$.

Assumption 9. $p_n / (q_n n \alpha) \to 0$ and $p_n^2/(q_n n) \to 0$ as n goes to infinity and α goes to zero.

Assumption 9 is very mild given that in Proposition 10 we will require $n\alpha^2 \to \infty$, and also from Lemma 9 in Carrasco and Florens (2000), it is known that if there exist $0 < \gamma < 1$ and some positive constant c such that $p_n \sim c\alpha^{-\gamma}$, then $q_n \sim e\alpha^{-\gamma}$ for some positive constant e (see also remark 3 below).

Proposition 10 *Suppose Assumptions 1, 4-9 hold. Assume that* $L_\omega \eta \in \mathcal{H}_K$ *and*

$$\frac{n b_n^2}{\sqrt{q_n}} \to d \text{ for some constant } d. \tag{23}$$

Under H_{2n}, *we have*

$$J_{\alpha_n}(\hat{\theta}, \hat{K}) \xrightarrow{d} \mathcal{N}(d \|L_\omega \eta\|_K^2, 1)$$

as $n \to \infty$, $\alpha \to 0$, *and* $n\alpha^2 \to \infty$.

Remarks.

1. Under H_0, $J_{\alpha_n}(\hat{\theta}, \hat{K})$ converges to the same pivotal distribution for any consistent estimator satisfying Assumption 6. Hence, this test is robust to parameter uncertainty.

2. As the asymptotic distribution of $J_{\alpha_n}(\hat{\theta}, \hat{K})$ is a standard normal distribution, critical values from normal tabls can be readily used.

3. The condition (23) indicates the rate of b_n, which is related to the rate of the eigenvalues $\lambda_{j\omega}$ through q_n. Let us consider an example where $\lambda_{j\omega} = j^{-m}$. Then $q_n \sim \alpha^{-1/(2m)}$ (see Carrasco

and Florens (2000, Example 2) for the case $m = 1$ and Wahba (1975) for the general case). So condition (23) can be rewritten as $b_n \sim n^{-1/2}\alpha^{-1/(8m)}$.

4. The test $J_{\alpha_n}(\hat{\theta}, \hat{K})$ has nontrivial power against local alternatives $\psi_0(t, \theta_0) + cb_n\eta(t)$ provided $L_\omega \eta \neq 0$. In the case of the GMM estimator, this condition requires that $\eta \neq v' D_0$ for any vector v.

5. The fact that $J_{\alpha_n}(\hat{\theta}, \hat{K})$ has trivial power against $1/\sqrt{n}$ alternatives is linked to the rescaling of the statistic. In fact, all tests involving centering and rescaling exhibit the same lack of power against contiguous alternatives. This includes Neyman's smooth test with an increasing number of polynomials (see Lehmann and Romano), the chi-square type test for conditional moments (De Jong and Bierens, 1994), the goodness-of-fit tests considered by Eubank and LaRiccia (1992), Härdle and Mammen (1993) and the one considered by Aït-Sahalia, Bickel, and Stocker (2001), among others.

6. In this section, we assumed θ unknown. If θ is known, one can use the test $J_{\alpha_n}(\hat{\theta}, \hat{K})$ after replacing $\hat{\theta}$ by the true value θ_0 in the expression of the test statistic. The asymptotic distribution remains the same. One could also use the known eigenvalues and eigenfunctions of K instead of the estimated ones, but again the asymptotic distribution would not be altered.

7. Carrasco and Florens (2000) derived the asymptotic null distribution of $J_{\alpha_n}(\hat{\theta}, \hat{K})$ under a stronger assumption (Assumption 15: $q_n\sqrt{\alpha_n} \to \infty$). This assumption requires that the eigenvalues go to zero very slowly, which is not realistic here. On the contrary, the eigenvalues of K are likely to go to zero very fast, as illustrated in Figures 2e and 2f. For that reason, we propose a new proof which relaxes this assumption.

8. The lack of power of $J_{\alpha_n}(\hat{\theta}, \hat{K})$ against contiguous alternatives may speak in favor of tests such that T_B, which have power against contiguous alternatives. However, the test $J_{\alpha_n}(\hat{\theta}, \hat{K})$ may have higher power than T_B for higher frequency alternatives (case $j \to \infty$ in Proposition 8); see Theorem 3 in Eubank and LaRiccia (1992). The next remark considers this issue from a different angle.

9. Proposition 10 establishes the asymptotic distribution of $J_{\alpha_n}(\hat{\theta}, \hat{K})$ for η such that $\|\eta\|_K^2 < \infty$. However, this condition is not necessarily satisfied, so it is of special interest to look at what happens when it does not hold. Specifically, consider the case where

$$\frac{1}{\sqrt{q_n}} \sum_{l=1}^{n} \frac{a_{l\omega}^2}{\lambda_{l\omega}} \left|\langle L_\omega \eta, \phi_{l\omega}\rangle\right|^2 \to \infty. \tag{24}$$

The proof of Proposition 10 implies that the right rate for the alternatives H_{2n} is such that

$$nb_n^2 \frac{1}{\sqrt{q_n}} \sum_{l=1}^{n} \frac{a_{l\omega}^2}{\lambda_{l\omega}} \left|\langle L_\omega \eta, \phi_{l\omega}\rangle\right|^2 \to d$$

for some constant d. It follows from (24) that $nb_n^2 \to 0$. Hence the test $J_{\alpha_n}(\hat{\theta}, \hat{K})$ has power against local alternatives which approach the null hypothesis at a faster rate than $n^{-1/2}$. For these alternatives, the power of the tests T_B presented earlier remain $n^{-1/2}$. So the test $J_{\alpha_n}(\hat{\theta}, \hat{K})$ is able to detect certain alternatives which are closer to the null than the tests based on a fixed α. This result is similar to what was observed by Fan and Li (2000) in the context of specification tests for nonparametric regression. In particular, they show that nonparametric specification tests such as that of Härdle and Mammen (1993) with a fixed bandwidth has analogous properties as the integrated conditional tests of Bierens (1982) and Bierens and Ploberger (1997). Further, they show that kernel based tests with bandwidth going to zero can detect specific alternatives (the so-called singular alternatives) at a faster rate than $n^{-1/2}$. As we mentioned before, we can interpret α as a bandwidth in our tests.

5.2 Numerical invariance to moment transformations

As is well known, the traditional J test corresponding to the continuous unpdated estimator (CUE) is invariant to parameter-dependent linear transformations of the moments (see Hansen, Heaton and Yaron (1995)). To illustrate this fact, let $\underline{\hat{h}}(\theta)$ be the sample average of a vector of moments and M_θ be a (possibly complex-valued) square invertible matrix. Then, it is easy to check that the J-test based on $\underline{\hat{h}}(\theta)$ is the same as the J-test based on $M_\theta \underline{\hat{h}}(\theta)$ because:

$$J = n\underline{\hat{h}}(\theta)^\star M_\theta^\star (M_\theta \hat{\mathcal{K}}_\theta M_\theta^\star)^{-1} M_\theta \underline{\hat{h}}(\theta) = n\underline{\hat{h}}(\theta)^\star \hat{\mathcal{K}}_\theta^{-1} \underline{\hat{h}}(\theta).$$

When one uses regularization to invert the covariance matrix, this result is not true in general. Indeed, we have that

$$n\underline{\hat{h}}(\theta)^\star M_\theta^\star (M_\theta \hat{\mathcal{K}}_\theta M_\theta^\star)^{1/2} [(M_\theta \hat{\mathcal{K}}_\theta M_\theta^\star)^2 + \alpha I]^{-1} (M_\theta \hat{\mathcal{K}}_\theta M_\theta^\star)^{1/2} M_\theta \underline{\hat{h}}(\theta)$$

is not usually equal to

$$n\underline{\hat{h}}(\theta)^\star \hat{\mathcal{K}}_\theta^{1/2} (\hat{\mathcal{K}}_\theta^2 + \alpha I)^{-1} \hat{\mathcal{K}}_\theta^{1/2} \underline{\hat{h}}(\theta)$$

unless M_θ is unitary, that is $M_\theta M_\theta^\star = M_\theta^\star M_\theta = I$, in which case the two expressions coincide.

When there is a continuum of moment conditions, an analogous result turns out to be true for unitary transformations of h.

Define U_θ as a nonrandom linear operator from $L^2(\pi)$ into $L^2(\pi)$. Let U_θ^* be the adjoint of

U_θ. By the Riesz representation theorem, there is a unique $g_\theta(.,s)$ such that

$$(U_\theta \varphi)(s) = \langle g_\theta(.,s), \varphi(.)\rangle = \int g_\theta(t,s)\varphi(t)\pi(t)\,dt.$$

Let K_θ be the covariance operator of $h_i(.;\theta)$ and $\tilde{K}_\theta$ be the covariance operator of $U_\theta h_i(.;\theta)$. The kernel of $\tilde{K}_\theta$ is such that

$$\begin{aligned}
\tilde{k}_\theta(s_1,s_2) &= E\left[(U_\theta h_i(.;\theta))(s_1)(U_\theta h_i(.;\theta))(s_2)^\star\right] \\
&= E\left[\int g_\theta(t,s_1)h_i(t;\theta)\pi(t)\,dt \int \overline{g_\theta(u,s_2)h_i(u;\theta)}\pi(u)\,du\right] \\
&= \int g_\theta(t,s_1)\left\{\int E[h_i(t;\theta)\overline{h_i(u;\theta)}]\overline{g_\theta(u,s_2)}\pi(u)\,du\right\}\pi(t)\,dt \\
&= \langle g_\theta(.,s_1), K_\theta g_\theta(.,s_2)\rangle.
\end{aligned}$$

Then, we can characterize $\tilde{K}_\theta$:

$$\begin{aligned}
\left(\tilde{K}_\theta\varphi\right)(\tau) &= \int\int g_\theta(t,\tau)\left\{\int E[h_i(t;\theta)\overline{h_i(u;\theta)}]\overline{g_\theta(u,s)}\pi(u)\,du\right\}\pi(t)\,dt\varphi(s)\pi(s)\,ds \\
&= (U_\theta K_\theta U_\theta^*\varphi)(\tau).
\end{aligned}$$

Proposition 11 *Let U_θ be an unitary operator from $L^2(\pi)$ to $L^2(\pi)$ i.e. $U_\theta^* U_\theta = U_\theta U_\theta^* = I$. Then, the following equality holds:*

$$\left\|U_\theta \hat{h}(\theta)\right\|_{(U_\theta K_\theta U_\theta^*)^\alpha} = \left\|\hat{h}(\theta)\right\|_{K_\theta^\alpha} \tag{25}$$

regardless of the sample size n.

This means that the CUE versions of tests T_B with $B = (K^\alpha)^{-1/2}$ and $J_{\alpha_n}(\hat{\theta}, \hat{K})$ are invariant to unitary transformations of h. For non unitary transformations, the result is no longer true because of the regularization. In contrast, T_B with $B = I$ for instance is not even invariant to unitary transformations.

6 Monte Carlo experiments

In this section, we assess the finite sample performance of our proposed tests by means of an extensive Monte Carlo exercise. In addition, we compare them to several popular nonparametric tests based on the empirical distribution function, as well as to directional tests that target specific parametric alternatives to the null. In all cases, our sample size is $n = 100$.

6.1 Testing univariate normality

The first design we consider is a univariate normal distribution, which is by far the most common null hypothesis in distributional tests. In order to make our tests numerically invariant to affine transformations of the observations, we systematically centre and standardize them using the sample mean and standard deviation (with denominator n), which are the ML estimators under the null. As proved by Carrasco and Florens (2014), an asymptotically equivalent procedure would estimate the mean and variance by minimizing the continuum of moment conditions criterion function, but this would result in an increase of the computational costs. Either way, we can set the true mean and variance to 0 and 1, respectively, without loss of generality.

We consider three versions of our test, which differ on the way the covariance operator is estimated. The first one uses the theoretical covariance operator for a standard normal, which we presented in Section 2. In turn, the second and third versions rely on the centred and uncentred sample estimators using expressions (11) and (12), respectively, with the matrices C and $\mathring{C}$ computed using the analytical integrals in Appendix B.2. Given that these two sample versions produce very similar results, we only report the centred one in what follows. Importantly, the test that uses the theoretical covariance operator offers two notable computational advantages: i) the calculation of its eigenvalues and eigenfunctions depends on the number of grid points M, which we set to 1,000, but not on the sample size, so it can be used with very large datasets; and ii) we only need to compute those eigenelements once regardless of the number of Monte Carlo simulations.

In view of the discussion in Section 2, we look at two values of the Tikhonov regularization parameter α (.1 and .01) and two values for the scale parameter of the $\mathcal{N}(0,\omega^2)$ density defining inner products (1 and $\sqrt{10}$). As we have previously discussed, increasing ω not only changes the eigenvalues and eigenfunctions, but more intuitively, it pays relative more attention to the characteristic function for large (in absolute terms) values of its argument t.

In this univariate context, it is straightforward to compute the Cramer von Mises (CvM), Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) statistics on the basis of the probability integral transforms (PIT) of the standardized observations obtained through the standard normal cdf (see Appendix B.3 for details). Their usual asymptotic distributions are invalid, though, because those PITs make use of the sample mean and variance.

Further, we also compute two moment-based tests: one focusing on the fourth Hermite polynomial $(z^4 - 3z^2 + 1)/\sqrt{24}$ and another one that simultaneously looks at the third Hermite polynomial $(z^3 - 3z)/\sqrt{6}$ too. The advantage of working with Hermite polynomials is that they are asymptotically invariant to parameter estimation under the null (see e.g. Bontemps and

Meddahi (2005)). As is well known, these two statistics can be derived as Lagrange multiplier tests against a variety of non-normal distributions (see e.g. Jarque and Bera (1980) or Mencía and Sentana (2012)). Finally, we also compute the Bierens and Wang (2012) test described in Appendix B.3 using a Matlab translation of their C+ code.

The first thing we do is to compute all the aforementioned tests for 10,000 simulated samples generated under the null, whence we obtain finite sample critical values. This parametric bootstrap procedure automatically generates size-adjusted rejection rates, as forcefully argued by Horowitz and Savin (2000); see also Dufour (2006) for a discussion of Monte Carlo tests.

Panels A-F of Table 1 contain those rejection rates for six different alternatives: a symmetric Student t with 12 degrees of freedom; an asymmetric Student t with the same number of degrees of freedom but skewness parameter $\beta = -.75$; a scale mixture of two normals with the same kurtosis as the symmetric t, 3.75, and mixture probability $\lambda = .1$ (outlier case); another scale mixture with the same kurtosis but $\lambda = .75$ (inlier case); a location-scale mixture constructed in such a way that it has same skewness and kurtosis as the normal and $E(x^5) = -1$, $E(x^6) = 18$; and finally the second order Hermite expansion of the normal density mentioned in Lemma 7 with parameters $a = .4$ and $b = .5$. Details on how we simulate those distributions can be found in Appendix B.4. Figure 4 presents the densities of these alternative distributions once they have been standardized so that they all have 0 means and unit standard deviations in the population.

The first four columns of each panel in Table 1 report the results for the test that is based on the theoretical covariance operator, $J(\hat{\theta}, K_{\hat{\theta}})$, for the different values of α and ω that we consider. In turn, the next four columns contain the same figures for the test $J(\hat{\theta}, \mathring{K})$ which uses centred sample estimator of the covariance operator. As can be seen across the different panels, in all cases the results seem robust to the choice of the regularization parameter α. For the majority of the DGPs, $J(\hat{\theta}, K_{\hat{\theta}})$ has more power when $\omega = 1$ while the performance of $J(\hat{\theta}, \mathring{K})$ is better with $\omega = \sqrt{10}$. In addition, they generally outperform the other consistent tests that we consider, with AD being the most powerful of them. Somewhat surprisingly, this is also true when the DGP is the second order Hermite expansion of the normal mentioned in Lemma 7 (Panel F). Nevertheless, it is important to remember that this lemma refers to local alternatives, while our test is consistent versus fixed alternatives. Not surprisingly, the LM tests are the most powerful testing procedures when the distribution under the alternative is the one they are designed to detect. Specifically, S–t, the LM test against symmetric Student t alternatives, in Panel A, and A–t, the LM test against asymmetric Student t alternatives, in Panel B.

In summary, our proposed tests display good power against a variety of alternatives.

6.2 Testing uniformity

The second design we consider is a uniform distribution. Although this distribution does not often arise as a model for natural phenomena, it plays a fundamental role in statistics for two reasons: most computer-based pseudo-random number generators aim to draw uniform variates, and the PITs of any continuous random variables are uniform. To facilitate the comparison with the normal distribution, we transform the standard uniform random numbers by subtracting from them their population mean (.5) and scaling them up by their population standard deviation ($\sqrt{12}$), so that the resulting distribution will become standardized.

We consider exactly the same versions of our tests as in Section 6.1, but with the expressions for the population kernel and the centred and uncentred sample versions modified accordingly, as explained in Appendix B.2. We also compute the three non-parametric tests based on the CDF, as well as the Bierens and Wang (2012) test. As for directional tests, we consider two possibilities. The first one is the LM test of uniform vs beta proposed by Sefton (1992), which exploits the fact that a beta distribution with shape parameters $a = b = 1$ becomes uniform. This test is based on the average scores with respect to the beta parameters evaluated under the null, which are $1 + \ln(u)$ and $1 + \ln(1 - u)$, respectively.[1] The second directional test is a moment test based on the first two Jacobi polynomials evaluated again under the null, namely $\sqrt{3}(2u - 1)$ and $\sqrt{3}(6u^2 - 6u + 1)$, which was proposed by Bontemps and Meddahi (2012). As is well known, those polynomials constitute an orthonormal basis for the beta random variable.

The three panels of Table 2 contain the parametric bootstrap rejection rates for three different alternatives. The first one is a symmetric, unimodal beta distribution with parameters $a = b = 1.1$. The second one is an asymmetric unimodal concave beta distribution with parameters $a = 1.1$ and $b = 1$. Finally, the last distribution is generated as the standard Gaussian PITs of observations drawn from the same asymmetric Student t distribution with 12 degrees of freedom and asymmetric parameter in Section 6.1. The motivation for including this alternative is that we can use it to compare the direct application of our proposed tests to the original observations and to a monotonic transformation of them.

The first four columns of each panel report the results for the test that based on the theoretical covariance operator, $J(\theta_0, K)$, for the different values of α and ω that we consider, while the next four columns focus on $J(\theta_0, \mathring{K})$. As in Section 6.1, the rejection rates of our tests seem robust to the choice of the regularization parameter α. But in this case they are also

[1] The asymptotic variance for the scores reported by Sefton (1992) seems to be incorrect. As a result, we use instead 1 for the two asymptotic variances and $(6 - \pi^2)/6$ for the covariance. Hence, the LM test is $T/2$ times the square of the difference between the two scores divided by $\pi^2/6$ plus the square of their sum divided by $(12 - \pi^2/6)$.

less sensitive to the choice of ω. As before, the test based on K outperforms the one that uses centred sample estimator of the covariance operator $\mathring{K}$. Interestingly, both of them outperform the competitors when the DGP is either a symmetric beta or the Gaussian PITs of observations drawn from an asymmetric Student t. In contrast, CvM and AD are slightly more powerful when the alternative is the asymmetric beta. Somewhat surprisingly, the LM test is not particularly powerful.

6.3 Testing bivariate normality

Our next design is a bivariate normal distribution, which is by far the most common null hypothesis in multivariate distributional tests. Once again, we make our tests numerically invariant to affine transformations of the observations by systematically centring and standardizing them using the sample mean and the Cholesky decomposition of the sample covariance matrix (with denominator n), which are the ML estimators under the null.[2] Thus, we can set the true means and standard deviations to 0 and 1, respectively, and the correlation coefficient to 0 without loss of generality.

We consider exactly the same versions of our tests as in the Section 6.1, but with the expressions for the population kernel and the centred and uncentred sample versions modified accordingly (see Appendix B.2). However, we do not compute any classical non-parametric tests because there is no consensus on distribution-free multivariate generalization of the CvM, KS and AD statistics based on the joint distribution function. Nevertheless, we continue to apply the Bierens and Wang (2012) test. By analogy with the univariate normal case in section 6.1, we also consider two directional tests: the LM test of a multivariate normal against a multivariate Student t in Fiorentini, Sentana and Calzolari (2003) (denoted S–t), which effectively focuses on Mardia's (1970) coefficient of multivariate excess kurtosis, and the LM test against a generalized hyperbolic distribution in Mencía and Sentana (2012) (denoted A–t), which also looks at third moments in order to capture asymmetries in the multivariate distribution. By construction, both tests are asymptotically invariant to parameter estimation under the null.

The three panels of Table 3 contain the parametric bootstrap rejection rates for three different alternatives. The first one is a multivariate Student t with 12 degrees of freedom. The second one is an asymmetric Student t with the same degrees of freedom and vector of asymmetric parameters $(-.75, -.75)$. Finally, the third alternative is a spherically symmetric bivariate version of the outlier distribution considered in Section 6.1.

[2] As we mentioned before, an asymptotically equivalent procedure would estimate the two means and variances as well as the covariance by minimising the continuum of moment conditions criterion function, but this would result in a huge increase of the computational cost.

As in Table 1, the first four columns of each panel in Table 3 report the results for the test $J(\hat{\theta}, K_{\hat{\theta}})$ for the different values of α and ω and the next four columns correspond to same figures for the test $J(\hat{\theta}, \mathring{K})$. As can be seen in Table 3, in all cases the results seem robust to the choice of the regularization parameter α. Moreover, for the DGPs we consider $J(\hat{\theta}, K_{\hat{\theta}})$ has more power when $\omega = 1$ while the performance of $J(\hat{\theta}, \mathring{K})$ is better with $\omega = \sqrt{10}$, as in the univariate case. Interestingly, $J(\hat{\theta}, K_{\hat{\theta}})$ beats the S–t LM test when the DGPs is asymmetric Student t and there is a tie between $J(\hat{\theta}, \mathring{K})$ and S–t LM test when the alternative is a discrete-scale mixture of normals.

6.4 Testing chi-square

The final design that we consider is a chi-square distribution with two degrees of freedom. Like the uniform, the chi-square distribution does not often arise as a model for natural phenomena. But it also plays a fundamental role in statistics because it is the distribution of the (square) Mahalanobis distance of a multivariate normal random variable from its mean. In other words, it corresponds to the distribution of $(y_i - \mu)'\Sigma^{-1}(y_i - \mu)$ when $y_i \sim \mathcal{N}(\mu, \Sigma)$.

We consider exactly the same versions of our tests as in Sections 6.1 and 6.2, but with the expressions for the population kernel and the centred and uncentred sample versions in Appendix B.2 suitably modified. In that regard, the main difference is that we define inner products using a uniform density over $[-\omega, \omega]$ for tractability, for values of ω equal to 1 and $\sqrt{10}$. Although we standardize again the random draws by subtracting their population mean (=2) and scaling them down by their population standard deviation (=2), their distribution remains asymmetric, which implies that both the CF and the eigenfunctions of the associated covariance operator are complex, as explained in Section 2. This creates a normalization problem because any complex vector of unit length remains so after scaling its elements by any complex scalar on the unit circle, e^{iv}, where $v \in [0, 2\pi)$. Nevertheless, our proposed tests are numerically invariant to any chosen normalization.

We also compute the three non-parametric tests, as well as the Bierens and Wang (2012) test. As for directional tests, we consider two possibilities. The first one is the LM test of chi square with N degrees of freedom versus F with the same number of degrees of freedom in the numerator but ν degrees of freedom in the denominator proposed by Fiorentini, Sentana and Calzolari (2003). This test is based on the average score with respect to the reciprocal of ν evaluated under the null, which coincides with the second order Laguerre polynomial

$$\frac{1}{4}\varsigma^2 - 2\varsigma + 2,$$

whose asymptotic variance for $N = 2$ is 4 under the null. The second directional test is the LM test against a gamma distribution with mean N but shape parameter $\alpha \neq N/2$ developed in Amengual and Sentana (2012). In this case, the score is proportional to

$$\left(\frac{\varsigma}{2} - 1\right) - \left[\ln\left(\frac{\varsigma}{2}\right) - \psi(1)\right],$$

whose asymptotic variance is $\psi'(1) - 1$, where $\psi(.)$ and $\psi'(.)$ are the digamma and trigamma functions, respectively.

The three panels of Table 4 contain the parametric bootstrap rejection rates for three different alternatives. The first one is an F distribution with 12 degrees of freedom in the denominator, while the second one is a gamma distribution with shape parameter $\alpha = 2/3$ and scale parameter $\beta = 3$. Finally, the last distribution is generated as the square norm of observations drawn from a bivariate asymmetric Student t distribution with 12 degrees of freedom. Once again, the motivation for including this alternative is that we can use it to compare the direct application of our proposed bivariate Gaussian tests to the original observations or to a transformation of them which implicitly imposes spherical symmetry. In that regard, the F distribution would correspond to a bivariate Student t while the gamma to a Kotz distribution.

As in Table 2, the first four columns of each panel of Table 4 report the results for the test $J(\theta_0, K)$, for the different values of α and ω that we consider, while the next four columns contain the same figures for $J(\theta_0, \mathring{K})$. Once again, the results seem robust to the choice of the regularization parameter α, but at the same time they are less sensitive to the choice of ω. Still, for $J(\theta_0, K)$ the value $\omega = 1$ delivers higher rejection rates. As before, the test based on the theoretical covariance operator outperforms the one using centred sample estimator of the covariance operator. Interestingly, $J(\theta_0, K)$ has more power than its competitors, except when the DGP is Gamma.

7 Conclusion

In this paper, we propose goodness-of-fit tests based on comparing the empirical and theoretical characteristic functions. Our proposals are based on the continuum of moment conditions analogue to the usual overidentifying restrictions test, and therefore take into account the correlation between the influence functions for different argument values.

We consider different versions depending on whether the parameter vector θ is known in advance or replaced by a consistent estimator, and whether we make use of the analytical expression for the covariance operator or estimate it. Relying on the theoretical covariance operator offers substantial computational gains because the calculation of its eigenvalues and eigenvectors does not depend on the sample size, which allows its use with very large datasets.

We derive the asymptotic distribution of our proposed tests for fixed regularization parameter and when this vanishes with the sample size. Both types of tests have very different asymptotic properties. The fixed α J test has a nonstandard asymptotic distribution which depends on nuisance parameters but has power against $1/\sqrt{n}$ alternatives. In contrast, the vanishing α J test has a standard normal asymptotic distribution but generally fails to reject local $1/\sqrt{n}$ alternatives, except for some specific alternatives which it can detect at a faster rate.

Our theoretical study of power sheds some light on the alternatives for which each test is more powerful. While there is no test whose power dominates overall, it seems that fixing α at a small positive value is a good compromise. An extensive Monte Carlo exercise confirms this point by showing that our proposed tests display good power in finite samples against a variety of alternatives.

Although we have focused on a random sample framework for pedagogical reasons, versions of our tests robust to serial or cross-sectional dependence in the observations should be relatively straightforward. The analysis of conditional distributions would also constitute a very valuable but non-trivial addition with many potentially interesting empirical applications.

Appendix

A Proofs and auxiliary results

Proof of Lemma 1. K is self-adjoint positive definite because it is a covariance operator ($k(s,t) = \overline{k(t,s)}$) and its null space is reduced to 0, i.e. $Kf = 0 \Rightarrow f = 0$ (see the proof of Proposition A.1, condition A.5(i) in Carrasco, Chernov, Florens, and Ghysels, 2007). K is a Hilbert-Schmidt operator because its kernel is square integrable, indeed

$$\int\int |k(s,t)|^2 \pi(s)\, ds \pi(t)\, dt < \infty.$$

Consequently, K admits an infinite spectrum of positive eigenvalues. Let $\{\lambda_j, \varphi_j\}$ be the eigenvalues arranged in decreasing order and eigenfunctions (the eigenfunctions are taken orthonormal in $L^2(\pi)$) of K. By Mercer's formula (see Carrasco, Florens, and Renault, 2007, Theorem 2.42),

$$k(t,s) = \sum_j \lambda_j \varphi_j(t) \varphi_j(s).$$

By setting $s = t$, we have

$$\sum \lambda_j = \int k(t,t) \pi(t)\, dt.$$

Here $k(t,s) = \psi(t-s) - \psi(t)\psi(-s)$. Hence $k(t,t) = 1 - |\psi(t)|^2 \leq 1$. It follows that $\sum \lambda_j \leq 1$ and therefore because the operator is self-adjoint positive definite $0 \leq \lambda_j \leq 1$. Therefore $\lambda_j^2 \leq \lambda_j$ and hence $\sum \lambda_j^2 \leq 1$. So the Hilbert Schmidt norm of K is also bounded by 1:

$$\|K\|_{HS}^2 = \int\int |k(t,s)|^2 \pi(s)\, ds \pi(t)\, dt = \sum \lambda_j^2 \leq 1,$$

as desired. □

Proof of Proposition 2. We check the conditions (a) to (c) of Lemma 3.1 of Chen and White (1998) on

$$W_{nj} = \frac{1}{\sqrt{n}} \left(h_j - \frac{c\eta}{\sqrt{n}} \right).$$

Checking (a): We need to check that for all $\varphi \in L^2(\pi)$, $\sum_{j=1}^n \langle W_{nj}, \varphi \rangle \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\varphi)\right)$ where $\sigma^2(\varphi) > 0$. To do so, first notice that under H_{1n}, $W_{nj} = \frac{1}{\sqrt{n}}[e^{itX_j} - \psi_n(t)]$. We have

$E\left[\langle W_{nj},\varphi\rangle\right]=0$ and $\langle W_{nj},\varphi\rangle$, $j=1,2,...,n$ are independent. Moreover,

$$\begin{aligned}
E[|\langle W_{nj},\varphi\rangle|^2] &= E[\langle W_{nj},\varphi\rangle\,\overline{\langle W_{nj},\varphi\rangle}] \\
&= E\int\int W_{nj}(s)\,\overline{\varphi(s)W_{nj}(t)}\varphi(t)\,\pi(s)\,ds\pi(t)\,dt \\
&= \int\int E[W_{nj}(s)\,\overline{W_{nj}(t)}]\overline{\varphi(s)}\varphi(t)\,\pi(s)\,ds\pi(t)\,dt \\
&= \frac{1}{n}\langle\varphi,K_n\varphi\rangle
\end{aligned}$$

where K_n is the integral operator with kernel

$$\begin{aligned}
& k_n(s,t) \\
= \;& \psi_n(s-t)-\psi_n(s)\,\psi_n(-t) \\
= \;& \psi_0(s-t)-\psi_0(s)\,\psi_0(-t)+\frac{c\eta(s-t)}{\sqrt{n}}-\frac{c\eta(s)}{\sqrt{n}}\psi_0(-t)-c\psi_0(s)\,\frac{\eta(-t)}{\sqrt{n}}+c^2\frac{\eta(s)\,\eta(-t)}{n}.
\end{aligned}$$

Interchanging the order of integration is justified by the fact that $\frac{1}{n}\langle\varphi,K_n\varphi\rangle<\infty$. Now, we check the conditions of Lindeberg-Feller central limit theorem (van der Vaart (1998), Proposition 2.27) to establish $\sum_{j=1}^{n}\langle W_{nj},\varphi\rangle\xrightarrow{d}\mathcal{N}\left(0,\sigma^2(\varphi)\right)$ with $\sigma^2(\varphi)=\langle\varphi,K\varphi\rangle>0$. Let $Y_{nj}=\langle W_{nj},\varphi\rangle$. Here Y_{nj} are independent scalar random variables with zero mean and finite variance. The two conditions for the CLT are

$$\begin{aligned}
\text{(i)}\ \sum_{j=1}^{n}E[|Y_{nj}|^2\,I\{|Y_{nj}|>\varepsilon\}] &\to 0 \text{ for every } \varepsilon>0, \text{ and} \\
\text{(ii)}\ \sum_{j=1}^{n}V(Y_{nj}) &\to \sigma^2(\varphi).
\end{aligned}$$

Note that

$$\begin{aligned}
|Y_{nj}|^2 &= |\langle W_{nj},\varphi\rangle|^2 \\
&\le \left\|\frac{1}{\sqrt{n}}[e^{itx_j}-\psi_n(t)]\right\|^2\|\varphi\|^2 \\
&\le \frac{C}{n}\|\varphi\|^2
\end{aligned}$$

for some fixed constant C. Hence,

$$\begin{aligned}
& \sum_{j=1}^{n} E[|Y_{nj}|^2 I\{|Y_{nj}| > \varepsilon\}] \\
\leq \; & \frac{C\|\varphi\|^2}{n} \sum_{j=1}^{n} P[|Y_{nj}| > \varepsilon] \\
\leq \; & \frac{C\|\varphi\|^2}{n} \sum_{j=1}^{n} \frac{E[|Y_{nj}|^2]}{\varepsilon^2} \\
\leq \; & \frac{C^2}{n\varepsilon^2} \|\varphi\|^4
\end{aligned}$$

by Markov inequality. So condition (i) is satisfied. For (ii), we use the results above which give

$$\begin{aligned}
\sum_{j=1}^{n} V(Y_{nj}) &= \sum_{j=1}^{n} E[|\langle W_{nj}, \varphi\rangle|^2] \\
&= \langle \varphi, K_n \varphi\rangle \\
&\to \langle \varphi, K\varphi\rangle,
\end{aligned}$$

and hence, (ii) is also satisfied.

Checking (b) and (c): By Remark 3.3 (ii) of Chen and White (1998), conditions (b) and (c) can be replaced by the following condition:

$W_{n,j}$ strictly stationary and

$$\lim_{n\to\infty} E\left|\sum_{j=1}^{n} W_{nj}\right|^2 \leq C < \infty. \tag{A1}$$

We have

$$\begin{aligned}
E\left|\sum_{j=1}^{n} W_{nj}\right|^2 &= E\left\langle \sum_{j=1}^{n} W_{nj}, \sum_{l=1}^{n} \overline{W_{nl}} \right\rangle \\
&= \sum_{j=1}^{n} E\left\langle W_{nj}, \overline{W_{nj}}\right\rangle \\
&= \sum_{j=1}^{n} k_n(s,s) \\
&= 1 - |\psi_0(s)|^2 - \frac{c\eta(s)}{\sqrt{n}}\psi_0(-s) - \psi_0(s)\frac{c\eta(-s)}{\sqrt{n}} + \frac{c^2|\eta(s)|^2}{n}
\end{aligned}$$

which is bounded because $|\psi_0(s)|^2 \leq 1$ by the property of CFs and $|\eta(s)| < C$ by assumption. Therefore, (A1) is satisfied and $\sum_{j=1}^{n} W_{nj}$ is tight.

It follows that $\sqrt{n}\hat{h} = \sum_{j=1}^{n} W_{nj} + c\eta \Rightarrow \mathcal{N}(c\eta, K)$. □

Proof of Proposition 3. As B is bounded, we have (Chen and White (1992, working paper))

$$B_n\sqrt{n}\hat{h} \Rightarrow \mathcal{N}\left(cB\eta, BKB^*\right)$$

where B^* is the adjoint of B. Let Z denote a Gaussian process $\mathcal{N}\left(0, BKB^*\right)$. By the continuous mapping theorem,

$$\left\|B\sqrt{n}\hat{h}\right\|^2 \xrightarrow{d} \|cB\eta + Z\|^2 = \sum_j \left(c\left\langle B\eta, \phi_j\right\rangle + \left\langle Z, \phi_j\right\rangle\right)^2$$

where the equality uses the so-called Karhunen-Loeve representation of Gaussian processes:

$$Z = \sum_{j=1}^{\infty} \left\langle Z, \phi_j\right\rangle \phi_j = \sum_{j=1}^{\infty} \sqrt{a_j}\frac{\left\langle Z, \phi_j\right\rangle}{\sqrt{a_j}}\phi_j$$

and the fact that the ϕ_j form an orthonormal basis of $L^2(\pi)$. Moreover, $\left\langle Z, \phi_j\right\rangle/\sqrt{a_j}$ are *iid* $\mathcal{N}(0,1)$. □

Proof of Proposition 4. The proof is similar to those of Neuhaus (1976, Theorem 2.2.) and Escanciano (2009, Theorem 1) and is not repeated here. □

Proof of Proposition 5. Under our assumptions,

$$\left\|B_n\hat{h}\left(.;\theta\right)\right\| \xrightarrow{P_0} \left\|BE^{P_0}[h_j\left(.;\theta\right)]\right\|$$

uniformly in θ. (The uniformity part comes from the fact that $\hat{h}\left(.;\theta\right) - E[h_j\left(.;\theta\right)] = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j} - \psi_0\left(t;\theta_0\right)$ does not depend on θ.) Moreover, $E[h_j\left(.;\theta\right)] = \psi_0\left(.;\theta_0\right) - \psi_0\left(.;\theta\right)$. By the identification assumption, the objective function reaches its minimum at $\theta = \theta_0$. Hence, $\hat{\theta}$ is consistent under H_0.

We turn our attention toward the asymptotic normality. To simplify the notation, we write $\psi_0(\theta)$ for $\psi_0(.;\theta)$ and $\hat{h}(\theta)$ for $\hat{h}(.;\theta)$, and $\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}$ for $\left.\frac{\partial\psi_0(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}$. The first order condition of the minimization problem gives

$$\begin{aligned}&\left\langle B_n\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}, B_n\hat{h}(\hat{\theta})\right\rangle \\ =\ & 0 = \left\langle B_n\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}, B_n\hat{h}\left(\theta_0\right)\right\rangle - \left\langle B_n\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}, B_n\frac{\partial\psi_0(\tilde{\theta})}{\partial\theta}(\hat{\theta}-\theta_0)\right\rangle\end{aligned}$$

where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}$. It follows that

$$\sqrt{n}(\hat{\theta}-\theta_0) = \left\langle B_n\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}, B_n\frac{\partial\psi_0(\tilde{\theta})}{\partial\theta}\right\rangle^{-1}\left\langle B_n\frac{\partial\psi_0(\hat{\theta})}{\partial\theta}, B_n\sqrt{n}\hat{h}\left(\theta_0\right)\right\rangle$$

By the continuity of $\frac{\partial\psi_0}{\partial\theta}$ and the consistency of $\hat{\theta}$, we have

$$\sqrt{n}(\hat{\theta}-\theta_0) = \left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\left(.;\theta_0\right)\right\rangle + o_{P_0}(1). \tag{A2}$$

The asymptotic normality follows from Proposition 1.

For the convergence of $\hat{\theta}$ to θ_1 under H_1, we use the same arguments as for the consistency under H_0. The existence of the minimum comes from the fact that $\psi_0(.;\theta)$ is continuous in θ and Θ is compact. □

Proof of Proposition 6.

(i) The mean value theorem gives

$$\begin{aligned}\sqrt{n}\hat{h}(\hat{\theta}) &= \sqrt{n}\hat{h}(\theta_0) - \frac{\partial\psi_0(\tilde{\theta})}{\partial\theta}\sqrt{n}(\hat{\theta}-\theta_0) \\ &= \sqrt{n}\hat{h}\left(\theta_0\right) - D_0\sqrt{n}(\hat{\theta}-\theta_0) + o_{P_0}\left(1\right) \\ &= \sqrt{n}\hat{h}\left(\theta_0\right) - D_0\left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\left(\theta_0\right)\right\rangle + o_{P_0}\left(1\right)\end{aligned}$$

by Equation (A2). By the contiguity of P_n to P_0, it follows that

$$\sqrt{n}\hat{h}(\hat{\theta}) - \sqrt{n}\hat{h}\left(\theta_0\right) + D_0\left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\left(\theta_0\right)\right\rangle \overset{P_n}{\to} 0. \tag{A3}$$

By Proposition 2, we have under H_{1n}

$$\sqrt{n}\hat{h}\left(\theta_0\right) - D_0\left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\left(\theta_0\right)\right\rangle \Rightarrow \mathcal{N}(L\eta, \tilde{K}). \tag{A4}$$

Combining Equations (A3) and (A4) yields $\sqrt{n}\hat{h}(\hat{\theta}) \Rightarrow \mathcal{N}(L\eta, \tilde{K})$ under H_{1n}. The kernel of $\tilde{K}$ can be computed explicitly as follows:

$$\begin{aligned}&\tilde{k}\left(s,t\right) \\ = \; & E\Big[(\sqrt{n}\hat{h}\left(s\right) - D_0\left(s\right)\left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\right\rangle) \\ &\times(\overline{\sqrt{n}\hat{h}\left(t\right)} - \overline{D_0\left(t\right)}\left\langle BD_0, BD_0\right\rangle^{-1}\overline{\left\langle B^*BD_0, \sqrt{n}\hat{h}\right\rangle})\Big].\end{aligned}$$

Detailing the calculation for one of the 4 terms gives

$$\begin{aligned}&E\left[D_0\left(s\right)\left\langle BD_0, BD_0\right\rangle^{-1}\left\langle B^*BD_0, \sqrt{n}\hat{h}\right\rangle\overline{\sqrt{n}\hat{h}\left(t\right)}\right] \\ = \; & D_0\left(s\right)\left\langle BD_0, BD_0\right\rangle^{-1}E\left[\int B^*BD_0\left(u\right)\sqrt{n}\hat{h}\left(u\right)\pi\left(u\right)du\overline{\sqrt{n}\hat{h}\left(t\right)}\right] \\ = \; & D_0\left(s\right)\left\langle BD_0, BD_0\right\rangle^{-1}\int B^*BD_0\left(u\right)E[h_j\left(u\right)\overline{h_j\left(t\right)}]\pi\left(u\right)du \\ = \; & D_0\left(s\right)\left\langle BD_0, BD_0\right\rangle^{-1}\left(KB^*BD_0\right)\left(t\right).\end{aligned}$$

The other terms can be computed similarly.

(ii) The proof of (ii) is similar to that of Proposition 3 and hence omitted. □

Proof of Lemma 7. The CF of a $\mathcal{N}(\mu,\sigma^2)$ is $\psi_0(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$. Let $\theta = (\mu,\sigma^2)'$, then

$$D_0 = \frac{\partial \psi_0}{\partial \theta} = \begin{pmatrix} it\psi_0(t) \\ -\frac{t^2}{2}\psi_0(t) \end{pmatrix}.$$

Let $v = (a,b)$ and $\eta(t) = v' D_0 = \left(ait - \frac{bt^2}{2}\right)\psi_0(t)$. Now consider $\psi_n(t) = \left(1 + \frac{ait}{\sqrt{n}} - \frac{bt^2}{2\sqrt{n}}\right)\psi_0(t)$. Observe that $\psi_n(0) = 1$, $\overline{\psi_n(t)} = \psi_n(-t)$. We need $|\psi_n(t)| < 1$ which will be satisfied if $b > 0$ (and possibly for $b < 0$ and n large enough). So ψ_n satisfies the necessary conditions to be a CF, however these conditions are not sufficient. Necessary and sufficient conditions for a function ψ_n to be CF are that (a) $\psi_n(0) = 1$, and (b) ψ_n is non-negative definite (see Theorem 4.2.2 of Lukacs (1960)). It can be shown that, given ψ_0 is a CF, ψ_n will satisfy (b) for n large enough. So ψ_n is a CF.

Moreover, $\psi_n(t)$ is absolutely integrable so the density (f_n) corresponding to ψ_n satisfies:

$$\begin{aligned}
f_n(x) &= \frac{1}{2\pi}\int e^{-itx}\psi_n(t)\,dt \\
&= \frac{1}{2\pi}\int e^{-itx}\left(1 + \frac{ait}{\sqrt{n}} - \frac{bt^2}{2\sqrt{n}}\right)\psi_0(t)\,dt \\
&= \frac{1}{2\pi}\int e^{-itx}\psi(t)\,dt + \frac{ai}{\sqrt{n}}\frac{1}{2\pi}\int te^{-itx}\psi_0(t)\,dt - \frac{b}{2\sqrt{n}}\frac{1}{2\pi}\int e^{-itx}t^2\psi_0(t)\,dt.
\end{aligned}$$

Note that

$$\frac{i}{2\pi}\int te^{-itx}\psi_0(t)\,dt = \frac{\partial \frac{1}{2\pi}\int e^{-itx}\psi_0(t)\,dt}{\partial \mu}$$

$$\frac{1}{2\pi}\int e^{-itx}t^2\psi_0(t)\,dt = -2\frac{\partial \frac{1}{2\pi}\int e^{-itx}\psi_0(t)\,dt}{\partial \sigma^2}.$$

On the other hand

$$\frac{1}{2\pi}\int e^{-itx}\psi(t)\,dt = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \equiv f_0(x).$$

$$\begin{aligned}
\frac{\partial f_0(x)}{\partial \sigma^2} &= -\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma^3}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&\quad + \frac{(x-\mu)^2}{2\sigma^4}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \left[\frac{(x-\mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right] f_0(x).
\end{aligned}$$

$$\frac{\partial f_0(x)}{\partial \mu} = \frac{(x-\mu)}{\sigma^2} f_0(x)$$

It follows that $f_n(x) = \left\{1 + \frac{a}{\sqrt{n}}\frac{(x-\mu)}{\sigma^2} + \frac{b}{2\sqrt{n}}\left[\frac{(x-\mu)^2 - \sigma^2}{2\sigma^4}\right]\right\} f_0(x)$. □

Proof of Proposition 8. Under H_1, $\hat{h}(.;\hat{\theta}) \xrightarrow{P_1} E^{P_1} h_j(.;\theta_1) = \psi(.) - \psi_0(.;\theta_1) \neq 0$, where

$\psi(.)$ is the CF of X_j under H_1, and then the result follows. □

Preliminary results to the proof of Proposition 10.

The following lemmas will be used in the proof of Proposition 10.

Let $h_{in}(s) = e^{isX_i} - \psi_n(s)$ and $Y_{in}(s)$ be the process defined as

$$Y_{in}(s) = h_{in}(s) - D_0(s) H^{-1} \langle G, h_{in} \rangle .$$

Because h_{in} $i = 1, 2, ...$ are *iid* with mean zero, Y_{in} $i = 1, 2, ...$ are *iid* with mean 0 and covariance $E[Y_{in}(s) \overline{Y_{in}(t)}] \equiv k_{\omega n}(s,t)$ under H_{2n}. Let $K_{\omega n}$ be the integral operator with kernel $k_{\omega n}$ and $(\lambda_{l,n}, \phi_{l,n})$ be the eigenvalues and eigenfunctions of $K_{\omega n}$. Note that $K_{\omega n}$ converges to K_ω when n goes to infinity and similarly $\phi_{l,n}$ converges to $\phi_{l,\omega}$ as n goes to infinity.

Lemma 12 *Under H_{2n}, $\left(\frac{\langle Y_{in}, \phi_{l,n} \rangle}{\sqrt{\lambda_{l,n}}}\right)$, $l = 1, 2, ...$ are uncorrelated across l with zero mean and variance equal to 1.*

Proof of Lemma 12. We have

$$\begin{aligned} E\left[\langle Y_{in}, \phi_{l,n} \rangle \overline{\langle Y_{in}, \phi_{l',n} \rangle}\right] &= E \int Y_{in}(s) \overline{\phi_{l,n}(s)} \pi(s) \, ds \int \overline{Y_{in}(t)} \phi_{l',n}(t) \pi(t) \, dt \\ &= \int \overline{\phi_{l,n}(s)} \int E\left[Y_{in}(s) \overline{Y_{in}(t)}\right] \phi_{l',n}(t) \pi(t) \, dt \pi(s) \, ds \\ &= \overline{\langle \phi_{l,n}, K_{\omega n} \phi_{l',n} \rangle} \\ &= \begin{cases} \lambda_{l,n} \text{ if } l = l', \\ 0 \text{ otherwise,} \end{cases} \end{aligned}$$

as desired. □

The following lemma is taken from Eubank and LaRiccia (1992) and is reproduced here for convenience.

Lemma 13 *(Lemma 2 of Eubank and LaRiccia (1992)) Let $\{Y_{in}\}_{i=1}^n$, $n = 1, 2, ...$ be a triangular array of random variables that are iid within rows. Set $w_{ijn} = w_{ijn}(Y_{in}, Y_{jn}) + w_{ijn}(Y_{jn}, Y_{in})$ for some function $w_{ijn}(.,.)$ and assume that $E[w_{ijn} | Y_{in}] = 0$ for all $i, j \leq n$. Define*

$$\begin{aligned} w(n) &= \sum_{1 \leq i < j \leq n} w_{ijn}, \\ \sigma(n)^2 &= Var(w(n)) = \sum_{1 \leq i < j \leq n} E\left(w_{ijn}^2\right), \\ G_I &= \sum_{1 \leq i < j \leq n} E\left(w_{ijn}^4\right), \\ G_{II} &= \sum_{1 \leq i < j < k \leq n} \left[E\left(w_{ijn}^2 w_{ikn}^2\right) + E\left(w_{jin}^2 w_{jkn}^2\right) + E\left(w_{kin}^2 w_{kjn}^2\right)\right], \end{aligned}$$

and

$$G_{IV} = \sum_{1\leq i<j<k<m\leq n} \left[E\left(w_{ijn}w_{ikn}w_{mjn}w_{mkn}\right) + E\left(w_{ijn}w_{imn}w_{kjn}w_{kmn}\right) + E\left(w_{imn}w_{ikn}w_{jkn}w_{jmn}\right)\right].$$

Then, if G_I, G_{II}, and G_{IV} are all of smaller order than $\sigma(n)^4$,

$$\frac{w(n)}{\sigma(n)} \xrightarrow{d} \mathcal{N}(0,1).$$

Lemma 14 *Let $a_{l,n} = \frac{\lambda_{l,n}^2}{\lambda_{l,n}^2+\alpha}$, $p_{n,n} = \sum_{j=1}^n a_{l,n}$, $q_{n,n} = 2\sum_{j=1}^n a_{l,n}^2$. Under H_{2n} :*

$$\frac{\sum_{l=1}^n \frac{a_{l,n}}{\lambda_{l,n}} \left\langle \sqrt{n}\hat{h}\left(.;\hat{\theta}\right), \phi_{l,n}\right\rangle^2 - p_{n,n}}{\sqrt{q_{n,n}}} \xrightarrow{d} \mathcal{N}\left(c\left\|L_\omega\eta\right\|_K^2, 1\right).$$

Proof of Lemma 14. Our proof draws from the proof of Theorem 1 in Eubank and LaRiccia (1992). Using the notation $\hat{h}(\hat{\theta}) \equiv \hat{h}(.;\hat{\theta})$ and dropping the subscript n from $a_{l,n}$, $\lambda_{l,n}$, $\phi_{l,n}$, $p_{n,n}$, and $q_{n,n}$, we obtain

$$\begin{aligned} & \frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}\hat{h}(\hat{\theta}), \phi_l\right\rangle^2 - p_n}{\sqrt{q_n}} \\ = & \frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}\{\hat{h}(\hat{\theta}) - E[\hat{h}(\hat{\theta})] + E[\hat{h}(\hat{\theta})]\}, \phi_l\right\rangle^2 - p_n}{\sqrt{q_n}} \\ = & \frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}\{\hat{h}(\hat{\theta}) - E[\hat{h}(\hat{\theta})]\}, \phi_l\right\rangle^2 - p_n}{\sqrt{q_n}} + R_n \end{aligned}$$

where

$$R_n = \frac{2\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}\{\hat{h}(\hat{\theta}) - E[\hat{h}(\hat{\theta})]\}, \phi_l\right\rangle \left\langle \sqrt{n}E[\hat{h}(\hat{\theta})], \phi_l\right\rangle}{\sqrt{q_n}} + \frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}E[\hat{h}(\hat{\theta})], \phi_l\right\rangle^2}{\sqrt{q_n}}.$$

In a first step, we will show that R_n converges to $d\left\|L_\omega\eta\right\|_K^2$ in probability under H_{2n} as n goes to infinity and α goes to zero. In a second step, we will show that

$$\frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \sqrt{n}\{\hat{h}(\hat{\theta}) - E[\hat{h}(\hat{\theta})]\}, \phi_l\right\rangle^2 - p_n}{\sqrt{q_n}} \xrightarrow{d} \mathcal{N}(0,1)$$

under H_{2n}.

First step. By Assumption 8, we have

$$E(R_n) = \frac{nb_n^2}{\sqrt{q_n}} \sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle L_\omega\eta, \phi_l\right\rangle^2 + \frac{\sum_{l=1}^n \frac{a_l}{\lambda_l}\left\langle \nu_n, \phi_l\right\rangle^2}{\sqrt{q_n}}$$

where the second term goes to zero. Moreover, $\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle L_\omega\eta,\phi_l\right\rangle^2 \to \sum_{l=1}^{\infty}\frac{1}{\lambda_l}\left\langle L_\omega\eta,\phi_l\right\rangle^2 = \|L_\omega\eta\|_K^2$ and $\frac{nb_n^2}{\sqrt{q_n}} \to d$ as n goes to infinity and α goes to zero. Therefore, $E\left(R_n\right) \to d\left\|L_\omega\eta\right\|_K^2$.

Now we show that the variance of R_n goes to zero. Using the notation $Y_i = Y_{in}$, we have

$$\begin{aligned}
V\left(R_n\right) &= V\left[\frac{2\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle\frac{\sum_{i=1}^{n}Y_i}{\sqrt{n}},\phi_l\right\rangle\left\langle\sqrt{n}b_nL_\omega\eta,\phi_l\right\rangle}{\sqrt{q_n}}\right]+o_{p_n}(1)\\
&= \frac{4nb_n^2}{q_n}V\left[\frac{\sum_{i=1}^{n}}{\sqrt{n}}\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle Y_i,\phi_l\right\rangle\left\langle L_\omega\eta,\phi_l\right\rangle\right]\\
&= \frac{4nb_n^2}{q_n}V\left[\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle Y_i,\phi_l\right\rangle\left\langle L_\omega\eta,\phi_l\right\rangle\right]
\end{aligned}$$

because Y_i, $i = 1, 2, ..., n$ are *iid*. As $\langle Y_i, \phi_l\rangle$, $l = 1, 2...$ are uncorrelated by Lemma 12, we obtain

$$\begin{aligned}
V\left(R_n\right) &= \frac{4nb_n^2}{q_n}\sum_{l=1}^{n}\frac{a_l^2}{\lambda_l}\left|\left\langle L_\omega\eta,\phi_l\right\rangle\right|^2\\
&\leq \frac{4nb_n^2}{q_n}\left\|L_\omega\eta\right\|_K^2 \to 0.
\end{aligned}$$

It follows that R_n converges to $d\left\|L_\omega\eta\right\|_K^2$ in probability under H_{2n}.

Second step. We have

$$\begin{aligned}
\frac{\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle\sqrt{n}\{\hat{h}(\hat{\theta})-E[\hat{h}(\hat{\theta})]\},\phi_l\right\rangle^2-p_n}{\sqrt{q_n}} &= \frac{\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i,\phi_l\right\rangle^2-p_n}{\sqrt{q_n}}+o_{p_{2n}}(1)\\
&= \frac{w_1(n)+w(n)}{\sqrt{q_n}}+o_{p_{2n}}(1)
\end{aligned}$$

where

$$\begin{aligned}
w_1(n) &= \frac{1}{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\frac{a_l}{\lambda_l}\left\langle Y_i,\phi_l\right\rangle^2-p_n,\\
w(n) &= \frac{2}{n}\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\sum_{1\leq i<j\leq n}\left\langle Y_i,\phi_l\right\rangle\left\langle Y_j,\phi_l\right\rangle = \sum_{1\leq i<j\leq n}w_{ijn}
\end{aligned}$$

with

$$w_{ijn} = \frac{2}{n}\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\left\langle Y_i,\phi_l\right\rangle\left\langle Y_j,\phi_l\right\rangle.$$

First, we show that $w_1(n)/\sqrt{q_n} \xrightarrow{P} 0$. We have

$$\begin{aligned}
E[w_1(n)] &= \frac{1}{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\frac{a_l}{\lambda_l}E[\langle Y_i,\phi_l\rangle^2]-p_n\\
&= \sum_{l=1}^{n}a_l-p_n = 0.
\end{aligned}$$

As $\langle Y_i, \phi_l \rangle^2$ are independent across i, we have

$$
\begin{aligned}
& V[w_1(n)] \\
= & \frac{1}{n} V\left[\sum_{l=1}^{n} a_l \left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} - 1\right)\right] \\
= & \sum_{l=1}^{n} \frac{a_l^2}{n} E\left[\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} - 1\right)^2\right] + \sum_{l \neq l'} \frac{a_l a_{l'}}{n} E\left[\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} - 1\right)\left(\frac{\langle Y_i, \phi_{l'}\rangle^2}{\lambda_{l'}} - 1\right)\right] \quad \text{(A5)}
\end{aligned}
$$

Using Lemma 12, we have

$$
\begin{aligned}
& \sum_{l \neq l'} \frac{a_l a_{l'}}{n} E\left[\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} - 1\right)\left(\frac{\langle Y_i, \phi_{l'}\rangle^2}{\lambda_{l'}} - 1\right)\right] \\
= & \sum_{l \neq l'} \frac{a_l a_{l'}}{n} E\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} \frac{\langle Y_i, \phi_{l'}\rangle^2}{\lambda_{l'}}\right) \quad \text{(A6)} \\
& - \sum_{l \neq l'} \frac{a_l a_{l'}}{n}. \quad \text{(A7)}
\end{aligned}
$$

Consider (A7):

$$
\frac{1}{q_n} \sum_{l \neq l'} \frac{a_l a_{l'}}{n} \leq \frac{p_n^2}{q_n n}
$$

which goes to zero by Assumption 9. To deal with the term (A6), we exploit the fact that for n large enough, $|Y_i| = \left|e^{itX_i} - \psi_n(t)\right| \leq \left|e^{itX_i}\right| + |\psi_n(t)| = 2$, hence $\|Y_i\|^2 \leq 4$ and $|\langle Y_i, \phi_l\rangle|^2 \leq 4$ by Cauchy-Schwarz and $\|\phi_l\| = 1$. Therefore, by Lemma 12,

$$
E\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} \frac{\langle Y_i, \phi_{l'}\rangle^2}{\lambda_{l'}}\right) \leq \frac{4}{\lambda_l \lambda_{l'}} E\left(\langle Y_i, \phi_{l'}\rangle^2\right) = \frac{4}{\lambda_l}.
$$

Hence,

$$
\sum_{l \neq l'} \frac{a_l a_{l'}}{n} E\left(\frac{\langle Y_i, \phi_l\rangle^2}{\lambda_l} \frac{\langle Y_i, \phi_{l'}\rangle^2}{\lambda_{l'}}\right) \leq \sum_{l \neq l'} \frac{a_l a_{l'}}{n \lambda_l} = \frac{p_n}{n} \sum_l \frac{a_l}{\lambda_l}.
$$

Note that

$$
\sum_l \frac{a_l}{\lambda_l} = \sum_l \frac{\lambda_l}{\lambda_l^2 + \alpha} \leq \frac{1}{\alpha} \sum_l \lambda_l.
$$

So, we obtain:

$$
\frac{p_n}{q_n n} \sum_l \frac{a_l}{\lambda_l} \leq \frac{p_n}{q_n n \alpha}
$$

which goes to zero under Assumption 9.

The first term in (A5) can be treated in the same manner. So that $V[w_1(n)]/q_n \to 0$ under our assumptions and hence $w_1(n)/\sqrt{q_n} \xrightarrow{P} 0$.

Second, we show that

$$
\frac{w(n)}{\sqrt{q_n}} \xrightarrow{d} \mathcal{N}(0, 1).
$$

To establish this result, we check all the conditions of Lemma 13.

$$\sigma(n)^2 = V(w_n) = \sum_{1\le i<j\le n} E(w_{ijn}^2)$$

where

$$\begin{aligned}
E\left(w_{ijn}^2\right) &= \frac{4}{n^2}E\left[\left(\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\langle Y_i,\phi_l\rangle\langle Y_j,\phi_l\rangle\right)^2\right] \\
&= \frac{4}{n^2}\sum_{l=1}^{n}\frac{a_l^2}{\lambda_l^2}E[\langle Y_i,\phi_l\rangle^2\langle Y_j,\phi_l\rangle^2] \\
&= \frac{4}{n^2}\sum_{l=1}^{n}\frac{a_l^2}{\lambda_l^2}E[\langle Y_i,\phi_l\rangle^2]E[\langle Y_j,\phi_l\rangle^2] \\
&= \frac{4}{n^2}\sum_{l=1}^{n}a_l^2 = \frac{2q_n}{n^2}
\end{aligned}$$

because the $\langle Y_i,\phi_l\rangle$ are uncorrelated across l and independent across i. Hence,

$$\sigma(n)^2 \sim q_n.$$

Consider now the term G_I:

$$G_I = \sum_{1\le i<j\le n} E(w_{ijn}^4).$$

We have

$$\begin{aligned}
w_{ijn}^4 &= \frac{16}{n^4}\left(\sum_{l=1}^{n}\frac{a_l}{\lambda_l}\langle Y_i,\phi_l\rangle\langle Y_j,\phi_l\rangle\right)^4 \\
&= \frac{16}{n^4}\sum_{l=1}^{n}\frac{a_l^4}{\lambda_l^4}\langle Y_i,\phi_l\rangle^4\langle Y_j,\phi_l\rangle^4 && \text{(A8a)} \\
&\quad+\frac{16}{n^4}\sum_{l\ne l'}\frac{a_l^3}{\lambda_l^3}\frac{a_{l'}}{\lambda_{l'}}\langle Y_i,\phi_l\rangle^3\langle Y_j,\phi_l\rangle^3\langle Y_i,\phi_{l'}\rangle\langle Y_j,\phi_{l'}\rangle && \text{(A8b)} \\
&\quad+\frac{16}{n^4}\sum_{l\ne l'}\frac{a_l^2}{\lambda_l^2}\frac{a_{l'}^2}{\lambda_{l'}^2}\langle Y_i,\phi_l\rangle^2\langle Y_j,\phi_l\rangle^2\langle Y_i,\phi_{l'}\rangle^2\langle Y_j,\phi_{l'}\rangle^2. && \text{(A8c)}
\end{aligned}$$

Consider (A8a): Using $|\langle Y_i,\phi_l\rangle|^2 \le 4$ as before, we get $E|\langle Y_i,\phi_l\rangle|^4 \le 4E|\langle Y_i,\phi_l\rangle|^2 = 4\lambda_l$. Therefore,

$$\sum_{l=1}^{n}\frac{a_l^4}{\lambda_l^4}\langle Y_i,\phi_l\rangle^4\langle Y_j,\phi_l\rangle^4 \le 16\sum_{l=1}^{n}\frac{a_l^4}{\lambda_l^2}$$

and

$$\begin{aligned}\sum_{l=1}^{n}\frac{a_l^4}{\lambda_l^2} &= \sum_{l=1}^{n}\frac{\lambda_l^6}{(\lambda_l^2+\alpha)^4}\\ &= \sum_{l=1}^{n}\frac{\lambda_l^4}{(\lambda_l^2+\alpha)^2}\frac{\lambda_l^2}{(\lambda_l^2+\alpha)^2}\\ &\leq \sum_{l=1}^{n}\frac{\lambda_l^2}{(\lambda_l^2+\alpha)^2}\leq\frac{1}{\alpha^2}\sum_{l=1}^{n}\lambda_l^2.\end{aligned}$$

Hence,

$$\frac{\sum_{1\leq i<j\leq n}(A8a)}{q_n^2}\leq\frac{C}{\alpha^2 n^2 q_n^2}\to 0$$

as $n\alpha^{3/2}\to\infty$.

Consider (A8b):

$$E(A8b)=\frac{16}{n^4}\sum_{l\neq l'}\frac{a_l^3}{\lambda_l^3}\frac{a_{l'}}{\lambda_{l'}}E[\langle Y_i,\phi_l\rangle^3\langle Y_i,\phi_{l'}\rangle]E[\langle Y_j,\phi_l\rangle^3\langle Y_j,\phi_{l'}\rangle].$$

By Cauchy-Schwarz,

$$\begin{aligned}E[\langle Y_i,\phi_l\rangle^3\langle Y_i,\phi_{l'}\rangle] &\leq \sqrt{E[\langle Y_i,\phi_l\rangle^6]E[\langle Y_i,\phi_{l'}\rangle^2]}\\ &\leq 4\sqrt{E[\langle Y_i,\phi_l\rangle^2]E[\langle Y_i,\phi_{l'}\rangle^2]}\\ &= 4\sqrt{\lambda_l}\sqrt{\lambda_{l'}}.\end{aligned}$$

Hence,

$$E(A8b)\leq\frac{C}{n^4}\sum_{l\neq l'}\frac{a_l^3}{\lambda_l^2}a_{l'}\leq C\frac{p_n}{n^4}\sum_l\frac{a_l^3}{\lambda_l^2}.$$

Moreover,

$$\sum_l\frac{a_l^3}{\lambda_l^2}=\sum_l\frac{\lambda_l^4}{(\lambda_l^2+\alpha)^3}\leq\sum_l\frac{1}{(\lambda_l^2+\alpha)}\leq\frac{n}{\alpha}.$$

It follows that

$$\frac{\sum_{1\leq i<j\leq n}(A8b)}{q_n^2}\leq C\frac{p_n}{q_n^2 n\alpha}\to 0.$$

Now, consider (A8c):

$$\begin{aligned}E\,(A8c) &= \frac{16}{n^4}\sum_{l\neq l'}\frac{a_l^2}{\lambda_l^2}\frac{a_{l'}^2}{\lambda_{l'}^2}E[\langle Y_i,\phi_l\rangle^2\langle Y_i,\phi_{l'}\rangle^2]E[\langle Y_j,\phi_l\rangle^2\langle Y_j,\phi_{l'}\rangle^2]\\ &\leq \frac{C}{n^4}\sum_{l\neq l'}\frac{a_l^2}{\lambda_l^2}\frac{a_{l'}^2}{\lambda_{l'}^2}\lambda_l\lambda_{l'}\\ &\leq \frac{C}{n^4}\left(\sum_l\frac{a_l^2}{\lambda_l}\right)^2.\end{aligned}$$

Moreover,

$$\begin{aligned}\sum_l \frac{a_l^2}{\lambda_l} &= \sum_l \frac{\lambda_l^3}{(\lambda_l^2+\alpha)^2} = \sum_l \frac{\lambda_l^2}{(\lambda_l^2+\alpha)} \frac{\lambda_l}{(\lambda_l^2+\alpha)} \\ &\leq \sum_l \frac{\lambda_l}{(\lambda_l^2+\alpha)} \leq \frac{\sum_l \lambda_l}{\alpha}.\end{aligned}$$

Therefore,

$$\frac{\sum_{1\leq i<j\leq n}(A8c)}{q_n^2} \leq \frac{C}{\alpha^2 n^2 q_n^2} \to 0.$$

It follows that $G_I = o(\sigma(n)^4)$.

Now consider G_{II}:

$$\begin{aligned}E\left(w_{ijn}^2 w_{ikn}^2\right) &= \frac{1}{n^4} E\left[\left(\sum_{l=1}^n \frac{a_l}{\lambda_l}\langle Y_i,\phi_l\rangle\langle Y_j,\phi_l\rangle\right)^2 \left(\sum_{l'=1}^n \frac{a_{l'}}{\lambda_{l'}}\langle Y_i,\phi_{l'}\rangle\langle Y_k,\phi_{l'}\rangle\right)^2\right] \\ &= \frac{1}{n^4}\sum_{l,l'} \frac{a_l^2 a_{l'}^2}{\lambda_l^2\lambda_{l'}^2} E[\langle Y_i,\phi_l\rangle^2\langle Y_j,\phi_l\rangle^2\langle Y_i,\phi_{l'}\rangle^2\langle Y_k,\phi_{l'}\rangle^2]\end{aligned}$$

because the cross products equal zero. We have

$$\begin{aligned}E[\langle Y_i,\phi_l\rangle^2\langle Y_j,\phi_l\rangle^2\langle Y_i,\phi_{l'}\rangle^2\langle Y_k,\phi_{l'}\rangle^2] &\leq 4E[\langle Y_i,\phi_l\rangle^2]E[\langle Y_j,\phi_l\rangle^2]E[\langle Y_k,\phi_{l'}\rangle^2] \\ &= 4\lambda_l^2\lambda_{l'}.\end{aligned}$$

Hence,

$$E\left(w_{ijn}^2 w_{ikn}^2\right) \leq \frac{C}{n^4}\sum_l a_l^2 \sum_{l'} \frac{a_{l'}^2}{\lambda_{l'}} \leq \frac{Cq_n}{n^4}\frac{\sum \lambda_l}{\alpha}.$$

$$\frac{\sum_{1\leq i<j\leq n} E[w_{ijn}^2 w_{ikn}^2]}{q_n^2} \leq \frac{C}{n^2\alpha q_n} \to 0.$$

The other terms of G_{II} have the same form. Therefore, $G_{II} = o(\sigma(n)^4)$.

Consider G_{IV}:

$$\begin{aligned}&E\left(w_{ijn}w_{ikn}w_{mjn}w_{mkn}\right) \\ =\ & \frac{1}{n^4}E\left[\left(\sum_{l=1}^n \frac{a_l}{\lambda_l}\langle Y_i,\phi_l\rangle\langle Y_j,\phi_l\rangle\right)\left(\sum_{l'=1}^n \frac{a_{l'}}{\lambda_{l'}}\langle Y_i,\phi_{l'}\rangle\langle Y_k,\phi_{l'}\rangle\right)\right. \\ &\left.\left(\sum_{g=1}^n \frac{a_g}{\lambda_g}\langle Y_m,\phi_g\rangle\langle Y_j,\phi_g\rangle\right)\left(\sum_{g'=1}^n \frac{a_{g'}}{\lambda_{g'}}\langle Y_m,\phi_{g'}\rangle\langle Y_k,\phi_{g'}\rangle\right)\right] \\ =\ & \frac{1}{n^4}\sum_{l=1}^n \frac{a_l^4}{\lambda_l^4}E[\langle Y_i,\phi_l\rangle^2\langle Y_j,\phi_l\rangle^2\langle Y_m,\phi_l\rangle^2\langle Y_k,\phi_l\rangle^2]\end{aligned}$$

because $\langle Y_i, \phi_l \rangle$, $l = 1, 2...$ are uncorrelated across l. As Y_i, $i = 1, 2, ...,$ are iid, we have

$$
\begin{aligned}
E\left(w_{ijn}w_{ikn}w_{mjn}w_{mkn}\right) &= \frac{1}{n^4}\sum_{l=1}^{n}\frac{a_l^4}{\lambda_l^4}E[\langle Y_i, \phi_l\rangle^2]^4 \\
&= \frac{1}{n^4}\sum_{l=1}^{n}a_l^4 \\
&\leq \frac{q_n}{n^4}.
\end{aligned}
$$

It follows that

$$
\frac{1}{q_n^2}\sum_{1\leq i<j<k<m\leq n} E\left(w_{ijn}w_{ikn}w_{mjn}w_{mkn}\right) \leq \frac{1}{q_n} \to 0.
$$

As the other terms in G_{IV} have the same form, we can conclude that $G_{IV} = o(\sigma(n)^4)$.

All the conditions of Lemma 13 are satisfied and the result follows. □

Proof of Proposition 10. As in Carrasco and Florens (2000, proof of Theorem 10), the proof proceeds in three steps.

Step 1. Let P_n denote the projection which associates to an operator K the operator K_2 defined by the first n eigenvalues and eigenfunctions of K. We show that

$$
\frac{1}{\sqrt{q_n}}\left\{\left\|(\hat{K}^{\alpha})^{-1/2}\sqrt{n}\hat{h}(\hat{\theta})\right\| - \left\|P_n\left(K_{\omega n}^{\alpha}\right)^{-1/2}\sqrt{n}\hat{h}(\hat{\theta})\right\|\right\} \xrightarrow{P} 0 \tag{A9}
$$

under H_{2n}.

We have

$$
\begin{aligned}
&\frac{1}{\sqrt{q_n}}\left\{\left\|(\hat{K}^{\alpha})^{-1/2}\sqrt{n}\hat{h}(\hat{\theta})\right\| - \left\|P_n\left(K_{\omega n}^{\alpha}\right)^{-1/2}\sqrt{n}\hat{h}(\hat{\theta})\right\|\right\} \\
\leq\ & \frac{1}{\sqrt{q_n}}\left\|[(\hat{K}^{\alpha})^{-1/2} - P_n\left(K_{\omega n}^{\alpha}\right)^{-1/2}]\sqrt{n}\hat{h}(\hat{\theta})\right\| \\
\leq\ & \frac{1}{\sqrt{q_n}}\|P_n\|\left\|(\hat{K}^{\alpha})^{-1/2} - \left(K_{\omega n}^{\alpha}\right)^{-1/2}\right\|\left\|\sqrt{n}\hat{h}(\hat{\theta})\right\| \\
=\ & O_p\left(\frac{1}{\sqrt{q_n}n^{1/2}\alpha^{3/4}}\right)
\end{aligned}
$$

because $\|P_n\| \leq 1$, $\left\|\sqrt{n}\hat{h}(\hat{\theta})\right\| = O_p(1)$ and $\left\|(\hat{K}^{\alpha})^{-1/2} - \left(K_{\omega n}^{\alpha}\right)^{-1/2}\right\| = O_p\left(1/(n^{1/2}\alpha^{3/4})\right)$ by Lemma B.2. of Carrasco, Chernov, Florens, and Ghysels (2007). Therefore (A9) is satisfied.

Step 2. Show that

$$
\hat{p}_n - p_{n,n} \xrightarrow{P} 0 \text{ and } \hat{q}_n - q_{n,n} \xrightarrow{P} 0
$$

under H_{2n} as $n\alpha^2 \to \infty$. Under Assumption 7, this result can be established using a proof similar to those of Theorems 4 and 10 in Carrasco and Florens (2000).

Step 3. By Lemma 14, we have

$$\frac{\left\|P_n\left(K_{\omega n}^{\alpha}\right)^{-1/2}\sqrt{n}\hat{h}(\hat{\theta})\right\| - p_{n,n}}{\sqrt{q_{n,n}}} = \frac{\sum_{l=1}^{n}\frac{a_{l,n}}{\lambda_{l,n}}\left\langle\sqrt{n}\hat{h}(\hat{\theta}),\phi_{l,n}\right\rangle^2 - p_{n,n}}{\sqrt{q_{n,n}}} \xrightarrow{d} \mathcal{N}(c\left\|L_\omega\eta\right\|_K^2, 1).$$

Using steps 1 and 2, we obtain the desired result. □

Proof of Proposition 11. Let $\{\phi_j, \lambda_j\}$ be the eigenfunctions and eigenvalues of K_θ. Let ψ_j such that $\phi_j = U_\theta^*\psi_j$ and consequently $U_\theta\phi_j = U_\theta U_\theta^*\psi_j = \psi_j$. We have

$$\begin{aligned} U_\theta K_\theta U_\theta^*\psi_j &= U_\theta K_\theta \phi_j \\ &= \lambda_j U_\theta\phi_j \\ &= \lambda_j\psi_j. \end{aligned}$$

Therefore, $\{\psi_j, \lambda_j\}$ are the eigenfunctions and eigenvalues of $\tilde{K}$. It follows that

$$\begin{aligned} \left\|U_\theta\hat{h}(\theta)\right\|^2_{\left(U_\theta K_\theta U_\theta^*\right)^\alpha} &= \sum_j \frac{\lambda_j}{\lambda_j^2+\alpha}\left\langle U_\theta\hat{h}(\theta), \psi_j\right\rangle^2 \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2+\alpha}\left\langle \hat{h}(\theta), U_\theta^*\psi_j\right\rangle^2 \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2+\alpha}\left\langle \hat{h}(\theta), \phi_j\right\rangle^2 \\ &= \left\|\hat{h}(\theta)\right\|^2_{K_\theta^\alpha}, \end{aligned}$$

as desired. □

B Computational details

B.1 Theoretical covariance operator

B.1.1 Eigenvalues and eigenfunctions

As we mentioned in Section 2, the eigenvalues and eigenfunctions of the covariance operator K are the solutions to the functional equations

$$(K\phi_j)(s) = \int[\psi_0(t-s) - \psi_0(t)\psi_0(-s)]\phi_j(t)\pi(t)dt = \lambda_j\phi_j(s).$$

Given that it is not possible to find the analytical solution to this equation for arbitrary distributions, we solve for $\phi_j(s)$ at a very fine but discrete grid of M points over a finite range of values of the characteristic function argument t as follows. For the sake of brevity we describe the case in which t is scalar. Let $F(.)$ and $Q(.)$ denote the cdf and quantile functions, respectively, associated with the continuous density function $\pi(t)$, which we assume integrates to 1 over (t_l, t_u). Then, if we define $\nu = F(t)$, the usual change of variable formula immediately implies that the integral between t_l and t_u of any function $g(t)$ weighted by $\pi(t)dt$ coincides with the integral between 0 and 1 of $g[Q(\nu)]d\nu$. We exploit this equivalence to numerically approximate all the required integrals using the rectangle method over M equidistant points between 0 and 1 regardless of $\pi(.)$.

Let $\mathcal{K}$ be an $M \times M$ matrix whose elements are

$$\psi[Q(\nu_i) - Q(\nu_j)] - \psi[Q(\nu_i)]\psi[-Q(\nu_j)], \quad i, j = 1, \ldots, M,$$

so that $\mathcal{K}$ effectively gives us the asymptotic covariance matrix of the sample average of an $M \times 1$ vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, $j = 1, \ldots, M$.

Given that the eigenvalues of $\mathcal{K}$ increase with M, we work with $M^{-1}\mathcal{K}$, whose eigenvalues stabilize. In this context, we take the decreasingly ordered eigenvalues of this scaled matrix as an approximation to the decreasingly ordered eigenvalues of the theoretical covariance operator K. Similarly, we also take the normalized eigenvectors of $M^{-1}\mathcal{K}$ multiplied by $\sqrt{M}$ as an approximation to the eigenfunctions of the covariance operator scaled so that they have unit norm.

B.1.2 Test statistic

We compute the (scaled by $\sqrt{n}$) average values of the "population principal components" of the vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, $j = 1, \ldots, M$ by premultiplying the scaled sample average of this vector by the eigenfunctions previously computed and dividing the resulting expression by M.

Finally, we compute the T_B test statistic as a linear combination of the square norm of the scaled average values of those principal components weighted by $\frac{\lambda_j}{\lambda_j^2 + \alpha}$. In effect, this T_B is numerically identical to the overidentifying restriction statistic of a discrete GMM procedure based on the $M \times 1$ vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, in which we replace the inverse of the asymptotic covariance matrix $M^{-1}\mathcal{K}$ by its Tikhonov regularized inverse, as in (10).

B.2 Analytical expressions for c_{il}

B.2.1 Univariate normal

Given that the CF of the standard normal is $\psi(t) = e^{-\frac{1}{2}t^2}$, $h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i - x_l)} - e^{-\frac{1}{2}t^2 + ix_i t} - e^{-\frac{1}{2}t^2 - ix_l t} + e^{-t^2}.$$

Using a $\mathcal{N}(0, \omega^2)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int e^{it(x_i - x_l)} \pi(t) dt = e^{-\frac{1}{2}\omega^2 (x_i - x_l)^2},$$

$$c_2(x) = \int e^{-\frac{1}{2}t^2 + ixt} \pi(t) dt = \frac{e^{-\frac{\omega^2 x^2}{2(1+\omega^2)}}}{\sqrt{1+\omega^2}},$$

and

$$c_3 = \int e^{-t^2} \pi(t) dt = \frac{1}{\sqrt{1+2\omega^2}}.$$

B.2.2 Standardized uniform

Given that the CF of the standardized uniform is

$$\psi(t) = \frac{i}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}),$$

$h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i - x_l)} - \frac{ie^{ix_i t}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}) - \frac{ie^{-ix_l t}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}) - \frac{e^{-2i\sqrt{3}t}(e^{2i\sqrt{3}t} - 1)^2}{12t^2}.$$

Using a $\mathcal{N}(0, \omega^2)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int e^{it(x_i - x_l)} \pi(t) dt = e^{-\frac{1}{2}\omega^2 (x_i - x_l)^2},$$

$$\begin{aligned} c_2(x) &= \int \frac{ie^{ixt}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}) \pi(t) dt \\ &= \frac{1}{2\omega}\sqrt{\frac{\pi}{6}} \left\{ \operatorname{erf}\left[\frac{\omega(\sqrt{3}-x)}{\sqrt{2}}\right] + \operatorname{erf}\left[\frac{\omega(\sqrt{3}+x)}{\sqrt{2}}\right] \right\}, \end{aligned}$$

where erf is the error function i.e. $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, and

$$c_3 = -\int \frac{e^{-2i\sqrt{3}t}(e^{2i\sqrt{3}t} - 1)^2}{12t^2} \pi(t) dt = \frac{e^{-6\omega^2} - 1 + \sqrt{6\pi}\omega \operatorname{erf}(\sqrt{6}\omega)}{6\omega^2}.$$

B.2.3 Bivariate standard normal

Given that the CF of the bivariate normal with zero mean and identity covariance matrix is $\psi(t_1,t_2) = e^{-\frac{1}{2}(t_1^2+t_2^2)}$, $h_i(t_1,t_2)\overline{h_l(t_1,t_2)}$ has the following four terms

$$e^{i[t_1(x_{1i}-x_{1l})+t_2(x_{2i}-x_{2l})]} - e^{-\frac{1}{2}(t_1^2+t_2^2)+i(t_1x_{1i}+t_2x_{i2})} - e^{-\frac{1}{2}(t_1^2+t_2^2)-i(t_1x_{1i}+t_2x_{i2})} + e^{-t_1^2-t_2^2}.$$

Using two independent $\mathcal{N}(0,\omega^2)$ densities as weighting functions π for both t_1 and t_2, we obtain

$$c_{il} = c_1(x_i,x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i,x_l) = \int\int e^{i[t_1(x_{1i}-x_{1l})+t_2(x_{2i}-x_{2l})]}\pi(t_1)\pi(t_2)dt_1dt_2 = e^{-\frac{1}{2}\omega^2[(x_{1i}-x_{1l})^2+(x_{2i}-x_{2l})^2]},$$

$$c_2(x) = \int\int e^{-\frac{1}{2}(t_1^2+t_2^2)+i(t_1x_1+t_2x_2)}\pi(t_1)\pi(t_2)dt_1dt_2 = \frac{e^{-\frac{\omega^2(x_1^2+x_2^2)}{2(1+\omega^2)}}}{(1+\omega^2)},$$

and

$$c_3 = \int\int e^{-t_1^2-t_2^2}\pi_1(t_1)\pi_2(t_2)dt_1dt_2 = \frac{1}{1+2\omega^2}.$$

B.2.4 Standardized chi-square with 2 degrees of freedom

Given that the CF of the standardized $\chi^2(2)$ is $\psi(t) = ie^{-it}/(i+t)$, $h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i-x_l)} - \frac{ie^{it(1+x_i)}}{i+t} - \frac{ie^{it(1+x_i)}}{i-t} - \frac{ie^{-it(1+x_l)}}{i+t} - \frac{1}{(i-t)(i+t)}.$$

Using a $U(-\omega,\omega)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i,x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i,x_l) = \int e^{it(x_i-x_l)}\pi(t)dt = \frac{\sin[\omega(x_i-x_l)]}{\omega(x_i-x_l)},$$

$$\begin{aligned} c_2(x) &= \int \frac{ie^{it(1+x)}}{i+t}\pi(t)dt \\ &= \frac{e^{-(1+x)}}{2\omega}\left\{\pi - i\,\mathrm{Ci}[(\omega-i)(1+x)] + i\,\mathrm{Ci}[(\omega+i)(1+x)]\right\} \\ &\quad + \frac{e^{-(1+x)}}{2\omega}\left\{\mathrm{Si}[(\omega+i)(1+x)] - \mathrm{Si}[(i-\omega)(1+x)]\right\}, \end{aligned}$$

where Si is the sine integral function $\text{Si}(z) = \int_0^z \frac{\sin(t)}{t}dt$, Ci is the cosine integral function $\text{Ci}(z) = -\int_z^\infty \frac{\cos(t)}{t}dt$, and

$$c_3 = -\int \frac{\pi(t)dt}{(i-t)(i+t)} = \frac{\arctan(\omega)}{\omega}.$$

B.3 Classical goodness of fit tests

We briefly review below some classical goodness of fit tests (see for instance, Lehmann and Romano (2005)), which serve as benchmarks in our Monte Carlo exercise. For convenience, we present them for scalar X.

For testing $H_0 : F = F_0$ versus $H_1 : F \neq F_0$, the classical Kolmogorov-Smirnov (KS) test is based on a sup norm of the difference between the empirical distribution function $\hat{F}_n$ and the distribution function:

$$KS = \sup_{x\in\mathbb{R}} \sqrt{n} \left| \hat{F}_n(x) - F_0(x) \right|.$$

On the other hand, the Cramer-von-Mises (CvM) test is based on the L^2 norm of the difference:

$$CvM = n \int_{-\infty}^{\infty} [\hat{F}_n(x) - F_0(x)]^2 dF_0(x).$$

Finally, the Anderson-Darling (AD) test differs from the Cramer-von-Mises by the weight:

$$AD = n \int_{-\infty}^{\infty} \frac{[\hat{F}_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x).$$

So far, F_0 was completely specified. For testing normality with unknown mean and variance, the KS test is usually computed as

$$KS = \sup_{x\in\mathbb{R}} \sqrt{n} \left| \hat{F}_n(x) - \Phi\left(\frac{x - \bar{X}}{\hat{\sigma}}\right) \right|$$

where Φ is the distribution function of the standard normal and $\bar{X}$ and $\hat{\sigma}^2$ are the maximum likelihood estimators of the mean and variance. This version of the KS test is often referred to as the Lilliefors test. The other tests can be similarly modified. A multivariate extension is proposed in Andrews (1997).

Consider now the case $X_j \in \mathbb{R}^q$. To test $H_0 : \psi = \psi_0(.;\theta_0)$ versus $H_1 : \psi \neq \psi_0(.;\theta_0)$, Bierens and Wang (2012) consider a L^2 test based on the empirical characteristic function and a uniform weight:

$$BW = \int_\Upsilon \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [e^{i\tau' X_j} - \psi_0(.;\hat{\theta})] \right|^2 \frac{d\tau}{2^q \Pi_{l=1}^q \overline{\tau}_l}$$

where $\Upsilon = \times_{l=1}^q [-\overline{\tau}_l, \overline{\tau}_l]$, $\overline{\tau}_l > 0$ and $\hat{\theta}$ is a consistent estimator of θ.

These four tests are consistent against any fixed alternative to the null hypothesis and have power against $1/\sqrt{n}$ alternatives too. However, for testing general distributions with unknown parameter, their asymptotic distributions are not nuisance parameter free.

B.4 On simulating distributions

We simulate all the distributions under the null, as well as the symmetric Student t, gamma and beta distributions, using the available MATLAB routines. Namely, we use `rand.m` for the uniform, `randn.m` (`mvnrnd.m`) for the univariate (bivariate) normal, `chi2rnd.m` for the $\chi^2(2)$, `trnd.m` (`mvtrnd.m`) times $\sqrt{(\nu-2)/2}$ where ν denotes the degrees of freedom for the univariate (bivariate) symmetric Student t, `gamrnd.m` for the gamma and `betarnd.m` for the beta distribution. As for the remaining ones, the procedure is as follows.

B.4.1 Asymmetric Student *t*

The asymmetric t distribution is a special case of the Generalized Hyperbolic family with $\gamma = 0$ and $-\infty < \nu < -2$ (see Mencía and Sentana (2012)). As explained by these authors, if the number of degrees of freedom exceeds 4, we can easily simulate a standardized (zero mean, unit variance) version of a univariate asymmetric Student t distribution by exploiting its representation as a location-scale mixture of normals,

$$X_i = c(\beta,\nu,\gamma)\beta\left[\frac{(1-2\eta)}{\eta\xi_t}-1\right]+\sqrt{\frac{(1-2\eta)}{\eta\xi_i}}\sqrt{c(\beta,\nu,\gamma)}Z_i, \tag{B10}$$
$$c(\beta,\nu,\gamma)=\frac{1-4\eta}{2\eta}\frac{\sqrt{1+8\beta'\beta\eta/(1-4\eta)}-1}{2\beta'\beta}$$

where $\eta = -1/(2\nu)$, ξ_i is distributed *iid* gamma with mean η^{-1} and variance $2\eta^{-1}$, and $Z_i|\xi_i$ is *iid* $\mathcal{N}(0,1)$.

If we further assume that $\eta < 1/8$, then the skewness and kurtosis coefficients of the asymmetric t distribution will be

$$E(X_i^3) = 16c^3(\beta,\nu,\gamma)\frac{\eta^2}{(1-4\eta)(1-6\eta)}\beta^3+6c^2(\beta,\nu,\gamma)\frac{\eta}{1-4\eta}\beta$$

and

$$E(X_i^4) = 12c^4(\beta,\nu,\gamma)\frac{\eta^2(10\eta+1)}{(1-4\eta)(1-6\eta)(1-8\eta)}\beta^4$$
$$+12c^3(\beta,\nu,\gamma)\frac{\eta(2\eta+1)}{(1-4\eta)(1-6\eta)}\beta^2+3\frac{1-2\eta}{1-4\eta}c^2(\beta,\nu,\gamma).$$

Not surprisingly, we can obtain maximum asymmetry for a given kurtosis by letting $|\beta|\to\infty$.

In contrast, a standardized version of the usual symmetric Student t with $1/\eta$ degrees of freedom is achieved when $\beta = 0$ for $\eta < 1/2$. Since $\lim_{\beta \to 0} c(\beta, \nu, \gamma) = 1$, in that case the coefficient of kurtosis becomes

$$E(X_i^4) = 3\frac{1-2\eta}{1-4\eta}$$

for any $\eta < 1/4$, while the coefficient of asymmetry is obviously 0.

In the bivariate case the same location scale intepretation in (B10) applies but with $Z_{it}|\xi_i$ being *iid* $\mathcal{N}(0, I)$. However, since the elements of the resulting random vector are correlated when $\beta \neq 0$, we use the standardization procedure in Mencía and Sentana (2012).

We chose 12 degrees of freedom and $\beta = -0.75$ to avoid having too much power for both the univariate and bivariate cases. According to the above calculations, in the univariate case $E(X_i^4) = 3.75$ for the symmetric Student t, while for its asymmetric version, $E(X_i^3) = -0.54$ and $E(X_i^4) = 4.62$.

B.4.2 Discrete location-scale mixtures of normals

Univariate discrete location-scale mixtures of normals (DLSMN) Let s_i denote an *iid* Bernoulli variate with $P(s_i = 1) = \lambda$. If $z_i|s_i$ is *iid* $N(0, 1)$, then

$$X_i = \frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_i - \lambda) + \frac{s_i + (1-s_i)\sqrt{\varkappa}}{\sqrt{\lambda + (1-\lambda)\varkappa}} Z_i\right],$$

where $\delta \in \mathbb{R}$ and $\varkappa > 0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t - \lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t + (1-s_t)\sqrt{\varkappa}}{\sqrt{\lambda + (1-\lambda)\varkappa}} Z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t - \lambda)$, the sum of the two random variables will have variance $1 + \lambda(1 - \lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardized random variable is as follows

$$X_i = \begin{cases} N[\mu_1^*(\eta), \sigma_1^{*2}(\eta)] \text{ with probability } \lambda \\ N[\mu_2^*(\eta), \sigma_2^{*2}(\eta)] \text{ with probability } 1-\lambda \end{cases} \quad \text{(B11)}$$

where $\eta = (\delta, \varkappa, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\eta) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\eta) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\eta), \\
\sigma_1^{*2}(\eta) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\varkappa]}, \\
\sigma_2^{*2}(\eta) &= \frac{\varkappa}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\varkappa]} = \varkappa\sigma_1^{*2}(\eta).
\end{aligned}$$

Therefore, we can immediately interpret $\varkappa$ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\eta) - \mu_1^*(\eta)}{\sqrt{\lambda\sigma_1^{*2}(\eta) + (1-\lambda)\sigma_1^{*2}(\eta)}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components.

We can trivially extended this procedure to define and simulate standardized mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (B11) by a k-component normal mixture with mean and variance given by $\mu_1^*(\eta)$ and $\sigma_1^{*2}(\eta)$, respectively, then the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

In the case of two-component Gaussian mixtures, the parameters λ, δ and $\varkappa$ determine the higher order moments of X_i through the relationship

$$E(X_i^j) = \lambda E(x_i^j|s_i = 1) + (1-\lambda)E(x_i^j|s_i = 0),$$

where $E(X_i^j|s_i = 1)$ can be obtained from the usual normal expressions

$$\begin{aligned}
&E(X_i|s_t = 1) = \mu_1^*(\eta) \\
&E(X_i^2|s_t = 1) = \mu_1^{*2}(\eta) + \sigma_1^{*2}(\eta) \\
&E(X_i^3|s_t = 1) = \mu_1^{*3}(\eta) + 3\mu_1^*(\eta)\sigma_1^{*2}(\eta) \\
&E(X_i^4|s_t = 1) = \mu_1^{*4}(\eta) + 6\mu_1^{*2}(\eta)\sigma_1^{*2}(\eta) + 3\sigma_1^{*4}(\eta) \\
&E(X_i^5|s_t = 1) = \mu_1^{*5}(\eta) + 10\mu_1^{*3}(\eta)\sigma_1^{*2}(\eta) + 15\mu_1^*(\eta)\sigma_1^{*4}(\eta) \\
&E(X_i^6|s_t = 1) = \mu_1^{*6}(\eta) + 15\mu_1^{*4}(\eta)\sigma_1^{*2}(\eta) + 45\mu_1^{*2}(\eta)\sigma_1^{*4}(\eta) + 15\sigma_1^{*6}(\eta)
\end{aligned}$$

etc. But since $E(X_i) = 0$ and $E(X_i^2) = 1$ by construction, straightforward algebra shows that the skewness and kurtosis coefficients will be given by

$$E(X_i^3) = \frac{3\delta\lambda(1-\lambda)(1-\varkappa)}{[\lambda+(1-\lambda)\varkappa][1+\lambda(1-\lambda)\delta^2]^{3/2}} + \frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{3/2}} = a(\delta,\kappa,\lambda) \qquad \text{(B12)}$$

and

$$
\begin{aligned}
E(X_i^4) \quad = \quad & \frac{3[\lambda+(1-\lambda)\varkappa^2]}{[\lambda+(1-\lambda)\varkappa]^2[1+\lambda(1-\lambda)\delta^2]^2} + \frac{6\delta^2\lambda(1-\lambda)[(1-\lambda)+\varkappa\lambda]}{[\lambda+(1-\lambda)\varkappa][1+\lambda(1-\lambda)\delta^2]^2} \\
& + \frac{\delta^4\lambda(1-\lambda)[1-3\lambda(1-\lambda)]}{[1+\lambda(1-\lambda)\delta^2]^2} = b(\delta,\kappa,\lambda). \qquad \text{(B13)}
\end{aligned}
$$

Two issues are worth pointing out. First, $a(\delta, \kappa, \lambda)$ is an odd function of δ, which means that δ and $-\delta$ yield the same skewness in absolute value. In this sense, if we set $\delta = 0$ then we will obtain a discrete scale mixture of normals, which is always symmetric but leptokurtic.[3] Second, $b(\delta, \kappa, \lambda)$ is an even function of δ, which implies that δ and $-\delta$ give rise to the same kurtosis. For that reason, in what follows we mostly consider the case of $\delta \geq 0$.

For the symmetric alternatives, we calibrate the parameters by matching the kurtosis coefficient to that one of the Student t with 12 degrees of freedom ($E(X_i^4) = 3.75$). Since there are two parameters, we arbitrarily set the probability λ to 1/10 for the so-called "outlier case" (Panel C of Table 1) and to 3/4 for the so-called "inlier case" (Panel B of Table 1), delivering values of $\varkappa$ equal to 1/3 and $(15 - 8\sqrt{3})/11$, respectively.

As for the asymmetric mixture of three normals, we impose the same skewness and kurtosis as the normal, and fix the fifth and sixth moments to -1 and 18 (as a reference, they are 0 and 15, respectively, in the Gaussian case), which together with arbitrary weights of 0.3, 0.3, and 0.4, allow us to fully characterize the corresponding alternative.

Multivariate scale mixture of two normals $X_i = \sqrt{\varsigma_i} U_i$, with U_i being uniform on the unit sphere surface in $\mathbb{R}^N$, is distributed as a two-point discrete mixture of normals (DSMN) if and only if

$$\varsigma_i \equiv X_i' X_i = \frac{s_i + (1 - s_i)\varkappa}{\lambda + (1 - \lambda)\varkappa} \varsigma_i^o$$

where s_i is an *iid* Bernoulli variate with $P(s_i = 1) = \lambda$, $\varkappa$ is the variance ratio of the two components, which for identification purposes we restrict to be in the range $(0, 1]$ and ς_i^o is an independent $\chi^2(N)$. The DSMN approaches the multivariate normal when $\varkappa \to 1$, $\alpha \to 1$ or $\alpha \to 0$. Near the limit, though, the distributions can be radically different. For instance, given that $\varkappa \in (0, 1]$ when $\alpha \to 0^+$ there are very few observations with very large variance ("outliers case"), while when $\alpha \to 1^-$ the opposite happens, very few observations with very small variance ("inliers case"). As all scale mixtures of normals, the distribution of x_i is leptokurtic.

We calibrate the bivariate outlier distribution (Panel C of Table 3) by following the same steps as in the univariate case.

[3] Another way of obtaining discrete normal mixture distributions that are symmetric is by making $\lambda = \frac{1}{2}$ and $\varkappa = 1$.

B.4.3 Standardized second order Hermite expansion of the standard normal

The standardized version of the density in Lemma 7 that we use as alternative to the univariate normal can be written as

$$f(x;a,b) = \frac{e^{-\frac{1}{2}(a+cx)^2} c[1+a^2+acx]}{\sqrt{2\pi}} + \frac{e^{-\frac{1}{2}(a+cx)^2} bc[(a+cx)^2-1]}{2\sqrt{\pi}}$$

where $c = \sqrt{1-a^2+b\sqrt{2}}$. Moreover, we can obtain an analytical expression for the corresponding cdf in terms of the error function erf,

$$F(x;a,b) = \frac{1}{2}\left[1+\text{erf}\left(\frac{a+cx}{\sqrt{2}}\right)\right] - \frac{e^{-\frac{1}{2}(a+cx)^2}[a(b+\sqrt{2})+bcx]}{2\sqrt{\pi}},$$

which is the basis for simulating from this distribution. Specifically, we generate a uniform random number u between 0 and 1 and then numerically find the root x to the equation $F(x;a,b) = u$.

B.4.4 Scaled *F*

If we assume that X_i is *iid* as a standardized symmetric multivariate t with ν degrees of freedom, then

$$X_i = \sqrt{\frac{(\nu-2)\zeta_i}{\xi_i}} U_i$$

where U_i is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_i is a $\chi^2(N)$, ξ_i is a $\chi^2(\nu)$, and u_i, ζ_i, and ξ_i are mutually independent. Therefore, we can easily generate a scaled F random variable with mean N from the square Euclidean norm of an N-variate Student t with finite degrees of freedom.

References

Aït-Sahalia, Y., P.J. Bickel and T.M. Stoker (2001): "Goodness-of-fit tests for regression using kernel methods", *Journal of Econometrics* 105, 363–412.

Amengual, D. and E. Sentana (2012): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Andrews, D. (1997): "A conditional Kolmogorov test", *Econometrica* 65, 1097–1128.

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

Bierens, H. (1982) "Consistent model specification tests", *Journal of Econometrics* 20, 105-134.

Bierens, H. and W. Ploberger (1997) "Asymptotic theory of integrated conditional moment tests", *Econometrica* 65,1129-1152.

Bierens, H.J. and L. Wang (2012): "Integrated conditional moment tests for parametric conditional distributions", *Econometric Theory* 28, 328–362.

Bontemps, C. and N. Meddahi (2005): "Testing normality: A GMM approach", *Journal of Econometrics* 124, 149–186.

Bontemps, C. and N. Meddahi (2012): "Testing distributional assumptions: A GMM approach", *Journal of Applied Econometrics* 27, 978–1012.

Carrasco, M., M. Chernov, J.P. Florens, and E. Ghysels (2007): "Efficient estimation of general dynamic models with a continuum of moment conditions", *Journal of Econometrics* 140, 529–573.

Carrasco, M. and J.P. Florens (2000): "Generalization of GMM to a continuum of moment conditions", *Econometric Theory* 16, 797–834.

Carrasco, M. and J.P. Florens (2014): "On the asymptotic efficiency of GMM", *Econometric Theory* 30, 372–406.

Carrasco, M., J.P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization", In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics* vol. 6B, pp. 5633–575. Elsevier.

Chen, B. and Y. Hong (2010): "Characteristic function-based testing for multifactor continuous-time Markov models via nonparametric regression", *Econometric Theory* 26, 1115–1179.

Chen, X. and H. White (1992): "Central limit and functional central limit theorems for Hilbert space-valued dependent processes", Working Paper, University of California, San Diego.

Chen, X. and H. White (1998): "Central limit and functional central limit theorems for Hilbert space-valued dependent processes", *Econometric Theory* 14, 260–284.

De Jong, R.M. and H.J. Bierens (1994): "On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity", *Econometric Theory* 9, 70–90.

Dufour, J.-M. (2006): "Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics", *Journal of Econometrics* 133, 443–477.

Dufour, J.-M. and P. Valery (2016): "Hypothesis tests when rank conditions fail: A smooth regularization approach", mimeo, HEC Montréal.

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

Escanciano, J.C. (2009): "On the lack of power of omnibus specification tests", *Econometric Theory* 25, 162–194.

Eubank, R.L. and V.N. LaRiccia (1992): "Asymptotic comparison of Cramer-von Mises and nonparametric function estimation techniques for resting goodness-of-fit", *The Annals of Statistics* 20, 2071–2086.

Fan, Y. (1997): "Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function", *Journal of Multivariate Analysis* 62, 36–63.

Feuerverger, A. and R.A. Mureika (1977): "The empirical characteristic function and its applications", *The Annals of Statistics* 5, 88–97.

Fiorentini G., E. Sentana and G. Calzolari (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations", *Journal of Business and Economic Statistics* 21, 532–546.

Genest C., W. Huang, and J.-M. Dufour (2013): "A regularized goodness-of-fit test for copulas", *Journal de la Société Française de Statistique*, 154, 64–77.

Ghosh, S. and F. Ruymgaart (1982): "Applications of empirical characteristic functions in some multivariate problems", *The Canadian Journal of Statistics* 20, 429–440.

Hall, P. and A.H. Welsh (1983): "A test for normality based on the empirical characteristic function", *Biometrika* 70, 485–489.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029–1054.

Hansen, L.P., J. Heaton, and A. Yaron (1996): "Finite-sample properties of some alternative GMM estimators", *Journal of Business & Economic Statistics* 14, 262-280.

Härdle, W. and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits", *Annals of Statistics* 21, 1926-1947.

Hong, Y. (1999): "Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach", *Journal of the American Statistical Association* 1201–1220.

Horowitz, J. and N.E. Savin (2000): "Empirically relevant critical values for hypothesis tests: a bootstrap approach", *Journal of Econometrics* 95, 375–389.

Imhof, J.P. (1961): "Computing the distribution of quadratic forms in normal variables", *Biometrika* 48, 419–426.

Janssen, A. (2000): "Global power functions of goodness of fit tests", *Annals of Statistics* 28, 239–253.

Jarque, C.M. and A. Bera (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economics Letters* 6, 255–259.

Kiefer, N.M. and T.J. Vogelsang (2002): "Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size", *Econometric Theory* 18, 1350–1366.

Kress, R. (1999): *Linear integral equations*. Springer-Verlag.

Fan, Y. and Q. Li (2000) "Consistent model specification Tests", *Econometric Theory* 16, 1016-1041.

Lehmann, E.L. and J.P. Romano (2005): *Testing statistical hypotheses*. Springer.

Leucht, A. (2012): "Characteristic function-based hypothesis tests under weak dependence", *Journal of Multivariate Analysis* 108, 67–89.

Lukacs, E. (1960): *Characteristic functions.* New York, Hafner Pub. Co.

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

Mencía J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with normal and Student t innovations", *Review of Economics and Statistics* 94, 133–152.

Neuhaus, G. (1976): "Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives", *Journal of Multivariate Analysis* 6, 95–110.

Neyman, J. (1937): "Smooth test for goodness of fit", *Skand. Aktuarietidskr.* 20, 150–199.

Rayner, J. and D. Best (1989): *Smooth test of goodness of fit.* Oxford University Press, Oxford.

Sefton, M. (1992): "A lagrange multiplier test of uniformity against beta alternatives", *Journal of Statistical Computation and Simulation* 44, 17–25.

Singleton, K. (2001): "Estimation of affine pricing models using the empirical characteristic function", *Journal of Econometrics* 102, 111–141.

Su, L. and H. White (2007): "A consistent characteristic function-based test for conditional independence", *Journal of Econometrics* 141, 807–834.

Tuvaandorj, P. and V. Zinde-Walsh (2014): "Limit theory and inference about conditional distributions", *Advances in Econometrics* 33, 397–423.

van der Vaart, A.W. (1998): *Asymptotic statistics.* Cambridge University Press.

van der Vaart, A.W. and J.A. Wellner (1996): *Weak convergence and empirical processes, with applications to statistics.* Springer-Verlag.

Wahba, G. (1975): "Smoothing noisy data with spline functions", *Numerische Mathematik* 24, 383–393.

Table 1: Bootstrap-based size corrected rejection rates at 1%, 5%, and 10% significance levels for univariate Gaussian null hypothesis

	$J(\hat{\theta}, K_{\hat{\theta}})$				$J(\hat{\theta}, \mathring{K})$				Other consistent tests				LM tests	
	$\alpha = .1$		$\alpha = .01$		$\alpha = 0.01$		$\alpha = 0.01$							
ω	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	KS	CvM	AD	BW	S t	A t
Panel A: Student t with 12 degrees of freedom														
10%	15.0	21.9	15.0	23.9	26.8	8.8	28.6	8.5	15.1	16.9	20.1	12.5	31.4	32.3
5%	8.0	13.9	8.0	15.5	16.8	4.3	19.3	4.7	8.8	9.5	11.7	6.4	25.4	24.3
1%	2.5	4.5	2.3	5.9	5.5	1.1	7.0	1.3	1.7	3.1	4.1	1.1	12.9	12.4
Panel B: Scale mixture of two normals. the outliers case														
10%	16.3	23.4	15.7	24.2	23.9	11.9	26.0	11.6	17.5	18.7	20.3	13.3	30.9	33.3
5%	9.4	15.3	9.1	15.6	14.8	6.4	17.5	6.3	10.8	11.3	12.3	7.1	24.8	25.7
1%	3.0	5.5	2.8	6.4	5.2	1.7	6.3	1.7	2.9	4.0	4.8	1.6	13.6	14.0
Panel C: Scale mixture of two normals, the inliers case														
10%	21.2	21.2	19.8	22.9	23.8	5.1	19.2	4.0	21.4	21.7	21.4	7.8	21.7	23.7
5%	13.4	13.3	12.6	14.6	14.8	2.4	12.1	1.6	14.0	13.9	13.5	3.5	15.6	15.4
1%	5.4	4.1	4.7	5.3	4.6	0.4	3.4	0.2	4.2	5.3	5.3	0.7	5.5	4.9

Notes: Results based on 10,000 samples of size $n = 100$. Critical values are computed using parametric bootstrap. $J(\hat{\theta}, K_{\hat{\theta}})$ and $J(\hat{\theta}, \mathring{K})$ denote the test that uses the theoretical covariance operator and the centred sample estimator of the covariance operator, respectively. α denotes the regularisation parameter and ω is the scale parameter of the $\mathcal{N}(0, \omega^2)$ density defining inner products. S t and A t are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t, respectively. KS, CvM and AD denote the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling tests while BW denotes the test in Bierens and Wang (2012).

Table 1 (cont.): Bootstrap-based size corrected rejection rates at 1%, 5%, and 10% significance levels for univariate Gaussian null hypothesis

	$J(\hat{\theta}, K_{\hat{\theta}})$				$J(\hat{\theta}, \mathring{K})$									
	$\alpha = .1$		$\alpha = .01$		$\alpha = 0.01$		$\alpha = 0.01$		Other consistent tests				LM tests	
ω	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	KS	CvM	AD	BW	S t	A t
Panel D: Asymmetric Student t with 12 degrees of freedom and $\beta = -0.75$														
10%	25.6	43.8	24.2	43.2	35.3	13.9	39.1	17.1	28.7	33.7	37.7	17.4	40.9	51.3
5%	16.6	34.1	15.3	33.1	24.5	8.0	29.4	11.2	19.7	23.3	27.5	10.3	34.9	41.6
1%	6.2	17.4	5.3	17.4	9.6	2.8	13.3	4.9	6.7	10.8	13.9	3.0	21.8	25.2
Panel E: Asymmetric (but with zero skewess), mesokurtic, location-scale mixture of three normals														
10%	16.7	25.3	20.4	25.1	13.4	29.4	20.9	37.3	15.6	20.3	25.9	12.6	21.1	22.5
5%	7.7	14.7	11.4	13.9	7.7	15.6	13.0	23.4	7.6	9.9	14.4	6.5	9.0	9.7
1%	1.4	3.6	2.8	2.8	2.6	3.7	3.7	6.5	1.1	1.8	3.7	1.2	1.1	2.4
Panel F: Standardized second order Hermite expansion of the standard normal														
10%	45.2	57.7	40.6	56.0	16.4	18.2	20.7	31.0	45.1	52.0	53.3	24.7	9.2	38.0
5%	32.3	45.3	28.7	42.6	9.0	10.6	12.8	19.5	32.8	39.1	40.2	16.1	3.2	18.6
1%	14.9	20.8	11.9	19.3	2.2	3.0	3.6	6.7	12.5	19.4	20.1	5.2	0.4	2.2

Results based on 10,000 samples of size $n = 100$. Critical values are computed using parametric bootstrap. $J(\hat{\theta}, K_{\hat{\theta}})$ and $J(\hat{\theta}, \mathring{K})$ denote the test that uses the theoretical covariance operator and the centred sample estimator of the covariance operator, respectively. α denotes the regularisation parameter and ω is the scale parameter of the $\mathcal{N}(0, \omega^2)$ density defining inner products. S t and A t are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t, respectively. KS, CvM and AD denote the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling tests while BW denotes the test in Bierens and Wang (2012).

Table 2: Bootstrap-based size corrected rejection rates at 1%, 5%, and 10% significance levels for univariate Uniform null hypothesis

	$J(\theta_0, K)$				$J(\theta_0, \mathring{K})$				Other consistent tests				Directional tests	
	$\alpha = .1$		$\alpha = .01$		$\alpha = 0.01$		$\alpha = 0.01$							
ω	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	KS	CvM	AD	BW	LM	BM
Panel A: Symmetric Beta with parameters $a = b = 1.1$														
10%	22.6	24.6	25.1	25.5	17.7	15.0	21.5	20.9	10.4	10.3	10.0	10.7	12.6	13.7
5%	13.5	14.6	15.2	15.9	9.5	7.6	11.9	11.8	5.0	4.5	4.4	5.5	5.6	7.7
1%	4.0	4.5	4.5	5.0	2.1	1.7	3.1	3.1	0.9	0.8	0.8	1.0	0.8	1.6
Panel B: Asymmetric Beta with parameters $a = 1.1$ and $b = 1$														
10%	19.3	19.2	19.5	17.3	17.4	18.2	19.4	21.9	18.9	20.9	20.7	15.0	18.2	18.3
5%	11.5	11.0	11.4	9.8	9.7	10.9	11.1	13.2	11.7	12.4	12.4	8.5	9.8	11.2
1%	3.4	3.3	3.1	2.5	2.3	3.3	2.9	4.1	2.9	3.4	3.7	2.6	2.5	2.8
Panel C: Standard Gaussian PITs of obs. drawn from a univariate asymmetric Student t with 12 df and β =-.75														
10%	27.4	26.8	26.4	27.1	16.6	11.6	18.7	18.5	13.8	13.7	14.5	11.8	16.6	14.2
5%	17.5	16.3	16.9	16.7	9.4	6.5	10.8	10.9	8.0	6.8	6.8	6.4	9.6	8.1
1%	6.6	5.8	6.1	5.9	2.6	1.7	3.3	3.4	1.6	1.4	1.3	1.2	3.8	1.7

Results based on 10,000 samples of size $n = 100$. Critical values are computed using parametric bootstrap. $J(\theta_0, K)$ and $J(\theta_0, \mathring{K})$ denote the test that uses the theoretical covariance operator and the centred sample estimator of the covariance operator, respectively. α denotes the regularisation parameter and ω is the scale parameter of the $\mathcal{N}(0, \omega^2)$ density defining inner products. LM is the LM test of uniform vs beta proposed by Selton (1992) and BM is a moment test based on the first two Jacobi polynomials proposed by Bontemps and Meddahi (2012). KS, CvM and AD denote the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling tests while BW denotes the test in Bierens and Wang (2012).

Table 3: Bootstrap-based size corrected rejection rates at 1%, 5%, and 10% significance levels for bivariate Gaussian null hypothesis

	$J(\hat{\theta}, K_{\hat{\theta}})$				$J(\hat{\theta}, \mathring{K})$				LM tests	
	$\alpha = .1$		$\alpha = .01$		$\alpha = 0.01$		$\alpha = 0.01$			
ω	$\sqrt{10}$	1	$\sqrt{10}$	$\omega = 1$	$\sqrt{10}$	1	$\sqrt{10}$	1	S t	A t
Panel A: Student t with 12 degrees of freedom										
10%	20.5	30.1	20.0	32.1	31.0	17.1	35.5	43.7	45.7	47.5
5%	13.1	20.3	13.1	22.7	19.7	10.1	23.7	33.4	39.7	38.8
1%	3.8	8.5	3.7	11.1	5.4	3.4	8.0	16.8	25.9	23.5
Panel B: Scale mixture of two normals, the outliers case										
10%	16.8	27.2	16.5	28.4	23.7	22.1	27.4	44.5	44.4	48.3
5%	10.1	18.1	9.9	20.0	14.2	13.6	17.1	33.7	38.8	39.8
1%	2.5	7.2	2.4	9.3	3.5	4.6	5.1	17.2	25.4	24.8
Panel C: Asymmetric Student t with 12 degrees of freedom and $\beta_1 = \beta_2$ =-0.75										
10%	39.7	65.9	38.4	64.7	40.8	33.1	46.7	63.6	59.6	75.0
5%	29.1	55.6	28.7	54.9	27.7	23.5	34.0	54.0	54.4	67.4
1%	12.1	37.2	11.6	38.8	9.2	11.3	14.0	36.1	40.7	49.5

Results based on 10,000 samples of size $n = 100$. Critical values are computed using parametric bootstrap. $J(\hat{\theta}, K_{\hat{\theta}})$ and $J(\hat{\theta}, \mathring{K})$ denote the test that uses the theoretical covariance operator and the centred sample estimator of the covariance operator, respectively. α denotes the regularisation parameter and ω is the scale parameter of the $\mathcal{N}(0, \omega^2)$ density defining inner products. S t and A t are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t, respectively.

Table 4: Bootstrap-based size corrected rejection rates at 1%, 5%, and 10% significance levels for univariate χ^2_2 null hypothesis

	$J(\theta_0, K)$				$J(\theta_0, \mathring{K})$				Other consistent tests				LM tests	
	$\alpha = .1$		$\alpha = .01$		$\alpha = .1$		$\alpha = .01$							
ω	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	$\sqrt{10}$	1	KS	CvM	AD	BW	F	G
Panel A: Scaled F distribution with 2 and 12 degrees of freedom														
10%	31.9	35.6	31.7	28.5	20.7	7.3	13.0	5.4	22.8	23.4	25.1	14.9	51.8	34.1
5%	20.0	22.8	20.3	17.4	13.5	4.7	6.9	3.2	14.3	14.4	15.4	8.5	43.9	25.6
1%	6.1	7.8	6.2	5.7	4.6	1.6	2.2	1.1	4.3	4.4	4.7	2.3	26.5	12.0
Panel B: Gamma distribution with shape parameter $\alpha = 2/3$ and scale parameter $\beta = 3$														
10%	29.0	41.8	31.8	27.0	22.1	5.6	7.2	3.5	59.3	61.3	80.0	29.6	67.4	95.7
5%	17.9	29.0	20.7	16.8	15.3	4.0	3.7	1.9	45.7	46.2	68.1	18.9	56.2	93.2
1%	5.6	12.0	6.6	5.7	6.4	1.4	1.3	0.8	22.1	22.2	41.8	6.0	28.8	84.0
Panel C: Square norm of observations drawn from a bivariate asymmetric Student t with 12 df and β =-.75ℓ														
10%	40.6	47.1	40.5	39.4	28.7	9.2	18.7	6.9	29.0	29.8	32.1	17.9	62.1	40.9
5%	27.1	32.5	27.6	26.1	19.4	7.0	11.4	4.6	18.7	20.0	21.3	10.4	55.6	32.3
1%	9.3	13.2	9.4	9.9	7.3	3.2	4.1	2.2	6.5	7.1	7.9	3.3	39.7	17.9

Results based on 10,000 samples of size $n = 100$. Critical values are computed using parametric bootstrap. $J(\theta_0, K)$ and $J(\theta_0, \mathring{K})$ denote the test that uses the theoretical covariance operator and the centred sample estimator of the covariance operator, respectively. α denotes the regularisation parameter and ω is the scale parameter of the $U[-\omega, \omega]$ density defining inner products. F is the LM test of chi square with N degrees of freedom versus F with the same number of degrees of freedom in the numerator but degrees of freedom in the denominator proposed by Fiorentini, Sentana and Calzolari (2003) and G the LM test against a gamma distribution with mean N but shape parameter $\alpha \neq N/2$ developed in Amengual and Sentana (2012). KS, CvM and AD denote the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling tests while BW denotes the test in Bierens and Wang (2012).

Figure 1: Examples of characteristic functions

Figure 1a: Standard normal

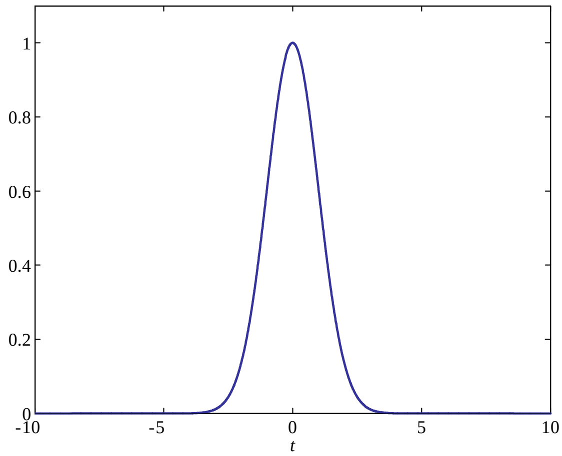


Figure 1b: Standardized uniform

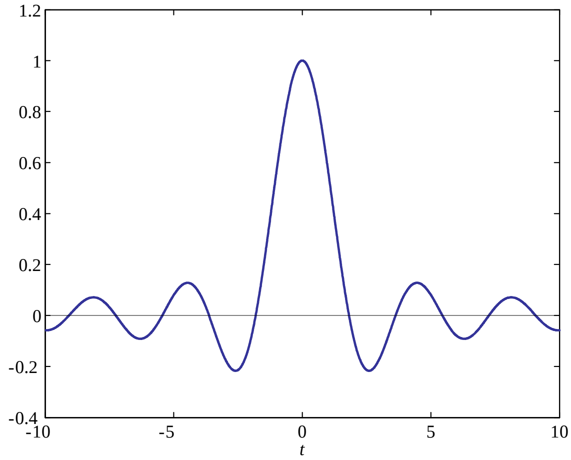


Figure 1c: Standardized $\chi^2(2)$

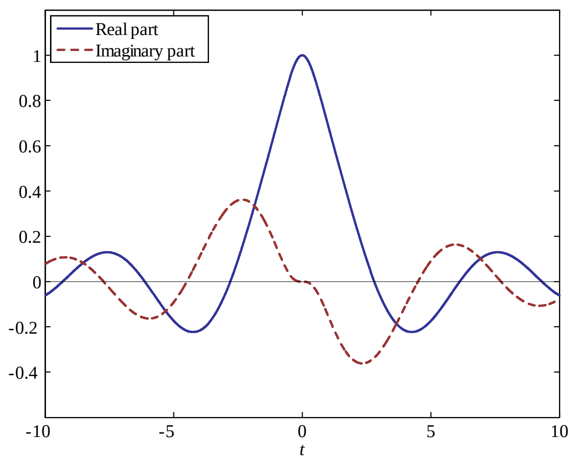


Figure 2: Eigenvalues and eigenfunctions of the covariance operator K

Figure 2a: 1^{st} eigenfunction of K for the standard normal

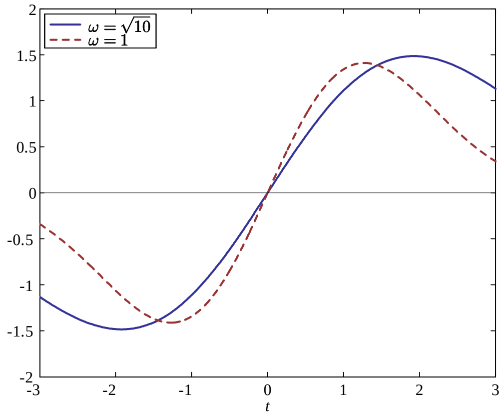


Figure 2b: 1^{st} eigenfunction of K for the (standardized) uniform

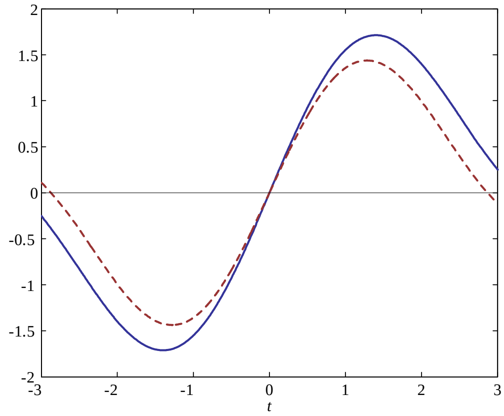


Figure 2c: 2^{nd} eigenfunction of K for the standard normal

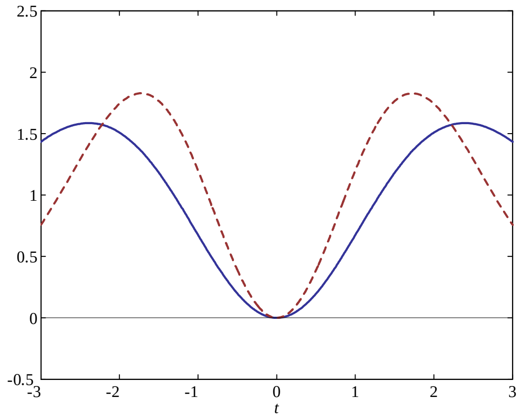


Figure 2d: 2^{nd} eigenfunction of K for the (standardized) uniform

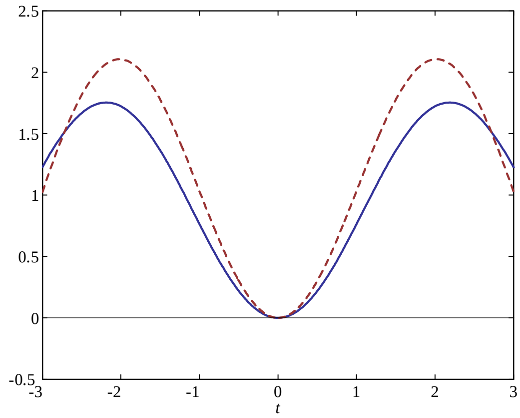


Figure 2e: Eigenvalues of K (in logs) for the standard normal

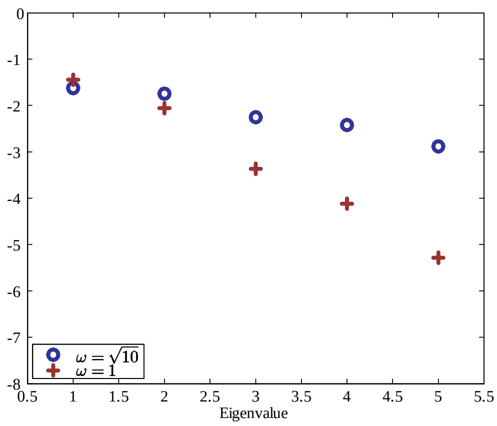
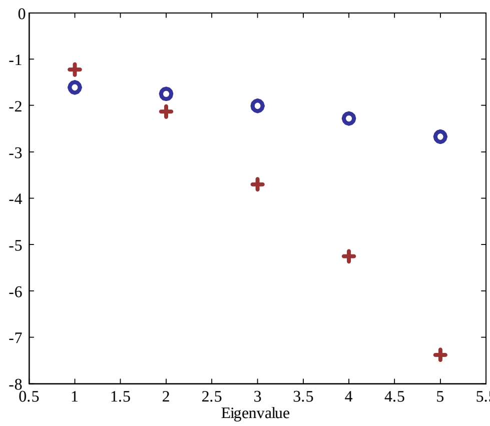
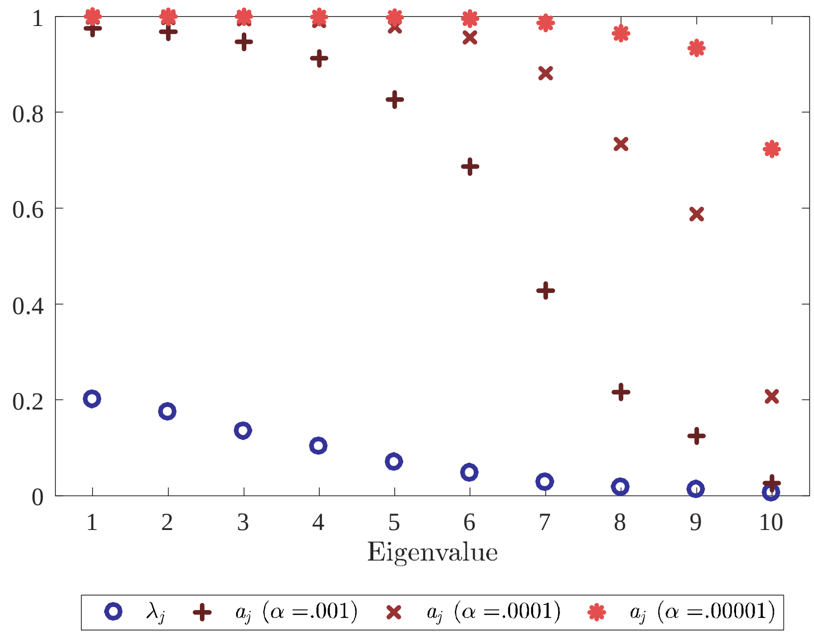


Figure 2f: Eigenvalues of K (in logs) for the (standardized) uniform



Notes: Eigenvalues and eigenfunctions are computed following the procedure described in Appendix B.1 with a grid of 1,000 points.

Figure 3: Eigenvalues (λ_j's) and weights (a_j's) of the covariance K for the standardized Uniform distribution



Notes: Eigenvalues are computed following the procedure described in Appendix B.1 with a grid of 1,000 points.

Figure 4: Densities of alternatives to the univariate normal

Figure 4a: Symmetric Student t

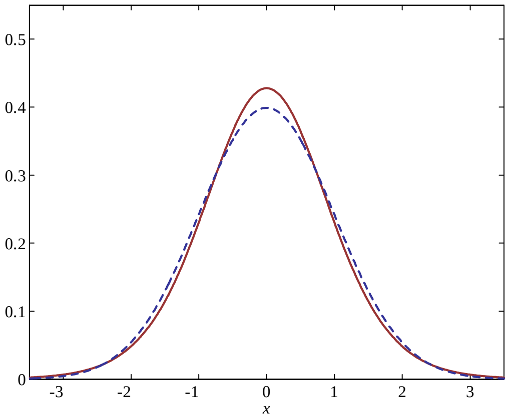


Figure 4b: Asymmetric Student t

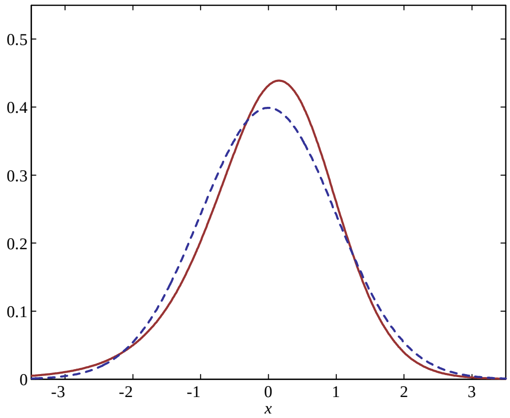


Figure 4c: Scale mixture of two normals (outliers case)

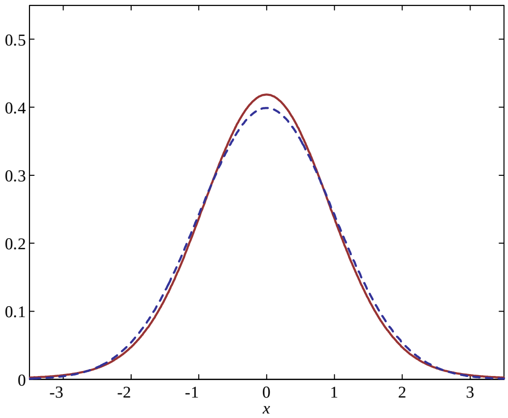


Figure 4d: Third-moment symmetric and mesokurtic Gaussian mixture

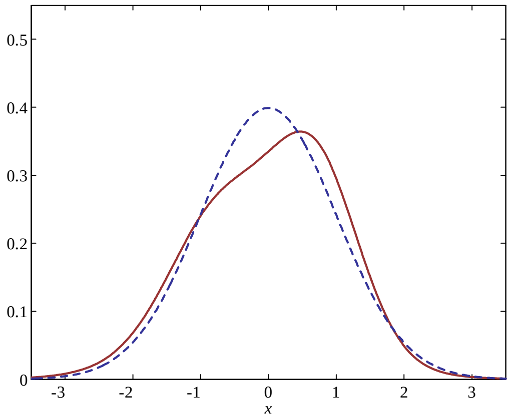


Figure 4c: Scale mixture of two normals (inliers case)

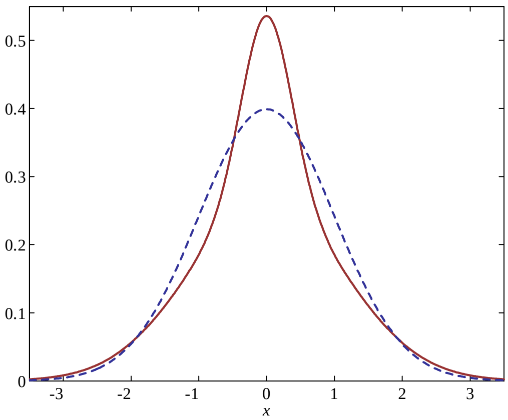
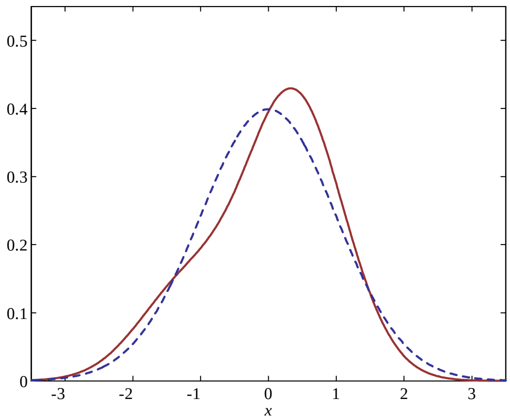


Figure 2f: Second-order Hermite expansion of the normal



Notes: Figure 4a: Student t with 12 degrees of freedom. Figure 4b: Asymmetric t with 12 degrees of freedom and skewness parameter $\beta = -.75$. Figure 4c: Discrete scale mixture with same kurtosis as the symmetric t, 3.75, and $\lambda = 0.1$ (outlier). Figure 4d: Discrete location-scale mixture of three normals with same skewness and kurtosis as the normal and $E(x^5) = -1$, $E(x^6) = 18$. Figure 4e: Discrete scale mixture with kurtosis 3.75 and $\lambda = 0.75$ (inlier). Figure 4f: Second order expansion with $a = 0.4$ and $b = 0.5$. See Appendix B.3 for parameter definitions.

Figure 5: Densities of alternatives to the uniform distribution

Figure 5a: Symmetric beta (with parameters $a = b = 1.1$)

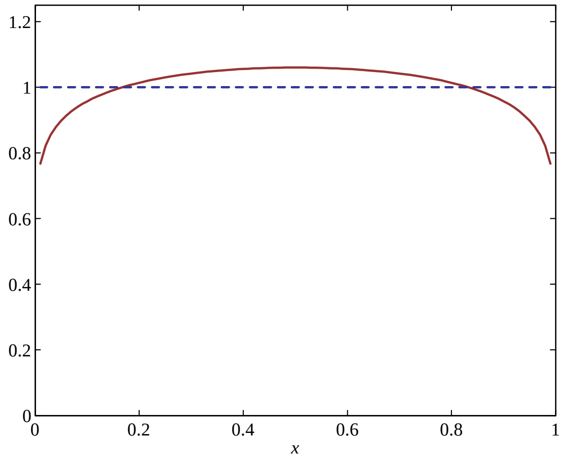


Figure 5b: Asymmetric beta (with parameters $a = 1.1$, $b = 1$)

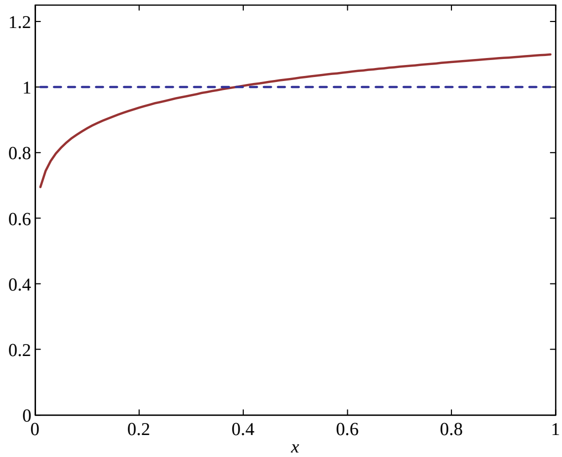
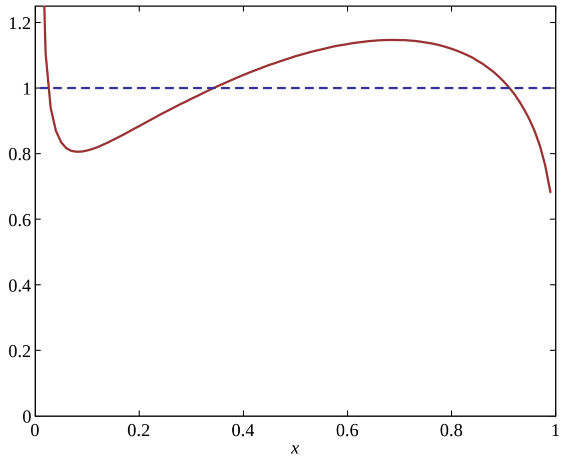


Figure 5c: Gaussian PITs of observations
drawn from an asymmetric Student t



Notes: Figure 5c: asymmetric Student t distribution with 12 degrees of freedom and skewness parameter $\beta = -.75$. Density of Figure 5c is computed as the ratio of the pdfs of the asymmetric Student t and the normal.

Figure 6: Alternative distributions to the bivariate normal

Figure 6a: Symmetric Student t density

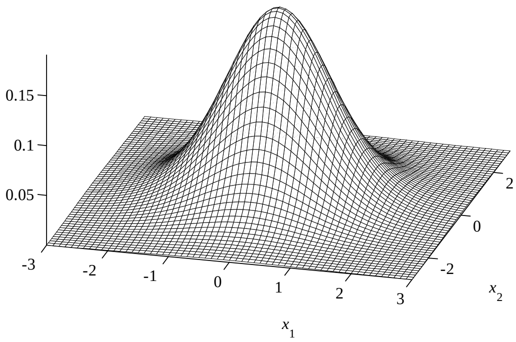


Figure 6b: Contours of a symmetric Student t

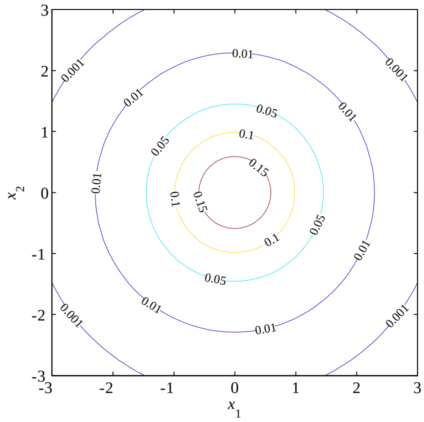


Figure 6c: Scale mixture of two normals (outliers case) density

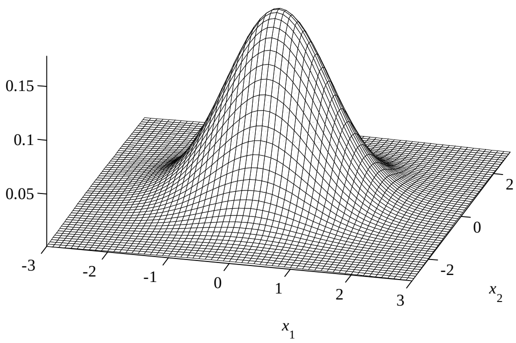


Figure 6d: Contours of a scale mixture of two normals (outliers case)

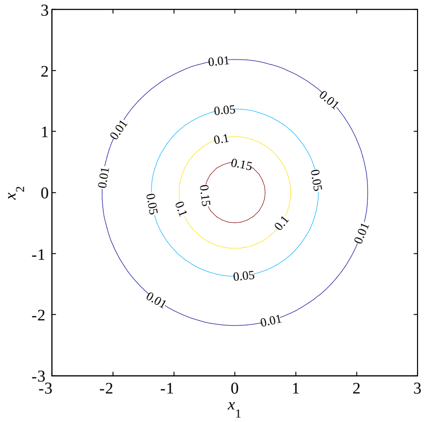


Figure 6e: Asymmetric Student t density

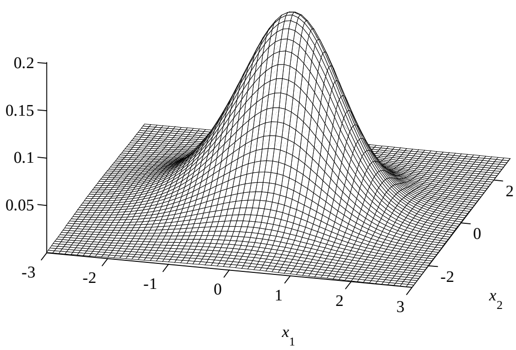
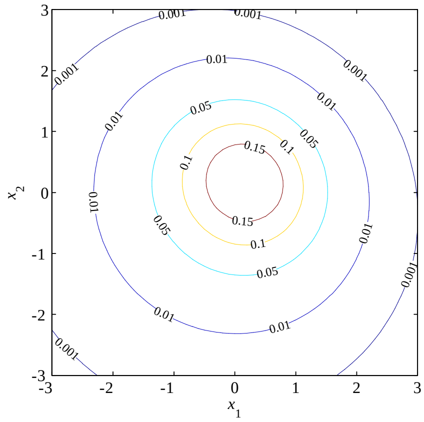


Figure 6f: Contours of an Asymmetric Student t



Notes: Figures 4a–b: Student t with 12 degrees of freedom. Figures 4c–d: Scale mixture with same Mardia's excess kurtosis coefficient as the symmetric t, 0.5, and $\lambda = 0.1$. Figures 4e–f: Asymmetric t with 12 degrees of freedom and skewness parameter $\beta = -.75\ell$. See Appendix B.3 for parameter definitions.

Figure 7: Densities of alternatives to the $\chi^2(2)$

Figure 7a: Scaled F with 2 and 12 degrees of freedom

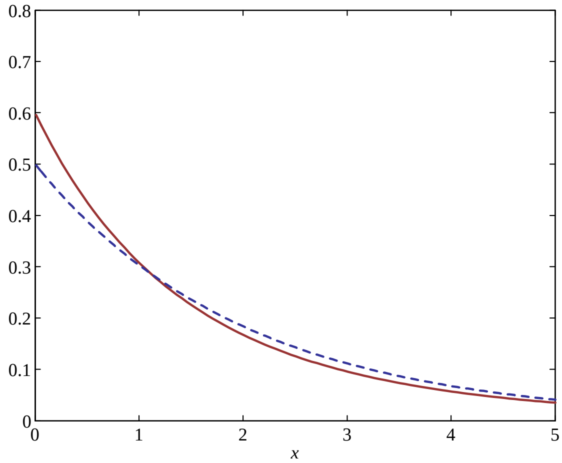


Figure 7b: Gamma with parameters $\alpha = 2/3$ and $\beta = 3$

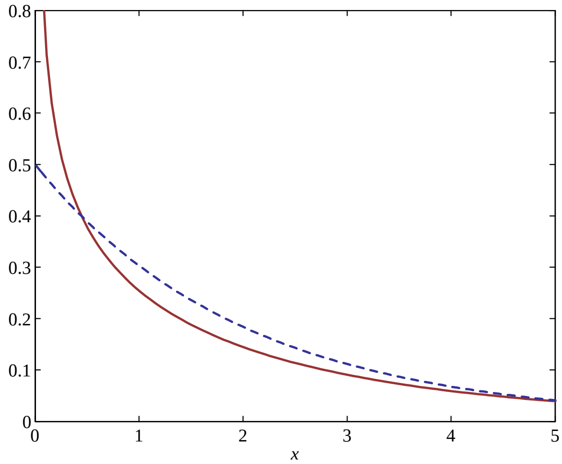
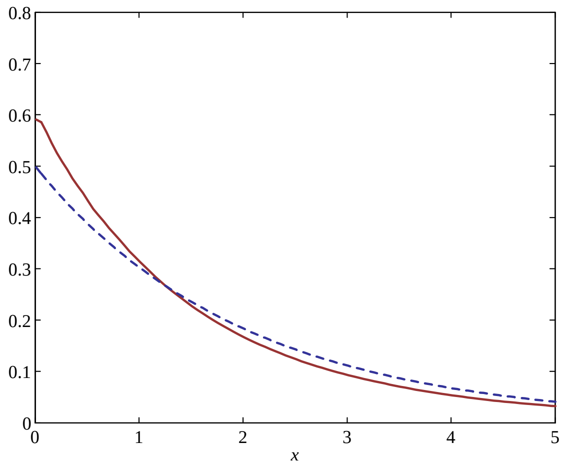


Figure 7c: Square norm of bivariate draws from asymmetric Student t



Notes: Figure 7c: asymmetric Student t distribution with 12 degrees of freedom and skewness parameter vector $\beta = -.75\ell$. Density of Figure 7c was computed by nonparametric estimation of a simulated sample of size 5,000,000.