

Estimation and Testing of Dynamic Models with Generalised Hyperbolic Innovations*

Javier Mencía

CEMFI

<mencia@cemfi.es>

Enrique Sentana

CEMFI

<sentana@cemfi.es>

February 2004
Revised: July 2005

Abstract

We analyse the Generalised Hyperbolic distribution adequacy to model kurtosis and asymmetries in multivariate conditionally heteroskedastic dynamic regression models. We standardise this distribution, obtain analytical expressions for the log-likelihood score, and explain how to evaluate the information matrix. We also 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 application to five NASDAQ sectorial stock returns that indicates that their conditional distribution is asymmetric and leptokurtic, which can be successfully exploited for risk management purposes.

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

JEL: C52, C32, G11

*We are grateful to Manuel Arellano, Ole Barndorff-Nielsen, Luc Bauwens, Anil Bera, Rob Engle, Gabriele Fiorentini, Eric Ghysels, Vassilis Hajivassiliou, Sébastien Laurent, Eric Renault, Luis Seco, Neil Shephard and Mervyn Sylvapulle, as well as seminar audiences at Aarhus, Alicante, CORE, Duke, ESRC Econometric Study Group, LSE FMG, MEFF, Monash, UIB, UPN, the 2004 CIRANO-CIREQ Financial Econometrics Conference (Montreal), the 2004 European Meetings of the Econometric Society (Madrid), the XII Finance Forum (Barcelona) and the XXIX Symposium on Economic Analysis (Pamplona) for helpful comments and suggestions. 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 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 $\alpha < 1/2$. 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 (2003) (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 rather flexible Generalised Hyperbolic (GH) distribution introduced by Barndorff-Nielsen (1977). 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). Although the GH distribution has been used to model the unconditional distribution of financial returns (see e.g. Prause, 1998), to the best of our knowledge it has not yet been used in its more general form for modelling the conditional distribution of financial time series, which is the relevant one from a risk management perspective.

Our approach is related to Bera and Premaratne (2002), who also nest the Student t distribution by using Pearson’s type IV distribution in univariate static models. However, they 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 implementation of their technique becomes increasingly difficult in large dimensions, as the number of orthants is 2^N , where N denotes the number of assets. Similarly, semi-parametric procedures, including Hermite polynomial expansions, become infeasible for moderately large N , unless one maintains the assumption of elliptical symmetry, and the same is true of copulae methods.

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. Section 4 presents the results of several Monte Carlo experiments. Finally we include an empirical application to NASDAQ sectorial stock returns in section 5, 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 $\nu, \delta, \gamma, \boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\Upsilon}$, which we write as $\mathbf{u} \sim GH_N(\nu, \delta, \gamma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Upsilon})$, then its density will be given by

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

where $\nu \in \mathbb{R}$, $\delta, \gamma \in \mathbb{R}^+$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^N$, $\boldsymbol{\Upsilon}$ 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, p. 374, as well as appendix C), and $q[\delta^{-1}(\mathbf{u} - \boldsymbol{\alpha})] = \sqrt{1 + \delta^{-2}(\mathbf{u} - \boldsymbol{\alpha})'\boldsymbol{\Upsilon}^{-1}(\mathbf{u} - \boldsymbol{\alpha})}$.

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

$$\mathbf{u} = \boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi^{-1} + \xi^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\mathbf{r}, \quad (1)$$

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

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 $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ as follows:

¹Although (1) is written in terms of the reciprocal of ξ by analogy with the usual presentation of the Student t distribution, we could alternatively work with $\chi = \xi^{-1}$ without loss of generality since the reciprocal of a GIG random variable is also GIG (see appendix D).

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

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

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

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

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

Thus, we can achieve a leptokurtic and asymmetric standardised multivariate distribution with only $N + 2$ additional parameters.

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 $\nu \rightarrow -\infty$ or (ii) $\nu \rightarrow +\infty$, regardless of the values of γ and $\boldsymbol{\beta}$; and (iii) when $\gamma \rightarrow \infty$ irrespective of the values of ν and $\boldsymbol{\beta}$.
- **Symmetric Student t** , obtained when $-\infty < \nu < -2$, $\gamma = 0$ and $\boldsymbol{\beta} = \mathbf{0}$.
- **Asymmetric Student t** , which is like its symmetric counterpart except that the vector $\boldsymbol{\beta}$ of skewness parameters is no longer zero.
- **Asymmetric Normal-Gamma**, which is obtained when $\gamma = 0$ and $0 < \nu < \infty$.
- **Normal Inverse Gaussian**, for $\nu = -.5$ (see Aas, Dimakos, and Haff, 2004).
- **Hyperbolic**, for $\nu = 1$ (see Chen, Härdle, and Jeong, 2004)

More generally, the distribution of $\boldsymbol{\varepsilon}^*$ becomes a simple scale mixture of normals, and thereby spherical, when $\boldsymbol{\beta}$ 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 5 confirms.

Another important feature of the standardised *GH* distribution is that, although the elements of $\boldsymbol{\varepsilon}^*$ 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 $\boldsymbol{\varepsilon}^*$ 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 β appropriately. Specifically, we can make the probability of extremely low realisations of several 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. This is confirmed by Figure 2, which represents the so-called exceedance correlation between the uncorrelated marginal components in Figure 1. Hence, the GH distribution could capture the empirical observation that there is higher tail dependence across stock returns in market downturns (see Longin and Solnik, 2001).

Finally, Blæsild (1981) showed that the marginal distributions of linear combinations of GH variables (including the individual components) are also GH .² In our notation, his result can be stated as follows:

Proposition 2 *Let $\boldsymbol{\varepsilon}^*$ be distributed as a $N \times 1$ standardised GH random vector with parameters ν , γ and $\boldsymbol{\beta}$. Then, for any vector $\mathbf{w} \in \mathbb{R}^N$, with $\mathbf{w} \neq \mathbf{0}$, $s^* = \mathbf{w}'\boldsymbol{\varepsilon}^*/\sqrt{\mathbf{w}'\mathbf{w}}$ is distributed as a standardised GH scalar random variable with parameters ν , γ and*

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

Note that only the skewness parameter, $\beta(\mathbf{w})$, is affected, as it becomes a function of the weights, \mathbf{w} . This result is particularly useful in financial applications, since the returns to any conceivable portfolio of a collection of assets will be a linear combination of the returns on those primitive assets. It also implies that skewness is a “common feature” of the GH distribution, in the Engle and Kozicki (1993) sense, as we can generate full-rank linear combinations of $\boldsymbol{\varepsilon}^*$ with the asymmetry confined to a single element.

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

$$\left. \begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*, \\ \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}), \end{aligned} \right\} \quad (4)$$

²Likewise, the distribution of a subvector of $\boldsymbol{\varepsilon}^*$ conditional on the values taken by a different subvector is also GH (see Blæsild, 1981, for a proof, and Mencía, 2005, for financial applications).

where $\boldsymbol{\mu}(\cdot)$ and $\text{vech}[\boldsymbol{\Sigma}(\cdot)]$ are N and $N(N+1)/2$ -dimensional vectors of functions known up to the $p \times 1$ vector of true parameter values, $\boldsymbol{\theta}_0$, \mathbf{z}_t are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of \mathbf{y}_t and \mathbf{z}_t , $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some $N \times N$ “square root” matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2'}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t^*$ is a vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$. As a consequence, $E(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and $V(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$.

In this context, FSC assumed that $\boldsymbol{\varepsilon}_t^*$ 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 $\boldsymbol{\varepsilon}_t^*$.

Importantly, given that $\boldsymbol{\varepsilon}_t^*$ is not generally observable, the choice of “square root” matrix is not irrelevant except in univariate *GH* models, or in multivariate *GH* models in which either $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is time-invariant or $\boldsymbol{\varepsilon}_t^*$ is spherical (i.e. $\boldsymbol{\beta} = \mathbf{0}$), a fact that previous efforts to model multivariate skewness in dynamic models have overlooked (see e.g. Bauwens and Laurent, 2002). Therefore, if there were reasons to believe that $\boldsymbol{\varepsilon}_t^*$ were not only a martingale difference sequence, but also serially independent, then we could in principle try to estimate the “unique” orthogonal rotation underlying the “structural” shocks. However, since we believe that such an identification procedure would be neither empirically plausible nor robust, we prefer the conditional distribution of \mathbf{y}_t *not* to depend on whether $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is a symmetric or lower triangular matrix, nor on the order of the observed variables in the latter case. This can be achieved by making $\boldsymbol{\beta}$ a function of past information and a new vector of parameters \mathbf{b} in the following way:

$$\boldsymbol{\beta}_t(\boldsymbol{\theta}, \mathbf{b}) = \boldsymbol{\Sigma}_t^{\frac{1}{2}'}(\boldsymbol{\theta})\mathbf{b}. \quad (5)$$

It is then straightforward to see that the resulting *GH* log-likelihood function will not depend on the choice of $\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})$.³ Finally, it is analytically convenient to replace ν and γ by η and ψ , where $\eta = -.5\nu^{-1}$ and $\psi = (1 + \gamma)^{-1}$, although we continue to use ν and γ in some equations for notational simplicity.⁴

³Nevertheless, it would be fairly easy to adapt all our subsequent expressions to the alternative assumption that $\boldsymbol{\beta}_t(\boldsymbol{\theta}, \mathbf{b}) = \mathbf{b} \forall t$ (see Mencía, 2003).

⁴An undesirable aspect of this reparametrisation is that the log-likelihood is continuous but non-

2.3 The log-likelihood function

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \eta, \psi, \mathbf{b})'$ denote the parameters of interest. The log-likelihood function of a sample of size T takes the form $L_T(Y_T|\boldsymbol{\phi}) = \sum_{t=1}^T l(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi})$, where $l(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi})$ is the conditional log-density of \mathbf{y}_t given \mathbf{z}_t , I_{t-1} and $\boldsymbol{\phi}$. Given the non-linear nature of the model, a numerical optimisation procedure is usually required to obtain maximum likelihood (ML) estimates of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$ say. Assuming that all the elements of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are twice continuously differentiable functions of $\boldsymbol{\theta}$, we can use a standard gradient method in which the first derivatives are numerically approximated by re-evaluating $L_T(\boldsymbol{\phi})$ with each parameter in turn shifted by a small amount, with an analogous procedure for the second derivatives. Unfortunately, such numerical derivatives are sometimes unstable, and moreover, their values may be rather sensitive to the size of the finite increments used. 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 Vinod, 1999). Moreover, a fast and numerically reliable procedure for the computation of the score for any value of $\boldsymbol{\phi}$ 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 $s_t(\boldsymbol{\phi}) = \partial l(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}) / \partial \boldsymbol{\phi}$. The idea is based on the following dual decomposition of the joint log-density (given \mathbf{z}_t , I_{t-1} and $\boldsymbol{\phi}$) of the observable process \mathbf{y}_t and the latent mixing process ξ_t :

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

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

differentiable with respect to η at $\eta = 0$, even though it is continuous and differentiable with respect to ν for all values of ν . The problem is that at $\eta = 0$, we are pasting together the extremes $\nu \rightarrow \pm\infty$ into a single point. Nevertheless, it is still worth working with η instead of ν when testing for normality.

and ϕ) of the observable and unobservable processes, respectively. If we differentiate both sides of the previous identity with respect to ϕ , and take expectations given the full observed sample, I_T , then we will end up with:

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

because $E[\partial l(\xi_t | \mathbf{y}_t, \mathbf{z}_{t-1}, I_{t-1}; \phi) / \partial \phi | I_T; \phi] = \mathbf{0}$ by virtue of the Kullback inequality. In this way, we decompose $s_t(\phi)$ 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).⁵

For the purposes of developing our testing procedures in section 3, it is convenient to obtain closed-form expressions for $s_t(\phi)$ 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 $\eta \rightarrow 0^+$ or (ii) $\eta \rightarrow 0^-$ regardless of the values of \mathbf{b} and ψ ; and (iii) when $\psi \rightarrow 0$, irrespective of η and \mathbf{b} . Therefore, it is not surprising that the Gaussian scores with respect to η or ψ are 0 when these parameters are not identified, and also, that $\lim_{\eta, \psi \rightarrow 0} s_{\mathbf{b}t}(\phi) = \mathbf{0}$. Similarly, the limit of the score with respect to the mean and variance parameters, $\lim_{\eta, \psi \rightarrow 0} s_{\theta t}(\phi)$, coincides with the usual Gaussian expressions (see e.g. Bollerslev and Wooldridge (1992)). Further, we can show that for fixed $\psi > 0$,

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

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \mu_t(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, which confirms the non-differentiability of the log-likelihood function with respect to η at $\eta = 0$ (see footnote 4). Finally, we can show that for $\eta \neq 0$, $\lim_{\psi \rightarrow 0} s_{\psi t}(\phi)$ 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 $\psi \rightarrow 1$ and $\mathbf{b} \rightarrow 0$ of the general score function. Not surprisingly, the score with respect to $\boldsymbol{\pi}$, where $\boldsymbol{\pi} = (\boldsymbol{\theta}', \eta)'$, coincides with

⁵It is possible to show that the latent variable ξ_t could be fully recovered from observations on \mathbf{y}_t as $N \rightarrow \infty$, which would greatly simplify the calculations implicit in expression (6).

the formulae in FSC. But our more general GH assumption introduces two additional terms: the score with respect to \mathbf{b} ,

$$s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, 0) = \frac{\eta [\varsigma_t(\boldsymbol{\theta}) - (N + 2)]}{1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \quad (8)$$

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 $s_t(\boldsymbol{\phi})$ evaluated at $\boldsymbol{\phi}_0$ has the martingale difference property under certain regularity conditions. In addition, his results also imply that under additional regularity conditions (which in particular require that $\boldsymbol{\phi}_0$ 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

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

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

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

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 $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_{T_s}$, where the symbol $\hat{}$ indicates that the data has been generated using the GH maximum likelihood estimates $\hat{\boldsymbol{\phi}}_T$. Then, if we denote by $s_{t_s}(\hat{\boldsymbol{\phi}}_T)$ the value of the score function for each simulated observation, our proposed estimator of the information matrix is

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

where we can get arbitrarily close in a numerical sense to the value of the asymptotic information matrix evaluated at $\hat{\boldsymbol{\phi}}_T$, $\mathcal{I}(\hat{\boldsymbol{\phi}}_T)$, 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, 1998). We shall investigate the finite sample performance of these alternative estimators of the sampling variance of the ML estimators in section 4. In some special cases, though, it is also possible to estimate $\mathcal{I}(\boldsymbol{\phi}_0)$ as the sample average of the conditional information matrix $\mathcal{I}_t(\boldsymbol{\phi}) = \text{Var}[s_t(\boldsymbol{\phi}) | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}]$. In particular, analytical expressions for $\mathcal{I}_t(\boldsymbol{\phi})$ 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 $\eta \rightarrow 0^+$,

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

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

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

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

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

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

2.5.2 The conditional information matrix under Student t innovations

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

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

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

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. $\eta \rightarrow 0^+$, $\eta \rightarrow 0^-$ or $\psi \rightarrow 0$, 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 $\psi \rightarrow 0$, while ψ and \mathbf{b} are unidentified when $\eta \rightarrow 0^\pm$.

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). 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, the motivation being that this is precisely what the ordinary likelihood ratio (LR) test will do in those circumstances at a larger computational cost (see Hansen, 1991). 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.

3.1.1 LM test for fixed values of the unidentified parameters

Let $\tilde{\boldsymbol{\theta}}_T$ denote the ML estimator of $\boldsymbol{\theta}$ obtained by maximising a Gaussian log-likelihood function. For the case in which normality is achieved as $\eta \rightarrow 0^+$, 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 $\boldsymbol{\theta}$ and η , $\bar{s}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b})$ and $\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b})$, with 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 $\bar{s}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T, 0^+, \psi, \mathbf{b}) = \mathbf{0}$ by definition of $\tilde{\boldsymbol{\theta}}_T$, and $\mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\theta}_0, 0^+, \psi, \mathbf{b}) = 0$, we can show that

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

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

Proposition 5 (LM normality test)

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

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

Under standard regularity conditions, $LM(\tilde{\boldsymbol{\theta}}_T, \mathbf{b})$ will be asymptotically chi-square with one degree of freedom for a given \mathbf{b} under the null hypothesis of normality, which effectively imposes the single restriction $\eta \cdot \psi = 0$ on the parameter space.

3.1.2 The supremum LM test

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

Proposition 6 (Supremum test)

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

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

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

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. $LM_k(\tilde{\boldsymbol{\theta}}_T)$, 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 $\varsigma_t(\boldsymbol{\theta}_0)$ implicit in

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

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

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

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:

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

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

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

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 also assesses if (14) holds, we do not discuss it separately.

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

By analogy with Bontemps and Meddahi (2005), one possible way to adjust Mardia's (1970) formulae is to replace $\varepsilon_{it}^{*3}(\boldsymbol{\theta})$ by $H_3[\varepsilon_{it}^*(\boldsymbol{\theta})]$ and $\varepsilon_{it}^{*2}(\boldsymbol{\theta})\varepsilon_{jt}^*(\boldsymbol{\theta})$ by $H_2[\varepsilon_{it}^*(\boldsymbol{\theta})]H_2[\varepsilon_{jt}^*(\boldsymbol{\theta})]$ ($i \neq j$) in the moment conditions (14), where $H_k(\cdot)$ is the Hermite polynomial of order k . Alternatively, we can correct the asymptotic size by treating (14) as moment conditions, with the Gaussian scores defining the PML estimators $\tilde{\boldsymbol{\theta}}_T$ (see Newey (1985) and Tauchen (1985) for the general theory, and appendix F for specific details).

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

3.1.3 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

$$E[m_{kt}(\boldsymbol{\theta}_0) | \boldsymbol{\theta}_0, \eta_0 > 0, \psi_0 > 0] > 0. \quad (16)$$

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 by making it asymptotically equivalent to the (sup) LR test (see also Andrews, 2001). Specifically, we recommend the use of

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

where $\mathbf{1}(\cdot)$ is the indicator function, and $\bar{m}_{kT}(\boldsymbol{\theta})$ the sample mean of $m_{kt}(\boldsymbol{\theta}_0)$. Asymptotically, the probability that $\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T)$ becomes negative is .5 under the null. Hence,

$KT(\tilde{\boldsymbol{\theta}}_T)$ 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 $\Pr(X > c) = 1 - .5F_{\chi_N^2}(c) - .5F_{\chi_{N+1}^2}(c)$ (see e.g. Demos and Sentana, 1998)

3.1.4 Power of the normality test

Although we shall investigate the finite sample properties of the different multivariate normality tests in section 4, it is interesting to study their asymptotic power properties. However, since the block-diagonality of the information matrix between $\boldsymbol{\theta}$ and the other parameters is generally lost under the alternative of GH innovations, for the purposes of this exercise we only consider models in which $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are constant but otherwise unrestricted, so that we can analytically compute the information matrix. In more complex parametrisations, though, the results are likely to be qualitatively similar.

The results at the usual 5% significance level are displayed in Figures 3a to 3d for $\psi = 1$ and $T = 5,000$ (see appendix F for details). In Figures 3a to 2b, we have represented η on the x -axis. We can see in Figure 3a that for $\mathbf{b} = \mathbf{0}$ and $N = 3$, the test with the highest power is the one-sided kurtosis test, followed by its two-sided counterpart, the KT test, the sup test, and finally the skewness test. On the other hand, if we consider asymmetric alternatives in which \mathbf{b} is proportional to a vector of ones $\boldsymbol{\iota}$, such as in Figure 3b, which is not restrictive because the power of our normality test only depends on \mathbf{b} through its Euclidean norm, the skewness component of the normality test becomes important, and eventually makes the KT test, the sup test and even the skewness test itself more powerful than both kurtosis tests. Not surprisingly, we can also see from these figures that if we apply our tests to a single component of the random vector, their power is significantly reduced.

In contrast, we have represented b_i on the x -axis in Figures 3c and 3d. There we can clearly see the effects on power of the fact that \mathbf{b} is not identified in the limiting case of $\eta = 0$. When η is very low, \mathbf{b} is almost unidentified, which implies that large increases in b_i have a minimum impact on power, as shown in Figure 3c for $\eta = .005$ and $N = 3$. However, when we give η a larger value such as $\eta = .01$ (see Figure 3d), we can see how the power of those normality test that take into account skewness rapidly increases with the asymmetry of the true distribution. Hence, we can safely conclude that, once we get away from the immediate vicinity of the null, the inclusion of the skewness component

of our test can greatly improve its power. On the other hand, the power of the kurtosis test, which does not account for skewness, is less sensitive to increases in b_i . Similar results are obtained for $N = 1$, which we do not present to avoid cluttering the pictures.

Finally, we have also compared the power of our sup test with those of the moment versions of Mardia's (1970) and Lütkepohl (1993) tests, where this time we have assumed that $\mathbf{b} = (b_1, 0, 0)'$ under the alternative for computational simplicity. The results show the superiority of our proposed test against both symmetric and asymmetric GH alternatives (see Figures 3e and 3f, respectively), which confirms the fact that it is testing the most relevant moment conditions.

3.2 Student t tests

3.2.1 Student t vs symmetric GH innovations

As we saw before, the Student t distribution is nested in the GH family when $\psi = 1$ and $\mathbf{b} = \mathbf{0}$. We can use this fact to test the validity of the distributional assumptions made by FSC and other authors. In this respect, a test of $H_0 : \psi = 1$ under the maintained hypothesis that $\mathbf{b} = \mathbf{0}$ would be testing that the tail behaviour of the multivariate t distribution adequately reflects the kurtosis of the data.

As we mentioned in section 2.4.2, though, it turns out that $s_{\psi t}(\boldsymbol{\pi}, 1, \mathbf{0}) = 0 \forall t$, which means that we cannot compute an LM test for $H_0 : \psi = 1$. To deal with this unusual type of testing situation, Lee and Chesher (1986) propose to replace the LM test by what they call an "extremum test" (see also Bera, Ra, and Sarkar, 1998). Given that the first-order conditions are identically 0, their suggestion is to study the restrictions that the null imposes on higher order conditions. In our case, we will use a moment test based on the second order derivative

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

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

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

under the alternative of standardised symmetric GH innovations.

Let $\bar{\boldsymbol{\pi}}_T = (\bar{\boldsymbol{\theta}}_T', \bar{\eta}_T)'$ denote the parameters estimated by maximising the symmetric Student t log-likelihood function. The statistic that we propose to test for $H_0 : \psi = 1$

versus $H_1 : \psi \neq 1$ under the maintained hypothesis that $\mathbf{b} = \mathbf{0}$ is given by

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

where $\hat{V}[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]$ is a consistent estimator of the asymptotic variance of $s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ that takes into account the sampling variability in $\bar{\boldsymbol{\pi}}_T$. Under the null hypothesis of Student t innovations with more than 4 degrees of freedom,⁶ it is easy to see that the asymptotic distribution of $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ will be $N(0, 1)$. The required expression for $V[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]$ is given in the following result:

Proposition 7 (Student t symmetric test) *If $\boldsymbol{\varepsilon}_t^*$ is conditionally distributed as a standardised Student t with $\eta_0^{-1} > 4$ degrees of freedom, then*

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

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

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

and

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

where

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

But given (18), and the fact that 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.

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

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

but

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

⁶If $.25 < \eta_0 < .5$ the variance of $[N - \zeta_t(\boldsymbol{\theta})]$ becomes unbounded. Given that the expected value of this term remains 0 under the alternative hypothesis, the obvious solution is to base the test on the first component of (17) only. The exact implementation details are available on request.

3.2.2 Student t vs asymmetric GH innovations

By construction, the previous test maintains the assumption that $\mathbf{b} = \mathbf{0}$. However, it is straightforward to extend it to incorporate this symmetry restriction as an explicit part of the null hypothesis. The only thing that we need to do is to include $E[s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, \mathbf{0})] = \mathbf{0}$ as an additional condition in our moment test, where $s_{\mathbf{b}t}(\boldsymbol{\pi}, 1, \mathbf{0})$ is defined in (8). The asymptotic joint distribution of the two moment conditions that takes into account the sampling variability in $\bar{\boldsymbol{\pi}}_T$ is given in the following result

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

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

where

$$\begin{aligned} \mathcal{V}(\boldsymbol{\pi}_0) &= \begin{bmatrix} \mathcal{V}_{\mathbf{b}\mathbf{b}}(\boldsymbol{\pi}_0) & \mathcal{V}_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) \\ \mathcal{V}'_{\mathbf{b}\psi}(\boldsymbol{\pi}_0) & \mathcal{V}_{\psi\psi}(\boldsymbol{\pi}_0) \end{bmatrix} = \left\{ \begin{array}{cc} \mathcal{I}_{\mathbf{b}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{0} \\ \mathbf{0}' & V[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})] \end{array} \right\} \\ &- \begin{bmatrix} \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \\ \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{M}(\boldsymbol{\pi}_0) \end{bmatrix}, \quad (20) \end{aligned}$$

$\mathcal{I}_{\pi\pi}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\pi\pi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ is the Student t information matrix derived in FSC, $\mathcal{I}_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\pi\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ and $\mathcal{I}_{\mathbf{b}\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) = E[\mathcal{I}_{\mathbf{b}\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ are defined in Proposition 4, and $\mathcal{M}(\boldsymbol{\pi}_0)$ and $V[s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0})]$ are given in Proposition 7.

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

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

which is distributed as a chi-square with $N + 1$ degrees of freedom under the null of Student t innovations. Alternatively, we can again exploit the one-sided nature of the ψ -component of the test. However, since $\mathcal{V}(\boldsymbol{\pi}_0)$ 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 LR test (see e.g. Silvapulle and Silvapulle, 1995). Specifically, instead of using directly the score with respect to \mathbf{b} , we consider

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

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

It is then straightforward to see that the asymptotic distribution of

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

is another 50:50 mixture of chi-squares with N and $N + 1$ degrees of freedom under the null, because asymptotically, the probability that $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ is negative will be .5 if $\psi_0 = 1$. Such a one-sided test benefits from the fact that a non-positive $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ gives no evidence against the null, in which case we only need to consider the orthogonalised skewness component. In contrast, when $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})$ is positive, (22) numerically coincides with (21).

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

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

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, $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ 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 $LM_s(\tilde{\boldsymbol{\theta}}_T)$ by multiplying it by a consistent estimator of the adjusting factor $[(1 - 4\eta_0)(1 - 6\eta_0)]/[1 + (N - 2)\eta_0 + 2(N + 4)\eta_0^2]$. Alternatively, we can run the univariate regression of 1 on $m_{st}(\bar{\boldsymbol{\theta}}_T)$, or the multivariate regression of $\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\theta}}_T)$ on $\varsigma_t(\bar{\boldsymbol{\theta}}_T) - (N + 2)$, although in the latter case we should use standard errors that are robust to heteroskedasticity. Not surprisingly, we can show that these three procedures to test (15) are asymptotically equivalent under the null. However, they are generally less powerful against local alternatives of the form $\mathbf{b}_{0T} = \mathbf{b}_0/\sqrt{T}$ than $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ in (23), which is the proper LM test for symmetry.

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

4 A Monte Carlo comparison of finite sample size properties and standard error estimators

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

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

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

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

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

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

where $\boldsymbol{\mu}' = (\mu_1, \mu_2, \mu_3)$, $\mathbf{c}' = (c_1, c_2, c_3)$, and $\boldsymbol{\Gamma}_t = \text{diag}(\gamma_{1t}, \gamma_{2t}, \gamma_{3t})$. As for parameter values, we have chosen $\mu_i = .2$, $c_i = 1$, $\alpha_1 = \phi_1 = .1$, $\alpha_2 = \phi_2 = .85$, $\alpha_0 = 1 - \alpha_1 - \alpha_2$ and $\phi_0 = 1 - \phi_1 - \phi_2$. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 1000$ observations based on 15,000 Monte Carlo replications. Further details are available on request.

Figures 4a-4c summarise our findings for the different multivariate normality tests, as well as their asymmetric and kurtosis components, by means of Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. As expected, the tests based on the original Mardia (1970) and Lütkepohl (1993) expressions, which were derived under the assumption that the covariance matrix of the innovations is constant, unrestricted and does not affect the conditional mean, suffer very large size distortions in a conditionally heteroskedastic model such as this. At the same time, if we interpret their tests as moment tests, and adjust them appropriately, their size distortions essentially disappear. More importantly for our purposes, the actual finite sample sizes of the one-sided

and two-sided versions of our proposed multivariate normality test also seem to be very close to their nominal levels, with the possible exception of the one-sided version of the kurtosis test, which seems to be somewhat conservative for larger nominal sizes.

Finite sample sizes of the Student t tests In this case we maintain the conditional mean and variance specification in (24), but generate the standardised innovations ε_t^* from a trivariate Student t distribution with 10 degrees of freedom. Figure 5a shows the p-value discrepancy plots of the one- and two-sided versions of the Student t tests discussed in section 3.2, together with those of their asymmetric and kurtosis components. The most striking feature of these graphs is the fact that the actual sizes of the “kurtosis” tests based on $\tau_{kT}(\bar{\pi}_T)$, which is defined in (19), are well below their nominal sizes. This is due to the fact that the sampling distribution of $\tau_{kT}(\bar{\pi}_T)$ is not well approximated by a standard normal, as illustrated in Figure 5b. In contrast, the actual sizes of the asymmetry component are very much on target, while those of the joint tests inherit part of the size distortions of the kurtosis tests.

Finite sample properties of standard error estimators in GH models Finally, we assess the performance of three possible ways of estimating the standard errors in GH models, namely, outer-product of the gradient (O), numerical Hessian (H) and information (I) matrix, which we obtain by simulation using the ML estimators as if they were the true parameter values, as suggested in section 2.5. Once again, we maintain the conditional mean and variance specification in (24), but this time we generate the standardised innovations ε_t^* from a trivariate asymmetric Student t distribution with $\eta = .1$ and $\mathbf{b} = -.1\boldsymbol{\iota}$. Since the purpose of this exercise is to guide empirical work, our target is the sampling covariance matrix of the ML estimators, $V_T(\hat{\phi}_T)$, which we estimate as the Monte Carlo covariance matrix of $\hat{\phi}_T$ in 30,000 samples of 1,000 observations each. Given the large number of parameters involved, we summarise the performance of the estimators of $V_T(\hat{\phi}_T)$ by looking at the sampling distributions of $vech'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vech[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$ and $vecd'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vecd[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$, where E is either O, H or I.⁷ The results, which are presented in Figures 6a and 6b, respectively, show that the I standard errors seem to be systematically more reliable than either the O or numerical H counterparts.

⁷In the case of a single parameter, the mean of the sampling distribution of these two norms reduce to the mean square error of the different estimators of its sampling variance.

5 Empirical application

We now apply the methodology derived in the previous sections to the returns on five sectorial NASDAQ indices, namely banks, other finance, industrials, insurance and transportation, which we obtained from www.finance.yahoo.com. Specifically, our data consists on daily excess returns for the period January 5th, 1982 - June 3rd, 2005 (5898 observations), where we have used the Eurodollar overnight interest rate as safe rate (Datastream code ECUSDST). The model used is a generalisation of the one in the previous section (see (24)), in which the mean dynamics are captured by a diagonal VAR(1) model with drift, and the covariance dynamics by a conditionally heteroskedastic single factor model in which the conditional variances of both common and specific factors follow GQARCH(1,1) processes to allow for leverage effects (see Sentana, 1995).

We have estimated this model under three different conditional distributional assumptions on the standardised innovations ε_t^* : Gaussian, Student t and GH . We first estimated the model by Gaussian PML, and then computed the Kuhn-Tucker normality test $KT(\tilde{\boldsymbol{\theta}}_T)$ described in section 3.1.3, which is reported in Table 1. Notice that we can easily reject normality because both the skewness and kurtosis components of the test lead to this conclusion. Next, we estimated a multivariate Student t model using the analytical formulae that FSC provide. The results show that the estimate for the tail thickness parameter η , which corresponds to slightly more than 8 degrees of freedom, is significantly larger than 0. This finding is confirmed by the LR test. Then, on the basis of the Student t ML estimates, we have computed the statistics $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$ and $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$ presented in section 3.2, with $\hat{V}[s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0})]$ estimated with the analogue of the simulation procedure described in section 2.5. The results in Table 1 show that we can reject the Student t assumption because of the value we obtain for the skewness component $\tau_{aT}(\bar{\boldsymbol{\pi}}_T)$. However, the one-sided version of the ψ component of the test is completely unable to reject the Student t specification against the alternative hypothesis of symmetric GH innovations because $\bar{s}_{\psi\psi T}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) < 0$.

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 five-dimensional vector \mathbf{b} . Since the ML estimate of ψ is 1, and $\hat{\eta}_T$ is positive, the estimated conditional distribution is effectively an asymmetric t . Therefore, it is not surprising that the com-

mon parameter estimates are very similar to those of the symmetric Student t model. In any case, and an LR 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.

The rejection of the null of normality that we find could be exacerbated by misspecification of the first and especially second conditional moments. If our specification of the model dynamics is correct, however, the marked distributional differences that we have found should not affect the consistency of the Gaussian PML estimators of θ . With this in mind, we have computed the conditional (log) standard deviations that the three estimated models generate for the portfolio of the sectorial returns that best replicates (in the usual tracking error sense) the excess returns on the aggregate NASDAQ index. Reassuringly, the correlation coefficients between the three series are extremely high (almost 1 between the two non-Gaussian models, and over .99 between each of them and the Gaussian one). Still, the results depicted in Figure 7a indicate that both the symmetric and asymmetric t distributions tend to produce somewhat less extreme values.

Another way to check that our multivariate dynamic specification is adequate is to compare the conditional variance that it implies for the NASDAQ tracking portfolio with the one we could obtain by fitting a univariate model to its excess returns. In this respect, Figure 7b compares the conditional (log) standard deviation implied by the multivariate Gaussian PML estimates with the ones obtained by fitting an AR(1)-GQARCH(1,1) model to this series. Once again, the temporal evolution is very similar, although the correlation between the two series is slightly lower (.955), which reflects the fact that the univariate model cannot differentiate common from idiosyncratic shocks.

As mentioned in the introduction, one of the main reasons for using the GH distribution is to compute the quantiles of the one-period-ahead predictive distributions of portfolio returns required in V@R calculations. To determine to what extent the GH distribution is more useful than either the normal or t distribution in this respect across all conceivable quantiles, we have computed the empirical cumulative distribution function of the probability integral transforms of the NASDAQ tracking portfolio returns generated by the three fitted distributions (see Diebold, Gunther and Tay 1998). Figure 8a shows the difference between the corresponding cumulative distributions and the 45

degree line. Under correct specification, those differences should tend to 0 asymptotically. Unfortunately, a size-corrected version of the usual Kolmogorov-type test that takes into account the sample uncertainty in the estimates of the underlying parameters is rather difficult to obtain in this case. Nevertheless, the graph clearly indicates that the multivariate asymmetric t distribution does indeed provide a better fit than its symmetric counterpart, or indeed the normal. In this respect, it is important to emphasize that our estimating criterion is multivariate, and not particularly targeted to this portfolio.

However, it could be that the encouraging results that we have found were specific of the particular portfolio considered. For this reason, we have repeated the same exercise with 5,000 different portfolio whose weights were randomly chosen from a uniform distribution on the unit sphere. Figure 8b, which compares the empirical distribution function across those 5,000 portfolios of the Kolmogorov statistics, clearly shows that the multivariate asymmetric t distribution dominates in the first-order stochastic sense the symmetric t distribution, which in turn dominates the multivariate normal. Hence, we can argue that the GH distribution provides a more adequate representation of the conditional distribution of NASDAQ portfolios than the multivariate Student t , and especially than the multivariate Gaussian distribution, but without substantially affecting the estimated conditional standard deviations.

6 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 normal-

ity 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.

We also assess the finite sample size properties of the different testing procedures that we propose by means of extensive Monte Carlo exercises. Our results indicate that the asymptotic sizes of our normality tests are very reliable in finite samples. However, we also find that the kurtosis component of the Student t test is too conservative. In addition, we show that the (simulated) information matrix systematically provides more accurate standard errors than either the OPG or numerical Hessian expressions.

Finally, we present an empirical application to NASDAQ sectorial stock returns, which suggests that their conditional distribution is asymmetric and leptokurtic. More importantly, we show that the GH distribution is more adequate than the symmetric Student t , and especially the Gaussian distribution from a risk management perspective.

A fruitful avenue for future research would be to consider bootstrap procedures (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 and Sentana (2005). 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

- Aas, K., X. Dimakos, and I. Haff (2004). Risk estimation using the multivariate normal inverse gaussian distribution. mimeo.
- 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 maintained hypothesis. *Econometrica* 69, 683–734.
- Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc.* 353, 401–419.
- Barndorff-Nielsen, O. and N. Shephard (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*, pp. 319–351. Allied Publishers.
- Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to Johansen’s bean data. *Biometrika* 68, 251–263.
- Bollerslev, T. and J. Wooldridge (1992). Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143–172.
- Bontemps, C. and N. Meddahi (2005). Testing normality: a GMM approach. *Journal of Econometrics* 124, 149–186.
- Chen, Y., W. Härdle, and S. Jeong (2004). Nonparametric risk management with Generalised Hiperbolic distributions. mimeo, CASE, Humboldt University.

- Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society, Series B* 38, 45–53.
- Dagpunar, J. (1998). *Principles of random variate generation*. New York: Oxford University Press.
- 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.
- Engle, R. and S. Kozicki (1993). Testing for common features. *Journal of Business and Economic Statistics* 11, 369–380.
- Farebrother, R. (1990). The distribution of a quadratic form in normal variables (Algorithm AS 256.3). *Applied Statistics* 39, 294–309.
- Fiorentini, G. and E. Sentana (2005). 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 (2003). Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student t innovations. *Journal of Business and Economic Statistics* 21, 532–546.
- Fiorentini, G., E. Sentana, and G. Calzolari (2004). On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models. *Economics Letters* 83, 307–312.
- Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12, 657–681.
- Hansen, B. (1991). A comparison of tests for parameter instability: an examination of asymptotic local power. Mimeo.
- 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.
- King, M., E. Sentana, and S. Wadhvani (1994). Volatility and links between national stock markets. *Econometrica* 62, 901–933.
- Koerts, J. and A. P. J. Abrahamse (1969). *On the theory and application of the general linear model*. Rotterdam: Rotterdam University Press.
- Lee, L. F. and A. Chesher (1986). Specification testing when score test statistics are identically zero. *Journal of Econometrics* 31, 121–149.
- Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. *The Journal of Finance* 56, 649–676.
- Lütkepohl, 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, J. (2003). Modeling fat tails and skewness in multivariate regression models. Unpublished Master Thesis CEMFI.
- Mencía, J. (2005). Testing dependence between financial returns. an application to the hedge fund industry. mimeo CEMFI.
- Newey, W. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047–1070.
- Newey, W. K. and D. G. Steigerwald (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–599.
- 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 Technical document.
- Sargan, J. D. (1983). Identification and lack of identification. *Econometrica* 51, 1605–1634.
- Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies* 62, 639–661.

- Sentana, E. (2004). Factor representing portfolios in large asset markets. *Journal of Econometrics* 119, 257–289.
- Silvapulle, M. and P. Silvapulle (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* 90, 342–349.
- Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.

Appendix

A Proofs of Propositions

Proposition 1

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

$$\boldsymbol{\varepsilon}^* = c(\boldsymbol{\beta}, \nu, \gamma) \boldsymbol{\beta} \left[\frac{\gamma \xi^{-1}}{R_\nu(\gamma)} - 1 \right] + \sqrt{\frac{\gamma \xi^{-1}}{R_\nu(\gamma)}} \left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}' \boldsymbol{\beta}} \boldsymbol{\beta} \boldsymbol{\beta}' \right]^{\frac{1}{2}} \mathbf{r} \quad (\text{A1})$$

Then, we can use the independence of ξ and \mathbf{r} , together with the fact that $E(\mathbf{r}) = \mathbf{0}$, and the expression for $E(\xi^{-1})$ from (D1) to show that $\boldsymbol{\varepsilon}^*$ will also have zero mean. Analogously, we will have that

$$V(\boldsymbol{\varepsilon}^*) = \frac{\gamma^2 c^2(\boldsymbol{\beta}, \nu, \gamma) E(\xi^{-2})}{R_\nu^2(\gamma)} \boldsymbol{\beta} \boldsymbol{\beta}' + \sqrt{\frac{\gamma}{R_\nu(\gamma)}} E(\xi^{-2}) \left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}' \boldsymbol{\beta}} \boldsymbol{\beta} \boldsymbol{\beta}' \right],$$

where $E(\xi^{-2})$ is also obtained from (D1). Replacing the expressions for $E(\xi^{-1})$ and $E(\xi^{-2})$, and substituting $c(\boldsymbol{\beta}, \nu, \gamma)$ by (3), we can finally show that $V(\boldsymbol{\varepsilon}^*) = \mathbf{I}_N$. \square

Proposition 2

Using (A1), we can write s^* as

$$s^* = c(\boldsymbol{\beta}, \nu, \gamma) \frac{\mathbf{w}' \boldsymbol{\beta}}{\sqrt{\mathbf{w}' \mathbf{w}}} \left[\frac{\gamma \xi^{-1}}{R_\nu(\gamma)} - 1 \right] + \sqrt{\frac{\gamma \xi^{-1}}{R_\nu(\gamma)}} \frac{\mathbf{w}'}{\sqrt{\mathbf{w}' \mathbf{w}}} \left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}' \boldsymbol{\beta}} \boldsymbol{\beta} \boldsymbol{\beta}' \right]^{\frac{1}{2}} \mathbf{r}_t.$$

But since the third term in this expression is the product of the square root of a *GIG* variable times a univariate normal variate, r_t say, we can also rewrite s^* as

$$s^* = c(\boldsymbol{\beta}, \nu, \gamma) \frac{\mathbf{w}' \boldsymbol{\beta}}{\sqrt{\mathbf{w}' \mathbf{w}}} \left[\frac{\gamma \xi^{-1}}{R_\nu(\gamma)} - 1 \right] + \sqrt{\frac{\gamma \xi^{-1}}{R_\nu(\gamma)}} \sqrt{1 + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}' \boldsymbol{\beta}} \frac{(\mathbf{w}' \boldsymbol{\beta})^2}{\mathbf{w}' \mathbf{w}}} r_t \quad (\text{A2})$$

Given that s^* is a standardised variable by construction, if we compare (A2) with the general formula for standardised *GH* variables in (A1), then we will conclude that the parameters η and ψ are the same as in the multivariate distribution, while the skewness parameter is now a function of the vector \mathbf{w} . Finally, the exact formula for $\beta(\mathbf{w})$ can be easily obtained from the relationships

$$\begin{aligned} c[\beta(\mathbf{w}), \nu, \gamma] \beta(\mathbf{w}) &= c(\boldsymbol{\beta}, \nu, \gamma) \frac{\mathbf{w}' \boldsymbol{\beta}}{\sqrt{\mathbf{w}' \mathbf{w}}}, \\ c[\beta(\mathbf{w}), \nu, \gamma] &= 1 + \frac{c(\boldsymbol{\beta}, \nu, \gamma) - 1}{\boldsymbol{\beta}' \boldsymbol{\beta}} \frac{(\mathbf{w}' \boldsymbol{\beta})^2}{\mathbf{w}' \mathbf{w}}, \end{aligned}$$

where $c[\beta(\mathbf{w}), \nu, \gamma] = [-1 + \sqrt{1 + 4(D_{\nu+1}(\gamma) - 1)\beta^2(\mathbf{w})}] / \{2[D_{\nu+1}(\gamma) - 1]\beta^2(\mathbf{w})\}$. \square

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 $\psi \rightarrow 0$, 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 $\boldsymbol{\varepsilon}_t^*$ is distributed as a multivariate standard normal, then it can be written as $\boldsymbol{\varepsilon}_t^* = \sqrt{\zeta_t} \mathbf{u}_t$, where \mathbf{u}_t is uniformly distributed on the unit sphere surface in \mathbb{R}^N , ζ_t is a chi-square random variable with N degrees of freedom, and \mathbf{u}_t and ζ_t are mutually independent. \square

Proposition 4

The proof is straightforward if we rely on the results in the appendix of Fiorentini and Sentana (2005), who indicate that when $\boldsymbol{\varepsilon}_t^*$ is distributed as a standardised multivariate Student t with $1/\eta_0$ degrees of freedom, it can be written as $\boldsymbol{\varepsilon}_t^* = \sqrt{(1 - 2\eta_0)\zeta_t/(\xi_t\eta_0)} \mathbf{u}_t$, where \mathbf{u}_t is uniformly distributed on the unit sphere surface in \mathbb{R}^N , ζ_t is a chi-square random variable with N degrees of freedom, ξ_t is a gamma variate with mean η_0^{-1} and variance $2\eta_0^{-1}$, and the three variates are mutually independent. These authors also exploit the fact that $X = \zeta_t/(\zeta_t + \xi_t)$ has a beta distribution with parameters $a = N/2$ and $b = 1/(2\eta_0)$ to show that

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

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

Proposition 5

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

Proposition 6

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

$$LM(\tilde{\boldsymbol{\theta}}_T, \mathbf{b}) = \frac{T\mathbf{b}^+\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)\bar{m}_T(\tilde{\boldsymbol{\theta}}_T)\mathbf{b}^+}{(N+2)\mathbf{b}^+\mathbf{D}_T\mathbf{b}^+}, \quad (\text{A3})$$

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

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

But since the maximisation of (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 $N/[2\bar{m}_{kT}(\tilde{\boldsymbol{\theta}}_T)]$ as the constant of proportionality, then we can make sure that the first element in \mathbf{b}^+ is equal to one. Substituting this value in the formula of $LM(\tilde{\boldsymbol{\theta}}_T, \mathbf{b})$ yields the required result. Finally, the asymptotic distribution of the sup test follows directly from the fact that $\sqrt{T}\bar{m}_{kT}(\boldsymbol{\theta}_0)$ and $\sqrt{T}\bar{m}_{sT}(\boldsymbol{\theta}_0)$ are asymptotically orthogonal under the null, with asymptotic variances $N(N+2)/2$ and $2(N+2)\boldsymbol{\Sigma}$, respectively. \square

Propositions 7 and 8

We can use again the results of Fiorentini and Sentana (2005) mentioned in the proof of Proposition 4, together with the results in Crowder (1976), to show that

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

where

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

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

$$\begin{aligned} g_{2t}(\bar{\boldsymbol{\pi}}_T) &= \begin{bmatrix} s_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{bmatrix} - \begin{bmatrix} \mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ \mathcal{M}'(\boldsymbol{\pi}_0) \end{bmatrix} \mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0}) s_{\pi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ &= \begin{bmatrix} -\mathcal{I}'_{\pi\mathbf{b}}(\boldsymbol{\pi}_0, 1, \mathbf{0})\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{I}_N & \mathbf{0} \\ -\mathcal{M}'(\boldsymbol{\pi}_0)\mathcal{I}_{\pi\pi}^{-1}(\boldsymbol{\pi}_0, 1, \mathbf{0}) & \mathbf{0}' & 1 \end{bmatrix} \begin{bmatrix} s_{\pi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{bmatrix} = \mathcal{A}_2(\boldsymbol{\pi}_0) \begin{bmatrix} s_{\pi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{bmatrix}. \end{aligned}$$

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

$$\begin{aligned} \mathbf{0} &= \frac{\sqrt{T}}{T} \sum_t g_{2t}(\bar{\boldsymbol{\pi}}_T) = \frac{\sqrt{T}}{T} \sum_t \begin{bmatrix} s_{\mathbf{b}t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \\ s_{\psi\psi t}(\bar{\boldsymbol{\pi}}_T, 1, \mathbf{0}) \end{bmatrix} = \mathcal{A}_2(\boldsymbol{\pi}_0) \frac{\sqrt{T}}{T} \sum_t \begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{bmatrix} \\ &\quad + \mathcal{A}_2(\boldsymbol{\pi}_0) E \left[\frac{\partial}{\partial \boldsymbol{\pi}'} \begin{pmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\mathbf{b}t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \\ s_{\psi\psi t}(\boldsymbol{\pi}_0, 1, \mathbf{0}) \end{pmatrix} \right] \sqrt{T} (\bar{\boldsymbol{\pi}}_T - \boldsymbol{\pi}_0) + o_p(1), \end{aligned}$$

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

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

As a result

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

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

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 $l(\mathbf{y}_t | \xi_t, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi})$ and $l(\xi_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$. Then, in the expectation step, we take the expected value of these derivatives given $I_T = \{(\mathbf{z}_1, \mathbf{y}_1), (\mathbf{z}_2, \mathbf{y}_2), \dots, (\mathbf{z}_T, \mathbf{y}_T)\}$ and the parameter values.

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

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

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

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

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

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

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

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

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 $z_t, I_{t-1}; \phi$ to show that

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

which implies that

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

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

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

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

$$\begin{aligned} \frac{\partial l(\mathbf{y}_t|Y_{t-1}; \phi)}{\partial \theta'} &= -\frac{1}{2} \text{vec}'[\Sigma_t^{-1}(\theta)] \frac{\partial \text{vec}[\Sigma_t(\theta)]}{\partial \theta'} - f(I_T, \phi) \mathbf{p}_t' \Sigma_t^{*-1}(\phi) \frac{\partial \mathbf{p}_t}{\partial \theta'} \\ &- \frac{1}{2} \frac{c_t(\phi) - 1}{c_t(\phi) \mathbf{b}' \Sigma_t(\theta) \mathbf{b} \sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \Sigma_t(\theta) \mathbf{b}}} \text{vec}'(\mathbf{b}\mathbf{b}') \frac{\partial \text{vec}[\Sigma_t(\theta)]}{\partial \theta'} + \mathbf{b}' \frac{\partial \mathbf{p}_t}{\partial \theta'} \\ &\quad + \frac{1}{2} f(I_T, \phi) [\mathbf{p}_t' \Sigma_t^{*-1}(\phi) \otimes \mathbf{p}_t' \Sigma_t^{*-1}(\phi)] \frac{\partial \text{vec}[\Sigma_t^*(\phi)]}{\partial \theta'} \\ &- \frac{1}{2} \frac{g(I_T, \phi)}{\sqrt{1 + 4(D_{\nu+1}(\gamma) - 1) \mathbf{b}' \Sigma_t(\theta) \mathbf{b}}} \text{vec}'(\mathbf{b}\mathbf{b}') \frac{\partial \text{vec}[\Sigma_t(\theta)]}{\partial \theta'}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\mathbf{y}_t|Y_{t-1}; \boldsymbol{\phi})}{\partial \mathbf{b}} &= -\frac{c_t(\boldsymbol{\phi}) - 1}{c_t(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\sqrt{1+4(D_{\nu+1}(\gamma) - 1)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \\
&\quad - f(I_T, \boldsymbol{\phi})c_t(\boldsymbol{\phi})\mathbf{p}'_t + \boldsymbol{\varepsilon}'_t + f(I_T, \boldsymbol{\phi})\frac{c_t(\boldsymbol{\phi}) - 1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}(\mathbf{b}'\mathbf{p}_t) \\
&\quad \times \left\{ \frac{[c_t(\boldsymbol{\phi}) - 1](\mathbf{b}'\mathbf{p}_t)}{c_t^2(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\sqrt{1+4(D_{\nu+1}(\gamma) - 1)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \right. \\
&\quad \left. + \frac{\mathbf{p}'_t}{c_t(\boldsymbol{\phi})} - \frac{1}{\sqrt{1+4(D_{\nu+1}(\gamma) - 1)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \right\} \\
&\quad + \frac{[2 - g(I_T, \boldsymbol{\phi})]}{\sqrt{1+4(D_{\nu+1}(\gamma) - 1)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\mathbf{y}_t|Y_{t-1}; \boldsymbol{\phi})}{\partial \eta} &= \frac{N}{2}\frac{\partial \log R_\nu(\gamma)}{\partial \eta} + \left(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{1}{2c_t(\boldsymbol{\phi})} \right) \frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} + \frac{\log(\gamma)}{2\eta^2} \\
&\quad - \frac{\partial \log K_\nu(\gamma)}{\partial \eta} - \frac{1}{2\eta^2}E[\log \xi_t|Y_T; \boldsymbol{\phi}] - \frac{f(I_T, \boldsymbol{\phi})}{2} \left\{ \frac{\partial \log R_\nu(\gamma)}{\partial \eta} \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi}) \mathbf{p}_t \right. \\
&\quad \left. + \frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} \left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{(\mathbf{b}'\boldsymbol{\varepsilon}_t)^2}{c_t^2(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}} \right] \right\} \\
&\quad - \frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{2}g(I_T, \boldsymbol{\phi}) \left\{ \frac{\partial c_t(\boldsymbol{\phi})}{\partial \eta} - c_t(\boldsymbol{\phi})\frac{\partial \log R_\nu(\gamma)}{\partial \eta} \right\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial l(\mathbf{y}_t|Y_{t-1}; \boldsymbol{\phi})}{\partial \psi} &= \frac{N}{2}\frac{\partial \log R_\nu(\gamma)}{\partial \psi} + \frac{N}{2\psi(1-\psi)} + \left(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{1}{2c_t(\boldsymbol{\phi})} \right) \frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi} \\
&\quad + \frac{1}{2\eta\psi(1-\psi)} - \frac{\partial \log K_\nu(\gamma)}{\partial \psi} - \frac{f(I_T, \boldsymbol{\phi})}{2} \left\{ \left[\frac{\partial \log R_\nu(\gamma)}{\partial \psi} + \frac{1}{\psi(1-\psi)} \right] \mathbf{p}'_t \boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi}) \mathbf{p}_t \right. \\
&\quad \left. + \frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi} \left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} - \frac{(\mathbf{b}'\boldsymbol{\varepsilon}_t)^2}{c_t^2(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}} \right] \right\} \\
&\quad - \frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{2}g(I_T, \boldsymbol{\phi}) \left\{ -\frac{c_t(\boldsymbol{\phi})}{\psi(1-\psi)} + \frac{\partial c_t(\boldsymbol{\phi})}{\partial \psi} - c_t(\boldsymbol{\phi})\frac{\partial \log R_\nu(\gamma)}{\partial \psi} \right\} + g(I_T, \boldsymbol{\phi})\frac{R_\nu(\gamma)}{\psi^2},
\end{aligned}$$

where

$$\begin{aligned}
f(I_T, \boldsymbol{\phi}) &= \gamma^{-1}R_\nu(\gamma)E(\xi_t|I_T; \boldsymbol{\phi}), \\
g(I_T, \boldsymbol{\phi}) &= \gamma R_\nu^{-1}(\gamma)E(\xi_t^{-1}|I_T; \boldsymbol{\phi}),
\end{aligned}$$

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

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

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

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

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

and

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

C Modified Bessel function of the third kind

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

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\pi\nu)}. \quad (\text{C1})$$

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

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

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

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

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

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

where $\nu > 0$ and $l = [1 + (x/\nu)^2]^{-\frac{1}{2}}$. 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 $n = 0, 1, 2, \dots$ due to the sine that appears in the denominator. In our experience, it is much better to use the following Taylor expansion for small $|\nu - n|$:

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

where for *integer* ν :

$$\begin{aligned} \frac{\partial K_\nu(x)}{\partial \nu} &= \frac{1}{4 \cos(\pi n)} \left[\frac{\partial^2 I_{-\nu}(x)}{\partial \nu^2} - \frac{\partial^2 I_\nu(x)}{\partial \nu^2} \right] + \pi^2 [I_{-\nu}(x) - I_\nu(x)], \\ \frac{\partial^2 K_\nu(x)}{\partial \nu^2} &= \frac{1}{6 \cos(\pi n)} \left[\frac{\partial^3 I_{-\nu}(x)}{\partial \nu^3} - \frac{\partial^3 I_\nu(x)}{\partial \nu^3} \right] + \frac{\pi^2}{3 \cos(\pi n)} \left[\frac{\partial I_{-\nu}(x)}{\partial \nu} - \frac{\partial I_\nu(x)}{\partial \nu} \right] - \frac{\pi^2}{3} K_n(x), \end{aligned}$$

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

and

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

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

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

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

where

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

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

where

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

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

where

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

and finally,

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

where

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

with

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

and

$$\begin{aligned} f_b(z) &= -5\psi^4(1-z) + 10\pi^2\psi^2(1-z) - 30\psi^2(1-z)\psi'(1-z) \\ &\quad - 20\psi(1-z)\psi''(1-z) - 15[\psi'(1-z)]^2 + 10\pi^2\psi'(1-z) - 5\psi'''(1-z) - \pi^4. \end{aligned}$$

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 $\delta, \gamma \geq 0$, $\nu \in \mathbb{R}$, $x > 0$. Two important properties of this distribution are $X^{-1} \sim GIG(-\nu, \gamma, \delta)$ and $(\gamma/\delta)X \sim GIG(\nu, \sqrt{\gamma\delta}, \sqrt{\gamma\delta})$. For our purposes, the most useful moments of X when $\delta\gamma > 0$ are

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

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

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 $k \geq |\nu|$. 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

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

and

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

where

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 1
Maximum likelihood estimates of a conditionally heteroskedastic factor model for 5
NASDAQ sectorial indices

Parameter	Gaussian		Student t		Asymmetric t	
	SE		SE		SE	
η	0	-	0.1239	0.005	0.131	0.005
ψ	0	-	0	-		
b_1	0	-	0	-	-0.029	0.020
b_2	0	-	0	-	0.070	0.027
b_3	0	-	0	-	-0.077	0.014
b_4	0	-	0	-	-0.018	0.017
b_5	0	-	0	-	0.041	0.013
Log-likelihood	-64581.7382		-63566.4432		-63555.1033	

Normality tests

Test		p-value
Kurtosis	44809.590	0.000
Skewness	314.775	0.000
Sup-LM	45124.370	0.000
Kuhn-Tucker	45124.370	0.000

Student t tests

Test		p-value
$s_{\psi\psi}$	-7.512	1.000
Skewness	13.072	0.023
Kuhn-Tucker	13.070	0.032

Note: $s_{\psi\psi}$ denotes the “kurtosis” test based on $\tau_{kT}(\bar{\boldsymbol{\pi}}_T)$.

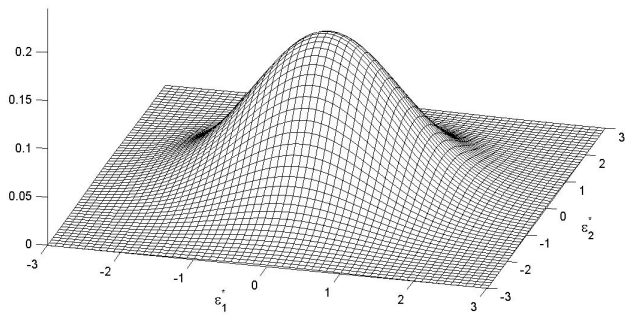


Figure 1a: Standardised bivariate normal density

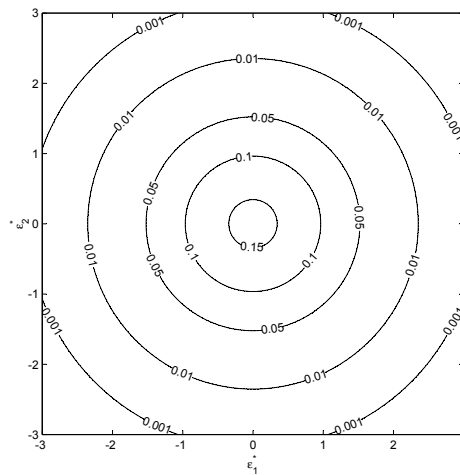


Figure 1b: Contours of a standardised bivariate normal density

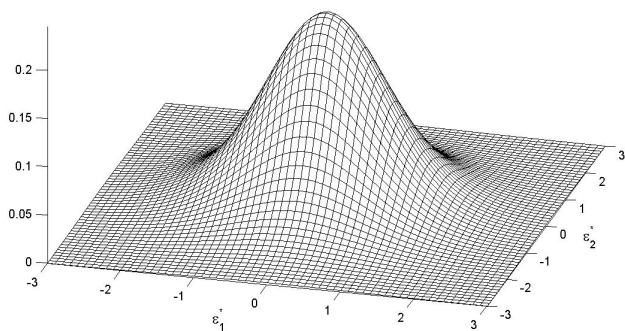


Figure 1c: Standardised bivariate Student t density with 10 degrees of freedom ($\eta = .1$)

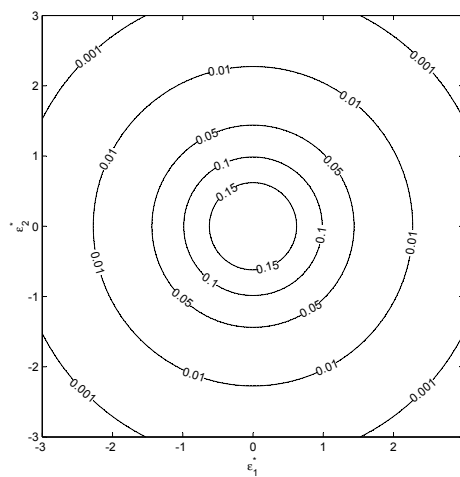


Figure 1d: Contours of a standardised bivariate Student t density with 10 degrees of freedom ($\eta = .1$)

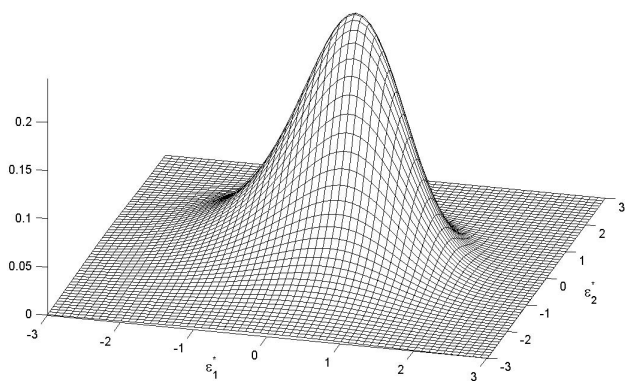


Figure 1e: Standardised bivariate asymmetric Student t density with 10 degrees of freedom ($\eta = .1$) and $\beta = (-3, -3)'$

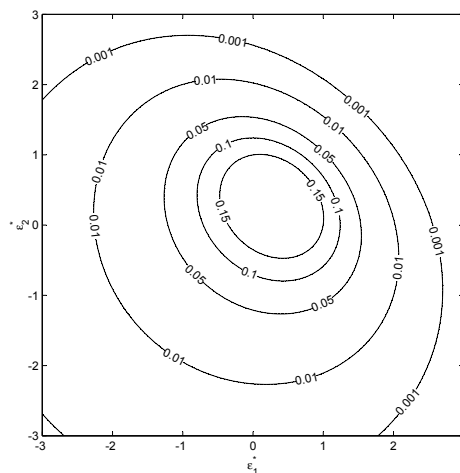


Figure 1f: Contours of a standardised bivariate asymmetric Student t density with 10 degrees of freedom ($\eta = .1$) and $\beta = (-3, -3)'$

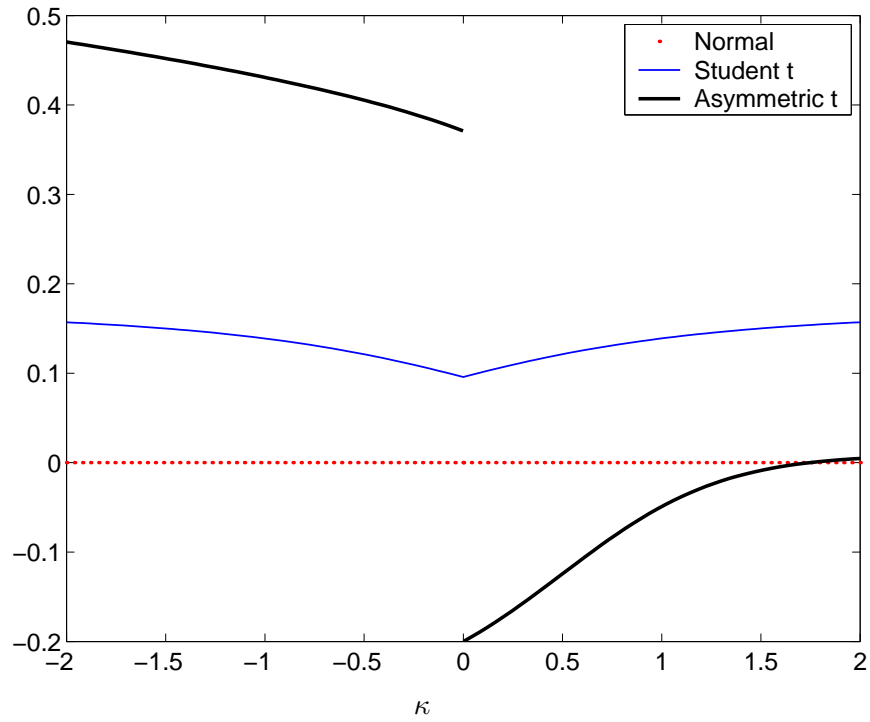


Figure 2: Exceedance correlation for a standardised bivariate normal distribution, Student t distribution with 10 degrees of freedom ($\eta = .1$) and Asymmetric t distribution with $\eta = .1$ and $\beta = (-3, -3)$

Notes: The exceedance correlation between two variables ε_1^* and ε_2^* is defined as $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \kappa, \varepsilon_2^* > \kappa)$ for positive κ and $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* < \kappa, \varepsilon_2^* < \kappa)$ for negative κ (see Longin and Solnik, 2001)

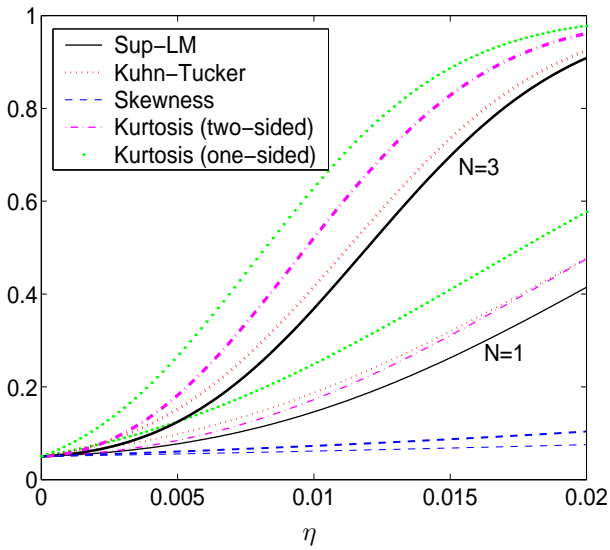


Figure 3a: Power of the normality tests under symmetric t alternatives

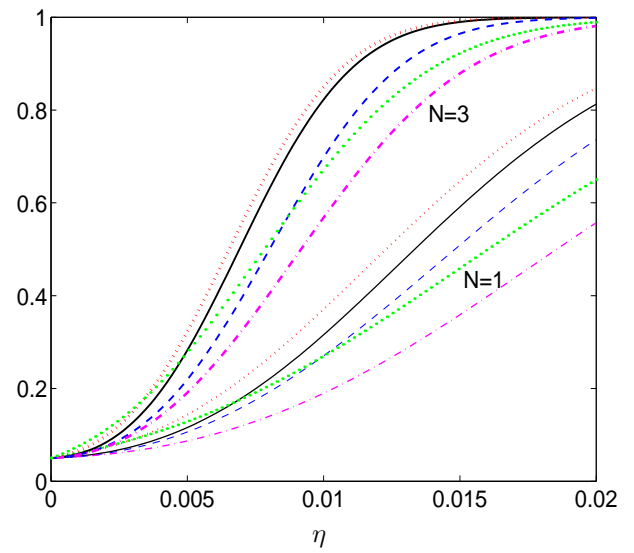


Figure 3b: Power of the normality tests under asymmetric t alternatives ($b_i = .75, \forall i$)

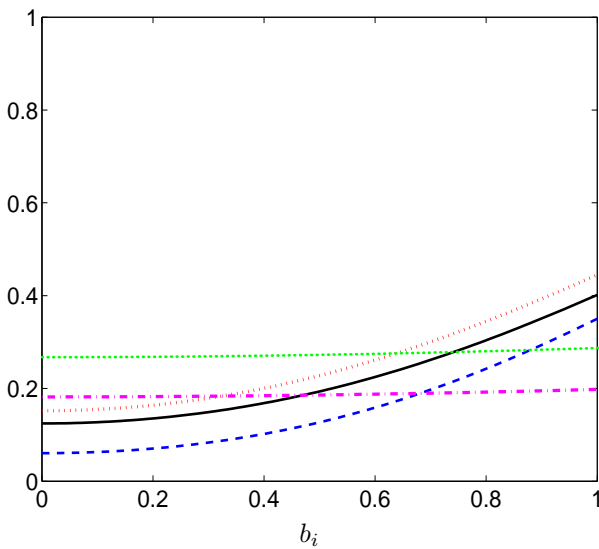


Figure 3c: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .005, N = 3$)

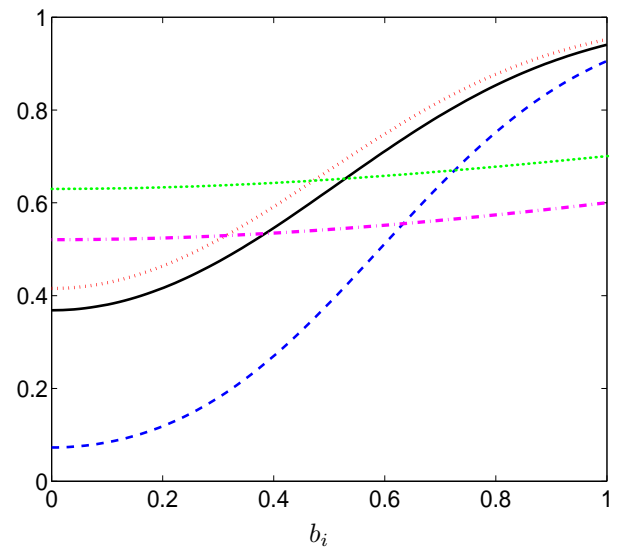


Figure 3d: Power of the multivariate normality tests against asymmetric t alternatives with increasing skewness ($\eta = .01, N = 3$)

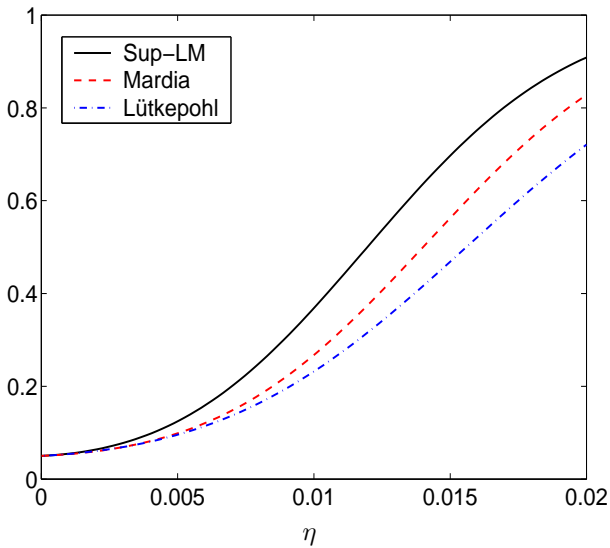


Figure 3e: Power of Sup-LM, Mardia and Lütkepohl normality tests against symmetric t alternatives ($N = 3$).

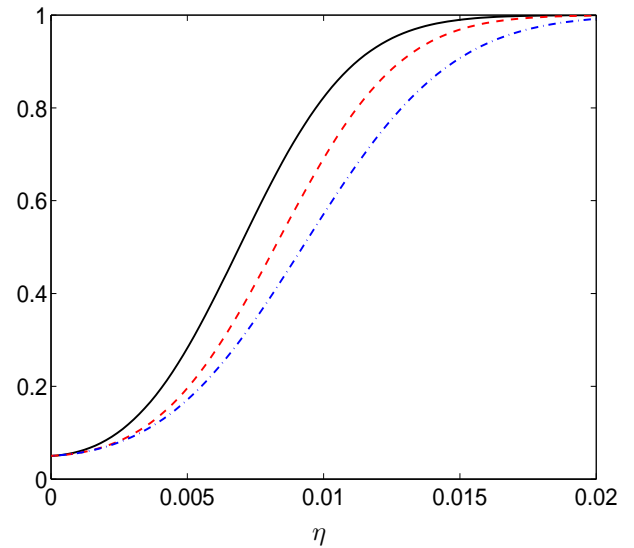


Figure 3f: Power of Sup-LM, Mardia and Lütkepohl normality tests against asymmetric t alternatives ($N = 3$).

Notes: Thicker lines represent the power of the trivariate tests. Figures 3b-3d share the legend of figure 3a, while figure 3f shares the legend of figure 3e.

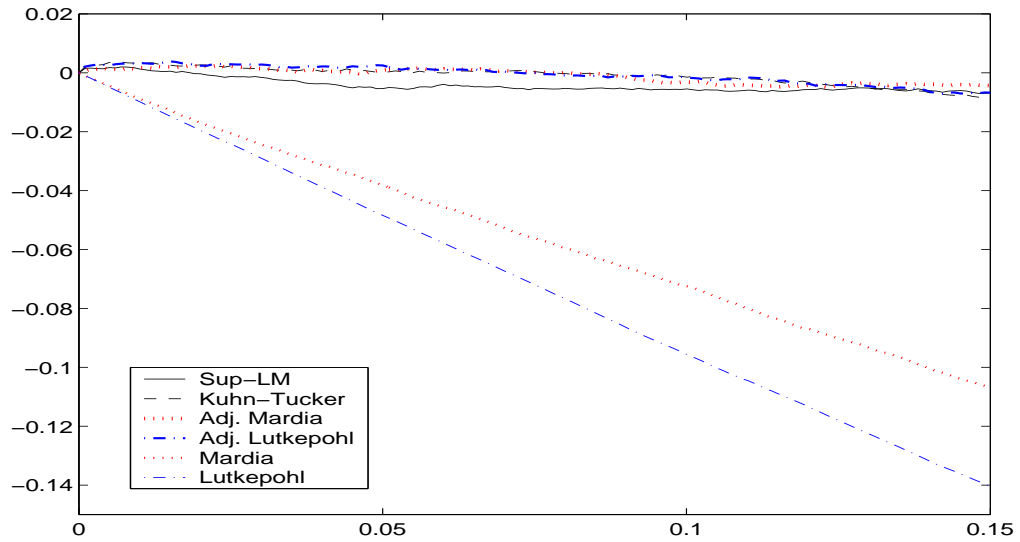


Figure 4a: p-value discrepancy plots of the joint normality tests

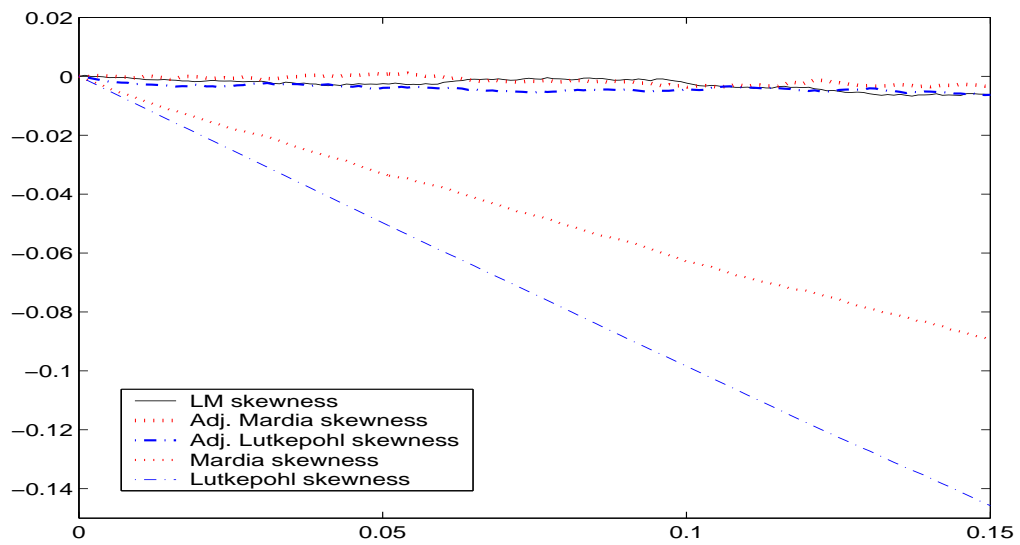


Figure 4b: p-value discrepancy plots of the skewness components of the joint normality tests

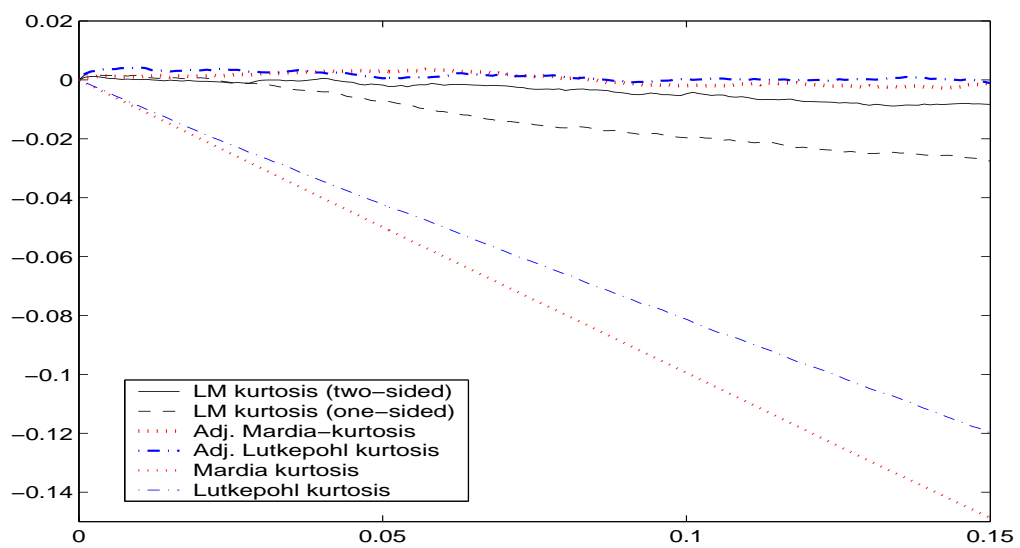


Figure 4c: p-value discrepancy plots of the kurtosis components of the joint normality tests

Notes: p-value discrepancy plots obtained from a Monte Carlo study with 15,000 simulations of samples with $T = 1,000$. Size distortions in Mardia and Lütkepohl tests have been corrected in their adjusted versions and plotted with thicker lines.

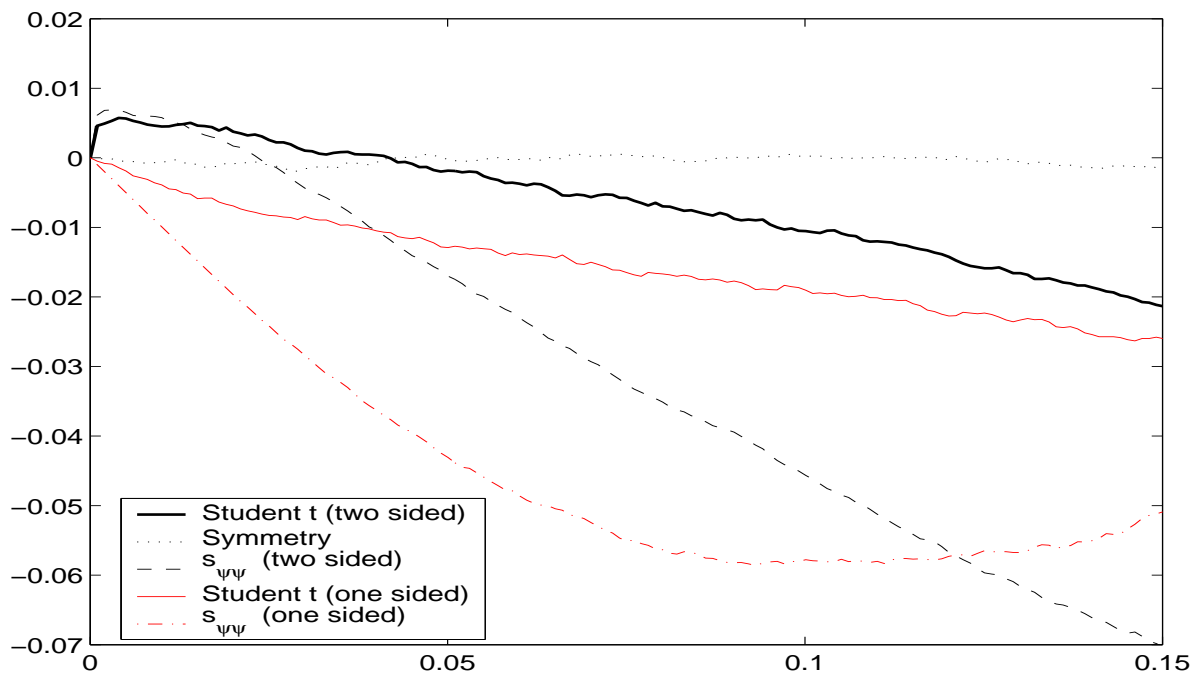


Figure 5a: p-value discrepancy plots of the Student t tests

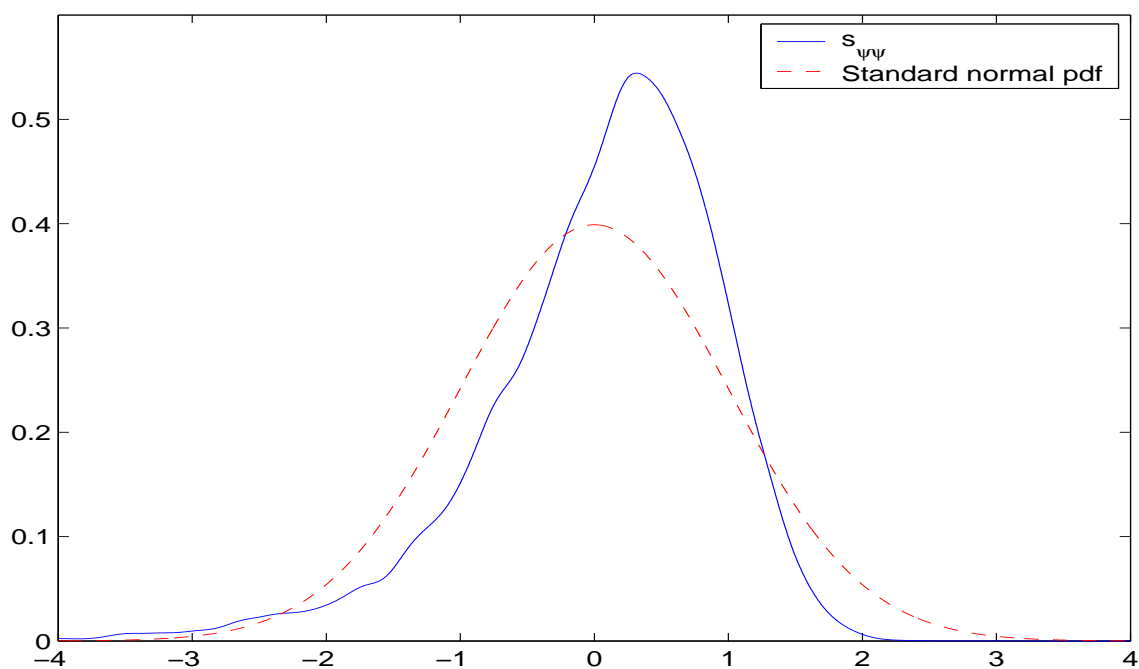


Figure 5b: kernel estimation of the density of the symmetric Student t test

Notes: p-value discrepancy plots obtained from a Monte Carlo study with 15,000 simulations with $T = 1,000$. $s_{\psi\psi}$ denotes the “kurtosis” test based on $\tau_{kT}(\bar{\pi}_T)$.

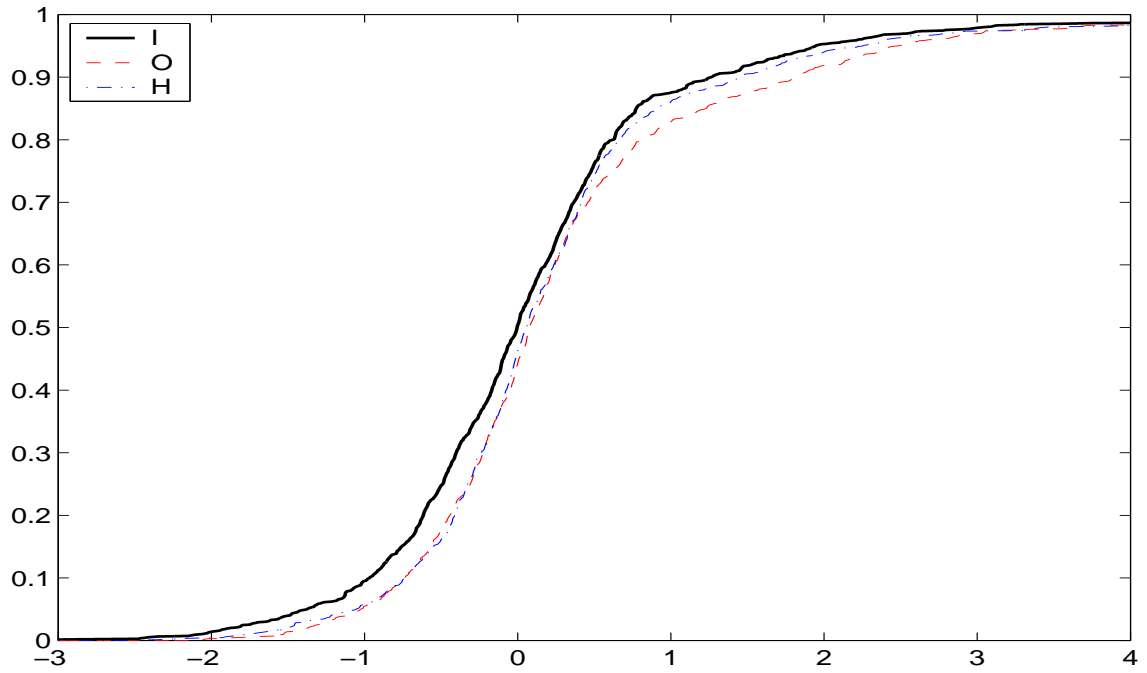


Figure 6a: Sampling distribution of $vech'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vech[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$

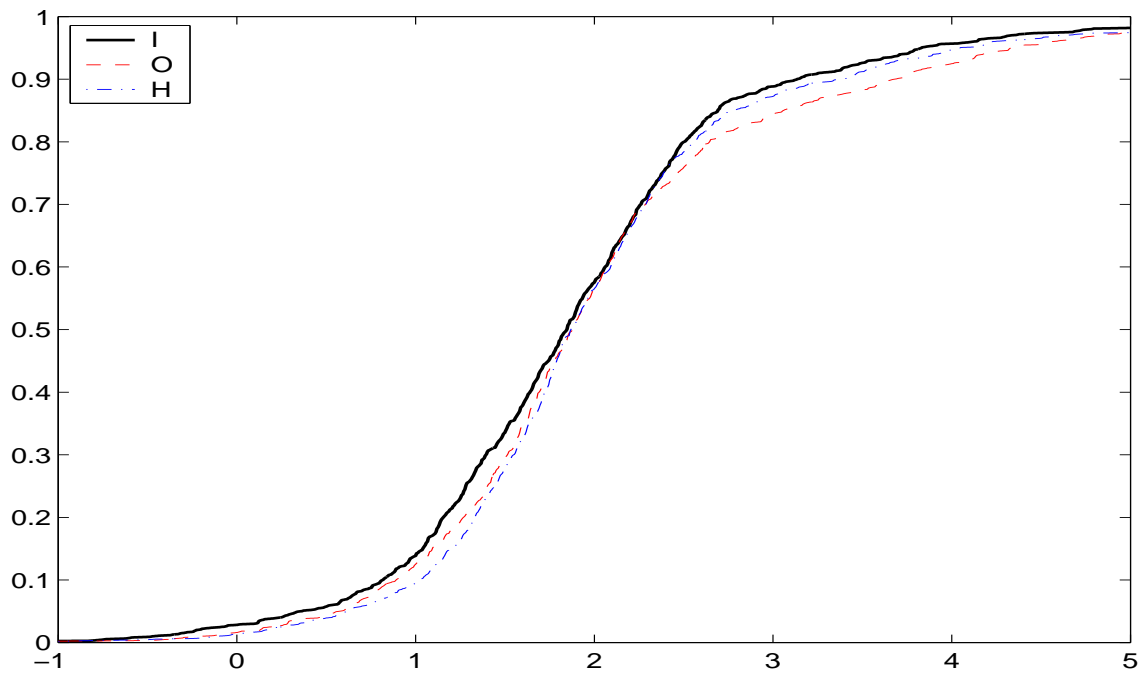


Figure 6b: Sampling distribution of $vecd'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vecd[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$

Notes: Obtained from a Monte Carlo study with 1,000 replications of sample size $T = 1,000$, except $V_T(\hat{\phi}_T)$, which is the sampling variance of the ML estimators in 30,000 samples of the same size. E refers to the standard errors obtained by either the outer-product of the gradient (O), numerical Hessian (H), or the simulated unconditional information matrix.

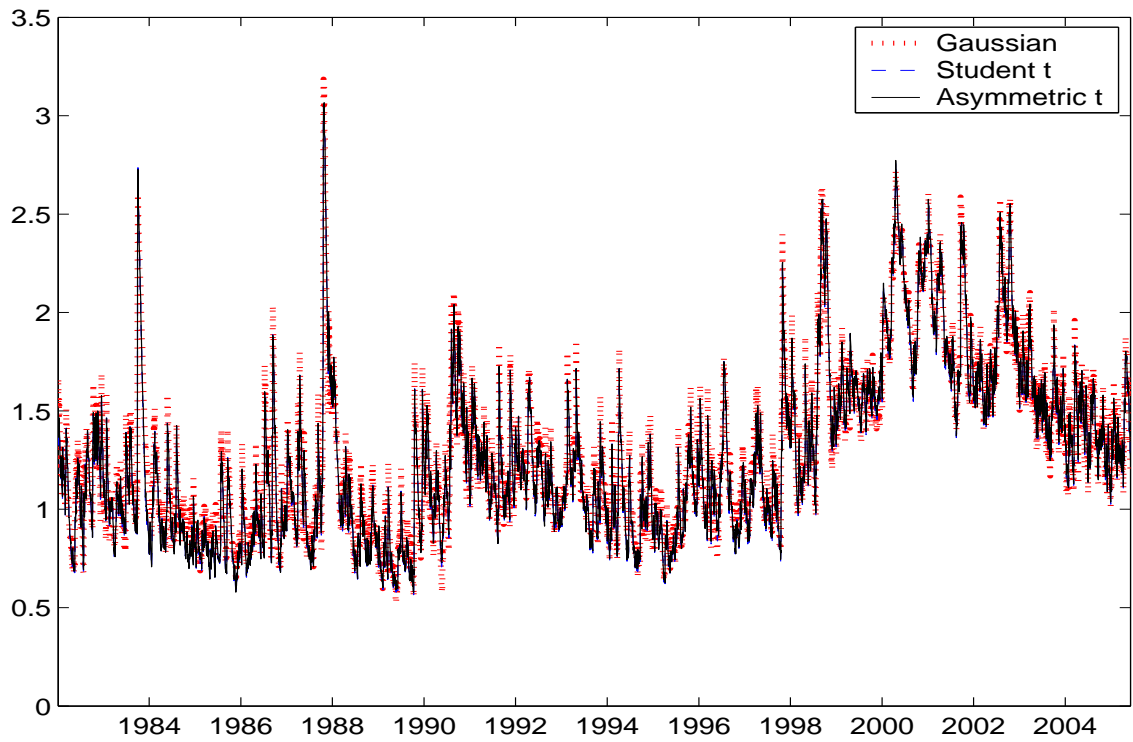


Figure 7a: Comparison of the multivariate estimates of the (log)standard deviation of the NASDAQ tracking portfolio

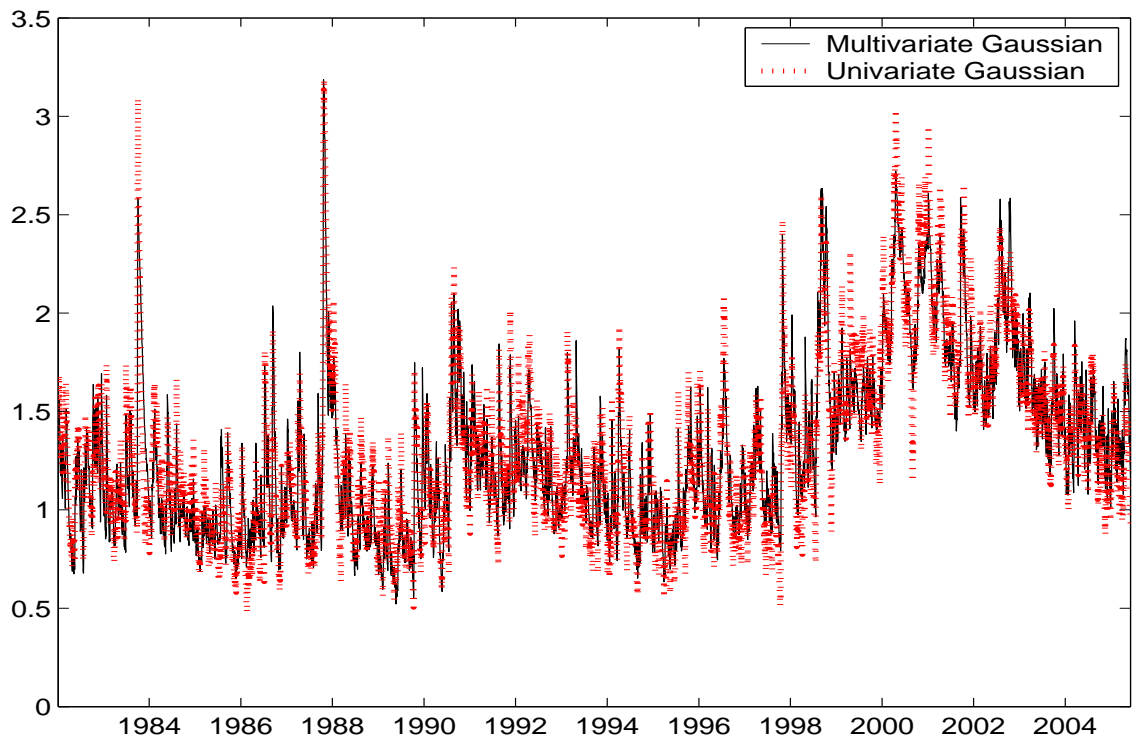


Figure 7b: Comparison of univariate and multivariate Gaussian estimates of the (log)standard deviation of the NASDAQ tracking portfolio

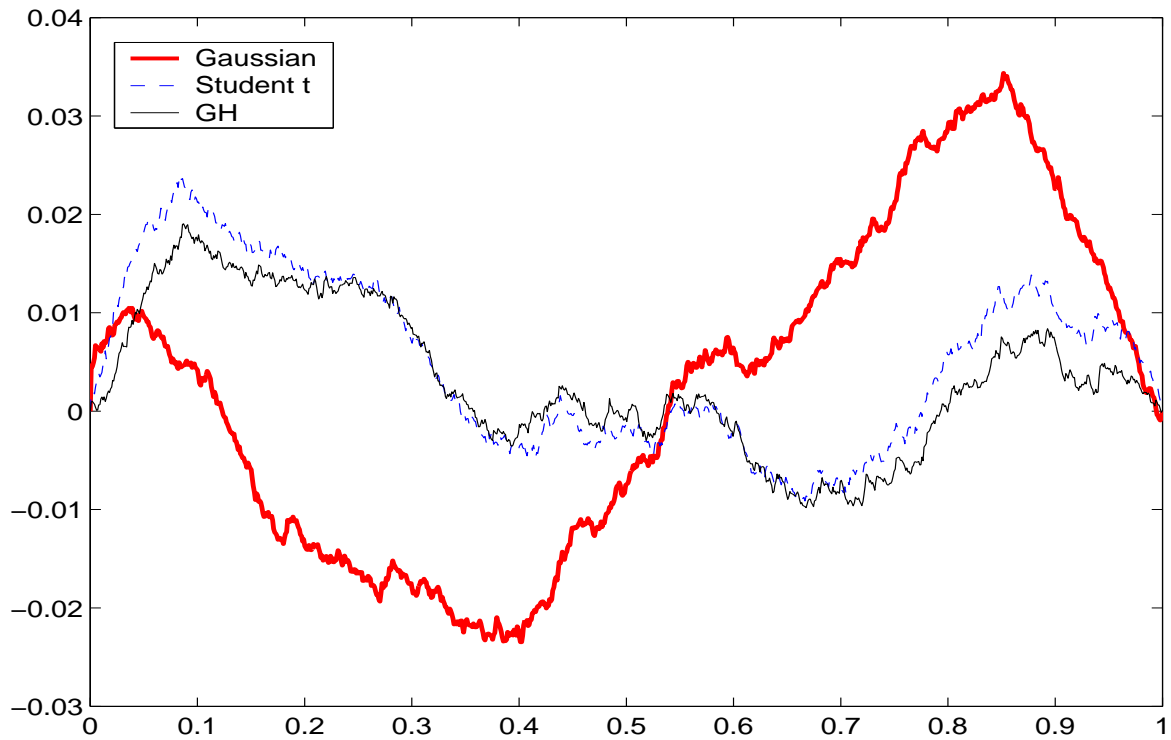


Figure 8a: P-value discrepancy plots of the empirical cumulative distribution function of probability integral transform of returns on the NASDAQ tracking portfolio

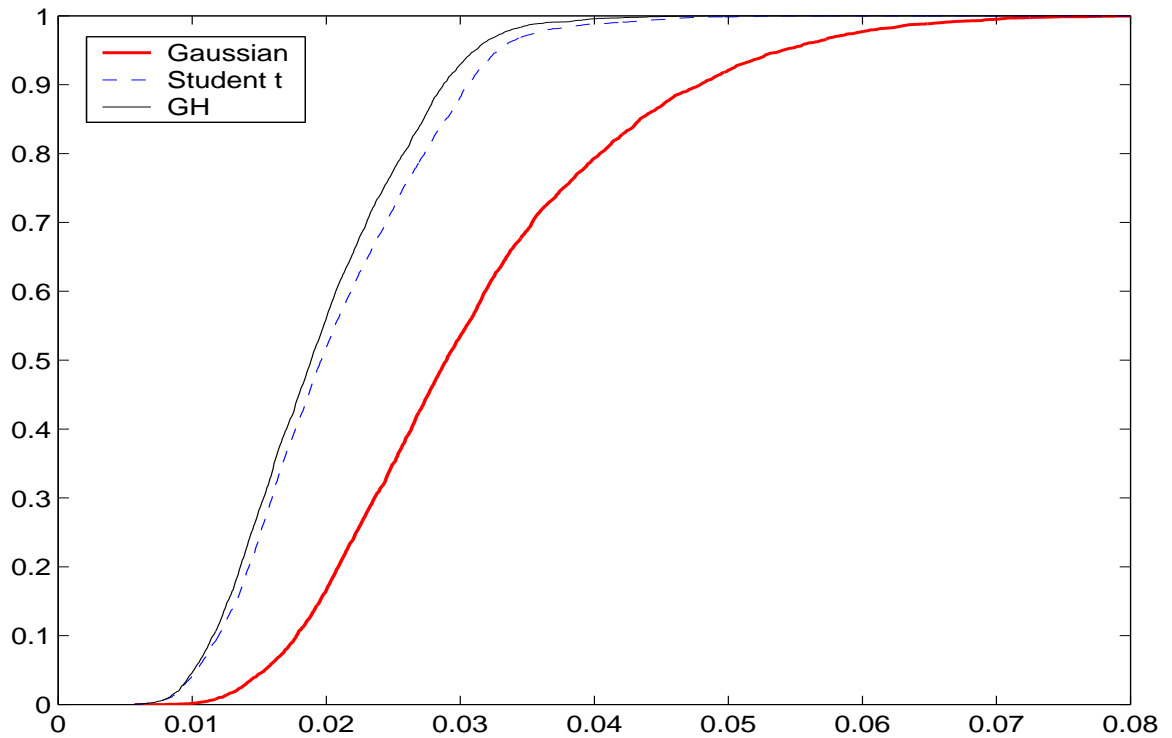


Figure 8b: Empirical cumulative distribution function of Kolmogorov tests for 5,000 random portfolios

Notes: The weights of the random portfolios are sampled from a uniform distribution on the unit sphere