Distributional tests in multivariate dynamic models with Normal and Student t innovations*

Javier Mencía
Bank of Spain
<javier.mencia@bde.es>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

June 2010

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. We also obtain more powerful one-sided Kuhn-Tucker versions that are equivalent to the Likelihood Ratio test, whose asymptotic distribution we provide. Finally, we conduct detailed Monte Carlo exercises to study the size and power properties of our proposed tests in finite samples.

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

JEL: C12, C52, C32

*This paper partly supersedes our earlier work circulated under the title "Estimation and testing of dynamic models with Generalised Hyperbolic innovations". We are grateful to Dante Amengual, Manuel Arellano, Ole Barndorff-Nielsen, Luc Bauwens, Anil Bera, Rob Engle, Gabriele Fiorentini, Eric Ghysels, Vassilis Hajivassiliou, Sébastien Laurent, Eric Renault, Luis Seco, Neil Shephard and Mervyn Sylvapulle, as well as seminar audiences at Aarhus, Alicante, CORE, Duke, ESRC Econometric Study Group, LSE FMG, MEFF, Monash, UIB, UPN, the 2004 CIRANO-CIREQ Financial Econometrics Conference (Montreal), the 2004 European Meetings of the Econometric Society (Madrid), the XII Finance Forum (Barcelona) and the XXIX Symposium on Economic Analysis (Pamplona) for helpful comments and suggestions. The input of Mark W. Watson and three anonymous referees has also greatly improved the paper. Of course, the usual caveat applies. Financial support from Fundación Ramón Areces (Mencía) and the Spanish Ministry of Science and Innovation through grant ECO2008-00280 (Sentana) is gratefully acknowledged. Address for correspondence: Casado del Alisal 5, E-28014 Madrid, Spain, tel: +34 91 429 05 51, fax: +34 91 429 1056.

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 numerically invariant to the ordering of the factorisation of the joint density into marginal and conditional components, 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, 2009).

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. Finally, we also study the finite sample properties of our proposed tests with an extensive Monte Carlo analysis.

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, 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

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

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 C). 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.

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

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

where

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

$R_\nu\,(\gamma) = K_{\nu+1}\,(\gamma)\,/K_\nu\,(\gamma)$, $D_{\nu+1}\,(\gamma) = K_{\nu+2}\,(\gamma)\,K_\nu\,(\gamma)\,/K_{\nu+1}^2\,(\gamma)$ and $K_\nu\,(\cdot)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun, 1965, p. 374, as well as Appendix C). 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,

[1] See Appendix C for further details.

the Normal distribution 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 $\boldsymbol{\beta}$. Analogously, the Student t is obtained when $-\infty < \nu < -2$, $\gamma = 0$ and $\boldsymbol{\beta} = \mathbf{0}$.

Importantly, since $\boldsymbol{\varepsilon}_t^*$ 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] From an asset allocation perspective, one interesting feature of (5) is that $\mathbf{b}$ can be interpreted as the weights of the portfolio that yields maximum asymmetry (see Mencía and Sentana, 2009). In what follows, we maintain the assumption that $\mathbf{b}$ is time-invariant (see Appendix A for a generalisation of (5) that allows for time varying asymmetry parameters). 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 C for explicit expressions). As we mentioned before, we can achieve normality in three different ways: (i) when $\eta \to 0^+$ or (ii) $\eta \to 0^-$ regardless of the values of $\mathbf{b}$ and ψ; and (iii) when $\psi \to 0^+$, irrespective of η and $\mathbf{b}$. Therefore, it is not surprising that the Gaussian scores with respect to η 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,

[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 $\nu \to \pm\infty$ into a single point. Nevertheless, it is still worth working with η instead of ν when testing for normality. See the proof of Proposition 4 for an alternative reparametrisation.

$\lim_{\eta\cdot\psi\to 0} s_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$, 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 $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ 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 $\eta \to 0^+$,

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

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* $\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) = (N+2)[.5N + 2\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}]$, *so that* $E[\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b})] = (N+2)[.5N + 2\mathbf{b}'\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{b}]$ *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 The supremum tests

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 underidentified parameters under the null involves fixing the underidentified parameters to some arbitrary values, and then computing the appropriate test statistic for those values.

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, $\bar{s}_{\boldsymbol{\theta}T}\big(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b}\big)$ and $\bar{s}_{\eta T}\big(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b}\big)$, with weighting matrix the inverse of the unconditional information matrix, which can be obtained as the unconditional expected value

[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 $b_1 \to \pi$.

of the conditional information matrix in Proposition 1. But since $\bar{s}_{\boldsymbol{\theta}T}\big(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b}\big) = \mathbf{0}$ by definition of $\tilde{\boldsymbol{\theta}}_T$, and $\mathcal{I}_{\boldsymbol{\theta}\eta t}\left(\boldsymbol{\theta}_0, 0^+, \psi, \mathbf{b}\right) = \mathbf{0}$, we can show that

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

We can operate analogously for the other two limits, thereby obtaining the test statistic $LM_2\big(\tilde{\boldsymbol{\theta}}_T, \psi, \mathbf{b}\big)$ for the null $\eta \to 0^-$, and $LM_3\big(\tilde{\theta}_T, \eta, \mathbf{b}\big)$ 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:*

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

   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*

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

   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.

Testing for fixed values of the underidentified parameters 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, we now 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*

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

$$LM_k(\boldsymbol{\theta}_0) = \frac{2}{N(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{N(N+2)}{4} \right] \right\}^2, \tag{10}$$

$$LM_s(\boldsymbol{\theta}_0) = \frac{1}{2(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)\left[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)\right] \right\}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \times \left\{ \frac{\sqrt{T}}{T} \sum_t \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)\left[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)\right] \right\}, \tag{11}$$

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

$$E\left[\frac{N(N+2)}{4}-\frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0)+\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0)\right]=E[m_{kt}(\boldsymbol{\theta}_0)]=0, \tag{12}$$

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

$$E\left[m_{kt}(\boldsymbol{\theta}_0)\right|\boldsymbol{\theta}_0,\eta_0>0,\psi_0>0]>0. \tag{13}$$

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:*

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

   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 $\Pr(X > c) = 1 - .5F_{\chi^2_N}(c) - .5F_{\chi^2_{N+1}}(c)$ 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:

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

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.*

[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.

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

By construction, both the LR and the KT tests will be unable to yield power for symmetric departures from normality with tails thinner than the normal. In those cases, though, the sup LM test (9) will retain non-trivial power, since it does not maintain the assumption of non-negative excess kurtosis under the alternative hypothesis. Proposition 7 also implies that our test would capture departures from normality as long as the coefficients of the Hermite expansion (15) of the density are different from zero. Hence, our approach would not yield power if (15) included some additional orthogonal terms but the coefficients on $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left(\varsigma_t(\boldsymbol{\theta})-(N+2)\right)$ and $0.25\varsigma_t^2(\boldsymbol{\theta})-0.5(N+2)\varsigma_t(\boldsymbol{\theta})+0.25N\left(N+2\right)$ were zero.

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

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

which are the relevant ones against GH innovations. In order to interpret these moment conditions, we can rewrite the N elements of $\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\mathbf{m}_{st}(\boldsymbol{\theta}_0)$ as

$$\sqrt{6}H_3(\varepsilon_{it}^*(\boldsymbol{\theta}))+\sqrt{2}\sum_{\substack{j=1\\ j\neq i}}^{N}H_1(\varepsilon_{it}^*(\boldsymbol{\theta}))H_2(\varepsilon_{jt}^*(\boldsymbol{\theta})), \tag{18}$$

for $i=1,\cdots,N$, where $H_k(\cdot)$ is the Hermite polynomial of order k. Hence, (17) takes into account both the univariate skewness of each variable and its co-skewness $H_1(\varepsilon_{it}^*(\boldsymbol{\theta}))H_2(\varepsilon_{jt}^*(\boldsymbol{\theta}))$ with the remaining variables. Compared to (16), though, (18) does not consider terms of the form $H_1(\varepsilon_{it}^*(\boldsymbol{\theta}))H_1(\varepsilon_{jt}^*(\boldsymbol{\theta}))H_1(\varepsilon_{lt}^*(\boldsymbol{\theta}))$ for $i\neq j\neq l$. Similarly, we can obtain the following analogous decomposition for the kurtosis component $\mathbf{m}_{kt}(\boldsymbol{\theta}_0)$:

$$\sqrt{24}\sum_{i=1}^{N}H_4(\varepsilon_{it}^*(\boldsymbol{\theta}))+4\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N}}_{i\neq j}H_2(\varepsilon_{it}^*(\boldsymbol{\theta}))H_2(\varepsilon_{jt}^*(\boldsymbol{\theta})). \tag{19}$$

Thus, the kurtosis component considers the marginal kurtosis of each element as well as all the co-kurtosis terms $H_2(\varepsilon^*_{it}(\boldsymbol{\theta}))H_2(\varepsilon^*_{jt}(\boldsymbol{\theta}))$.

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.

The test proposed by Mardia (1970) was 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 $\boldsymbol{\theta}_0$.

By analogy with Bontemps and Meddahi (2005, 2010), one possible way to adjust Mardia's (1970) formulae is to replace $\varepsilon^{*3}_{it}(\boldsymbol{\theta})$ by $H_3[\varepsilon^*_{it}(\boldsymbol{\theta})]$ and $\varepsilon^{*2}_{it}(\boldsymbol{\theta})\varepsilon^*_{jt}(\boldsymbol{\theta})$ by $H_2[\varepsilon^*_{it}(\boldsymbol{\theta})]H_1[\varepsilon^*_{jt}(\boldsymbol{\theta})]$ $(i \neq j)$ in the moment conditions (16). 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 C.6 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 $\varepsilon^*_{it}(\tilde{\boldsymbol{\theta}}_T)$.

3.4 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 C 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.

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.

[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.

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 $\mathbf{b}$,

$$s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, 0) = \frac{\eta\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right]}{1 - 2\eta + \eta\varsigma_t(\theta)}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \tag{20}$$

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 $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) = V\left[s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}, 1, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\pi}, 1, \mathbf{0}\right]$. 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 $\mathcal{I}_{\eta\mathbf{b}t}(\boldsymbol{\theta}, \eta > 0, 1, \mathbf{0}) = \mathbf{0}$,*

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

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 $\mathcal{I}^{-1}_{\boldsymbol{\pi\pi}}$, where $\mathcal{I}_{\boldsymbol{\pi\pi}}$ is the unconditional information matrix under Student t innovations.[7]

4.3 Student t vs symmetric 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

$$s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = \frac{\eta^2}{(1-2\eta)(1-4\eta)} \frac{\varsigma_t(\boldsymbol{\theta}) - N(1-2\eta)}{1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})} + \frac{\eta^2 [N - \varsigma_t(\boldsymbol{\theta})]}{(1-2\eta)(1+(N-2)\eta)}, \quad (21)$$

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

$$E[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) | \boldsymbol{\pi}_0, \psi_0 < 1, \mathbf{b}_0 = \mathbf{0}] > 0 \quad (22)$$

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

$$\tau_{kT}(\bar{\boldsymbol{\pi}}_T) = \frac{\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})}{\sqrt{\hat{V}[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]}}, \quad (23)$$

where $\hat{V}[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]$ is a consistent estimator of the asymptotic variance of $s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ that takes into account the sampling variability in $\bar{\boldsymbol{\pi}}_T$. Under the null

[7] In particular, Crowder (1976) requires: (i) $\boldsymbol{\pi}_0 \in \text{int}\,\boldsymbol{\Pi}$ is locally identified, where $\boldsymbol{\Pi}$ is a bounded subset of $\mathbb{R}^{p+1}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\pi}_0$; (iii) there is uniform convergence of the integrals involved in the computation of the mean vector and covariance matrix of $s_t(\boldsymbol{\pi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \partial s_t(\boldsymbol{\pi})/\partial\boldsymbol{\pi}'\right] T^{-1}\sum_t \partial s_t(\boldsymbol{\pi})/\partial\boldsymbol{\pi}' \xrightarrow{p} \mathbf{I}_{p+1}$, where $E^{-1}\left[-T^{-1}\sum_t \partial s_t(\boldsymbol{\pi})/\partial\boldsymbol{\pi}'\right]$ is positive definite on a neighbourhood of $\boldsymbol{\pi}_0$.

hypothesis of Student t innovations with more than 4 degrees of freedom, it is easy to see that the asymptotic distribution of $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ will be $N(0,1)$. The required expression for $V[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})]$ 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*

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

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

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

2. *If in addition the regularity conditions in Crowder (1976) hold, then*

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

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

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

with

$$\begin{aligned}\mathcal{M}_{\boldsymbol{\theta}t}(\boldsymbol{\pi}_0) &= \frac{4(N+2)\eta_0^4(1-2\eta_0)^{-1}(1-4\eta_0)^{-1}}{[1+(N+2)\eta_0][1+(N-2)\eta_0]}\frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)],\\ \mathcal{M}_{\eta t}(\boldsymbol{\pi}_0) &= \frac{-2N(N+2)\eta_0^3(1-2\eta_0)^{-2}(1-4\eta_0)^{-1}}{(1+N\eta_0)[1+(N+2)\eta_0]}.\end{aligned}$$

Lee and Chesher (1986, page 145) show the equivalence between (23) 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 (22). Hence, the statistic $\tau_{kT}^2(\bar{\boldsymbol{\pi}}_T)\mathbf{1}[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})>0]$ 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.[8]

[8] In our case, the equivalence between the LR and the extremum test can be formally proved by reparametrising the GH distribution in terms of $\psi^* = .5(\psi-1)^2$, instead of ψ. This change of variables does not affect the LR test, but the score with respect to ψ^* will be $s_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})$ under Student t innovations. Hence, $\tau_{kT}^2(\bar{\boldsymbol{\pi}}_T)\mathbf{1}[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})>0]$ can be interpreted as a KT test under this reparametrisation, which is equivalent to the LR test. See Lee and Chesher (1986) for more details.

Finally, it is also important to mention that although $s_{\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{b}) = 0 \ \forall t$, 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 $E[s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})] = \mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})$ is defined in (20). 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*

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

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 $V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right]$ are defined in Propositions 8 and 9, respectively.

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

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

where

$$\mathcal{V}(\boldsymbol{\pi}_0) = \begin{bmatrix} \mathcal{V}_{\mathbf{bb}}(\boldsymbol{\pi}_0) & \mathcal{V}_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) \\ \mathcal{V}'_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) & \mathcal{V}_{\psi\psi}(\boldsymbol{\pi}_0) \end{bmatrix} = \left\{ \begin{matrix} \mathcal{I}_{\mathbf{bb}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{0} \\ \mathbf{0}' & V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] \end{matrix} \right\}$$
$$- \begin{bmatrix} \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \\ \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \end{bmatrix}, \quad (24)$$

$\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ 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

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

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

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

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 $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{b}}(\boldsymbol{\pi}_0, 1, 0) = \mathbf{0}$ (see Proposition 8).

It is then straightforward to see that the asymptotic distribution of

$$\begin{aligned}\tau_{oT}(\bar{\boldsymbol{\pi}}_T) = T\bar{s}^{\perp\prime}_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \left[\mathcal{V}_{\mathbf{bb}}(\bar{\boldsymbol{\pi}}_T) - \frac{\mathcal{V}_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)\mathcal{V}'_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)}{\mathcal{V}_{\psi\psi}(\bar{\boldsymbol{\pi}}_T)}\right]^{-1} \bar{s}^{\perp}_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ +\tau^2_{kT}(\bar{\boldsymbol{\pi}}_T)\,\mathbf{1}\left[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) > 0\right]\end{aligned} \tag{26}$$

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, (26) numerically coincides with (25). 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 (26) 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

$$\tau_{aT}(\bar{\boldsymbol{\pi}}_T) = \sqrt{T}\bar{s}'_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})\,\mathcal{V}^{-1}_{\mathbf{bb}}(\bar{\boldsymbol{\pi}}_T)\sqrt{T}\bar{s}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}), \tag{27}$$

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

$$\frac{2N(N+2)}{E(\varsigma_t^3(\boldsymbol{\theta}_0)) - 2(N+2)E(\varsigma_t^2(\boldsymbol{\theta}_0)) + N(N+2)^2} \tag{28}$$

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.[9] 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 (27), 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

[9] This approach extends to the multivariate case the results of Godfrey and Orme (1991), who test for asymmetry of regression residuals without assuming normality.

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 Gaussian model:

$$y_{it} = \mu_i + c_i f_t + v_{it} \quad i = 1, \cdots, N,$$

where $f_t = \lambda_t^{1/2} f_t^*$, $v_{it} = \gamma_{it}^{1/2} v_{it}^*$ $(i = 1, \cdots, N)$,

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

$(f_t^*, v_{1t}^*, \cdots, v_{Nt}^*)|I_{t-1} \sim N(\mathbf{0}, \mathbf{I}_{N+1})$, 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

$$\begin{aligned}\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{cc}'\lambda_t + \boldsymbol{\Gamma}_t,\end{aligned} \tag{29}$$

where $\boldsymbol{\mu}' = (\mu_1, \cdots, \mu_N)$, $\mathbf{c}' = (c_1, \cdots, c_N)$, and $\boldsymbol{\Gamma}_t = diag(\gamma_{1t}, \cdots, \gamma_{Nt})$. 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 $\phi_0 = 1 - \phi_1 - \phi_2$. We report results for $N = 3$, $N = 10$, $T = 1,000$ and $T = 10,000$ based on 10,000 Monte Carlo replications, which allows us to precisely estimate actual sizes.[10] 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, p. 404) shows that the size of bootstrap tests remains asymptotically valid when some of the parameters are on the boundary of the parameter space, even though the sampling distribution of the estimators provided by the bootstrap is invalid. 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.[11] 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

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

[11]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.

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).[12]

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 parameters 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. Specifically, we orthogonalise the Mardia (1970) and Lütkephohl (1993) expressions with respect to the Gaussian scores of $\boldsymbol{\theta}$. Thanks to this orthogonality, we do not need to reestimate $\boldsymbol{\theta}$ in each bootstrap sample, which would be computationally infeasible in this Monte Carlo exercise. The main drawback of this approach is that it does not benefit from the higher order refinements that the bootstrap provides. In this sense, we recommend reestimating $\boldsymbol{\theta}$ in actual empirical applications where such computational considerations are generally irrelevant.

Figures 2-4 summarise our findings for the different multivariate normality tests in the trivariate model with $T = 1,000$. We use 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 those distortions to some extent.[13] 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.

We investigate in Figures 5-7 the impact of either increasing the sample size or using

[12]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.

[13]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.

a larger cross-sectional dimension on the asymptotic p-values. We observe that a larger sample size generally reduces the size distortions. Nevertheless, although the distortions of the LR test become slightly smaller for $T = 10,000$, they are still larger than those of the KT test. In contrast, we observe larger size distortions when we set $N = 10$, with the LR suffering the most. In any case, the performance of the LR test would improve if we could consider a sufficiently dense grid of initial values for η, ψ and $\mathbf{b}$.

Finite sample size of the Student t tests In this case we maintain the conditional mean and variance specification in (29), but generate the standardised innovations $\boldsymbol{\varepsilon}_t^*$ from a Student t distribution. 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 $\bar{\boldsymbol{\pi}}_T$, 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 8 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 when $N = 3$, $\nu = 10$ and $T = 1,000$, shown in Figure 8a, is the fact that the actual sizes of the "kurtosis" tests based on $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$, which is defined in (23), are well below their nominal sizes. This is due to the fact that the sampling distribution of $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ is not well approximated by a standard normal unless the sample size is rather large, as illustrated in Figure 9. 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. Figure 8b confirms that the parametric bootstrap is able to yield p-values that are much closer to the nominal ones.[14] In turn, Figure 8c shows that the asymptotic p-values are more reliable for larger sample sizes, especially for the $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ test, as also illustrated in Figure 9. Finally, Figures 8d and 8e show that we generally obtain larger distortions for higher

[14] 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.

values of η as well as for larger cross-sectional dimensions.

Finite sample power of the normality tests We have repeated the normality tests using the same mean and variance specification as in (29), but generating the 10,000 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 10a 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 10b 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 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.

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. It could also be useful to study the empirical relevance of asymmetric deviations with time varying features, such as the ones mentioned in Appendix A. 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.

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.

Bontemps, C. and N. Meddahi (2010). Testing distributional assumptions: a GMM approach. mimeo University of Toulouse.

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

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. 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.

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.

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.

Godfrey, L. G. and C. D. Orme (1991). Testing for skewness of regression disturbances. *Economic Letters 37*, 31–34.

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.

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.

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 (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics 153*, 105–121.

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

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

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

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 Normality test with a time varying vector of asymmetry parameters

Consider the following parametrisation of $\boldsymbol{\beta}_t$:

$$\boldsymbol{\beta}_t(\boldsymbol{\theta}, \mathbf{B}, \mathbf{z}_{t-1}) = \boldsymbol{\Sigma}_t^{\frac{1}{2}\prime}(\boldsymbol{\theta})\mathbf{B}\mathbf{z}_{t-1},$$

where $\mathbf{B}$ is a $N \times k$ matrix and $\mathbf{z}_{t-1}$ is a vector of k covariates known at $t-1$. This parametrisation is a generalisation of (5) that also ensures that the log-likelihood does not depend on the choice of $\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})$. It can be shown that the score with respect to η under normality becomes in this case

$$\begin{aligned}\lim_{\eta\to 0^+} s_{\eta t}(\boldsymbol{\phi}) &= -\lim_{\eta\to 0^-} s_{\eta t}(\boldsymbol{\phi}) = \left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N\,(N+2)}{4}\right] \\ &\quad + \mathbf{z}_{t-1}'\mathbf{B}'\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right]. \qquad \text{(A1)}\end{aligned}$$

Similarly, we can show that for $\eta \neq 0$, $\lim_{\psi\to 0^+} s_{\psi t}(\boldsymbol{\phi})$ is exactly one half of (A1).

We can express the conditional variance of (A1) under normality as

$$V\left[\lim_{\eta\to 0^+} s_{\eta t}(\boldsymbol{\phi})\right] = \frac{N(N+2)}{2} + 2(N+2)vec'(\mathbf{B})E[\mathbf{z}_{t-1}\mathbf{z}_{t-1}' \otimes \boldsymbol{\Sigma}_t(\boldsymbol{\theta})]vec(\mathbf{B}).$$

Hence, the LM test for given values of the unidentified parameters can be expressed as

$$\begin{aligned}LM\left(\boldsymbol{\theta}_0, \mathbf{B}, \mathbf{z}_{t-1}\right) &= (N+2)^{-1}\left(\frac{N}{2} + 2vec'(\mathbf{B})E[\mathbf{z}_{t-1}\mathbf{z}_{t-1}' \otimes \boldsymbol{\Sigma}_t(\boldsymbol{\theta})]vec(\mathbf{B})\right)^{-1} \\ &\times \left\{\frac{\sqrt{T}}{T}\sum_t \left[\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{N\,(N+2)}{4}\right]\right. \\ &\left. + vec'(\mathbf{B})\frac{\sqrt{T}}{T}\sum_t \left[\mathbf{z}_{t-1}\otimes\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\right]\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right]\right\}^2. \qquad \text{(A2)}\end{aligned}$$

Once again, we obtain the same formula regardless of the testing direction in which we approach the null of normality. Using analogous arguments as in Proposition 3, it can be shown that this test converges asymptotically to a chi-square with one degree of freedom under normality. Also as in the simpler case, we can obtain the sup of (A2) by expressing this maximisation as an eigenvalue problem (see proof of Proposition 5). Specifically, we obtain

$$\sup_{\mathbf{B}\in\mathbb{R}^{N\times k}} LM(\boldsymbol{\theta}_0) = LM_k(\boldsymbol{\theta}_0) + LM_s(\boldsymbol{\theta}_0), \qquad \text{(A3)}$$

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

$$LM_s(\boldsymbol{\theta}_0) = \frac{1}{2(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t [\mathbf{z}_{t-1} \otimes \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})] [\varsigma_t(\boldsymbol{\theta}) - (N+2)] \right\}' E^{-1}[\mathbf{z}_{t-1}\mathbf{z}_{t-1}' \otimes \boldsymbol{\Sigma}_t(\boldsymbol{\theta})]$$
$$\times \left\{ \frac{\sqrt{T}}{T} \sum_t [\mathbf{z}_{t-1} \otimes \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})] [\varsigma_t(\boldsymbol{\theta}) - (N+2)] \right\}.$$

The asymptotic distribution of (A3) under normality is a chi-square with $NK+1$ degrees of freedom. Furthermore, if the regularity conditions of Proposition 2 hold, then the above results will remain true if we substitute $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.

B 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 C). 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. □

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. □

Proposition 4

The first thing to note is that, although it may seem that the null hypotheses of $\eta \rightarrow 0^{\pm}$ are interior points, η is in fact on the boundary of the parameter space in the three cases. The reason is that, by using the change of variables $\eta = -.5\nu^{-1}$, we are "pasting" together the two limits $\nu \rightarrow \pm\infty$. In this sense, not that if we had used $\eta^* = [1+\exp(-\nu)]^{-1}$ instead, then normality would be obtained for $\eta^* = 0$ (corresponding to $\eta \rightarrow 0^+$) and $\eta^* = 1$ ($\eta \rightarrow 0^-$), which are clearly on the boundary of the admissible parameter space. As a result, the log-likelihood would be right-continuously differentiable at $\eta^* = 0$ and left-continuously differentiable at $\eta^* = 1$. In that context, we could use the results of Andrews (2001) to prove the equivalence of the LR test and the KT tests on the three testing directions, which are based on the directed score.

Consider initially the situation in which we fix $\mathbf{b}$ and ψ, and only allow η to be positive under the alternative. Note 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 Andrews (2001) imply that the LR test will not be asymptotically equivalent to the corresponding LM test $LM_1\big(\boldsymbol{\theta}_0, \psi, \mathbf{b}\big)$, but rather to the Kuhn-Tucker test

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

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

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

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

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

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:

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

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. □

Proposition 5

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

$$LM\big(\boldsymbol{\theta}_0, \mathbf{b}\big) = \frac{T\mathbf{b}^{+\prime}\bar{m}_T(\boldsymbol{\theta}_0)\bar{m}_T(\boldsymbol{\theta}_0)\mathbf{b}^{+}}{(N+2)\,\mathbf{b}^{+\prime}\mathbf{D}_T\mathbf{b}^{+}}, \tag{B4}$$

where $\mathbf{b}^{+} = (1, \mathbf{b}')'$, $\bar{m}_T(\boldsymbol{\theta}_0) = \big[\bar{m}_{kT}(\boldsymbol{\theta}_0), \bar{m}_{sT}(\boldsymbol{\theta}_0)\big]$, $\bar{m}_{kT}(\boldsymbol{\theta})$ 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 (B4) 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 $LM\big(\boldsymbol{\theta}_0, \mathbf{b}\big)$ 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. □

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

$$\lim_{\eta \to 0^+} s_{\eta^{\ddagger}t}(\boldsymbol{\phi}) = \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4}, \tag{B5}$$

$$\lim_{\eta \to 0^+} s_{\mathbf{b}^{\ddagger}t}(\boldsymbol{\phi}) = \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right], \tag{B6}$$

under normality. Note that the maximum likelihood estimate of $\eta^{\ddagger}$, which cannot be negative, will be zero when (B5) 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 (B5) and (B6) 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

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

where $\psi(\cdot)$ is the digamma function and $B(\cdot,\cdot)$ the usual beta function. □

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

$$V_{t-1}\begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0,1,\mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0,1,\mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0}) \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\pi}_0,1,\mathbf{0}) & \mathcal{I}_{\boldsymbol{\pi}\boldsymbol{b}t}(\boldsymbol{\pi}_0,1,\mathbf{0}) & \mathcal{M}_t(\boldsymbol{\pi}_0) \\ \mathcal{I}'_{\boldsymbol{\pi}\boldsymbol{b}t}(\boldsymbol{\pi}_0,1,\mathbf{0}) & V_{t-1}\left[s_{\mathbf{b}t}(\boldsymbol{\pi}_0,1,\mathbf{0})\right] & 0 \\ \mathcal{M}'_t(\boldsymbol{\pi}_0) & 0' & V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})\right] \end{bmatrix} \quad \text{(B7)}$$

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 (B7). Parts 1 of the two propositions follow immediately from the $(N+1)\times(N+1)$ submatrix of (B7) that yield the variances of the test moment conditions. To account for parameter uncertainty, consider the function

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

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

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

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

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

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

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

Hence

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

As a result

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

from which we can obtain the asymptotic distributions in the Propositions. □

C Supplementary results

Supplementary results associated with this article can be found at

ftp://ftp.cemfi.es/pdf/papers/es/gh_testing_extra_appendix.pdf

Figure 1a: Power of the normality tests under symmetric t alternatives

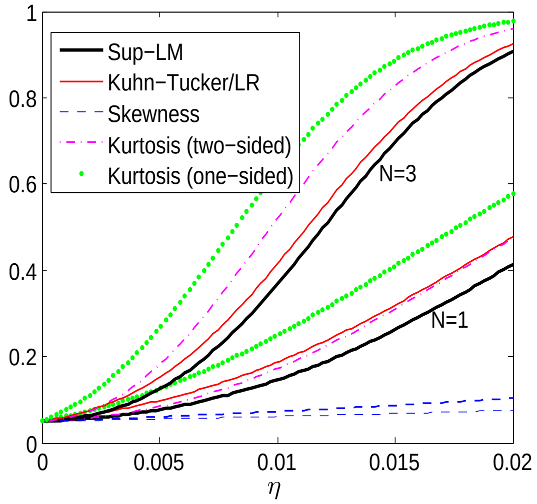


Figure 1b: Power of the normality tests under asymmetric t alternatives ($b_i = .75, \forall i$)

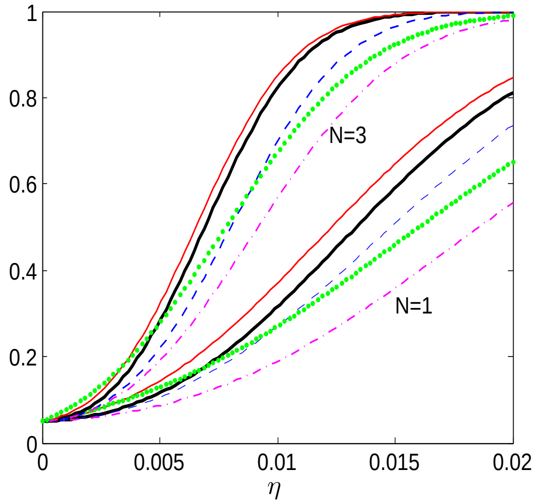


Figure 1c: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .005$, $N = 3$)

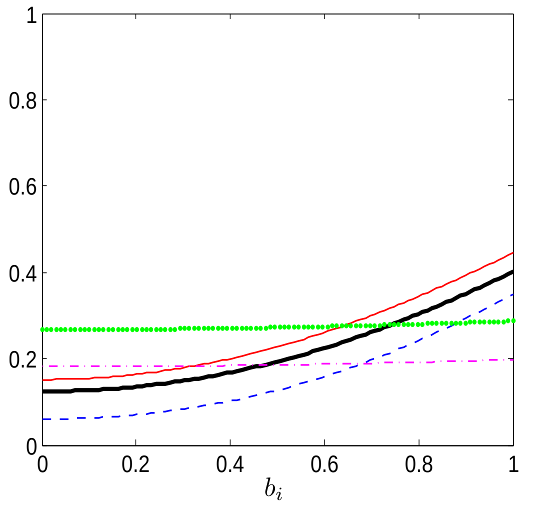


Figure 1d: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .01$, $N = 3$)

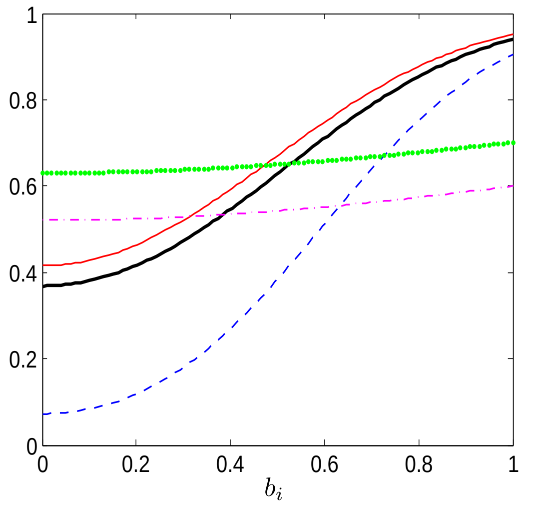


Figure 1e: Power of Sup-LM, Mardia and Lütkepohl normality tests against symmetric t alternatives ($N = 3$).

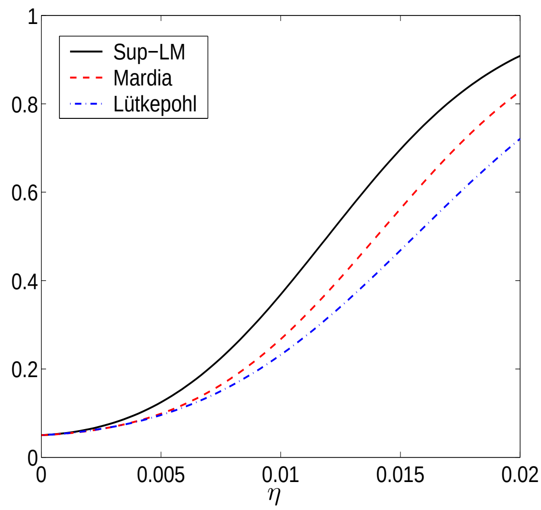
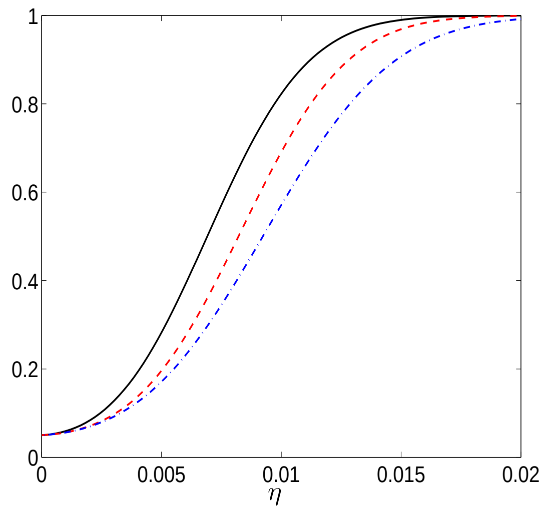


Figure 1f: Power of Sup-LM, Mardia and Lütkepohl normality tests against asymmetric t alternatives ($N = 3$).



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

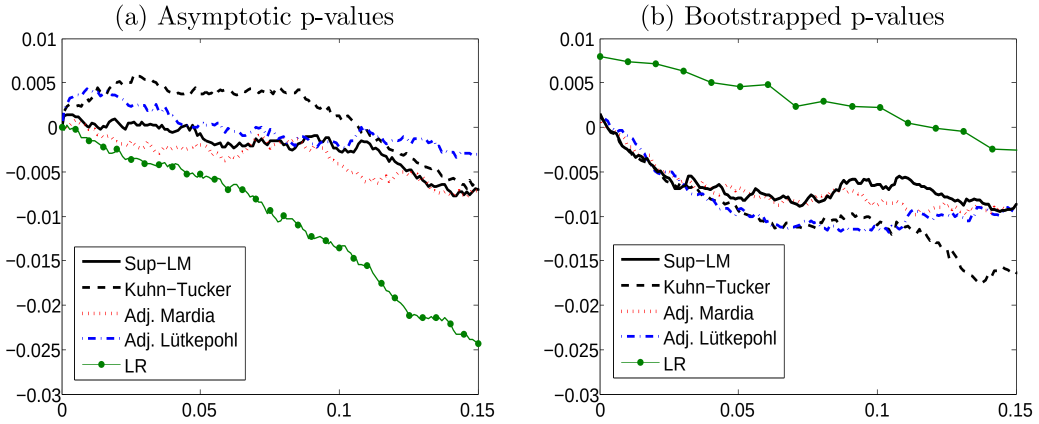


Figure 3: p-value discrepancy plots of the skewness components of the joint normality tests

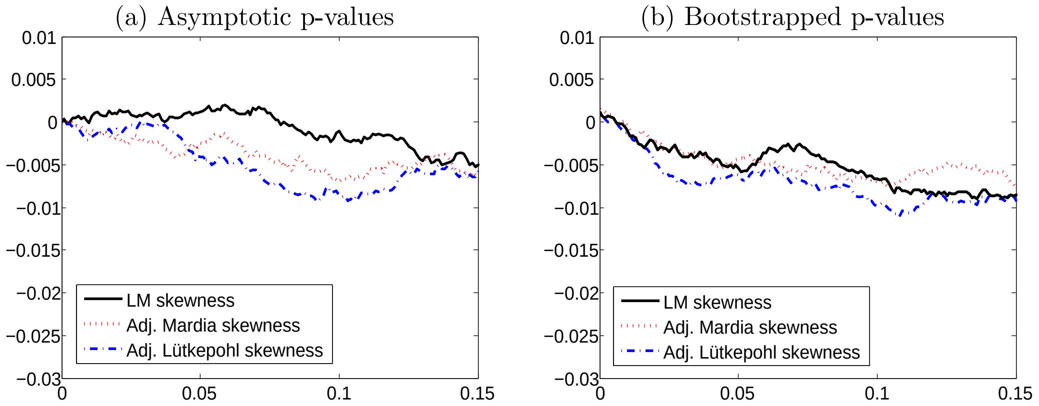
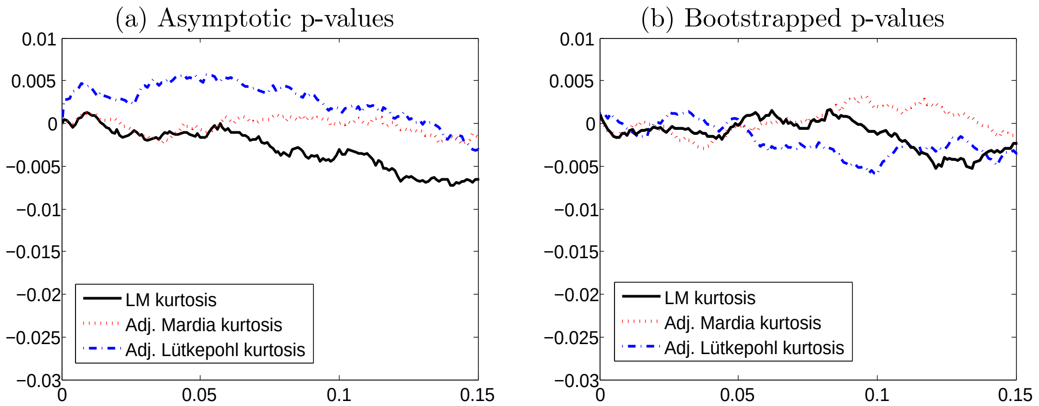


Figure 4: p-value discrepancy plots of the kurtosis components of the joint normality tests



Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations with T=1,000 and N=3. 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 joint normality tests under the null of normality

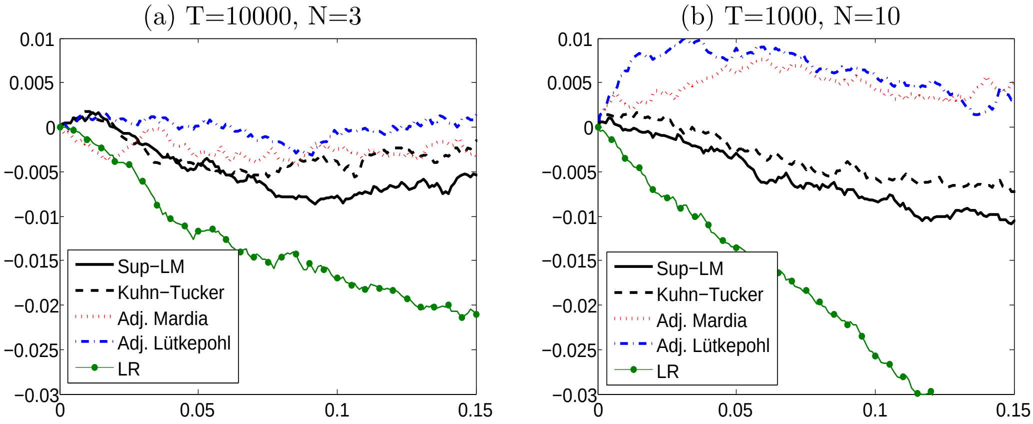


Figure 6: p-value discrepancy plots of the skewness components of the joint normality tests

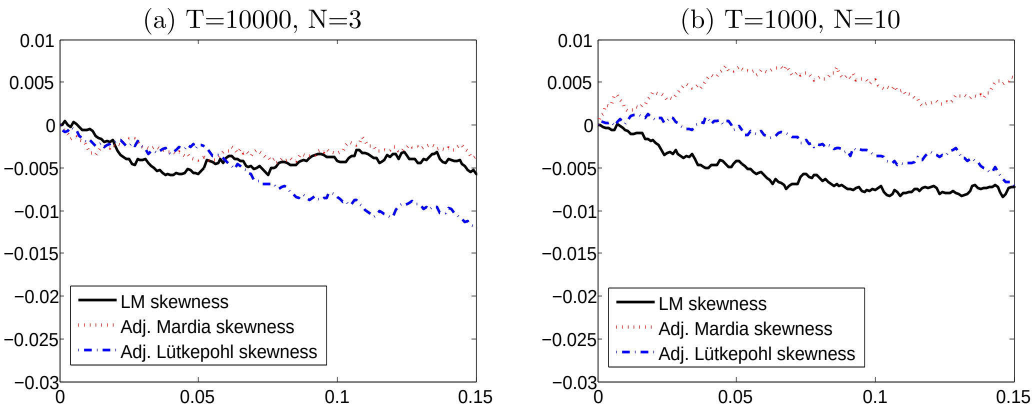
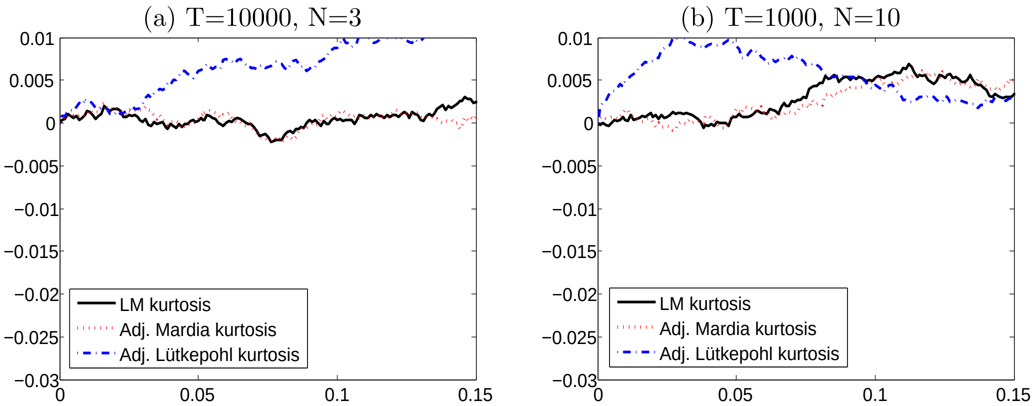


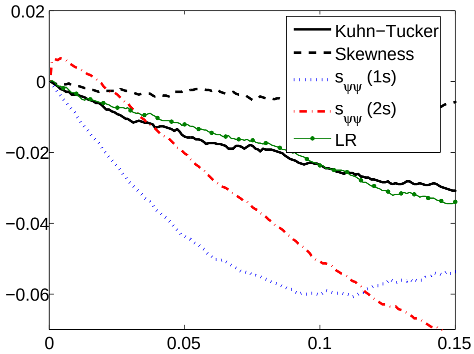
Figure 7: p-value discrepancy plots of the kurtosis components of the joint normality tests



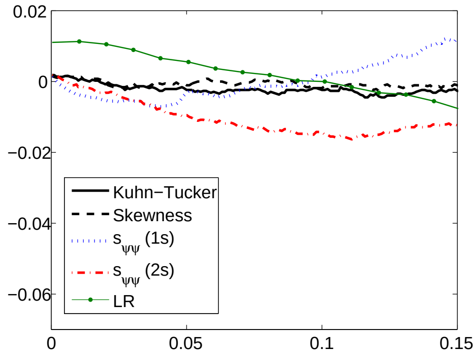
Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations.

Figure 8: p-value discrepancy plots of the Student t tests under the null of Student t innovations

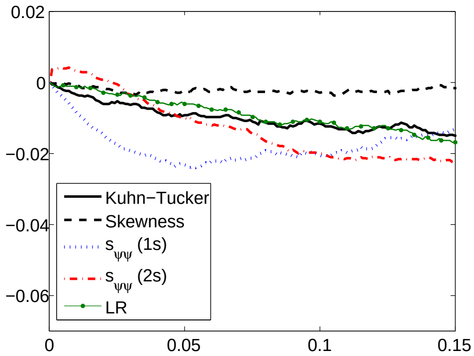
(a) Asymptotic. T=1000, N=3, $\eta = 0.1$



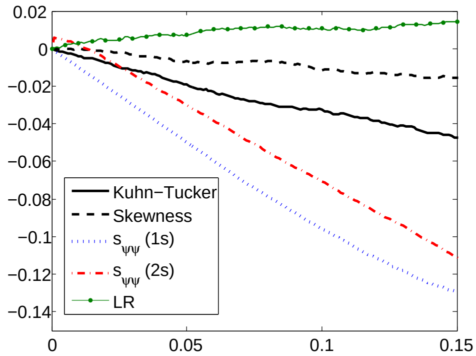
(b) Bootstrap. T=1000, N=3, $\eta = 0.1$



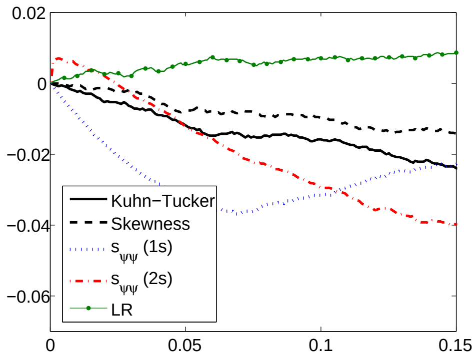
(c) Asymptotic. T=10000, N=3, $\eta = 0.1$



(d) Asymptotic. T=1000, N=3, $\eta = 0.2$

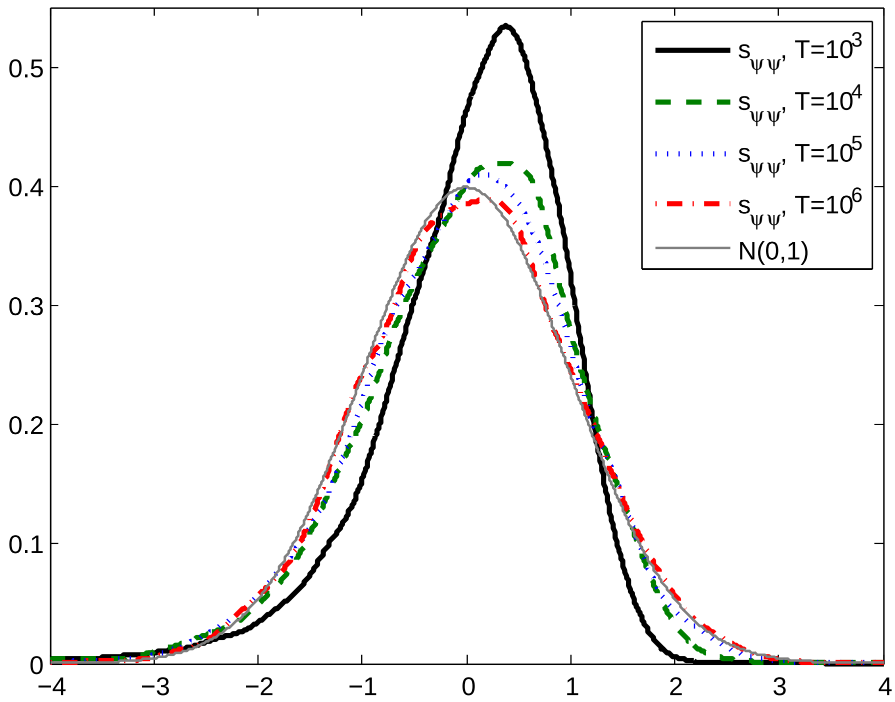


(e) Asymptotic. T=1000, N=10, $\eta = 0.1$



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

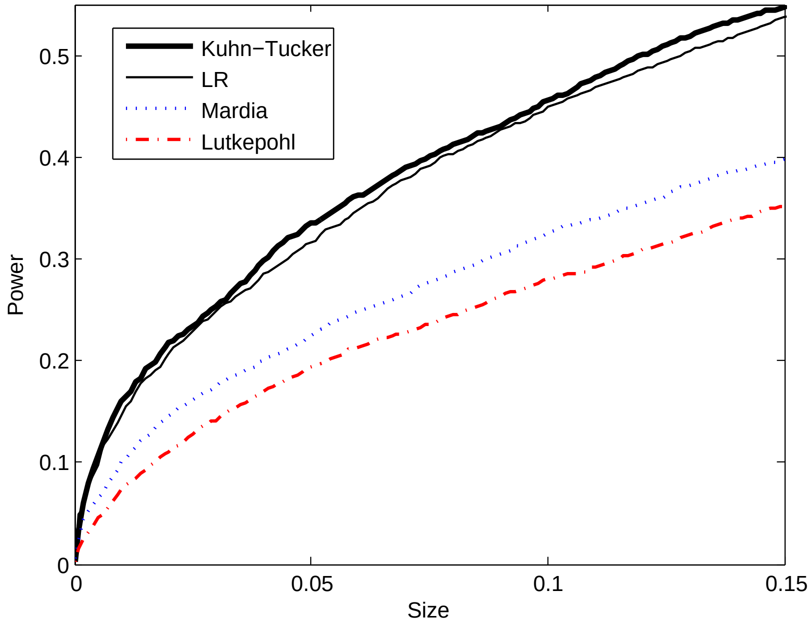
Figure 9: Kernel estimation of the density of the symmetric Student t test



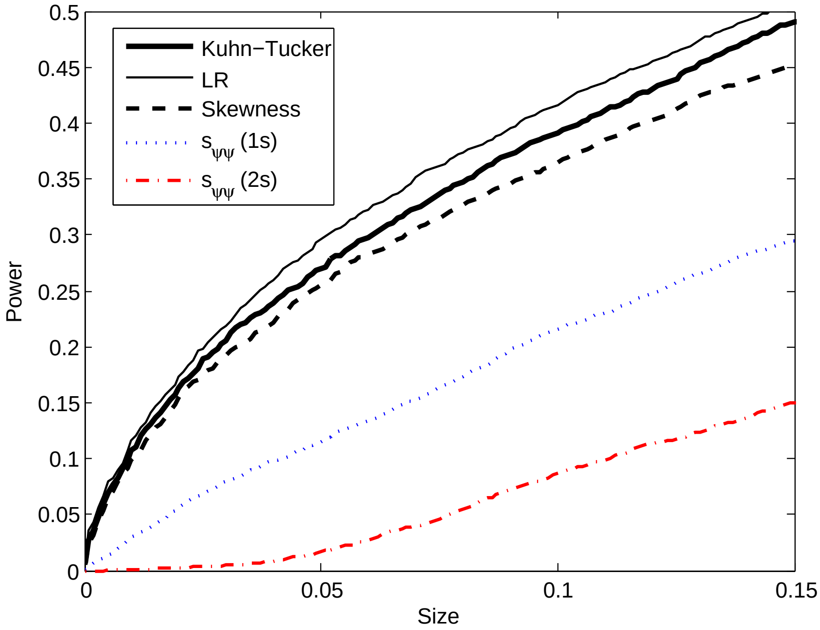
Notes: Monte Carlo study with 10,000 simulations. For T=1,000, T=10,000 and T=100,000, the test statistics have been obtained with estimated parameters. For computational reasons, the test for T=1,000,000 is based on the orthogonalised moment conditions evaluated at the true parameters. Both approaches yield almost identical kernel densities for T=10,000 and T=100,000.

Figure 10: 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.