DISTRIBUTIONAL TESTS IN MULTIVARIATE DYNAMIC MODELS WITH NORMAL AND STUDENT *t* INNOVATIONS

Javier Mencía and Enrique Sentana

2009

Documentos de Trabajo
N.º 0929

BANCO DE ESPAÑA
Eurosistema

DISTRIBUTIONAL TESTS IN MULTIVARIATE DYNAMIC MODELS WITH NORMAL AND STUDENT *t* INNOVATIONS

DISTRIBUTIONAL TESTS IN MULTIVARIATE DYNAMIC MODELS WITH NORMAL AND STUDENT *t* INNOVATIONS (*)

Javier Mencía

BANCO DE ESPAÑA

Enrique Sentana

CEMFI

Abstract

We derive Lagrange Multiplier and Likelihood Ratio specification tests for the null hypotheses of multivariate normal and Student t innovations using the Generalised Hyperbolic distribution as our alternative hypothesis. We decompose the corresponding Lagrange Multiplier-type tests into skewness and kurtosis components, from which we obtain more powerful one-sided Kuhn-Tucker versions that are equivalent to the Likelihood Ratio test, whose asymptotic distribution we provide. We conduct detailed Monte Carlo exercises to study our proposed tests in finite samples. Finally, we present an empirical application to ten US sectoral stock returns, which indicates that their conditional distribution is mildly asymmetric and strongly leptokurtic.

Keywords: Bootstrap, Inequality Constraints, Kurtosis, Normality Tests, Skewness, Supremum Test, Underidentifed parameters.

JEL classification: C12, C52, C32.

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, as long as both moments are correctly specified. However, the normality assumption does not guarantee consistent estimators of other features of the conditional distribution, such as its quantiles. This is particularly true 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, especially in larger dimensions.

For most practical purposes, departures from normality can be attributed to two different sources: excess kurtosis and skewness. In this sense, Fiorentini, Sentana and Calzolari (2003) (FSC) discuss the use of the multivariate Student t distribution to model excess kurtosis. Despite its attractiveness, though, the multivariate Student t, which is a member of the elliptical family, rules out any potential asymmetries in the conditional distribution of asset returns. Such a shortcoming is more problematic than it may seem, because ML estimators based on incorrectly specified non-Gaussian distributions may lead to inconsistent parameter estimates (see Newey and Steigerwald, 1997; and Fiorentini and Sentana, 2007).

The main objective of our paper is to provide specification tests that assess the adequacy of the multivariate Gaussian and Student t distributional assumptions. There already exist some well known multivariate normality tests based on the skewness and kurtosis of the data, such as the one in Mardia (1970). This test was originally intended for models with homoskedastic disturbances and unrestricted covariance matrices. In general dynamic models, though, it may suffer from asymptotic size distortions (see Fiorentini, Sentana, and Calzolari, 2004; Bontemps and Meddahi, 2005). In addition, the number of moment conditions of the skewness component of Mardia's test is of order N^3, where N is the multivariate dimension. Hence, this test may show further size distortions and low power when the cross-sectional dimension is relatively large. In this paper, we avoid the curse of dimensionality by considering a family of distributions that allow for

both excess kurtosis and asymmetries in the innovations, which at the same time nest the multivariate normal and Student t. Specifically, we will use the rather flexible Generalised Hyperbolic *(GH)* distribution introduced by Barndorff-Nielsen (1977), which nests other well known cases as well, such as the Hyperbolic, the Normal Inverse Gaussian, the Normal Gamma associated to the Variance Gamma process, the Multivariate Laplace and their asymmetric generalisations, and whose empirical relevance has already been widely documented in the literature (see e.g. Madan and Milne, 1991; Chen, Härdle, and Jeong, 2004; Aas, Dimakos, and Haff, 2005; or Cajigas and Urga, 2007). Therefore, we focus on those departures from both normal and Student t distributions that seem to be relevant from an empirical point of view.

Our approach is related to Bera and Premaratne (2002), who also nest the Student t by using Pearson's type IV distribution in univariate static models. However, they do not explain how to extend their approach to multivariate contexts, nor do they consider dynamic models explicitly. Our choice also differs from Bauwens and Laurent (2005), who introduce skewness by "stretching" the multivariate Student t distribution differently in different orthants. However, the implementation of their technique becomes increasingly difficult in large dimensions, as the number of orthants is 2^N. Similarly, semi-parametric procedures, including Hermite polynomial expansions, become infeasible for moderately large N, unless one maintains the assumption of elliptical symmetry, and the same is true of copulae methods. An alternative approach is followed by Bai (2003), who tests parametric conditional univariate distributions by coupling the Kolmogorov test with Khmaladze's transformation. Unfortunately, its multivariate extension in Bai and Zhihong (2008), which is not invariant to the chosen orthogonalisation, is difficult to implement for N greater than 2 when the shape parameters are unknown.

In contrast, given that the *GH* distribution can be understood as a location-scale mixture of a multivariate Gaussian vector with a positive mixing variable that follows a Generalised Inverse Gaussian *(GIG)* distribution (see Jørgensen, 1982, and Johnson, Kotz, and Balakrishnan, 1994 for details), the number of additional parameters that we have to introduce simply grows linearly with the cross-sectional dimension. In addition, the mixture of normals interpretation also makes the *GH* distribution analytically rather tractable, as illustrated by Blæsild (1981). This mixture interpretation has important implications from an asset allocation point of view too, because it implies that the

distribution of the returns to any portfolio will exclusively depend on its first three moments, thereby replacing the traditional mean-variance paradigm by mean-variance-skewness analysis (see Mencía and Sentana, 2008).

In this framework, we obtain closed form expressions for the score tests and show the asymptotic equivalence of their one-sided Kuhn-Tucker versions to the likelihood ratio tests. We use this equivalence to derive the common asymptotic distribution of the likelihood ratio and Kuhn-Tucker tests, which turns out to be standard despite the non-standard features of the problem. We also study the finite sample properties of our proposed tests with an extensive Monte Carlo analysis. Finally, we employ our proposed tests to assess the empirical fit of the Gaussian an Student t distributions in an empirical application to ten US sectoral stock returns that takes into account time-variation in volatilities and correlations. We consider both daily and weekly data to see whether the results depend on the frequency of the observations.

The rest of the paper is organised as follows. Section 2 describes the econometric model and the GH distribution. We derive the normality tests in section 3, and the Student t tests in section 4. Section 5 presents the results of our Monte Carlo experiments. Finally, we include our empirical application in section 6, followed by our conclusions. Proofs and auxiliary results can be found in appendices.

2 The dynamic econometric model and the alternative hypothesis

Discrete time models for financial time series are usually characterised by an explicit dynamic regression model with time-varying variances and covariances. Typically, the N dependent variables in $\mathbf{y}_t$ are assumed to be generated as

$$\left.\begin{array}{l} \mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*, \\ \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}\left(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}\right), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\left(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}\right), \end{array}\right\} \tag{1}$$

where $\boldsymbol{\mu}()$ and $vech\left[\boldsymbol{\Sigma}()\right]$ are N and $N(N+1)/2$-dimensional vectors of 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 $N \times N$ "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 vector martingale difference sequence satisfying

$E(\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$. As a consequence, $E(\mathbf{y}_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and $V(\mathbf{y}_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$.

In practice, the multivariate Gaussian and Student t have been the two most popular choices to model the distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$. For the purposes of conducting specification tests of those two distributions, we postulate that under the alternative $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a GH random vector, which nests both Normal and Student t as particular cases. In addition, it also includes other well known and empirically relevant special cases, such as symmetric and asymmetric versions of the Hyperbolic (Chen, Härdle, and Jeong, 2004), Normal Gamma (Madan and Milne, 1991), Normal Inverse Gaussian (Aas, Dimakos, and Haff, 2005) and Laplace distributions (Cajigas and Urga, 2007).[1]

We can gain some intuition about the parameters of the GH distribution by considering its interpretation as a location-scale mixture of normals. If $\boldsymbol{\varepsilon}_t^*$ is a GH vector, then it can be expressed as

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi_t^{-1} + \xi_t^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\mathbf{r}_t, \qquad (2)$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^N$, $\boldsymbol{\Upsilon}$ is a positive definite matrix of order N and $\mathbf{r}_t \sim iid\, N(\mathbf{0}, \mathbf{I}_N)$. The positive mixing variable ξ_t is an independent $iid\, GIG$ with parameters $-\nu$, γ and δ, or $\xi_t \sim GIG(-\nu, \gamma, \delta)$ for short, where $\nu \in \mathbb{R}$, $\delta, \gamma \in \mathbb{R}^+$ (see Appendix B.4). Since $\boldsymbol{\varepsilon}_t^*$ given ξ_t is Gaussian with conditional mean $\boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi_t^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi_t^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. There is a further scale parameter, δ; two other scalars, ν and γ, to allow for flexible tail modelling; and the vector $\boldsymbol{\beta}$, which introduces skewness in this distribution.

Like any mixture of normals, though, the GH distribution does not allow for thinner tails than the normal. Nevertheless, financial returns are typically leptokurtic in practice, as section 6 confirms.

In order to ensure that the elements of $\boldsymbol{\varepsilon}_t^*$ are uncorrelated with zero mean and unit variance by construction, we consider a standardised version. Specifically, we set $\delta = 1$, $\boldsymbol{\alpha} = -c(\boldsymbol{\beta}, \nu, \gamma)\boldsymbol{\beta}$ and

$$\boldsymbol{\Upsilon} = \frac{\gamma}{R_\nu(\gamma)}\left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right], \qquad (3)$$

---

[1]See Appendix B.1 for further details.

where

$$c\left(\boldsymbol{\beta},\nu,\gamma\right)=\frac{-1+\sqrt{1+4[D_{\nu+1}\left(\gamma\right)-1]\boldsymbol{\beta}'\boldsymbol{\beta}}}{2[D_{\nu+1}\left(\gamma\right)-1]\boldsymbol{\beta}'\boldsymbol{\beta}}, \tag{4}$$

$R_{\nu}\left(\gamma\right)=K_{\nu+1}\left(\gamma\right)/K_{\nu}\left(\gamma\right)$, $D_{\nu+1}\left(\gamma\right)=K_{\nu+2}\left(\gamma\right)K_{\nu}\left(\gamma\right)/K_{\nu+1}^{2}\left(\gamma\right)$ and $K_{\nu}\left(\cdot\right)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun, 1965, p. 374, as well as Appendix B.3). Thus, the distribution of $\boldsymbol{\varepsilon}_{t}^{*}$ depends on two shape parameters, ν and γ, and a vector of N skewness parameters, denoted by $\boldsymbol{\beta}$. Under this parametrisation, the Normal distribution can be achieved in three different ways: (i) when $\nu\rightarrow-\infty$ or (ii) $\nu\rightarrow+\infty$, regardless of the values of γ and $\boldsymbol{\beta}$; and (iii) when $\gamma\rightarrow\infty$ irrespective of the values of ν and $\boldsymbol{\beta}$. Analogously, the Student t is obtained when $-\infty<\nu<-2$, $\gamma=0$ and $\boldsymbol{\beta}=\mathbf{0}$.

Importantly, given that $\boldsymbol{\varepsilon}_{t}^{*}$ is not generally observable, the choice of "square root" matrix is not irrelevant except in univariate GH models, or in multivariate GH models in which either $\boldsymbol{\Sigma}_{t}(\boldsymbol{\theta})$ is time-invariant or $\boldsymbol{\varepsilon}_{t}^{*}$ is spherical (i.e. $\boldsymbol{\beta}=\mathbf{0}$). But, if we parametrise $\boldsymbol{\beta}$ as a function of past information and a new vector of parameters $\mathbf{b}$ in the following way:

$$\boldsymbol{\beta}_{t}(\boldsymbol{\theta},\mathbf{b})=\boldsymbol{\Sigma}_{t}^{\frac{1}{2}\prime}(\boldsymbol{\theta})\mathbf{b}, \tag{5}$$

then it is straightforward to see that the resulting distribution of $\mathbf{y}_{t}$ conditional on I_{t-1} will not depend on the choice of $\boldsymbol{\Sigma}_{t}^{\frac{1}{2}}(\boldsymbol{\theta})$.[2] Finally, it is analytically convenient to replace ν and γ by η and ψ, where $\eta=-.5\nu^{-1}$ and $\psi=(1+\gamma)^{-1}$, although we continue to use ν and γ in some equations for notational simplicity.[3]

3 Multivariate normality versus GH innovations

3.1 The score under Gaussianity

Let $s_{t}'(\boldsymbol{\phi})=[s_{\boldsymbol{\theta}t}'(\boldsymbol{\phi}),s_{\eta t}(\boldsymbol{\phi}),s_{\psi t}(\boldsymbol{\phi}),s_{\mathbf{b}t}'(\boldsymbol{\phi})]$ denote the score vector of the GH log-likelihood function, where $\boldsymbol{\phi}'=(\boldsymbol{\theta}',\eta,\psi,\mathbf{b}')$ (see Appendix B.2 for explicit expressions). As we mentioned before, we can achieve normality in three different ways: (i) when $\eta\rightarrow 0^{+}$ or (ii) $\eta\rightarrow 0^{-}$ regardless of the values of $\mathbf{b}$ and ψ; and (iii) when $\psi\rightarrow 0^{+}$, irrespective of η and $\mathbf{b}$. Therefore, it is not surprising that the Gaussian scores with respect to η

---

[2]Nevertheless, it would be fairly easy to adapt all our subsequent expressions to the alternative assumption that $\boldsymbol{\beta}_{t}(\boldsymbol{\theta},\mathbf{b})=\mathbf{b}\ \forall t$ (see Mencía, 2003).

[3]An undesirable aspect of this reparametrisation is that the log-likelihood is continuous but non-differentiable with respect to η at $\eta=0$, even though it is continuous and differentiable with respect to ν for all values of ν. The problem is that at $\eta=0$, we are pasting together the extremes $\nu\rightarrow\pm\infty$ into a single point. Nevertheless, it is still worth working with η instead of ν when testing for normality.

or ψ are 0 when these parameters are not identified, and also, that $\lim_{\eta \cdot \psi \to 0} s_{\mathbf{b}t}(\boldsymbol{\phi}) = \mathbf{0}$. Similarly, the limit of the score with respect to the mean and variance parameters, $\lim_{\eta \cdot \psi \to 0} s_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$, coincides with the usual Gaussian expressions (see e.g. Bollerslev and Wooldridge (1992)). Further, we can show that for fixed $\psi > 0$,

$$
\begin{aligned}
\lim_{\eta \to 0^+} s_{\eta t}(\boldsymbol{\phi}) &= -\lim_{\eta \to 0^-} s_{\eta t}(\boldsymbol{\phi}) = \left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4} \right] \\
&\quad + \mathbf{b}' \left\{ \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) \left[\varsigma_t(\boldsymbol{\theta}) - (N+2) \right] \right\},
\end{aligned}
\tag{6}
$$

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, which confirms the non-differentiability of the log-likelihood function with respect to η at $\eta = 0$ (see footnote 3). Finally, we can show that for $\eta \neq 0$, $\lim_{\psi \to 0^+} s_{\psi t}(\boldsymbol{\phi})$ is exactly one half of (6).

3.2   The conditional information matrix under Gaussianity

Again, we must study separately the three possible ways to achieve normality. First, consider the conditional information matrix $\mathcal{I}_t(\boldsymbol{\phi})$ when $\eta \to 0^+$,

$$
\begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) & \mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) \\ \mathcal{I}'_{\boldsymbol{\theta}\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) & \mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) \end{bmatrix} = \lim_{\eta \to 0^+} V \left[\begin{matrix} s_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \eta, \psi, \mathbf{b}) \\ s_{\eta t}(\boldsymbol{\theta}, \eta, \psi, \mathbf{b}) \end{matrix} \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right],
$$

where we have excluded the terms corresponding to $\mathbf{b}$ and ψ because both $s_{\mathbf{b}t}(\boldsymbol{\phi})$ and $s_{\psi t}(\boldsymbol{\phi})$ are 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). Moreover, we can show that

Proposition 1 *The conditional information matrix of the GH distribution when $\eta \to 0^+$ is characterised by $\mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) = \mathbf{0}$ and $\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) = (N+2)[.5N + \mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}]$, so that $E[\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b})] = (N+2)[.5N + \mathbf{b}'\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{b}]$ where $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = E[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]$ denotes the unconditional covariance matrix of the data.*

Not surprisingly, these expressions reduce to the ones in FSC for $\mathbf{b} = \mathbf{0}$.

Similarly, when $\eta \to 0^-$ we will have exactly the same conditional information matrix because $\lim_{\eta \to 0^-} s_{\eta t}(\boldsymbol{\theta}, \eta, \psi, \mathbf{b}) = -\lim_{\eta \to 0^+} s_{\eta t}(\boldsymbol{\theta}, \eta, \psi, \mathbf{b})$, as we saw before.

Finally, when $\psi \to 0^+$, we must exclude $s_{\mathbf{b}t}(\boldsymbol{\phi})$ and $s_{\eta t}(\boldsymbol{\phi})$ from the computation of the information matrix for the same reasons as above. However, due to the proportionality of the scores with respect to η and ψ under normality, it is trivial to see that $\mathcal{I}_{\boldsymbol{\theta}\psi t}(\boldsymbol{\theta}, \eta, 0, \mathbf{b}) = \mathbf{0}$, and that $\mathcal{I}_{\psi\psi t}(\boldsymbol{\theta}, \eta, 0^+, \mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^-, \psi, \mathbf{b})$.

Importantly, once we estimate the mean and variance parameters $\boldsymbol{\theta}$, we can use the previous closed form expressions to evaluate the information matrix without resorting to either the outer product of the score or the Hessian matrix.

Let $\tilde{\boldsymbol{\theta}}_T$ denote the ML estimator of $\boldsymbol{\theta}$ obtained by maximising the Gaussian log-likelihood function. Since our normality tests will require the root T consistency of this estimator, we will rely on the following result.

Proposition 2 *Let $\tilde{\boldsymbol{\theta}}_T$ denote the Gaussian ML estimator of $\boldsymbol{\theta}$. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\theta}_0$ is iid $N(\mathbf{0}, \mathbf{I}_N)$ and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N(\mathbf{0}, \mathcal{I}_{\boldsymbol{\theta\theta}}^{-1})$, where $\mathcal{I}_{\boldsymbol{\theta\theta}}$ is the unconditional information matrix under normality.*

We use such high level regularity conditions because we want to leave unspecified the conditional mean vector and covariance matrix in order to maintain full generality. Primitive conditions for specific multivariate models can be found for instance in Ling and McAleer (2003).

3.3   Tests for fixed values of the underidentified parameters

The derivation of the Lagrange multiplier (LM) and Likelihood Ratio (LR) tests for multivariate normality versus GH innovations is complicated by two unusual features. First, since the GH distribution can approach the normal distribution along three different paths in the parameter space, i.e. $\eta \to 0^+$, $\eta \to 0^-$ or $\psi \to 0^+$, the null hypothesis can be posed in three different ways. In addition, some of the other parameters become increasingly underidentified along each of those three paths. In particular, η and $\mathbf{b}$ are not identified in the limit when $\psi \to 0^+$, while ψ and $\mathbf{b}$ are underidentified when $\eta \to 0^\pm$. Unfortunately, the reparametrisation of η and ψ in terms of either hyperbolic or polar coordinates, as suggested by King and Shively (1993), does not reduce the multiplicity of testing directions in our case.[4]

One standard solution in the literature to deal with testing situations with under-identified parameters under the null involves fixing the underidentified parameters to some arbitrary values, and then computing the appropriate test statistic for those given values.

---

[4]Under hyperbolic coordinates, $a_0 = \eta\psi$, $a_1 = -.5\log(\psi/\eta)$ for $\eta > 0$ and $a_1 = -.5\log(-\psi/\eta)$ for $\eta < 0$, which would yield normality for $a_0 = 0$ or $a_1 \to \pm\infty$ for the two signs of η. With polar coordinates, $\eta = b_0\cos(b_1)$ and $\psi = b_0\sin(b_1)$, which yield normality for $b_0 \to 0$, $b_1 \to 0$, $b_1 \to \pi/2$ or $b_1 \to \pi$.

For the case in which normality is achieved as $\eta \to 0^+$, we can use the results in sections 3.1 and 3.2 to show that for given values of ψ and $\mathbf{b}$, the LM test will be the usual quadratic form in the sample averages of the scores corresponding to $\boldsymbol{\theta}$ and η, $\bar{s}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b})$ and $\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b})$, with weighting matrix the inverse of the unconditional information matrix, which can be obtained as the unconditional expected value of the conditional information matrix in Proposition 1. But since $\bar{s}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b}) = \mathbf{0}$ by definition of $\tilde{\boldsymbol{\theta}}_T$, and $\mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\theta}_0, 0^+, \psi, \mathbf{b}) = \mathbf{0}$, we can show that

$$
LM_1(\tilde{\boldsymbol{\theta}}_T, \psi, \mathbf{b}) = \frac{\left[\sqrt{T}\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b})\right]^2}{E[\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}_0, 0^+, \psi, \mathbf{b})]}.
$$

We can operate analogously for the other two limits, thereby obtaining the test statistic $LM_2(\tilde{\boldsymbol{\theta}}_T, \psi, \mathbf{b})$ for the null $\eta \to 0^-$, and $LM_3(\tilde{\boldsymbol{\theta}}_T, \eta, \mathbf{b})$ for $\psi \to 0^+$. Somewhat remarkably, all these test statistics share the same formula, which only depends on $\mathbf{b}$.

Proposition 3

1. *The LM Normality tests for fixed values of the underidentified parameters and known $\boldsymbol{\theta}_0$ can be expressed as:*

$$
LM_1(\boldsymbol{\theta}_0, \psi, \mathbf{b}) = LM_2(\boldsymbol{\theta}_0, \psi, \mathbf{b}) = LM_3(\boldsymbol{\theta}_0, \eta, \mathbf{b}) = LM(\boldsymbol{\theta}_0, \mathbf{b})
$$
$$
= (N+2)^{-1}\left(\frac{N}{2} + 2\mathbf{b}'\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\mathbf{b}\right)^{-1}\left\{\frac{\sqrt{T}}{T}\sum_t\left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{N(N+2)}{4}\right]\right.
$$
$$
\left. + \mathbf{b}'\frac{\sqrt{T}}{T}\sum_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)]\right\}^2, \tag{7}
$$

    which converges in distribution to a chi-square with one degree of freedom for a given $\mathbf{b}$ under the null hypothesis of normality.

2. *If in addition the regularity conditions of Proposition 2 hold, then the above results will remain true if we substitute $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.*

The fact that we obtain the same test regardless of the path that we follow to obtain normality is worth remarking, as this feature is not shared by tests of normality vs. a discrete mixture of normals (see Cho and White, 2007). The rationale is that the null hypothesis of normality effectively imposes the single restriction $\eta \cdot \psi = 0$ on the parameter space. Importantly, note that (7) is numerically invariant to the chosen factorisation of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, as expected from (5).

Perhaps not surprisingly, we can prove the following result for the corresponding LR test:

Proposition 4

1. *Under the null of normality and sequences of local alternatives, the LR Normality tests for fixed values of the unidentified parameters $\mathbf{b}$ is asymptotically equivalent to the Kuhn-Tucker (KT) test*

$$KT(\boldsymbol{\theta}_0, \mathbf{b}) = \mathbf{1}(\bar{s}_{\eta T}(\boldsymbol{\theta}_0, 0, \mathbf{b}) \geq 0) \cdot LM(\boldsymbol{\theta}_0, \mathbf{b}), \tag{8}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

2. *In addition, if the regularity conditions of Proposition 2 hold, then the above results will remain true if we substitute $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.*

But since in large samples $\mathbf{1}(\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0, \mathbf{b}) \geq 0)$ will be 0 approximately half the time under the null, the common asymptotic distribution of the LR and KT tests will be a 50:50 mixture of 0 and a chi-square with one degree of freedom. Once again, note that the single degree of freedom reflects the fact that normality effectively imposes the restriction $\eta \cdot \psi = 0$. This is confirmed by the fact that the log-likelihood contours are parallel to the axes in η, ψ space for values of η or ψ close to 0.

3.4 The supremum tests

The approach described in the previous subsection is plausible in situations where there are values of the underidentified parameters that make sense from an economic or statistical point of view. Unfortunately, it is not at all clear a priori what values of $\mathbf{b}$ are likely to prevail under the alternative of GH innovations. For that reason, in this subsection we follow a second approach, which consists in computing either the LR or the LM test statistic for the whole range of values of the underidentified parameters, which are then combined to construct an overall test statistic (see Andrews, 1994). In our case, we compute these tests for all possible values of $\mathbf{b}$ for each of the three testing directions, and then take the supremum over those parameter values.

Let us start with the LM test. It turns out that we can maximise $LM(\tilde{\boldsymbol{\theta}}_T, \mathbf{b})$ with respect to $\mathbf{b}$ in closed form, and also obtain the asymptotic distribution of the resulting test statistic:

Proposition 5

1. *The supremum of the LM Normality test (7) with respect to $\mathbf{b}$ can be expressed as*

$$\sup_{\mathbf{b} \in \mathbb{R}^N} LM(\boldsymbol{\theta}_0) = LM_k(\boldsymbol{\theta}_0) + LM_s(\boldsymbol{\theta}_0), \tag{9}$$

$$LM_k(\boldsymbol{\theta}_0) = \frac{2}{N(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{N(N+2)}{4} \right] \right\}^2,$$

(10)

$$LM_s(\boldsymbol{\theta}_0) = \frac{1}{2(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0) \left[\varsigma_t(\boldsymbol{\theta}_0) - (N+2) \right] \right\}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$$

$$\times \left\{ \frac{\sqrt{T}}{T} \sum_t \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0) \left[\varsigma_t(\boldsymbol{\theta}_0) - (N+2) \right] \right\},$$

(11)

which converges in distribution to a chi-square random variable with $N+1$ degrees of freedom under the null hypothesis of normality.

2. In addition, if the regularity conditions of Proposition 2 hold, then the above results will remain true if we substitute $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.

The first component of the sup LM test, i.e. $LM_k(\tilde{\boldsymbol{\theta}}_T)$, is numerically identical to the LM statistic derived by FSC to test multivariate normal versus Student t innovations. These authors reinterpret (10) as a specification test of the restriction on the first two moments of $\varsigma_t(\boldsymbol{\theta}_0)$ implicit in

$$E \left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) \right] = E[m_{kt}(\boldsymbol{\theta}_0)] = 0,$$

(12)

and show that it numerically coincides with the kurtosis component of Mardia's (1970) test for multivariate normality in the models he considered (see below). Hereinafter, we shall refer to $LM_k(\tilde{\boldsymbol{\theta}}_T)$ as the kurtosis component of our multivariate normality test.

In contrast, the second component of the sup LM test, $LM_s(\tilde{\boldsymbol{\theta}}_T)$, arises because we also allow for skewness under the alternative hypothesis. This symmetry component is asymptotically equivalent under the null and sequences of local alternatives to T times the uncentred R^2 from either a multivariate regression of $\boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)$ on $\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - (N+2)$ (Hessian version), or a univariate regression of 1 on $\left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - (N+2) \right]\boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)$ (Outer product version). Nevertheless, we would expect a priori that $LM_s(\tilde{\boldsymbol{\theta}}_T)$ would be the version of the LM test with the smallest size distortions (see Davidson and MacKinnon, 1983).

As we discussed in Section 2, the class of GH distributions can only accommodate fatter tails than the normal. In terms of the kurtosis component of our sup LM multivariate normality test, this implies that as we depart from normality, we will have

$$E \left[m_{kt}(\boldsymbol{\theta}_0) \,|\, \boldsymbol{\theta}_0, \eta_0 > 0, \psi_0 > 0 \right] > 0.$$

(13)

While a (sup) LR test will take this feature into account by construction in maximising the *GH* log-likelihood function, we need to modify the sup LM test if we want to reflect the one sided nature of its kurtosis component, as FSC do in the case of the Student t. For that reason, we would recommend a KT multiplier version of the sup LM test that exploits (13) in order to increase its power and make it asymptotically equivalent to the (sup) LR test (see also Hansen, 1991 and Andrews, 2001). More formally:

Proposition 6

1. *The (sup) LR test of Gaussian vs. GH innovations is asymptotically equivalent under the null of normality and sequences of local alternatives to the following (sup) Kuhn-Tucker test:*

$$KT(\boldsymbol{\theta}_0) = LM_k(\boldsymbol{\theta}_0)\mathbf{1}\big(\bar{m}_{kT}(\boldsymbol{\theta}_0) > 0\big) + LM_s(\boldsymbol{\theta}_0), \qquad (14)$$

   where $\mathbf{1}(\cdot)$ is the indicator function, and $\bar{m}_{kT}(\boldsymbol{\theta}_0)$ the sample mean of $m_{kt}(\boldsymbol{\theta}_0)$.

2. *If the regularity conditions of Proposition 2 hold, then the above results will remain true if we substitute $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.*

Asymptotically, the probability that $\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T)$ becomes negative is .5 under the null. Consequently, $KT(\tilde{\boldsymbol{\theta}}_T)$ and the (sup) LR test will be distributed as a 50:50 mixture of chi-squares with N and $N+1$ degrees of freedom because the information matrix is block diagonal under normality. In practice, the LR test is computationally more burdensome. Given that the underidentifiability of η, ψ and $\mathbf{b}$ under the null implies that the *GH* log-likelihood function is numerically rather flat in the neighbourhood of the normality region, in principle we would need to estimate the model under the alternative hypothesis starting from a dense grid of values for those $N + 2$ parameters. In practice, however, it will not be possible to consider a grid of values for $\mathbf{b}$ even in small cross-sectional dimensions. In this sense, the main advantage of the sup KT test is that it only requires the estimation of the model under the null hypothesis. In any case, we can use the expression $\Pr(X > c) = 1 - .5F_{\chi_N^2}(c) - .5F_{\chi_{N+1}^2}(c)$ to obtain p-values for the sup KT and sup LR tests (see e.g. Demos and Sentana, 1998).

As in other testing situations (see Engle, 1984, page 804), the score tests will retain their optimal power against certain non normal alternatives other than the *GH*. For instance, consider a multivariate distribution with the following density function:

$$f(\mathbf{y}_t|I_{t-1};\boldsymbol{\theta}) = \frac{\exp(-\varsigma_t(\boldsymbol{\theta})/2)}{(2\pi)^{N/2}\,|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{1/2}} \left[\begin{array}{l} 1 + \eta\left(\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4}\right) \\ +\eta\mathbf{b}'\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left(\varsigma_t(\boldsymbol{\theta}) - (N+2)\right) \end{array} \right]. \qquad (15)$$

This distribution can be interpreted as a multivariate Hermite expansion of the normal distribution in which asymmetry is a common feature.[5] In this case, normality is also obtained for $\eta = 0$, regardless of $\mathbf{b}$. More formally:

Proposition 7 *If the conditional distribution of $\mathbf{y}_t$ is given by (15), then the LM and KT tests for fixed $\mathbf{b}$ will be given by (7) and (8), respectively. In addition, the sup LM test and the (sup) KT test will be given by (9) and (14), respectively.*

In contrast, the (sup) LR test should require the maximisation of (15) under the alternative hypothesis.

It is also useful to compare our symmetry test with the existing ones. In particular, the skewness component of Mardia's (1970) test can be interpreted as checking that all the (co)skewness coefficients of the standardised residuals are zero, which can be expressed by the $N(N+1)(N+2)/6$ non-duplicated moment conditions of the form:

$$E[\varepsilon_{it}^*(\boldsymbol{\theta}_0)\varepsilon_{jt}^*(\boldsymbol{\theta}_0)\varepsilon_{kt}^*(\boldsymbol{\theta}_0)] = 0, \quad i, j, k = 1, \cdots, N \tag{16}$$

But since $\varsigma_t(\boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)$, it is clear that (11) is also testing for co-skewness. Specifically, $LM_s(\tilde{\boldsymbol{\theta}}_T)$ is testing the N alternative moment conditions

$$E\{\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)]\} = E[m_{st}(\boldsymbol{\theta}_0)] = \mathbf{0}, \tag{17}$$

which are the relevant ones against GH innovations.

A less well known multivariate normality test was proposed by Bera and John (1983), who considered multivariate Pearson alternatives instead. However, since the asymmetric component of their test also assesses if (16) holds, we do not discuss it separately.

All these tests were derived for nonlinear regression models with conditionally homoskedastic disturbances estimated by Gaussian ML, in which the covariance matrix of the innovations, $\boldsymbol{\Sigma}$, is unrestricted and does not affect the conditional mean, and the conditional mean parameters, $\boldsymbol{\varrho}$ say, and the elements of $vech(\boldsymbol{\Sigma})$ are variation free. In more general models, though, they may suffer from asymptotic size distortions, as pointed out in a univariate context by Bontemps and Meddahi (2005) and Fiorentini, Sentana, and Calzolari (2004). An important advantage of our proposed normality test is that its asymptotic size is always correct because both $m_{kt}(\boldsymbol{\theta}_0)$ and $m_{st}(\boldsymbol{\theta}_0)$ are orthogonal by construction to the Gaussian score with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_0$.

---

[5]See Kiefer and Salmon (1983) for the analogous reinterpretation of the Jarque and Bera (1980) test as a test against a univariate Hermite expansion of the normal density.

By analogy with Bontemps and Meddahi (2005), one possible way to adjust Mardia's (1970) formulae is to replace $\varepsilon_{it}^{*3}(\boldsymbol{\theta})$ by $H_3[\varepsilon_{it}^*(\boldsymbol{\theta})]$ and $\varepsilon_{it}^{*2}(\boldsymbol{\theta})\varepsilon_{jt}^*(\boldsymbol{\theta})$ by $H_2[\varepsilon_{it}^*(\boldsymbol{\theta})]H_1[\varepsilon_{jt}^*(\boldsymbol{\theta})]$ $(i \neq j)$ in the moment conditions (16), where $H_k(\cdot)$ is the Hermite polynomial of order k. Alternatively, we can correct the asymptotic size by treating (16) as moment conditions, with the Gaussian scores defining the PML estimators $\tilde{\boldsymbol{\theta}}_T$ (see Newey, 1985 and Tauchen, 1985 for the general theory, and appendix D for specific details).

Finally, note that both $LM_k(\tilde{\boldsymbol{\theta}}_T)$ and $LM_s(\tilde{\boldsymbol{\theta}}_T)$ are again numerically invariant to the way in which the conditional covariance matrix is factorised, unlike the statistics proposed by Lütkephohl (1993), Doornik and Hansen (1994) or Kilian and Demiroglu (2000), who apply univariate Jarque and Bera (1980) tests to $\varepsilon_{it}^*(\tilde{\boldsymbol{\theta}}_T)$.

3.5   Power of the normality test

Although we shall investigate the finite sample properties of the different multivariate normality tests in section 5, it is interesting to study their asymptotic power properties. However, since the block-diagonality of the information matrix between $\boldsymbol{\theta}$ and the other parameters is generally lost under the alternative of GH innovations, for the purposes of this exercise we only consider models in which $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are constant but otherwise unrestricted, so that we can analytically compute the information matrix. In more complex parametrisations, though, the results are likely to be qualitatively similar.

The results at the usual 5% significance level are displayed in Figures 1a to 1d for $\psi = 1$ and $T = 5,000$ (see appendix D for details). In Figures 1a and 1b we have represented η on the x-axis. We can see in Figure 1a that for $\mathbf{b} = \mathbf{0}$ and $N = 3$, the test with the highest power is the one-sided kurtosis test, followed by its two-sided counterpart, the KT test, the sup LM test, and finally the skewness test.[6] On the other hand, if we consider asymmetric alternatives in which $\mathbf{b}$ is proportional to a vector of ones $\boldsymbol{\iota}$, such as in Figure 1b, which is not restrictive because the power of our normality test only depends on $\mathbf{b}$ through its Euclidean norm, the skewness component of the normality test becomes important, and eventually makes the KT test, the sup LM test and even the skewness test itself more powerful than both kurtosis tests. Not surprisingly, we can also see from these figures that if we apply our tests to a single component of the random vector, their power is significantly reduced.

---

[6]Given that the asymptotic power of the sup LR and sup KT test will be identical under local alternatives such as the ones that we are implicitly considering in these figures, we have drawn them together.

In contrast, we have represented b_i on the x-axis in Figures 1c and 1d. There we can clearly see the effects on power of the fact that $\mathbf{b}$ is not identified in the limiting case of $\eta = 0$. When η is very low, $\mathbf{b}$ is almost underidentified, which implies that large increases in b_i have a minimum impact on power, as shown in Figure 1c for $\eta = .005$ and $N = 3$. However, when we give η a larger value such as $\eta = .01$ (see Figure 1d), we can see how the power of those normality tests that take into account skewness rapidly increases with the asymmetry of the true distribution. Hence, we can safely conclude that, once we get away from the immediate vicinity of the null, the inclusion of the skewness component of our test can greatly improve its power. On the other hand, the power of the kurtosis test, which does not account for skewness, is less sensitive to increases in b_i. Similar results are obtained for $N = 1$, which we do not present to avoid cluttering the pictures.

Finally, we have also compared the power of our tests with those of the moment versions of Mardia's (1970) and Lütkephohl (1993) tests, where this time we have assumed that $\mathbf{b} = (b_1, 0, 0)'$ under the alternative for computational simplicity. The results show the superiority of our proposed tests against both symmetric and asymmetric GH alternatives (see Figures 1e and 1f, respectively), which confirms the fact that they are testing the most relevant moment conditions.

4 Student t tests

As we saw before, the Student t distribution is nested in the GH family when $\eta > 0$ $\psi = 1$ and $\mathbf{b} = \mathbf{0}$. In this particular case, η can be interpreted as the reciprocal of the degrees of freedom of the Student t distribution. We can use this fact to test the validity of the distributional assumptions made by FSC and other authors. Again, we will consider both LR and score tests.

4.1 The score under Student t innovations

In this case, we have to take the limit as $\psi \to 1^-$ and $\mathbf{b} \to \mathbf{0}$ of the general score function. Not surprisingly, the score with respect to $\boldsymbol{\pi}$, where $\boldsymbol{\pi} = (\boldsymbol{\theta}', \eta)'$, coincides with the formulae in FSC. But our more general GH assumption introduces two additional terms: the score with respect to $\mathbf{b}$,

$$s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, 0) = \frac{\eta\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right]}{1 - 2\eta + \eta\varsigma_t(\theta)} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \qquad (18)$$

which we will use for testing the Student t distribution versus asymmetric alternatives; and the score with respect to ψ, which in this case is identically zero despite the fact that ψ is locally identified. We shall revisit this issue in section 4.3.

4.2 The conditional information matrix under Student t innovations

Since $s_{\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = 0 \quad \forall t$, the only interesting components of the conditional information matrix under Student t innovations correspond to $s_{\boldsymbol{\theta} t}(\boldsymbol{\phi})$, $s_{\eta t}(\boldsymbol{\phi})$ and $s_{\mathbf{b} t}(\boldsymbol{\phi})$. In this respect, we can use Proposition 1 in FSC to obtain $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi} t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) = V[s_{\boldsymbol{\pi} t}(\boldsymbol{\pi}, 1, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\pi}, 1, \mathbf{0}]$. As for the remaining elements, we can show that:

Proposition 8 *The information matrix of the GH distribution, evaluated at $\eta > 0$ and $\psi = 1$ is characterised by $\mathcal{I}_{\eta \mathbf{b} t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) = \mathbf{0}$,*

$$
\begin{aligned}
\mathcal{I}_{\boldsymbol{\theta} \mathbf{b} t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) &= \frac{-2(N+2)\eta^2}{(1-2\eta)(1+(N+2)\eta)} \frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
\mathcal{I}_{\mathbf{b} \mathbf{b} t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) &= \frac{2(N+2)\eta^2}{(1-2\eta)(1+(N+2)\eta)} \boldsymbol{\Sigma}_t(\boldsymbol{\theta}).
\end{aligned}
$$

As in the case of normality, we can use the previous closed form expressions to evaluate the information matrix without resorting to either the outer product of the score or the Hessian matrix.

Let $\bar{\boldsymbol{\pi}}_T = (\bar{\boldsymbol{\theta}}_T', \bar{\eta}_T)'$ denote the parameters estimated by maximising the symmetric Student t log-likelihood function. We will assume throughout this section that the regularity conditions in Crowder (1976) apply, so that $\sqrt{T}(\bar{\boldsymbol{\pi}}_T - \boldsymbol{\pi}_0)$ is asymptotically normal with mean zero and covariance matrix $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}$, where $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}$ is the unconditional information matrix under Student t innovations. See Proposition 1 in FSC for a formal result.

4.3 Student t vs symmetric $\boldsymbol{GH}$ innovations

A test of $H_0 : \psi = 1$ under the maintained hypothesis that $\mathbf{b} = \mathbf{0}$ would be testing that the tail behaviour of the multivariate t distribution adequately reflects the kurtosis of the data. As we mentioned in section 4.1, though, it turns out that $s_{\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = 0$ $\forall t$, which means that we cannot compute the usual LM test for $H_0 : \psi = 1$. To deal with this unusual type of testing situation, Lee and Chesher (1986) propose to replace the LM test by what they call an "extremum test" (see also Bera, Ra, and Sarkar, 1998).

Given that the first-order conditions are identically 0, their suggestion is to study the restrictions that the null imposes on higher order conditions. In our case, we will use a moment test based on the second order derivative

$$s_{\psi\psi t}\left(\boldsymbol{\pi}, 1, \mathbf{0}\right) = \frac{\eta^2}{\left(1-2\eta\right)\left(1-4\eta\right)}\frac{\varsigma_t(\boldsymbol{\theta}) - N\left(1-2\eta\right)}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} + \frac{\eta^2\left[N-\varsigma_t(\boldsymbol{\theta})\right]}{\left(1-2\eta\right)\left(1+\left(N-2\right)\eta\right)}, \quad (19)$$

the rationale being that $E\left[s_{\psi\psi t}\left(\boldsymbol{\pi}_0, 1, \mathbf{0}\right)|\mathbf{z}_t, I_{t-1}, \boldsymbol{\pi}_0, \psi_0=1, \mathbf{b}_0=\mathbf{0}\right] = 0$ under the null of standardised Student t innovations with η_0^{-1} degrees of freedom, while

$$E\left[s_{\psi\psi t}\left(\boldsymbol{\pi}_0, 1, \mathbf{0}\right)|\boldsymbol{\pi}_0, \psi_0<1, \mathbf{b}_0=\mathbf{0}\right] > 0 \quad (20)$$

under the alternative of standardised symmetric GH innovations.

The statistic that we propose to test for $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$ under the maintained hypothesis that $\mathbf{b} = \mathbf{0}$ is given by

$$\tau_{kT}\left(\bar{\boldsymbol{\pi}}_T\right) = \frac{\sqrt{T}\bar{s}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)}{\sqrt{\hat{V}\left[s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)\right]}}, \quad (21)$$

where $\hat{V}\left[s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)\right]$ is a consistent estimator of the asymptotic variance of $s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)$ that takes into account the sampling variability in $\bar{\boldsymbol{\pi}}_T$. Under the null hypothesis of Student t innovations with more than 4 degrees of freedom, it is easy to see that the asymptotic distribution of $\tau_{kT}\left(\bar{\boldsymbol{\pi}}_T\right)$ will be $N\left(0, 1\right)$. The required expression for $V\left[s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)\right]$ is given in the following result:

Proposition 9

1. *If $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with $\eta_0^{-1} > 4$ degrees of freedom, then*

$$\sqrt{T}\bar{s}_{\psi\psi T}\left(\boldsymbol{\pi}_0, 1, \mathbf{0}\right) \xrightarrow{d} N\left\{0, V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right]\right\},$$

   for known $\boldsymbol{\pi}_0$, where

$$V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] = \frac{8N\left(N+2\right)\eta_0^6}{\left(1-2\eta_0\right)^2\left(1-4\eta_0\right)^2\left(1+\left(N+2\right)\eta_0\right)\left(1+\left(N-2\right)\eta_0\right)}.$$

2. *If in addition the regularity conditions of Proposition 1 in FSC hold, then*

$$\sqrt{T}\bar{s}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \xrightarrow{d} N\left\{0, V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] - \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0)\right\},$$

   where $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ is the Student t information matrix in FSC and

$$\mathcal{M}(\boldsymbol{\pi}_0) = E\left[\begin{array}{c}\mathcal{M}_{\boldsymbol{\theta}t}(\boldsymbol{\pi}_0)\\\mathcal{M}_{\eta t}(\boldsymbol{\pi}_0)\end{array}\right] = E\left[\begin{array}{c}E\left[s_{\boldsymbol{\theta}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\pi}_0, 1, \mathbf{0}\right]\\E\left[s_{\eta t}(\boldsymbol{\pi}_0, 1, \mathbf{0})s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\pi}_0, 1, \mathbf{0}\right]\end{array}\right],$$

with

$$\mathcal{M}_{\boldsymbol{\theta}t}(\boldsymbol{\pi}_0) = \frac{4\,(N+2)\,\eta_0^4\,(1-2\eta_0)^{-1}\,(1-4\eta_0)^{-1}}{[1+(N+2)\,\eta_0][1+(N-2)\,\eta_0]}\frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)],$$

$$\mathcal{M}_{\eta t}(\boldsymbol{\pi}_0) = \frac{-2N\,(N+2)\,\eta_0^3\,(1-2\eta_0)^{-2}\,(1-4\eta_0)^{-1}}{(1+N\eta_0)\,[1+(N+2)\,\eta_0]}.$$

Lee and Chesher (1986, page 145) show the equivalence between (21) and the corresponding LR test under the null and sequences of local alternatives in unrestricted contexts. However, similarly to what occurs to the normality tests, we can only compare the LR test with a one-sided Extremum test that exploits (20). Hence, the statistic $\tau_{kT}^2\left(\bar{\boldsymbol{\pi}}_T\right)\mathbf{1}\left[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})>0\right]$ will be asymptotically equivalent to a LR test of symmetric Student t vs. symmetric GH innovations, and their asymptotic distribution will be a chi-square with one degree of freedom with probability $1/2$ and 0 otherwise. For this reason, we again recommend the one sided version over the two sided counterpart.

Finally, it is also important to mention that although $s_{\psi t}\left(\boldsymbol{\pi}_0,1,\mathbf{b}\right)=0\ \forall t$, we can show that ψ is third-order identifiable at $\psi=1$, and therefore locally identifiable, even though it is not first- or second-order identifiable (see Sargan, 1983). More specifically, we can use the Barlett identities to show that

$$E\left[\frac{\partial^2 s_{\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})}{\partial\psi^2}|\boldsymbol{\pi}_0,1,\mathbf{0}\right] = -E\left[\frac{\partial s_{\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})}{\partial\psi}\cdot s_{\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})|\boldsymbol{\pi}_0,1,\mathbf{0}\right] = 0,$$

but

$$E\left[\frac{\partial^3 s_{\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})}{\partial\psi^3}|\boldsymbol{\pi}_0,1,\mathbf{0}\right] = -3V\left[\frac{\partial s_{\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})}{\partial\psi}|\boldsymbol{\pi}_0,1,\mathbf{0}\right]\neq 0.$$

4.4   Student t vs asymmetric *GH innovations*

By construction, the previous test maintains the assumption that $\mathbf{b}=\mathbf{0}$. However, it is straightforward to extend it to incorporate this symmetry restriction as an explicit part of the null hypothesis. The only thing that we need to do is to include $E[s_{\mathbf{b}t}\left(\boldsymbol{\pi}_0,1,\mathbf{0}\right)]=\mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}\left(\boldsymbol{\pi}_0,1,\mathbf{0}\right)$ is defined in (18). The asymptotic joint distribution of the two moment conditions that takes into account the sampling variability in $\bar{\boldsymbol{\pi}}_T$ is given in the following result

Proposition 10

1. *If $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with $\eta_0^{-1}>4$ degrees of freedom, then*

$$\left[\begin{array}{c}\sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}\left(\boldsymbol{\pi}_0,1,\mathbf{0}\right)\\ \sqrt{T}\bar{s}_{\psi\psi T}\left(\boldsymbol{\pi}_0,1,\mathbf{0}\right)\end{array}\right]\overset{d}{\rightarrow}N\left[0,\left[\begin{array}{cc}\mathcal{I}_{\mathbf{bb}}(\boldsymbol{\pi}_0,1,\mathbf{0}) & \mathbf{0}\\ \mathbf{0}' & V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})\right]\end{array}\right]\right],$$

for known $\boldsymbol{\pi}_0$, *where* $\mathcal{I}_{\mathbf{bb}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\mathbf{bb}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ *and* $V[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ *are defined in Propositions 8 and 9, respectively.*

2. *If in addition the regularity conditions of Proposition 1 in FSC hold, then*

$$\left[\begin{array}{c} \sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{array} \right] \xrightarrow{d} N[0, \mathcal{V}(\boldsymbol{\pi}_0)],$$

where

$$\mathcal{V}(\boldsymbol{\pi}_0) = \left[\begin{array}{cc} \mathcal{V}_{\mathbf{bb}}(\boldsymbol{\pi}_0) & \mathcal{V}_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) \\ \mathcal{V}'_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) & \mathcal{V}_{\psi\psi}(\boldsymbol{\pi}_0) \end{array} \right] = \left\{ \begin{array}{cc} \mathcal{I}_{\mathbf{bb}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{0} \\ \mathbf{0}' & V[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})] \end{array} \right\}$$
$$- \left[\begin{array}{cc} \mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \\ \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \end{array} \right],$$

$$(22)$$

$\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi} t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ *is the Student t information matrix derived in FSC,* $\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ *is defined in Proposition 8 and* $\mathcal{M}(\boldsymbol{\pi}_0)$ *is given in Proposition 9.*

Therefore, if we consider a two-sided test, we will use

$$\tau_{gT}(\bar{\boldsymbol{\pi}}_T) = \left[\begin{array}{c} \sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{array} \right]' \mathcal{V}^{-1}(\bar{\boldsymbol{\pi}}_T) \left[\begin{array}{c} \sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{array} \right], \qquad (23)$$

which is distributed as a chi-square with $N + 1$ degrees of freedom under the null of Student t innovations. However, we must again exploit the one-sided nature of the ψ-component of the test to obtain a statistic that is asymptotically equivalent to a LR test of Symmetric Student t vs. Asymmetric GH innovations. Since $\mathcal{V}(\boldsymbol{\pi}_0)$ is not block diagonal in general, we must orthogonalise the moment conditions (see e.g. Silvapulle and Silvapulle, 1995). Specifically, instead of using directly the score with respect to $\mathbf{b}$, we consider

$$s_{\mathbf{b}t}^{\perp}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) = s_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) - \mathcal{V}_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)\mathcal{V}_{\psi\psi}^{-1}(\bar{\boldsymbol{\pi}}_T)s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}),$$

whose sample average is asymptotically orthogonal to $\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ by construction. Note, however, that there is no need to do this orthogonalisation when $E[\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0] = \mathbf{0}$, since in this case $\mathcal{V}_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) = \mathbf{0}$ because $\mathcal{I}_{\boldsymbol{\pi}b}(\boldsymbol{\pi}_0, 1, 0) = \mathbf{0}$ (see Proposition 8).

It is then straightforward to see that the asymptotic distribution of

$$\tau_{oT}(\bar{\boldsymbol{\pi}}_T) = T\bar{s}_{\mathbf{b}t}^{\perp\prime}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \left[\mathcal{V}_{\mathbf{bb}}(\bar{\boldsymbol{\pi}}_T) - \frac{\mathcal{V}_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)\mathcal{V}'_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)}{\mathcal{V}_{\psi\psi}(\bar{\boldsymbol{\pi}}_T)} \right]^{-1} \bar{s}_{\mathbf{b}t}^{\perp}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$$
$$+ \tau_{kT}^2(\bar{\boldsymbol{\pi}}_T)\mathbf{1}[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) > 0] \qquad (24)$$

will be another 50:50 mixture of chi-squares with N and $N + 1$ degrees of freedom under the null, because asymptotically, the probability that $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ is negative

will be .5 if $\psi_0 = 1$. Such a one-sided test benefits from the fact that a non-positive $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ gives no evidence against the null, in which case we only need to consider the orthogonalised skewness component. In contrast, when $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ is positive, (24) numerically coincides with (23). The asymptotic null distribution of the LR test of Symmetric Student t vs. Asymmetric GH innovations will be the same. Importantly, note once more that (24) is numerically invariant to the chosen factorisation of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, as expected from (5).

On the other hand, if we only want to test for symmetry, we may use

$$\tau_{aT}(\bar{\boldsymbol{\pi}}_T) = \sqrt{T}\bar{s}'_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})\,\mathcal{V}^{-1}_{\mathbf{bb}}(\bar{\boldsymbol{\pi}}_T)\,\sqrt{T}\bar{s}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}), \tag{25}$$

which can be interpreted as a regular LM test of the Student t distribution versus the GH distribution under the maintained assumption that $\psi = 1$. In this particular case, the GH distribution is known as the Asymmetric t (see Mencía, 2003). As a result, $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ will be asymptotically distributed as a chi-square distribution with N degrees of freedom under the null of Student t innovations, and it will be asymptotically equivalent to a LR test of Symmetric Student t vs. Asymmetric t innovations.

Given that we can show that the moment condition (17) remains valid for any elliptical distribution, the symmetry component of our proposed normality test provides an alternative consistent test for $H_0 : \mathbf{b} = \mathbf{0}$, which is however incorrectly sized when the innovations follow an elliptical but non-Gaussian distribution. To avoid size distortions, one possibility would be to scale $LM_s(\tilde{\boldsymbol{\theta}}_T)$ by multiplying it by a consistent estimator of the adjusting factor

$$\frac{2N(N+2)}{E(\varsigma_t^3(\boldsymbol{\theta}_0)) - 2(N+2)E(\varsigma_t^2(\boldsymbol{\theta}_0)) + N(N+2)^2} \tag{26}$$

which becomes $[(1 - 4\eta_0)(1 - 6\eta_0)]/[1 + (N-2)\eta_0 + 2(N+4)\eta_0^2]$ for the Student t. Alternatively, we can run the univariate regression of 1 on $m_{st}(\bar{\boldsymbol{\theta}}_T)$, or the multivariate regression of $\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\theta}}_T)$ on $\varsigma_t(\bar{\boldsymbol{\theta}}_T) - (N+2)$, although in the latter case we should use standard errors that are robust to heteroskedasticity. Not surprisingly, we can show that these three procedures to test (17) are asymptotically equivalent under the null. However, they are only valid if there are finite moments up to the sixth order (i.e. $\eta < 1/6$), and will be generally less powerful against local alternatives of the form $\mathbf{b}_{0T} = \mathbf{b}_0/\sqrt{T}$ than $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ in (25), which is the proper LM test for symmetry.

Nevertheless, an interesting property of a moment test for symmetry based on (17) is that $\sqrt{T}\bar{m}_{sT}(\bar{\boldsymbol{\theta}}_T)$ and $\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ are asymptotically independent under the null of symmetric Student t innovations, which means that there is no need to resort to orthogonalisation in order to obtain a one-sided version that combines both of them.

5 A Monte Carlo comparison of finite sample size and power properties

In this section, we assess the finite sample size and power properties of the testing procedures discussed above by means of several extensive Monte Carlo exercises, with an experimental design borrowed from Sentana (2004), which aimed to capture some of the main features of the conditionally heteroskedastic factor model in King, Sentana, and Wadhwani (1994).

Finite sample size of the normality tests   We first simulate the following trivariate Gaussian model:
$$y_{it} = \mu_i + c_i f_t + v_{it} \quad i = 1, 2, 3,$$
where $f_t = \lambda_t^{1/2} f_t^*$, $v_{it} = \gamma_{it}^{1/2} v_{it}^*$ $(i = 1, 2, 3)$,

$$\lambda_t = \alpha_0 + \alpha_1(f_{t-1|t-1}^2 + \omega_{t-1|t-1}) + \alpha_2 \lambda_{t-1},$$
$$\gamma_{it} = \phi_0 + \phi_1 \left[(y_{it-1} - \mu_i - c_i f_{t-1|t-1})^2 + c_i^2 \omega_{t-1|t-1} \right] + \phi_2 \gamma_{it-1}, \quad i = 1, 2, 3,$$

$(f_t^*, v_{1t}^*, v_{2t}^*, v_{3t}^*)|I_{t-1} \sim N(\mathbf{0}, \mathbf{I}_4)$, and $f_{t-1|t-1}$ and $\omega_{t-1|t-1}$ are the conditional Kalman filter estimate of f_t and its conditional MSE, respectively. Hence, the conditional mean vector and covariance matrix functions associated with this model are of the form

$$\begin{aligned}\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{c}\mathbf{c}'\lambda_t + \boldsymbol{\Gamma}_t,\end{aligned} \tag{27}$$

where $\boldsymbol{\mu}' = (\mu_1, \mu_2, \mu_3)$, $\mathbf{c}' = (c_1, c_2, c_3)$, and $\boldsymbol{\Gamma}_t = diag(\gamma_{1t}, \gamma_{2t}, \gamma_{3t})$. As for parameter values, we have chosen $\mu_i = .2$, $c_i = 1$, $\alpha_1 = \phi_1 = .1$, $\alpha_2 = \phi_2 = .85$, $\alpha_0 = 1 - \alpha_1 - \alpha_2$ and $\phi_0 = 1 - \phi_1 - \phi_2$. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 1,000$ observations based on 10,000 Monte Carlo replications, which allows us to estimate actual sizes with high precision.[7] Further details are available on request.

Given that the asymptotic distributions that we have derived in previous sections may be unreliable in finite samples, we compute both asymptotic and bootstrap p-values. In this regard, it is important to note that Andrews (2000) shows that the size

---

[7]For instance, the 95% confidence interval for a nominal size of 5% would be (4.6%,5.4%).

of bootstrap tests remains asymptotically valid when some of the parameters are on the boundary of the parameter space, even though the usual bootstrap standard errors are not reliable in those circumstances. We consider a parametric bootstrap procedure with 1,000 samples for all tests except the LR test, for which we could only use 100 samples for computational reasons.[8] Given that the *GH* log-likelihood function is very flat around the normality region, one has to be very careful in choosing starting values. We consider a fine grid of 20×5 different initial values for the pair (η, ψ) to maximise the likelihood under the alternative. But since it was computationally infeasible to implement a similar grid search for the vector of asymmetry parameters, we only considered a single initial **b** given by the value that leads to the sup LM test (see the proof of Proposition 5).[9]

Proposition 1 implies that both the sup LM and the LR tests are asymptotically independent of the Gaussian PML estimators of the conditional mean and variance para-meters regardless of the model specification. In contrast, the original Mardia (1970) and Lütkephohl (1993) expressions were derived under the assumption that the covariance matrix of the innovations is constant but otherwise unrestricted, and does not affect the conditional mean. To deal with this problem, we have interpreted those tests as moment tests, and adjusted them appropriately so that their size distortions disappear. Specific-ally, we orthogonalise the Mardia (1970) and Lütkephohl (1993) expressions with respect to the Gaussian scores of $\boldsymbol{\theta}$. This orthogonality allows us to save substantial computer time because we do not need to reestimate $\boldsymbol{\theta}$ in each bootstrap sample.

Figures 2-4 summarise 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. The left panels show the discrepancy plots of the asymptotic p-values, while the right panels show the corresponding results obtained with the parametric bootstrap. Figure 2a shows that the LR test seems to be too conservative in general, especially for large nominal sizes. In this sense, we can observe in Figure 2b that the parametric bootstrap is able to reduce

---

[8]Even so, the computation of the bootstrap p-value of the LR test took about 15 minutes in a MS Windows PC node with a 2.8GHz processor. To speed up the computations, we employed a cluster of ten such nodes, which limited the computational time to approximately two weeks per Monte Carlo design. Using 1,000 bootstrap samples would provide more reliable results but at the cost of increasing the computational burden tenfold.

[9]Despite our careful choice of initial values, the LR turned out to be negative approximately 10% of the time. In those cases, we simply set it to 0.

those distortions to some extent.[10] As for the remaining tests, the actual finite sample sizes seem to be fairly close to their nominal levels, with the possible exception of the one-sided version of the kurtosis test (see Figure 4a), which seems to be also somewhat conservative for larger nominal sizes. But again, Figure 4b shows that the bootstrap can substantially reduce the distortions.

Finite sample size of the Student t tests   In this case we maintain the conditional mean and variance specification in (27), but generate the standardised innovations ε_t^* from a trivariate Student t distribution with 10 degrees of freedom. As before, we compare the asymptotic p-values of the tests with their bootstrapped counterparts. Again, we consider 1,000 bootstrap samples for the LM-type test, but we can only afford 100 samples for the LR test. Since we can easily orthogonalise the moment conditions of the LM test with respect to $\bar{\bar{\pi}}_T$, we did not need to reestimate the model to carry out a parametric bootstrap. Unfortunately, in the case of the LR test we have to reestimate $\boldsymbol{\theta}$ under the null and the alternative hypothesis in each bootstrap sample, which makes these computations even slower than those of the normality test.

Figure 5 shows the p-value discrepancy plots of the one- and two-sided versions of the Student t tests discussed in section 4, together with those of their asymmetric and kurtosis components, and the LR test. The most striking feature of the results for the asymptotic p-values, shown in Figure 5a, is the fact that the actual sizes of the "kurtosis" tests based on $\tau_{kT}(\bar{\bar{\pi}}_T)$, which is defined in (21), are well below their nominal sizes. This is due to the fact that the sampling distribution of $\tau_{kT}(\bar{\bar{\pi}}_T)$ is not well approximated by a standard normal, as illustrated in Figure 6. In contrast, the actual sizes of the asymmetry component are very much on target. The joint tests inherit part of the size distortions of the kurtosis tests, while the LR test is also somewhat conservative. Finally, Figure 5b confirms that the parametric bootstrap is able to yield p-values that are much closer to the nominal ones.[11]

Finite sample power of the normality tests   We have repeated the normality tests using the same mean and variance specification as in (27), but generating the 10,000

---

[10]The apparent higher distortions of the bootstrapped p-values of the LR test for very small nominal sizes is simply due to the limited accuracy that we can obtain from just 100 bootstrap samples.

[11]Once again, the bootstrapped p-values of the LR test are not very accurate for very small nominal sizes due to the small number of bootstrap samples that we can use.

Monte Carlo replications from a standardised GH distribution with $\eta = .01, \psi = 1$ and $\mathbf{b} = (-.05, -.05, -.05)'$, which corresponds to an asymmetric t distribution. Figure 7a shows the size-power curves proposed by Davidson and MacKinnon (1998) using the empirical distribution that we have obtained under the null. The results indicate that the LR and KT tests seem to display similar power even though we are not in the immediate vicinity of the null. In this sense, it is worth mentioning that these two tests are consistent for fixed alternatives but diverge to infinity separately. Finally, note that the power of Mardia's and especially Lütkepohl's tests is smaller.

Finite sample power of the Student t tests   Finally, we generate the standardised innovations from a GH distribution with $\eta = .2, \psi = .3$ and $\mathbf{b} = (-.05, -.05, -.05)'$. These parameter values are far away from the null of $\psi = 1$, which implies that the local equivalence between the LR and KT test no longer applies. Note that we do not consider an asymmetric t alternative in this case, because then we could only assess the power of the symmetry component of the test. Once again, we use the size-power curves of Davidson and MacKinnon (1998) using the empirical distribution under a null generated with the pseudo true parameter values, which in this case differ from the true ones (see Fiorentini and Sentana, 2007). As we cannot obtain those values in closed form, we use the average Student t estimates of $\boldsymbol{\pi}$ obtained from the 10,000 replications simulated under the alternative. As Figure 7b shows, the LR and KT tests also yield similar power in this case, although the LR test seems to be slightly more powerful for large sizes. As expected, the one-sided kurtosis component of the test displays less power, because it only relies on one moment condition. Finally, the two sided kurtosis test almost has no power, which confirms the convenience of considering its one sided counterpart.

6   Empirical application

We now apply the tests derived in the previous sections to the returns on the ten US sectoral stock indices from Datastream.[12] Specifically, our data consists on daily excess returns for the period January 4th, 1988 - October 12th, 2007 (4971 observations), where we have used the Eurodollar overnight interest rate as safe rate (Datastream code ECUSDST). The model used is a generalisation of the one in the previous section

---

[12]Namely, Basic Materials, Consumer Goods, Consumer Services, Financials, Health Care, Industrials, Oil and Gas, Technology, Telecommunications and Utilities.

(see (27)), in which the mean dynamics are captured by a diagonal VAR(1) model with drift, and the covariance dynamics by a conditionally heteroskedastic single factor model in which the conditional variances of both common and specific factors follow GQARCH(1,1) processes to allow for leverage effects (see Sentana, 1995).

We have estimated this model under three different conditional distribution assumptions on the standardised innovations ε_t^*: Gaussian, Student t and GH. We first estimated the model by Gaussian PML and then computed the sup LM and Kuhn-Tucker normality tests described in section 3.4, whose asymptotic and parametric bootstrap p-values are reported in Table 1b. In this sense, it worth mentioning that the skewness component of Mardia's test would have 220 degrees of freedom in our ten dimensional application. Our tests show that skewness and excess kurtosis are both very significant, although the kurtosis component is one order of magnitude larger than the skewness test. Nevertheless, the skewness test is affected by the presence of excess kurtosis. We can control for this effect by using the scaling factor in (26) in estimating the appropriate weighting matrix in (11). When we carry out this adjustment, the skewness test turns out to be 18.10, with a p-value of 0.053. Therefore, while the evidence about the presence excess kurtosis is clearly overwhelming, skewness is less important once we take into account excess kurtosis.

In order to shed further light on this issue, we estimated a multivariate Student t model using the analytical formulae for the score that FSC provide. The results in Table 1a show that the estimate for the tail thickness parameter η, which corresponds to slightly more than 10 degrees of freedom, is significantly larger than 0. Then, on the basis of the Student t ML estimates, we have computed the statistics $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ and $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ introduced in section 4. The results in Table 1c show that we can reject the Student t assumption at conventional levels because of the value we obtain for the skewness component $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$. However, the one-sided version of the ψ component of the test is unable to reject the Student t specification against the alternative hypothesis of symmetric GH innovations because $\mathbf{1}(\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) > 0) = 0$.

Finally, we re-estimated the model under the assumption that the conditional distribution of the innovations is GH using the analytical expressions for the score that we derive in Appendix B.2. In this case, the GH log-likelihood introduces as additional parameters ψ and the ten-dimensional vector $\mathbf{b}$. Since the ML estimate of ψ reported in

Table 1a is 1, and $\hat{\eta}_T$ is positive, the estimated GH conditional distribution effectively corresponds to an asymmetric t.

The KT results are confirmed by the LR tests. Specifically, the LR test of Gaussian versus symmetric Student t innovations yields a value of 2246.63, which is highly significant, despite being more than four times smaller than the corresponding KT test. Note, though, that the asymptotic equivalence of the KT and LR tests only holds under the null of Gaussianity and sequences of local alternatives, which is clearly not the case in the data. Similarly, the LR test of Student t vs. asymmetric GH innovations also rejects the null, although the gains in fit obtained by allowing for asymmetry are not as important as those previously obtained by generalising the normal distribution in the leptokurtic dimension. This fact is also likely to explain why the LR and KT test are now commensurate. Interestingly, the asymptotic and bootstrapped p-values are fairly similar in all cases.

Conceivably, though, the rejection of the null hypotheses of normal and Student t innovations that we find could be exacerbated by misspecification of the first and especially second conditional moments. If our specification of the model dynamics is correct, however, the marked distributional differences that we have found should not affect the consistency of the Gaussian PML estimators of $\boldsymbol{\theta}$. With this in mind, we compare the multivariate Gaussian estimate of the conditional variance with the one obtained with a univariate model for the equally weighted portfolio. Specifically, the univariate model is a Gaussian AR(1)-GQARCH(1,1) model. Reassuringly, Figure 8a shows that the (log)standard deviations of the two series display a very similar pattern, although the univariate estimates are somewhat noisier. An alternative check of our dynamic specification is to compare the multivariate Gaussian and GH estimates of the covariance matrix. The Gaussian estimates should remain consistent even in the absence of normality, while the GH ones will only be consistent if this distribution is correctly specified. Figure 8b shows the (log)standard deviations of the equally weighted portfolio that we obtain with these two distributions, which again display a very similar pattern.

As mentioned in the introduction, one of the main reasons for taking into account departures from normality is to avoid biases in the estimation of the quantiles of the distribution, which are required for instance in V@R calculations. To determine to what extent the Student t and the GH distributions are more useful than the Gaussian

assumption across all conceivable quantiles, we have computed the empirical cumulative distribution function of the one-period-ahead probability integral transforms of the equally weighted portfolio of Datastream indices generated by the three fitted distributions (see Diebold, Gunther, and Tay, 1998). Figure 9a shows the difference between the corresponding cumulative distributions and the 45 degree line. Under correct specification, those differences should tend to 0 asymptotically. Unfortunately, a size-corrected version of the usual Kolmogorov or Cramer von Mises tests that takes into account the sample uncertainty in the estimation of the underlying parameter estimates is rather difficult to obtain in this case. Nevertheless, the graph clearly indicates that the Student t and asymmetric t distributions clearly provide a better fit than the normal. At the same time, the two non-normal distributions are hard to distinguish.

Since these results are specific to one particular portfolio, we repeat the same exercise for 5,000 different portfolios whose weights are randomly chosen from a uniform distribution on the unit sphere. Figure 10, which compares the empirical distribution function across those 5,000 portfolios of the Cramer von Mises goodness of fit statistics, clearly shows that the symmetric and asymmetric t dominate the normal distribution in the first-order stochastic sense. Once again, though, it is difficult to distinguish the two fat tailed distributions. In order to magnify the differences between the symmetric and asymmetric t distributions, we look at the most asymmetric portfolio, whose weights are proportional to $\mathbf{b}$ (see Mencía and Sentana, 2008, for a formal justification). In this sense, Figure 9b shows that the asymmetric t provides a slightly better fit to the empirical distribution of this benchmark portfolio. Note, however, that even in this case the gains from allowing for asymmetry are smaller than those from taking into account excess kurtosis.

Finally, an interesting question is whether the results that we have obtained are specific to the daily frequency. We assess this issue by repeating our tests for weekly returns of the same 10 Datastream US indices in excess of the Eurodollar weekly interest rates, with the same parametrisation of the conditional mean vector and covariance matrix. Table 2 shows the results. Again, normality is easily rejected, mainly because of the presence of excess kurtosis.[13] In this case, though, the Student t distribution cannot be rejected at the 5% level, although the LR tests are significant at the 10% level. This

---

[13]The skewness score based test becomes 12.42 once we adjust the weighting matrix in (11) by assuming ellipticity but not Gaussianity.

can be due to the smaller sample size (1043 observations). Notice also that the LR test of asymmetric t vs. asymmetric GH innovations is positive but not significant at conventional levels.

7 Conclusions

In this paper, we propose LM and LR specification tests of multivariate normality and multivariate Student t against alternatives with GH innovations, which is a rather flexible multivariate asymmetric distribution that also nests as particular cases many other well known and empirically realistic examples. Methodologically, our main contribution is to explain how to overcome the identification problems that the use of the GH distribution as an embedding model entails. We derive closed form expressions for the score based tests and decompose our proposed statistics into skewness and kurtosis components. From these expressions, we obtain more powerful one-sided KT versions and show their asymptotic equivalence to LR tests. For this reason, we would recommend the KT instead of the LM tests. We also exploit this equivalence to obtain the common asymptotic distributions of the LR and KT tests, which turn out to be standard despite the non-standard features of the problem.

We assess the finite sample size properties of the testing procedures that we propose and previously suggested methods by means of detailed Monte Carlo exercises. Our results indicate that the asymptotic sizes of our normality tests are very reliable in finite samples. However, we also find that the kurtosis component of the Student t test is too conservative, and the same is true of the corresponding LR test. Nevertheless, we show that one can correct those distortions by means of a parametric bootstrap, although obtaining reliable p-values for the LR test is computationally time consuming. In finite samples, we find that the LR and KT tests yield very similar power in both the Gaussian and Student t cases for parameter configurations that cannot be regarded as local to the null.

Finally, we present an empirical application to the ten US sectoral daily excess stock returns from Datastream. We can easily reject normality because the skewness and especially kurtosis components of our tests are highly significant. And while a multivariate symmetric Student t seems to fit well the kurtosis of the data, the skewness component of the Student t is still significant. In sum, our results suggest that the conditional

distribution of the returns on those US indices is mildly asymmetric and strongly leptokurtic. From a risk management perspective, the quantiles of arbitrary portfolios are much better captured by either the Student t or the GH distribution than by the normal for most cases, while the GH distribution displays a slightly better fit than the Student t in the case of the most asymmetric portfolio. We also compute tests for the same data at the weekly frequency, finding similar results, although in this case asymmetry seems to be even less important, which may be due to the smaller sample size.

An interesting extension of our results would be to test multivariate normality against a general location-scale mixture of normals, although the resulting tests will also be affected by the same type of underidentification problems under the null. Alternatively, we could consider as our null hypothesis other special cases of the GH distribution, such as the symmetric normal-gamma. Finally, one could use the test statistics that we have derived to improve the efficiency of indirect estimators along the lines suggested by Calzolari, Fiorentini, and Sentana (2004).

References

Aas, K., X. Dimakos, and I. Haff (2005). Risk estimation using the multivariate normal inverse gaussian distribution. *Journal of Risk 8*, 39–60.

Abramowitz, M. and A. Stegun (1965). *Handbook of mathematical functions.* New York: Dover Publications.

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

Andrews, D. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica 68*, 399–405.

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

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics 85*, 531–549.

Bai, J. and C. Zhihong (2008). Testing multivariate distributions in GARCH models. *Journal of Econometrics 143*, 19–36.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. 353*, 401–419.

Barndorff-Nielsen, O. and N. Shephard (2001). Normal modified stable processes. *Theory of Probability and Mathematical Statistics 65*, 1–19.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics 23*, 346–354.

Bera, A. and S. John (1983). Tests for multivariate normality with Pearson alternatives. *Communications in Statistics - Theory and Methods 12*, 103–117.

Bera, A. and G. Premaratne (2002). Modeling asymmetry and excess kurtosis in stock return data. Working Paper 00-0123, University of Illinois.

Bera, A., S. Ra, and N. Sarkar (1998). Hypothesis testing for some nonregular cases in econometrics. In D. Coondoo and R. Mukherjee (Eds.), *Econometrics: Theory and Practice*, pp. 319–351. Allied Publishers.

Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika 68*, 251–263.

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

Bontemps, C. and N. Meddahi (2005). Testing normality: a GMM approach. *Journal*

of Econometrics 124, 149–186.

Cajigas, J. and G. Urga (2007). A risk management analysis using the AGDCC model with asymmetric multivariate Laplace distribution of innovations. mimeo Cass Business School.

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

Chen, Y., W. Härdle, and S. Jeong (2004). Nonparametric risk management with Generalised Hiperbolic distributions. mimeo, CASE, Humboldt University.

Cho, J. S. and H. White (2007). Testing for regime switching. *Econometrica 75*, 1671–1720.

Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society, Series B 38*, 45–53.

Davidson, R. and J. G. MacKinnon (1983). Small sample properties of alternative forms of the Lagrange Multiplier test. *Economics Letters 12*, 269–275.

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

Demos, A. and E. Sentana (1998). Testing for GARCH effects: a one-sided approach. *Journal of Econometrics 86*, 97–127.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*, 863–883.

Doornik, J. A. and H. Hansen (1994). An omnibus test for univariate and multivariate normality. Working Paper W4&91, Nuffield College, Oxford.

Engle, R. F. (1984). Wald, likelihood ratio, and lagrange multiplier tests in econometrics. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 775–826. North-Holland.

Farebrother, R. (1990). The distribution of a quadratic form in normal variables (Algorithm AS 256.3). *Applied Statistics 39*, 294–309.

Fiorentini, G. and E. Sentana (2007). On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models. CEMFI Working Paper No. 0713.

Fiorentini, G., E. Sentana, and G. Calzolari (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.

Fiorentini, G., E. Sentana, and G. Calzolari (2004). On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models. *Economics Letters 83*, 307–312.

Gouriéroux, C., A. Holly, and A. Monfort (1980). Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters. Harvard Institute of Economic Research Discussion Paper 770.

Hansen, B. (1991). Lagrange multiplier tests for parameter instability in non-linear models. Mimeo.

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

Johnson, N., S. Kotz, and N. Balakrishnan (1994). *Continuous univariate distributions*, Volume 1. New York: John Wiley and Sons.

Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution.* New York: Springer-Verlag.

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

Kilian, L. and U. Demiroglu (2000). Residual-based test for normality in autoregressions: asymptotic theory and simulation evidence. *Journal of Business and Economic Statistics 18*, 40–50.

King, M., E. Sentana, and S. Wadhwani (1994). Volatility and links between national stock markets. *Econometrica 62*, 901–933.

King, M. L. and T. S. Shively (1993). Locally optimal testing when a nuisance parameter is present only under the alternative. *The Review of Economics and Statistics 75*, 1–7.

Koerts, J. and A. P. J. Abrahamse (1969). *On the theory and application of the general linear model.* Rotterdam: Rotterdam University Press.

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 M. McAleer (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory 19*, 280–310.

Lütkephohl, H. (1993). *Introduction to multiple time series analysis* (2nd ed.). New York: Springer-Verlag.

Madan, D. B. and F. Milne (1991). Option pricing with V.G. martingale components. *Mathematical Finance 1*, 39–55.

Magnus, J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Économie et de Statistique 4*, 95–109.

Magnus, J. R. and H. Neudecker (1988). *Matrix differential calculus with applications in statistics and econometrics.* Chichester: John Wiley and Sons.

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

Mencía, J. (2003). Modeling fat tails and skewness in multivariate regression models.

Unpublished Master Thesis CEMFI.

Mencía, J. and E. Sentana (2008). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. mimeo.

Newey, W. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*, 1047–1070.

Newey, W. K. and D. G. Steigerwald (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica 65*, 587–599.

Sargan, J. D. (1983). Identification and lack of identification. *Econometrica 51*, 1605–1634.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies 62*, 639–661.

Sentana, E. (2004). Factor representing portfolios in large asset markets. *Journal of Econometrics 119*, 257–289.

Silvapulle, M. and P. Silvapulle (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association 90*, 342–349.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics 30*, 415–443.

A   Proofs of Propositions

Proposition 1

To compute the score when η goes to zero, we must take the limit of the score function after substituting the modified Bessel functions by the appropriate expansion (see Appendix B.3). We operate in a similar way when $\psi \to 0^+$. Then, the conditional information matrix under normality can be easily derived as the conditional variance of the score function by using the property that, if $\boldsymbol{\varepsilon}_t^*$ is distributed as a multivariate standard normal, then it can be written as $\boldsymbol{\varepsilon}_t^* = \sqrt{\zeta_t}\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_t is a chi-square random variable with N degrees of freedom, and $\mathbf{u}_t$ and ζ_t are mutually independent. $\qquad\square$

Proposition 2

This proposition is a particular case of Theorem 2.1 in Bollerslev and Wooldridge (1992), where we impose that Gaussianity is satisfied.

Proposition 3

For fixed $\mathbf{b}$ and ψ and known $\boldsymbol{\theta}_0$, the LM_1 test is based on the average score with respect to η evaluated at $\eta \to 0^+$. The proportionality of the log-likelihood scores corresponding to η evaluated at $0^\pm$ and the score corresponding to ψ evaluated at 0^+ leads to (7). Then, a standard central limit theorem for martingale difference sequences can be used to show that the LM test has the expected asymptotic distribution.

If we introduce $\tilde{\boldsymbol{\theta}}_T$, the test will in principle be based on the scores with respect to either η and $\boldsymbol{\theta}$ or ψ and $\boldsymbol{\theta}$. But since the average score with respect to $\boldsymbol{\theta}$ will be zero at those parameter values, and the conditional information matrix is block-diagonal, the formula of the test does not change. In addition, we can exploit the root T consistency of $\tilde{\boldsymbol{\theta}}_T$ to perform the usual Taylor expansion of the test moment conditions around $\boldsymbol{\theta}_0$. Then, using the asymptotic orthogonality between these moment conditions and the score with respect to $\boldsymbol{\theta}$ we can easily obtain the required result. $\qquad\square$

Proposition 4

Consider initially the situation in which we fix $\mathbf{b}$ and ψ, and only allow η to be positive under the alternative. The first thing to note is that such a LR ratio will be identically 0 if the sample average of (6) is negative, which happens approximately half the time in large samples. Therefore, the results in Gouriéroux, Holly, and Monfort (1980) imply

that the LR test will not be asymptotically equivalent to the corresponding LM test $LM_1(\boldsymbol{\theta}_0, \psi, \mathbf{b})$, but rather to the Kuhn-Tucker test

$$KT_1(\boldsymbol{\theta}_0, \psi, \mathbf{b}) = \mathbf{1}(\bar{s}_{\eta T}(\boldsymbol{\theta}_0, 0^+, \psi, \mathbf{b}) \geq 0) \cdot LM_1(\boldsymbol{\theta}_0, \psi, \mathbf{b}),$$

which does not depend on ψ.

Similarly, if we fix $\mathbf{b}$ and ψ, but this time we only allow η to be negative under the alternative, we will have that the LR test will be asymptotically equivalent to

$$KT_2(\boldsymbol{\theta}_0, \psi, \mathbf{b}) = \mathbf{1}(\bar{s}_{\eta T}(\boldsymbol{\theta}_0, 0^-, \psi, \mathbf{b}) \leq 0) \cdot LM_2(\boldsymbol{\theta}_0, \psi, \mathbf{b})$$

Finally, it is not surprising that if we fix $\mathbf{b}$ and η then the LR test is asymptotically equivalent to the Kuhn-Tucker test

$$KT_3(\boldsymbol{\theta}_0, \eta, \mathbf{b}) = \mathbf{1}(\bar{s}_{\psi T}(\boldsymbol{\theta}_0, \eta, 0^+, \mathbf{b}) \geq 0) \cdot LM_3(\boldsymbol{\theta}_0, \eta, \mathbf{b}),$$

which does not depend on η.

But since those three Kuhn-Tucker tests numerically coincide in any given sample, we will have that the LR that estimates over both η and ψ for a given value of $\mathbf{b}$ will be asymptotically equivalent under the null to the following test statistic:

$$KT(\boldsymbol{\theta}_0, \mathbf{b}) = \mathbf{1}(\bar{s}_{\eta T}(\boldsymbol{\theta}_0, 0, \mathbf{b}) \geq 0) \cdot LM(\boldsymbol{\theta}_0, \mathbf{b}),$$

as required. Finally, given the root T consistency of $\tilde{\boldsymbol{\theta}}_T$, the second part of the proposition follows from the same arguments as in Proposition 3. $\qquad\square$

Proposition 5

$LM(\boldsymbol{\theta}_0, \mathbf{b})$ can be trivially expressed as

$$LM(\boldsymbol{\theta}_0, \mathbf{b}) = \frac{T\mathbf{b}^{+\prime}\bar{m}_T(\boldsymbol{\theta}_0)\bar{m}_T(\boldsymbol{\theta}_0)\mathbf{b}^+}{(N+2)\,\mathbf{b}^{+\prime}\mathbf{D}_T\mathbf{b}^+}, \tag{A1}$$

where $\mathbf{b}^+ = (1, \mathbf{b}')'$, $\bar{m}_T(\boldsymbol{\theta}_0) = [\bar{m}_{kT}(\boldsymbol{\theta}_0), \bar{m}_{sT}(\boldsymbol{\theta}_0)]$, $\bar{m}_{kT}(\boldsymbol{\theta})$ and $\bar{m}_{sT}(\boldsymbol{\theta})$ are the sample means of $m_{kt}(\boldsymbol{\theta})$ and $m_{st}(\boldsymbol{\theta})$, which are defined in (12) and (17), respectively, and

$$\mathbf{D}_T = \begin{bmatrix} N/2 & \mathbf{0} \\ \mathbf{0}' & 2\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \end{bmatrix}.$$

But since the maximisation of (A1) with respect to $\mathbf{b}^+$ is a well-known generalised eigenvalue problem, its solution will be proportional to $\mathbf{D}_T^{-1}\bar{m}_T$. If we select $N/[2\bar{m}_{kT}(\boldsymbol{\theta}_0)]$ as the constant of proportionality, then we can make sure that the first element in $\mathbf{b}^+$

is equal to one. Substituting this value in the formula of $LM(\boldsymbol{\theta}_0, \mathbf{b})$ yields the required result. Based on a standard central limit theorem for martingale difference sequences, the asymptotic distribution of the sup test follows directly from the fact that $\sqrt{T}\bar{m}_{kT}(\boldsymbol{\theta}_0)$ and $\sqrt{T}\bar{m}_{sT}(\boldsymbol{\theta}_0)$ are asymptotically orthogonal under the null, with asymptotic variances $N(N+2)/2$ and $2(N+2)\boldsymbol{\Sigma}$, respectively.

Finally, given the root T consistency of $\tilde{\boldsymbol{\theta}}_T$, the second part of the proposition follows from the same arguments as in Proposition 3. $\qquad\square$

Proposition 6

For the sake of simplicity, let us consider the asymmetric t distribution, which is a particular case of the GH distribution in which $\eta > 0$ and $\psi = 1$. Hence, normality will be obtained when $\eta = 0$. Under normality, the score with respect to $\mathbf{b}$ is zero, while the score with respect to η is given by (6). Now, consider a reparametrisation in terms of $\eta^{\ddagger}$ and $\mathbf{b}^{\ddagger}$, where $\eta^{\ddagger} = \eta$ and $\mathbf{b}^{\ddagger} = \mathbf{b}\eta$. This reparametrisation is such that under normality both $\eta^{\ddagger}$ and $\mathbf{b}^{\ddagger}$ will be zero, while under local alternatives of the form $\eta_T^{\ddagger} = T^{-1/2}\bar{\eta}^{\ddagger}$ and $\mathbf{b}_T^{\ddagger} = T^{-1/2}\bar{\mathbf{b}}^{\ddagger}$ we will have an asymmetric student t distribution with parameters $\eta_T = T^{-1/2}\bar{\eta}$ and $\mathbf{b}_T = \bar{\mathbf{b}}$. If we apply the chain rule we can express the score with respect to the new parameters as

$$\lim_{\eta\to 0^+} s_{\eta^{\ddagger}t}(\boldsymbol{\phi}) = \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4}, \tag{A2}$$

$$\lim_{\eta\to 0^+} s_{\mathbf{b}^{\ddagger}t}(\boldsymbol{\phi}) = \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right], \tag{A3}$$

under normality. Note that the maximum likelihood estimate of $\eta^{\ddagger}$, which cannot be negative, will be zero when (A2) is negative, which approximately happens half the time in large samples. Hence, we need to consider the partially one-sided test (14) to obtain a test equivalent to the LR test. Furthermore, a standard central limit theorem for martingale difference sequences can be used to show that (A2) and (A3) will be asymptotically independent under normality.

Finally, given the root T consistency of $\tilde{\boldsymbol{\theta}}_T$, the second part of the proposition follows from the same arguments as in Proposition 3.

Proposition 7

It is straightforward to check that the scores of the log of (15) with respect to $\boldsymbol{\theta}$ and η evaluated at $\eta = 0$ for fixed $\mathbf{b}$ are equal to the corresponding ones of the GH

distribution. Based on this result, we can use the same procedure followed for the GH distribution to obtain the LM and KT tests for this distribution.

Proposition 8

The proof is straightforward if we rely on the results in the appendix of Fiorentini and Sentana (2007), who indicate that when $\boldsymbol{\varepsilon}_t^*$ is distributed as a standardised multivariate Student t with $1/\eta_0$ degrees of freedom, it can be written as $\boldsymbol{\varepsilon}_t^* = \sqrt{(1-2\eta_0)\zeta_t/(\xi_t\eta_0)}\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_t is a chi-square random variable with N degrees of freedom, ξ_t is a gamma variate with mean η_0^{-1} and variance $2\eta_0^{-1}$, and the three variates are mutually independent. These authors also exploit the fact that $X = \zeta_t/(\zeta_t + \xi_t)$ has a beta distribution with parameters $a = N/2$ and $b = 1/(2\eta_0)$ to show that

$$
\begin{aligned}
E\left[X^p\left(1-X\right)^q\right] &= \frac{B\left(a+p, b+q\right)}{B\left(a, b\right)}, \\
E\left[X^p\left(1-X\right)^q \log\left(1-X\right)\right] &= \frac{B\left(a+p, b+q\right)}{B\left(a, b\right)}\left[\psi\left(b+q\right) - \psi\left(a+b+p+q\right)\right],
\end{aligned}
$$

where $\psi\left(\cdot\right)$ is the digamma function and $B\left(\cdot, \cdot\right)$ the usual beta function. $\qquad\square$

Propositions 9 and 10

We can use standard central limit theory for martingale difference sequences to show the asymptotic joint normality of

$$
\frac{\sqrt{T}}{T}\sum_t \begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{bmatrix} \xrightarrow{d} N\left[0, E\left\{V_{t-1}\begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{bmatrix}\right\}\right],
$$

where

$$
V_{t-1}\begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{I}_{\boldsymbol{\pi}\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{M}_t(\boldsymbol{\pi}_0) \\ \mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & V_{t-1}\left[s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] & 0 \\ \mathcal{M}'_t(\boldsymbol{\pi}_0) & 0' & V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] \end{bmatrix} \tag{A4}
$$

under the null hypothesis of Student t innovations. In addition, we can again exploit the results of Fiorentini and Sentana (2007) mentioned in the proof of Proposition 8 to obtain the expressions for the elements of (A4). Parts 1 of the two propositions follow immediately from the $(N+1)\times(N+1)$ submatrix of (A4) that yield the variances of the

test moment conditions. To account for parameter uncertainty, consider the function

$$g_{2t}\left(\bar{\boldsymbol{\pi}}_T\right) = \left[\begin{array}{c} s_{\mathbf{b}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \end{array}\right] - \left[\begin{array}{c} \mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ \mathcal{M}'(\boldsymbol{\pi}_0) \end{array}\right] \mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) s_{\boldsymbol{\pi}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)$$

$$= \left[\begin{array}{ccc} -\mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{I}_N & \mathbf{0} \\ -\mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{0}' & 1 \end{array}\right] \left[\begin{array}{c} s_{\boldsymbol{\pi}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\mathbf{b}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \end{array}\right] = \mathcal{A}_2(\boldsymbol{\pi}_0) \left[\begin{array}{c} s_{\boldsymbol{\pi}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\mathbf{b}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \end{array}\right].$$

Using the root T consistency of $\bar{\boldsymbol{\pi}}_T$, we can now derive the required asymptotic distribution by means of the usual Taylor expansion around the true values of the parameters

$$\frac{\sqrt{T}}{T}\sum_t g_{2t}\left(\bar{\boldsymbol{\pi}}_T\right) = \frac{\sqrt{T}}{T}\sum_t \left[\begin{array}{c} s_{\mathbf{b}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \end{array}\right] = \mathcal{A}_2(\boldsymbol{\pi}_0)\frac{\sqrt{T}}{T}\sum_t \left[\begin{array}{c} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{array}\right]$$

$$+ \mathcal{A}_2(\boldsymbol{\pi}_0)E\left[\frac{\partial}{\partial\boldsymbol{\pi}'}\left(\begin{array}{c} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{array}\right)\right]\sqrt{T}\left(\bar{\boldsymbol{\pi}}_T - \boldsymbol{\pi}_0\right) + o_p\left(1\right),$$

where we have used the fact that $\sum_t s_{\boldsymbol{\pi}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) = 0$. It can be tediously shown by means of the Barlett identities that

$$E\left[\frac{\partial}{\partial\boldsymbol{\pi}'}\left(\begin{array}{c} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{array}\right)\right] = -\left(\begin{array}{c} \mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ \mathcal{I}'_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ \mathcal{M}'(\boldsymbol{\pi}_0) \end{array}\right).$$

Hence

$$\mathcal{A}_2(\boldsymbol{\pi}_0)E\left[\frac{\partial}{\partial\boldsymbol{\pi}'}\left(\begin{array}{c} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{array}\right)\right] = \mathbf{0}.$$

As a result

$$\frac{\sqrt{T}}{T}\sum_t \left[\begin{array}{c} s_{\mathbf{b}t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \\ s_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right) \end{array}\right] = \mathcal{A}_2(\boldsymbol{\pi}_0)\frac{\sqrt{T}}{T}\sum_t \left[\begin{array}{c} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{array}\right],$$

from which we can obtain the asymptotic distributions in the Propositions. $\qquad\square$

B The Generalised Hyperbolic distribution

B.1 The density function

If the mixing variable ξ_t appearing in (2) follows a $GIG\left(-\nu, \gamma, \delta\right)$ distribution, then the density of the $N \times 1$ GH random vector $\boldsymbol{\varepsilon}^*$ will be given by

$$f_{GH}(\boldsymbol{\varepsilon}^*) = \frac{\left(\frac{\gamma}{\delta}\right)^\nu}{(2\pi)^{\frac{N}{2}}\left[\boldsymbol{\beta}'\boldsymbol{\Upsilon}\boldsymbol{\beta} + \gamma^2\right]^{\nu-\frac{N}{2}}|\boldsymbol{\Upsilon}|^{\frac{1}{2}}K_\nu\left(\delta\gamma\right)}\left\{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Upsilon}\boldsymbol{\beta} + \gamma^2}\delta q\left[\delta^{-1}(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha})\right]\right\}^{\nu-\frac{N}{2}}$$

$$\times K_{\nu-\frac{N}{2}}\left\{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Upsilon}\boldsymbol{\beta} + \gamma^2}\delta q\left[\delta^{-1}(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha})\right]\right\}\exp\left[\boldsymbol{\beta}'\left(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha}\right)\right],$$

where $-\infty < \nu < \infty$, $\gamma > 0$, $q\left[\delta^{-1}(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha})\right] = \sqrt{1 + \delta^{-2}(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha})'\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\varepsilon}^* - \boldsymbol{\alpha})}$ and $K_\nu(\cdot)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun, 1965, p. 374, as well as appendix B.3).

Given that δ and $\boldsymbol{\Upsilon}$ are not separately identified, Barndorff-Nielsen and Shephard (2001) set the determinant of $\boldsymbol{\Upsilon}$ equal to 1. However, it is more convenient to set $\delta = 1$ instead in order to reparametrise the GH distribution so that it has mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_N$.

One of the most attractive properties of the GH distribution is that it contains as particular cases several of the most important multivariate distributions already used in the literature. For the sake of concreteness, consider the case in which $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ are such that $\boldsymbol{\varepsilon}^*$ is a standardised random vector. Then, the most important particular cases are:

• **Normal**, which can be achieved in three different ways: (i) when $\nu \to -\infty$ or (ii) $\nu \to +\infty$, regardless of the values of γ and $\boldsymbol{\beta}$; and (iii) when $\gamma \to \infty$ irrespective of the values of ν and $\boldsymbol{\beta}$.

• **Symmetric Student** t, obtained when $-\infty < \nu < -2$, $\gamma = 0$ and $\boldsymbol{\beta} = \mathbf{0}$.

• **Asymmetric Student** t, which is like its symmetric counterpart except that the vector $\boldsymbol{\beta}$ of skewness parameters is no longer zero.

• **Asymmetric Normal-Gamma**, associated to the **Variance Gamma** process, which is obtained when $\gamma = 0$ and $0 < \nu < \infty$ (see Madan and Milne, 1991).

• **Normal Inverse Gaussian**, for $\nu = -.5$ (see Aas, Dimakos, and Haff, 2005).

• **Hyperbolic**, for $\nu = 1$ (see Chen, Härdle, and Jeong, 2004)

• **Asymmetric Laplace**, for $\nu = 1$ and $\gamma = 0$ (see Cajigas and Urga, 2007).

B.2   The score function

We can use EM algorithm - type arguments to obtain analytical formulae for the score function $s_t(\boldsymbol{\phi}) = \partial l\left(\mathbf{y}_t|I_{t-1};\boldsymbol{\phi}\right)/\partial\boldsymbol{\phi}$. The idea is based on the following dual decomposition of the joint log-density (given I_{t-1} and $\boldsymbol{\phi}$) of the observable process $\mathbf{y}_t$ and the latent mixing process ξ_t:

$$
\begin{aligned}
l\left(\mathbf{y}_t, \xi_t | I_{t-1}; \boldsymbol{\phi}\right) &\equiv l\left(\mathbf{y}_t | \xi_t, I_{t-1}; \boldsymbol{\phi}\right) + l\left(\xi_t | I_{t-1}; \boldsymbol{\phi}\right) \\
&\equiv l\left(\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi}\right) + l\left(\xi_t | \mathbf{y}_t, I_{t-1}; \boldsymbol{\phi}\right),
\end{aligned}
$$

where $l\left(\mathbf{y}_t | \xi_t, I_{t-1}; \boldsymbol{\phi}\right)$ is the conditional log-likelihood of $\mathbf{y}_t$ given ξ_t, I_{t-1} and $\boldsymbol{\phi}$; $l\left(\xi_t | \mathbf{y}_t, I_{t-1}; \boldsymbol{\phi}\right)$ is the conditional log-likelihood of ξ_t given $\mathbf{y}_t$, I_{t-1} and $\boldsymbol{\phi}$; and finally

$l\left(\mathbf{y}_t|I_{t-1};\boldsymbol{\phi}\right)$ and $l\left(\xi_t|I_{t-1};\boldsymbol{\phi}\right)$ are the marginal log-densities (given I_{t-1} and $\boldsymbol{\phi}$) of the observable and unobservable processes, respectively. If we differentiate both sides of the previous identity with respect to $\boldsymbol{\phi}$, and take expectations given the full observed sample, I_T, then we will end up with:

$$s_t(\boldsymbol{\phi}) = E\left(\left.\frac{\partial l\left(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}}\right| I_T;\boldsymbol{\phi}\right) + E\left(\left.\frac{\partial l\left(\xi_t|I_{t-1};\boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}}\right| I_T;\boldsymbol{\phi}\right) \qquad \text{(C1)}$$

because $E\left[\partial l\left(\xi_t|\mathbf{y}_t, I_{t-1};\boldsymbol{\phi}\right)/\partial \boldsymbol{\phi}\right| I_T;\boldsymbol{\phi}] = \mathbf{0}$ by virtue of the Kullback inequality. Thus, the EM-type procedure that we follow is divided in two parts. In the maximisation step, we derive $l\left(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi}\right)$ and $l\left(\xi_t|I_{t-1};\boldsymbol{\phi}\right)$ with respect to $\boldsymbol{\phi}$. Then, in the expectation step, we take the expected value of these derivatives given $I_T = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T\}$ and the parameter values.

Conditional on ξ_t, $\mathbf{y}_t$ is the following multivariate normal:

$$\mathbf{y}_t|\xi_t, I_{t-1} \sim N\left[\boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t(\boldsymbol{\theta})c_t(\boldsymbol{\phi})\mathbf{b}\left[\frac{\gamma}{R_\nu(\gamma)}\frac{1}{\xi_t} - 1\right], \frac{\gamma}{R_\nu(\gamma)}\frac{1}{\xi_t}\boldsymbol{\Sigma}_t^*(\boldsymbol{\phi})\right],$$

where $c_t(\boldsymbol{\phi}) = c[\boldsymbol{\Sigma}_t^{\frac{1}{2}'}(\boldsymbol{\theta})\mathbf{b}, \nu, \gamma]$ and

$$\boldsymbol{\Sigma}_t^*(\boldsymbol{\phi}) = \boldsymbol{\Sigma}_t(\theta) + \frac{c_t(\boldsymbol{\phi}) - 1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{bb}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$$

If we define $\mathbf{p}_t = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}) + c_t(\boldsymbol{\phi})\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$, then we have the following log-density

$$\begin{aligned}
l\left(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi}\right) &= \frac{N}{2}\log\left[\frac{\xi_t R_\nu(\gamma)}{2\pi\gamma}\right] - \frac{1}{2}\log|\boldsymbol{\Sigma}_t^*(\boldsymbol{\phi})| - \frac{\xi_t}{2}\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\mathbf{p}_t \\
&\quad + \mathbf{b}'\mathbf{p}_t - \frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{2\xi_t}\frac{\gamma c_t(\boldsymbol{\phi})}{R_\nu(\gamma)}.
\end{aligned}$$

Similarly, ξ_t is distributed as a GIG with parameters $\xi_t|I_{t-1} \sim GIG\left(-\nu, \gamma, 1\right)$, with a log-likelihood given by

$$l\left(\xi_t|I_{t-1};\boldsymbol{\phi}\right) = \nu\log\gamma - \log 2 - \log K_\nu(\gamma) - (\nu+1)\log\xi_t - \frac{1}{2}\left(\xi_t + \gamma^2\frac{1}{\xi_t}\right).$$

In order to determine the distribution of ξ_t given all the observable information I_T, we can exploit the serial independence of ξ_t given $I_{t-1};\boldsymbol{\phi}$ to show that

$$f\left(\xi_t|I_T;\boldsymbol{\phi}\right) = \frac{f\left(\mathbf{y}_t, \xi_t|I_{t-1};\boldsymbol{\phi}\right)}{f\left(\mathbf{y}_t|I_{t-1};\boldsymbol{\phi}\right)} \propto f\left(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi}\right)f\left(\xi_t|I_{t-1};\boldsymbol{\phi}\right)$$

$$\propto \xi_t^{\frac{N}{2}-\nu-1} \times \exp\left\{\frac{-1}{2}\left[\left(\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\mathbf{p}_t + 1\right)\xi_t + \left(\frac{\gamma c_t(\boldsymbol{\phi})}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} + \gamma^2\right)\frac{1}{\xi_t}\right]\right\},$$

which implies that

$$\xi_t | I_T; \phi \sim GIG\left(\frac{N}{2} - \nu, \sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2}, \sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\phi) \mathbf{p}_t + 1}\right).$$

From here, we can use (C5) and (C6) to obtain the required moments. Specifically,

$$E(\xi_t | I_T; \phi) = \frac{\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2}}{\sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}}$$

$$\times R_{\frac{N}{2}-\nu}\left[\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2} \sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}\right],$$

$$E\left(\frac{1}{\xi_t}\Bigg| I_T; \phi\right) = \frac{\sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}}{\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2}}$$

$$\times \frac{1}{R_{\frac{N}{2}-\nu-1}\left[\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2} \sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}\right]},$$

$$E(\log \xi_t | I_T; \phi) = \log\left(\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2}\right) - \log\left(\sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}\right)$$

$$+ \frac{\partial}{\partial x} \log K_x \left[\sqrt{\frac{\gamma c_t(\phi)}{R_\nu(\gamma)} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} + \gamma^2} \sqrt{\frac{R_\nu(\gamma)}{\gamma} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1} \mathbf{p}_t + 1}\right]\Bigg|_{x=\frac{N}{2}-\nu}.$$

If we put all the pieces together, we will finally have that

$$\frac{\partial l(\mathbf{y}_t | I_{t-1}; \phi)}{\partial \boldsymbol{\theta}'} = -\frac{1}{2} vec'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} - f(I_T, \phi) \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\phi) \frac{\partial \mathbf{p}_t}{\partial \boldsymbol{\theta}'}$$

$$-\frac{1}{2} \frac{c_t(\phi) - 1}{c_t(\phi) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} \sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} vec'(\mathbf{b}\mathbf{b}') \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \mathbf{b}' \frac{\partial \mathbf{p}_t}{\partial \boldsymbol{\theta}'}$$

$$+\frac{1}{2} f(I_T, \phi)[\mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\phi) \otimes \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\phi)] \frac{\partial vec[\boldsymbol{\Sigma}_t^*(\phi)]}{\partial \boldsymbol{\theta}'}$$

$$-\frac{1}{2} \frac{g(I_T, \phi)}{\sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} vec'(\mathbf{b}\mathbf{b}') \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'},$$

$$\frac{\partial l(\mathbf{y}_t | I_{t-1}; \phi)}{\partial \mathbf{b}} = -\frac{c_t(\phi) - 1}{c_t(\phi) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} \sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$$

$$-f(I_T, \phi) c_t(\phi) \mathbf{p}'_t + \varepsilon'_t + f(I_T, \phi) \frac{c_t(\phi) - 1}{\mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}} (\mathbf{b}' \mathbf{p}_t)$$

$$\times \left\{\frac{[c_t(\phi) - 1](\mathbf{b}' \mathbf{p}_t)}{c_t^2(\phi) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} \sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right.$$

$$\left.+\frac{\mathbf{p}'_t}{c_t(\phi)} - \frac{1}{\sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right\}$$

$$+\frac{[2 - g(I_T, \phi)]}{\sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b}}} \mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}),$$

$$\frac{\partial l\left(\mathbf{y}_t|\, I_{t-1};\boldsymbol{\phi}\right)}{\partial \eta} = \frac{N}{2}\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \eta} + \left(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{1}{2c_t(\boldsymbol{\phi})}\right)\frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} + \frac{\log\left(\gamma\right)}{2\eta^2}$$

$$-\frac{\partial \log K_\nu\left(\gamma\right)}{\partial \eta} - \frac{1}{2\eta^2}E\left[\log \xi_t|Y_T;\boldsymbol{\phi}\right] - \frac{f\left(I_T,\boldsymbol{\phi}\right)}{2}\left\{\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \eta}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\mathbf{p}_t\right.$$

$$\left.+\frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{\left(\mathbf{b}'\boldsymbol{\varepsilon}_t\right)^2}{c_t^2(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\right]\right\}$$

$$-\frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{2}g\left(I_T,\boldsymbol{\phi}\right)\left\{\frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} - c_t(\boldsymbol{\phi})\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \eta}\right\},$$

and

$$\frac{\partial l\left(\mathbf{y}_t|\, I_{t-1};\boldsymbol{\phi}\right)}{\partial \psi} = \frac{N}{2}\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \psi} + \frac{N}{2\psi\left(1-\psi\right)} + \left(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{1}{2c_t(\boldsymbol{\phi})}\right)\frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi}$$

$$+\frac{1}{2\eta\psi\left(1-\psi\right)} - \frac{\partial \log K_\nu\left(\gamma\right)}{\partial \psi} - \frac{f\left(I_T,\boldsymbol{\phi}\right)}{2}\left\{\left[\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \psi} + \frac{1}{\psi\left(1-\psi\right)}\right]\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\mathbf{p}_t\right.$$

$$\left.+\frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{\left(\mathbf{b}'\boldsymbol{\varepsilon}_t\right)^2}{c_t^2(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\right]\right\}$$

$$-\frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{2}g\left(I_T,\boldsymbol{\phi}\right)\left\{-\frac{c_t(\boldsymbol{\phi})}{\psi\left(1-\psi\right)} + \frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi} - c_t(\boldsymbol{\phi})\frac{\partial \log R_\nu\left(\gamma\right)}{\partial \psi}\right\} + g\left(I_T,\boldsymbol{\phi}\right)\frac{R_\nu\left(\gamma\right)}{\psi^2},$$

where

$$f\left(I_T,\boldsymbol{\phi}\right) = \gamma^{-1}R_\nu\left(\gamma\right)E\left(\xi_t|I_T;\boldsymbol{\phi}\right),$$

$$g\left(I_T,\boldsymbol{\phi}\right) = \gamma R_\nu^{-1}\left(\gamma\right)E\left(\xi_t^{-1}|I_T;\boldsymbol{\phi}\right),$$

$$\frac{\partial vec[\boldsymbol{\Sigma}_t^*(\boldsymbol{\phi})]}{\partial \boldsymbol{\theta}'} = \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \frac{c_t(\boldsymbol{\phi})-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\left\{\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}' \otimes I_N\right] + \left[I_N \otimes \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\right]\right\}\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}$$

$$+\frac{c_t(\boldsymbol{\phi})-1}{[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}]^2}\left\{\frac{1}{\sqrt{1 + 4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}} - 1\right\}$$

$$\times vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]vec'\left(\mathbf{b}\mathbf{b}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'},$$

$$\frac{\partial \mathbf{p}_t}{\partial \boldsymbol{\theta}'} = -\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + c_t(\boldsymbol{\phi})\left[\mathbf{b}' \otimes I_N\right]\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}$$

$$+\frac{c_t(\boldsymbol{\phi})-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\frac{1}{\sqrt{1 + 4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}vec'\left(\mathbf{b}\mathbf{b}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'},$$

$$\frac{\partial c_t(\boldsymbol{\phi})}{\partial\left(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right)} = \frac{c_t(\boldsymbol{\phi})-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\frac{1}{\sqrt{1 + 4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}},$$

$$\frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} = \frac{c_t(\boldsymbol{\phi})-1}{\left[D_{\nu+1}\left(\gamma\right)-1\right]\sqrt{1 + 4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\frac{\partial D_{\nu+1}\left(\gamma\right)}{\partial \eta},$$

and

$$\frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi} = \frac{c_t(\boldsymbol{\phi})-1}{\left[D_{\nu+1}\left(\gamma\right)-1\right]\sqrt{1 + 4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\frac{\partial D_{\nu+1}\left(\gamma\right)}{\partial \psi}.$$

B.3 Modified Bessel function of the third kind

The modified Bessel function of the third kind with order ν, which we denote as $K_\nu(\cdot)$, is closely related to the modified Bessel function of the first kind $I_\nu(\cdot)$, as

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\pi\nu)}. \tag{C2}$$

Some basic properties of $K_\nu(\cdot)$, taken from Abramowitz and Stegun (1965), are $K_\nu(x) = K_{-\nu}(x)$, $K_{\nu+1}(x) = 2\nu x^{-1} K_\nu(x) + K_{\nu-1}(x)$, and $\partial K_\nu(x)/\partial x = -\nu x^{-1} K_\nu(x) - K_{\nu-1}(x)$. For small values of the argument x, and ν fixed, it holds that

$$K_\nu(x) \simeq \frac{1}{2}\Gamma(\nu)\left(\frac{1}{2}x\right)^{-\nu}.$$

Similarly, for ν fixed, $|x|$ large and $m = 4\nu^2$, the following asymptotic expansion is valid

$$K_\nu(x) \simeq \sqrt{\frac{\pi}{2x}} e^{-x} \left\{ 1 + \frac{m\text{-}1}{8x} + \frac{(m\text{-}1)(m\text{-}9)}{2!(8x)^2} + \frac{(m\text{-}1)(m\text{-}9)(m\text{-}25)}{3!(8x)^3} + \cdots \right\}. \tag{C3}$$

Finally, for large values of x and ν we have that

$$K_\nu(x) \simeq \sqrt{\frac{\pi}{2\nu}} \frac{\exp(-\nu l^{-1})}{l^{-2}} \left[\frac{(x/\nu)}{1+l^{-1}}\right]^{-\nu} \left[1 - \frac{3l\text{-}5l^3}{24\nu} + \frac{81l^2\text{-}462l^4+385l^6}{1152\nu^2} + \cdots\right], \tag{C4}$$

where $\nu > 0$ and $l = \left[1 + (x/\nu)^2\right]^{-\frac{1}{2}}$. Although the existing literature does not discuss how to obtain numerically reliable derivatives of $K_\nu(x)$ with respect to its order, our experience suggests the following conclusions:

• For $\nu \leq 10$ and $|x| > 12$, the derivative of (C3) with respect to ν gives a better approximation than the direct derivative of $K_\nu(x)$, which is in fact very unstable.

• For $\nu > 10$, the derivative of (C4) with respect to ν works better than the direct derivative of $K_\nu(x)$.

• Otherwise, the direct derivative of the original function works well.

We can express such a derivative as a function of $I_\nu(x)$ by using (C2) as:

$$\frac{\partial K_\nu(x)}{\partial \nu} = \frac{\pi}{2\sin(\nu\pi)} \left[\frac{\partial I_{-\nu}(x)}{\partial \nu} - \frac{\partial I_\nu(x)}{\partial \nu}\right] - \pi\cot(\nu\pi) K_\nu(x)$$

However, this formula becomes numerically unstable when ν is near any non-negative integer $n = 0, 1, 2, \cdots$ due to the sine that appears in the denominator. In our experience, it is much better to use the following Taylor expansion for small $|\nu - n|$:

$$\begin{aligned}
\frac{\partial K_\nu(x)}{\partial \nu} &= \left.\frac{\partial K_\nu(x)}{\partial \nu}\right|_{\nu=n} + \left.\frac{\partial^2 K_\nu(x)}{\partial \nu^2}\right|_{\nu=n}(\nu - n) \\
&+ \left.\frac{\partial^3 K_\nu(x)}{\partial \nu^3}\right|_{\nu=n}(\nu - n)^2 + \left.\frac{\partial^4 K_\nu(x)}{\partial \nu^4}\right|_{\nu=n}(\nu - n)^3,
\end{aligned}$$

where for *integer* ν:

$$\frac{\partial K_\nu(x)}{\partial \nu} = \frac{1}{4\cos(\pi n)}\left[\frac{\partial^2 I_{-\nu}(x)}{\partial \nu^2} - \frac{\partial^2 I_\nu(x)}{\partial \nu^2}\right] + \pi^2\left[I_{-\nu}(x) - I_\nu(x)\right],$$

$$\frac{\partial^2 K_\nu(x)}{\partial \nu^2} = \frac{1}{6\cos(\pi n)}\left[\frac{\partial^3 I_{-\nu}(x)}{\partial \nu^3} - \frac{\partial^3 I_\nu(x)}{\partial \nu^3}\right] + \frac{\pi^2}{3\cos(\pi n)}\left[\frac{\partial I_{-\nu}(x)}{\partial \nu} - \frac{\partial I_\nu(x)}{\partial \nu}\right] - \frac{\pi^2}{3}K_n(x),$$

$$\frac{\partial^3 K_\nu(x)}{\partial \nu^3} = \frac{1}{8\cos(\pi n)}\left\{\left[\frac{\partial^4 I_{-\nu}(x)}{\partial \nu^4} - \frac{\partial^4 I_\nu(x)}{\partial \nu^4}\right]\right.$$
$$\left. -4\pi^2\left[\frac{\partial^2 I_{-\nu}(x)}{\partial \nu^2} - \frac{\partial^2 I_\nu(x)}{\partial \nu^2}\right] - 12\pi^4\left[I_{-\nu}(x) - I_\nu(x)\right]\right\} + 3\pi^2\frac{\partial K_n(x)}{\partial \nu},$$

and

$$\frac{\partial^4}{\partial \nu^4}K_\nu(x) = \frac{1}{8\cos(\pi n)}\left\{\frac{3}{2}\left[\frac{\partial^5 I_{-\nu}(x)}{\partial \nu^5} - \frac{\partial^5 I_\nu(x)}{\partial \nu^5}\right]\right.$$
$$\left. -10\pi^2\left[\frac{\partial^3 I_{-\nu}(x)}{\partial \nu^3} - \frac{\partial^3 I_\nu(x)}{\partial \nu^3}\right] - 4\pi^4\left[\frac{\partial I_{-\nu}(x)}{\partial \nu} - \frac{\partial I_\nu(x)}{\partial \nu}\right]\right\} + 6\pi^2\frac{\partial^2 K_n(x)}{\partial \nu^2} - \pi^4 K_n(x).$$

Let $\psi^{(i)}(\cdot)$ denote the polygamma function (see Abramowitz and Stegun, 1965). The first five derivatives of $I_\nu(x)$ for any real ν are as follows:

$$\frac{\partial I_\nu(x)}{\partial \nu} = I_\nu(x)\log\left(\frac{x}{2}\right) - \left(\frac{x}{2}\right)^\nu\sum_{k=0}^{\infty}\frac{Q_1(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k,$$

where

$$Q_1(z) = \begin{cases} \psi(z)/\Gamma(z) \text{ if } z > 0 \\ \pi^{-1}\Gamma(1-z)\left[\psi(1-z)\sin(\pi z) - \pi\cos(\pi z)\right] \text{ if } z \leq 0 \end{cases}$$

$$\frac{\partial^2 I_\nu(x)}{\partial \nu^2} = 2\log\left(\frac{x}{2}\right)\frac{\partial I_\nu(x)}{\partial \nu} - I_\nu(x)\left[\log\left(\frac{x}{2}\right)\right]^2 - \left(\frac{x}{2}\right)^\nu\sum_{k=0}^{\infty}\frac{Q_2(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k,$$

where

$$Q_2(z) = \begin{cases} \left[\psi'(z) - \psi^2(z)\right]/\Gamma(z) \text{ if } z > 0 \\ \pi^{-1}\Gamma(1-z)\left[\pi^2 - \psi'(1-z) - \left[\psi(1-z)\right]^2\right]\sin(\pi z) \\ +2\Gamma(1-z)\psi(1-z)\cos(\pi z) \text{ if } z \leq 0 \end{cases}$$

$$\frac{\partial^3 I_\nu(x)}{\partial \nu^3} = 3\log\left(\frac{x}{2}\right)\frac{\partial^2 I_\nu(x)}{\partial \nu^2} - 3\left[\log\left(\frac{x}{2}\right)\right]^2\frac{\partial I_\nu(x)}{\partial \nu} + \left[\log\left(\frac{x}{2}\right)\right]^3 I_\nu(x)$$
$$- \left(\frac{x}{2}\right)^\nu\sum_{k=0}^{\infty}\frac{Q_3(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k,$$

where

$$Q_3(z) = \begin{cases} \left[\psi^3(z) - 3\psi(z)\psi'(z) + \psi''(z)\right]/\Gamma(z) \text{ if } z > 0 \\ \pi^{-1}\Gamma(1-z)\left\{\psi^3(1-z) - 3\psi(1-z)\left[\pi^2 - \psi'(1-z)\right] + \psi''(1-z)\right\}\sin(\pi z) \\ +\Gamma(1-z)\left\{\pi^2 - 3\left[\psi^2(1-z) + \psi'(1-z)\right]\right\}\cos(\pi z) \text{ if } z \leq 0 \end{cases}$$

$$\frac{\partial^4 I_\nu(x)}{\partial \nu^4} = 4\log\left(\frac{x}{2}\right)\frac{\partial^3 I_\nu(x)}{\partial \nu^3} - 6\left[\log\left(\frac{x}{2}\right)\right]^2\frac{\partial^2 I_\nu(x)}{\partial \nu^2} + 4\left[\log\left(\frac{x}{2}\right)\right]^3\frac{\partial I_\nu(x)}{\partial \nu}$$

$$- \left[\log\left(\frac{x}{2}\right)\right]^4 I_\nu(x) - \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{Q_4(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k,$$

where

$$Q_4(z) = \begin{cases} \left[-\psi^4(z) + 6\psi^2(z)\psi'(z) - 4\psi(z)\psi''(z) - 3\left[\psi'(z)\right]^2 + \psi'''(z)\right]/\Gamma(z) \text{ if } z > 0 \\ \pi^{-1}\Gamma(1-z)\{-\psi^4(1-z) + 6\pi^2\psi^2(1-z) - 6\psi^2(1-z)\psi'(1-z) \\ -4\psi(1-z)\psi''(1-z) - 3\left[\psi'(1-z)\right]^2 + 6\pi^2\psi'(1-z) \\ -\psi'''(1-z) - \pi^4\} \sin(\pi z) + \Gamma(1-z)\, 4\psi^3(1-z) - 4\pi^2\psi(1-z) \\ +12\psi(1-z)\psi'(1-z) + 4\psi''(1-z)\cos(\pi z) \text{ if } z \leq 0 \end{cases}$$

and finally,

$$\frac{\partial^5 I_\nu(x)}{\partial \nu^5} = 5\log\left(\frac{x}{2}\right)\frac{\partial^4 I_\nu(x)}{\partial \nu^4} - 10\left[\log\left(\frac{x}{2}\right)\right]^2\frac{\partial^3 I_\nu(x)}{\partial \nu^3} + 10\left[\log\left(\frac{x}{2}\right)\right]^3\frac{\partial^2 I_\nu(x)}{\partial \nu^2}$$

$$-5\left[\log\left(\frac{x}{2}\right)\right]^4\frac{\partial I_\nu(x)}{\partial \nu} + \left[\log\left(\frac{x}{2}\right)\right]^5 I_\nu(x) - \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{Q_5(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k,$$

where

$$Q_5(z) = \begin{cases} \{\psi^5(z) - 10\psi^3(z)\psi'(z) + 10\psi^2(z)\psi''(z) + 15\psi(z)\left[\psi'(z)\right]^2 \\ -5\psi(z)\psi'''(z) - 10\psi'(z)\psi''(z) + \psi^{(iv)}(z)\}/\Gamma(z) \text{ if } z > 0 \\ \pi^{-1}\Gamma(1-z)f_a(z)\sin(\pi z) + \Gamma(1-z)f_b(z)\cos(\pi z) \text{ if } z \leq 0 \end{cases}$$

with

$$f_a(z) = \psi^5(1-z) - 10\pi^2\psi^3(1-z) + 10\psi^3(1-z)\psi'(1-z) + 10\psi^2(1-z)\psi''(1-z)$$
$$+15\psi(1-z)\left[\psi'(1-z)\right]^2 + 5\psi(1-z)\psi'''(1-z) + 5\pi^4\psi(1-z)$$
$$-30\pi^2\psi(1-z)\psi'(1-z) + 10\psi'(1-z)\psi''(1-z) - 10\pi^2\psi''(1-z) + \psi^{(iv)}(1-z),$$

and

$$f_b(z) = -5\psi^4(1-z) + 10\pi^2\psi^2(1-z) - 30\psi^2(1-z)\psi'(1-z)$$
$$-20\psi(1-z)\psi''(1-z) - 15\left[\psi'(1-z)\right]^2 + 10\pi^2\psi'(1-z) - 5\psi'''(1-z) - \pi^4.$$

B.4 Moments of the GIG distribution

If $X \sim GIG(\nu, \delta, \gamma)$, its density function will be

$$\frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}x^{\nu-1}\exp\left[-\frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right],$$

where $K_\nu(\cdot)$ is the modified Bessel function of the third kind and $\delta, \gamma \geq 0$, $\nu \in \mathbb{R}$, $x > 0$. Two important properties of this distribution are $X^{-1} \sim GIG(-\nu, \gamma, \delta)$ and

$(\gamma/\delta)X \sim GIG\left(\nu, \sqrt{\gamma\delta}, \sqrt{\gamma\delta}\right)$. For our purposes, the most useful moments of X when $\delta\gamma > 0$ are

$$E\left(X^k\right) = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{\nu+k}\left(\delta\gamma\right)}{K_\nu\left(\delta\gamma\right)} \tag{C5}$$

$$E\left(\log X\right) = \log\left(\frac{\delta}{\gamma}\right) + \frac{\partial}{\partial\nu}K_\nu\left(\delta\gamma\right). \tag{C6}$$

The GIG nests some well-known important distributions, such as the gamma ($\nu > 0$, $\delta = 0$), the reciprocal gamma ($\nu < 0$, $\gamma = 0$) or the inverse Gaussian ($\nu = -1/2$). Importantly, all the moments of this distribution are finite, except in the reciprocal gamma case, in which (C5) becomes infinite for $k \geq |\nu|$. A complete discussion on this distribution can be found in Jørgensen (1982), who also presents several useful Gaussian approximations based on the following limits:

$$\sqrt{\delta\gamma}[(\gamma x/\delta) - 1] \overset{\delta\gamma\to\infty}{\to} N(0,1)$$

$$\sqrt{\delta\gamma}\log\left(\gamma x/\delta\right) \overset{\delta\gamma\to\infty}{\to} N(0,1)$$

$$\frac{\gamma^2}{2\sqrt{\nu}}\left[x - \frac{2\nu}{\gamma^2}\right] \overset{\nu\to+\infty}{\to} N(0,1)$$

$$\frac{-2\nu^{3/2}}{\delta^2}\left[x + \frac{\delta^2}{2\nu}\right] \overset{\nu\to-\infty}{\to} N(0,1)$$

C Skewness and kurtosis of GH distributions

We can tediously show that

$$E\left[vec\left(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\right)\boldsymbol{\varepsilon}^{*\prime}\right] = E\left[\left(\boldsymbol{\varepsilon}^* \otimes \boldsymbol{\varepsilon}^*\right)\boldsymbol{\varepsilon}^{*\prime}\right]$$
$$= c^3(\boldsymbol{\beta},\nu,\gamma)\left[\frac{K_{\nu+3}\left(\gamma\right)K_\nu^2\left(\gamma\right)}{K_{\nu+1}^3\left(\gamma\right)} - 3D_{\nu+1}\left(\gamma\right) + 2\right]vec\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)\boldsymbol{\beta}'$$
$$+c(\boldsymbol{\beta},\nu,\gamma)\left[D_{\nu+1}\left(\gamma\right)\text{-}1\right]\left(\mathbf{K}_{NN}\text{+}\mathbf{I}_{N^2}\right)\left(\boldsymbol{\beta}\otimes\mathbf{A}\right)\mathbf{A}'+c(\boldsymbol{\beta},\nu,\gamma)\left[D_{\nu+1}\left(\gamma\right)\text{-}1\right]vec\left(\mathbf{A}\mathbf{A}'\right)\boldsymbol{\beta}',$$

and

$$E\left[vec\left(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\right)vec'\left(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\right)\right] = E\left[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\otimes\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\right]$$
$$= c^4(\boldsymbol{\beta},\nu,\gamma)\left[\frac{K_{\nu+4}\left(\gamma\right)K_\nu^3\left(\gamma\right)}{K_{\nu+1}^4\left(\gamma\right)} - 4\frac{K_{\nu+3}\left(\gamma\right)K_\nu^2\left(\gamma\right)}{K_{\nu+1}^3\left(\gamma\right)} + 6D_{\nu+1}\left(\gamma\right) - 3\right]vec\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)vec'\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)$$
$$+c^2(\boldsymbol{\beta},\nu,\gamma)\left[\frac{K_{\nu+3}\left(\gamma\right)K_\nu^2\left(\gamma\right)}{K_{\nu+1}^3\left(\gamma\right)} - 2D_{\nu+1}\left(\gamma\right) + 1\right]$$
$$\times\left\{vec\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)vec'\left(\mathbf{A}\mathbf{A}'\right) + vec\left(\mathbf{A}\mathbf{A}'\right)vec'\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right) + \left(\mathbf{K}_{NN}\text{+}\mathbf{I}_{N^2}\right)\left[\boldsymbol{\beta}\boldsymbol{\beta}'\otimes\mathbf{A}\mathbf{A}'\right]\left(\mathbf{K}_{NN}\text{+}\mathbf{I}_{N^2}\right)\right\}$$
$$+D_{\nu+1}\left(\gamma\right)\left\{\left[\mathbf{A}\mathbf{A}'\otimes\mathbf{A}\mathbf{A}'\right]\left(\mathbf{K}_{NN} + \mathbf{I}_{N^2}\right) + vec\left(\mathbf{A}\mathbf{A}'\right)vec'\left(\mathbf{A}\mathbf{A}'\right)\right\},$$

where

$$\mathbf{A} = \left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}'\boldsymbol{\beta}} \boldsymbol{\beta}\boldsymbol{\beta}' \right]^{\frac{1}{2}},$$

and $\mathbf{K}_{NN}$ is the commutation matrix (see Magnus and Neudecker, 1988). In this respect, note that Mardia's (1970) coefficient of multivariate excess kurtosis will be -1 plus the trace of the fourth moment above divided by $N(N+2)$.

Under symmetry, the distribution of the standardised residuals $\boldsymbol{\varepsilon}^*$ is clearly elliptical, as it can be written as $\boldsymbol{\varepsilon}^* = \sqrt{\zeta/\xi}\sqrt{\gamma/R_\nu(\gamma)}\mathbf{u}$, where $\zeta \sim \chi_N^2$ and $\xi^{-1} \sim GIG(\nu, 1, \gamma)$. This is confirmed by the fact that the third moment becomes 0, while

$$E\left[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime} \otimes \boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}\right] = D_{\nu+1}(\gamma)\left\{[\mathbf{I}_N \otimes \mathbf{I}_N](K_{NN} + \mathbf{I}_{N^2}) + vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\right\}.$$

In the symmetric case, therefore, the coefficient of multivariate excess kurtosis is simply $D_{\nu+1}(\gamma)$-1, which is always non-negative, but monotonically decreasing in γ and $|\nu|$.

D Power of the normality tests

We can determine the power of the sup test by rewriting it as a quadratic form in

$$\left[\begin{array}{cc} 2/[N(N+2)] & \mathbf{0}' \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}^{-1}/[2(N+2)] \end{array} \right]$$

evaluated at $\bar{m}_T(\tilde{\boldsymbol{\theta}}_T) = [\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T), \bar{m}'_{sT}(\tilde{\boldsymbol{\theta}}_T)]'$, where $\tilde{\boldsymbol{\theta}}_T$ must be interpreted as a PML estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}'_0, vech'(\boldsymbol{\Sigma}_0))'$ under the alternative of GH innovations. Hence, its asymptotic distribution will be given by the robust formulae provided by Bollerslev and Wooldridge (1992), which, in terms of the Gaussian score can be written as

$$\sqrt{T}\left[\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right] = \mathcal{A}^{-1}(\boldsymbol{\theta}_0)\sqrt{T}\bar{s}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0, 0, 0, \mathbf{0}) + o_p(1),$$

where

$$\mathcal{A}(\boldsymbol{\phi}_0) = \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} + \frac{1}{2} \frac{\partial vec'\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}} \left[\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right] \frac{\partial vec\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}}.$$

Hence, the usual Taylor expansion around the true parameter values yields

$$\sqrt{T}\bar{m}_T\left(\tilde{\boldsymbol{\theta}}_T\right) = \left[\begin{array}{cc} -\mathcal{B}(\boldsymbol{\theta}_0)\mathcal{A}^{-1}(\boldsymbol{\theta}_0) & \mathbf{I}_{N+1} \end{array} \right]\sqrt{T}\left[\begin{array}{c} \bar{s}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0, 0, 0, \mathbf{0}) \\ \bar{m}_T(\boldsymbol{\theta}_0) \end{array} \right] + o_p(1), \quad \text{(F1)}$$

where $\mathcal{B}(\boldsymbol{\theta}_0) = -E\left[\partial \bar{m}_T(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'\right]$

Fortunately, $\mathcal{A}(\boldsymbol{\phi}_0)$, $\mathcal{B}(\boldsymbol{\theta}_0)$, as well as the mean and variance of $\bar{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0)$ and $\bar{m}_T(\boldsymbol{\theta}_0)$ under the alternative can be computed analytically by using the location-scale mixture of normals interpretation of the GH distribution. In particular, we can write

$$\boldsymbol{\varepsilon}_t^* = c(\boldsymbol{\phi})\mathbf{b}(h_t - 1) + \sqrt{h_t}\mathbf{A}\mathbf{r}_t,$$

$$\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^* = c^2(\boldsymbol{\phi})\,(h_t - 1)^2\,\mathbf{b}'\mathbf{b} + 2c(\boldsymbol{\phi})\sqrt{h_t}\,(h_t - 1)\,\mathbf{b}'\mathbf{A}\mathbf{r}_t + h_t\mathbf{r}_t'\mathbf{A}'\mathbf{A}\mathbf{r}_t,$$

with $h_t = \xi_t^{-1}\gamma/R_\nu(\gamma)$, and

$$\mathbf{A} = \left[\mathbf{I}_N + \frac{c(\boldsymbol{\phi}, \nu, \gamma) - 1}{\mathbf{b}'\mathbf{b}}\mathbf{b}\mathbf{b}'\right]^{\frac{1}{2}},$$

where $\mathbf{r}_t|\mathbf{z}_t, I_{t-1} \sim N(0, \mathbf{I}_N)$ and $\xi_t|\mathbf{z}_t, I_{t-1} \sim GIG[.5\eta^{-1}, \psi^{-1}(1 - \psi), 1]$ are mutually independent. But since both ξ_t and $\mathbf{r}_t$ are *iid*, then $\boldsymbol{\varepsilon}_t^*$ and $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$ will also be *iid*. As a result, given that all the moments of normal and *GIG* random variables are finite (except when $\psi = 1$, in which case some moments may become unbounded for large enough η; see Jørgensen, 1982), we can apply the Lindeberg-Lévy Central Limit Theorem to show that the asymptotic distribution of $\sqrt{T}\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)$ is $N[m(\boldsymbol{\theta}_0, \eta, \psi, \mathbf{b}), V(\boldsymbol{\theta}_0, \eta, \psi, \mathbf{b})]$, where the required expressions can be computed from (F1). In particular, we can use Magnus (1986) to evaluate the moments of quadratic forms of normals, such as $\mathbf{r}_t'\mathbf{A}'\mathbf{A}\mathbf{r}_t$.

Finally, we can use Koerts and Abrahamse's (1969) implementation of Imhof's procedure for evaluating the probability that a quadratic form of normals is less than a given value (see also Farebrother, 1990).

To obtain the power of the KT test, we will use the following alternative formulation

$$\frac{KT}{T} = \frac{2}{N(N + 2)}\bar{m}_{kT}^2(\tilde{\boldsymbol{\theta}}_T) \cdot \mathbf{1}\big(\bar{m}_{kT}\big(\tilde{\boldsymbol{\theta}}_T\big) \geq 0\big) + \frac{1}{2(N + 2)}\bar{m}_{sT}'(\tilde{\boldsymbol{\theta}}_T)\hat{\boldsymbol{\Sigma}}^{-1}\bar{m}_{sT}(\tilde{\boldsymbol{\theta}}_T).$$

Hence, the distribution function of the KT statistic can be expressed as

$$\Pr\left(\frac{KT}{T} < x\right) = \int_{-\infty}^{\infty}\Pr\left(\frac{KT}{T} < x\,\bigg|\,\bar{m}_{kt} = l\right)f_k(l)\,dl, \qquad \text{(F2)}$$

where $f_k(\cdot)$ is the pdf of the distribution of the kurtosis component. But since the joint asymptotic distribution of $\sqrt{T}\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)$ is normal, so that the conditional distribution of $\sqrt{T}\bar{m}_{sT}(\tilde{\boldsymbol{\theta}}_T)$ given $\sqrt{T}\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T)$ will also be normal, the KT test can also be written as a quadratic form of normals for each value of the kurtosis component. As a result, we can use Imhof's procedure again to evaluate

$$\Pr\left[\frac{1}{2(N + 2)}\bar{m}_{sT}(\tilde{\boldsymbol{\theta}}_T)\hat{\boldsymbol{\Sigma}}^{-1}\bar{m}_{sT}(\tilde{\boldsymbol{\theta}}_T) < x - \frac{2}{N(N + 2)}l^2 \cdot \mathbf{1}(l \geq 0)\,\bigg|\,\bar{m}_{kt} = l\right]$$
$$= \Pr\left(\frac{KT}{T} < x\,\bigg|\,\bar{m}_{kt} = l\right).$$

Once we know this conditional probability, we can evaluate the integral in (F2) by numerical integration with a standard quadrature algorithm.

<div align="center">

Table 1

Maximum likelihood estimates of a conditionally heteroskedastic single factor model for
10 Datastream US sectoral stock indices. Daily excess returns.

</div>

	Gaussian		Student t		Asym. t		Sym. GH		Asym. GH	
		SE		SE		SE		SE		SE
η	0	-	0.095	0.003	0.095	0.004	0.095	0.003	0.095	0.004
ψ	0	-	1	-	1	-	1	-	1	-
Log-lik.	-53132.29		-52008.98		-51997.25		-52008.98		-51997.25	

<div align="center">

(b) Normality tests

</div>

		Test	p-value	
Score based			Asymptotic	Bootstrap
	Kurtosis	9289.32	0.000	0.000
	Skewness	204.34	0.000	0.000
	Sup-LM	9493.66	0.000	0.000
	Kuhn-Tucker	9493.66	0.000	0.000
LR			Asymptotic	Bootstrap
	H_1: sym. GH	2246.63	0.000	0.000
	H_1: asym. t	2270.09	0.000	0.000
	H_1: asym. GH	2270.09	0.000	0.000

<div align="center">

(c) Student t tests

</div>

		Test	p-value	
Score based			Asymptotic	Bootstrap
	Kurtosis	0.00	1.000	1.000
	Skewness	25.35	0.005	0.007
	Joint	25.35	0.006	0.007
LR			Asymptotic	Bootstrap
	H_1: sym. GH	0.00	1.000	1.000
	H_1: asym. t	23.45	0.012	0.010
	H_1: asym. GH	23.45	0.012	0.010

Note: The Student t score based Kurtosis test denotes the one-sided test of Student t vs. symmetric
GH innovations. In the LR tests, "H_1: sym. GH", "H_1: asym. t" and "H_1: asym. GH" indicate
whether the alternative hypothesis is symmetric GH, asymmetric t or asymmetric GH, respectively.
Bootstrapped p-values have been obtained from a parametric bootstrap with 1,000 samples, except
for the LR p-values, where only 100 samples have been considered.

Table 2
Maximum likelihood estimates of a conditionally heteroskedastic single factor model for 10 Datastream US sectoral stock indices. Weekly excess returns.

	Gaussian		Student t		Asym. t		Sym. GH		Asym. GH	
		SE		SE		SE		SE		SE
η	0	-	0.101	0.008	0.098	0.008	0.586	0.072	0.137	0.072
ψ	0	-	1	-	1	-	0.209	0.109	0.264	0.110
Log-lik.	-20052.46		-19857.74		-19849.08		-19856.39		-19848.91	

(b) Normality tests

		Test	p-value	
Score based			Asymptotic	Bootstrap
	Kurtosis	842.63	0.000	0.000
	Skewness	53.98	0.000	0.000
	Sup-LM	896.61	0.000	0.000
	Kuhn-Tucker	896.61	0.000	0.000

LR			Asymptotic	Bootstrap
	H_1: sym. GH	392.14	0.000	0.000
	H_1: asym. t	406.77	0.000	0.000
	H_1: asym. GH	407.10	0.000	0.000

(c) Student t tests

		Test	p-value	
Score based			Asymptotic	Bootstrap
	Kurtosis	1.00	0.158	0.124
	Skewness	11.21	0.341	0.315
	Joint	12.22	0.309	0.270

LR			Asymptotic	Bootstrap
	H_1: sym. GH	2.70	0.050	0.070
	H_1: asym. t	17.33	0.067	0.080
	H_1: asym. GH	17.67	0.075	0.090

Note: The Student t score based Kurtosis test denotes the one-sided test of Student t vs. symmetric GH innovations. In the LR tests, "H_1: sym. GH", "H_1: asym. t" and "H_1: asym. GH" indicate whether the alternative hypothesis is symmetric GH, asymmetric t or asymmetric GH, respectively. Bootstrapped p-values have been obtained from a parametric bootstrap with 1,000 samples, except for the LR p-values, where only 100 samples have been considered.

Figure 1a: Power of the normality tests under symmetric t alternatives
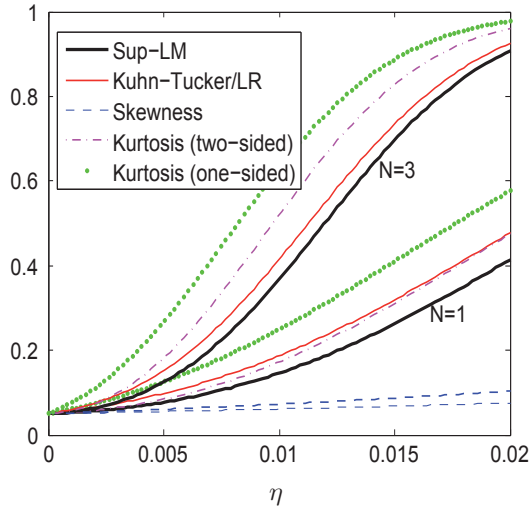
Figure 1b: Power of the normality tests under asymmetric t alternatives ($b_i = .75$, $\forall i$)
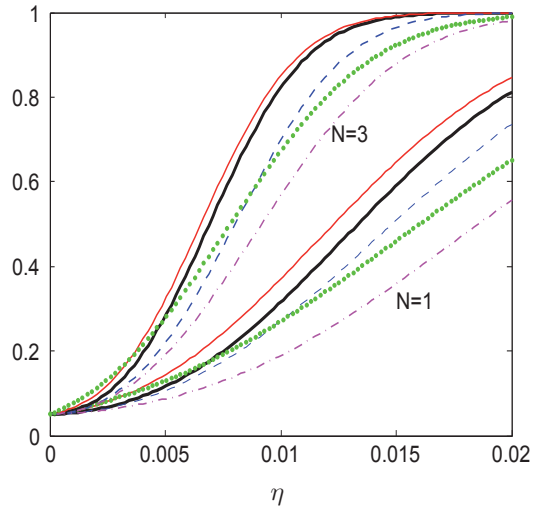
Figure 1c: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .005$, $N = 3$)
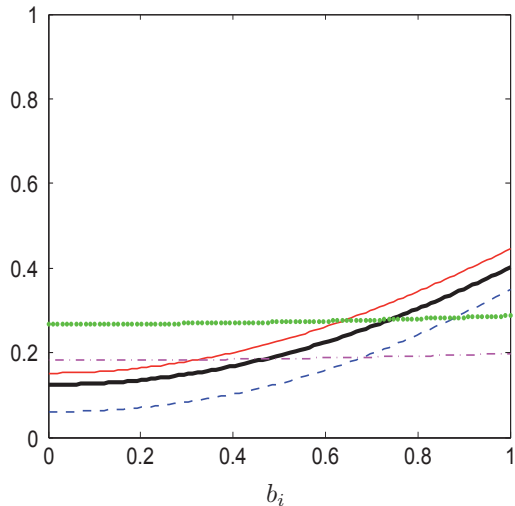
Figure 1d: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .01$, $N = 3$)
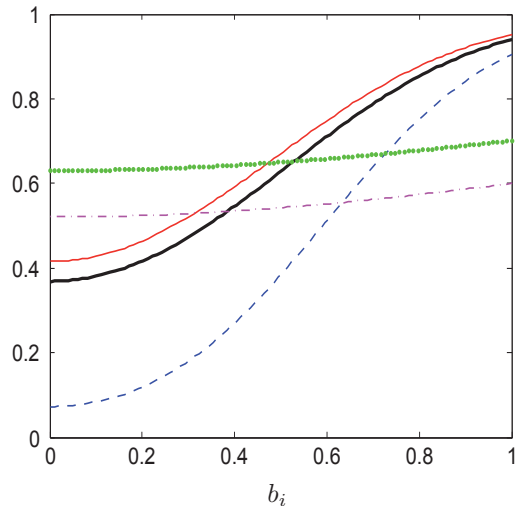
Figure 1e: Power of Sup-LM, Mardia and Lütkepohl normality tests against symmetric t alternatives ($N = 3$).
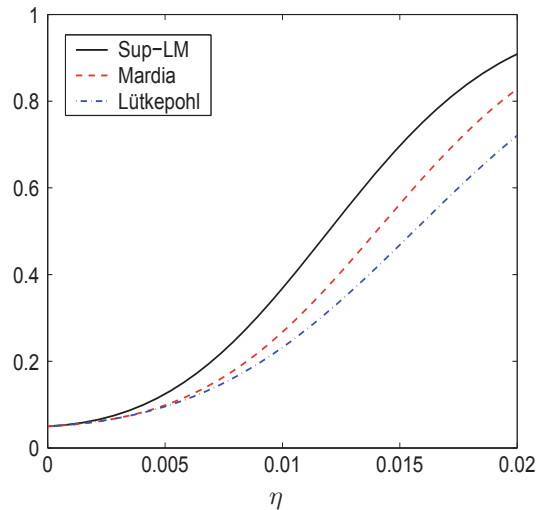
Figure 1f: Power of Sup-LM, Mardia and Lütkepohl normality tests against asymmetric t alternatives ($N = 3$).
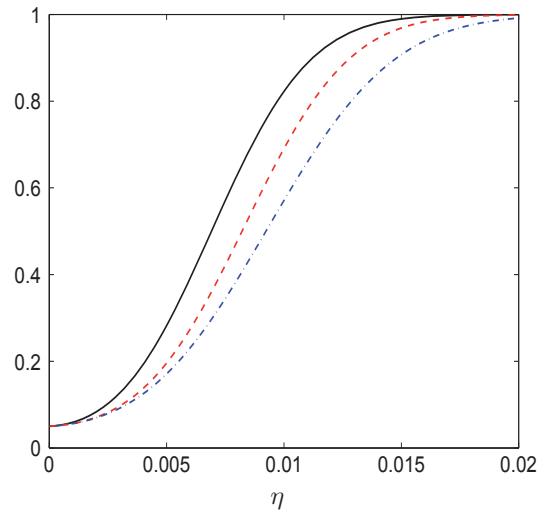
Notes: Thicker lines represent the power of the trivariate tests. Figures 1b-1d share the legend of Figure 1a, while Figure 1f shares the legend of figure 1e.

Figure 2: P-value discrepancy plots of the joint normality tests under the null of normality

(a) Asymptotic p-values

(b) Bootstrapped p-values
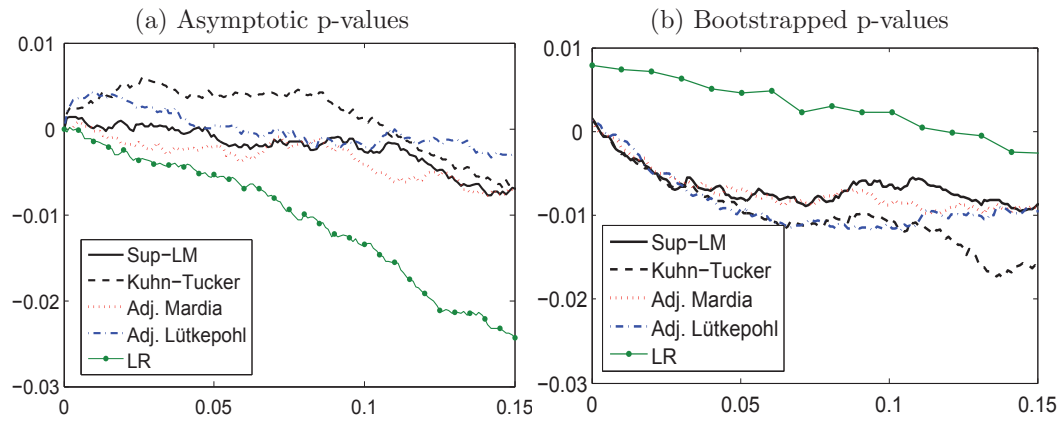
Figure 3: p-value discrepancy plots of the skewness components of the joint normality tests

(a) Asymptotic p-values

(b) Bootstrapped p-values
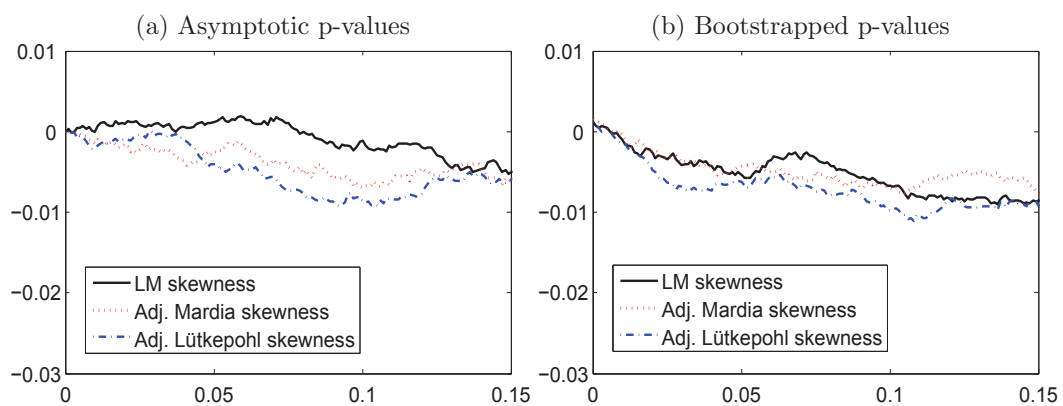
Figure 4: p-value discrepancy plots of the kurtosis components of the joint normality tests
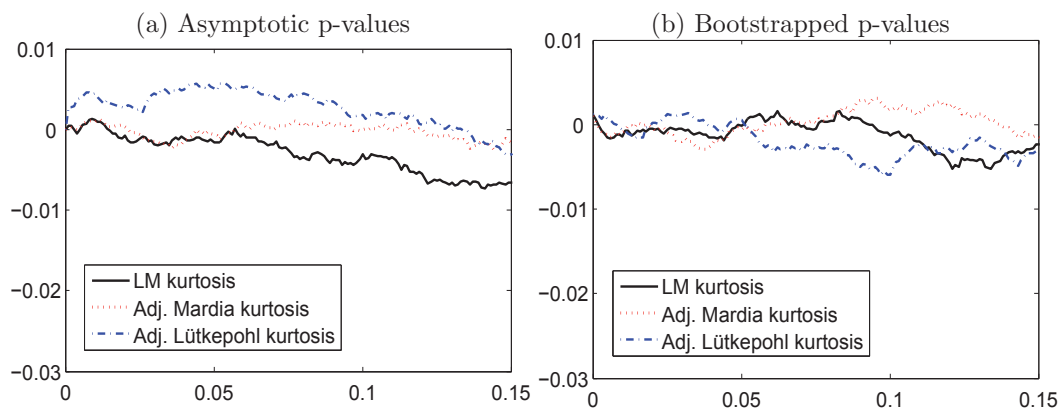
(a) Asymptotic p-values

(b) Bootstrapped p-values

Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations with T=1,000. Parametric bootstraped p-values are computed from 1,000 samples for all the tests except the LR, which is based on 100 only.

Figure 5: p-value discrepancy plots of the Student t tests under the null of Student t innovations

(a) Asymptotic p-values
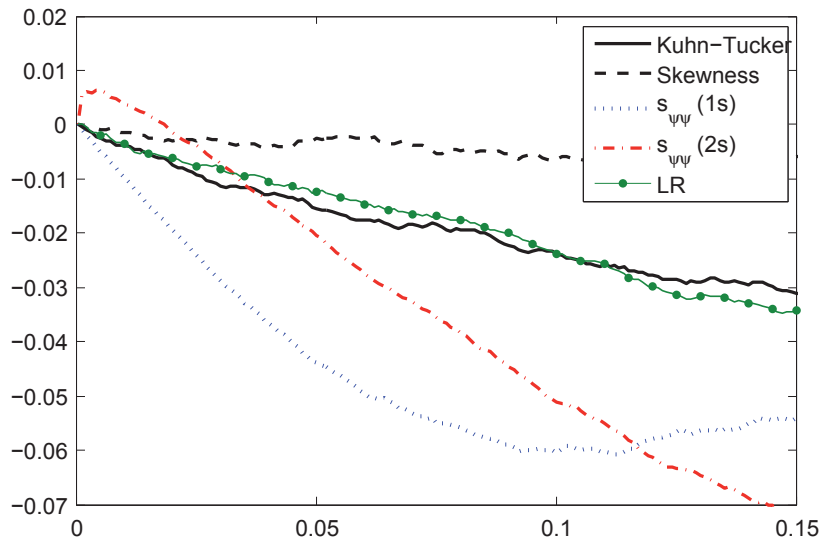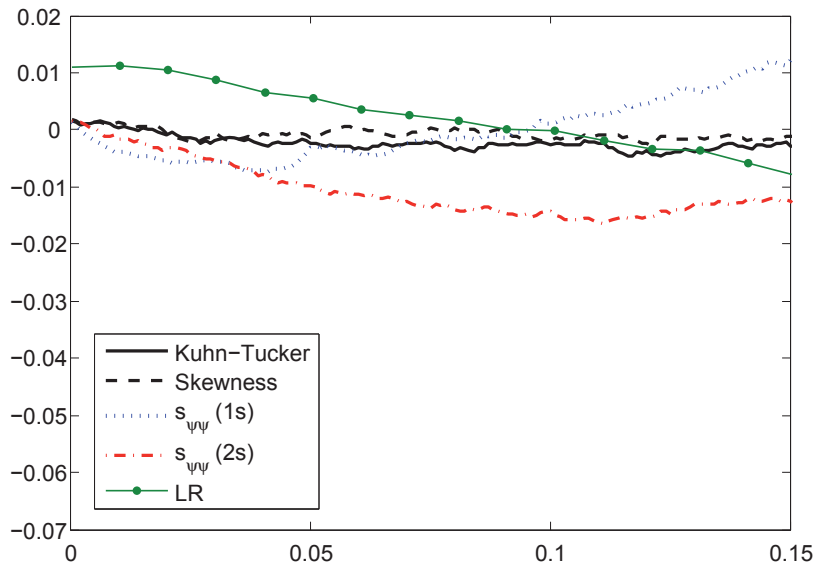
(b) Bootstrapped p-values

Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations with T=1,000. Parametric bootstraped p-values are computed from 1,000 samples for all the tests except the LR, which is based on 100 only.

Figure 6: Kernel estimation of the density of the symmetric Student t test
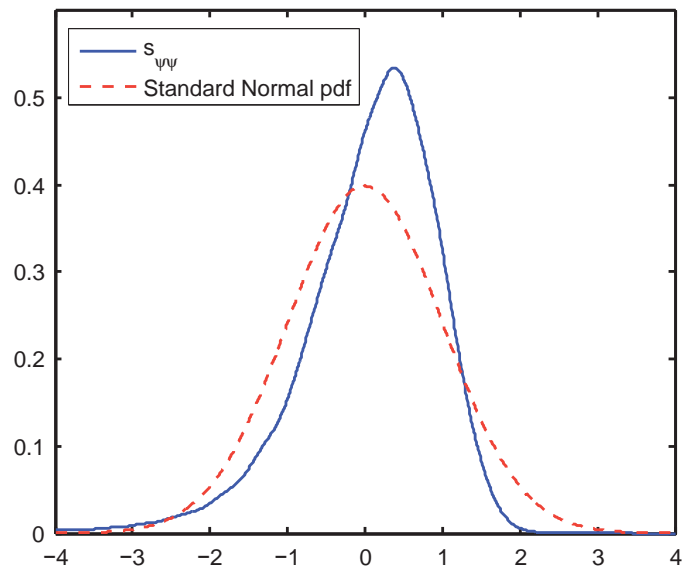
Notes: Monte Carlo study with 10,000 simulations with T=1,000.

Figure 7: Size-power plots under GH alternative hypotheses

(a) Normality tests
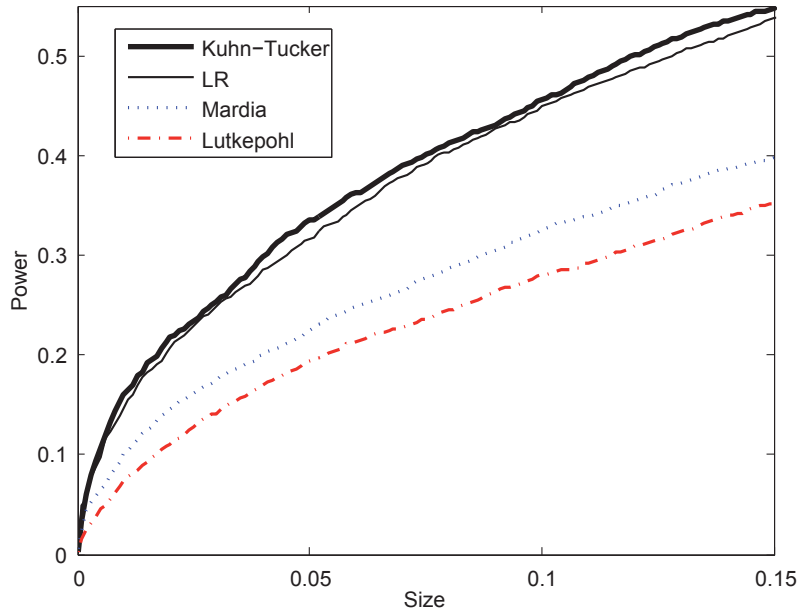
(b) Student t tests
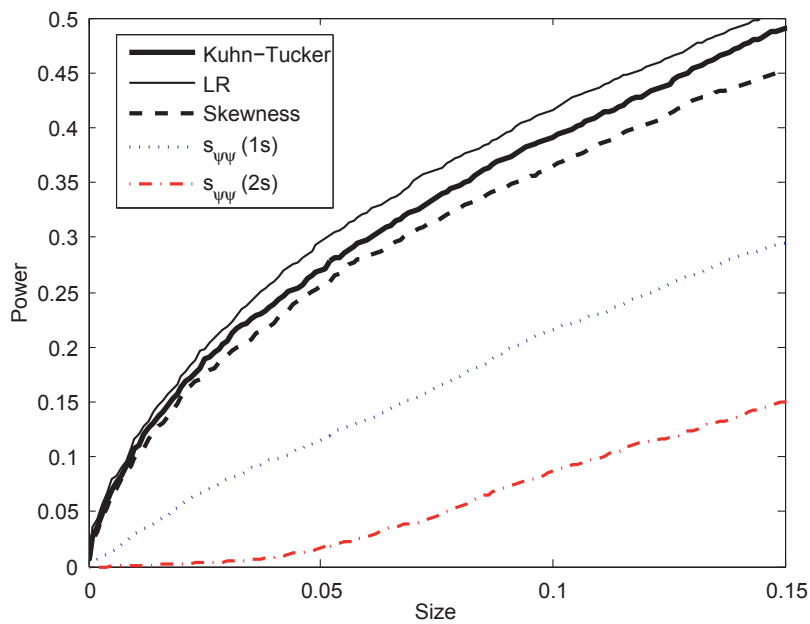
Notes: Monte Carlo study with 10,000 simulations with T=1,000. The data generating process in (a) is a GH distribution with $\eta = .01$, $\psi = 1$ and $\mathbf{b} = (-.05, -.05, -05)'$, while in (b) it is a GH distribution with $\eta = .2$, $\psi = .3$ and $\mathbf{b} = (-.05, -.05, -05)'$. In the Student t case, nominal sizes have been corrected by computing the p-values with the finite sample distribution of the tests under the null, which has been obtained from 10,000 simulations using the pseudo true values of the parameters to generate the data.

Figure 8a: Comparison of the univariate and multivariate Gaussian estimates of the (log)standard deviation for the equally weighted portfolio of the US Datastream sectoral indices
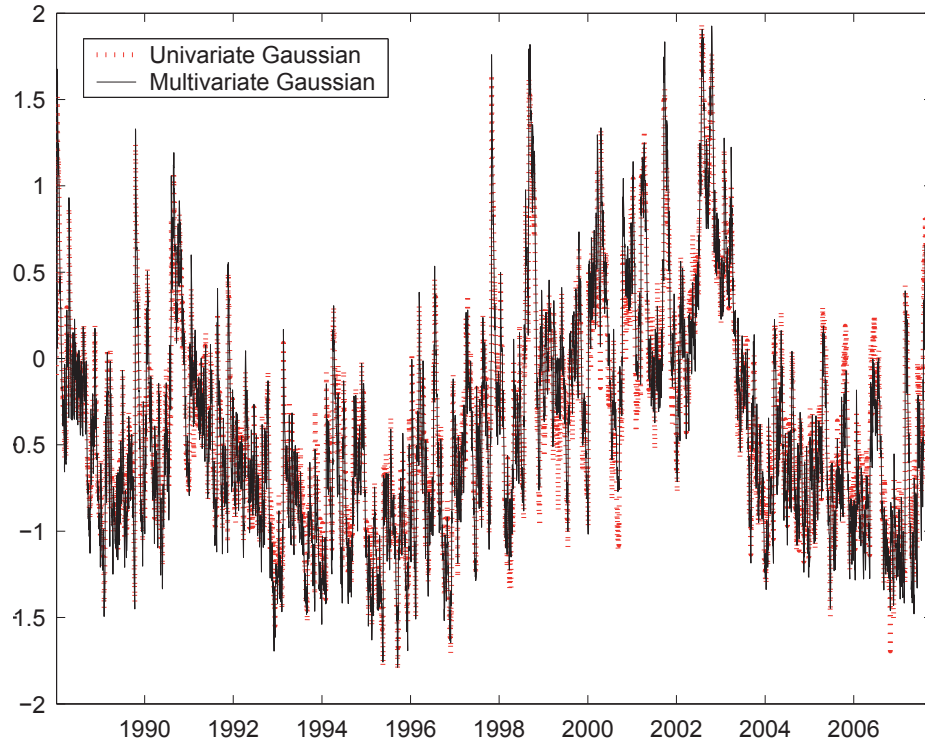
Figure 8b: Comparison of the multivariate Gaussian and *GH* estimates of the (log)standard deviation for the equally weighted portfolio of the US Datastream sectoral indices
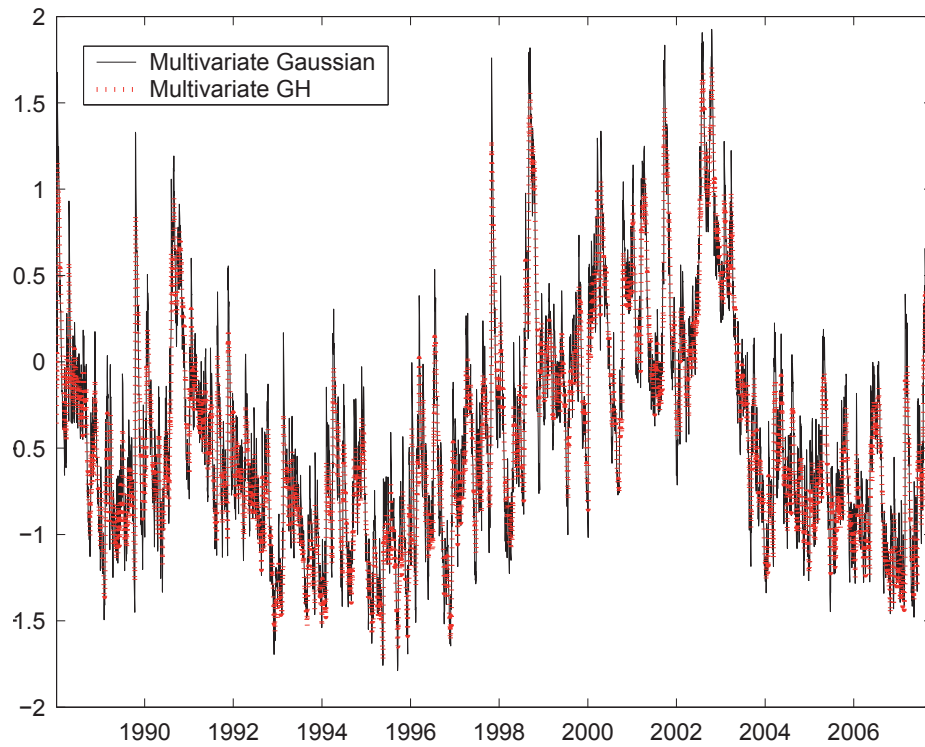
Figure 9: Discrepancy plots of the empirical cumulative distribution function of probability integral transforms of daily returns on portfolios of Datastream US indices.

(a) Equally weighted portfolio

(b) Maximum asymmetry portfolio
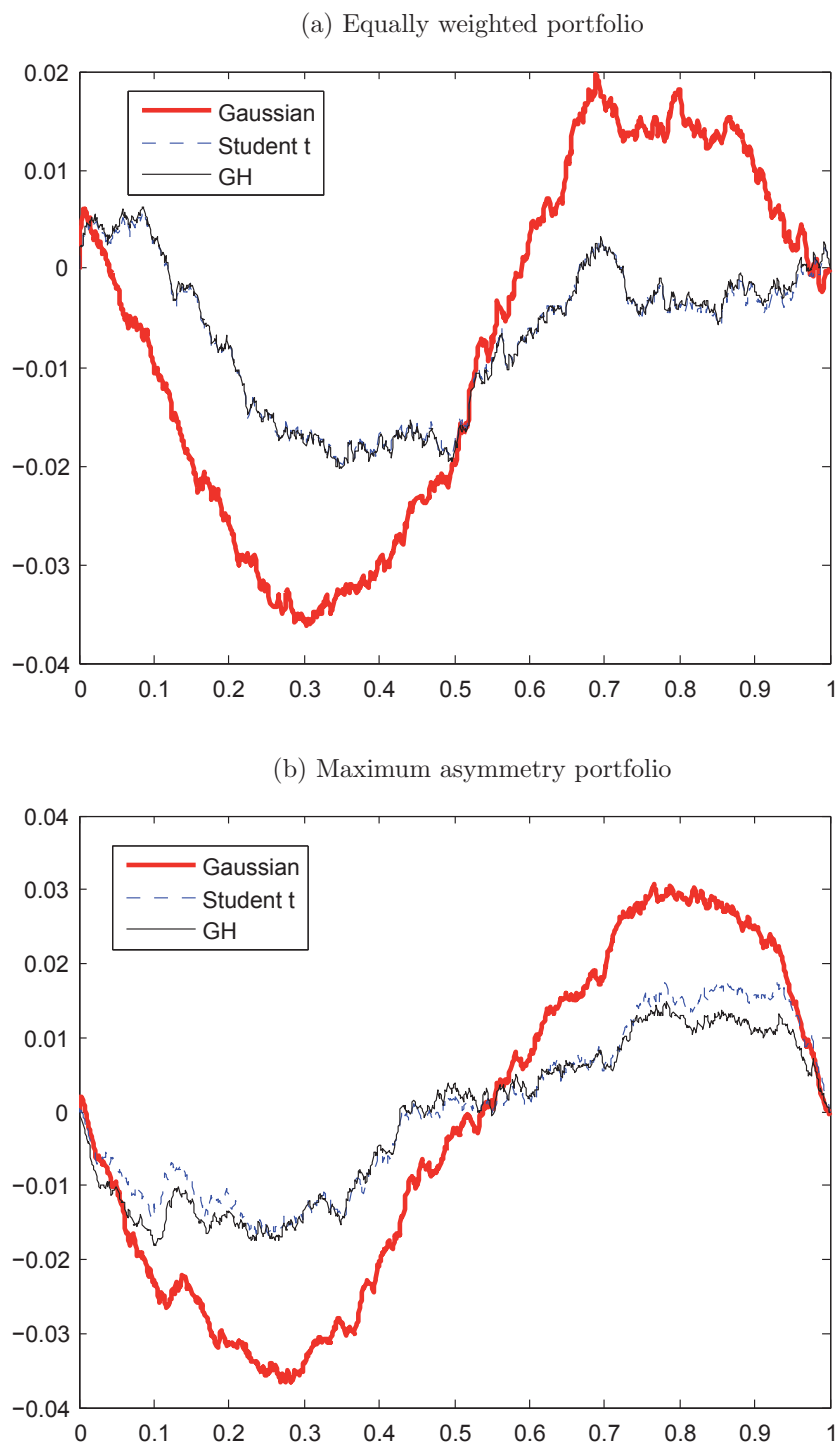
Note: Maximum asymmetry is obtained by considering portfolios with weights proportional to **b**, the asymmetry parameter of the Generalised Hyperbolic distribution.

Figure 10: Empirical cumulative distribution function of the Cramer von Mises test for 5,000 random portfolios that use the US Datastream sectoral daily data
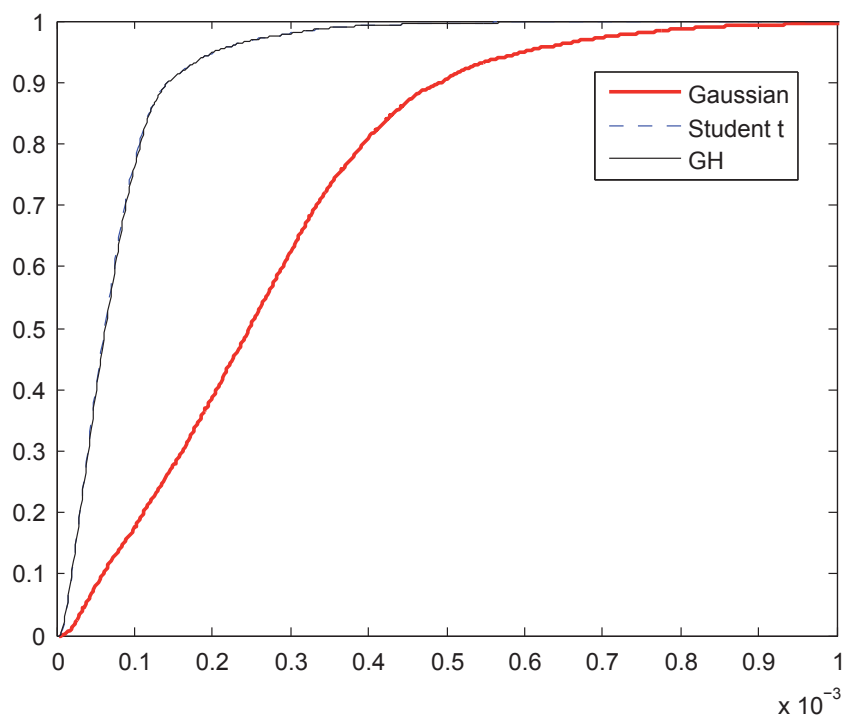
Note: The weights of the random portfolios are sampled from a uniform distribution on the 10-dimensional unit sphere.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS[1]

0821 GABRIEL JIMÉNEZ, JOSÉ A. LÓPEZ AND JESÚS SAURINA: Empirical analysis of corporate credit lines.

0822 RAMÓN MARÍA-DOLORES: Exchange rate pass-through in new Member States and candidate countries of the EU.

0823 IGNACIO HERNANDO, MARÍA J. NIETO AND LARRY D. WALL: Determinants of domestic and cross-border bank acquisitions in the European Union.

0824 JAMES COSTAIN AND ANTÓN NÁKOV: Price adjustments in a general model of state-dependent pricing.

0825 ALFREDO MARTÍN-OLIVER, VICENTE SALAS-FUMÁS AND JESÚS SAURINA: Search cost and price dispersion in vertically related markets: the case of bank loans and deposits.

0826 CARMEN BROTO: Inflation targeting in Latin America: Empirical analysis using GARCH models.

0827 RAMÓN MARÍA-DOLORES AND JESÚS VAZQUEZ: Term structure and the estimated monetary policy rule in the eurozone.

0828 MICHIEL VAN LEUVENSTEIJN, CHRISTOFFER KOK SØRENSEN, JACOB A. BIKKER AND ADRIAN VAN RIXTEL: Impact of bank competition on the interest rate pass-through in the euro area.

0829 CRISTINA BARCELÓ: The impact of alternative imputation methods on the measurement of income and wealth: Evidence from the Spanish survey of household finances.

0830 JAVIER ANDRÉS AND ÓSCAR ARCE: Banking competition, housing prices and macroeconomic stability.

0831 JAMES COSTAIN AND ANTÓN NÁKOV: Dynamics of the price distribution in a general model of state-dependent pricing.

0832 JUAN A. ROJAS: Social Security reform with imperfect substitution between less and more experienced workers.

0833 GABRIEL JIMÉNEZ, STEVEN ONGENA, JOSÉ LUIS PEYDRÓ AND JESÚS SAURINA: Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?

0834 ENRIQUE ALBEROLA AND JOSÉ MARÍA SERENA: Sovereign external assets and the resilience of global imbalances.

0835 AITOR LACUESTA, SERGIO PUENTE AND PILAR CUADRADO: Omitted variables in the measure of a labour quality index: the case of Spain.

0836 CHIARA COLUZZI, ANNALISA FERRANDO AND CARMEN MARTÍNEZ-CARRASCAL: Financing obstacles and growth: An analysis for euro area non-financial corporations.

0837 ÓSCAR ARCE, JOSÉ MANUEL CAMPA AND ÁNGEL GAVILÁN: asymmetric collateral requirements and output composition.

0838 ÁNGEL GAVILÁN AND JUAN A. ROJAS: Solving Portfolio Problems with the Smolyak-Parameterized Expectations Algorithm.

0901 PRAVEEN KUJAL AND JUAN RUIZ: International trade policy towards monopoly and oligopoly.

0902 CATIA BATISTA, AITOR LACUESTA AND PEDRO VICENTE: Micro evidence of the brain gain hypothesis: The case of Cape Verde.

0903 MARGARITA RUBIO: Fixed and variable-rate mortgages, business cycles and monetary policy.

0904 MARIO IZQUIERDO, AITOR LACUESTA AND RAQUEL VEGAS: Assimilation of immigrants in Spain: A longitudinal analysis.

0905 ÁNGEL ESTRADA: The mark-ups in the Spanish economy: international comparison and recent evolution.

0906 RICARDO GIMENO AND JOSÉ MANUEL MARQUÉS: Extraction of financial market expectations about inflation and interest rates from a liquid market.

0907 LAURA HOSPIDO: Job changes and individual-job specific wage dynamics.

0908 M.ª DE LOS LLANOS MATEA AND JUAN S. MORA: La evolución de la regulación del comercio minorista en España y sus implicaciones macroeconómicas.

0909 JAVIER MENCÍA AND ENRIQUE SENTANA: Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation.

0910 ALICIA GARCÍA-HERRERO, SERGIO GAVILÁ AND DANIEL SANTABÁRBARA: What explains the low profitability of Chinese banks?

0911 JAVIER MENCÍA: Assessing the risk-return trade-off in loans portfolios.

0912 MAXIMO CAMACHO AND GABRIEL PEREZ-QUIROS: Ñ-STING: España Short Term INdicator of Growth.

0913 RAQUEL VEGAS, ISABEL ARGIMÓN, MARTA BOTELLA AND CLARA I. GONZÁLEZ: Retirement behaviour and retirement incentives in Spain.

---

1. Previously published Working Papers are listed in the Banco de España publications catalogue.