Inference in multivariate dynamic models with elliptical innovations*

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

15 February 2011

Preliminary and incomplete.
Please do not cite without permission.

Abstract

We obtain analytical expressions for the score of conditionally heteroskedastic dynamic regression models when the conditional distribution is elliptical. We pay special attention not only to the Student t and Kotz distributions, but also to flexible families such as discrete scale mixtures of normals and polynomial expansions. We derive score tests for multivariate normality versus those elliptical distributions. The alternative tests for multivariate normality present power properties that differ substantially under different alternative hypotheses. Finally, we illustrate the small sample performance of the alternative tests through Monte Carlo simulations.

Keywords: Financial Returns, Elliptical Distributions, Normality Tests.

JEL: C12, C13, C51, C52

*We would like to thank Manuel Arellano, Olivier Faugeras, Gabriele Fiorentini, Javier Mencía, Francisco Peñaranda and David Veredas, as well as participants at CEMFI Econometrics Workshop, the 2010 Toulouse School of Economics Financial Econometrics Conference, the XVIII Finance Forum (Elche) and the XXXV Symposium on Economic Analysis (Madrid) for useful comments and suggestions. Carlos González provided excellent research assistance. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 is gratefully acknowledged.

1 Introduction

Many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering effects. Nevertheless, the Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) remain consistent for the conditional mean and variance parameters in those circumstances, so long as those moments are correctly specified.

However, a non-normal distribution may be indispensable when one is interested in features of the distribution of asset returns beyond its conditional mean and variance. For instance, empirical researchers and financial market practitioners are often interested in the so-called Value at Risk of an asset, which is the positive threshold value V such that the probability of the asset suffering a reduction in wealth larger than V equals some pre-specified level $\varkappa < 1/2$. In addition, they are sometimes interested in the probability of the joint occurrence of several extreme events, which is regularly underestimated by the multivariate normal distribution, especially in larger dimensions. This naturally leads one to specify a parametric leptokurtic distribution for the standardised innovations, such as the multivariate Student t analysed in Fiorentini, Sentana and Calzolari (2003) (hereinafter FSC), and to estimate not only the conditional mean and variance parameters but also the parameters characterising the shape of the assumed distribution. Elliptical distributions such as the multivariate t are attractive in this context because they relate mean-variance analysis with expected utility maximisation (see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997)). Moreover, they generalise the multivariate normal distribution, but at the same time they retain its analytical tractability irrespective of the number of assets, as opposed to e.g. non-parametric methods, which become infeasible when the number of assets is moderately large.

Despite its attractiveness, though, the multivariate Student t distribution, which is a member of the elliptical family, rules out platykurtic distribution. Moreover, a multivariate t with ν degrees of freedom has unbounded moments of order higher or equal to ν, a property that may not be desirable in some applications. For that reasons, the main purpose of this paper is to extend the results in FSC in the following dimensions.

First, we consider other elliptical distributions that also nest the normal, including the Kotz distribution, as well as some flexible families such as scale mixtures of normals and polynomial expansions of the multivariate normal density. In this sense, we provide numerically reliable analytical expressions for the score vector, which can be used to obtain numerically accurate extrema of the objective function, as well as reliable standard errors.

We also use our analytical expressions to develop score tests for multivariate normality when a dynamic model for the conditional mean and variance is fully specified, but the model is estimated under the Gaussianity null. The alternative tests for multivariate normality we obtain present power properties that differ substantially under different alternative hypotheses. Moreover, even though ellipticity is a maintained assumption for the aforementioned testing procedures, it should be noted that they can still be understood as kurtosis tests even if the distribution under the alternative hypothesis is asymmetric (see Mencía and Sentana (2010)).

The rest of the paper is organised as follows. In section 2, we discuss some relevant properties of elliptical distributions, and present closed-form expressions for the score vector and conditional information matrix. Then, in section 3 we introduce our proposed tests for multivariate normality. A Monte Carlo evaluation of the different testing procedures can be found in section 4. Finally, we present our conclusions in section 5. Proofs and auxiliary results are gathered in appendices.

2 Theoretical background

2.1 Elliptical distributions

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^*$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^* = e_t \mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is a non-negative random variable independent of $\mathbf{u}_t$, whose distribution determines the distribution of $\boldsymbol{\varepsilon}_t^*$. The variables e_t and $\mathbf{u}_t$ are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^*$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^*) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*) = \mathbf{I}_N$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis $\boldsymbol{\varepsilon}_t^*$ (see Mardia (1970)), κ, which is given by $E(e_t^4)/[N(N+2)] - 1$, will also be bounded.[1]

In what follows, we briefly describe the alternative distributions that we consider for the $N \times 1$ random vector $\boldsymbol{\varepsilon}_t^*$:

Gaussian: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ is distributed as a standardised multivariate normal if and only if ς_t is an independent chi-square random variable with N degrees of freedom.

Student *t*: $\boldsymbol{\varepsilon}_t^* = \sqrt{\nu - 2} \times \sqrt{\zeta_t/\xi_t}\mathbf{u}_t$ is distributed as a standardised multivariate Student t if and only if ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν and variance 2ν, with $\mathbf{u}_t$, ζ_t and ξ_t mutually independent.

Kotz: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ is distributed as a standardised Kotz if and only if ς_t is a gamma random

[1] In this respect, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^*$ is proportional to $\mathbf{u}_t$, then $\kappa \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

variable with mean N and variance $N[(N+2)\kappa+2]$.

Discrete scale mixture of normals: $\varepsilon_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ is distributed as a discrete mixture of normals (DSMN) if and only if

$$\varsigma_t = \frac{s_t + (1-s_t)\varkappa}{\alpha + (1-\alpha)\varkappa}\varsigma_t^o$$

where s_t is an independent Bernoulli variate with $P(s_t = 1) = \alpha$, $\varkappa$ is the variance ratio of the two components, which for identification purposes we restrict to be in the range $(0, 1]$ and ς_t^o is an independent chi-square random variable with N degrees of freedom.

Polynomial expansion: $\varepsilon_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ is distributed as a J^{th} order polynomial expansion if and only if ς_t has a density defined by

$$h(\varsigma_t) = \frac{1}{2^{N/2}\Gamma(N/2)}\varsigma_t^{N/2-1}\exp\left(-\frac{1}{2}\varsigma_t\right)P_J(\varsigma_t)$$

with

$$P_J(\varsigma_t) = \left[1 + \sum_{j=2}^{J} c_j p_{N/2-1,j}(\varsigma_t)\right],$$

where $p_{N/2-1,j}(.)$ denotes the generalized Laguerre polynomial of order j and parameter $N/2-1$.[2] For instance, the second and third order standardized Laguerre polynomials are

$$\begin{aligned}
p_{N/2-1,2}(\varsigma) &= \sqrt{\frac{2}{N(N+2)}}\left[\frac{N(N+2)}{4} - \left(\frac{N+2}{2}\right)\varsigma + \frac{1}{4}\varsigma^2\right], \text{ and} \\
p_{N/2-1,3}(\varsigma) &= \sqrt{\frac{12}{N(N+2)(N+4)}} \\
&\quad \times \left[\frac{N(N+2)(N+4)}{24} - \frac{(N+2)(N+4)}{8}\varsigma + \frac{N+4}{8}\varsigma^2 - \frac{1}{24}\varsigma^3\right].
\end{aligned}$$

The problem with polynomial expansions is that $P_J(\varsigma)$ will not be a proper density unless we restrict the coefficients c_j's so that it cannot become negative. For that reason, in Appendix B.1 we explain how to obtain restrictions on those coefficients to guarantee the positivity of $P_J(\varsigma)$ for all ς. Figure 1 describes the region in the (c_2, c_3) space for which densities of the 3^{rd} order expansion are well defined for all $\varsigma \geq 0$.

Importantly, all these examples nest the Gaussian distribution. In particular, the multivariate Student t approaches the multivariate normal as $\nu \to \infty$, but has generally fatter tails.

[2] The Rodrigues equation for the generalized Laguerre polynomials is

$$p_n^\alpha(x) = \frac{x^{-\alpha}e^{-x}}{n!}\frac{d^n}{dx^n}(e^{-x}x^{n+\alpha}).$$

These polynomials are orthogonal to each other with respect to the weighting function

$$f(x;\alpha) = \frac{x^\alpha e^{-x}}{\Gamma(1+\alpha)} \quad \text{for} \quad x > 0,\ \alpha > -1.$$

For that reason, we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions. Similarly, the Kotz distribution nests the multivariate normal distribution for $\kappa = 0$, but it can also be either platykurtic ($\kappa < 0$) or leptokurtic ($\kappa > 0$). Although such a nesting provides an analytically convenient generalisation of the multivariate normal, the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view. As for the DSMN, it approaches the multivariate normal when $\varkappa \to 1$, $\alpha \to 1$ or $\alpha \to 0$. As can be seen in Figure 2a-c, 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), while when $\alpha \to 1^-$ the opposite happens, very few observations with very small variance (inliers). More generally, ς_t will be a two-component scale mixture of $\chi_N^{2\prime}s$. As all scale mixtures of normals, the distribution of $\boldsymbol{\varepsilon}_t^*$ is leptokurtic. Finally, the polynomial expansion reduces to the spherical normal when $c_j = 0$ for all $j \in \{2, ..., J\}$. Interestingly, as can be seen in Figure 1, while the distribution of $\boldsymbol{\varepsilon}_t^*$ is leptokurtic for a 2^{nd} order expansion, it is possible to generate platykurtic random variables with a 3^{rd} order expansion.

Figure 3 plots the densities of a normal, a Student t, a platykurtic Kotz distribution, a discrete scale mixture of normals and a 3^{rd} order polynomial expansion in the bivariate case. Although they all have concentric circular contours because we have standardised and orthogonalised the two components, their densities can differ substantially in shape, and in particular, in the relative importance of the centre and the tails.

They also differ in the degree of cross-sectional "tail dependence" between the components, the normal being the only example in which lack of correlation is equivalent to stochastic independence. Allowing for dependence beyond correlation is particularly important in the context of multiple financial assets, in which the probability of the joint occurrence of several extreme events is regularly underestimated by the multivariate normal distribution (see Longin and Solnik, 2001). Figure 4 plots the so-called exceedance correlation between the uncorrelated marginal components in Figure 3. It can be noted the flexibility of the distributions we consider to generate different shapes of exceedance correlation.

For our purposes, it is also convenient to study the higher order moments of elliptical distributions. In this sense, it is easy to combine the representation of elliptical distributions above with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of a spherically symmetric distribution with $V(\boldsymbol{\varepsilon}_t^*) = \mathbf{I}_N$ are given by

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*\prime} \otimes \boldsymbol{\varepsilon}_t^*) &= \mathbf{0}, \\ E(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*\prime} \otimes \boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*\prime}) &= E[vec(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*\prime}) vec'(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^*)] = (\kappa+1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)]. \end{aligned}$$

An alternative characterization can be based on the higher order moment parameter of spherical random variables introduced by Berkane and Bentler (1986), $\tau_m(\boldsymbol{\eta})$, which Maruyama and Seo (2003) relate to higher order moments as

$$E[\varsigma_t^m|\boldsymbol{\eta}] = [1+\tau_m(\boldsymbol{\eta})]E[\varsigma_t^m|\mathbf{0}] \text{ where } E[\varsigma_t^m|\mathbf{0}] = 2^m\prod_{j=1}^{m}(N/2+j-1).$$

For the elliptical examples mentioned above, the $1+\tau_m(\boldsymbol{\eta})$'s –which are derived in Appendix B.2– become:

Student t:

$$1+\tau_m^t(\eta) = (1-2\eta)^{m-1}\prod_{j=2}^{m}\frac{1}{(1-2j\eta)} \quad \text{if} \quad m<\frac{1}{\eta}.$$

Kotz:

$$1+\tau_m^k(\kappa,N) = \left(\frac{(N+2)\kappa+2}{2}\right)^m\prod_{j=1}^{m}\frac{N/[(N+2)\kappa+2]+j-1}{N/2+j-1}.$$

Two-component scale mixture of normals:

$$1+\tau_m^{ds}(\alpha,\varkappa) = \frac{\alpha+(1-\alpha)\,\varkappa^m}{[\alpha+(1-\alpha)\varkappa]^m}.$$

3rd-order polynomial expansion:

$$1+\tau_m^{pe}(\alpha,\varkappa) = \frac{N(N+2)(N+4)+2c_2m(m-1)(N+4)-4c_3m[2+m(m-3)]\mathbf{1}\{m\geq 3\}}{N(N+2)(N+4)}.$$

Figure 5 show the different patterns in which these distributions departure from the Gaussian distribution in terms of $\tau_2(\boldsymbol{\eta})$ and $\tau_3(\boldsymbol{\eta})$. Notice also that with the exception of the Student t, a noteworthy property of the remaining distributions that we consider is that their moments are always bounded. In this respect, Appendix B.3 contains the moment generating functions for the Kotz, the two-component scale mixture of normals and the 3^{rd} order polynomial expansion.

2.2 The dynamic econometric model

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N dependent variables, $\mathbf{y}_t$, is typically assumed to be generated as:

$$\begin{gathered}\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)+\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*,\\ \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}),\\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}),\end{gathered}$$

where $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are $N\times 1$ and $N(N+1)/2\times 1$ vector functions known up to the $p\times 1$ vector of true parameter values $\boldsymbol{\theta}_0$, $\mathbf{z}_t$ are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and $\mathbf{z}_t$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t^*$ is a martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$. Hence,

$$\left.\begin{array}{l} E(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0) \\ V(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) \end{array}\right\}. \tag{1}$$

To complete the model, we need to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. We shall assume that, conditional on $\mathbf{z}_t$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as some particular member of the elliptical distributions described in the previous subsection, say $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim$ $i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ for short, where $\boldsymbol{\eta}$ are some q additional parameters that determine the shape of the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$.

2.3 The log-likelihood function

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $p+q$ parameters of interest, which we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T based on a particular parametric spherical assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

FSC provide expressions for $c(\eta)$ and $g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]$ in the multivariate Student t case, which are obviously such that $L_T(\boldsymbol{\theta}, 0)$ collapses to a conditionally Gaussian log-likelihood. As for the remaining distributions that we consider, the corresponding expressions are:

Kotz:

$$\begin{aligned} c(\kappa) &= \ln\Gamma\left(\frac{N}{2}\right) - \frac{N}{2}\ln\pi - \ln\Gamma\left(\frac{N}{b(\kappa)}\right) - \frac{N}{b(\kappa)}\ln b(\kappa) \\ g(\varsigma_t, \kappa) &= -\frac{1}{b(\kappa)}\varsigma_t + N\left[\frac{1}{b(\kappa)} - \frac{1}{2}\right]\ln\varsigma_t, \end{aligned}$$

where $b(\kappa) = 2 + \kappa(N+2)$.

Two-component scale mixtures:

$$\begin{aligned} c(\alpha, \varkappa) &= -\frac{N}{2}\ln\left(2\varpi\pi\right), \\ g[\varsigma, \alpha, \varkappa] &= \ln\left[\alpha\exp\left(-\frac{1}{2\varpi}\varsigma\right) + (1-\alpha)\varkappa^{-N/2}\exp\left(-\frac{1}{2\varpi\varkappa}\varsigma\right)\right], \end{aligned}$$

with $\varpi = [\alpha + \varkappa(1-\alpha)]^{-1}$.

3^{rd} order polynomial expansion:

$$\begin{aligned} c(\boldsymbol{\eta}) &= -\frac{N}{2}\ln\pi, \\ g[\varsigma_t, \boldsymbol{\eta}] &= -\frac{1}{2}\varsigma_t + \ln\left[1 + \sum_{j=1}^{J} c_j p_{N/2-1,j}(\zeta_t)\right], \end{aligned}$$

with $\boldsymbol{\eta} = (c_2, ...c_J)'$.

Given the nonlinear nature of the model, a numerical optimization procedure is usually required to obtain maximum likelihood (ML) estimates of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$ say. Assuming that all of the elements of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are twice continuously differentiable functions of $\boldsymbol{\theta}$, we can use a standard gradient method in which the first derivatives are numerically approximated by re-evaluating $L_T(\boldsymbol{\phi})$, with each parameter in turn shifted by a small amount, with an analogous procedure for the second derivatives. Unfortunately, such numerical derivatives are sometimes unstable, and, moreover, their values may be rather sensitive to the size of the finite increments used. As we will show in the next section, though, it is possible to obtain simple analytical expressions for the score vector.[3] The use of analytical derivatives in the estimation routine, as opposed to their numerical counterparts, should improve the accuracy of the resulting estimates considerably (McCullough and Vinod (1999)). Moreover, a fast and numerically reliable procedure for computing the score for any value of $\boldsymbol{\eta}$ is of paramount importance in the implementation of the score-based indirect inference procedures introduced by Gallant and Tauchen (1996). (See Calzolari, Fiorentini, and Sentana (2003) for an application to a discrete-time, stochastic volatility model.)

2.4 The score vector

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. Then, it is straightforward to show that if $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank, and $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ are differentiable

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varsigma}\frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})\right]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \quad (2)\\
\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \quad (3)
\end{aligned}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N)\\
\partial \varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, \quad (4)
\end{aligned}$$

[3] Nevertheless, it is important to stress that because both $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are often recursively defined (as in autoregressive moving average or generalized autoregressive conditional heteroscedasticity (GARCH) models), it may be necessary to choose some initial values to start up the recursions. As pointed out by Fiorentini, Calzolari, and Panattoni (1996), this fact should be taken into account in computing the analytic score, to make the results exactly comparable with those obtained by using numerical derivatives.

$$\begin{aligned}
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}),\\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})],\\
\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), &(5)\\
\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right\}, &(6)\\
\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] &= -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma, &(7)
\end{aligned}$$

and $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}'$ depend on the particular specification adopted while $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ and $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$ depend on the specific distribution assumed for estimation purposes.[4]

As for

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma,$$

which can be understood as the damping factor associated that reflects the kurtosis of the specific distribution assumed for estimation purposes, it reduces to:

Student t:

$$(N\eta+1)/[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})].$$

Kotz:

$$[N(N+2)\kappa\varsigma_t^{-1}(\boldsymbol{\theta})+2]/[(N+2)\kappa+2].$$

Two-component mixture:

$$[\pi+(1-\pi)\varkappa]\cdot\frac{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{[\pi+(1-\pi)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{[\pi+(1-\pi)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}. \tag{8}$$

Polynomial expansion

$$1-\frac{\sum_{j=1}^{J}c_j p_{N/2,j}[\varsigma_t(\boldsymbol{\theta})]}{1+\sum_{j=1}^{J}c_j p_{N/2-1,j}[\varsigma_t(\boldsymbol{\theta})]}.$$

Given that $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ is equal to 1 under Gaussianity, it is straightforward to check that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

$$\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})=\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right]=\left\{\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right]\end{array}\right\}.$$

2.5 PML and ML estimators under Gaussianity

If the interest of the researcher lied exclusively in $\boldsymbol{\theta}$, which are the parameters characterising the conditional mean and variance functions, then one attractive possibility would be to estimate

[4]Note that while both $\mathbf{Z}_t(\boldsymbol{\theta})$ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ depend on the specific choice of square root matrix $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ does not, a property that inherits from $l_t(\boldsymbol{\phi})$. The same result is not generally true for non-elliptical distributions (see Mencía and Sentana (2010)).

an equality restricted version of the model in which $\boldsymbol{\eta}$ is set to zero. Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote such a PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments; and, of course, it is efficient under Gaussian innovations.

Alternately, one could consider an s-based ML estimator of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$, which, in principle would be more efficient under correct specification of the distributional assumption s. The information matrix that characterizes the asymptotic variance of the s-based ML estimator of $\boldsymbol{\phi}$ under normality, which is a particular case of Proposition 1 in Fiorentini and Sentana (2010), is stated in the following result:

Proposition 1 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0$ is i.i.d. $N(\mathbf{0}, \mathbf{I}_N)$, then the information matrix of the feasible ML estimator that assumes $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ will be $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$, with $\mathcal{I}_t(\boldsymbol{\phi}) = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}_t'(\boldsymbol{\theta})$,*

$$\mathbf{Z}_t(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},$$

and

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & V[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,(\boldsymbol{\theta}_0', \mathbf{0}')'] \end{pmatrix}, \tag{9}$$

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n.

The block diagonality of $\mathcal{I}(\boldsymbol{\phi}_0)$ between the parameters of the conditional mean and variance functions, $\boldsymbol{\theta}$, and the shape parameters $\boldsymbol{\eta}$ under Gaussianity will repeatedly prove useful below.

3 Normality tests

3.1 Student t-based tests

As we mentioned before, FSC derived the Lagrange Multiplier (LM, or efficient score) test statistic to test multivariate normal versus Student t innovations on the basis of the value of the score of the log-likelihood function evaluated at the restricted parameter estimates $\tilde{\boldsymbol{\phi}}_T' = (\tilde{\boldsymbol{\theta}}_T', 0)'$. The following Proposition summarizes their result.

Proposition 2 *The Student t-based LM normality test*

$$LM_t(\tilde{\boldsymbol{\theta}}_T) = \frac{2}{N(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t s_\eta(\tilde{\boldsymbol{\theta}}_T, 0) \right\}^2$$

where

$$s_\eta(\boldsymbol{\theta}, 0) = \frac{N(N+2)}{4} - \frac{N+2}{2} \varsigma_t(\boldsymbol{\theta}) + \frac{1}{4} \varsigma_t^2(\boldsymbol{\theta})$$

is asymptotically distributed as a chi-square random variable with one degree of freedom under the null hypothesis of normality.

The $LM_t(\tilde{\boldsymbol{\theta}}_T)$ can be reinterpreted as a specification test of the restriction on the first two moments of $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$ implied by the Laguerre polynomial of second order. In addition, it numerically coincides with the kurtosis component of Mardia's (1970) test for multivariate normality in the models he considered.

Mencía and Sentana (2010) show that this score test will retain its optimal power against certain non-normal alternatives other than the Student t. Specifically, $LM_t(\tilde{\boldsymbol{\theta}}_T)$ is also the score tests against symmetric Generalized Hyperbolic (GH) distributions. As we shall see below, this is also the case for certain DSMN, as well as for distributions generated from a fourth order multivariate Hermite expansion of the normal distribution in which symmetry is assumed.

It is important to mention that the fact that $\eta = 0$ lies at the boundary of the parameter space in the case of the Student t invalidates the usual χ^2_1 distribution of the likelihood ratio or Wald tests, which under the null will be more concentrated toward the origin (see Andrews (2001) and references therein, as well as the simulation evidence in Bollerslev (1987)).[5] For that reason FSC recommend the Kuhn-Tucker multiplier test

$$KT_t(\tilde{\boldsymbol{\theta}}_T) = \mathbf{1}(\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0) > 0) \cdot LM_t(\tilde{\boldsymbol{\theta}}_T)$$

where $\mathbf{1}(.)$ is the indicator function. In this context, we would reject H_0 at the $100\tau\%$ significance level if the average score with respect to η evaluated at the Gaussian PML estimator is positive and $LM_t(\tilde{\boldsymbol{\theta}}_T)$ exceeds the $100(1-2\tau)$ percentile of a χ^2_1 distribution.

3.2 Kotz-based LM test

We can easily derive the LM test statistic to test multivariate normal versus Kotz innovations. To do so, it is necessary to find the value of $s_\kappa(\boldsymbol{\theta}, 0)$, i.e. the score with respect to κ when $\kappa = 0$ since we can base an LM test for normality on it. But given the previous expressions, it is not difficult to see that

$$s_\kappa(\boldsymbol{\theta}, 0) = \frac{N(N+2)}{4} \left\{ \left[\psi\left(\frac{N}{2}\right) + \ln 2 - \ln \varsigma_t(\boldsymbol{\theta}) \right] + \left(\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right) \right\}.$$

This expression has mean zero under normality since in that case $\varsigma_t \sim \chi^2_N \equiv$Gamma$(N/2, 2)$. Moreover, the block diagonality of the information matrix under normality stated in Proposition 1 allow us to base our LM test on this moment without need to make any adjustment for the

[5]The same happens with the class of symmetric Generalized Hyperbolic distributions since they can only accomodate fatter tails than the normal (see Mencía and Sentana (2010)).

fact that $\boldsymbol{\theta}$ will be replaced by the PML estimator $\tilde{\boldsymbol{\theta}}_T$, which is estimated on the basis of a moment condition that depends on $\varsigma_t/N - 1$.

Proposition 3 *The Kotz-based LM normality test*

$$LM_{kotz}(\tilde{\boldsymbol{\theta}}_T) = \left[\psi'\left(\frac{N}{2}\right) - \frac{2}{N}\right]^{-1}\left\{\frac{\sqrt{T}}{T}\sum_t s_\kappa(\tilde{\boldsymbol{\theta}}_T, 0)\right\}^2$$

is asymptotically distributed as a chi-square random variable with one degree of freedom under the null hypothesis of normality.

Interestingly, $s_\kappa(\boldsymbol{\theta}, 0)$ is not proportional to any Laguerre polynomial, unlike in the Student t case. To gain some intuition, we can project $s_{\kappa t}(\boldsymbol{\theta}, \kappa)$ onto Laguerre polynomials and look at the projection coefficients (see Appendix B.4.1), which for $j \geq 2$ are given by:

$$\begin{aligned} w_j^k(\kappa) &= E[s_{\kappa t}(\boldsymbol{\theta}, \kappa) \cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]/E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] \\ &= \frac{N+2}{j \cdot [(N+2)\kappa + 2]} \prod_{i=2}^{j} \left[\frac{[(N+2)\kappa + 2]i}{N + [(N+2)\kappa + 2](i-1)}\right]. \end{aligned}$$

Under the null of Gaussianity, the first three coefficients of $w_j^k(0)$ are

$$w_2^k(0) = \frac{1}{2},\ w_3^k(0) = \frac{2}{N+4}, \text{ and } w_4^k(0) = \frac{12}{(N+4)(N+6)},$$

which implies that the convergence in J of $\sum_{j=0}^{J}[w_j^k(0)]^2 E\{p^2_{N/2-1,j}[\varsigma_t(\boldsymbol{\theta})]\}$ to $E[s_\kappa^2(\boldsymbol{\theta}, 0)]$ is quite slow for N small. Therefore, the Kotz distribution behaves approximately as a polynomial expansion of a relatively high order when it approaches Gaussianity. This could be explained by the fact that the leading term in $s_\kappa(\boldsymbol{\theta}, 0)$ is $\ln \varsigma_t(\boldsymbol{\theta})$, which requires high number of terms in the Laguerre expansion in order to obtain a reasonable approximation.

3.3 DSMN-based LM tests

When the innovations are distributed as a discrete scale mixture of normals, we can achieve normality in three different ways: (i) when $\alpha \to 0^+$ or (ii) $\alpha \to 1^-$ regardless of the value of $\varkappa$ and (iii) when $\varkappa \to 1$ irrespective of α. Therefore, it is not surprising that the Gaussian scores with respect to α and $\varkappa$ are 0 when these parameters are not identified. Similarly, the limit of the score with respect to the mean and variance parameters,

$$\lim_{\alpha\cdot(1-\alpha)\cdot(1-\varkappa)\to 0} \mathbf{s}_{\boldsymbol{\theta} t}(\boldsymbol{\phi}),$$

coincides with the usual Gaussian expressions (see e.g. Bollerslev and Wooldridge (1992)).

Under $H_0 : \varkappa = 1$ we observe that $\lim_{\varkappa \to 1^-} s_{\varkappa t}(\boldsymbol{\phi}) = 0$ for $\alpha \in (0, 1)$, so we cannot use the first-order derivative to derive a normality test. As in Lee and Chesher (1986), though, we find that reparameterising the model appropriately solves the problem. Specifically, if we re-write the score in terms of υ, which is implicitly defined by $\varkappa = 1 - \sqrt{\upsilon}$, we obtain

$$\lim_{\upsilon \to 0^+} s_{\varpi t}(\boldsymbol{\theta},\alpha,\varpi) = \frac{1}{2}\alpha(1-\alpha)\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta})\right]$$

which again is proportional to the second order Laguerre polynomial. Thus, for any value of $\alpha \in (0,1)$ $LM_{\varkappa=1}(\tilde{\boldsymbol{\theta}}_T)$ also coincides with the Lagrange Multiplier test derived in FSC.[6]

Furthermore, we can show that for fixed $\varkappa \in (0,1)$, the relevant score in the case of "outliers" is given by

$$\lim_{\alpha \to 0^+} s_{\alpha t}(\boldsymbol{\theta},\alpha,\varkappa) = \varkappa^{N/2}\exp\left(\frac{1-\varkappa}{2}\varsigma_t(\boldsymbol{\theta})\right) - 1 - \frac{1-\varkappa}{2\varkappa}(\varsigma_t(\boldsymbol{\theta}) - N)$$

while in the case of "inliers" it will be given by

$$\lim_{\alpha \to 1^-} s_{\alpha t}(\boldsymbol{\theta},\alpha,\varkappa) = 1 - \varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right) - \frac{1-\varkappa}{2}(\varsigma_t(\boldsymbol{\theta}) - N).$$

We must study separately these possible ways to achieve normality. For instance, consider the conditional information matrix when $\alpha \to 0^+$, i.e.

$$\lim_{\alpha \to 0^+} V\left[\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\alpha,\varkappa) \\ s_{\alpha t}(\boldsymbol{\theta},\alpha,\varkappa) \end{array}\right]\middle| \mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right],$$

where we have excluded the term corresponding to $\varkappa$ because $s_{\varkappa t}(\boldsymbol{\phi})$ is identically zero in the limit. As expected, the conditional variance of the component of the score corresponding to the conditional mean and variance parameters $\boldsymbol{\theta}$ coincides with the expression obtained by Bollerslev and Wooldridge (1992). Importantly, we can show that the conditional information matrix of the DSMN distribution when $\alpha \to 0^+$ is finite only if $\varkappa > \frac{1}{2}$, in which case it is characterized by

$$V[s_{\alpha t}(\boldsymbol{\theta},0^+,\varkappa)] = \left(\frac{\varkappa^2}{2\varkappa-1}\right)^{N/2} - \frac{N}{2}\left(\varkappa^{-1}-1\right)^2 - 1.$$

Again, to gain some intuition on the reason why normality tests based on $s_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})$ are well defined only for a subset of the nuisance parameter $\varkappa$, we can project $s_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})$ onto Laguerre polynomials and look at its coefficients (see Appendix B.4.2):

$$\begin{aligned} w_j^{ds}(\varkappa,0) &= E[s_{\alpha t}(\boldsymbol{\theta},0^+,\varkappa)\cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]/E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] \\ &= \frac{1}{2}\left(\frac{1-\varkappa}{\varkappa}\right)^j \quad \text{for } j \geq 2. \end{aligned}$$

[6] Alternately, Neyman and Scott (1966) considered similar problems in the context of the $C(\alpha)$ statistic and recommend approximating the log-likelihood function using higher order derivatives. In our situation, the limit of the derivative of the score with respect to $\varkappa$,

$$\lim_{\varkappa \to 1}\frac{\partial s_{\varkappa t}(\boldsymbol{\theta},\alpha,\varkappa)}{\partial\varkappa} = \alpha(1-\alpha)\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta})\right],$$

coincides with $\lim_{\varpi \to 0^+} s_{\varpi t}(\boldsymbol{\theta},\alpha,\varpi)$.

As expected, the sequence $w_j^{ds}(\varkappa,0)$ diverges if $\varkappa \leq 1/2$.[7]

On the other hand, we find that

$$\begin{aligned} w_j^{ds}(\varkappa,1) &= E[s_{\alpha t}(\boldsymbol{\theta},1^-,\varkappa)\cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]/E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] \\ &= \frac{(-1)^{j-1}}{2}(1-\varkappa)^j \quad \text{for} \;\; j \geq 2, \end{aligned}$$

is convergent for any value of $\varkappa \in (0,1]$. The reason for the difference in the projections comes from the fact that when a discrete scale mixture of normals achieve normality when $\alpha \to 0^+$ there are very few observations with very large variance (outliers), while when $\alpha \to 1^-$ the opposite happens, very few observations with very small variance may (inliers). Because in the first case $V[s_{\alpha t}(\boldsymbol{\theta},0^+,\varkappa)]$ is only well defined for $\varkappa > \frac{1}{2}$, in what follows we focus on the case $\alpha \to 1^-$.

As is well known, the relative scale parameter $\varkappa$ becomes underidentified in the limit. One standard solution in the literature to deal with testing situations with underidentified parameters under the null involves fixing the remaining parameter to some arbitrary value, and then computing the appropriate test statistic for that given value.

Let $\tilde{\boldsymbol{\theta}}_T$ denote the ML estimator of obtained by maximising the Gaussian log-likelihood function. For the case in which normality is achieved as $\alpha \to 1^-$, we can use the results above to show that for a given value of $\varkappa \in (0,1)$, the LM test will be the usual quadratic form in the sample averages of the scores corresponding to $\boldsymbol{\theta}$ and α, $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T,1^-,\varkappa)$ and $\bar{s}_{\alpha T}(\tilde{\boldsymbol{\theta}}_T,1^-,\varkappa)$, with weighting matrix the inverse of the unconditional information matrix, which can be obtained as the unconditional expected value of the conditional information matrix. But since $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T,1^-,\varkappa) = \mathbf{0}$ by definition of $\tilde{\boldsymbol{\theta}}_T$, we can show the following result:

Proposition 4 *The DSMN$_1$-based LM normality test*

$$LM_{DSMN_1}(\tilde{\boldsymbol{\theta}}_T,\varkappa) = \left[\left(\frac{1}{2\varkappa-\varkappa^2}\right)^{N/2} - \frac{N}{2}(1-\varkappa)^2 - 1\right]^{-1}\left\{\frac{\sqrt{T}}{T}\sum_t s_\alpha(\tilde{\boldsymbol{\theta}}_T,1^-,\varkappa)\right\}^2$$

where

$$s_\alpha(\boldsymbol{\theta},1^-,\varkappa) = 1 - \varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right) - \frac{1-\varkappa}{2}(\varsigma_t(\boldsymbol{\theta})-N)$$

is asymptotically distributed as a chi-square random variable with one degree of freedom under the null hypothesis of normality.

[7] More importantly, under sequences of local alternatives the aforementioned constraint becomes tigther. For a fixed value of the scale parameter $\bar{\varkappa}$ (i) the expectation of $s_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})$ is finite only if $\bar{\varkappa} > 1-\varkappa_0$, where $\varkappa_0$ denotes the true value of the scale parameter $\varkappa$. Moreover, its second moment $E[s^2_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})|\,\varkappa_0]$ is finite only if $\bar{\varkappa} > 1-\varkappa_0/2$. It can also be shown that under Kotz innovations, $E[s_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})|\,\kappa_0]$ are $E[s^2_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})|\,\kappa_0]$ are finite if $\bar{\varkappa} > \kappa_0(N+2)/(2+2\kappa_0+2N)$ and $\bar{\varkappa} > (1+2\kappa_0+2N)/(2+2\kappa_0+2N)$, respectively. Finally, under Student t innovations $E[s_{\alpha t}(\boldsymbol{\theta},0^+,\bar{\varkappa})|\,\nu_0]$ diverges.

The approach described above is plausible in situations where there are values of the underidentified parameter $\varkappa$ that make sense from an economic or statistical point of view. In other situations, we could follow a second approach, which consists in computing either the LM test statistic for the whole range of values of the underidentified parameter, which are then combined to construct an overall test statistic (see Andrews, 1994).

3.4 Higher order Laguerre polynomials-based tests

Kiefer and Salmon (1983) proposed a simple test normality based on an Edgeworth expansion for the null hypothesis that the errors in an univariate econometric model follow a normal distribution (see also Bontemps and Meddahi (2005)). In the same spirit, under the alternative of a non Gaussian elliptical distribution, we can expand the density of ς_t in terms of Laguerre polynomials around a Gamma variate with parameters $N/2$ and 2 (see Bontemps and Meddahi (2010) for related ideas in a univariate context). Orthogonality of Laguerre polynomial-based moment conditions under the null hypothesis of normality with respect to the mean and variance parameters follows directly. Moreover, computing the limiting distribution of moment conditions involving $p_{N/2-1,j}(\varsigma_t)$ is straightforward.

Proposition 5 *Let $p_{N/2-1,j}(\varsigma_t)$ denote the j^{th} order standardised Laguerre polynomial with parameter $N/2-1$. Then*

$$L_j(\tilde{\boldsymbol{\theta}}_T) = \left\{ \frac{\sqrt{T}}{T} \sum_t p_{N/2-1,j}(\varsigma_t(\tilde{\boldsymbol{\theta}}_T)) \right\}^2$$

is asymptotically distributed as a chi-square random variable with one degree of freedom under the null hypothesis of normality.

The asymptotic independence of the successive chi-square variates follows from the orthogonality of the Laguerre polynomials under the normal measure. In consequence it can be seen that multiple order tests against the null of normality could be based just on the sum of the individual tests, say $L_{\Sigma J}(\tilde{\boldsymbol{\theta}}_T) = \sum_{j=2}^{J} L_j(\tilde{\boldsymbol{\theta}}_T)$ which is asymptotically distributed as a chi-square random variable with $J-1$ degrees of freedom under the null hypothesis of normality.

Similarly to the Student t case, Laguerre polynomial-based tests can be understood as normality tests against polynomial expansion alternatives. Hence, again more powerful versions of the tests can be obtained by imposing the positivity constraints of the polynomial expansion i.e. $sign(c_J) = (-1)^J$. A 2^{nd} order expansion yields an analogous result to $KT_t(\tilde{\boldsymbol{\theta}}_T)$. As for the 3^{rd} order expansion the relevant restriction becomes c_3 being non-positive so that we suggest using

$$KT^{pe}_{\Sigma 3}(\tilde{\boldsymbol{\theta}}_T) = \mathbf{1}(p_{N/2-1,3}(\varsigma_t(\tilde{\boldsymbol{\theta}}_T)) < 0) \cdot L_{\Sigma 3}(\tilde{\boldsymbol{\theta}}_T) + \mathbf{1}(p_{N/2-1,3}(\varsigma_t(\tilde{\boldsymbol{\theta}}_T)) \geq 0) \cdot L_2(\tilde{\boldsymbol{\theta}}_T)$$

which will be distributed as a $50:50$ mixture of chi-squared distributions with 1 and 2 degrees of freedom under the null of Gaussianity. In this respect, it is important to mention that when there is a single restriction, such as in our case, those one-sided tests would be asymptotically locally more powerful (Andrews 2001).

3.5 Reinterpretation of the multivariate normality tests as univariate tests

As we saw before, the assumption of ellipticity implies that the only difference between the Gaussian distribution and the other distributions that we are considering lies in the distribution of ς_t. Therefore, a natural way of testing for multivariate normality would be to test the null that distribution of ς_t is a χ^2_N against the different alternatives. It turns out that this is precisely what our score tests are doing. More formally,

Proposition 6 *Let*

$$h_\varsigma(\varsigma_t;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp\{c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})\}$$

denote the marginal density of ς_t. Then

$$\frac{\partial \ln f_{\mathbf{y}}(\mathbf{y}_t)}{\partial \boldsymbol{\eta}} = \frac{\partial \ln h_\varsigma(\varsigma_t)}{\partial \boldsymbol{\eta}}.$$

To prove this result, it is convenient to use the fact that the density of $\mathbf{u}_t$ is

$$f_{\mathbf{u}}(\mathbf{u}_t) = \frac{\Gamma(N/2)}{2\pi^{N/2}}\mathbf{1}\{\mathbf{u}_t'\mathbf{u}_t = 1\}.$$

Hence, we can write[8]

$$f_{\mathbf{y}|\varsigma}(\mathbf{y}_t|\varsigma_t;\boldsymbol{\theta},\boldsymbol{\eta}) = \frac{2}{|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{1/2}\varsigma_t^{N/2-1}}\frac{\Gamma(N/2)}{2\pi^{N/2}}\mathbf{1}\left\{[\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})[\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})]=\varsigma_t\right\}$$

On this basis, we can formally write the density of $\mathbf{y}_t$ as follows

$$f_{\mathbf{y}}(\mathbf{y}_t;\boldsymbol{\theta},\boldsymbol{\eta}) = \int f_{\mathbf{y}|\varsigma}(\mathbf{y}_t|\varsigma_t)h_\varsigma(\varsigma_t)d\varsigma_t = \int f_{\mathbf{y}|\sqrt{\varsigma}}(\mathbf{y}_t|\sqrt{\varsigma_t})h_{\sqrt{\varsigma}}(\sqrt{\varsigma_t})d\sqrt{\varsigma_t}$$

whence we obtain the required result by noticing that $f_{\mathbf{y}|\varsigma}(\mathbf{y}_t|\varsigma_t;\boldsymbol{\theta},\boldsymbol{\eta})$ has nonzero density only for the $\mathbf{y}_t^o$'s that satisfy $[\mathbf{y}_t^o-\boldsymbol{\mu}_t(\boldsymbol{\theta})]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})[\mathbf{y}_t^o-\boldsymbol{\mu}_t(\boldsymbol{\theta})] = \varsigma_t^o$, and therefore those $\mathbf{y}_t^o$'s ca be summarized through ς_t^o.

Finally, note that we can use the same expression to write

$$\frac{\partial \ln f_{\mathbf{y}}(\mathbf{y}_t)}{\partial \boldsymbol{\theta}} = \frac{\partial \ln f_{\mathbf{y}|\varsigma}(\mathbf{y}_t|\varsigma_t)}{\partial \boldsymbol{\theta}} + \frac{\partial \ln f_{\mathbf{y}|\varsigma}(\mathbf{y}_t|\varsigma_t)}{\partial \varsigma_t}\frac{\partial \varsigma_t}{\partial \boldsymbol{\theta}} + \frac{\partial \ln h_\varsigma(\varsigma_t)}{\partial \boldsymbol{\theta}}.$$

[8] Equivalently,

$$f_{\mathbf{y}|\sqrt{\varsigma}}(\mathbf{y}_t|\sqrt{\varsigma_t}) = \frac{1}{|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{1/2}(\sqrt{\varsigma_t})^{N-1}}\frac{\Gamma(N/2)}{2\pi^{N/2}}\mathbf{1}\left\{[\mathbf{y_t}-\boldsymbol{\mu}_\mathbf{t}(\boldsymbol{\theta})]\boldsymbol{\Sigma}_\mathbf{t}^{-\mathbf{1}}(\boldsymbol{\theta})[\mathbf{y_t}-\boldsymbol{\mu}_\mathbf{t}(\boldsymbol{\theta})]=(\sqrt{\varsigma_t})^2\right\};$$

and

$$h_{\sqrt{\varsigma}}(\sqrt{\varsigma_t}) = \frac{2\pi^{N/2}}{\Gamma(N/2)}(\sqrt{\varsigma_t})^{N-1}\exp\left\{c(\boldsymbol{\eta})+g\left[(\sqrt{\varsigma_t})^2,\boldsymbol{\eta}\right]\right\}.$$

3.6 Power of the normality tests

Although we shall investigate the finite sample properties of the different multivariate normality tests in section 6, it is interesting to study their asymptotic power properties. But given that the block-diagonality of the information matrix is generally lost under the alternative of $\boldsymbol{\eta} \neq \mathbf{0}$, and its exact form is unknown, we can get only closed-form expressions for the case in which the standardized innovations, $\boldsymbol{\varepsilon}_t^*$, are directly observed. In more realistic cases, though, the results are likely to be qualitatively similar.

In this section we exploit asymptotic expressions for the non-centrality parameters of:

- LM normality test against Student t
- LM normality test against Kotz
- LM normality test against DSMN
- Higher order Laguerre polynomial normality tests,

under each of the following alternative hypotheses:

- $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim t(\mathbf{0}, \mathbf{I}_N, \nu_0)$
- $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim Kotz(\mathbf{0}, \mathbf{I}_N, \kappa_0)$
- $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim DSMN(\mathbf{0}, \mathbf{I}_N, \alpha_0, \varkappa_0)$, and
- $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim p.e.(\mathbf{0}, \mathbf{I}_N, c_{20}, c_{30})$.

The relevant asymptotic expressions are reported in Appendix C. Specifically, Proposition C.1 extends Proposition 3 in FSC to the case in which the true distribution is either Kotz, discrete scale mixture of normals or polynomial expansion. The asymptotic distribution of the Kotz-based LM test under the same hypotheses of Proposition 3 is provided in Proposition C.2; while Proposition C.3 and Proposition C.4 do the same for the DSMN$_1$-based LM test and the 3^{rd} order standardized Laguerre polynomial-based LM test, respectively.

On the basis of these Propositions, we can obtain the asymptotic power of the several normality tests described above for any possible significance level k under the alternative distributions.

The results at the usual 5% level for a sample size $T = 1,000$ are plotted in Figure 6. Panel 6a reports departures from normality towards a Student t, for η_0 in the range $0 \leq \eta_0 \leq 0.015$. Not surprisingly, the power of the several tests uniformly increases as we depart from the null for a given sample size. As expected, the one-sided $KT_t(\tilde{\boldsymbol{\theta}}_T)$ test is the most powerful. The increase of degrees of freedom when higher order Laguerre expansions are considered seem to generate some decrease in power.

Departures from normality towards a Kotz distribution are plotted in Panel 6b. The Kotz LM-based test dominates the tests that consider not only the 2^{nd} but also the 3^{rd} Laguerre polynomial as well as the DSMN-based LM test. Interestingly, $L_{\Sigma 3}(\tilde{\boldsymbol{\theta}}_T)$ outperforms the 2^{nd} Laguerre polynomial-based test, which is in line with the observation that the Kotz distribution behaves approximately as a polynomial expansion of a relatively high order when it approaches Gaussianity.

As for the discrete mixture of normals alternative, from Panel 6c is clear that the ranking becomes:

$$LM_{DSMN_1}(\tilde{\boldsymbol{\theta}}_T, \varkappa) \succ LM_{kotz}(\tilde{\boldsymbol{\theta}}_T) \succ L_{\Sigma 3}(\tilde{\boldsymbol{\theta}}_T) \succ LM_t(\tilde{\boldsymbol{\theta}}_T) \succ L_2(\tilde{\boldsymbol{\theta}}_T).$$

Finally, Panel 6d reports departures from normality towards a 3^{rd} oder polynomial expansion distribution with $c_2 = 0$. In this case, $L_{\Sigma 3}(\tilde{\boldsymbol{\theta}}_T)$ and $LM_{DSMN_1}(\tilde{\boldsymbol{\theta}}_T, \varkappa)$ are the most powerful. Not surprisingly, both the 2^{nd} order Laguerre polynomial-based test and the $KT_t(\tilde{\boldsymbol{\theta}}_T)$ have no power since $c_2 = 0$ implies the kurtosis of the normal for the polynomial expansion distribution.

To investigate how the several distributions approach the normal distribution along local alternatives we could compare the higher order moment parameters of spherical random variables introduced in section 2.1, $\tau_s(\boldsymbol{\eta})$, when they achieve normality. To do so, we compare the limiting behavior of $E[\varsigma_t^m|\boldsymbol{\eta}]$ for $m > 2$ when $\tau_m(\boldsymbol{\eta}) \to 0$ but with the kurtosis, $\tau_2(\boldsymbol{\eta})$, being the same for any pair of distributions. Setting the Student t as benchmark, we find that for the Kotz distribution,[9]

$$\lim_{\kappa \to 0} \frac{\tau_{m+1}^k(\kappa)}{\tau_m^k(\kappa)} < \frac{m+1}{m-1} = \lim_{\eta \to 0^+} \frac{\tau_{m+1}^t(\eta^\circ(\kappa))}{\tau_m^t(\eta^\circ(\kappa))},$$

which implies that $\tau_m^k(\kappa) < \tau_m^k(\eta^\circ(\kappa))$, where $\eta^\circ(\kappa)$ is found by solving $\tau_2^t(\eta^\circ(\kappa)) = \tau_2^k(\kappa)$.

As for the discrete scale mixture of normals, for fixed $\varkappa \in (0, 1)$ we have

$$\lim_{\alpha \to 0^+} \frac{\tau_m^{ds}(\alpha, \varkappa)}{\tau_m^t[\eta^\circ(\alpha, \varkappa)]} = \frac{2}{\varkappa^{(m-2)} m(m-1)} \sum_{j=1}^{m-1} j \varkappa^{j-1}$$

and

$$\lim_{\alpha \to 1^-} \frac{\tau_m^{ds}(\alpha, \varkappa)}{\tau_m^t[\eta^\circ(\alpha, \varkappa)]} = \frac{2}{m(m-1)} \sum_{j=1}^{m-1} (m-j) \varkappa^{j-1},$$

which implies that when $\varkappa \to 1^-$, then $\tau_m^{ds}(\alpha, \varkappa) = \tau_m^t[\eta^\circ(\alpha, \varkappa)]$ for all m and that when $\alpha \to 0^+$ ($\alpha \to 1^-$), then $\tau_m^{ds}(\alpha, \varkappa) > \tau_m^t[\eta^\circ(\alpha, \varkappa)]$ ($\tau_m^{ds}(\alpha, \varkappa) < \tau_m^t[\eta^\circ(\alpha, \varkappa)]$) for $m > 2$. Then

[9] In fact, one can show that

$$\lim_{N \to \infty} \left\{ \lim_{\kappa \to 0} \frac{\tau_{m+1}^k(\kappa, N)}{\tau_m^k(\kappa, N)} \right\} = \frac{m+1}{m-1},$$

which is consistent with the fact that the coefficients of the projection of $s_\kappa(\boldsymbol{\theta}, \kappa)$ onto Laguerre polynomials converge to zero as $N \to \infty$, except for the second polynomial.

it is not surprising that $LM_{DSMN_1}(\tilde{\boldsymbol{\theta}}_T, \varkappa)$ does a good job when the true distribution is Kotz and similarly, $LM_{kotz}(\tilde{\boldsymbol{\theta}}_T)$ outperforms the $LM_t(\tilde{\boldsymbol{\theta}}_T)$ when the true distribution is a discrete mixture of normals, since the higher order moments of those distributions grow at a lower than the Student t rate when they approach normality.

4 Monte Carlo Evidence

In this section, we assess the finite sample performance of the different estimators and testing procedures discussed above by means of several extensive Monte Carlo exercises. Since our focus is on distributional assumptions, in these excercises we treat $\boldsymbol{\theta}$ as known and focus on the performance of the alternative tests as if the $\boldsymbol{\varepsilon}_t^*$'s were observed.

We sample Gaussian, Student t, Kotz innovations, and discrete scale mixture of normals exploiting the decomposition presented in section 2.1. Specifically, we simulate standardised versions of all these distributions by appropriately mixing a N-dimensional spherical normal vector with a univariate gamma random variable, and, in the case of discrete scale mixture of normals, a draw from a scalar uniform, which we obtain from the IMSL library routines DRNNOR, DRNGAM and DRNUN, respectively. To draw polynomial expansion innovations we again use the same decomposition first and then the probability integral transform in order to obtain a simulated ς since we can easily obtain its cdf, which is given by

$$\begin{aligned} F(\varsigma, c_2, c_3, N) &= 1 - \frac{\Gamma(N/2, \varsigma/2)}{\Gamma(N/2)} - c_2 \times \frac{\varsigma^{N/2} e^{-\varsigma/2}}{2^{N/2+2}\Gamma(N/2+2)}(\varsigma - 2 - N) \\ &\quad + c_3 \times \frac{\varsigma^{N/2} e^{-\varsigma/2}}{2^{N/2+3}\Gamma(N/2+3)}\left[\varsigma^2 - 2d(N+4) + (N+2)(N+4)\right]. \end{aligned}$$

To do so, we sample a gamma random variable with parameters $N/2$ and 2, d say, that we used as starting value to find the solution to $F(\varsigma, c_2, c_3, N) = F_{\chi^2_N}(d)$.

Figures 7 summarises our findings for the different multivariate normality tests by means of Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. It can be noted that the higher order Laguerre polynomials-based tests, and in particular, their Kuhn-Tucker versions, seem to be too conservative in general, especially for large nominal sizes. As for the remaining tests, the actual finite sample sizes seem to be fairly close to their nominal levels.

Table 1 reports actual Power at the usual 5% and 1% level under the following alternative hypotheses:

1. Student-t with 100 degrees of freedom

2. Kotz with the same kurtosis

3. DSMN with the same kurtosis and $\alpha = 0.5$ and same kurtosis

4. Polynomial expansion with zero excess kurtosis and $c_3 = -0.1$.

The results indicate that the LM and KT tests based on the true distribution are the most powerful, with the exception of the LM based on the score of discrete scale mixture of normals. It is worth mentioning that, consistent with the projection coefficients of the score of the true distribution onto Laguerre polynomials, 2^{nd} Laguerre dominates the higher order ones when the distribution is Student t but not when is either Kotz or discrete mixture of normals.

5 Conclusions

In the context of a general multivariate dynamic regression model with time-varying variances and covariances considered in Bollerslev and Wooldridge (1992), we provide numerically reliable analytical expressions for the score vector when the distribution of the innovations is assumed to be elliptical, paying special attention not only to the Student t and Kotz distributions, but also to flexible families such as discrete scale mixtures of normals and polynomial expansions.

We develop score tests for multivariate normality when a dynamic model for the conditional mean and variance is fully specified, but the model is estimated under the Gaussianity null. The limiting null distribution of our proposed tests is correct regardless of the model used and the additional computational cost is negligible We find that the alternative tests for multivariate normality we obtain present power properties that differ substantially under different alternative hypotheses. Finally, our Monte Carlo study suggest that the actual finite sample sizes seem to be fairly close to their nominal levels. For all those reasons, we recommend their use.

Appendix

A Proofs

Lemmata

Lemma 1 *Let ς be distributed as Gamma random variable with parameters k and λ and let $b < 1/\lambda$, then*

$$E\left[\varsigma^a \exp\left(b\varsigma\right)\right] = \frac{\Gamma(k+a)}{\Gamma(k)}\left(\frac{1}{1-\lambda b}\right)^{k+a}\lambda^a.$$

Proof. Write

$$\begin{aligned}
E\left[\varsigma^a \exp\left(b\varsigma\right)\right] &= \int_0^\infty \varsigma^a \exp\left(b\varsigma\right)\frac{1}{\Gamma(k)\lambda^k}\varsigma^{k-1}\exp\left(-\varsigma/\lambda\right)d\varsigma \\
&= \frac{\Gamma(k+a)}{\Gamma(k)\lambda^k}\left(\frac{\lambda}{1-\lambda b}\right)^{k+a} \\
&\quad \times \int_0^\infty \frac{1}{\Gamma(k+a)}\left(\frac{1-\lambda b}{\lambda}\right)^{k+a}\varsigma^{k-1+a}\exp\left(-\frac{\varsigma}{\lambda/(1-\lambda b)}\right)d\varsigma.
\end{aligned}$$

The result follows from noting that the integrand in the above equation is the probability density function of a Gamma variate with parameters $k+a$ and $\lambda/(1-\lambda b)$.

Lemma 2 *Let ς be distributed as a Gamma random variate with parameters k and λ, then*

$$E\left[\ln\varsigma\right] = \psi\left(k\right) + \ln\lambda,$$

$$E\left[\ln^2\varsigma\right] = \psi'\left(k\right) + \psi^2\left(k\right) + 2\psi\left(k\right)\ln\lambda + \ln^2\lambda,$$

$$E\left[\varsigma^j\ln\varsigma\right] = \left[\psi\left(k+j\right) + \ln\lambda\right]\cdot\prod_{i=1}^{j}\left[k\lambda + \lambda\left(i-1\right)\right];$$

and, provided $k > 1$,

$$E\left[\frac{1}{\varsigma}\ln\varsigma\right] = \frac{1}{\lambda}\left[\frac{\Gamma(k-1)}{\Gamma(k)}\psi\left(k-1\right) + \frac{1}{k-1}\ln\lambda\right],$$

where $\psi\left(x\right) = \partial\ln\Gamma(x)/\partial x$ is the so-called Digamma function, or the Gauss' psi function (see Abramowitz and Stegun 1964.)

Proof. If $\varsigma \sim$Gamma(k,λ), then $t = \frac{1}{\lambda}\varsigma \sim$Gamma$(k,1)$. Next, from

$$\frac{d^n\Gamma(x)}{dx^n} = \int_0^\infty \ln^n t\cdot t^{k-1}e^{-t}dt,$$

and the definition of the polygamma function, we know that

$$E[\ln t] = \frac{1}{\Gamma(k)}\int_0^\infty \ln t\cdot t^{k-1}e^{-t}dt = \psi\left(k\right),$$

$$E[\ln^2 t] = \frac{1}{\Gamma(k)} \int_0^\infty \ln^2 t \cdot t^{k-1} e^{-t} dt = \psi'(k) + \psi^2(k),$$

and, for $k > -j$,

$$\begin{aligned} E[t^j \ln t] &= \frac{\Gamma(k+j)}{\Gamma(k)} \frac{1}{\Gamma(k+j)} \int_0^\infty \ln t \cdot t^{k+j-1} e^{-t} dt \\ &= \frac{\Gamma(k+j)}{\Gamma(k)} \psi(k+j). \end{aligned}$$

Hence, because $\varsigma = \lambda t$ we have

$$E[\ln \varsigma] = E[\ln t] + \ln \lambda,$$

$$E\left[\ln^2 \varsigma\right] = E\left[\ln^2 t\right] + 2 \ln \lambda E[\ln t] + \ln^2 \lambda,$$

and, since $E(t^j) = \Gamma(k+j)/\Gamma(k)$,

$$E\left[\varsigma^j \ln \varsigma\right] = \lambda^j \left\{ E\left[t^j \ln t\right] + \frac{\Gamma(k+j)}{\Gamma(k)} \ln \lambda \right\},$$

from where the first three results follow directly. Finally, as

$$E\left[\frac{1}{\varsigma} \ln \varsigma\right] = \frac{1}{\lambda} E\left[\frac{1}{t} \ln t\right] + \frac{\ln \lambda}{\lambda} E\left[\frac{1}{t}\right],$$

the last result follows from $E[t^{-1}] = (k-1)^{-1}$ (see Cressie, N.A.C., A.S. Davis, J.L. Folks and G.E. Policello (1981)).

Lemma 3 *Let ς be distributed as $N(\nu-2)/\nu$ times an F variate with N and ν degrees of freedom and let $b < 0$; then*

$$E[\exp(b\varsigma)] = \frac{\Gamma[(N+\nu)/2]}{\Gamma(\nu/2)} U\left[\frac{N}{2}, 1 - \frac{\nu}{2}, -b(\nu-2)\right]$$

$$E[\varsigma \exp(b\varsigma)] = N \frac{\nu-2}{2} \frac{\Gamma[(N+\nu)/2]}{\Gamma(\nu/2)} U\left[\frac{N}{2} + 1, 2 - \frac{\nu}{2}, -b(\nu-2)\right]$$

$$E\left[\varsigma^2 \exp(b\varsigma)\right] = N(N+2) \frac{(\nu-2)^2}{2} \frac{\Gamma[(N+\nu)/2]}{\Gamma(\nu/2)} U\left[\frac{N}{2} + 2, 3 - \frac{\nu}{2}, -b(\nu-2)\right]$$

$$E\left[\varsigma^j \exp(b\varsigma)\right] = \frac{\Gamma[(N+\nu)/2]}{\Gamma(\nu/2)} \frac{\Gamma(N/2+j)}{\Gamma(N/2)} (\nu-2)^j U\left[\frac{N}{2} + j, j + 1 - \frac{\nu}{2}, -b(\nu-2)\right]$$

where U denotes Confluent Hypergeometric Function of the Second Kind (see Abramowitz and Stegun (1964)).

Proof. Let $\varsigma = \frac{N(\nu-2)}{\nu} x$ where x is $F_{N,\nu}$ variate so that its probability density function can be written as

$$f_X(x) = \frac{\Gamma[(N+\nu)/2]}{\Gamma(N/2)\Gamma(\nu/2)} \frac{1}{x} \left(\frac{x}{x+\nu/N}\right)^{N/2} \left(\frac{\nu/N}{x+\nu/N}\right)^{\nu/2}.$$

Hence,

$$\begin{aligned}
E\left[\varsigma^j \exp(b\varsigma)\right] &= E\left\{\left[\frac{N(\nu-2)}{\nu}x\right]^j \exp\left[\frac{N(\nu-2)}{\nu}x\right]\right\} \\
&= \frac{\Gamma[(N+\nu)/2]}{\Gamma(N/2)\Gamma(\nu/2)}\left[\frac{(\nu-2)N}{\nu}\right]^j \\
&\quad \times \int_0^\infty x^j \exp\left[b\frac{(\nu-2)N}{\nu}x\right]\frac{1}{x}\left(\frac{x}{x+\nu/N}\right)^{N/2}\left(\frac{\nu/N}{x+\nu/N}\right)^{\nu/2} dx \\
&= \frac{\Gamma[(N+\nu)/2]}{\Gamma(N/2)\Gamma(\nu/2)}\left[\frac{(\nu-2)N}{\nu}\right]^j \left(\frac{\nu}{N}\right)^{j-1} \\
&\quad \times \int_0^\infty \exp\left[b\frac{(\nu-2)N}{\nu}x\right]\left(\frac{x}{\nu/N}\right)^{N/2+j-1}\left(\frac{x}{\nu/N}+1\right)^{-(N+\nu)/2-j+1} dx.
\end{aligned}$$

Then, making the change of variable $t = xN/\nu$ (with $dx = \nu/N \cdot dt$),

$$\begin{aligned}
E\left[\varsigma^j \exp(b\varsigma)\right] &= \frac{\Gamma[(N+\nu)/2]}{\Gamma(\nu/2)}\frac{\Gamma(N/2+j)}{\Gamma(N/2)}(\nu-2)^j \\
&\quad \times \int_0^\infty \frac{\exp[b(\nu-2)t]}{\Gamma(N/2+j)} t^{N/2+j-1}(t+1)^{-(N+\nu)/2-j+1} dt
\end{aligned}$$

we can recognize that

$$U\left[\frac{N}{2}+j; 1+j-\frac{\nu}{2}; -b(\nu-2)\right] = \int_0^\infty \frac{1}{\Gamma(N/2+j)}\exp[b(\nu-2)t]\, t^{N/2+j-1}(t+1)^{-(N+\nu)/2-j+1} dt$$

is the Confluent Hypergeometric Function of the Second Kind; and the result follows directly.

Proposition 1

See Fiorentini and Sentana (2010).

Proposition 2

See FSC.

Proposition 3

We can easily compute the hessian

$$h_{\kappa\kappa}(\kappa) = \frac{\partial^2 c(\kappa)}{\partial\kappa^2} + \frac{\partial^2 g[\varsigma_t;\kappa]}{\partial\kappa^2}$$

where

$$\frac{\partial^2 c(\kappa)}{\partial\kappa^2} = \frac{N(N+2)^2}{b^3(\kappa)}\left\{\left[-\psi'\left(\frac{N}{b(\kappa)}\right)\frac{N}{b(\kappa)}+1\right] - 2\left[\psi\left(\frac{N}{b(\kappa)}\right)+\ln b(\kappa)-1\right]\right\}$$

and

$$\frac{\partial^2 g[\varsigma_t;\kappa]}{\partial\kappa^2} = \frac{2(N+2)^2}{b^3(\kappa)}\left[N\ln\varsigma_t - \varsigma_t\right].$$

Then, as

$$E\left[\left.\frac{\partial^2 g[\varsigma_t;\kappa]}{\partial\kappa^2}\right|\kappa=0\right]=\frac{2N(N+2)^2}{b^3(\kappa)}\left[\psi\left(\frac{N}{b(\kappa)}\right)+\ln b(\kappa)-1\right]$$

we have that

$$\begin{aligned}E\left[-h_{\kappa\kappa}(\kappa)\right] &= \frac{N(N+2)}{b^2(\kappa)}\left[\psi'\left(\frac{N}{b(\kappa)}\right)\frac{Nb'(\kappa)}{b^2(\kappa)}-\frac{b'(\kappa)}{b(\kappa)}\right]\\ &= \frac{N(N+2)^2}{b^3(\kappa)}\left[\frac{N}{b(\kappa)}\psi'\left(\frac{N}{b(\kappa)}\right)-1\right].\end{aligned}$$

Finally, when the innovations are Gaussian $b(\kappa)=2$ so that

$$E\left[-h_{\kappa\kappa}(0)\right]=\frac{N(N+2)^2}{8}\left[\frac{N}{2}\psi'\left(\frac{N}{2}\right)-1\right].$$

Proposition 4

Since

$$\lim_{\alpha\to 1^-}s_{\alpha t}(\boldsymbol{\phi})=\left[\varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma\right)-1\right]+\frac{1-\varkappa}{2}(\varsigma-N)$$

we have that

$$\begin{aligned}\lim_{\alpha\to 1^-}s^2_{\alpha t}(\boldsymbol{\phi}) &= \frac{(1-\varkappa)^2}{4}(\varsigma-N)^2+1-2\varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma\right)+\varkappa^{-N}\exp\left(\frac{\varkappa-1}{\varkappa}\varsigma\right)\\ &\quad +2\left[\varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma\right)-1\right]\frac{1-\varkappa}{2}(\varsigma-N).\end{aligned}$$

To obtain the variance of $s_{\alpha t}(\boldsymbol{\phi})$ we can use Lemma 1 to obtain

$$E\left[\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma\right)\right]=\varkappa^{N/2},\ \ E\left[\exp\left(\frac{\varkappa-1}{\varkappa}\varsigma\right)\right]=\left(\frac{\varkappa}{2-\varkappa}\right)^{N/2}$$

and

$$E\left[\varsigma\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma\right)\right]=N\varkappa^{N/2+1}.$$

Rearranging terms yields the desired result.

Proposition 5

Trivial.

Proposition 6

It is in the body of the paper.

B Auxiliar results

B.1 Positivity of Laguerre expansions

To identify the region in the $\boldsymbol{\eta}$-hyperplane for which $P_J(\varsigma) = 1 + \sum_{j=2}^{J} c_j(j) \cdot p_j(\mathbf{t}, N) \geq 0$ consider a given value of $\mathbf{t} \in \mathbb{R}^{J-2}$. For each such value the equation $P_J(\varsigma) = 0$ defines a straight line in the $\boldsymbol{\eta}$-hyperplane. To determine the envelope i.e. set of $\boldsymbol{\eta}$, as a function of $\mathbf{t}$, such that $P_J(\varsigma)$ remains zero for small variations of $\mathbf{t}$, we should also impose $\partial P_J(\varsigma)/\partial\varsigma = 0$ by solving

$$\begin{cases} 1 + \sum_{j=2}^{J} c_j \cdot p_j(\mathbf{t}, N) = 0 \\ \sum_{j=2}^{J} c_j \cdot \partial p_j(\mathbf{t}, N)/\partial \mathbf{t} = 0. \end{cases}$$

The first equation defines a straight line in the $\boldsymbol{\eta}$-space such that in any neighbourhood of the solution we will find positive and negative densities. In contrast, the second equation guarantees that we remain on the frontier as we move in the $\boldsymbol{\eta}$-space. Once this bound is found it remains to determine the subregion in which $P_J(\varsigma) \geq 0$ holds.

B.1.1 Second order expansion

In the simplest case $1 + c_2 \cdot p_2(t, N) \geq 0$ we can obtain the region in $\mathbb{R}$ directly: it is determined by those values of c_2 for which the polynomial $1 + c_2 \cdot p_2(t, N)$ has either complex roots or a doble root.

B.1.2 Third order expansion

For a given ς, the 3^{rd} order polynomial frontier that guarantees positivity must satisfy the following two equations $P_3(\varsigma) = 0$ and $\partial P_3(\varsigma)/\partial\varsigma = 0$; specifically,

$$\begin{cases} c_2 \cdot p_2(t, N) + c_3 \cdot p_3(t, N) + 1 = 0 \\ c_2 \cdot \partial p_2(t, N)/\partial t - c_3 \cdot \partial p_3(t, N)/\partial t = 0 \end{cases}$$

involves a system of two equations in two unknowns so that

$$c_2(t) = \frac{8 + 6N + N^2 - 8t - 2Nt + t^2}{8A(N, t)} \quad \text{and} \quad c_3(t) = \frac{N + 2 - t}{2A(N, t)}$$

with

$$\begin{aligned} A(N, t) &= \frac{N^3 t + N t^3 - 5N^2}{24} + \frac{t^3 - N^3}{12} - \frac{N^4 + t^4}{96} \\ &\quad + \frac{Nt - 2N}{3} + \frac{N^2 t - N t^2 - t^2}{4} - \frac{N^2 t^2}{16}. \end{aligned}$$

Nevertheless, these conditions are overly restrictive because they do not take into account the non-negativity of ς, and hence we have to focus on the envelope defined by ς taking values on the positive real line.

B.2 Higher order moments

The higher order moment parameter of spherical random variables, $\tau_m(\boldsymbol{\eta})$, which satisfy

$$E[\varsigma_t^m|\boldsymbol{\eta}] = [1 + \tau_m(\boldsymbol{\eta})]E[\varsigma_t^m|\mathbf{0}] \text{ where } E[\varsigma_t^m|\mathbf{0}] = 2^m \prod_{j=1}^{m}(N/2 + j - 1),$$

are given as follows.

(a) Student t distribution with $\nu = 1/\eta$ degrees of freedom:

$$1 + \tau_m^t(\eta) = (1 - 2\eta)^{m-1} \prod_{j=2}^{m} \frac{1}{(1 - 2j\eta)} \quad \text{when } \eta < (2m)^{-1}.$$

(b) Kotz distribution with excess kurtosis κ:

$$1 + \tau_m^k(\kappa, N) = \left(\frac{(N+2)\kappa + 2}{2}\right)^m \prod_{j=1}^{m} \frac{N/[(N+2)\kappa + 2] + j - 1}{N/2 + j - 1}.$$

(c) DSMN distribution with mixing probability α and variance ratio $\varkappa$:

$$1 + \tau_m^{ds}(\alpha, \varkappa) = \frac{\alpha + (1-\alpha)\,\varkappa^m}{[\alpha + (1-\alpha)\varkappa]^m}.$$

(d) 3^{rd}-order polynomial expansion distribution with parameters c_2 and c_3:

$$1 + \tau_m^{pe}(\alpha, \varkappa) = 1 + \frac{2m(m-1)}{N(N+2)}c_2 - \frac{4m(m-1)(m-2)\mathbf{1}\{m \geq 3\}}{N(N+2)(N+4)}c_3.$$

Derivation of the results:

(a) If ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν and variance 2ν, with ζ_t and ξ_t mutually independent, then the uncentered moments of integer order r of $(\nu/N) \times (\zeta_t/\xi_t)$ are given by

$$E\left[\left(\frac{\zeta_t/N}{\xi_t/\nu}\right)^r\right] = \left(\frac{\nu}{N}\right)^r \frac{r-1+N/2}{-1+\nu/2}\frac{r-2+N/2}{-2+\nu/2} \times \cdots \times \frac{1+N/2}{-(r-1)+\nu/2}\frac{N/2}{-r+\nu/2}$$

(Mood, Graybill and Boes, 1974). Given that $\varsigma_t = (\nu-2)\zeta_t/\xi_t$, it is straightforward to see that

$$E\left[\left((\nu-2)\frac{\zeta_t}{\xi_t}\right)^m\right] = \frac{N}{2}\left[\frac{2(\nu-2)}{\nu}\right]^{m-1} \prod_{j=2}^{m} \frac{(N/2+j-1)\nu}{\nu - 2j}$$

from where the result follows directly.

(b) The result follows from the higher order moments of the Gamma distribution, which are given in Lemma 1 when $b = 0$. Then, using the fact that when $\boldsymbol{\varepsilon}_t^*$ is distributed as a standardised Kotz ς_t is a gamma random variable with mean N and variance $N[(N+2)\kappa + 2]$, we obtain the desired result.

(c) When $\boldsymbol{\varepsilon}_t^*$ is distributed as a discrete mixture of normals, ς_t is a two-component scale mixture of χ_N^2's, so that conditioning on the mixing variate s,

$$E[\varsigma_t^m|s=1] = \left(\frac{1}{\alpha + (1-\alpha)\varkappa}\right)^m E[(\varsigma_t^0)^m] \text{ and } E[\varsigma_t^m|s=0] = \left(\frac{\varkappa}{\alpha + (1-\alpha)\varkappa}\right)^m E[(\varsigma_t^0)^m]$$

where ς_t^0 is a χ_N^2 variate. Then, the result follows directly.

(d) Since $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t)|\mathbf{0}] = 0$ for $m < j$, we only need to compute $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t)|\mathbf{0}]$ for for $m \geq j$, which can be written in terms of the higher order moments of the Gaussian distribution. For the 2^{nd}-order Laguerre polynomial we have

$$\begin{aligned}
E[\varsigma_t^m p_{N/2-1,2}(\varsigma_t)|\mathbf{0}] &= \frac{1}{2}E[\varsigma_t^m|\mathbf{0}] - \frac{1}{N}E[\varsigma_t^{m+1}|\mathbf{0}] + \frac{1}{2N\,(N+2)}E[\varsigma_t^{m+2}|\mathbf{0}] \\
&= \left[\frac{1}{2} - \frac{2(N/2+m+1)}{N} + \frac{4(N/2+m+1)(N/2+m+2)}{2N\,(N+2)}\right]E[\varsigma_t^m|\mathbf{0}] \\
&= \frac{2m(m-1)}{N(N+2)}E[\varsigma_t^m|\mathbf{0}].
\end{aligned}$$

The same procedure applied to the 3^{rd}-order Laguerre polynomial yields the result.

B.3 Moment generating functions

The moment generating function of spherical random variables, $\Upsilon_{\boldsymbol{\eta}}(\tau) \equiv E[e^{\tau\varsigma}|\boldsymbol{\eta}]$, are given as follows.

(a) Kotz distribution with excess kurtosis κ:

$$\Upsilon_k(\tau) \equiv E[e^{\tau\varsigma}|\kappa] = \{1 - [(N+2)\kappa + 2]\,\tau\}^{-N/[(N+2)\kappa+2]}\,.$$

(b) DSMN distribution with mixing probability α and variance ratio $\varkappa$:

$$\Upsilon_{ds}(\tau) \equiv E[e^{\tau\varsigma_t}|(\alpha,\varkappa)'] = \alpha\left[1 - \frac{2\tau}{\alpha+(1-\alpha)\varkappa}\right]^{-N/2} + (1-\alpha)\left[1 - \frac{2\varkappa\tau}{\alpha+(1-\alpha)\varkappa}\right]^{-N/2}.$$

(c) 3^{rd}-order polynomial expansion with parameters c_2 and c_3:

$$\Upsilon_{pe}^{J=3}(\tau) \equiv E[e^{\tau\varsigma_t}|(c_2,c_3)'] = (1-2\tau)^{-N/2}\left[1 + \frac{2\tau^2}{(1-2\tau)^2}c_2 - \frac{4\tau^3}{(1-2\tau)^3}c_3\right].$$

Derivation of the results:

(a) The result follows directly from the moment generating function of the Gamma distribution.

(b) Since ς_t is a two-component scale mixture of $\chi_N^2{}'s$, conditioning on s we can compute $E[e^{\tau\varsigma_t}|(\alpha,\varkappa)']$ from $\varsigma_t|s=1$ and $\varsigma_t|s=0$ which are Gamma variates with shape parameter $N/2$ and scale parameters

$$\frac{2}{\alpha+(1-\alpha)\varkappa} \quad \text{and} \quad \frac{2\varkappa}{\alpha+(1-\alpha)\varkappa}$$

respectively.

(c) The moment generating function of the polynomial expansion distribution can be easily obtained by applying Lemma 1. For the 2^{nd}-order Laguerre polynomial we have

$$\begin{aligned}
E[e^{\tau\varsigma_t}|(c_2,c_3)'] &= E[e^{\tau\varsigma_t}p_{N/2-1,2}(\varsigma_t)|\mathbf{0}] \\
&= \frac{1}{2}E[e^{\tau\varsigma_t}|\mathbf{0}] - \frac{1}{N}E[\varsigma_t e^{\tau\varsigma_t}|\mathbf{0}] + \frac{1}{2N\,(N+2)}E[\varsigma_t^2 e^{\tau\varsigma_t}|\mathbf{0}] \\
&= \frac{1}{2}\left(\frac{1}{1-2\tau}\right)^{N/2} - \left(\frac{1}{1-2\tau}\right)^{N/2+1} + \frac{1}{2}\left(\frac{1}{1-2\tau}\right)^{N/2+2} \\
&= (1-2\tau)^{-N/2}\left[\frac{(1-2\tau)^2 - 2(1-2t)+1}{2(1-2\tau)^2}\right] \\
&= (1-2\tau)^{-N/2}\,\frac{2\tau^2}{(1-2\tau)^2}.
\end{aligned}$$

The same procedure applied to the 3^{rd}-order Laguerre polynomial yields the result.

B.4 Projection of the score onto Laguerre polynomials

B.4.1 Kotz

The purpose here is to compute the coefficients

$$w_j(\kappa) = \frac{E[s_\kappa(\boldsymbol{\theta},\kappa)\cdot p_{N/2-1,\kappa,j}(\varsigma_t(\boldsymbol{\theta}))]}{E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]}.$$

Obviously, $E[s_\kappa(\boldsymbol{\theta},\kappa)\cdot p_{N/2-1,\kappa,0}(\varsigma_t(\boldsymbol{\theta}))] = 0$; while $E[s_\kappa(\boldsymbol{\theta},\kappa)\cdot p_{N/2-1,\kappa,1}(\varsigma_t(\boldsymbol{\theta}))] = 0$ in light of

$$E\left\{\frac{\varsigma_t^2}{N} - \varsigma_t\ln\varsigma_t + \varsigma_t\left[\psi\left(\frac{N}{b(\kappa)}\right) + \ln b(\kappa) - 1\right]\right\} = 0.$$

To obtain $E[s_\kappa(\boldsymbol{\theta},\kappa)\cdot p_{N/2-1,\kappa,j}(\varsigma_t(\boldsymbol{\theta}))]$ for $j\geq 2$, we only need to compute $E[\varsigma_t^j\ln\varsigma_t]$ and $E[\varsigma_t^j]$. The latter is simply

$$E[\varsigma_t^j] = \prod_{i=1}^{j}\left[N + b(\kappa)\,(i-1)\right];$$

as for the first one, from Lemma 2 we have

$$E[\varsigma_t^j\ln\varsigma_t] = \left[\psi\left(\frac{N}{b(\kappa)} + j\right) + \ln b(\kappa)\right]\cdot\prod_{i=1}^{j}\left[N + b(\kappa)\,(i-1)\right].$$

Hence, we can show that for $j\geq 2$,

$$E[s_\kappa(\boldsymbol{\theta},0)\cdot p_{N/2-1,\kappa,j}(\varsigma_t(\boldsymbol{\theta}))] = \frac{N\,(N+2)}{j\cdot b(\kappa)},$$

which together with

$$E[p_{N/2-1\kappa,j}(\varsigma_t(\boldsymbol{\theta}))] = b^2(\kappa)\prod_{i=1}^{j}\left[\frac{N + b(\kappa)\,(i-1)}{b(\kappa)i}\right],$$

yield the following weights for $j \geq 2$,

$$\begin{aligned}
w_j(\kappa) &= \frac{N(N+2)}{j \cdot b(\kappa)} \bigg/ \left\{ b^2(\kappa) \prod_{i=1}^{j} \left[\frac{N + b(\kappa)(i-1)}{b(\kappa) i} \right] \right\} \\
&= \frac{N+2}{j \cdot b^2(\kappa)} \prod_{i=2}^{j} \left[\frac{b(\kappa) i}{N + b(\kappa)(i-1)} \right].
\end{aligned}$$

For instance, the first three coefficients are

$$\begin{aligned}
w_2(\kappa) &= \frac{2+N}{b(\kappa)\left[N + b(\kappa)\right]}, \; w_3(\kappa) = \frac{2(2+N)}{\left[N + b(\kappa)\right]\left[N + 2b(\kappa)\right]}, \text{ and} \\
w_4(\kappa) &= \frac{6b(\kappa)(2+N)}{\left[N + b(\kappa)\right]\left[N + 2b(\kappa)\right]\left[N + 3b(\kappa)\right]}.
\end{aligned}$$

B.4.2 Discrete scale mixture of normals

Similarly, in this subsection, we compute $w_j^{ds}(\varkappa, I)$ defined as

$$w_j^{ds}(\varkappa, I) = \frac{E[s_{\alpha t}(\boldsymbol{\phi};\alpha = I) \cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]}{E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]}$$

with $I \in \{0, 1\}$. As in the Kotz case, we can show that $w_0^{ds}(\varkappa, I) = w_1^{ds}(\varkappa, I) = 0$. Next, to obtain $E[s_{\alpha t}(\boldsymbol{\phi};\alpha = I) \cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))]$ for $j \geq 2$, we only need to compute $E[\varsigma_t^j \exp(b\varsigma_t)]$ and $E[\varsigma_t^j]$. The latter is simply

$$E[\varsigma_t^j] = \prod_{i=1}^{j} \left[N + 2(i-1)\right];$$

as for the first one, we can use Lemma 1, to show that

$$\begin{aligned}
E[\varsigma_t^j \exp(b\varsigma_t)] &= \frac{\Gamma(N/2 + j)}{\Gamma(N/2)} \left(\frac{1}{2}\right)^{N/2} \cdot \left(\frac{2}{1-2b}\right)^{N/2+j} \\
&= \left(\frac{1}{1-2b}\right)^{N/2+j} \prod_{i=1}^{j} \left[N + 2(i-1)\right].
\end{aligned}$$

On this basis, we can show that for $j \geq 2$,

$$E[s_{\alpha t}(\boldsymbol{\phi};\alpha = 0) \cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] = \left(\frac{1-\varkappa}{\varkappa}\right)^j 2 \prod_{i=1}^{j} \left[\frac{N + 2(i-1)}{2i}\right]$$

and

$$E[s_{\alpha t}(\boldsymbol{\phi};\alpha = 1) \cdot p_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] = (-1)^{j-1} (1-\varkappa)^j \, 2 \prod_{i=1}^{j} \left[\frac{N + 2(i-1)}{2i}\right],$$

which together with

$$E[p^2_{N/2-1,j}(\varsigma_t(\boldsymbol{\theta}))] = 4 \prod_{i=1}^{j} \left[\frac{N + 2(i-1)}{2i}\right],$$

yield the following weights $w_j^{ds}(\varkappa, I)$ for $j \geq 2$,

$$w_j^{ds}(\varkappa, 0) = \frac{1}{2} \left(\frac{1-\varkappa}{\varkappa}\right)^j \quad \text{and} \quad w_j^{ds}(\varkappa, 1) = \frac{(-1)^{j-1}}{2} (1-\varkappa)^j.$$

C Additional Propositions

Proposition C.1 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$*, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then*

$$\frac{\sqrt{T}}{T}\sum_t \left[\frac{s_\eta(\boldsymbol{\theta}_0,0) - E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}{V^{1/2}[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}\right] \xrightarrow{d} N(0,1)$$

where $E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ *and* $E[s_\eta^2(\boldsymbol{\theta}_0,0)\big|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ *are given below.*

(a) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ *with* $\nu_0 > 8$ *then*

$$E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\left(\frac{\nu_0-2}{\nu_0-4}-1\right)$$

and

$$\begin{aligned} E[s_\eta^2(\boldsymbol{\theta}_0,0)\big|\,\boldsymbol{\phi}_0] &= -\frac{3N^2(N+2)^2}{16} + \frac{N(N+2)^2(3N+4)}{8}\frac{\nu_0-2}{\nu_0-4} \\ &\quad -\frac{N(N+2)^2(N+4)}{4}\frac{(\nu_0-2)^2}{(\nu_0-4)(\nu_0-6)} \\ &\quad +\frac{N(N+2)(N+4)(N+6)}{16}\frac{(\nu_0-2)^3}{(\nu_0-4)(\nu_0-6)(\nu_0-8)}. \end{aligned}$$

(b) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim Kotz(\mathbf{0}, \mathbf{I}_N, \kappa_0)$*, then*

$$E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\kappa_0$$

and

$$E[s_\eta^2(\boldsymbol{\theta}_0,0)\big|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{16}\left[8 + 8(N+3)\kappa_0 + (N+2)(3N+20)\kappa_0^2 + 6(N+2)^2\kappa_0^3\right].$$

(c) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim DSMN(\mathbf{0}, \mathbf{I}_N, \alpha_0, \varkappa_0)$*, then*

$$E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\left\{\frac{\alpha_0 + (1-\alpha_0)\varkappa_0^2}{[\alpha_0 + (1-\alpha_0)\varkappa_0]^2} - 1\right\}$$

and

$$\begin{aligned} E[s_\eta^2(\boldsymbol{\theta}_0,0)\big|\,\boldsymbol{\phi}_0] &= -\frac{3N^2(N+2)^2}{16} + \frac{N(N+2)^2(3N+4)}{8}\frac{\alpha_0 + (1-\alpha_0)\varkappa_0^2}{[\alpha_0 + (1-\alpha_0)\varkappa_0]^2} \\ &\quad -\frac{N(N+2)^2(N+4)}{4}\frac{\alpha_0 + (1-\alpha_0)\varkappa_0^3}{[\alpha_0 + (1-\alpha_0)\varkappa_0]^3} \\ &\quad -\frac{N(N+2)(N+4)(N+6)}{16}\frac{\alpha_0 + (1-\alpha_0)\varkappa_0^4}{[\alpha_0 + (1-\alpha_0)\varkappa_0]^4}. \end{aligned}$$

(d) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim p.e.(\mathbf{0}, \mathbf{I}_N, c_{20}, c_{30})$*, then*

$$E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = c_{20}$$

and

$$E[s_\eta^2(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{2} + 2(N+8)c_{20} - 24c_{30}.$$

Proof. Writing the expectations in terms of the higher order moment parameter of a spherical random variable $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$, $\tau_m^s(\boldsymbol{\eta}_0)$,

$$E[s_\eta(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\tau_2^s(\boldsymbol{\eta}_0)$$

and

$$\begin{aligned} E[s_\eta^2(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] &= -\frac{3N^2(N+2)^2}{16} + \frac{N(N+2)^2(3N+4)}{8}[1+\tau_2^s(\boldsymbol{\eta}_0)] \\ &\quad -\frac{N(N+2)^2(N+4)}{4}[1+\tau_3^s(\boldsymbol{\eta}_0)] + \frac{N(N+2)(N+4)(N+6)}{16}[1+\tau_4^s(\boldsymbol{\eta}_0)], \end{aligned}$$

the results follow from the expressions for $\tau_m(\boldsymbol{\eta}_0)$ given in Appendix B.2.

Proposition C.2 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then*

$$\frac{\sqrt{T}}{T}\sum_t \left[\frac{s_\kappa(\boldsymbol{\theta}_0,0) - E[s_\kappa(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}{V^{1/2}[s_\kappa(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}\right] \xrightarrow{d} N(0,1)$$

where $E[s_\kappa(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ and $E[s_\kappa^2(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ are given below.

(a) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim Kotz(\mathbf{0}, \mathbf{I}_N, \kappa_0)$, then

$$E[s_\kappa(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\left[\psi\left(\frac{N}{2}\right) - \psi\left(\frac{N}{b(\kappa_0)}\right) + \ln\left(\frac{2}{b(\kappa_0)}\right)\right]$$

and

$$\begin{aligned} E[s_\kappa^2(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] &= \frac{N^2(N+2)^2}{16}\left[\psi'\left(\frac{N}{b(\kappa_0)}\right) - \frac{b(\kappa_0)}{N}\right] \\ &\quad + \frac{N^2(N+2)^2}{16}\left[\psi\left(\frac{N}{b(\kappa_0)}\right) - \psi\left(\frac{N}{2}\right) + \ln\frac{b(\kappa_0)}{2}\right]^2. \end{aligned}$$

(b) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim t(\mathbf{0}, \mathbf{I}_N, \nu_0)$, then

$$E[s_\kappa(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)}{4}\left[\psi\left(\frac{\nu_0}{2}\right) - \ln\left(\frac{\nu_0-2}{2}\right)\right]$$

and

$$\begin{aligned} E[s_\kappa^2(\boldsymbol{\theta}_0,0)|\,\boldsymbol{\phi}_0] &= \frac{N^2(N+2)^2}{16}\left\{\frac{2(\nu_0-6)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-2)} + 2\ln 2\,[1-\ln(\nu_0-2)] + \ln^2(\nu_0-2)\right. \\ &\quad \left. + \psi'\left(\frac{N}{2}\right) + \psi'\left(\frac{\nu_0}{2}\right) + \psi\left(\frac{N}{2}\right)\left[\psi\left(\frac{N}{2}\right) - \ln\left(\frac{\nu_0-2}{2}\right)\right]\right\} \end{aligned}$$

(c) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim DSMN(\mathbf{0}, \mathbf{I}_N, \alpha_0, \varkappa_0)$*, then*

$$E[s_\kappa(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0] = -\frac{N(N+2)}{4}\left\{(1-\alpha_0)\ln\varkappa_0 + \ln[\alpha_0 + (1-\alpha_0)\varkappa_0]\right\}$$

and

$$\begin{aligned} E[s_\kappa^2(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0] &= \frac{N(N+2)^2}{16}\left\{(N+2)\frac{\alpha_0 + (1-\alpha_0)\varkappa_0^2}{[\alpha_0 + (1-\alpha_0)\varkappa_0]^2} - (N+4)\right\} \\ &+ \frac{N^2(N+2)^2}{16}\left\{\psi'\left(\frac{N}{2}\right) + \ln^2[\alpha_0 + (1-\alpha_0)\varkappa_0] + (1-\alpha_0)\ln^2\varkappa_0\right\} \\ &+ \frac{N^2(N+2)^2}{8}\frac{\alpha_0(1-\alpha_0)(1-\varkappa_0)}{\alpha_0 + (1-\alpha_0)\varkappa_0}\ln\varkappa_0. \end{aligned}$$

(d) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim p.e.(\mathbf{0}, \mathbf{I}_N, c_{20}, c_{30})$*, then*

$$E[s_\kappa(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0] = \frac{1}{2}c_{20} + \frac{2}{(N+4)}c_{30}$$

and

$$\begin{aligned} E[s_\kappa^2(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0] &= \frac{N}{32(N+4)}\Big[64 + 80N - 8(N+2)^2(N+4) + N^3\left(4 + 24\log^2 2 - \log 16\log 64\right) \\ &+ N^2\left(32 + 48\log^2 2 - \log 64\log 256\right) + 2N(N+2)^2(N+4)\psi^{(1)}\left(\frac{N}{2}\right)\Big] + \frac{1}{N}c_{20}. \end{aligned}$$

Proof. (a) Given that $E\left[(\varsigma_t/N - 1)\,|\kappa_0\right] = 0$ regardless of the value of κ_0, we can use the fact that ς_t is a gamma random variable with mean N and variance $N[(N+2)\kappa_0 + 2]$ and apply the results in Lemma 2 to obtain the first expression. Similarly, expanding $s_\kappa^2(\boldsymbol{\theta}, 0)$,

$$\begin{aligned} s_\kappa^2(\boldsymbol{\theta}, 0) &= \frac{N^2(N+2)^2}{16}\left\{\left(\frac{\varsigma_t}{N} - 1\right)^2 + \left[\ln\varsigma_t - \left(\psi\left(\frac{N}{2}\right) + \ln 2\right)\right]^2\right. \\ &\left. - 2\left(\frac{\varsigma_t}{N} - 1\right)\left[\ln\varsigma_t - \left(\psi\left(\frac{N}{2}\right) + \ln 2\right)\right]\right\}, \end{aligned}$$

the second expression follows from

$$E\left[\left(\frac{\varsigma_t}{N} - 1\right)^2\middle|\,\kappa_0\right] = \frac{b(\kappa_0)}{N},$$

and, again applying Lemma 2, to obtain

$$\begin{aligned} E\left\{\left[\ln\varsigma_t - \left(\psi\left(\frac{N}{2}\right) + \ln 2\right)\right]^2\middle|\,\kappa_0\right\} &= \psi'\left(\frac{N}{b(\kappa_0)}\right) + \psi^2\left(\frac{N}{b(\kappa_0)}\right) + 2\psi\left(\frac{N}{b(\kappa_0)}\right)\ln b(\kappa_0) \\ &+ \ln^2 b(\kappa_0) + \left(\psi\left(\frac{N}{2}\right) + \ln 2\right)^2 \\ &- 2\left[\psi\left(\frac{N}{2}\right) + \ln 2\right]\left[\psi\left(\frac{N}{b(\kappa_0)}\right) + \ln b(\kappa_0)\right], \end{aligned}$$

and

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\ln\varsigma_t-\left(\psi\left(\frac{N}{2}\right)+\ln 2\right)\right]\Big|\kappa_0\right\}=E\left\{\left(\frac{\varsigma_t}{N}-1\right)\ln\varsigma_t\Big|\kappa_0\right\}=\frac{b(\kappa_0)}{N}.$$

(b) To obtain the relevant quantities we exploit the fact that the squared Euclidean norm of the standardized Student t innovations, ς_t, is independently and identically distributed as $N(\nu-2)/\nu$ times an F variate with N and ν degrees of freedom when $\nu<\infty$, say $\varsigma_t=\frac{N(\nu-2)}{\nu}\frac{Y_1/N}{Y_2/\nu}$ with $Y_1\sim\chi^2_N$ and $Y_2\sim\chi^2_\nu$. Then, Lemma 2 applied to $Y=2t$ where $t\sim$Gamma$(m/2,1)$ yields $E\left[\ln\left(Y/m\right)\right]=\psi\left(m/2\right)+\ln\left(2/m\right)$ so that

$$E[\ln\varsigma_t|\nu_0]=\psi\left(\frac{N}{2}\right)-\psi\left(\frac{\nu_0}{2}\right)+\ln\left(\nu_0-2\right)$$

and the first expression follows directly. Next, we can again apply Lemma 2 to obtain

$$\begin{aligned}E\left[\ln^2\varsigma_t|\nu_0\right]&=E\left\{\left[\ln\left(\frac{N(\nu_0-2)}{\nu_0}\right)+\ln\left(\frac{Y_1}{N}\right)-\ln\left(\frac{Y_2}{\nu_0}\right)\right]^2\right\}\\&=\psi'\left(\frac{N}{2}\right)+\psi'\left(\frac{\nu_0}{2}\right)+\left[\ln\left(\nu_0-2\right)+\psi\left(\frac{N}{2}\right)-\psi\left(\frac{\nu_0}{2}\right)\right]^2\end{aligned}$$

$$E\left[\varsigma_t\ln\varsigma_t|\nu_0\right]=N\left[\psi\left(\frac{N}{2}+1\right)-\psi'\left(\frac{\nu_0}{2}-1\right)+\ln\left(\nu_0-2\right)\right]$$

and

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\ln\varsigma_t\Big|\nu_0\right\}=\frac{2}{N}+\frac{2}{\nu_0-2}.$$

Rearranging terms yields the second expression.

(c) Conditioning on the mixing variate s –which determines the scale parameter ξ_s– Lemma 2 implies

$$E\left[\ln\xi_s\varsigma_t\right]=\psi\left(\frac{N}{2}\right)+\ln(2\xi_s),$$

and then

$$\begin{aligned}E[\ln\varsigma_t|(\alpha_0,\varkappa_0)']&=\alpha_0E[\ln\varsigma_t|s=1]+(1-\alpha_0)\,E[\ln\varsigma_t|s=0]\\&=\psi\left(\frac{N}{2}\right)+\ln\left[\frac{2}{\alpha_0+(1-\alpha_0)\varkappa_0}\right]+(1-\alpha_0)\ln\varkappa_0,\end{aligned}$$

from where we obtain the first expression. Similarly,

$$E\left[\ln^2(\xi_s\varsigma_t)\right]=\psi'\left(\frac{N}{2}\right)+\psi^2\left(\frac{N}{2}\right)+2\psi\left(\frac{N}{2}\right)\ln(2\xi_s)+\ln^2(2\xi_s),$$

so that

$$\begin{aligned}E[\ln^2\varsigma_t|(\alpha_0,\varkappa_0)']&=\alpha_0E[\ln^2\varsigma_t|s=1]+(1-\alpha_0)\,E[\ln^2\varsigma_t|s=0]\\&=\psi'\left(\frac{N}{2}\right)+\psi^2\left(\frac{N}{2}\right)+2\psi\left(\frac{N}{2}\right)\left\{\ln\left[\frac{2}{\alpha_0+(1-\alpha_0)\varkappa_0}\right]+(1-\alpha_0)\ln\varkappa_0\right\}\\&\quad+\alpha_0\ln^2\left[\frac{2}{\alpha_0+(1-\alpha_0)\varkappa_0}\right]+(1-\alpha_0)\ln^2\left[\frac{2\varkappa_0}{\alpha_0+(1-\alpha_0)\varkappa_0}\right];\end{aligned}$$

Analogous computations yield $E[\varsigma_t \ln \varsigma_t|(\alpha_0, \varkappa_0)']$. Rearranging terms we obtain the expressions for $E[s_\kappa(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0]$ and $E[s^2_\kappa(\boldsymbol{\theta}_0, 0)|\,\boldsymbol{\phi}_0]$.

(d) Expanding the density of the polynomial expansion appropriately, which can be written in terms of the higher order moments of the Gaussian distribution –or equivalently in terms of Gamma random variables with different scale parameters– we can again apply Lemma 2 to obtain $E[\varsigma_t^j \ln \varsigma_t|\mathbf{0}]$ for $j \geq 0$. Then, using the recurrence relation for integer and half-integer values of the Digamma function (see Abramowitz and Stegun, 1964) we obtain

$$E[\ln \varsigma_t|(c_{20}, c_{30})'] = \psi\left(\frac{N}{2}\right) + \ln 2 - c_{20} - \frac{2}{3}c_{30}$$

and

$$E[\varsigma_t \ln \varsigma_t|(c_{20}, c_{30})'] = N\left[\psi\left(\frac{N}{2}+1\right) + 3\ln 2 + c_{20} + \frac{1}{3}c_{30}\right].$$

As for $E[\ln \varsigma_t^2|(c_{20}, c_{30})']$, using the recursive relation for integer and half-integer values of the Trigamma function (see Abramowitz and Stegun, 1964) it is tedious but otherwise straightforward to show that

$$\begin{aligned} E[\ln \varsigma_t^2|(c_{20}, c_{30})'] &= \ln^2 2 + 2\ln 2\psi\left(\frac{N}{2}\right) + \psi^2\left(\frac{N}{2}\right) + \psi'\left(\frac{N}{2}\right) \\ &\quad +2\left[1 - \ln 2 - \psi\left(\frac{N}{2}\right)\right]c_{20} + \left[2 - \frac{4}{3}\left(\ln 2 + \psi\left(\frac{N}{2}\right)\right)\right]c_{30}, \end{aligned}$$

from where the results can be easily derived.

Proposition C.3 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then*

$$\frac{\sqrt{T}}{T}\sum_t \left[\frac{s_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa}) - E[s_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa})|\,\boldsymbol{\theta}_0, \boldsymbol{\eta}_0]}{V^{1/2}[s_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa})|\,\boldsymbol{\theta}_0, \boldsymbol{\eta}_0]}\right] \xrightarrow{d} N(0,1)$$

where $E[s_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa})|\,\boldsymbol{\theta}_0, (\alpha_0, \varkappa_0)']$ and $E[s^2_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa})|\,\boldsymbol{\theta}_0, (\alpha_0, \varkappa_0)']$ are given below.

(a) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim DSMN(\mathbf{0}, \mathbf{I}_N, \alpha_0, \varkappa_0)$, then

$$E[s_\alpha(\boldsymbol{\theta}, 1^-, \bar{\varkappa})|\,\boldsymbol{\phi}_0] = 1 - \alpha_0\left[\frac{\alpha_0 + (1-\alpha_0)\varkappa_0}{1 - (1-\alpha_0)(1-\varkappa_0)\overline{\varkappa}}\right]^{N/2} - (1-\alpha_0)\left[\frac{\alpha_0 + (1-\alpha_0)\varkappa_0}{\varkappa_0 + \alpha_0(1-\varkappa_0)\overline{\varkappa}}\right]^{N/2}$$

and

$$
\begin{aligned}
E[s_{\alpha}^{2}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] &= 1+\frac{N(N+2)}{4}\left\{\frac{\alpha_0+(1-\alpha_0)\varkappa_0^2}{[(1-\varkappa_0)\alpha_0+\varkappa_0]^2}-\frac{N}{N+2}\right\}(1-\overline{\varkappa})^2\\
&\quad -\alpha_0\left[\frac{\alpha_0+(1-\alpha_0)\varkappa_0}{1-(1-\alpha_0)(1-\varkappa_0)\overline{\varkappa}}\right]^{N/2}\left\{\frac{N(1-\overline{\varkappa})\overline{\varkappa}}{1-(1-\alpha_0)(1-\varkappa_0)\overline{\varkappa}}-[2+N(1-\overline{\varkappa})]\right\}\\
&\quad +(1-\alpha_0)\left[\frac{\alpha_0+(1-\alpha_0)\varkappa_0}{\varkappa_0+\alpha_0(1-\varkappa_0)\overline{\varkappa}}\right]^{N/2}\left\{\frac{N\varkappa_0(1-\overline{\varkappa})\overline{\varkappa}}{\varkappa_0+\alpha_0(1-\varkappa_0)\overline{\varkappa}}-[2+N(1-\overline{\varkappa})]\right\}\\
&\quad +\frac{\alpha_0}{\overline{\varkappa}}\left\{\frac{\alpha_0+(1-\alpha_0)\varkappa_0}{2(1-\overline{\varkappa})+[\alpha_0+(1-\alpha_0)\varkappa_0]\overline{\varkappa}}\right\}^{N/2}\\
&\quad +\frac{1-\alpha_0}{\overline{\varkappa}}\left\{\frac{\alpha_0+(1-\alpha_0)\varkappa_0}{2\varkappa_0(1-\overline{\varkappa})+[\alpha_0+(1-\alpha_0)\varkappa_0]\overline{\varkappa}}\right\}^{N/2}.
\end{aligned}
$$

(b) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim t(\mathbf{0},\mathbf{I}_N,\nu_0)$, *then*

$$
E[s_{\alpha}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] = 1-\overline{\varkappa}^{-N/2}\frac{\Gamma\left[(N+\nu_0)/2\right]}{\Gamma(\nu_0/2)}U\left(\frac{N}{2},1-\frac{\nu_0}{2},\frac{(\nu_0-2)(1-\overline{\varkappa})}{2\overline{\varkappa}}\right)
$$

and

$$
\begin{aligned}
E[s_{\alpha}^{2}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] &= 1+\frac{N}{2}\frac{N+\nu_0-2}{\nu_0-4}(1-\overline{\varkappa})^2\\
&\quad -\left[2+N\left(1-\overline{\varkappa}\right)\right]\overline{\varkappa}^{-N/2}\frac{\Gamma\left[(N+\nu_0)/2\right]}{\Gamma(\nu_0/2)}U\left(\frac{N}{2},1-\frac{\nu_0}{2},\frac{(\nu_0-2)(1-\overline{\varkappa})}{2\overline{\varkappa}}\right)\\
&\quad +N\left(1-\overline{\varkappa}\right)\overline{\varkappa}^{-N/2}\frac{\Gamma\left[(N+\nu_0)/2\right]}{\Gamma(\nu_0/2-1)}U\left(\frac{N}{2}+1,2-\frac{\nu_0}{2},\frac{(\nu_0-2)(1-\overline{\varkappa})}{2\overline{\varkappa}}\right)\\
&\quad +\overline{\varkappa}^{-N}\frac{\Gamma\left[(N+\nu_0)/2\right]}{\Gamma(\nu_0/2)}U\left(\frac{N}{2},1-\frac{\nu_0}{2},\frac{(\nu_0-2)(1-\overline{\varkappa})}{\overline{\varkappa}}\right).
\end{aligned}
$$

(c) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim Kotz(\mathbf{0},\mathbf{I}_N,\kappa_0)$, *then*

$$
E[s_{\alpha}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] = 1-\overline{\varkappa}^{-N/2}\left[\frac{2\overline{\varkappa}}{2-\kappa_0(N+2)(\overline{\varkappa}-1)}\right]^{N/[(N+2)\kappa_0+2]}
$$

and

$$
\begin{aligned}
E[s_{\alpha}^{2}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] &= 1+\frac{N\left[(N+2)\,\kappa_0+2\right]}{4}(1-\overline{\varkappa})+\left[\frac{2\overline{\varkappa}}{2+(N+2)\kappa_0(1-\overline{\varkappa})}\right]^{N/[(N+2)\kappa_0+2]}\\
&\quad \times\left\{\frac{2N\overline{\varkappa}\,(1-\overline{\varkappa})}{2+(N+2)\kappa_0(1-\overline{\varkappa})}-\left[2+N\left(1-\overline{\varkappa}\right)\right]\right\}\overline{\varkappa}^{N/2}\\
&\quad +\left[\frac{\overline{\varkappa}}{2+(N+2)\kappa_0(1-\overline{\varkappa})+\overline{\varkappa}}\right]^{N/[(N+2)\kappa_0+2]}\overline{\varkappa}^{N}.
\end{aligned}
$$

(d) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim p.e.(\mathbf{0},\mathbf{I}_N,c_{20},c_{30})$, *then*

$$
E[s_{\alpha}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] = -\frac{(1-\overline{\varkappa})^2}{2}\left[c_{20}+c_{30}(1-\overline{\varkappa})\right]
$$

and

$$
\begin{aligned}
E[s_{\alpha}^{2}(\boldsymbol{\theta},1^{-},\bar{\varkappa})|\,\boldsymbol{\phi}_{0}] &= 1+\frac{N\left(N+2\right)}{4}\left(1-\overline{\varkappa}\right)^2\left[\frac{4\left[\left(2-\overline{\varkappa}\right)\overline{\varkappa}\right]^{-N/2}}{\left(2-\overline{\varkappa}\right)^2}-N\left(1-\overline{\varkappa}\right)^2-4\overline{\varkappa}^2\right]c_{20}\\
&\quad +\frac{N\left(N+2\right)\left(N+4\right)}{24}\left(1-\overline{\varkappa}\right)^3\left[\frac{8\left[\left(2-\overline{\varkappa}\right)\overline{\varkappa}\right]^{-N/2}}{\left(2-\overline{\varkappa}\right)^3}-N\left(1-\overline{\varkappa}\right)^2-6\overline{\varkappa}^2-2\right]c_{30}
\end{aligned}
$$

Proof. (a) Given that $E[(\varsigma_t/N-1)]=0$ regardless of the value of $\boldsymbol{\eta}_0$, we can compute the expectation of the remaining part of $s_\alpha(\boldsymbol{\theta},1^-,\overline{\varkappa})$ by conditioning on s –which determines the scale parameter ξ_s–, for $\varsigma_t|s=1$ and $\varsigma_t|s=0$ which are Gamma variates with shape parameter $N/2$ and scale parameters

$$\xi_0=\frac{2}{\alpha+(1-\alpha)\varkappa}\quad\text{and}\quad\xi_1=\frac{2\varkappa}{\alpha+(1-\alpha)\varkappa}.$$

Specifically, using Lemma 1,

$$E\left[\overline{\varkappa}^{-N/2}\exp\left(\frac{\overline{\varkappa}-1}{2\overline{\varkappa}}\xi_s\varsigma_t\right)\middle|\mathbf{0}\right]=\overline{\varkappa}^{-N/2}\left(1-\frac{\overline{\varkappa}-1}{\overline{\varkappa}}\xi_s\right)^{-N/2}=[\overline{\varkappa}+(1-\overline{\varkappa})\xi_s]^{-N/2},$$

taking expectations with respect to s and rearranging yields the first expression. As for the second expression, we can proceed in the same manner. To do so, it is convenient to expand $s_\alpha^2(\boldsymbol{\theta},1^-,\overline{\varkappa})$ into

$$\frac{1}{4}\left[2+N\left(1-\overline{\varkappa}\right)\right]^2-\frac{1}{2}\left[2+N\left(1-\overline{\varkappa}\right)\right]\left(1-\overline{\varkappa}\right)\varsigma_t+\frac{1}{4}\left(1-\overline{\varkappa}\right)^2\varsigma_t^2$$

whose expectation can be easily computed using the higher order moments of Appendix B.2, and then, the expected value of the remaining terms, which involve exponentials of ς_t, which are obtained using Lemma 1 as before

(b) The relevant quantities described in (a) can be easily computed using the fact that the squared Euclidean norm of the standardized Student t innovations, ς_t, is independently and identically distributed as $N(\nu-2)/\nu$ times an F variate with N and ν degrees of freedom when $\nu<\infty$, say $\varsigma_t=\frac{N(\nu-2)}{\nu}\frac{Y_1/N}{Y_2/\nu}$ with $Y_1\sim\chi_N^2$ and $Y_2\sim\chi_\nu^2$. Then, use of the results of Lemma 3 and rearranging yield the results.

(c) and (d) follow from direct application of Lemma 1 to the several terms involving $s_\alpha(\boldsymbol{\theta},1^-,\bar{\varkappa})$ and $s_\alpha^2(\boldsymbol{\theta},1^-,\bar{\varkappa})$, and rearranging appropiately.

Proposition C.4 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then*

$$\frac{\sqrt{T}}{T}\sum_t\left[\frac{p_{N/2-1,3}(\boldsymbol{\theta}_0)-E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}{V^{1/2}[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]}\right]\xrightarrow{d}N(0,1)$$

where $E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ and $E[p_{N/2-1,3}^2(\boldsymbol{\theta}_0)\Big|\,\boldsymbol{\theta}_0,\boldsymbol{\eta}_0]$ are given below.

(a) If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1}\sim p.e.(\mathbf{0},\mathbf{I}_N,c_{20},c_{30})$, then

$$E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\phi}_0]=c_{30}$$

and

$$E[p^2_{N/2-1,3}(\boldsymbol{\theta}_0)\big|\,\boldsymbol{\phi}_0] = \frac{N\,(N+2)\,(N+4)}{12} + \frac{(N+4)\,(N+14)}{2}c_{20} - 4(3N+22)c_{30}.$$

(b) If $\boldsymbol{\varepsilon}^*_t|\mathbf{z}_t, I_{t-1} \sim t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ *with* $\nu_0 > 12$*, then*

$$E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\phi}_0] = -\frac{N(N+2)(N+4)}{3}\frac{2}{(\nu_0-6)(\nu_0-4)}$$

and

$$\begin{aligned}E[p^2_{N/2-1,3}(\boldsymbol{\theta}_0)\big|\,\boldsymbol{\phi}_0] &= \frac{1}{72}\frac{N(N+2)(N+4)}{(\nu_0-12)\,(\nu_0-10)\,(\nu_0-8)\,(\nu_0-6)\,(\nu_0-4)}\\ &\times\left[-43392 - 15536N + 1776N^2 + 740N^3\right.\\ &+\left(20352 + 5176N + 3252N^2 + 800N^3\right)\nu_0\\ &+\left(-4944 + 4548N + 1476N^2 + 15N^3\right)\nu_0^2\\ &+\left.\left(2760 + 738N + 27N^2\right)\nu_0^3 + (12+18N)\,\nu_0^4 + 6\nu_0^5\right].\end{aligned}$$

(c) If $\boldsymbol{\varepsilon}^*_t|\mathbf{z}_t, I_{t-1} \sim Kotz(\mathbf{0}, \mathbf{I}_N, \kappa_0)$*, then*

$$E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\phi}_0] = -\frac{N(N+2)}{12}\left[(N+2)\kappa_0 - 2\right]\kappa_0$$

and

$$\begin{aligned}E[p^2_{N/2-1,3}(\boldsymbol{\theta}_0)\big|\,\boldsymbol{\phi}_0] &= \frac{1}{576}N(N+2)\\ &\times\{48(N+4) + 24(3N(N+8)+40)\kappa_0\\ &+2(N+2)(N(27N+368)+768)\kappa_0^2\\ &+(N+2)^2(N(15N+488)+1344)\kappa_0^3\\ &+26(N+2)^3(5N+24)\kappa_0^4 + 120(N+2)^4\kappa_0^5\}.\end{aligned}$$

(d) If $\boldsymbol{\varepsilon}^*_t|\mathbf{z}_t, I_{t-1} \sim DSMN(\mathbf{0}, \mathbf{I}_N, \alpha_0, \varkappa_0)$*, then*

$$E[p_{N/2-1,3}(\boldsymbol{\theta}_0)|\,\boldsymbol{\phi}_0] = \frac{N(N+2)(N+4)}{24}\frac{\left[\alpha_0(1-\alpha_0)(1-2\alpha_0)-1\right](1-\varkappa_0)^3}{\left[(1-\varkappa_0)\alpha_0+\varkappa_0\right]^3}$$

and

$$\begin{aligned}E[p^2_{N/2-1,3}(\boldsymbol{\theta}_0)\big|\,\boldsymbol{\phi}_0] &= \frac{N\,(N+2)^2\,(N+4)^2}{576}\left\{-5N + 3\,(5N+6)\,\frac{\alpha_0+(1-\alpha_0)\varkappa_0^2}{\left[\alpha_0+(1-\alpha_0)\varkappa_0\right]^2}\right.\\ &-4\,(5N+18)\,\frac{\alpha_0+(1-\alpha_0)\varkappa_0^3}{\left[\alpha_0+(1-\alpha_0)\varkappa_0\right]^3} + 3\frac{(N+6)\,(5N+16)}{(N+2)}\frac{\alpha_0+(1-\alpha_0)\varkappa_0^4}{\left[\alpha_0+(1-\alpha_0)\varkappa_0\right]^4}\\ &+\left.\frac{(N+6)\,(N+8)}{(N+2)}\left[-6\frac{\alpha_0+(1-\alpha_0)\varkappa_0^5}{\left[\alpha_0+(1-\alpha_0)\varkappa_0\right]^5} + \frac{N+10}{N+4}\frac{\alpha_0+(1-\alpha_0)\varkappa_0^6}{\left[\alpha_0+(1-\alpha_0)\varkappa_0\right]^6}\right]\right\}.\end{aligned}$$

Proof. Writing the expectations in terms of the higher order moment parameter of a spherical random variable $\varepsilon_t^*|\mathbf{z}_t, I_{t-1} \sim s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$, $\tau_m^s(\boldsymbol{\eta}_0)$,

$$
\begin{aligned}
E[p_{N/2-1,3}(\boldsymbol{\theta}_0, 0)|\, \boldsymbol{\theta}_0, \boldsymbol{\eta}_0] &= \frac{N(N+2)}{12}\left\{-(N+4) + \frac{3N+12}{2}[1+\tau_2^s(\boldsymbol{\eta}_0)] - \frac{N+4}{2}[1+\tau_3^s(\boldsymbol{\eta}_0)]\right\} \\
&= \frac{N(N+2)(N+4)}{24}\left[3\tau_2^s(\boldsymbol{\eta}_0) - \tau_3^s(\boldsymbol{\eta}_0)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
E[s_\eta^2(\boldsymbol{\theta}_0, 0)|\, \boldsymbol{\theta}_0, \boldsymbol{\eta}_0] = \; & \frac{7N^2(N+2)^2(N+4)^2}{576} + \frac{(N+2)(N+4)^2(5N+6)}{192}[1+\tau_2^s(\boldsymbol{\eta}_0)] \\
& - \frac{(N+2)(N+4)(5N+18)}{144}[1+\tau_3^s(\boldsymbol{\eta}_0)] + \frac{(N+4)(5N+16)}{192}[1+\tau_4^s(\boldsymbol{\eta}_0)] \\
& - \frac{(N+4)}{96}[1+\tau_5^s(\boldsymbol{\eta}_0)] + \frac{1}{576}[1+\tau_6^s(\boldsymbol{\eta}_0)],
\end{aligned}
$$

the results follow from the expressions for $\tau_m(\boldsymbol{\eta}_0)$ given in Appendix B.2.

References

Abramowitz, M. and Stegun I.A. (1964). *Handbook of mathematical functions*, AMS 55, National Bureau of Standards.

Andrews, D. (1994). "Empirical process methods in econometrics", In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, 2247–2294. North Holland.

Andrews, D. (2001). "Testing when a parameter is on the boundary of the mantained hypothesis", *Econometrica* 69, 683–734.

Balestra, P. and Holly, A. (1990). "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.

Berk, J. (1997). "Necessary conditions for the CAPM", *Journal of Economic Theory* 73, 245-257.

Berkane, M. and Bentler, P.M. (1986). "Moments of elliptically distributed random variates", *Statistics and Probability Letters* 4, 333-335.

Berndt, E. R., Hall, B. H., Hall, R. E. and Hausman, J. A. (1974). "Estimation and inference in nonlinear structural models", *Annals of Economic and Social Measurement* 3, 653–665.

Bollerslev, T. (1987). "A conditionally heteroskedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics* 69, 542–547.

Bollerslev, T. and Wooldridge, J. M. (1992). "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.

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

Bontemps, C. and Meddahi, N. (2010). "Testing distributional assumptions: A GMM approach", mimeo, Toulouse School of Economics.

Calzolari, G., Fiorentini, G. and Sentana, E. (2004). "Constrained indirect estimation", *Review of Economic Studies* 71, 945-973.

Chamberlain, G. (1983). "A characterization of the distributions that imply mean-variance utility functions", *Journal of Economic Theory* 29, 185-201.

Cressie, N.A.C., Davis A.S., Folks J.L. and Policello G.E. (1981). "The moment generating function and negative integer moments", *American Statistician* 35, 148-150.

Davidson, R. and MacKinnon, J.G. (1998). "Graphical methods for investigating the size and power of tests statistics", *The Manchester School* 66, 1-26.

Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric multivariate and related distributions*, Chapman and Hall.

Fiorentini, G., Calzolari, G. and Panattoni, L. (1996). "Analytical Derivatives and the

Computation of GARCH Models," *Journal of Applied Econometrics* 11, 399–417.

Fiorentini, G. and Sentana, E. (2010). "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working Paper 0713, revised October 2010.

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

Gallant, A. R. and Tauchen, G. (1996). "Which moments to match?," *Econometric Theory* 12, 657–681.

Hall, P. and Yao, Q. (2003). "Inference in Arch and Garch models with heavy-tailed errors", *Econometrica* 71, 285-317.

Kiefer, N.M. and Salmon M. (1983). "Testing normality in econometric models", *Economics Letters* 11, 123-127.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G. (1989). "Robust statistical modeling using the t distribution", *Journal of the American Statistical Association* 84, 881-896.

Lee, L. F. and A. Chesher (1986). "Specification testing when score test statistics are identically zero", *Journal of Econometrics* 31, 121–149.

Ling, S. and McAleer, M. (2003). "Asymptotic theory for a vector Arma-Garch model", *Econometric Theory* 19, 280-310.

Longin, F. and Solnik, B. (2001). "Exreme correlation of international equity markets", *Journal of Finance* 56, 649-676.

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

Maruyama, Y. and Seo, T. (2003). "Estimation of moment parameter in elliptical distributions", *Journal of the Japan Statistical Society* 33, 215-229.

McCullough, B. D. and Vinod, H. D. (1999). "The numerical reliability of econometric software", *Journal of Economic Literature* 37, 633–665.

Mencía, J. and Sentana, E. (2010). "Distributional tests in multivariate dynamic models with normal and Student t innovations", CEMFI Working Paper 0804, revised June 2010, forthcoming in the *Review of Economics and Statistics*.

Neyman, J. and Scott, E.L. (1966). "On the use of optimal C_α-tests of composite hypotheses", *Bulletin of the International Statistical Institute* 41, 477–497.

Owen, J. and Rabinovitch R. (1983): "On the class of elliptical distributions and their applications to the theory of. portfolio choice", *Journal of Finance* 38, 745–752.

Table 1

Power properties of normality tests in finite samples

Panel A: 5% size-adjusted rejection rates

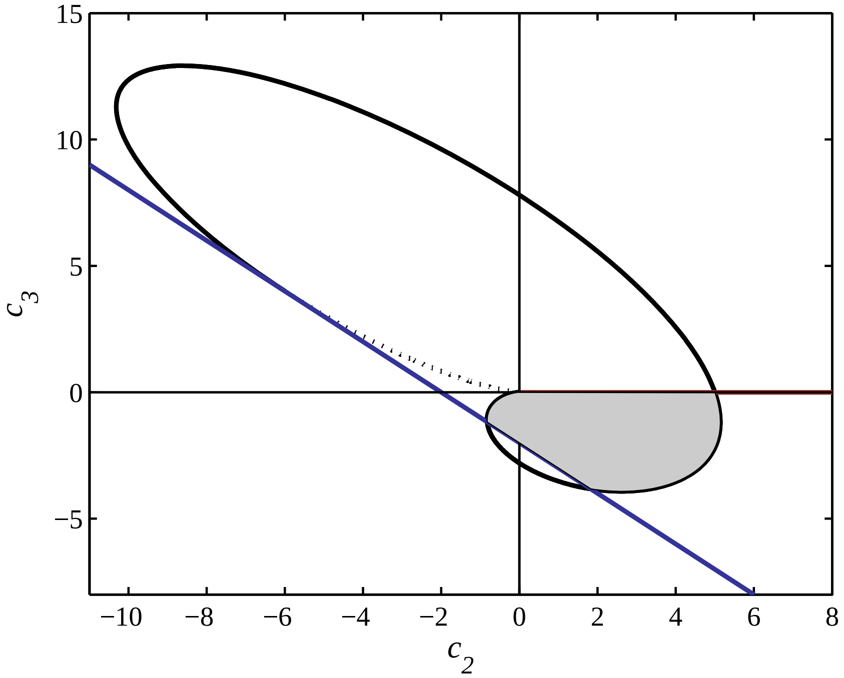
	DGP				
			DSMN		
Test	Student-t	Kotz	$\varkappa = 0.5$	$\varkappa = 0.25$	Expansion
Student-t KT	0.386	0.381	0.118	0.336	0.071
2^{nd} Laguerre	0.291	0.279	0.082	0.246	0.068
2^{nd} & 3^{rd} L.	0.269	0.295	0.092	0.324	0.121
2^{nd} & 3^{rd} KT	0.204	0.084	0.091	0.045	0.147
Kotz LM	0.198	0.413	0.060	0.576	0.008
DSMN $LM_{\varkappa=\varkappa_0}$	0.196	0.399	0.059	0.634	0.061

Panel B: 1% size-adjusted rejection rates

	DGP				
			DSMN		
Test	Student-t	Kotz	$\varkappa = 0.5$	$\varkappa = 0.25$	Expansion
Student-t KT	0.163	0.151	0.029	0.123	0.016
2^{nd} Laguerre	0.137	0.125	0.023	0.098	0.015
2^{nd} & 3^{rd} L.	0.123	0.111	0.035	0.118	0.053
2^{nd} & 3^{rd} KT	0.108	0.040	0.037	0.021	0.061
Kotz LM	0.070	0.206	0.014	0.343	0.010
DSMN $LM_{\varkappa=\varkappa_0}$	0.035	0.169	0.015	0.403	0.001

Notes: Sample: 10.000 replications with $T = 1,000$ and $N = 5$. Alternative hypotheses: Student t with 100 degrees of freedom, Kotz with the same kurtosis, discrete scale mixture of normals (DSMN) with the same kurtosis and $\alpha = 0.5$, and Polynomial expansion with zero excess kurtosis and $c_3 = -0.1$. Test statistics are described in section 3.

Figure 1: Positivity region of a 3^{rd} order polynomial expansion



Notes: For a given ς, the 3^{rd} order polynomial frontier that guarantees positivity must satisfy the following two equations $P_3(\varsigma) = 0$ and $\partial P_3(\varsigma)/\partial\varsigma = 0$ where $P_3(\varsigma)$ is defined in section 2. The first equation defines a straight line in (c_2, c_3) space such that in any neighbourhood of the solution we will find positive and negative densities. In contrast, the second equation guarantees that we remain on the frontier as we move in (c_2, c_3) space. Nevertheless, these conditions are overly restrictive because they do not take into account the non-negativity of ς (dashed line versus blue line). Finally, the red line represents the tangent of $P_3(0) = 0$. The grey area defines the admissible set in the (c_2, c_3) space.

Figure 2a: Densities of standardized univariate Discrete scale mixture of normals with $\varkappa = 0.05$ and α approaching 0

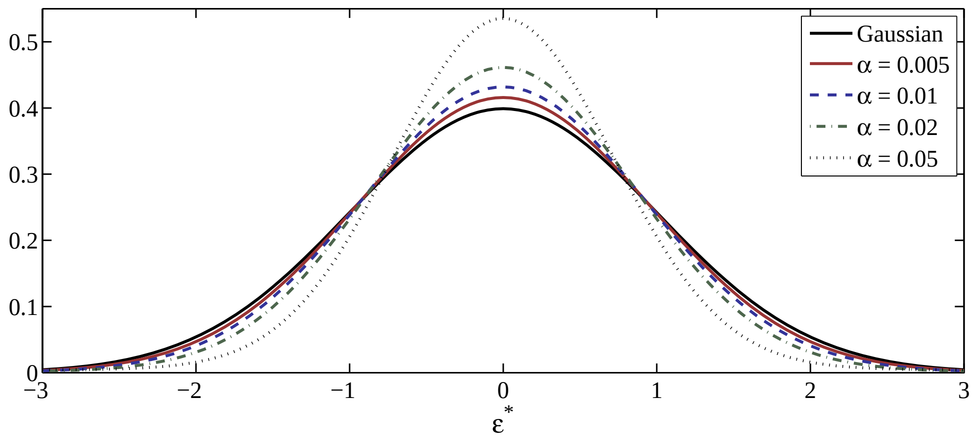


Figure 2b: Densities of standardized univariate Discrete scale mixture of normals with $\varkappa = 0.05$ and α approaching 1

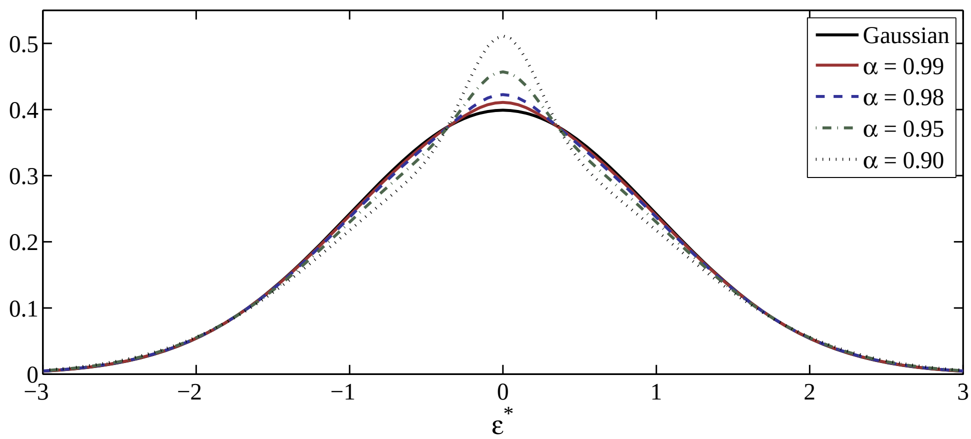


Figure 2c: Densities of standardized univariate Discrete scale mixture of normals with $\alpha = 0.5$ and $\varkappa$ approaching 1

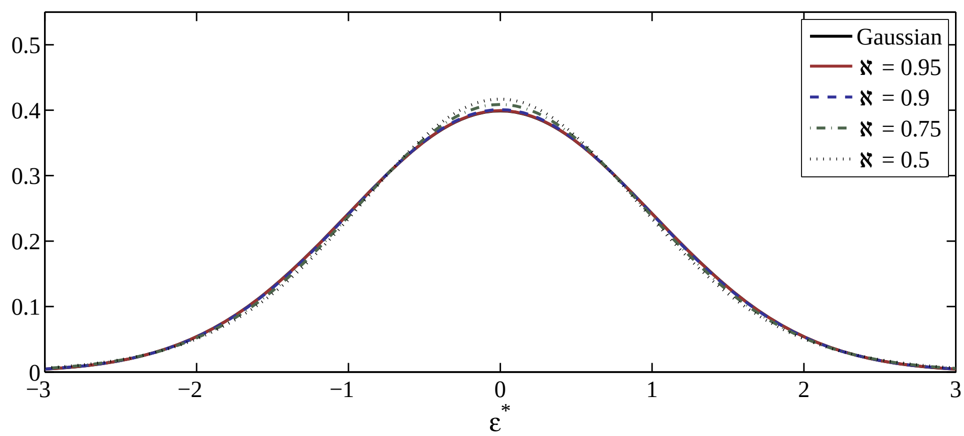


Figure 3a: Standardized bivariate normal density

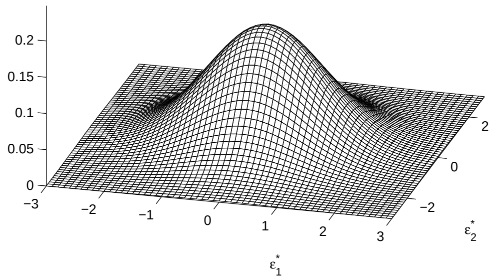


Figure 3b: Contours of a standardized bivariate normal density

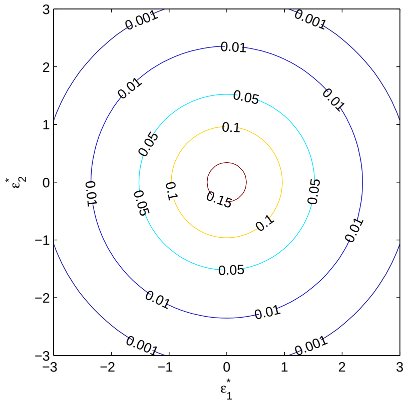


Figure 3c: Standardized bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

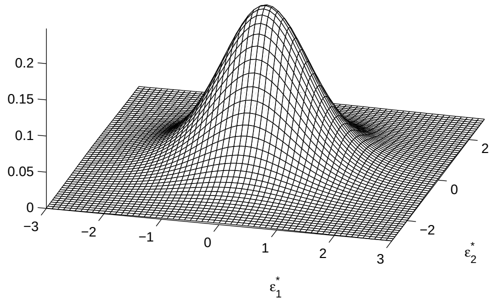


Figure 3d: Contours of a standardized bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

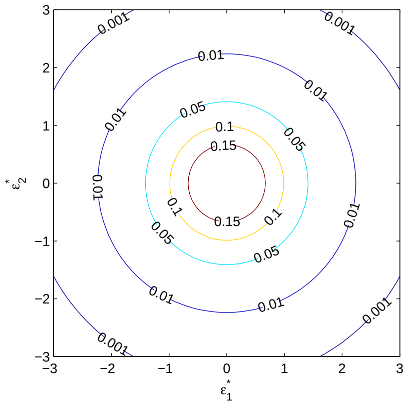


Figure 3e: Standardized bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

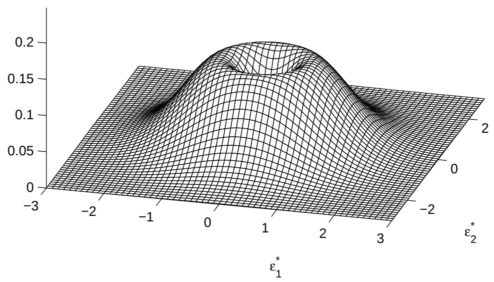


Figure 3f: Contours of a standardized bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

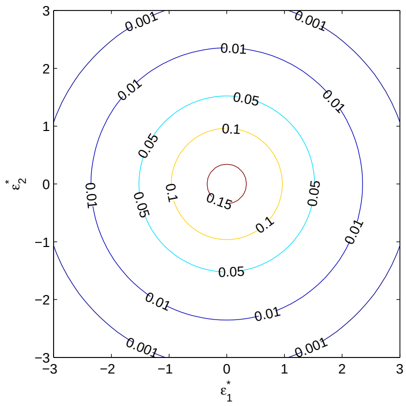


Figure 3g: Standardized bivariate Discrete scale mixture of normals density with multivariate excess kurtosis $\kappa = 0.125$ ($\pi = 0.5$)

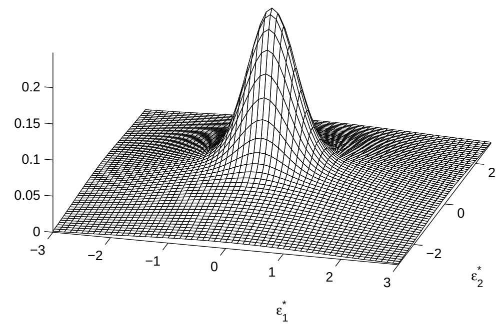


Figure 3h: Contours of a standardized bivariate Discrete scale mixture of normals density with multivariate excess kurtosis $\kappa = 0.125$ ($\pi = 0.5$)

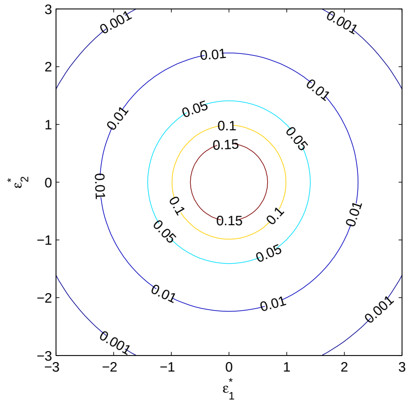


Figure 3i: Standardized bivariate 3^{rd} order polynomial expansion with parameters $c_2 = 0$ and $c_3 = -0.2$

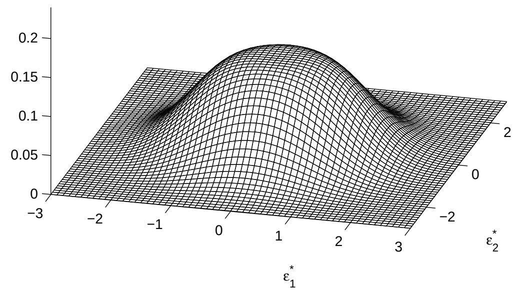


Figure 3j: Contours of a standardized 3^{rd} order polynomial expansion with parameters $c_2 = 0$ and $c_3 = -0.2$

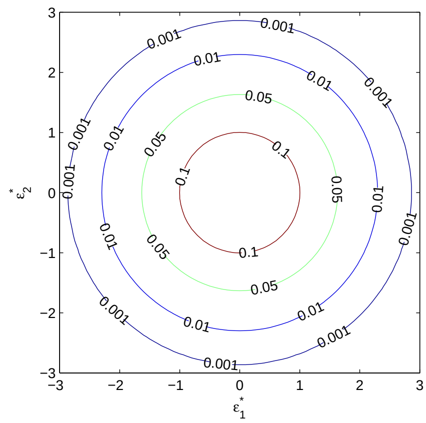
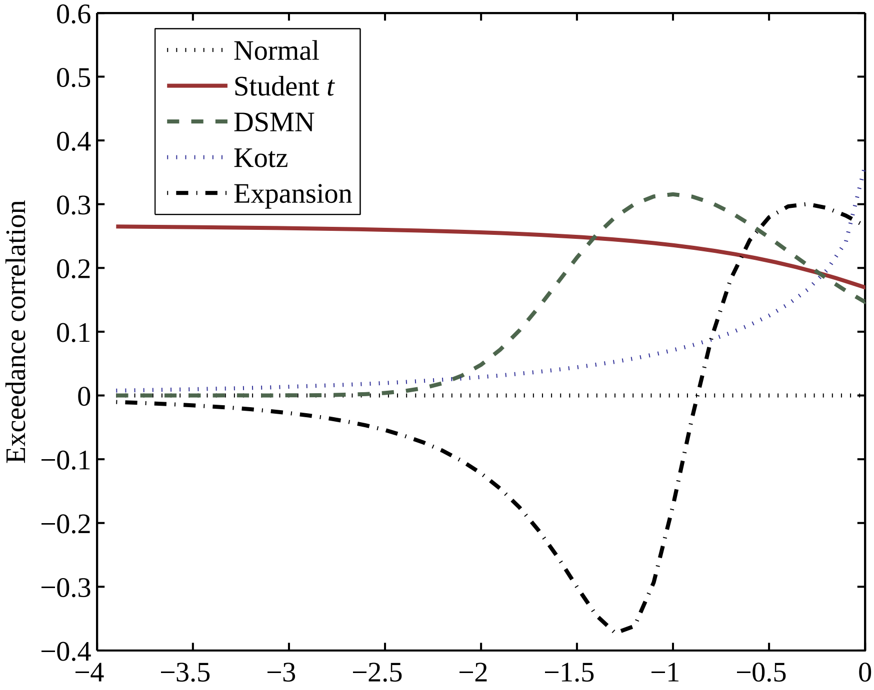
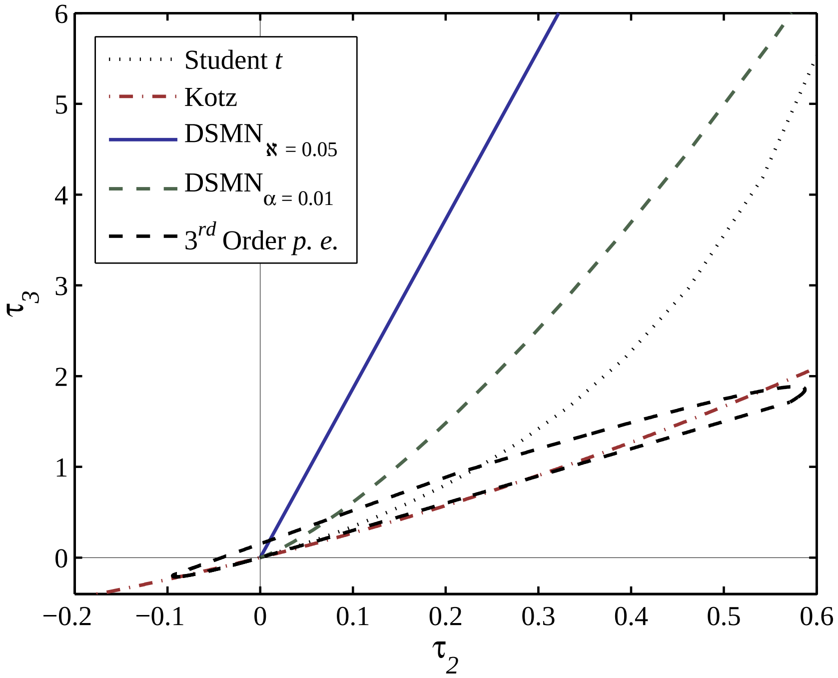


Figure 4: Exceedance correlation for alternative distributions



Notes: The exceedance correlation between two variables ε_1^* and ε_2^* is defined as $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \varrho, \varepsilon_2^* > \varrho)$ for positive ϱ and $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \varrho, \varepsilon_2^* > \varrho)$ for negative ϱ (see Longin and Solnik, 2001). Because all the distributions we consider are elliptical, we only report results for $\varrho < 0$. Student t distribution with 6 degrees of freedom, Kotz distribution with $\kappa = 1$, discrete scale mixture of normals with parameters $\alpha = 0.05$ and the same kurtosis and polynomial expansion with the same kurtosis and $c_3 = -0.5$.

Figure 5: Higher order moment parameters



Notes: The m-moment parameter of spherical random variables, $\tau_m(\boldsymbol{\eta})$, is defined as $E[\varsigma_t^m|\boldsymbol{\eta}] = [1 + \tau_m(\boldsymbol{\eta})]E[\varsigma_t^m|\mathbf{0}]$ where $E[\varsigma_t^m|\mathbf{0}]$ denotes the m-uncentered moment of a chi-squared random variables with N degrees of freedom. Analytical expressions for $\tau_m(\boldsymbol{\eta})$'s are provided in section 2.1.

Figure 6: Power of the normality tests

Figure 6a: Student t distribution

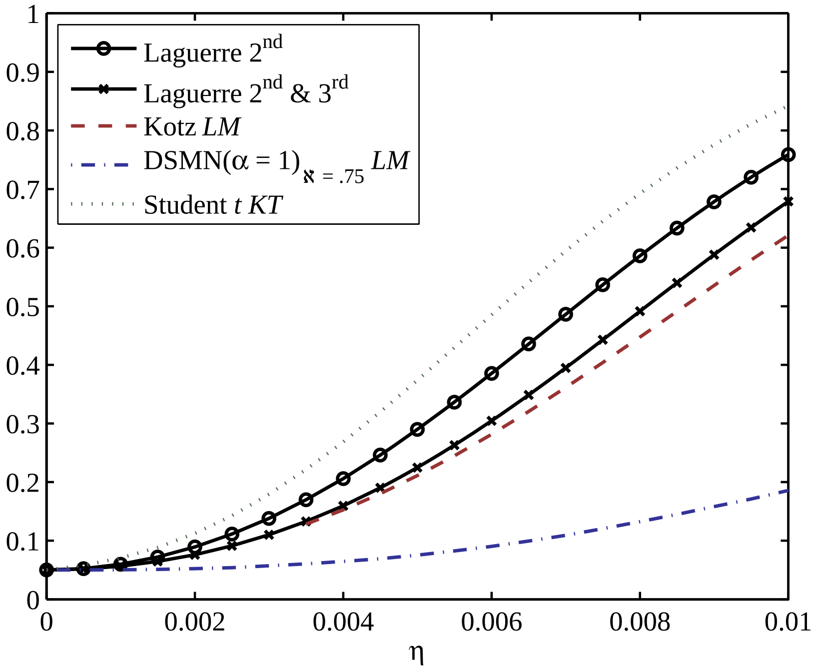
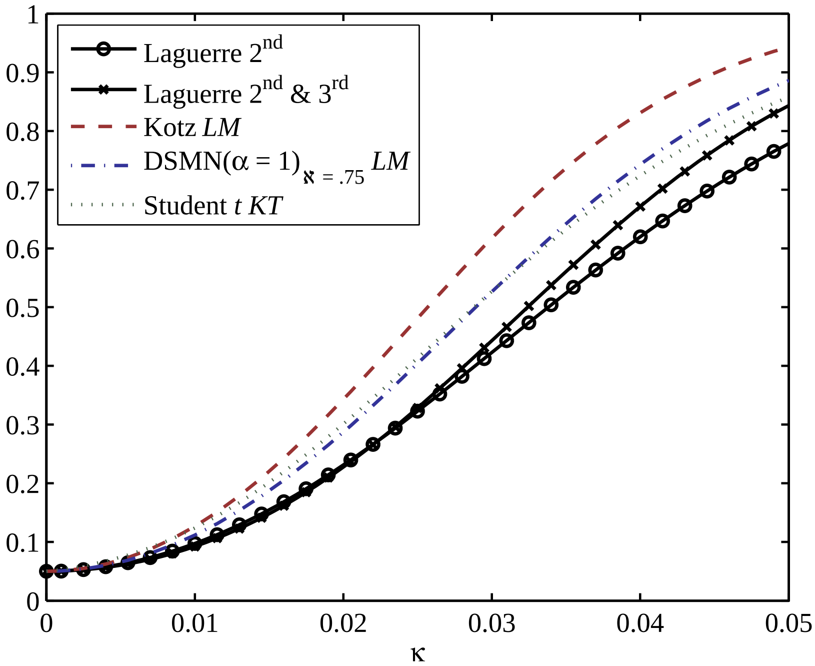


Figure 6b: Kotz distribution



Notes: Results at the 5% level. $T = 1,000$ and $N = 5$. For t innovations with ν degrees of freedom, $\eta = 1/\nu$. For Kotz innovations, κ denotes the coefficient of multivariate excess kurtosis. α is the mixing probability while $\varkappa$ is the variance ratio of the two components in the Discrete scale mixture of normals. For polynomial expansion innovations, c_2 and c_3 are the coefficients associated to the 2^{nd} and 3^{3d} Laguerre polynomials, respectively. Test statistics are described in section 3.

Figure 6: Power of the normality tests

Figure 6c: Discrete scale mixture of normals ($\varkappa = 0.75$)

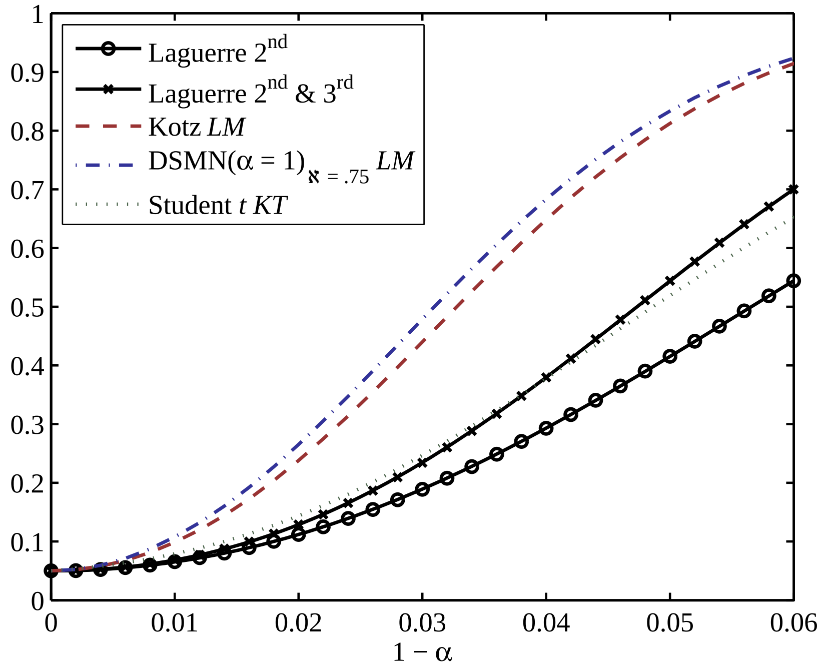
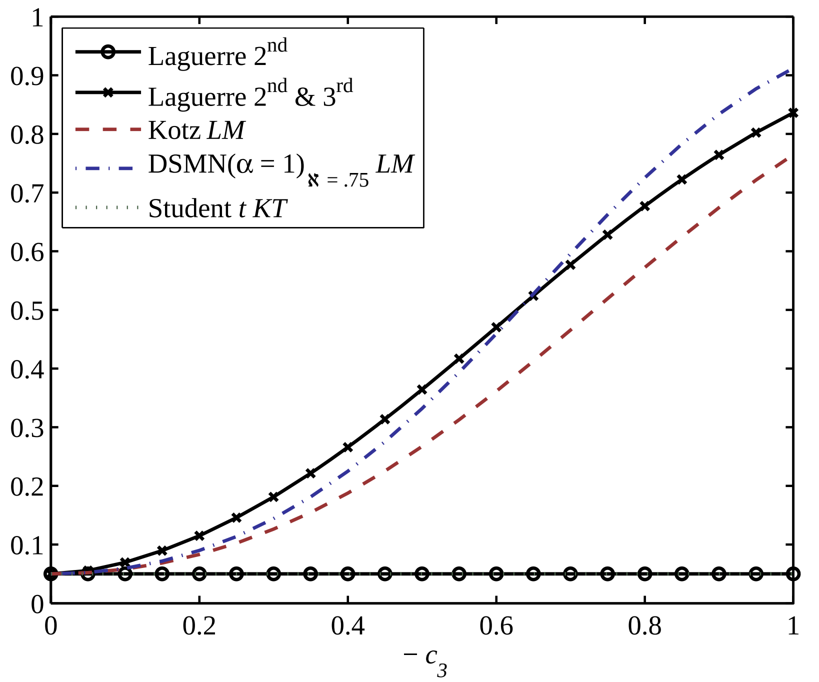
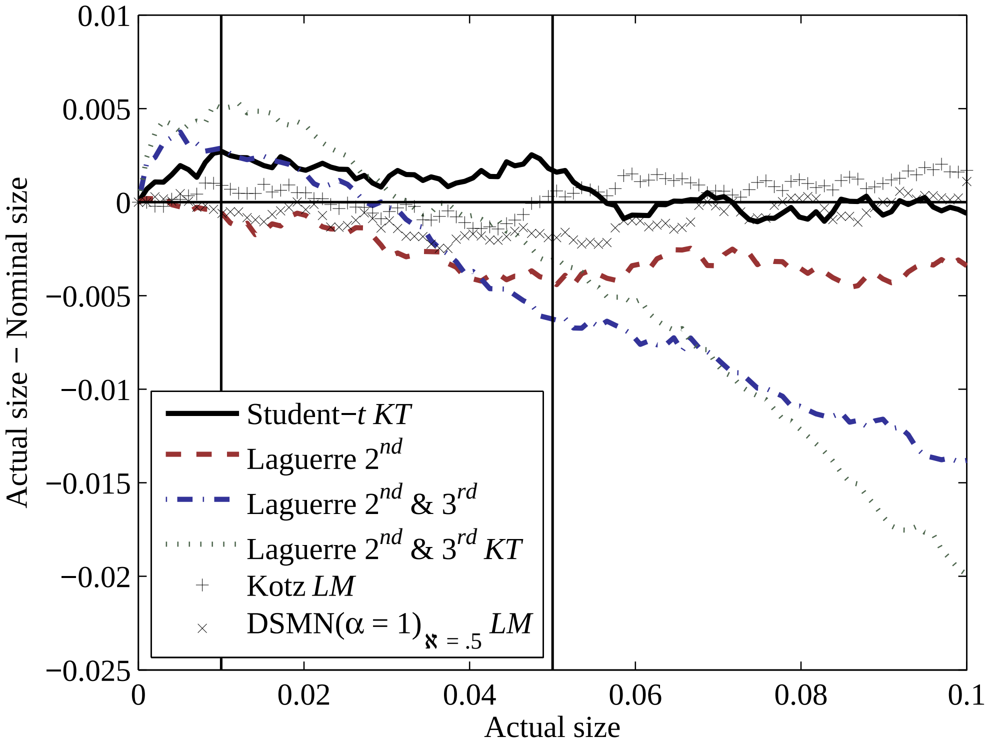


Figure 6d: 3^{rd} order polynomial expansion ($c_2 = 0$)



Notes: Results at the 5% level. $T = 1,000$ and $N = 5$. For t innovations with ν degrees of freedom, $\eta = 1/\nu$. For Kotz innovations, κ denotes the coefficient of multivariate excess kurtosis. α is the mixing probability while $\varkappa$ is the variance ratio of the two components in the Discrete scale mixture of normals. For polynomial expansion innovations, c_2 and c_3 are the coefficients associated to the 2^{nd} and 3^{3d} Laguerre polynomials, respectively. Test statistics are described in section 3.

Figure 7: p-value discrepancy plot



Notes: 10,000 replications. $T = 1000$, $N = 5$. Vertical lines denote the usual 1% and 5% levels. α is the mixing probability while $\varkappa$ is the variance ratio of the two components in the Discrete scale mixture of normals. Test statistics for the normality tests are described in section 3.