Estimation and Testing of Dynamic Models with Generalised Hyperbolic Innovations*

F. Javier Mencía
CEMFI
<mencia@cemfi.es>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

February 2004
Revised: June 2004

Abstract

We analyse the Generalised Hyperbolic distribution as a model for fat tails and asymmetries in multivariate conditionally heteroskedastic dynamic regression models. We provide a standardised version of this distribution, obtain analytical expressions for the log-likelihood score, and explain how to evaluate the information matrix. In addition, we derive tests for the null hypotheses of multivariate normal and Student t innovations, and decompose them into skewness and kurtosis components, from which we obtain more powerful one-sided versions. Finally, we present an empirical illustration with UK sectorial stock returns, which suggests that their conditional distribution is asymmetric and leptokurtic.

Keywords: Inequality Constraints, Kurtosis, Multivariate Normality Test, Skewness, Student t, Tail Dependence.

JEL: C52, C22, C32

*We are grateful to Manuel Arellano, Ole Barndorff-Nielsen, Anil Bera, Rob Engle, Gabriele Fiorentini, Eric Ghysels, Eric Renault, Neil Shephard and Mervyn Sylvapulle, as well as seminar audiences at Monash and the 2004 CIRANO-CIREQ Financial Econometrics Conference for helpful suggestions. Of course, the usual caveat applies. Mencía acknowledges financial support from Fundación Ramón Areces. Address for correspondence: Casado del Alisal 5, E-28014 Madrid, Spain, tel: +34 91 429 05 51, fax: +34 91 429 1056.

1 Introduction

The Basel Capital Adequacy Accord, now under revision, forced banks and other financial institutions to develop models to quantify all their risks accurately. In practice, most institutions chose the so-called Value at Risk (V@R) framework to determine the capital necessary to cover their market risk exposure. As is well known, the V@R of a portfolio is defined as the positive threshold value V such that the probability of the portfolio suffering a reduction in wealth larger than V over some fixed time interval equals some prespecified level $\varkappa < 1/2$. Undoubtedly, the most successful V@R methodology was developed by the RiskMetrics Group (1996). A key assumption of this methodology, though, is that the distribution of the returns on primitive assets, such as stocks and bonds, can be well approximated by a multivariate normal distribution after controlling for predictable time-variation in their covariance matrix. However, many empirical studies with financial time series data indicate that the distribution of asset returns is clearly non-normal even after taking volatility clustering effects into account. And although it is true that we can obtain consistent estimators of the conditional mean and variance parameters irrespective of the validity of the assumption of normality by using the Gaussian pseudo-maximum likelihood (PML) procedure advocated by Bollerslev and Wooldridge (1992) among others, the resulting V@R estimates could be substantially biased if the extreme tails accumulate more density than a normal distribution can allow for. 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. Excess kurtosis implies that extraordinary gains or losses are more common than what a normal distribution predicts. Analogously, if we assume zero mean returns for simplicity, positive (negative) skewness indicates a higher (lower) probability of experiencing large gains than large losses of the same magnitude. Therefore, the effects of non-normality are especially noticeable in the tails of the distribution. In a recent paper, Fiorentini, Sentana and Calzolari (2003a) (FSC) discuss the use of the multivariate Student t distribution to model excess kurtosis. Despite its attractiveness, though, the multivariate Student t distribution, 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 often lead to inconsistent parameter estimates (see Newey and Steigerwald, 1997). In this context, the main objective of our paper is to assess the adequacy of the distributional assumption made by FSC and other authors by considering an alternative family of distributions which allows for both excess kurtosis and asymmetries in the innovations, but which at the same time nests the multivariate Student t and Gaussian distributions. Specifically, we will use the Generalised Hyperbolic *(GH)* distribution (see Barndorff-Nielsen and Shephard, 2001a and Prause, 1998), which is a rather flexible asymmetric multivariate distribution that to the best of our knowledge has not yet been used for modelling the conditional distribution of financial time series. Formally, the *GH* distribution can be understood as a location-scale mixture of a multivariate Gaussian vector, in which the positive mixing variable follows a Generalised Inverse Gaussian *(GIG)* distribution (see Jørgensen, 1982, and Johnson, Kotz, and Balakrishnan, 1994, for details, as well as appendix D).

Our approach differs from Bera and Premaratne (2002), who also nest the Student t distribution by using Pearson's type IV distribution in univariate static models. However, these authors do not explain how to extend their approach in multivariate contexts, nor do they consider dynamic models explicitly. Our approach also differs from Bauwens and Laurent (2002), who introduce skewness by "stretching" the multivariate Student t distribution differently in different orthants. However, the larger the dimension of the random vectors, the more difficult the implementation of their technique becomes, as the number of orthants is 2^N, where N denotes the number of assets.

The rest of the paper is organised as follows. We first give an overview of the original *GH* distribution in section 2.1, and explain how to reparametrise it so that it has zero mean and unitary covariance matrix. Then, in section 2.2 we describe the econometric model under analysis, while in sections 2.3, 2.4 and 2.5 we discuss the computation of the log-likelihood function, its score, and the information matrix, respectively. Section 3 focuses on testing distributional assumptions. In particular, we develop tests for both multivariate normal and multivariate Student t innovations against *GH* alternatives in sections 3.1 and 3.2, respectively. Finally we include an illustrative empirical application to 26 U.K. sectorial stock returns in section 4, followed by our conclusions. Proofs and auxiliary results can be found in the appendix.

2 Maximum likelihood estimation

2.1 The *Generalised Hyperbolic* distribution

If the $N \times 1$ random vector $\mathbf{u}$ follows a *GH* distribution with parameters $\nu, \delta, \gamma, \quad ,$ and $\boldsymbol{\Upsilon}$, which we write as $\mathbf{u} \sim GH_N(\nu, \delta, \gamma, \quad, \quad, \boldsymbol{\Upsilon})$, then its density will be given by

$$f_{GH}(\mathbf{u}) = \frac{\left(\frac{\gamma}{\delta}\right)^{\nu}}{(2\pi)^{\frac{N}{2}} \left[\quad '\boldsymbol{\Upsilon} \quad + \gamma^2\right]^{\nu - \frac{N}{2}} |\boldsymbol{\Upsilon}|^{\frac{1}{2}} K_{\nu}(\delta\gamma)} \left\{\sqrt{\quad '\boldsymbol{\Upsilon} \quad + \gamma^2}\delta q\left[\delta^{-1}(\mathbf{u} - \quad)\right]\right\}^{\nu - \frac{N}{2}}$$
$$\times K_{\nu - \frac{N}{2}}\left\{\sqrt{\quad '\boldsymbol{\Upsilon} \quad + \gamma^2}\delta q\left[\delta^{-1}(\mathbf{u} - \quad)\right]\right\} \exp\left[\quad '(\mathbf{u} - \quad)\right],$$

where $\nu \in \mathbb{R}$, $\delta, \gamma \in \mathbb{R}^+$, $\quad, \quad \in \mathbb{R}^N$, $\boldsymbol{\Upsilon}$ is a positive definite matrix of order N, $K_\nu(\cdot)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun (1965), as well as appendix C), and $q\left[\delta^{-1}(\mathbf{u} - \quad)\right] = \sqrt{1 + \delta^{-2}(\mathbf{u} - \quad)'\boldsymbol{\Upsilon}^{-1}(\mathbf{u} - \quad)}$.

To gain some intuition on the role that each parameter plays in the *GH* distribution, it is useful to write $\mathbf{u}$ as the following location-scale mixture of normals

$$\mathbf{u} = \quad + \boldsymbol{\Upsilon} \quad \xi^{-1} + \xi^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\mathbf{r}, \tag{1}$$

where $\mathbf{r}$ is a spherical normal random vector, and the positive mixing variable ξ is an independent *GIG* with parameters $-\nu$, γ and δ, or $\xi \sim GIG(-\nu, \gamma, \delta)$ for short. Since $\mathbf{u}$ given ξ is Gaussian with conditional mean $\quad + \boldsymbol{\Upsilon} \quad \xi^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi^{-1}$, it is clear that 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 , which introduces skewness in this distribution.

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

Proposition 1 (Standardisation) *Let* $\quad^* \sim GH_N(\nu, \delta, \gamma, \quad, \quad, \boldsymbol{\Upsilon})$. *If* $\delta = 1$, $\quad = -c(\quad, \nu, \gamma) \quad$, *and*

$$\boldsymbol{\Upsilon} = \frac{\gamma}{R_\nu(\gamma)}\left[\mathbf{I}_N + \frac{c(\quad, \nu, \gamma) - 1}{\quad '} \quad '\right], \tag{2}$$

where $R_\nu(\gamma) = K_{\nu+1}(\gamma)/K_\nu(\gamma)$, $D_{\nu+1}(\gamma) = K_{\nu+2}(\gamma)K_\nu(\gamma)/K^2_{\nu+1}(\gamma)$ *and*

$$c(\quad, \nu, \gamma) = \frac{-1 + \sqrt{1 + 4[D_{\nu+1}(\gamma) - 1] \quad '}}{2[D_{\nu+1}(\gamma) - 1] \quad '}, \tag{3}$$

then $E(\quad^*) = \mathbf{0}$ *and* $V(\quad^*) = \mathbf{I}_N$.

One of the most attractive properties of the *GH* distribution is that it contains as particular cases several of the most important multivariate distributions already used in the literature. The most important ones are:

- **Normal**, which can be achieved in three different ways: (i) when $\nu \to -\infty$ or (ii) $\nu \to +\infty$, regardless of the values of γ and ; and (iii) when $\gamma \to \infty$ irrespective of the values of ν and .
- **Symmetric Student** t, obtained when $-\infty < \nu < -2$, $\gamma = 0$ and $= \mathbf{0}$.
- **Asymmetric Student** t, which is like its symmetric counterpart except that the vector of skewness parameters is no longer zero.
- **Asymmetric Normal-Gamma**, which is obtained when $\gamma = 0$ and $0 < \nu < \infty$.
- **Normal Inverse Gaussian**, for $\nu = -.5$ (see Eriksson, Forsberg, and Ghysels, 2003).

More generally, the distribution of * becomes a simple scale mixture of normals, and thereby spherical, when is zero, with a coefficient of multivariate kurtosis that is monotonically decreasing in both γ and $|\nu|$ (see appendix E). Like any scale mixture of normals, though, the *GH* distribution does not allow for thinner tails than the normal. Nevertheless, financial returns are very often leptokurtic in practice, as section 4 confirms.

Another important feature of the standardised *GH* distribution is that, although the elements of * are uncorrelated, they are not independent except in the multivariate normal case. In general, the *GH* distribution induces "tail dependence", which operates through the positive *GIG* variable in (1). Intuitively, ξ forces the realisations of all the elements in * to be very large in magnitude when it takes very small values, which introduces dependence in the tails of the distribution. In addition, we can make this dependence stronger in certain regions by choosing appropriately. Specifically, we can make the probability of extremely low realisations of both variables much higher than what a Gaussian variate can allow for, as illustrated in Figures 1a-f, which compare the densities of standardised bivariate normal with symmetric and asymmetric Student t distributions. Hence, the *GH* distribution could capture the empirical observation that there is higher tail dependence across stock returns in market downturns.

Finally, we can show that linear combinations of *GH* variables are also *GH*:

Proposition 2 *Let * be distributed as a $N \times 1$ standardised GH random vector with parameters ν, γ and . Then, for any vector $\mathbf{w} \in \mathbb{R}^N$, $s^* = \mathbf{w}'\ ^*/\sqrt{\mathbf{w}'\mathbf{w}}$ is distributed as a standardised GH scalar random variable with parameters ν, γ and*

$$\beta(\mathbf{w}) = \frac{c(\ ,\nu,\gamma)\,(\mathbf{w}'\ )\,\sqrt{\mathbf{w}'\mathbf{w}}}{\mathbf{w}'\mathbf{w} + [c(\ ,\nu,\gamma) - 1]\,(\mathbf{w}'\ )^2/(\ '\ )}.$$

Note that only the skewness parameter, $\beta(\mathbf{w})$, is affected, as it becomes a function of the weights of the linear combination, $\mathbf{w}$.

2.2 The dynamic econometric model

Barndorff-Nielsen and Shephard (2001a) use the (non-standardised) *GH* distribution in the previous section to capture the unconditional distribution of returns on assets whose price dynamics are generated by continuous time stochastic volatility models in which the instantaneous volatility follows an Ornstein-Uhlenbeck process with Lévy innovations. Discrete time models for financial time series, in contrast, are usually characterised by an explicit dynamic regression model with time-varying variances and covariances. Typically, the N dependent variables in $\mathbf{y}_t$ are assumed to be generated as

$$\left.\begin{array}{l} \mathbf{y}_t = \boldsymbol{\mu}_t(\;) + \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\;)\;{}_t^*, \\ \boldsymbol{\mu}_t(\;) = \boldsymbol{\mu}\left(\mathbf{z}_t, I_{t-1};\;\right), \\ \boldsymbol{\Sigma}_t(\;) = \boldsymbol{\Sigma}\left(\mathbf{z}_t, I_{t-1};\;\right), \end{array}\right\} \tag{4}$$

where $\boldsymbol{\mu}()$ and $vech\left[\boldsymbol{\Sigma}()\right]$ are N and $N(N+1)/2$-dimensional vectors of functions known up to the $p \times 1$ vector of true parameter values, ${}_0$, $\mathbf{z}_t$ 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 $\mathbf{z}_t$, $\boldsymbol{\Sigma}_t^{1/2}(\;)$ is some $N \times N$ "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\;)\boldsymbol{\Sigma}_t^{1/2\prime}(\;) = \boldsymbol{\Sigma}_t(\;)$, and ${}_t^*$ is a vector martingale difference sequence satisfying $E({}_t^*|\mathbf{z}_t, I_{t-1};\;{}_0) = \mathbf{0}$ and $V({}_t^*|\mathbf{z}_t, I_{t-1};\;{}_0) = \mathbf{I}_N$. As a consequence, $E(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\;{}_0) = \boldsymbol{\mu}_t({}_0)$ and $V(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\;{}_0) = \boldsymbol{\Sigma}_t({}_0)$.

In this context, FSC assumed that ${}_t^*$ followed a standardised multivariate Student t distribution with ν_0 degrees of freedom conditional on $\mathbf{z}_t$ and I_{t-1}. Instead, we will assume that the conditional distribution of the standardised innovations belongs to the more general *GH* class. Hence, we will be able to assess the adequacy of their assumption by allowing for both skewness and more flexible excess kurtosis in the distribution of ${}_t^*$.

However, we must be particularly careful in making sure that our parametrisation is invariant to the choice of "square root" factorisation of $\boldsymbol{\Sigma}_t(\;)$ because ${}_t^*$ is not generally observable. In particular, we do not want the conditional distribution of $\mathbf{y}_t$ to depend on whether $\boldsymbol{\Sigma}_t^{1/2}(\;)$ is a symmetric or lower triangular matrix, nor in the order of the observed variables in the latter case. Although such a dependence does not arise in univariate *GH* models, or in multivariate *GH* models in which either $\boldsymbol{\Sigma}_t(\;)$ is time-invariant or $\;= \mathbf{0}$, it is a problem that previous efforts to model multivariate skewness have not fully solved (see e.g. Bauwens and Laurent, 2002). In this paper, we circumvent this undesirable feature by making a function of past information and a new vector of parameters $\mathbf{b}$ in the following way:

$$_t(\quad, \mathbf{b}) = \mathbf{\Sigma}_t^{\frac{1}{2}\prime}(\quad)\mathbf{b}. \tag{5}$$

It is then straightforward to see that the resulting *GH* log-likelihood function will not depend on the choice of $\Sigma_t^{\frac{1}{2}}(\quad)$.[1]

Finally, it is analytically convenient to replace ν and γ by η and ψ, where $\eta = -.5\nu^{-1}$ and $\psi = (1+\gamma)^{-1}$.[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.

2.3 The log-likelihood function

Let $= (\;', \eta, \psi, \mathbf{b})'$ denote the parameters of interest. The log-likelihood function of a sample of size T takes the form $L_T(Y_T|\quad) = \sum_{t=1}^{T} l(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\quad)$, where $l(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\quad)$ is the conditional log-density of $\mathbf{y}_t$ given $\mathbf{z}_t$, I_{t-1} and . Given the non-linear nature of the model, a numerical optimisation procedure is usually required to obtain maximum likelihood (ML) estimates of , $\hat{}_T$ say. Assuming that all the elements of $\boldsymbol{\mu}_t(\quad)$ and $\mathbf{\Sigma}_t(\quad)$ are twice continuously differentiable functions of , we can use a standard gradient method in which the first derivatives are numerically approximated by re-evaluating $L_T(\quad)$ with each parameter in turn shifted by a small amount, with an analogous procedure for the second derivatives. Unfortunately, such numerical derivatives are sometimes unstable, and moreover, their values may be rather sensitive to the size of the finite increments used. This is particularly true in our case, because even if the sample size T is large, the *GH* log-likelihood function is often rather flat for values of the parameters that are close to the Gaussian case (see FSC). Fortunately, in this case it is possible to obtain analytical expressions for the score vector (see appendix B), which should considerably improve the accuracy of the resulting estimates (McCullough and

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

[2] We continue to use ν and γ in some equations for notational simplicity. However, we always interpret them as functions of η and ψ, and not as parameters of interest.

Vinod, 1999). Moreover, a fast and numerically reliable procedure for the computation of the score for any value of   is of paramount importance in the implementation of the score-based indirect estimation procedures introduced by Gallant and Tauchen (1996).

2.4 The score vector

We can use EM algorithm - type arguments to obtain analytical formulae for the score function $s_t(\ ) = \partial l\left(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\ \right)/\partial\ $. The idea is based on the identity:

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

where $l\left(\mathbf{y}_t, \xi_t|\mathbf{z}_t, I_{t-1};\ \right)$ is the joint log-density function of $\mathbf{y}_t$ and ξ_t (given $\mathbf{z}_t$, I_{t-1} and); $l\left(\mathbf{y}_t|\xi_t, \mathbf{z}_t, I_{t-1};\ \right)$ is the conditional log-likelihood of $\mathbf{y}_t$ given ξ_t ($\mathbf{z}_t$, I_{t-1} and); $l\left(\xi_t|\mathbf{y}_t, \mathbf{z}_t, I_{t-1};\ \right)$ is the conditional log-likelihood of ξ_t given $\mathbf{y}_t$ ($\mathbf{z}_t$, I_{t-1} and); and finally $l\left(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\ \right)$ and $l\left(\xi_t|\mathbf{z}_t, I_{t-1};\ \right)$ are the marginal log-densities of $\mathbf{y}_t$ and ξ_t (given $\mathbf{z}_t$, I_{t-1} and), respectively. If we differentiate both sides of the previous identity with respect to , and take expectations, then we will end up with:

$$s_t(\ ) = E\left(\left.\frac{\partial l\left(\mathbf{y}_t|\xi_t, \mathbf{z}_t, I_{t-1};\ \right)}{\partial\ }\right| Y_T;\ \right) + E\left(\left.\frac{\partial l\left(\xi_t|\mathbf{z}_t, I_{t-1};\ \right)}{\partial\ }\right| Y_T;\ \right) \tag{6}$$

because $E\left[\partial l\left(\xi_t|\mathbf{y}_t, \mathbf{z}_{t-1}, I_{t-1};\ \right)/\partial\ |Y_T;\ \right] = \mathbf{0}$ by virtue of the Kullback inequality. In this way, we decompose $s_t(\ )$ as the sum of the expected values of (i) the score of a multivariate Gaussian log-likelihood function, and (ii) the score of a univariate *GIG* distribution, both of which are easy to obtain (see appendix B for details).

For the purposes of developing our testing procedures in section 3, it is convenient to obtain closed-form expressions for $s_t(\ )$ under the two important special cases of multivariate Gaussian and Student t innovations.

2.4.1 The score under Gaussianity

As we saw before, we can achieve normality in three different ways: (i) when $\eta \to 0^+$ or (ii) $\eta \to 0^-$ regardless of the values of $\mathbf{b}$ and ψ; and (iii) when $\psi \to 0$, irrespective of η and $\mathbf{b}$. Therefore, it is not surprising that the Gaussian scores with respect to η or ψ are 0 when these parameters are not identified, and also, that $\lim\limits_{\eta\cdot\psi\to 0} s_{\mathbf{b}t}(\ ) = \mathbf{0}$. Similarly, the limit of the score with respect to the mean and variance parameters, $\lim_{\eta\cdot\psi\to 0} s\ _t(\ )$, coincides with the usual Gaussian expressions (see e.g. Bollerslev and

Wooldridge (1992)). Further, we can show that for fixed $\psi > 0$,

$$\begin{aligned}\lim_{\eta\to 0^+} s_{\eta t}(\ ) &= -\lim_{\eta\to 0^-} s_{\eta t}(\ ) = \left[\frac{1}{4}\varsigma_t^2(\ ) - \frac{N+2}{2}\varsigma_t(\ ) + \frac{N(N+2)}{4}\right] \\ &\quad + \mathbf{b}'\left\{\ _t(\ )\left[\varsigma_t(\ ) - (N+2)\right]\right\}, \end{aligned} \tag{7}$$

where $\ _t(\ ) = \mathbf{y}_t - \boldsymbol{\mu}_t(\ )$, $\ ^*_t(\ ) = \boldsymbol{\Sigma}_t^{-\frac{1}{2}}\ _t(\ )$ and $\varsigma_t(\ ) = \ ^{*\prime}_t(\ )\ ^*_t(\ )$, which confirms the non-differentiability of the log-likelihood function with respect to η at $\eta = 0$. Finally, we can also show that for fixed $\eta \neq 0$, $\lim_{\psi\to 0} s_{\psi t}(\ )$ is exactly one half of (7).

2.4.2 The score under Student t innovations

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

$$s_{\mathbf{b}t}(\ , 1, 0) = \frac{\eta\left[\varsigma_t(\ ) - (N+2)\right]}{1 - 2\eta + \eta\varsigma_t(\ )}\ _t(\ ), \tag{8}$$

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 3.2.

2.5 The information matrix

Given correct specification, the results in Crowder (1976) imply that the score vector $s_t(\ )$ evaluated at the true parameter values $\ _0$ has the martingale difference property. In addition, his results also imply that under additional regularity conditions (which in particular require that $\ _0$ is locally identified and belongs to the interior of the parameter space), the ML estimator will be asymptotically normally distributed with a covariance matrix which is the inverse of the usual information matrix

$$\mathcal{I}(\ _0) = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} s_t(\ _0)s_t'(\ _0) = E[s_t(\ _0)s_t'(\ _0)]. \tag{9}$$

The simplest consistent estimator of $\mathcal{I}(\ _0)$ is the sample outer product of the score:

$$\hat{\mathcal{I}}_T(\hat{\ }_T) = \frac{1}{T}\sum_{t=1}^{T} s_t(\hat{\ }_T)s_t'(\hat{\ }_T).$$

However, the resulting standard errors and tests statistics can be badly behaved in finite samples, especially in dynamic models (see e.g. Davidson and MacKinnon, 1993). We can evaluate much more accurately the integral implicit in (9) in pure time series models by generating a long simulated path of size T_s of the postulated process $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \cdots, \hat{\mathbf{y}}_{T_s}$, where the symbol ˆ indicates that the data has been generated using the *GH* maximum likelihood estimates $\hat{\ }_T$. Then, if we denote by $s_{t_s}(\hat{\ }_T)$ the value of the score function for each simulated observation, our proposed estimator of the information matrix is

$$\tilde{\mathcal{I}}_{T_s}(\hat{\ }_T) = \frac{1}{T_s}\sum_{t_s=1}^{T_s} s_{t_s}(\hat{\ }_T)s'_{t_s}(\hat{\ }_T),$$

where we can get arbitrarily close in a numerical sense to the value of the asymptotic information matrix evaluated at $\hat{\ }_T$, $\mathcal{I}(\hat{\ }_T)$, as we increase T_s. Our experience suggests that $T_s = 100,000$ yields reliable results. In this respect, the simplest way to simulate a *GH* variable is to exploit its mixture-of-normals interpretation in (1) after sampling from a multivariate normal and a scalar *GIG* distribution (see Dagpunar, 1989).

In some important special cases, though, it is also possible to estimate $\mathcal{I}(\ _0)$ as the sample average of the conditional information matrix $\mathcal{I}_t(\ ) = Var\left[s_t(\ )\right|\mathbf{z}_t, I_{t-1};\ ]$. In particular, analytical expressions for $\mathcal{I}_t(\ )$ can be obtained in the case of Gaussian and Student t innovations.

2.5.1 The conditional information matrix under Gaussianity

In principle, we must study separately the three possible ways to achieve normality. First, consider the conditional information matrix when $\eta \to 0^+$,

$$\begin{bmatrix} \mathcal{I}\ \ _t(\ ,0^+,\psi,\mathbf{b}) & \mathcal{I}\ \ _t(\ ,0^+,\psi,\mathbf{b}) \\ \mathcal{I}'\ \ _t(\ ,0^+,\psi,\mathbf{b}) & \mathcal{I}\ \ _t(\ ,0^+,\psi,\mathbf{b}) \end{bmatrix} = \lim_{\eta\to 0^+} V\left[\begin{array}{c} s\ _t(\ ,\eta,\psi,\mathbf{b}) \\ s_{\eta t}(\ ,\eta,\psi,\mathbf{b}) \end{array}\middle|\mathbf{z}_t, I_{t-1};\ \right], \quad (10)$$

where we have not considered either $s_{\mathbf{b}t}(\ )$ or $s_{\psi t}(\ )$ because they 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 coincides with the expression obtained by Bollerslev and Wooldridge (1992). Moreover, we can show that

Proposition 3 $\mathcal{I}\ _{\eta t}(\ ,0^+,\psi,\mathbf{b}) = \mathbf{0}$ and $\mathcal{I}_{\eta\eta t}(\ ,0^+,\psi,\mathbf{b}) = (N+2)\left[.5N + \mathbf{b}'\mathbf{\Sigma}_t(\ )\mathbf{b}\right]$.

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 $\lim_{\eta\to 0^-} s_{\eta t}(\ ,\eta,\psi,\mathbf{b}) = -\lim_{\eta\to 0^+} s_{\eta t}(\ ,\eta,\psi,\mathbf{b})$, as we saw before.

Finally, when $\psi \to 0$, we must exclude $s_{\mathbf{b}t}(\ )$ and $s_{\eta t}(\ )$ 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 $\mathcal{I}\ _{\psi t}(\ ,\eta,0,\mathbf{b}) = \mathbf{0}$, and that $\mathcal{I}_{\psi\psi t}(\ ,\eta,0,\mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\ ,0^+,\psi,\mathbf{b}) = \frac{1}{4}\mathcal{I}_{\eta\eta t}(\ ,0^-,\psi,\mathbf{b})$.

2.5.2 The conditional information matrix under Student t innovations

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

Proposition 4 $\mathcal{I}_{\eta\mathbf{b}t}(\ ,\eta > 0,1,\mathbf{0}) = \mathbf{0}$,

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

$$\mathcal{I}_{\mathbf{bb}t}(\ ,\eta > 0,1,\mathbf{0}) = \frac{2\,(N+2)\,\eta^2}{(1-2\eta)\,(1+(N+2)\,\eta)}\boldsymbol{\Sigma}_t(\ ).$$

3 Testing the distributional assumptions

3.1 Multivariate normality versus *GH* innovations

The derivation of a Lagrange multiplier (LM) test for multivariate normality versus *GH*-distributed 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 unidentified 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 unidentified when $\eta \to 0^{\pm}$.

There are two standard solutions in the literature to deal with testing situations with unidentified parameters under the null. One approach involves fixing the unidentified parameters to some arbitrary values, and then computing the appropriate test statistic for those given values. This approach is plausible in situations where there are values for the unidentified parameters which make sense from an economic or statistical point of view. Unfortunately, it is not at all clear a priori what values for $\mathbf{b}$ and ψ or η are likely to prevail under the alternative of *GH* innovations. For that reason, we follow here the second approach, which consists in computing the LM test statistic for the whole range of values of the unidentified parameters, which are then combined to construct an overall test statistic (see Andrews, 1994 for a formal justification). In our case, we compute LM 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. As we will show in the next subsections, we can obtain closed-form analytical expressions for the supremum of the LM test statistics, as well as for its asymptotic distribution, in contrast to what happens in the general case (see again Andrews, 1994).

3.1.1 LM test for fixed values of the unidentified parameters

Let $\tilde{\ }_T$ denote the ML estimator of obtained by maximising a Gaussian log-likelihood function. For the case in which normality is achieved as $\eta \to 0^+$, we can use the results in sections 2.4.1 and 2.5.1 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 and η, $\bar{s}\ _T\left(\tilde{\ }_T, 0^+, \psi, \mathbf{b}\right)$ and $\bar{s}_{\eta T}\left(\tilde{\ }_T, 0^+, \psi, \mathbf{b}\right)$, with the inverse of the unconditional information matrix as weighting matrix, which can be obtained as the unconditional expected value of the conditional information matrix (10). But since $\bar{s}\ _T\left(\tilde{\ }_T, 0^+, \psi, \mathbf{b}\right) = \mathbf{0}$ by definition of $\tilde{\ }_T$, and $\mathcal{I}\ _{\eta t}(\ _0, 0^+, \psi, \mathbf{b}) = 0$, we can show that

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

We can operate analogously for the other two limits, thereby obtaining the test statistic $LM_2\left(\tilde{\ }_T, \psi, \mathbf{b}\right)$ for the null $\eta \to 0^-$, and $LM_3\left(\tilde{\ }_T, \eta, \mathbf{b}\right)$ for $\psi \to 1$. Somewhat remarkably, all these test statistics share the same formula, which only depends on $\mathbf{b}$:

Proposition 5 (LM normality test)

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

where $\hat{\mathbf{\Sigma}}$ *is some consistent estimator of* $\mathbf{\Sigma}(\ _0) = E\left[\mathbf{\Sigma}_t(\ _0)\right]$.

Under standard regularity conditions, $LM\left(\tilde{\ }_T, \mathbf{b}\right)$ 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.

3.1.2 The supremum LM test

By maximising $LM\left(\tilde{\ }_T, \mathbf{b}\right)$ with respect to $\mathbf{b}$, we obtain the following result:

Proposition 6 (Supremum test)

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

$$LM_k(\tilde{\ }_T) = \frac{2}{N(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \left[\frac{1}{4}\varsigma_t^2(\tilde{\ }_T) - \frac{N+2}{2}\varsigma_t(\tilde{\ }_T) + \frac{N(N+2)}{4} \right] \right\}^2, \quad (11)$$

$$LM_s(\tilde{\ }_T) = \frac{1}{2(N+2)} \left\{ \frac{\sqrt{T}}{T} \sum_t \ _t(\tilde{\ }_T) \left[\varsigma_t(\tilde{\ }_T) - (N+2) \right] \right\}' \hat{\boldsymbol{\Sigma}}^{-1}$$
$$\times \left\{ \frac{\sqrt{T}}{T} \sum_t \ _t(\tilde{\ }_T) \left[\varsigma_t(\tilde{\ }_T) - (N+2) \right] \right\}, \quad (12)$$

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 test, i.e. $LM_k(\tilde{\ }_T)$, is numerically identical to the LM statistic derived by FSC to test multivariate normal versus Student t innovations. These authors reinterpret (11) as a specification test of the restriction on the first two moments of $\varsigma_t(\ _0)$ implicit in

$$E\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\ _0) + \frac{1}{4}\varsigma_t^2(\ _0) \right] = E[m_{kt}(\ _0)] = 0, \quad (13)$$

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 $LM_k(\tilde{\ }_T)$ as the kurtosis component of our multivariate normality test.

In contrast, the second component of our test, $LM_s(\tilde{\ }_T)$, 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 $\ _t(\tilde{\ }_T)$ on $\varsigma_t(\tilde{\ }_T) - (N+2)$ (Hessian version), or a univariate regression of 1 on $\left[\varsigma_t(\tilde{\ }_T) - (N+2)\right] \ _t(\tilde{\ }_T)$ (Outer product version). Nevertheless, we would expect a priori that $LM_s(\tilde{\ }_T)$ would be the version of the LM test with the smallest size distortions (see Davidson and MacKinnon, 1983).

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:

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

But since $\varsigma_t(\ _0) = \ _t^{*\prime}(\ _0) \ _t^*(\ _0)$, it is clear that (12) is also testing for co-skewness. Specifically, $LM_s(\tilde{\ }_T)$ is testing the N alternative moment conditions

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

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 simply assesses whether (14) holds for $i = j = k = 1, \cdots, N$, we shall 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 (2004) and Fiorentini, Sentana, and Calzolari (2004). An important advantage of our proposed normality test is that its asymptotic size is always correct because both (13) and (15) are orthogonal by construction to the Gaussian score corresponding to   evaluated at $_0$.

By analogy with Bontemps and Meddahi (2004), one possible way to adjust Mardia's (1970) formulae is to replace $\varepsilon_{it}^{*3}(\ )$ by $H_3[\varepsilon_{it}^{*}(\ )]$ and $\varepsilon_{it}^{*2}(\ )\varepsilon_{jt}^{*}(\ )$ by $H_2[\varepsilon_{it}^{*}(\ )]H_2[\varepsilon_{it}^{*}(\ )]$ $(i \neq j)$ in the moment conditions (14), where $H_k(\cdot)$ is the Hermite polynomial of order k. Unfortunately, this will make his test numerically dependent on the chosen orthogonalisation of $_t(\ )$. In this respect, note that both $LM_k(\tilde{\ }_T)$ and $LM_s(\tilde{\ }_T)$ are 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 $\varepsilon_{it}^{*}(\tilde{\ }_T)$.

3.1.3 A one-sided, Kuhn-Tucker multiplier version of the supremum test

As we discussed in section 2.1, the class of GH distributions can only accommodate fatter tails than the normal. In terms of the kurtosis component of our multivariate normality test, this implies that as we depart from normality, we will have

$$E\left[m_{kt}(\ _0)|\ _0, \eta_0 > 0, \psi_0 > 0\right] > 0. \tag{16}$$

In view of the one-sided nature of the kurtosis component, we will follow FSC and suggest a Kuhn-Tucker (KT) multiplier version of the supremum test that exploits (16) in order to increase its power (see also Andrews, 2001). Specifically, we recommend the use of

$$KT(\tilde{\ }_T) = LM_k(\tilde{\ }_T)\mathbf{1}\left(\bar{m}_{kT}(\tilde{\ }_T) > 0\right) + LM_s(\tilde{\ }_T),$$

where $\mathbf{1}(\cdot)$ is the indicator function, and $\bar{m}_{kT}(\ )$ the sample mean of $m_{kt}(\ _0)$. Asymptotically, the probability that $\bar{m}_{kT}(\tilde{\ }_T)$ becomes negative is .5 under the null. Hence, $KT(\tilde{\ }_T)$ 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. To obtain p-values for this test, we can use the expression $\Pr(X > c) = 1 - .5F_{\chi^2_N}(c) - .5F_{\chi^2_{N+1}}(c)$ (see e.g. Demos and Sentana, 1998)

3.1.4 Power of the normality test

It is interesting to study the power properties of the multivariate normality tests derived in the previous sections. However, given that the block-diagonality of the information matrix between and the other parameters is generally lost under the alternative of GH innovations, and its exact form is unknown, we can only get closed form expressions for the case in which the standardised innovations $^*_t$ are directly observed. Thus, for the purposes of this exercise we will only consider models in which $\boldsymbol{\mu}_t(\ ) = \mathbf{0}$, $\boldsymbol{\Sigma}_t(\ ) = \mathbf{I}_N$, and $^*_t$ is a standardised GH variable with parameters η, ψ and $\mathbf{b}$. In more realistic cases, though, the results are likely to be qualitatively similar.

In addition, we only consider alternatives in which $\mathbf{b}$ is proportional to a vector of ones. Although this may seem a restrictive assumption, we can show that the power of the test only depends on $\mathbf{b}$ through its Euclidean norm when the residuals are observed.

The results at the usual 5% significance level are displayed in Figures 2a to 2f for $\psi = 1$ and $T = 100$ (see appendix F for details). In Figures 2a to 2d, we have represented η on the x-axis. We can see in Figure 2a that for $N = 1$ and $|\mathbf{b}| = 0$, the test with the highest power is the one-sided kurtosis test, followed by the KT test. On the other hand, if we consider asymmetric alternatives, the skewness component of the normality test becomes important, and eventually makes the KT test more powerful than the kurtosis test (see e.g. Figure 2b for $|\mathbf{b}| = 1$). Note also that the KT test displays higher power than the supremum test, which is its two-sided counterpart, under alternatives very close to the null. Not surprisingly, we can also see from Figures 2c and 2d that power increases with the dimension N.[3] Those figures also confirm our previous conclusions on the relative power of the kurtosis and supremum tests continue to hold for $N > 1$.

In contrast, we have represented the norm of $\mathbf{b}$ on the x-axis in Figures 2e and 2f.

[3] We do not report the power of the one-sided tests for N greater than 1 due to the numerical unreliability in higher dimensions of the quadrature integration methods described in appendix F.

There we can clearly see the effects on power of the fact that **b**

Let $\bar{\ }_T = (\bar{\ }'_T, \bar{\eta}_T)'$ 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

$$\tau_{kT}\left(\bar{\ }_T\right) = \frac{\sqrt{T}\bar{s}_{\psi\psi T}\left(\bar{\ }_T, 1, \mathbf{0}\right)}{\hat{V}\left[s_{\psi\psi t}\left(\bar{\ }_T, 1, \mathbf{0}\right)\right]}, \tag{17}$$

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

Proposition 8 (*Student t* symmetric test) *If $\ ^*_t$ is conditionally distributed as a standardised Student t with η_0^{-1} degrees of freedom, then*

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

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

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

and

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

where

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

But since $E\left[s_{\psi\psi t}\left(\ _0, \eta_0, 1, \mathbf{0}\right) \mid\ _0, \psi_0 < 1, \mathbf{b}_0 = \mathbf{0}\right] > 0$ (see Proposition 7), and 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

$$\tau_{kT}\left(\bar{\ }_T\right)\mathbf{1}\left[\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0}) > 0\right].$$

Finally, it is also important to mention that although $s_{\psi t}\left(\ _0, \psi, \mathbf{b}\right) = 0\ \forall t$, we can combine Propositions 7 and 8 to 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

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

but

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

3.2.2 Student t vs asymmetric *GH innovations*

By construction, the 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 $E[s_{\mathbf{b}t}(\ ,1,\mathbf{0})]=\mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}(\ ,1,\mathbf{0})$ is defined in (8). The asymptotic joint distribution of the two moment conditions that takes into account the sampling variability in $\bar{\ }_T$ is given in the following result

Proposition 9 (*Student t* asymmetric test) *If $\ ^*_t$ is conditionally distributed as a standardised Student t with η_0^{-1} degrees of freedom, then*

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

where

$$\mathcal{V}(\ _0)=\begin{bmatrix}\mathcal{V}_{\mathbf{bb}}(\ _0) & \mathcal{V}_{\mathbf{b}\psi}(\ _0)\\ \mathcal{V}'_{\mathbf{b}\psi}(\ _0) & \mathcal{V}_{\psi\psi}(\ _0)\end{bmatrix}=\left\{\begin{matrix}\mathcal{I}_{\mathbf{bb}}(\ _0,1,\mathbf{0}) & \mathbf{0}\\ \mathbf{0}' & V[s_{\psi\psi t}(\ _0,1,\mathbf{0})]\end{matrix}\right\}$$
$$-\begin{bmatrix}\mathcal{I}'_{\ \mathbf{b}}(\ _0,1,\mathbf{0})\mathcal{I}^{-1}(\ _0,1,\mathbf{0})\mathcal{I}_{\ \mathbf{b}}(\ _0,1,\mathbf{0}) & \mathcal{I}'_{\ \mathbf{b}}(\ _0,1,\mathbf{0})\mathcal{I}^{-1}(\ _0,1,\mathbf{0})\mathcal{M}(\ _0)\\ \mathcal{M}'(\ _0)\mathcal{I}^{-1}(\ _0,1,\mathbf{0})\mathcal{I}_{\ \mathbf{b}}(\ _0,1,\mathbf{0}) & \mathcal{M}'(\ _0)\mathcal{I}^{-1}(\ _0,1,\mathbf{0})\mathcal{M}(\ _0)\end{bmatrix},\quad (18)$$

$\mathcal{I}\ \ (\ _0,1,\mathbf{0})=E[\mathcal{I}\ \ _t(\ _0,1,\mathbf{0})]$ is the Student t information matrix derived in FSC, $\mathcal{I}_{\ \mathbf{b}}(\ _0,1,\mathbf{0})=E[\mathcal{I}_{\ \mathbf{b}t}(\ _0,1,\mathbf{0})]$ and $\mathcal{I}_{\mathbf{bb}}(\ _0,1,\mathbf{0})=E[\mathcal{I}_{\mathbf{bb}t}(\ _0,1,\mathbf{0})]$ are defined in Proposition 4, and $\mathcal{M}(\ _0)$ and $V[s_{\psi\psi t}(\ _0,1,\mathbf{0})]$ are given in Proposition 8.

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

$$\tau_{gT}(\bar{\ }_T)=\begin{bmatrix}\sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\ }_T,1,\mathbf{0})\\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\ }_T,1,\mathbf{0})\end{bmatrix}'\mathcal{V}^{-1}(\bar{\ }_T)\begin{bmatrix}\sqrt{T}\bar{\mathbf{s}}_{\mathbf{b}T}(\bar{\ }_T,1,\mathbf{0})\\ \sqrt{T}\bar{s}_{\psi\psi T}(\bar{\ }_T,1,\mathbf{0})\end{bmatrix},\quad (19)$$

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 described in Proposition 7. However, since $\mathcal{V}(\ _0)$ is not block diagonal in general, we must orthogonalise the moment conditions to obtain a partially one-sided moment test which is asymptotically equivalent to the likelihood ratio test (see e.g. Silvapulle and Silvapulle, 1995). Specifically, instead of using directly the score with respect to $\mathbf{b}$, we consider

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

whose sample average is asymptotically orthogonal to $\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0})$ by construction. Note, however, that there is no need to do this orthogonalisation when $E[\partial\boldsymbol{\mu}_t(\ _0)/\partial\ _0] = \mathbf{0}$, since in this case $\mathcal{V}_{\mathbf{b}\psi}(\ _0) = \mathbf{0}$ because $I\ _{\mathbf{b}}(\ _0, 1, 0) = \mathbf{0}$ (see Proposition 4).

It is then straightforward to see that the asymptotic distribution of

$$\tau_{oT}(\bar{\ }_T) = T\bar{s}^{\perp\prime}_{\mathbf{b}t}(\bar{\ }_T, 1, \mathbf{0})\left[\mathcal{V}_{\mathbf{bb}}(\bar{\ }_T) - \frac{\mathcal{V}_{\mathbf{b}\psi}(\bar{\ }_T)\mathcal{V}'_{\mathbf{b}\psi}(\bar{\ }_T)}{\mathcal{V}_{\psi\psi}(\bar{\ }_T)}\right]^{-1}\bar{s}^{\perp}_{\mathbf{b}t}(\bar{\ }_T, 1, \mathbf{0})$$
$$+\tau_{kT}(\bar{\ }_T)\,\mathbf{1}\left[\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0}) > 0\right], \tag{20}$$

is another 50:50 mixture of chi-squares with N and $N+1$ degrees of freedom under the null, because asymptotically, the probability that $\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0})$ is negative will be .5 if $\psi_0 = 1$. Such a one-sided test benefits from the fact that a non-positive $\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0})$ gives no evidence against the null, in which case we only need to consider the orthogonalised skewness component. In contrast, when $\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0})$ is positive, (20) numerically coincides with (19).

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

$$\tau_{aT}(\bar{\ }_T) = \sqrt{T}\bar{s}'_{\mathbf{b}T}(\bar{\ }_T, 1, \mathbf{0})\,\mathcal{V}^{-1}_{\mathbf{bb}}(\bar{\ }_T)\sqrt{T}\bar{s}_{\mathbf{b}T}(\bar{\ }_T, 1, \mathbf{0}), \tag{21}$$

which can be interpreted as a regular LM test of the Student t distribution versus the asymmetric t distribution under the maintained assumption that $\psi = 1$ (see Mencía, 2003). As a result, $\tau_{aT}(\bar{\ }_T)$ will be asymptotically distributed as a chi-square distribution with N degrees of freedom under the null of Student 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. One possibility would be to scale $LM_s(\tilde{\ }_T)$ 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 $m_{st}(\bar{\ }_T)$, or the multivariate regression of $\ _t(\bar{\ }_T)$ on $\varsigma_t(\bar{\ }_T) - (N+2)$, 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 generally less powerful against local alternatives of the form $\mathbf{b}_{0T} = \mathbf{b}_0/\sqrt{T}$ than $\tau_{aT}(\bar{\ }_T)$ in (21), which is the proper LM test for symmetry.

Nevertheless, an interesting property of a moment test for symmetry based on (15) is that $\sqrt{T}\bar{m}_{sT}(\bar{\ }_T)$ and $\sqrt{T}\bar{s}_{\psi\psi T}(\bar{\ }_T, 1, \mathbf{0})$ are asymptotically independent under the null of symmetric Student t innovations, which means that there is no need to orthogonalise them in order to obtain a one-sided version that combines the two of them.

4 Empirical illustration

We now apply the methodology derived in the previous sections to the empirical application reported by FSC to assess the validity of the multivariate Student t assumption they made. Their data consists on monthly excess returns on 26 U.K. sectorial indices for the period 1971:2-1990:10 (237 observations). The vector of conditional mean returns $\boldsymbol{\mu}_t(\ )$ was assumed to be 0 because the returns had been demeaned prior to estimation. As for the conditional covariance matrix, they considered

$$\boldsymbol{\Sigma}_t(\ ) = \boldsymbol{\Gamma} + \lambda_t \mathbf{cc}',$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix, $\mathbf{c}$ a vector of dimension $N = 26$, and

$$\lambda_t = 1 - \alpha_1 - \alpha_2 - \alpha_1 v^2 + \alpha_1 \left[\left(f_{t-1|t-1} - v \right)^2 + \omega_{t-1|t-1} \right] + \alpha_2 \lambda_{t-1},$$

with $\omega_{t|t} = \left[\lambda_t^{-1} + \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c} \right]^{-1}$, $f_{t|t} = \omega_{t|t}\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{y}_t$. Such a covariance matrix structure corresponds to a conditionally heteroskedastic single factor model in which the conditional covariance of the latent factor follows a univariate GQARCH(1,1)-type process (see FSC for details). FSC estimated the model subject to the constraints $v = \rho\sqrt{(1 - \alpha_1 - \alpha_2)/\alpha_1}$, $-1 \leq \rho \leq 1$, and $0 \leq \alpha_2 \leq 1 - \alpha_1 \leq 1$, which ensure that $\lambda_t \geq 0$ for all t, and solve the scale indeterminacy of the vector $\mathbf{c}$ by implicitly setting $E[\lambda_t] = 1$.

We have re-estimated this model under three different conditional distributional assumptions on the standardised innovations $\ _t^*$: Gaussian, Student t and *GH*. We first estimated the model by Gaussian PML. The estimates for α_1, α_2, and ρ are reported in Table 1, together with robust standard errors, which we have obtained by using the formulae in Bollerslev and Wooldridge (1992) with analytical expressions for the derivatives. Then, on the basis of these PML estimators, we have computed the Kuhn-Tucker normality test $KT(\tilde{\ }_T)$ described in section 3.1.3, which is reported in the first panel of Table 2. Notice that we can easily reject normality because both the skewness and kurtosis components of the test lead to this conclusion.

Next, we estimated the same Student t model as FSC using the analytical formulae for the score and the conditional information matrix they provide. The results, also reported in Table 1, show that the estimate for the tail thickness parameter η, which corresponds to slightly less than 10 degrees of freedom, is significantly larger than 0. This result can be confirmed by comparing the log-likelihood functions under Gaussian and Student t innovations, which implies that we would also reject normality with a likelihood ratio test. Then, on the basis of these estimates, we have computed the Student t test statistics $\tau_{kT}(\bar{}_T)$ and $\tau_{aT}(\bar{}_T)$ presented in section 3.2 (see also Table 2). The results show that we can easily reject the Student t assumption because of the high value we obtain for the skewness component $\tau_{aT}(\bar{}_T)$. However, the one-sided version of the ψ component of the test is completely unable to reject the Student t specification against the alternative hypothesis of symmetric *GH* innovations because $\bar{s}_{\psi\psi T}(\bar{}_T, 1, \mathbf{0}) < 0$. This, together with the fact that the conditional mean is assumed to be 0, implies that the KT version of the Student t test in (20) numerically coincides with $\tau_{aT}(\bar{}_T)$.

Finally, we re-estimated the model under the assumption that the conditional distribution of the innovations is *GH* by using the analytical formulae for the score provided in appendix B, which introduces as additional parameters ψ and the vector $\mathbf{b}$. The results for α_1, α_2, ρ, η, also reported in Table 1, are very similar to those of the Student t model. However, since the ML estimate of ψ is 1, the estimated conditional distribution is effectively an asymmetric t. Again, a likelihood ratio test would also reject the Student t specification, although the gains in fit obtained by allowing for asymmetry (as measured by the increments in the log-likelihood function) are not as important as those obtained by generalising the normal distribution in the leptokurtic direction.

5 Conclusions

In this paper we develop a rather flexible parametric framework that allows us to account for the presence of skewness and kurtosis in multivariate dynamic heteroskedastic regression models. In particular, we assume that the standardised innovations of the model have a conditional Generalised Hyperbolic (*GH*) distribution, which nests as particular cases the multivariate Gaussian and Student t distributions, as well as other potentially asymmetric alternatives. To do so, we first standardise the usual *GH* distribution by imposing restrictions on its parameters. Importantly, we make sure that our model is invariant to the orthogonalisation used to compute the square root of the conditional covariance matrix. Then, we give analytical formulae for the log-likelihood score, which simplify its computation, and at the same time make it more reliable. In addition, we explain how to evaluate the unconditional information matrix.

On the basis of these first and second derivatives, we obtain multivariate normality and Student t tests against alternatives with GH innovations. In this respect, we show how to overcome the identification problems that the use of the GH distribution entails. Moreover, we decompose both our proposed test statistics into skewness and kurtosis components, which we exploit to derive more powerful one-sided versions. We also evaluate in detail the power of several versions of the normality tests against GH alternatives, and conclude that the inclusion of the skewness component of our test yields substantial power gains unless we are very close to the null hypothesis.

Finally, we revisit the empirical application to UK sectorial stock returns presented in FSC. Testing the distributional assumption is particularly important in the Student t case because ML estimators based on incorrectly specified non-Gaussian distributions often lead to inconsistent parameter estimates (see Newey and Steigerwald, 1997). In this respect, we find clear evidence of conditional skewness in the FSC dataset.

Because the existing simulation evidence indicates that the finite-sample size properties of many LM tests could be different from the nominal one, a fruitful avenue for future research would be to consider bootstrap procedures to reduce size distortions (see e.g. Kilian and Demiroglu, 2000). In addition, it would be interesting to develop sequential estimators of the asymmetry and kurtosis parameters introduced by the GH assumption, which would keep constant the conditional mean and variance parameters at their Gaussian PML estimators along the lines of Fiorentini, Sentana, and Calzolari (2003b). At the same time, it would also be useful to assess the biases of the Student t-based ML estimators of the conditional mean and variance parameters when the true conditional distribution of the innovations is in fact a different member of the GH family. Finally, although in order to derive our distributional specification tests we have maintained the implicit assumption that the first and second moments adequately capture all the model dynamics, it would also be worth extending Hansen's (1994) approach to a multivariate context, and explore time series specifications for the parameters characterising the higher order moments of the GH distribution.

References

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

Barndorff-Nielsen, O. and N. Shephard (2001a). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 167–241.

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

Bauwens, L. and S. Laurent (2002). A new class of multivariate skew densities, with application to GARCH models. *Journal of Business and Economic Statistics*, forthcoming.

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.*

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 (2004). Testing normality: a GMM approach. *Journal of Econometrics*, forthcoming.

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

Dagpunar, J. (1989). An easily implemented, generalized inverse gaussian generator. *Communications in Statistics - Simulation 18*, 703–710.

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 (1993). *Estimation and inference in econometrics.* Oxford, U.K.: Oxford University Press.

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.

Eriksson, A., L. Forsberg, and E. Ghysels (2003). Approximating the probability distribution of functions of random variables: A new approach. mimeo University of North Carolina.

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

Fiorentini, G., E. Sentana, and G. Calzolari (2003a). 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 (2003b). The relative efficiency of pseudo-maximum likelihood estimation and inference in conditionally heteroskedastic dynamic regression models. mimeo CEMFI.

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.

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

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review 35*, 705–730.

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.

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.

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

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

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

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

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

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

Prause, K. (1998). The generalised hyperbolic models: estimation, financial derivatives and risk measurement. Unpublished Ph.D. thesis, Mathematics Faculty, Freiburg University.

RiskMetrics Group (1996). Riskmetrics Technichal document.

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

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

Appendix

A Proofs of Propositions

Proposition 1

If we impose the parameter restrictions of Proposition 1 in equation (1), we get

$$^{*} = c(\ ,\nu,\gamma)\quad\left[\frac{\gamma\xi^{-1}}{R_\nu(\gamma)} - 1\right] + \sqrt{\;}$$

Proposition 3

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 expansion (C3), which is valid in this case. We operate in a similar way when $\psi \to 0$, but in this case the appropriate expansion is (C2). 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 $^*_t$ is distributed as a multivariate standard normal, then it can be written as $^*_t = \sqrt{\zeta_t}\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_t is a chi-square random variable with N degrees of freedom, and $\mathbf{u}_t$ and ζ_t are mutually independent. □

Proposition 4

The proof is straightforward if we rely on the results in the appendix of Fiorentini, Sentana, and Calzolari (2003b), who indicate that when $^*_t$ is distributed as a standardised multivariate Student t with $1/\eta_0$ degrees of freedom, it can be written as $^*_t = \sqrt{(1-2\eta_0)\zeta_t/(\xi_t\eta_0)}\mathbf{u}_t$, 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 $X = \zeta_t/\left(\zeta_t + \xi_t\right)$ has a beta distribution with parameters $a = N/2$ and $b = 1/\left(2\eta_0\right)$ to show that

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

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

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

Proposition 5

For fixed $\mathbf{b}$ and ψ, the LM_1 test is based on the average scores with respect to η and evaluated at 0^+ and the Gaussian maximum likelihood estimates $\tilde{\ }_T$. But since the average score with respect to 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 $\tilde{\ }_T$ with the score corresponding to ψ evaluated at 0 and $\tilde{\ }_T$ leads to the desired result. □

Proposition 6

$LM\left(\tilde{}_T, \mathbf{b}\right)$ can be trivially expressed as

$$LM\left(\tilde{}_T, \mathbf{b}\right) = \frac{T\mathbf{b}^{+\prime}\bar{m}_T(\tilde{}_T)\bar{m}_T(\tilde{}_T)\mathbf{b}^+}{(N+2)\,\mathbf{b}^{+\prime}\mathbf{D}_T\mathbf{b}^+}, \tag{A3}$$

where $\mathbf{b}^+ = (1, \mathbf{b}')'$, $\bar{m}_T(\tilde{}_T) = \left[\bar{m}_{kT}(\tilde{}_T), \bar{m}_{sT}(\tilde{}_T)\right]$, $\bar{m}_{kT}(\ )$ and $\bar{m}_{sT}(\ )$ are the sample means of $m_{kt}(\ )$ and $m_{st}(\ )$, which are defined in (13) and (15), respectively, and

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

But since the maximisation of (A3) 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 $N/[2\bar{m}_{kT}(\tilde{}_T)]$ 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 $LM\left(\tilde{}_T, \mathbf{b}\right)$ yields the required result. Finally, the asymptotic distribution of the sup test follows directly from the fact that $\sqrt{T}\bar{m}_{kT}(\ _0)$ and $\sqrt{T}\bar{m}_{sT}(\ _0)$ are asymptotically orthogonal under the null, with asymptotic variances $N(N+2)/2$ and $2(N+2)\mathbf{\Sigma}$, respectively. □

Proposition 7

The average Hessian matrix at the restricted parameter estimates $\bar{\phi}_T = (\bar{}'_T, 1, \mathbf{0})'$ is

$$\frac{1}{T}\sum_t \frac{\partial^2 l_t(\bar{}_T)}{\partial\ \partial\ '} = \frac{\partial^2 \bar{l}_T(\bar{}_T)}{\partial\ \partial\ '} = \begin{bmatrix} \partial^2 \bar{l}_T(\bar{}_T)/\partial\ \partial\ ' & \mathbf{0} \\ \mathbf{0}' & \bar{s}_{\psi\psi T}(\bar{}) \end{bmatrix} \tag{A4}$$

because $\partial^2 l_t(\ )/\partial\psi\partial\pi$ will be zero when $\psi = 1$ given that $s_{\psi t}(\ )$ is identically zero at that value. If $\bar{}_T$ does not maximize the GH log-likelihood, then the matrix in (A4) cannot be negative definite. However, since $\partial^2 \bar{l}_T(\bar{}_T)/\partial\ \partial\ '$ is negative definite because $\bar{}_T$ maximizes the Student t log-likelihood, then $\bar{s}_{\psi\psi T}(\bar{}_T)$ must be positive for $\bar{}_T$ not to be a maximum. This conclusion also applies asymptotically. Thus, under the alternative hypothesis, the expected value of $s_{\psi\psi t}(\ _0, 1, \mathbf{0})$ must be necessarily positive. In contrast, we can use a conditional version of the Barlett identities to show that $E\left[s_{\psi\psi t}(\ _0, 1, \mathbf{0})|\mathbf{z}_t, I_{t-1},\ _0, 1, \mathbf{0}\right] = 0$. □

Propositions 8 and 9

We can use again the results of Fiorentini, Sentana, and Calzolari (2003b) mentioned in the proof of Proposition 4, together with the results in Crowder (1976), to show that

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

where

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

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

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

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

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

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

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

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

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

As a result

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

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

B The score using the EM algorithm

The EM-type procedure that we follow is divided in two parts. In the maximisation step, we derive $l(\mathbf{y}_t|\xi_t,\mathbf{z}_t,I_{t-1};\ \ )$ and $l(\xi_t|\mathbf{z}_t,I_{t-1};\ \ )$ with respect to . Then, in the expectation step, we take the expected value of these derivatives given $I_T=\{(\mathbf{z}_1,\mathbf{y}_1),(\mathbf{z}_2,\mathbf{y}_2),\cdots,(\mathbf{z}_T,\mathbf{y}_T)\}$ and the parameter values.

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

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

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

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

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

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

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

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

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

$$\begin{aligned} f\left(\xi_t|I_T;\ \right) &= \frac{f\left(\mathbf{y}_t, \xi_t|\mathbf{z}_t, I_{t-1};\ \right)}{f\left(\mathbf{y}_t|\mathbf{z}_t, I_{t-1};\ \right)} \propto f\left(\mathbf{y}_t|\xi_t, \mathbf{z}_t, I_{t-1};\ \right) f\left(\xi_t|\mathbf{z}_t, I_{t-1};\ \right) \\ &\propto \xi_t^{\frac{N}{2}-\nu-1} \times \exp\left\{\frac{-1}{2}\left[\left(\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\ )\mathbf{p}_t + 1\right)\xi_t + \left(\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2\right)\frac{1}{\xi_t}\right]\right\}, \end{aligned}$$

which implies that

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

From here, we can use (D1) and (D2) to obtain the required moments. Specifically,

$$\begin{aligned} E\left(\xi_t|I_T;\ \right) &= \frac{\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}}{\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}} \\ &\quad \times R_{\frac{N}{2}-\nu}\left[\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}\right], \\ E\left(\left.\frac{1}{\xi_t}\right|I_T;\ \right) &= \frac{\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}}{\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}} \\ &\quad \times \frac{1}{R_{\frac{N}{2}-\nu-1}\left[\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}\right]}, \\ E\left(\log\xi_t|\,Y_T;\ \right) &= \log\left(\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}\right) - \log\left(\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}\right) \\ &\quad + \frac{\partial}{\partial x}\log K_x\left[\sqrt{\frac{\gamma c_t(\ )}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\ )\mathbf{b} + \gamma^2}\sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}\mathbf{p}_t + 1}\right]\Bigg|_{x=\frac{N}{2}-\nu}. \end{aligned}$$

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

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

$$-\frac{1}{2} \frac{c_t(\quad) - 1}{c_t(a \quad) \mathbf{b}' \boldsymbol{\Sigma}_t(\quad) \mathbf{b} \sqrt{1 + 4 \left(D_{\nu+1}(\gamma) - 1 \right) \mathbf{b}}}$$

where

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

$$\frac{\partial vec[\boldsymbol{\Sigma}_t^*(\ \ )]}{\partial\ \ '}=\frac{\partial vec[\boldsymbol{\Sigma}_t(\ \ )]}{\partial\ \ '}+\frac{c_t(\ \ )-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}}\left\{[\boldsymbol{\Sigma}_t(\ \ )\mathbf{bb}'\otimes I_N]+[I_N\otimes\boldsymbol{\Sigma}_t(\ \ )\mathbf{bb}']\right\}\frac{\partial vec[\boldsymbol{\Sigma}_t(\ \ )]}{\partial\ \ '}$$
$$+\frac{c_t(\ \ )-1}{[\mathbf{b}'\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}]^2}\left\{\frac{1}{\sqrt{1+4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}}}-1\right\}$$
$$\times vec\left[\boldsymbol{\Sigma}_t(\ \ )\mathbf{bb}'\boldsymbol{\Sigma}_t(\ \ )\right]vec'\left(\mathbf{bb}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\ \ )]}{\partial\ \ '},$$

$$\frac{\partial\mathbf{p}_t}{\partial\ \ '}=-\frac{\partial\boldsymbol{\mu}_t(\ \ )}{\partial\ \ '}+c_t(\ \ )\left[\mathbf{b}'\otimes I_N\right]\frac{\partial vec[\boldsymbol{\Sigma}_t(\ \ )]}{\partial\ \ '}$$
$$+\frac{c_t(\ \ )-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}}\frac{1}{\sqrt{1+4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}}}\boldsymbol{\Sigma}_t(\ \ )\mathbf{b}vec'\left(\mathbf{bb}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\ \ )]}{\partial\ \ '},$$

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

and

$$\frac{\partial c_t(\ \ )}{\partial\psi}=\frac{c_t(\ \ )-1}{D_{\nu+1}\left(\gamma\right)-1}\frac{\partial D_{\nu+1}\left(\gamma\right)}{\partial\psi}.$$

C Modified Bessel function of the third kind

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

$$K_\nu\left(x\right)=\frac{\pi}{2}\frac{I_{-\nu}\left(x\right)-I_\nu\left(x\right)}{\sin\left(\pi\nu\right)}. \tag{C1}$$

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

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

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

$$K_\nu\left(x\right)\simeq\sqrt{\frac{\pi}{2x}}e^{-x}\left\{1+\frac{m\text{-}1}{8x}+\frac{\left(m\text{-}1\right)\left(m\text{-}9\right)}{2!\left(8x\right)^2}+\frac{\left(m\text{-}1\right)\left(m\text{-}9\right)\left(m\text{-}25\right)}{3!\left(8x\right)^3}+\cdots\right\}. \tag{C2}$$

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

$$K_\nu(x) \simeq \sqrt{\frac{\pi}{2\nu}} \frac{\exp\left(-\nu l^{-1}\right)}{l^{-2}} \left[\frac{(x/\nu)}{1+l^{-1}}\right]^{-\nu} \left[1\text{-}\frac{3l\text{-}5l^3}{24\nu}+\frac{81l^2\text{-}462l^4+385l^6}{1152\nu^2}+\cdots\right], \qquad \text{(C3)}$$

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

• For $\nu \leq 10$ and $|x| > 12$, the derivative of (C2) with respect to ν gives a better approximation than the direct derivative of $K_\nu(x)$, which is in fact very unstable.

• For $\nu > 10$, the derivative of (C3) with respect to ν works better than the direct derivative of $K_\nu(x)$.

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

We can express such a derivative as a function of $I_\nu(x)$ by using (C1) as:

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

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

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

where for *integer* ν:

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

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

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

and

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

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

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

where

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

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

where

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

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

where

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

$$\begin{aligned} \frac{\partial^4 I_\nu(x)}{\partial \nu^4} &= 4\log\left(\frac{x}{2}\right)\frac{\partial^3 I_\nu(x)}{\partial \nu^3} - 6\left[\log\left(\frac{x}{2}\right)\right]^2\frac{\partial^2 I_\nu(x)}{\partial \nu^2} + 4\left[\log\left(\frac{x}{2}\right)\right]^3\frac{\partial I_\nu(x)}{\partial \nu} \\ &\quad - \left[\log\left(\frac{x}{2}\right)\right]^4 I_\nu(x) - \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{Q_4(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k, \end{aligned}$$

where

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

and finally,

$$\begin{aligned} \frac{\partial^5 I_\nu(x)}{\partial \nu^5} &= 5\log\left(\frac{x}{2}\right)\frac{\partial^4 I_\nu(x)}{\partial \nu^4} - 10\left[\log\left(\frac{x}{2}\right)\right]^2\frac{\partial^3 I_\nu(x)}{\partial \nu^3} + 10\left[\log\left(\frac{x}{2}\right)\right]^3\frac{\partial^2 I_\nu(x)}{\partial \nu^2} \\ &\quad -5\left[\log\left(\frac{x}{2}\right)\right]^4\frac{\partial I_\nu(x)}{\partial \nu} + \left[\log\left(\frac{x}{2}\right)\right]^5 I_\nu(x) - \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{Q_5(\nu+k+1)}{k!}\left(\frac{1}{4}x^2\right)^k, \end{aligned}$$

where

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

with

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

and

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

D Moments of the *GIG* distribution

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

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

where $K_{\nu}(\cdot)$ is the modified Bessel function of the third kind and $\delta, \gamma \ge 0$, $\nu \in \mathbb{R}$, $x > 0$. Two important properties of this distribution are $X^{-1} \sim GIG(-\nu, \gamma, \delta)$ and $(\gamma/\delta)X \sim GIG\left(\nu, \sqrt{\gamma\delta}, \sqrt{\gamma\delta}\right)$. For our purposes, the most useful moments of X when $\delta\gamma > 0$ are

$$E\left(X^k\right) = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{\nu+k}(\delta\gamma)}{K_{\nu}(\delta\gamma)} \tag{D1}$$

$$E(\log X) = \log\left(\frac{\delta}{\gamma}\right) + \frac{\partial}{\partial\nu} K_{\nu}(\delta\gamma). \tag{D2}$$

The *GIG* nests some well-known important distributions, such as the gamma ($\nu > 0$, $\delta = 0$), the reciprocal gamma ($\nu < 0$, $\gamma = 0$) or the inverse Gaussian ($\nu = -1/2$). Importantly, all the moments of this distribution are finite, except in the reciprocal gamma case, in which (D1) becomes infinite for $k \ge |\nu|$. A complete discussion on this distribution, including proofs of (D1) and (D2), can be found in Jørgensen (1982).

E Skewness and kurtosis of *GH* distributions

We can tediously show that

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

and

$$E\left[vec\left(\ ^{*}\ ^{*\prime}\right)vec'\left(\ ^{*}\ ^{*\prime}\right)\right]=E\left[\ ^{*}\ ^{*\prime}\otimes\ ^{*}\ ^{*\prime}\right]$$

$$=c^{4}(\ ,\nu,\gamma)\left[\frac{K_{\nu+4}(\gamma)K_{\nu}^{3}(\gamma)}{K_{\nu+1}^{4}(\gamma)}-4\frac{K_{\nu+3}(\gamma)K_{\nu}^{2}(\gamma)}{K_{\nu+1}^{3}(\gamma)}+6D_{\nu+1}(\gamma)-3\right]vec(\ \ ')\,vec'(\ \ ')$$

$$+c^{2}(\ ,\nu,\gamma)\left[\frac{K_{\nu+3}(\gamma)K_{\nu}^{2}(\gamma)}{K_{\nu+1}^{3}(\gamma)}-2D_{\nu+1}(\gamma)+1\right]$$

$$\times\left\{vec(\ \ ')\,vec'(\mathbf{AA}')+vec(\mathbf{AA}')\,vec'(\ \ ')+(\mathbf{K}_{NN}+\mathbf{I}_{N^{2}})\left[\ '\otimes\mathbf{AA}'\right](\mathbf{K}_{NN}+\mathbf{I}_{N^{2}})\right\}$$

$$+D_{\nu+1}(\gamma)\left\{\left[\mathbf{AA}'\otimes\mathbf{AA}'\right](\mathbf{K}_{NN}+\mathbf{I}_{N^{2}})+vec(\mathbf{AA}')\,vec'(\mathbf{AA}')\right\},$$

where

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

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

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

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

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

F Power of the normality tests

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

$$\begin{bmatrix}2/[N(N+2)] & \mathbf{0}'\\ \mathbf{0} & 1/[2(N+2)]\end{bmatrix}$$

evaluated at $\bar{m}_{T}=[\bar{m}_{kT},\bar{m}_{sT}']'$. To obtain the asymptotic distribution of $\sqrt{T}\bar{m}_{T}$ under the alternative of GH innovations, we can use the fact that when $\boldsymbol{\mu}_{t}(\ )=\mathbf{0}$ and $\boldsymbol{\Sigma}_{t}(\ )=\mathbf{I}_{N}$, we can write

$$\ _{t}^{*}=c(\ )\mathbf{b}(h_{t}-1)+\sqrt{h_{t}}\mathbf{Ar}_{t},$$

$$\varsigma_{t}=\ _{t}^{*\prime}\ _{t}^{*}=c^{2}(\ )(h_{t}-1)^{2}\mathbf{b}'\mathbf{b}+2c(\ )\sqrt{h_{t}}(h_{t}-1)\mathbf{b}'\mathbf{Ar}_{t}+h_{t}\mathbf{r}_{t}'\mathbf{A}'\mathbf{Ar}_{t},$$

with $h_{t}=\xi_{t}^{-1}\gamma/R_{\nu}(\gamma)$, and

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

where $\mathbf{r}_t|\mathbf{z}_t, I_{t-1} \sim iid\ N(0, \mathbf{I}_N)$ and $\xi_t|\mathbf{z}_t, I_{t-1} \sim iid\ GIG[.5\eta^{-1}, \psi^{-1}(1-\psi), 1]$ are mutually independent. Hence, since both ξ_t and $\mathbf{r}_t$ are iid , then ${}^{*}_{t}$ and $\varsigma_t = {}^{*\prime}_{t}\ {}^{*}_{t}$ 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 appendix D), we can apply the Lindeberg-Lévy Central Limit Theorem to show that the asymptotic distribution of $\sqrt{T}\bar{m}_T$ is $N[m(\eta, \psi, \mathbf{b}), V(\eta, \psi, \mathbf{b})]$, where the required expressions can be computed analytically. 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

$$\frac{KT}{T} = \frac{2}{N(N+2)}\bar{m}_{kT}^2 \cdot \mathbf{1}(\bar{m}_{kT} \geq 0) + \frac{1}{2(N+2)}\bar{m}_{sT}'\bar{m}_{sT}.$$

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

$$\Pr\left(\frac{KT}{T} < x\right) = \int_{-\infty}^{\infty} \Pr\left(\frac{KT}{T} < x \middle| \bar{m}_{kt} = l\right) f_k(l)\, dl, \tag{F1}$$

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

$$\Pr\left(\frac{KT}{T} < x \middle| \bar{a}_{kT}\right) = \Pr\left[\frac{1}{2(N+2)}\bar{m}_{sT}\bar{m}_{sT} < x - \frac{2}{N(N+2)}\bar{a}_{kT}^2 \cdot \mathbf{1}(\bar{a}_{kT} \geq 0) \middle| \bar{a}_{kT}\right].$$

Once we know this conditional probability, we can evaluate the integral in (F1) by numerical integration with a standard quadrature algorithm.

Table 1
Maximum likelihood estimates of a conditionally heteroskedastic model for 26 U.K. sectorial indices

	Gaussian		Student t		Generalised Hyperbolic	
Parameter		SE		SE		SE
α_1	.111	.075	.053	.026	.053	.033
α_2	.670	.258	.675	.120	.668	.135
ρ	.951	.629	1.0		1.0	
η	-	-	.103	.012	.113	.012
ψ	-	-	-		1.0	
Log-likelihood	-4,471.216		-4,221.162		-4,192.209	

Note: Monthly excess returns 1971:2-1990:10 (237 observations)

Table 2
Tests of Gaussian and Student t distributional assumptions

Normality tests

Test		p-value
Kurtosis component	2,962.6	.000
Skewness component	625.2	.000
Kuhn-Tucker	3,587.7	.000

Student t tests

Test		p-value
ψ component	0	1.000
Skewness-component	63.5	.000
Kuhn-Tucker	63.5	.000

Notes: Since the Kurtosis component of the normality test is positive, the supremum test is numerically identical to the Kuhn-Tucker test. The ψ component is the one-sided test of Student t vs symmetric GH, and the skewness component is the two-sided test of a Student t against an asymmetric Student t distribution.

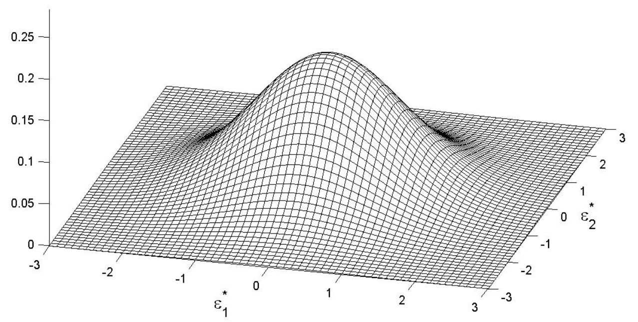

Figure 1a: Standardised bivariate normal density

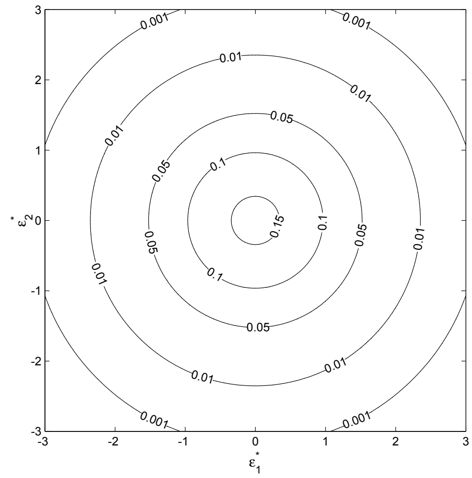

Figure 1b: Contours of a standardised bivariate normal density

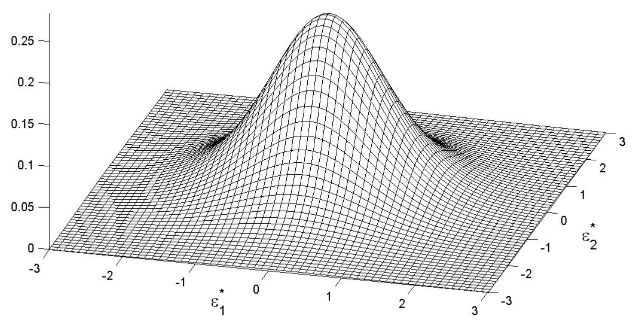

Figure 1c: Standardised bivariate Student t density with 8 degrees of freedom ($\eta = .125$)

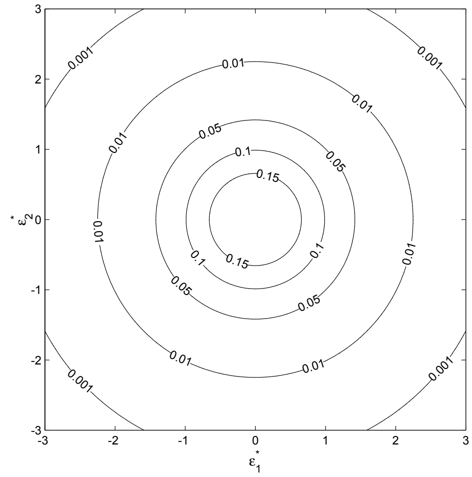

Figure 1d: Contours of a standardised bivariate Student t density with 8 degrees of freedom ($\eta = .125$)

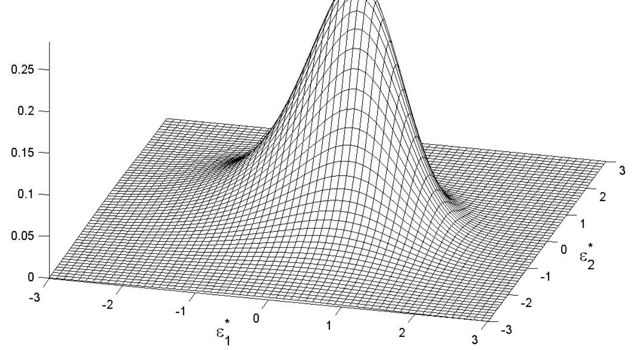

Figure 1e: Standardised bivariate asymmetric Student t density with 8 degrees of freedom ($\eta = .125$) and $\boldsymbol{\beta} = (-2, -2)'$

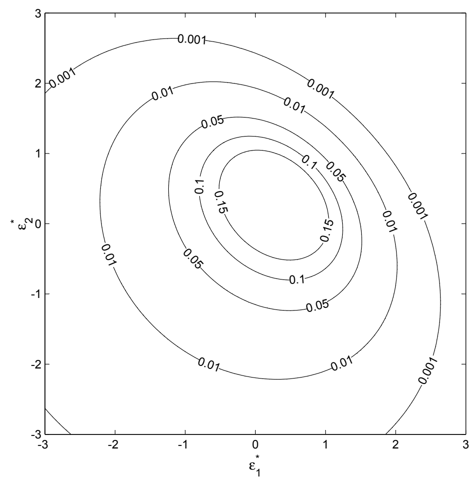

Figure 1f: Contours of a standardised bivariate asymmetric Student t density with 8 degrees of freedom ($\eta = .125$) and $\boldsymbol{\beta} = (-2, -2)'$

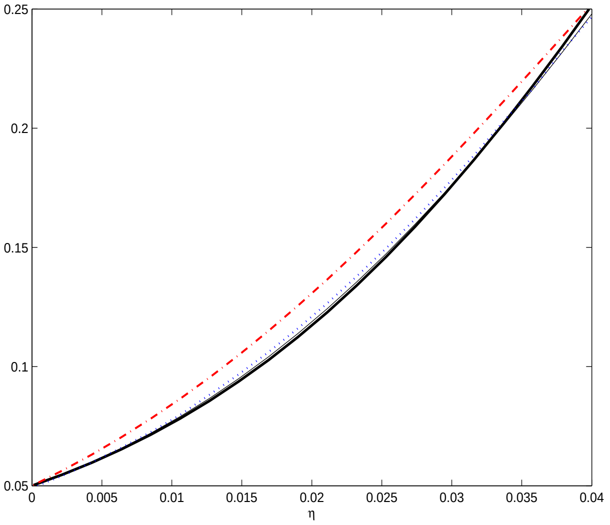

Figure 2a: Power of the univariate normality tests under symmetric alternatives ($N = 1$, $b = 0$)

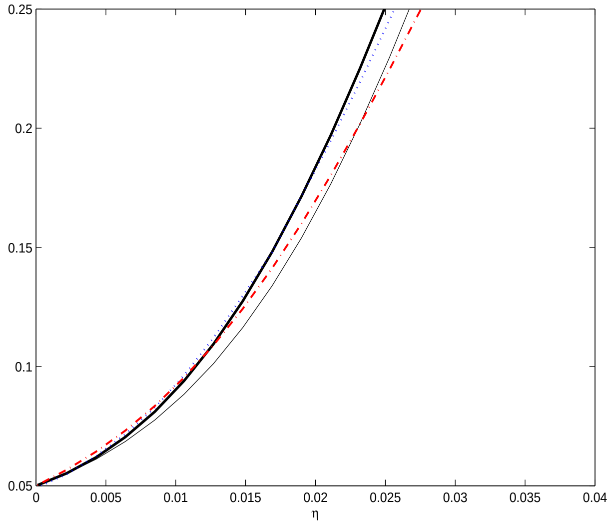

Figure 2b: Power of the univariate normality tests under asymmetric alternatives ($N = 1$, $b = 1$)

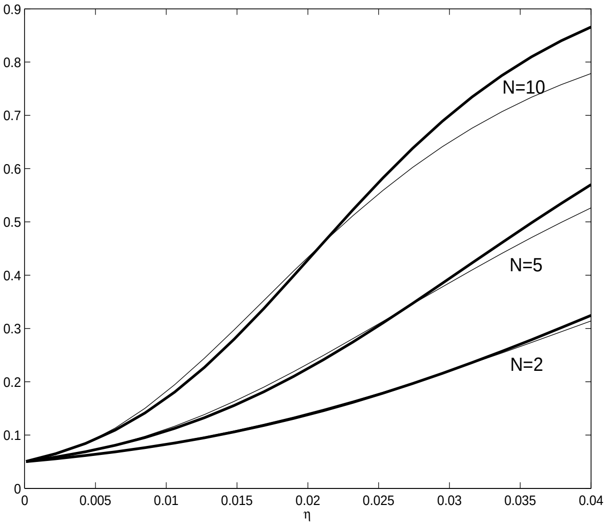

Figure 2c: Power of the multivariate normality tests under symmetric alternatives ($\mathbf{b} = \mathbf{0}$)

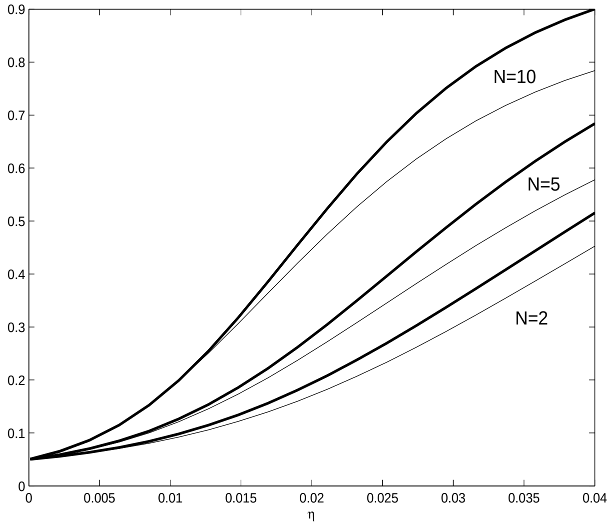

Figure 2d: Power of the multivariate normality tests under asymmetric alternatives ($\mathbf{b} = \mathbf{1}$)

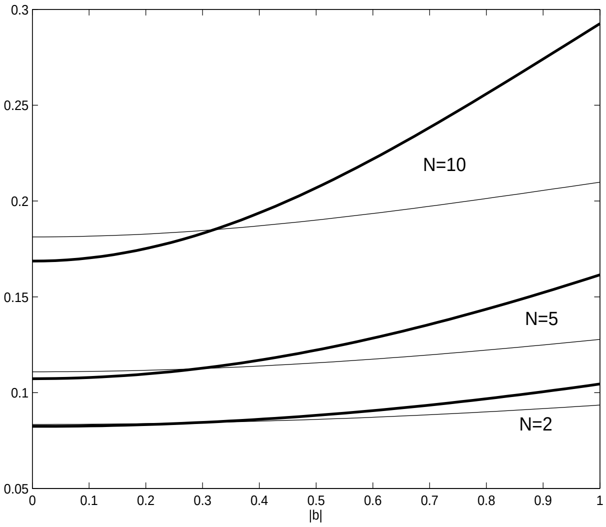

Figure 2e: Power of the normality tests against increasing skewness near normality ($\eta = .01$)

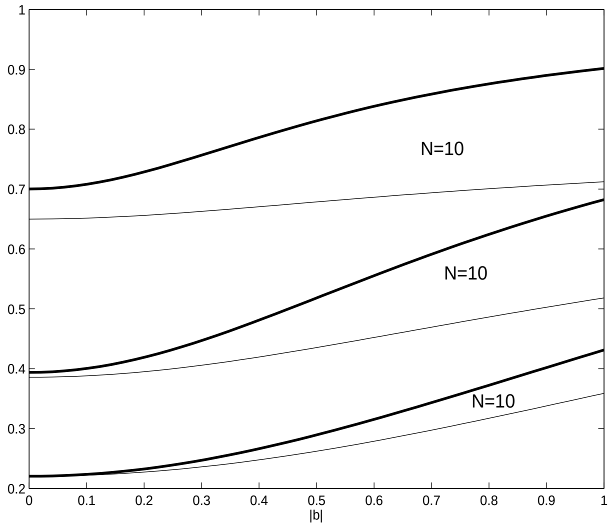

Figure 2f: Power of the normality test against increasing skewness ($\eta = .03$)

Supremum — Kurtosis(two-sided) — Kuhn-Tucker — Kurtosis(one-sided)

Notes: Size = 5%, $T = 100$, $\psi = 1$, $\boldsymbol{\mu}_t(\boldsymbol{\theta}_0) = \mathbf{0}$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \mathbf{I}_N$. Kurtosis tests mean tests of Normal vs symmetric Student t innovations.