Nothing is normal in finance!

On Tail Correlations and Robust Higher Order Moments in Normal Portfolio Frameworks

February 16, 2012

Author:
Jan Engshagen

Supervisors:
Filip Lindskog (KTH)
Jesper Sandin (Lynx)

Abstract

This thesis project is divided in two parts. The first part examines the possibility that correlation matrix estimates based on an outlier sample would contain information about extreme events. According to my findings, such methods do not perform better than simple shrinkage methods where robust shrinkage targets are used. The method tested is especially outperformed when it comes to the extreme events, where a shrinkage of the correlation matrix towards the identity matrix seems to give the best result.

The second part is about valuation of skewness in marginal distributions and the penalizing of heavy tails. I argue that it is reasonable to use a degrees of freedom parameter instead of kurtosis and a certain regression parameter, that I develop, instead of skewness due to robustness issues. When minimizing the one period draw-down is our target, the "value" of skewness seems to have a linear relationship with expected returns. Re-valuing of expected returns, in terms of skewness, in the standard Markowitz framework will tend to lower expected shortfall (ES), increase skewness and lower the realized portfolio variance. Penalizing of heavy tails will most times in the same way lower ES, lower kurtosis and realized portfolio variance. The results indicate that the parameters representing higher order moments in some way characterize the assets and also reflect their future behaviour. These properties can be used in a simple optimization framework and seem to have a positive impact even on portfolio level.

Keywords: Correlation Matrix, Tail Correlations, Skewness, Robust Skewness, Kurtosis, Higher Order Moments, Portfolio Optimization, Expected Shortfall, Value at Risk

Acknowledgements

I would especially like to thank Jesper Sandin, my supervisor at Lynx Asset Management, for encouragement, ideas and great feedback. I would also like to thank the rest of the staff at Lynx, who have shown interest and encouraged me in my work. Extra thanks goes to Daniel for help with the data and Anders for the extra motivation. Without my supervisor at KTH, Filip Lindskog, I would not have been able to do this project. Thank you for valuable tips on writing and how to work with a project.

Stockholm, February 2012

Jan Engshagen

Contents

Chapter 1

Introduction

The popularization of models based on assumptions of normal distributions date back to the beginning of modern portfolio theory. Modern portfolio theory can be considered to have it's origin in 1952, when Markowitz [1952] publicized his famous article *"Portfolio Selection"*. For his contribution in to the topic, Markowitz received the Nobel Memorial Price in Economic Sciences. The result is a quadratic optimization framework, where expected returns are weighted against the expected portfolio variance, that is a risk measure based on the properties of a normal distribution. However, asset returns are very seldom normally distributed[1], if not to say never. Many studies have shown this. For example Mandelbrot [1963] and Fama [1965] has gotten early attention for pointing out the rather fractal distributions of asset returns. Other distributions have been suggested as for example the students t-distribution by Blattberg and Gonedes [1974]. The fact that the distribution of data disagrees with common risk parameters used in the simple optimization frameworks can be troubling. For example, investors will take on larger risks than what the models assume if they don't use caution.

Since normal-based frameworks are common in practice, and asset return distributions often look different, I have as an aim in this project, to look at different ways of analysing the properties that don't fit the assumption of a normal distribution and incorporate them in normal-based frameworks. During this semester I have had the opportunity to make this thesis project at the Swedish hedge fund Lynx Asset Management AB. Lynx is a trend following hedge fund, that rely on model based investment decisions in a portfolio of futures contracts. Since I will work for Lynx, the assets to be studied are futures contracts.

[1]For a more popular scientific critique of the assumptions of normal-based investment models and assumption, read "*The Black Swan*"[Taleb, 2007]. It's written by a quantitative analyst and contain many good sources for further reading for the technically interested.

The project is divided into two parts. In the first part I will look at the dependence structure of data in terms of correlations among "rare" events. Some observations, usually observed in times of market turbulence or observations like a market shock, will seem more extreme than the usual observations. Chow et al. [1999] argues that these outlier observations will have a different dependence structure than the ones in the inside sample. It is therefore interesting to test if these extreme observations contain any information that will help to make better predictions of the dependence structures of future extreme events. To test this I will look at the likelihood of observations made from the perspective of different correlation matrices. A correlation matrix based on the standard estimation methods will be used as a benchmark and compared to various set-ups of blended matrices, based on outliers and inside sample estimates. As a comparison more robust matrix estimates will also be used.

In the second part I will instead look at the higher order moments of the asset return distributions. The normal distribution is only characterized by it's first two moments[Blom et al., 2005], mean and variance, the asset returns however show clear signs of asymmetry (skewness) and a much higher likeliness of extreme events/tail thickness (kurtosis). The actual higher order moments tend to be quite unreliable when it comes to estimating them from data, as for example pointed out by Harvey et al. [2004]. Therefore I will devote some time to find fitting substitutes for them. Substitutes that should be easy to interpret in similar ways as the standard definitions of the moments. It is of great interest to see, how these parameters can be used to enhance portfolio performance. To narrow it down I will try to incorporate them in standard normal based frameworks, which are still common in use. As a performance measure it will be natural to look at realized portfolio risk[2], since it is not return prediction I will be focusing on. It is often mentioned that investors actually value skewness in terms of returns and to investigate how it can be done I will conduct a simulation study. To test the results I will run portfolio optimizations on historical data.

The following chapter contains the theory and explanations about the terms and the mathematics that will be used in chapter 3, 4 and 5. Chapter 3 contains the description of the data sets and a more well motivated background and problem formulation for the two different parts of the project. Chapter 4 contains a description of the methods in use, the different set-ups and choices of parameters. As well all the empirical results and findings are presented here. In chapter 5 the results are summarized, interpreted ans

[2] With realized risk I mean the risk estimates one can make based on historical observations of portfolio returns.

discussed.

The aim of this paper is that you should be able to read and understand it even if you don't have a background in finance. If you however lack basic knowledge of calculus, linear algebra and some statistics it will be hard. Proofs are mostly omitted and if the reader should have any interest in reading them I refer to the literature in the bibliography. I will instead try to give the reader an intuitive understanding of the theory and provide the the central results and formulas.

Chapter 2

Theory

Here follows the theory part. It contains the mathematics and most of the theory that are used throughout the empirical parts. It also takes up some terms that might not be known by the readers that do not have a background in finance. However, I assume that the reader has a general knowledge of calculus, linear algebra and basic statistics.

2.1 Distributions

A probability distribution can be expressed as a function that tells us about the probability of an observation. Let us denote a random variable (RV) X, that takes values in $\mathbb{R}$, and the probability of a certain event be denoted $p(\text{event})$. If X behaves in a nice way, it will have a probability density function (PDF) $f(x)$. The density function can be defined by the equality:

$$p(X \leq x) = \int_{-\infty}^{x} f(u)du = F(x)$$

Where $F(x)$ is called the cumulative distribution function (CDF).[1]

2.1.1 Normal and multivariate normal distribution

In [Blom et al., 2005], we can read about the normal distribution. It is a distribution that has the density function

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

If a RV, X, has an expected value, $E(X) = \mu$, a standard deviation of σ and the density function given above, we say that $X \sim N(\mu, \sigma)$. That is, normally distributed with parameters μ and σ.

[1] Note that in the discrete case $f(x)$ is the actual probability for a certain event.

This might be the most popular and frequently assumed distribution. First of all it has very appealing computational properties, e.g. most problems have relatively easy analytical solutions, sums of normally distributed RVs are also normal etc. And second, many things in nature tend to be normally distributed.

The multivariate normal distribution is actually the one of greatest interest here, since it is the base for the most common portfolio optimization models. Consider the random vector $X = [X_1\, X_2 \ldots X_n]$, with $E(X) = \mu = [\mu_1\, \mu_2 \ldots \mu_n]$ and covariance matrix

$$Cov(X) = \Sigma = \begin{bmatrix} var(X_1) & cov(X_1, X_2) & \cdots & cov(X_1, X_n) \\ cov(X_2, X_1) & var(X_2) & \cdots & cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n, X_1) & cov(X_n, X_2) & \cdots & var(X_n) \end{bmatrix}$$

It is multivariate normal, $X \sim N(\mu, \Sigma)$, if it has the density [Gut, 1995]

$$f(x) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} \exp\left(\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

note that it is required for the covariance matrix to be positive definite and symmetric, which it always will be if nothing else is mentioned.

2.1.2 Students t-distribution

In Blom et al. [2005] we also find the student's t-distribution. It has a shape that reminds about the normal distribution, it is a symmetric distribution that looks like it has the shape of a bell. The advantage with the t-distribution is that it takes care of thick tails. The PDF looks like:

$$t_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu-1}{2}}$$

(2.1)

The parameter ν, tells us how many degrees of freedom the distribution has. It actually derives from the case where we observe n independent and identically normally distributed RVs with unknown σ and μ. Then $\frac{X-\mu_e}{\sigma_e}$, where μ_e and σ_e are sample estimates, will be t-distributed with $\nu = n - 1$. Tail thickness is determined by ν and we see directly from the PDF that it converges as a polynomial rather than an exponential as the normal distribution do.

In the rest of the text when I mention the t-distribution, I will actually refer to the t-location scale. It has two additional parameters, μ and σ, such that a RV, $\frac{X-\mu}{\sigma}$, will be t-distributed with a parameter ν. When fitting a t-distribution to data, these parameter estimates differ from the estimates of sample mean, μ_e, and standard deviation, σ_e, above.

2.1.3 Polynomials of symmetric distributions and QQ-plots

Suppose that we have a symmetric PDF, $f(x)$ fitted to a data set with n observations. A very nice way to graphically display how well the distribution fit with data is to make a QQ-plot. It plots the theoretical quantiles implied by the distribution versus the empirical ones[2]. For example, if we have 100 observations, the empirical 4%-quantile, $Q_{E,0.04}$, is the 4^{th} smallest observation and the theoretical 4%-quantile is $F^{-1}(0.04)$. If we have a perfect fit, they should be equally large and form a straight line. As an illustration see Figure2.1, where quantiles from an empirical thick-tailed distribution is

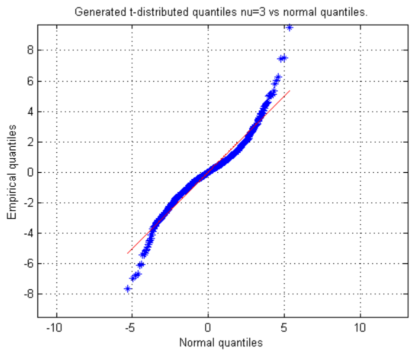

Figure 2.1: *1500 observations generated from a t-distribution with $\nu = 3$ vs theoretical normal quantiles*

plotted against a normal distribution.

The most eye-catching behaviour here is the large deviation of the tails of the t-distribution. Note also that the graph still has a very smooth shape, that reminds about a third degree polynomial. A very nice approach on this is made by Hult et al. [2011], with normal distributions. If we in such a case take a new set of parameters, $\theta = \theta_0 : \theta_3$, the quantile levels, $LvL_i = \frac{i}{n}$, $i = 1 : n$ and make a new set-up:

$$Q_{E,LvL_i} = \theta_0 + \theta_1 F^{-1}(LvL_i) + \theta_2 F^{-1}(LvL_i)^2 + \theta_3 F^{-1}(LvL_i)^3$$

(2.2)

[2]When I mention an empirical distribution I refer to a set of observations or a generated sample

Where $F(x)$ is the CDF of the quantiles on the x-axis. This will give us a new distribution and the parameters can be estimated using various methods such as least squares (LS) or maximum likelihood (ML). It is important to be able to interpret the parameters. θ_0 is simply a constant to ensure that the mean of the distribution does not change. θ_1 tells something about how well the original fitted distribution fits the data[3]. θ_2 will in some way tell us about skew-properties of the data, it is easy to realize that if data is totally symmetrically distributed, this parameter will be 0. To illustrate this take a look at Figure 2.2 , where generated skewed normal quantiles, where one

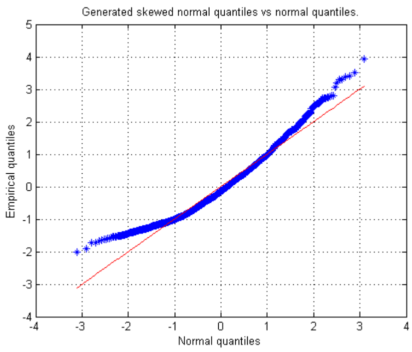

Figure 2.2: *1500 observations generated from a skewed normal distribution vs theoretical normal quantiles*

tail is made heavier than the other, is plotted against normal quantiles. The skewed normal distribution in question is generated according to (2.3) in the next section. It looks much more like a parable, as we would expect from the term $F^{-1}(LvL_i)^2$. The last one, θ_3, will of course tell us something about the thickness of the tails in comparison to $f(x)$. As we will see later on, it is very appealing to choose F in (2.2) to be the t-distribution and only use the first three parameters, since ν will describe the tail thickness of the data analysed pretty well.

2.1.4 A new type of multivariate distributions

Sometimes it is useful to have a distribution where higher order moments can be varied and mean and covariance can be kept constant, when sampling from it. One way of doing so is to use a type of multivariate skewed distributions developed by Ferreira and Steel [2007]

If f denotes a univariate pdf symmetric around zero and γ is a positive

[3]If $\theta_1 = 1$ and the other parameters 0, we have a perfect fit for the original distribution

scalar, we can construct the skewed univariate pdf:

$$p(\epsilon|\gamma, f) = \frac{2}{\gamma + \frac{1}{\gamma}} f\left(\epsilon \gamma^{sign(\epsilon)}\right)$$

(2.3)

For a multivariate distribution with N variables, the multivariate distribution with independent components is given by

$$p(\epsilon|\gamma, f) = \prod_{n=1}^{N} p(\epsilon_n|\gamma_n, f_n)$$

To add mean and a covariance structure to the distribution, we define a new variable

$$\eta = A'\epsilon + \mu$$

(2.4)

μ is a $N \times 1$-vector and A is a non-singular $N \times N$-matrix. The nice thing about the distribution of η, is that there are explicit expressions of the expected value and covariance matrix.

$$E(\eta) = \mu + M_1 A' \begin{pmatrix} \gamma_1 - \frac{1}{\gamma_1} \\ \vdots \\ \gamma_N - \frac{1}{\gamma_N} \end{pmatrix}$$

(2.5)

$$Cov(\eta) = A' \left\{ Diag \left[(M_{2,j} - M_{1,j}^2) \left(\gamma_j^2 - \frac{1}{\gamma_j^2} \right) + 2M_{1,j}^2 - M_{2,j} \right]_{j=1,\ldots,N} \right\} A$$

(2.6)

$M_{r,i}$ denotes the $r^{th]}$ order moment of f_i:

$$M_{r,i} = \int_0^\infty s^r 2 f_i(s) ds$$

The drawback with the distribution in question is that it is not closed under marginalisation. That means that after the dependence structure is applied the marginal distributions will not have the same form as before the transformation. When using this distribution for the generation of data this will however not be an issue as long as we keep it in mind.

Generating a sample

To generate a sample from this distribution I use the following approach: If we have decided on the mean, μ_{wanted} and covariance structure, Σ_{wanted} that we want. When choosing the A and the μ, a very simple approach is to let the computer solve for A and μ such that $|A_{wanted} - Cov(\eta)|_F$ is minimized and then the norm of $\mu_{wanted} - E(\eta)$. $||_F$ here denotes the Frobenius norm and is simply the square root of the sum of the squares of all matrix elements. This will give us the parameters we seek.

If we want to simulate a sample of data, T observations of N assets, we first have to generate T random numbers for each one of the N univariate skewed PDFs with a distribution according to (2.3), before we can transform the data with the parameters A and μ. For a given univariate, symmetric centred PDF, f, we have if $\epsilon < 0$:

$$P(\epsilon < x|\gamma, f) = \frac{2}{\gamma + \frac{1}{\gamma}} \int_{-\infty}^{\epsilon} f(\gamma x) dx =$$

$$= \frac{2}{\gamma + \frac{1}{\gamma}} \int_{-\infty}^{\gamma\epsilon} f(t)\gamma dt = \frac{2\gamma^2}{\gamma^2 + 1} F(\gamma\epsilon)$$

From that we can solve for ϵ:

$$\epsilon = \frac{1}{\gamma} F^{-1} \left(\frac{P \cdot (\gamma^2 + 1)}{2\gamma^2} \right)$$

(2.7)

in a similar way we can derive that for $\epsilon > 0$ we have that:

$$\epsilon = \gamma F^{-1} \left(\frac{P \cdot (\gamma^2 + 1) - 1}{2\gamma^2} + \frac{1}{2} \right)$$

(2.8)

also note that $P(\epsilon > 0|\gamma, f) = \frac{\gamma^2}{1+\gamma^2}$. Hence if we want to generate a random sample, we first generate uniformly distributed numbers, P, from the interval $[0, 1]$, corresponding to probabilities. Then we compute ϵ, using (2.7) if $P < \frac{\gamma^2}{1+\gamma^2}$ and (2.8) if $P > \frac{\gamma^2}{1+\gamma^2}$. If f is a common distribution, F^{-1} will be easy to access in for example Matlab.

Now we will have samples of independent random numbers from N independent distributions. Arrange the data in a matrix N columns with T observations numbers each. Each row now correspond to an observation, that we transform according to (2.4) to get the desired linear dependences and expected values

2.2 On Outliers and Correlation

2.2.1 Outliers

According to Stock and Watson [2008] an outlier is an extreme observation. What an extreme observation is, is a little bit fuzzy, but certainly something that we don't see every day. Typically it is referred to something that we would not expect to see if the assumption about a normal distribution was true. In this paper however when I refer to an outlier, and an outlier at a certain level, I will mean tail observations. By identifying outliers I mean identifying if observations belong to the outer part of the tails of the distribution or not, where outer part has to be specifyed by some level.

For the univariate case identifying an outlier is not a problem. If we decide on a certain level, say $q\%$, it means that we consider the top $q\%$ most deviating observations. For a RV, X, with CDF $F(x)$, the criteria for an outlier can be given by: $x\ s.t.\ F(x) \leq q/2$ and $x\ s.t.\ F(x) \geq 1 - q/2$. If the distribution is symmetric around mean, this can be represented with a certain distance from mean.

For a multivariate distribution it however becomes a little bit more complicated. Chow et al. [1999] give us a tool for accomplishing the task. If data is assumed to be normally distributed, for an observation, $x = [x_1\ x_2\ ...\ x_n]$ of the RV X, we can look at the distance measure

$$d = (x - \mu)\Sigma^{-1}(x - \mu)'$$

(2.9)

where $\mu = E(X)$ and $\Sigma = cov(X)$. This is actually a very intuitive definition of a multivariate distance. Note that if we would be dealing with normalized RVs, that is $X_{n,normalized} = \frac{X_n}{std(X_n)}$, Σ would then be a correlation matrix (since $var(X_{n,normalized}) = 1$) and d would be the number of standard deviations from mean, in a multivariate sense. If we have a series of data, we can simply choose a certain distance that represent the $q\%$ largest values of d to identify outliers.

2.2.2 Tail correlations and blended covariance matrices

There are some empirical results that indicate that correlations and volatilities estimated on outlier samples differ significantly from estimates based on the whole sample. Since financial data usually isn't normally distributed it would be a brave assumption that the tail of the distribution would behave as the inside sample in terms of linear correlation. Chow et al. [1999] suggest that it's useful to create a blend of the two covariance matrices, so that weight can be put on the outlier matrix in turbulent times and give a better protection against risks. If we use the notation:

$\Sigma =$ covariance matrix based on the whole sample

$\Sigma_G =$ covariance matrix of the inside sample

$\Sigma_B =$ covariance matrix of the outlier sample

$\Sigma_{GB} =$ blended covariance matrix

$q_{out} =$ probability for being in the outlier sample

$\lambda_G =$ inside risk aversion

$\lambda_B =$ outlier risk aversion, $\lambda_B + \lambda_G = 2$ must hold

By identifying outliers with the distance measure in (2.9) and choosing a level of q, the blended covariance matrix will be

$$\Sigma_{GB} = \lambda_G(1-q)\Sigma_G + \lambda_B q \Sigma_B$$

(2.10)

By changing the risk aversion, we can put more weight on the tail covariances or the inside covariances.

2.2.3 The Most Likely Correlation Structure

When it comes to identifying multivariate outliers d is a good measure. But when testing if a series of observations has a certain correlation structure it is not a good measure. Assume we have a time series of T centred returns, for N assets, denoted dX. Assume further that we have two different suggestions for the covariance matrix of the distribution of returns, Σ_1 and Σ_2 (with the same variance). It is tempting to use the sum $\sum_{i=1}^{n} dX_t \Sigma^{-1} dX_t^T$ and see which covariance matrix that minimizes it. An interpretation is, that the minimizing correlation structure gives the smallest average multivariate distance from mean for the observed returns.

This approach however does not tell us about if it is the most likely correlation structure that will be the minimizer. Just consider the two-dimensional case with:

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}$$

A certain level of d will in the first case be represented of a circle with radius $\sqrt{d}$, in the dX-plane, whilst in the second case will be represented by an ellipse. This is easily realized if we look at

$$d = dX\Sigma^{-1}dX^T = dX\frac{1}{1-c^2}\begin{bmatrix} 1 & -c \\ -c & 1 \end{bmatrix}dX^T =$$

$$= \frac{1}{1-c^2}(dX_1^2 - 2c \cdot dX_1dX_2 + dX_2^2)$$

which is an ellipse. For a c close to one, the minor axis will be very narrow and the distance measure, d, will be very large for outliers in this direction. The average d will be misleading, since it does not tell us anything about the likelihood of the observation.

Let us instead construct a likelihood function. First, consider the eigen-decomposition of Σ,

$$\Sigma = VDV^{-1}$$

where V is the orthogonal matrix of eigenvectors and D is the diagonal matrix with the corresponding eigenvalues [Anton and Rorres, 2005], λ_n. The decomposition can always be done since a covariance matrix is positive definite [Gut, 1995]. Now let $dF = dXV$ and note that since V is orthogonal $V^{-1} = V^T$. Due to the fact that D is a diagonal matrix, we can now rewrite the expression for d.

$$d = dX\Sigma^{-1}dX^T = dXVD^{-1}V^{-1}dX^T = dFD^{-1}dF^T$$

Note that we now can construct a likelihood function for dF as if it came from independent normal distributions with variances corresponding to the eigenvalues in D. The distribution function of dF_i will then be

$$f(dF_n) = N(0, dF_n) = \frac{1}{\sqrt{2\pi\lambda_n}}e^{-\frac{dF_n^2}{2\lambda_n}}$$

and we will then get the likelihood function

$$L(dF) = \prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\lambda_n}}e^{-\frac{dF_n^2}{2\lambda_n}}$$

It is much more convenient to look at the log-likelihood function since then we can work with sums instead if we want find a maximizer.

$$L_{log}(dF) = \sum_{n=1}^{N}\log\left(\frac{1}{\sqrt{2\pi\lambda_n}}e^{-\frac{dF_n^2}{2\lambda_n}}\right) = \sum_{n=1}^{N} -\left(log(\sqrt{2\pi\lambda_n}) + \frac{dF_n^2}{2\lambda_n}\right) =$$

$$= \sum_{n=1}^{N} -\frac{1}{2}\left(log(\lambda_n) + log(2\pi) + \frac{dF_n^2}{\lambda_n}\right)$$

(2.11)

This measure says something about the likelihood of an observation compared to a certain correlation structure. And by looking on an average across the time series we will be able to compare different models for correlation structure.

2.2.4 Shrinkage

To make more robust estimates of a covariance or correlation matrix it is common to shrink the covariance matrix against some target. To shrink the covariance matrix simply means mixing the estimate with another (hopefully) more robust estimate. So that:

$$\Sigma = \alpha\Sigma_{Prior} + (1-\alpha)\Sigma_{Target}$$

where $\alpha \in [0, 1]$, is called the shrinkage parameter, Σ_{Prior} is the prior estimate and Σ_{Target} is the shrinkage target. It can be useful if the covariance matrix is estimated on a small sample, with a size comparable to the number of parameters that need to be estimated.

When I later on will use shrunken matrices as a comparison to the outlier based estimates, I will look at a whole range of α, therefore I am not interested in looking at optimal shrinkage parameter. It can however be worth mentioning that for some targets we can find an optimal shrinkage parameter. For more on the topic, see for example Ledoit and Wolf [2004] where a one-factor model with CAPM regression parameters is used. I will leave out this model to avoid the discussion about what to chose as market returns for the futures portfolio when looking at absolute return series.

For my purposes it will be sufficient to use the identity matrix and a factor-model based on eigenvectors as shrinkage target. The factor model I mention is practically based on Principal Component Analysis (PCA). A tutorial on PCA is given by Shlens [2009]. The idea of PCA builds on the assumption that we can observe different factors that explain most of the variance and that the factors that contribute very little to the variance are produced by noise. By removing these factors we can create a more robust estimate. If the eigendecomposition of a estimated $N \times N$-covariance matrix, Σ, based on a set of multivariate observations X. Then Σ can be expressed as $\Sigma = V'DV$, where V is the matrix of eigenvectors and D a diagonal matrix with the eigenvalues. We can by changing the $N - n$ lowest eigenvalues to 0 to get an estimate of Σ based on only the n most significant factors. The factors we observe actually derives from the transformed data series XV that have D as a covariance matrix.

2.3 Moments

We all have our moments and so do random variables, but in a somewhat different matter. For a RV, X, with a PDF, f, the r^{th} order moment and centred moment can respectively according to Gut [1995] be defined as

$$E\left(X^r\right) \quad \text{and} \quad E\left((X - E(X))^r\right)$$

The first moment and the second order centred moment are known to us as mean and variance. In multivariate distributions we also have comoments, covariance is such an example of the second order.

2.3.1 Higher order moments, skewness and kurtosis

With higher order moments it is commonly meant, moments of order three or greater. Of particular interest are the third and fourth order moments. The typical version of these that are frequently used are skewness and kurtosis [Stock and Watson, 2008]. If σ denotes the standard deviation of a RV and μ the mean, X, then

$$\text{Skewness}(X) = E\left(\frac{(X-\mu)^3}{\sigma^3}\right) \quad \text{and} \quad \text{Kurtosis}(X) = E\left(\frac{(X-\mu)^4}{\sigma^4}\right)$$

(2.12)

If we look at the definition, it is easy to realize that the skewness of a distribution will tell us about how thick the tails of a distribution are compared to one other. If extreme negative events are more likely than positive events (thicker negative tail), the distribution will have a negative skewness and the other way around. A symmetric distribution will have a skewness of zero.

Kurtosis on the other hand can be interpreted in two ways. One is that higher valuer imply that the distribution thicker tails and the other is that the "peakiness" of the distribution is high. With peakiness I mean that most observations are centred around mean, leading to a small standard deviation with a comparable high impact from the fourth power of the deviation. As a an example, the normal distribution has a kurtosis of exactly three. Kurtosis exceeding three are often referred to as excess kurtosis.

Robustness problems

Skewness and kurtosis are not very stable when it comes to estimates in empirical data. They are easily biased by outliers, and are as well not even defined for distributions with low degrees of freedom. For financial time series of returns it is not uncommon to observe a parameter $\nu \approx 3$, when a t-distribution is fitted to data.[4] It is important to mention that kurtosis is undefined if $\nu \leq 4$, as well skewness is undefined when $\nu \leq 3$, which is quite obvious if we look at the definition (2.12) and the t-distribution (2.1). The polynomial properties of the distribution will make the expected value diverge at the mentioned limits since $\int \frac{1}{x} dx = ln(x)$ does not converge at infinity. The empirical estimation of kurtosis of a sample will show this very clearly by a diverging behaviour when sample sizes get large. To get an illustration of this, take a look at Figure 2.3 where kurtosis is estimated

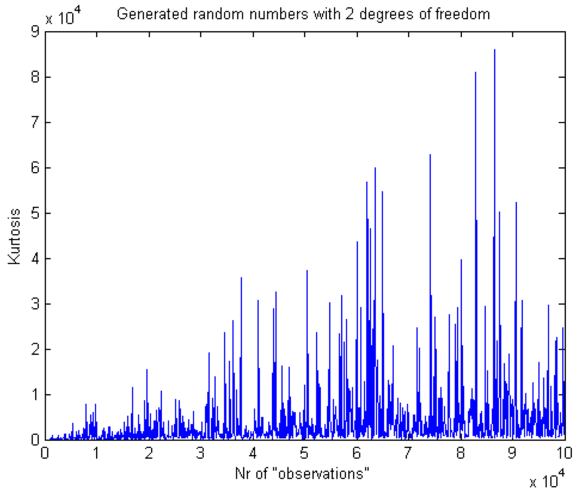

Figure 2.3: *Number of random t-numbers with $\nu = 2$ in a sample vs. estimated kurtosis for the generated sample*

for an increasingly large sample of pseudo random numbers, drawn from a t-distribution with $\nu = 2$, the behaviour is clearly divergent.

2.3.2 Usual Moment Estimates

The most usual estimates for the central moments are for a series of T observations, $x = [x_1, \ x_2, \ , ..., x_T]'$

$$M_r = \frac{1}{T-1} \sum_{i=1}^{T} (x_i - \mu)^r$$

[4]See the section about Data set 2 in chapter 3

(2.13)

where μ is the sample mean. In the case of covariance of variables in a random vector with dependent variables, $X = [X_1,\ X_2,\ ,...,X_N]$ with T rows of observations, $x_i = [x_{i,1},\ x_{i,2},\ ,...,x_{i,N}]$, The most common estimate [Stock and Watson, 2008], can be written on matrix form:

$$Cov(X) = \frac{1}{T-1}\sum_{i=1}^{T}(x_i-\mu)'(x_i-\mu)$$

where μ in this case is the sample mean vector.

An exponential rolling window

When dealing with time series there is a very reasonable assumption that recent observations play a larger role than the ones made ten years ago. For a chosen time constant say α, corresponding to a certain number of days, we construct the multiplier $a = 0.5^{\frac{1}{\alpha}}$. For a time series with returns denoted dX_t, the estimates of the covariance matrix and the mean vector at time t is then:

$$\mu(dX_t) = a\mu(dX_{t-1}) + (1-a)dX_t$$
$$Cov(dX_t) = aCov(dX_{t-1}) + (1-a)(dX_t-\mu)'(dX_t-\mu)$$

A quick look at the formula reveals that in this frame, after an additional time period of α days, half of the information of $Cov(dX_t)$ is still there, and after another α days only a fourth and so on.

To have this rolling frame is very convenient, since it is very computationally effective, every new day we only have to care about the new observation and not the whole sample, by simply updating the old matrix. As well as the estimate will contain information from the entire sample[5].

2.3.3 A Robust Estimate of Volatility

In my portfolio optimizations I will estimate the correlation separately from the volatility. With the access to Open, High, Low and Close prices there is a volatility estimate developed by Yang and Zhang [2000] that are 8 times more efficient than estimates of the type described in (2.13). At day t, let O_t, H_t, L_t and C_t denote the prices observed. Then the estimated (daily) volatility based on n days is

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=t-n}^{t}\left[\left(ln\frac{O_i}{O_{i-1}}\right)^2 + \frac{1}{2}\left(ln\frac{H_i}{L_i}\right)^2 - (2ln2-1)\left(ln\frac{C_i}{O_i}\right)^2\right]}$$

[5](The earliest observations will however play an insignificant role due to the exponential decay)

Off course we can as well use it in an exponential rolling framework just by using that

$$b_t^2 = \left(ln\frac{O_t}{O_{t-1}}\right)^2 + \frac{1}{2}\left(ln\frac{H_t}{L_t}\right)^2 - (2ln2-1)\left(ln\frac{C_t}{O_t}\right)^2$$

b_t^2 is the representation of $(C_t)^2$, the terms that build up the second order moment estimate in the simplest case, but now with some more parameters. Then let the estimated variance at time t, be: $\sigma_t^2 = a\sigma_{t-1}^2 + (1-a)b_t^2$.

An important note to make is that this estimator is drift independent[6]. When the drift is of the same magnitude as the volatility this estimator will tend to overestimate the volatility.

2.4 Risk Measures

Risk measures are of great concern when it comes to asset management. To be able to measure and estimate risk is important as well as a part of measuring performance and as well how large risks a company is exposed to. It is far from trivial which methods that are superior.

In this paper I will only mention risk measures that can be derived from based on historical data and distribution. Otherwise there are a lot of other types of risks than the actual risk of the drop in value of an asset. In Litterman [2003] such other risks as Credit risk and Liquidity risk are described an discussed, but it will not be of concern of this paper.

2.4.1 Volatility

Volatility is probably the most frequently used measure of risk. Volatility is a measure of the magnitude with which asset prices change [Litterman, 2003]. The most common estimator is the standard deviation of asset returns. It can be considered a risk measure since it measures the average deviation from the mean. The higher the deviation is, the larger the risk is, however it is a really bad measure when it comes to give a fair picture of the extreme risks e.g. the 5% largest losses. Since it practically assumes a normal distribution it underestimates these kind of risks. The frequent use in practice is due to that it is so easy to use as well as almost every one knows it.

2.4.2 Value at Risk and Expected Shortfall

Two other risk measures are Value at Risk (VaR) and Expected Shortfall[7] (ES). They are risk measures that focuses on extreme events by measuring

[6]That means that it ignores trends, or that μ is assumed to be zero

[7]Even known as Conditional value at risk (CVaR)

tail risk. These risk measures are always computed on a certain level, say q. The notation will then be VaR_q and ES_q. Hult et al. [2011] gives a great tutorial on different approaches for computing these risk measures.

In words if for example $q = 0.05$, $VaR_{0.05}$ is the best outcome of the 5% worst outcomes. Or as an other example, if $VaR_{0.05} = 10000$SEK it means that there is a 5% probability that losses will not be 10000SEK or greater. ES is instead the expected value of the 5% worst outcomes. If we assume that the returns in question, can be seen as observations of a RV, X, that have the PDF $f(x)$. Then VaR_q and ES_q could be expressed as

$$VaR_q(X) = -F^{-1}(q) \quad ES_q = \frac{1}{q}\int_{-\infty}^{q} VaR_t(X)dt$$

If estimated in an empirical sample the simplest way possible, $VaR_{0.05}$ is the 5^{th}% largest loss observed and ES the average of the 5% largest losses.

These measures are not without criticism, since it is even here hard to account for the risks of very rare events. If we for example would like to estimate $ES_{0.0001}$ we would face some serious problems problems. Think about how reliable would data actually be on events as rare as one in 10000, there would most certainly not be any observations reflecting such an event, however there are approaches such as extreme value theory, see Hult et al. [2011]. VaR and ES are however better measures of risk in many more aspects than volatility.

2.4.3 Draw-Downs

A draw-down is the difference between the last maximum of cumulative returns and the value at the time in question. Draw-downs are often used as a measure when models are tested to see that they perform well over time. If there are some periods where draw-downs are large, the model should be handled with care since it then show tendencies of being able to make large losses in certain cases.

2.5 Futures markets

Since Lynx are trading with different models on futures markets I just want to mention a couple of things about futures markets, without digging too deep into technical details. A futures contract is very similar to a forward contract but with certain terms and conditions. Let's start there.

A forward contract is a binding contract between a buyer (who is said to have a long position) and a seller (who has a short position) that states that at a certain date in the future, the buyer will buy a certain quantity of the underlying asset, from the seller for a certain price called the forward price. These contracts are generally not traded on regulated exchanges, but are custom made products traded "Over The Counter"[8].

As well as when it comes to forward contracts, a futures contract also consist of the obligation to sell a certain asset at a certain date at a certain price (the futures price). The futures price is as well as the spot price of the underlying asset determined by the laws of supply and demand. The futures price is simply the markets expectation of the spot price of the underlying asset at the delivery date. One of the major differences between the contracts, is that futures contracts are traded at regulated exchanges, for example in the USA there is the Commodity Futures Trading Commission that has been active since 1974. Futures are very important hedging instruments in many industries, and it has great economical value if the trading of these contracts are regulated by a reliable institution according to Hull [2009]. The other major difference is the marginal account, which gives a daily settlement of the contract. To enter into a futures contract you have to have a marginal account with a certain amount of money on it, specified by policies of the exchange. If the futures price at closing time day t is denoted F_t then at the end of the day, the amount $F_t - F_{t-1}$ shall be added to the account. This daily settlement ensures that there is always enough money to close out the contract. The marginal account will also pay a fair interest rate to investors.

The contracts are usually issued a couple of years before delivery. It is however common with delivery every three months, if the underlying asset is not strictly related to some seasonal effects, as it could be with certain agricultural products. The contract with shortest time to delivery is in most cases the contract with highest trading activity.

2.5.1 History of futures

Futures markets has changed dramatically over the last decades. When looking at the amount of contracts traded, and which types of contracts that are traded, you get the feeling that it has walked away from hedging purposes, to be dominated by speculative trading. From about 1980 to 2003 the number of contracts traded on US exchanges increased with about a factor 10 [Kolb and Overdahl, 2006]. According to CTFC annual reports, the CFTC agricultural products, in the seventies, made up the majority of contracts traded on the CTFC and there was a small amount of precious metal-contracts traded. In 1980, still about $\frac{3}{5}$ of the volume was agricultural, financial instruments and currencies now representing a fourth of the volume.

[8]OTC products are those that are not traded at regulated exchanges. For example they can be issued by a bank, or be tailored for costumers with certain needs.

In 1990 we had a dramatic change, where futures on financial instruments alone took up half the trading volume. Today it is even more, about $\frac{3}{4}$ of the trades on CFTC are futures on financial instruments, and agricultural products have a about a tenth of the volume. I mention this to make the reader aware that it might be a good idea to use caution when building models based on old data, since the market has changed a lot over time.

2.5.2 Why futures?

The futures contracts are very appealing investment alternatives to investing in the underlying assets. First of all we can easily build a portfolio of different asset classes traded at the same kind of exchanges. We can have interest rate futures, commodity futures, equity futures and currency futures and we will basically be dealing with the same type of asset. It is as well really nice with futures, since there is always a long and a short position in a contract, therefore there will (almost) never arise a problem with, for example, short selling equity if you believe in a stock market crash. At last it is really convenient since the money is never bound in the asset in question, it can rest safely at a bank account, earning interest. Since the marginal account yields interest, the returns of a futures contract can actually therefore be treated as excess returns. Excess returns are usually referred to as the returns that exceed the risk free interest rate. This is nice since many portfolio models use excess returns in their frameworks.

Rolling into a new contract

When speaking about rolling into a new contract, it is meant to close out the old contract and take corresponding position in a new one of the same type. Since contracts are daily settled this is not a problem and the return series will be as the one of a single asset, instead of numerous contracts. This task is typically performed about two weeks before delivery and is done automatically when downloading the time series. I will mention more about the databases in the next chapter.

2.6 Portfolio optimization

The modern portfolio theory can roughly be said to have been founded by Harry Markowitz in the 50's, with his work on min-variance optimization. His findings were later in the 60's developed by people for example, William F. Sharpe and Jan Mossin to the famous Capital Asset Pricing Model, CAPM[Litterman, 2003].

What CAPM provides, is how we can view the equity market in equilibrium if we have rational investors, with the same information, using the same min-variance frameworks and as well under the assumption that we have an efficient market. Not that very realistic, but it gives us a reference to about how we can think of investments. The idea with the model is that riskier assets will have greater expected returns, but to the price of a higher risk. In an n-dimensional asset space, for the risk premium of asset i, $\mu_{RP,i}$ the model says [Litterman, 2003]:

$$\mu_{RP,i} = \frac{Cov(dX_i, dX_m)}{Var(dX_i)}\mu_m = \beta_i\mu_{RP,m}$$

Where dX_i and dX_m are asset and market returns respectively and $\mu_{RP,m}$ is the market risk premium. The risk premium, or equilibrium excess return, is what you can think of as the award for choosing a risky asset. It will have a better pay off than a non-risky asset, otherwise you would not choose to invest in it. A lot can be said about estimating the market risk premium, refinements of the CAPM etc. but it is a little bit out of topic in this report, since it will be dealing with futures markets. To mention the CAPM is however important since it reflects the views of many actors in the financial industry.

2.6.1 Quadratic optimization

If we make the (unrealistic but comfortable) assumption that asset returns can be well described by a multivariate normal distribution, that is, only characterized by its first two moments. First let the row vector, μ, denote the expected returns for N assets and Σ their covariance matrix. If we let the $1 \times N$-vector, w, denote the weights an investor put in each asset and make another assumption, namely that the utility of an investor can be described by the function

$$\mu w - \frac{1}{2}\lambda w\Sigma w'$$

then we have the set-up for the simplest and maybe most famous portfolio optimization problem. We want to maximize this utility. The term μw is the expected portfolio return and $\frac{1}{2}\lambda w\Sigma w'$ is the cost of risk, where λ is a risk awareness parameter and $w\Sigma w'$, the portfolio variance. Of course the holdings can not exceed the investors wealth and some times we have to add restrictions like $w \geq 0$, if we don't have the possibility to take short positions.

Since we will only be dealing with futures portfolios the problem above will not be of interest. Neither will constraints about short selling, efficient frontiers etc. be interesting. The only constraint that actually will matter is a risk target. At a fund, there is almost always a certain risk level that is not allowed to be exceeded. Usually this risk level is the portfolio variance, let the constraint be $w\Sigma w' = \sigma^2_{max}$. There is an equality sign because if $w\Sigma w' \leq \sigma^2_{max}$, then we could increase the risk taken in some asset and gain some more expected return. Let index t, for a parameter denote the estimate of that parameter at day t. Then our optimal weights at day t, will be decided by

$$\max_{w_t} \ \mu_{t-1} w_t'$$
$$s.t. \quad w_t \Sigma_{t-1} w_t' = \sigma_{max}^2$$

If we think about it this is exactly the same problem as

$$\min_{w_t} \ \frac{1}{2} w_t \Sigma_{t-1} w_t'$$
$$s.t. \quad \mu_{t-1} w_t' = k$$

(2.14)

for some constant k. This is easy to realize that it's just a matter of scaling since

$$\min_{w_t} \ \frac{1}{2} \frac{w_t}{k} \Sigma_{t-1} \frac{w_t'}{k}$$
$$s.t. \quad \mu_{t-1} \frac{w_t'}{k} = 1$$

always yield the same solution for $\frac{w_t}{k}$. We just have to chose k so that the optimal weights yield $w_t \Sigma_{t-1} w_t = \sigma_{max}^2$. By using Lagranges method, see Sasane and Svanberg [2009], it is easy to solve problems on the form

$$\min_{w} \ \frac{1}{2} w' \Sigma w + c' w + c_0$$
$$s.t. \quad Aw = b$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is symmetric, $c \in \mathbb{R}^n$, $c_0 \in \mathbb{R}$, $A \in \mathbb{R}^{M \times n}$, and $b \in \mathbb{R}^m$. The system

$$\begin{bmatrix} \Sigma & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

where $\lambda \in \mathbb{R}^m$ is a vector of Lagrange-multipliers, then gives the solution of w and λ. In our case this means that

$$\begin{bmatrix} \Sigma_{t-1} & mu_{t-1}' \\ mu_{t-1} & 0 \end{bmatrix} \begin{bmatrix} w_t' \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ k \end{bmatrix}$$

We can then solve for w_t,

$$\begin{bmatrix} w_t' \\ \lambda \end{bmatrix} = \begin{bmatrix} \Sigma_{t-1} & mu_{t-1}' \\ mu_{t-1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ k \end{bmatrix}$$

(2.15)

and choose k so that $w_t \Sigma_{R,t-1} w_t' = \sigma^2_{max}$.

Chapter 3

Background, problem formulation and data

In general, asset returns that have a normal distribution are rare, if not totally absent as supported by for example Fama [1965] and Blattberg and Gonedes [1974]. During some periods normal behaviour might be observed, but in the long run the market will experience chocks, periods of high volatility and periods where things behave in unexpected ways. Due to the non-normal properties of return series, and the generally normal oriented optimization frameworks that are popular and easy to use, it is interesting to see if, or how one can incorporate tail dependencies and higher order moments in these models for enhanced performance.

As well it is important to stay close to the frameworks that are used in practice at Lynx. That's why the current data sets are chosen and as well why some of the frameworks are used. I will also emphasize that the evaluation of the results will be focused on risk measures and not returns. That's because return prediction is a more fuzzy business and as well as it is the crucial part where you can make the money. To be able to make a report where everything can be publicized a more general focus towards risk and traditional models will be used.

3.1 The Data

The data consist of two different data sets, retrieved from a tick-database. Both sets are containing daily open, high, low and close prices. Since the assets studied are in fact contracts, that are rolled over to new ones approximately every third month, the raw data would consist of information regarding different contracts on the same underlying asset. When extracted from the database, to get data that is useful without too much time consuming manipulation, there are functions that extract the data, automatically rolled into new contracts. The data extracted will then be based on the return series of futures prices rather than actual futures prices.

3.1.1 Dataset 1

During the first part of this project, "Good times and bad". I worked with a set based on arithmetic return series for 52 futures contracts, divided into four different asset classes, equity, bonds, commodities and exchange rates. The data ranges from October 1999 until today. This was the data set given to me as I first came to the company, since it resembles the data they usually are dealing with. There are some issues with this data set. Since it is based on arithmetic return series, the prices observed, are not the real futures prices but cumulative sums of the returns, which does not allow for the construction of log-return series[1]. Volume data were not available in this dataset. There are two good reasons for using this dataset. One: Pretty much the same dataset is used in actual models. Two: For convenience, so that I would not have to spend a lot of time to extract new data that would resemble the set that I already got. See Appendix A for a complete list of the contracts.

3.1.2 Dataset 2

The second dataset, I extracted my self from the tick-database when I realized that I needed longer return series if I were about to test my ideas with higher order moments. There is much less data on equity class futures than for example commodities, in which producers and buyers have a long tradition of using the contracts as hedging instruments. Speculative trading in futures started to grow first around the 80' and early 90's as mentioned. I ended up with a set of 16 contracts ranging from 1993 until today. Luckily I managed to download the data in terms of relative returns, enabling log-returns. See Appendix A for a complete list of the contracts.

3.2 Good Times and Bad

On the topic of tail correlation, there are many approaches and theories that are interesting and might be useful. For example I first looked into the area of copulas, especially the t-copula. By using a copula structure one can much better take into account properties as tail correlation, in comparison to a normal framework[2]. On the other hand, to use the copula for optimization

[1] Some assets even have negative prices in the beginning of the arithmetic time series due to the loss of compounding effects.

[2] For the reader interested in copulas see Frees and Valdez [1996] for a tutorial

does not allow (in most cases) for standard optimization routines, which is my aim to use. The methods would be really computational intensive, especially for the higher dimensional portfolios we will be dealing with here.

A more "primitive" approach would be to use two or more different covariance matrices, one for the tails and one for the central observations. Chow et al. [1999] chooses to make this approach. The main idea is that in times of market turbulence we will observe multivariate outliers in our return data. These effects might be the cause of individual assets experiencing chocks or the combined effects of many assets responding simultaneously to certain events. What the authors claim is that in these turbulent, or as the title of the article suggest, bad times, we will observe a different dependence structure with higher correlations and higher volatilities. In the empirical part, their asset space is based on a portfolio with eight different asset classes, that can further be divided into equities, bonds and commodities. Monthly returns are used and the data spans from January 1988 to September 1998, giving 129 observation dates.

It is very interesting to see if the method described in the article is a good approach when it comes to a higher dimensional portfolio of futures contracts with daily returns. In the framework I will be using, there are robust short term estimates of the volatility, therefore I will only be interested in looking at the ability of a "good/bad times-model" in means of prediction of the correlation structure.

3.2.1 Problem Statement 1

A number of questions arise:

- Is there a good measure for the predictive ability of the correlation structure?
- Is the "Good and Bad times" idea a good approach when it comes to daily returns in a high dimensional futures portfolio?
- Does the tails of the distribution actually contain valuable information, in terms of tail correlation, that can be used to improve prediction of the correlation structure of outliers?

3.3 Valuation of Higher Order Moments

Many sources suggest that higher order moments of asset returns will affect portfolio behaviour. For example, Harvey et al. [2004] discuss the shortcomings of the traditional Markowitz framework and compares it to a model where the skew-normal distribution is used for optimization. They show that their method yield higher expected utility than the models generally used in practice. Two other articles dealing with skewness are Chunhachinda et al. and Sun and Yan [2003b]. They both use the same polynomial goal programming framework, applied on weekly and monthly stock market returns, concluding that the addition of skewness in the optimization framework creates major changes in the portfolio construction. Their results indicate that investors trade expected return for skewness.

Not only the skewness is of interest when studying asset returns. The heaviness of the tails are also of general concern when it comes to the construction of a portfolio. Christie-David and Chaudhry [2001] develop an optimization framework that practically is a four-moment extension of the CAPM, with market comoments. The results show an increase in explanatory power when it comes to explaining returns in futures markets. An other study Gioulekas and Djechiche [2009], emphasizes the importance of maximizing expected returns when minimizing the draw-downs rather than variance. T-copulas are used, and penalizing assets with heavy tails are done by scaling the standard deviation with $\frac{1}{\sqrt{\nu}}$, where, ν , represents the degrees of freedom parameter in a t-distribution.

The suggestions of approaches seem to be endless. Once again the copulas seem like a tempting approach when one would like to study multi variate distributions with non-normal behaviour. The ability to combine marginal distributions and complex dependence structures, comes with a cost of computational ineffectiveness. Is it still be possible to catch some of these effects and use them, even if we are working in a normal framework? There is also the issue of interpreting and estimating the higher order moments, Sun and Yan [2003a] argue that the usual measures of skewness and kurtosis are bad and should be replaced with more robust measures.

3.3.1 Problem Statement 2

Having studied the above articles, still with the ambition to use traditional optimization frameworks, it is quite appealing to value skewness in terms of expected returns, since both are odd moments. Kurtosis or tail thickness is appealing to value in terms of standard deviation, since they represent even moments. There are some certain problems that have to be studied further:

- Due to the instability of higher order moments, caused by their sensitivity to large outliers, are there better and more stable measures? E.g. degrees of freedom and θ_2.

- Is it possible to affect portfolio performance in a positive way, by only using the HOM in marginal distributions, ignoring higher order comoments?

- Is it reasonable to value skewness in terms of expected returns and how can it be expressed in a standard optimization framework?
- Is it reasonable to penalize the more heavy-tailed assets, in term of scaling their volatility in a normal framework?
- If implemented, can the valuing and penalizing have positive effects on portfolio returns, in means of risk, if the target is to reduce the one period draw-downs?

Chapter 4

Methods and Empirical Results

In this chapter I will present my work progress and my findings. As the main tool for data analysis I have been using Matlab. Matlab is a high-level programming language specialized for technical computing.

Some of the analysis will contain large correlation matrices. To display them I will use heat-maps. Each element in the matrix will be represented by a coloured square, making it easy to overview as well as it is easy to see how the different asset classes interact. See Figure 4.1 for an example.

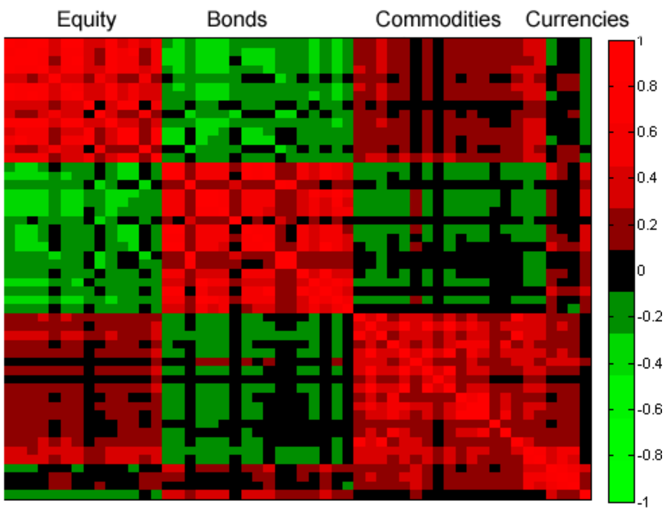

Figure 4.1: *Example of a heat map of a correlation matrix estimated on the whole sample of 3100 days and all the 52 assets in data set 1, with asset classes marked out. A bright red color implies strong correlation, black no correlation and green negative correlation*

4.1 The marginal distributions of data

In general the data tend to be almost t-distributed with some skew tendencies and some deviating tail behaviour. A typical histogram with a fitted t-distribution, on normalized absolute returns of schatz-futures (Short term German interest), is shown in Figure 4.2. The t-distribution fits rather well

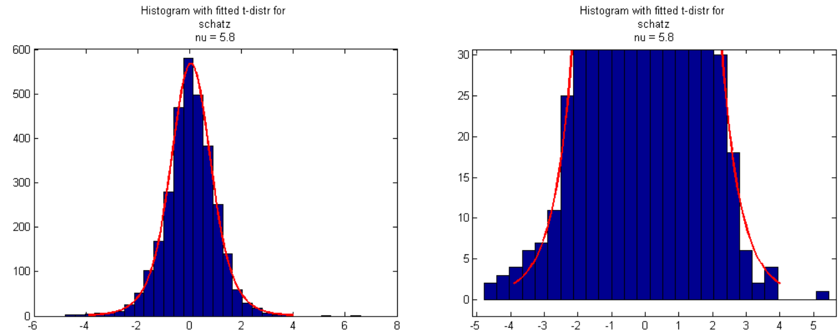

Figure 4.2: *Return series for schatz-futures with fitted t-distribution, $\nu = 5.8$. The histogram to the right is the same as the one to the left, but zoomed in*

and we can see that there are tendencies of negative skewness, the negative tail is a little bit thicker than the positive tail. How well a distribution fits the empirical distribution of data is easier to illustrate in QQ-plots. Take for example the return series of orange juice-futures in Figure 4.3 it is easy

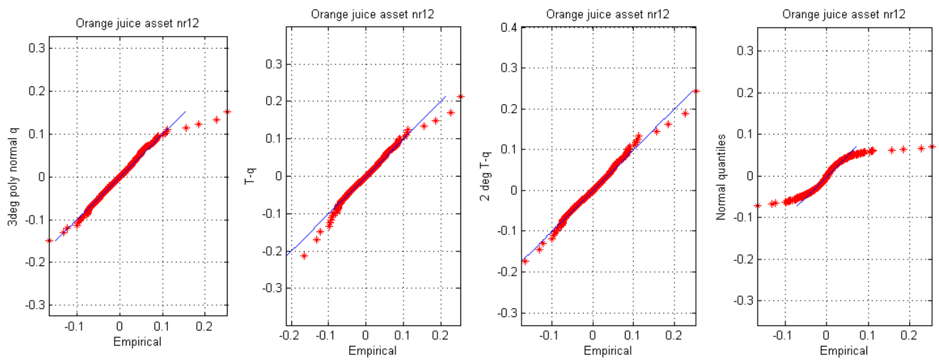

Figure 4.3: *QQ-plots for Orange Juice contract. From the left, Empirical quantiles vs: 3-degree polynomial of normal distribution, t-distribution, 2-deg polynomial of t-distribution and normal distribution.*

to see that the distribution is non-normal. As well it tends to be skewed compared to a t-distribution. The 2^{nd}-degree polynomial of t-distributions catches the tails pretty well. This kind of behaviour is consistent for the marginal distributions across the different assets. Some assets for example futures on 2-year US treasury notes, does not at all seem to be skewed as

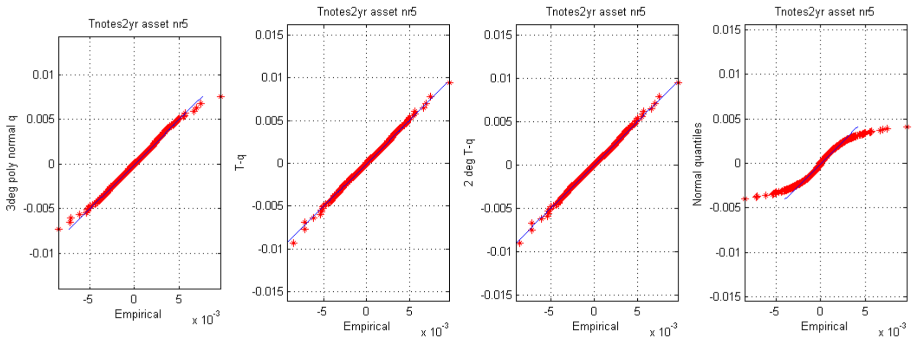

Figure 4.4: *QQ-plots for 2-year US treasury note contract. From the left, Empirical quantiles vs: 3-degree polynomial of normal distribution, t-distribution, 2-deg polynomial of t-distribution and normal distribution.*

can be seen in Figure 4.4.

I have of course studied QQ-plots for all asset return series, but since there are 68 of them it would not make sense displaying them all[1].

4.1.1 ML vs LS

When fitting distributions to data I usually feel most comfortable with using the maximum likelihood method since it gives the most likely estimates. When available, I use the built in methods in Matlab to estimate parameters, and they are always based on ML-estimates. Sometimes however, the methods collapse. When trying to fit the return data of the Eurodollar futures to a t-distribution, the ML-method encounters a problem with a non-invertible hessian, and the resulting QQ-plots are displayed in Figure 4.5. Instead I am forced to use the LS-estimate. One reason might be the large number of observations of returns of value 0, it is 603 out of the 4873 observations. The most probable reason for the many zeros is that the tick size[2] is large in comparison to the volatility of the contract. If we for example look at the parameter estimate ν for the assets in data set 2 and compare the LS and ML-estimates, we see that there are some inconsistencies.(Figure 4.6) Estimates seem to be able to differ up to one degree of freedom. This would only be alarming when it concerns estimates with low degrees of freedom. I will however use the ML-estimate as long as the the QQ-plots don't behave as in Figure 4.5.

[1]The reader who for any reason would be interested in that should e-mail me, and will then recieve a little video containing all the QQ-plots

[2]The tick size is the minimal amount the prize of a contract can be adjusted with, any bids must also be a multiple of this amount

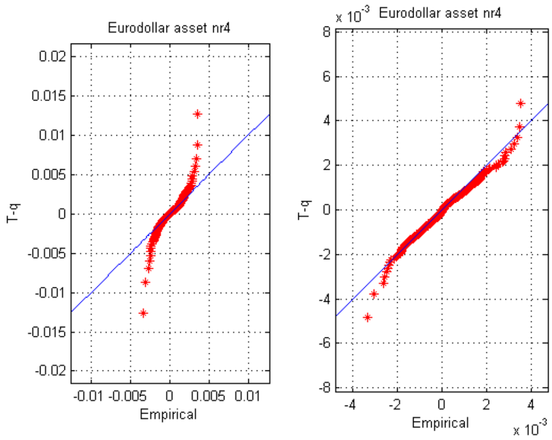

Figure 4.5: *QQ-plots for Eurodollar contract. From the left, Empirical quantiles vs: t-distribution ML-estimate and t-distribution LS-estimate. In this case the ML estimate has failed*

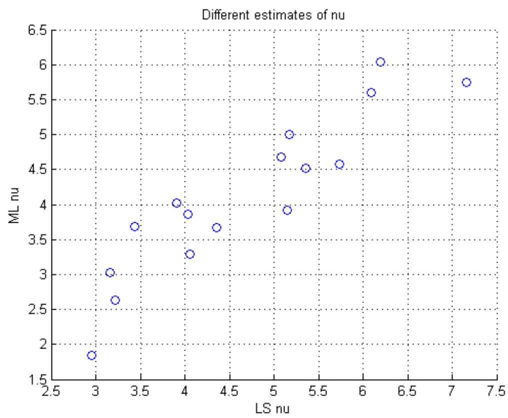

Figure 4.6: *Estimates of parameter ν for assets in data set 2, LS vs ML-estimates*

For the fit of the polynomial coefficient parameters in (2.2) I will simply use the least squares method for one simple reason; it is faster. It is very time consuming to make rolling estimates of the parameters for the whole time series and if I would use my laptop it would take too long time. As well, the eventual small precision gained in choice of method would be eaten up directly by the large uncertainty in the parameter estimates.

4.1.2 The effect of using the OHLC-volatility estimator

Another thing that is worth noticing, is that when we look at the distribution of an asset over the whole time series, as we have done in the section above, σ is estimated on the entire sample. In the portfolio models implemented later, the robust short term estimate of volatility will be used. What effect does it have on the degrees of freedom of the distribution?

Take a look at Figure 4.7, where the original estimates of ν are compared

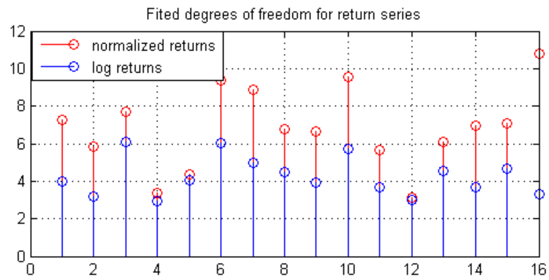

Figure 4.7: *Estimates of parameter ν for assets in data set 2, LS vs ML-estimates*

with estimates of ν, where the time series have been normalized with the daily short term estimates of volatility. For most assets the degrees of freedom drop with about 2 to 3. For three of the assets however there seem to be no effect. An interesting observation is that one of the differences between the time series that are not affected and those who are. The structure of the return series not affected, seem to not change namely if they are normalized. In Figure 4.8, we can see that the asset returns affected (SP500 futures in

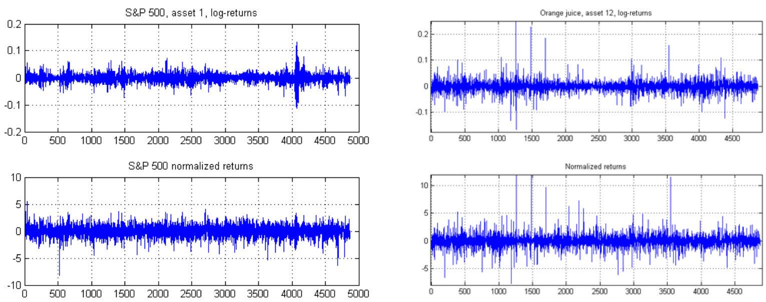

Figure 4.8: *To the left, log-returns and volatility-normalized returns of SP500 futures. To the right, log-returns and volatility-normalized returns of orange juice futures*

this example) have periods of high volatility that are smoothed out by the normalization. This observation is consistent for the other assets as well, but has to be investigated further if any conclusions should be drawn. Maybe it would make sense to divide assets into different classes based on these properties.

4.2 Good times and bad

With the aim of using using outliers to make better predictions of the correlation structure, it is natural to see how it changes over time. An exponential rolling window will be used, since the assumption that correlation is constant over time is clearly wrong. This becomes really obvious if you look at the development of the correlation matrix over the time period studied, there are clear distinctions between different periods. I made a short film showing the development over time, where periods of stronger and weaker correlation can easily be spotted, the difference is clearly visible. To get an illustration

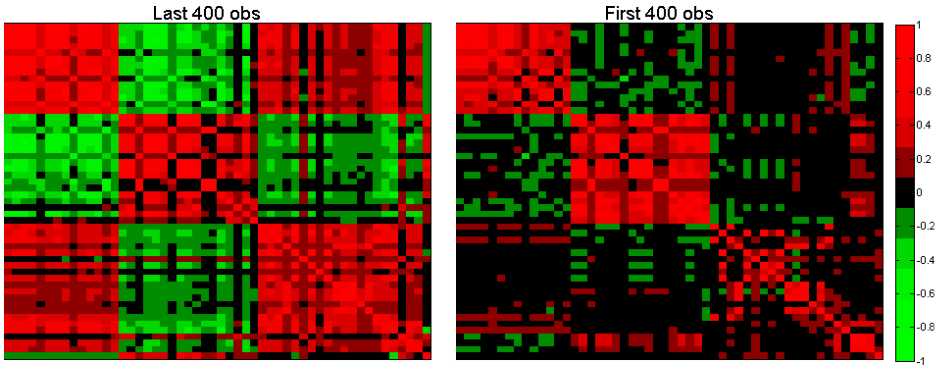

Figure 4.9: *Heatmaps of correlation matrices estimated on the first and last 400 days of the sample*

of this look at Figure 4.9 where you can see two correlation matrices, one estimated on the first 400 days and one on the last 400 days. Due to this kind of behaviour, and since it is commonly used in practice, I will use an exponentially rolling window.

4.2.1 Outliers and periods of turbulence

If we use the notation

$\Sigma_t =$ covariance matrix of the whole time series day t

$\Sigma_{G,t} =$ covariance matrix of the inside sample day t

$\Sigma_{B,t} =$ covariance matrix of the outlier sample day t

$p_{out} =$ probability for being in the outlier sample

$dX_t =$ row vector of returns for each asset at day t

$\lambda_G =$ inside risk aversion

$\lambda_B =$ outlier risk aversion, $\lambda_B + \lambda_G = 2$

$$\Sigma_{GB,t} = \lambda_G(1-p)\Sigma_{G,t} + \lambda_B p \Sigma_{B,t} = \quad \text{the blended covariance matrix}$$

Recall from the theory section that the measure, $d_t = dX_t \Sigma_t^{-1} dX_t'$, tells us about the distance from the mean of a normal distribution. In the exponential rolling framework that I use, the time constant for correlation estimates is chosen to be 130 days. To ensure that the possible change of magnitude of return sizes over time will not change the the d's can be normalized with an exponential rolling average of themselves. It could be an issue since we are dealing with absolute returns, and not log-returns. In Figure 4.10 there

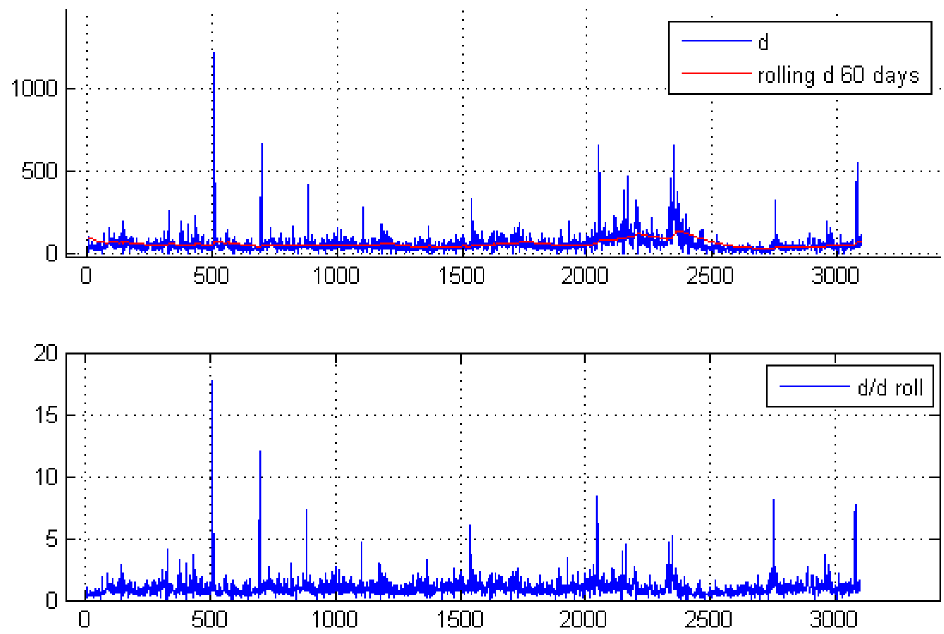

Figure 4.10: *The distance measure d, for the whole time series. The second graph displays the normalized series of d*

is a graph, displaying the d's for the entire time series.

We have to chose an outlier level, p_{out}. Since the asset space has 52 dimensions, it is important to choose a level where there are enough observations to get an estimate of the covariance matrix of the outliers. We have about 3100 observations of dX, choosing p of 2% will give us 62 observations for the estimation, which seems almost too few. I will try some different values between the top 2% and 25%.

The test period, and the covariance matrices

The last 1000 observations will serve as our test period, and the first 2101 observations will be used to get initial values of $\Sigma_{G,n}$, $\Sigma_{B,n}$ and Σ_n for the rolling estimates. To get to the initial values, I first chose Σ_1 to be the covariance matrix of the 400 first observations, $\Sigma_{G,1}$ to be the covariance matrix of the first 400 inside observations and $\Sigma_{B,1}$ to be the covariance matrix of all the outlier observations in the initial period (To ensure that there are enough observations for estimating $\Sigma_{B,1}$). Then I use a exponentially rolling window to update the matrices. The time constants are chosen to be $t = t_G = 130$ and $t_B = \frac{t_G}{2}$. t_B is chosen to be smaller since for small p, we will have to few outlier observations.

The Different Set-Ups

The measure to be studied will be the negative log-likelihood measure in (2.11), we look at the negative version of it because I feel more comfortable, looking for minima rather than maxima. I will look at a couple of different set-ups and compare their performance. I have chosen four different set-ups in terms of outlier awareness:

$$\begin{array}{llll} \lambda_G = \ 1 & \lambda_G = \ 1.5 & \lambda_G = \ 0.5 & \lambda_G = \ 2 \\ \lambda_B = \ 1 & \lambda_B = \ 0.5 & \lambda_B = \ 1.5 & \lambda_B = \ 0 \end{array} \tag{4.1}$$

I will try these set-ups for different values of p, and compare the average likelihood of the observations to the one produced by the use of Σ estimated on the whole sample. To compare the performance of the Good/Bad times-model with something, I will try different shrinkage approaches, for example shrinkage towards the identity matrix and shrinkage towards factor models, where eigenvectors are used.

4.2.2 The Most Likely Estimates, an Empirical Evaluation

First of all, to get a little bit more comfortable, lets take a look at the negative log-Likelihood function, derived in (2.11) versus d as a measure of identifying the most likely model from a already known distribution. From the first 1000 observations of daily returns I estimated a covariance matrix, which I used to generate a series of 1000 random normal data. With the known correlation matrix I looked at the two different measures for shrinkage towards the identity matrix, compared in Figure 4.11. As we can see, there is a less average multivariate distance from mean, the closer we get to a correlation structure, corresponding to the identity, but if we look at the likelihood instead, the unlikelyness of a event is penalized more. This leads to an increase, the further away we get from the true matrix. If we shrink the matrix towards a factor model based on the first three eigenvectors, we get a graph that looks practically the same but with about half of the deviation from the non-shrunken matrix. This observation does give some credibility to the ML-measure (Even if it does not prove anything).

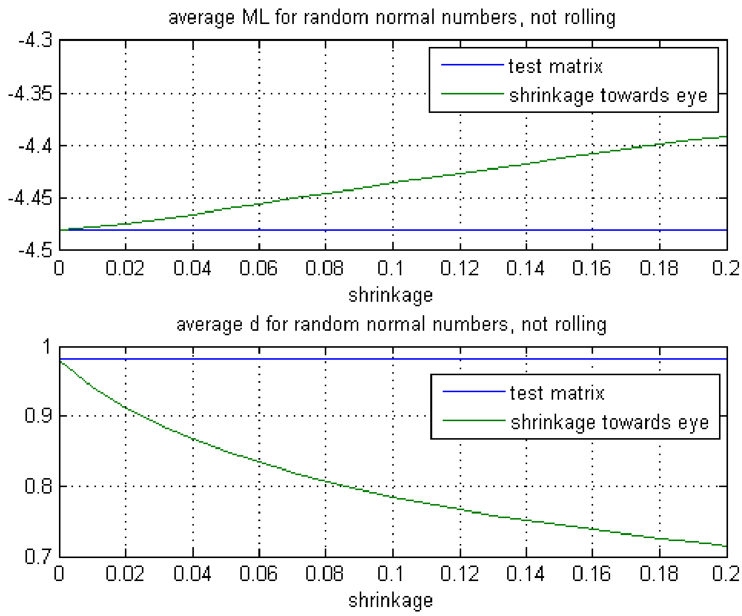

Figure 4.11: *Average negative log-Likelihood measure and average d, for different shrinkage intensities. "test matrix" indicates the original matrix without shrinkage, "shrinkage towards eye" indicates the effect on a correlation matrix shrunken towards the identity*

Weighing between the inside and outlier sample

In Figure 4.12 we can see the results for different levels of p for the different set-ups in (4.1). The inside aware, as well as the neutral set-up, show an increase in average likelihood for all values of p. The outlier aware set-up where the outlier matrix receives three times more awareness than the inside-estimated matrix, performs only better than the reference, for low values of p. The same applies to the only inside aware matrix which represents a scenario where we exclude outliers from our estimation of $\Sigma_{GB,n}$ this could have been a good approach if outliers do not behave in similar ways and hence only add noise to the estimates.

Shrinkage methods as a comparison

It is well known that more robust estimates of the covariance matrix can be made by the use of different shrinkage methods. Both in signal processing and finance this is used. A simple illustration of this is if I as in Figure 4.11, use the same inputs, except from that we use estimates of the covariance matrix, based on the sample of random numbers, instead of the known matrix. Figure 4.13 shows how a moderate shrinkage towards a factor model based on three eigenvectors, can be used to yield more likely estimates.

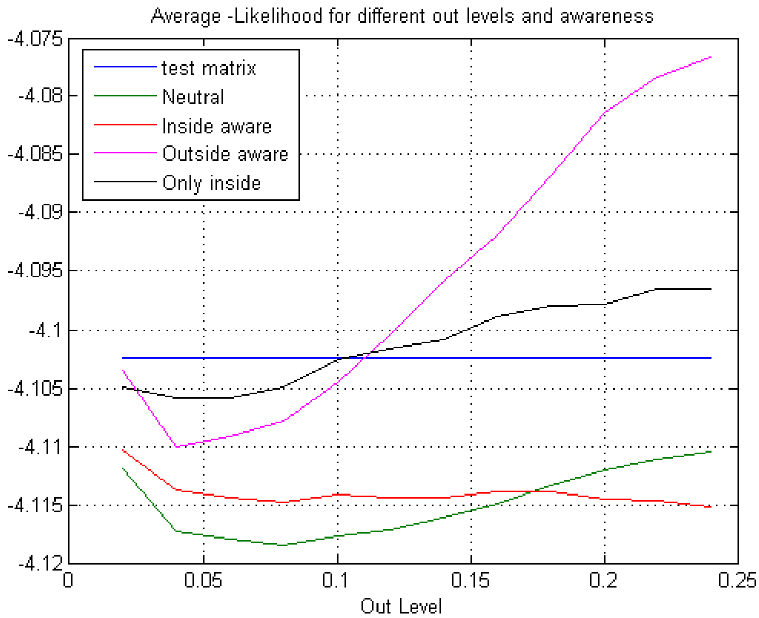

Figure 4.12: *Average negative log-Likelihood measure, for different levels of p (Out Level). "test matrix" indicates that Σ_n is used for day n and the others $\Sigma_{GB,n}$ based on the set-ups (4.1) in order*

This behaviour is as well something that is observable in our estimates on the real data. If we consider the test period, of the last 1000 days and look at the average likelihood of the observations, for different shrinkage constants towards ether, identity as correlation structure, or a 3-factor model of eigenvectors, we observe the same pattern. As can be seen in Figure 4.14 the average likelihood of the observations are better modelled with a shrunken Σ_n. Note that for the optimal shrinkages, in this context, the level of average likelihood is about the same as for the Good/Bad models. Can we in some way distinguish between the effect of the robustness in a shrinkage method and the possible event that the outlier sample contains valuable information? Or is the Good/Bad times-model in this case just another way of making more robust estimates?

Does the outlier aware models provide extra information?

If the increase in likelihood for the Good/Bad-models is an effect of extra information contained in the tail-matrix we should be able to spot a difference in the likelihood of the outlier observations. If we make a comparison of the

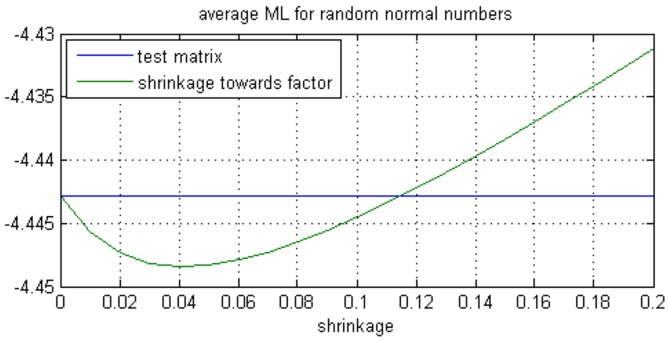

Figure 4.13: *Average negative log-Likelihood measure, for different shrinkage intensities. "test matrix" indicates the reference when shrinkage = 0, "shrinkage towards factor" refers to a eigenvector 3-factor model*

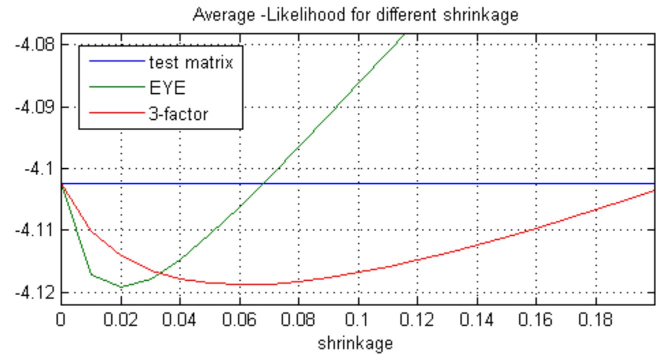

Figure 4.14: *Average negative log-Likelihood measure, for different shrinkage intensities. "test matrix" indicates the reference when shrinkage = 0, EYE is for identity as a shrinkage target*

reference matrix, estimated on the whole sample and the different set-ups, by simply, day for day, looking at the difference between the negative likelihood measure. In Figure 4.15 and Figure 4.16, these daily differences are displayed. Just by observing these graphs we can get many answers. In the first period of the test period, we see that the outlier aware set-up performs better than the neutral one, when having a better likelihood for the observed outliers is what is valued. On the other hand in the last period it performs worse on outliers, implied by the negative difference. This is not what we would expect if the outlier matrix estimates would carry information about other outliers.

If we instead look at the two different shrinkage set-ups in Figure 4.16, not only do we get larger differences for the outliers, but it is consistent over the whole sample. Also note that for the largest outliers, a strong shrinkage towards the identity gives the most likely results when it comes to predict

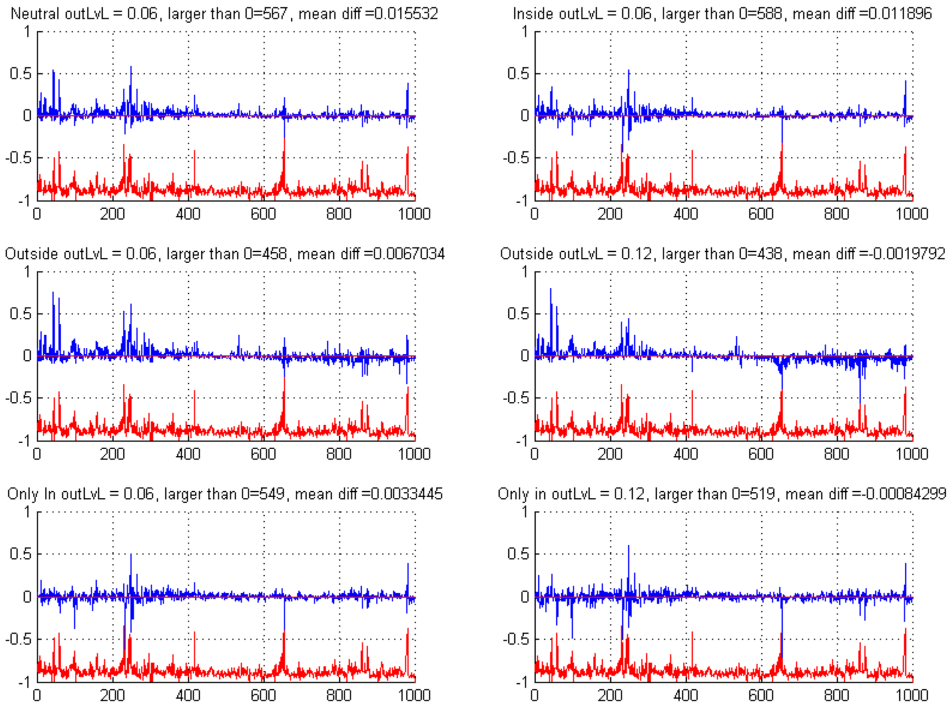

Figure 4.15: *The blue graphs shows the difference in the negative likelihood between the reference set-up and different kinds of awareness and values of p (OutLvL). The red ones displays* $\frac{d}{max(d)} - 1$ *to make it easy to see which observations that are outliers. The set-ups corresponds in order: Neutral, Inside, Outside and Only in, to the ones in (4.1). Mean diff, means the average difference from the reference and larger than 0 means the number of observations where the Good/Bad times-model outperforms the reference model.*

the correlation structure of outliers.

One way to interpret this is that in a high-dimensional futures portfolio with different asset classes, future outliers will often be of a different character than earlier observed ones. That is not such a stupid assumption, since even if stock-markets tend to crash together, that must not be true for commodities or other asset classes. Especially since there are so many different reasons to why outliers may occur, it only feels natural that all the outliers would not have a similar correlation structure. The conclusion is however, that using the kind of models where linear tail-correlation is used, is not a good approach when it comes to make outlier aware covariance matrix estimates.

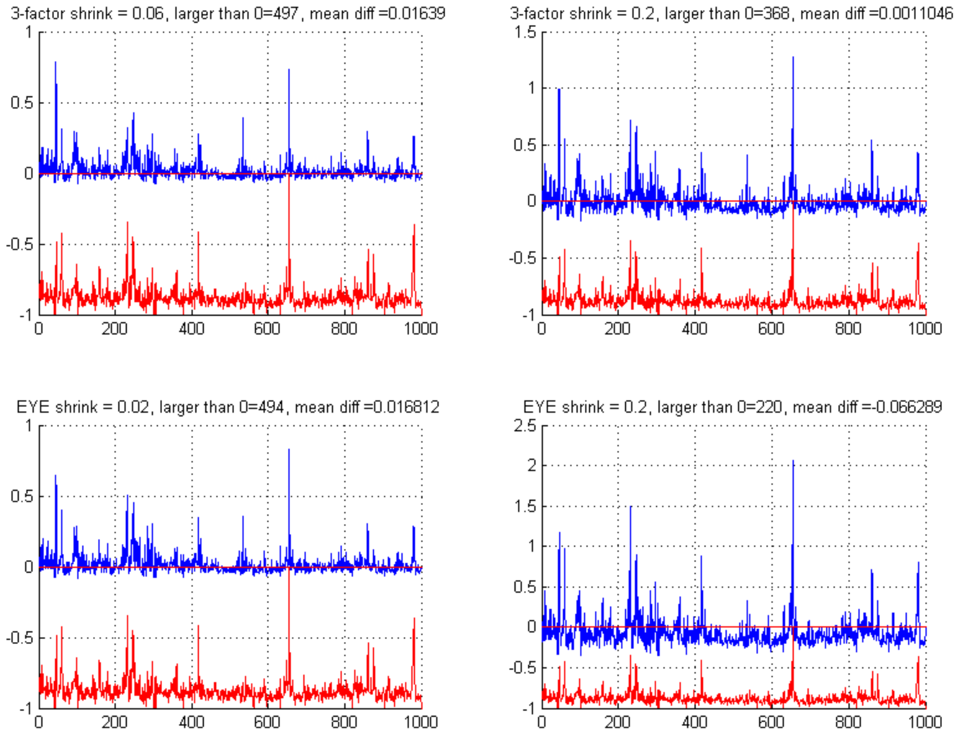

Figure 4.16: *The blue graphs shows the difference in the negative likelihood between the reference set-up and different shrinkage towards a 3-factor model or towards the identity matrix. The red ones displays* $\frac{d}{max(d)} - 1$ *to make it easy to see which observations that are outliers. "Mean diff", means the average difference from the reference and larger than 0 means the number of observations where the Good/Bad times-model outperforms the reference model.*

4.3 Valuation of higher order moments

The whole idea with using higher order moments is that they might say very much about an assets general characteristics. For example, assets with positive skewness are typically more likely to have large positive outcomes, rather than large negative outcomes. Imagine some kind of commodity where single political events or accidents might make prices rise quickly, like natural gas. There are few suppliers on the market and a wide range of consumers with an approximately constant demand. If the Russians for example would make up some new policies, making their gas less available or if a major pipeline would break, positive shocks in natural gas prices would be observed. On the other hand, there are not as many likely scenarios that would make prices go down very rapidly[3]. If that is true, natural gas would be an asset with positive skewness. On the contrary, take a stock index as an example. A market crash, is typically larger and more rapid than a market boom, a typical negatively skewed asset.

To be able to characterize the asset it is important to use a long timespan. Observe that this is an assumption that long term historical behaviour concerning these measures, can be used to characterize the future behaviour of the asset. For this purpose it is good if the time period chosen at least covers a normal economical business cycle. According to Baxter and King [1999] a normal business cycle is less than 8 years. However there is a conflict between having a long test period and a long initial period to estimate initial values of parameters.

4.3.1 Empirical evaluation of the skewness parameter

If we have a polynomial of degree 2 of inverse t-distributions, according to (2.2), fitted to the return series of data set 2. Observe that the skewness parameter, θ_2 will need to be scaled with the standard deviation of the return series, σ, to be easy to interpret in term of skewness. The term $F^{-1}(LvL_i)^2$ will have the same standard deviation as the data squared. It follows from that, $F^{-1}(LvL_i)$, will have standard deviation σ and for a RV, X, it holds that $Var(X^2) = 2E(X)^2 Var(X) + Var(X)^2$ and since $E(X)$ is usually small with respect to $Var(X)$ when dealing with financial data, $E(X)^2$ is even much smaller so $Std(X^2) \approx \sigma^2$. It means that the original θ_2 will be of order σ^{-1}. To make it "dimensionless" as the skewness is, we have to scale it with σ. In Figure 4.17, I plot skewness against $\theta_2\sigma$ and the relationship seems to be linear. Bearing in mind the robustness problems with empirical skewness estimates, the fit is actually surprisingly linear looking.

For an illustration of the robustness of the parameter compared to empirical estimates we can take a look at the return series of natural gas futures. In the beginning of 2003 there is a huge outlier corresponding to 16 standard deviations[4]. If we set up a rolling frame where 700 days at the time are used to estimate skewness and θ_2 we can see in Figure 4.18, that the impact of the outlier, greatly affects both estimates. When the observation however is not included any more we see that the effect on the empirical skewness is about three times as great.

[3] At least not in my imagination at the moment, only to be able to support my point, see it as a thought experiment if you don't find it reasonable.

[4] According to people I spoke with at Lynx, it was the result of some kind of mistake, and the price is not representative for the actual price that trading day

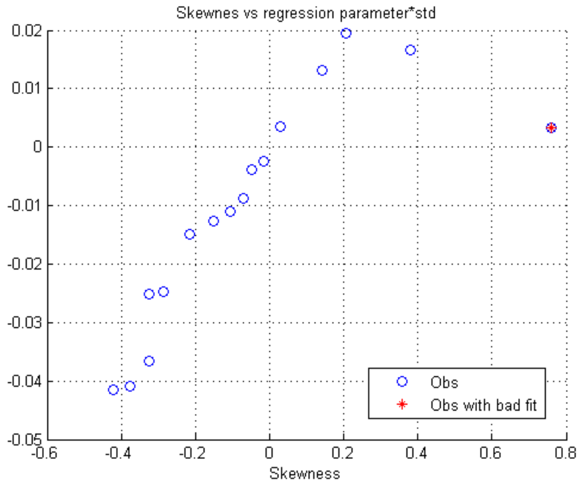

Figure 4.17: *Empirical skewness vs. $\theta_2\sigma$, for the 16 different assets in dataset 2*

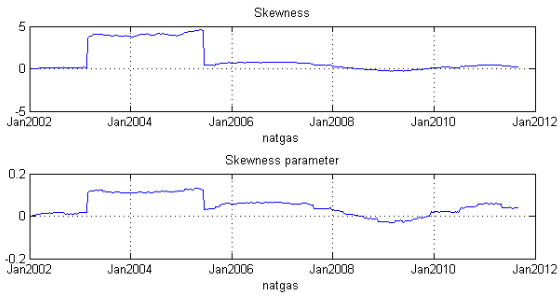

Figure 4.18: *Empirical skewness and estimates of $\theta_2\sigma$, for the natural gas contract*

4.3.2 Valuation of skewness

That investors value skewness is well established, but how is it valued? I will here try to make an approach to value it in terms of mean returns. When managing a portfolio, it is natural to have as an objective to minimize draw-downs. To minimize draw-downs for one period, you want to minimize things like VaR or ES.

Let's make a thought experiment where we have a portfolio with N assets, that have constant correlation among assets, two assets A_1 and A_2 such that Skewness(A_1) > Skewness(A_2), that are equal in aspects of mean returns, $\mu = [\mu_1\ \mu_2\ \mu_3\ ...]$ and volatility. When constructing an unconstrained mean-variance portfolio, with optimal solution $w_{before} = [w_1 \quad w_2 \quad w_3 \quad ...]$, these assets will have the same weights, $w_1 = w_2$. In terms of VaR or ES, the portfolio is not optimal. By changing the weights of the portfolio by choosing d_w, s.t. $w_{after} = [(w_1 + d_w) \quad (w_2 - d_w) \quad w_3 \quad ...]$ minimizes VaR or ES, we have reached a more optimal portfolio when minimizing the one period draw-down is the target. Note that changing the weights of A_1 and A_2 in this way will not change the expected return of the portfolio since $\mu_1 = \mu_2$. Let us now re-value the expected returns by choosing d_μ s.t. $\mu_{revalued} = [(\mu_1 + d_\mu) \quad (\mu_2 - d_\mu) \quad \mu_3 \ ...]$, inserted in the mean-variance optimization will give the weights w_{after}. We have now constructed a new portfolio, more optimal in a matter of minimizing single period draw-downs, by re-valuing skewness in terms of expected returns and without changing the expected returns for the portfolio. The d_μ, is then in some terms the value of the difference in skewness between the two different assets.

Simulation of skewed distributions

The question is, how will the relation between d_μ and Skewness(A_2) − Skewness(A_1) look? To get a picture of if, I will conduct a simulation study. What I need is some kind of multivariate distributions where I can vary higher order moments and keep the covariance matrix and mean constant. By using type of multivariate distributions described in the theory section, this can be done. Another appealing approach is to use skewed t-copulas. But to get explicit expressions for mean and covariance it would require marginal distributions with the same degrees of freedom.

For the set-up of the simulations I will use generated data series corresponding to five simulated assets where two are dominating the others in terms of mean returns. One set-up will consist of a choice of desired mean returns, covariance matrix, the skewness parameters, γ_i, and the degrees of freedom for each one of the univariate distributions that the simulation starts with. The desired vector of mean returns for all set-ups will be

$$\mu_{wanted} = [0.05 \quad 0.01 \quad 0.01 \quad 0.01 \quad 0.05]$$

I will as well use four different covariance matrices:

$$\Sigma_{wanted,1} = \begin{bmatrix} 1.0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1.0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1.0 \end{bmatrix}$$

$$\Sigma_{wanted,2} = \begin{bmatrix} 1.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1.0 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1.0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1.0 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1.0 \end{bmatrix}$$

$$\Sigma_{wanted,3} = \begin{bmatrix} 1.0 & 0.9 & 0.9 & 0.9 & 0.9 \\ 0.9 & 1.0 & 0.9 & 0.9 & 0.9 \\ 0.9 & 0.9 & 1.0 & 0.9 & 0.9 \\ 0.9 & 0.9 & 0.9 & 1.0 & 0.9 \\ 0.9 & 0.9 & 0.9 & 0.9 & 1.0 \end{bmatrix}$$

$$\Sigma_{wanted,4} = \begin{bmatrix} 1.0 & 0.5 & 0 & -0.2 & -0.2 \\ 0.5 & 1.0 & 0 & -0.2 & -0.2 \\ 0 & 0 & 1.0 & 0 & 0 \\ -0.2 & -0.2 & 0 & 1.0 & 0.5 \\ -0.2 & -0.2 & 0 & 0.5 & 1.0 \end{bmatrix}$$

For all of the different covariance matrices, I will use five different combinations of degrees of freedom corresponding to the following excess kurtosis

$$K_1 = [5.0\ 0.1\ 0.1\ 0.1\ 5.0]$$

$$K_2 = [0.1\ 0.1\ 0.1\ 5.0\ 5.0]$$

$$K_3 = [5.0\ 5.0\ 0.1\ 0.1\ 0.1]$$

$$K_4 = [2.0\ 2.0\ 2.0\ 2.0\ 2.0]$$

$$K_5 = [0.1\ 0.1\ 0.1\ 0.1\ 0.1]$$

Which gives us 20 different set-ups. The chosen skewness parameters for all set-ups are:

$$\gamma_a = [0.7\ 0.7\ 1.0\ 1.5\ 1.5]$$

To try out some different skewness parameters I will make 15 additional set-ups based on $\Sigma_{wanted,1}$ and

$$\gamma_b = [1.0\ 1.0\ 1.0\ 1.0\ 1.0]$$

$$\gamma_c = [0.5\ 0.5\ 1.0\ 2.0\ 2.0]$$

$$\gamma_d = [0.9\ 0.9\ 1.0\ 1.2\ 1.2]$$

Note that the parameters are quite arbitrary chosen, but chosen to give a variety in the different set-ups.

The first step is to generate the data needed. A series of 1250 returns is generated with the chosen parameters. To resemble a portfolio optimization the first 1000 returns are used for estimating a vector of mean returns and a covariance matrix to chose optimal weights. These weights are then used to form a portfolio, and the performance is measured for the portfolio of the last 250 returns, when varying these weights. This procedure is then repeated 2000 times to gain some stability. Look at Figure 4.19 where some outputs from the set-up $[\mu_{wanted},\ \Sigma_{wanted,1},\ K_2,\ \gamma_a]$ are shown as an example.

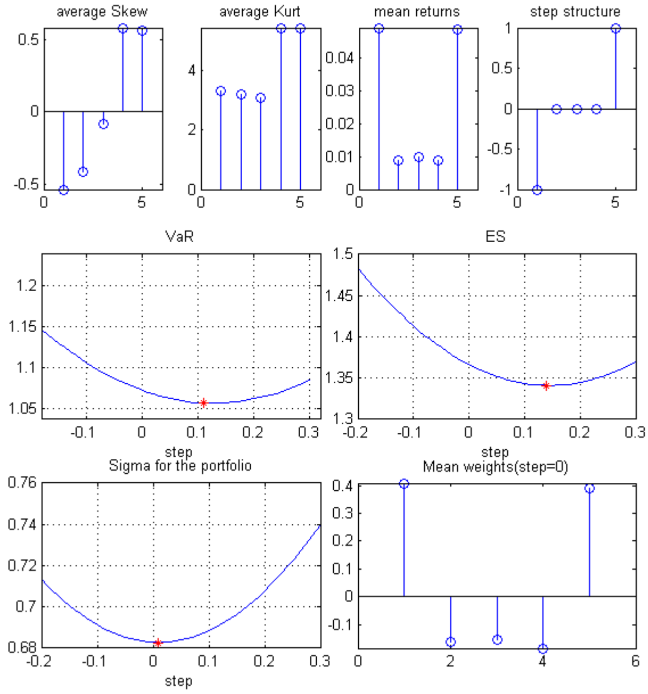

Figure 4.19: *For the set-up* $[\mu_{wanted}, \Sigma_{wanted,1}, K_2, \gamma_a]$. *The first graph shows the average empirical Skewness for the set-up. The second displays the average empirical kurtosis. The third graph shows the mean returns, also used as expected returns in the optimization. The last graph shows the average resulting weights,* $w_{average}$, *of the optimization. The fourth graph simply displays the step structure, that is* $D = [-1\ 0\ 0\ 0\ 1]$. *The fifth, sixth and seventh graph, displays the average* $VaR_{0.05}$, $ES_{0.05}$ *and* σ *for the resulting portfolio when varying the weights according to* $w_{after} = w_{opt} + d_w \cdot D$ *where* w_{opt} *is the weights gotten by the optimization and* d_w *is the variable "step" displayed on the x-axis*

Now, if we take all the set-ups and put them together, we find all the average weights that minimizes the average $VaR_{0.05}$ and $ES_{0.05}$. Now look at d_μ, the adjustment needed to be done on average to the expected returns to get w_{after}. As well look at the difference between the skewness of return series 1 and 5, and plot it against $2d_\mu$, see Figure 4.20. The relationship seems to be linear, at least for this range of values of skewness. Since values of larger magnitude of skewness are not observed in any of the datasets I use, I now feel comfortable with building a model where I use a linear relationship when the difference in skewness among assets is valued in terms of mean returns. Note that I have used the same volatility for the assets in these simulations. If applied to assets with large differences in volatility, the relationship, will also be dependent on the volatility, the magnitude of the expected returns

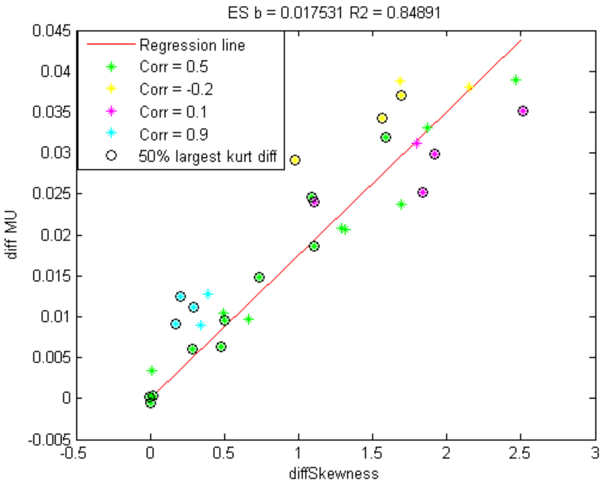

Figure 4.20: *The result from the different set-ups, skewness difference vs.* $2d_\mu$*, that minimizes* $ES_{0.05}$

are typically proportional to volatility in the long run. I am aware that I have not really proven anything by this simulation study, but it indicates that I'm not on the totally wrong track with my approach.

4.3.3 Empirical Tests with Portfolios

I will build a set of different test portfolios of the assets in dataset 2, containing log-returns from the 4^{th} of January 1993 until today. First let me introduce some notation that I intend to use:

$O_t, H_t, L_t, X_t =$ Vectors of Open-, High-, Low- and Close prices for contracts day t.

$C_t =$ correlation matrix of the whole time series day t, in an exponential rolling frame with time constant $T_{cov} = 250$

$dX_t = X_t - X_{t-1} =$ vector of returns for each asset at day t

$\sigma_t =$ OHLC-volatility estimate, exponentially rolling with time constant $T_{vol} = 15$

$\sigma_{t,T_{days}} =$ Estimated volatility in an exponentially rolling frame based on close prices, time constant $= T_{days}$

$\nu_t =$ vector of ML-estimate of degrees of freedom for marginal distributions of all asset returns based on the 1500 observations before day t.

$\theta_{2,t} =$ vector of estimated skewness regression parameter times $\sigma_{t,T_{days}}$ of returns, for marginal distribution of all assets based on the 1500 returns before day t.

$\mu_{E,t} =$ vector of expected returns day t

$\mu_{R,t} =$ vector of re-valued expected returns day t

$\sigma_{R,t} =$ re-valued/penalized volatility.

$\Sigma_{R,t} =$ re-valued/penalized covariance matrix.

$\sigma^2_{max} =$ target portfolio variance for the optimization $= 10^{-5}$.

$\Sigma_t = \mathbf{diag}(\sigma_t) C_t \mathbf{diag}(\sigma_t)$ Covariance matrix used in optimization to scale the weights to achieve σ^2_{max}

In a futures portfolio there are (mostly) no restrictions on taking short or long positions. There are of course transaction costs, slippage and a fact that contracts requires discrete positions of determined sizes. In the evaluation of portfolio performance, I will neglect these issues. Note that I have also omitted the analysis of volume data, this is due to the fact that when the speculative market is rolling into the new contracts the volume traded will be unproportionally large to what it "actually" should have been. There are no indicators in the time series for when the contract is rolled and it would require some very sophisticated filtering of the data to get something useful out of it.

The optimal weights

The first 1500 days of the time series will be used to estimate the first values of $\theta_{2,t}$ and ν_t. So the actual test period will be from observation 1501 and forward (30^{th} October 1998).

In the simulated portfolio I will re-balance every 5^{th} trading day. This is a reasonable and quite realistic time frame for rebalancing according to practice. The weights, w_t at day t, can be derived from the corresponding result in (2.15).

To test performance of a portfolio when the valuation of skewness and the penalizing of heavy tails are taken into account, I first want to try it out on a lot of different levels on many different set-ups. Let (element wise)

$$\sigma_{R,t} = \sigma_t \left(\frac{1}{\sqrt{\nu_t}} \right)^{\epsilon} \quad \text{and} \quad \mu_{R,t} = \mu_{E,t} + 0.1\sigma_{T_{250}}\delta \left(\frac{\theta_{2,t} - \text{mean}(\theta_{2,t})}{0.06} \right)$$

ϵ and δ are parameters that can be varied. For the value of $\mu_{R,t}$, $\theta_{2,t} - \text{mean}(\theta_{2,t})$ is the deviation from the mean of the skewness parameter at day t. The denominator, 0.06 is simply the maximal observed $\theta_{2,t}$, and it makes sure that the deviation from the mean of $\theta_{2,t}$ is normalized, so δ is easier to interpret. For example $\delta = 0.1$ means that a deviation of 0.06 from the

Set-up nr.	$\mu_{E,t}$	Details
1	$\mu_{E,t,equal}$	$= 0.1\sigma_{t,T_{250}}$, proportional to exponentially rolling volatility estimates with time constant 250
2	$\mu_{E,t,tr500}$	Trend model: Exponentially rolling average of dX $T = 500$
3	$\mu_{E,t,tr200}$	Trend model: Exponentially rolling average of dX $T = 200$
4	$\mu_{E,t,tr100}$	Trend model: Exponentially rolling average of dX $T = 100$
5	$\mu_{E,t,tr25}$	Trend model: Exponentially rolling average of dX $T = 25$
6	$\mu_{E,t,rnd500}$	$\mu_{E,t}$ is a random process, time constant $T = 500$ according to (4.2)
7	$\mu_{E,t,rnd50}$	$\mu_{E,t}$ is a random process, time constant $T = 50$ according to (4.2)

Table 4.1: *For set-up 1-5 different periods of the time series will be studied for many different values of δ and ϵ. For a few of these values, I will look at draw-downs and return performance. Set-up 5 and 6 will be repeated many times to see how the methods in average affect the portfolio behaviour by simply choosing expected returns to be random.*

mean is valued to 1% of the standard deviation of that asset, in means of returns. Long term excess returns usually vary between 1% and 10% of σ. That implies that $\delta = 1$ is a quite large value where the valuation of skewness start to dominate the $\mu_{E,t}$.

As for ϵ, the findings that the factor $\frac{1}{\sqrt{\nu}}$ is suitable for penalizing heavy tails by re-scaling return series, it is natural to use it to the power of ϵ, to study the effects of heavier penalizing. Note that when penalizing the heavy tails, weights are still re-scaled so that $w\Sigma_t w' = \sigma^2_{max}$.

The different set-ups

To get a better understanding of how the re-valuation/penalizing will affect a portfolio, it is important to try out many different set-ups of different $\mu_{E,t}$ see Table 4.1 for an explanation of all the set-ups I will use.

The first set-up can be called a benchmark set-up. The expected return is assumed to be proportional to volatility, and it resembles the equilibrium idea of the CAPM. What proportion we assume here is irrelevant since it's just a matter of scaling, the optimal weights will still have the same proportions to each other. But the constant is chosen to be 0.1 so δ will be easy to interpret in the revaluation. Set-up $2-5$ are simple trend following set-ups. The value of $\mu_{E,t}$ is decided by the exponentially rolling average of returns with different time steps. In set-up 6 and 7 I will chose a random $\mu_{E,t}$. $\mu_{E,1}$ is then generated from a uniformly distributed variable on the interval $(-0.1\sigma_{1,T_{250}}, 0.1\sigma_{1,T_{250}})$. It is also updated daily by

$$\mu_{E,t} = 0.5^{\frac{1}{T}}\mu_{E,t-1} + (1-0.5^{\frac{1}{T}}) \cdot \text{rnd}(-0.1\sigma_{t-1,T_{250}}, 0.1\sigma_{t-1,T_{250}}) \tag{4.2}$$

for some time constant T where $\text{rnd}(a,b)$ denotes a uniformly distributed random number from the interval (a,b). This means that $\mu_{E,t}$ is a random process and it will change over time. In set-up 6, the process is slow moving with a time constant of $T = 500$ and in 7, faster moving with $T = 50$.

For the set-ups $1-5$ I will also divide the test period into three different periods. The first 1120 days, the middle 1120 days and the last 1120 days to see if effects are consistent over time on the risk measures.

Results

For set-up 1-5 it's easiest to illustrate the simultaneous effects of δ and ϵ in 3d-graphs. For the first set-up in period 1 the results are displayed in Figure 4.21. As another example take a look at Figure 4.22 where results from set-up 3, period 2 are displayed, that is a trend-model based on 200 days as a time constant. It would be very tiring and inconvenient to show all the results in graphs like Figure 4.22. What I can say generally about the results is that I have not found a case so far where $\mu_{R,t}$ is dominated by the valuation of skewness, that increases the $ES_{0.05}$ or $ES_{0.025}$. As well as high values of δ in every case seems to have a positive effect on the skewness of portfolio returns. Another very interesting observation is that the decreases in ES seem to be closely connected to a decrase in realized portfolio standard deviation.

For the random set-ups I have made 100 optimizations and looked at the average outcome when varying δ and ϵ separately. In Figure 4.23 average outcomes as a function of δ are displayed. The pattern seems to be consistent for random $\mu_{E,t}$ as well. To not be fooled by the averages it might be a good idea to look at the outcomes for a high value of δ. In Figure 4.24 I have chosen $\delta = 3$, that means $\mu_{R,t}$ is dominated by the valuation of skewness and plotted the difference between the outcomes of $\delta = 0$ and $\delta = 3$. The results for ES, skewness and realized standard deviation are quite convincing. There are only about 1-3 cases out of the 100 where it has not changed in the desirable direction, and the overlap is those cases very

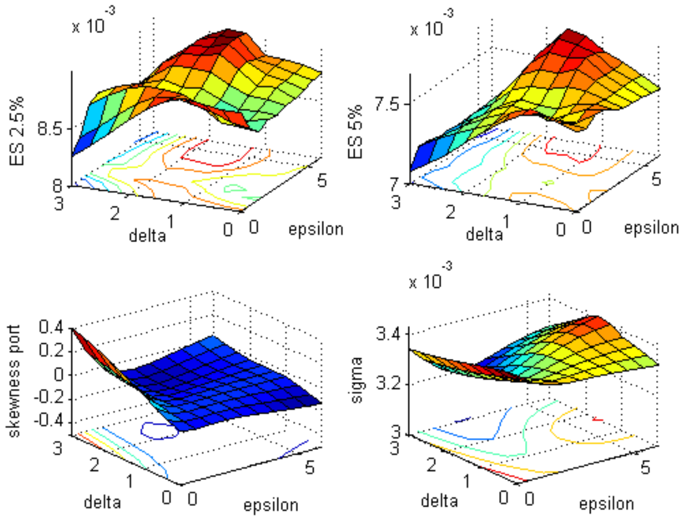

Figure 4.21: *The result from set-up 1 period 1. First and second graph shows* $ES_{0.025}$ *and* $ES_{0.05}$*, empirical estimates of portfolio returns. The third graph shows portfolio return skewness and the fourth, the realized portfolio standard deviation. To get a more comprehensive picture of the graphs, it can be good to bear in mind that the risk are expressed in term of log-returns and the target standard deviation,* $\sigma_{max} \approx 3.16 \cdot 10^{-3}$

marginal. I have as well observed similar results for varying ϵ and I have put those in Appendix B. The essential difference is that the penalizing of heavy tails instead tend to have a clearly visible effect of lowering kurtosis instead of increasing skewness.

Even though is seems like being aware of skewness and heavy tails tend to lower the expected single period draw-down. It is important to see to the total effect. In Figure 4.3.3 the draw-downs for set-up 1 based on the whole test-period, with $\delta = 0$ and $\delta = 3$ are displayed. In the period in the middle $\delta = 3$ leads to much larger draw-downs. Still the ES is lowered for that period in the set up (see tables in Appendix B). The result is consistent with the other observations, the lowered ES does not necessarily have a cumulative effect on draw-downs.

To give the reader an overview of the results I have put together the results in tables in Appendix B For all the different set-ups I have computed values of all parameters, for high values of δ and ϵ representing an extreme belief in skewness as a value and heavy penalizing of tail thickness. As well as moderate values, so the optimization is influenced, not dominated by the results and zero values as a benchmark.

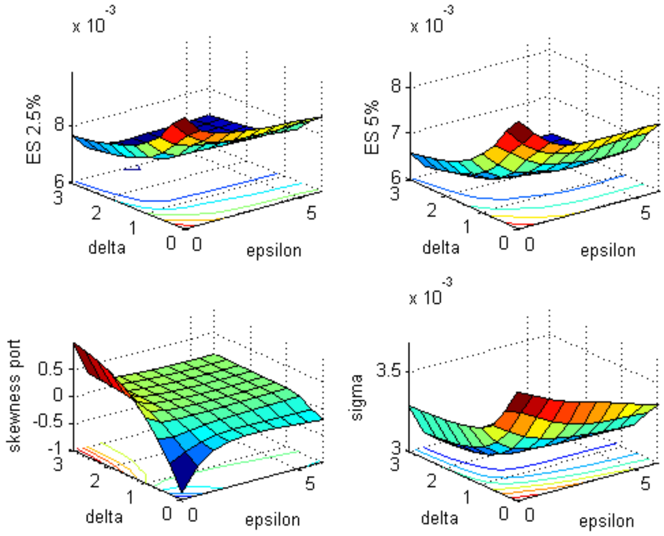

Figure 4.22: *The result from set-up 3 period 2. First and second graph shows* $ES_{0.025}$ *and* $ES_{0.05}$*, empirical estimates of portfolio returns. The third graph shows portfolio return skewness and the fourth realized portfolio standard deviation*

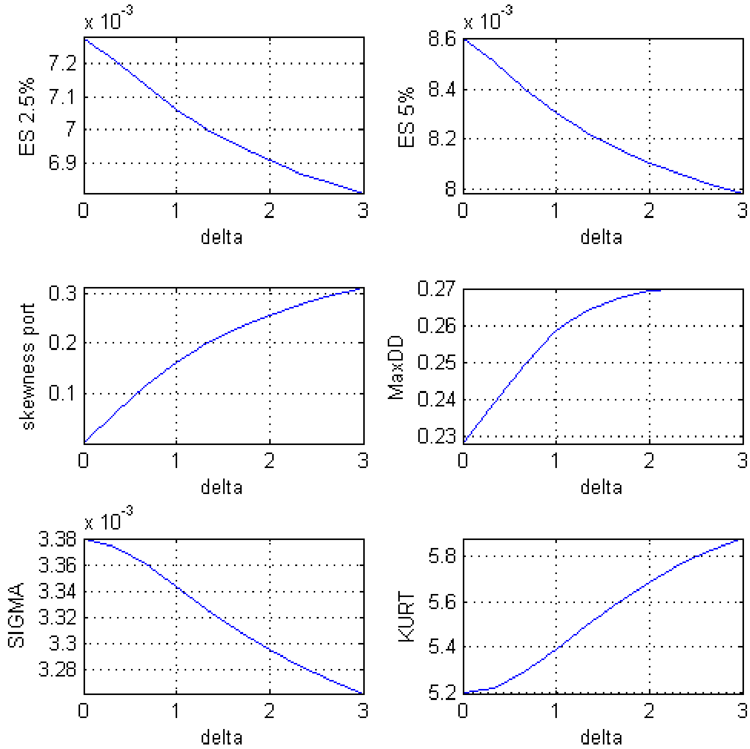

Figure 4.23: *The result from set-up 6 based on the whole sample as functions of δ. First and second graph shows $ES_{0.025}$ and $ES_{0.05}$. Third graph shows the skewness and the fourth the maximum draw-down. Fifth graph shows realized standard deviation and the sixth the kurtosis for portfolio returns.*

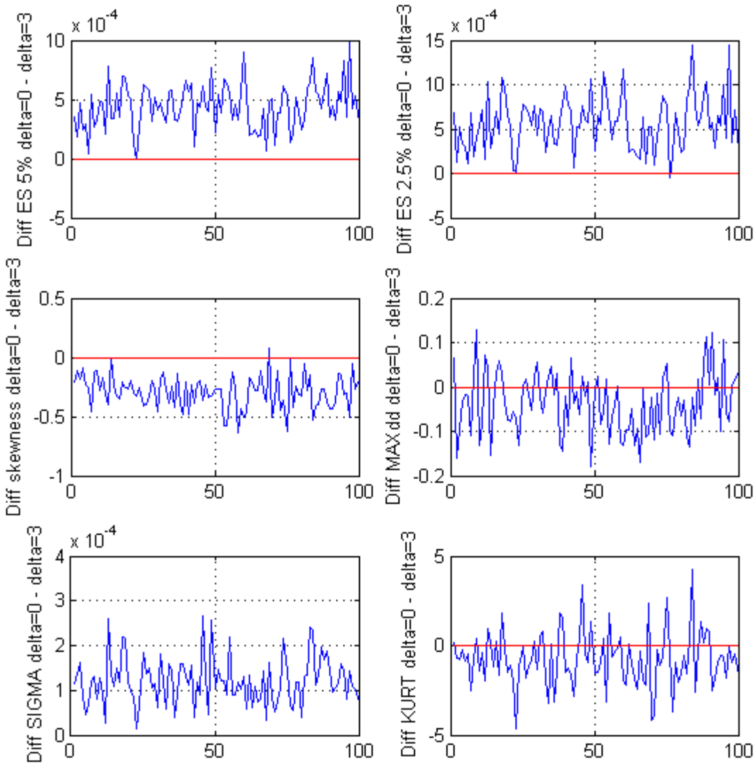

Figure 4.24: *The result from set-up 6, differences for $\delta = 0$ and $\delta = 3$ for all optimization runs. First and second graph shows $ES_{0.025}$ and $ES_{0.05}$. Third graph shows the skewness and the fourth the maximum draw-down. Fifth graph shows realized standard deviation and the sixth the kurtosis for portfolio returns.*

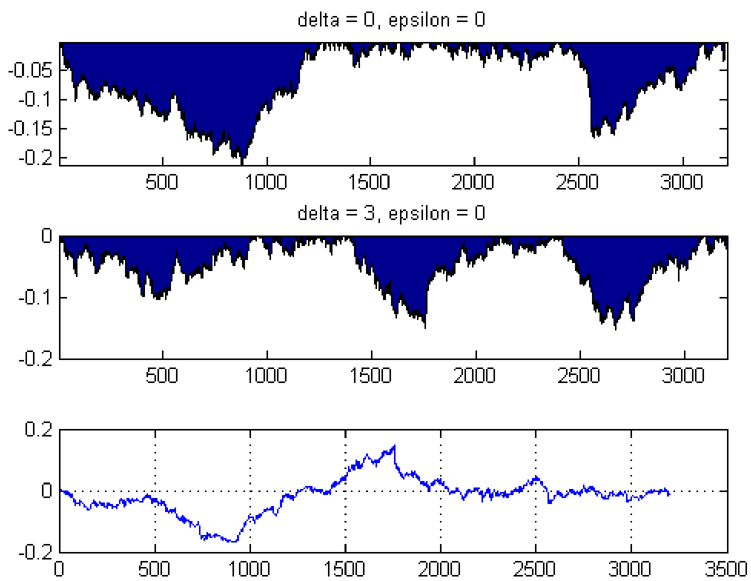

Figure 4.25: *Draw-downs for set-up 1, with $\delta = 0$ above, $\delta = 3$ in the middle and the third graph display the difference between the both*

Chapter 5

Conclusions and discussion

5.1 Good Times and Bad

The empirical results showed that even if many of the correlation matrices based on blended matrices would increase the average likelihood of observations made, it was not due to the fact that estimated tail correlation would be a good predictor of the dependence structure of future outliers. The analysis of the data showed that if we instead used shrinkage towards a factor model based on eigenvectors or even a identity matrix, we could achieve essentially the same results. These shrinkage methods seemed to have a consequently better performance in predicting the correlation structure of large outliers than the reference matrix estimates, a property the outlier-inside-based estimates did not have. A strong shrinkage towards the identity matrix even seemed to be the best estimate, when it came to the most extreme outliers.

One way to interpret the results is that when it comes to extreme events, it is hard to predict behaviour. At least when it comes to describe the dependence with linear correlation. That the shrinkage towards the identity seemed to be the superior method to describe the dependence among the most rare events[1] indicate that outliers have weaker correlations, or at least a very different correlation structure, than both the original correlation matrix, and the one based on outliers. However, the assets I have been looking at are futures contracts and their behaviour might differ from the underlying asset.

The results suggest that the outlier correlation structure is bad at predicting future outcomes in terms of outlier correlation. This supports the belief that outliers can occur for many different and complex reasons, affect different markets in different ways and be very different from time to time. I would not recommend to use it as a method, but recommend to use some kind of

[1]The most rare events, given this framework based on normal distributions

shrinkage based approach instead.

5.2 Higher Order Moments

The empirical results of this part clearly shows that paying attention to the higher order moments can have positive effects on portfolio returns in different ways. These results are consistent with earlier research in this area, namely it tells us that we can use these higher order moments to get better portfolio performance when it comes to estimated risk based on the realized portfolio returns. The frameworks I have been using are only the most simple quadratic optimization frameworks of classic modern portfolio theory.

The simulation study implies that it is reasonable to value the product of skewness and standard deviation, linearly in terms of expected returns, if the target is to minimize one period draw-downs. In the optimization part however, skewness is valued linearly but the weights of the portfolio is always re-balanced to maintain a certain target portfolio variance. My purpose was to study the impact on realized portfolio risk, not to find an optimal valuation of skewness.

For the part concerning the regression parameter θ_2 as a robust alternative for skewness, is something new, that I have not seen in any other research before. It has however to be investigated further, as well analytically and quantitative if one should be able to draw any conclusions about it. The results in this paper indicate that it has at least approximately a linear relationship with skewness, in the area of realistic asset return skewness. It also has the appealing property of taking all observations into account, which most quantile based estimates don,t and it seems more robust than the standard definition of skewness. The drawback is that it is much more computational intensive than the standard skewness measure.

The results form the optimization suggest that both penalizing heavy tails by re-scaling volatilities and re-valuing expected returns in terms of skewness will reduce portfolio risk. Not only does it seem to lower the expected shortfall of the portfolio it also seem to lower the realized portfolio variance. It is a nice and to me rather surprising effect, that I have no good explanation for, except that it might just be a coincidence. That the realized portfolio variance is lowered is however interesting from a portfolio managers point of view, since it means that he/she can take on more risk in terms of expected portfolio variance. An other interesting observation was that a portfolio where skewness is valued high, got high increases in skewness in the portfolio returns, as well as kurtosis was lowered for portfolios with high tail awareness. First it is interesting that these properties can be transferred to portfolio level. Second it's almost even more interesting that this result supports the hypothesis that different assets are characterized by their higher order moments and that they can be used to give an idea about future behaviour.

It is important to note that even if the effects are positive it does not seem like there are any cumulative effects on draw-downs. The results are anyhow interesting, but will require some more research to be more reliable. It is hopefully possible to use the parameters together with other market signals, to affect the cumulative draw-downs as well.

For further studies I would recommend to use other frameworks like the Black-Litterman model on which a intuitive tutorial is given by Cheung [2009], where valuation of skewness could be used in the view-matrix and the tail thickness could be used in the uncertainty part. To have an effect on long term draw-downs, maybe it would be better to try the approach with weekly or monthly returns? Note also that I only have run these set-ups on various portfolios of the same 16 assets. It is important to try this set-up on a wide range of different assets before jumping to any general conclusions.

5.3 Final comments

The results discussed above are interesting and (hopefully) useful from the perspective of a portfolio manager, in the quantitative area. That outliers and their dependence structure would be hard to predict, is not a big surprise, but it should at least encourage to use shrinkage methods in correlation estimates, if such events are of concern. The combination of higher order moments combined with other market signals is an area that would be of interest to study further, especially for those who have risk reduction in portfolios as a target. However there is a lot that has to be investigated further and tested empirically on a wider scale.

Appendix A

Assets

Equity	Bonds	Commodities	Currencies
dax	canada60	aluminium	aud
estoxx	10ynote	brent	cad
ftse	5ynote	copper	chf
hangseng	aus10y	corn	eur
ibex	bobl	crude	gbp
mib	bund	gold	jpy
nasdaq	gilt	zinc	
nikkeiose	jgb	wheat	
omx	tbond	hoil	
russel	cgb10y	natgas	
sp	2ynote	rbob	
spi	aus3y	soybeanmeal	
cac	ausbax	soybeanoil	
	cadbax	soybeans	
	euribor	sugar	
	eurodollar		
	schatz		
	ss		

Table A.1: *Contract names of dataset 1*

Equity	Bonds	Commodities	Currencies
Nikkei 225	Canadian 10yr	NY crude light oil	aud
SP 500	Eurodollar	Soy bean oil	gbp
	Tnotes2yr	Corn	cad
	Tbonds30yr	Feeder Cattle	
		HG copper	
		Orange juice	
		Sugar	

Table A.2: *Contract names of dataset 2*

Appendix B

Graphs and Tables

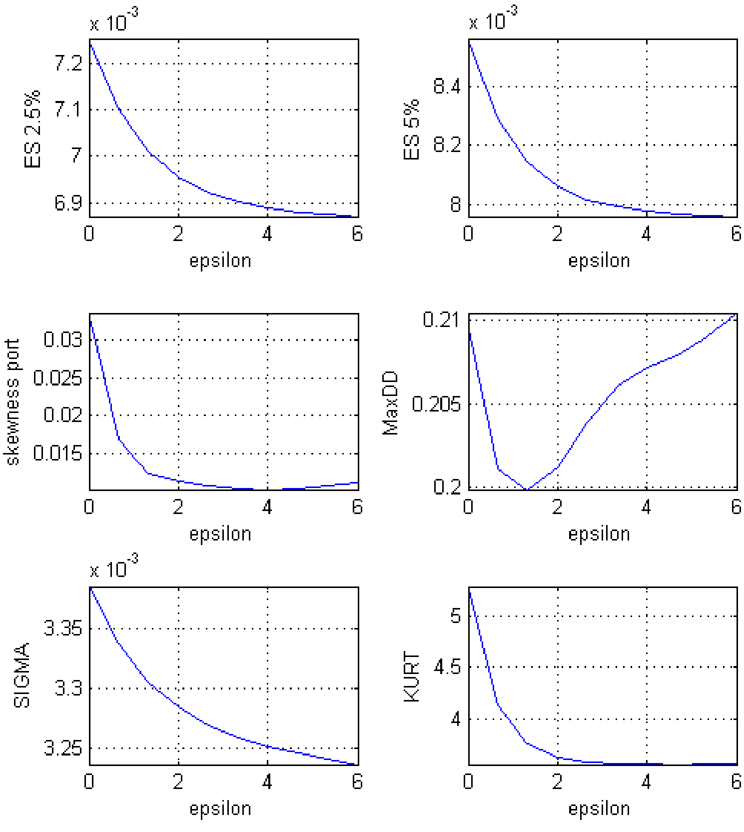

Figure B.1: *The result from set-up 6 based on the whole sample as functions of* ϵ*. First and second graph shows* $ES_{0.025}$ *and* $ES_{0.05}$*. Third graph shows the skewness and the fourth the maximum draw-down. Fifth graph shows realized standard deviation and the sixth the kurtosis for portfolio returns.*

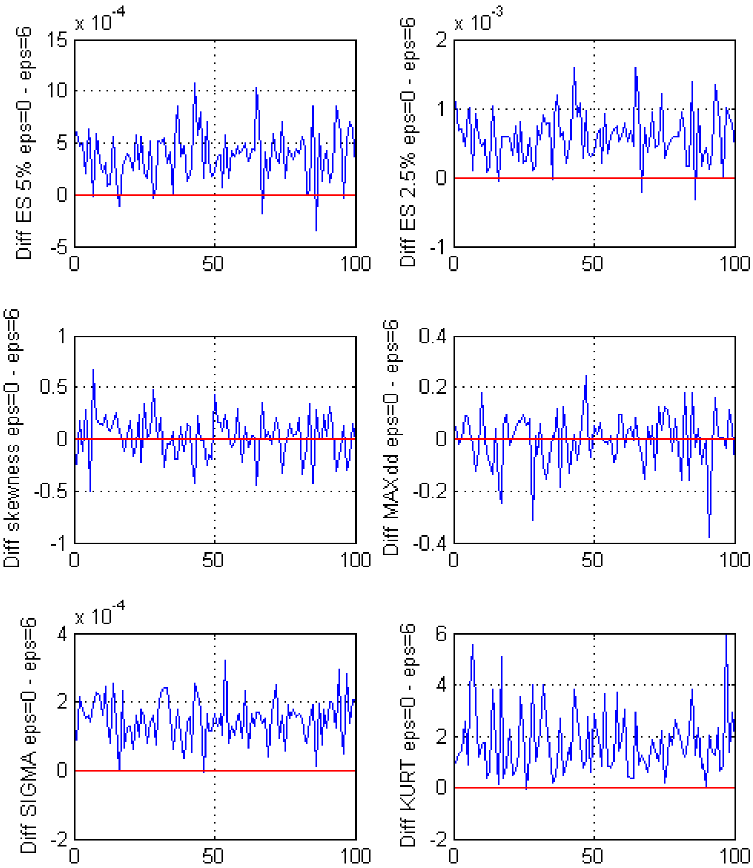

Figure B.2: *The result from set-up 6, differences for $\epsilon = 0$ and $\epsilon = 6$ for all optimization runs. First and second graph shows $ES_{0.025}$ and $ES_{0.05}$. Third graph shows the skewness and the fourth the maximum draw-down. Fifth graph shows realized standard deviation and the sixth the kurtosis for portfolio returns.*

Period	[δ ϵ]	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
1	[0 0]	8.67	7.37	-0.06	4.00	3.38	0.218
1	[0.2 0]	8.74	7.46	0.01	4.40	3.39	0.190
1	[3 0]	8.33	7.18	0.32	6.26	3.34	0.105
1	[0 1]	8.45	7.33	-0.08	3.74	3.31	0.266
1	[0 6]	8.54	7.28	-0.14	3.65	3.26	0.158
1	[3 6]	8.18	6.94	0.09	3.43	3.26	0.249
2	[0 0]	9.01	7.70	-0.02	3.63	3.63	0.052
2	[0.2 0]	9.15	7.71	-0.08	4.72	3.69	0.055
2	[3 0]	8.19	6.82	-0.10	11.84	3.49	0.151
2	[0 1]	8.51	7.50	0.22	3.50	3.51	0.046
2	[0 6]	7.86	6.70	1.14	3.65	3.25	0.057
2	[3 6]	7.28	6.23	0.05	4.07	3.08	0.092
3	[0 0]	10.94	8.76	-0.60	5.93	3.50	0.166
3	[0.2 0]	11.37	8.87	-0.66	6.61	3.53	0.178
3	[3 0]	10.88	8.57	-0.50	6.28	3.48	0.153
3	[0 1]	9.64	7.91	-0.37	4.41	3.40	0.153
3	[0 6]	8.22	7.08	-0.09	3.27	3.31	0.182
3	[3 6]	8.96	7.72	-0.06	3.86	3.48	0.226

Table B.1: *Results for set-up 1.*

Period	[δ ϵ]	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
1	[0 0]	8.61	7.29	0.13	5.24	3.40	0.121
1	[0.2 0]	9.05	7.57	0.19	5.98	3.46	0.120
1	[3 0]	7.70	6.63	0.45	5.74	3.34	0.094
1	[0 1]	8.51	7.47	-0.03	4.08	3.39	0.113
1	[0 6]	8.19	6.96	-0.10	3.42	3.31	0.050
1	[3 6]	7.32	6.31	0.31	3.52	3.18	0.140
2	[0 0]	9.47	8.07	-0.10	4.52	3.51	0.103
2	[0.2 0]	9.05	7.52	0.87	11.62	3.48	0.138
2	[3 0]	7.66	6.48	1.52	15.73	3.32	0.199
2	[0 1]	9.17	7.84	-0.26	3.53	3.49	0.083
2	[0 6]	8.24	7.09	-0.28	3.65	3.23	0.104
2	[3 6]	7.03	6.16	0.09	3.32	3.04	0.106
3	[0 0]	10.43	8.46	-0.87	8.76	3.43	0.121
3	[0.2 0]	10.78	8.49	-0.85	10.1	3.40	0.147
3	[3 0]	10.05	8.04	-0.40	5.90	3.41	0.109
3	[0 1]	9.96	8.40	-0.78	7.76	3.47	0.105
3	[0 6]	8.53	7.25	-0.10	3.63	3.45	0.109
3	[3 6]	8.00	6.96	-0.08	3.38	3.20	0.142

Table B.2: *Results for set-up 2.*

Period	$[\delta\ \epsilon]$	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
1	[0 0]	8.65	7.34	0.23	5.67	3.46	0.092
1	[0.2 0]	8.67	7.42	0.29	5.65	3.48	0.101
1	[3 0]	7.76	6.65	0.47	5.52	3.37	0.093
1	[0 1]	8.48	7.51	0.02	3.94	3.46	0.085
1	[0 6]	8.88	7.41	-0.16	3.60	3.42	0.107
1	[3 6]	7.37	6.36	0.30	3.42	3.24	0.156
2	[0 0]	10.21	8.49	-0.91	10.06	3.76	0.097
2	[0.2 0]	9.21	7.82	-0.28	4.18	3.51	0.083
2	[3 0]	7.65	6.62	0.84	9.45	3.28	0.189
2	[0 1]	9.18	7.78	-0.31	4.42	3.65	0.055
2	[0 6]	8.41	7.37	-0.17	3.35	3.36	0.091
2	[3 6]	7.08	6.17	0.07	3.30	3.04	0.105
3	[0 0]	11.11	9.09	-0.65	6.84	3.70	0.088
3	[0.2 0]	10.83	8.65	-0.65	8.36	3.55	0.110
3	[3 0]	10.13	8.05	-0.43	6.36	3.42	0.103
3	[0 1]	10.43	8.69	-0.59	6.04	3.67	0.065
3	[0 6]	8.64	7.35	-0.14	3.48	3.52 0.053	
3	[3 6]	8.01	6.92	-0.03	3.41	3.25	0.113

Table B.3: *Results for set-up 3.*

Period	$[\delta\ \epsilon]$	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
1	[0 0]	8.86	7.51	0.18	5.47	3.53	0.083
1	[0.2 0]	8.92	7.47	0.25	5.54	3.53	0.089
1	[3 0]	7.97	6.73	0.44	5.33	3.41	0.091
1	[0 1]	8.51	7.53	0.01	3.80	3.54	0.078
1	[0 6]	8.51	7.32	-0.09	3.41	3.45	0.120
1	[3 6]	7.29	6.33	0.32	3.48	3.24	0.191
2	[0 0]	10.44	8.57	-0.65	7.93	3.90	0.067
2	[0.2 0]	9.41	7.86	-0.12	4.00	3.69	0.0651
2	[3 0]	7.72	6.69	0.56	7.11	3.29	0.183
2	[0 1]	9.25	7.85	-0.15	3.86	3.75	0.058
2	[0 6]	8.32	7.25	-0.16	3.17	3.37	0.67
2	[3 6]	7.06	6.15	0.08	3.25	3.05	0.107
3	[0 0]	10.73	8.88	-0.44	5.60	3.83	0.086
3	[0.2 0]	10.73	8.65	-0.48	6.46	3.71	0.102
3	[3 0]	10.01	7.98	-0.40	6.05	3.45	0.097
3	[0 1]	10.44	8.68	-0.46	5.23	3.80	0.078
3	[0 6]	8.91	7.53	-0.12	3.49	3.56	0.052
3	[3 6]	8.33	7.08	0.03	3.45	3.40	0.089

Table B.4: *Results for set-up 4.*

Period	$[\delta\ \epsilon]$	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
1	[0 0]	8.69	7.35	0.19	4.51	3.58	0.104
1	[0.2 0]	8.64	7.39	0.23	4.59	3.60	0.110
1	[3 0]	8.52	7.11	0.32	5.02	3.52	0.109
1	[0 1]	8.56	7.57	0.05	3.51	3.54	0.134
1	[0 6]	8.38	7.32	0.00	3.34	3.44	0.189
1	[3 6]	7.16	6.35	0.33	3.41	3.24	0.245
2	[0 0]	10.56	8.55	-0.46	5.99	3.92	0.070
2	[0.2 0]	9.77	8.14	-0.17	3.89	3.81	0.080
2	[3 0]	8.28	7.18	0.21	4.85	3.43	0.179
2	[0 1]	9.71	8.10	-0.19	3.74	3.81	0.059
2	[0 6]	8.50	7.30	-0.13	3.28	3.46	0.098
2	[3 6]	7.39	6.31	0.06	3.55	3.15	0.122
3	[0 0]	10.94	9.29	-0.25	4.52	4.11	0.115
3	[0.2 0]	10.95	9.12	-0.29	4.80	4.05	0.110
3	[3 0]	9.88	8.27	-0.32	5.13	3.66	0.106
3	[0 1]	10.62	9.10	-0.19	4.10	4.03	0.103
3	[0 6]	9.16	7.86	-0.10	3.64	3.62	0.064
3	[3 6]	9.08	7.74	0.01	3.39	3.67	0.083

Table B.5: *Results for set-up 5.*

T	$[\delta\ \epsilon]$	$ES_{0.025} \cdot 10^{-3}$	$ES_{0.05} \cdot 10^{-3}$	Skew	Kurt	$\sigma \cdot 10^{-3}$	Max DD
500	[0 0]	8.59	7.28	0.00	5.15	3.39	0.239
500	[0.2 0]	8.46	7.20	0.07	5.22	3.37	0.256
500	[3 0]	8.03	6.84	0.29	5.91	3.26	0.276
500	[0 1]	8.23	7.07	0.00	3.88	3.35	0.213
500	[0 6]	8.02	6.91	0.00	3.52	3.28	0.222
50	[0 0]	8.62	7.30	-0.02	5.18	3.38	0.253
50	[0.2 0]	8.48	7.20	0.06	5.24	3.38	0.268
50	[3 0]	8.01	6.83	0.31	5.98	3.27	0.280
50	[0 1]	8.24	7.08	0.00	3.86	3.32	0.219
50	[0 6]	8.02	6.91	0.00	3.55	3.24	0.206

Table B.6: *Results for set-up 6 and 7.*

Bibliography

URL `http://www.cftc.gov/About/CFTCReports/index.htm`. Annual reports of CFTC.

Howard Anton and Chris Rorres. *Elementary Linear Algebra*. Anton Textbooks Inc., 9 edition, 2005. pages 369-371.

Marianne Baxter and Robert G. King. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics ans Statistics*, 81(4):575–593, 1999.

R. C. Blattberg and N. J. Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, 47, 1974.

Gunnar Blom, Jan Enger, Gunnar Englund, Jan Grandell, and Lars Holst. *Sannolikhetsteori och statistikteori med tillampningar*. Studentlitteratur, 2005. Ch. 6 and pages 292-293.

Wing Cheung. The blackâlitterman model explained. *Journal of Asset Management*, 11, 2009.

George Chow, Eric Jacquier, Mark Kritzman, and Kenneth Lowry. Optimal portfolios in good times and bad. *Financial Analysts Journal*, 27, May/June 1999.

Rohan Christie-David and Mukesh Chaudhry. Coskewness and cokurtosis in futures markets. *Journal of Empirical Finance*, 8, 2001.

Pornchai Chunhachinda, Krishnan Dandapani, Shahid Hamid, and Arun J. Prakash. Portfolio selection and skewness, evidence from international stock markets. *Journal of Banking and Finance*, 21.

E. Fama. The behavior of stock market pricess. *Journal of Business*, 38, 1965.

JosÂ´e T. A. S. Ferreira and Mark F. J. Steel. A new class of skewed multivarite distributions with applications to regression analysis. *Statistica Sinica*, 17, 2007.

Edward W. Frees and Emiliano A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2, 1996.

Alex Gioulekas and Boualem Djechiche. Tail risk optimisation, 2009. URL http://www.ipm.se/Upload/Research/Insights201120-20Q4202009.pdf. Published by IPM Informed Portfolio Management.

Allan Gut. *An Intermediate Course in Probability.* Springer, 1995. Ch. 5 and pages 122 and 8-9.

Campbell R. Harvey, John C. Liechty, Merrill W. Liechty, and Peter Muller. Portfolio selection with higher moments, December 2004. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=634141. Working Paper Series.

John C. Hull. *Options Futures and Other Derivatives.* Pearson Education Inc, 2009. pages x.

H. Hult, F. Lindskog, O. Hammarlid, and C.J. Rehn. Risk and portfolio analysis, 2011. To be bublished by Springer. Available at the institution of mathematics at KTH. pages 219 and 133-186 and 222-236.

Robert W. Kolb and James A. Overdahl. *Understanding Futures Markets.* Blackwell Publishing, 6 edition, 2006. pages 2-3.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *Journal of Portfolio Management*, 31(1), 2004.

Bob Litterman. *Modern Investment Management.* John Wiley and Sons Inc, 2003. pages 36 and 42.

B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 26, 1963.

Harry M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

Amol Sasane and Krister Svanberg. Optimization, 2009. pages 95. Course material available on the institution of mathematics at KTH.

Jonathon Shlens. A tutorial on principial component analysis, April 2009. URL http://www.mendeley.com/research/a-tutorial-on-principal-components-analysis/.

James H. Stock and Mark W. Watson. *Introduction to Econometrics.* Pearson Education Inc, 2008. pages 25 and 93.

Qian Sun and Yuxing Yan. Robust measures of tail weight. *Journal of Banking and Finance*, 27, 2003a.

Qian Sun and Yuxing Yan. Skewness persistence with optimal portfolio selection. *Journal of Banking and Finance*, 27, 2003b.

Nassim Nicholas Taleb. *The Black Swan.* Random House, 2007.

Dennis Yang and Qiang Zhang. Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, 73, 2000.