# Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation*

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

**Enrique Sentana**
*CEMFI*
<sentana@cemfi.es>

January 2008

**Abstract**

We show that the distribution of any portfolio whose components jointly follow a location-scale mixture of normals can be characterised solely by its mean, variance and skewness. Under this distributional assumption, we derive the mean-variance-skewness frontier in closed form, and show that it can be spanned by three funds. For practical purposes, we derive a standardised distribution, provide analytical expressions for the log-likelihood score and explain how to evaluate the information matrix. Finally, we present an empirical application in which we obtain the mean-variance-skewness frontier generated by the ten Datastream US sectoral indices, and conduct spanning tests.

**Keywords:** Generalised Hyperbolic Distribution, Maximum Likelihood, Portfolio Frontiers, Spanning Tests, Tail Dependence.

**JEL**: C52, C32, G11

---

*We are grateful to Francisco Peñaranda for helpful comments and suggestions. Of course, the usual caveat applies. Address for correspondence: Casado del Alisal 5, E-28014 Madrid, Spain, tel: +34 91 429 05 51, fax: +34 91 429 1056.

# 1 Introduction

Despite its simplicity, mean-variance analysis remains the most widely used asset allocation method. There are several reasons for its popularity. First, it provides a very intuitive assessment of the relative merits of alternative portfolios, as their risk and expected return characteristics can be compared in a two-dimensional graph. Second, mean-variance frontiers are spanned by only two funds, which simplifies their calculation and interpretation. Finally, mean-variance analysis becomes the natural approach if we assume Gaussian or elliptical distributions, because then it is fully compatible with expected utility maximisation regardless of investor preferences (see e.g. Chamberlain, 1983; Owen and Rabinovitch, 1983; and Berk, 1997).

At the same time, mean-variance analysis also suffers from important limitations. Specifically, it neglects the effect of higher order moments on asset allocation. In this sense, Patton (2004) uses a bivariate copula model to show the empirical importance of asymmetries in asset allocation. Further empirical evidence has been provided by Jondeau and Rockinger (2006) and Harvey et al. (2002). From the theoretical point of view, Athayde and Flôres (2004) derive several useful properties of mean-variance-skewness frontiers, and obtain their shape for some examples by simulation techniques.

In this paper, we make mean-variance-skewness analysis fully operational by working with a rather flexible family of multivariate asymmetric distributions, known as location-scale mixtures of normals ($LSMN$), which nest as particular cases several important elliptically symmetric distributions, such as the Gaussian or the Student $t$, and also some well known asymmetric distributions like the Generalised Hyperbolic ($GH$) introduced by Barndorff-Nielsen (1977). The $GH$ distribution in turn nests many other well known distributions, such as symmetric and asymmetric versions of the Hyperbolic, Normal Gamma, Normal Inverse Gaussian or Multivariate Laplace (see Appendix C), whose empirical relevance has already been widely documented in the literature (see e.g. Madan and Milne, 1991; Chen, Härdle, and Jeong, 2004; Aas, Dimakos, and Haff, 2005; and Cajigas and Urga, 2007). In addition, $LSMN$ nest other interesting examples, such as finite mixtures of normals, which have been shown to be a flexible and empirically plausible device to introduce non-Gaussian features in high dimensional multivariate distributions (see e.g. Kon, 1984), but which at the same time remain analytically tractable.

In terms of portfolio allocation, our first result is that if the distribution of asset returns can be expressed as a $LSMN$, then the distribution of any portfolio that combines those assets will be uniquely characterised by its mean, variance and skewness. Therefore, under rather mild assumptions on investors' preferences, optimal portfolios will be located on the mean-variance-skewness frontier, which we are able to obtain in closed form. Furthermore, we will show that the efficient part of this frontier can be spanned by three funds: the two funds that generate the usual mean-variance frontier, plus an additional fund that spans the skewness-variance frontier.

For practical purposes, we study several aspects related to the maximum likelihood estimation of a general conditionally heteroskedastic dynamic regression model whose innovations have a *LSMN* representation. In particular, we obtain analytical expressions for the score by means of the EM algorithm. We also describe how to evaluate the unconditional information matrix by simulation, and confirm the accuracy of our proposed technique in a Monte Carlo exercise.

Finally, we apply our methodology to obtain the frontier generated by the ten US sectoral indices in Datastream. Our results illustrate several interesting features of the resulting mean-variance-skewness frontier. Specifically, we find that, for a given variance, important gains in terms of positive skewness can be obtained with very small reductions in expected returns. We also analyse the effect of considering additional assets in our portfolios. In particular, we formally test whether the Datastream World-ex US index is able to improve the investment opportunity set in the traditional mean-variance sense, as well as in the skewness-variance sense.

The rest of the paper is organised as follows. We define *LSMN* in section 2.1, and explain how to reparametrise them so that their mean is zero and their covariance matrix the identity. Then, we analyse portfolio allocation in section 3, and discuss maximum likelihood estimation in section 4. Section 5 presents the results of our empirical application, which are followed by our conclusions. Proofs and auxiliary results can be found in appendices.

# 2 Distributional assumptions

## 2.1 Location-scale mixtures of normals

Consider the following $N$-dimensional random vector $\mathbf{u}$, which can be expressed in terms of the following Location-Scale Mixture of Normals ($LSMN$):

$$\mathbf{u} = \boldsymbol{\alpha} + \xi^{-1}\boldsymbol{\Upsilon}\boldsymbol{\beta} + \xi^{-1/2}\boldsymbol{\Upsilon}^{1/2}\mathbf{r}, \tag{1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $N$-dimensional vectors, $\boldsymbol{\Upsilon}$ is a positive definite matrix of order $N$, $\mathbf{r} \sim N(\mathbf{0}, \mathbf{I}_N)$, and $\xi$ is an independent positive mixing variable. For the sake of concreteness, we will denote the distribution function of $\xi$ as $F(\cdot;\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is a vector of $q$ shape parameters. Since $\mathbf{u}$ given $\xi$ 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. The parameters $\boldsymbol{\tau}$ allow for flexible tail modelling, while the vector $\boldsymbol{\beta}$ introduces skewness in this distribution.

We will refer to the distribution of $\mathbf{u}$ as $LSMN_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Upsilon}, \boldsymbol{\tau})$. To obtain a version that we can use to model the standardised residuals of any conditionally heteroskedastic, dynamic regression model, we need to restrict $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ in (1) as follows:

**Proposition 1** *Let* $\boldsymbol{\varepsilon}^* \sim LSMN_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Upsilon}, \boldsymbol{\tau})$ *and* $\pi_k(\boldsymbol{\tau}) = E\left(\xi^{-k}\right)$. *If* $\pi_k(\boldsymbol{\tau}) < \infty$ *for* $k = 1, 2$, $\boldsymbol{\alpha} = -c(\boldsymbol{\beta}, \boldsymbol{\tau})\boldsymbol{\beta}$ *and*

$$\boldsymbol{\Upsilon} = \frac{1}{\pi_1(\boldsymbol{\tau})}\left[\mathbf{I}_N + \frac{c(\boldsymbol{\beta}'\boldsymbol{\beta}, \boldsymbol{\tau}) - 1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right],$$

*where*

$$c(x, \boldsymbol{\tau}) = \frac{-1 + \sqrt{1 + 4xc_v^2(\boldsymbol{\tau})}}{2xc_v^2(\boldsymbol{\tau})}, \tag{2}$$

*and*

$$c_v(\boldsymbol{\tau}) = \frac{\sqrt{\pi_2(\boldsymbol{\tau}) - \pi_1^2(\boldsymbol{\tau})}}{\pi_1(\boldsymbol{\tau})},$$

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

As expected, the scale of $\xi^{-1}$, which can be fully characterised by $\pi_1(\boldsymbol{\tau})$, is arbitrary, and can be set to 1 without loss of generality. Notice also that the distribution of $\boldsymbol{\varepsilon}^*$ becomes a simple scale mixture of normals, and thereby spherical, when $\boldsymbol{\beta}$ is zero. Like any scale mixture of normals, though, this 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 a $LSMN$ is that, although the elements of $\boldsymbol{\varepsilon}^*$ are uncorrelated, they are not independent except in the multivariate normal case. In general, the $LSMN$ induces "tail dependence", which operates through the positive *mixing* variable in (1). Intuitively, $\xi$ 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 $\boldsymbol{\beta}$ appropriately. Specifically, we can make the joint 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 density of the standardised bivariate normal with those of two asymmetric examples: a particular case of the $GH$ distribution known as the asymmetric $t$ (see Appendix C) and a *LSMN* whose mixing variable is Bernoulli.[1] We can observe in Figures 1c and 1e that the non-Gaussian densities are much more peaked around their mode than the Gaussian one. In addition, the contour plots of the asymmetric examples show that we have introduced much fatter tails in the third quadrant by considering negative values for all the elements of $\boldsymbol{\beta}$. This is confirmed in Figure 2, which represents the so-called exceedance correlation between the uncorrelated marginal components in Figure 1. Therefore, a *LSMN* could capture the empirical observation that there is higher tail dependence across stock returns in market downturns (see Longin and Solnik, 2001). In this sense, the examples that we consider illustrate the flexibility of a *LSMN* to generate different shapes for the exceedance correlation, which could be further enhanced by assuming a multinomial distribution for $\xi$.

It is possible to show that the marginal distributions of linear combinations of a *LSMN* (including the individual components) can also be expressed as a *LSMN*:

**Proposition 2** *Let $\boldsymbol{\varepsilon}^*$ be distributed as a $N \times 1$ standardised LSMN random vector with parameters $\boldsymbol{\tau}$ 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 LSMN scalar random variable with parameters $\boldsymbol{\tau}$ and*

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

*where $c(\cdot, \cdot)$ is defined in (2).*

[1] Interestingly, the *LSMN* driven by the Bernoulli mixing variable in Figures 1 and 2 can be interpreted as a mixture of two multivariate normal distributions with different mean vectors but proportional covariance matrices.

Proposition 2 generalises an analogous result obtained by Blæsild (1981) for the *GH* distribution. Note that only the skewness parameter, $\beta(\mathbf{w})$, is affected, as it becomes a function of the weights, $\mathbf{w}$. As we shall see in section 3, this is particularly useful for asset allocation purposes, since the returns to any conceivable portfolio of a collection of assets is a linear combination of the returns on those primitive assets. For the same reason, Proposition 2 is very useful for risk management purposes, since we can easily compute in closed form the Value at Risk of any portfolio from the parameters of the joint distribution. Finally, it also implies that skewness is a "common feature" of *LSMN*, in the Engle and Kozicki (1993) sense, as we can generate a full-rank linear transformation of $\boldsymbol{\varepsilon}^*$ with the asymmetry confined to a single element.

## 2.2 Dynamic econometric specifications

We will analyse investments in a risk-free asset and a set of $N$ risky assets with excess returns $\mathbf{y}_t$. To accommodate flexible specifications, we assume that those excess returns are generated by the following conditionally heteroskedastic dynamic regression model:

$$\left.\begin{array}{c}\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*, \\ \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}\left(I_{t-1};\boldsymbol{\theta}\right), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\left(I_{t-1};\boldsymbol{\theta}\right), \end{array}\right\} \tag{3}$$

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

In this context, we will assume that the distribution of $\boldsymbol{\varepsilon}_t^*$ is a *LSMN* conditional on $I_{t-1}$. Importantly, given that the standardised innovations are not generally observable, the choice of "square root"matrix is not irrelevant except in univariate models, or in multivariate 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, 2005). 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 $\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ 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:

$$\boldsymbol{\beta}_t(\boldsymbol{\theta}, \mathbf{b}) = \mathbf{\Sigma}_t^{\frac{1}{2}\prime}(\boldsymbol{\theta})\mathbf{b}. \tag{4}$$

It is then straightforward to see that the distribution of $\mathbf{y}_t$ conditional on $I_{t-1}$ will not depend on the choice of $\mathbf{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})$.[2]

# 3 Portfolio allocation

## 3.1 The investor's problem

Consider an investor whose wealth at time $t-1$ is $A_{t-1}$. If she allocates her wealth among the $N+1$ available assets, then her wealth at $t$ can be expressed as:

$$A_t = A_{t-1}\left(1 + r_t + \mathbf{w}_t'\mathbf{y}_t\right),$$

where $r_t$ is the risk free rate, and $\mathbf{w}_t$ is the vector of allocations to the risky assets, both of which are known at $t-1$. She will choose the allocations that maximise her expected utility at $t-1$. That is,

$$\mathbf{w}_t^* = \arg\max_{\mathbf{w}_t \in \mathbb{R}^N} E\left[U(A_t)|I_{t-1}\right], \tag{5}$$

where $U(\cdot)$ is her utility function and $I_{t-1}$ denotes the information set available at $t-1$. In this context, we can show the following property for any *LSMN*:

**Proposition 3** *Let $\mathbf{y}_t$ be conditionally distributed as a $N \times 1$ LSMN random vector with conditional mean $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, conditional covariance matrix $\mathbf{\Sigma}_t(\boldsymbol{\theta})$, and shape parameters $\boldsymbol{\tau}$ and $\mathbf{b}$. Then, for any vector $\mathbf{w}_t \in \mathbb{R}^N$ known at $t-1$, the conditional distribution of $\mathbf{w}_t'\mathbf{y}_t$ can be fully characterised as a function of its mean, variance and skewness.*

Proposition 3 implies that, if the distribution of asset returns is a *LSMN*, then any portfolio is completely described just by its mean, variance and skewness. Hence, no matter what preferences we consider, the expected utility of any portfolio will be a

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

function of its first three moments. In this sense, it is straightforward to show that the first two moments of $A_t$ can be expressed as:

$$
\begin{aligned}
E_{t-1}(A_t) &= A_{t-1}\left[1 + r_t + \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})\right], \\
E_{t-1}\left\{\left[A_t - E_{t-1}(A_t)\right]^2\right\} &= A_{t-1}^2\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t.
\end{aligned}
$$

As for the third moment, we can use the results in Appendix B to show that

$$
E_{t-1}\left[\left(A_t - E_{t-1}(A_t)\right)^3\right] = A_{t-1}^3\varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau})
$$

where

$$
\varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) = (s_{1t} + 3s_{2t}s_{3t})\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^3 + 3s_{2t}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right], \tag{6}
$$

and

$$
\begin{aligned}
s_{1t} &= \frac{E\left\{\left[\xi^{-1} - \pi_1(\boldsymbol{\tau})\right]^3\right\}}{\pi_1^3(\boldsymbol{\tau})} c^3[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}], \\
s_{2t} &= c_v^2(\boldsymbol{\tau})c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}], \\
s_{3t} &= \left\{c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}] - 1\right\} / \left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right].
\end{aligned}
$$

Since in line with most of the literature we are implicitly assuming that the investment technology shows constant returns to scale, we can normalise the above moments by setting $A_{t-1} = 1$ without loss of generality. In addition, we will systematically consider all portfolio returns in excess of the risk free rate in what follows.

## 3.2 Mean-variance and skewness-variance frontiers

Consider an investor who, ceteris paribus, prefers high expected returns and positive skewness but dislikes high variances. Under this fairly mild assumption, a portfolio whose returns can be expressed as a *LSMN* will only be optimal if it is located on the mean-variance-skewness frontier. Given that Proposition 3 shows that only the first three moments matter in this context, it will always be possible to improve the investor's utility at any interior point by either increasing the expected return or the positive skewness of her portfolio, or reducing its variance.

The mean-variance-skewness frontier is a generalisation of the mean-variance frontier:

$$
\mu_{0t} = \sigma_{0t}\sqrt{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}. \tag{7}
$$

which we obtain by maximising expected return $\mu_{0t}$ for every possible standard deviation $\sigma_{0t}$. As is well known, the mean-variance frontier (7) can be spanned by just two funds: the risk-free asset and a portfolio with weights proportional to $\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$.

Similarly, we can obtain a skewness-variance frontier by maximising skewness subject to a variance constraint:

**Proposition 4** *If*

$$\frac{s_{2t}}{s_{1t}+3s_{2t}s_{3t}}\left[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}+\frac{s_{2t}}{s_{1t}+3s_{2t}s_{3t}}\right]>0. \tag{8}$$

*then the solution to the problem*

$$\max_{\mathbf{w}_t\in\mathbb{R}^N}\varphi_t(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})\quad s.t.\quad \mathbf{w}_t'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t=\sigma_{0t}^2 \tag{9}$$

*will be*

$$\left[\varphi_t(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})\right]^{1/3}=\Lambda_1(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})\sigma_{0t}, \tag{10}$$

*where*

$$\Lambda_1(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})=\left\{(s_{1t}+3s_{2t}s_{3t})\left[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^{3/2}+3s_{2t}\left[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^{1/2}\right\}^{1/3}, \tag{11}$$

*which is achieved by*

$$\mathbf{w}_t^{\dagger}=\frac{\sigma_{0t}}{\sqrt{\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}. \tag{12}$$

*Otherwise the solution to (9) will be*

$$\left[\varphi_t(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})\right]^{1/3}=\max\left\{\Lambda_1(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau}),\Lambda_2(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})\right\}\sigma_{0t},$$

*where*

$$\Lambda_2(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau})=2^{1/3}\sqrt{s_{2t}}\left[-s_{1t}-3s_{2t}s_{3t}\right]^{-1/6}, \tag{13}$$

*which is obtained by portfolios that satisfy*

$$\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^{\ddagger}=\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}}. \tag{14}$$

Therefore, there are two cases. If (8) is satisfied, then there will be a unique solution to the skewness-variance frontier given by (10). In this case, we can interpret $\mathbf{b}$ as a "skewness-variance" efficient portfolio, since every portfolio on this frontier will be proportional to $\mathbf{b}$. However, when (8) is not satisfied, (12) will not necessarily yield maximum skewness, because there is another local maximum characterised by (14). In addition, whereas there will be just one portfolio satisfying (12) for any given variance, there might be an infinite number of portfolios that satisfy (14), all of them yielding exactly the same variance and skewness but different expected returns. Therefore, we must take their expected returns into account in order to decide which of them will be preferred by a rational investor. Specifically, for any investor who, ceteris paribus, prefers high to low expected returns, it will only be optimal to chose the portfolio satisfying (14) that maximises expected return. In this sense, we can show that:

**Proposition 5** *If (8) does not hold, then the solution to the problem*

$$\arg\max_{\mathbf{w}_t\in\mathbb{R}^N} \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) \quad s.t. \quad \left\{ \begin{array}{c} \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2 \\ \mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}\sqrt{\dfrac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}} \end{array} \right. \tag{15}$$

*can be expressed as a linear combination of the "skewness-variance" efficient portfolio* $\mathbf{b}$ *and the "mean-variance" efficient portfolio* $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$.

Once again, it is important to emphasise that (15) only has a solution if condition (8) is not satisfied. In that case, the asymmetry-variance frontier will be spanned by the risk free asset, $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$, and $\mathbf{b}$ if in addition (13) is greater than (11). Otherwise, we will only need two funds: the risk-free asset and $\mathbf{b}$.

### 3.3 Mean-variance-skewness frontiers

The efficient portion of the mean-variance-skewness frontier yields the maximum asymmetry for every feasible combination of mean and variance. We can express this problem as follows:

$$\max_{\mathbf{w}_t\in\mathbb{R}^N} \varphi_t(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau}) \quad \text{s.t.} \quad \left\{ \begin{array}{c} \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mu_{0t} \\ \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2 \end{array} \right. \tag{16}$$

Obviously, there are other approaches to obtain this frontier. For instance, Athayde and Flôres (2004) maximise expected returns subject to constraints on the variance and asymmetry, as in Proposition 5. However, we prefer the formulation in (16) because it is straightforward to ensure the feasibility of the target expected return and variance. Specifically, we can exploit the fact that, for a given expected return $\mu_{0t}$, the target variance $\sigma_{0t}^2$ must be greater or equal than that of the mean-variance frontier (7), that is

$$\sigma_{0t}^2 \geq \frac{\mu_{0t}^2}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}. \tag{17}$$

We can solve (16) by forming the Lagrangian

$$\mathcal{L} = \varphi_t(\boldsymbol{\theta},\mathbf{b},\boldsymbol{\tau}) + \gamma_1\left[\mu_{0t} - \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] + \gamma_2\left[\sigma_{0t}^2 - \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right], \tag{18}$$

and differentiating it with respect to the portfolio weights, thereby obtaining the following first order conditions:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_t} = \left\{ 3(s_{1t} + 3s_{2t}s_{3t}) \left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]^2 + 3s_{2t}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right] \right\} \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$$
$$+6s_{2t}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t - \gamma_1\boldsymbol{\mu}_t(\boldsymbol{\theta}) - 2\gamma_2\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t. \tag{19}$$

We can explicitly obtain in closed-form the set of portfolio weights that satisfy these conditions:

**Proposition 6** *The efficient mean-variance-skewness portfolios that solve (19) can be expressed as either*

$$\mathbf{w}_{1t}^* = \frac{\mu_{0t} + \Delta_t^{-1}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{1}{\Delta_t}\mathbf{b}, \tag{20}$$

*or*

$$\mathbf{w}_{2t}^* = \frac{\mu_{0t} - \Delta_t^{-1}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) + \frac{1}{\Delta_t}\mathbf{b}, \tag{21}$$

*where*

$$\Delta_t = \sqrt{\frac{(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2}{\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2}}.$$

Thus, there are two potential solutions,[3] both of which can be expressed as a linear combination of the mean-variance efficient portfolio $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and the skewness-variance efficient portfolio $\mathbf{b}$. Hence, Proposition 6 shows that the efficient region of the mean-variance-skewness frontier can be spanned by the aforementioned three funds. In addition, it can be shown that if (8) holds, then not only the efficient section but also the whole frontier will be spanned by those three funds.

In order to obtain an explicit equation for the frontier, let $j = -1, +1$ and define $\varphi_{0t}(j)$ as the third centred moment that results from introducing (20) or (21) in (6),

[3]In order to assess whether (20) or (21) yields the efficient part of the frontier, we can check for which of the two solutions the Hessian matrix,

$$6(s_{1t} + 3s_{2t}s_{3t})(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t)\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})'$$
$$+6s_{2t}\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\mathbf{b}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})'\right]$$
$$+\left[6s_{2t}(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t) - 2\gamma_2\right]\boldsymbol{\Sigma}_t(\boldsymbol{\theta}),$$

is negative definite.

respectively. It is straightforward to show that $\varphi_{0t}(j)$ can be expressed as:

$$\begin{aligned}\varphi_{0t}(j) &= (s_{1t}+3s_{2t}s_{3t})h_{1t}(4h_{1t}^2-3h_{2t})\mu_{0t}^3 \\ &+3\left\{(s_{1t}+3s_{2t}s_{3t})h_{1t}(h_{2t}-h_{1t}^2)\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right]+s_{2t}h_{1t}\right\}\mu_{0t}\sigma_{0t}^2 \\ &+j\sqrt{(h_{2t}-h_{1t}^2)\{\sigma_0^2\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right]-\mu_{0t}^2\}} \\ &\times\left(\begin{array}{c}(s_{1t}+3s_{2t}s_{3t})(4h_{1t}^2-h_{2t})\mu_{0t}^2 \\ +\left\{(s_{1t}+3s_{2t}s_{3t})(h_{2t}-h_{1t}^2)\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right]+3s_{2t}\right\}\sigma_{0t}^2\end{array}\right)\end{aligned} \tag{22}$$

where

$$\begin{aligned} h_{1t} &= \frac{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}, \\ h_{2t} &= \frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}. \end{aligned}$$

If (17) is satisfied with equality, which only occurs on the mean variance frontier, then one can show that $\mathbf{w}_{1t}^* = \mathbf{w}_{2t}^*$ and $\varphi_{0t}(-1) = \varphi_{0t}(1)$. Interestingly, if $\mathbf{b} = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$, then the mean-variance and skewness-variance frontiers will coincide, and (22) will collapse to

$$\varphi_{0t} = (s_{1t}+3s_{2t}s_{3t})\mu_{0t}^3 + 3s_{2t}\mu_{0t}\sigma_{0t}^2,$$

where (17) holds with equality.

It is not difficult to show that (22) satisfies the set of properties obtained by Athayde and Flôres (2004) for general distributions. The two most important ones are homothecy and linearity along directions in which the Sharpe ratio remains constant. Homothecy states that if a portfolio with weights $\mathbf{w}_t^*$ belongs to the frontier, then $k\mathbf{w}_t^*$ will also be on the frontier. Moreover, if we consider a direction in which $\sigma_{0t}$ is proportional to $\mu_{0t}$, $\sigma_{0t} = k'\mu_{0t}$ say, then the cubic root of the asymmetry will also be proportional to $|\mu_{0t}|$ along this direction.

Figures 3 and 4 show the shape of the mean-variance-skewness frontier for two examples with three risky assets. In Figure 3 we have chosen $\mathbf{b}$ so that (8) is satisfied. The three dimensional plot of the frontier is displayed in Figure 3a. In addition, we also compute the three types of contour plots. Figure 3b shows the well known mean-variance frontier, but it also includes several iso-skewness lines along which $\varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau})$ is constant. Note that the efficient section of the mean-variance frontier corresponds to negative skewness in this example.

We focus on the mean-skewness space in Figure 3c, where we plot the iso-variance lines and include the efficient parts of both mean-variance and asymmetry-variance frontiers, whose linearity on this space is due to the homothecy property discussed above. Note that the mean-variance frontier is located on the eastern part of the space, while the asymmetry-variance frontier is on the northern half. This is a general result because for a given variance the former contains the points with highest expected return, which is displayed on the x-axis, while the latter maximises skewness (on the y-axis). Furthermore, for the same reason the asymmetry-variance frontier will always be above the mean-variance line. In this sense, an investor who prefers higher expected returns and positive skewness will choose a portfolio that is located to the right of the skewness-variance frontier and above the mean-variance one. Otherwise, she will be worse off in terms of either expected return or skewness. Thus, if she only cares about the mean, she will choose some point on the mean-variance frontier, while if she only cares about asymmetry, she will choose some point on the skewness-variance frontier. In general, though, she will choose an intermediate combination.

We consider the skewness-variance space in Figure 3d, where we can confirm the linearity of the skewness-variance frontier (see Proposition 4).

Finally we display in Figure 4 the analogous graphs for a case in which condition (8) is not satisfied and (13) is larger than (11). As expected, in this case the iso-variance contours have a flat region with maximum skewness. However, only the points of this region with highest expected return will be relevant in practice, as the vertical part of the iso-skewness contours in Figure 4b show.

# 4 Maximum likelihood estimation

In the previous sections, we have assumed that we know the true values of the parameters of interest, $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\tau})'$. Of course, this is not the case in practice. Given that we are considering a specific family of distributions, it seems natural to estimate $\boldsymbol{\phi}$ by maximum likelihood.

The log-likelihood function of a sample of size $T$ takes the form

$$L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l(\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi}),$$

where $l(\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi})$ is the conditional log-density of $\mathbf{y}_t$ given $I_{t-1}$ and $\boldsymbol{\phi}$. We can generally express this log-density as

$$l(\mathbf{y}_t|I_{t-1};\boldsymbol{\phi}) = \log\left[E\left[f(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi})\,|I_T;\boldsymbol{\phi}\right]\right],$$

where $f(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi})$ is the Gaussian likelihood of $\mathbf{y}_t$ given $\xi_t$, $I_{t-1}$ and $\boldsymbol{\phi}$. Given the nonlinear 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. Fortunately, it is possible to obtain analytical expressions for the score vector of our model, 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).

## 4.1 The score vector

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

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

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

$$s_t(\boldsymbol{\phi}) = E\left(\left.\frac{\partial l(\mathbf{y}_t|\xi_t, I_{t-1};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right| I_T;\boldsymbol{\phi}\right) + E\left(\left.\frac{\partial l(\xi_t|I_{t-1};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right| I_T;\boldsymbol{\phi}\right) \tag{23}$$

because $E\left[\partial l\left(\xi_t|\mathbf{y}_t, I_{t-1};\boldsymbol{\phi}\right)/\partial\boldsymbol{\phi}|\, I_T;\boldsymbol{\phi}\right] = \mathbf{0}$ by virtue of the Kullback inequality. This result was first noted by Louis (1982); see also Ruud (1991) and Tanner (1996, p. 84). In this way, we decompose $s_t(\boldsymbol{\phi})$ as the sum of the expected values of (i) the score of a multivariate Gaussian log-likelihood function, and (ii) the score of the distribution of the mixing variable.[4] We illustrate this procedure in Appendix C for the particular case of the $GH$ distribution.

## 4.2 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 standard 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

$$\mathcal{I}(\boldsymbol{\phi}_0) = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} s_t(\boldsymbol{\phi}_0)s_t'(\boldsymbol{\phi}_0) = E[s_t(\boldsymbol{\phi}_0)s_t'(\boldsymbol{\phi}_0)]. \tag{24}$$

In general, though, (24) cannot be obtained in closed form.[5] 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 (24) in pure time series models by generating a long simulated path of size $T_s$ of the postulated process $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \cdots, \hat{\mathbf{y}}_{T_s}$, where the symbol $\hat{}$ indicates that the data has been generated using the maximum likelihood estimates $\hat{\boldsymbol{\phi}}_T$. This path can be easily generated by exploiting (1). 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

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

[4]It is possible to show that $\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*/N$ converges in mean square to $1/[\pi_1(\boldsymbol{\tau})\xi]$ as $N\to\infty$. This means that in the limit the latent variable $\xi_t$ could be fully recovered from observations on $\mathbf{y}_t$, which would greatly simplify the calculations implicit in expression (23).

[5]Exact formulas for the conditional information matrix are known, for instance, for the Gaussian (see Bollerslev and Wooldridge, 1992) and the Student $t$ distributions (see Fiorentini, Sentana, and Calzolari, 2003).

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.

We have compared the finite sample performance of our technique with the accuracy of other alternative estimators of the sampling variance of the ML estimators. In our Monte Carlo exercise, we use a trivariate experimental design borrowed from Sentana (2004), which aimed to capture some of the main features of the conditionally heteroskedastic factor model in King, Sentana, and Wadhwani (1994). Specifically, we model the standardised residuals with the *GH* distribution, while the conditional mean and variance specifications are given by:

$$\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{cc}'\lambda_t + \boldsymbol{\Gamma}_t, \end{aligned} \tag{25}$$

where $\boldsymbol{\mu}' = (\mu_1, \mu_2, \mu_3)$, $\mathbf{c}' = (c_1, c_2, c_3)$, $\boldsymbol{\Gamma}_t = diag(\gamma_{1t}, \gamma_{2t}, \gamma_{3t})$,

$$\lambda_t = \alpha_0 + \alpha_1(f^2_{t-1|t-1} + \omega_{t-1|t-1}) + \alpha_2\lambda_{t-1}, \tag{26}$$

$$\gamma_{it} = \phi_0 + \phi_1\left[(y_{it-1} - \mu_i - c_i f_{t-1|t-1})^2 + c_i^2\omega_{t-1|t-1}\right] + \phi_2\gamma_{it-1}, \quad i = 1, 2, 3, \tag{27}$$

$f_{t|t} = \omega_{t|t}\mathbf{c}'\boldsymbol{\Gamma}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))$ and $\omega_{t|t} = [\lambda_t^{-1} + \mathbf{c}'\boldsymbol{\Gamma}_t^{-1}\mathbf{c}]^{-1}$. This parametrisation can be interpreted in terms of a latent factor model where (26) would be the variance of the latent factor, while (27) would correspond to the idiosyncratic effects. 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.

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 before.[6] Since the purpose of this exercise is to guide empirical work, our target is the sampling covariance matrix of the ML estimators, $V_T(\hat{\boldsymbol{\phi}}_T)$, which we estimate as the Monte Carlo covariance matrix of $\hat{\boldsymbol{\phi}}_T$ in 30,000 samples of 1,000 observations each. Given the large number of parameters involved, we summarise the performance of the estimators of $V_T(\hat{\boldsymbol{\phi}}_T)$ by looking at the sampling distributions of the logs of $vech'[V_T^E(\hat{\boldsymbol{\phi}}_T) - V_T(\hat{\boldsymbol{\phi}}_T)]vech[V_T^E(\hat{\boldsymbol{\phi}}_T) - V_T(\hat{\boldsymbol{\phi}}_T)]$ and

[6] We choose $\eta = .1$, $\psi = 1$ and $\mathbf{b} = -.1\boldsymbol{\iota}$ as the shape parameters of the *GH* distribution. See appendix C.

$vecd'[V_T^E(\hat{\boldsymbol{\phi}}_T)-V_T(\hat{\boldsymbol{\phi}}_T)]vecd[V_T^E(\hat{\boldsymbol{\phi}}_T)-V_T(\hat{\boldsymbol{\phi}}_T)]$, where E is either O, H or I.[7] The results, which are presented in Figures 5a and 5b, respectively, show that the I standard errors seem to be systematically more reliable than either the O or numerical H counterparts.

# 5 Empirical application

We now apply the methodology derived in the previous sections to the ten Datastream main sectoral indices for the US.[8] Specifically, our dataset consists of daily excess returns for the period January 4th, 1988 - October 12th, 2007 (4971 observations), where we have used the Eurodollar overnight interest rate as safe rate (Datastream code ECUSDST). The model used is a generalisation of the one in the previous section (see (25)), 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 borrowed this application from Mencía and Sentana (2008), who find that these indices are asymmetric and leptokurtic. We have estimated this model by maximum likelihood under the assumption that the conditional distribution of the innovations is *GH*. Although this distribution has already 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 our perspective. We use the formulae for the score provided in Appendix C, and compute the standard errors by simulation as explained in section 4.2.

The first column of Table 1 shows the estimates of the shape parameters of this distribution. Although not all of the asymmetry parameters are individually significant, Mencía and Sentana (2008) report that symmetry is rejected at conventional levels. In particular, a joint LR test of symmetric vs. asymmetric *GH* innovations yields 23.45 (p-value=0.012), while the result of an analogous LM symmetry test is 25.35 (p-value=0.005).

One potential concern is whether we are able to correctly capture the dynamics of the

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

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

data. If our model were misspecified, then it could introduce severe distortions in the results. However, if our specification of the model dynamics is correct, the departure from normality that we have found should not affect the consistency of the Gaussian PML estimators of $\boldsymbol{\theta}$. With this in mind, we compare the estimates of the conditional variances obtained with a univariate Gaussian AR(1)-GQARCH(1,1) model for the equally weighted portfolio with the ones obtained from the Gaussian version of our multivariate model. Reassuringly, Figure 6a shows that the (log) standard deviations of the two series display a very similar pattern, although the univariate estimates are somewhat noisier. Another way to check the adequacy of our specification is to compare the multivariate Gaussian and *GH* estimates. In this sense, Figure 6b shows that the (log) standard deviations implied by the two distributional assumptions for the equally weighted portfolio are extremely similar.

From an investor's point of view, an important question is whether the addition of some assets improves the trade-offs that they face. Given that we have only considered investments in the US so far, it seems natural to test whether the mean-variance-skewness frontier remains unchanged when we also allow for investments outside the US, which we proxy by the Datastream World ex-US index. Notice that this test generalises the usual mean-variance spanning tests, because it also takes into account the effect of the World ex-US index on the skewness-variance frontier.

As is well known (see e.g. Gibbons, Ross, and Shanken, 1989), the additional asset does not lead to any change in the mean-variance frontier if and only if the conditional mean of the additional asset satisfies

$$\mu_{2t}(\boldsymbol{\theta}) = \mathbf{d}'_{12t}(\boldsymbol{\theta})\boldsymbol{\mu}_{1t}(\boldsymbol{\theta}), \tag{28}$$

where $\boldsymbol{\mu}_{1t}(\boldsymbol{\theta})$ and $\mu_{2t}(\boldsymbol{\theta})$ denote, respectively, the vector of (conditional) expected excess returns on the ten US indices, and the expected excess return of the Word ex-US index, while $\mathbf{d}_{12t}(\boldsymbol{\theta})$ denotes the coefficients of the conditional regression of the World ex-US index excess returns on those of the US sectoral indices. Therefore, we can follow Gibbons, Ross, and Shanken (1989), and check (28) by introducing an intercept in this expression and assessing whether it equals zero in practice.

Similarly, the World ex-US index will only expand the skewness-variance frontier if its skewness parameter is significantly different from zero (see (20) and (21)). We analyse these two effects in Table 2 by means of Wald and LR tests. While we are unable to reject the mean-variance spanning restriction (28), the World ex-US index seems to introduce significant skewness in the investment opportunity set of a US investor. As a consequence, we reject the joint null. This result has interesting implications. In particular, for the set of assets that we consider, a US investor that only cares about mean-variance efficiency will not be willing to invest outside the US. In contrast, if this investor takes skewness into account in making her portfolio decisions, then she will find significant gains by investing part of her wealth outside the US.

Figure 7 illustrates these gains by showing the mean-variance-skewness frontier before and after considering the additional asset. The results of this figure correspond to a representative day whose mean vector and covariance matrix are set to their unconditional values. We can observe the differences between the two frontiers in Figure 7a, where we consider a three dimensional plot in which we include the positions of the individual indices. We can also observe in Figure 7b that the mean-variance frontier is almost unaffected, which is consistent with (28) being satisfied. Nevertheless, the iso-skewness lines have moved to the left, which implies that, for given levels of expected return and skewness, we can obtain a lower standard deviation if we invest in the World ex-US index. Figures 7c and 7d confirm this effect on the iso-variance and the skewness-variance frontiers, respectively. We can also notice in Figure 7c that the iso-variance lines are rather flat with respect to skewness. Hence, if we start from some point on the mean-variance frontier and follow the corresponding iso-variance line, we can substantially increase skewness without hardly deteriorating expected returns. Finally, note that the third column of Table 1 shows that the estimates of the shape parameters of the $GH$ distribution remain fairly stable when we include the additional asset.

# 6 Conclusions

In this paper, we make mean-variance-skewness analysis fully operational by working with a rather flexible family of multivariate asymmetric distributions, known as location-scale mixtures of normals ($LSMN$), which nest as particular cases several popular and empirically relevant distributions that account for asymmetry and tail dependence with a rather flexible and parsimonious structure. Specifically, we assume that, conditional on the information that agents have at the time they make their investment decisions, the standardised innovations of excess returns can be expressed as a $LSMN$.

In this context, we show that the distribution of any portfolio of the original assets can be fully characterised in terms of its mean, variance and skewness. Hence, investors who like high means and positive asymmetry but dislike high variances will only choose among portfolios on the mean-variance-skewness frontier regardless of their specific preferences. In this sense, our result extends previous results by Chamberlain, 1983; Owen and Rabinovitch, 1983 and Berk, 1997, which justify the use of mean-variance analysis with elliptically distributed returns. In addition, we are able to obtain analytical expressions for the mean-variance-skewness frontier, and show that its efficient part can always be spanned by three funds: the risk-free asset, a mean-variance efficient portfolio, and a skewness-variance efficient portfolio.

We also study the maximum likelihood estimation of dynamic models for excess returns with $LSMN$ innovations. In particular, we provide analytical expressions for the score on the basis of the EM algorithm, and explain how to evaluate the information matrix by simulation. A detailed Monte Carlo exercise confirms that our method yields more accurate standard errors than the Hessian matrix or the sample outer product of the score.

Finally, we estimate the mean-variance-skewness frontier generated by the ten Datastream main sectoral indices for the US for the particular case of $GH$ innovations. We find that by moving away from the traditional mean-variance frontier, we can increase skewness for a given variance without hardly reducing expected returns. We also analyse whether including the Datastream World ex-US index can improve the investment opportunity set of a US investor. We find that this additional asset does not have a significant impact from a mean-variance perspective. In contrast, it does indeed offer substantial improvements once we take into account its effect on skewness.

It would be interesting to check whether our empirical results are robust to replacing the $GH$ assumption by a nonparametric specification for the distribution of the mixing variable $\xi_t$. Another fruitful avenue for future research would be to assess the asset pricing implications of our model. In particular, we could relate our framework to the extensions of the CAPM based on the first three moments of returns (see e.g. Kraus and Litzenberger, 1976; Barone-Adesi, 1985; and Lim, 1989). Similarly, it would be useful to explore the implications of our model at different time horizons. As a starting point, we could exploit the properties of specific examples such a the Variance Gamma process, which generates Asymmetric Normal Gamma returns at any investment horizon (see e.g. Madan and Milne, 1991; and Madan, Carr, and Chang, 1998). Finally, it would also be interesting to derive a specification test of the "common feature" in skewness implicit in our model, and, if needed, relax that assumption by allowing for several skewness factors.

# References


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

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

Athayde, G. M. d. and R. G. Flôres (2004). Finding a maximum skewness portfolio- a general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control 28*, 1335–1352.

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

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

Barone-Adesi, G. (1985). Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis 20*, 299–313.

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

Berk, J. (1997). Necessary conditions for the CAPM. *Journal of Economic Theory 73*, 245–257.

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

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

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

Chamberlain, G. (1983). A characterisation of the distributions that imply mean-variance utility functions. *Journal of Economic Theory 29*, 185–201.

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

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

Davidson, R. and J. G. MacKinnon (1993). *Estimation and inference in econometrics.* Oxford, U.K.: Oxford University Press.

Engle, R. and S. Kozicki (1993). Testing for common features. *Journal of Business and*

*Economic Statistics 11*, 369–380.

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.

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

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*, 1121–1152.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2002). Portfolio selection with higher moments. Duke University Working Paper.

Jondeau, E. and M. Rockinger (2006). Optimal portfolio allocation under higher moments. *European Financial Management 12*, 29–55.

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

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

Kon, S. J. (1984). Models of stock returns-A comparison. *The Journal of Finance 39*, 147–165.

Kraus, A. and R. H. Litzenberger (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance 31*, 1085–1100.

Lim, K. G. (1989). A new test of the three-moment capital asset pricing model. *Journal of Financial and Quantitative Analysis 24*, 205–216.

Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. *The Journal of Finance 56*, 649–676.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *Journal of the Royal Statistical Society, Series B 44*, 98–103.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The Variance Gamma process and option pricing. *European Finance Review 2*, 79–105.

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

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. and E. Sentana (2008). Distributional tests in multivariate dynamic models with normal and student $t$ innovations. mimeo.

Owen, J. and R. Rabinovitch (1983). On the class of elliptical distributions and their

applications to the theory of portfolio choice. *The Journal of Finance 38*, 745–752.
Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics 2*, 130–168.
Prause, K. (1998). The generalised hyperbolic models: estimation, financial derivatives and risk measurement. Unpublished Ph.D. thesis, Mathematics Faculty, Freiburg University.
Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics 49*, 305–341.
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.
Tanner, M. A. (1996). *Tools for statistical inference: methods for exploration of posterior distributions and likelihood functions* (Third ed.). New York: Springer-Verlag.

# A Proofs of Propostions

## Proposition 1

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

$$\boldsymbol{\varepsilon}^* = c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\boldsymbol{\beta}\left[\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right]+\sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\left[\mathbf{I}_N+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right]^{\frac{1}{2}}\mathbf{r} \tag{A1}$$

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

$$V\left(\boldsymbol{\varepsilon}^*\right) = c_v^2(\boldsymbol{\tau})c^2\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\boldsymbol{\beta}\boldsymbol{\beta}'+\mathbf{I}_N+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}',$$

Substituting $c\left(\boldsymbol{\beta},\nu,\gamma\right)$ by (2), we can finally show that $V\left(\boldsymbol{\varepsilon}^*\right)=\mathbf{I}_N$. □

## Proposition 2

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

$$\begin{aligned} s^* &= c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\frac{\mathbf{w}'\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{w}}}\left[\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right] \\ &\quad+\sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\frac{\mathbf{w}'}{\sqrt{\mathbf{w}'\mathbf{w}}}\left[\mathbf{I}_N+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right]^{\frac{1}{2}}\mathbf{r}. \end{aligned}$$

But since the second term in this expression can be written as the product of the square root of the mixing variable times a univariate normal variate, $r$ say, we can also rewrite $s^*$ as

$$\begin{aligned} s^* &= c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\frac{\mathbf{w}'\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{w}}}\left[\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right] \\ &\quad+\sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\sqrt{1+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\frac{\left(\mathbf{w}'\boldsymbol{\beta}\right)^2}{\mathbf{w}'\mathbf{w}}}\,r \end{aligned} \tag{A2}$$

Given that $s^*$ is a standardised variable by construction, if we compare (A2) with the general formula for a standardised *LSMN* in (A1), then we will conclude that the parameters $\boldsymbol{\tau}$ 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

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

□

## Proposition 3

If we introduce the results of Proposition 1 in (3), we can express $\mathbf{y}_t$ as:

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\mu}_t(\boldsymbol{\theta}) + c(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau})\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\left[\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})} - 1\right] \\ &+ \sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\left\{\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) + \frac{c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}] - 1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right\}^{\frac{1}{2}}\mathbf{r}_t\end{aligned}$$

where $\xi_t \sim iid\ F(\cdot; \boldsymbol{\tau})$ and $\mathbf{r}_t \sim iid\ N(\mathbf{0}, \mathbf{I}_N)$ are independent. Hence, $\mathbf{w}_t'\mathbf{y}_t$ can be expressed as:

$$\begin{aligned}\mathbf{w}_t'\mathbf{y}_t = \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) + c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}]\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\left[\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})} - 1\right] \\ + \sqrt{\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})}}\left\{\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t + \frac{c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}] - 1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^2\right\}^{\frac{1}{2}}\mathbf{r}_t\end{aligned} \tag{A3}$$

We can observe that $\mathbf{w}_t'\mathbf{y}_t$ is a *LSMN* that can be characterised in terms of its mean $\mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})$, its variance $\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t$ and the bi-linear form $\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$. Finally, the bijective relationship between $\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$ and the third centred moment of $\mathbf{w}_t'\mathbf{y}_t$ (see 6) proves the required result. □

## Propositions 4 and 5

We can solve (9) by forming the Lagrangian

$$\mathcal{L} = \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) + \gamma_2\left(\sigma_{0t}^2 - \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right). \tag{A4}$$

If we differentiate (A4) with respect to the portfolio weights, we obtain the following first order conditions:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{w}_t} &= \left\{3(s_{1t} + 3s_{2t}s_{3t})[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t]^2 + 3s_{2t}\sigma_{0t}^2\right\}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} \\ &+ \left\{6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t] - 2\gamma_2\right\}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = 0\end{aligned}$$

There are two possible situations. First, assume that

$$3(s_{1t} + 3s_{2t}s_{3t})[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t]^2 + 3s_{2t}\sigma_{0t}^2 \tag{A5}$$

is different from zero. In this case, we can express the optimal portfolio weights as $\mathbf{w}_t = \kappa\mathbf{b}$ for some constant $\kappa$. Then, if we impose the variance constraint by choosing $\kappa$ appropriately, we obtain (12). However, an additional solution will be obtained if the scalars (A5) and $6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t]-2\gamma_2$ are both zero. This solution will be characterised by

$$
\begin{aligned}
\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t &= \pm\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}}, \qquad (A6)\\
\mathbf{w}_t\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t &= \sigma_{0t}^2. \qquad (A7)
\end{aligned}
$$

However, we will choose the positive sign because it is the one that yields positive skewness. Condition (A6) defines a plane. Thus, this solution will only exist if this plane intersects the ellipse defined by (A7). We need to find under what conditions (A6) and (A7) are both satisfied. If this solution exists, there will be an infinite number of portfolios with the same asymmetry and standard deviation but different expected returns. We can consider the one that has maximum expected return by solving (15). In this case, the Lagrangian can be expressed as

$$
\begin{aligned}
\mathcal{L} &= \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})+\gamma_1\left[\sigma_{0t}^2-\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\\
&\quad+\gamma_2\left[\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}}-\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]. \qquad (A8)
\end{aligned}
$$

If we differentiate (A8) with respect to $\mathbf{w}_t$, we obtain:

$$
\mathbf{w}_t=\frac{1}{2\gamma_1}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})-\gamma_2\mathbf{b}\right] \qquad (A9)
$$

It is straightforward to show that

$$
\gamma_1=\pm\frac{\sqrt{\boldsymbol{\mu}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})-2\gamma_2\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta})+\gamma_2^2(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})}}{2\sigma_{0t}} \qquad (A10)
$$

ensures that (A7) holds. If we introduce (A9) and (A10) in (A6), we obtain the following restriction:

$$
\frac{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})-\gamma_2\mathbf{b}}{\sqrt{\boldsymbol{\mu}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})-2\gamma_2\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta})+\gamma_2^2[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t]}}=\pm\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}}
$$

If we square the above expression, it is straightforward to show that it can be expressed as a second order equation which will only have real solutions if (8) does not hold. □

## Proposition 6

In what follows we maintain the assumption that (A5) is different from zero, since the equality case is treated in Propositions 4 and 5. If we set (19) to zero, we can express the optimal portfolio weights as:

$$\begin{aligned}\mathbf{w}_t^* &= \frac{\gamma_1}{6s_{2t}[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad - \frac{\{3(s_{1t} + 3s_{2t}s_{3t})[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]^2 + 3s_{2t}\sigma_{0t}^2\}}{6s_{2t}[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{b} \qquad \text{(A11)}\end{aligned}$$

If we pre-multiply (A11) by $\mathbf{b}'\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})$, we obtain:

$$\begin{aligned}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^* &= \frac{\gamma_1}{6s_{2t}[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &- \frac{\{3(s_{1t} + 3s_{2t}s_{3t})[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]^2 + 3s_{2t}\sigma_{0t}^2\}}{6s_{2t}[\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{b}'\mathbf{\Sigma}\mathbf{b} \qquad \text{(A12)}\end{aligned}$$

Hence, we can express (A11) as

$$\mathbf{w}_t^* = \frac{\gamma_1}{6s_{2t}z^* - 2\gamma_2}\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]}{6s_{2t}z^* - 2\gamma_2}\mathbf{b} \qquad \text{(A13)}$$

where $z^*$ is the solution of the following equation:

$$\begin{aligned}&[6s_{2t} + 3(\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\,(s_{1t} + 3s_{2t}s_{3t})]\,z^2 \\ &-2\gamma_2 z + \left[3s_{2t}(\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\sigma_{0t}^2 - \gamma_1\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] = 0. \qquad \text{(A14)}\end{aligned}$$

The equality restrictions of our problem can then be written as:

$$\begin{aligned}\mu_0 &= \frac{\gamma_1}{6s_{2t}z^* - 2\gamma_2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad - \frac{[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_0^2]}{6s_{2t}z^* - 2\gamma_2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \qquad \text{(A15)}\end{aligned}$$

$$\begin{aligned}\sigma_{0t}^2 &= \frac{\gamma_1^2}{[6s_{2t}z^* - 2\gamma_2]^2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad + \frac{[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]^2}{[6s_{2t}z^* - 2\gamma_2]^2}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} \\ &\quad -2\frac{\gamma_1\,[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]}{[6s_{2t}z^* - 2\gamma_2]^2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \qquad \text{(A16)}\end{aligned}$$

Thus, we must find $z^*$, $\gamma_1$ and $\gamma_2$ such that (A14), (A15) and (A16) are satisfied. From (A15), it is straightforward to express $\gamma_1$ as:

$$\begin{aligned}\gamma_1 &= \frac{\mu_0}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\,[6s_{2t}z^* - 2\gamma_2] \\ &\quad + \frac{[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_0^2]}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \qquad \text{(A17)}\end{aligned}$$

If we introduce (A17) in (A16), we will obtain after some algebraic manipulations that:

$$\begin{aligned}[6s_{2t}z^* - 2\gamma_2]^2 &= \frac{(\mathbf{b}'\boldsymbol{\Sigma}\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}'\mathbf{b})^2}{\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2} \\ &\quad \times \left[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2\right]^2\end{aligned}$$

From condition (17) $\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2 \geq 0$, whereas $(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})$ is also non-negative because of the Cauchy-Schwarz inequality. Therefore, we can express $\gamma_2$ as:

$$\begin{aligned}\gamma_2 &= 3s_{2t}z^* \\ &\pm\frac{1}{2}\sqrt{\frac{(\mathbf{b}'\boldsymbol{\Sigma}\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2}{\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_0^2}}\left[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2\right],\end{aligned}$$

whence

$$\begin{aligned}\gamma_1 &= \frac{[3(s_{1t} + 3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \\ &\quad \times\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \pm \mu_{0t}\sqrt{\frac{(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2}{\sigma_0^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_0^2}}\right].\end{aligned}$$

If we introduce these expressions in (A14), we obtain the following "non-trivial" solutions:

$$\begin{aligned}z^* &= \mu_{0t}\frac{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \\ &\mp\frac{\sqrt{\left[(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2\right]\left[\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2\right]}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\end{aligned} \tag{A18}$$

There are potentially two other solutions characterised by $3(s_{1t}+3s_{2t}s_{3t})z^{*2}+3s_{2t}\sigma_{0t}^2 = 0$. However, it can be checked that those two solutions belong to the inefficient frontier mentioned in Proposition 5.

Finally, we obtain the required result by introducing (A18) in (A13). □

# B Third and fourth moments of a LSMN

Consider $\mathbf{w}_t \in \mathbb{R}^N$. Then,

$$\begin{aligned}E\left[(\mathbf{w}_t'(\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))^3 | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\tau}\right] &= vec(\mathbf{w}_t\mathbf{w}_t')'\boldsymbol{\Phi}_t(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{w}_t = \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}), \\ E\left[(\mathbf{w}_t'(\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))^4 | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\tau}\right] &= vec(\mathbf{w}_t\mathbf{w}_t')'\mathbf{K}_t(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{w}_t,\end{aligned}$$

where

$$\begin{aligned}
\mathbf{\Phi}_t(\boldsymbol{\theta},\boldsymbol{\tau}) &= E\left[vec\left[(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'\right](\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'|I_{t-1};\boldsymbol{\theta},\boldsymbol{\tau}\right] \\
&= s_{1t}vec\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta}) \\
&\quad + s_{2t}vec\left[\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right]\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta}) \\
&\quad + s_{2t}\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\otimes\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right],
\end{aligned}$$

$$\begin{aligned}
&\mathbf{K}_t(\boldsymbol{\theta},\boldsymbol{\tau}) = \\
&= E\left[vec\left[(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'\right]vec'\left[(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'\right]|I_{t-1};\boldsymbol{\theta},\boldsymbol{\tau}\right] \\
&= \kappa_{1t}vec\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]vec'\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right] \\
&\quad + \kappa_{2t}\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)\left(\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\otimes\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right) \\
&\quad + \kappa_{2t}\left[vec\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]vec'\left[\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right]+vec\left[\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right]vec'\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]\right] \\
&\quad + \kappa_{3t}\left[\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)\left(\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\otimes\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right)+vec\left(\mathbf{\Sigma}_t^*(\boldsymbol{\theta})\right)vec'(\mathbf{\Sigma}_t^*(\boldsymbol{\theta}))\right],
\end{aligned}$$

$\mathbf{K}_{NN}$ is the duplication matrix, and

$$\begin{aligned}
\kappa_{1t} &= \frac{E\left[\left(\xi^{-1}-\pi_1(\boldsymbol{\tau})\right)^4\right]}{\pi_1^4(\boldsymbol{\tau})}c^4(\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\boldsymbol{\tau}), \\
\kappa_{2t} &= \frac{E\left[\left(\xi^{-1}-\pi_1(\boldsymbol{\tau})\right)^2\xi^{-1}\right]}{\pi_1^3(\boldsymbol{\tau})}c^2(\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\boldsymbol{\tau}), \\
\kappa_{3t} &= \frac{\pi_2(\boldsymbol{\tau})}{\pi_1^2(\boldsymbol{\tau})}, \\
\mathbf{\Sigma}_t^*(\boldsymbol{\theta}) &= \mathbf{\Sigma}_t(\boldsymbol{\theta})+s_{3t}\mathbf{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\mathbf{\Sigma}_t(\boldsymbol{\theta}).
\end{aligned}$$

# C The Generalised Hyperbolic distribution

## C.1 The density function

If the mixing variable $\xi$ appearing in (1) follows a $GIG\,(-\nu,\gamma,\delta)$ distribution, then the density of the $N\times 1$ $GH$ random vector $\mathbf{u}$ will be given by

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

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

Given that $\delta$ and $\boldsymbol{\Upsilon}$ are not separately identified, Barndorff-Nielsen and Shephard (2001) set the determinant of $\boldsymbol{\Upsilon}$ equal to 1. However, it is more convenient to set $\delta = 1$ instead in order to reparametrise the $GH$ distribution so that it has mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_N$. Hence, if $\xi \sim GIG(-\nu, \gamma, 1)$, then $\boldsymbol{\tau} = (\nu, \gamma)'$, $\pi_1(\boldsymbol{\tau}) = R_\nu(\gamma)/\gamma$, and $c_v(\boldsymbol{\tau}) = \sqrt{D_{\nu+1}(\gamma) - 1}$, where $R_\nu(\gamma) = K_{\nu+1}(\gamma)/K_\nu(\gamma)$ and $D_{\nu+1}(\gamma) = K_{\nu+2}(\gamma) K_\nu(\gamma)/K^2_{\nu+1}(\gamma)$. It is then straightforward to use Proposition 1 to obtain a standardised $GH$ distribution.

One of the most attractive properties of the $GH$ distribution is that it contains as particular cases several of the most important multivariate distributions already used in the literature. For the standardised vector $\boldsymbol{\varepsilon}^*$, the most important ones are:

- **Normal**, which can be achieved in three different ways: (i) when $\nu \to -\infty$ or (ii) $\nu \to +\infty$, regardless of the values of $\gamma$ and $\boldsymbol{\beta}$; and (iii) when $\gamma \to \infty$ irrespective of the values of $\nu$ 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 $0 < \nu < \infty$ (see Madan and Milne, 1991).
- **Normal Inverse Gaussian**, for $\nu = -.5$ (see Aas, Dimakos, and Haff, 2005).
- **Hyperbolic**, for $\nu = 1$ (see Chen, Härdle, and Jeong, 2004)
- **Asymmetric Laplace**, for $\nu = 1$ and $\gamma = 0$ (see Cajigas and Urga, 2007).

## C.2 The score function

For our purposes, it is analytically convenient to replace $\nu$ and $\gamma$ by $\eta$ and $\psi$, where $\eta = -.5\nu^{-1}$ and $\psi = (1+\gamma)^{-1}$, although we continue to use $\nu$ and $\gamma$ in some equations for notational simplicity. In this context, the EM-type procedure that we follow is divided in two parts. In the maximisation step, we derive $l(\mathbf{y}_t|\xi_t, I_{t-1}; \boldsymbol{\phi})$ and $l(\xi_t|I_{t-1}; \boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$. Then, in the expectation step, we take the expected value of these derivatives given $I_T = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T\}$ and the parameter values.

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

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

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

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

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

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

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

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

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

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

which implies that

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

From here, we can use (C4) and (C5) to obtain the required moments. Specifically,

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

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

$$\frac{\partial l(\mathbf{y}_t|\, I_{t-1};\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} = -\frac{1}{2}vec'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} - f(I_T,\boldsymbol{\phi})\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\frac{\partial\mathbf{p}_t}{\partial\boldsymbol{\theta}'}$$
$$-\frac{1}{2}\frac{c_t(\boldsymbol{\phi})-1}{c_t(\boldsymbol{\phi})\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\sqrt{1+4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}vec'\left(\mathbf{b}\mathbf{b}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} + \mathbf{b}'\frac{\partial\mathbf{p}_t}{\partial\boldsymbol{\theta}'}$$
$$+\frac{1}{2}f(I_T,\boldsymbol{\phi})[\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\otimes\mathbf{p}_t'\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{*}(\boldsymbol{\phi})]}{\partial\boldsymbol{\theta}'}$$
$$-\frac{1}{2}\frac{g(I_T,\boldsymbol{\phi})}{\sqrt{1+4\left(D_{\nu+1}\left(\gamma\right)-1\right)\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}vec'\left(\mathbf{b}\mathbf{b}'\right)\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'},$$

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

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

and

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

where

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

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

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

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

and

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

## C.3 Modified Bessel function of the third kind

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

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

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

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

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

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

Finally, for large values of $x$ and $\nu$ we have that

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

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

- For $\nu \le 10$ and $|x| > 12$, the derivative of (C2) with respect to $\nu$ 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 $\nu$ 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 $\nu$ is near any non-negative integer $n = 0, 1, 2, \cdots$ due to the sine that appears in the denominator. In our experience, it is much better to use the following Taylor expansion for small $|\nu - n|$:

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

where for *integer* $\nu$:

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

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

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

and

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

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

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

where

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

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

where

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

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

where

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

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

where

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

and finally,

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

where

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

with

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

and

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

## C.4 Moments of the GIG distribution

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

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

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

$$E\left(X^k\right) = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{\nu+k}(\delta\gamma)}{K_\nu(\delta\gamma)} \tag{C4}$$

$$E(\log X) = \log\left(\frac{\delta}{\gamma}\right) + \frac{\partial}{\partial\nu} K_\nu(\delta\gamma). \tag{C5}$$

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 (C4) becomes infinite for $k \geq |\nu|$. A complete discussion on this distribution can be found in Jørgensen (1982), who also presents several useful Gaussian approximations based on the following limits:

$$\begin{aligned} &\sqrt{\delta\gamma}[(\gamma x/\delta) - 1] \overset{\delta\gamma\to\infty}{\to} N(0,1) \\ &\sqrt{\delta\gamma}\log(\gamma x/\delta) \overset{\delta\gamma\to\infty}{\to} N(0,1) \\ &\frac{\gamma^2}{2\sqrt{\nu}}\left[x - \frac{2\nu}{\gamma^2}\right] \overset{\nu\to+\infty}{\to} N(0,1) \\ &\frac{-2\nu^{3/2}}{\delta^2}\left[x + \frac{\delta^2}{2\nu}\right] \overset{\nu\to-\infty}{\to} N(0,1) \end{aligned}$$

**Table 1**

Maximum likelihood estimates of a conditionally heteroskedastic single factor model for ten Datastream sectoral indices for the US

| | Ten indices | | Extended model | |
|---|---|---|---|---|
| Parameter | | SE | | SE |
| $\eta$ | 0.095 | 0.004 | 0.091 | 0.003 |
| $\psi$ | 1 | - | 1 | - |
| **b** | | | | |
| Basic Materials | -0.100 | 0.038 | -0.088 | 0.040 |
| Consumer Goods | 0.068 | 0.066 | 0.053 | 0.070 |
| Consumer Services | 0.077 | 0.091 | 0.093 | 0.091 |
| Financials | 0.009 | 0.052 | 0.048 | 0.050 |
| Health Care | -0.033 | 0.078 | -0.082 | 0.083 |
| Industrials | -0.096 | 0.084 | -0.080 | 0.089 |
| Oil and Gas | 0.116 | 0.056 | 0.130 | 0.058 |
| Technology | -0.091 | 0.066 | -0.092 | 0.066 |
| Telecommunications | 0.067 | 0.074 | 0.062 | 0.082 |
| Utilities | -0.027 | 0.037 | -0.034 | 0.042 |
| World ex-US | | | -0.163 | 0.052 |

Note: Extended model denotes the model based on the ten US indices and the World ex-US index.

**Table 2:**

Spanning tests. Improvement in the investment opportunity set caused by the introduction of the World ex-US index

| Null hypothesis | Wald | | LR | |
|---|---|---|---|---|
| | | p-value | | p-value |
| Mean-variance efficiency | 1.00 | 0.317 | 1.05 | 0.306 |
| Skewness-variance efficiency | 9.64 | 0.002 | 9.79 | 0.002 |
| Joint | 13.57 | 0.001 | 13.72 | 0.001 |

Notes: The mean-variance efficiency test denotes a test of the null hypothesis $\mu_{2t}(\boldsymbol{\theta}) = \mathbf{d}'_{12t}\boldsymbol{\mu}_{1t}(\boldsymbol{\theta})$, where $\boldsymbol{\mu}_{1t}(\boldsymbol{\theta})$ and $\mu_{2t}(\boldsymbol{\theta})$ denote, respectively, the vector of expected excess returns of the 10 US indices and the expected excess return of the World ex-US index, while $\mathbf{d}_{12t}$ denotes the coefficients of the conditional regression of the excess returns of the World ex-US index on those of the 10 US sectoral indices. The skewness-variance efficiency test denotes a test of the null hypothesis that the element of the skewness vector $\mathbf{b}$ corresponding to the World ex-US index is zero.

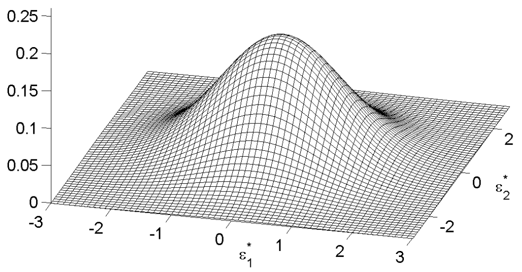



Figure 1a: Standardised bivariate normal density

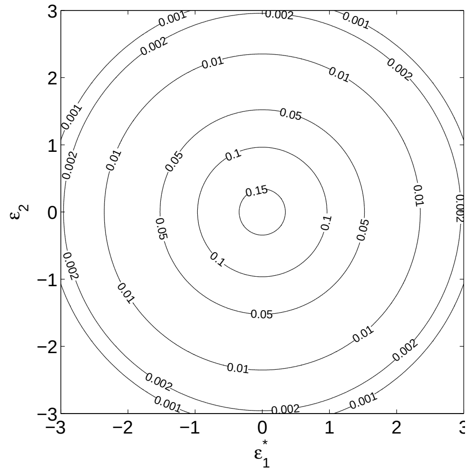



Figure 1b: Contours of a standardised bivariate normal density

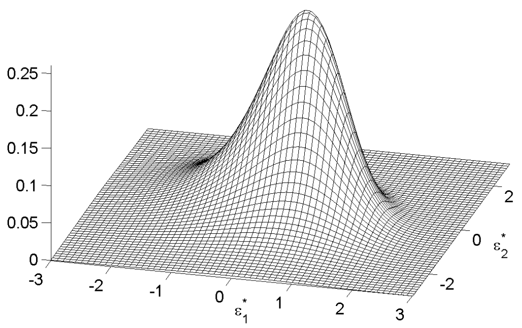



Figure 1c: Standardised bivariate asymmetric Student $t$ density with 10 degrees of freedom ($\eta = .1$) and $\boldsymbol{\beta} = (-3, -3)'$

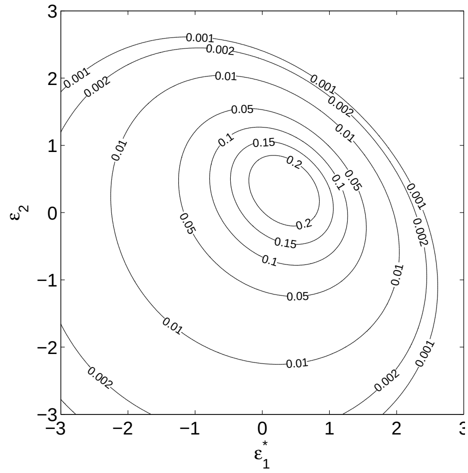



Figure 1d: Contours of a standardised bivariate asymmetric Student $t$ density with 10 degrees of freedom ($\eta = .1$) and $\boldsymbol{\beta} = (-3, -3)'$

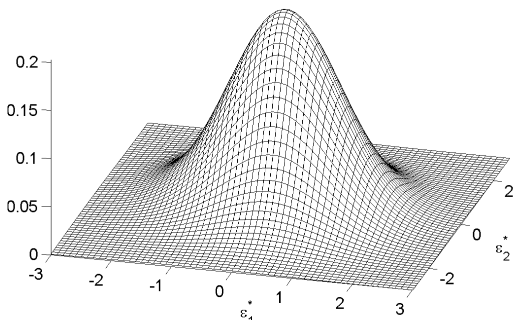



Figure 1e: Standardised bivariate LMSN with a Bernoulli mixing variable and $\boldsymbol{\beta} = (-3, -3)'$

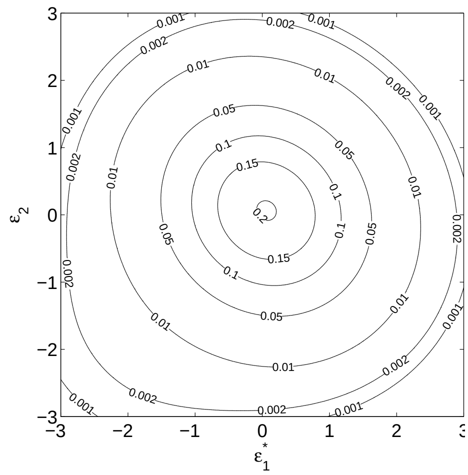
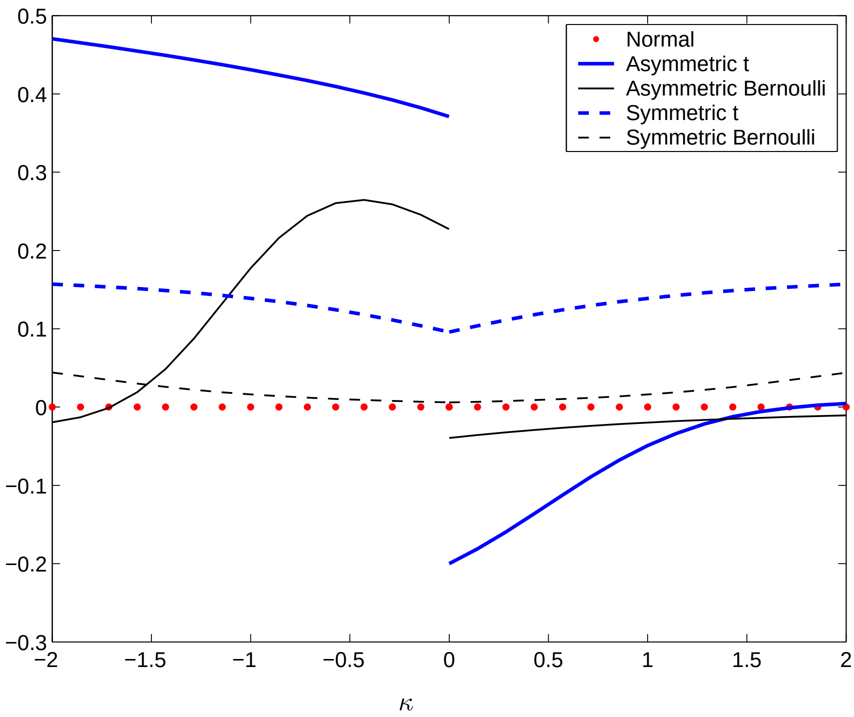



Figure 1f: Contours of a standardised bivariate LMSN with a Bernoulli mixing variable and $\boldsymbol{\beta} = (-3, -3)'$

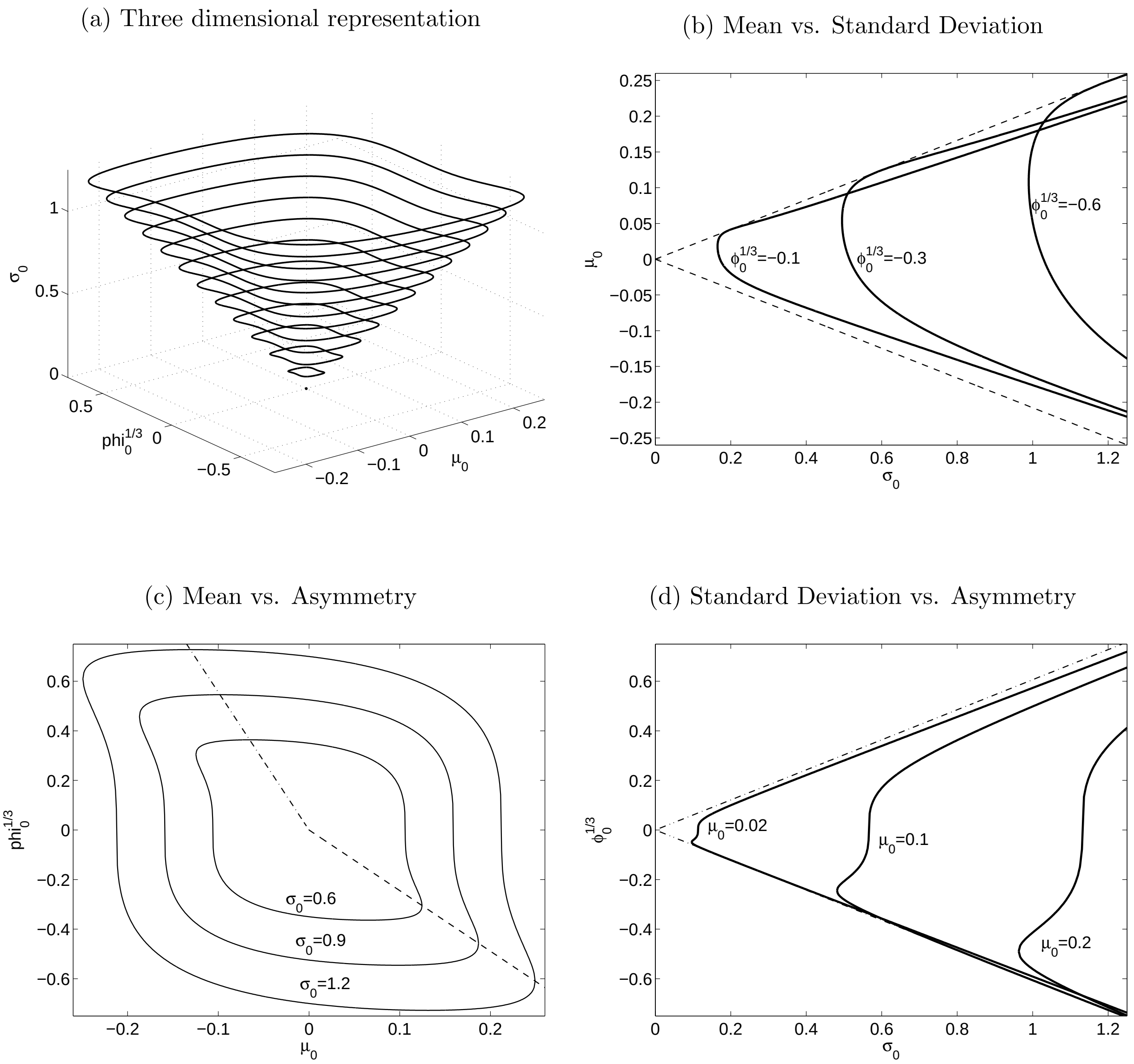
Notes: The Bernoulli mixing variable of Figures 1e and 1f is such that it has mean $E(\xi) = 1$ and $Pr(\xi = 0.6) = 0.04$.

Figure 2: Exceedance correlation for symmetric and asymmetric location-scale mixtures of normals




Notes: The exceedance correlation between two variables $\varepsilon_1^*$ and $\varepsilon_2^*$ is defined as $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \kappa, \varepsilon_2^* > \kappa)$ for positive $\kappa$ and $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* < \kappa, \varepsilon_2^* < \kappa)$ for negative $\kappa$ (see Longin and Solnik, 2001). Symmetric $t$ distribution with 10 degrees of freedom ($\eta = .1$) and Asymmetric $t$ distribution with $\eta = .1$ and $\beta = (-3, -3)$. Asymmetric Bernoulli denotes a location-scale mixture of normals with $\beta = (-3, -3)$ and mixing variable such that it has mean $E(\xi) = 1$ and $Pr(\xi = 0.6) = 0.04$.

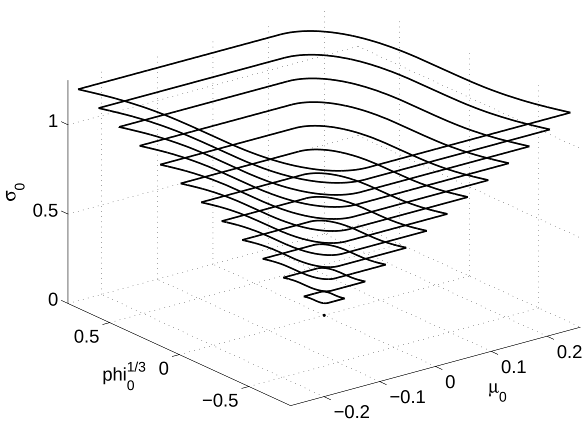
Figure 3: Mean-Variance-Skewness frontier of a LSMN. Example 1.



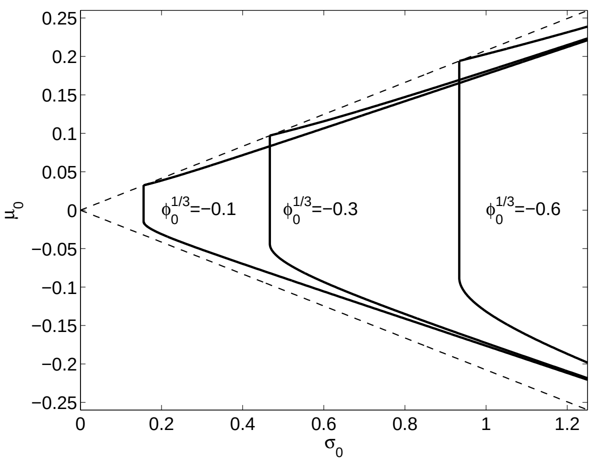

Notes: The mean-variance frontier is plotted with dotted lines, while dash-dot lines are used for the skewness-variance frontier.

Figure 4: Mean-Variance-Skewness frontier of a LSMN. Example 2.

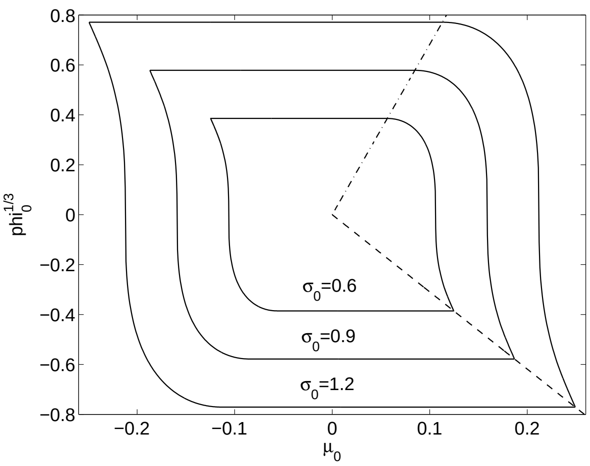
(a) Three dimensional representation



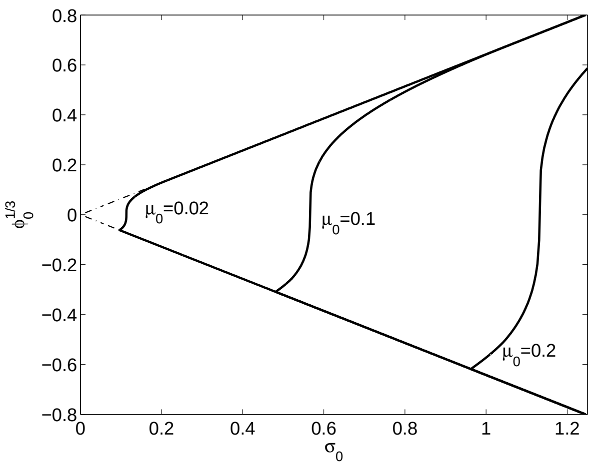

(b) Mean vs. Standard Deviation




(c) Mean vs. Asymmetry




(d) Standard Deviation vs. Asymmetry




Notes: The mean-variance frontier is plotted with dotted lines, while dash-dot lines are used for the skewness-variance frontier.

Figure 5a: Sampling distribution of the log of $vech'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vech[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$

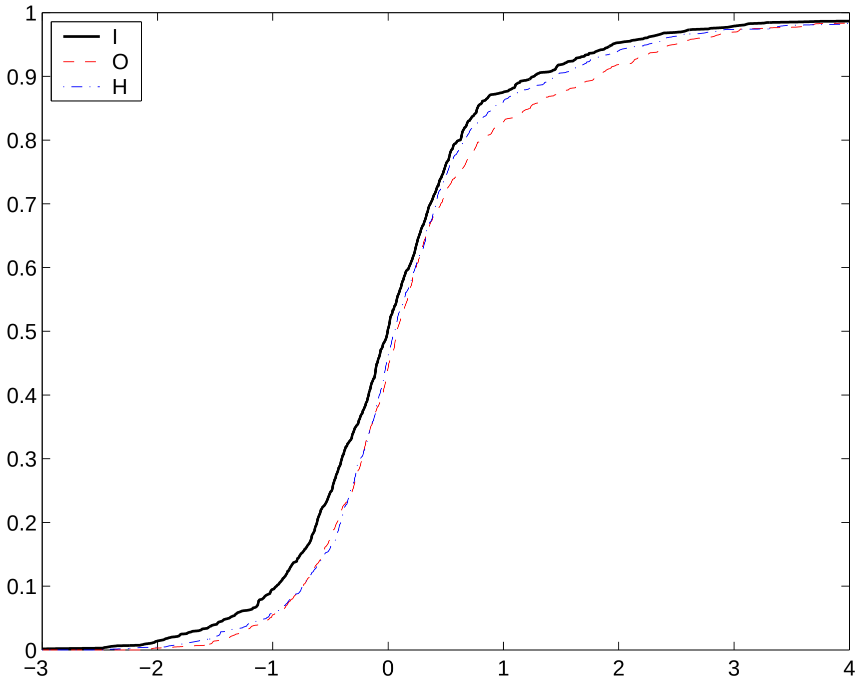
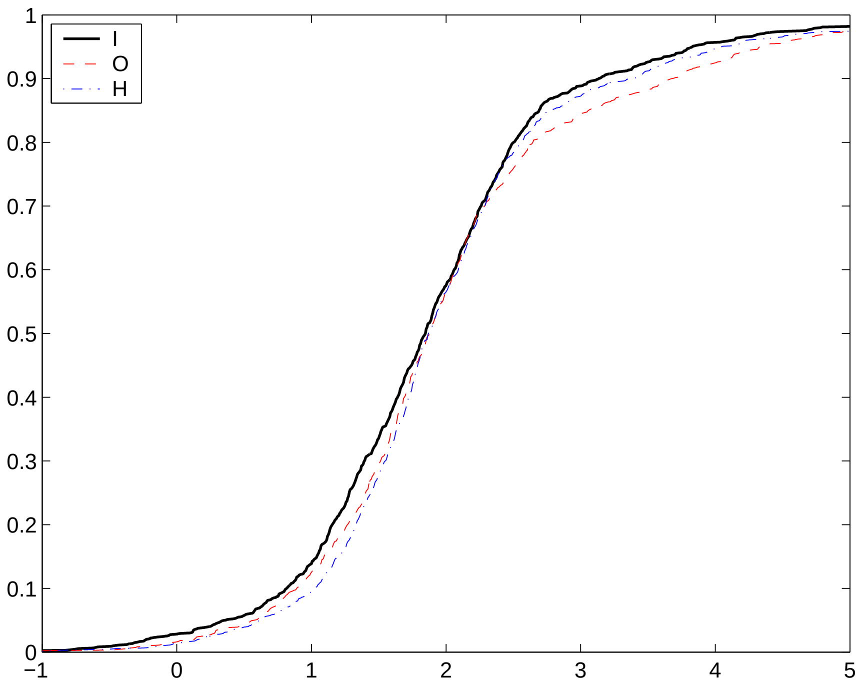



Figure 5b: Sampling distribution of the log of $vecd'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vecd[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$




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 6a: Comparison of the univariate and multivariate Gaussian estimates of the equally weighted portfolio of the US Datastream sectoral indices

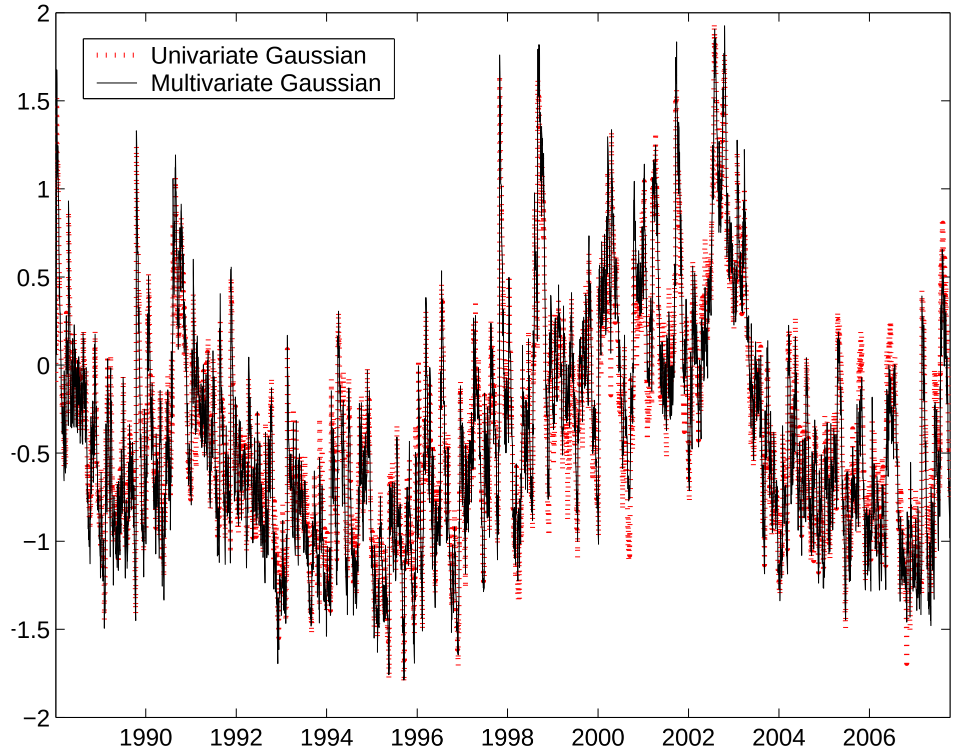



Figure 6b: Comparison of the multivariate Gaussian and *GH* estimates of the equally weighted portfolio of the US Datastream sectoral indices

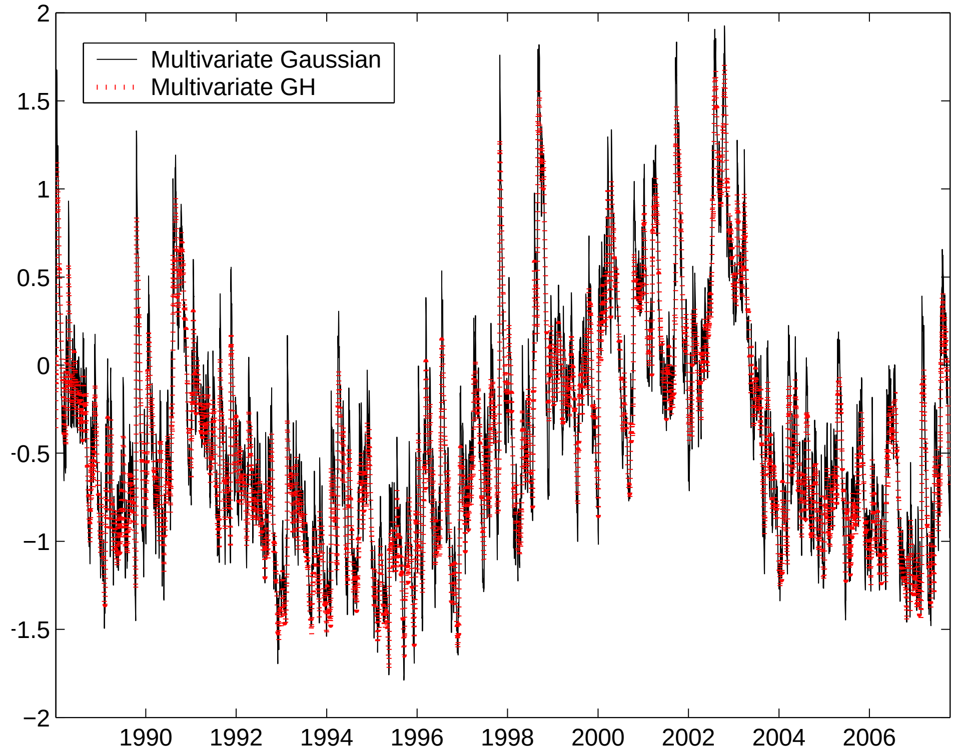
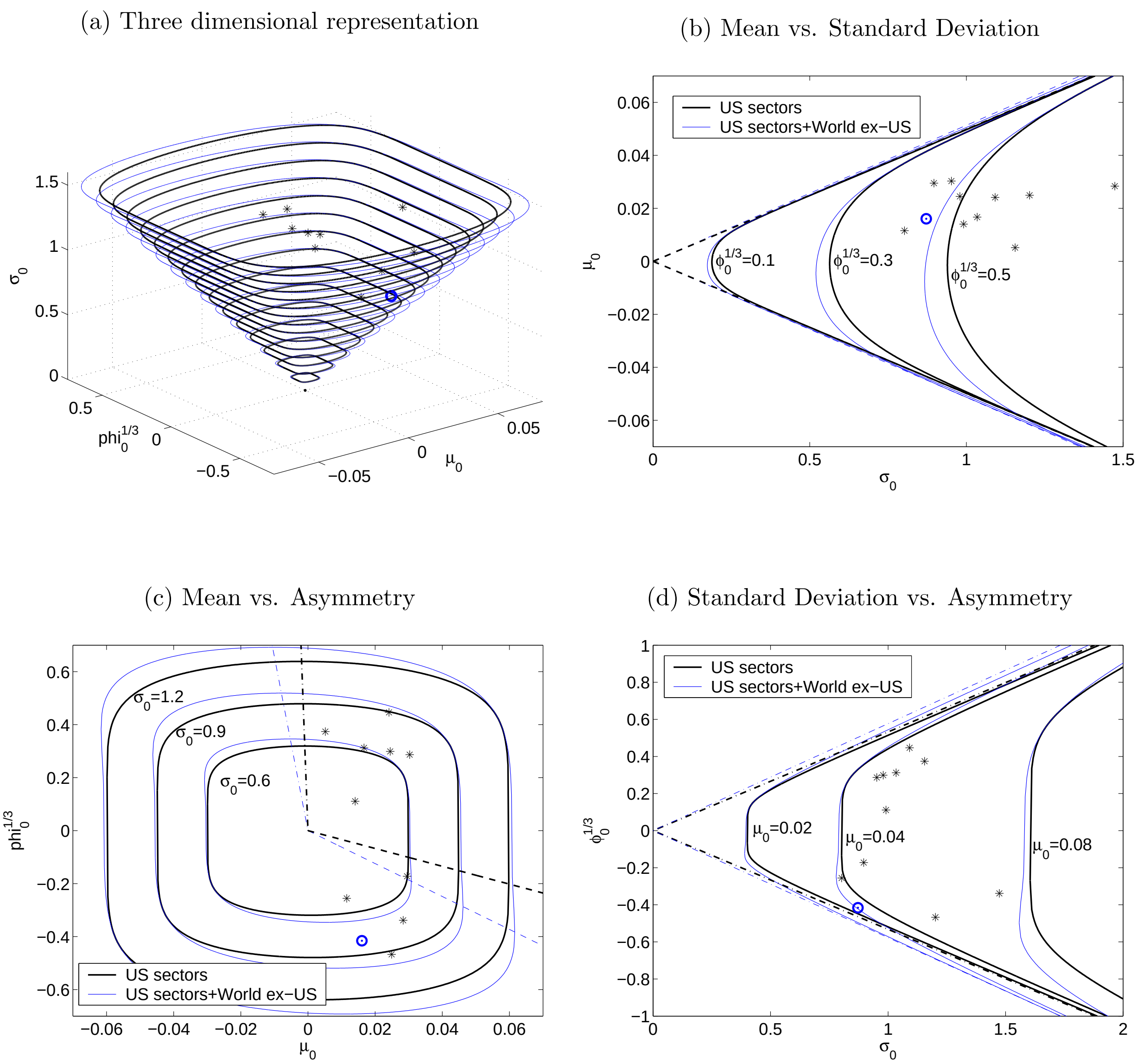


Figure 7: Mean-Variance-Skewness frontier of the the US Datastream indices and change induced by adding the World ex-US Datastream index




Notes: Frontier of excess returns with moments expressed in percent terms. Thick lines represent the contours obtained from the ten US indices, while the contours of the frontier obtained including the World ex-US index are represented with thin lines. The mean-variance frontier is plotted with dotted lines, while dash-dot lines are used for the skewness-variance frontier. Asterisks (circle) are used to plot the positions of the individual US indices (World ex-US index). The results correspond to a representative day whose mean vector and covariance matrix are set to their unconditional values.