Optimal Bayesian portfolios of hedge funds

Jean-François Bacmann[a] and Saverio Massi Benedetti[a]

RMF, a core investment manager of Man Investments

This Version: February 2006

Preliminary draft

[a] Quantitative Analysis, RMF Investment Management, Huobstrasse 16, 8808 Pfäffikon/SZ, Switzerland

Email: Jean-Francois.Bacmann@rmf.ch

Saverio.Massibenedetti@rmf.ch

Optimal Bayesian portfolios of hedge funds

Abstract:

Hedge fund returns are not normally distributed. Hedge fund styles related to arbitrage strategies exhibit negative skewness while more directional styles like managed futures and global macro are more positively skewed. We implement and test a Bayesian framework for portfolio optimisation process, in order to take these characteristics as well as the estimation risk into account. Hedge fund returns are modelled using multivariate skew elliptical distributions. The first three predictive estimates are used in a truncated utility function to obtain sets of optimal portfolios. We show that the choice of the underlying distribution as well as the modelling of co-skewnesses has an important impact on the final optimal portfolios.

Keywords: hedge funds, higher moments, estimation risk, skew elliptical distributions, Bayesian allocation.

JEL code: C11, C61, G11.

1 Introduction

The risks embedded in hedge funds are in essence much more complex than that of traditional investments, i.e. stocks and bonds. The absence of normal distribution and the skewed and leptokurtic behaviour of returns have been documented by several researches; see among others Schmidhuber and Moix (2001) and Brooks and Kat (2002). Following Fung and Hsieh (1999), a large body of contributions have been examining the validity of the traditional mean-variance framework and have proposed alternative measures of risk to the volatility. Most of the researches have focused on the inclusion of the skewness and the kurtosis expressions in the risk definition. For example, Berenyi (2002) suggests using a Taylor series expansion of a general utility function linking the expected utility to the different moments of the distribution. Signer and Favre (2002) and Lamm (2003) favour the delta-gamma approximation via the Cornish-Fisher expansion. Davies, Kat and Lu (2004) propose a Polynomial Goal Programming optimisation in the mean-variance-skewness-kurtosis framework.

While these new approaches represent a step forward in the construction of optimal portfolio of hedge funds, the addition of higher moments is overshadowed by the errors embedded in the estimation of the various statistics. Estimation risk is a well known problem leading to instability in the optimal portfolios; see among others Bawa, Brown and Klein (1979). In the context of hedge funds, the addition of higher moments exacerbates the instability problem as the estimation errors on higher moments are bigger than that of volatility. Indeed, the low number of data available for hedge funds, i.e. monthly returns, is responsible for the large confidence intervals observed for the skewness and kurtosis. The use of estimated skewness and kurtosis may also lead to an overemphasis of certain events, like the LTCM crisis, and may bias portfolio allocation with respect to these events.

This article contributes to the growing literature on higher moment optimisation in two ways. Firstly, we apply a Bayesian framework previously advocated by Harvey, Liechty, Liechty, and Müller (2002) to the hedge fund context. This Bayesian framework permits the modelling of higher moments and of estimation risk via the derivation of a predictive distribution of hedge fund returns. The first three predictive moments are then incorporated in the optimisation of a truncated utility function. Secondly, we analyse the impact of the choice of the underlying distribution within the family of skew elliptical distribution. This family entails in particular the skew normal distribution used by Harvey et al. (2002) and the skew-t distribution. We also consider different parametrisation with respect to the skewnesses and co-skewnesses.

Our paper is organised as follows. Section 2 presents the model for hedge fund returns. In section 3, we discuss the estimation of a predictive distribution in a Bayesian context. The results of the empirical application to hedge funds are reported in section 4. Finally, section 5 concludes this article.

2 Modelling hedge fund returns

Given the asymmetry and strong departure from normality exhibited by hedge fund returns, a skewed probability distribution is needed in order to model these return series. In recent years an increasing number of publications have proposed different methods for the definition of both univariate and multivariate skewed distributions.

Azzalini (1985, 1986) was among the first to introduce and investigate the univariate skew-normal distribution. This family of skewed distributions was later extended by Azzalini and Della Valle (1996) to the multivariate case. The approach used to obtain this kind of distributions is based on conditioning on a latent variable. Using this approach, different definitions and generalizations of skewed distributions have been proposed. A survey of these "hidden truncation" models is provided by Arnold and Beaver (2002). A class of multivariate skew-elliptical distributions was proposed by Branco and Dey (2001) and then improved by Sahu, Branco and Dey (2003) by extending the number of conditioning arguments to match the number of observed variables[2].

The analysis developed in this paper will focus on the class of Skew Elliptical Distributions introduced by Sahu et al. Up to now the use of skewed distribution has been mainly restricted to the skew-normal distribution and applications of more general skew-elliptical distributions have been quite seldom.

General model

Sahu et al. (2003) define a skew-elliptical class of distributions by transforming elliptically symmetric distributed random variables and then conditioning on some latent variables.
The starting point is thus the family of elliptical distributions that can be defined as follows:

$$\mathbf{X} \sim El_{\mathrm{p}}(\boldsymbol{\theta}, \Sigma; g^{(p)})$$

if the p-dimensional random vector X has multivariate density equal to

$$\mathrm{f}_{El_p}(\mathbf{x}; \boldsymbol{\theta}, \Sigma, \mathrm{g}^{(\mathrm{p})}) = |\Sigma|^{-1/2}\, \mathrm{g}^{(\mathrm{p})}\left\{(\mathbf{x}-\boldsymbol{\theta})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})\right\}$$

[2] A more flexible family of skewed distributions has been suggested in Ferreira and Steel (2004).

where

$$g^{(p)}(u) = \frac{\Gamma(p/2)}{p^{p/2}} \frac{g(u;p)}{\int_0^\infty r^{p/2-1} g(r;p)dr}, \ a \geq 0$$

is the so-called density generator function of the random variable x.

and where $g(u;p)$ is a function from $\Re^+$ to $\Re^+$ such that the integral $\int_0^\infty r^{p/2-1}\, g(u;\, p)dr$ exists.

The multivariate normal distribution is a special case of the previous model by setting

$$g(u;\ p) \ = \ g_{Norm} \ := \ e^{-u/2}.$$

Similarly, when choosing

$$g(u;\ p) \ = \ g_t \ := \left(1 + \frac{u}{v}\right)^{-(v+p)/2} \quad \text{with } v \ > \ 0$$

the multivariate *t* distribution is obtained.

Skew elliptical distribution (Sahu et al.)

To obtain the skew-elliptical distribution defined in Sahu et al. (2003) we first consider the transformation:

$$\mathbf{X} = \boldsymbol{\mu} + D\mathbf{Z} + \boldsymbol{\varepsilon}$$

where $\mathbf{Z}$ is a vector of unobserved (latent) random variables whose distribution is elliptical with zero mean and identity covariance matrix I_p.

$\boldsymbol{\mu} \in \Re^p$,

D is a ***p×p*** full matrix

and $\boldsymbol{\varepsilon} \ \sim \ El_p(\mathbf{0},\, \boldsymbol{\Sigma},\, g^{(p)})$

where El_p identifies the elliptical distribution and $g^{(p)}$ the corresponding density generator.

As a consequence, the random variable $\mathbf{Y} = (\mathbf{X}|\mathbf{Z} > \mathbf{0})$ has a multivariate skew-elliptical distribution. The p-dimensional density of the random variable $\mathbf{Y}$ is given by

$$f(\mathbf{y};\, \boldsymbol{\mu},\, \Sigma,\, D,\, g^{(p)}) \ = \ 2^p \ f_{El_p}(\mathbf{y};\, \boldsymbol{\mu},\, \Sigma \, + \, DD',\, g^{(p)})\ P(\mathbf{V} \ > \ \mathbf{0})$$

where f_{El_p} is the pdf of a p-dimensional elliptical distribution,

$$\mathbf{V} \sim El_p(D'(\Sigma + DD')^{-1}(\mathbf{y} - \boldsymbol{\mu}), I_p - D'(\Sigma + DD')^{-1}D; g^{(p)}_{q(y-\mu)}),$$

$$g_a^{(p)}(u) = \frac{\Gamma(p/2)}{p^{p/2}} \frac{g(a+u;2p)}{\int_0^\infty r^{p/2-1} g(a+r;2p)dr}, \; a \geq 0$$

and $q(\mathbf{y}-\boldsymbol{\mu}) = (\mathbf{y}-\boldsymbol{\mu})'(\Sigma + DD')^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

An important feature of this class of distributions is that for any subset of components of $\mathbf{Y}$ the marginal distribution has the same form as the distribution of $\mathbf{Y}$. This coherence property is very important in a portfolio optimisation process. In fact, it ensures that we will obtain the same solution for the portfolio weights even if we remove the assets for whom the weight was calculated to be zero.

Skew-normal distribution

The skew-normal distribution is obtained by defining

$$g(u;p) = g_{Norm} := e^{-u/2}.$$

Thus

$$\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

Skew-*t* distribution

The skew-*t* distribution is obtained by setting:

$$g(u;2p) = g_t := \left(1 + \frac{u}{v}\right)^{-(n+2p)/2}$$

where v is some integer representing the degrees of freedom.

Thus

$$\boldsymbol{\varepsilon} \sim t_{p,v}(\mathbf{0}, \boldsymbol{\Sigma})$$

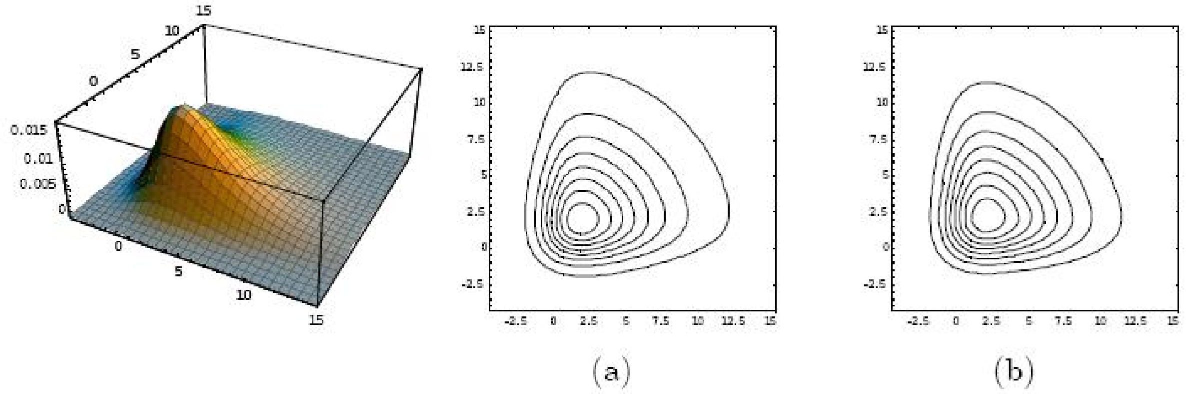


Figure 2-1. Bivariate skew-*t* distribution with ν = 4 , μ = (0, 0), Σ ={(1, 0) ; (0, 1)} , D={(5, 0) ; (0, 5)}; The contour plot of the skew-*t* (a) is compared to the contour plot of a skew-normal distribution (b) with same μ , Σ and D.

Bayesian Inference for Skew Elliptical models

It is now possible to set up a Bayesian inferential procedure where the data is assumed to follow a skew-normal or a skew-*t* distribution. The specification of a model for the MCMC framework can be done by considering a hierarchical model where the likelihood for the *i*-th data observation $\mathbf{Y}_i$ conditional on $\mathbf{Z}_i = \mathbf{z}_i$ is given by

$$\mathbf{Y}_i \big| \mathbf{Z}_i \ = \ \mathbf{z}_\mathrm{i} \ \sim \ \mathrm{El}_\mathrm{p}(\boldsymbol{\mu} \ + \ D\mathbf{z}_i, \ \boldsymbol{\Sigma}; \ g^{(p)}_{q(\mathbf{z}_i)})$$

with $q(\mathbf{z}_\mathrm{i}) = \mathbf{z}_i{}'\mathbf{z}_i$

and where the marginal specification for $\mathbf{Z}_i$ will be a truncated elliptical distribution[3] whose values are restricted to be positive, i.e.

$$\mathbf{Z}_i \ \sim \ \mathrm{El}_\mathrm{p}(\mathbf{0}, \ I_p; \ g^{(p)})\mathbf{1}(\mathbf{Z}_i \ > \ \mathbf{0}).$$

Taking

$$\mathrm{g}(u;2p) = \mathrm{g}_t := \left(1 + \frac{u}{\nu}\right)^{-(n+2p)/2},$$

the skew-*t* model is obtained. Representing the *t* distribution as a scale mixture of normal distributions, the likelihood for each observation can be specified as

$$\mathbf{Y}_i \big| \mathbf{Z}_i \ = \ \mathbf{z}_\mathrm{i} \ \sim \ N_\mathrm{p}(\boldsymbol{\mu} \ + \ D\mathbf{z}_i, \frac{\boldsymbol{\Sigma}}{w_i})$$

---

[3] In particular we will need $\mathbf{Z} \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}) \cdot \mathbf{1}(\mathbf{z} > \mathbf{0})$ for the Skew Normal model and $\mathbf{Z} \sim t_{p,n+p}(\mathbf{0}, \mathbf{I}) \cdot \mathbf{1}(\mathbf{z} > \mathbf{0})$ for the Skew t model. For a proof see Sahu et al. (2003).

$$\mathbf{Z}_i \sim \mathrm{N}_p(\mathbf{0}, \frac{I_p}{w_i}) \cdot \mathbf{1}(\mathbf{Z}_i > \mathbf{0})$$

where[4]

$$w_i \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

and where we assume the following conjugate prior densities for the unknown parameters:

$$\boldsymbol{\mu} \sim \mathrm{N_m}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda})$$

$$\Sigma \sim \mathrm{InvWishart_m}(q, \mathbf{K})$$

$$vec(D)' \equiv \boldsymbol{\delta} \sim \mathrm{N_m}(d_0, \boldsymbol{\Gamma})$$

$$\nu \sim \Gamma(g, \Sigma_\nu)\mathbf{1}(\nu > 2)$$

where $vec(\cdot)'$ returns the vector obtained by stacking the columns of a matrix.

Note that this specification for the skew-*t* model includes the skew-normal as a special case. To obtain the setup for the latter we just need to set $w_i = 1$, $\forall\ i$ and remove the last conjugate prior density.
Diffuse priors on the parameters are set in the model by choosing

$b_0 = 0$, $\quad \boldsymbol{\Lambda} = 100 I_p$

$d_0 = 0$, $\quad \boldsymbol{\Gamma} = 100 I_{p^2}$

$q = 2p$, $\quad \mathbf{K} = pI_p$

$g = 1$, $\quad \Sigma_n = 0.1$

Bayesian estimation can be undertaken using Monte Carlo Markov chains (MCMC) methods such as the Gibbs sampler proposed by Geman and Geman (1984). This method allows producing a Markov chain whose output corresponds to a sample from the joint posterior distribution. In our case we will implement a Gibbs sampler returning samples of the model parameters $\boldsymbol{\mu}, \Sigma, D$ (and ν in the skew-*t* case). In order to implement the sampler we need to draw samples from the full conditional distributions which are specified in Appendix A.
With the same model specification both the skew normal and the skew-*t* models are fitted. In addition, two different settings are specified for each model using a diagonal D matrix and a full D matrix. The full D models are accommodating for co-skewnesses therefore they are expected to provide a better fit when the observed data exhibits significant levels of co-skewness.

---

[4] The parameterisation used is such that $\mathrm{E}(w_i) = 1$.

3 Estimation of predictive distribution

Addressing estimation risk

A limitation of the traditional Markovitz (1952) asset allocation model is that the choice is made between alternative probability distributions where the parameters are assumed to be known. In practice, the distributions are assumed to belong to a certain family and, since the parameters characterizing the model are in general not known, the usual approach consists of replacing the true parameters with their sample estimates that are then plugged in the optimal portfolio formulas. This approach completely ignores the estimation risk that arises both from the parameter uncertainty and from the assumption made on the probability model[5].

Many studies addressing this issue have tried to incorporate estimation risk into the portfolio choice problem. Zellner and Chetty (1965) and Klein and Bawa (1976) are amongst the first to argue for the use of a Bayesian framework in order to improve the estimate using predictive distributions of portfolio returns[6]. Many other authors have focused on the use of the Bayesian predictive approach to account for parameter uncertainty (see Jobson and Korkie (1980), Jorion (1985) and Frost and Savarino (1986) among others). While the mentioned papers are all in the context of i.i.d. returns, Kandel and Stambaugh (1996) and Barberis (2000) point out the importance of recognizing parameter uncertainty also in the context of portfolio allocation with predictable returns and the implications for investors with different investment horizons.

Black and Litterman (1990,1992) and Pástor (2000) improve the bayesian approach by mixing the equilibrium implications from an asset pricing model with the investor views. Pástor and Stambaugh (2000) compare different asset pricing models from the perspective of investors who center their prior beliefs on the models and then update those beliefs with data.

Other authors address the estimation risk issue from different perspectives. Michaud (1998) suggests using of resampling from the estimated distribution in order to deal with estimation error[7]. Xia (2001) studies the effect of parameter uncertainty in a dynamic continuous time context. Kan and Zhou (2003) are providing an analytical comparison of alternative decision rules under estimation risk.

---

[5] See Michaud(1989) for a detailed discussion on the problems arising in implementing mean-variance optimal portfolios.

[6] Bawa, Brown and Klein(1979) provide an extensive survey of the early work on the subject

[7] Scherer (2002) describes this approach and some of its limitations in detail.

The Bayesian approach

The Bayesian approach based on the predictive distributions pioneered by Zellner and Chetty (1965) provides a general framework that integrates estimation risk into the analysis. In the Bayesian decision rule, uncertainty about the parameters is summarised by the posterior distribution of the parameters given the observed returns $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$. Integrating out the parameters over this distribution gives the predictive distribution for future asset returns.

Let W_0 be the initial investor's wealth and $W(\omega, \mathbf{y}_{n+1}) = (1 + \omega' \mathbf{y}_{n+1})W_0$ the next period wealth. Denoting the investor's utility function by $u[W(\omega, \mathbf{y}_{n+1})]$ and the conditional distribution of the future returns by $\mathrm{f}(\mathbf{y}_{n+1}|\boldsymbol{\theta})$, the conditional expected utility of portfolio ω is given by

$$E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\Big[u[W(\omega, \mathbf{y}_{n+1})] \mid \boldsymbol{\theta}\Big] \equiv \int_{\mathbf{y}_{n+1}} u[W(\omega, \mathbf{y}_{n+1})] \; f(\mathbf{y}_{n+1}|\boldsymbol{\theta}) \; d\mathbf{y}_{n+1} .$$

Assuming that the sample parameters will exactly match the true parameters (i.e.,$q = \hat{q}$), the optimal portfolio weights $\hat{\omega}$ in the classical framework will be obtained as

$$\hat{\omega} = \underset{\omega \in \boldsymbol{C}_\omega}{\operatorname{argmax}} \left\{ E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\Big[u[W(\omega, \mathbf{y}_{n+1})] \mid \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}\Big] \right\}$$

$$\equiv \underset{\omega \in \boldsymbol{C}_\omega}{\operatorname{argmax}} \left\{ \int_{\mathbf{y}_{n+1}} u[W(\omega, \mathbf{y}_{n+1})] \; f(\mathbf{y}_{n+1}|\hat{\boldsymbol{\theta}}) \; d\mathbf{y}_{n+1} \right\}$$

where $\boldsymbol{C}_\omega$ is a given set of constraints on the portfolio weights.

In contrast, the Bayesian approach considers the predictive distribution of future returns given the observed data:

$$p(\mathbf{y}_{n+1}|\mathbf{y}) \equiv E_{\boldsymbol{\theta}}\left[f(\mathbf{y}_{n+1}|\boldsymbol{\theta})\right] = \int_{\boldsymbol{\theta}} p(\mathbf{y}_{n+1}|\boldsymbol{\theta}) \; p(\boldsymbol{\theta}|\mathbf{y}) \; d\boldsymbol{\theta}$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution of $\boldsymbol{\theta}$.

The investor's optimal portfolio solution in the Bayesian framework is given by

$$\hat{\omega}_{Bayes} = \underset{\omega \in \boldsymbol{C}_\omega}{\operatorname{argmax}} \left\{ E_{\boldsymbol{\theta}}\left[E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\Big[u[W(\omega, \mathbf{y}_{n+1})] \mid \boldsymbol{\theta}\Big]\right] \right\}$$

$$\equiv \underset{\omega \in \boldsymbol{C}_\omega}{\operatorname{argmax}} \int_{\mathbf{y}_{n+1}} \int_{\boldsymbol{\theta}} u\left[W(\omega, \mathbf{y}_{n+1})\right] \, f(\mathbf{y}_{n+1} | \boldsymbol{\theta}) \; f(\boldsymbol{\theta} | \mathbf{y}) \; d\boldsymbol{\theta} \; d\mathbf{y}_{n+1}$$

$$\equiv \underset{\omega \in \boldsymbol{C}_\omega}{\operatorname{argmax}} \int_{\mathbf{y}_{n+1}} u\left[W(\omega, \mathbf{y}_{n+1})\right] \, p(\mathbf{y}_{n+1} | \mathbf{y}) \; d\mathbf{y}_{n+1} \; .$$

The investor which follows a Bayesian allocation decision is maximising expected utility of the future wealth by calculating the expectation with respect to the predictive distribution of the future returns. The Bayesian framework is explicitly considering the estimation risk since the parameters are modelled as random variables themselves. As shown by Bawa and Klein (1976) the introduction of estimation risk alters the optimal portfolio choice. Kan and Zhou (2003) show that in the mean variance framework, the Bayesian solution for the portfolio decision problem is more conservative for risk averse individuals than the case where the parameters are known. In the mean-variance framework a Bayesian approach will suggest investing more on the riskless asset. Intuitively, the Bayesian decision rule recognizes the estimation risk and hence is considering an additional source of risk on the risky assets. Thus, the riskless asset becomes more attractive.

Predictive estimates.

The predictive moments from the means of the posterior moments can be calculated as follows:

$$m_p = \overline{m}$$

$$V_p = \overline{V} + Var(m | \mathbf{y})$$

$$S_p = \overline{S} + 3 \cdot E(V \otimes m | \; \mathbf{y}) - 3 \cdot E(V | \; \mathbf{y}) \otimes m_p - E\left[(m - m_p) \otimes (m - m_p) | \; \mathbf{y}\right]$$

where

$$m_p = \int_{-\infty}^{\infty} \mathbf{y}_{n+1} \; p(\mathbf{y}_{n+1} | \mathbf{y}) d\mathbf{y}_{n+1} \; ,$$

V_p **and** S_p **are the second and third[8] predictive central moments respectively and** $\overline{m}$ **,** $\overline{V}$ **,** $\overline{S}$ **are the means of the posterior moments** m **,** V **and** S **.**

Truncated utility function

In order to capture the effect of skewness, a proper utility function must be considered. The portfolio return for the period is defined as $r_p = \omega' \, \mathbf{y}_{n+1}$**. Setting the initial wealth equal to one**

---

[8] The same representation as in Athayde and Flores (2001) is used for the third central moment tensor.

gives for the next period investor's wealth $W = (1 + r_p)$. The allocation problem for the investor maximizing his expected utility can be stated as

$$\max_{\omega \in \mathbf{C}_\omega} E(u(W))$$

under the constraint $\mathbf{C}_\omega = \{\omega'\mathbf{1} = 1, \omega \geq \mathbf{0}\}$.

In order to describe the agents' preferences a family of linear utility functions is considered, i.e.

$$u(1 + \omega' \mathbf{y}_{n+1}) = \omega' \mathbf{y}_{n+1} - \lambda\left[\omega' (\mathbf{y}_{n+1} - m_p)\right]^2 + \gamma\left[\omega' (\mathbf{y}_{n+1} - m_p)\right]^3$$

where $m_p = \int_{-\infty}^{\infty} \mathbf{y}_{n+1} \; p(\mathbf{y}_{n+1}|\mathbf{y}) d\mathbf{y}_{n+1}$ is the predictive mean of future returns $\mathbf{y}_{n+1}$.

The expected utility calculated with respect to the predictive density is:

$$E(u(W)) = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega S_p \omega \otimes \omega$$

This can be estimated by generating a set of G draws $\mathbf{y}_{n+1}^{(g)} \sim p(\mathbf{y}_{n+1}|\mathbf{y})$ from the posterior predictive distribution and then estimating the predictive moments over the generated sample, i.e.,

$$m_p = \int_{-\infty}^{\infty} \mathbf{y}_{n+1} \; p(\mathbf{y}_{n+1}|\mathbf{y}) d\mathbf{y}_{n+1} \approx \frac{1}{n} \sum_{g=1}^{G} \mathbf{y}_{n+1}^{(g)}$$

For an arbitrary utility function, the expected utility can be approximated via Monte Carlo simulation by averaging the utility over the predictive samples, i.e.

$$E(u(W)) = E\left[u(1 + \omega' \mathbf{y}_{n+1})\right] = \int_{-\infty}^{\infty} u(1 + \omega' \mathbf{y}_{n+1}) \; p(\mathbf{y}_{n+1}|\mathbf{y}) \; d(\mathbf{y}_{n+1}) \approx \frac{1}{n} \sum_{g=1}^{G} u(1 + \omega' \mathbf{y}_{n+1}^{(g)})$$

The expected utility can then be optimised with numerical methods.

Specifying additional information in the model

One of the advantages of the Bayesian approach is that it allows reflecting subjective views or financial information about the future returns. Many studies have already taken advantage of this feature of the Bayesian approach by the inclusion of subjective information in the priors. In the framework described by Black and Litterman (1990, 1992) the investors combine individual views with market equilibrium to select their portfolios. Pástor (2000) proposes a portfolio selection methodology that includes in the priors the investor's degree of confidence in an asset pricing model.

In the hedge fund context this feature allows to include in the asset allocation procedure all the information arising from the analysis of the strategies adopted in the return-generating process by the hedge fund managers. In particular, information concerning the direction of the skewness of the returns generated by the different hedge funds styles is available from the analysis of the various hedge fund strategies. By appropriately adjusting the prior on the parameter that regulates the skewness is then possible to include this information in the model.

This feature is particularly valuable in a context where the data is scarce. Here, often, the full behaviour of the return generating process has not yet been disclosed to the data. Some of this undisclosed information can be read from the analysis of the process strategy and included in the asset allocation procedure. In a Bayesian framework, this prior information will be combined with the information arising from the data to produce the posterior predictive estimates for the parameters of the model.

4 Application to hedge fund returns

Data description

We consider a set of returns on hedge funds indices provided by Hedge Fund Research, Inc. (HFRI). The data provided by HFRI is based on a database containing the returns of 3'700 funds. The monthly data is composed by 4 non-overlapping HFRI strategy indices, representing the equally weighted returns, net of fees, of hedge funds classified in each strategy. The indices selected are:

- HFRI Equity Hedge Index (EH),
- HFRI Relative Value Arbitrage Index (RV),
- HFRI Event-Driven Index (ED),
- HFRI Macro Index (GM).

In addition, to represent an investment opportunity in the managed futures strategy, the Stark 300 Trader Index (MF) is included in the dataset.

The data series consists of the monthly observations for the above indices between January 1994 and March 2005. As a preliminary investigation of the data some summary statistics of the returns are provided in Table 1.

Table 1. Summary statistics for the hedge fund monthly returns.

	Equity hedge	Relative value	Event driven	Global macro	Managed futures
Mean	1.17%	0.79%	1.09%	0.84%	0.63%
St.dev.	2.63%	0.91%	1.88%	2.14%	2.86%
Skewness	0.324	-2.663	-1.272	0.058	0.290
Kurtosis	1.875	18.960	5.468	0.861	-0.030

Estimation results

In the Bayesian framework, the predictive moments have been calculated for both the skew-normal and the skew-*t* distributions. For both models two different settings for the parameter ***D*** are used. In the first setting the parameter is defined as a diagonal matrix and in the second setting as a full matrix. The Bayesian estimation is thus performed for four models:

§ Skew-normal (diagonal D)
§ Skew-normal (full D)
§ Skew-*t* (diagonal D)
§ Skew-*t* (full D)

The models where the D parameter is defined as a diagonal matrix have the advantage that the definition of informative priors for the skewness of the distribution is easily implemented as the diagonal elements of D are the only parameters adjusting the amount of skewness of the distribution. When D is specified as a full matrix, the co-skewnesses are modelled more accurately although the definition of priors on the parameter becomes difficult to implement.

The estimation is done using a Gibbs sampler implemented in WinBUGS[9]. A total of 12'000 iterations have been undertaken for each model with a burn-in of 20'000 iterations each. After this number of iterations, MC-errors for the estimated parameters are usually small and convergence for the MCMC sampler seems to have been achieved.

In order to check the goodness of fit of each model to the observed data we rely on the approach based on posterior predictive assessment. This approach, introduced by Guttman (1967) and Rubin (1984) and further discussed in Gelman and Meng (1996) and Gelman, Meng and Stern (1993), provides an intuitive and easy to implement method to estimate the p-values of the observed data given a certain model.

The method relies on the draws $\theta_1, \ldots, \theta_G$ from the posterior distribution generated by the MCMC. Replications of the data $\mathbf{y}_g^{rep} = (\mathbf{y}_{g,1}, \ldots, \mathbf{y}_{g,n}),\ g = 1, \ldots, G$ are generated for each simulated parameter by drawing from the sampling distribution given the parameters (predictive distribution). The simulated replications are then compared to the observed values $\mathbf{y}$ by computing the estimated p-value of a test statistic $T(\mathbf{y})$. The estimated p-value is calculated as the proportion of cases in which the simulated test statistics exceed the realized value:

$$p_{val} = \frac{1}{G}\sum_{i=1}^{G} \mathbf{1}\left\{T(\mathbf{y}_i^{rep}) > T(\mathbf{y})\right\}$$

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

A p-value too close to 1 or to 0 will be the evidence that the model is not accurately fitting some aspects of the data related to the test statistic which is used.

[9] Spiegelhalter et al. (2003)

Given the complexity of the models involved and the high number of parameters to be fitted, we believe that this kind of assessment is particularly needed in the current framework and may be used to discriminate between the various models. The results of the analysis on the four models fitted to our hedge fund data are confirming this intuition highlighting a statistically significant lack of fit for some of the models. In the context discussed in this paper, some lack of fit concerning particularly important statistics like the first three moments, may be used as additional evidence to give more strength to the choice in favour of one model rather than another from a model selection perspective.

Table 2 displays the p-values for the four models implemented. Mean, variance and skewness together with minimum and maximum values are used as test statistics. The results show that models with a diagonal D matrix have substantial difficulties in fitting negatively skewed data. In fact, for the relative value and event driven styles, all the generated samples have higher skewness values than the (negative) skewness of the observed data.

When a full matrix is used for D, the fit is improved although the p-values obtained for minimum value and skewness for the relative value and event driven styles are still high. In this regard, we might consider that our data is including the observations relative to a severe hedge fund crisis occurred in the year 1998 (LCTM crisis). The crisis affected negatively both the event driven and the relative value styles managers and had a more severe impact on the last ones. Since the returns produced during this period are reflecting the outcome of an extreme event, the fact that only a small percentage of the data replications are exhibiting lower minimum value and skewness than the observed data may not be entirely considered as lack of fit. This intuition is confirmed by the p-values obtained using a measure of skewness which is less sensitive to outliers[10].

Table 2. Model checking results

Skew-normal (diagonal *D*) p-values of observed statistics

	EH	RV	ED	GM	MF
Mean	0.509	0.505	0.504	0.507	0.507
Min.value	0.971	1.000	1.000	0.926	0.391
Max.value	0.047	0.846	0.868	0.323	0.464
Variance	0.368	0.571	0.514	0.513	0.500
Skewness	0.204	1.000	1.000	0.368	0.100
Robust Skew.	0.559	0.588	0.959	0.047	0.141

Skew-t (diagonal *D*) p-values of observed statistics

	EH	RV	ED	GM	MF
Mean	0.537	0.568	0.581	0.494	0.446
Min.value	0.866	0.996	0.968	0.721	0.232
Max.value	0.272	0.815	0.862	0.588	0.738
Variance	0.327	0.200	0.308	0.519	0.643
Skewness	0.476	0.997	0.978	0.479	0.340
Robust Skew.	0.670	0.623	0.938	0.073	0.227

Skew-normal (full *D*) p-values of observed statistics

	EH	RV	ED	GM	MF
Mean	0.504	0.501	0.506	0.497	0.499
Min.value	0.484	0.989	0.768	0.888	0.573
Max.value	0.257	0.682	0.650	0.444	0.772
Variance	0.569	0.695	0.587	0.595	0.568
Skewness	0.244	0.999	0.851	0.501	0.524
Robust Skew.	0.606	0.327	0.837	0.089	0.393

Skew-t (full *D*) p-values of observed statistics

	EH	RV	ED	GM	MF
Mean	0.490	0.557	0.557	0.523	0.516
Min.value	0.582	0.980	0.797	0.759	0.441
Max.value	0.295	0.785	0.783	0.565	0.808
Variance	0.498	0.311	0.483	0.542	0.656
Skewness	0.337	0.993	0.847	0.559	0.555
Robust Skew.	0.646	0.530	0.856	0.130	0.415

[10] The robust measure of skewness used here is the one proposed by Groeneveld and Meeden (1984).

Model selection

To compare the models we compute the deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002) and implemented in the BUGS software. Although there is some criticism concerning the use of the DIC to select the best fitting model, the results are in line with the conclusions drawn from the p-values used to assess the goodness of fit of the models. DIC values for the four model specifications are displayed in table 3.

Table 3. Deviance information criterion of the implemented models

Model	DIC
Skew-normal (diagonal D)	3127.13
Skew-t (diagonal D)	2770.50
Skew-normal (full D)	1832.84
Skew-t (full D)	1890.14

The models based on a full D matrix which allows for co-skewnesses are providing a better fit. The skew-normal model (full D) seems to be the best model for the data.

The posterior means for the parameters of the skew-normal (full D) are displayed in Table 4. Appendix B shows the averages of the first three moments over 12’000 replications of the data sampled from the fitted skew-normal (full D) model.

Table 4. Posterior means of the skew-normal (full D) parameters

Posterior mean of μ				
0.6391	1.397	2.371	0.5682	-2.173

Posterior mean of D				
-0.9621	0.3578	1.198	0.8726	-0.7969
-0.5043	-0.02614	0.08169	0.02091	-0.3758
-1.064	-0.1779	0.4158	0.1492	-1.001
-0.5518	0.2026	0.891	0.377	-0.5827
0.6699	0.6405	0.8925	0.5674	0.8406

Posterior mean of Σ				
1.171	0.1118	0.5328	0.5645	0.2456
0.1118	0.3771	0.18	0.09973	0.06598
0.5328	0.18	0.8583	0.4836	0.4192
0.5645	0.09973	0.4836	1.909	1.034
0.2456	0.06598	0.4192	1.034	2.528

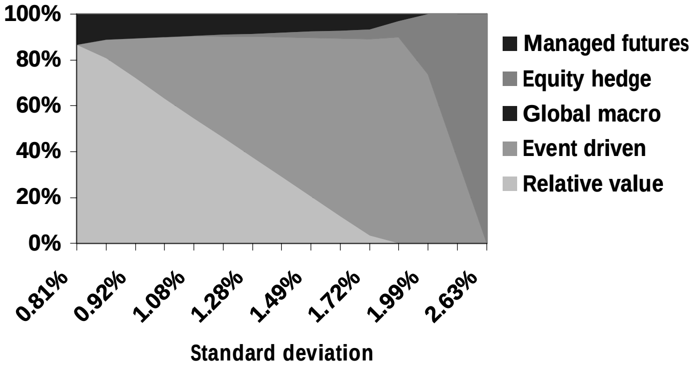
Optimisation results

In this section, the efficient frontier portfolios from the classical mean-variance framework computed using the hedge fund monthly return data are compared with the efficient frontier allocations resulting from the optimisation framework based on the first three moments.

Table 5 displays the allocation chart, mean return and standard deviation (monthly figures) of the fifteen portfolios used to approximate the efficient frontier in a mean-variance optimisation framework.

Table 5. Mean-variance efficient frontier

Portfolio	Mean return	Standard deviation	Skewness
1	0.76%	0.81%	-1.44
2	0.79%	0.86%	-1.67
3	0.82%	0.92%	-1.62
4	0.85%	1.00%	-1.52
5	0.88%	1.08%	-1.42
6	0.91%	1.18%	-1.32
7	0.94%	1.28%	-1.23
8	0.97%	1.38%	-1.15
9	1.00%	1.49%	-1.08
10	1.03%	1.61%	-1.03
11	1.05%	1.72%	-0.99
12	1.08%	1.84%	-1.08
13	1.11%	1.99%	-0.90
14	1.14%	2.26%	-0.24
15	1.17%	2.63%	0.32

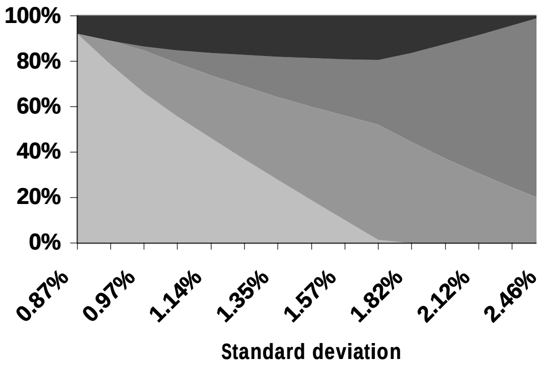


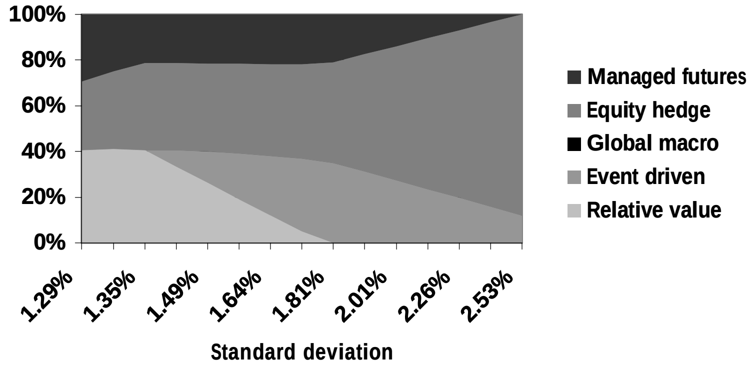
The three moment optimisation is based on the predictive estimates of the mean, covariance matrix and third central moment tensor obtained from the skew-normal full D model in the Bayesian framework. To approximate the efficient frontier, the minimum and maximum third central moment values attainable from a portfolio satisfying the budget and no-short-selling constraints[11] are calculated. Then, the interval of the attainable third central moment values is spanned by picking the fifteen values dividing the difference between the minimum and the maximum in equal intervals. A mean-variance efficient frontier is then obtained for each fixed value of the third central moment. This is done by imposing an additional constraint on the third central moment of the portfolios which is constrained to match the selected fixed value.

Each constrained mean-variance frontier is approximated by fifteen portfolios spanning the attainable returns. Thus, we end up with a three dimensional efficient frontier approximated by 15^2 portfolios. As an example, Table 6 displays the allocation charts and some statistics for two of the fifteen efficient frontiers constrained on the third central moment value.

[11] This corresponds to the constraints set $\boldsymbol{C}_{\omega} = \{\omega'\mathbf{1} = 1, \omega \geq \mathbf{0}\}$

Table 6. Efficient frontier portfolios for two different values of the third central moment





Portfolio	Mean return	Standard deviation	3rd central moment	Skewness
1	0.77%	0.87%	-3.8E-07	-0.58
2	0.80%	0.90%	-3.8E-07	-0.52
3	0.83%	0.97%	-3.8E-07	-0.42
4	0.86%	1.05%	-3.8E-07	-0.33
5	0.88%	1.14%	-3.8E-07	-0.26
6	0.91%	1.24%	-3.8E-07	-0.20
7	0.94%	1.35%	-3.8E-07	-0.15
8	0.97%	1.46%	-3.8E-07	-0.12
9	0.99%	1.57%	-3.8E-07	-0.10
10	1.02%	1.69%	-3.8E-07	-0.08
11	1.05%	1.82%	-3.8E-07	-0.06
12	1.08%	1.96%	-3.8E-07	-0.05
13	1.10%	2.12%	-3.8E-07	-0.04
14	1.13%	2.28%	-3.8E-07	-0.03
15	1.16%	2.46%	-3.8E-07	-0.03

Portfolio	Mean return	Standard deviation	3rd central moment	Skewness
1	0.85%	1.29%	6.0E-07	0.28
2	0.88%	1.30%	6.0E-07	0.27
3	0.90%	1.35%	6.0E-07	0.24
4	0.92%	1.42%	6.0E-07	0.21
5	0.94%	1.49%	6.0E-07	0.18
6	0.96%	1.56%	6.0E-07	0.16
7	0.99%	1.64%	6.0E-07	0.13
8	1.01%	1.72%	6.0E-07	0.12
9	1.03%	1.81%	6.0E-07	0.10
10	1.05%	1.91%	6.0E-07	0.09
11	1.07%	2.01%	6.0E-07	0.07
12	1.10%	2.13%	6.0E-07	0.06
13	1.12%	2.26%	6.0E-07	0.05
14	1.14%	2.39%	6.0E-07	0.04
15	1.16%	2.53%	6.0E-07	0.04

The minimum third central moment portfolios are fully allocated to the event driven style and the maximum third central moment portfolios are 100% allocated to managed futures. When the constraint on the third central moment is introduced the resulting portfolios are in general loosing their efficiency in the mean-variance space. Although, comparing the efficient frontier portfolios obtained in the mean-variance framework with the ones from the three moment optimisation from Table 6, we notice that the latter exhibit a higher diversification for small differences in mean and standard deviation. The results in Table 6 show that in order to increase skewness the portfolio standard deviation is increased or the expected return reduced. This reflects the fact that the investor may give up something on the risk-return side in order to improve the skewness of the portfolio.

5 Conclusions

This article proposes an application of Bayesian allocation techniques to the portfolio selection problem in the hedge fund context. Given the strong departures from normality of the hedge fund returns and the short data series available in the hedge fund context, both the inclusion of higher moments and the parameter uncertainty need to be addressed in the portfolio selection task.

Parameter estimates for the hedge fund return distribution are obtained from four different model specifications.

- § Skew-normal (diagonal D)
- § Skew-normal (full D)
- § Skew-*t* (diagonal D)
- § Skew-*t* (full D)

In our analysis the estimates produced by the full D models are providing a better fit to the data than the diagonal D models. In the latter models, in fact, co-skewnesses are not modelled accurately, although these models have the advantage that the definition of priors on the D parameter becomes easier and more intuitive. The skew-normal full D model is providing the best fit to the hedge fund return data used in the analysis. One of the drawbacks of the full D models is that the specification of investor's views in the prior, one of the most important features of the Bayesian framework, is more difficult to implement. An improvement that could be addressed in future research is the development of a model based on the skewed class of distributions recently proposed by Ferreira and Steel (2004). This family of skewed distributions is more flexible than the one proposed by Sahu et al. (2003) and its characteristics could provide an improvement in fitting the return distributions.

Our results confirm that introducing skewness in the asset allocation task will produce a different allocation for investors with skewness preference. An investor wanting to improve the skewness profile of his portfolio, will loose the portfolio efficiency in the risk-return space. Although, the additional constraint imposed on the third central moment may produce more diversified allocations if compared to the allocations obtained in a mean-variance framework for similar risk-return characteristics.

References

Arnold, B. C. and R. J. Beaver , 2002, "Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion)," Sociedad de Estadistica e Investigacion Operativa, Test, 11, 1, 7-54.

Athayde, G. M., and R. G. Flores Jr., 2001, "Finding a maximum skewness portfolio," Computing in Economics and Finance 2001, 273.

Azzalini, A., 1985, "A class of distributions which includes the normal ones," Scandinavian Journal of Statistics, 12,171-17.

Azzalini, A., 1986, "Further results on a class of distributions which includes the normal ones," Statistica, 46, 199-208.

Azzalini, A.,and A. Dalla Valle, 1996, "The Multivariate Skew-Normal Distribution," Biometrika, 83, 715-726.

Barberis, N., 2000, "Investing for the long run when returns are predictable," Journal of finance, 55, 225-264

Bawa, V. S., S. J. Brown and R. W. Klein, 1979, "Estimation Risk and Optimal Portfolio Choice," Amsterdam: North-Holland.

Berenyi, Z., 2002, "Measuring hedge fund risk with multi-moment risk measures", Working paper (Universtiy of Munich).

Black, F. and R. Litterman, 1990, "Asset Allocation: Combining Investor Views with Market Equilibrium," Discussion paper, Goldman, Sachs & Co.

Black, F. and R. Litterman, 1992, "Global Portfolio optimisation," Financial Analysts Journal, 48, 28-43.

Branco, M. D.,and D. K. Dey, 2001, "A General Class of Multivariate Skew-Elliptical Distributions," Journal of Multivariate Analysis, 79, 99-113.

Brooks, C. and H. Kat, 2002, "The statistical properties of hedge fund index returns and their implications for investors," Journal of Alternative Investments 5, 2.

Davies, R., Kat H. and S. Lu, 2004, "Fund of hedge funds portfolio selection: a multiple objective approach", Working paper (City University London).

Fung, W. and D. Hsieh, 1999, "Is mean-variance analysis applicable to hedge funds?", Economic Letters 62.

Frost, P. A. and J. E.Savarino,1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," Journal of Financial and Quantitative Analysis, 21, 3, 293-305.

Ferreira, J.T.A.S. and M.F.J. Steel, 2004, 2Bayesian multivariate regression analysis with a new class of skewed distributions," Mimeo, University of Warwick.

Gelman, A. and Meng, X. L., 1996, "Model checking and model improvement," In *Practical Markov Chain Monte Carlo*, (Edited by W. Gilks, S. Richardson, and D. Spiegelhalter),189-201. Chapman and Hall, London.

Gelman, A., X. L. Meng and Stern, H. S., 1993, "Bayesian model checking using tail area probabilities," Technical report, Department of Statistics, University of California, Berkeley.

Geman, S. and D. Geman, 1984, "Stochastic relaxation, Gibbs Distributions and the Bayesian restoration of Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

Groeneveld, R.A. and Meeden, G. ,1984, "Measuring Skewness and Kurtosis," The Statistician, 33, 391-399.

Guttman, I., 1967, "The use of the concept of a future observation in goodness-of-fit problems," Journal of the Royal Statistical Society Series B, 29, 83-100.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller, 2002, "Portfolio Selection with Higher Moments: A Bayesian Decision Theoretic Approach," Working Paper, Duke University Durham.

Jobson, J. D. and B. Korkie, 1980, "Estimation of Markowitz efficient portfolios," Journal of the American Statistical Association, 75, 544-554.

Jorion, P., 1985, "International portfolio diversification with estimation risk," Journal of Business, 58, 259-278.

Kan, R.and G. Zhou, 2004, "Optimal Estimation for Economic Gains:Portfolio Choice with Parameter Uncertainty,"Working paper,Washington University,St.Louis.

Kandel, S.and R. F. Stambaugh, 1996, "On the predictability of stock returns: An asset allocation Perspective," Journal of Finance, 51, 2, 385-424.

Klein, R. W. and V. Bawa, 1976, "The effect of estimation risk on optimal portfolio choice," Journal of Financial Economics, 3, 215-231.

Lamm, McFall R., 2003, "Asymmetric returns and optimal hedge fund portfolios", Journal of Alternative Investments 6 (2).

Markowitz, H., 1952, "Mean-variance analysis in portfolio choice and capital markets," Journal of Finance, 7, 77-91.

Michaud, R., 1998, "Efficient Asset Management," Cambridge: Harvard Business School Press

Pástor, L., 2000, "Portfolio selection and asset pricing models," Journal of Finance, 55, 179 — 223.

Pástor, L., and R. F. Stambaugh, 2000, "Comparing asset pricing models: An investment perspective," Journal of Financial Economics, 56, 335 —381.

Polson, N. G. and B. V. Tew, 2000, "Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970-1996," Journal of Business and Economic Statistics, 18, 2, 164 -73.

Rubin, D.B., 1984, "Bayesianly justifiable and relevant frequency calculations for the applied statistician," Annals of Statistics, 12, 1151-1172.

Sahu, S. K., D. K. Dey, and M. D. Branco, 2003, "A new class of multivariate skew distributions with applications to Bayesian Regression Models," The Canadian Journal of Statistics, 31, 129-150.

Scherer, B., 2002, "Portfolio Resampling: Review and Critique," Financial Analysts Journal, 58, 6.

Schmidhuber, C. and P. Y. Moix, 2001, "Fat tail risk: the case for hedge funds," AIMA Newsletter (sept-dec).

Signer, A. and L. Favre, 2002, "The difficulties of measuring the benefit of hedge funds", Journal of Alternative Investments 5 (1).

Spiegelhalter, D. J., N.G. Best and Carlin, B. P. and van der Linde, A., 2002, "Bayesian measures of model complexity and fit (with discussion)," Journal of the Royal Statistical Society Series B, 64, 583-639.

Spiegelhalter, D. J., A. Thomas and N. G. Best, 2003, "WinBUGS Version 1.4 User Manual," Cambridge: Medical Research Council Biostatistics Unit. (Available from http://www.mrc-bsu.cam.ac.uk/bugs).

Xia, Y., 2001, "Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation," Journal of Finance, 56, 205-246.

Zellner, A. and V. K. Chetty, 1965, "Prediction and decision problems in regression models from the Bayesian point of view," Journal of the American Statistical Association, 60, 608-616.

Appendix A

Full conditional distributions for the skew-*t* model

Define: $\xi_i = (\mathbf{I}_m, \mathbf{z}_i' \otimes \mathbf{I}_m)$ **and** $\boldsymbol{\beta}^* = (\boldsymbol{\mu}', \mathrm{vec}(D))$

Then

$$\xi_i \boldsymbol{\beta}^* = \boldsymbol{\mu} + D\mathbf{z}_i$$

Since

$\boldsymbol{\beta}^* \sim \mathrm{N}_{m(m+1)}(\boldsymbol{\beta}_0^*, \boldsymbol{\Lambda}^*)$, **with** $\boldsymbol{\beta}_0^* = (\boldsymbol{\beta}_0, d_0)$ **and** $\boldsymbol{\Lambda}^* = \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix}$

we obtain:

$$\boldsymbol{\beta}^* \big| \mathbf{Y}, \mathbf{Z}, \boldsymbol{\Sigma}, w_i, n \sim \mathrm{N}_{m(m+1)}\{\mathbf{B}^{-1}\mathbf{b}, \mathbf{B}^{-1}\}$$

$$\boldsymbol{\Sigma} \big| \mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, D, w_i, n \sim \mathrm{InvWishart}_{\mathrm{m}} \left\{ (q+n), \left(\mathbf{K} + \sum_{i=1}^{n} \mathbf{y}_i^{*\prime} w_i \mathbf{y}_i^* \right) \right\}$$

$$\mathbf{z}_i \big| \mathbf{Y}, \boldsymbol{\mu}, D, \boldsymbol{\Sigma}, w_i, n \sim \mathrm{N}_{\mathrm{m}}(\mathbf{A}^{-1}\mathbf{a}, \mathbf{A}^{-1}) \cdot 1(\mathbf{Z} > \mathbf{0})$$

$$w_i \big| \mathbf{Y}, \boldsymbol{\mu}, D, \boldsymbol{\Sigma}, w_i, n \sim \Gamma\left(\left(m + \frac{n}{2} \right), \left(\frac{\mathbf{y}_i^{*\prime} \Sigma^{-1} \mathbf{y}_i^* + \mathbf{z}_i' \mathbf{I} \mathbf{z}_i}{2} \right) \right)$$

$$\mathrm{p}(n \big| \mathbf{Y}, \mathbf{z}_i, \boldsymbol{\mu}, D, \boldsymbol{\Sigma}, w_i) \propto \left(\frac{n}{2} \right)^{\frac{nn}{2}} \frac{1}{\Gamma(n/2)} \prod_{i=1}^{n} \left[w_i^{(n/2-1)} \exp\left\{ -w_i \frac{n}{2} \right\} \right] \exp\{-n\Sigma_n\} \cdot n^{(g-1)}$$

with

$$\mathbf{y}_i^* = [\mathbf{y}_i - (\boldsymbol{\mu} + D\mathbf{z}_i)]$$

$$\mathbf{a} = [D\boldsymbol{\Sigma}^{-1} w_i (\mathbf{y}_i - \boldsymbol{\mu} - D\mathbf{z}_i)]$$

$$\mathbf{A}^{-1} = (D'\boldsymbol{\Sigma}^{-1} w_i D + w_i \mathbf{I})^{-1}$$

$$\mathbf{b} = \sum_{i=1}^{n} \xi_i' \boldsymbol{\Sigma}^{-1} w_i \mathbf{y}_i + \boldsymbol{\Lambda}^{*-1} \boldsymbol{\beta}_0^*$$

$$\mathbf{B}^{-1} = \left(\sum_{i=1}^{n} \xi_i{}' \boldsymbol{\Sigma}^{-1} w_i \xi_i + \boldsymbol{\Lambda}^{*-1} \right)^{-1}$$

Note:

The full conditional of n is not a known distribution: a sampling algorithm (e.g., Slice sampling or rejection envelope method) is needed in order to draw the samples.

The specification for the skew-*t* model includes as a special case also the model for the skew-normal. The full conditionals for the latter are obtained by setting $w_i = 1, \forall\ i$. The full conditional distributions for w_i and n are not needed in the skew-normal case.

Appendix B

Skew-normal (full D) model

Averages of the first three moments over 12'000 replications of the data (Y_{rep}) compared with sample moments of the data (Y)

Mean(Y_{rep})/1E-02

EH	RV	ED	GM	MF
1.172	0.786	1.094	0.842	0.625

Mean(Y)/1E-02

EH	RV	ED	GM	MF
1.171	0.786	1.092	0.843	0.625

Covariance(Y_{rep})/1E-04

	EH	RV	ED	GM	MF
EH	7.086	1.436	4.051	3.451	-0.518
RV	1.436	0.907	1.256	0.777	-0.484
ED	4.051	1.256	3.641	2.392	-0.526
GM	3.451	0.777	2.392	4.785	2.753
MF	-0.518	-0.484	-0.526	2.753	8.390

Covariance(Y)/1E-04

	EH	RV	ED	GM	MF
EH	6.940	1.422	4.008	3.399	-0.527
RV	1.422	0.837	1.239	0.766	-0.488
ED	4.008	1.239	3.536	2.352	-0.551
GM	3.399	0.766	2.352	4.592	2.682
MF	-0.527	-0.488	-0.551	2.682	8.165

Skewness(Y_{rep})

EH	RV	ED	GM	MF
0.140	-0.879	-0.894	0.055	0.303

Skewness(Y)

EH	RV	ED	GM	MF
0.320	-2.633	-1.258	0.058	0.287

Third central moment: $m_3(Y_{rep})$ /1E-06

	EH					RV					ED					GM					MF				
	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF
EH	2.642	-1.893	-3.530	1.671	7.192	-1.893	-1.334	-2.661	-0.768	2.839	-3.530	-2.661	-5.512	-1.098	6.639	1.671	-0.768	-1.098	1.132	3.365	7.192	2.839	6.639	3.365	-4.225
RV	-1.893	-1.334	-2.661	-0.768	2.839	-1.334	-0.764	-1.483	-0.629	1.319	-2.661	-1.483	-2.965	-1.128	2.837	-0.768	-0.629	-1.128	-0.329	1.288	2.839	1.319	2.837	1.288	-2.389
ED	-3.530	-2.661	-5.512	-1.098	6.639	-2.661	-1.483	-2.965	-1.128	2.837	-5.512	-2.965	-6.229	-2.078	6.407	-1.098	-1.128	-2.078	-0.353	2.891	6.639	2.837	6.407	2.891	-5.100
GM	1.671	-0.768	-1.098	1.132	3.365	-0.768	-0.629	-1.128	-0.329	1.288	-1.098	-1.128	-2.078	-0.353	2.891	1.132	-0.329	-0.353	0.573	1.848	3.365	1.288	2.891	1.848	-0.992
MF	7.192	2.839	6.639	3.365	-4.225	2.839	1.319	2.837	1.288	-2.389	6.639	2.837	6.407	2.891	-5.100	3.365	1.288	2.891	1.848	-0.992	-4.225	-2.389	-5.100	-0.992	7.417

Third central moment: $m_3(Y)$/1E-06

	EH					RV					ED					GM					MF				
	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF	EH	RV	ED	GM	MF
EH	5.828	-2.514	-4.439	3.402	7.097	-2.514	-2.619	-4.017	-1.165	3.585	-4.439	-4.017	-7.349	-0.945	7.879	3.402	-1.165	-0.945	2.393	3.464	7.097	3.585	7.879	3.464	-3.395
RV	-2.514	-2.619	-4.017	-1.165	3.585	-2.619	-2.010	-3.111	-1.495	2.170	-4.017	-3.111	-4.713	-1.737	4.262	-1.165	-1.495	-1.737	-0.201	2.136	3.585	2.170	4.262	2.136	-2.765
ED	-4.439	-4.017	-7.349	-0.945	7.879	-4.017	-3.111	-4.713	-1.737	4.262	-7.349	-4.713	-8.332	-2.083	8.707	-0.945	-1.737	-2.083	0.378	3.767	7.879	4.262	8.707	3.767	-5.359
GM	3.402	-1.165	-0.945	2.393	3.464	-1.165	-1.495	-1.737	-0.201	2.136	-0.945	-1.737	-2.083	0.378	3.767	2.393	-0.201	0.378	0.566	1.488	3.464	2.136	3.767	1.488	-0.938
MF	7.097	3.585	7.879	3.464	-3.395	3.585	2.170	4.262	2.136	-2.765	7.879	4.262	8.707	3.767	-5.359	3.464	2.136	3.767	1.488	-0.938	-3.395	-2.765	-5.359	-0.938	6.676