Hedge Fund portfolio selection with higher moments

Diploma thesis submitted to the
ETH ZÜRICH and UNIVERSITY OF ZÜRICH

for the degree of
MASTER OF ADVANCED STUDIES IN FINANCE

Author: Saverio Massi Benedetti

Date:28/12/2004

Supervisors:

Prof. Dr. Alexander McNeil
(Department of Mathematics - ETH Zürich)

Dr. Jean Francois Bacmann
(RMF Investment Management - Quantitative Analysis Team)

This thesis was written while doing an internship at "RMF Investment Management" in Pfaeffikon, Switzerland. I am grateful to Dr. Jean Francois Bacmann for his guidance throughout the work and for the many insights into the subject. I would like to thank Prof. Alexander McNeil for his support and precious suggestions. Finally, I am grateful to my wife Paola for her emotional support and encouragement.

1 Introduction.

Financial assets returns are in general not normally distributed. Strong empirical evidence against the normality of the returns has been reported in the past years by several authors. This evidence suggests that the probability distribution followed by the returns is often characterized by skewness and leptokurtic behaviour. This departure from the normal distribution usually exhibited by the returns of many financial assets is even more accentuated in the hedge fund environment. In this context, the returns distribution is strictly linked to the return generating process adopted by the fund manager and the strategy employed to exploit market opportunities is strongly affecting the return profile. In an asset allocation context, the presence of asymmetry and fat tails violates the assumption of elliptically distributed asset returns that underlies the traditional mean-variance analysis (see Ingersoll 1987).

The asymmetry in the return distribution will have an impact on the portfolio selection task for the investors that have preference for skewness. A study by Arrow (1971) suggests that non increasing absolute risk aversion is a desirable property of a utility function. It is possible to show that under non increasing absolute risk aversion the investors will exhibit preference for positively skewed portfolios. Scott and Horvath (1980) point out that under the assumptions of positive marginal utility, decreasing absolute risk aversion and strict consistency for moment preference (i.e. all moments are always associated with the same preference direction for any wealth level), investors will exhibit preference for odd moments and aversion for even moments. Several researchers have proposed advances to the traditional mean variance theory in order to include higher moments in the portfolio optimisation task (see Athayde and Flôres(2001), Adcock (2002), Jondeau and Rockinger (2004) among others). In the hedge fund context, recent research has proposed new methods to include higher moments in the hedge fund portfolio selection. A work by Bacmann and Bosshard (2003) suggests using an asymmetric risk measure in order to penalise fat negative tailed investments and reward investments with fat positive tails.

Moreover, in order to select the optimal portfolio, a preliminary assessment of the relevant beliefs about future performances of the available assets is needed. The method proposed by Markovitz (1952) is addressing the selection of the optimal portfolio using as inputs the parameters of the return distributions of the available assets. Since in general the true values for the parameters of the future returns distribution are not known, an estimate of these values must be used. This generates an estimation risk that must be included in the portfolio selection task. Many studies are addressing the estimation risk issue but remain in a mean variance framework. In this context, a well-studied solution (see Bawa and Klein(1976), Jobson and Korkie (1980), Jorion(1986) and Polson and Tew (2000) among others) is given by the use of Bayesian estimation techniques where parameter uncertainty is explicitly considered and predictive parameters can be used to describe future return distributions.

A recent work by Harvey, Liechty et al.(2004) propose a method to address both estimation risk and the inclusion of higher moments in the portfolio selection. They suggest specifing a Bayesian probability model for the joint distribution of the asset returns when these returns are driven by a Skew Normal distribution. This allows us to capture the asymmetry of the returns and include it in the portfolio selection task. In such a Bayesian framework the expected utilities are then maximized using predictive returns. This thesis applies such techniques in a hedge fund portfolio selection context. To our knowledge this approach was never applied in such environment. The heterogeneity of different styles and hedge funds characteristics produces a great variety also in the behaviour of the higher moments of the return distributions and this makes this context particularly interesting to assess the impact of higher moments on the asset allocation process. In addition, we extend the method proposed by Harvey Liechty et al. using together with the Skew Normal model, also a model based on the Skew t distribution. The Skew t distribution was already used in other studies to model financial assets returns as in Patton (2004). However such a distribution was never applied in the Bayesian portfolio optimisation approach where also the estimation risk is taken into account. Given the characteristics of the hedge fund returns, we find opportune trying to model the returns not just with a Skew Normal distribution but also with a Skew t distribution. Comparing the two models we find that, in this environment, the Skew t distribution is providing strong improvement in fitting the observed data. In this context, in fact, the Skew Normal model is probably not flexible enough to fit the data observed and in particular to provide reliable estimates for all of the first three moments.

The rest of the thesis is organized as follows: Section 2 describes the hedge fund environment with emphasis on the analysis of the different managing styles and on how the distribution of the fund returns are affected by the strategy adopted. In Section 3 the model is described. The class of Skew elliptical distributions proposed by Sahu et al. (2003) is defined together with the Skew Normal and Skew t special cases. Then the Bayesian inference technique is explained and specification of the models for Bayesian estimation is provided. Section 4 addresses the estimation of the predictive distribution in order to take into account for parameter uncertainty and estimation risk. In Section 5 the empirical results obtained from the application of the proposed models to a dataset consisting of hedge fund strategies returns are presented. A preliminary assessment of the non-normality of hedge fund returns is also provided.

2 Hedge Fund environment.

Hedge Funds are private investment vehicles where the manager is free to operate in a variety of markets using investment strategies not restricted in short exposures or leverage. Typically the legal form adopted is the limited partnership, where the manager is usually the general partner and investors are limited partners. The main differences with traditional long-only mutual funds are given by the differences in trading strategies and regulation. The freedom in the managing process and the lack of regulation are probably the main reasons for the heterogeneity and variety of characteristics of the hedge fund industry.

Following Fung and Hsieh (1997), hedge funds can be classified on the basis of three key characteristics: location, strategy and leverage. The location identifies the markets on which the funds are specialised or the financial instruments they mainly trade as for example equity or bond markets. The strategy specifies the trading methods adopted by the managers and the kind of approach used to take advantage of market opportunities. Strategies can be based on exploiting arbitrage opportunities, assuming long and/or short positions in financial assets or adopting a trend following approach. Finally, hedge funds can be distinguished on the basis of the amount of leverage used.

In this classification, a traditional mutual fund can be characterised as operating in equity and/or bond markets, having a buy and hold strategy and no leverage. In the mutual fund universe, the location represents the only distinguishing element. Applying the same classification setup, any of the traditional hedge fund styles can easily be characterised. Hedge funds offer more variety and therefore the hedge fund universe is usually segmented in styles.

The principal hedge fund styles can be described as follows:

For the **Relative value** style the strategies rely on the identification of arbitrages opportunities on the price of the securities traded. Within this group, the *Convertible Arbitrage* strategy is located in convertible bonds markets and seeks to capture mispricings of the options embedded in these instruments. The risk-return characteristics of convertible arbitrage funds are analysed in detail in a recent work by Agarwal, Fung et al. (2004). The *Fixed Income arbitrage* strategy operates within global fixed income markets and is characterised by high levels of leverage and mainly adopting an approach based on convergence trading. Convergence trading consists in assuming offsetting positions in securities with similar characteristics but different prices in order to profit when the prices will narrow. The returns generated by these strategies are negatively skewed. In fact, the profits arising from single arbitrage opportunities are limited by the spread size but occasionally, under particular circumstances, the strategy may fail as the spread widens instead of disappearing. Since in general the strategy is long on the more illiquid securities, which can hardly be sold in bad circumstances, the occasional failure of the strategy can result in very high losses for the manager. An example of the extreme market conditions that can cause similar strategies to fail is the occurrence of a *flight to quality*. During this kind of market circumstances the prices of high rating securities will rise while the prices of credit sensitive securities will drop. A convergence trading strategy involving a long position in low rating securities and a short position in treasury bonds will suffer significant losses in such situation. The returns are thus expected to present a large number of positive and relatively small observations with few large negative events. A study by Fung and Hsieh (2002*)* focused on fixed income strategies points out the fact that a convergence trading strategy can be modelled as a short position in a lookback straddle.

Event Driven style relies on the identification and analysis of securities that can benefit from the occurrence of extraordinary transactions such as mergers, liquidations, bankruptcies or similar events. Event driven strategies are put in place in order to take advantage of valuation disparities produced by corporate events. Within this style the *Distressed Securities* strategies invest in undervalued securities of companies experiencing financial distress. An approach adopted by the funds operating in these strategies is based on investing in equity or bonds of selected distressed firms that are expected to recover. Another approach is addressing firms in much more advanced distressed situation by buying a consistent portion of the firm's debt and trying to get rid of the shareholders in order to gain the control of the reorganization process.

The *Merger Arbitrage* strategies rely on the identification and analysis of securities that can benefit from the occurrence of mergers and acquisitions. The funds operating within the Merger arbitrage style exploit the arbitrage opportunities created in these corporate events typically by buying stocks of the target company and shorting the stocks of the acquirer, trying to capture the merger spread. However, if the deal fails, the arbitrageur will face a loss that is usually much greater than the profits obtained in case of success of the strategy. Mitchell and Pulvino (2001) provide a detailed study of this kind of strategies and point out that the payoff generated from Merger arbitrage strategies is equivalent to the payoff generated by selling uncovered put options. Such payoff will produce a strong asymmetry in the distribution of the returns generated by these strategies.

Global Macro style relies on macroeconomic analysis to take bets on major risk factors such as currencies, interest rates, stock indices and commodities. *Macro strategies* exploit changes affecting global markets and economies. Hedge funds operating in this style usually pursue a base strategy like futures trend following and then add to this strategy some large scale high leveraged directional bets in other markets when the opportunities show up. For example, George Soros' Quantum fund made large gains by betting that the British pound would drop out of the European Rate Mechanism in 1992

Emerging markets strategies focus on equity and fixed income investments in emerging markets. This style is highly volatile since emerging markets are more volatile themselves and offer limited instruments for hedging risk. Due to short selling restrictions and the lack of sophisticated derivatives these funds are usually investing in the long only side. These funds are exposed to market risk and credit risk from the countries in which they are invested.

Equity hedged style seeks to anticipate movements on the equity markets from signals generated from statistical, fundamental and technical analysis.

These strategies are characterised by being located on the equity market and exploiting techniques such as pairs trading (see Gatev et al.,1999), contrarian and momentum trading[1]. *Equity Market-Neutral* strategies exploits inefficiencies in equity markets and is designed to have zero correlation with the market. Arbitrage opportunities are identified by statistical or fundamental analysis. Long and short positions are adjusted in order to eliminate all sources of risk except the ones on which the strategy is focused on.

Long/Short Equity funds combine long as well as short equity positions. They can add value in both directions by selection of undervalued stocks to buy or overvalued stocks to sell. Thus, short positions are not only taken to hedge systematic risk but also to benefit from opportunities. These funds can adopt consistent or variable net long or net short exposure. They tend to build portfolios that are much more concentrated than traditional mutual funds.

Managed Futures (Commodity Trading Advisors) adopt mainly trend following strategies investing in worldwide futures markets. Funds operating within this style are trading futures and derivatives on financial instruments and tangible commodities. Fung and Hsieh (2001) show that the payoff to a "trend following" hedge fund strategy is related to a payoff from an investment in a lookback straddle. The skewness of returns generated by funds in this category tend to be positive since this strategy, as a lookback straddle, is providing downside protection while allowing for large gains during extreme up and down moves of the markets.

In contrast with the mutual funds environment, in the hedge fund context the variety of different strategies and levels of leverage used, together with the location, will become very important in the analysis of the return generating process. In particular, the strategy adopted is here a very important determinant of return characteristics. A study conducted by Agarwal and Naik (2001) point out that a factor model consisting of a trading strategy factor and a location factor is able to explain a significant portion of the variation in Event Driven and Relative Value arbitrage returns over time. The hedge fund's return generating process is strictly linked to both the location and the style or strategy followed by the manager. Thus, also the distribution of the returns will depend on the location and strategy adopted. These two factors can combine in different ways to yield significant performance differences. This link between the returns profile and the hedge funds characteristics allows to retrieve significant information on the distribution of future returns not only from the past returns time series, but also from a careful analysis of the strategies.

Due to the strategies used, hedge funds returns exhibit typically strong deviations from the normal distribution. A more detailed analysis of the normality departures is presented in section 5. Here, the normality of returns is strongly rejected. Studies such as Brooks and Kat (2002) and Schmidhuber and Moix (2001) show that in general hedge fund returns exhibit skewness and leptokurtic behaviour.

[1] for a reference on these trading techniques see De Bondt and Thaler (1985), Conrad and Kaul (1998) and Jegadeesh and Titman (1993).

3 The Model

3.1 Skewed Distributions.

In order to model Hedge Fund returns the probability distribution needs to represent the existence of skewness. 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 good 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 the conditioning arguments to match the number of observed variables[2].

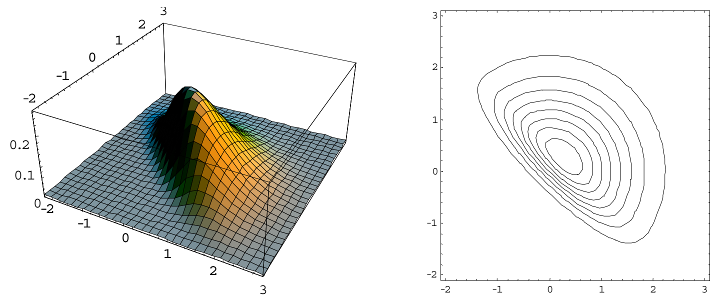


Figure 3.1 - Probability density function for the Skew Normal Distribution obtained by Azzalini and Della Valle (1996);

The analysis developed in this thesis will focus on the class of Skew Elliptical Distributions introduced by Sahu et al. Two distributions of this class will be fitted to the hedge fund returns data. 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.

3.1.1 General Model

Sahu et al. (2003) define a skew elliptical class of distributions by transforming elliptically symmetric distributed random variables and then conditioning

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

on some latent variables. The starting point is thus the family of elliptical distributions that can be defined as follows:

$\mathbf{X} \sim El_p(\boldsymbol{\theta}, \Sigma; g^{(p)})$ if the p-dimensional random vector $\mathbf{X}$ has multivariate density equal to

$$f(\mathbf{x}; \boldsymbol{\theta}, \Sigma, g^{(p)}) = |\Sigma|^{-1/2} g^{(p)} \left\{(\mathbf{x} - \boldsymbol{\theta})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\theta})\right\} \qquad \mathbf{x} \in \mathbb{R}^p$$

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

and where $g(u;p)$ is a function from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that the integral

$\int_0^\infty r^{p/2-1} g(r;p)dr$ 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 $g(u;p) = g_t := \left(1 + \frac{u}{\nu}\right)^{-(\nu+p)/2}$ with $\nu > 0$ the Multivariate t distribution is obtained.

Skew Elliptical distribution (Sahu et al.)

To obtain the skew elliptical distribution we first consider the transformation:

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

where
$\mathbf{Z}$ is a vector of unobserved (latent) random variables whose distribution is elliptical with zero mean and identity covariance matrix I_p.
$\boldsymbol{\mu} \in \mathbb{R}^p$,
D is a $p \times p$ diagonal matrix[3] with entries $\delta_1, ..., \delta_p$
and $\varepsilon \sim El_p(\mathbf{0}, \Sigma; g^{(p)})$.

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

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

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

$$\mathbf{V} \sim El_p(D(\Sigma + D^2)^{-1}(\mathbf{y} - \boldsymbol{\mu}), I - D(\Sigma + D^2)^{-1}D; g^{(p)}_{q(\mathbf{y}-\boldsymbol{\mu})}),$$

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

[3]A different version of this model can be obtained by taking D as a full matrix as in Harvey, Liechty et al.(2004). Though in our case this would result in a number of paramaters too high compared with the limited number of observations in the data.

and $q(\mathbf{y}-\boldsymbol{\mu}) = (\mathbf{y}-\boldsymbol{\mu})'(\Sigma + D^2)^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

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.

3.1.2 Special cases

Skew-Normal Distribution

The skew-Normal distribution is obtained by defining

$g(u;p) = e^{(-u/2)}$. Thus $g^{(p)}(u) = \frac{e^{-u/2}}{(2\pi)^{p/2}}$.

The density of this distribution will be[4]

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, D) = 2^p \mid \Sigma + D^2 \mid^{-1/2} \phi_p \left\{(\Sigma + D^2)^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right\} P(\mathbf{V} > \mathbf{0})$$

where ϕ_p is the density of the p-dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix identity

and $\mathbf{V} \sim N_p\left\{D(\Sigma + D^2)^{-1}(\mathbf{y}-\boldsymbol{\mu}), \mathbf{I} - D(\Sigma + D^2)^{-1}D\right\}$

The Moment Generating function is given by

$$M_{\mathbf{y}}(\mathbf{t}) = 2^p \Phi_p(D\mathbf{t})\; e^{\left\{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'(\Sigma + D^2)\mathbf{t}/2\right\}}$$

where Φ_p is the cumulative density function of the p-dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix identity.

The mean vector and the covariance matrix can be calculated as:

$$\mathbf{E}(\mathbf{Y}) = \boldsymbol{\mu} + \left(\tfrac{2}{\pi}\right)^{1/2} D\mathbf{1}$$

$$\text{cov}(\mathbf{Y}) = \Sigma + \left(\tfrac{2}{\pi}\right)^{1/2} D^2$$

For the third central moment tensor we obtain non zero values only for the *iii coordinates* entries. This values can be calculated as

$$\mu^{(3)}_{iii}(\mathbf{Y}) = -\frac{2^{1/2}\delta_i^3(\pi-4)}{\pi^{3/2}}.$$

In Figure 3.2 and 3.3 the densities of two bivariate Skew Normal Distributions (Sahu et al. version) are plotted. Note that conditioning on the two

[4]Note that $g^{(2p)}_{q(\mathbf{y}-\boldsymbol{\mu})}$ is not depending on $q(\mathbf{y}-\boldsymbol{\mu})$.

elements of the vector $\mathbf{Z}$ being positive allows for two lines to bound the tail of the distribution as opposed to the Skew Normal density obtained by Azzalini and Della Valle (see Figure 3.1) where, since the conditioning is on just a scalar random variable, only one line is limiting the tail of the distribution.

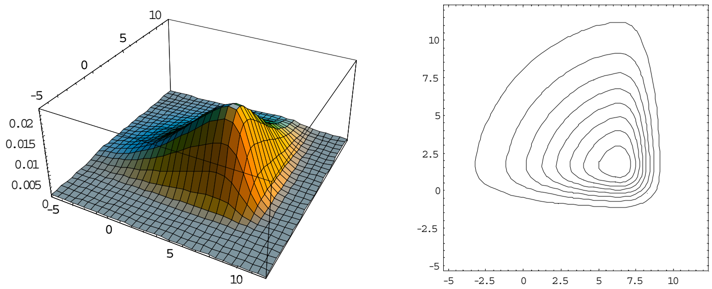


Figure 3.2 - Bivariate Skew Normal distribution with $\mu = (8, 0)$, $\Sigma = \{(0.5, 0.1)\,; (0.1, 0.5)\}\,, \delta_1 = -5$ and $\delta_2 = 5$;
(In this setting we obtain the two marginal distributions with same mean and variance.)

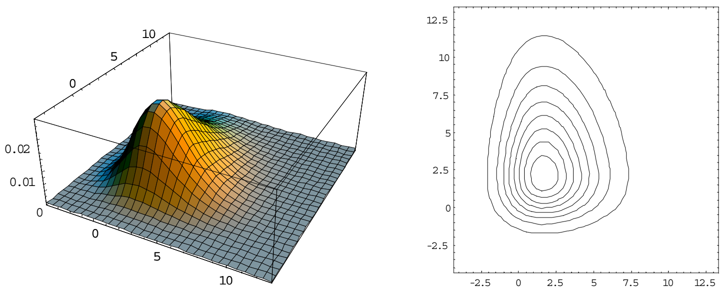


Figure 3.3 - Bivariate Skew Normal distribution with $\boldsymbol{\mu} = (0, 0)$, $\Sigma = \{(1, 0)\,; (0, 1)\}\,, \delta_1 = 3$ and $\delta_2 = 5$;

Skew-t Distribution

Let $g(u; 2p) = \left(1 + \frac{u}{\nu}\right)^{-(\nu+2p)/2}$ where ν is some integer representing the degrees of freedom.

As shown in Sahu et al.(2003), the following density for the skew-t distribution is obtained:

$$2^p t_{p,\nu}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma + D^2) P(\mathbf{V} > \mathbf{0})$$

where $\mathbf{V}$ follows the t-distribution $t_{p,\nu+p}$.

Using standard results on the representation of the student-t distribution as a scale mixture of normal distribution we can obtain the Moment generating function as

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^p \int_0^\infty e^{\mathbf{t}'\boldsymbol{\mu}+\mathbf{t}'(\Sigma+D)\mathbf{t}/(2w)} \Phi_p(Dw^{-1/2}\mathbf{t})dG(w)$$

where

$W \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ and $G(w)$ denotes the cumulative distribution function[5] of $\Gamma(\frac{\nu}{2}, \frac{\nu}{2})$.

Then

$$\mathbf{E}(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)} D\mathbf{1}$$

$$\text{cov}(\mathbf{Y}) = \left[\Sigma + (1-\tfrac{2}{\pi})D^2 + \tfrac{2}{\pi}D\mathbf{1}(D\mathbf{1})'\right] \frac{\nu}{(\nu-2)} - D\mathbf{1}(D\mathbf{1})'\frac{\nu}{\pi}\left[\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right]^2$$

Using the same notation as Athayde and Flôres(2002) to represent the third central moment tensor as a matrix of order $n \times n^2$, gives

$$\mu^{(3)}(\mathbf{Y}) =$$

$$= \frac{\nu^{3/2}\left\{4(\nu-2)\cdot\Gamma\left(\frac{\nu-1}{2}\right)^3\Delta+\Gamma\left(\frac{\nu}{2}\right)^2\left[2\Delta+[(\pi-2)H+(4-\pi)L]+\pi\cdot\Omega(\nu-2)\Gamma\left(\frac{\nu-3}{2}\right)-2\Gamma\left(\frac{\nu-3}{2}\right)(6\Delta+(\pi-2)H+\pi\cdot\Omega)\right]\right\}}{2\pi^{3/2}(\nu-2)\Gamma\left(\frac{\nu}{2}\right)^3}$$

where

$\delta = D\mathbf{1}$

$\Delta = \delta \otimes \delta\delta'$

$\Omega = \delta' \otimes \Sigma + \Sigma \otimes \delta' + \text{vec}(\Sigma)' \otimes \delta$

$H = \delta' \otimes D^2 + D^2 \otimes \delta' + \text{vec}(D^2)' \otimes \delta$

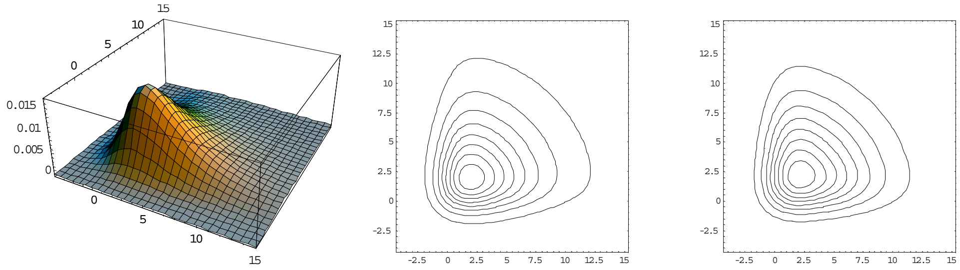
(vec($\cdot$) creates a vector by stacking the columns of a matrix)

and

$L = (D \otimes \delta'). * (\delta' \otimes I_{d\times d})$

Here the symbol $\otimes$ is used to indicate the Kronecker product, $I_{d\times d}$ is the identity matrix and $.*$ is the "element by element" matrix multiplication.

[5]The parametrization used is such that $\mathbf{E}(W) = 1$.



(a) (b)

Figure 3.4 - Bivariate Skew t distribution with $\nu = 4$, $\mu = (0, 0)$, $\Sigma = \{(1, 0)\,; (0, 1)\}$, $\delta_1 = 5$ and $\delta_2 = 5$;
The contour plot of the Skew t (a) is compared to the contour plot of a Skew Normal distribution (b) with same $\boldsymbol{\mu}, \Sigma$ and $\boldsymbol{\delta}$.

3.2 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 data $\mathbf{Y}$ conditional on $\mathbf{Z} = \mathbf{z}$ is given by

$$\mathbf{Y} \mid \mathbf{Z} = \mathbf{z} \sim El_p\left(\boldsymbol{\mu} + D\mathbf{z}, \Sigma, g^{(p)}_{q(\mathbf{z})}\right)$$

with $q(\mathbf{z}) = \mathbf{z}'\mathbf{z}$

and where the marginal specification for $\mathbf{Z}$ will be a truncated elliptical distribution[6] whose values are restricted to be positive, i.e.

$$\mathbf{Z} \sim El_p(\mathbf{0}, I, g^{(p)})\ \mathbb{I}(\mathbf{z} > \mathbf{0}).$$

Taking $g(u, 2p)$ as g_t, we obtain the Skew t model. Since we can represent the t distribution as a mixture of Normal distributions[7], the likelihood for each observation can be specified as

$$\mathbf{Y}_i \mid \mathbf{z}_i, \boldsymbol{\mu}, \Sigma, D, w_i \sim N_p(\boldsymbol{\mu} + D\mathbf{z}_i, \tfrac{\Sigma}{w_i})$$

where

[6] In particular we will need $\mathbf{Z} \sim N_p(\mathbf{0}, I)\ \mathbb{I}(\mathbf{z} > \mathbf{0})$ for the Skew Normal model and $\mathbf{Z} \sim t_{p,\nu+p}(\mathbf{0}, I)\ \mathbb{I}(\mathbf{z} > \mathbf{0})$ for the Skew t model. For a proof see Sahu et al. (2003)

[7] See for example, McNeil, Frey and Embrechts (2005).

$\mathbf{Z}_i \sim N_p(\mathbf{0}, I_p)\ \mathbb{I}(\mathbf{z} > \mathbf{0})$,

$W_i \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$

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

$\boldsymbol{\mu} \sim N_p(\mathbf{m}, \Sigma_\mu)$

$\Sigma \sim IW_p(c_\Sigma, \Omega_\Sigma)$

$D\mathbf{1} \equiv \boldsymbol{\delta} \sim N_p(\mathbf{d}, \Sigma_\delta)$

$\nu \sim \Gamma(\gamma, \Sigma_\nu)\ \mathbb{I}(\nu > 2)$

with IW_p denoting the Inverse-Wishart distribution.

Note that this specification for the Skew t model includes as a special case also the model for the Skew Normal. To obtain the setup for the latter we just need to set $W_i = 1,\ \forall\ i$ and remove the last conjugate prior density.

Uninformative priors on the parameters can be given in the model by setting
$\mathbf{m} = \mathbf{0}$, $\Sigma_\mu = 100 I_p$,
$\mathbf{d} = \mathbf{0}$, $\Sigma_\delta = 100 I_p$,
$c_\Sigma = p$, $\Omega_\Sigma = p I_p$,
$\gamma = 1$ and $\Sigma_\nu = 0.1$

Informative priors can be set on one or more parameters. For example, to set informative priors on $\boldsymbol{\delta}$, which is the parameter adjusting the amount of skewness of the distribution, we can specify in the prior mean vector $\mathbf{d}$ the values which reflect our prior information. It is also possible to reflect the degree of confidence in the prior information by choosing lower values for the diagonal elements of Σ_δ (prior variances). Note that setting $\Sigma_\delta = \frac{1}{n} I_p$, where n equals the number of observations in the data, the model will produce a posterior distribution which is obtained by equally weighting the information coming from the data and the information included in the priors.

See the Appendix A.1 for a detailed specification of the full conditionals.

3.3 Bayesian Estimation using MCMC and Gibbs sampling.

Bayesian estimation can be undertaken by using MCMC methods such as the algorithm introduced by Metropolis et al. (1953) and generalised by Hastings (1970) or the Gibbs sampler proposed by Geman and Geman (1984). These methods allow to produce a Markov chain whose output corresponds to a sample from the joint posterior distribution.

Metropolis-Hastings algorithm

Suppose that we want to generate a sample from a distribution $p(\boldsymbol{\theta}) = f(\boldsymbol{\theta})/K$, for which we know f but not the normalizing constant K. The Metropolis algorithm is based on generating a Markov chain by

1.) drawing samples v from a symmetric[8] proposal density $q(\cdot, u)$ such that $q(v, u)$ is the probability of having a proposed value v given the current state of the Markov chain $u = \boldsymbol{\theta}^{(t-1)}$.

2.) computing the ratio $\alpha = \frac{p(v)}{p(u)} = \frac{f(v)}{f(u)}$.

3.) accepting the new candidate value v for the Markov chain, i.e. setting $\boldsymbol{\theta}^{(t)} = v$, if $\alpha \geq 1$. Otherwise, if $\alpha < 1$ setting $\boldsymbol{\theta}^{(t)} = v$ with probability α and $\boldsymbol{\theta}^{(t)} = u$ with probability $1 - \alpha$.

Under mild conditions[9] the algorithm converges to a stationary distribution, i.e.

$$\boldsymbol{\theta}^{(t)} \xrightarrow{d} \boldsymbol{\theta} \sim p \quad as \quad t \longrightarrow \infty.$$

A generalization to this algorithm was introduced by Hastings(1970) who proposed to modify the acceptance ratio as

$$\alpha = \frac{p(v)q(u,v)}{p(u)q(v,u)} = \frac{f(v)q(u,v)}{f(u)q(v,u)}$$

allowing to drop the requirement of q being symmetric.

Gibbs Sampling

This method is a special case of Metropolis Hastings sampling where the random value generated by the proposal is always accepted. To implement the Gibbs sampler we need to be able to sample from the full conditional distributions $\{p_i(\boldsymbol{\theta}_i | \ \boldsymbol{\theta}_{j \neq i}), \ i = 1, .., K\}$ for all the parameters $\boldsymbol{\theta}_i$, $i = 1, ..., k$. The algorithm proceeds by drawing iteratively from these distributions starting from an arbitrary set of values $\left(\boldsymbol{\theta}_1^{(0)}, ..., \boldsymbol{\theta}_k^{(0)}\right)$, i.e. draw

$$\boldsymbol{\theta}_1^{(1)} \sim p_1(\boldsymbol{\theta}_1 | \ \boldsymbol{\theta}_2^{(0)}, ..., \boldsymbol{\theta}_k^{(0)})$$

$$\boldsymbol{\theta}_2^{(1)} \sim p_2(\boldsymbol{\theta}_2 | \ \boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_3^{(0)}, ..., \boldsymbol{\theta}_k^{(0)})$$

...

The k draws above will form one iteration of the algorithm.

As showed by Geman and Geman (1984), for the $\left(\boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_k^{(t)}\right)$ sample obtained after t iterations we have:

[8] i.e. such that $q(v, u) = q(u, v)$ for every v, u.

[9] See Tierney(1994).

(a) $\left(\boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_k^{(t)}\right) \xrightarrow{d} (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k) \sim p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k) \qquad as\ t \longrightarrow \infty.$

(b) The convergence in (a) is exponential in t using the L_1 norm.

In the Bayesian analysis context the distribution of interest is the joint posterior distribution, $p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k | \ \mathbf{y})$. Then the Gibbs sampler will require draws from the full conditional posterior distributions $p_i(\boldsymbol{\theta}_i | \ \boldsymbol{\theta}_{j \neq i}, \mathbf{y})$.

3.4 Model selection

In the Bayesian context the most used method to choose between two competing models is the Bayes Factor[10]. This factor is defined as the ratio of the observed marginal densities for a pair of models. Thus, given the observed data $\mathbf{y}$ and two models M_1 and M_2, the Bayes Factor is:

$$BF = \frac{p(\mathbf{y}| \ M_1)}{p(\mathbf{y}| \ M_2)}$$

Bayes Factor represents the evidence provided by the data in favour of a certain model and naturally takes into account for both the explanatory power and the complexity of the models.

In order to estimate Bayes Factors, the approach proposed by Newton and Raftery (1994) will be used.

[10]See Jeffreys (1935,1961), Kass and Raftery (1995) and Lavine and Schervish(1999).

4 Estimation of predictive distribution.

4.1 Addressing estimation risk.

A limitation of the traditional 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 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[11].

Many studies addressing this issue have tried to incorporate estimation risk into the portfolio choice problem. 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[12]. 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.

Black and Litterman (1990,1992) and Pástor(2000) improve this approach by using the equilibrium implications from an asset pricing model to define the priors for the model. 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[13]. Xia(2001) studies the effect of parameter uncertainty in a dynamic continuous time context[14]. Kan and Zhou(2003) are providing an analytical comparison of alternative decision rules under estimation risk.

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

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

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

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

[14] His findings suggest that the uncertainty about the predictive parameters introduces dynamic learning into the model. In addition, his results indicate that the optimal stock allocation may increase or decrease with the level of parameter uncertainty, depending on the importance of the hedge for parameter uncertainty relative to the importance of the hedge for stochastic predictive variable.

given the observed returns $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$. Integrating out the parameters $\boldsymbol{\theta}$ 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 $f(\mathbf{y}_{n+1}|\,\boldsymbol{\theta})$, the conditional expected utility of portfolio ω is given by

$$E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\left[U[W(\omega, \mathbf{y}_{n+1})]|\,\boldsymbol{\theta}\right] \equiv \int_{\mathbf{y}_{n+1}} U[W(\omega, \mathbf{y}_{n+1})]\ f(\mathbf{y}_{n+1}|\,\boldsymbol{\theta})\ d\mathbf{y}_{n+1}$$

Assuming that the parameter equals the estimate from the data means conditioning on $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, the optimal portfolio weights $\hat{\omega}$ will be obtained as

$$\hat{\omega} = \arg\max_{\omega} E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\left[U[W(\omega, \mathbf{y}_{n+1})]|\,\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}\right]$$

$$\equiv \arg\max_{\omega} \int_{\mathbf{y}_{n+1}} U[W(\omega, \mathbf{y}_{n+1})]\ f(\mathbf{y}_{n+1}|\,\hat{\boldsymbol{\theta}})\ d\mathbf{y}_{n+1}$$

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

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

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

The investor's optimal portfolio solution will be

$$\hat{\omega}^{Bayes} = \arg\max_{\omega} E_{\boldsymbol{\theta}}\left\{E_{\mathbf{y}_{n+1}|\boldsymbol{\theta}}\left[U[W(\omega, \mathbf{y}_{n+1})]|\,\boldsymbol{\theta}\right]\right\} =$$

$$= \arg\max_{\omega} \int_{\mathbf{y}_{n+1}} \int_{\boldsymbol{\theta}} U[W(\omega, \mathbf{y}_{n+1})]\ f(\mathbf{y}_{n+1}|\,\boldsymbol{\theta})\ p(\boldsymbol{\theta}|\,\mathbf{y})\ d\boldsymbol{\theta}\ d\mathbf{y}_{n+1}$$

$$= \arg\max_{\omega} \int_{\mathbf{y}_{n+1}} \int_{\boldsymbol{\theta}} U[W(\omega, \mathbf{y}_{n+1})]\ f(\mathbf{y}_{n+1}, \boldsymbol{\theta}|\,\mathbf{y})\ d\boldsymbol{\theta}\ d\mathbf{y}_{n+1}$$

$$= \arg\max_{\omega} \int_{\mathbf{y}_{n+1}} U[W(\omega, \mathbf{y}_{n+1})]\ p(\mathbf{y}_{n+1}|\,\mathbf{y})\ d\mathbf{y}_{n+1}$$

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.

To obtain the predictive moments we follow the setting proposed by Polson and Tew (2000) in the mean-variance context and extended to skewness in

Harvey, Liechty et al.(2004). Thus we can calculate the predictive moments from the means of the posterior moments as follows:

$$m_p = \bar{m}$$

$$V_p = \bar{V} + \text{Var}(m|\,\mathbf{y})$$

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

where m_p, V_p, S_p are the predictive (central) moments and $\bar{m}, \bar{V}, \bar{S}$ are the posterior means for the moments.

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 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

$$\max_{\omega} E(U(W))$$

under the constraint $\omega' \mathbf{1} = 1$ and $\omega \geq \mathbf{0}$.

Following Harvey et al.(2004) we consider a family of linear utility functions to describe the agents' preferences, i.e.

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

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

So that the expected utility (with respect to the predictive density) will be

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

This is particularly useful for the analysis of the optimal portfolio changes due to different levels of the second or third moment sensitivities.

As suggested by Harvey, Liechty et al. (2004), for a general utility function, the Bayesian framework can be used to evaluate and optimize the expected utilities. The expected utility can be estimated via Monte Carlo simulation by generating a set of G draws $\mathbf{y}_{n+1}^{(g)} \sim p\left(\mathbf{y}_{n+1}|\ \mathbf{y}\right)$ from the posterior predictive distribution and then averaging the utility over the values obtained, i.e.

$$E[U(W)] = E[U(1 + \omega' \mathbf{y}_{n+1})] =$$

$$= \int_{-\infty}^{\infty} U(1 + \omega' \mathbf{y}_{n+1}) p\left(\mathbf{y}_{n+1}|\ \mathbf{y}\right)\, d(\mathbf{y}_{n+1}) \approx \frac{1}{n} \sum_{g=1}^{G} U(1 + \omega' \mathbf{y}_{n+1}^{(g)})$$

The expected utility can then be optimised with numerical methods. To perform the optimisation, Harvey, Liechty et al. propose a Metropolis-Hastings algorithm to explore the expected utility as a function of the weights ω.

4.2 Specifying additional information in the model.

One of the advantages of the Bayesian approach is that it allows to reflect 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. This gives a framework where investors combine individual views with market equilibrium to select their portfolios. Then these portfolios, based on their composition or performance, lead the investor to update their initial opinions. An example in which the market views of the investor are used to affect the asset-allocation process is provided by Black and Litterman (1993). In their model the investor's views are combined with CAPM-implied expected returns to select optimal 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. As shown in Section 2, a consistent amount of information can be retrieved from the analysis of the different strategies. In particular, information concerning the direction of the skewness of the returns generated by the different hedge funds styles is available. By appropriately adjusting the prior on the parameter that regulates the skewness is then possible to include this information in the model.

Moreover 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.

5 Application to Hedge Fund returns.

5.1 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 (EQHg);
HFRI Relative Value Arbitrage Index (RVAr);
HFRI Event-Driven Index (EvDr);
HFRI Macro Index (Mac.);

In addition, to represent an investment opportunity in the managed futures strategy, the Stark 300 Trader Index (St300) is included in the dataset.

The data series consists of the monthly observations for the above indices between January 1990 and September 2004. As a preliminary investigation of the data some summary statistics of the returns are provided in Table 1.

Table 1 - Sample moments

	EQHg	RVAr	EvDr	Mac.	St300
mean	1.39	0.98	1.16	1.29	0.77
var.	6.73	1.11	3.66	6.03	11.24
skew.	0.18	-0.90	-1.34	0.31	0.82
kurt.	1.34	10.75	4.85	0.44	2.97

5.2 Presence of skewness. Non-normality assessment.

Test for Multivariate Normality.

In order to assess the joint non normality of the returns Mardia's test for multivariate normality is performed on the data.

This test involves comparing Mardia's measures of skewness (b_d) and kurtosis (k_d) with two reference asymptotic distributions. The measures for skewness and kurtosis for a sample $\{\mathbf{Y}_1, ..., \mathbf{Y}_n\}$ are defined respectively as

$$b_d = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij}^3$$

$$k_d = \frac{1}{n} \sum_{i=1}^{n} M_i^4$$

where M_i is the mahalanobis distance, thus $\left\{M_i = (\mathbf{Y}_i - \bar{\mathbf{Y}})' S^{-1} (\mathbf{Y}_i - \bar{\mathbf{Y}}), i = 1, ..., n\right\}$

and M_{ij} is the mahalanobis angle, thus $\{M_{ij} = (\mathbf{Y}_i - \bar{\mathbf{Y}})' S^{-1} (\mathbf{Y}_j - \bar{\mathbf{Y}}), \quad i, j = 1, ..., n\}$

($\bar{\mathbf{Y}}$ and S are respectively the mean vector and covariance matrix estimators).

Under the null hypothesis of multivariate normality the asymptotic distributions of these statistics as $n \longrightarrow \infty$ are

$\frac{n}{6} b_d \sim \chi^2_{d(d+1)(d+2)/6}$

$\frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim N(0, 1)$.

The test rejects the hypothesis of joint normality with a p value very close to zero for both the multivariate skewness and kurtosis measures (see Table 2).

Table 2 - Mardia's test

	b_d	*test stat*	*p value*
Skewness	6.16	181.92	0

	k_d	*test stat*	*p value*
Kurtosis	56.61	17.18	0

An additional confirmation that the joint distribution is not normal is coming from the QQ plot of the mahalanobis distances against a χ^2_5 distribution (Figure 5.1). For a p-dimensional Multivariate normal distribution the mahalanobis distance points given by $\{M_i;\ i = 1, ..., n\}$ should have approximately a χ^2_p distribution. As the figure shows, the distribution of the M_i's is significantly different from a χ^2_p . This can be assessed also by the Kolmogorov-Smirnov test for the M_i being distributed as a χ^2_p. Here the null hypothesis is again rejected with a p value of 0. Figure 5.2 displays a scatterplot of the data.

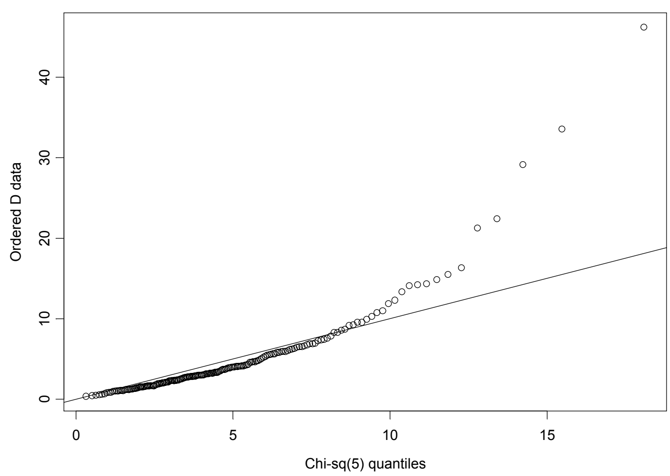


Figure 5.1 - Multivariate Normality plot. QQ plot of the mahalanobis distance points against a χ^2_5 distribution.

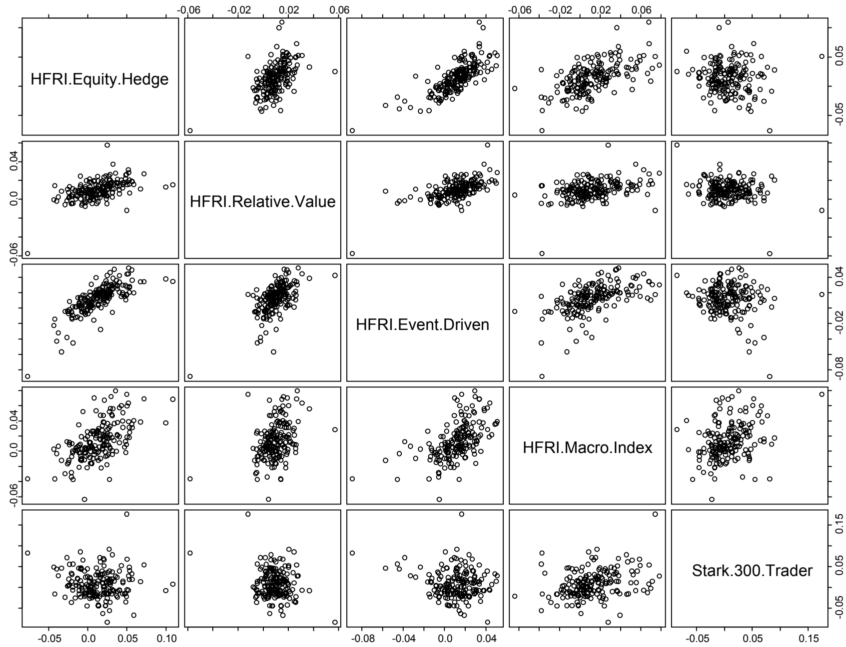
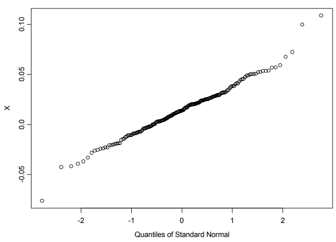


Figure 5.2 - Scatterplot of the five principal hedge fund styles' returns

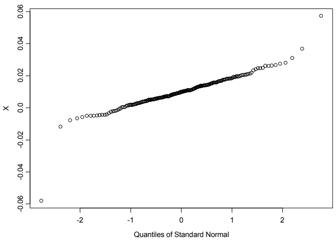
Univariate tests for Normality.

In addition to the joint normality tests we also perform the tests on each return series in order to check for univariate normality. A QQ Plot for each of the HFRI Indices series against the normal distribution quantiles is shown in Table 3. Most return series exhibit significant departures from normality. Other evidence against the normality of the data is obtained by performing a Jarque-Bera test for univariate skewness and kurtosis on the HFRI and Stark 300 return series. Table 4 shows the results for the test statistic under the null hypothesis that the data is normally distributed and the corresponding p values. All the strategies returns except Macro Index are failing the test at 99% significance level.

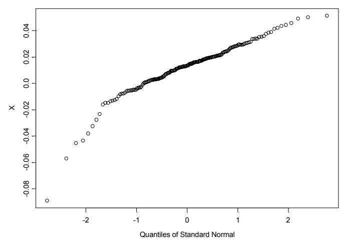
Table 3.- QQ Plots for the 4 HFRI Hedge Fund Indices return series and for the Stark300 Index.



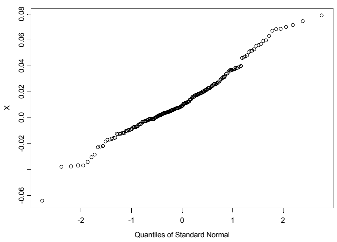
HFRI Equity Hedge



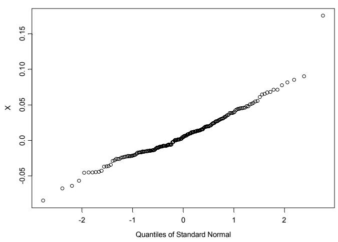
HFRI Relative Value Arbitrage



HFRI Event-Driven



HFRI Macro



Stark 300 Trader Index

The QQ Plots for these return series exhibit in most cases significative departures from normality. In particular, the arbitrage strategies returns are presenting a similar non-normal behaviour which corresponds to the presence of negative skewness. In contrast, the Stark300 Trader Index return series present an opposite behaviour which corresponds to positive skewness.

Table 4 - Jarque-Bera Test results.

Index	*Test Stat*	*p values*
HFRI Equity Hedge	12.74	0.002
HFRI Relative Value Arbitrage	823.09	0.000
HFRI Event Driven	214.18	0.000
HFRI Macro Index	3.95	0.138
Stark 300 Trader	79.94	0.000

5.3 Definition of priors

We want to assess the impact of the introduction of the third moment on the optimal allocation in hedge fund strategies. Parameter estimation and portfolio optimisation is performed for both the Multivariate Skew Normal model and the Multivariate Skew t model in the family of skew elliptical distributions proposed by Sahu et al. (2003).

For each model three different sets of priors are used to include additional degrees of information coming from the analysis of the strategies:

- Uninformative priors;
- Informative priors;
- Highly informative priors.

The first set considers uninformative priors. Thus the parameters of the prior distribution on the asymmetry parameter D are set to $\mathbf{d} = \mathbf{0}$ and $\Sigma_\delta = 100 I_p$. In the second set some information on the asymmetry parameter is specified taking $\mathbf{d} = (0.5, -1, -1, 1, 2.5)'$ and $\Sigma_\delta = I_p$. This corresponds to our prior beliefs on the distribution of the returns from the strategies. Based on the analysis of the return generating process adopted by the managers in these styles our priors assume slightly positively skewed returns for the Equity hedge style, negatively skewed returns for the arbitrage based Relative value and Event driven styles, positively skewed returns for the Macro style and strongly positively skewed returns for the trend following Managed futures. We will refer to the model defined by this set as the "Informative priors" model. Finally, a "Highly Informative priors" model is defined by using $\mathbf{d}$ as in the Informative priors set and then increasing the degree of confidence in the prior information by setting $\Sigma_\delta = \frac{1}{n} I_p$, where n is the number of observations in the data.

5.4 Estimation Results

In the Bayesian framework, the predictive moments have been calculated for both the skewed distributions with the different sets of priors. The estimation is done using a Gibbs sampler implemented in WinBUGS[15]. A total of 20'000 iterations have been undertaken for each model with a burn-in of 10'000 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. We also considered estimates coming from drawing a shorter chain[16] from the sampler in order to assess the effect of the convergence on the allocation.

[15] Spiegelhalter et al. (2003)

[16] The shorter chain is obtained with 10'000 iterations and a burn-in of 1'000 iterations.

The inclusion of prior information is improving the stability of the estimates and convergence of the sampler. The posterior densities produced are smoothed and MC errors reduced. Using a model with informative priors is improving the quality of the estimates and thus reducing the estimation risk. Table 5 illustrates the parameters' posterior statistics and densities produced by WinBUGS for the Skew t models with uninformative priors and highly informative priors.

Table 5 - Parameter statistics and densities.

Parameter statistics - Skew t model with uninformative priors.

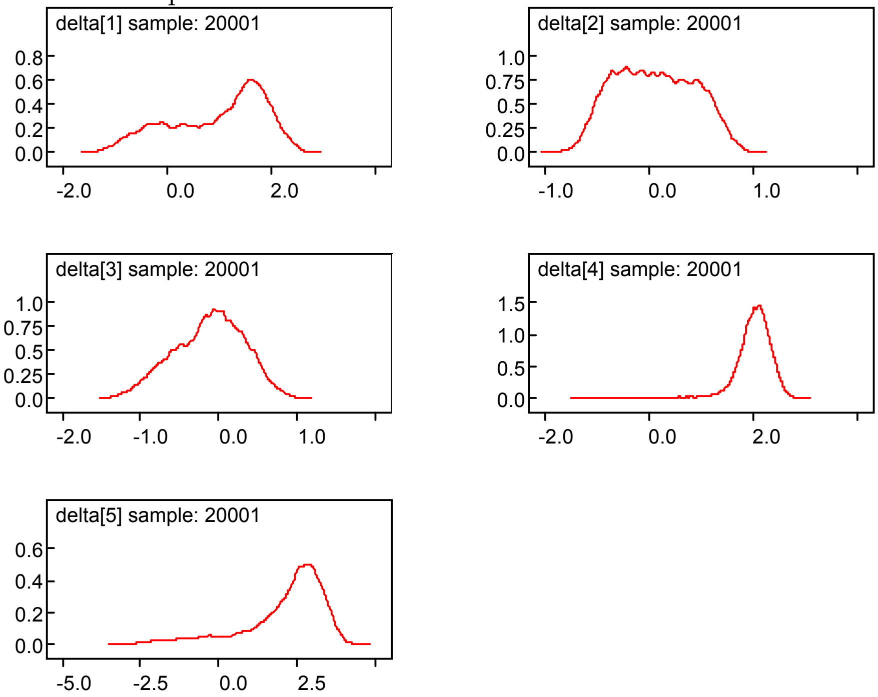
param.	mean	st.dev.	MC error	2.5%	median	97.5%
ν	5.74	1.28	0.04	3.77	5.57	8.64
$\mu[1]$	0.60	0.71	0.05	-0.44	0.41	2.07
$\mu[2]$	0.92	0.31	0.02	0.38	0.93	1.46
$\mu[3]$	1.36	0.37	0.02	0.70	1.34	2.07
$\mu[4]$	-0.31	0.33	0.02	-0.80	-0.35	0.49
$\mu[5]$	-1.08	1.04	0.08	-2.35	-1.37	1.82
$\delta[1]$	0.98	0.91	0.07	-0.87	1.23	2.30
$\delta[2]$	0.06	0.38	0.02	-0.59	0.05	0.74
$\delta[3]$	-0.12	0.45	0.03	-1.03	-0.09	0.67
$\delta[4]$	2.00	0.41	0.03	0.92	2.05	2.55
$\delta[5]$	2.14	1.33	0.10	-1.57	2.52	3.76
$\Sigma[1,1]$	4.20	0.73	0.03	2.90	4.17	5.72
$\Sigma[1,2]$	0.88	0.17	0.00	0.58	0.87	1.25
$\Sigma[1,3]$	2.49	0.37	0.01	1.84	2.47	3.28
$\Sigma[1,4]$	2.46	0.44	0.01	1.69	2.43	3.39
$\Sigma[1,5]$	-0.15	0.48	0.01	-1.11	-0.15	0.80
$\Sigma[2,1]$	0.88	0.17	0.00	0.58	0.87	1.25
$\Sigma[2,2]$	0.63	0.10	0.00	0.45	0.62	0.83
$\Sigma[2,3]$	0.70	0.13	0.00	0.48	0.69	0.97
$\Sigma[2,4]$	0.62	0.15	0.00	0.36	0.61	0.93
$\Sigma[2,5]$	-0.12	0.18	0.00	-0.48	-0.12	0.24
$\Sigma[3,1]$	2.49	0.37	0.01	1.84	2.47	3.28
$\Sigma[3,2]$	0.70	0.13	0.00	0.48	0.69	0.97
$\Sigma[3,3]$	2.23	0.31	0.01	1.68	2.21	2.92
$\Sigma[3,4]$	1.66	0.30	0.01	1.14	1.64	2.30
$\Sigma[3,5]$	-0.25	0.34	0.01	-0.92	-0.24	0.41
$\Sigma[4,1]$	2.46	0.44	0.01	1.69	2.43	3.39
$\Sigma[4,2]$	0.62	0.15	0.00	0.36	0.61	0.93
$\Sigma[4,3]$	1.66	0.30	0.01	1.14	1.64	2.30
$\Sigma[4,4]$	3.13	0.58	0.02	2.17	3.07	4.42
$\Sigma[4,5]$	1.89	0.53	0.02	0.94	1.86	3.01
$\Sigma[5,1]$	-0.15	0.48	0.01	-1.11	-0.15	0.80
$\Sigma[5,2]$	-0.12	0.18	0.00	-0.48	-0.12	0.24
$\Sigma[5,3]$	-0.25	0.34	0.01	-0.92	-0.24	0.41
$\Sigma[5,4]$	1.89	0.53	0.02	0.94	1.86	3.01
$\Sigma[5,5]$	5.57	1.46	0.08	3.05	5.46	8.58

Parameter statistics - Skew t model with highly informative priors.

param.	mean	st.dev.	MC error	2.5%	median	75.0%	97.5%
ν	6.03	1.44	0.04	3.86	5.80	6.78	9.49
$\mu[1]$	0.87	0.19	0.00	0.50	0.87	1.00	1.25
$\mu[2]$	1.89	0.09	0.00	1.72	1.89	1.95	2.06
$\mu[3]$	2.25	0.14	0.00	1.97	2.25	2.34	2.52
$\mu[4]$	0.33	0.18	0.00	-0.03	0.33	0.45	0.68
$\mu[5]$	-1.40	0.23	0.00	-1.86	-1.40	-1.25	-0.93
$\delta[1]$	0.50	0.08	0.00	0.35	0.50	0.55	0.65
$\delta[2]$	-1.30	0.07	0.00	-1.44	-1.30	-1.25	-1.16
$\delta[3]$	-1.44	0.07	0.00	-1.58	-1.44	-1.39	-1.30
$\delta[4]$	1.02	0.08	0.00	0.87	1.02	1.07	1.17
$\delta[5]$	2.51	0.07	0.00	2.36	2.51	2.55	2.65
$\Sigma[1,1]$	4.80	0.64	0.01	3.67	4.76	5.21	6.17
$\Sigma[1,2]$	0.93	0.19	0.00	0.59	0.92	1.05	1.34
$\Sigma[1,3]$	2.64	0.40	0.01	1.94	2.61	2.89	3.49
$\Sigma[1,4]$	2.57	0.46	0.01	1.76	2.54	2.87	3.55
$\Sigma[1,5]$	-0.31	0.50	0.01	-1.32	-0.30	0.03	0.68
$\Sigma[2,1]$	0.93	0.19	0.00	0.59	0.92	1.05	1.34
$\Sigma[2,2]$	0.51	0.09	0.00	0.35	0.50	0.56	0.71
$\Sigma[2,3]$	0.69	0.14	0.00	0.45	0.68	0.77	0.98
$\Sigma[2,4]$	0.63	0.17	0.00	0.32	0.62	0.73	0.98
$\Sigma[2,5]$	-0.28	0.20	0.00	-0.70	-0.27	-0.14	0.10
$\Sigma[3,1]$	2.64	0.40	0.01	1.94	2.61	2.89	3.49
$\Sigma[3,2]$	0.69	0.14	0.00	0.45	0.68	0.77	0.98
$\Sigma[3,3]$	2.02	0.31	0.01	1.47	1.99	2.22	2.68
$\Sigma[3,4]$	1.71	0.32	0.00	1.14	1.69	1.92	2.39
$\Sigma[3,5]$	-0.38	0.36	0.00	-1.11	-0.37	-0.14	0.31
$\Sigma[4,1]$	2.57	0.46	0.01	1.76	2.54	2.87	3.55
$\Sigma[4,2]$	0.63	0.17	0.00	0.32	0.62	0.73	0.98
$\Sigma[4,3]$	1.71	0.32	0.00	1.14	1.69	1.92	2.39
$\Sigma[4,4]$	3.98	0.57	0.01	2.99	3.94	4.33	5.19
$\Sigma[4,5]$	1.92	0.50	0.01	1.01	1.89	2.23	2.98
$\Sigma[5,1]$	-0.31	0.50	0.01	-1.32	-0.30	0.03	0.68
$\Sigma[5,2]$	-0.28	0.20	0.00	-0.70	-0.27	-0.14	0.10
$\Sigma[5,3]$	-0.38	0.36	0.00	-1.11	-0.37	-0.14	0.31
$\Sigma[5,4]$	1.92	0.50	0.01	1.01	1.89	2.23	2.98
$\Sigma[5,5]$	5.58	0.96	0.02	3.94	5.51	6.17	7.69

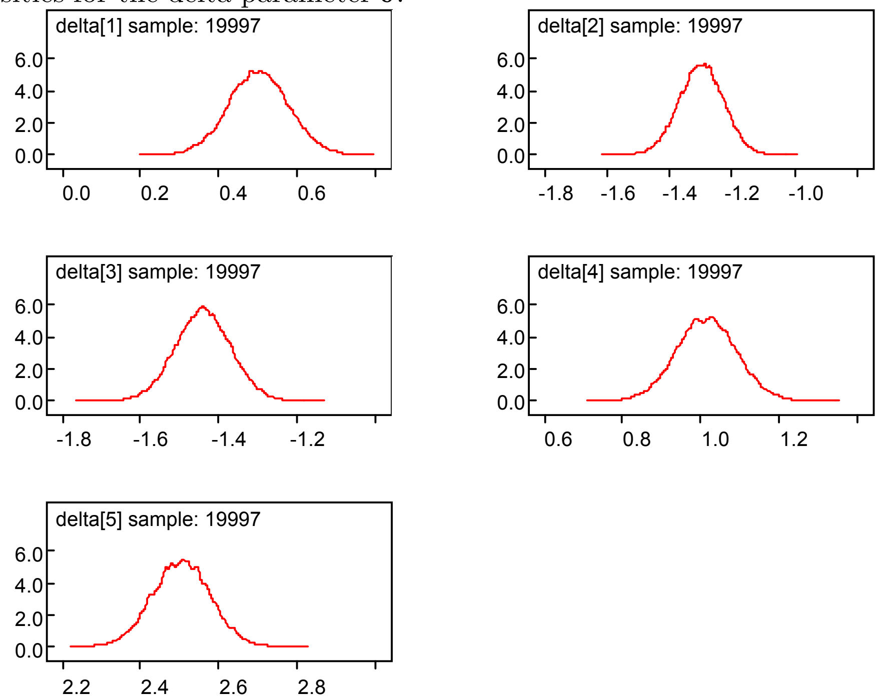
Skew t model with uninformative priors

Densities for the parameter $\boldsymbol{\delta}$:



Skew t model with highly informative priors

Densities for the delta parameter $\boldsymbol{\delta}$:



We obtain the estimates of the predictive moments using the expressions shown in section 4.1. Table 6 contains the estimates for the first three predictive moments for both the skewed model adopted. The estimates obtained with the shorter chain are displayed in Table 7. For the covariance matrix only the diagonal elements are displayed and for the skewness tensor only the *iii coordinates* elements are reported. The main differences between the estimates provided from the two models are concerning the skewness and thus the third central moment tensor. This suggests that the model choice will be particularly important for the asset allocation problem when the third moment is taken into account. Indeed the investors with a higher preference for skewness will be the most affected by the model choice. The predictive values are more conservative than the posterior means leading to smaller values for the skewness estimates and higher values for the variances. The difference in those estimates is varying according to the uncertainty exhibited by the parameters. The higher the estimation error on the parameter, the higher the difference between predictive moment and posterior mean.

The Bayes Factor analysis indicates that the Skew t model fits the observed data much better than the Skew Normal model. The probability of the data coming from a Skew t model is far higher than the probability that the observed data are coming from a Skew Normal model. The values obtained in the Skew Normal model for the return skewnesses are quite small. By comparing these values with those calculated from the sample we get the impression that the Skew Normal model is unable to completely fit the data even when a stronger informative prior is applied.

Table 6 - Estimates for the hedge fund dataset for all the models.

Skew Normal Models (5 Styles) MCMC with 20'000 iterations

Uninformative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.41	0.98	1.15	1.31	0.80
$\bar{V}$	6.55	1.16	3.64	6.03	11.00
Vp	6.58	1.16	3.66	6.07	11.07
$\bar{S}$	0.11	0.01	-0.03	0.12	0.19
Sp	0.10	-0.01	-0.04	0.11	0.19

Informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.38	0.97	1.13	1.28	0.77
$\bar{V}$	6.70	1.17	3.63	5.95	10.66
V_p	6.73	1.18	3.65	5.98	10.72
$\bar{S}$	0.01	-0.09	-0.08	0.04	0.13
S_p	0.00	-0.10	-0.09	0.03	0.12

Highly informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.38	0.88	1.13	1.29	0.77
$\bar{V}$	6.77	1.48	3.65	5.99	10.59
V_p	6.80	1.49	3.67	6.03	10.59
$\bar{S}$	0.00	-0.30	-0.10	0.02	0.10
S_p	-0.01	-0.31	-0.11	0.01	0.10

Skew *t* Models (5 Styles) MCMC with 20'000 iterations

Uninformative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.51	0.98	1.25	1.55	0.93
$\bar{V}$	7.98	1.11	3.71	7.99	13.53
V_p	8.05	1.12	3.73	8.04	13.65
$\bar{S}$	0.56	0.14	-0.10	1.31	1.19
S_p	0.56	0.12	-0.12	1.30	1.19

Informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.38	0.87	1.03	1.37	0.97
V	7.77	1.21	4.17	7.26	13.48
V_p	7.81	1.21	4.19	7.29	13.54
$\bar{S}$	0.34	-1.05	-0.90	0.92	1.35
S_p	0.33	-1.07	-0.91	0.91	1.34

Highly informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.34	0.69	0.92	1.27	0.92
$\bar{V}$	7.63	1.97	4.56	6.88	13.02
V_p	7.66	1.98	4.58	6.91	13.02
$\bar{S}$	0.21	-1.65	-0.96	0.50	1.02
S_p	0.20	-1.66	-0.97	0.50	1.02

Table 7 - Estimates for the hedge fund dataset for all the models obtained with less simulations.

Skew Normal Models (5 Styles) MCMC with 10'000 iterations

Uninformative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.42	0.98	1.16	1.31	0.80
$\bar{V}$	6.56	1.16	3.66	6.04	11.12
V_p	6.60	1.17	3.68	6.07	11.18
$\bar{S}$	0.11	0.02	-0.02	0.11	0.15
S_p	0.10	0.01	-0.03	0.10	0.15

Informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.37	0.96	1.13	1.28	0.77
$\bar{V}$	6.70	1.17	3.63	5.95	10.67
V_p	6.74	1.18	3.65	5.99	10.73
$\bar{S}$	0.01	-0.09	-0.08	0.04	0.12
S_p	0.00	-0.10	-0.09	0.03	0.12

Highly informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.38	0.88	1.13	1.29	0.76
$\bar{V}$	6.78	1.49	3.65	6.01	10.61
V_p	6.82	1.49	3.67	6.04	10.67
$\bar{S}$	0.00	-0.30	-0.10	0.02	0.10
S_p	-0.01	-0.31	-0.11	0.01	0.10

Skew t Models (5 Styles) MCMC with 10'000 iterations

Uninformative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.44	0.98	1.26	1.59	0.91
$\bar{V}$	7.95	1.13	3.74	8.19	13.95
V_p	8.02	1.14	3.76	8.24	14.11
$\bar{S}$	0.29	0.20	-0.07	1.45	1.14
S_p	0.28	0.19	-0.08	1.45	1.15

Informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.37	0.87	1.04	1.38	0.97
$\bar{V}$	7.73	1.20	4.15	7.24	13.48
V_p	7.77	1.21	4.17	7.28	13.54
$\bar{S}$	0.28	-0.94	-0.81	0.87	1.26
S_p	0.27	-0.96	-0.82	0.86	1.26

Highly informative priors

	EQHg	RVAr	EvDr	Mac.	St300
$\bar{m}$	1.34	0.68	0.91	1.27	0.93
$\bar{V}$	7.68	1.98	4.59	6.92	13.04
V_p	7.71	1.99	4.61	6.95	13.09
$\bar{S}$	0.22	-1.53	-0.98	0.48	0.97
S_p	0.21	-1.54	-0.99	0.48	0.97

5.5 Portfolio weights

Numerical methods are applied to optimise the expected utility. Optimal portfolio weights are calculated using the expected predicted utility. The estimates coming from the Skew t and Skew Normal models are calculated with the different levels of information included in the priors. For all models we consider linear utility functions with two different levels of risk aversion ($\lambda = 10$ and $\lambda = 20$) and various levels of preference for skewness (γ). The optimal portfolio weights calculated for the six models considered are shown in Table 8. Table 9 displays the optimal portfolios obtained using the Markov chain with a smaller number of simulations. The charts are displaying the optimal weights for different levels of the third central moment coefficient γ while the sensitivity coefficient for the covariances λ is fixed. The first bar of each chart, corresponding to $\gamma = 0$, would be the optimal portfolio solution in a Bayesian framework where the third moment is not taken into account.

Considering the third moment in the portfolio optimisation is producing different allocations depending on the preference for skewness of the investor. Moreover the model choice is strongly affecting the composition of the portfolios. The allocation obtained for the two models is different in size and in the predominant hedge fund styles used to build the portfolio. These differences, while present even when the third moment is not taken into account (first bar of each chart), become more important as the λ increases. Thus the differences are triggered by the skewness estimation. In the Skew Normal model more weight is given to Relative Value strategies with Equity Hedge and Event Driven also playing an important role. In the Skew t models, the wealth allocated in Event Driven strategies is reduced and the Macro style becomes more important. In addition the Skew t model is able to exploit Managed Futures strategies characteristic of positively skewed returns by increasing the allocation in this style as the preference for the third moment increases.

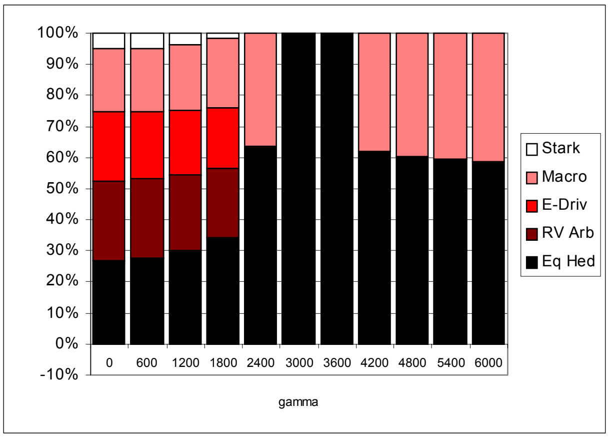
We also notice that in the model assuming a Skew t distribution for the returns the allocation is already affected by small changes in the values of the skewness sensitivity coefficient γ. In the Skew Normal model the changes in the optimal allocation are only obtained with much stronger shifts of γ. This implies that in this model an investor must have a higher preference for skewness to be affected by the introduction of the third moment in the portfolio optimisation problem.

Comparing the different allocations obtained as the priors are becoming more informative, we notice that with uninformative priors the portfolios are more concentrated. Placing informative priors on the model we achieve very well diversified models even without imposing any restriction on the weights. The additional information that we can include in the model by using informative priors can improve the quality of the estimates and therefore decrease the estimation risk. This can provide better results in the portfolio selection.

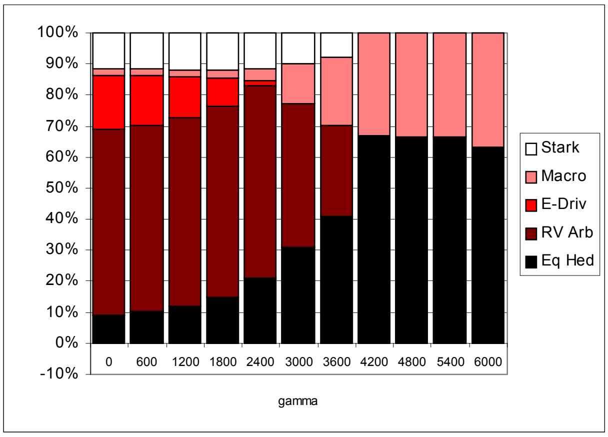
Table 8 - Portfolios obtained by changing the value of the skewness (sensitivity) coefficient γ.

Skew Normal Model (5 Styles) MCMC with 20'000 iterations

Uninformative priors

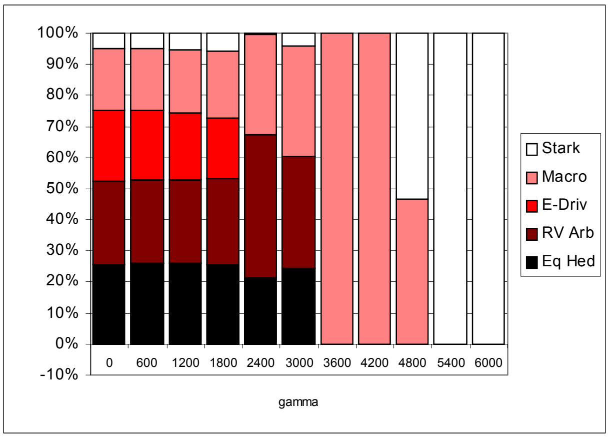


$\lambda = 10$

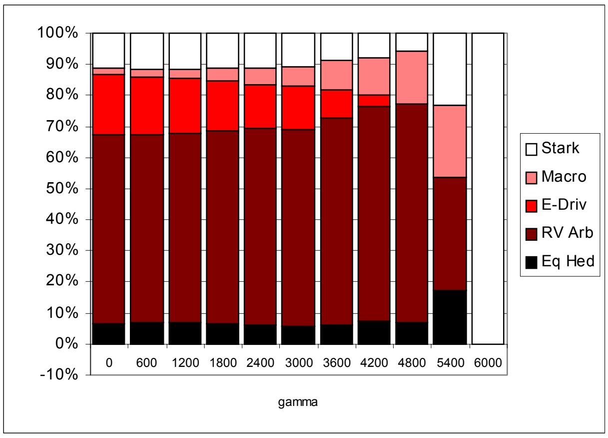


$\lambda = 20$

Informative priors

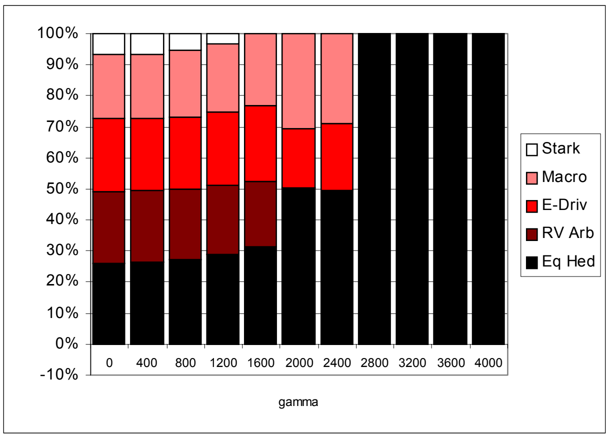


$\lambda = 10$

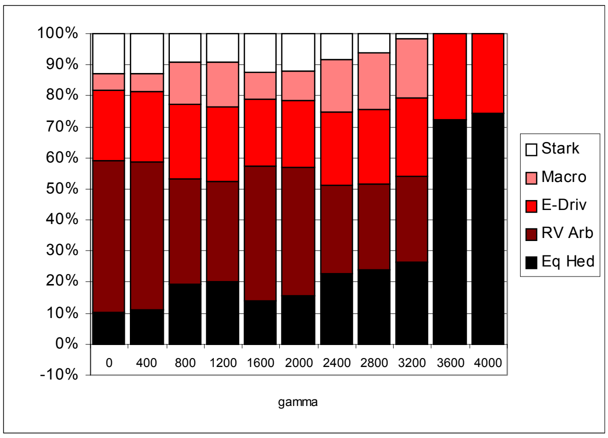


$\lambda = 20$

Highly informative priors



$\lambda = 10$



$\lambda = 20$

Expected Utility and Moments for some of the above displayed portfolios.

Skew Normal - Uninformative priors

λ	10	10	10	10	10
γ	0	1800	3600	5400	6000
EU	0.009	0.010	0.017	0.020	0.023
m_{ptf}	1.19	1.24	1.41	1.37	1.37
var_{ptf}	2.43	2.91	6.51	4.94	4.93
Sk_{ptf}	0.01	0.06	0.15	0.19	0.21

λ	20	20	20	20	20
γ	0	1800	3600	5400	6000
EU	0.008	0.008	0.009	0.015	0.018
m_{ptf}	1.04	1.05	1.21	1.38	1.37
var_{ptf}	1.16	1.18	2.59	5.08	5.01
Sk_{ptf}	0.01	0.04	0.13	0.19	0.21

Skew Normal - Informative priors

λ	10	10	10	10	10
γ	0	1800	3600	5400	6000
EU	0.009	0.009	0.011	0.025	0.029
m_{ptf}	1.16	1.16	1.28	0.77	0.77
var_{ptf}	2.43	2.39	5.91	10.60	10.60
Sk_{ptf}	0.00	-0.01	0.08	0.15	0.15

λ	20	20	20	20	20
γ	0	1800	3600	5400	6000
EU	0.008	0.008	0.008	0.007	0.018
m_{ptf}	1.01	1.01	1.02	1.07	0.77
var_{ptf}	1.12	1.11	1.14	1.94	10.60
Sk_{ptf}	-0.02	-0.03	-0.04	0.05	0.15

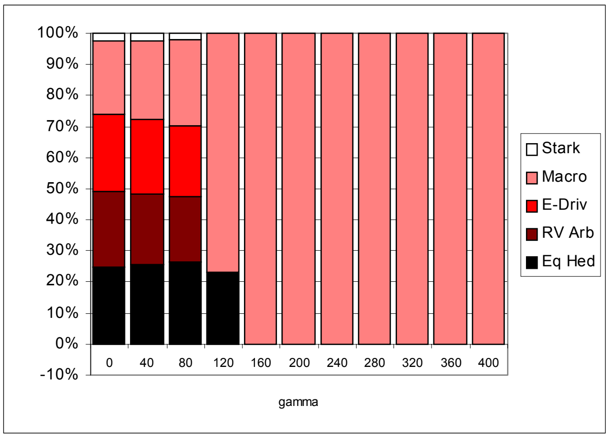
Skew Normal - Highly informative priors

λ	10	10	10	10	10
γ	0	1800	3600	5400	6000
EU	0.009	0.010	0.016	0.026	0.030
m_{ptf}	1.15	1.20	1.38	1.38	1.38
var_{ptf}	2.50	3.08	6.73	6.73	6.73
Sk_{ptf}	0.00	0.06	0.13	0.20	0.22

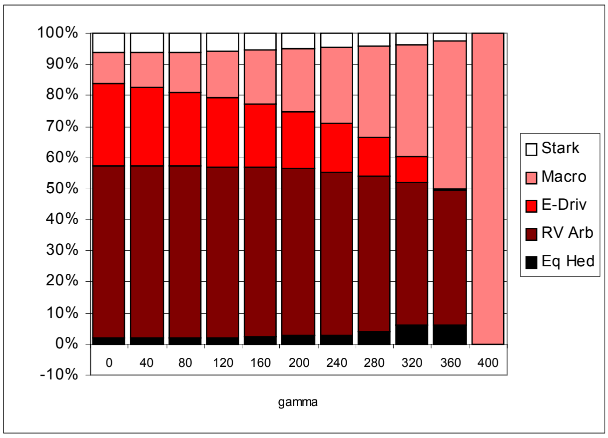
λ	20	20	20	20	20
γ	0	1800	3600	5400	6000
EU	0.007	0.007	0.009	0.019	0.024
m_{ptf}	1.00	1.03	1.31	1.38	1.38
var_{ptf}	1.37	1.56	5.28	6.73	6.73
Sk_{ptf}	-0.04	0.02	0.15	0.20	0.22

Skew t model (5 Styles) MCMC with 20'000 iterations

Uninformative priors

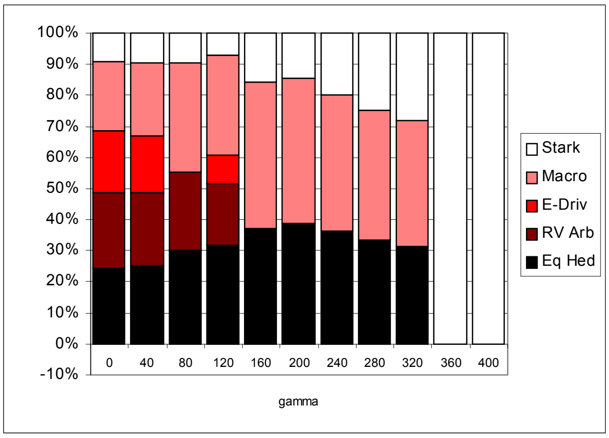


$\lambda = 10$

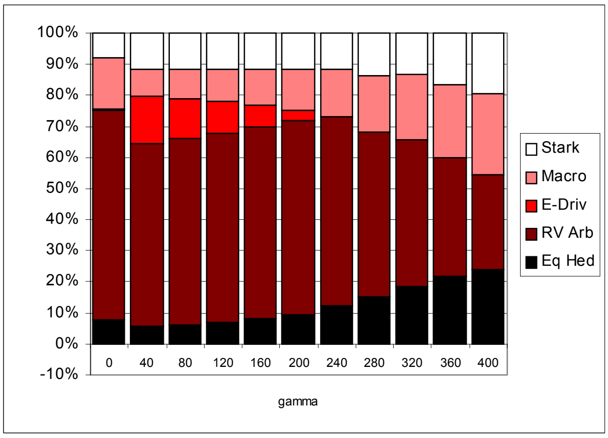


$\lambda = 20$

Informative priors

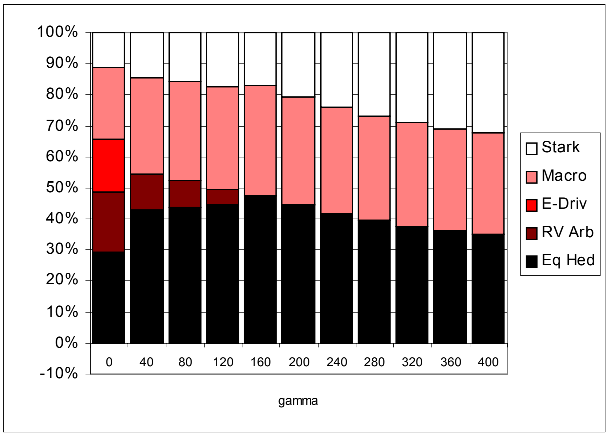


$\lambda = 10$

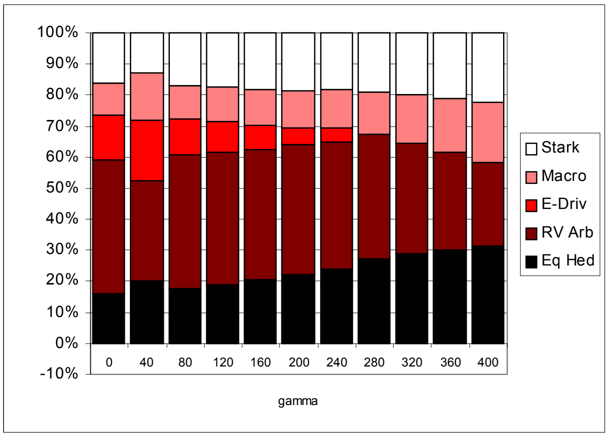


$\lambda = 20$

Highly informative priors



$\lambda = 10$



$\lambda = 20$

Moments for some of the above displayed portfolios.

Skew t - Uninformative priors

λ	10	10	10	10	10
γ	0	120	240	360	400
EU	0.010	0.011	0.015	0.018	0.020
m_{ptf}	1.31	1.54	1.55	1.55	1.55
var_{ptf}	2.95	6.60	7.95	7.95	7.95
Sk_{ptf}	0.67	1.18	1.35	1.35	1.35

λ	20	20	20	20	20
γ	0	120	240	360	400
EU	0.008	0.008	0.009	0.009	0.012
m_{ptf}	1.11	1.13	1.17	1.28	1.55
var_{ptf}	1.41	1.49	1.74	2.88	7.95
Sk_{ptf}	0.48	0.59	0.77	1.12	1.35

Skew t - Informative priors

λ	10	10	10	10	10
γ	0	120	240	360	400
EU	0.009	0.009	0.011	0.020	0.023
m_{ptf}	1.15	1.22	1.30	0.97	0.97
var_{ptf}	2.66	3.36	4.84	13.42	13.42
Sk_{ptf}	0.32	0.53	0.99	1.36	1.36

λ	20	20	20	20	20
γ	0	120	240	360	400
EU	0.007	0.007	0.007	0.007	0.007
m_{ptf}	1.00	0.99	1.02	1.12	1.14
var_{ptf}	1.37	1.32	1.49	2.34	2.69
Sk_{ptf}	0.04	-0.01	0.26	0.69	0.81

Skew t - Highly informative priors

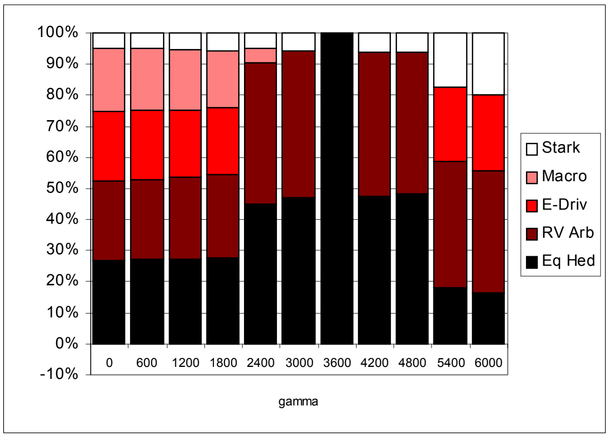
λ	10	10	10	10	10
γ	0	120	240	360	400
EU	0.008	0.008	0.009	0.010	0.010
m_{ptf}	1.08	1.22	1.22	1.19	1.18
var_{ptf}	2.92	4.27	4.51	4.54	4.57
Sk_{ptf}	0.10	0.55	0.68	0.77	0.78

λ	20	20	20	20	20
γ	0	120	240	360	400
EU	0.006	0.006	0.006	0.006	0.006
m_{ptf}	0.92	0.94	0.97	1.04	1.06
var_{ptf}	1.80	1.88	2.03	2.53	2.73
Sk_{ptf}	-0.21	-0.08	0.07	0.35	0.42

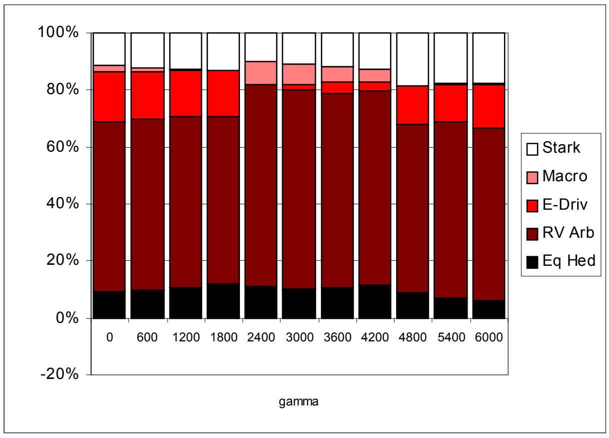
Table 9 - Charts of the optimal portfolios obtained using a Markov chain with a smaller number of simulations.

Skew Normal Models (5 Styles) MCMC with 10'000 iterations

Uninformative priors

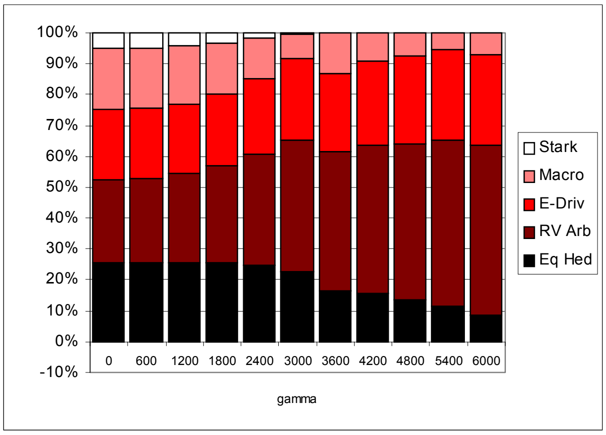


$\lambda = 10$

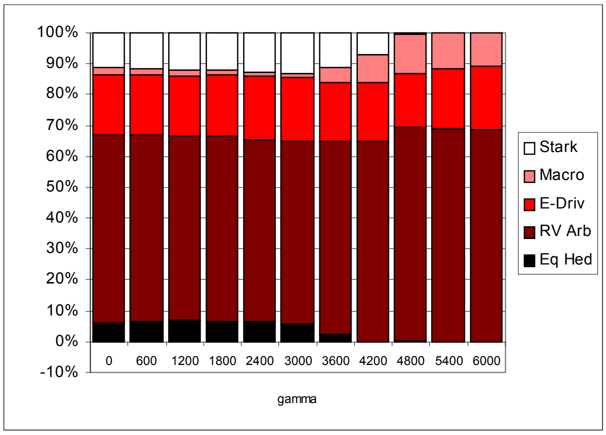


$\lambda = 20$

Informative priors

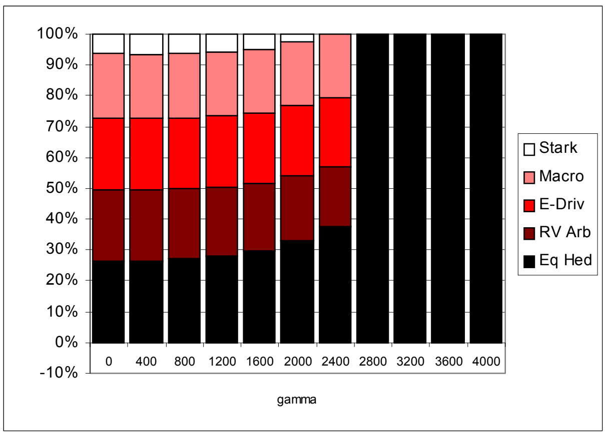


$\lambda = 10$

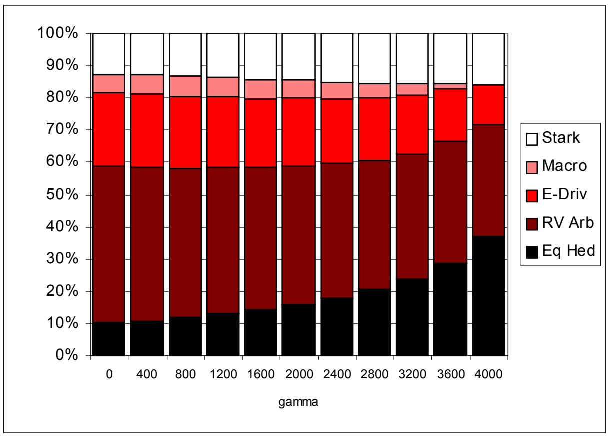


$\lambda = 20$

Highly Informative priors



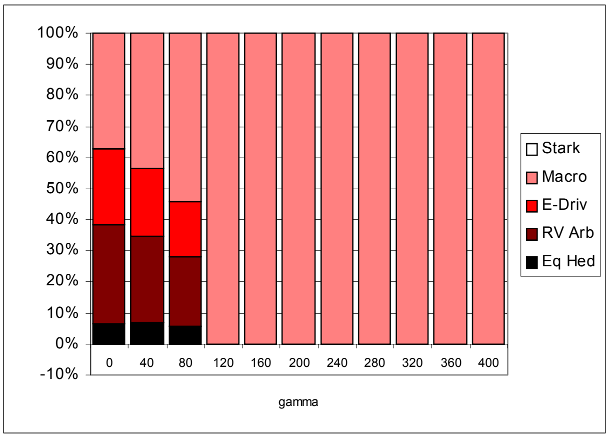
$\lambda = 10$



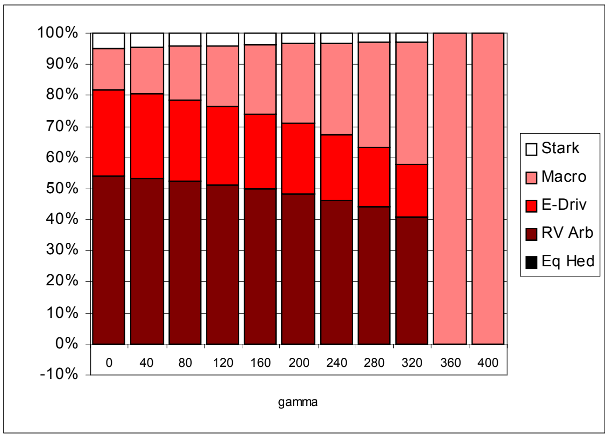
$\lambda = 20$

Skew t models (5 Styles) MCMC with 10'000 iterations

Uninformative priors

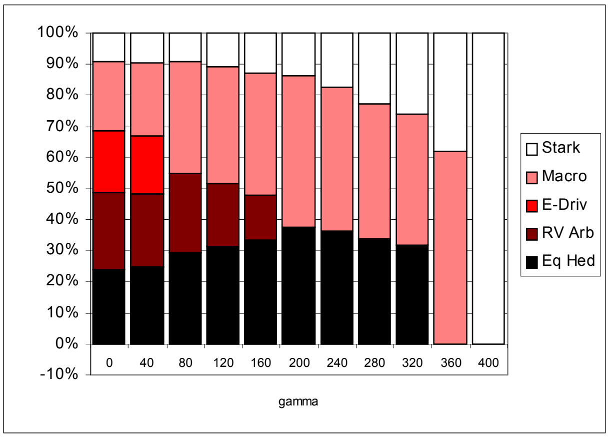


$\lambda = 10$

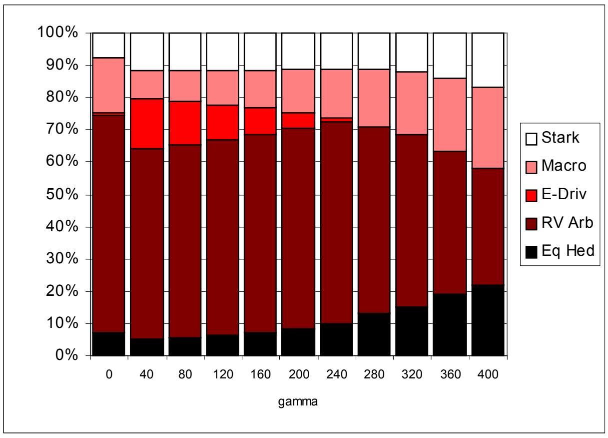


$\lambda = 20$

Informative priors

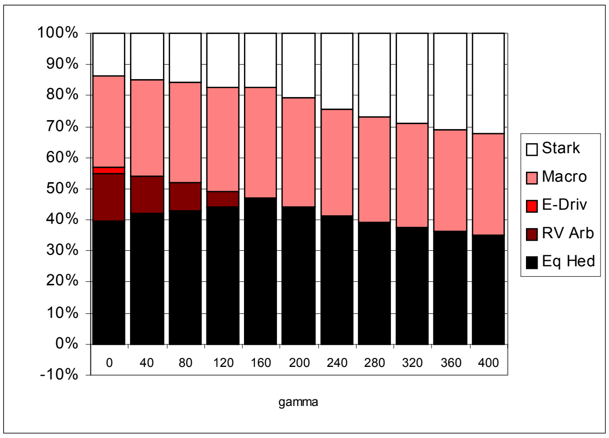


$\lambda = 10$

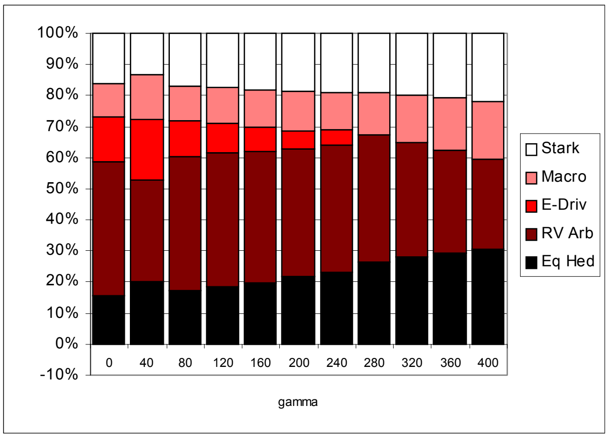


$\lambda = 20$

Highly informative priors



$\lambda = 10$



$\lambda = 20$

6 Conclusions

This thesis provides an application of recent improvements in portfolio selection methods that address the inclusion of higher moments and parameter uncertainty in the portfolio selection task. In particular, the Bayesian approach developed by Harvey, Liechty et al.(2004) is applied to hedge fund portfolio selection and extended to include together with a model based on the Skew Normal distribution, also a model assuming a Skew t distribution for hedge fund returns. Our results confirm that introducing skewness in the asset allocation task will produce a different allocation for investors with skewness preference. The results obtained by fitting this skewed distributions to hedge fund returns data suggest that the Skew t model can improve significantly the description of the return distribution in the hedge fund context. In addition, the inclusion of the third moment in the portfolio selection process becomes much more important if returns are assumed to be driven by a Skew t distribution. In this case, in fact, the allocation is affected by the inclusion of the third moment for lower coefficients of the skewness preference.

An additional feature of the model is exploited by setting informative priors reflecting additional information that in this context can be retrieved from the analysis of the strategies adopted in the different hedge fund styles. We show that the use of informative priors can improve the quality of the estimates and thus reduce the estimation risk.

The portfolio choice problem examined is a static one, which is not taking into account the investment horizon and other issues considered by the growing research in the field of dynamic asset allocation. However, the analysis provided is important in understanding issues in portfolio selection such as the inclusion of higher moments and parameter uncertainty. Another limitation of the model proposed is the use of a diagonal matrix D in the derivation of the skewed distributions. Using an approach assuming a full matrix for D could provide some additional insights, in particular concerning the evaluation and analysis of the coskewnesses.

Another 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 characterictics could provide an improvement in fitting the return distributions.

Appendix

A.1 - Full conditionals for the Skew Normal model.

Assuming n independent observations for $\mathbf{Y}$. The full conditionals for this model can be calculated as:

$$\mathbf{Z}_i \mid \mathbf{y}, \boldsymbol{\mu}, D, \Sigma \sim N_p(A^{-1}\mathbf{a}_i, A^{-1})\ \mathbb{I}(\mathbf{z}_i > \mathbf{0})$$

$$\boldsymbol{\mu} \mid \mathbf{y}, \mathbf{Z}, D, \Sigma \sim N_p(M_n^{-1}\mathbf{m}_n, M_n^{-1})$$

$$\boldsymbol{\delta} \mid \mathbf{y}, \mathbf{Z}, \boldsymbol{\mu}, \Sigma \sim N_p(B^{-1}\mathbf{b}, B^{-1})$$

$$\Sigma \mid \mathbf{y}, \mathbf{Z}, \boldsymbol{\mu}, D \sim IW_p(c_\Sigma + n, C)$$

where

$$\mathbf{a}_i = D'\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\ ;$$

$$A = D'\Sigma^{-1}D + I_p\ ;$$

$$\mathbf{m}_n = \left(\Sigma_\mu^{-1} + n\Sigma^{-1}\right)^{-1}\left(\Sigma_\mu^{-1}\mathbf{m} + \sum_{i=1}^{n}\Sigma^{-1}(\mathbf{y}_i - D\mathbf{z}_i)\right);$$

$$M_n = \left(\Sigma_\mu^{-1} + n\Sigma^{-1}\right)^{-1};$$

$$B = \Sigma_\delta^{-1} + \sum_{i=1}^{n} diag(\mathbf{z}_i)\Sigma^{-1}diag(\mathbf{z}_i)$$

$$\mathbf{b} = \sum_{i=1}^{n} diag(\mathbf{z}_i)\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) + \Sigma_\delta^{-1}\mathbf{d}$$

$$C = \sum_{i=1}^{n}\left((\mathbf{y}_i - (\boldsymbol{\mu}+D\mathbf{z}_i))\,(\mathbf{y}_i - (\boldsymbol{\mu}+D\mathbf{z}_i))' + \Omega_\Sigma\right)$$

References

Adcock, C. J., 2002, "Asset Pricing and Portfolio Selection Based on the Multivariate Skew-Student Distribution," paper presented at the Non-linear Asset Pricing Workshop, April, Paris.

Agarwal, V., and N. Y. Naik, 2004, "Risks and Portfolio Decisions involving Hedge Funds," Review of Financial Studies, 17, 1, 63-98.

Agarwal, V., W. H. Fung, Y. Cheng Loon, N. Y. Naik, 2004, "Risks in Hedge Fund Strategies: Case of Convertible Arbitrage," Working paper. Georgia State University and London Business School.

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.

Arrow, K. J., 1971, "Essays in the Theory of Risk Bearing," Chicago: Markham Publishing. Chicago.

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.

Bacmann, J. F. and U. Bosshard, 2004, "Optimising portfolios with asymmetric return distributions: An application to hedge fund styles".

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

Bawa, V. S., J. B. Stephen, and W. K. Roger, 1979, "Estimation Risk and Optimal Portfolio Choice," Amsterdam: North-Holland.

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.

Conrad, J. and G. Kaul, 1998, “An anatomy of trading strategies,” Review of financial studies, 11, 489-519.

De Bondt, W.F.M. and R. Thaler, 1985, “Does the stock market overreact?,” Journal of finance, 40, 3, 793-805.

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.

Fung, W., and D. A. Hsieh, 1997, “Empirical characteristics of dynamic trading strategies: The case of hedge funds,” Review of Financial Studies, 10, 275-302.

Gatev, E. G., W. N. Goetzmann and K. G. Rouwenhorst, 1999, “Pairs Trading: performance of a relative value arbitrage rule,” Working Paper, Yale School of Management.

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.

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.

Hastings, W. K., 1970, “Monte Calo Sampling Methods Using Markov Chains and their Applications,” Biometrika, 57, 97-109.

Ingersoll, J. E. Jr., 1987, “Theory of Financial Decision Making,” Baltimore: Rowman and Littlefield.

Jeffreys, H., 1935, “Some Tests of Significance, Treated by Theory of Probability,” Proceeding of the Cambridge Philosophical Society, 31, 203-222.

Jegadeesh N. and S. Titman, 1993, “Returns to buying winners and selling losers: Implications for stock market efficiency,” Journal of Finance, 48, 65-91.

Jobson, J. D. and B. Korkie, 1980, “Estimation of Markowitz e .cient portfolios,” Journal of the American Statistical Association, 75, 544-554.

Jondeau, E. and G. M. Rockinger , 2004, “Optimal Portfolio Allocation Under Higher Moments,” EFMA 2004 Basel Meetings Paper.

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

Jorion, P., 1986, “Bayes-Stein estimation for portfolio analysis,” Journal of Financial and Quantitative Analysis, 21, 279–292.

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.

Kass, R. E. and A. E. Raftery, 1995, “Bayes Factors,” Journal of the American Statistical Association, 90, 773-795.

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

Lavine, M. and M. J. Schervish, 1999, “Bayes factors: what they are and what they are not,” The American Statistician, 53, 119-122.

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

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953, “Equations of State Calculations by Fast Computing Machines,” Journal of Chemical Physics, 21, 1087-1091.

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

Mitchell, M., and T. Pulvino, 2001, “Characteristics of Risk in Risk Arbitrage,” The Journal of Finance, 56, 6, 2135-75.

McNeil, A. J., R. Frey and P.Embrechts, forthcoming, 2005, “Quantitative Risk Management: Concepts, Techniques and Tools”.

Newton, M. A.and A. E. Raftery, 1994, “Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion),” Journal of the Royal Statistical Society, Series B, 56, 3-48.

Pastor, L., 2000, “Portfolio selection and asset pricing models,” Journal of Finance, 55, 179 –223.

Pastor, L., and R. F. Stambaugh, 2000, “Comparing asset pricing models: An investment perspective,” Journal of Financial Economics, 56, 335 –381.

Patton, A. J., 2004, “On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation,” Journal of Financial Econometrics, 2, 1, 130-168.

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.

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).

Scott, R. C. and P. A. Horvath, 1980, "On the Direction of Preference for Moments of Higher Order Than Variance," Journal of Finance, 35, 4, 915-919.

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.