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

Javier Mencía, Enrique Sentana

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

Javier Mencía, Enrique Sentana

PII: S0304-4076(09)00126-2
DOI: 10.1016/j.jeconom.2009.05.001
Reference: ECONOM 3203

To appear in: *Journal of Econometrics*

Received date: 2 March 2008
Revised date: 20 April 2009
Accepted date: 6 May 2009

Please cite this article as: Mencía, J., Sentana, E., Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.05.001

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

Javier Mencía
Bank of Spain, Alcalá 48, E-28014 Madrid, Spain
<javier.mencia@bde.es>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

January 2008
Revised: March 2009

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, Sortino Ratio, Spanning Tests, Tail Dependence.

JEL: C52, C32, G11

*Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain. Tel.: +34 91 429 0551, fax: +34 91 429 1056, e-mail: sentana@cemfi.es.

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 Chamberlain, 1983; Owen and Rabinovitch, 1983; Berk, 1997; as well as Ross, 1978 for a related discussion).

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). Unfortunately, it is rather difficult to obtain general results for mean-variance-skewness frontiers unless one introduces some structure in the $N(N+1)(N+2)/6$ non-redundant third moments, where N is the cross-sectional dimension. In this sense, Athayde and Flôres (2004) derive several useful properties of mean-variance-skewness frontiers, and obtain their shape for some examples by simulation techniques. Similarly, Briec, Kerstens, and Jokung (2007) propose an optimisation algorithm that, starting from a specific portfolio, obtains the efficient mean-variance-skewness portfolio along a given direction (see also Jurczenko, Maillet, and Merlin, 2006).

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 includes other well known and empirically relevant special cases, such as symmetric and asymmetric versions of the Hyperbolic (Chen, Härdle, and Jeong, 2008), Normal Gamma (Madan

and Milne, 1991), Normal Inverse Gaussian (Aas, Dimakos, and Haff, 2005) and Multivariate Laplace distributions (Cajigas and Urga, 2007).[1] 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 parameter. Therefore, the differences between any two portfolios can be fully explained by the discrepancies between those three parameters.

Then, we analyse these moments to characterise the feasible investment opportunity set, and obtain the mean-variance-skewness frontier 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. Moreover, we show that one can replace this last fund by the portfolio that maximises the Sortino ratio. Lastly, we show that we can continue to span the frontier with three funds when no risk-free asset exists.

For practical purposes, we study several aspects related to the maximum likelihood estimation of a general multivariate 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 including additional assets in our

[1] Barndorff-Nielsen and Shephard (2001) also use the *GH* distribution to capture the unconditional distribution of returns on assets whose price dynamics are generated by continuous time stochastic volatility models in which the instantaneous volatility follows an Ornstein-Uhlenbeck process with Lévy innovations.

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. For the sake of robustness, we check our results using two non-nested members of the *LSMN* family: the *GH* distribution, and a location-scale mixture of two Gaussian vectors.

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 ξ is an independent positive mixing variable. As usual, the conditional normality given the mixing variable could be justified by appealing to the central limit theorem. For the sake of concreteness, we will denote the distribution function of ξ as $F(\cdot;\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is a vector of q shape parameters. For instance, ξ will be a Generalised Inverse Gaussian variate in the *GH* case (see Jørgensen, 1982). Since $\mathbf{u}$ given ξ is Gaussian with conditional mean $\boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. 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 ξ^{-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, ξ forces the realisations of all the elements in $\boldsymbol{\varepsilon}^*$ to be very large in magnitude when it takes very small values, which introduces dependence in the tails of the distribution. In addition, we can make this dependence stronger in certain regions by choosing $\boldsymbol{\beta}$ 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 and a *LSMN* whose mixing variable is Bernoulli.[2] 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,

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

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. However, it is important to mention that while the *GH* distribution can display non-zero values at the extreme tails, any finite *LSMN* cannot be distinguished from the Gaussian distribution in terms of its extreme value exceedance correlation, as Figure 2 illustrates.

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).

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 compute in closed form the Value at Risk, Expected Shortfall or Marginal Value at Risk of any portfolio from the parameters of the joint distribution (see Jorion, 2006, for formal definitions of these statistics). Finally, Proposition 2 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

Let $\mathbf{y}_t$ be a vector of excess returns on N risky assets. To accommodate flexible specifications, we assume that those excess returns are generated by the following con-

ditionally 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 $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is a symmetric or lower triangular matrix, nor on the order of the observed variables in the latter case. This can be achieved by making $\boldsymbol{\beta}$ a function of past information and a new vector of parameters $\mathbf{b}$ in the following way:

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

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 $\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\boldsymbol{\theta})$.

3 Portfolio allocation

3.1 The investor's problem

Consider an investor whose wealth at time $t-1$ is A_{t-1}. Assuming that a risk-free asset with returns r_t exists (see Appendix A for the no riskless asset case), 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 $\mathbf{w}_t$ is the vector of allocations to the risky assets, which together with r_t are known at $t-1$. In this context, we can show the following property for any portfolio of *LSMN* returns:

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 $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, 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 parameter $\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$.*

Proposition 3 implies that, from an investor's point of view, the relative attractiveness of any two portfolios can always be explained by their expected returns, variances and skewness because the higher order moments depend on the lower ones and the common tail parameters $\boldsymbol{\tau}$. Hence, we only need to characterise the investment opportunity set in terms of these moments to fully describe the investor's available strategies. This in turn involves studying the mean-variance-skewness frontier, which limits the feasible combinations of the first three moments of portfolios. In this sense, it is straightforward to show that the expected return of A_t can be expressed as $A_{t-1}[1 + r_t + \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})]$ while the conditional variance is $A_{t-1}^2\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t$. As for the third centred moment, we can use the results in Appendix D to show that it can be expressed as A_{t-1}^3 times

$$\varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) = \left(s_{1t} + 3s_{2t}s_{3t}\right)\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^3 + 3s_{2t}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right], \tag{5}$$

where

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

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.

3.2 Mean-variance-skewness frontiers

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

$$\mu_{0t} = \sigma_{0t}\sqrt{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}, \tag{6}$$

which we obtain by maximising expected return μ_{0t} for every possible standard deviation σ_{0t}. As is well known, the mean-variance frontier (6) can be spanned by just two funds: the risk-free asset and a portfolio with weights proportional to $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$.

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

$$\max_{\mathbf{w}_t \in \mathbb{R}^N} \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) \quad \text{s.t.} \quad \begin{cases} \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mu_{0t} \\ \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2 \end{cases} \tag{7}$$

Obviously, there are other equivalent approaches to obtain this frontier. For instance, Athayde and Flôres (2004) maximise expected returns subject to constraints on the variance and asymmetry. However, we prefer the formulation in (7) because it is straightforward to ensure the feasibility of the target expected return and variance. Specifically, we know that any feasible portfolio must satisfy

$$\frac{\mu_{0t}}{\sigma_{0t}} \leq \sqrt{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}, \tag{8}$$

since its Sharpe ratio cannot be greater than the Sharpe ratio corresponding to the mean-variance frontier (6).

We can solve (7) by forming the Lagrangian

$$\mathcal{L} = \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) + \gamma_1 \left[\mu_{0t} - \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] + \gamma_2 \left[\sigma_{0t}^2 - \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right], \tag{9}$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}_t} = &\left\{3(s_{1t} + 3s_{2t}s_{3t})\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]^2 + 3s_{2t}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\right\}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} \\ &+6s_{2t}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t - \gamma_1\boldsymbol{\mu}_t(\boldsymbol{\theta}) - 2\gamma_2\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t. \end{aligned} \tag{10}$$

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

Proposition 4 *The efficient mean-variance-skewness portfolios that solve (10) can be expressed as either*

$$\mathbf{w}_{1t}^{*} = \frac{\mu_{0t} + \Delta_t^{-1}(\mu_{0t}, \sigma_{0t})\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{1}{\Delta_t(\mu_{0t}, \sigma_{0t})}\mathbf{b}, \tag{11}$$

or

$$\mathbf{w}_{2t}^{*} = \frac{\mu_{0t} - \Delta_t^{-1}(\mu_{0t}, \sigma_{0t})\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) + \frac{1}{\Delta_t(\mu_{0t}, \sigma_{0t})}\mathbf{b}, \tag{12}$$

where

$$\Delta_t(\mu_{0t}, \sigma_{0t}) = \sqrt{\frac{(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}\right)^2}{\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2}}. \tag{13}$$

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 a portfolio with weights $\mathbf{b}$. This second vector can be interpreted as an asymmetry-variance efficient portfolio, because we can maximise asymmetry for a given standard deviation by considering portfolios with weights proportional to $\mathbf{b}$ (see Appendix B for details). Hence, Proposition 4 shows that the efficient region of the mean-variance-skewness frontier can be spanned by the aforementioned three funds.

In order to obtain an explicit equation for the frontier in mean-variance-skewness space, let $j = -1, +1$ and define $\varphi_{0t}(j)$ as the third centred moment that results from introducing (11) or (12) in (5), respectively. It is straightforward to show that $\varphi_{0t}(j)$ can be expressed as:

$$\begin{aligned}
\varphi_{0t}(j) &= (s_{1t} + 3s_{2t}s_{3t})h_{1t}(4h_{1t}^2 - 3h_{2t})\mu_{0t}^3 \\
&+3\left\{(s_{1t} + 3s_{2t}s_{3t})h_{1t}(h_{2t} - h_{1t}^2)\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] + s_{2t}h_{1t}\right\}\mu_{0t}\sigma_{0t}^2 \\
&+j\sqrt{(h_{2t} - h_{1t}^2)\{\sigma_0^2\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] - \mu_{0t}^2\}} \\
&\times\left(\begin{array}{c}(s_{1t} + 3s_{2t}s_{3t})(4h_{1t}^2 - h_{2t})\mu_{0t}^2 \\ +\left\{(s_{1t} + 3s_{2t}s_{3t})(h_{2t} - h_{1t}^2)\left[\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] + 3s_{2t}\right\}\sigma_{0t}^2\end{array}\right)
\end{aligned} \tag{14}$$

where

$$h_{1t} = \frac{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \text{ and } h_{2t} = \frac{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}.$$

It is not difficult to show that (14) 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

[3] In order to assess whether (11) or (12) yields the efficient part of the frontier, we can either check the second order conditions or simply choose the solution with the highest asymmetry, which can be computed using (14). But if (8) is satisfied with equality, which only occurs on the mean variance frontier, then we can show that $\mathbf{w}_{1t}^{*} = \mathbf{w}_{2t}^{*}$ and $\varphi_{0t}(-1) = \varphi_{0t}(1)$.

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 σ_{0t} is proportional to μ_{0t}, $\sigma_{0t} = k'\mu_{0t}$ say, then the cubic root of the asymmetry will also be proportional to $|\mu_{0t}|$ along this direction.

In addition, we can show the convexity of the mean-variance-skewness frontier in terms of portfolio weights.

Proposition 5 *If the conditional distribution of $\mathbf{y}_t$ is a LSMN, then any linear combination of two mean-variance-skewness efficient portfolios will also be an efficient mean-variance-skewness portfolio.*

Figure 3 shows the shape of the mean-variance-skewness frontier for an example with five risky assets. 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. In contrast, the asymmetry-variance frontier, which contains those portfolios that yield maximum asymmetry for given values of standard deviation and whose weights are proportional to $\mathbf{b}$, is on the northern half. These relative positions hold in general because, for a given variance, the mean-variance frontier contains the points with highest expected return, which is displayed on the x-axis, while the asymmetry-variance frontier maximises skewness (on the y-axis). Finally, we consider the skewness-variance space in Figure 3d, where we can observe the linearity of the skewness-variance frontier (see Appendix B for details).

We have also plotted in this figure the portfolios chosen by investors who minimise the second lower partial moment,

$$S(\mathbf{w}_t) = \sqrt{E_{t-1}\left[\min(0, \mathbf{w}_t'\mathbf{y}_t)^2\right]},$$

for any given target return. The portfolios that solve this problem maximise the Sortino ratio, which is defined as $\mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})/S(\mathbf{w}_t)$ (see e.g. Pedersen and Satchell, 2002). We

can span those portfolios by scaling any of them with a *positive* scalar (see Hogan and Warren, 1974). In the case of *LSMN*, we can show the following additional property.

Proposition 6 *If the conditional distribution of* $\mathbf{y}_t$ *is a LSMN, then the portfolios that maximise the Sortino ratio will be along a straight line on the mean-variance-skewness frontier.*

By combining Propositions 5 and 6, we can show that the mean-variance-asymmetry frontier can also be spanned if we combine a portfolio that maximises the Sortino ratio with the risk-free asset and a mean-variance efficient portfolio. In addition, Proposition 6 implies that we can restrict our attention to the frontier portfolios in Proposition 4 in order to maximise the Sortino ratio. This feature makes the Sortino ratio maximisation much easier to handle, since it only requires choosing the free parameter σ_{0t}^2 for a given target expected return μ_{0t}, regardless of N. As we can see in Figure 3c, this type of investors would choose portfolios with higher (positive) skewness than mean-variance investors for the same variance.

The departure from the mean-variance solution is even more remarkable in Figure 4, which shows another example with five assets under a more asymmetric parameter configuration. Notice that in this case the the iso-variance contours have a flat region with maximum constant skewness. However, we can arbitrarily define the asymmetry-variance frontier as the line with highest expected returns. Interestingly, it can be shown that the frontier in terms of the first two moments and the asymmetry parameter $\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t$ is a cone whose iso-variance contours are ellipses.

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

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

where $f(\mathbf{y}_t|\xi_t, I_{t-1}; \boldsymbol{\phi})$ is the Gaussian likelihood of $\mathbf{y}_t$ given ξ_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 ξ_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 ξ_t, I_{t-1} and $\boldsymbol{\phi}$; $l(\xi_t|\mathbf{y}_t, I_{t-1}; \boldsymbol{\phi})$ is the conditional log-likelihood of ξ_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:

$$s_t(\boldsymbol{\phi}) = E\left(\frac{\partial l(\mathbf{y}_t|\xi_t, I_{t-1}; \boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\middle| I_T; \boldsymbol{\phi}\right) + E\left(\frac{\partial l(\xi_t|I_{t-1}; \boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\middle| I_T; \boldsymbol{\phi}\right) \tag{15}$$

because $E\left[\partial l(\xi_t|\mathbf{y}_t, I_{t-1}; \boldsymbol{\phi})/\partial \boldsymbol{\phi}\middle| I_T; \boldsymbol{\phi}\right] = \mathbf{0}$ 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] In many cases of practical interest, it is straightforward to compute the required expectations. For example, if the distribution of excess returns is *GH*, then the distribution of ξ_t given the full observed sample will be a Generalised Inverse Gaussian (*GIG*) with parameters,

$$GIG\left(\frac{N}{2}-\nu, \sqrt{\frac{\gamma c(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\nu,\gamma)}{R_\nu(\gamma)}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}+\gamma^2}, \sqrt{\frac{R_\nu(\gamma)}{\gamma}\mathbf{p}_t'(\boldsymbol{\phi})\boldsymbol{\Sigma}_t^{*-1}(\boldsymbol{\phi})\mathbf{p}_t(\boldsymbol{\phi})+1}\right),$$

where

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

$\mathbf{p}_t(\boldsymbol{\phi}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}) + c(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\nu,\gamma)\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$, $c(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\nu,\gamma)$ is defined in (2) with $c_v^2(\nu,\gamma) = D_{\nu+1}(\gamma)-1$, $R_\nu(\gamma) = K_{\nu+1}(\gamma)/K_\nu(\gamma)$, $D_{\nu+1}(\gamma) = K_{\nu+2}(\gamma)K_\nu(\gamma)/K_{\nu+1}^2(\gamma)$ and $K_\nu(\cdot)$ is the modified Bessel function of the third kind (see Abramowitz and Stegun, 1965).

Analogously, if ξ_t is multinomial, then $\xi_t|I_T;\phi$ will also be multinomial, where the probability for each possible value can be easily obtained using Bayes rule.

4.2 The information matrix

Given correct specification, the results in Crowder (1976) imply that under standard regularity conditions,[5] the score vector $s_t(\boldsymbol{\phi})$ evaluated at $\boldsymbol{\phi}_0$ has the martingale difference property and consequently the ML estimator will be asymptotically normally distributed with a covariance matrix which is the inverse of the usual information matrix

$$\mathcal{I}(\boldsymbol{\phi}_0) = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}s_t(\boldsymbol{\phi}_0)s_t'(\boldsymbol{\phi}_0) = E[s_t(\boldsymbol{\phi}_0)s_t'(\boldsymbol{\phi}_0)]. \tag{16}$$

In general, though, (16) cannot be obtained in closed form.[6] The simplest consistent

[4] It is possible to show that $\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*/N$ converges in mean square to $1/[\pi_1(\boldsymbol{\tau})\xi_t]$ as $N\to\infty$. This means that in the limit the latent variable ξ_t could be fully recovered from observations on $\mathbf{y}_t$, which would greatly simplify the calculations implicit in (15).

[5] In particular, Crowder (1976) requires: (i) $\boldsymbol{\phi}_0 \in \text{int}\,\boldsymbol{\Phi}$ is locally identified, where $\boldsymbol{\Phi}$ is a bounded subset of $\mathbb{R}^{p+q+N}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\phi}_0$; (iii) there is uniform convergence of the integrals involved in the computation of the mean vector and covariance matrix of $s_t(\boldsymbol{\phi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \partial s_t(\boldsymbol{\phi})\partial\boldsymbol{\phi}\right]T^{-1}\sum_t \partial s_t(\boldsymbol{\phi})\partial\boldsymbol{\phi} \xrightarrow{p} \mathbf{I}_{p+q+N}$, where $E^{-1}\left[-T^{-1}\sum_t \partial s_t(\boldsymbol{\phi})\partial\boldsymbol{\phi}\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$.

[6] 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).

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 (16) 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

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

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:

$$\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{c}\mathbf{c}'\lambda_t + \boldsymbol{\Gamma}_t, \end{array} \tag{17}$$

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

$$\lambda_t = \alpha_0 + \alpha_1(f_{t-1|t-1}^2 + \omega_{t-1|t-1}) + \alpha_2\lambda_{t-1}, \tag{18}$$

$$\gamma_{it} = \phi_0 + \phi_1\left[(y_{it-1} - \mu_i - c_i f_{t-1|t-1})^2 + c_i^2\omega_{t-1|t-1}\right] + \phi_2\gamma_{it-1}, \quad i = 1,2,3, \tag{19}$$

$f_{t|t} = \omega_{t|t}\mathbf{c}'\boldsymbol{\Gamma}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))$ 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 (18) would be the variance of the latent factor, while (19) would correspond to the idiosyncratic effects. In this sense, note

that the conditional correlation coefficient between any two elements i and j of $\mathbf{y}_t$ is given by

$$\rho_{ijt} = \frac{c_i c_j \lambda_t}{\sqrt{c_i^2 \lambda_t + \gamma_{it}}\sqrt{c_j^2 \lambda_t + \gamma_{jt}}},$$

which implies that periods when the volatility of the unobservable factor rises are also those when, ceteris paribus, individual assets exhibit greater intercorrelation (see King, Sentana, and Wadhwani, 1994).

As for parameter values, we have chosen $\mu_i = .2$, $c_i = 1$, $\alpha_1 = \phi_1 = .1$, $\alpha_2 = \phi_2 = .85$, $\alpha_0 = 1 - \alpha_1 - \alpha_2$ and $\phi_0 = 1 - \phi_1 - \phi_2$.

We 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.[7] 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 $T = 1,000$ observations each (results for other sample sizes are available on request). 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 $vecd'[V_T^E(\hat{\boldsymbol{\phi}}_T) - V_T(\hat{\boldsymbol{\phi}}_T)]vecd[V_T^E(\hat{\boldsymbol{\phi}}_T) - V_T(\hat{\boldsymbol{\phi}}_T)]$, where E is either O, H or I.[8] 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. Finally, note that our simulation-based approach to evaluate the information matrix is not only valid under our *LSMN* assumption, but it can also be applied more generally to any parametric time series model.

5 Empirical application

We now apply the methodology derived in the previous sections to the ten Datastream main sectoral indices for the US.[9] Specifically, our dataset consists of daily excess returns

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

[8] 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.

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

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 (17)), 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 (2009), who find that these US indices are significantly asymmetric and leptokurtic even after controlling for volatility clustering. 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. In addition, we assess the robustness of our results by considering a *LSMN* with a Bernoulli mixing variable as well. In both cases, we derive the score following the approach described in section 4.1 and compute the standard errors by simulation as explained in section 4.2.

The first column of Table 1a shows the estimates of the asymmetry parameters for the *GH* distribution. Although not all of the asymmetry parameters are individually significant, 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 score-based symmetry test is 25.35 (p-value=0.005). Note that the null of this test allows for fat tails in the conditional distribution (see Mencía and Sentana, 2009). The first column of Table 1b shows that the discrete mixing variable yields similar asymmetry parameters, although the overall fit is worse in this case, as the smaller log-likelihood confirms.

One potential concern is whether we are able to correctly capture the dynamics of the 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 have compared the estimates of the con-

ditional 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, 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, we also find that the (log) standard deviations implied by the two distributional assumptions for the equally weighted portfolio are extremely similar (see Mencía and Sentana, 2009, for further details).

Figure 6 shows the mean-variance-skewness frontier that we obtain for the US indices with the *GH* distribution. 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 in Figure 6c that a mean-variance investor would implicitly choose portfolios with negative asymmetry. In this sense, the rather vertical shape of the iso-variance contours around the mean-variance optimal line indicate that, for a given variance, mean-variance investors could obtain important gains in skewness in exchange for only minor reductions in expected returns. We have also plotted the line that maximises the Sortino ratio. This line is located in the sector between the mean-variance and asymmetry-variance frontiers, although much closer to the first one.

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

$$\mu_{2t}(\boldsymbol{\theta}) = \mathbf{d}'_{12t}(\boldsymbol{\theta})\boldsymbol{\mu}_{1t}(\boldsymbol{\theta}), \tag{20}$$

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 (20) 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 (11) and (12)). We analyse these two effects in Table 2 by means of Wald and LR tests for the two distributions that we consider. While we are unable to reject the mean-variance spanning restriction (20), the World ex-US index seems to introduce significant additional skewness in the investment opportunity set of a US investor. These results remain true regardless of whether we consider the *GH* distribution or the *LSMN* with a Bernoulli mixing variable. As a consequence, we reject the joint null. Hence, 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. Intuitively, investors concerned about skewness can probably use the additional asset to diversify the asymmetry of their US exposures, which is statistically significant, as we have already mentioned. From a different perspective, though, these potential benefits of foreign investments would tend to exacerbate the empirically observed home bias puzzle (see e.g. French and Poterba, 1991).

Figure 6 also shows the changes in the mean-variance-skewness investment opportunity set before and after considering the additional asset for the *GH* case.[10] We can observe the differences between the three-dimensional plots of the two frontiers in Figure 6a. We can also observe in Figure 6b that the mean-variance frontier is almost unaffected, which is consistent with (20) 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 6c and 6d confirm this effect on the iso-variance and the skewness-variance frontiers, respectively. This graphical intuition is confirmed by the formal statistical tests in Table 2, which explicitly take into account the sampling uncertainty surrounding the parameter estimates. Finally, note that the third column of Tables 1a and 1b shows that the estimates of the shape parameters of the *GH* distribution remain fairly stable when

[10] This frontier is quantitatively very similar when we consider the Bernoulli mixing variable. We do not include these results for the sake of brevity, but they are available from the authors on request.

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. 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 our case, we are able to obtain analytical expressions for the mean-variance-skewness frontier, which encloses the feasible investment opportunity set, and show that its efficient part can always be spanned by three funds: the two funds that span the mean-variance frontier 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 when the distribution of the standardised innovations comes from either the *GH* distribution or a *LSMN* with a Bernoulli mixing variable. 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. For the two members of the *LSMN* family that we consider, we find that this additional asset does not have a significant impact

from a mean-variance perspective, but 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 considering a nonparametric specification for the distribution of the mixing variable ξ_t. In particular, we could consider either a semiparametric expansion of some known distribution, such as the Gamma, or a flexible multinomial distribution.

It could also be interesting to introduce dynamic features in higher order moments. In this sense, at least two possibilities might be worth exploring: either considering time varying shape parameters τ_t and $\mathbf{b}_t$, in line with Jondeau and Rockinger (2003), or introducing a regime switching process for ξ_t, following Guidolin and Timmermann (2007). These extensions might be helpful for skewness timing (see Jondeau and Rockinger, 2008).

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 as 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). 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.

Finally, it is important to note that our results are largely driven by the fact that the third centred moment of a *LSMN* satisfies (5). However, there are other asymmetric distributions that satisfy this property. Specifically, Simaan (1993) shows that if one adds an independent scalar asymmetric variable times a vector to an elliptical random vector, then (5) holds with $s_{2t} = 0$ (see also Gamba and Rossi, 1998; Pressacco and Stucchi, 2000). In addition, it is possible to show that (5) would also hold with $s_{1t}+3s_{2t}s_{3t} = 0$ for a multivariate Hermite expansion in which asymmetry is a common feature (see Mencía and Sentana, 2009, for a formal definition of this density). In this sense, it would be very helpful to characterise a broader class of distributions that satisfy (5).

Acknowledgements

We are grateful to Enrico Biffis, Christian Bontemps, Mariano González, Emmanuel Jurczenko, Francisco Peñaranda, Michael Rockinger, Steve Ross as well as seminar audiences at Cass Business School, ESCP-EAP Business School, HEC Lausanne, HEC Paris, Imperial College Business School, University of Pennsylvania, Yale University, the 2008 European Meeting of the Econometric Society (Milan), the 2008 Finance Forum (Barcelona) and the 2008 Symposium on Economic Analysis (Zaragoza) for helpful comments and suggestions. The input of two anonymous referees has also greatly improved the paper. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 (Sentana) is gratefully acknowledged.

Appendix

A Mean-variance-asymmetry frontier without a risk-free asset

If r_t is not a risk-free asset, then we could understand the analysis in Section 3 as describing the investment opportunity set of an investor with zero net wealth who can only invest in arbitrage portfolios. For investors with positive net wealth, consider the vector of gross returns $\mathbf{x}_t = \mathbf{y}_t + r_t \boldsymbol{\iota}_N$, where $\boldsymbol{\iota}_N$ is a vector of N ones. Then, we can obtain the portfolios on the mean-variance-asymmetry frontier from the following problem:

$$\max_{\mathbf{w}_t \in \mathbb{R}^N} \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) \quad \text{s.t.} \quad \left\{ \begin{array}{c} \mathbf{w}_t' \boldsymbol{\nu}_t(\boldsymbol{\theta}) = \nu_{0t} \\ \mathbf{w}_t' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t = \sigma_{0t}^2 \\ \mathbf{w}_t' \boldsymbol{\iota}_N = 1 \end{array} \right. \tag{A1}$$

where $\boldsymbol{\nu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + r_t$. Compared to (7), we have introduced an additional restriction to ensure the unit cost of the portfolio. We can solve (7) by forming the Lagrangian

$$\mathcal{L} = \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) + \gamma_1 \left[\nu_{0t} - \mathbf{w}_t' \boldsymbol{\nu}_t(\boldsymbol{\theta}) \right] + \gamma_2 \left[\sigma_{0t}^2 - \mathbf{w}_t' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t \right] + \gamma_3 \left[1 - \mathbf{w}_t' \boldsymbol{\iota}_N \right],$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}_t} = & \left\{ 3(s_{1t} + 3 s_{2t} s_{3t}) \left[\mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t \right]^2 + 3 s_{2t} \left[\mathbf{w}_t' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t \right] \right\} \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{b} \\ & + 6 s_{2t} \left[\mathbf{b}' \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t \right] \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t - \gamma_1 \boldsymbol{\nu}_t(\boldsymbol{\theta}) - 2 \gamma_2 \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{w}_t - \gamma_3 \boldsymbol{\iota}_N. \end{aligned} \tag{A2}$$

Once again, we can explicitly obtain in closed-form the set of portfolio weights that satisfy these conditions:

Proposition 7 *The efficient mean-variance-skewness portfolios that solve (A2) can be expressed as either*

$$\begin{aligned}\mathbf{w}_{1t}^{\diamond} &= \left[\frac{C\nu_{0t}-A}{D}+\frac{C\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta})-A\mathbf{b}'\boldsymbol{\iota}_N}{D}\sqrt{\frac{\Pi_2(\nu_{0t},\sigma_{0t}^2)}{\Pi_1}}\right]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}_t(\boldsymbol{\theta})\\ &+\left[\frac{B-A\nu_{0t}}{D}+\frac{B\mathbf{b}'\boldsymbol{\iota}_N-A\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta})}{D}\sqrt{\frac{\Pi_2(\nu_{0t},\sigma_{0t}^2)}{\Pi_1}}\right]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\iota}_N\\ &-\sqrt{\frac{\pi_2(\nu_{0t},\sigma_{0t}^2)}{\pi_1}}\mathbf{b} \qquad \text{(A3)}\end{aligned}$$

or

$$\begin{aligned}\mathbf{w}_{2t}^{\diamond} &= \left[\frac{C\nu_{0t}-A}{D}-\frac{C\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta})-A\mathbf{b}'\boldsymbol{\iota}_N}{D}\sqrt{\frac{\Pi_2(\nu_{0t},\sigma_{0t}^2)}{\Pi_1}}\right]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}_t(\boldsymbol{\theta})\\ &+\left[\frac{B-A\nu_{0t}}{D}-\frac{B\mathbf{b}'\boldsymbol{\iota}_N-A\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta})}{D}\sqrt{\frac{\Pi_2(\nu_{0t},\sigma_{0t}^2)}{\Pi_1}}\right]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\iota}_N\\ &+\sqrt{\frac{\Pi_2(\nu_{0t},\sigma_{0t}^2)}{\Pi_1}}\mathbf{b}. \qquad \text{(A4)}\end{aligned}$$

where $A=\boldsymbol{\iota}_N\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $B=\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $C=\boldsymbol{\iota}_N'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\iota}_N$, $D=BC-A^2$,

$$\begin{aligned}\Pi_1 &= \mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\\ &-\frac{1}{D}\begin{pmatrix}\mathbf{b}'\boldsymbol{\iota}_N & \mathbf{b}_t'\boldsymbol{\nu}_t(\boldsymbol{\theta})\end{pmatrix}\begin{bmatrix}B & -A\\ -A & C\end{bmatrix}\begin{pmatrix}\mathbf{b}'\boldsymbol{\iota}_N\\ \mathbf{b}_t'\boldsymbol{\nu}_t(\boldsymbol{\theta})\end{pmatrix}\geq 0, \qquad \text{(A5)}\end{aligned}$$

and

$$\Pi_2(\nu_{0t},\sigma_{0t}^2)=\sigma_{0t}^2-\frac{1}{D}\begin{pmatrix}1 & \nu_{0t}\end{pmatrix}\begin{bmatrix}B & -A\\ -A & C\end{bmatrix}\begin{pmatrix}1\\ \nu_{0t}\end{pmatrix}\geq 0. \qquad \text{(A6)}$$

Hence, there are two potential solutions, both of which can be expressed as a linear combination of the portfolios with weights $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\iota}_N$, and $\mathbf{b}$.

It is also interesting to analyse the characteristics of the tangency curve between the frontiers for $\mathbf{x}_t$ and $\mathbf{y}_t$. In this sense, we can show the following result.

Proposition 8

1. *The tangency portfolios between the mean-variance-skewness frontiers with and without risk-free asset have weights*

$$\mathbf{w}_t^{\star}=\frac{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}-\mu_{0t}\mathbf{b}'\boldsymbol{\iota}_N}{A\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}-B\mathbf{b}'\boldsymbol{\iota}_N}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})-\frac{B-\mu_{0t}A}{A\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}-B\mathbf{b}'\boldsymbol{\iota}_N}\mathbf{b}$$

on the risky assets.

2. The tangency curve can be expressed as

$$\sigma_{0t}^2 = \frac{(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\, B - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2}{B[A\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} - B\mathbf{b}'\boldsymbol{\iota}_N]^2}(B - \mu_{0t}A)^2 + \frac{\mu_{0t}^2}{B}. \tag{A7}$$

B Variance-asymmetry frontier

If we ignore expected returns, we can obtain a skewness-variance frontier by maximising skewness subject to a variance constraint:

Proposition 9 *If*

$$\frac{s_{2t}}{s_{1t} + 3s_{2t}s_{3t}}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} + \frac{s_{2t}}{s_{1t} + 3s_{2t}s_{3t}}\right] > 0. \tag{B8}$$

then the solution to the problem

$$\max_{\mathbf{w}_t \in \mathbb{R}^N} \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) \quad s.t. \quad \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2 \tag{B9}$$

will be

$$[\varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau})]^{1/3} = \Lambda_1(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau})\sigma_{0t}, \tag{B10}$$

where

$$\Lambda_1(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) = \left\{(s_{1t} + 3s_{2t}s_{3t})\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^{3/2} + 3s_{2t}\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^{1/2}\right\}^{1/3}, \tag{B11}$$

which is achieved by

$$\mathbf{w}_t^{\dagger} = \frac{\sigma_{0t}}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}}\mathbf{b}. \tag{B12}$$

Otherwise the solution to (B9) will be

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

where

$$\Lambda_2(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) = 2^{1/3}\sqrt{s_{2t}}\left[-s_{1t} - 3s_{2t}s_{3t}\right]^{-1/6}, \tag{B13}$$

which is obtained by portfolios that satisfy

$$\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^{\ddagger} = \sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t} + 3s_{2t}s_{3t}}}. \tag{B14}$$

Hence, we can interpret $\mathbf{b}$ as a "skewness-variance" efficient portfolio, since every portfolio on this frontier will be proportional to $\mathbf{b}$ when (B8) is satisfied. However, when (B8) is not satisfied, (B12) will not necessarily yield maximum skewness. In fact, there might be an infinite number of portfolios that satisfy (B14), all of them yielding exactly the same variance and skewness but different expected returns. One way of solving this indeterminacy is to choose the portfolio with maximum expected return. In this sense, we can show that:

Proposition 10 *If (B8) does not hold, then the solution to the problem*

$$\arg\max_{\mathbf{w}_t\in\mathbb{R}^N} \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) \quad s.t. \quad \begin{cases} \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2 \\ \mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}\sqrt{\dfrac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}} \end{cases} \tag{B15}$$

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

C Proofs of Propositions

Proposition 1

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

$$\boldsymbol{\varepsilon}^* = c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\boldsymbol{\beta}\left[\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right]+\sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\left[\mathbf{I}_N+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right]^{\frac{1}{2}}\mathbf{r} \tag{C16}$$

Then, we can use the independence of ξ 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 (C16), 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

$$\begin{aligned} s^* &= c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)\frac{\mathbf{w}'\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{w}}}\left[\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right] \\ &\quad +\sqrt{\frac{\xi^{-1}}{\pi_1(\boldsymbol{\tau})}}\sqrt{1+\frac{c\left(\boldsymbol{\beta}'\boldsymbol{\beta},\boldsymbol{\tau}\right)-1}{\boldsymbol{\beta}'\boldsymbol{\beta}}\frac{\left(\mathbf{w}'\boldsymbol{\beta}\right)^2}{\mathbf{w}'\mathbf{w}}}r \end{aligned} \tag{C17}$$

Given that s^* is a standardised variable by construction, if we compare (C17) with the general formula for a standardised *LSMN* in (C16), 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

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

□

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:

$$\begin{aligned}\mathbf{w}_t'\mathbf{y}_t &= \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) + c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\boldsymbol{\tau}]\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\left[\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})}-1\right] \\ &+\sqrt{\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})}}\left\{\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t + \frac{c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b},\boldsymbol{\tau}]-1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}\left[\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\right]^2\right\}^{\frac{1}{2}}\mathbf{r}_t \qquad \text{(C18)}\end{aligned}$$

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 $\mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$. □

Proposition 4

In what follows we maintain the assumption that

$$3(s_{1t}+3s_{2t}s_{3t})[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t]^2 + 3s_{2t}\sigma_{0t}^2 \qquad \text{(C19)}$$

is different from zero, since the equality case is treated in Propositions 9 and 10. If we set (10) to zero, we can express the optimal portfolio weights as:

$$\begin{aligned}\mathbf{w}_t^* &= \frac{\gamma_1}{6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]-2\gamma_2}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &-\frac{\{3(s_{1t}+3s_{2t}s_{3t})[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]^2+3s_{2t}\sigma_{0t}^2\}}{6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]-2\gamma_2}\mathbf{b} \qquad \text{(C20)}\end{aligned}$$

If we pre-multiply (C20) by $\mathbf{b}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})$, we obtain:

$$\begin{aligned}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^* &= \frac{\gamma_1}{6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &- \frac{\{3(s_{1t}+3s_{2t}s_{3t})[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*]^2 + 3s_{2t}\sigma_{0t}^2\}}{6s_{2t}[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*] - 2\gamma_2}\mathbf{b}'\boldsymbol{\Sigma}\mathbf{b} \end{aligned} \tag{C21}$$

Hence, we can express (C20) as

$$\mathbf{w}_t^* = \frac{\gamma_1}{6s_{2t}z^* - 2\gamma_2}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{[3(s_{1t}+3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]}{6s_{2t}z^* - 2\gamma_2}\mathbf{b} \tag{C22}$$

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

$$\begin{aligned}&[6s_{2t} + 3(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\,(s_{1t}+3s_{2t}s_{3t})]\,z^2 \\ &-2\gamma_2 z + \left[3s_{2t}(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\sigma_{0t}^2 - \gamma_1\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta})\right] = 0. \end{aligned} \tag{C23}$$

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

$$\begin{aligned}\mu_0 &= \frac{\gamma_1}{6s_{2t}z^* - 2\gamma_2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad - \frac{[3(s_{1t}+3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_0^2]}{6s_{2t}z^* - 2\gamma_2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \end{aligned} \tag{C24}$$

$$\begin{aligned}\sigma_{0t}^2 &= \frac{\gamma_1^2}{[6s_{2t}z^* - 2\gamma_2]^2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad + \frac{[3(s_{1t}+3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]^2}{[6s_{2t}z^* - 2\gamma_2]^2}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b} \\ &\quad - 2\frac{\gamma_1[3(s_{1t}+3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_{0t}^2]}{[6s_{2t}z^* - 2\gamma_2]^2}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \end{aligned} \tag{C25}$$

Thus, we must find z^*, γ_1 and γ_2 such that (C23), (C24) and (C25) are satisfied. From (C24), it is straightforward to express γ_1 as:

$$\begin{aligned}\gamma_1 &= \frac{\mu_0}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\left[6s_{2t}z^* - 2\gamma_2\right] \\ &\quad + \frac{[3(s_{1t}+3s_{2t}s_{3t})z^{*2} + 3s_{2t}\sigma_0^2]}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} \end{aligned} \tag{C26}$$

If we introduce (C26) in (C25), we will obtain after some algebraic manipulations that:

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

From condition (8) $\sigma_{0t}^2\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \mu_{0t}^2 \geq 0$, whereas $(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}\right)^2$ is also non-negative because of the Cauchy-Schwarz inequality. Therefore, we can express γ_2 as:

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

whence

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

If we introduce these expressions in (C23), we obtain the following "non-trivial" solutions:

$$z^* = \mu_{0t}\frac{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \mp \frac{\sqrt{\left[(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}\right)^2\right]\left[\sigma_{0t}^2\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - \mu_{0t}^2\right]}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \tag{C27}$$

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 10.

Finally, we obtain the required result by introducing (C27) in (C22). □

Proposition 5

Consider two mean-variance-skewness efficient portfolios with weights

$$\begin{aligned}\mathbf{w}_{at}^* &= \frac{\mu_{0at} + i_a\Delta_t^{-1}(\mu_{0at},\sigma_{0at})\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{i_a}{\Delta_t(\mu_{0at},\sigma_{0at})}\mathbf{b}, \\ \mathbf{w}_{bt}^* &= \frac{\mu_{0bt} + i_b\Delta_t^{-1}(\mu_{0bt},\sigma_{0bt})\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \frac{i_b}{\Delta_t(\mu_{0bt},\sigma_{0bt})}\mathbf{b},\end{aligned}$$

where i_a and i_b are either 1 or -1. Then, for any $k_a, k_b \in \mathbb{R}$ we can express the linear combination $\mathbf{w}_{ct}^* = k_a\mathbf{w}_{at}^* + k_b\mathbf{w}_{bt}^*$ as

$$\begin{aligned}\mathbf{w}_{at}^* &= \frac{\mu_{0ct} + \left[k_a i_a\Delta_t^{-1}(\mu_{0at},\sigma_{0at}) + k_b i_b\Delta_t^{-1}(\mu_{0bt},\sigma_{0bt})\right]\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b}}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ &\quad - \left[k_a i_a\Delta_t^{-1}(\mu_{0at},\sigma_{0at}) + k_b i_b\Delta_t^{-1}(\mu_{0bt},\sigma_{0bt})\right]\mathbf{b},\end{aligned}$$

where $\mu_{0ct} = k_a\mu_{0at} + k_b\mu_{0bt}$. Hence, we only need to show that we can express

$$k_a i_a \Delta_t^{-1}(\mu_{0at}, \sigma_{0at}) + k_b i_b \Delta_t^{-1}(\mu_{0bt}, \sigma_{0bt})$$

as $\pm\Delta_t^{-1}(\mu_{0ct}, \sigma_{0ct})$ for some σ_{0ct} to obtain the required result. In this sense, it is straightforward to show that

$$\begin{aligned}\sigma_{0ct}^2 &= \frac{\mu_{0ct}^2}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})} \\ &\quad + \frac{\left[k_a i_a \Delta_t^{-1}(\mu_{0at}, \sigma_{0at}) + k_b i_b \Delta_t^{-1}(\mu_{0bt}, \sigma_{0bt})\right]^2}{\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\left[(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})\left(\boldsymbol{\mu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta})\right) - (\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b})^2\right]^{-1}} > 0\end{aligned}$$

satisfies this restriction. □

Proposition 6

We can exploit conditional normality of the *LSMN* family to express the second lower partial moment as

$$S(\mathbf{w}_t) = \int E\left[\min(0, \mathbf{w}_t'\mathbf{y}_t)^2 | \xi_t\right] dF(\xi_t; \boldsymbol{\tau}),$$

where

$$\begin{aligned}E\left[\min(0, \mathbf{w}_t'\mathbf{y}_t)^2 | \xi_t\right] &= \left[\mu_t^2(\xi_t) + \sigma_t^2(\xi_t)\right]\Phi\left(-\frac{\mu_t(\xi_t)}{\sigma_t(\xi_t)}\right) \\ &\quad -\mu_t(\xi_t)\sigma_t(\xi_t)\phi\left(-\frac{\mu_t(\xi_t)}{\sigma_t(\xi_t)}\right),\end{aligned} \tag{C28}$$

$$\begin{aligned}\mu_t(\xi_t) &= \mu_{0t} + c(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau})z_{0t}\left[\frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})} - 1\right], \\ \sigma_t^2(\xi_t) &= \frac{\xi_t^{-1}}{\pi_1(\boldsymbol{\tau})}\left[\sigma_{0t}^2 + \frac{c[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}, \boldsymbol{\tau}] - 1}{\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}}z_{0t}^2\right],\end{aligned}$$

$\mu_{0t} = \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\sigma_{0t}^2 = \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t$, and $z_{0t} = \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}$ and $\Phi(\cdot)$ is the standard normal cdf. In addition, it can be shown that

$$\frac{\partial S(\mathbf{w}_t)}{\partial \sigma_{0t}^2} = \int \frac{\partial E\left[\min(0, \mathbf{w}_t'\mathbf{y}_t)^2 | \xi_t\right]}{\partial \sigma_{0t}^2} dF(\xi_t; \boldsymbol{\tau}) > 0,$$

since

$$\frac{\partial E\left[\min(0, \mathbf{w}_t'\mathbf{y}_t)^2 | \xi_t = \xi_i\right]}{\partial \sigma_0^2} = \Phi\left(-\frac{\mu_t(\xi_t)}{\sigma_t(\xi_t)}\right)\frac{\xi_i^{-1}}{\pi_1(\boldsymbol{\tau})} > 0.$$

Hence, for any interior point of the feasible mean-variance-skewness opportunity set, we can always diminish the second lower partial moment by reducing the standard deviation

keeping constant mean and skewness until we reach the mean-variance-skewness frontier. Finally, we can show that the portfolios with maximum Sortino ratio will be along a straight line by combining the homothecy property of the mean-variance-skewness frontier with the fact that we can span these portfolios by scaling any of them with a *positive* scalar . □

Proposition 7

It is not difficult to check that (A3) and (A4) satisfy the first order conditions (A2) with γ_1, γ_2 and γ_3 such that

$$\begin{aligned}\frac{\gamma_1}{6s_{2t}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^* - 2\gamma_2} &= -\frac{A - C\nu_0}{D} + \frac{C\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta}) - A\mathbf{b}'\boldsymbol{\iota}_N}{D}\sqrt{\frac{\pi_2(\nu_0, \sigma_0^2)}{\pi_1}},\\ \frac{\gamma_3}{6s_{2t}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^* - 2\gamma_2} &= \frac{B - A\nu_0}{D} + \frac{B\mathbf{b}'\boldsymbol{\iota}_N - A\mathbf{b}'\boldsymbol{\nu}_t(\boldsymbol{\theta})}{D}\sqrt{\frac{\pi_2(\nu_0, \sigma_0^2)}{\pi_1}}.\end{aligned}$$

and

$$\frac{6s_{2t}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^* - 2\gamma_2}{3(s_{1t} + 3s_{2t}s_{3t})\left[\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t^*\right]^2 + 3s_{2t}\sigma_{0t}^2} = \pm\sqrt{\frac{\pi_1}{\pi_2(\nu_0, \sigma_0^2)}},$$

where the positive (negative) sign corresponds to $\mathbf{w}_{1t}^{\diamond}$ ($\mathbf{w}_{2t}^{\diamond}$). It is necessary, though, to show that (A5) and (A6) are both non-negative so that the optimal portfolio weights are well defined. It can be checked that $\Pi_2(\nu_{0t}, \sigma_{0t}^2)$ is the difference between σ_{0t}^2 and the variance of a portfolio on the mean-variance frontier that has the same expected return. Hence, this difference cannot be negative. As for Π_1, we can express (A5) as $\Pi_1 = \mathbf{b}'\boldsymbol{\Omega}_t(\boldsymbol{\theta})\mathbf{b}$, where

$$\begin{aligned}\boldsymbol{\Omega}_t(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\\ &\quad -\begin{pmatrix}\boldsymbol{\iota}_N & \boldsymbol{\nu}_t(\boldsymbol{\theta})\end{pmatrix}\begin{bmatrix}C & A\\ A & B\end{bmatrix}^{-1}\begin{pmatrix}\boldsymbol{\iota}_N'\\ \boldsymbol{\nu}_t'(\boldsymbol{\theta})\end{pmatrix}.\end{aligned}$$

But since $\boldsymbol{\Omega}_t(\boldsymbol{\theta})$ is the residual variance of the regression of $\mathbf{y}_t$ on $\boldsymbol{\iota}_N'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{y}_t$ and $\boldsymbol{\nu}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{y}_t$, it must be positive semidefinite, which ensures that $\Pi_1 \geq 0$. Finally, if there are portfolios such that

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

then there might be other combinations of portfolio weights that satisfy (A2). As in the frontier with a risk-free asset, though, these portfolios will have the same asymmetry and standard deviation but different expected returns. It is possible to follow the same

procedure as in Proposition 10 to show that the portfolio in this set with maximum expected return can also be expressed in terms of either (A3) or (A4). □

Proposition 8

The tangency portfolios will be such that $1 = \mathbf{w}_{it}^{*\prime}\boldsymbol{\iota}_N$, for $i = 1, 2$. If we introduce this restriction in (11) and (12), we obtain:

$$\Delta_t(\mu_{0t}, \sigma_{0t}) = \pm\frac{A\boldsymbol{\mu}_t'(\boldsymbol{\theta})\mathbf{b} - B\mathbf{b}'\boldsymbol{\iota}_N}{B - \mu_{0t}A}. \tag{C29}$$

By introducing (13) in (C29), we can obtain (A7). In addition, $\mathbf{w}_t^\star$ follows directly from introducing (C29) in either (11) or (12). □

Propositions 9 and 10

We can solve (B9) by forming the Lagrangian

$$\mathcal{L} = \varphi_t(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\tau}) + \gamma_2\left(\sigma_{0t}^2 - \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right). \tag{C30}$$

If we differentiate (C30) 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 (C19) is different from zero. In this case, we can express the optimal portfolio weights as $\mathbf{w}_t = \kappa\mathbf{b}$ for some constant κ. Then, if we impose the variance constraint by choosing κ appropriately, we obtain (B12). However, an additional solution will be obtained if the scalars (C19) 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

$$\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \pm\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t} + 3s_{2t}s_{3t}}}, \tag{C31}$$

$$\mathbf{w}_t\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t = \sigma_{0t}^2. \tag{C32}$$

However, we will choose the positive sign because it is the one that yields positive skewness. Condition (C31) defines a plane. Thus, this solution will only exist if this plane intersects the ellipse defined by (C32). We need to find under what conditions

(C31) and (C32) 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 (B15). In this case, the Lagrangian can be expressed as

$$\begin{aligned}\mathcal{L} &= \mathbf{w}_t'\boldsymbol{\mu}_t(\boldsymbol{\theta}) + \gamma_1\left[\sigma_{0t}^2 - \mathbf{w}_t'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right] \\ &\quad +\gamma_2\left[\sigma_{0t}\sqrt{\frac{-s_{2t}}{s_{1t}+3s_{2t}s_{3t}}} - \mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t\right]. \end{aligned} \tag{C33}$$

If we differentiate (C33) with respect to $\mathbf{w}_t$, we obtain:

$$\mathbf{w}_t = \frac{1}{2\gamma_1}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - \gamma_2\mathbf{b}\right] \tag{C34}$$

It is straightforward to show that

$$\gamma_1 = \pm\frac{\sqrt{\boldsymbol{\mu}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_t(\boldsymbol{\theta}) - 2\gamma_2\mathbf{b}'\boldsymbol{\mu}_t(\boldsymbol{\theta}) + \gamma_2^2(\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b})}}{2\sigma_{0t}} \tag{C35}$$

ensures that (C32) holds. If we introduce (C34) and (C35) in (C31), 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 (B8) does not hold. □

D Third and fourth moments of a *LSMN*

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

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

where

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

$$\mathbf{K}_t(\boldsymbol{\theta},\boldsymbol{\tau}) =$$

$$= E\left[vec\left[(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'\right]vec'\left[(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))(\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta}))'\right]|I_{t-1};\boldsymbol{\theta},\boldsymbol{\tau}\right]$$

$$= \kappa_{1t} vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]$$

$$+\kappa_{2t}\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)\left(\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)$$

$$+\kappa_{2t}\left[vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]vec'\left[\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\right]+vec\left[\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\right]vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]\right]$$

$$+\kappa_{3t}\left[\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)\left(\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\right)+vec\left(\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta})\right)vec'(\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta}))\right],$$

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

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

$$\boldsymbol{\Sigma}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) + s_{3t}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}).$$

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, Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society, Series B 63, 167–241.

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.

Briec, W., K. Kerstens, and O. Jokung, 2007, Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach. Management Science 53, 135–149.

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, 2008, Nonparametric risk management with Generalized Hyperbolic distributions. Journal of the American Statistical Association 103, 910–923.

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.

French, K. R. and J. M. Poterba, 1991, Investor diversification and international equity markets. American Economic Review 81, 222–226.

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

Gamba, A. and F. A. Rossi, 1998, A three-moment based portfolio selection model. Rivista di matematica per le scienze economiche e sociali 21, 25–48.

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

Guidolin, M. and A. Timmermann, 2007, Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control 31, 3503–3544.

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

Hogan, W. W. and J. M. Warren, 1974, Toward the development of an equilibrium capital-market model based on semivariance. Journal of Financial and Quantitative Analysis 9, 1–11.

Jondeau, E. and M. Rockinger, 2003, Conditional volatility, skewness and kurtosis: existence, persistence and comovements. Journal of Economics Dynamics and Control 27, 1699–1737.

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

Jondeau, E. and M. Rockinger, 2008, The economic value of distributional timing. mimeo Swiss Finance Institute.

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

Jorion, P., 2006, Value at risk: the new benchmark for managing financial risk. New York: McGraw-Hill.

Jurczenko, E., B. Maillet, and P. Merlin, 2006, Hedge funds portfolio selection with higher order moments: a non-parametric mean-variance-skewness-kurtosis efficient

frontier. In E. Jurczenko and B. Maillet (Eds.), Multi-moment asset allocation and pricing models, pp. 51–66. John Wiley & Sons.

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. and E. Sentana, 2009, 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.

Pedersen, C. S. and S. E. Satchell, 2002, On the foundation of performance measures under asymmetric returns. mimeo University of Cambridge.

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

Pressacco, F. and P. Stucchi, 2000, Linearity properties of a three-moments portfolio model. Decisions in economics and finance 23, 133–150.

Ross, S. A., 1978, Mutual fund separation in financial theory - The separating distributions. Journal of Economic Theory 17, 254–286.

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.

Simaan, Y., 1993, Portfolio selection and asset pricing-three-parameter framework. Management Science 39, 568–577.

Tanner, M. A., 1996, Tools for statistical inference: methods for exploration of posterior distributions and likelihood functions (Third ed.). New York: Springer-Verlag.

Table 1
Maximum likelihood estimates of a conditionally heteroskedastic single factor model for ten Datastream sectoral indices for the US.

Daily excess returns (January 4th, 1988 - October 12th, 2007)

(a) Generalised Hyperbolic distribution

	Ten indices		Extended model	
Asymmetry parameters **b**		SE		SE
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
Log-likelihood	-51997.250		-57588.015	

(b) *LSMN* with a Bernoulli mixing variable

	Ten indices		Extended model	
Asymmetry parameters **b**		SE		SE
Basic Materials	-0.123	0.041	-0.122	0.063
Consumer Goods	0.095	0.066	0.101	0.110
Consumer Services	0.009	0.089	-0.004	0.138
Financials	0.036	0.061	0.083	0.120
Health Care	-0.004	0.078	-0.047	0.121
Industrials	-0.156	0.092	-0.125	0.135
Oil and Gas	0.125	0.055	0.139	0.086
Technology	-0.055	0.054	-0.063	0.121
Telecommunications	0.088	0.067	0.079	0.121
Utilities	-0.009	0.039	-0.016	0.075
World ex-US	-	-	-0.168	0.068
Log-likelihood	-52142.065		-57752.316	

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.

Daily excess returns (January 4th, 1988 - October 12th, 2007)

(a) Generalised Hyperbolic distribution

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

(b) *LSMN* with a Bernoulli mixing variable

Null hypothesis	Wald		LR	
		p-value		p-value
Mean-variance efficiency	0.99	0.320	1.08	0.298
Skewness-variance efficiency	6.12	0.013	8.58	0.003
Joint	10.20	0.006	12.10	0.002

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.

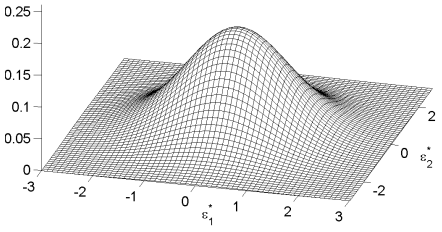


Figure 1a: Standardised bivariate normal density

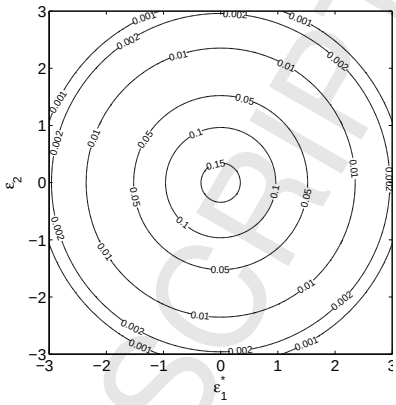


Figure 1b: Contours of a standardised bivariate normal density

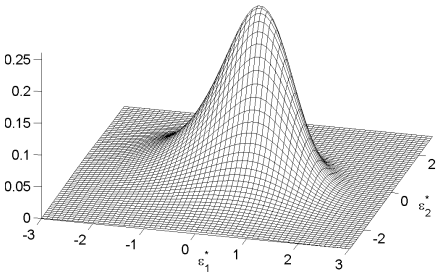


Figure 1c: Standardised bivariate asymmetric Student t density with 10 degrees of freedom ($\eta = .1$) and $\boldsymbol{\beta} = (-3, -3)'$

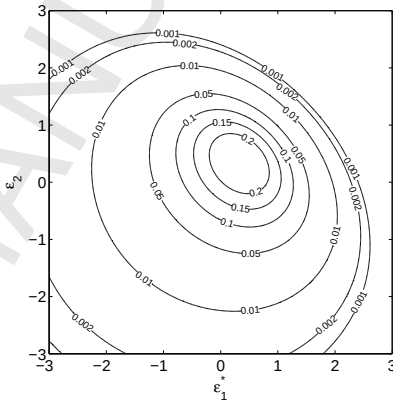


Figure 1d: Contours of a standardised bivariate asymmetric Student t density with 10 degrees of freedom ($\eta = .1$) and $\boldsymbol{\beta} = (-3, -3)'$

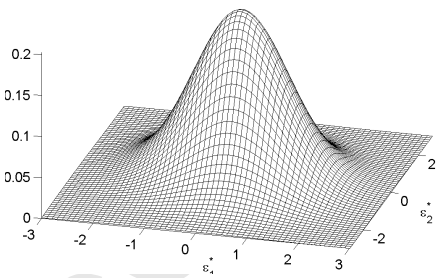


Figure 1e: Standardised bivariate *LSMN* with a Bernoulli mixing variable and $\boldsymbol{\beta} = (-3, -3)'$

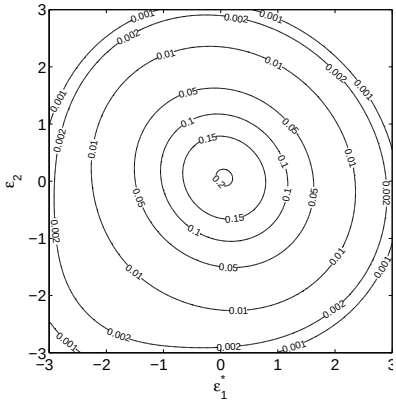
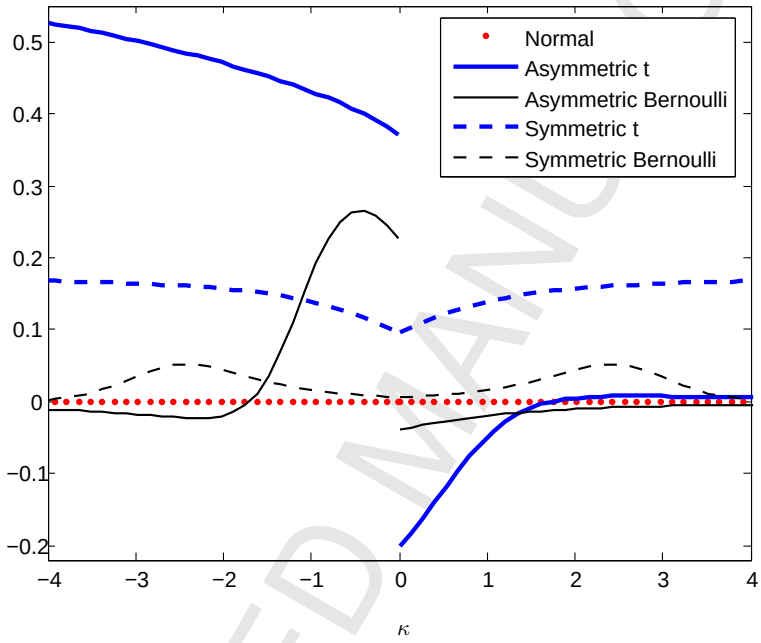


Figure 1f: Contours of a standardised bivariate *LSMN* with a Bernoulli mixing variable and $\boldsymbol{\beta} = (-3, -3)'$

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

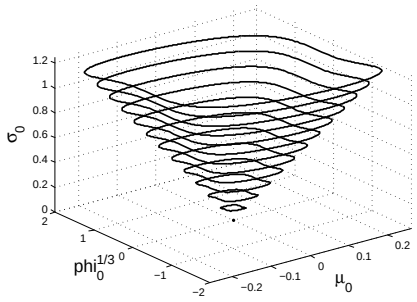
Figure 2: Exceedance correlation for symmetric and asymmetric location-scale mixtures of normals



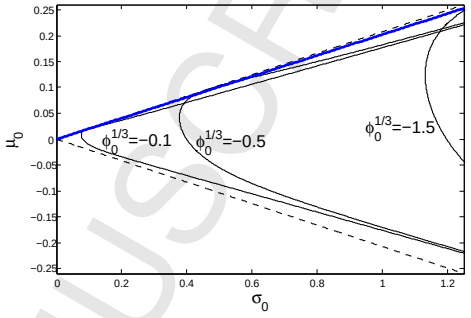
Notes: The exceedance correlation between two variables ε_1^* and ε_2^* is defined as $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \kappa, \varepsilon_2^* > \kappa)$ for positive κ and $\text{corr}(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* < \kappa, \varepsilon_2^* < \kappa)$ for negative κ (see Longin and Solnik, 2001). 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 $Pr(\xi = 0.6) = 0.04$. By construction, $\text{corr}(\varepsilon_1^*, \varepsilon_2^*) = 0$ in all cases.

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

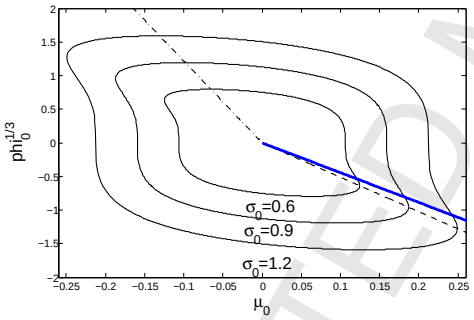
(a) Three dimensional representation



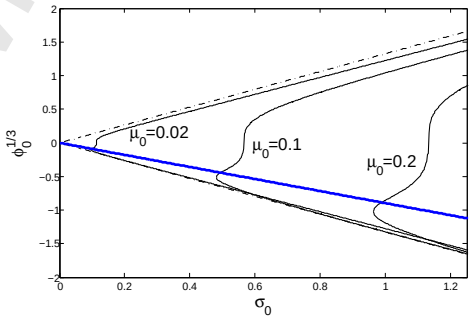
(b) Mean vs. Standard Deviation



(c) Mean vs. Asymmetry



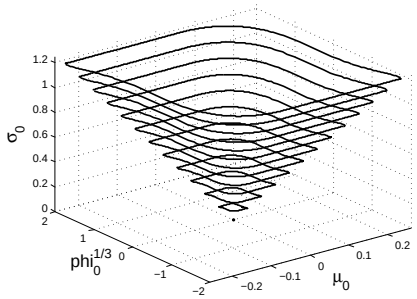
(d) Standard Deviation vs. Asymmetry



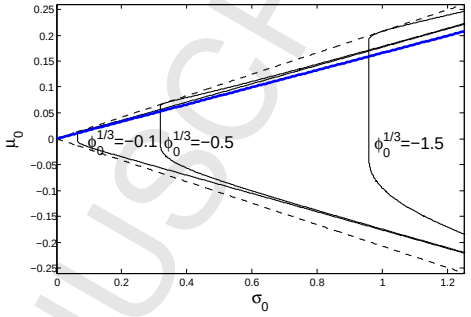
Notes: The mean-variance frontier is plotted with dotted lines. Dash-dot lines are used for the skewness-variance frontier, which maximises asymmetry for given standard deviations. In this case, the asymmetry-variance frontier portfolio weights are proportional to $-\mathbf{b}$. Thin solid lines indicate several iso-asymmetry, iso-variance and iso-mean contours on panels (b), (c) and (d), respectively. Thick solid lines represent the location of the optimal portfolios that an investor who maximises the Sortino ratio for different target expected returns would choose. This example has been obtained with a five-dimensional *LSMN* with $\mathbf{b} = -\boldsymbol{\iota}_5$ in which the mixing variable is the Bernoulli variable $\xi = 2, 10$, with $\Pr(\xi = 2) = .1$.

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

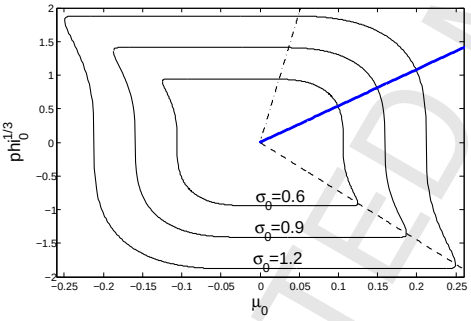
(a) Three dimensional representation



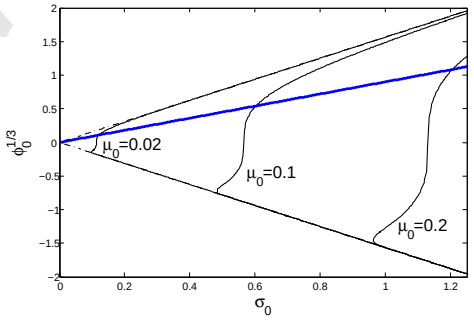
(b) Mean vs. Standard Deviation



(c) Mean vs. Asymmetry



(d) Standard Deviation vs. Asymmetry



Notes: The mean-variance frontier is plotted with dotted lines. Dash-dot lines are used for the skewness-variance frontier, which maximises asymmetry for given standard deviations. Thin solid lines indicate several iso-asymmetry, iso-variance and iso-mean contours on panels (b), (c) and (d), respectively. Thick solid lines represent the location of the optimal portfolios that an investor who maximises the Sortino ratio for different target expected returns would choose. This example has been obtained with a five-dimensional *LSMN* with $\mathbf{b} = -\boldsymbol{\iota}_5$ in which the mixing variable is the Bernoulli variable $\xi = .1, 10$, with $\Pr(\xi = .1) = .1$.

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

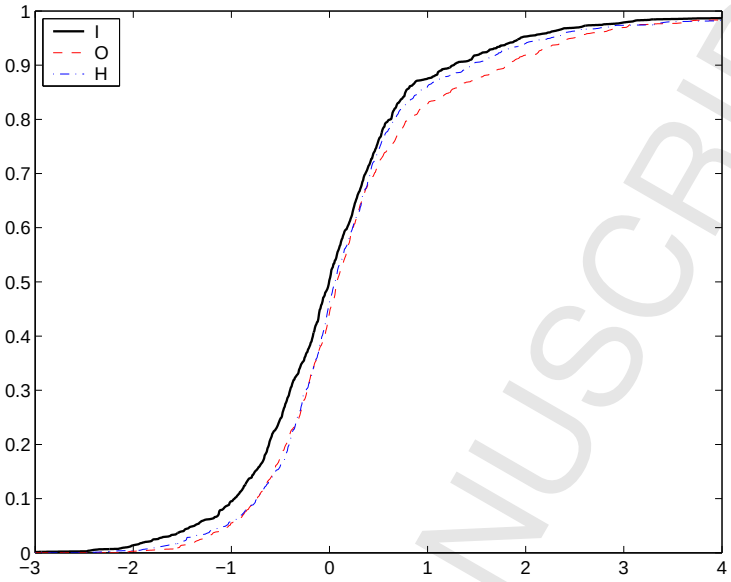
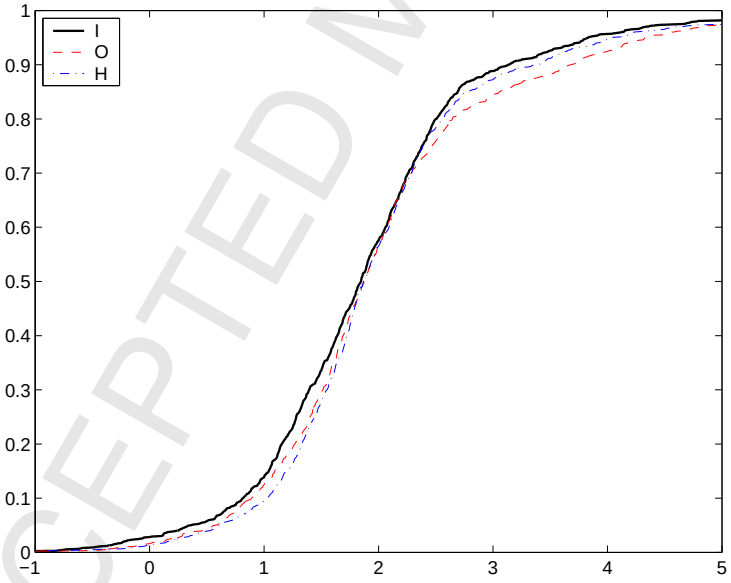


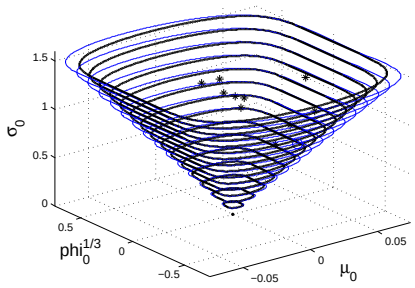
Figure 5b: Sampling distribution of the log of $vecd'[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]vecd[V_T^E(\hat{\phi}_T) - V_T(\hat{\phi}_T)]$



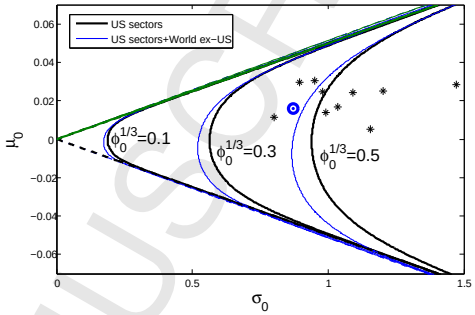
Notes: Obtained from a Monte Carlo study with 1,000 replications of sample size $T = 1,000$, except $V_T(\hat{\phi}_T)$, which is the sampling variance of the ML estimators in 30,000 samples of the same size. E refers to the standard errors obtained by either the outer-product of the gradient (O), numerical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

Figure 6: Mean-Variance-Skewness frontier of the the US Datastream indices and change induced by adding the World ex-US Datastream index

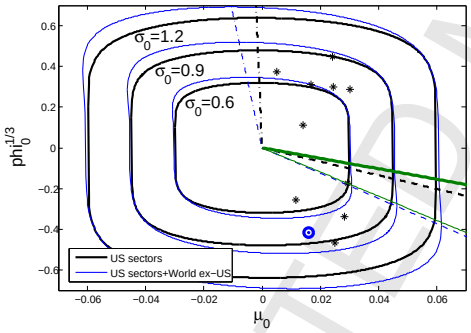
(a) Three dimensional representation



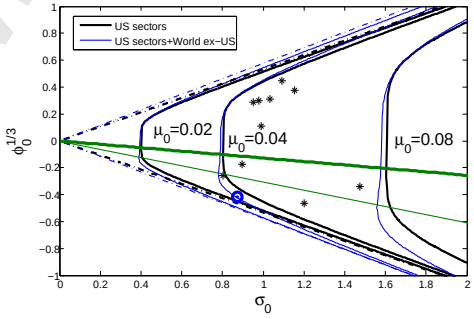
(b) Mean vs. Standard Deviation



(c) Mean vs. Asymmetry



(d) Standard Deviation vs. Asymmetry



Notes: Frontier of daily excess returns with moments expressed in percent terms. Results based on the *GH* distribution, whose parameters have been estimated by maximum likelihood. 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. Straight lines represent the location of the optimal portfolios that an investor who maximises the Sortino ratio for different target expected returns would choose, where the thin (thick) line is based on the 10 US sectoral indices (10 US sectoral indices and the World ex-US index).