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

Javier Mencía and Enrique Sentana

CEMFI Working Paper No. 0804

April 2008

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551 Fax (34) 914 291 056
Internet: www.cemfi.es

A previous draft of this paper was 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, S´ebastien 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 recommendations of the associate editor and two anonymous referees have also substantially improved the paper. Of course, the usual caveat applies. Financial support from Fundación Ramón Areces (Mencía) and the Spanish Ministry of Education and Science through grant SEJ2005-08880 (Sentana) is gratefully acknowledged.

CEMFI Working Paper 0804
April 2008

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

Abstract

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

JEL Codes: C12, C52, C32.
Keywords: Bootstrap, Inequality Constraints, Kurtosis, Normality Tests, Skewness, Supremum Test, Underidentified parameters.

Javier Mencía
Bank of Spain
javier.mencía@bde.es

Enrique Sentana
CEMFI
sentana@cemfi.es

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. As our alternative hypothesis, we consider a family of distributions that allow for both excess kurtosis and asymmetries in the innovations, but 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 Gamma, the Normal Inverse Gaussian, 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).

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, where N denotes the number of assets. 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. 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) and Mencía and Sentana (2008).

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

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

where $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ 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).

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 $\nu \in \mathbb{R}$, $\delta, \gamma \in \mathbb{R}^+$. Since $\boldsymbol{\varepsilon}_t^*$ given ξ_t is Gaussian with conditional mean $\boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi_t^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi_t^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. There is a further scale parameter, δ; two other scalars, ν and γ, to allow for flexible tail modelling; and the vector $\boldsymbol{\beta}$, which introduces skewness in this distribution.

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

In order to ensure that the elements of $\boldsymbol{\varepsilon}_t^*$ are uncorrelated with zero mean and unit variance by construction, we consider the standardised *version* in Mencía and Sentana

(2008). Specifically, we set $\delta = 1$, $\boldsymbol{\alpha} = -c\left(\boldsymbol{\beta}, \nu, \gamma\right)\boldsymbol{\beta}$ and

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

where

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

$R_{\nu}\left(\gamma\right) = K_{\nu+1}\left(\gamma\right)/K_{\nu}\left(\gamma\right)$, $D_{\nu+1}\left(\gamma\right) = K_{\nu+2}\left(\gamma\right)K_{\nu}\left(\gamma\right)/K^2_{\nu+1}\left(\gamma\right)$ and $K_{\nu}\left(\cdot\right)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun, 1965, p. 374). Thus, the distribution of $\boldsymbol{\varepsilon}^*_t$ depends on two shape scalar parameters, ν and γ, and a vector of N skewness parameters, denoted by $\boldsymbol{\beta}$. Under this parametrisation, the Normal distribution can be achieved in three different ways: (i) when $\nu \to -\infty$ 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, given that $\boldsymbol{\varepsilon}^*_t$ is not generally observable, the choice of "square root" matrix is not irrelevant except in univariate *GH* models, or in multivariate *GH* models in which either $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is time-invariant or $\boldsymbol{\varepsilon}^*_t$ is spherical (i.e. $\boldsymbol{\beta} = \mathbf{0}$). As discussed by Mencía and Sentana (2008), though, 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 $\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})$.[1] 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.[2]

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,[3] where $\boldsymbol{\phi}' = (\boldsymbol{\theta}', \eta, \psi, \mathbf{b}')$. As we mentioned before, we can achieve

[1]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).

[2]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.

[3]See Mencía and Sentana (2008) for explicit expressions.

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, $\lim_{\eta \cdot \psi \to 0} s_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$, coincides with the usual Gaussian expressions (see e.g. Bollerslev and Wooldridge (1992)). Further, we can show that for fixed $\psi > 0$,

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

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, which confirms the non-differentiability of the log-likelihood function with respect to η at $\eta = 0$ (see footnote 2). 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 $\eta \to 0^+$,

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

where we have excluded the terms corresponding to $\mathbf{b}$ and ψ because both $s_{\mathbf{b}t}(\boldsymbol{\phi})$ and $s_{\psi t}(\boldsymbol{\phi})$ are identically zero in the limit. As expected, the conditional variance of the component of the score corresponding to the conditional mean and variance parameters $\boldsymbol{\theta}$ coincides with the expression obtained by Bollerslev and Wooldridge (1992). Moreover, we can show that

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

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 $\mathcal{I}_{\psi\psi t}(\boldsymbol{\theta}, \eta, 0^+, \mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^+, \psi, \mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\boldsymbol{\theta}, 0^-, \psi, \mathbf{b})$.

3.3 Tests for fixed values of the underidentified parameters

The derivation of the Lagrange multiplier (LM) and Likelihood Ratio (LR) tests for multivariate normality versus *GH* innovations is complicated by two unusual features. First, since the *GH* distribution can approach the normal distribution along three different paths in the parameter space, i.e. $\eta \to 0^+$, $\eta \to 0^-$ or $\psi \to 0^+$, the null hypothesis can be posed in three different ways. In addition, some of the other parameters become increasingly underidentified along each of those three paths. In particular, η and $\mathbf{b}$ are not identified in the limit when $\psi \to 0^+$, while ψ and $\mathbf{b}$ are underidentified when $\eta \to 0^\pm$.

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 given values.

Let $\tilde{\boldsymbol{\theta}}_T$ denote the ML estimator of $\boldsymbol{\theta}$ obtained by maximising the Gaussian log-likelihood function. 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 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 2 *The LM Normality tests for fixed values of the underidentified parameters can be expressed as:*

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

where $\hat{\mathbf{\Sigma}}$ *is some consistent estimator of* $\mathbf{\Sigma}(\boldsymbol{\theta}_0) = E\left[\mathbf{\Sigma}_t(\boldsymbol{\theta}_0)\right]$, *such as* $\frac{1}{T}\sum_t \boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)\boldsymbol{\varepsilon}'_t(\tilde{\boldsymbol{\theta}}_T)$.

Under standard regularity conditions, $LM(\tilde{\boldsymbol{\theta}}_T, \mathbf{b})$ will be asymptotically chi-square with one degree of freedom for a given $\mathbf{b}$ under the null hypothesis of normality, which 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 $\mathbf{\Sigma}_t(\boldsymbol{\theta})$, as expected from (5).

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

Proposition 3 *The LR Normality tests for fixed values of the unidentified parameters* $\mathbf{b}$ *is asymptotically equivalent to the Kuhn-Tucker (KT) test*

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

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

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

3.4 The supremum tests

The approach described in the previous subsection is plausible in situations where there are values of the underidentified parameters that make sense from an economic or statistical point of view. Unfortunately, it is not at all clear a priori what values of $\mathbf{b}$ and ψ or η are likely to prevail under the alternative of *GH* innovations. For that reason, in this subsection we follow a second approach, which consists in computing either the LR or the LM test statistic for the whole range of values of the underidentified parameters, which are then combined to construct an overall test statistic (see Andrews, 1994). In our case, we compute these tests for all possible values of $\mathbf{b}$ and ψ or η 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 4 *The supremum of the LM Normality test (7) with respect to* $\mathbf{b}$ *can be expressed as*

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

$$LM_k(\tilde{\boldsymbol{\theta}}_T) = \frac{2}{N(N+2)}\left\{\frac{\sqrt{T}}{T}\sum_t \left[\frac{1}{4}\varsigma_t^2(\tilde{\boldsymbol{\theta}}_T) - \frac{N+2}{2}\varsigma_t(\tilde{\boldsymbol{\theta}}_T) + \frac{N(N+2)}{4}\right]\right\}^2, \tag{9}$$

$$\begin{aligned} LM_s(\tilde{\boldsymbol{\theta}}_T) = {} & \frac{1}{2(N+2)}\left\{\frac{\sqrt{T}}{T}\sum_t \boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)\left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - (N+2)\right]\right\}' \hat{\boldsymbol{\Sigma}}^{-1} \\ & \times \left\{\frac{\sqrt{T}}{T}\sum_t \boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)\left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - (N+2)\right]\right\}, \end{aligned} \tag{10}$$

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

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 (9) as a specification test of the restriction on the first two moments of $\varsigma_t(\boldsymbol{\theta}_0)$ implicit in

$$E\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0)\right] = E[m_{kt}(\boldsymbol{\theta}_0)] = 0, \tag{11}$$

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

$$E\left[m_{kt}(\boldsymbol{\theta}_0)\middle|\, \boldsymbol{\theta}_0, \eta_0 > 0, \psi_0 > 0\right] > 0. \tag{12}$$

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 consider a KT multiplier version of the sup LM test that exploits (12) 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 5 *The (sup) LR test of Gaussian vs. GH innovations is asymptotically equivalent under the null of normality to the following (sup) Kuhn-Tucker test:*

$$KT(\tilde{\boldsymbol{\theta}}_T) = LM_k(\tilde{\boldsymbol{\theta}}_T)\mathbf{1}\big(\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T) > 0\big) + LM_s(\tilde{\boldsymbol{\theta}}_T), \tag{13}$$

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

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. However, the LR test is computationally much 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 LM and sup KT tests is that they only require 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).

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:

$$E[\varepsilon^*_{it}(\boldsymbol{\theta}_0)\varepsilon^*_{jt}(\boldsymbol{\theta}_0)\varepsilon^*_{kt}(\boldsymbol{\theta}_0)] = 0, \quad i,j,k = 1,\cdots,N \tag{14}$$

But since $\varsigma_t(\boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)$, it is clear that (10) is also testing for co-skewness. Specifically, $LM_s(\tilde{\boldsymbol{\theta}}_T)$ is testing the N alternative moment conditions

$$E\{\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)]\} = E[m_{st}(\boldsymbol{\theta}_0)] = \mathbf{0}, \tag{15}$$

which are the relevant ones against GH innovations.

A less well known multivariate normality test was proposed by Bera and John (1983), who considered multivariate Pearson alternatives instead. However, since the asymmetric component of their test also assesses if (14) holds, we do not discuss it separately.

All these tests were derived for nonlinear regression models with conditionally homoskedastic disturbances estimated by Gaussian ML, in which the covariance matrix of the innovations, $\mathbf{\Sigma}$, is unrestricted and does not affect the conditional mean, and the conditional mean parameters, $\boldsymbol{\varrho}$ say, and the elements of $vech(\mathbf{\Sigma})$ 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), one possible way to adjust Mardia's (1970) formulae is to replace $\varepsilon_{it}^{*3}(\boldsymbol{\theta})$ by $H_3[\varepsilon_{it}^{*}(\boldsymbol{\theta})]$ and $\varepsilon_{it}^{*2}(\boldsymbol{\theta})\varepsilon_{jt}^{*}(\boldsymbol{\theta})$ by $H_2[\varepsilon_{it}^{*}(\boldsymbol{\theta})]H_1[\varepsilon_{jt}^{*}(\boldsymbol{\theta})]$ $(i \neq j)$ in the moment conditions (14), where $H_k(\cdot)$ is the Hermite polynomial of order k. Alternatively, we can correct the asymptotic size by treating (14) 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 for specific details).

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

3.5 Power of the normality test

Although we shall investigate the finite sample properties of the different multivariate normality tests in section 5, it is interesting to study their asymptotic power properties. However, since the block-diagonality of the information matrix between $\boldsymbol{\theta}$ and the other parameters is generally lost under the alternative of GH innovations, for the purposes of this exercise we only consider models in which $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\mathbf{\Sigma}_t(\boldsymbol{\theta})$ 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.[4] 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.

[4] 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 LM 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 $\mathbf{b}$,

$$s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, 0) = \frac{\eta\left[\varsigma_t(\boldsymbol{\theta}) - (N+2)\right]}{1 - 2\eta + \eta\varsigma_t(\theta)}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \tag{16}$$

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

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

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

$$s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = \frac{\eta^2}{(1-2\eta)(1-4\eta)} \frac{\varsigma_t(\boldsymbol{\theta}) - N(1-2\eta)}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} + \frac{\eta^2 [N - \varsigma_t(\boldsymbol{\theta})]}{(1-2\eta)(1+(N-2)\eta)}, \tag{17}$$

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

$$E[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) | \boldsymbol{\pi}_0, \psi_0 < 1, \mathbf{b}_0 = \mathbf{0}] > 0 \tag{18}$$

under the alternative of standardised symmetric GH innovations.

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. 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

$$\tau_{kT}(\bar{\boldsymbol{\pi}}_T) = \frac{\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})}{\sqrt{\hat{V}[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]}}, \tag{19}$$

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 hypothesis of Student t innovations with more than 4 degrees of freedom,[5] 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 7 *If $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with $\eta_0^{-1} > 4$ degrees of freedom, then*

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

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,

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

[5] If $.25 < \eta_0 < .5$ the variance of $[N - \varsigma_t(\boldsymbol{\theta})]$ becomes unbounded. Given that the expected value of this term remains 0 under the alternative hypothesis, the obvious solution is to base the test on the first component of (17) only. Exact implementation details can be found in Appendix A.

and

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

where

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

Lee and Chesher (1986) show the equivalence between (19) and the corresponding LR test 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 (18). 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.

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 $E[s_{\mathbf{b}t}(\boldsymbol{\pi},1,\mathbf{0})]=\mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}(\boldsymbol{\pi},1,\mathbf{0})$ is defined in (16). 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 8 *If $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with $\eta_0^{-1}>4$ degrees of freedom, then*

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

where

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

$$-\left[\begin{array}{cc} \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0,1,\mathbf{0}) & \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \\ \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0,1,\mathbf{0}) & \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}^{-1}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0,1,\mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \end{array}\right], \quad (20)$$

$\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi}_0,1,\mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}t}(\boldsymbol{\pi}_0,1,\mathbf{0})]$ *is the Student t information matrix derived in FSC,* $\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}}(\boldsymbol{\pi}_0,1,\mathbf{0}) = E[\mathcal{I}_{\boldsymbol{\pi}\mathbf{b}t}(\boldsymbol{\pi}_0,1,\mathbf{0})]$ *and* $\mathcal{I}_{\mathbf{bb}}(\boldsymbol{\pi}_0,1,\mathbf{0}) = E[\mathcal{I}_{\mathbf{bb}t}(\boldsymbol{\pi}_0,1,\mathbf{0})]$ *are defined in Proposition 6, and* $\mathcal{M}(\boldsymbol{\pi}_0)$ *and* $V\left[s_{\psi\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})\right]$ *are given in Proposition 7.*

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

$$\tau_{gT}(\bar{\boldsymbol{\pi}}_T) = \left[\begin{array}{c} \sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}) \\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}) \end{array}\right]' \mathcal{V}^{-1}(\bar{\boldsymbol{\pi}}_T) \left[\begin{array}{c} \sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}) \\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}) \end{array}\right], \quad (21)$$

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 $E\left[\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\right] = \mathbf{0}$, 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 6).

It is then straightforward to see that the asymptotic distribution of

$$\tau_{oT}(\bar{\boldsymbol{\pi}}_T) = T\bar{s}^{\perp\prime}_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})\left[\mathcal{V}_{\mathbf{bb}}(\bar{\boldsymbol{\pi}}_T) - \frac{\mathcal{V}_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)\mathcal{V}'_{\mathbf{b}\psi}(\bar{\boldsymbol{\pi}}_T)}{\mathcal{V}_{\psi\psi}(\bar{\boldsymbol{\pi}}_T)}\right]^{-1}\bar{s}^{\perp}_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})$$
$$+\tau^2_{kT}(\bar{\boldsymbol{\pi}}_T)\,\mathbf{1}\left[\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}) > 0\right] \quad (22)$$

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, (22) numerically coincides with (21). 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 (22) is numerically invariant to the chosen factorisation of $\mathbf{\Sigma}_t(\boldsymbol{\theta})$, as expected from (5).

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

$$\tau_{aT}\left(\bar{\boldsymbol{\pi}}_T\right) = \sqrt{T}\bar{s}'_{\mathbf{b}T}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)\mathcal{V}^{-1}_{\mathbf{bb}}\left(\bar{\boldsymbol{\pi}}_T\right)\sqrt{T}\bar{s}_{\mathbf{b}T}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right), \tag{23}$$

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, $\tau_{aT}\left(\bar{\boldsymbol{\pi}}_T\right)$ 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 (15) 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 a Student t. 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 $[(1-4\eta_0)(1-6\eta_0)]/[1+(N-2)\eta_0+2(N+4)\eta_0^2]$. Alternatively, we can run the univariate regression of 1 on $m_{st}(\bar{\boldsymbol{\theta}}_T)$, or the multivariate regression of $\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\theta}}_T)$ on $\varsigma_t(\bar{\boldsymbol{\theta}}_T) - (N+2)$, although in the latter case we should use standard errors that are robust to heteroskedasticity. Not surprisingly, we can show that these three procedures to test (15) 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 $\tau_{aT}\left(\bar{\boldsymbol{\pi}}_T\right)$ in (23), which is the proper LM test for symmetry.

Nevertheless, an interesting property of a moment test for symmetry based on (15) is that $\sqrt{T}\bar{m}_{sT}(\bar{\boldsymbol{\theta}}_T)$ and $\sqrt{T}\bar{s}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}\right)$ 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 properties

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

Finite sample sizes of the normality tests The trivariate model that we simulate and estimate under Gaussianity is given by the following equations:

$$y_{it} = \mu_i + c_i f_t + v_{it} \quad i = 1, 2, 3,$$

where $f_t = \lambda_t^{1/2} f_t^*$, $v_{it} = \gamma_{it}^{1/2} v_{it}^*$ $(i = 1, 2, 3)$,

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

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

$$\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}, \\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{cc}'\lambda_t + \boldsymbol{\Gamma}_t,
\end{aligned} \tag{24}$$

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

Given that the asymptotic distributions that we have derived in previous sections may be unreliable in finite samples, we compute both asymptotic and bootstrap p-values. In this regard, it is important to note that Andrews (2000) shows that the size of bootstrap tests remains asymptotically valid when some of the parameters are on the boundary of the parameter space, even though the usual bootstrap standard errors are not reliable in those circumstances. We consider a parametric bootstrap procedure with 1,000 samples for all tests except the LR test, for which we could only use 100 samples for computational reasons.[6] Given that the GH log-likelihood function is very flat around the normality region, we consider a grid of 20×5 different initial values for the pair (η, ψ) to maximise the likelihood under the alternative. But since it was computationally infeasible to implement a similar grid search for the vector of asymmetry parameters, we only considered a single initial $\mathbf{b}$ given by the value that leads to the sup LM test (see the proof of Proposition 4).[7]

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 $\boldsymbol{\theta}$. This orthogonality allows us to save substantial computer time because we do not need to reestimate $\boldsymbol{\theta}$ in each bootstrap sample.

Figures 2-4 summarise our findings for the different multivariate normality tests by means of Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. The left panels show the discrepancy plots of the asymptotic p-values, while the right panels show the corresponding results obtained with the parametric bootstrap. Figure 2a shows that the LR test seems to be too conservative in general, especially for large nominal sizes. In this sense, we can observe in Figure 2b that the parametric bootstrap is able to reduce those distortions to some extent.[8] 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

[6] 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. In our empirical application, in contrast, we used 1,000 bootstrap samples.

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

[8] 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.

substantially reduce the distortions.

Finite sample sizes of the Student t tests In this case we maintain the conditional mean and variance specification in (24), but generate the standardised innovations $\boldsymbol{\varepsilon}_t^*$ from a trivariate Student t distribution with 10 degrees of freedom. As before, we compare the asymptotic p-values of the tests with their bootstrapped counterparts. Again, we consider 1,000 bootstrap samples for the LM-type test, but we can only afford 100 samples for the LR test. Since we can easily orthogonalise the moment conditions of the LM test with respect to $\bar{\boldsymbol{\pi}}_T$, 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 5 shows the p-value discrepancy plots of the one- and two-sided versions of the Student t tests discussed in section 4, together with those of their asymmetric and kurtosis components, and the LR test. The most striking feature of the results for the asymptotic p-values, shown in Figure 5a, is the fact that the actual sizes of the "kurtosis" tests based on $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$, which is defined in (19), 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, as illustrated in Figure 6. In contrast, the actual sizes of the asymmetry component are very much on target. The joint tests inherit part of the size distortions of the kurtosis tests, while the LR test is also somewhat conservative. Finally, Figure 5b confirms that the parametric bootstrap is able to yield p-values that are much closer to the nominal ones.[9]

6 Empirical application

We now apply the tests derived in the previous sections to the returns on the ten US sectoral stock indices from Datastream.[10] Specifically, our data consists on daily excess returns for the period January 4th, 1988 - October 12th, 2007 (4971 observations), where we have used the Eurodollar overnight interest rate as safe rate (Datastream

[9]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.

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

code ECUSDST). The model used is a generalisation of the one in the previous section (see (24)), in which the mean dynamics are captured by a diagonal VAR(1) model with drift, and the covariance dynamics by a conditionally heteroskedastic single factor model in which the conditional variances of both common and specific factors follow GQARCH(1,1) processes to allow for leverage effects (see Sentana, 1995).

We have estimated this model under three different conditional distribution assumptions on the standardised innovations $\boldsymbol{\varepsilon}_t^*$: Gaussian, Student t and *GH*. We first estimated the model by Gaussian PML and then computed the sup LM and Kuhn-Tucker normality tests described in section 3.4, whose asymptotic and parametric bootstrap p-values are reported in Table 1b. These tests show that skewness and excess kurtosis are both very significant, although the kurtosis component is one order of magnitude larger than the skewness test.

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

Finally, we re-estimated the model under the assumption that the conditional distribution of the innovations is *GH* by using the analytical expressions for the score in Mencía and Sentana (2008). In this case, the *GH* log-likelihood introduces as additional parameters ψ and the ten-dimensional vector $\mathbf{b}$. Since the ML estimate of ψ reported in Table 1a is 1, and $\hat{\eta}_T$ is positive, the estimated *GH* conditional distribution effectively corresponds to an asymmetric t.

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

Conceivably, though, the rejection of the null hypotheses of normal and Student t innovations that we find could be exacerbated by misspecification of the first and especially second conditional moments. If our specification of the model dynamics is correct, however, the marked distributional differences that we have found should not affect the consistency of the Gaussian PML estimators of $\boldsymbol{\theta}$. With this in mind, we compare, the multivariate Gaussian estimate of the conditional variance with the one obtained with a univariate model for the equally weighted portfolio. Specifically, the univariate model is a Gaussian AR(1)-GQARCH(1,1) model. Reassuringly, Figure 7 shows that the (log)standard deviations of the two series display a very similar pattern, although the univariate estimates are somewhat noisier.

7 Conclusions

In this paper, we derive LM-type 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 also decompose our proposed LM test statistics into skewness and kurtosis components, from which we obtain more powerful one-sided KT versions that are asymptotically equivalent to the LR test.

We assess the finite sample size properties of both 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.

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

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

References

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

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

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

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

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

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.

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

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

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

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.

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A Student t tests for $\eta \geq .25$

A.1 Student t vs symmetric *GH innovations*

When $\eta \geq .25$, $s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0})$ has infinite variance. Thus, we can no longer apply a central limit theorem to find the asymptotic distribution of $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$. The source of this problem is the presence of $N - \varsigma_t(\boldsymbol{\theta})$ in the second component of $s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0})$. Note that this second component has zero mean both under the null and under the alternative. We will develop a modified version of $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$, which will be based exclusively on the first part of $s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0})$ multiplied by $(1 - 4\eta)$:

$$ms_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = \frac{\eta^2}{(1-2\eta)} \frac{\varsigma_t(\boldsymbol{\theta}) - N(1-2\eta)}{1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})}, \tag{A1}$$

This test will not only be valid for $\eta \geq .25$, but also for lower values. However, we recommend using $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ in the latter case, since it will be more powerful against *GH* alternatives. In what follows, we will derive the Student t tests using (A1) instead of (17). The proofs are analogous to those of the results in Section 4.

It is not difficult to show that the expected value of $ms_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) - s_{\psi\psi t}(\boldsymbol{\pi}, 1, \mathbf{0})$ is zero under the null and under the alternative. Hence, under the null of standardised Student t innovations with η_0^{-1} degrees of freedom,

$$E\left[ms_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\,|\mathbf{z}_t, I_{t-1}, \boldsymbol{\pi}_0, \psi_0 = 1, \mathbf{b}_0 = \mathbf{0}\right] = 0,$$

while under the alternative of standardised symmetric *GH* innovations

$$E\left[ms_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\,|\boldsymbol{\pi}_0, \psi_0 < 1, \mathbf{b}_0 = \mathbf{0}\right] > 0. \tag{A2}$$

Furthermore, it can be shown that, if $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with η_0^{-1} degrees of freedom, then, for $\eta < .5$,

$$\sqrt{T}\overline{ms}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \xrightarrow{d} N\left\{0, V\left[ms_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})\right] - \mathcal{M}_m'(\boldsymbol{\pi}_0)\mathcal{I}_{\boldsymbol{\pi\pi}}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}_m(\boldsymbol{\pi}_0)\right\}, \tag{A3}$$

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

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

and

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

where

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

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

Thus, the modified 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

$$\tau_{mkT}\left(\bar{\boldsymbol{\pi}}_T\right) = \frac{\sqrt{T}\overline{ms}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)}{\hat{V}^{\frac{1}{2}}\left[ms_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\right]},$$

where $\hat{V}\left[ms_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T\right)\right]$ is a consistent estimator of the asymptotic variance of $ms_{\psi\psi t}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)$ given in (A3). Under the null hypothesis of Student t innovations, it is easy to see that the asymptotic distribution of $\tau_{mkT}\left(\bar{\boldsymbol{\pi}}_T\right)$ will be $N(0,1)$. But given (A2), since ψ can only be less than 1 under the alternative, a one-sided test against $H_1:\psi<1$ should again be more powerful in this context (see Andrews, 2001). Specifically, we should use

$$\tau_{mkT}\left(\bar{\boldsymbol{\pi}}_T\right)\mathbf{1}\left[\sqrt{T}\overline{ms}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})>0\right].$$

A.2 Student t vs asymmetric *GH innovations*

By construction, the modified extremum test discussed in the previous subsection 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. In particular, the only thing that we need to do is to include $E[s_{\mathbf{b}t}\left(\boldsymbol{\pi},1,\mathbf{0}\right)]=\mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}\left(\boldsymbol{\pi},1,\mathbf{0}\right)$ is defined in (16). In this sense, it can be shown that, if $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with η_0^{-1} degrees of freedom, then

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

where

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

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

Proposition 6, and $V\left[ms_{\psi\psi t}(\boldsymbol{\pi}_0,1,\mathbf{0})\right]$ and $\mathcal{M}_m(\boldsymbol{\pi}_0)$ are given in Section A.1. Therefore, if we consider a two-sided test, we will use

$$\tau_{mgT}\left(\bar{\boldsymbol{\pi}}_T\right)=\left[\begin{array}{c}\sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\\ \sqrt{T}\overline{ms}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\end{array}\right]'\mathcal{V}_m^{-1}\left(\bar{\boldsymbol{\pi}}_T\right)\left[\begin{array}{c}\sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\\ \sqrt{T}\overline{ms}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\end{array}\right], \tag{A4}$$

which is distributed as a chi-square with $N+1$ degrees of freedom under the null of Student t innovations. Alternatively, we can again exploit the one-sided nature of the ψ-component of the test. However, since $\mathcal{V}_m\left(\boldsymbol{\pi}_0\right)$ is not block diagonal in general, we must first orthogonalise the moment conditions. Specifically, instead of using directly the score with respect to $\mathbf{b}$, we consider

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

whose sample average is asymptotically orthogonal to $\sqrt{T}\overline{ms}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)$ by construction. Note, however, that there is no need to do this orthogonalisation when $E\left[\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\right]=\mathbf{0}$, since in this case $\mathcal{V}_{m\mathbf{b}\psi}\left(\pi_0\right)=\mathbf{0}$ because $\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0,1,0)=\mathbf{0}$.

It is then straightforward to see that the asymptotic distribution of

$$\begin{aligned}\tau_{oT}\left(\bar{\boldsymbol{\pi}}_T\right)=T\bar{s}_{\mathbf{b}t}^{m,\perp\prime}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)&\left[\mathcal{V}_{\mathbf{bb}}\left(\bar{\boldsymbol{\pi}}_T\right)-\frac{\mathcal{V}_{m\mathbf{b}\psi}\left(\bar{\boldsymbol{\pi}}_T\right)\mathcal{V}_{m\mathbf{b}\psi}'\left(\bar{\boldsymbol{\pi}}_T\right)}{\mathcal{V}_{\psi\psi}\left(\bar{\boldsymbol{\pi}}_T\right)}\right]^{-1}\bar{s}_{\mathbf{b}t}^{m,\perp}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)\\ &+\tau_{kT}\left(\bar{\boldsymbol{\pi}}_T\right)\mathbf{1}\left[\overline{ms}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})>0\right],\end{aligned} \tag{A5}$$

is another 50:50 mixture of chi-squares with N and $N+1$ degrees of freedom under the null, because asymptotically, the probability that $\overline{ms}_{\psi\psi T}\left(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0}\right)$ is negative will be .5 if $\psi_0=1$. Such a one-sided test benefits from the fact that a non-positive $\overline{ms}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})$ gives no evidence against the null, in which case we only need to consider the orthogonalised skewness component. In contrast, when $\overline{ms}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T,1,\mathbf{0})$ is positive, (A5) numerically coincides with (A4).

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 Mencía and Sentana, 2008). 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. $\square$

Proposition 2

For fixed $\mathbf{b}$ and ψ, the LM_1 test is based on the average scores with respect to η and $\boldsymbol{\theta}$ evaluated at 0^+ and the Gaussian maximum likelihood estimates $\tilde{\boldsymbol{\theta}}_T$. But since the average score with respect to $\boldsymbol{\theta}$ will be 0 at those parameter values, and the conditional information matrix is block diagonal, the formula for the test is trivially obtained. The proportionality of the log-likelihood scores corresponding to η evaluated at $0^{\pm}$ and $\tilde{\boldsymbol{\theta}}_T$ with the score corresponding to ψ evaluated at 0^+ and $\tilde{\boldsymbol{\theta}}_T$ leads to the desired result. $\square$

Proposition 3

Consider initially the situation in which we fix $\mathbf{b}$ and ψ, and only allow η to be positive under the alternative. The first thing to note is that such a LR ratio will be identically 0 if the sample average of (6) evaluated at the Gaussian PML estimators is negative, which happens approximately half the time in large samples. Therefore, the results in Gouriéroux, Holly, and Monfort (1980) imply that the LR test will not be asymptotically equivalent to the corresponding LM test $LM_1\big(\tilde{\boldsymbol{\theta}}_T, \psi, \mathbf{b}\big)$, but rather to the Kuhn-Tucker test

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

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

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

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

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

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:

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

as required. □

Proposition 4

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

$$LM\big(\tilde{\boldsymbol{\theta}}_T,\mathbf{b}\big)=\frac{T\mathbf{b}^{+\prime}\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)\mathbf{b}^{+}}{(N+2)\,\mathbf{b}^{+\prime}\mathbf{D}_T\mathbf{b}^{+}},\tag{B6}$$

where $\mathbf{b}^{+}=(1,\mathbf{b}')'$, $\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)=\big[\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T),\bar{m}_{sT}(\tilde{\boldsymbol{\theta}}_T)\big]$, $\bar{m}_{kT}(\boldsymbol{\theta})$ and $\bar{m}_{sT}(\boldsymbol{\theta})$ are the sample means of $m_{kt}(\boldsymbol{\theta})$ and $m_{st}(\boldsymbol{\theta})$, which are defined in (11) and (15), respectively, and

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

But since the maximisation of (B6) 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 $N/[2\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T)]$ 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 $LM\big(\tilde{\boldsymbol{\theta}}_T,\mathbf{b}\big)$ yields the required result. Finally, 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. □

Proposition 5

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

$$\lim_{\eta\to 0^+} s_{\eta^{\ddagger}t}(\boldsymbol{\phi}) = \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta})-\frac{N+2}{2}\varsigma_t(\boldsymbol{\theta})+\frac{N\,(N+2)}{4},\tag{B7}$$

$$\lim_{\eta\to 0^+} s_{\mathbf{b}^{\ddagger}t}(\boldsymbol{\phi}) = \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left[\varsigma_t(\boldsymbol{\theta})-(N+2)\right],\tag{B8}$$

under normality. Note that the maximum likelihood estimate of $\eta^{\ddagger}$, which cannot be negative, will be zero when (B7) is negative, which approximately happens half the time in large samples. Hence, we need to consider the partially one-sided test (13) to obtain a test equivalent to the LR test. Furthermore, (B7) and (B8) will be asymptotically independent under normality.

Proposition 6

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. $\square$

Propositions 7 and 8

We can use again the results of Fiorentini and Sentana (2007) mentioned in the proof of Proposition 6, together with the results in Crowder (1976), to show that

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

where

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

under the null hypothesis of Student t innovations. To account for parameter uncertainty, consider the function

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

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

We can now derive the required asymptotic distribution by means of the usual Taylor expansion around the true values of the parameters

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

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

where it can be tediously shown by means of the Barlett identities that

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

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 Power of the normality tests

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 1

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

	Gaussian		Student t		Asymmetric t	
Parameter		SE		SE		SE
η	0	-	0.095	0.003	0.095	0.004
ψ	0	-	1	-	1	-
Log-likelihood	-53132.29		-52008.98		-51997.25	

(b) Normality tests

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

(c) Student t tests

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

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

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

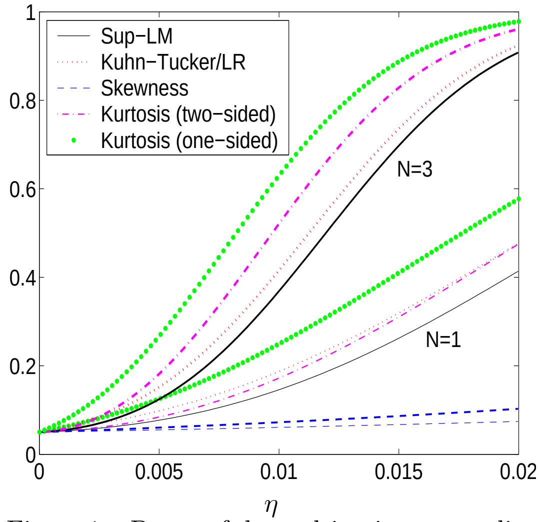


Figure 1b: Power of the normality tests under asymmetric t alternatives ($b_i = .75$, $\forall i$)

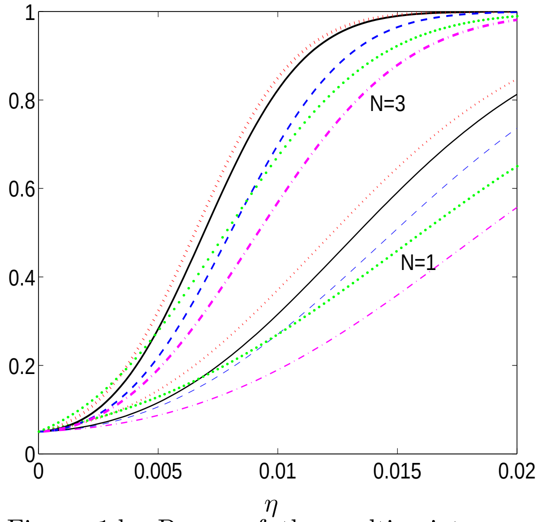


Figure 1c: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .005$, $N = 3$)

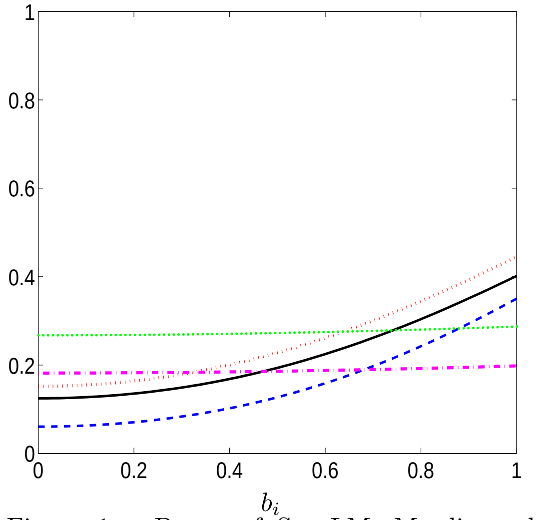


Figure 1d: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .01$, $N = 3$)

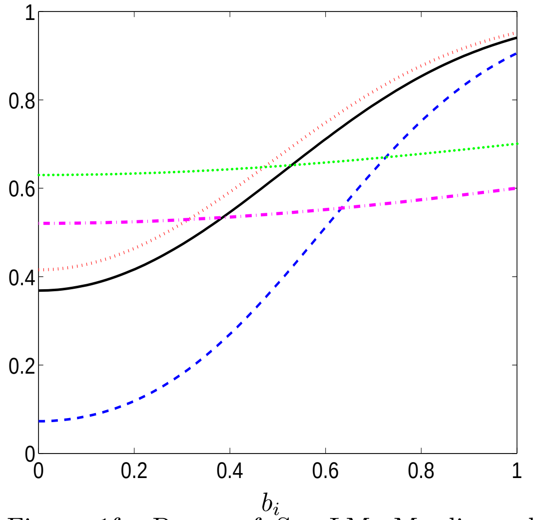


Figure 1e: Power of Sup-LM, Mardia and Lütkepohl normality tests against symmetric t alternatives ($N = 3$).

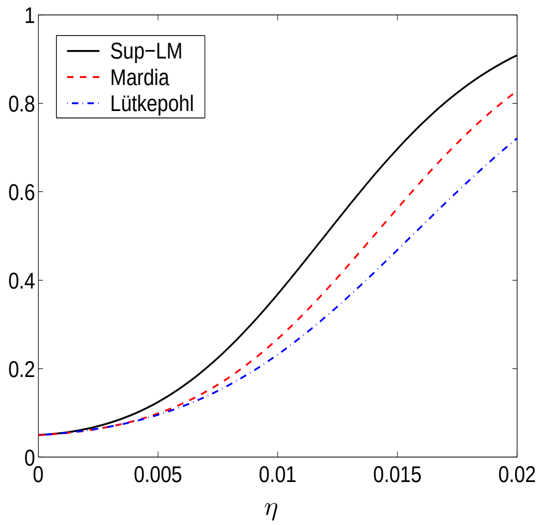
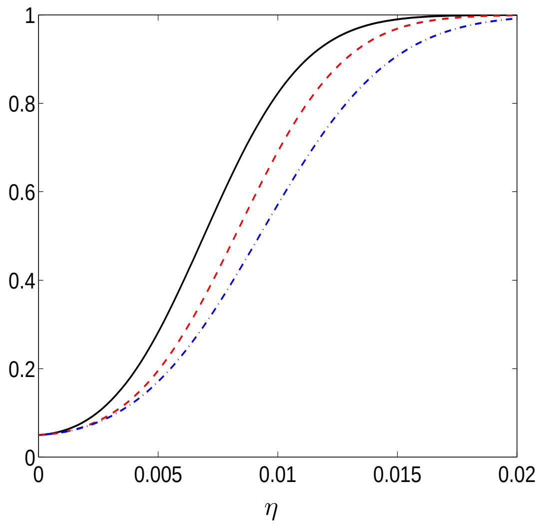


Figure 1f: Power of Sup-LM, Mardia and Lütkepohl normality tests against asymmetric t alternatives ($N = 2$).



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

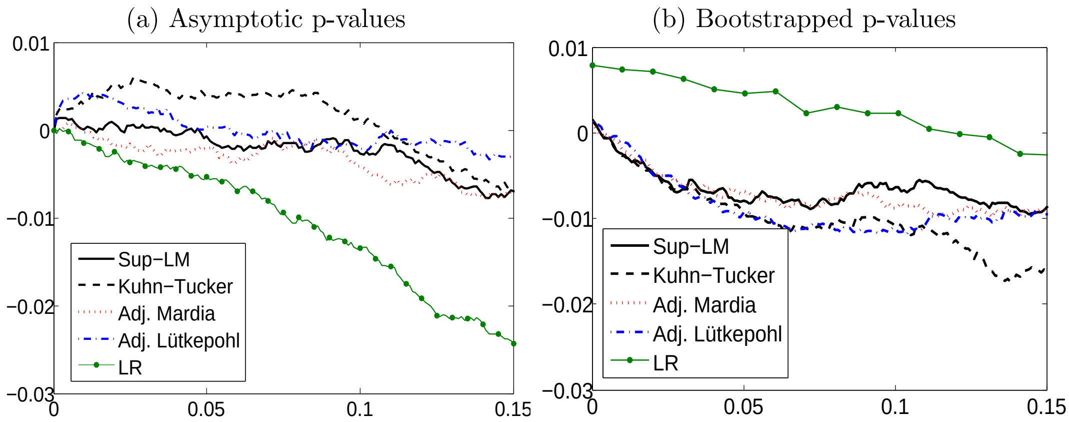


Figure 3: p-value discrepancy plots of the skewness components of the joint normality tests

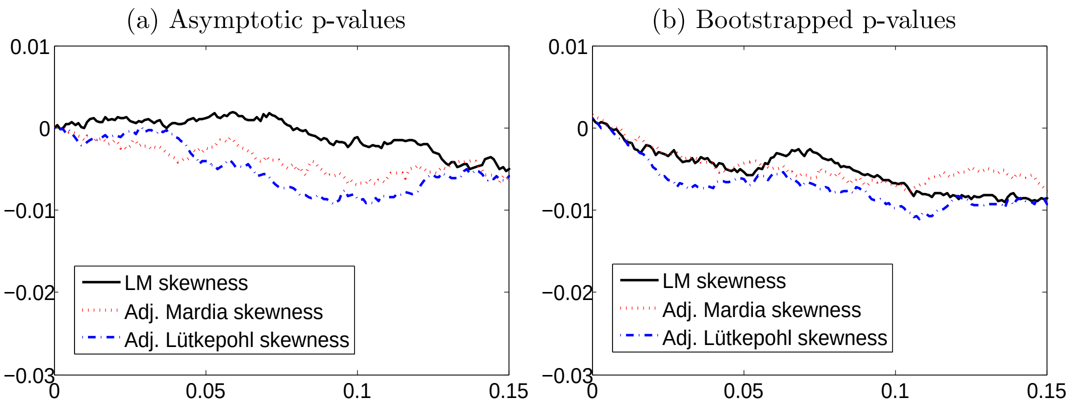
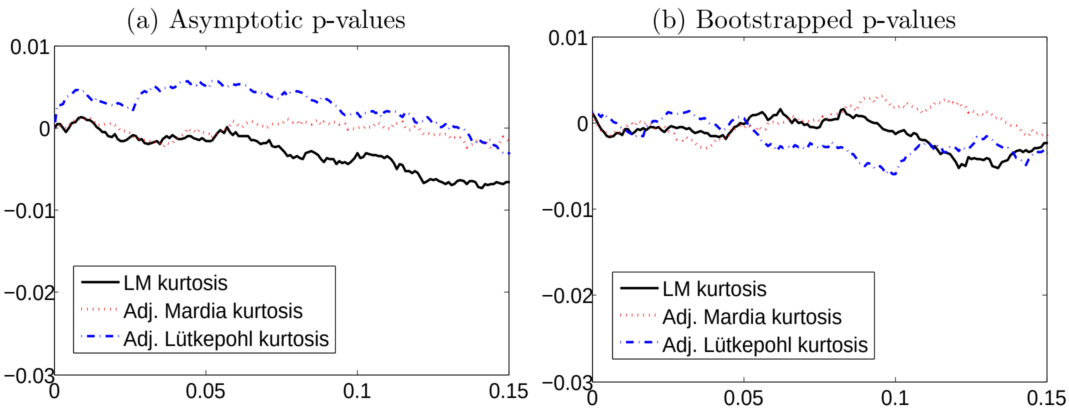


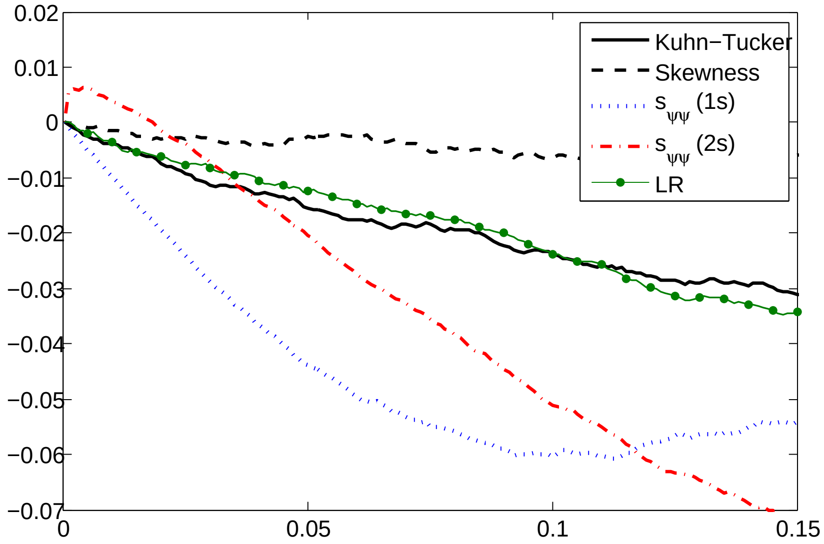
Figure 4: p-value discrepancy plots of the kurtosis components of the joint normality tests



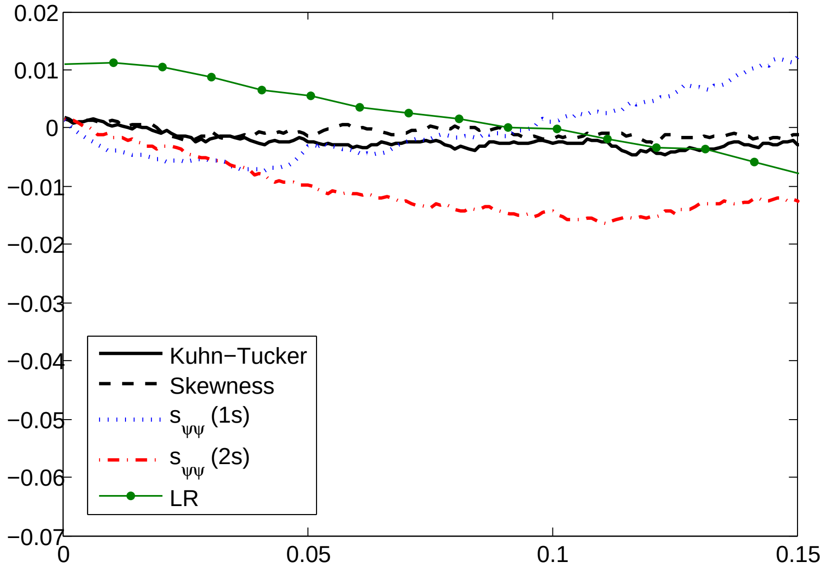
Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations with T=1,000. Parametric bootstraped p-values are computed from 1,000 samples for all the tests except the LR, which is based on 100 only.

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

(a) Asymptotic p-values

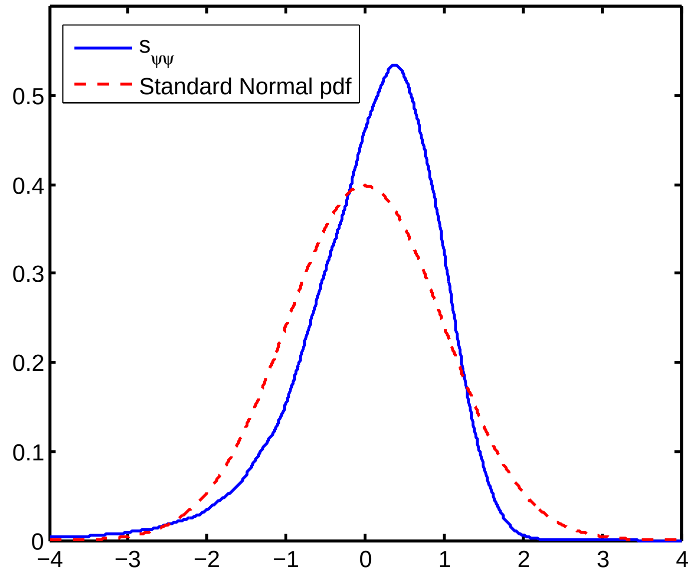


(b) Bootstrapped p-values



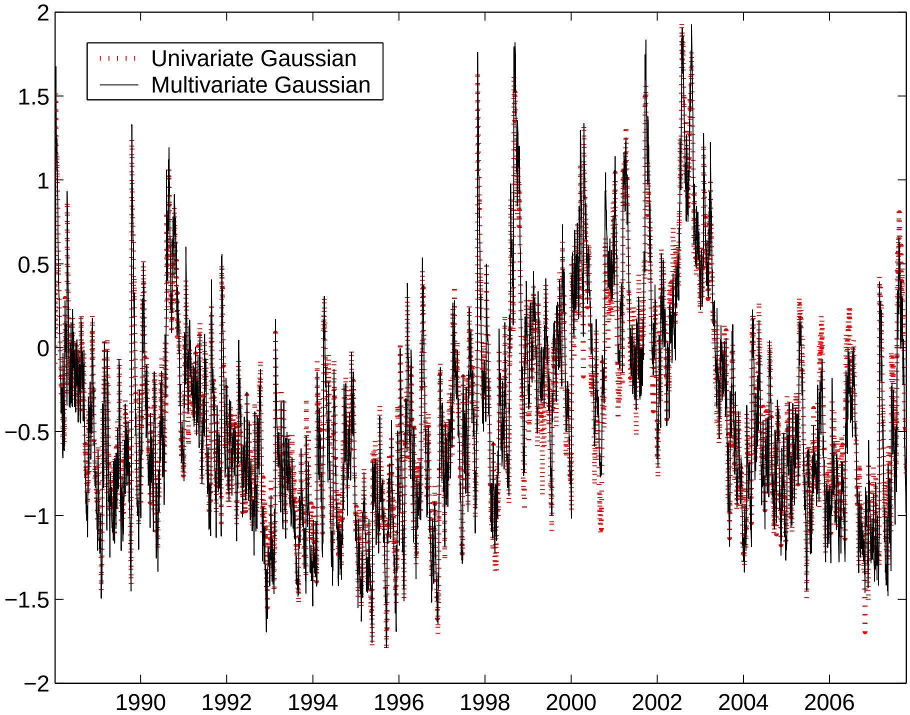
Notes: p-value discrepancy plots obtained from a Monte Carlo study with 10,000 simulations with T=1,000. Parametric bootstraped p-values are computed from 1,000 samples for all the tests except the LR, which is based on 100 only.

Figure 6: Kernel estimation of the density of the symmetric Student t test



Notes: Monte Carlo study with 10,000 simulations with T=1,000.

Figure 7: Comparison of univariate and multivariate Gaussian estimates of the (log)standard deviation of the Datastream equally weighted portfolio



CEMFI WORKING PAPERS

0601 *Beatriz Domínguez, Juan José Ganuza and Gerard Llobet*: “R&D in the pharmaceutical industry: a world of small innovations”.

0602 *Guillermo Caruana and Liran Einav*: “Production targets”.

0603 *Jose Ceron and Javier Suarez*: “Hot and cold housing markets: International evidence”.

0604 *Gerard Llobet and Michael Manove*: “Network size and network capture”.

0605 *Abel Elizalde*: “Credit risk models I: Default correlation in intensity models”.

0606 *Abel Elizalde*: “Credit risk models II: Structural models”.

0607 *Abel Elizalde*: “Credit risk models III: Reconciliation reduced – structural models”.

0608 *Abel Elizalde*: “Credit risk models IV: Understanding and pricing CDOs”.

0609 *Gema Zamarro*: “Accounting for heterogeneous returns in sequential schooling decisions”.

0610 *Max Bruche*: “Estimating structural models of corporate bond prices”.

0611 *Javier Díaz-Giménez and Josep Pijoan-Mas*: “Flat tax reforms in the U.S.: A boon for the income poor”.

0612 *Max Bruche and Carlos González-Aguado*: “Recovery rates, default probabilities and the credit cycle”.

0613 *Manuel Arellano and Jinyong Hahn*: “A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects”.

0614 *Manuel Arellano and Stéphane Bonhomme*: “Robust priors in nonlinear panel data models”.

0615 *Laura Crespo*: “Caring for parents and employment status of European mid-life women”.

0701 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: “Royalty stacking in high tech industries: separating myth from reality”.

0702 *Anne Layne-Farrar, A. Jorge Padilla and Richard Schmalensee*: “Pricing patents for licensing in standard setting organizations: Making sense of *FRAND* commitments”.

0703 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: “The *ex ante* auction model for the control of market power in standard setting organizations”.

0704 *Abel Elizalde*: “From Basel I to Basel II: An analysis of the three pillars”.

0705 *Claudio Michelacci and Josep Pijoan-Mas*: “The effects of labor market conditions on working time: the US-UE experience”.

0706 *Robert J. Aumann and Roberto Serrano*: “An economic index of *riskiness*”.

0707 *Roberto Serrano*: “El uso de sistemas dinámicos estocásticos en la Teoría de Juegos y la Economía”.

0708 *Antonio Cabrales and Roberto Serrano*: “Implementation in adaptive better-response dynamics”.

0709 *Roberto Serrano*: “Cooperative games: Core and Shapley value”.

0710 *Allan M. Feldman and Roberto Serrano*: “Arrow’s impossibility theorem: Preference diversity in a single-profile world”.

0711 *Victor Aguirregabiria and Pedro Mira*: “Dynamic discrete choice structural models: A Survey”.

0712 *Rene Saran and Roberto Serrano*: “The evolution of bidding behaviour in private-values auctions and double auctions”.

0713 *Gabriele Fiorentini and Enrique Sentana:* “On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models”.

0714 *Antonio Díaz de los Ríos and Enrique Sentana:* “Testing uncovered interest parity: A continuous-time approach”.

0715 *Francisco Peñaranda and Enrique Sentana:* “Duality in mean-variance frontiers with conditioning information”.

0716 *Stefano Gagliarducci, Tommaso Nannicini and Paolo Naticchioni:* “Electoral rules and politicians’ behavior: A micro test”.

0717 *Laura Hospido*: “Modelling heterogeneity and dynamics in the volatility of individual wages”.

0718 *Samuel Bentolila, Juan J. Dolado and Juan F. Jimeno*: “Does immigration affect the Phillips curve? Some evidence for Spain”.

0719 *Enrique Moral-Benito*: “Determinants of economic growth: A Bayesian panel data approach”.

0801 *David Martinez-Miera and Rafael Repullo*: “Does competition reduce the risk of bank failure?”.

0802 *Joan Llull*: “The impact of immigration on productivity”.

0803 *Cristina López-Mayán:* “Microeconometric analysis of residential water demand”.

0804 *Javier Mencía and Enrique Sentana:* “Distributional tests in multivariate dynamic models with Normal and Student *t* innovations”.