Asymmetric dependence between multiple asset returns:
an Expectation-Maximization algorithm for the Skewed t copula

Master Thesis

Daniël van den Brand - 290798

Erasmus University Rotterdam
Erasmus School of Economics
Master Econometrics and Management Science
Specialisation: Quantitative Finance

Supervisor EUR:
Lennart Hoogerheide

Co-reader EUR:
Erik Kole

Supervisor PGGM:
Jacco Koopmans

October 22, 2009

Acknowledgements

What seemed as an unbridgeable final project turned into a most interesting and defiant study on dependence, risk and optimization. With the eternal interesting possibilities my supervisors handed me over to study, the end never even seemed to come close. However, during the course of the project, the pieces started getting together bit by bit. What started as an overwhelming project, developed into a very specific and most relevant project. However, this was not possible without the great ideas which my colleagues at PGGM - and especially Jacco Koopmans and Richard Kleijn - came up with initially. The discussions and their suggestions and comments helped me in finding the right direction, for which I am really thankful. Further, I am grateful to Thijs Markwat of Erasmus University for proposing the idea of analyzing the Skewed t copula and providing useful hints including many references to relevant literature. Next, I want to thank Erik Kole (in advance) for reading the thesis and preparing questions for the defense session.

However it must be said, that this entire project would not have become the Master Thesis that I am proud to present without the inexhaustible effort of my supervisor, Lennart Hoogerheide. He always provided me with new ideas when certain methods did not work out properly. And although he was my supervisor, our relationship was always very informal and friendly. All which I am really thankful for.

Besides my supervisors I also like to thank, my parents, oldest brother and all my friends who have had their input somehow. They all had their contribution to bringing this thesis to a good end.

"*Thanks to God who has made all things possible*" (Mark 10:27).

Contents

Executive Summary

Diversification is used among investors as an important risk management technique to reduce risk. However to make diversification effective and even a lucrative tool, one needs to understand the overall behavior of different markets. More specifically one needs to know which dependency exists between these different markets. Especially in times of crisis where huge market downturns at one particular market can cause disastrous effects on other markets as well. The lastest financial crisis has shown that at first sight seemingly unrelated markets, can display dependencies in times of extreme market situations.

The multivariate normal distribution is the most popular model to estimate and simulate market returns. However there are two problems concerning the use of this model. Firstly, it can not model extreme market behavior. Secondly, since markets display non-normal behavior, linear correlation is no longer an adequate measure of dependence. Because multivariate normal distributions can not reproduce extreme market behavior, it can also not display extreme (tail) dependence. An alternative approach to resolve both problems, is by using copula models. The great advantage of the use of a copula is that the behavior of (return) series can be modeled separately from their interdependence structure. However most developed copulas in the literature are either constructed to capture symmetric - non skewed - dependence in higher dimensions, or otherwise are able to capture asymmetric tail dependence, but they are not well constructed to fit the distribution of higher (than 2) dimensional relationships.

In this thesis we join together both desired characteristics by constructing a higher dimensional Skewed t copula to model asymmetric tail dependence between multiple asset return indices. Furthermore we use different kinds of models to fit each unique asset return series as well as possible. Moreover, we use a Skewed t distribution or a combination of multiple normal- or Student's t distributions to fit each series separately. Next we apply these combinations of models in the field of risk- and portfolio management.

Goodness of fit tests show that the most advanced model (i.e. a Skewed t copula with the marginals from either the Skewed t distribution or a mixture of Student's t distributed components), is preferred over others. By testing the model on a variety of risk measures (Value-at-Risk, Conditional Value-at-Risk and Maximum Drawdown), we find that the mixtures of normal distributions for each return series, form the best marginal distributions. The same result is obtained by considering portfolio optimization techniques. When we look purely at the obtained tail dependence by the Skewed t copula, the results show that the Skewed t copulas does identify asymmetric tail dependence between High Yield and Real Estate. Our model also confirms that Treasury is not negative dependent with any of the other asset categories, during times of crisis. However looking at the overall performance of

the different copulas at the different applications, we can not conclude that there is a very large gain in performance by using the Skewed t copula. This is possibly due to the fact that there are only few joined tail events present in that sample.

The most surprising result comes from the performance of the most simple model, that is the multivariate normal distribution. Based on the goodness of fit tests this standard model shows the worst performance. Also based on the different risk measures it does underestimate heavy risk. However when it comes to portfolio optimization the multivariate normal model leads to impressively good portfolio selection. Even when advanced risk measures need to be minimized at a certain required return, the model still performs very well. Moreover, it even outperforms the most advanced model, and most of the other models as well. This is the most welcome news to all asset-, portfolio- and risk managers, as it seems that using the most simple model results in almost the same (in some cases even better) portfolio choice as more advanced models. It should be noted that the performed analysis is based on weekly data. For monthly data, tail dependence of contemporaneous returns may be stronger, possibly leading to a different conclusion. This is left as a topic for further research.

In conclusion, when it comes to dependence, the estimation results show that the higher dimensional Skewed t copula provides for a potential model improvement on higher dimensional problems. Moreover, based both on fit and the applications in the field of risk management and portfolio optimization, it really provides for a solution to improve the ability to estimate and reproduce the degree in which certain assets show tail dependence. And although the choice of the marginals has a big influence on the model outcomes, the dependence between the different assets is better captured and demonstrated by the use of a Skewed t copula model. For instance the knowledge that Treasury is not negatively dependent with the other assets, is a desirable result in the use of the Skewed t copula, especially in these times of crisis where no investment gives a guaranteed insurance. One could also consider further research on the use of so-called Markov regime-switching models which enables the use of different copulas in different time periods.

Abstract

During the financial crisis extreme market downturns occurred on several markets at the same time, causing deep tail dependence relationships to arise. Copula models are developed to provide for a solution to model interdependencies between different markets, in an adequate way. However most developed copulas in the literature are either constructed in a way that they are only capable of capturing symmetric dependence relationships in higher dimensions (i.e. the normal- of Student's t copula). Or otherwise they are able to capture asymmetric tail dependence, but are not well constructed to fit higher (than 2) dimensional relationships (i.e. the Clayton or Gumbel copula, among others). In this thesis we join together both desired characteristics by constructing a higher dimensional Skewed t copula to model asymmetric tail dependence between multiple asset return indices. Therefore we use an extended version of the EM algorithm of Hu (2005), for estimating multivariate Skewed t distributions, in combination with the simplex method, to set up a new framework; an EM-within-Simplex algorithm for estimating the Skewed t copula. Furthermore we apply the Skewed t copula in the field of risk- and portfolio management to track its performance. We identify in a portfolio of 6 assets (i.e. Equities, Commodities, EMD, Real Estate, Treasury and High Yield) asymmetric tail dependence between High Yield and Real Estate. Our model also confirms that Treasury is not negatively dependent with any of the other asset categories, during times of crisis.

Keywords: copula, marginals, skewed t, mixture components, tail dependence, correlation, risk, portfolio optimization, asset management

1 Introduction

The main objective of PGGM Investments is to generate a high and stable return for their clients. This is done through a judicious choice of a strategic investment mix and a tight control of this mix by intensive portfolio- and risk management. The main tool to achieve this goal is by ensuring a good diversification of risk by allocating the portfolio to different asset classes.[1] The main task of the Strategy department of PGGM is therefore to define a well diversified portfolio which tries to ensure these high and stable returns, but which is particularly robust in times of financial distress.

To make sure diversification will pay off, one needs to invest in different asset classes which are not linked. Moreover, if these different investments do not have any form of dependence in between, negative returns from one investment will be compensated by the gains of other investments. However, if markets are strongly linked together, a negative portfolio return on one market can cause portfolios on other markets to suffer as well.

Under 'normal' market conditions, dependence relationships of most financial markets are relatively easy to estimate and diversification will pay off. But the last financial crisis gives a classis example of changing interdependencies among assets. That is: financial distress on one market can lead to (growing) distress on other markets as well, which were seemingly not dependent in first place. Therefore it is crucial to understand the interdependence between different asset classes in different market seasons for defining a well diversified portfolio.

Yet this is only one important aspect of successful asset management. Another important issue to take into account is the ability to estimate size, frequency and impact of abnormal negative returns happening on certain financial markets. Specifically from the financial crisis, it has become clear that portfolio returns on most financial markets are increasingly exposed to extreme down movements than can be described by a normal distribution. The most well known measure of dependence is linear correlation. But since markets display non-normal behavior, linear correlation is no longer an adequate measure of dependence. This is because the use of linear correlation is only valid (in the sense that it tells the 'whole story' about the dependence) under the assumption of multivariate normally distributed returns. However, the normal distribution may not be the most desirable model to use in times of stress.

An alternative approach to model the dependence and overcome the disadvantages of using linear correlation and the shortcomings of normal distributions, is by using a copula. The great advantage of the use of a copula is that the behavior of each univariate asset returns series can be modeled separately from their interdependence structure. In short, the return series can be transformed into normally distributed variables after which one can measure

[1] See also http://www.pggm.nl

the dependence again by linear correlation. This time without any hurdles from the differences in marginal distributions of each asset.

With the introduction to copulas we have come to the aim of this thesis:

The aim of this thesis is to develop a copula model that is able to estimate both the dependence structure between multiple assets adequately and furthermore to model the marginal distribution of each asset in such a way that the asset returns' historical distributions (particularly extreme returns) are well captured within the model.

There are two reasons why we prefer to use this model instead of using the historical data itselves. First, because we do not expect that the historical combination of data points will repeat itself in the future. Moreover, we want a approximation of the historical- dependence and distribution, where small changes are allowed to occur. Secondly, we may want to add Bayesian prior information to this model (or a Markov chain model for example). To be more specific; during a crisis regime we do not know the possibility of a return to the normal regime.

There are several reasons why one may prefer to use a model instead of using directly the empirical distribution of the historical data. First, we do not expect that the historical combination of data points will exactly be repeated in the future. In the future, other returns (and combinations of returns) will be observed that are between or *beyond* historical data. Second, estimated model parameters can provide insight into processes that is not directly observable from historical data. These are the two main reasons for using models in this thesis. Third, the future returns may particularly depend on the latest observed data. This behavior can be incorporated in a model, for example by assuming autoregressive (AR) or generalized autoregressive conditional heteroskedasticity (GARCH) processes, or Markov regime-switching behavior. Without a model, it may be difficult to take such behavior of returns processes into account. Fourth, one may desire to add personal (expert) beliefs to the information stemming from historical data. That is, a Bayesian prior can be specified to incorporate prior beliefs. A formal Bayesian analysis is only possible within the context of a model. Especially, in a Markov-regime-switching, during a new crisis regime we can not estimate the probability of 'escaping' this regime from historical data, as no shift from the regime has yet been observed. That is, without prior information, one would estimate a zero probability of ever leaving the new crisis regime, which may be considered highly undesirable. The approaches mentioned in the latter two reasons will be left as topics for further research.

Nevertheless, in the sequel we will compare the estimated portfolio risk based on different models with results based on the empirical data. The reason for this is that we expect the difference between these results to be moderately small. Extreme differences are considered as an indication that the model is not well specified.

We will investigate different methods to model the dependence and distribution of certain assets. To make sure the model fits the assets well, we will test the model on different aspects. First of all, we will test the fit on standard criteria i.e. log-likelihood, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which will be explained further in the sequel. Secondly, since portfolio managers are very concerned with a detailed description of the tail-risk they face on their investments, we will evaluate some commonly used risk measures. These risk measures include Value-at-Risk, Conditional Value at Risk and Maximum Drawdown. Since each risk measure focuses on somewhat different aspects of the data, we are able to evaluate the models rigorously. At last, we will compare the performance of the different models by using different portfolio optimization techniques. Since PGGM is mainly interested in minimizing their overall risks, we build different optimization approaches to minimize portfolio risk given a certain required return.

In this thesis, we extend the recent academic literature in four ways. First, we build a higher dimensional Skewed t copula which is able to model skewed tail dependence that is not captured by normal or Student t copulas. For this purpose we use a modified version of the Expectation-Maximization (EM) algorithm of Hu (2005) which he has used to fit a multivariate Skewed t distribution. Secondly, we will use the extended EM algorithm of Hoogerheide (2009) to model distributions existing of a mixture of multiple Student t distributions to fit the marginal distributions of the different copulas. Thirdly, we will combine the different copula models and the different modelled marginal distributions to build a model that is best capable of capturing extreme dependence and further the occurrence of extreme returns in the marginal distributions. At last, we will use all different models to test their performance in the field of risk management and portfolio optimization.

The thesis is organized as follows: In section two, we will describe the used data. In the third section, we will make the shortcomings of the linear correlation more explicit. This will be followed by an introduction to copulas. In the fourth section, we will describe the estimation methods and algorithms to model both copulas and marginal distributions. In section five, we will test the different models on their fit. In section six, we will use risk measurement to test further the performance of these models. In section seven, we will evaluate the performance of the different models in the field portfolio optimization. Next, we will shortly evaluate the different model outcomes from their different aspects and conclude. We will finish with suggestions for further research.

2 Data

The data[2] used in this thesis consist of six total return index series taken from the data vendors Bloomberg and Barclays Capital. To investigate the behavior and dependence between different assets, PGGM suggested to use the following different index total return series of different markets as a proxy. They give a good approximation of the overall behavior of these different markets firstly, because these indexes are highly frequently dealt throughout the whole world. Secondly, because these index series include data over a large span of time.

The first series we use is the S&P 500 index, which acts as a proxy for equities. The S&P 500 index is a market-capitalization-weighted (large-cap) index. The S&P 500 includes a representative sample of 500 leading companies in leading industries of the US economy actively traded on the largest stock market companies; the NYSE, Euro Next and the NASDAQ OMX in the United States.

The second index series is the FTSE EPRA/NAREIT North America index presents public Real Estate. This index is derived by the National Association of Real Estate Investment Trusts (REIT) and consists of publicly quoted real estate companies from the United States and Canada.

Thirdly, we use the S&P GSCI index as a proxy for commodities. The S&P GSCI index is a composite index of commodity futures traded on the Chicago Mercantile Exchange. Moreover, it contains commodities from all commodity sectors i.e. energy products (about 75% of which 55% crude oil), industrial metals (7%), agricultural products (13%), livestock products (4%) and precious metals (2%).

Fourthly, as a proxy for the Emerging Markets Debt data, PGGM suggested to use the following two index series. Firstly, we use the JPMorgan Emerging Markets Bonds Index (EMBI). The JPMorgan EMBI Global diversified composite index is a market-capitalization-weighted index containing of U.S. dollar denominated Brady bonds, Eurobonds, traded loans, and local market debt instruments issued by sovereign and quasi-sovereign entities rated single A or lower. Because the origin of this index descends from 1994, we use a second index to fill in the gap from 1990 to 1994. The second index we use is the City Group Brady Bond local currency index which is an U.S. dollar-denominated bond index issued by emerging markets, particularly those in Latin America, and collateralized by U.S. Treasury zero-coupon bonds. Moreover, this index is discontinued since 1994.

The fifth index we use as a proxy for the US High Yield corporate bonds market, is the

[2]More information can be found on http://www2.standardandpoors.com/, http://www.epra.com, http://www.jpmorgan.com, https://live.barcap.com, http://www.ml.com/

	Equities	Commodities	EMD	Real Estate	Treasury	High Yield
Mean	0,00151	0,00126	0,00151	0,00159	0,00105	0,00131
Median	0,0028	0,00156	0,00205	0,00331	0,00122	0,00205
Maximum	0,11407	0,12176	0,09602	0,21948	0,02340	0,05335
Minimum	-0,20030	-0,21090	-0,16264	-0,21057	-0,03081	-0,11517
Std. Dev.	0,02376	0,03033	0,01741	0,02866	0,006021	0,009394
Skewness	-0,82485	-0,8666	-1,92084	-0,70525	-0,50578	-3,04443
Kurtosis	7,76663	4,56837	16,88074	16,09625	1,68778	34,90347
Jarque-Bera	1060,096	227,6575	8643,056	7229,22	114,3835	43954,4
Probability	0	0	0	0	0	0

Table 1: Descriptive statistics of the weekly log (total) return series

Merrill Lynch US High Yield Master II Index.

The sixth and last index we use, is the US Treasury security index. The US Treasury security index is an aggregated index of US-dollar denominated publicly issued Treasury Bonds and Notes with a maturity of at least one year. The Government Bonds and Notes must be rated investment-grade (Baa3/BBB-) or higher.

We use weekly closing prices ranging from 4/6/1990 to 5/29/2009 which we converted to weekly log returns. There are 1000 observations in total. Furthermore, to limit the scope of this thesis, we assume that the weekly data is i.i.d.[3]. Further research can be done to overcome the increasing characteristic hurdles i.e. serial correlation and volatility variation and clustering over time of financial data in higher frequencies such as daily data (see also the section Further Research). Next in table 1 the data characteristics are given. Moreover, since PGGM expects an other long term average than is determined by the sample mean over the given period, we have demeaned the data and added the expected long term return (see also Appendix A for further data details). In the remainder of this thesis when we refer to these mean adjusted data we mean the empirical data.

[3]For monthly data auto-correlations are low as a result of the time averaging of log returns. Weekly data may have some auto-correlation but are preferred over monthly data because otherwise there are too few observations in the dataset to be used.

3 Methods

3.1 Pitfalls of using linear correlation

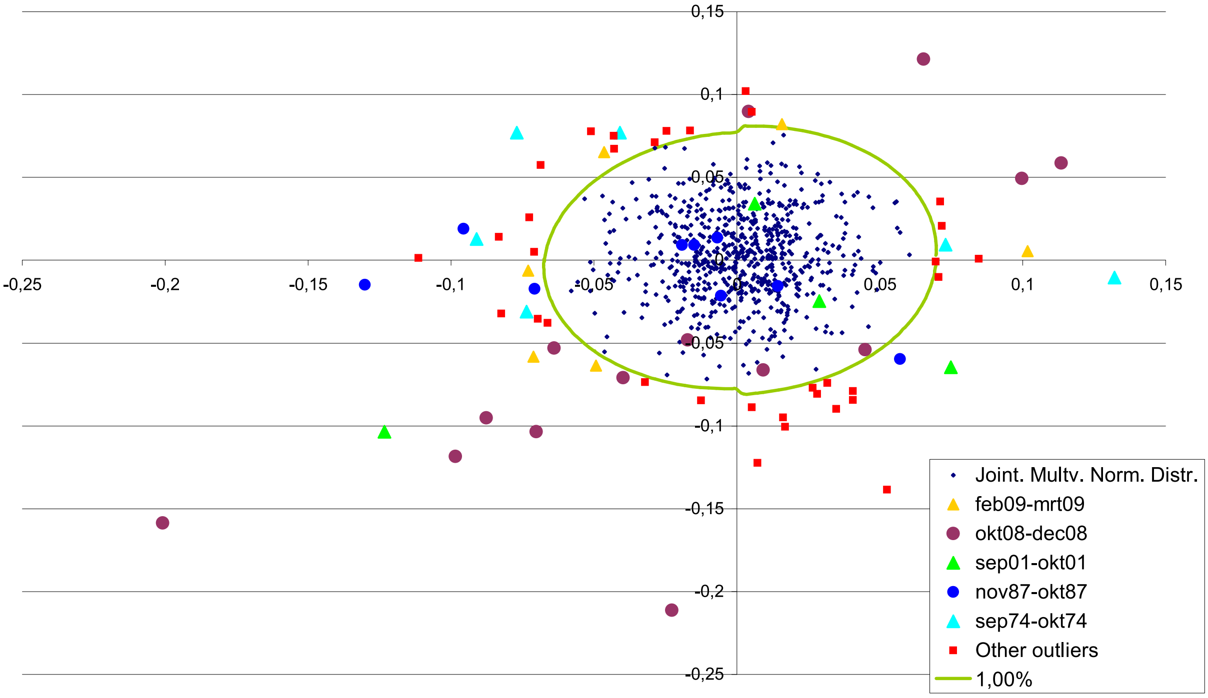
As we have noticed in the introduction of this thesis, the most well known and used dependence measure, linear correlation, is not always as adequate as one might think it is. The underlying assumptions and restrictions of the models using this measure of dependence, may be too restrictive to get a clear understanding. Wrong model assumptions will lead to wrong models, which will give misleading outcomes and wrong decisions. To be more specific, a necessary condition in the use of linear correlation is that the returns used in the model, should be multivariate normally distributed. Moreover, it is not enough to know the *marginal* distributions are normally distributed, but the returns should also be *jointly* multivariately normally distributed. Since it is commonly known that asset returns are not exactly normally distributed but rather contain a negative skewness and excess kurtosis, using linear correlation might be somewhat misleading. In other words when returns are not jointly multivariately normally distributed, the commonly used correlation matrix is no longer adequate to describe the inter-dependence.

This can also be verified from figure 1 where a scatter diagram from the returns series of Equities against Commodities is shown. Inside the circle returns can follow a joint multivariate normal distribution [4]. But it seems to be that the returns exhibit a larger dependence in the 3rd quadrant which can not be captured by a multivariate normal distribution. Alternative distributions where one can use the linear correlation as dependence measure are elliptical distributions (of which the multivariate normal is a special case). However in the case one uses the elliptical Student's t distribution, return series can be uncorrelated, that is have a correlation of zero. However it is not said that the two series are independent, because there still might exist some dependence in the tails of the distribution. This also becomes clear from figure 1. Only in the special case when two variables are jointly normally distributed, uncorrelatedness is equivalent to independence.

Another pitfall of linear correlation is that correlation is not invariant under transformations of the risks. For example, two random vectors X and Y generally do not have the same correlation as log(X) and log(Y). Moreover, the two random vectors X and Y could for example be perfectly positively (negatively) dependent and have a correlation of 1 (-1), but it is particularly not said that the transformed random vector, log(X) and log(Y), have the same perfect dependence. In this section it has become clear that the use of linear correlation is not as straightforward and justified in many cases. For further details and examples of the pitfalls and fallacies of the use of linear correlation see also Embrechts *et al.* (1999, 2002) or Rachev *et al.* (2005).

[4] see Appendix A for a derivation of the normal distributed ellipse

Figure 1: Scatter plot of Equities vs Commodities



3.2 Introduction to copulas

To be able to model the interdependence between return series in an adequate way one might consider the use of copulas. The word 'copula' originates from the Latin noun for a "link or tie" that connects two different things.[5]. The 'copula' is first used in the mathematical literature by Sklar (1959) to describe "a function that links a multidimensional distribution to its one-dimensional margins" Sklar (1996). Throughout the nowadays literature a copula is mainly used to refer to the dependence structure between univariate marginals. In this thesis we will refer to a copula in the same way.

We have already noticed that a great advantage of the copula is the separate modeling of the dependence and the marginal behavior of the univariate series. Another great advantage is that the marginal distributions do not have to be similar to each other so that each marginal distribution can be modeled separately. Most 'straightforward' copulas are the elliptical copulas. A great advantage in the use of elliptical copulas is that one is still able to use the linear correlation as a measure of dependence in a correct way. Most commonly used elliptical copulas are the normal copula (better known as Gaussian copula) and the Student's t copula (see for some applications Cherubini *et al.* (2004), or Kang *et al.* (2007), among others). These two copulas match close with the widely used multivariate normal- and Student's t distributions and are the most easy to use also in higher dimensions.

[5]See also http://dictionary.babylon.com/

However empirical evidence has shown that Gaussian copulas are not able to capture the dependence between extreme - non-normal - events in equity markets as shown by Lognin and Solnik (2001) and Ang and Chen (2002). Moreover, they also show that especially in times of crises - i.e. in times markets show non-normal behavior - portfolio returns exhibit larger conditional correlation during market downturns than during market upturns. This phenomenon is called financial contagion; crisis in one market leads to down movements in other markets as well while these markets should not be correlated based on fundamental linkage (moreover see for example Bae et al. (2003), Hartmann *et al.* (2004) or Rodrigues (2007)). Based on these empirical findings the Student's t copula is also disqualified because although it can be used to model tail dependence it is only good for modelling symmetric tail dependence.

Other (bivariate) copulas like the Gumbel copula, the Clayton copula (Nelsen (2006)) or the Joe-Clayton copula (Patton 2006a) have been introduced to better capture asymmetric dependence structure between return series. Moreover, Gumbel copulas features higher dependence at upper side with positive tail dependence and Clayton and Joe-Clayton Rotated Gumbel's copula features higher dependence at lower side with positive tail dependence. Unfortunately these latter copulas are not well designed for applications in higher dimensions due to computationally difficulties. Different possibilities has been proposed in the recent literature to build a multidimensional copulas. For example Kang (2007) build a n-dimensional hierarchical copula, in which multiple bivariate copulas are used to sum up the n-dimensional hierarchical copulas structure. Moreover, the n-dimensional Student's t hierarchical copulas yield the highest log-likelihood while hierarchical Archimedean copulas, like the Clayton, Gumbel and Frank copula, yields significantly lower log-likelihoods. However the asymmetric correlations between the stock and bond returns is *not* fitted well by both the normal and Student's t copula. Savu and Trede (2006) developed hierarchical Archimedean copulas which render more flexible parameters to characterize dependency between each pair of variables. In their model, each most related pair of variables is modelled by one copula of a particular Archimedean class and then these pairs are nested by copulas as well. Hu (2006) studied dependence structure between a number of pairs of major stock indices by a mixed copula approach that is a weighted sum of three copulas (normal, Gumbel and rotated Gumbel). Dep *et al.* (2006) apply in this way a trivariate hierarchical Archimedean copula structure to model sample selection and treatment effects with applications to family health care demand.

From the studied literature it becomes clear that Archimedean copulas, structured to capture asymmetric dependence, do not allow a higher dimensional framework in a direct way. Moreover, they need a hierarchical structure of multiple bivariate copulas in order to add up to higher dimensions. On the contrary Gaussian and Student's-t copulas do allow for a higher dimensional approach, but they are not capable to fit asymmetric dependence. It is however not said that asymmetric copulas always outperform the elliptical copulas when it comes to goodness-of-fit tests (Kole *et al.* 2006). Still it is a desired tool to be able to model the asymmetric tail dependence in higher dimensions.

By the introduction of the so-called Skewed t distribution of Demarta and McNeil (2005) and the Expectation-Maximization algorithm of Hu (2005), new possibilities have emerged in order to model asymmetric dependence. Moreover, Hu (2005) introduced an efficient manner of estimating the multivariate Skewed t distribution of Demarta and McNeil (2005). In this thesis we will also combine the studies from Demarta and McNeil (2005) and Hu (2005) to set up a new framework to construct a Skewed t copula from a Skewed t distribution. In this way we are able to build a higher dimensional copula model with asymmetric dependence.

3.3 Copula theory

Before we can continue the discussion about which copula performs best in terms of fit and which are preferred over others in some specific applications, we first need to define the copula theory behind these models. The theoretical copula framework, from which all different types of copulas are originated, has been introduced by Sklar (1959).

Definition 3.3.1 (following McNeil, Frey and Embrechts (2005))
A d-dimensional copula $C : [0,1]^d :\rightarrow [0,1]$ is a function which is a cumulative distribution function (cdf) with standard uniform marginal distribution functions. The copula notated as $C(\mathbf{u}) = C(u_1, ..., u_d)$ with the condition of being a distribution has the following properties:

1. $C(u_1, ..., u_d)$ is bounded on $[0,1]$ and increasing in each component u_i
2. $C(1, ..., 1, u_i, 1, ..., 1) = u_i$
3. For all $(a_1, ..., a_d), (b_1, ..., b_d) \in [0,1]^d$ with $a_i \leq b_i$ we have

$$\sum_{i_1=1}^{2} ... \sum_{i_d=1}^{2} (-1)^{i_1+...+i_d} C(u_{1,i_1}, ..., u_{d,i_d}) \geq 0,$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j \ \forall \ j \in 1, ..., d$.

The first property is required for any multivariate df since cdfs are always increasing. The second property is the requirement of uniform marginals. The third property so-called the rectangle property, ensures that if the random vector $(U_1, \ldots, U_d)'$ has cdf C then, $P(a_1 \leq U_1 \leq b_1, ..., a_d \leq U_d \leq b_d)$ is non-negative. These three properties characterize a copula; if a function fulfills them, then it is a copula. Choosing a copula and some marginals and structuring it in the right way, we will get a multivariate distribution function. This is due to the following theorem.

Sklar's Theorem 3.3.2 (Nelson (1999); following Demarta and McNeil (2004)) Let F be an n-dimensional distribution function with margins. Then there exists an n-copula C such that for all $x \in \mathbb{R}^n$:

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)), \tag{1}$$

for some copula C, which is uniquely determined on $[0, 1]$ for distributions F with absolutely continuous margins. As a consequence of representation (1) any copula can be used to join any set of univariate cdfs $F_1, ..., F_d$, to create a multivariate cdf F with marginals $F_1, ..., F_d$. For the purpose of this thesis we concentrate exclusively on random vectors $X = (X_1, ..., X_d)'$ whose marginal cdfs are continuous and strictly increasing. The copula function C of their joint cdfs may be extracted from (2) as follows

$$C(u) = C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)), \tag{2}$$

where the F_i^{-1}'s are the quantile functions of the margins. The copula C can be thought of as the cdf of the component wise probability transformed random vector $F_1^{-1}(u_1), ..., F_d^{-1}(u_d)$. The copula remains invariant under a standardization of the marginal distributions (in fact it remains invariant under any series of strictly increasing transformations of the components of the random vector X. For later estimation purposes it is useful to know that we can obtain the copula density from the representation in (2) by stating that the cdf is differentiable.

$$C(u) = C(u_1, ..., u_d) = \frac{f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot ... \cdot f_d(F_d^{-1}(u_d))}, \tag{3}$$

where f is the joint density and $f_i, i = 1, \ldots, d$ the marginal densities.

3.3.1 The Multivariate Normal distribution and its Copula

In the coming sections, we consider properties of the normal-, Student's t- and Skewed t multivariate distributions, and corresponding copulas. The usefulness of the presented formulas will be made clear in the sequel of this thesis. As we have seen from the multivariate distribution function we can easily define the normal copula function also named as the Gaussian copula.

Definition 3.3.1.1 Multivariate Normal distribution
The d-dimensional random vector $X = (X_1, ..., X_d)'$ is said to have a (non-singular) multivariate Normal distribution with mean vector μ and positive definite matrix Σ, denoted $X \sim N_d(\mu, \Sigma)$, if its density is given by

$$f(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right). \tag{4}$$

Definition 3.3.1.2 Normal Copula
From the previous we can see that the copula of a $N_d(\mu, \Sigma)$ is identical to that of a $N_d(0, P)$ distribution where P is the correlation matrix implied by the dispersion matrix Σ. The unique copula is thus given by

$$C_P^N(u) = \Phi_{P,d}(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)). \tag{5}$$

The density form of the copula is given by

$$c_P^N(u) = \frac{1}{(2\pi)^{N/2}|P|^{1/2}} \exp\left(-\frac{x'P^{-1}x}{2}\right), \tag{6}$$

where $x = \Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)'$ and u from $(0,1)^d$.

3.3.2 The Multivariate Student's *t* distribution and its Copula

Definition 3.3.2.1 Multivariate Student's t distribution
A d-dimensional random vector $X = (X_1, \ldots, X_d)'$ is said to have a (non-singular) multivariate Student's t distribution with mean vector μ, positive definite matrix Σ and with ν degrees of freedom, denoted $X \sim t_d(\mu, \Sigma, \nu)$, if its density is given by

$$f(x) = \frac{\frac{\Gamma(\nu+d)}{2}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{d/2}|\Sigma|^{1/2}} \exp\left(1 + \frac{(x-\mu)'\Sigma(x-\mu)}{\nu}\right)^{-\frac{\nu+d}{2}}. \tag{7}$$

The covariance of the Multivariate Student's t distribution is given by

$$Cov(X) = \frac{\nu}{\nu - 2}\Sigma. \tag{8}$$

From this we can see that the covariance matrix is only defined for $\nu > 2$. The multivariate Student's t distribution belongs to the class of multivariate normal variance mixtures and has the representation

$$X_d = \mu + \sqrt{W}Z, \tag{9}$$

where $Z \sim N_d(0, \Sigma)$ and W is independent of Z and satisfies $\nu/W \sim \chi^2_\nu$; equivalently W has an inverse gamma distribution $W \sim IG(\nu/2, \nu/2)$. That states,

$$W = \frac{1}{W'/\nu},$$

with W' χ^2_ν. For more information about the larger class of elliptically symmetric distributions to which the normal variance mixtures belong, see Fang, Kotz and Ng (1990) or Kelker (1970).

Definition 3.3.2.2 Student's *t* Copula
The Student's t copula differs from the normal copula from the difference in distribution.

The copula of a $t_d(\mu, \Sigma, \nu)$ is identical to that of a $t_d(0, P, \nu)$ distribution where P is the correlation matrix implied by the dispersion matrix Σ. The density of the copula is thus given by

$$c^t_{\nu,P} f(x) = \frac{\frac{\Gamma(\nu+d)}{2}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{d/2} \mid P \mid^{1/2}} \exp\left(1 + \frac{x'P^{-1}x}{\nu}\right)^{-\frac{\nu+d}{2}}, \tag{10}$$

where $x = (t_1^{-1}(u_1), ..., t_d^{-1}(u_d))'$ with t_d^{-1} denotes the quantile function of a standard univariate t_d distribution.

3.3.3 Multivariate Generalized Hyperbolic distributions and its Copula

Definition 3.3.3.1 Gamma distribution

The random variable X is said to have a gamma distribution, written as $X \sim Gamma(a, b)$, if its probability density function is:

$$f(x) = \beta^\alpha x^{-\alpha-1} \exp\left(-\beta/x\right)/\Gamma(\alpha) \quad x > 0, \alpha > 0, \beta > 0. \tag{11}$$

If $X \sim Gamma(a, b)$, then $1/X \sim InverseGamma(a, b)$. The mean and variance of the Gamma distribution are given by:

$$E[X] = \frac{\beta}{\alpha - 1}, \quad \text{if } \alpha > 1, \tag{12}$$

$$Var(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \text{if } \alpha > 2. \tag{13}$$

Definition 3.3.3.2 Generalized Inverse Gaussian distribution (GIG)

The random variable X is said to have a generalized Inverse Gaussian (GIG) distribution, if its probability density function is:

$$h(x; \lambda, \chi, \psi) = \frac{\chi^\lambda(\sqrt{\chi\psi})^\lambda}{2K_\lambda\sqrt{\chi\psi}} x^{\lambda-1} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), \quad x > 0, \tag{14}$$

where K_λ is a modified Bessel function of the second kind [6] . The integral presentation of the modified Bessel function of the second kind with index λ can be found in Barndorff-Nielsen *et al.* (1981) and Abramowitz and Stegun (1965, chapter 9 and 10).

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} \exp(-\frac{x}{2}(y + y^{-1}))dy, \quad x > 0. \tag{15}$$

[6]The Bessel function of the second kind is one of the solutions from the function $K_\lambda(x)$ with index λ which is derived from the *modified Bessels's differential equation*. The modified Bessel functions of the second kind are sometimes called the Basset functions, modified Bessel functions of the third kind (Spanier and Oldham 1987, p. 499), or Macdonald functions (Spanier and Oldham 1987, p. 499; Samko *et al.* 1993, p. 20). The modified Bessel function of the second kind is implemented in Matlab as *besselk*(λ, x). For more information and derivations of the modified bessel function of the second kind, see also the Wolfram Mathworld website of Weisstein (2009) at http://mathworld.wolfram.com/ModifiedBesselFunctionoftheSecondKind.html

In case of the GIG distribution the parameters of the Bessel function satisfy:

- $\chi > 0, \psi \geq 0$ if $\lambda < 0$,
- $\chi > 0, \psi\, 0$ if $\lambda = 0$,
- $\chi \geq 0, \psi > 0$ if $\lambda > 0$.

In short, we write $X \sim N^{-}(\lambda, \chi, \psi)$ if X is GIG distributed. The following formulas for GIG distributed variable X when $\chi > 0$ and $\psi > 0$ may be needed later,

$$E[X^{\alpha}] = \left(\frac{\chi}{\psi}\right)^{\alpha/2} = \frac{K_{\lambda+\alpha}(\sqrt{\chi\psi})}{K_{\lambda}\sqrt{\chi\psi}}. \tag{16}$$

Especially when $\alpha \pm 1$ and 2, and

$$E[\log(X)] = \left.\frac{\partial E[X^{\alpha}]}{\partial \alpha}\right|_{\alpha=0}, \tag{17}$$

where equation (3.4.7) needs to be evaluated numerically. More details about the limiting case of GIG can be found in Eberlein and Hammerstein (2003).

3.3.4 Skewed t distribution

When $\psi = 0$ and $\lambda < 0$, GIG becomes the so-called $InverseGamma$ distribution and it is denoted by $X \sim InverseGamma(-\lambda, \psi/2)$. This limiting case of GIG will lead to a limiting case of generalized hyperbolic distributions called the Skewed t distribution, under the conditions that $\lambda = -\nu/2$, $\chi = \nu$ and $\psi = 0$ so that we have a $IG(\nu/2, \nu/2)$ distribution as we will see soon. The following asymptotic formula is useful when calculating the limiting density of GIG and GH,

$$K_{\lambda}(x) \sim \Gamma(\lambda) 2^{-\lambda-1} x^{\lambda}, \quad as \quad x \downarrow 0. \tag{18}$$

Further we have the useful fact that,

$$K_{\lambda}(x) = K_{-\lambda}(x). \tag{19}$$

Theorem 3.3.4.1 Generalized Hyperbolic distributions (GH)
If the mixing variable $W \sim N^{-}(\lambda, \chi, \psi)$, then the joint density of a d-dimensional generalized hyperbolic distributions in the non-singular case (Σ has rank d) is given by,

$$f(x) = c\frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))(\psi + \gamma'\Sigma^{-1}\gamma)}\right)\exp\left((x-\mu)'\Sigma^{-1}\gamma\right)}{\left(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))(\psi + \gamma'\Sigma^{-1}\gamma)}\right)^{\frac{d}{2}-\lambda}}, \tag{20}$$

where the normalizing constant is given by

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}} \mid \Sigma \mid^{\frac{1}{2}} K_{\lambda}(\sqrt{\chi\psi})}. \tag{21}$$

A proof is given in Hu (2005).

3.3.5 Multivariate Skewed *t* distribution

The Skewed t distribution also belongs to the class of multivariate normal variance mixtures just like the multivariate Student's t distribution (8) except for an additional term for the skewness. A random dx1 vector X has a Skewed t distribution[7] if

$$X = \mu + \gamma g(W) + \sqrt{W}Z. \tag{22}$$

For some function $g : [0, \infty) > [0, \infty)$ and a d-dimensional parameter vector γ. Where $Z \sim N_d(0, \Sigma)$ and W is independent of Z and satisfies $\nu/W \sim X^2_{\nu}$; equivalently W has an $InverseGamma$ distribution $W \sim IG(\nu/2, \nu/2)$. When $gamma = 0$ we again get the elliptical symmetric distribution. However when $\gamma \neq 0$ it turns into a family of skewed, non elliptical asymmetric distributions. If we extend the Generalized Hyperbolic distribution (20) to the multivariate case and continue from section 3.4.1.4 by setting $\lambda = -\nu/2$, $\chi = \nu$ and $\psi = 0$ in $W \sim N^{-}(\lambda, \chi, \psi)$ and replace $\frac{\psi^{\lambda/2}}{K_{\lambda}(\sqrt{\chi\psi})}$ by $\frac{\nu^{\nu/4}}{\Gamma(\nu/2)2^{\nu/2-1}}$ in equation (20) we end up with the density function of the skewed multivariate Student's t distribution given by

$$f(x) = c\frac{K_{\frac{d+\nu}{2}}\left(\sqrt{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma}\right)\exp\left((x-\mu)'\Sigma^{-1}\gamma\right)}{\left(\sqrt{\nu + (x-\mu)'\Sigma^{-1}(x-\mu)\gamma'\Sigma^{-1}\gamma}\right)^{-\frac{d+\nu}{2}}\left(1 + \frac{(x-\mu)'\Sigma^{-1}()x-\mu}{\nu}\right)^{\frac{d+\nu}{2}}}, \tag{23}$$

where the normalizing constant is given by

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}} \mid \Sigma \mid^{\frac{1}{2}} K_{\lambda}(\sqrt{\chi\psi})}. \tag{24}$$

We denote this distribution by $X \sim t_d(\mu, \Sigma, \nu, \gamma)$. From the normal mixture structure of the distribution in (22) we also find that the random vector X conditioned on W has the following distribution

$$X|W \sim N_d(\mu + W\gamma, W\Sigma). \tag{25}$$

We can get the mean and covariance of the multivariate Skewed t distribution given by:

$$E[X] = E\left[E[X \mid W]\right] = \mu + E[W]\gamma = \mu + \frac{\nu}{\nu - 2}\gamma, \tag{26}$$

$$\text{Cov}(X) = E[\text{Var}(X \mid W)] + \text{Var}(E[X \mid W]) = \frac{\nu}{\nu - 2}\Sigma + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\gamma\gamma', \tag{27}$$

[7]Different Skewed t distributions exist. We stick to this definition throughout this thesis.

where the covariance is only defined when $\nu > 4$. This is in contrast with the Student's t distribution where we had the restriction that $\nu > 2$. This is due to the fact that we use the mean variance mixture of equation (22).
Moreover, the $InverseGamma(\alpha, \beta) = InverseGamma(\nu/2, \nu/2)$ density of W is given by

$$\begin{aligned} f(w) &= \beta^{\alpha}\, w^{-\alpha-1} \exp[-\beta/w]/\Gamma(\alpha) \\ &= (\nu/2)^{\nu/2} w^{-\nu/2-1} \exp[-(\nu/2)/w]/\Gamma(\nu/2), \end{aligned} \tag{28}$$

for $w > 0$. For this distribution we have:

$$E[W] = \frac{\beta}{\alpha-1} = \frac{\nu/2}{\nu/2-1} = \frac{\nu}{\nu-2}, \tag{29}$$

$$\mathrm{Var}(W) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}, \tag{30}$$

$$E[\log(W)] = \log(\beta) - \psi(\alpha) = \log(\nu/2) - \psi(\nu/2), \tag{31}$$

where $\psi(x) \equiv d\log(\Gamma(x))/dx$ is the digamma function.
Since $1/W$ has the $Gamma(\alpha, \beta) = Gamma(\nu/2, \nu/2)$ distribution, we have

$$E[1/W] = \frac{\alpha}{\beta} = 1. \tag{32}$$

Definition 3.3.5.1 Skewed t Copula
The Skewed t copula differs from the Student's t copula from the difference in distribution. The Skewed t copula denoted by $C^t_{P,\nu,\gamma}$ is identical to a $t_d(0, P, \nu, \gamma)$ distribution where P is the correlation matrix implied by the dispersion matrix Σ. The univariate margins are given by $t_i(0, 1, \nu, \gamma_i)$ distributions for $i = 1, ..., d$. The unique copula is given by

$$f(x) = c\frac{K_{\frac{d+\nu}{2}}\left(\sqrt{(\nu + x'P^{-1}x)\gamma'P^{-1}\gamma}\right)\exp\left(x'P^{-1}\gamma\right)}{\left(\sqrt{\nu + x'P^{-1}x\gamma'P^{-1}\gamma}\right)^{-\frac{d+\nu}{2}}\left(1 + \frac{x'P^{-1}x}{\nu}\right)^{\frac{d+\nu}{2}}}, \tag{33}$$

where the normalizing constant is given by

$$c = \frac{2^{\frac{2-(v+d)}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}}|P|^{\frac{1}{2}}}. \tag{34}$$

4 Estimation and Simulation

4.1 Introduction to the EM algorithm

To estimate the many parameters of Generalized Hyperbolic distributions in higher dimensions properly, we need to use an efficient framework which is able to estimate the many parameters within an acceptable time period. To be able to do so we can use the set up of the expectation-maximization (EM) algorithm of Dempster et al. (1977).

The EM algorithm of Dempster is an iterative optimization method for finding maximum likelihood (ML) estimates of the parameters of a model which are not directly observable. These parameters which can not directly be observed are the so-called latent variables. The EM algorithm consists out of two steps. The first E-step is to take the expectation of the log-likelihood of the model w.r.t. the latent variables. In the second M-step the parameter values are computed which maximize the expected log-likelihood which was found in the first E-step. The parameters computed in the M-step are now used to determine the expectation of the latent variables in the next E-step. This procedure is iteratively repeated until the value of the log-likelihood of the (parameters of the) model has converged.

Note that to start the algorithm, initial values are chosen for the parameters. Depending on these chosen initial values (and obviously the chosen model and the data), the EM algorithm may stop at a globally suboptimal local optimum.

Much research has been done in the use of EM algorithms for the estimation of generalized hyperbolic distributions (introduced by Barndorff-Nielsen (1977) and the skewed Student's t distribution as a special case (see also Barndorff-Nielsen *et al.* (1981)). Since then it has been a challenging task to be able to estimate the parameters of these Generalized Hyperbolic distributions. Blæsild and Sørensen (1992) used their computer program 'hyp' to estimate the parameters by ML up to the third dimension with a fixed λ. Prause (1999) proposed some extra restrictions to special cases of the GH, i.e. the multivariate hyperbolic and normal inverse Gaussian, to be able to estimate the parameters in a higher multidimensional framework. However the multivariate skewed hyperbolic distribution in higher dimensions than three was computationally still not tractable. Protassov (2004) proposed a ML method based on the EM algorithm of Dempster, Laird and Rubin (1977), to estimate the parameters for the multivariate GH distributions and the multivariate skewed GH distribution in arbitrary higher dimension.

Hu (2005) combined the knowledge of Liu and Rubin (1995), Protassov (2004) and McNeil, Frey and Embrechts (2005) to build a generalized framework for all the different cases of the multivariate GH distributions. He was the first one who was able to come up with

a stable EM algorithm with a fast calibration speed which he applies in the field of finance i.e. risk management, portfolio optimization and the pricing of portfolio credit risk. Further application of the skewed GH distribution can be found in for example Aas and Haff (2006) who demonstrates its superiority over some of its competitors in the field of financial risk management.
Generalized hyperbolic distributions have also been used in a copula framework. Schmid (2003a) used symmetric generalized hyperbolic distributions to create tail independent copulas. Demarta and McNeil (2005) and McNeil, Frey and Embrechts (2005) used the skewed Student's t distribution to build a bivariate Skewed t copula.

As we earlier noted in this paper, we now combine the frameworks of Demarta and McNeil(2005) and Hu(2005) to estimate higher dimensional Skewed t copulas. In the next section we will first describe the general framework of the EM algorithm for estimating multivariate Skewed t distributions. Next in section 4.1.2 and 4.1.3, we will give a short-handed notation of the recipe of the EM algorithm for both the multivariate Skewed t distribution and Student's t distribution respectively.

4.1.1 Estimation of Multivariate Skewed t distributions

For the estimation of the multivariate Skewed t distribution as a special case of the generalized hyperbolic distribution, we use the EM algorithm framework of Hu (2005). Assume that we have i.i.d. data $X_1, ..., X_n$, where $X_i \in \mathbb{R}_d$ and we want to fit these data by multivariate generalized hyperbolic distributions. The parameters are denoted by $\zeta = (\lambda, \chi, \psi, \Sigma, \mu, \gamma)$. The log-likelihood function that we want to maximize is

$$\log L(\zeta; x_1, ..., x_n) = \sum_{i=1}^{n} \log f_{X_i}(x_i; \zeta). \tag{35}$$

However we cannot maximize this function directly i.e. in higher dimensions it would require an immense amount of computing time to reach convergence (if any). Therefore we use the conditional normal distribution representation of the generalized hyperbolic random variable. Now we can estimate most of the parameters (Σ, μ, γ) given that the other parameters (λ, χ, ψ) are already known or assumed to be some value. More specific the setup of the EM algorithm introduces latent mixing variables $w_1, ..., w_n$ which are supposed to be observable at the beginning and are optimized later on. The log-likelihood function which we need to optimize with the included latent variables - the so called quasi or augmented log-likelihood function - is given by

$$\log \tilde{L}(\zeta; x_1, ..., x_n, w_1, ..., w_n) = \sum_{i=1}^{n} \log f_{X_i, W_i}(x_i; \zeta). \tag{36}$$

By the mean-variance mixture definition of generalized hyperbolic distributions, the log-likelihood function can be rewritten as

$$\begin{aligned} \log \tilde{L}(\zeta; x_1, ..., x_n, w_1, ..., w_n) &= \\ \sum_{i=1}^{n} \log f_{X_i|W_i}(x_i|w_i; \mu, \Sigma, \gamma) &+ \sum_{i=1}^{n} \log h_{W_i}(w_i; \lambda, \chi, \psi) = \\ L_1(\mu, \Sigma, \gamma; x_1, ..., x_n|w_1, ..., w_n) &+ L_2(\lambda, \chi, \psi; w_1, ..., w_n), \end{aligned} \quad (37)$$

where $X \mid W \sim N(\mu + W\gamma, W\Sigma)$ and $f_{X|W}(x \mid w)$ is the density of a conditional normal distribution, and $h(W)$ is the density function of a GIG distributed mixing random variable.

We can see from the above equation that the estimation of μ, Σ, γ and λ, χ, ψ can be separated by maximizing L_1, and L_2 respectively. Following the same procedure in the proof of Theorem 3.3.4.1 (GH) the density of conditional normal distribution can be rewritten as,

$$f_{X_i|W_i}(x \mid w) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} w_i^{\frac{d}{2}}} \exp\left((x-\mu)'\Sigma^{-1}\gamma\right) \exp\left(\frac{-p}{2w}\right) \exp\left(-\frac{w}{2}\gamma'\Sigma^{-1}\gamma\right), \quad (38)$$

where

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

From this we can get the log-likelihood function L_1:

$$\begin{aligned} &L_1(\mu, \Sigma, \gamma; x_1, ..., x_n|w_1, ..., w_n) = \\ &-\frac{n}{2} \log |\sigma| - \frac{d}{2} \sum_{i=1}^{n} \log w_i + \sum_{i=1}^{n} (x_i - \mu)'\sigma^{-1}\gamma \\ &-\frac{1}{2} \sum_{i=1}^{n} \frac{1}{w_i} p_i - \frac{1}{2}\gamma'\Sigma^{-1}\gamma \sum_{i=1}^{n} w_i. \end{aligned} \quad (39)$$

From equation (14), we can get the log-likelihood function L_2:

$$\begin{aligned} &L_2(\lambda, \chi, \psi; w_1, ..., w_n) = \\ &(\lambda - 1) \sum_{i=1}^{n} \log w_i - \frac{\chi}{2} \sum_{i=1}^{n} w_i^{-1} - \frac{n\lambda}{2} \log \chi \\ &+\frac{2\lambda}{2} \log \psi - n \log(2K_\lambda(\sqrt{\chi\psi})). \end{aligned} \quad (40)$$

If we however like to maximize the log-likelihood of the following function we need to know the mixing variables $w_1, ..., w_n$ which are not observable. In this stage an iterative procedure is needed, consisting of an estimation (E) step and a maximization (M) step. In this step, the conditional expectation of the augmented log-likelihood function given current parameter estimates and sample data is calculated. Suppose we are at step k, we need to calculate the following conditional expectation and get a new objective function to be maximized,

$$Q(\zeta; \zeta^k) = E[\log \tilde{L}(\zeta; x_1, ..., x_n, W_1, ..., W_n|x_1, ..., x_n; \zeta^k)].$$

In the M step, we maximize the above new objection function to get updated estimates ζ^{k+1}. From equation (39) and (40), we can see that it is equivalent to updating all the w_i, w_i^{-1} and $\log(w_i)$ in the augmented log-likelihood function by their conditional estimates $E[W_i|x_i;\zeta^k]$,$E[W_i^{-1}|x_i;\zeta^k]$, and $E[\log(W_i)|x_i;\zeta^k]$. In this way, $Q(\zeta;\zeta^k)$ is expressed by observations and known conditional expectations so that it can be maximized as we have just done. To calculate those conditional expectations, we need the following conditional density function,

$$f_{W|X}(w|x;\zeta) = \frac{f(x|w;\zeta)h(w;\zeta)}{f(x;\zeta)}.$$

From which we can get in case of the multivariate Skewed t distribution

$$W_i|X_i \sim N^-(-\frac{d+\nu}{2}, p_i+\nu, \gamma'\Sigma^{-1}\gamma), \tag{41}$$

by using equation (16) and (17) and where $\chi = \pi + \nu$, $\psi = \gamma'\Sigma^{-1}\gamma$ and $\lambda = -(d+\nu)/2$. This will give the following equation

$$E[w_i^\alpha] = \left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{\alpha/2} \frac{K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}, \tag{42}$$

with again,

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

From this we can derive the expressions for δ, η and ϵ, which we define as $\delta_i = E[W_i^{-1}|x_i;\zeta]$, $\eta_i = E[W_i|x_i;\zeta]$, $\epsilon_i = E[\log(W_i)|x_i;\zeta]$, so that

$$\eta_i \equiv E[w_i] = \left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{1/2} \frac{K_{-(d+\nu-2)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}, \tag{43}$$

and

$$\delta_i \equiv E[w_i^{-1}] = \left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{-1/2} \frac{K_{-(d+\nu+2)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}. \tag{44}$$

Differentiating (42) w.r.t. α yields:

$$\frac{\partial E[w_i^\alpha]}{\partial \alpha} = \frac{1}{2}\log\left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)\left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{\alpha/2} \frac{K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})} + \tag{45}$$

$$\left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{\alpha/2} \frac{\frac{\partial K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{\partial\alpha}}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}. \tag{46}$$

By using equation (17) we can obtain

$$\xi_i \equiv \left.\frac{\partial E[w_i^\alpha]}{\partial\alpha}\right|_{\alpha=0} = \frac{1}{2}\log\left(\frac{\rho_i+\nu}{\gamma'\Sigma^{-1}\gamma}\right) + \frac{\left.\frac{\partial K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}{\partial\alpha}\right|_{\alpha=0}}{K_{-(d+\nu)/2}(\sqrt{(\rho_i+\nu)\gamma'\Sigma^{-1}\gamma})}. \tag{47}$$

In case of the multivariate Student's t distribution these equations will be simplified since $\gamma = 0$. Then the conditional distribution of the latent variable w_i will have the form

$$W_i|X_i \sim InverseGamma\left(\frac{d+\nu}{2}, \frac{p_i+\nu}{2}\right).$$

From which we can derive the expressions for δ, η and ϵ in the same manner again. That is

$$\delta_i = \frac{\nu+d}{p_i+\nu} \quad (48)$$

$$\eta_i = \frac{p_i+\nu}{\nu+d-2} \quad (49)$$

$$\epsilon_i = \log\left(\frac{p_i+\nu}{2}\right) - \psi\left(\frac{d+\nu}{2}\right). \quad (50)$$

Estimations of Σ, μ, γ are obtained by maximizing L_1. Suppose that the latent mixing variables $w_1, ..., w_n$ are made observable by the previous estimates we can optimize the parameters by taking the partial derivative of L1 with respect to Σ, μ, γ

$$\frac{\partial L_1}{\partial \gamma} = 0,$$
$$\frac{\partial L_1}{\partial \mu} = 0,$$
$$\frac{\partial L_1}{\partial \Sigma} = 0.$$

From the above equation array and (43),(44) and (17) we can get the following estimations

$$\gamma = \frac{n^{-1}\sum_{i=1}^n w_i^{-1}(\bar{x}-x_i)}{n^{-2}(\sum_{i=1}^n w_i)(\sum_{i=1}^n w_i^{-1}) - 1} = \frac{n^{-1}\sum_{i=1}^n \delta_i(\bar{x}-x_i)}{\bar{\delta}\bar{\eta}-1}, \quad (51)$$

$$\mu = \frac{n^{-1}\sum_{i=1}^n w_i^{-1}x_i - \gamma}{n^{-1}(\sum_{i=1}^n w_i^{-1})} = \frac{n^{-1}\sum_{i=1}^n \delta_i x_i - \gamma}{\bar{\delta}}, \quad (52)$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^n w_i^{-1}(x_i-\mu)(x_i-\mu)' - \frac{1}{n}\sum_{i=1}^n w_i\gamma\gamma' =$$
$$\frac{1}{n}\sum_{i=1}^n \delta_i(x_i-\mu)(x_i-\mu)' - \bar{\eta}\gamma\gamma', \quad (53)$$

where

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^n \delta_i, \quad \bar{\eta} = \frac{1}{n}\sum_{i=1}^n \eta_i, \quad \bar{\epsilon} = \frac{1}{n}\sum_{i=1}^n \epsilon_i, \quad (54)$$

To obtain a more clear view on this, note that we can also rewrite the model used for estimating the parameters μ, γ and Σ given the data by a linear regression model.
Following Hoogerheide (2009) conditionally on w_i $(i = 1, \ldots, n)$, we rewrite

$$x_i = \mu + \gamma w_i + \tilde{\varepsilon}_i,$$

with $\tilde{\varepsilon}_i \sim N(0, w_i\Sigma)$ as

$$x_i/\sqrt{w_i} = \mu \frac{1}{\sqrt{w_i}} + \gamma \sqrt{w_i} + \varepsilon_i \quad \text{or}$$

$$\tilde{y}_i' = z_i'\beta + \varepsilon_i', \tag{55}$$

with $\varepsilon_i \sim N(0, \Sigma)$, a Seemingly Unrelated Regression (SUR) model for $\tilde{y}_i = x_i/\sqrt{w_i}$ with the same explanatory variables $z_i = (\frac{1}{\sqrt{w_i}}, \sqrt{w_i})'$ in all equations. The Maximum Likelihood Estimator (MLE) of the $2 \times d$ coefficients matrix $\beta = (\mu\ \gamma)'$, where $\begin{pmatrix} \hat{\mu}_{MLE|w} \\ \hat{\gamma}_{MLE|w} \end{pmatrix} = \hat{\beta}_{MLE|w} = (X'X)^{-1}X'y$ with $X = z_i'I$ and $y = \tilde{y}_i'$, is given by:[8]

$$\begin{aligned}
\hat{\beta}_{MLE|w} &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \frac{1}{w_i} & 1 \\ 1 & \frac{1}{n}\sum_{i=1}^n w_i \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n x_i'/w_i \\ \frac{1}{n}\sum_{i=1}^n x_i' \end{pmatrix} \\
&= \frac{1}{(\frac{1}{n}\sum_{i=1}^n \frac{1}{w_i})(\frac{1}{n}\sum_{i=1}^n w_i) - 1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n w_i & -1 \\ -1 & \frac{1}{n}\sum_{i=1}^n \frac{1}{w_i} \end{pmatrix} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n x_i'/w_i \\ \frac{1}{n}\sum_{i=1}^n x_i' \end{pmatrix} \\
&= \frac{1}{(\frac{1}{n}\sum_{i=1}^n \frac{1}{w_i})(\frac{1}{n}\sum_{i=1}^n w_i) - 1} \begin{pmatrix} (\frac{1}{n}\sum_{i=1}^n w_i)\,(\frac{1}{n}\sum_{i=1}^n x_i'/w_i) - \frac{1}{n}\sum_{i=1}^n x_i' \\ -(\frac{1}{n}\sum_{i=1}^n x_i'/w_i) + (\frac{1}{n}\sum_{i=1}^n \frac{1}{w_i})\frac{1}{n}\sum_{i=1}^n x_i' \end{pmatrix} \\
&= \frac{1}{(\frac{1}{n}\sum_{i=1}^n \frac{1}{w_i})(\frac{1}{n}\sum_{i=1}^n w_i) - 1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n x_i'(-1 + \frac{\bar{w}}{w_i}) \\ \frac{1}{n}\sum_{i=1}^n \frac{1}{w_i}(\bar{x} - x_i)' \end{pmatrix},
\end{aligned} \tag{56}$$

and

$$\begin{aligned}
\hat{\Sigma}_{MLE|w} &= \frac{1}{n}\sum_{i=1}^n (\tilde{y}_i - \hat{\beta}_{MLE|w}'\, z_i)\,(\tilde{y}_i - \hat{\beta}_{MLE|w}'\, z_i)' \\
&= \frac{1}{n}\sum_{i=1}^n \frac{1}{w_i}(x_i - \hat{\mu} - \hat{\gamma}w_i)\,(x_i - \hat{\mu} - \hat{\gamma}w_i)'.
\end{aligned} \tag{57}$$

These obtained MLE estimates are in accordance to the equations (51) to (53) from Hu (2005).

Estimation of λ, χ, ψ is obtained by maximizing L_2 from equation (40). To maximize L_2, we

[8] In the SUR model with the same explanatory variables in all equations, the MLE estimator of the coefficients equals the OLS estimator. (See e.g. Heij *et al.* (2004).)

take the partial derivative with respect to χ and ψ and solve the following equation array,

$$\frac{\partial L_2}{\partial \chi} = 0,$$
$$\frac{\partial L_2}{\partial \psi} = 0.$$

Solving the above equation array leads us to solve $\theta = \sqrt{\chi\psi}$ from the following equation first,

$$n^{-2}\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j^{-1} K_\lambda^2(\theta)\theta + 2\lambda K_{\lambda+1}(\theta)K_\lambda(\theta) - \theta K_\lambda(\theta) = 0. \tag{58}$$

We find θ by zero-finder routine in Matlab, fzero. Once θ is solved, we can get parameters, (χ, ψ)

$$\chi = \frac{n^{-1}\theta \sum_{i=1}^{n} w_i K_\lambda(\theta)}{K_{\lambda+1}(\theta)} \tag{59}$$

$$\psi = \frac{\theta^2}{\chi}. \tag{60}$$

Especially, when $\lambda = -0.5$, we have the normal inverse Gaussian distribution, and we are able to get θ since $K_-\lambda(x) = K_\lambda(x)$ for any λ

$$\theta = \frac{2\lambda}{1 - n^{-2}} \sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j^{-1}. \tag{61}$$

When $\psi = 0$ this will give an unknown distribution, however we will get the Skewed t distribution with ν degrees of freedom, by setting $\lambda = -\nu/2$ and $\chi = \nu$. In what follows ν can be solved from the equation

$$-\psi\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{2}\right) + 1 - n^{-1}\sum_{i=1}^{n} w_i - n^{-1}\sum_{i=1}^{n}\log(w_i), \tag{62}$$

where ψ is the di-gamma function. When we use the replacements for w from equation (43 to 17) we get the following for the Skewed t distribution

$$-\psi\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{2}\right) + 1 - \bar{\epsilon} - \bar{\delta} = 0. \tag{63}$$

Which we can solve for ν, the degrees of freedom parameter.

When the L_1 and L_2 are maximized by updating the parameters for k=1, the first step of the EM algorithm is completed

In the next subsection, we will take a specific look at the 'numerical recipes' for fitting a multivariate Skewed t distribution or a multivariate Student's t distribution, the two GH distributions that we focus on in this thesis.

4.1.2 Recipe of EM algorithm for the multivariate Skewed t distribution:

The EM estimation algorithm of the multivariate Skewed t distribution involves the following steps: Choose initial values for μ_j, σ_j, ν and γ_j for series $(j = 1, 2, ..., J)$. Take as reasonable starting values for μ, σ and γ respectively the sample mean, the sample covariance matrix, $\nu = 30$[9] and a near zero vector for γ.

Then iterate E-step and M-step until convergence, that is until the log-likelihood function is maximized:

- **E-step:** Compute for $j = 1, 2, \ldots, J$:

$$\boxed{\delta_{ij} = \left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{-1/2} \frac{K_{-(d+\nu+2)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

with $\rho_{ij} = (x_i - \mu_j)'\Sigma_j^{-1}(x_i - \mu_j)$.

$$\boxed{\eta_{ij} = \left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{1/2} \frac{K_{-(d+\nu-2)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

$$\boxed{\xi_{ij} = \frac{1}{2}\log\left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right) + \frac{\left.\frac{\partial K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{\partial\alpha}\right|_{\alpha=0}}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

- **M-step:** Compute for $j = 1, 2, \ldots, J$:

$$\boxed{\hat{\gamma}_j = \frac{n^{-1}\sum_{i=1}^n \delta_{ij}(\bar{x}-x_i)}{\bar{\delta}_j\bar{\eta}_j - 1}}_{M-step}$$

$$\boxed{\hat{\mu}_j = \frac{n^{-1}\sum_{i=1}^n \delta_{ij}x_i - \gamma_j}{\bar{\delta}_j}}_{M-step}$$

$$\boxed{\hat{\Sigma}_j = \frac{1}{n}\sum_{i=1}^n \delta_{ij}(x_i - \mu_j)(x_i - \mu_j)' - \bar{\eta}_j\gamma_j\gamma_j'}_{M-step}$$

Solve:

$$\boxed{-\psi\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{2}\right) + 1 - \bar{\epsilon}_j - \bar{\delta}_j = 0}_{M-step}$$

using the Matlab fzero command where $1 - \bar{\epsilon}_j - \bar{\delta}_j = 0$ is constant w.r.t. ν, so that it only has to be evaluated once in the process of solving the equation.

[9] We tried several starting values for ν and obtained robust results. Also the calculation time remains the same.

4.1.3 Recipe of EM algorithm for the multivariate t distribution:

The estimation of multivariate Student's t distribution, i.e. when $\gamma = 0$, can also use above procedures except that in step E-step the estimates of the parameters are obtained by formulas (48 to 51). For maximizing Σ in the M-step one needs to subtract the γ parameters. The EM estimation algorithm of the multivariate Skewed t distribution involves the following steps: Choose initial values for μ_j, σ_j, and ν for series ($j = 1, 2, ..., J$). Take as reasonable starting values for μ and σ respectively the sample mean and the sample covariance matrix, $\nu = 30$[10].
Then iterate E-step and M-step until convergence, that is until the log-likelihood function is maximized:

- **E-step:** Compute for $j = 1, 2, \ldots, J$:

$$\boxed{\delta_j = \frac{\nu+d}{\rho_{ij}+\nu}}_{E-step}$$

with $\rho_{ij} = (x_i - \mu_j)'\Sigma_j^{-1}(x_i - \mu_j)$.

$$\boxed{\eta_j = \frac{\rho_{ij}+\nu}{\nu+d-2}}_{E-step}$$

$$\boxed{\epsilon_j = \log\left(\frac{\rho_i j+\nu}{2}\right) - \psi\left(\frac{d+\nu}{2}\right)}_{E-step}$$

- **M-step:** Compute for $j = 1, 2, \ldots, J$:

$$\boxed{\hat{\mu}_j = \frac{n^{-1}\sum_{i=1}^{n}\delta_{ij}x_i}{\bar{\delta}_j}}_{M-step}$$

$$\boxed{\hat{\Sigma}_j = \frac{1}{n}\sum_{i=1}^{n}\delta_{ij}(x_i - \mu_j)(x_i - \mu_j)'}_{M-step}$$

Solve:

$$\boxed{-\psi\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{2}\right) + 1 - \bar{\epsilon}_j - \bar{\delta}_j = 0}_{M-step}$$

using the Matlab fzero command where $1 - \bar{\epsilon}_j - \bar{\delta}_j = 0$ is constant w.r.t. ν, so that it only has to be evaluated once in the process of solving the equation.

The estimation of the multivariate normal distribution involves no latent data. In this case, the sample mean and covariance matrix are simply the ML estimates.

[10]We tried several starting values for ν and obtained robust results. Also the calculation time remains the same.

4.1.4 Simulation from multivariate distributions

To compare the outcomes of the different risk measures we simulate 10.000 returns from the different multivariate distributions. To simulate from the multivariate Student's t distribution we use the formula[11] $X_d = \mu + \sqrt{WZ}$ from (9), where the second term is given by $\frac{Z_1,...,Z_d}{\sqrt{x^2(\nu)/\nu}}$ with $Z \sim N_d(0, \sigma)$.

For the multivariate Skewed t distribution we use the multivariate normal variance mixture representation of (22). Note that when $\gamma = 0$ this will reduce to the student t distributions.

4.2 Specifying the marginal distributions

An important feature of a copula is that the marginal distributions do not need to be in any way similar to each other, nor is the choice of copula restricted by the choice of the marginal distribution. This flexibility makes copulas a potentially useful tool for building econometric models.

To be able to model the dependence structure in a copula as good as possible, one needs to model the different marginals as accurately as possible. The univariate marginals of the copula can be derived in three different ways. The first method is to fit parametric distributions for each margin. Parametric methods for univariate copulas are considered by for example Joe (1997). Most used parametric models are the normal and Student's t distribution for modeling more extreme movements. But although the latter distribution does account for more extreme movements it is also a symmetric distribution like the normal, which does not take negative skewness (and possibly also excess kurtosis) well into account. To overcome these pitfalls Hu (2005) used different kinds of hyperbolic distributions. He found that the Skewed t distribution fits best in terms of log-likelihood since this distribution accounts best for the negative extreme events. On the other hand the drawback of this method is that the incorporated high kurtosis can also account for such extreme values that might not be realistic.

The second way to derive the copula marginals, is by using a non-parametrical distribution i.e. an empirical distribution function. The great advantage of this model is its perfect fit over the sample. The main disadvantage is the fact that the distribution is bounded by the interval of the most extreme returns movements in the past. When it comes to simulating, no larger extreme movements can appear in the simulation set than have been occurred in the past. Given an vector of $X_{i,1}, ..., X_{n,d}$ i.i.d. data with for the d^{th} data vector containing

[11] To simulate from the multivariate normal and Student's t distribution we use respectively the mvnrnd and mvtrnd command function in Matlab. Since the mvtrnd command function does not allow for a covariance matrix (only a correlation matrix), need to scale the returns with the Cholesky factorization decomposition of the covariance matrix.

n data points. The j^{th} marginal empirical cdf F_j is given by

$$\hat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{(X_{i,j}\leq x)}. \quad (64)$$

The pseudo-sample from the copula is then constructed by forming vectors $U_1, ..., U_n$ where

$$\hat{U}_i = (U_{i,1}, ..., U_{i,d})' = (\hat{F}_1(X_{i,j}), ..., \hat{F}_d(X_{i,d}))'. \quad (65)$$

Note that even if the original data vectors are i.i.d., the pseudo-sample data are dependent, because the marginal F_j estimates are constructed from all of the original data vectors through the univariate samples $X_{i,1}, ..., X_{n,d}$. Note also that division by $n+1$ in (64) keeps transformed points away from the boundary of the unit cube.

Many other distributions are proposed to model the extreme behavior of returns. In this way mixtures of parametrically and non-parametrically methods are combined to account for all characteristics of returns data in a distribution. This third method approximates the body of the distribution using a normal, student's t or empirical distribution. Both the tails are next modeled using generalized Pareto distribution and used in extreme value theory like Davison and Smith (1990) and Bouye (2005) did. Hotta *et al.* (2006) considered the use of extreme value theory to model only the left tail of the distribution. Moreover, they modeled the marginal distributions by the generalized Pareto distribution in the left tail and by empirical distribution otherwise. Sun *et al.* (2008) used so-called Levy processes and motions to model the tail of the distribution.

Yet another possibility is to use a mixture of parametric distributions to be able to obtain the best fit. Firstly, because by using a mixture of parametric distributions, a good approach of the fit of an empirical distribution is guaranteed. Secondly, because we expect more extreme market movements to happen in the future, we need to specify heavier tails than is possible by the empirical distribution. Moreover, by applying this method, misspecification of the model and their parameters are also less likely to happen, because the model is less restricted than normal parametric models. Hamilton (1994) introduced therefore the use of a mixture of normal distributions to model. To account for even more heavy tails Hoogerheide (2009) used a mixture of Student's t distributions. We will follow these latter methods of both Hamilton and Hoogerheide. Besides these parametric mixture models the Skewed t distribution is also used to fit the marginals. Moreover, an EM algorithm is used to find the mixture model that best fits the marginals.

4.2.1 Mixture of Normal distributions (Hamilton (1994))

Suppose the distriution for a d-dimensional vector x_i is a mixture of normal distributions. Then

$$x_i \sim N(\mu_j, \Sigma_j),$$

if observation i belongs to regime j ($j = 1, 2, \ldots, J$) with J the number of regimes. Define z_i as the latent J-dimensional vector indicating from which regime the observation x_i stems: if observation i stems from regime j, then $z_{ij} = 1$, $z_{ik} = 0$ for $k \neq j$. Now the model is rewritten as:

$$x_i \sim N(\mu_j, \Sigma_j) \quad \text{if } z_i = e_j \;\; (j = 1, 2, \ldots, J),$$

with e_j the j-th column of the $J \times J$ identity matrix. Here: $\Pr[z_i = e_j] = \pi_j$ with $\pi_j \geq 0$ ($j = 1, 2, \ldots, J$), $\sum_{k=1}^{J} \pi_k = 1$.
The Likelihood (for observed data x) is given by[12]:

$$p(x|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \left[\sum_{j=1}^{J} \Pr[z_i = e_j|\theta] \; p(x_i|z_i = e_j, \theta) \right] = \prod_{i=1}^{n} \left[\sum_{j=1}^{J} \pi_j \; \text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \right],$$

with $x = \{x_i | i = 1, 2, \ldots, n\}$, θ containing parameters μ_j, Σ_j, π_j ($j = 1, 2, \ldots, J$). The complete data likelihood (for observed and latent data) is:

$$p(x, z|\theta) = \prod_{i=1}^{n} p(x_i, z_i|\theta) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left[p(x_i|z_i = e_j, \theta) \; \Pr[z_i = e_j|\theta] \right]^{z_{ij}} = \prod_{i=1}^{n} \prod_{j=1}^{J} \left[\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j \right]^{z_{ij}},$$

with $z = \{z_i | i = 1, 2, \ldots, n\}$. Hence the complete data log-likelihood is:

$$\log p(x, z|\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} \left\{ z_{ij} \; \log \left[\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \right] + z_{ij} \log(\pi_j) \right\}. \tag{66}$$

The function

$$k(z_i|x_i, \theta) = \prod_{j=1}^{J} \left[\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j \right]^{z_{ij}},$$

is a kernel[13] of a probability function of a multinomial distribution for $z_i = (z_1, z_2, \ldots, z_J)$ given x_i and θ, with probabilities

$$\tilde{\pi}_j = \frac{\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k,\Sigma_k)}(x_i) \; \pi_k},$$

since this has density

$$\prod_{j=1}^{J} \left[\frac{\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k,\Sigma_k)}(x_i) \; \pi_k} \right]^{z_{ij}} = \prod_{j=1}^{J} \left\{ \left[\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j \right]^{z_{ij}} \; \frac{1}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k,\Sigma_k)}(x_i) \; \pi_k}^{z_{ij}} \right\}$$

$$= \prod_{j=1}^{J} \left\{ \left[\text{pdf}_{N(\mu_j,\Sigma_j)}(x_i) \; \pi_j \right]^{z_{ij}} \right\} \left(\frac{1}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k,\Sigma_k)}(x_i) \; \pi_k} \right)^{\sum_{j=1}^{J} z_{ij}}$$

$$= k(z_i|x_i, \theta) \; \frac{1}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k,\Sigma_k)}(x_i) \; \pi_k},$$

[12] Pr denotes the discrete probability. p denotes the density

[13] A kernel is a function proportional to a density or probability function

with $\sum_{j=1}^{J} z_{ij} = 1$, where the last factor does not depend on z_i.

The EM algorithm for the mixture of normal distributions proceeds as follows:

E-step for z_i given x_i and θ: Define $\tilde{z}_{ij} \equiv E[z_{ij}|x_{ij}, \theta]$ with

$$E[z_{ij}|x_{ij}, \theta] = \frac{\text{pdf}_{N(\mu_j, \Sigma_j)}(x_i)\ \pi_j}{\sum_{k=1}^{J} \text{pdf}_{N(\mu_k, \Sigma_k)}(x_i)\ \pi_k}.$$

M-step: Maximize expectation of (66):

$$\sum_{i=1}^{n} \sum_{j=1}^{J} \tilde{z}_{ij}\ \log \left[\text{pdf}_{N(\mu_j, \Sigma_j)}(x_i)\right] + \sum_{i=1}^{n} \sum_{j=1}^{J} \tilde{z}_{ij} \log(\pi_j).$$

M-step for π_j $(j = 1, \ldots, J)$: Maximize

$$\log L_2 \equiv \sum_{i=1}^{n} \sum_{j=1}^{J} \tilde{z}_{ij} \log(\pi_j),$$

w.r.t. π_j (subject to $\sum_{j=1}^{J} \pi_j = 1$ and $\pi_j \geq 0$ $(j = 1, \ldots, J)$). This yields:

$$\hat{\pi}_j = \frac{1}{n} \sum_{j=1}^{n} \tilde{z}_{ij}.$$

M-step for μ_j, Σ_j $(j = 1, \ldots, J)$: Maximize

$$\log L_1 = \sum_{i=1}^{n} \sum_{j=1}^{J} \tilde{z}_{ij}\ \log \left[\text{pdf}_{N(\mu_j, \Sigma_j)}(x_i)\right],$$

w.r.t. μ_j, Σ_j $(j = 1, \ldots, J)$. This yields

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n} \tilde{z}_{ij}\ x_i}{\sum_{i=1}^{n} \tilde{z}_{ij}} = \sum_{i=1}^{n} z^*_{ij}\ x_i,$$

with $z^*_{ij} \equiv \frac{\tilde{z}_{ij}}{\sum_{i=1}^{n} \tilde{z}_{ij}}$; and

$$\hat{\Sigma}_j = \sum_{i=1}^{n} z^*_{ij}\ (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)'.$$

The estimates $\hat{\mu}_j$ and $\hat{\Sigma}_j$ are the sample mean and covariance matrix in case z_i were known, with $\tilde{z}_i = E[z_i|x_i, \theta]$ substituted for z_i.

4.2.2 Mixture of Student's t distributions of Hoogerheide (2009)

Suppose the distribution for a d-dimensional vector x_i is a mixture of Student's t distributions. Then

$$x_i \sim t(\mu_j, \Sigma_j, \nu),$$

if observation i belongs to regime j $(j = 1, 2, \ldots, J)$ with J the number of regimes... [For estimation of this model see Hoogerheide (2009).]

4.3 Estimation of the dependence of the Copula

When we have obtained the optimal marginal distribution functions for each series separately from the section 4.2, we transform marginal data into data with univariate marginal distributions by using the cumulative distribution function. Once this is done we are able to estimate the copula. We follow Demarta and McNeil (2004) and Embrechts (2001) to estimate the dependence structure.

4.3.1 Normal and t Copula

The estimation of a normal copula is particularly easy. The mean and covariance matrix of the set of vectors $\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\right)$ are ML estimates. In case of estimating a t copula we can use the density taken from definition 3.3.2. The density of the t copula is given by

$$c^t_{P,\nu}(u) = \frac{f_{P,\nu}\left(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)\right)}{\prod_{i=1}^d f_\nu t_\nu^{-1}(u_i)}, \quad u \in \ (0,1)^d, \tag{67}$$

where $f_{P,\nu}$ is the joint density of a $t_d(0, P, \nu)$ distributed random vector and f_ν is the density of the univariate standard t distribution with ν degrees of freedom.
Since we know the density function we now can use Maximum Likelihood to estimate the parameters P and ν. The log-likelihood function can now be written as

$$\log L(P, \nu; U_1, ..., U_d) = \sum_{i=1}^n \log c^t_{P,\nu}(U_i), \tag{68}$$

with respect to P and ν, where $c^t_{P,\nu}$ denotes the density of the t copula in (67).
In case of the t copula in lower dimensions (that is lower than 3), we can use the BFGS Quasi-Newton method[14]to optimize the parameters 'directly' to obtain the maximum of the log-likelihood of the copula. Another possibility is to use the simplex method of Nelder and Mead (1965) for unconstrained nonlinear optimization.[15] Although the simplex method is

[14]We use the fminunc Matlab command, with medium scale optimization method, which uses the BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure. This quasi-Newton method uses the BFGS formula (see Fletcher (1970) or Shanno (1970) for more details) for updating the approximation of the Hessian matrix. Fminunc finds the minimum of an unconstrained multivariate function min f(x) where x is a vector and f(x) is a function that returns a scalar. In order to maximize our problem we set f(x) to -f(x) to get the desired result. A drawback of fminunc is that the command might only give a local solution. See also the Matlab help for more information. To overcome the optimizing issues of Demarta and McNeil (2004) in maximizing the correlation matrix P in higher dimensions, we transformed the symmetric correlation matrix P into a vector by taking the values of the upper triangular matrix of the Cholesky decomposition of P. Combining this with the degrees of freedom we get a vector of start values for a fast iteration process of the optimization function.

[15]The fminsearch Matlab command uses the simplex search method of Lagarias *et al.* (1998). This is a direct search method that does not use numerical or analytic gradients as in fminunc. If n is the length of x, a simplex in n-dimensional space is characterized by the n+1 distinct vectors that are its vertices. In two-space, a simplex is a triangle; in three-space, it is a pyramid. At each step of the search, a new point in or near the

generally less efficient than the BFGS Quasi-Newton method for problems of order larger than two, it can be more robust when the problem is highly discontinuous or multimodal. This is in contrast to BFGS Quasi-Newton method which demands that the function to be minimized should be continuous. A drawback of both methods is that they might only give local solutions. Especially in higher dimensions both methods become very inefficient in estimating the large numbers of parameters. This is why we have to set up an other framework to optimize the log-likelihood of the t copula.

current simplex is generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. fminsearch only minimizes over the real numbers, that is, x must only consist of real numbers and f(x) must only return real numbers. When x has complex variables, they must be split into real and imaginary parts. (See also the Matlab help or Lagarias *et al.* (1998) for more information about the properties behind the method.

4.3.2 EM-within-Simplex for the Student's t copula

We use a combination of the simplex search method and the EM algorithm to optimize the log-likelihood of the t copula. This is needed because the EM algorithm requires 'fixed data' as an input (and not transformed data depending on copula parameters) in order to fit a multivariate distribution. In case of the Student's t copula, we fit a multivariate distribution to the transformed data (transformed to have $t(0,1,\nu)$ marginals), instead of the original data. The transformed data depend on ν, so that we can only apply our EM algorithm for optimization of μ and Σ given ν. We optimize the simplex method in order to optimize ν. Within the simplex method a second optimization function is placed which uses the EM-algorithm to maximize the log-likelihood by optimizing the parameters μ and Σ given ν. The algorithm now exists in choosing initial values for ν, Σ_j, $(j = 1, 2, \ldots, J)$. E.g. μ_j = sample mean, Σ_j = sample covariance matrix times factor (smaller/larger than sample covariance matrix), $\nu = 6$.
Then apply the EM-within-simplex until convergence, that is until the log-likelihood function is maximized and ν is optimized:

- **simplex algorithm:** maximize concentrated log-likelihood of ν. Each evaluation of the concentrated log-likelihood[16] of ν requires optimization of μ and Σ given ν with an EM-within-simplex step:
- **EM-within-Simplex step:** maximize log-likelihood given ν by iteration of the following steps:
  - **E-step:** Compute for each variable $j = 1, 2, \ldots, J$:

$$\boxed{\delta_j = \frac{\nu+d}{\rho_{ij}+\nu}}_{E-step}$$

with $\rho_{ij} = (x_i - \mu_j)'\Sigma_j^{-1}(x_i - \mu_j)$.[17]

$$\boxed{\eta_j = \frac{\rho_{ij}+\nu}{\nu+d-2}}_{E-step}$$

$$\boxed{\epsilon_j = \log\left(\frac{\rho_i j+\nu}{2}\right) - \psi\left(\frac{d+\nu}{2}\right)}_{E-step}$$

  - **M-step:** Compute for each $j = 1, 2, \ldots, J$:

$$\boxed{\hat{\mu}_j = \frac{n^{-1}\sum_{i=1}^{n}\delta_{ij}x_i}{\delta_j}}_{M-step}$$

$$\boxed{\hat{\Sigma}_j = \frac{1}{n}\sum_{i=1}^{n}\delta_{ij}(x_i - \mu_j)(x_i - \mu_j)'}_{M-step}$$

After the optimal value for ν is found, we can use the EM-algorithm in a final run to maximize the log-likelihood further by optimizing the Σ and μ given the optimized ν.

[16]The concentrated log-likelihood of a subset of hte parameters is the log-likelihood where all other parameters have been optimized conditionally upon the values of the subset.

[17]x_i are data 'within the copula', transformed to the marginal $t(0,1,\nu)$ distribution.

4.3.3 Skewed t Copula

Next we continue with estimating the Skewed t copula. If we construct the density of the Skewed t copula from section 3.4.3 we get the following:

$$c^{t}_{P,\nu,\gamma}(u) = \frac{f_{P,\nu,\gamma}\left(t^{-1}_{\nu,\gamma_1}(u_1), ..., t^{-1}_{\nu,\gamma_d}(u_d)\right)}{\prod_{i=1}^{d} f_i t^{-1}_{\nu,\gamma_i}(u_i)}, \quad u \ \in \ (0,1)^d \tag{69}$$

Where $f_{P,\nu,\gamma}$ is the joint density of a $t_d(0, P, \nu, \gamma_d)$ distributed random vector and f_i is the density of the univariate standard Skewed t distribution $t_i(0, 1, \nu, \gamma_i)$.
Now two problems arise; first, we can not compute the inverse of a Skewed t cumulative distribution function using an efficient built-in standard function, since it does not exist.[18] Second, using a grid is also problematic because then you have to choose the cut off points arbitrarily to define near zero and near one inverse cdf values. The interval and the width of the subintervals (i.e. the number of grid points) of the Skewed t cumulative distribution function is then arbitrarily chosen which we want to avoid. This is why we approximate the inverse cdf by simulating 100000 $t_d(0, 1, \nu, \gamma)$ variables.

Another problem arises when we try to maximize the log-likelihood function of (69) because of the addition of the γ parameters, the total number of parameters in higher dimension is too large to be optimized at once. To overcome these problems we apply again a combination of the simplex search method and the EM algorithm of the Skewed t distribution from section 4.1.2 to estimate the parameters of the Skewed t copula. Before we can estimate the dependence structure we first need to estimate the uniform marginals which exists in the following steps:

- First, we use the methods described in section 4.2 to obtain the best marginal distributions with parameters estimated parameters μ, Σ, ν and γ
- Second, we simulate 100.000 draws from an $t_d(0, 1, \nu, \gamma)$ distribution using equation (22) with the estimated parameters from step 1 for each marginal $i = 1, ..., d$
- Next, we match margins from step 2 with the corresponding quantile functions using equation (64)

Now the dependence structure of the copula can be estimated.

[18] No function skewtinv (like the tinv function) exists in Matlab.

4.3.4 EM-within-Simplex algorithm for the Skewed t copula

We use a combination of the simplex search method and the EM algorithm to optimize the log-likelihood of the Skewed t copula. The simplex search method maximizes the log-likelihood by optimizing the parameter ν and γ. Within the simplex method a second optimization function is placed which uses the EM-algorithm to maximize the log-likelihood by optimizing the parameters μ and Σ. The algorithm now exists in choosing initial values for ν, Σ_j, $(j = 1, 2, \ldots, J)$. E.g. μ_j = sample mean, Σ_j = sample covariance matrix times factor (smaller/larger than sample covariance matrix), $\nu = 6$ and $\gamma = 15$.
Now apply the EM-within-simplex algorithm until convergence, that is until the log-likelihood function is maximized and ν and γ is optimized:

- **simplex algorithm:** concentrated loglikelihood of ν and γ. Each evaluation of the concentrated log-likelihood of ν and γ requires optimization of μ and Σ given ν and γ within an EM-within-Simplex step:
- **EM-within-Simplex step:** maximize log-likelihood given ν and γ by iterating the following steps:
  - **E-step:** Compute for each variable $j = 1, 2, \ldots, J$:

$$\boxed{\delta_{ij} = \left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{-1/2} \frac{K_{-(d+\nu+2)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

with $\rho_{ij} = (x_i - \mu_j)'\Sigma_j^{-1}(x_i - \mu_j)$.[19]

$$\boxed{\eta_{ij} = \left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right)^{1/2} \frac{K_{-(d+\nu-2)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

$$\boxed{\xi_{ij} = \frac{1}{2}\log\left(\frac{\rho_{ij}+\nu}{\gamma'\Sigma^{-1}\gamma}\right) + \frac{\left.\frac{\partial K_{-(d+\nu)/2+\alpha}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}{\partial\alpha}\right|_{\alpha=0}}{K_{-(d+\nu)/2}(\sqrt{(\rho_{ij}+\nu)\gamma'\Sigma^{-1}\gamma})}}_{E-step}$$

  - **M-step:** Compute for $j = 1, 2, \ldots, J$:

$$\boxed{\hat{\mu}_j = \frac{n^{-1}\sum_{i=1}^{n}\delta_{ij}x_i - \gamma_j}{\delta_j}}_{M-step}$$

$$\boxed{\hat{\Sigma}_j = \frac{1}{n}\sum_{i=1}^{n}\delta_{ij}(x_i - \mu_j)(x_i - \mu_j)' - \bar{\eta}_j\gamma_j\gamma_j'}_{M-step}$$

After the optimal value for ν and γ is found, we can use the EM-algorithm in final run to maximize the log-likelihood further by optimizing the Σ and μ given the optimized ν and γ.

[19]With x_i $t_j(0, 1, \nu, \gamma_j)$.

4.3.5 Simulation from the copulas

To simulate from a normal-, Student's t- or Skewed t copula, we generate a normal-, Student's t- or multivariate Skewed t distributed random vector $X \sim N_d$ (using the normal variance mixture construction of (9) or (22) in case of the Student's t- or multivariate Skewed t distribution) and then return a vector $U = (F(X_1), ..., F(X_d))$', where F denotes a standard univariate distribution. Schematically we have the following order:

- Simulate from the particular multivariate distribution
- Transform to simulation to $u[0, 1]$ variables
- Transform the $u[0, 1]$ variables to multiple univariate marginals again.

4.3.6 Evaluation of the copula performance

To measure and compare the performance of the different copula models, we will calculate the log-likelihood of each model. We also use Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC and BIC criteria are given as follows:

$$AIC = -2\log(L) + 2 \cdot \#\, parameters, \tag{70}$$

$$BIC = -2\log(L) + \log(\#\, observations) \cdot \#\, parameters. \tag{71}$$

The choice of the mixture model and the number of parameters to fit the marginals will be based on the AIC criterion.[20] To compare the impact of the crisis, we have evaluated two periods; the first period will run from Jan 1990 to May 2009 and the second period will end before the crisis began, that is from Jan 1990 until December 2007. We will first take a look at the fit of the marginals. Thereafter we will compare the performance of the different (copula) models. Finally we will say something about the simulation results.

[20]Since the number of parameters is restricted by the number of chosen components the number of parameters will grow very fast by adding more components to the mixture models, the BIC criteria is therefore not our desired criteria.

5 Model Results

In this section we will discuss the performance of the copulas and their marginals in terms of goodness of fit. In section 5.1 we will discuss the marginal fit of the copulas. In section 5.2 we will discuss the copula fit and the performance of the multivariate models. In section 5.3 we will discuss the scatterplots of the different models. We will conclude in section 5.4.

We have compared the performance of nine different models for the six series over two different data periods. The first set of three models consists of the multivariate normal- the Student's t- and the Skewed t models. The second set of models consists of the normal- Student's t- and Skewed t copula model with marginals of a mixture with normally distributed components. The third set of models exist of the normal- Student's t- and Skewed t copula model with marginals of a mixture with Student's t distributed components.

5.1 Marginal fit

The performance in terms of log-likelihood, AIC and BIC of each marginal distribution is shown in table 6 and 7 in the appendix [21]. The fit of the different marginals over the period until May 2009 is shown in figure 13 and 14. The parameter estimates of the different marginals over the period until May 2009 are shown in table 8 and 9. By taking a look at the time period until December 2007 we can see that based on the performance of the AIC we should choose 4 out of 6 marginals from the combination of multiple normal distributions. On the other hand by looking at the log-likelihood it is noticeable that only the marginal distribution containing only 1 component of the normal distribution performs worse than all other combinations of mixtures of that specific marginal. Moreover, except from taking one normal distribution as marginal, the performance of all other marginals differ only slightly from each other based on the AIC criterion. For the data until May 2009 it becomes clear that choosing a mixture of multiple marginals is - except for High Yield - always preferable over choosing only one distribution to fit each marginal (based in the AIC criterion). Still the table shows that the addition of the mixture of Student's t components offers the best performance in 4 out of 6 marginals over this data period. Only EMD remains committed to the Skewed t distribution in comparison with the other data period. Moreover, only EMD, Real Estate and High Yield change in numbers of components for the normal mixture case, that is 7 to 5 components, 3 to 2 components and 6 to 3 components respectively. Other mixture models do not change over the different periods.

[21] The results in table 6 and 7 show sometimes a decreasing log-likelihood by an increasing number of components, which should be impossible because the models are nested models. This however might be due to the fact that either a local solution is found or because the optimization procedure has converged preliminary. Either way the returned solution turns out to be an illusional convergence.

5.2 Copula fit

To obtain a good overview of the dependence in each copula, a graphical representation of the densities of the normal-, Student's t and Skewed t copula is given in figure 3 to 7 in the appendix. The figures all show the dependence of a two-dimensional (bivariate) copula for different parameter values.

Several phenomena are reflected by the figures. For large degrees of freedom ν ($\nu = 50$), we obviously observe that results for the Student's t copula are similar to results for the normal copula, since the Student's t distribution tends to the normal distribution for growing ν. This also holds for the Skewed t distribution, since for growing ν the variable W tends to the constant 1.

For correlation parameter $\rho = 0$, we observe independent uniform variables in case of the normal copula, whereas the Student's t and Skewed t copulas (with low ν) exhibit tail dependence. The reason is that large draws of W cause extreme values for both variables. The parameter ρ can be roughly interpreted as the level of dependence or co-movement in general: for higher ρ the points in the scatter plots come closer to the diagonal line of the case of perfect correlation. Note that also the normal copula can produce data sets with extreme losses occurring together, but only at the cost of a high dependence in all periods.

Only for moderate or low ν (that is, for truly non-normal copulas) we can observe tail dependence that differs from dependence in general. For the Student's t copula, the picture always shows that the situation is symmetric for large profits and losses. The Skewed t copula allows for asymmetric shapes. For $\gamma = 0$, the Skewed t copula obviously reduces to the Student's t copula. For $\gamma \neq 0$, asymmetric shapes are observed (i.e. as long as ν is small, since otherwise the Skewed t distribution reduces to a normal distribution for any γ).

If only one element of the 2-dimensional vector γ is non-zero, in our example the second element is -1, then this, roughly stated, allows the second variable to break free from the general dependence structure: extremely low values (losses) are also observed when the first variable has high values (profits), even if the correlation parameter $\rho = 0.9$.

If both elements of γ have the same sign, in our example both negative, then this, roughly stated, allows the variables to have a negative tail dependence that exceeds the level of general dependence; there is negative tail dependence even if the correlation parameter $\rho = 0$. The larger (in absolute sense) the γ gets, the stronger the negative tail dependence gets (*in addition to* the dependence in a Student's t copula with the same ρ and ν).

To measure the fit of the different copula models we take a look at their log-likelihoods.

Table 2: Log-likelihood of different models for data Jan90-Dec07

Model	LogL	AIC	BIC
multivariate normal	1,62E+08	3,76E+01	6,37E+01
multivariate t	1,67E+08	3,96E+01	6,66E+01
multivariate skewed t	**1,67E+08**	**5,16E+01**	**8,44E+01**
normal cop with mixnormal marg	3,33E+06	**-6,37E+06**	-5,65E+06
t cop with mixnormal marg	3,90E+06	-7,47E+06	**-6,70E+06**
st cop with mixnormal marg	**3,94E+06**	-7,44E+06	-6,38E+06
normal cop with optmix marg	3,34E+06	-6,38E+06	-5,66E+06
t cop with optmix marg	3,90E+06	-7,49E+06	**-6,71E+06**
st cop with optmix marg	**4,02E+06**	**-7,60E+06**	-6,54E+06

We have also used the AIC and BIC criteria to adjust the models for their number of parameters. In table 3 and 2 the performance in terms of log-likelihood and both criteria AIC and BIC are shown. Note that the log-likelihoods for the different copulas shown in these tables are only based on the joint copula structure (that is, the log-likelihoods of the different marginals are not added). Based on the total log-likelihood (by addition of the log-likelihood of the marginals) the skewed t copula always outperforms the other (copula) models. The AIC criteria shows the same result over the period until May 2009. Based on the BIC criteria the t copulas always should be chosen. The multivariate Skewed t distribution is the best model among the multivariate models over all criteria.

To be able to measure the tail dependence among the different variables within the Skewed t copulas, we look at the γ's. The γ of each marginal is shown in table 10 and 11 in the appendix. From the tables it becomes clear that Treasury is not dependent with the other marginals in the tail of the distribution. Especially the Skewed t model with marginals of a mixture of normals shows a large dependence between Real Estate and High Yield. From table 14 it becomes clear that all of the models have some difficulties to model the right means. The copula models with the normal mixture marginals show least deviations from the mean. The Skewed t copula models which performed best on the log-likelihood show here some poor fittings, especially for EMD for the period until May 2009. When we take a look at the σ's of the different models it becomes clear that the different models incorporate the standard deviation from the mean pretty well. That is except for the copulas with the optimal mixtures of marginals, which show a too large standard deviation for some marginals. This also becomes immediately clear from the fit in term of skewness and kurtosis shown in table 16 and 17 in the appendix. Remarkable is the fact that these same copula models have great difficulties in fitting the skewness and kurtosis of Real Estate over the whole data period. They show a much better fit over the period until December 2007. Also the multivariate models show the worst performance this time.

Table 3: Log-likelihood of different models for data Jan90-May09

Model	LogL	AIC	BIC
multivariate normal	1,66E+08	3,76E+01	6,46E+01
multivariate t	1,76E+08	3,95E+01	6,75E+01
multivariate skewed t	**1,76E+08**	**5,15E+01**	**8,55E+01**
normal cop with mixnormal marg	4,57E+06	**-8,84E+06**	-8,10E+06
t cop with mixnormal marg	5,69E+06	-1,11E+07	**-1,03E+07**
st cop with mixnormal marg	**5,82E+06**	-1,12E+07	-1,01E+07
normal cop with optmix marg	4,60E+06	-8,89E+06	-8,16E+06
t cop with optmix marg	5,72E+06	-1,11E+07	**-1,03E+07**
st cop with optmix marg	**5,88E+06**	**-1,13E+07**	-1,02E+07

5.3 Scatterplots

An other way to look at the performance and fit of the different models is to look at the scatterplots of different combinations of marginals. Since it is impossible to show all dependence between the marginals at the same time, we plotted the simulated log returns for Equities vs. Commodities, EMD vs. Treasury and Real Estate vs. High Yield. The outcomes are shown in figure 15 to 23 in the appendix. By comparing the different multivariate models we can see the grow in the number of outliers as we move from multivariate normal to Student's t and from Student's t to Skewed t. Also the heaviness on the (onesided) tails becomes clearer from the scatterplots of the multivariate Skewed t distributions. From the copula models with mixtures of normals the difference in fit becomes somewhat more clear than from the multivariate models. The normal copula with normal marginals performs worst in terms of tail dependence (which is by definition the case). Especially the dependence between Real Estate and High Yield is not covered at all. The t copula and the Skewed t copula show much better performance in this manner. By moving from marginals with mixtures of normals to marginals with mixtures of Student's t components or single Skewed t distributions we can see much difference. That is the presence of outliers becomes clear in the case of the scatterplot between Real Estate vs. High Yield. On the other hand the tail dependence is better shown in case of the t copula and Skewed t copula.

5.4 Conclusions on copulas

Based on all foregoing results it is hard to obtain an unambiguous and straightforward model choice. Actually we can draw some different conclusions from the different outcomes. By looking purely at the copula models we have to make a choice which marginals to use and which copula dependence one wants to use. The optimal choice of marginals based on the AIC criteria is to use a mixtures marginal model of t components or to use a single Skewed

t distribution. On the other hand by comparing the outcomes of the scatterplots it is probably not desirable to choose these marginals since the outliers are of great magnitude. This is however not due to the way these models are estimated because these mixtures with t components do perform well on log-likelihood and AIC. It is rather the consequence of using the normal variance mixtures at the simulation part. The impact of the additional χ^2 distributed random variables is disastrous since these models show bad performance in terms of the main characteristics that is mean, std, skewness and kurtosis. Moreover, based on these outcomes one should prefer the marginals of mixtures of normals.

To determine the dependence structure of the copula, we have looked at the performance in terms of log-likelihood, AIC and BIC. Based on the log-likelihood and AIC the Skewed t copula, with marginals of a mixture of Student's t components or single Skewed t distribution, is overall best. The second best performing is the Skewed t copula with normal mixture marginals, but due to the conclusions drawn from the marginal fit and scatters we prefer the latter.

6 Risk measures

Now that we know the goodness of fit of each discussed model, we will apply these different models in the field of risk management. To be able to compare the performance of these different models in terms of risk, we will first have to come up with a workable definition of risk. Thereafter we will be able to define a proper risk measure. Next we will describe some widely used risk measures. In the following sections we will evaluate the performance of our models with the use of these risk measures. In the last section we will conclude on risk measures.

Risk is defined as the quantifiable likelihood of loss or less-than-expected returns. Since unexpected losses occur quit often on financial markets, risk managers needs to be aware of the risks they run on their portfolios or on their entire asset exposure. To be able to define a good risk measure Artzner *et al.* (1997, 1999) introduced a list of properties a good risk measure should have. When all these requirements are met the risk measure is called coherent.

Definition 6.1: A risk measure $\rho : G \rightarrow R$ is called coherent if it is:

1. translation invariant: $\rho(X + ra) = \rho(X) - a$, $\forall\ X, Y, a \in \mathbb{R}$
2. sub-additive: $\rho(X + Y) \leq \rho(X) + \rho(Y)$, $\forall\ X, Y, X + Y \in \mathbb{R}$
3. positive homogeneous: $\rho(tX) = t(\rho(X)$, $\forall\ \rho \geq 0$ and for all X and $t \in \mathbb{R}$
4. monotonous: $\rho(X) \geq \rho(Y)$, if $X \leq Y\ \forall\ X, Y \in \mathbb{R}$

where G is a set of real-valued R random variables describing possible solutions. X and Y are random variables and a is the total rate of return on a risk free investment.

Property 1 states that adding or subtracting cash to a position X and investing it in a reference instrument against the risk free rate reduces the risk by that same amount. Property 2 reflects the idea of a diversified portfolio which will reduce the risk compared to separate investments. Or to put it in the words of Artzner (1997): "A merger does not create extra risk". Property 3 states that the risk of a position depends linearly on the size of the position (when we do not account for liquidity issues). The last property states that all positions that lead to higher losses in every state of the world require more risk capital.

One of the most well known risk measures is Value at Risk (VaR) introduced by J. P. Morgan in 1994. VaR is an estimate of the maximum loss that the portfolio could incur within a given time period and at a given confidence level. In other words, if the 99% daily VaR is 'X', then there is a 99% chance that the portfolio will not lose more than amount X

in the following day. VaR at 95% and 99% confidence levels are the most commonly used tail-risk measures. The definition of VaR is as follows:

$$VaR_\alpha(L) = q_\alpha(F_L) = F_L^{-1}(\alpha), \tag{72}$$

where $q_\alpha(F_L)$ is the quantile of the cdf of L and F_L^{-1} the inverse of F_L.

Because of its simplicity and its great suitability the Basel committee on Banking Supervision suggested in 1995 to use a 10 day Value at Risk at a 99% level which banks need to reserve for the possible occurrence of unexpected great losses of their asset exposure. Despite the major use of VaR, it is also widely known that it suffers from two inadequacies. First, it provides no information about losses at points beyond the confidence threshold. That is, if a loss beyond the VaR does occur, no information is available of how large that loss might be, which is of great concern for portfolio managers. Second, VaR does not completely capture the benefits of diversification in non-normal scenarios which is empirically confirmed by i.e. Hyung and De Vries (2005). This issue is put forward in the subadditivity property of a coherent risk measure defined by Artzner *et al.* (1999) as we have seen earlier. As Artzner *et al.* (1999) shows, VaR is not additive in some circumstances.

The most commonly used risk measure that resolves both the above problems with VaR is called Conditional VaR (CVaR) better known as Expected Shortfall (ES). Moreover on the coherence of ES see for example Artzner *et al.* (1999)or Acerbi and Tasche (2002), among others. The expected shortfall is the expected loss given that the VaR is exceeded. It is defined as follows:

$$CVaR_\alpha(L) = E[L|L > F_L^{-1}(\alpha)], \tag{73}$$

where $q_\alpha(F_L)$ is the quantile of the cdf of L and $F_L^{-1}(\alpha)$ the inverse of F_L. In other words for a 99% weekly VaR, the corresponding "99% weekly ES" would be the average loss in 1% of the weeks where losses greater than VaR have occurred.

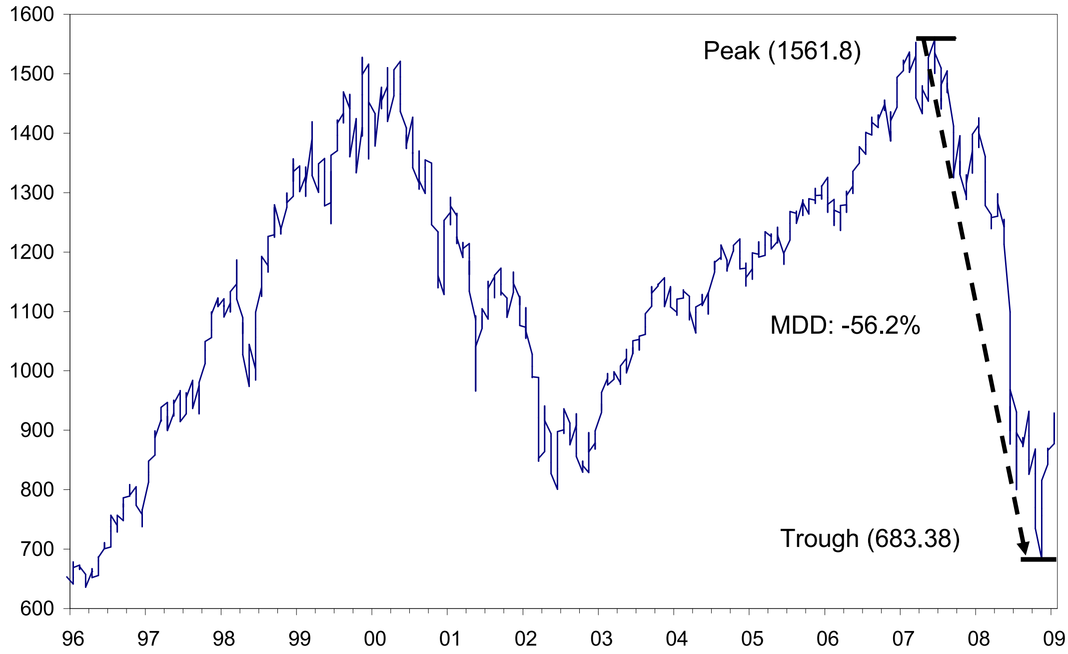
Yet another coherent risk measure in which PGGM is particularly interested, is the (Expected) Maximum Drawdown (MDD). This risk measure is not as widely known as VaR or ES, but is sometimes preferred over VaR en ES, see for example León and Laserna (2008). The Maximum Drawdown is a totally different risk measure since it focuses mainly on the performance of the relating series over time. MDD is defined as the maximum sustained percentage decline (peak to through) which has occurred in an asset or - portfolio investment within a certain period of time. The formal definition is as follows:

$$MDD_{[0,T]} = min\left(\frac{V_T - V_{max}}{V_{max}}, MDD_{[0,T-1]}\right), \tag{74}$$

where V is the index value and where $V_{max} = max\{V_{[0,T-1]}\}$. As can be seen from the figure below the Maximum Drawdown for the S&P 500 index from October 2007 until March 2009

was: -56.2%.
Furthermore, for an application of MDD see for example Rotundo and Navarrab (2007), who study the MDD during speculative bubbles. For a discussion about the coherence properties of MDD see Chekhlov *et al.* (2003) or Leal and Mendes (2005).

Figure 2: Maximum Drawdown of S&P500 index



6.1 Evaluation of the risk measures

In the next section we will evaluate the different risk measures over the nine different models describes in the section 3. We have tested these nine different models over two time periods. The first period contains weekly data from Jan 1990 until May 2009. The second time period contains weekly data from Jan 1990 until December 2007 to measure the impact of the financial crisis which started during 2008. We have calculated the historical VaR, ES and MDD over al the series and compared these with the 'historical' VaR, ES and MDD over the simulated series of all different models. It should be noted that we expect to gain some larger Maximum Drawdowns than the real data shows since we have simulated 10.000 log returns which returns in ten times more data points than the real data has. Since there are more simulated data points there is also a major chance in simulating a longer series of negative returns. Furthermore, to be able to compare the models in terms of portfolio risk, we consider a equal weighted portfolio of the different assets.

6.2 Risk measure Results

In the appendix in table 18 to 27 an overview is given of the different risk measures of the assets and the performance of the different models. By looking over the period until December 2007 at the 5% VaR in table 18, we do not see very much difference between the various models. All models seem to slightly underestimate the risk at the same degree. At the 1% VaR in table 19 the difference between the multivariate models becomes somewhat more clear although even the multivariate skewed t distribution still undervalues the risk. The same can be said for the copula models. Only Treasury (and EMD in lesser extent) are well being estimated by the different models. In case of 5% CVaR (table 20) the multivariate normal- and t distribution show most strain in estimating the risk. The copula models with the optimal mixtures components for the marginals show some overestimating results for EMD. But over the whole the copula models show similar results. When it comes to 1% CVaR (table 21) all models show large problems in estimating the risk of -14,7% of Real Estate. The multivariate Skewed t distribution has the best approximation with -8,3%. The copula models with the optimal marginals overvalue the risk for EMD in a major way. The multivariate normal distribution shows most trouble in estimating the risk. Also the multivariate Student's t distribution has difficulties. When we look at the performance of the different models for the Maximum Drawdown risk measure we find large deviations from the real data for some models. For Equities and Treasury the multivariate Student's t distribution shows the best results this time. No model is able to recall the small Drawdown of Treasury. Except for the copula models with optimal marginals, the Maximum Drawdown is best estimated for EMD. Moreover,especially at this asset the copula model with optimal marginals shows the worst performance. No model is able to reproduce the Max Drawdown of High Yield.

Next we consider the whole time period until May 2009. Table 23 shows the performance of the different models in terms of 5% Var. From this table we can see that for the 5% VaR most models do not have difficulties in reproducing the risk. This time the multivariate normal distribution even overvalues the risk of most assets somewhat. The models show a better reproduction of the risk of the total sample compared to the model estimated with data until December 2007. At 1% VaR (table 24) the multivariate models stay somewhat behind on the copula models. This is especially the case for Real Estate. At 5% CVaR (table 25) most models perform alright except for the multivariate normal distributions which undervalue the risk a little bit. For EMD and Real Estate the copula model with optimal marginals overvalue the risk a bit. Looking at 1% CVaR in table 26, the copula models with the normal mixture marginals show the best results in replicating the risk performance over the different assets. For EMD and Real Estate the multivariate models - and especially the normal and Student's t distribution - undervalue the risk, where the copula models with optimal marginals overvalue the risk somewhat. Looking at the performance of the different models for the Maximum Drawdown again it can be said that these models show better results than the same models over the period until December 2007. This time the risk for High Yield is better approximated by some copula models. On the other hand looking at EMD we can see again that this risk is mayorly overvalued by the copula models with optimal marginals.

In table 4 and 5 results are shown from the equal weighted portfolio. Looking at the results for the models based on the period until December 2007 it becomes clear that most models have some difficulties in reproducing the portfolio mean and standard deviation, that is except for the multivariate normal distribution and the normal copula model with a mixture of normal marginals. In terms of standard deviation also the Student's t and Skewed t copula model are doing well. However when we take a look at the performance on the different risk measures, we can see some contrary results. Except for the 5% VaR, the multivariate normal distribution is not capable of reproducing the more extreme risk types. The multivariate Student's t distribution performs really well this time. The multivariate Skewed t distribution overvalues the different risks somewhat. All different copula models reproduce the different risks very well. Only the Student's t- and the Skewed t copula models with mixtures of Student's t distributions or single Skewed t distributions overvalue the MDD quit a bit. This is however due to some extreme outliers in the simulated samples. Overall it can be said that the copula models with normal mixtures of marginals perform clearly the best.

Looking at the results on the period until May 2009 in table 5, the different models show more deviation from each other. This time only the Student's t copula with normal mixtures of marginals is able to reproduce the portfolio mean. The other models deviate somewhat more. When we look at the 5% VaR we can see that most models (with inclusion of the multivariate normal distribution) overestimate the risk a bit. Only the multivariate Student's t distribution reproduces the same risk as the empirical sample portfolio. When we look further to the other risk measures, the underperformance of the multivariate normal- and Student's t distribution becomes clear. Moreover,based on all remaining risk measures both models are not capable of reproducing the different risks. The multivariate Skewed t distribution and the different copula models are well able to reproduce the 1% VaR. Also the 5%CVaR does not cause much difficulties for these models. At 1%CVaR the normal- and Student's t copula with mixtures of normals shows some difficulties. Finally by looking at the MDD risk measure, we see the most deviated results. No model is really able to capture this risk. Almost all models underestimate the risk. Except the copula models with mixtures of Student's t distributions or single Skewed t distributions which overvalue the MDD again with big numbers.

Table 4: Results for equal weighted portfolio of different models for data Jan90-Dec07

Model	Mean	Std	5% VaR	1% VaR	5% CVaR	1% CVaR	MDD
Empirical data	0,00137	0,0088	1,3%	2,2%	1,9%	3,1%	17,0%
multivariate normal	0,00139	0,0088	1,3%	1,9%	1,7%	2,2%	15,3%
multivariate t	0,00200	0,0093	1,2%	2,2%	1,9%	2,9%	16,6%
multivariate skewed t	0,00141	0,0097	1,4%	2,7%	2,3%	4,1%	25,2%
normal cop with mixnormal marg	0,00136	0,0090	1,3%	2,3%	2,0%	3,0%	15,7%
t cop with mixnormal marg	0,00140	0,0091	1,3%	2,4%	2,0%	3,1%	17,1%
st cop with mixnormal marg	0,00143	0,0091	1,3%	2,6%	2,1%	3,6%	16,4%
normal cop with optmix marg	0,00126	0,0096	1,4%	2,4%	2,1%	3,8%	18,9%
t cop with optmix marg	0,00121	0,0110	1,4%	2,5%	2,3%	4,7%	30,4%
st cop with optmix marg	0,00119	0,0160	1,4%	2,6%	2,5%	5,5%	71,6%

Table 5: Results for equal weighted portfolio of different models for data Jan90-May09

Model	Mean	Std	5% VaR	1% VaR	5% CVaR	1% CVaR	MDD
Empirical data	0,00137	0,0121	1,4%	3,7%	2,9%	6,3%	43,0%
multivariate normal	0,00133	0,0120	1,9%	2,7%	2,4%	3,0%	26,8%
multivariate t	0,00212	0,0108	1,4%	2,6%	2,2%	3,9%	23,7%
multivariate skewed t	0,00146	0,0116	1,6%	3,3%	2,8%	5,5%	25,3%
normal cop with mixnormal marg	0,00135	0,0114	1,7%	3,4%	2,7%	4,6%	26,4%
t cop with mixnormal marg	0,00137	0,0116	1,6%	3,5%	2,8%	5,2%	22,6%
st cop with mixnormal marg	0,00132	0,0119	1,7%	4,0%	3,1%	6,1%	25,7%
normal cop with optmix marg	0,00119	0,0175	1,6%	3,5%	3,2%	7,1%	72,6%
t cop with optmix marg	0,00109	0,0290	1,6%	3,1%	3,5%	9,1%	93,7%
st cop with optmix marg	0,00122	0,0191	1,6%	3,6%	3,4%	8,5%	84,7%

6.3 Conclusions on risk measures

Based on the data set until December 2007 the diverse models show not much difference. In case of 5% VaR, 1% VaR and 5% CVaR the various models all show underperformance. The differences among the diverse models are minimal. Based on 1% CVaR and Max Drawdown there is not one model really outperforming the others. Overall can be said that the multivariate normal- and Student's t model perform worst of all models. Over the data until May 2009 the diverse models all perform better. However also this time the multivariate normal- and Student's t model perform the worst in the sense that they undervalue the risks most of the time. In case of the normal distribution this is to be expected because of the thin tails by definition. On the other hand the copula models with marginals of a mixture of the Student's t or Skewed t distributions sometimes overvalue the risk in case of the 1% CVaR and the Maximum Drawdown. In total it can be said that the copula models with the mixtures of normals perform best. Especially the Skewed t copula performs well, although the differences with the Student's t copula are small. From all the results it should be noted that there is not one model that always outperforms the other. Moreover,even the best model still shows sometimes major under- and overestimations.

By looking at the performance of the different equal weighted portfolios, the addition of a copula becomes really clear. All copula models show a great improvement in estimating the risk compared to the multivariate distributions. Moreover,among different multivariate distributions only the multivariate Skewed t distribution reproduces the risks. From the results it also becomes immediately clear that no model is really able to reproduce the MDD. Moreover,most models underestimate this risk. On the other hand the copula models with mixtures of Student's t distributions or single Skewed t distributions overvalue the MDD drastically. Overall can be said that the Skewed t copula with marginals of a mixture of normal distributions is best able to reproduce most different risk measures. On the other hand is should be noted that an equal weighted portfolio is generally not the most lucrative or efficient portfolio to hold on. This subject will now be discussed in the next section.

7 Portfolio optimization

Before we come to an overall conclusion we will also apply our different models in the field of portfolio optimization. We will describe different methods to minimize the risk of a portfolio at a required portfolio return by optimizing the weights of the assets within the portfolio. In section 7.2 we will discuss the optimized portfolios for each model. In section 7.3 we will conclude on portfolio optimization.

The investors' belief that higher expected returns come at the cost of greater risk has led to the introduction of the efficient frontier line by Markowitz (1952). He defined the efficient frontier as the maximum return given a level of risk or alternatively the minimized level of risk that must be taken at a required return. Markowitz suggested to use the standard deviation or variance as measure of risk. Variance however is an inappropriate measure of risk in case the returns are not normally distributed i.e. when returns are highly skewed and fat tailed. As we have seen earlier an appropriate risk measure i.e. a coherent risk measure is Expected Shortfall or Maximum Drawdown (see León and Laserna (2008) for more information about the MDD case) which will give a good indication of the risk. On the other hand Embrechts, Mc Neil and Straumann (2001) show that if the underlying distribution to model returns is elliptical (that is spherically symmetric), then the Markowitz minimum variance portfolio, for a given return, will be the same as the optimized portfolio. Moreover,the optimized portfolio will be obtained by minimizing any other risk measure which satisfies the positive homogeneous and translation invariant restrictions from Artzner *et al.* (1999). Embrechts *et al.* (2001) formulated the following proposition to summarize this:

Proposition 7.1: Efficient Frontier for Elliptical distributions
Suppose X is elliptically distributed and all univariate marginals have finite variance. For any $r \in \mathbb{R}$, let

$$Q = \left(Z = \sum_{i=1}^{d} w_i X_i | w_i \ \in \ \mathbb{R}, \sum_{i=1}^{d} w_i = 1, E[Z] = r \right), \tag{75}$$

be the set of all fully invested portfolios with weights w_i and expected return r. Then for any positively homogeneous, translation invariant risk measure ρ,

$$argmin_{Z\ Q} \rho(Z) = argmin_{Z\ Q} \sigma_Z^2. \tag{76}$$

In other words only a difference in the underlying distribution will lead to a different portfolio allocation at a given return. Hu and Kercheval (2007) propose to use the t distribution or the slightly better Skewed t distribution to obtain a better fit of the data. In this thesis we include portfolio return simulations from different multivariate distributions from foregoing sections. That is normal-, Student's t-, Skewed t- distributions. Furthermore, we use the simulations from different copula functions (i.e. the normal-, t- and Skewed t copula with marginals of a mixture of distributions).

In the following we construct the efficient frontier for different relevant risk measures including VaR, ES(CVaR) and MDD for different confidence levels. Portfolio optimization using the Drawdown is also considered in Chekhlov *et al.* (2005) and Magdon-Ismail and Atiya (2004). Since VaR is nonsmooth, nonconvex (and often not continuously twice differentiable) and therefore cannot be optimized using computationally efficient methods such as linear programming, many studies have been done to find an appropriate optimization algorithm. Many papers extract an empirical VaR from other efficient frontiers. For example Uryasev and Rockafellar (1999) developed a novel linear programming formulation to minimize Expected Shortfall. Mausser and Rosen (1999) and Clement (2003) construct the efficient frontiers for two risk measures (ES and Expected Regret (ER)) which are both tractable by solving linear programming problems, from which they extract the empirical VaR.

Another possibility to optimize a complex function like VaR is by using nonlinear optimization tools. In section 4.3 we already used the BFGS Quasi-Newton method and the simplex method of Nelder and Mead (1965). Both methods are for unconstrained nonlinear optimization. To be able to easily incorporate restrictions for the function to optimize we also can use the sequential quadratic programming (SQP) method[22] for constrained optimization. The efficient frontier is now obtained by the optimizing the following objective function with the corresponding constraints:

$$\begin{aligned} &\text{Min}_w f(\mathbf{w}) \\ &-\mathbf{w}'\boldsymbol{\mu} \leq -r, \\ &\sum_{i=1}^{d} w_i = 1, \\ &0 \leq w_i \leq 1 \end{aligned} \tag{77}$$

where $\mu = (\mu_1, ..., \mu_d)'$ is the return for asset i and r is the minimum required portfolio return. $\mathbf{w} = (w_1, ..., w_d)$ is the weight of the total invested capital in asset $i = 1, ..., d$, where we assume the initial invested capital is 1. Furthermore, we do not allow for short selling which is incorporated in the last constrained. The objective function to be optimized, $f(\mathbf{w})$, is the portfolio variance, VaR, ES (CVaR) or MDD. Due to liquidity constraints and policy

[22]We use the fmincon Matlab command, which uses the sequential quadratic programming (SQP) method. In this method, the function solves a quadratic programming (QP) subproblem at each iteration. An estimate of the Hessian of the Lagrangian is updated at each iteration using the BFGS formula. This is then used to generate a QP subproblem whose solution is used to form a search direction for a line search procedure. Shortcomings of the fmincon function is that the function to be minimized and the constraints must both be continuous. Just like fminsearch and fminunc also fmincon might only give local solutions. When the problem is infeasible, fmincon attempts to minimize the maximum constraint value. The objective function and constraint function must be real-valued; that is, they cannot return complex values. An overview of SQP is found in Fletcher(1980), Gill *et al.* (1981), Powell (1983) and Schittkowski (1983). See also the Matlab help or http://www.mathworks.com for more information.

it is often also not possible or desirable to invest more than a certain amount $\mathbf{a}$ in an asset. To incorporate this constraint we can easily change the last equation in (77) to $0 \leq w_i \leq \mathbf{a}$, where $\mathbf{a} = (a_1, ..., a_d)'$ is a vector of upper weights.
When we choose to use unconstrained nonlinear optimization methods we need to incorporate the constraints into the objective function. The efficient frontier is obtained by the optimizing the following objective function with the incorporated constraints:

$$\text{Min}_{\tilde{w}} F(\tilde{w}) \tag{78}$$
$$F(\tilde{w}) = f(\mathbf{w}) + 1_{(r-\mathbf{w}'\boldsymbol{\mu}<0)} \cdot c \cdot \exp(r - \mathbf{w}'\boldsymbol{\mu}) + \sum_{i=1}^{d} 1_{(\mathbf{w}-\mathbf{a}<0)} \cdot d \cdot \exp(a_i - w_i)$$
with
$$w_i = \frac{\exp(\tilde{w}_i)}{\sum_{i=1}^{d-1} \exp(\tilde{w}_i) + 1} \text{ for weight of asset } i = 1, ..., d-1$$
$$w_d = \frac{1}{\sum_{i=1}^{d-1} \exp(\tilde{w}_i) + 1}$$

where $f(w)$ is the same function as we have seen earlier, c and d are constants which are large enough to never end up with a solution that does not meet the requirements. $\mathbf{w} = (w_1, ..., w_d)$ are the weights as before. Furthermore, we have the temporary weights $\tilde{w} = (\tilde{w}_1, ..., \tilde{w}_d)$ which are the weights that will be optimized by iteration.[23] The indicator function incorporates the restrictions in the form of a penalty function. By adding the exponent, the penalty increases as the solution draws away from the optimal solution. The first part of the penalty is defined by the indicator function $1_{(r-\mathbf{w}'\boldsymbol{\mu}<0)}$ which starts to work when the solution returned by the optimizer is less than the required return. If we want to restrict the function of investing more than a certain amount $\mathbf{a}$ in an asset, we have to add the second indicator function $1_{(\mathbf{w}-\mathbf{a}<0)}$ into the objective function. To impose the no-short selling constraint we added the exponent in the last two equations in (78) to prevent the weights of becoming negative.

7.1 Evaluation of the portfolio optimization

This section we will evaluate the models of foregoing sections in portfolio optimization. Moreover,we will use the objective function (77) to minimize the risk at a required return. The risk is defined as one of the foregoing risk measures from section 5, that is the standard deviation, 5% VaR, 1% VaR, 5% ES, 1% ES and MDD. We will evaluate each model to the degree that they are able to reproduce the optimal portfolio of the sample data. All models are estimated over the total sample period until May 2009.

[23]Note that the real weights $\mathbf{w} = (w_1, ..., w_d)$ are obtained from the temporary weights $\tilde{w} = (\tilde{w}_1, ..., \tilde{w}_d)$.

7.2 Optimization Results

In the appendix in table 18 to 27 an overview is given of the optimization results. We have used the sequential quadratic programming (SQP) method to optimize the given portfolios. This is because this method gave the most desirable results. This is due to the fact that the constraints are explicitly incorporated into the objective function. On the other hand the unconstrained nonlinear optimization methods need the use of penalty functions which are very sensitive in their use. When the penalty is too high, the ability of gaining a higher return is overruled by the high amount of risk it will give in return. Or the opposite problem will occur, that is if the penalty is too loose the optimizer will always choose the portfolio which will result in the highest amount of return. Finding an appropriate penalty function which overcomes these issues takes much time. Besides when the unconstrained objective function is well specified, both optimization methods will converge to the same optimal solution.

In figure 24 an overview is given of the optimal portfolio weights obtained by the optimizations over different risk measures. The underlying sample data consist of weekly log returns from April 1990 until May 2009. We can see that the optimization results over all different risk measures return about the same optimal portfolio. Obvious is the outcome that Treasury and High Yield form the most safe portfolio, that is the least risk in terms of the different risk measures. Another remarkable point is the fact that over the total portfolio weight only a very small portion is allocated to Equities. Moreover, Real Estate and Emerging Market Debt are very popular when high returns are acquired albeit at the cost of a higher level of risk.

The figures 25 until 30 show the optimal portfolio weights over the different models from section 5 at each different type of risk measure. When we compare the results of the different optimizations shown from the different figures, it becomes clear that the similarity between the different optimizations per model is large. By consulting the different tables from the appendix we can clarify the outcomes of each optimization.

Looking at the performance of the multivariate normal distribution in figure 25, we can see that the optimal portfolio weights are very similar to the sample data. The only major difference comes from the choice in allocation to Commodities. This result is in accordance with table 14 where the choice in allocation to commodities is clarified by the high mean return, although be it at a high cost in terms of deviation (seen from table 15). The multivariate Student's t distribution shows about the same result as the multivariate normal distribution. The major difference can again be explained by the mean return results displayed in table 14. The smaller allocation to Treasury is explained by the low mean return of this asset. The still relatively major proportion to Treasury is clarified by the very low level of risk at the required return. When a high return is required the optimizer chooses a portfolio of Real Estate, EMD, Commodities and Equities. This is in accordance with the multivariate normal distribution, but differs from the sample portfolio in the fact that the sample portfolio does not incorporate any allocation to Commodities and Equities at higher required returns.

The multivariate Skewed t distribution stands out in the extreme high allocation to Equities at a high required return. This is due to the simulated high mean return of this asset (shown in table 14). Remarkable is also the fact that no optimal portfolio consists of an allocation to Real Estate. From table 15 it becomes clear that this is due to the high risk of this asset.

By looking at the performance of the normal copula model with the marginals of a mixture of normal distributions, we can find a relatively similar result to the optimal chosen portfolio of the sample data. The only main differences come from the major allocation to Real Estate at a higher required return, which comes at the cost of a bigger allocation to High Yield and EMD.

The Student's t copula with marginals of a mixture of normals looks very similar to the optimal allocation results of the normal distribution. That is except for the fact that the t copula has a bigger allocation to EMD at the cost of Real Estate. This is also the main difference between this model and the optimal portfolio of the sample data. Another difference is that this model chooses a major allocation to commodities in comparison with the original model.

When we take a look at the optimal portfolio choice for the skewed t copula model with marginals of mixtures of normals, we see some clear differences with the original model. The small allocation to High Yield is outstanding, but explained by the relative small mean return at relative high risk. The big allocation to EMD and Commodities is explained by the fact that these two assets provide the highest mean return in this portfolio.

At last we compare the performances of the copula models with marginals of either a mixture of t components or a Skewed t distribution. Looking at the results we can see a clear contrast with the other models and the sample portfolio. All three models allocate a huge part of their portfolio to Treasuries. The remaining part of the portfolio is almost totally allocated to High Yield. The t copula has small part invested in Equities and Commodities, which is in accordance to the sample model. The normal copula has somewhat more invested in these two assets but shows very similar results. The allocation of the Skewed t copula deviates a little bit more from the sample-, normal- and t copula model in the fact that the allocation towards Equities and Commodities grows substantially up to 60% of the portfolio at the highest required mean return. These results are explained by the relatively high deviation in mean return from the sample data, especially for EMD and Real Estate. These two assets also display an enormous standard deviation due to some big outliers in the simulated samples, which can be seen clearly from the scatterplots in the figures 21 to 23 in the appendix.

The efficient frontiers based on the minimizations of different risk measures over the different portfolio models are shown in the appendix. Figure 31 to 36 show the performance of all models at each risk criteria. Furthermore, the figures 37 until 54 show the same results, however this time in sets of three models in comparison with the performance of the sample model. As we look at the different optimizations in general, we can state that all models show almost each time the same relative performance towards each other. This was also verified with the optimal portfolio weights which we saw earlier. In general it can be said that the copula models with the marginals of a mixtures of normal distributions, perform best of all. Remarkable is the fact that the multivariate normal distribution performs very well over the different portfolio optimizations. The performance of the Skewed t distribution are also outstanding albeit in a negative way. This is due to the high mean return of Equities which we have already seen before. The minimization of the Maximum Drawdown seems to be the hardest portfolio optimizations of all, since the efficient frontier does not look really like a smoothed curve. This is probably due to the highly discontinuous function which needs to be optimized.

The same overall results become even more clear by zooming into the performances of each set of models. When we compare the different results of the multivariate models it becomes clear that the multivariate normal distributions always outperform the other models. This is remarkable since the log-likelihood of this model was clearly less than that of the multivariate Student's t- and Skewed t distribution. However when we compare the other statistics it makes more sense. For example from the calculated mean returns from table 14, we can see that especially the multivariate t distribution overestimates the mean return for each asset. When we compare the copula models with mixtures of normals we see very similar optimization results. Only at a high required return there is somewhat more dispersion among the different models. When we compare the performances of the copula models with their optimal marginal distributions, it becomes clear that the different models have somewhat more difficulties in copying the sample portfolio. Moreover, it is noticeable that all these optimal copula models always underestimate the risk.

7.3 Conclusions on Optimization

From the portfolio optimization results we can draw some preliminary conclusions that may be unexpected to some readers. First of all we have empirically verified that the underlying distribution is the most important cause of the differences in outcomes and not the choice of the risk measure. This is also verified by Hu and Kercheval (2007) among others. Moreover our empirical evidence suggests that we even may possibly extend the result of proposition 7.1 - which states that the choice of risk measure is not determining the optimal portfolio outcome, but choice in elliptical distribution is - to the case of any distribution than only elliptical distributions. The bad performance of some models becomes really clear from the results of the optimized portfolios. Overall it can be said that the multivariate normal distribution and the normal-,Student's t- and Skewed t copula with the marginals of a mixtures of normal distributions, perform best. Although it must be said that no model is able to precisely copy the results of the sample model. Moreover most models underestimate the risk and are not able to reach the empirical efficient frontier. Still maybe the most surprising fact comes from the outstanding performance of the multivariate normal model which clearly has far less difficulties in approximating the empirical efficient frontier (with mean adjusted to the preferences of PGGM) than the most advanced models. This is however mostly due to the fact that most (advanced) models are not well able to reproduce the real mean of each asset.

8 Overall Conclusion

The use of copula models in the field of risk management and portfolio optimization does not really lead to one straightforward overall conclusion. First of all we have tested the estimated models on their fit in terms of log-likelihood, AIC and BIC. Based on the log-likelihood and the AIC criteria the Skewed t copula with marginals of a mixture of Student's t components and a single Skewed t distribution performed best. Based on the BIC criteria we ended up with a Student's t copula. The multivariate Skewed t distribution showed best fit on all three criteria among the multivariate non-copula models. By looking further at the overall fit in terms of major statistics (i.e. mean, sigma, skewness and kurtosis) the copula models with normal mixture marginals performed best.

By comparing the scatter plots we preferred the Skewed t copula with marginals of a mixture of normal components which was the second best model based on the log-likelihood and the AIC criteria. This model is preferred over the copula models with optimal (mixtures of Student's t or Skewed t) marginals, since the performance of these latter models are enormously influenced by outliers. By looking at the multivariate models we did not see any large deviations from the data sample model in their plot.

Next we have applied the different models in risk management by testing in which way they were able to estimate the risk accurately in terms of different risk measures. By looking at the period until December 2007 it is a pity to see that all models had real trouble in reproducing the risk measures, 1% VaR, 1% ES and MDD, which concentrate on the tail events. Over the period until May 2009 all models showed a real improvement but it was interesting to see how the optimal copula models, in terms of fit, had overestimated some risks this time. Especially the 1% VaR and 1% ES were easily overestimated by these models. On the other hand the multivariate non-copula models also showed difficulties in reproducing the different risks. Only the multivariate Skewed t distribution was slightly better. Overall could be concluded that the copula models with mixtures of normal distributions as marginals were best in reproducing the different risks. It should be noted that it seems as if the choice of the marginals of the copulas causes the most impact on the risk performance. Moreover based on the risk results the choice which copula to use, seems of somewhat of less relevance.

At last we applied the different models in the field of portfolio management with the models estimated over the period until May 2009. We compared the models on their ability to reproduce the optimal portfolio for the historical sample of data. Also this time the copula model with the mixture of marginals showed the best results. On the other hand it was a remarkable fact that the multivariate normal model performed really well, where we expected that it would undervalue the risk. Continuing on this result it would seem as if no one really could blame a portfolio manager for using this standard model. It must be said that the underperformance of the multivariate t and Skewed t models and their copula variant is not due to the way they are estimated. It is rather because of the normal mixtures variance structure in the simulation part which allows for big outliers to occur via the additional random variables drawn from the χ^2 distribution. The 'stylized fact' of asymmetric dependence is not found for real portfolio relevance for PGGM, at least not for the considered data on weekly frequency.

In conclusion it can be said that the mixture of Student's t components does not really prove to be a good way of estimating marginal distributions, since it tends to overestimate the tails of the distribution. On the other hand the marginals with mixtures of normals do give improvement of the copula models.

When it comes to dependence, the estimation results show that the higher dimensional Skewed t copula provides for a potential model improvement on higher dimensional problems. Moreover, based both on fit and the applications in the field of risk management and portfolio optimization, it really provides for a solution to improve the ability to estimate and reproduce the degree in which certain assets show tail dependence. And although the choice of the marginals has a big influence on the model outcomes, the dependence between the different assets is better captured and demonstrated by the use of a Skewed t copula model. For instance the knowledge that Treasury is not negative dependent with the other assets, is a desirable result in the use of the Skewed t copula, especially in these times of crisis where no investment gives a guaranteed insurance.

9 Further Research

Although this thesis has shown some promising results, still a lot needs to be done in future research. The substantial improvement of choosing a Skewed t distribution or the Skewed t copula over the normal- or Student's t variant leaves room for debate, since no real substantial gain is obtained yet. Whether this is due to the possibly inappropriate way the models are built, or the choice of the sample data, is the question. The estimation and simulation process still leave some room for improvement, but it should also be noted that the performed analysis is based on weekly data. For monthly data, tail dependence of contemporaneous returns may be stronger, possibly leading to a different conclusion.

Moreover when we take a close look at the scatter plots of the different used index series in figure 1,10 and 11 it is not ensured that there really exists any structural tail dependence over time. Moreover, most of the possible evidence in the scatter of Equities vs. Commodities stems from only one tail observation. From the scatter between EMD and Real Estate in figure 10 we could come to a similar conclusion. From the scatter in figure 11 Treasury vs. High Yield no tail dependence should be expected.

The assumed model incorporates no serial correlation or lagged cross-effects. However, the current financial crisis did not start simultaneously at all different markets. It started with a credit crunch, which caused a domino effect onto other markets. Consequently certain tail events appeared in the sample with some delay, where our model assumes that these occur simultaneously. This caused certain tail dependence in subsequent periods, sometimes referred to as contagion, not to appear in the estimation results of our advanced models. Therefore, inclusion of these properties is also a desirable addition to possible further research.

When we take another look at the different figures it also must be noticed that it seems as if the (tail) dependence structure varies over time. To model the time varying dependence one might consider to use Markov regime-switching models. Moreover Patton (2006) uses conditional copulas in which he allows the parameters of a certain copula to vary over time. Another possibility is to fix the parameters of the copula and switch between different copulas over time (see for example Rodriquez (2007), Okimoto (2008), among others). Markwat *et al.* (2009) combine both methods to allow the copula to vary in 'strength' and 'structure' over time. Although these results look very promising, they are only studied for the bivariate case. Combining our higher dimensional (a)symmetric copulas in a time varying manner may provide for an interesting extension.

Another topic for further research can be found in the study of assets of which only few observations are available. The index return series we used in this thesis as proxy for the different assets all have a very rich history concerning the available data. Other assets, like Private Equity, might give somewhat more trouble in estimating dependencies. For this purpose one might consider also the use of Bayesian methods (see Gelman *et al.* (1995) or Sivia *et al.* (1996), among others) in a copula context for measuring dependence among assets of which only a few observations are available.

Where all foregoing methods focus on the return series themselves, much study has to be done on what are the real return and risk drivers. For example during the last crisis a lack of liquidity became a serious issue for many companies. This causes new dependence relationships to arise. Where many analysts could not imagine that liquidity would become a problem in times of crisis, it did. Moreover, also appropriate measures of risk liquidity issues should be considered. In this sense every crisis may have its one risk driver that causes the dependence.

Other further research might involve the choice of an appropriate optimization procedure. All the different methods we used i.e. the simplex search method of Nelder and Mead (1965), the BFGS Quasi-Newton method (see Fletcher (1970), among others) or the (SQP) method (see Fletcher(1980), among others), did not always provide for the best feasible solution. Moreover all these methods might return local solutions and thus can not guarantee that the globally optimal solution will always be found. Another possibility is to simply use more simulations from each model during optimization. Moreover because of calculation time we only simulated 10.000 returns series for each asset. However for example Hu (2005) and Hu and Kercheval (2007) use 1.000.000 simulations in their portfolio optimizations.

Another shortcoming of using only 10.000 simulations comes in the use of the mixtures models with normal- or Student's t components. Because if there are multiple components to fit but there is only a small chance of ending up in a particular distribution, then there are only few draws taken from that particular distribution. Looking at figure 13 for the commodities case, this becomes more clear. To fit the commodities series properly five Student's t distributions are used. However the fifth distribution is hardly seen because it contains very few observations. Moreover especially in the case of using Student's t distributions with a small degrees of freedom parameter, we could obtain very different results. Therefore more research has to be done in order to obtain optimal mixtures of both symmetric and asymmetric components within the same distribution to fit all desired properties. Using the mixtures models with Student's t components or a Skewed t distribution, we have tried to restrict the parameter of ν on becoming less than 4 to prevent very extreme returns to occur (not reported). However by restricting the ν parameter to to be larger than 4 the likelihood of the model descends much, so that a mixture of normals is then preferred over a mixture of t in every considered sense..

Finally it must be noted that although most advanced models show a better fit, they were also often not cable in reproducing the mean of the sample properly. Further research must be done to combine both flexibility and accurateness in constructing the optimal (marginal) distribution.

10 References

Aas K., Haff I. (2006), The Generalized Hyperbolic Skew Student's t distribution, *Journal of Financial Econometrics Advance Access* published online on January 26, 2006.

Acerbi, C., Tasche, D., (2002), On the coherence of expected shortfall, *Journal of Banking & Finance*, Vol 26, Is. 7, July 2002, P. 1487-1503.

Ang, A. and Chen, J. (2002), Asymmetric Correlations of Equity Portfolios, *Journal of Financial Economics*, 63(3), 443-494.

Abramowitz, M. and Stegun, I., eds. (1965), *Handbook of Mathematical Functions.* Dover Publications, Inc., New York.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., (1997). Thinking coherently. *Risk*, 10, 1997, p. 68-71.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., (1999). Coherent measures of risk., *Mathematical Finance*, Vol. 9, No. 3, 1999, p. 203-228.

Bae, K.-H., G.A. Karolyi and R.M. Stulz, (2003), A New Approach to Measuring Financial Contagion, *Review of Financial Studies*, 16(3), 717-764.

Barndorff-Nielsen O.E. (1977). Exponentially decreasing distributions for the logarithm of the particle size. Proceedings of the Royal Society. London. Series A. *Mathematical and Physical Sciences* 353: 401-419.

Barndorff-Nielsen, O.E. and Blæsid (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. In C. Taillie, G. Patil, and B. Baldessari(Eds.), *Statistical distributions in Scientific Work*, Vol. 4, Dordrecht: Reidel.

Blæsild P. and Sørensen M. (1992). "hyp"–a computer program for analyzing data by means of the hyperbolic distribution. Research Report 248, Department of Theoretical Statistics, Aarhus University.

Bouye, E. (2002). Multivariate extremes at work for portfolio risk management. Working paper, City University Business School, London, UK.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004), *Copula Methods in Finance*, John Wiley & Sons, England.

Chekhlov, A., S. Uryasev, M. Zabarankin (2005): Drawdown Measure in Portfolio Optimization," *International Journal of Theoretical and Applied Finance*, Vol. 8, No. 1, 13-58.

Clement, A. (2003). The empirical value-at-risk/expected return frontier: a useful tool of market risk managing. University of Rome. 2003.

Davison, A. and Smith, R. (1990). Models for exceedances over high thresholds (with discussion). J. R. Stat. Soc. Ser. B *Stat. Methodol.* 52, 393-442.

Demarta, S. and McNeil, A. (2005). The t copula and related copulas. *International Statistical Review.*

Dempster A.P., Laird N.M., and Rubin D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Series B 39: 1-37.

Trivedi, P.K. and Zimmer, D.M. (2006) Using Trivariate Copulas to Model Sample Selection and Treatment Effects. Journal of Business and Economic Statistics. January 1, 2006, 24(1), p 63-76.

Eberlein, E. and Hammerstein, E. (2003). Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes. *University of Freiburg.* Nr. 80.

Embrechts, P., A. McNeil and D. Straumann (1999), Correlation: Pitffalls and Alternatives, *RISK Magazine*, May, pp. 69-71.

Embrechts, P., A. McNeil and D. Straumann (2002), Correlation and Dependence in Risk Management: Properties and Pitfalls, in M. Dempster (ed.), Risk Management:Value at Risk and Beyond, Cambridge University Press, pp. 176-223.

Fang, K., Kotz, S., Ng, W. (1990). Symmetric multivariate and related distributions. Chapman & Hall.

Fletcher, R., (1970), "A New Approach to Variable Metric Algorithms," *Computer Journal*, Vol. 13, pp. 317-322, 1970.

Fletcher, R. (1980), "*Practical Methods of Optimization*," Vol. 1, Unconstrained Optimization, and Vol. 2, Constrained Optimization, John Wiley and Sons, 1980.

Gelman, A.; Carlin, J.; Stern, H.; and Rubin, D. (1995). *Bayesian Data Analysis.* Boca Raton, FL: Chapman & Hall, 1995.

Gill, P.E., W. Murray, and M.H.Wright, (1981). *Practical Optimization*, London, Academic Press, 1981.

Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

Hartmann, P., Straetmans, S., and de Vries, C. G. (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics*, 86(1).

Heij, C., De Boer, P., Franses, P.H., Kloek, T., and Van Dijk, H.K. (2004). *Econometric Methods with Applications in Business and Economics.* Oxford University Press. Oxford.

Hock, W. and K. Schittkowski, (1983). "A Comparative Performance Evaluation of 27 Nonlinear Programming Codes," *Computing*, Vol. 30, p. 335, 1983.

Hoogerheide, L. (forthcoming). Note on Expectation-Maximization (EM) algorithm for (multivariate) Student's *t*- and Skewed *t* distributions, and mixture models with Student's *t*- or Skewed *t* components. Unpublished manuscript.

Hu, W. (2005). Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization, and portfolio credit risk. Florida State University PhD Dissertation, 2005.

Hu L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics*, 2006, 16, 717-729

Hu, W. and Kercheval, A.N. (2007). Portfolio optimization for t and skewed-t returns. Florida State University 2007.

Hyung N. and De Vries, G., (2005). Portfolio Diversification Effects of Downside Risk. *Journal of Financial Econometrics.*2005; 3: 107-125

Kang, L. (2008). Modeling the Dependence Structure between Bonds and Stocks: A Multivariate Copula Approach (2008). Working paper

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhiā: The Indian Journal of Statistics* A 32, 419-430.

Kole, E., Koedijk, K., Verbeek, M. (2007). "Selecting copulas for risk management," *Journal of Banking & Finance*, Elsevier, vol. 31(8), pages 2405-2423, August.

Lagarias, J.C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, Vol. 9, Number 1, pp.112-147, 1998.

Leal, R.P.C., Mendes, B.V.M., (2006) Maximum drawdown: Models and Applications, The Journal of Alternative Investments (Spring 2005). B.V.M. Mendes, Maximum drawdown: Models and Applications, *The Journal of Alternative Investments* (Spring 2005).

León, C. and Laserna, J.M. (2008), "Efficient Portfolio Optimization in the Wealth Creation and Maximum Drawdown Space.", *Borradores de Economía*, no. 520, Banco de la Republica Colombia, august, 2008.

Liu, C. and Rubin, D.(1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica*, 5(1995), 19-39.

Longin, F. and Solnik, B. (2001), Extreme Correlation of International Equity Markets, *Journal of Finance*, 56(2), 649-676.

Markowitz, H.M. (1952). Portfolio Selection. *Journal of Finance* 7 (1): 77-91.

Magdon-Ismail M. and Atiya A. (2004) "Maximum Drawdown", *Risk*, October.

Mausser, H. and Rosen, D. (1999). Efficient Risk/Return Frontiers for Credit Risk. *Algo* *research Quarterly* vol. 2, no.4 December 1999.

McNeil, A., Frey, R. and Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. The Draft of June 21, 2004. Will be published by Princeton University Press. 2005.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization, *The Computer Journal*, 1965

Nelsen, R. B. (1999). *An Introduction to Copulas.* Springer, New York.

Patton, A.J., (2006), Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47(2), 527-556.

Powell, M.J.D., (1983). "Variable Metric Methods for Constrained Optimization," Mathematical Programming: The State of the Art, (A. Bachem, M. Grotschel and B. Korte, eds.) Springer Verlag, pp 288-311, 1983

Prause K. (1999). The generalized hyperbolic models: Estimation, financial derivatives and risk measurement. PhD Thesis, Mathematics Faculty, Freiburg University, Freiburg.

Protassov, Rostislav (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed. *Statistics and Computing.* Vol 14, issue 1.

Okimoto, T. (2008). New evidence of asymmetric dependence structures in international equity markets. Journal of Financial and Quantitative Analysis, 43,787815.

Ortobelli, S., Rachev, S.T, Stoyanov, S., Fabozzi, F.J., Biglova, A., (2005). The proper use of risk measures in the portfolio theory. *International Journal of Theoretical and Applied Finance* 8, 1107-1133.

Rodriguez, J.C., (2007), Measuring financial contagion: a copula approach, *Journal of Empirical Finance*, 14(3), 401-423

Rotundo, G. and Navarrab, M. (2007). On the maximum drawdown during speculative bubbles *Physica. A: Statistical Mechanics and its Applications.* Vol. 382, Is. 1, 1 August 2007, Pages 235-246.

Samko, S.G., Kilbas, A.A., and Marichev, O.I. (1993). Fractional Integrals and Derivatives. Yverdon, Switzerland: Gordon and Breach, p. 20.

Shanno, D.F., (1970), "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computing*, Vol. 24, pp. 647-656, 1970.

Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial.* New York: Oxford University Press, 1996.

Sklar, A. (1959). Fonctions de *ré*partition *à* n dimensions et leurs marges. *Publications de* *l'Institut de Statistique de L'Université de Paris* 8, 229-231.

Sklar, A. (1996). Random Variables, distribution Functions, and Copulas - a Personal Look Backward and Forward published in *Distributions with Fixed Marginals and Related Topics*, edited by L. Rüschendorf, B. Schweizer, and M.D. Taylor, Institute of Mathematical Statistics, Hayward, CA, pages 1-14.

Spanier, J. and Oldham, K.B. (1987). "The Basset." Ch. 51 in *An Atlas of Functions*. Washington, DC: Hemisphere, pp. 499-507.

Sun,W., Rachev, S., Stoyanov, S., Fabozzi, F. (2008).Multivariate skewed student's t copula in the analysis of nonlinear and asymmetric dependence in the German equity market. Studies in Nonlinear Dynamics and Econometrics, 12(2), 3.

Schmidt R.(2003a). Dependencies of extreme events in finance: modeling, statistics, and data analysis. Dissertation. ULM. 2003.

Savu, C. and Trede, M. (2006), Hierarchical Archimedean copulas. Working Paper, University of Munster.

Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2002). *Mathematical statistics with applications* Pacific Grove (California): Duxbury Press, Sixth Edition, 2002.

Weisstein, E.W. (2009) "Modified Bessel Function of the Second Kind." From *MathWorld* A Wolfram Web Resource. http://mathworld.wolfram.com/Modi¯edBesselFunctionoftheSecondKind.html

Other used websites (2009):

http://www.pggm.nl/,
http://www2.standardandpoors.com/,
http://www.epra.com/,
http://www.jpmorgan.com/,
https://live.barcap.com/,
http://www.ml.com/,
http://www.mathworks.com/,
http://dictionary.babylon.com/.

A Plotting a confidence ellipse: derivation of a normal distributed ellipse

If two random variables x and y are independent, normally distributed random variables with mean 0 and variance 1, then the random variables x and y are χ_2^2 distributed. From this we can compute an ellipse of the normal distribution in a scatter plot using the following formulae:

$$\hat{x}'\Sigma^{-1}\hat{x} = \chi^2_{2,\alpha} = d \tag{79}$$

where $\hat{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $\Sigma^{-1} = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$

From this we can see that

$$ax^2 + 2cxy + by^2 = d \tag{80}$$

Knowing that the angle $\gamma°$: $\tan\gamma = \frac{y}{x}$ so that

$$y = x\tan\gamma \tag{81}$$

Now we can use formulae (81) in (80) from which we can get

$$ax^2 + 2\tan\gamma cx^2 + b\tan^2\gamma x^2 = d \tag{82}$$

From which we can get the following for x:

$$x^2 = \frac{d}{a + 2\tan\gamma c + b\tan^2\gamma} \tag{83}$$

so,

$$x = \pm\sqrt{\frac{d}{a + 2\tan\gamma c + b\tan^2\gamma}} \tag{84}$$

B Normal-, Student's t- and Skewed t copula: a graphical illustration of different properties

Figure 3: Density of a two-dimensional normal(first 3)- and Student's t copula for different values of the correlation $\rho = 0; 0.5; 0.9$ and skewness parameter $\nu = 3; 8; 50$.

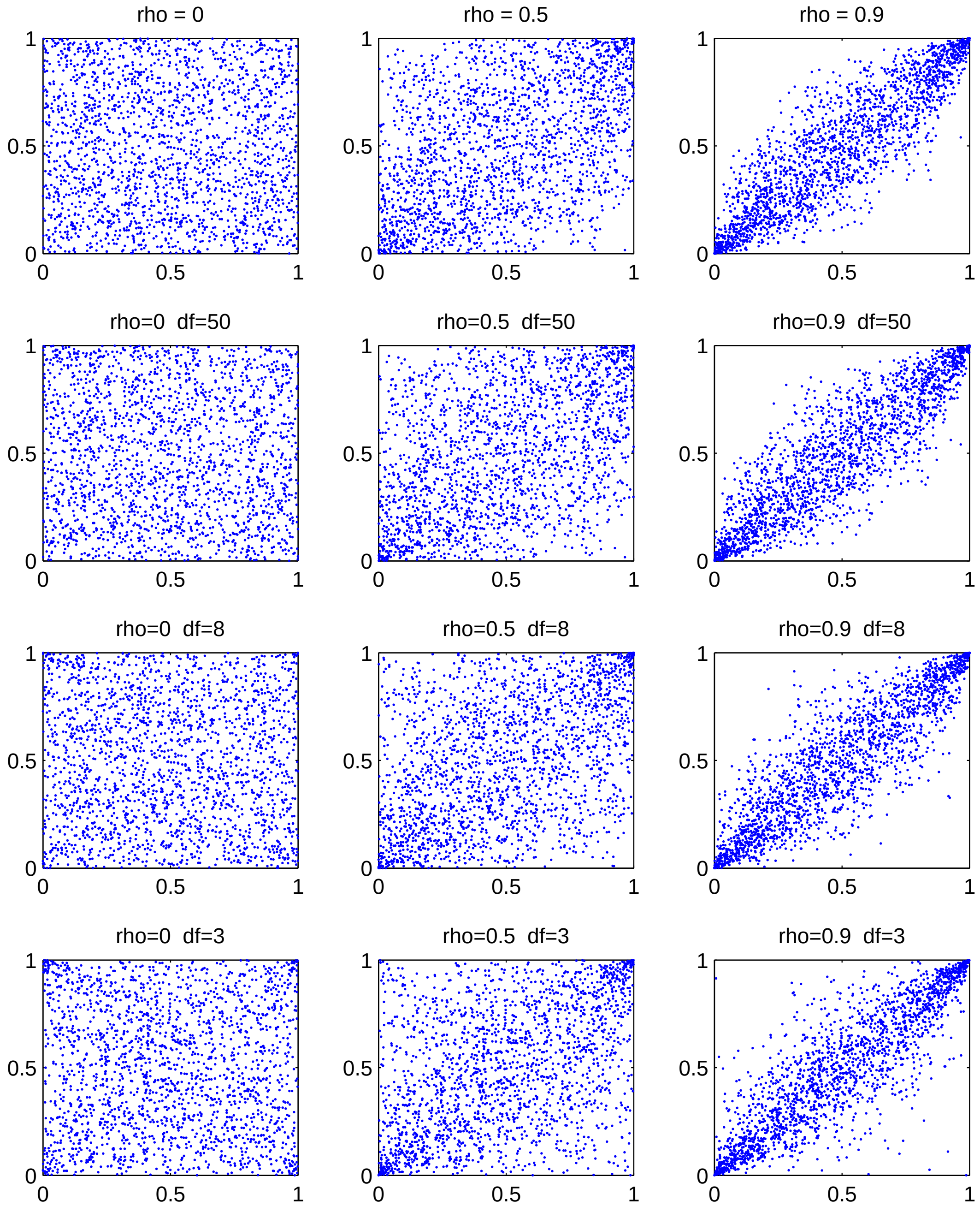


Figure 4: Density of a two-dimensional of bivariate Skewed t copula. With different values for the correlation $\rho = 0; 0.5; 0.9$ and skewness parameter $\nu = 3; 8; 50$. The parameter $\gamma = [0\ 0]$ determines the tail dependence. When the skewness is high (i.e. when ν is low) this will lead to skewed tail dependence if both γ parameters have the same sign and are nonzero. In this case no skewed tail dependence should be expected.

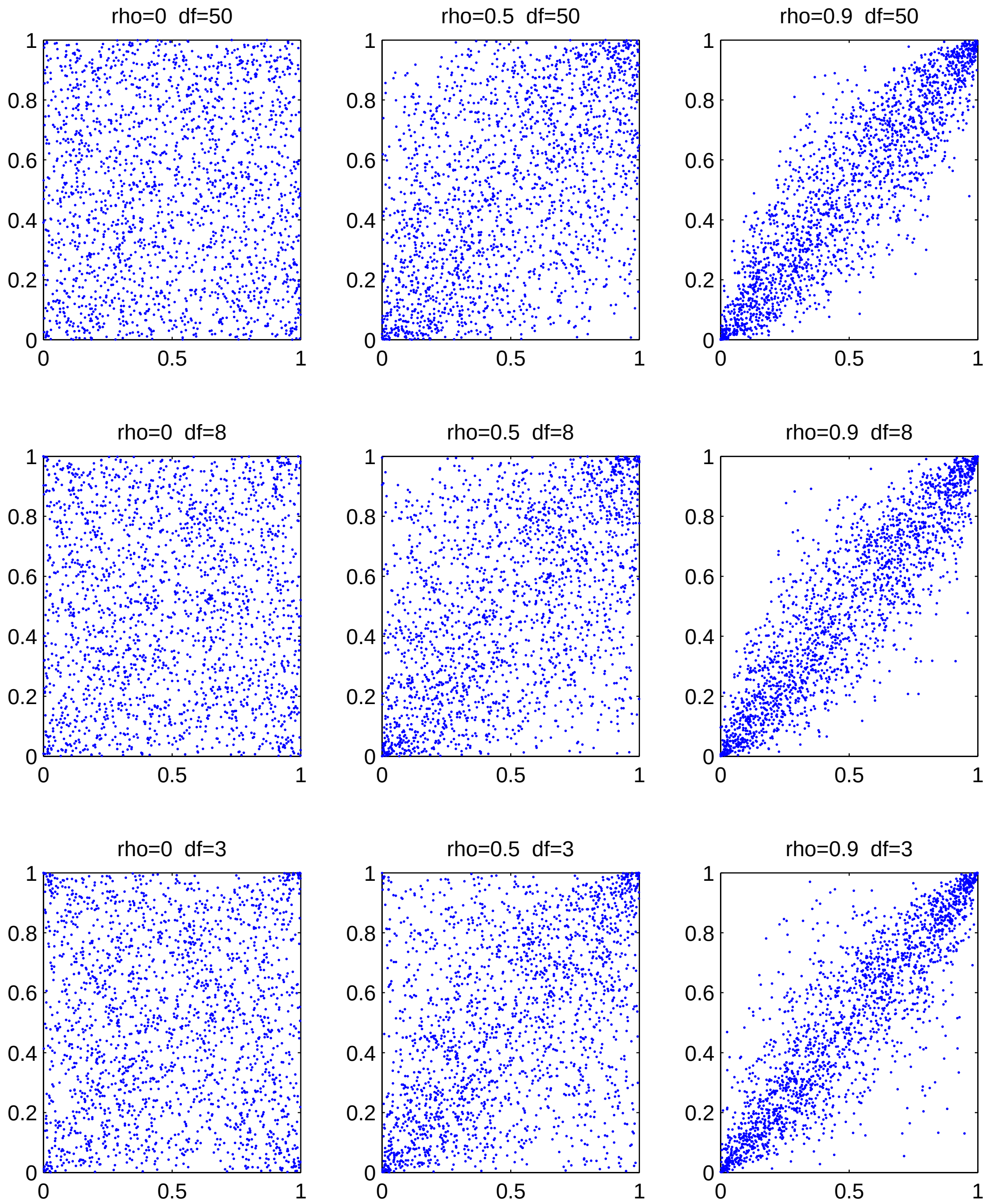


Figure 5: Density of a two-dimensional of bivariate Skewed t copula. With different values for the correlation $\rho = 0; 0.5; 0.9$ and skewness parameter $\nu = 3; 8; 50$. The parameter $\gamma = [0 - 1]$ determines the tail dependence. When the skewness is high (i.e. when ν is low) this will lead to tail skewed dependence if both γ parameters have the same sign and are nonzero. In this case no skewed tail dependence should be expected.

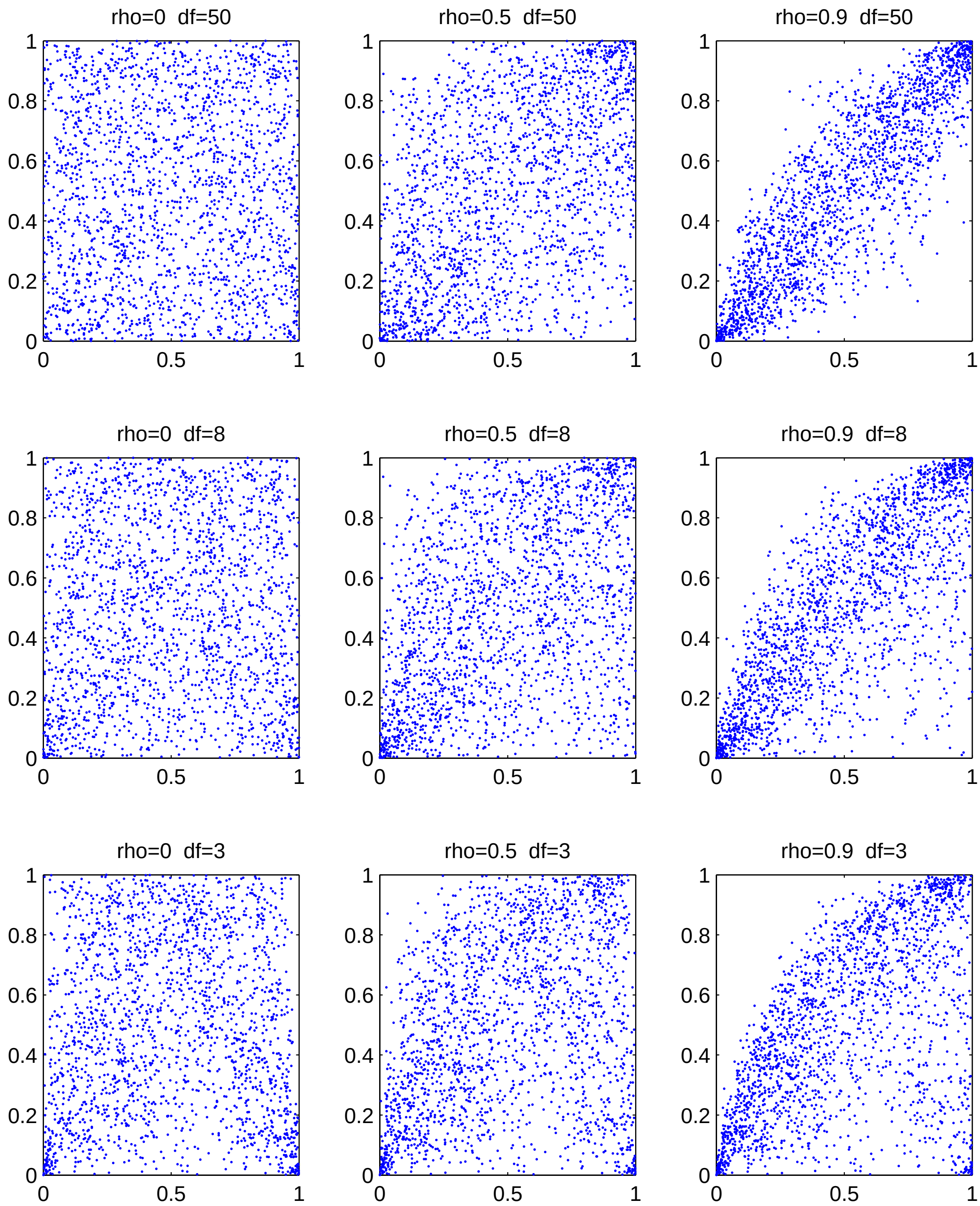


Figure 6: Density of a two-dimensional of bivariate Skewed t copula. With different values for the correlation $\rho = 0; 0.5; 0.9$ and skewness parameter $\nu = 3; 8; 50$. The parameter $\gamma = [-1 - 1]$ determines the tail dependence. When the skewness is high (i.e. when ν is low) this will lead to tail skewed dependence if both γ parameters have the same sign and are nonzero. In this case *little* skewed tail dependence should be expected.

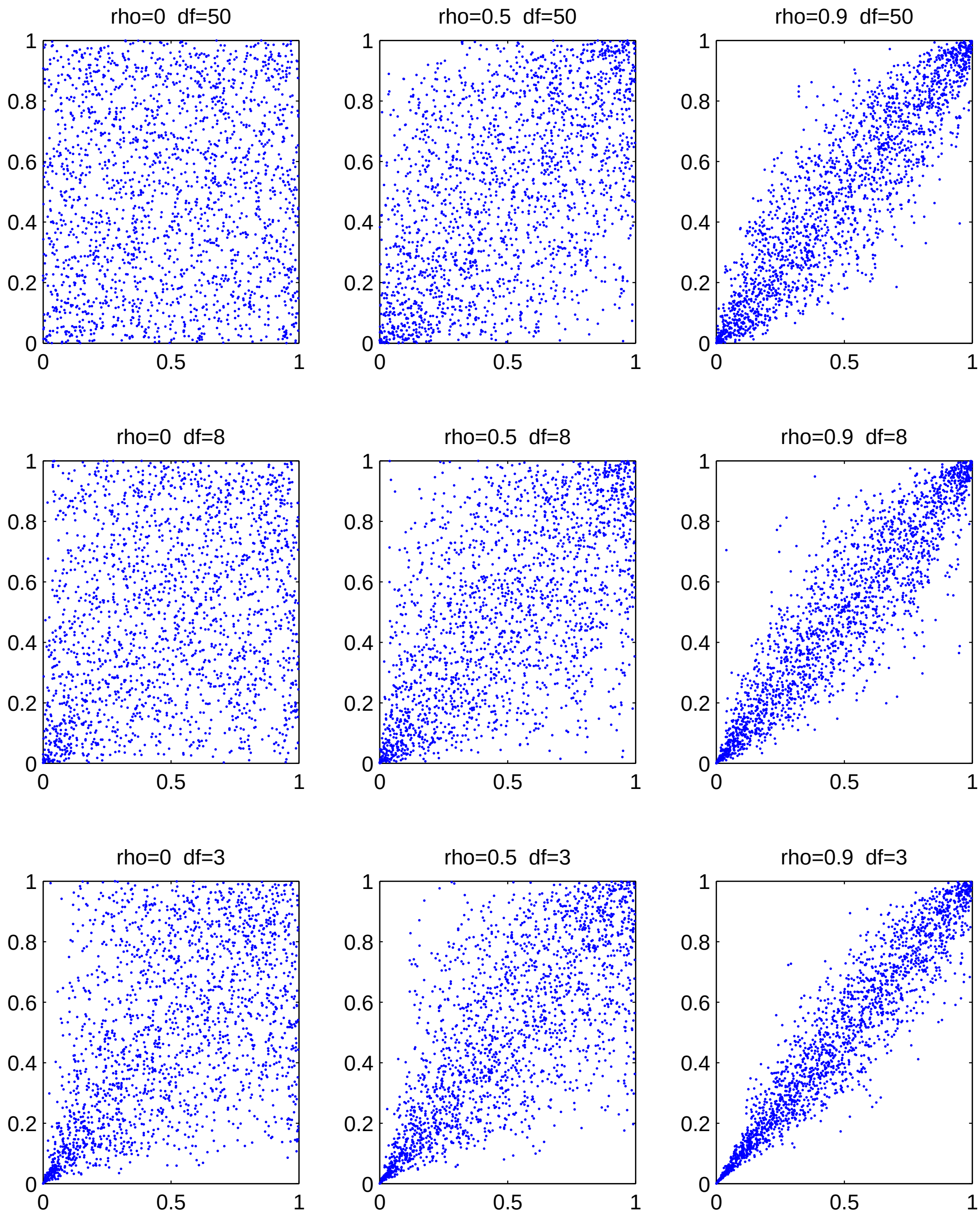
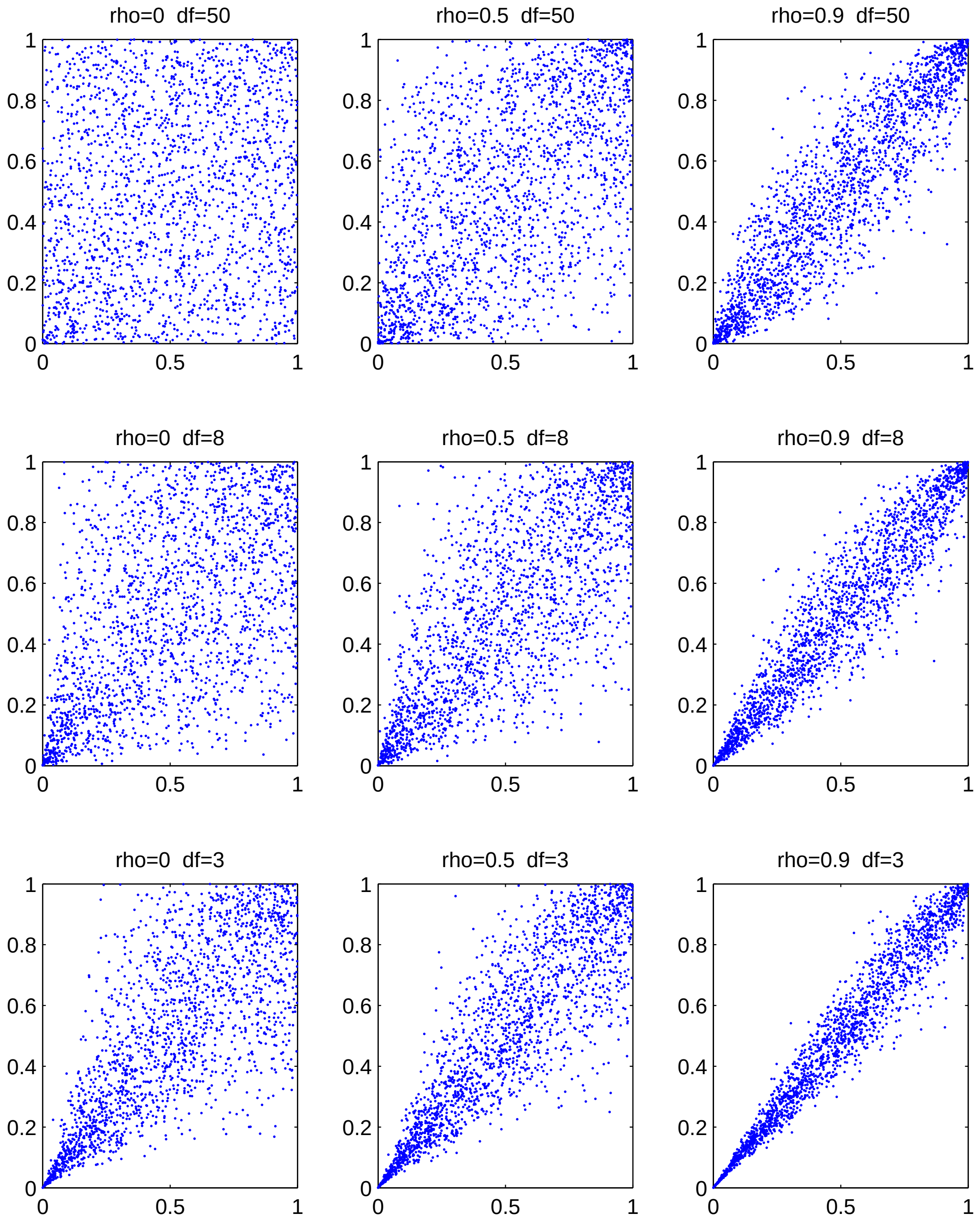


Figure 7: Density of a two-dimensional of bivariate Skewed t copula. With different values for the correlation $\rho = 0; 0.5; 0.9$ and skewness parameter $\nu = 3; 8; 50$. The parameter $\gamma = [-2 - 2]$ determines the tail dependence. When the skewness is high (i.e. when ν is low) this will lead to skewed tail dependence if both γ parameters have the same sign and are nonzero. In this case *strong* skewed tail dependence should be expected.



C Tables

Table 6: Optimal number of components per marginal for data Apr90-Dec07

Marginal	Equities				Commodities				EMD			
Mixture	no. comp.[a]	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC
normal	1	22.782	-45.426	-45.523	1	20.315	-40.494	-40.590	1	24.909	-49.682	-49.779
normal	2	23.125	-45.909	-46.150	2	20.599	-40.856	-41.098	2	26.517	-52.693	-52.934
normal	3	23.171	-45.796	-46.183	3*/**	20.690	-40.833	-41.220	3	26.647	-52.748	-53.135
normal	4*/**	23.235	-45.718	-46.250	4	20.695	-40.639	-41.170	4	26.672	-52.593	-53.125
normal	5	23.243	-45.529	-46.205	5	20.718	-40.480	-41.156	5	26.685	-52.414	-53.090
normal	6	23.242	-45.323	-46.144	6	20.740	-40.319	-41.140	6	26.685	-52.209	-53.030
normal	7	23.254	-45.143	-46.109	7	20.740	-40.114	-41.080	7**	26.770	-52.174	-53.140
normal	8	23.253	-44.935	-46.046	8	20.754	-39.937	-41.048	8	26.759	-51.947	-53.058
t	1	23.134	-46.062	-46.207	1	20.594	-40.982	-41.127	1	26.607	-53.008	-53.153
t	2	23.174	-45.871	-46.209	2	20.634	-40.789	-41.128	2	26.654	-52.829	-53.167
t	3	23.199	-45.646	-46.178	3	20.682	-40.612	-41.143	3	26.678	-52.605	-53.136
t	4	23.224	-45.424	-46.149	4	20.722	-40.419	-41.143	4	26.706	-52.387	-53.112
t	5	23.234	-45.170	-46.087	5	20.737	-40.176	-41.094	5	26.719	-52.140	-53.057
t	6	23.234	-44.898	-46.008	6	20.780	-39.990	-41.100	6	26.758	-51.945	-53.056
t	7	23.277	-44.709	-46.013	7	20.797	-39.750	-41.054	7	26.766	-51.687	-52.991
t	8	23.276	-44.434	-45.932	8	20.826	-39.534	-41.031	8	26.777	-51.437	-52.934
skewed t	1	23.144	-46.015	-46.209	1	20.601	-40.929	-41.122	1*	26.625	-52.977	-53.170

Marginal	Real Estate				Treasury				High Yield			
Mixture	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC
normal	1	23.416	-46.695	-46.792	1	34.498	-68.859	-68.956	1	33.137	-66.137	-66.233
normal	2*/**	24.029	-47.716	-47.957	2**	34.813	-69.284	-69.526	2	34.863	-69.384	-69.626
normal	3	24.035	-47.523	-47.909	3	34.823	-69.099	-69.485	3*/**	34.996	-69.446	-69.832
normal	4	24.035	-47.319	-47.850	4	34.825	-68.898	-69.430	4	34.999	-69.247	-69.778
normal	5	24.035	-47.114	-47.790	5	34.824	-68.693	-69.369	5	35.000	-69.043	-69.720
normal	6	24.035	-46.909	-47.730	6	34.824	-68.488	-69.309	6	35.008	-68.855	-69.676
normal	7	24.035	-46.704	-47.670	7	34.824	-68.282	-69.248	7	35.010	-68.654	-69.620
normal	8	24.035	-46.499	-47.609	8	34.824	-68.077	-69.188	8	35.045	-68.518	-69.629
t	1	23.990	-47.776	-47.921	1	34.742	-69.279	-69.424	1	34.904	-69.603	-69.748
t	2	24.031	-47.584	-47.922	2	34.832	-69.185	-69.523	2	34.974	-69.470	-69.808
t	3	24.031	-47.312	-47.843	3	34.836	-68.922	-69.453	3	34.975	-69.199	-69.730
t	4	24.033	-47.041	-47.766	4	34.859	-68.694	-69.418	4	35.004	-68.983	-69.708
t	5	24.032	-46.767	-47.685	5	34.862	-68.425	-69.343	5	35.015	-68.733	-69.651
t	6	24.032	-46.493	-47.604	6	34.906	-68.241	-69.352	6	35.009	-68.447	-69.558
t	7	24.032	-46.220	-47.524	7	34.922	-68.001	-69.305	7	35.032	-68.219	-69.523
t	8	24.032	-45.946	-47.443	8	34.934	-67.750	-69.247	8	35.032	-67.946	-69.444
skewed t	1	24.010	-47.747	-47.941	1*	34.807	-69.341	-69.535	1	34.946	-69.619	-69.813

[a]number of components per marginal. *(**) Indicates the optimal model (the optimal mixture of normals model) based on the AIC criteria.

Table 7: Optimal number of components per marginal for data Apr90-May09

Marginal	Equities				Commodities				EMD			
Mixture	no. comp.[a]	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC
normal	1	23.213	-46.288	-46.386	1	20.773	-41.408	-41.506	1	26.323	-52.507	-52.605
normal	2	24.038	-47.731	-47.977	2	21.399	-42.452	-42.698	2	28.497	-56.648	-56.893
normal	3	24.105	-47.658	-48.051	3**	21.495	-42.437	-42.830	3	28.706	-56.859	-57.252
normal	4**	24.154	-47.548	-48.088	4	21.501	-42.242	-42.782	4	28.739	-56.718	-57.257
normal	5	24.159	-47.351	-48.038	5	21.535	-42.102	-42.789	5**	28.775	-56.584	-57.271
normal	6	24.172	-47.170	-48.004	6	21.535	-41.896	-42.731	6	28.800	-56.427	-57.261
normal	7	24.225	-47.068	-48.050	7	21.537	-41.692	-42.673	7	28.808	-56.235	-57.217
normal	8	24.233	-46.877	-48.006	8	21.539	-41.490	-42.619	8	28.808	-56.027	-57.156
t	1	24.078	-47.948	-48.095	1	21.398	-42.588	-42.735	1	28.671	-57.134	-57.281
t	2*	24.125	-47.767	-48.111	2	21.429	-42.374	-42.718	2	28.692	-56.900	-57.244
t	3	24.126	-47.492	-48.032	3	21.494	-42.228	-42.768	3	28.733	-56.705	-57.245
t	4	24.155	-47.274	-48.010	4	21.531	-42.027	-42.763	4	28.753	-56.469	-57.206
t	5	24.167	-47.021	-47.954	5*	21.611	-41.910	-42.842	5	28.767	-56.222	-57.154
t	6	24.167	-46.745	-47.874	6	21.613	-41.638	-42.767	6	28.840	-56.092	-57.221
t	7	24.205	-46.546	-47.871	7	21.633	-41.400	-42.725	7	28.854	-55.842	-57.167
t	8	24.205	-46.269	-47.791	8	21.658	-41.174	-42.696	8	28.862	-55.582	-57.103
skewed t	1	24.094	-47.912	-48.109	1	21.420	-42.563	-42.759	1*	28.687	-57.097	-57.293

Marginal	Real Estate				Treasury				High Yield			
Mixture	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC	no. comp.	Log L	BIC	AIC
normal	1	21.339	-42.540	-42.638	1	36.940	-73.742	-73.840	1	32.493	-64.848	-64.946
normal	2	24.007	-47.668	-47.914	2**	37.257	-74.168	-74.413	2	35.961	-71.577	-71.822
normal	3**	24.152	-47.752	-48.144	3	37.261	-73.970	-74.362	3	36.287	-72.021	-72.414
normal	4	24.157	-47.553	-48.093	4	37.262	-73.763	-74.303	4	36.291	-71.823	-72.363
normal	5	24.157	-47.346	-48.034	5	37.262	-73.556	-74.243	5	36.359	-71.750	-72.437
normal	6	24.157	-47.139	-47.973	6	37.262	-73.349	-74.183	6**	36.369	-71.563	-72.398
normal	7	24.157	-46.932	-47.913	7	37.262	-73.142	-74.123	7	36.369	-71.357	-72.339
normal	8	24.157	-46.724	-47.853	8	37.262	-72.935	-74.063	8	36.369	-71.150	-72.279
t	1	24.094	-47.981	-48.128	1	37.190	-74.172	-74.319	1	36.241	-72.274	-72.421
t	2	24.129	-47.774	-48.117	2*	37.277	-74.071	-74.414	2	36.315	-72.147	-72.491
t	3	24.150	-47.539	-48.079	3	37.282	-73.805	-74.345	3*	36.326	-71.892	-72.432
t	4	24.150	-47.265	-48.001	4	37.283	-73.530	-74.266	4	36.331	-71.626	-72.362
t	5	24.188	-47.063	-47.995	5	37.279	-73.246	-74.178	5	36.345	-71.377	-72.310
t	6	24.186	-46.784	-47.913	6	37.279	-72.970	-74.099	6	36.364	-71.139	-72.268
t	7	24.195	-46.524	-47.849	7	37.319	-72.772	-74.098	7	36.372	-70.880	-72.205
t	8	24.211	-46.281	-47.802	8	37.361	-72.581	-74.103	8	36.401	-70.660	-72.181
skewed t	1*	24.113	-47.949	-48.145	1	37.242	-74.208	-74.405	1	36.273	-72.269	-72.466

[a]number of components per marginal. *(**) Indicates the optimal model (the optimal mixture of normals model) based on the AIC criteria.

Table 8: Statistics of optimal marginal distributions per mixture component for data Apr90-May09. The 1 component case indicates the Skewed t distribution. Multiple components indicate a mixture of Student's t distributions

	$\mu_{component\ 1}$	$\mu_{component\ 2}$	$\mu_{component\ 3}$	$\mu_{component\ 4}$	$\mu_{component\ 5}$
Equities	0,000745	0,011169	-	-	-
Commodities	-0,044399	0,005940	0,005404	-0,014182	-0,001643
EMD	0,002807	-	-	-	-
Real Estate	0,003884	-	-	-	-
Treasury	-0,015018	0,001498	-	-	-
High Yield	0,000466	0,002548	-	-	-

	$\sigma_{component\ 1}$	$\sigma_{component\ 2}$	$\sigma_{component\ 3}$	$\sigma_{component\ 4}$	$\sigma_{component\ 5}$
Equities	0,000310	0,000025	-	-	-
Commodities	0,001940	0,000647	0,000000	0,000009	0,000008
EMD	0,000071	-	-	-	-
Real Estate	0,000184	-	-	-	-
Treasury	0,000005	0,000024	-	-	-
High Yield	0,000046	0,000007	-	-	-

	$\nu_{component\ 1}$	$\nu_{component\ 2}$	$\nu_{component\ 3}$	$\nu_{component\ 4}$	$\nu_{component\ 5}$	γ
Equities	4,061427	4,689689	-	-	-	-
Commodities	5,852187	17,829934	9,180532	4,194016	8,352101	-
EMD	2,235211	-	-	-	-	-0,000280
Real Estate	2,307443	-	-	-	-	-0,000491
Treasury	1,920566	11,406669	-	-	-	-
High Yield	2,431360	7,405288	-	-	-	-

Table 9: Statistics of marginal distributions from mixture of normal distribution per mixture component for data Apr90-May09

	$\mu_{component\ 1}$	$\mu_{component\ 2}$	$\mu_{component\ 3}$	$\mu_{component\ 4}$	$\mu_{component\ 5}$
Equities	-0,0151	0,0024	-0,0149	0,0069	-
Commodities	-0,0323	0,0054	-0,0020	-	-
EMD	-0,1405	-0,0029	0,0022	0,0059	-0,0002
Real Estate	-0,0225	-0,0001	0,0038	-	-
Treasury	-0,0013	0,0018	-	-	-
High Yield	-0,0732	0,0001	0,0012	0,0010	0,0039

	$\sigma_{component\ 1}$	$\sigma_{component\ 2}$	$\sigma_{component\ 3}$	$\sigma_{component\ 4}$	$\sigma_{component\ 5}$
Equities	0,004329	0,000622	0,000042	0,000078	-
Commodities	0,003592	0,000750	0,000105	-	-
EMD	0,000255	0,001092	0,000185	0,000032	0,000012
Real Estate	0,010460	0,000932	0,000156	-	-
Treasury	0,000078	0,000020	-	-	-
High Yield	0,000944	0,000361	0,000041	0,000004	0,000003

Table 10: ν and γ of different models for data Apr90-Dec07

Model	ν	$\gamma_{Eq.}$	$\gamma_{Com.}$	γ_{EMD}	γ_{RE}	$\gamma_{Treas.}$	γ_{HY}
t cop with mixnormal marg	10.16						
t cop with optmix marg	9.75						
st cop with mixnormal marg	6.85	0.0054	-0.0146	-0.0183	-0.1535	0.0761	-0.4059
st cop with optmix marg	7.91	-0.0221	-0.0247	-0.0330	-0.1723	0.0961	-0.3686

Table 11: ν and γ of different models for data Apr90-May09

Model	ν	$\gamma_{Eq.}$	$\gamma_{Com.}$	γ_{EMD}	γ_{RE}	$\gamma_{Treas.}$	γ_{HY}
t cop with mixnormal marg	6.70						
t cop with optmix marg	6.67						
st cop with mixnormal marg	5.94	-0.1757	-0.0897	-0.1276	-0.3395	0.0782	-0.3245
st cop with optmix marg	6.47	-0.1149	-0.0946	-0.0336	-0.2398	0.1366	-0.2736

Table 12: Σ of different models for data Apr90-Dec07

Model	Σ	Equities	Commodities	EMD	RE	Treasury	High Yield
t cop with optmix marg	Equities	1	-0.0146	0.3239	0.4992	-0.0058	0.3018
	Commodities	-0.0146	1	0.0342	-0.0261	0.0236	-0.0042
	EMD	0.3239	0.0342	1	0.2294	0.2640	0.3702
	RE	0.4992	-0.0261	0.2294	1	0.0229	0.2787
	Treasury	-0.0058	0.0236	0.2640	0.0229	1	0.2677
	High Yield	0.3018	-0.0042	0.3702	0.2787	0.2677	1
normal cop with optmix marg	Equities	1	-0.0244	0.3186	0.5052	-0.0227	0.3002
	Commodities	-0.0244	1	0.0448	-0.0515	0.0069	-0.0065
	EMD	0.3186	0.0448	1	0.2401	0.2396	0.3599
	RE	0.5052	-0.0515	0.2401	1	0.0061	0.2821
	Treasury	-0.0227	0.0069	0.2396	0.0061	1	0.2190
	High Yield	0.3002	-0.0065	0.3599	0.2821	0.2190	1

		Equities	Commodities	EMD	RE	Treasury	High Yield
t cop with mixnormal marg	Equities	1	-0.0140	0.3233	0.4990	-0.0056	0.3021
	Commodities	-0.0140	1	0.0340	-0.0262	0.0237	-0.0040
	EMD	0.3233	0.0340	1	0.2301	0.2618	0.3697
	RE	0.4990	-0.0262	0.2301	1	0.0230	0.2784
	Treasury	-0.0056	0.0237	0.2618	0.0230	1	0.2676
	High Yield	0.3021	-0.0040	0.3697	0.2784	0.2676	1
normal cop with mixnormal marg	Equities	1	-0.0244	0.3186	0.5052	-0.0231	0.3002
	Commodities	-0.0244	1	0.0440	-0.0515	0.0093	-0.0065
	EMD	0.3186	0.0440	1	0.2411	0.2376	0.3596
	RE	0.5052	-0.0515	0.2411	1	0.0051	0.2821
	Treasury	-0.0231	0.0093	0.2376	0.0051	1	0.2177
	High Yield	0.3002	-0.0065	0.3596	0.2821	0.2177	1

Table 13: Σ of different models for data Apr90-May09

Model	Σ	Equities	Commodities	EMD	RE	Treasury	High Yield
t cop with optmix marg	Equities	1	0.0145	0.3361	0.5592	-0.0544	0.3613
	Commodities	0.0145	1	0.0496	-0.0144	0.0146	0.0134
	EMD	0.3361	0.0496	1	0.2465	0.2501	0.3929
	RE	0.5592	-0.0144	0.2465	1	-0.0238	0.3397
	Treasury	-0.0544	0.0146	0.2501	-0.0238	1	0.2024
	High Yield	0.3613	0.0134	0.3929	0.3397	0.2024	1
normal cop with optmix marg	Equities	1	0.0467	0.3494	0.5687	-0.0894	0.3777
	Commodities	0.0467	1	0.0840	-0.0009	-0.0242	0.0453
	EMD	0.3494	0.0840	1	0.2764	0.2090	0.4086
	RE	0.5687	-0.0009	0.2764	1	-0.0541	0.3551
	Treasury	-0.0894	-0.0242	0.2090	-0.0541	1	0.1286
	High Yield	0.3777	0.0453	0.4086	0.3551	0.1286	1

		Equities	Commodities	EMD	RE	Treasury	HY
t cop with mixnormal marg	Equities	1	0.0121	0.3367	0.5599	-0.0537	0.3618
	Commodities	0.0121	1	0.0499	-0.0155	0.0130	0.0134
	EMD	0.3367	0.0499	1	0.2467	0.2469	0.3917
	RE	0.5599	-0.0155	0.2467	1	-0.0263	0.3393
	Treasury	-0.0537	0.0130	0.2469	-0.0263	1	0.2027
	High Yield	0.3618	0.0134	0.3917	0.3393	0.2027	1
normal cop with mixnormal marg	Equities	1	0.0453	0.3494	0.5691	-0.0883	0.3750
	Commodities	0.0453	1	0.0847	-0.0014	-0.0258	0.0454
	EMD	0.3494	0.0847	1	0.2771	0.2053	0.4083
	RE	0.5691	-0.0014	0.2771	1	-0.0571	0.3505
	Treasury	-0.0883	-0.0258	0.2053	-0.0571	1	0.1263
	High Yield	0.3750	0.0454	0.4083	0.3505	0.1263	1

Table 14: μ of different models

Model	$\mu_{Eq.}$	$\mu_{Com.}$	μ_{EMD}	μ_{RE}	$\mu_{Treas.}$	μ_{HY}
Data Apr90-Dec07	0.0015	0.0013	0.0015	0.0016	0.0011	0.0013
multivariate normal	0.0019	0.0007	0.0016	0.0018	0.0010	0.0013
multivariate t	0.0019	0.0021	0.0024	0.0026	0.0011	0.0019
multivariate skewed t	0.0017	0.0012	0.0015	0.0018	0.0010	0.0013
normal cop with mixnormal marg	0.0013	0.0013	0.0015	0.0017	0.0011	0.0013
t cop with mixnormal marg	0.0013	0.0013	0.0016	0.0017	0.0011	0.0013
st cop with mixnormal marg	0.0015	0.0017	0.0016	0.0016	0.0011	0.0012
normal cop with optmix marg	0.0013	0.0013	0.0010	0.0017	0.0011	0.0013
t cop with optmix marg	0.0012	0.0013	0.0007	0.0016	0.0011	0.0013
st cop with optmix marg	0.0014	0.0014	0.0004	0.0016	0.0010	0.0012

Model	$\mu_{Eq.}$	$\mu_{Com.}$	μ_{EMD}	μ_{RE}	$\mu_{Treas.}$	μ_{HY}
Data Apr90-May09	0.0015	0.0013	0.0015	0.0016	0.0011	0.0013
multivariate normal	0.0014	0.0015	0.0013	0.0016	0.0009	0.0013
multivariate t	0.0024	0.0026	0.0023	0.0025	0.0011	0.0019
multivariate skewed t	0.0019	0.0014	0.0015	0.0016	0.0011	0.0013
normal cop with mixnormal marg	0.0013	0.0012	0.0015	0.0017	0.0011	0.0013
t cop with mixnormal marg	0.0013	0.0015	0.0015	0.0015	0.0011	0.0013
st cop with mixnormal marg	0.0012	0.0016	0.0016	0.0013	0.0011	0.0011
normal cop with optmix marg	0.0015	0.0013	0.0010	0.0007	0.0011	0.0015
t cop with optmix marg	0.0016	0.0014	-0.0005	0.0013	0.0011	0.0015
st cop with optmix marg	0.0017	0.0017	0.0006	0.0006	0.0011	0.0016

Table 15: σ of different models

Model	$\sigma_{Eq.}$	$\sigma_{Com.}$	σ_{EMD}	σ_{RE}	$\sigma_{Treas.}$	σ_{HY}
Data Apr90-Dec07	0,0206	0,0269	0,0164	0,0193	0,0058	0,0067
multivariate normal	0,0204	0,0270	0,0165	0,0190	0,0058	0,0068
multivariate t	0,0212	0,0289	0,0153	0,0200	0,0063	0,0059
multivariate skewed t	0,0214	0,0302	0,0155	0,0206	0,0063	0,0065
normal cop with mixnormal marg	0,0209	0,0276	0,0169	0,0191	0,0058	0,0068
t cop with mixnormal marg	0,0209	0,0279	0,0165	0,0191	0,0058	0,0069
st cop with mixnormal marg	0,0213	0,0271	0,0166	0,0191	0,0059	0,0069
normal cop with optmix marg	0,0209	0,0276	0,0248	0,0191	0,0058	0,0068
t cop with optmix marg	0,0212	0,0276	0,0360	0,0192	0,0058	0,0068
st cop with optmix marg	0,0210	0,0277	0,0786	0,0189	0,0059	0,0070

Model	$\sigma_{Eq.}$	$\sigma_{Com.}$	σ_{EMD}	σ_{RE}	$\sigma_{Treas.}$	σ_{HY}
Data Apr90-May09	0,0238	0,0303	0,0174	0,0287	0,0060	0,0094
multivariate normal	0,0237	0,0299	0,0175	0,0288	0,0060	0,0094
multivariate t	0,0240	0,0361	0,0165	0,0249	0,0072	0,0073
multivariate skewed t	0,0245	0,0352	0,0167	0,0261	0,0072	0,0079
normal cop with mixnormal marg	0,0241	0,0312	0,0180	0,0288	0,0060	0,0096
t cop with mixnormal marg	0,0244	0,0310	0,0181	0,0299	0,0061	0,0090
st cop with mixnormal marg	0,0246	0,0304	0,0174	0,0290	0,0061	0,0104
normal cop with optmix marg	0,0238	0,0314	0,0259	0,0807	0,0060	0,0114
t cop with optmix marg	0,0241	0,0308	0,1575	0,0404	0,0059	0,0100
st cop with optmix marg	0,0253	0,0304	0,0610	0,0510	0,0061	0,0116

Table 16: Skewness of different models

Model	Eq.	Com.	EMD	RE	Treas.	HY
Data Apr90-Dec07	-0,46	-0,49	-1,65	-0,54	-0,61	-1,49
multivariate normal	0,00	0,06	0,00	0,01	0,04	0,00
multivariate t	0,38	0,03	0,03	0,21	-0,12	-0,22
multivariate skewed t	-0,43	-0,64	-0,97	-0,74	-0,66	-3,49
normal cop with mixnormal marg	-0,39	-0,56	-1,87	-0,33	-0,69	-1,38
t cop with mixnormal marg	-0,33	-0,57	-1,48	-0,32	-0,59	-1,12
st cop with mixnormal marg	-0,50	-0,46	-1,79	-0,43	-0,55	-1,11
normal cop with optmix marg	-0,39	-0,56	-10,12	-0,33	-0,95	-1,38
t cop with optmix marg	-0,54	-0,55	-30,51	-0,29	-0,77	-1,44
st cop with optmix marg	-0,42	-0,65	-83,83	-0,42	-0,93	-1,34

Model	Eq.	Com.	EMD	RE	Treas.	HY
Data Apr90-May09	-0,82	-0,87	-1,92	-0,70	-0,51	-3,04
multivariate normal	-0,05	-0,04	0,00	-0,03	-0,01	-0,02
multivariate t	0,26	0,01	-0,81	0,06	-0,41	-0,62
multivariate skewed t	-0,32	-1,08	-1,41	-1,88	-0,14	-2,41
normal cop with mixnormal marg	-0,62	-0,89	-2,15	-1,22	-0,59	-3,62
t cop with mixnormal marg	-0,65	-0,73	-2,19	-1,50	-0,62	-2,23
st cop with mixnormal marg	-0,91	-0,85	-1,92	-1,93	-0,45	-3,87
normal cop with optmix marg	-0,11	-0,96	-9,75	-75,14	-0,63	-7,49
t cop with optmix marg	-0,14	-0,70	-90,31	-18,78	-0,49	0,97
st cop with optmix marg	0,52	-0,83	-77,94	-31,85	-0,47	10,12

Table 17: Kurtosis of different models

Model	Eq.	Com.	EMD	RE	Treas.	HY
Data Apr90-Dec07	6,1	5,2	17,2	6,1	4,9	13,8
multivariate normal	3,0	3,0	3,1	3,1	3,0	2,9
multivariate t	14,4	6,7	9,3	8,3	11,2	22,5
multivariate skewed t	7,9	10,0	15,7	9,5	14,8	71,8
normal cop with mixnormal marg	6,2	5,6	19,5	6,2	5,0	14,8
t cop with mixnormal marg	6,0	5,7	16,0	5,9	4,9	13,1
st cop with mixnormal marg	6,5	5,3	19,7	5,9	4,8	12,8
normal cop with optmix marg	6,2	5,6	243,2	6,2	7,4	14,8
t cop with optmix marg	6,9	5,5	1335,3	5,9	5,8	14,7
st cop with optmix marg	5,7	5,7	7801,3	6,0	7,8	13,2

Model	Eq.	Com.	EMD	RE	Treas.	HY
Data Apr90-May09	10,7	7,5	19,8	19,0	4,7	37,7
multivariate normal	2,9	2,9	3,0	3,0	3,0	3,0
multivariate t	11,9	28,4	20,5	15,1	26,1	14,8
multivariate skewed t	13,8	22,3	16,5	24,0	9,6	30,3
normal cop with mixnormal marg	9,6	7,5	21,0	24,4	4,8	48,3
t cop with mixnormal marg	10,8	7,0	20,5	24,0	5,2	28,7
st cop with mixnormal marg	10,1	7,0	20,2	25,3	4,6	45,2
normal cop with optmix marg	8,9	10,7	226,2	6754,4	5,3	272,8
t cop with optmix marg	9,2	6,2	8590,3	773,1	4,6	90,8
st cop with optmix marg	46,5	7,9	7104,6	1562,1	4,9	418,3

Table 18: 5% VaR for different models per marginal for data Apr90-Dec07

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-3,5%	-4,6%	-2,3%	-3,7%	-0,9%	-1,1%
mult var normal	-3,1%	-4,3%	-2,6%	-2,9%	-0,9%	-1,0%
mult var t	-3,1%	-4,3%	-2,1%	-2,8%	-0,8%	-0,7%
mult var st	-3,2%	-4,6%	-2,3%	-3,0%	-0,9%	-0,9%
normal cop with normal mix marg	-3,0%	-4,2%	-2,3%	-2,9%	-0,9%	-0,9%
t cop with normal mix marg	-3,1%	-4,2%	-2,4%	-3,0%	-0,9%	-0,9%
st cop with normal mix marg	-3,2%	-4,2%	-2,3%	-3,0%	-0,9%	-0,9%
normal cop with optimal mix marg	-3,0%	-4,2%	-2,2%	-2,9%	-0,9%	-0,9%
t cop with optimal mix marg	-3,0%	-4,3%	-2,3%	-3,0%	-0,9%	-0,9%
st cop with optimal mix marg	-3,1%	-4,3%	-2,2%	-3,0%	-0,9%	-1,0%

Table 19: 1% VaR for different models per marginal for data Apr90-Dec07

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-7,0%	-9,5%	-5,9%	-10,0%	-1,6%	-2,9%
mult var normal	-4,6%	-6,1%	-3,7%	-4,2%	-1,3%	-1,5%
mult var t	-5,2%	-7,4%	-3,8%	-5,1%	-1,5%	-1,3%
mult var st	-5,9%	-8,5%	-4,0%	-5,8%	-1,6%	-1,8%
normal cop with normal mix marg	-5,4%	-8,1%	-5,6%	-5,7%	-1,7%	-2,3%
t cop with normal mix marg	-5,3%	-8,5%	-5,7%	-5,7%	-1,6%	-2,4%
st cop with normal mix marg	-5,4%	-7,5%	-5,2%	-5,8%	-1,6%	-2,4%
normal cop with optimal mix marg	-5,4%	-8,1%	-6,0%	-5,7%	-1,7%	-2,3%
t cop with optimal mix marg	-5,4%	-8,0%	-6,1%	-5,7%	-1,7%	-2,1%
st cop with optimal mix marg	-5,5%	-8,3%	-5,4%	-5,6%	-1,6%	-2,4%

Table 20: 5% CVaR for different models per marginal for data Apr90-Dec07

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-5,7%	-7,6%	-4,6%	-7,4%	-1,4%	-2,4%
mult var normal	-4,0%	-5,4%	-3,3%	-3,7%	-1,1%	-1,3%
mult var t	-4,5%	-6,3%	-3,2%	-4,2%	-1,3%	-1,1%
mult var st	-4,8%	-7,1%	-3,6%	-4,8%	-1,4%	-1,5%
normal cop with normal mix marg	-4,7%	-6,6%	-4,5%	-4,6%	-1,4%	-1,7%
t cop with normal mix marg	-4,6%	-6,7%	-4,3%	-4,6%	-1,3%	-1,7%
st cop with normal mix marg	-4,8%	-6,2%	-4,3%	-4,7%	-1,4%	-1,8%
normal cop with optimal mix marg	-4,7%	-6,6%	-5,6%	-4,6%	-1,4%	-1,7%
t cop with optimal mix marg	-4,8%	-6,6%	-6,1%	-4,7%	-1,4%	-1,7%
st cop with optimal mix marg	-4,7%	-6,7%	-6,7%	-4,6%	-1,4%	-1,8%

Table 21: 1% CVaR for different models per marginal for data Apr90-Dec07

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-9,9%	-12,7%	-8,9%	-14,7%	-2,0%	-5,2%
mult var normal	-5,3%	-7,0%	-4,3%	-5,0%	-1,4%	-1,7%
mult var t	-6,8%	-9,9%	-5,3%	-6,6%	-2,1%	-1,8%
mult var st	-8,1%	-12,1%	-5,8%	-8,3%	-2,3%	-2,9%
normal cop with normal mix marg	-7,7%	-10,8%	-8,3%	-7,0%	-2,1%	-3,3%
t cop with normal mix marg	-7,4%	-11,1%	-7,8%	-6,9%	-2,0%	-3,3%
st cop with normal mix marg	-8,0%	-9,9%	-8,0%	-7,1%	-2,0%	-3,3%
normal cop with optimal mix marg	-7,7%	-10,8%	-14,3%	-7,0%	-2,2%	-3,3%
t cop with optimal mix marg	-8,2%	-10,7%	-16,7%	-6,9%	-2,2%	-3,3%
st cop with optimal mix marg	-7,5%	-11,0%	-20,5%	-6,9%	-2,3%	-3,4%

Table 22: Max Drawdown for different models per marginal for data Apr90-Dec07

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	54,6%	70,2%	36,6%	74,3%	6,9%	35,9%
mult var normal	44,3%	71,1%	32,1%	44,2%	17,1%	10,9%
mult var t	54,2%	63,1%	32,6%	43,8%	10,1%	10,5%
mult var st	47,4%	86,4%	33,1%	40,2%	11,0%	16,5%
normal cop with normal mix marg	70,6%	77,9%	43,9%	61,8%	10,3%	11,8%
t cop with normal mix marg	67,0%	77,3%	39,0%	55,8%	9,7%	13,7%
st cop with normal mix marg	52,7%	61,5%	36,2%	37,4%	13,6%	18,2%
normal cop with optimal mix marg	70,6%	77,9%	66,6%	61,8%	10,5%	11,8%
t cop with optimal mix marg	72,1%	71,9%	97,5%	56,0%	12,5%	14,0%
st cop with optimal mix marg	55,6%	67,1%	100,0%	34,4%	12,5%	16,7%

Table 23: 5% VaR for different models per marginal for data Apr90-May09

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-3,5%	-4,6%	-2,3%	-3,7%	-0,9%	-1,1%
mult var normal	-3,9%	-4,8%	-2,8%	-4,6%	-0,9%	-1,4%
mult var t	-3,3%	-4,6%	-2,2%	-3,4%	-0,9%	-0,9%
mult var st	-3,5%	-5,2%	-2,4%	-3,8%	-1,0%	-1,0%
normal cop with normal mix marg	-3,4%	-4,7%	-2,4%	-3,6%	-0,9%	-1,1%
t cop with normal mix marg	-3,5%	-4,6%	-2,4%	-3,6%	-1,0%	-1,1%
st cop with normal mix marg	-3,6%	-4,6%	-2,3%	-3,8%	-0,9%	-1,1%
normal cop with opt mix marg	-3,5%	-4,8%	-2,3%	-3,5%	-0,9%	-1,1%
t cop with optmix marg	-3,6%	-4,8%	-2,4%	-3,5%	-0,9%	-1,1%
st cop with optmix marg	-3,6%	-4,8%	-2,3%	-3,6%	-0,9%	-1,1%

Table 24: 1% VaR for different models per marginal for data Apr90-May09

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-7,0%	-9,5%	-5,9%	-10,0%	-1,6%	-2,9%
mult var normal	-5,3%	-6,8%	-3,9%	-6,7%	-1,4%	-2,1%
mult var t	-6,4%	-9,2%	-4,3%	-6,4%	-1,8%	-1,9%
mult var st	-6,6%	-9,9%	-4,9%	-7,4%	-2,0%	-2,4%
normal cop with normal mix marg	-7,0%	-10,5%	-6,0%	-9,4%	-1,7%	-3,1%
t cop with normal mix marg	-6,6%	-10,0%	-6,1%	-10,1%	-1,8%	-3,0%
st cop with normal mix marg	-7,5%	-10,2%	-5,8%	-9,8%	-1,7%	-3,3%
normal cop with opt mix marg	-6,5%	-10,0%	-6,4%	-9,1%	-1,7%	-2,7%
t cop with optmix marg	-6,4%	-9,6%	-5,7%	-8,6%	-1,6%	-2,8%
st cop with optmix marg	-6,6%	-9,1%	-5,6%	-8,8%	-1,6%	-2,9%

Table 25: 5% CVaR for different models per marginal for data Apr90-May09

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-5,7%	-7,6%	-4,6%	-7,4%	-1,4%	-2,4%
mult var normal	-4,7%	-6,0%	-3,5%	-5,8%	-1,2%	-1,8%
mult var t	-5,2%	-7,9%	-3,6%	-5,4%	-1,5%	-1,5%
mult var st	-5,6%	-8,4%	-4,1%	-6,5%	-1,6%	-2,0%
normal cop with normal mix marg	-5,7%	-8,0%	-4,8%	-7,3%	-1,4%	-2,4%
t cop with normal mix marg	-5,7%	-7,7%	-4,8%	-7,8%	-1,4%	-2,3%
st cop with normal mix marg	-6,1%	-7,6%	-4,5%	-7,7%	-1,4%	-2,7%
normal cop with opt mix marg	-5,4%	-7,9%	-5,8%	-9,3%	-1,4%	-2,3%
t cop with optmix marg	-5,4%	-7,7%	-8,8%	-8,2%	-1,4%	-2,2%
st cop with optmix marg	-5,7%	-7,5%	-6,5%	-9,1%	-1,4%	-2,2%

Table 26: 1% CVar for different models per marginal for data Apr90-May09

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	-9,9%	-12,7%	-8,9%	-14,7%	-2,0%	-5,2%
mult var normal	-6,1%	-7,8%	-4,6%	-7,7%	-1,5%	-2,4%
mult var t	-8,9%	-14,9%	-6,5%	-9,4%	-2,7%	-2,8%
mult var st	-9,7%	-14,8%	-7,8%	-12,1%	-2,8%	-3,8%
normal cop with normal mix marg	-10,4%	-13,4%	-9,8%	-15,3%	-2,1%	-5,4%
t cop with normal mix marg	-10,4%	-12,9%	-10,0%	-16,5%	-2,1%	-4,8%
st cop with normal mix marg	-11,0%	-13,0%	-9,2%	-16,4%	-2,0%	-6,4%
normal cop with opt mix marg	-9,0%	-13,4%	-15,2%	-25,8%	-2,0%	-5,3%
t cop with optmix marg	-8,9%	-12,3%	-30,3%	-20,5%	-1,9%	-4,8%
st cop with optmix marg	-9,8%	-12,3%	-19,1%	-24,0%	-2,0%	-4,6%

Table 27: Max Drawdown for different models per marginal for data Apr90-May09

Model	Equity	Com.	EMD	RE	Treas.	HY
Empirical data	54,6%	70,2%	36,6%	74,3%	6,9%	35,9%
mult var normal	55,0%	73,5%	45,0%	70,0%	10,2%	13,5%
mult var t	51,8%	74,9%	29,4%	45,9%	19,1%	10,5%
mult var st	59,0%	86,6%	35,3%	66,8%	20,3%	23,7%
normal cop with normal mix marg	77,7%	84,1%	46,8%	79,8%	11,2%	22,9%
t cop with normal mix marg	77,6%	83,9%	39,3%	82,2%	13,4%	18,2%
st cop with normal mix marg	57,7%	69,7%	44,2%	70,0%	14,5%	31,7%
normal cop with opt mix marg	73,5%	84,6%	68,2%	100,0%	10,9%	39,0%
t cop with optmix marg	69,3%	73,6%	100,0%	93,8%	12,0%	27,1%
st cop with optmix marg	53,9%	68,4%	99,7%	99,2%	13,6%	36,7%

D Figures

Figure 8: Index prices of assets

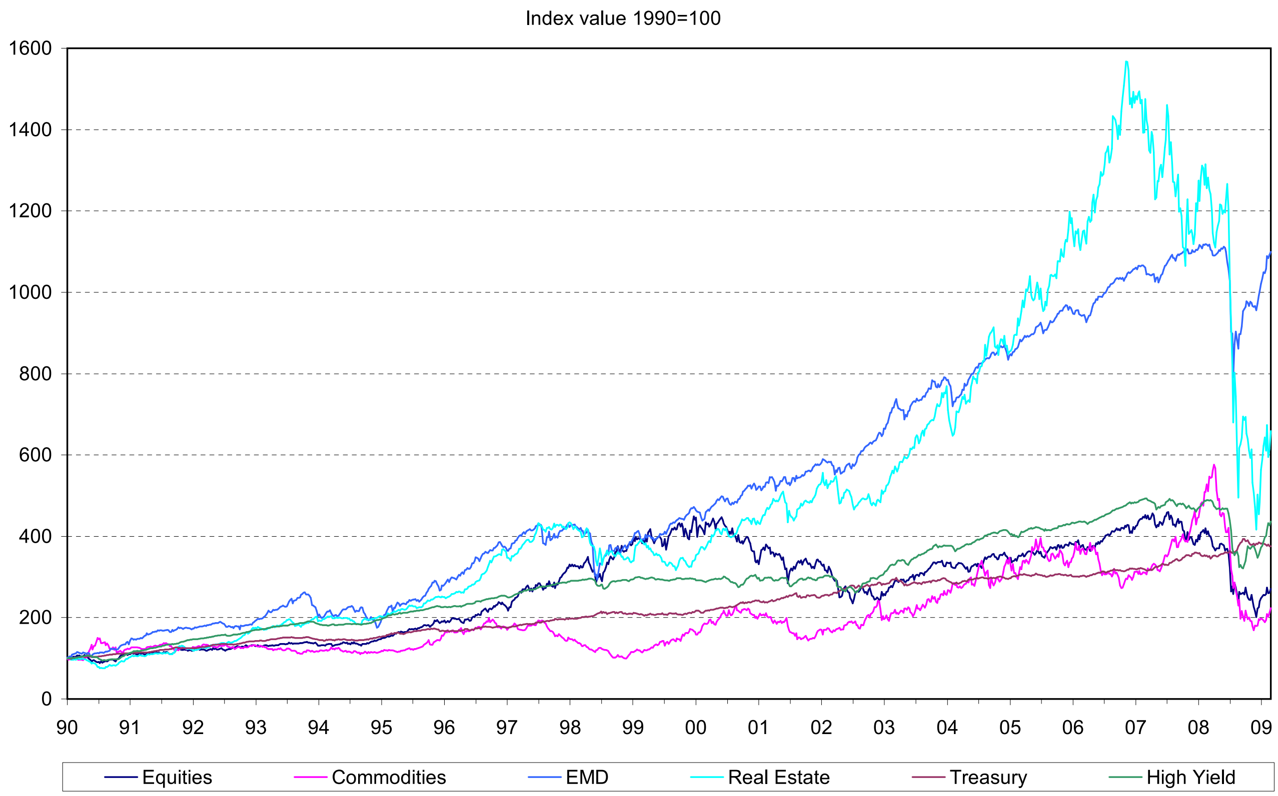


Figure 9: Weekly logreturns series

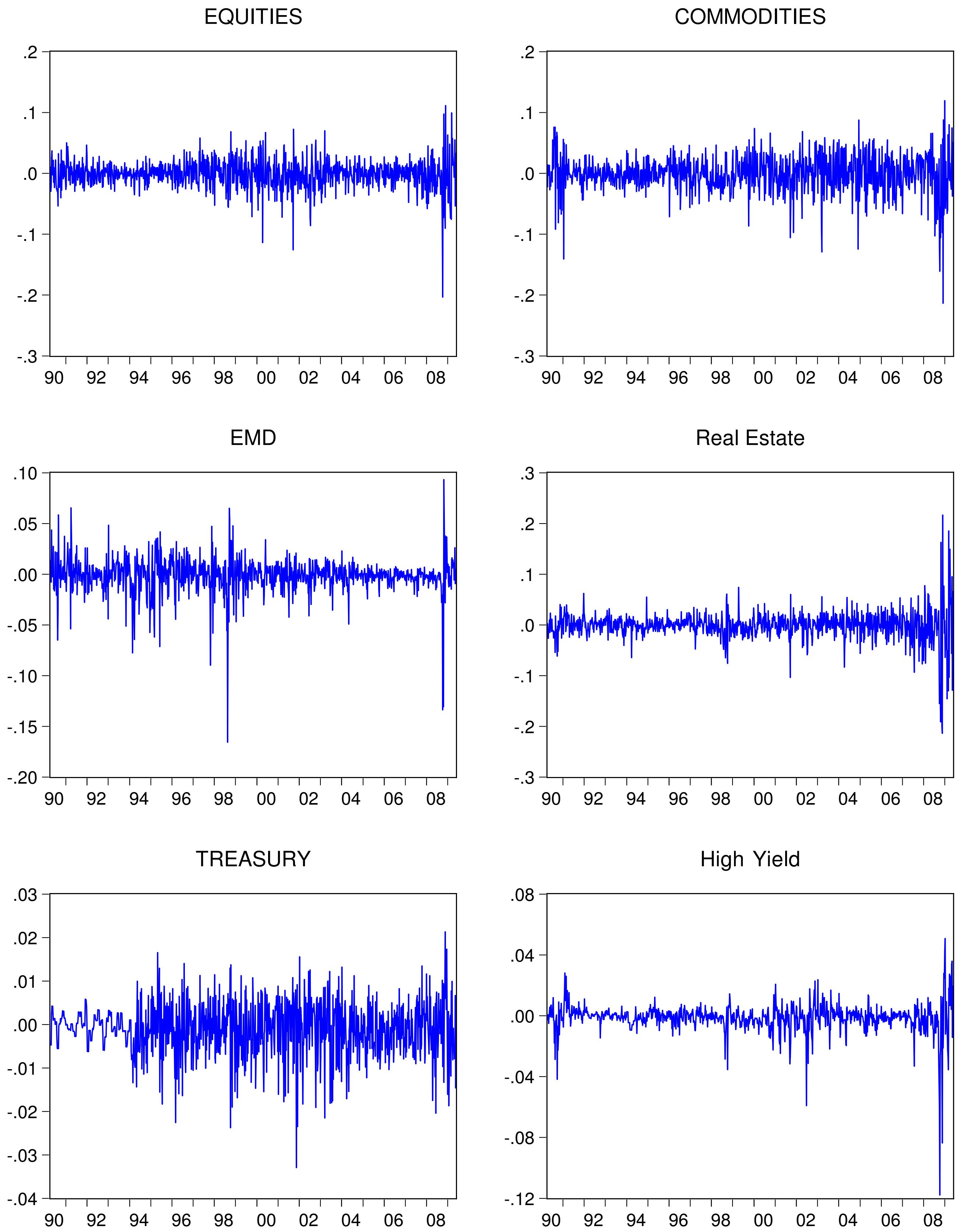


Figure 10: Scatterplot of EMD vs Real Estate

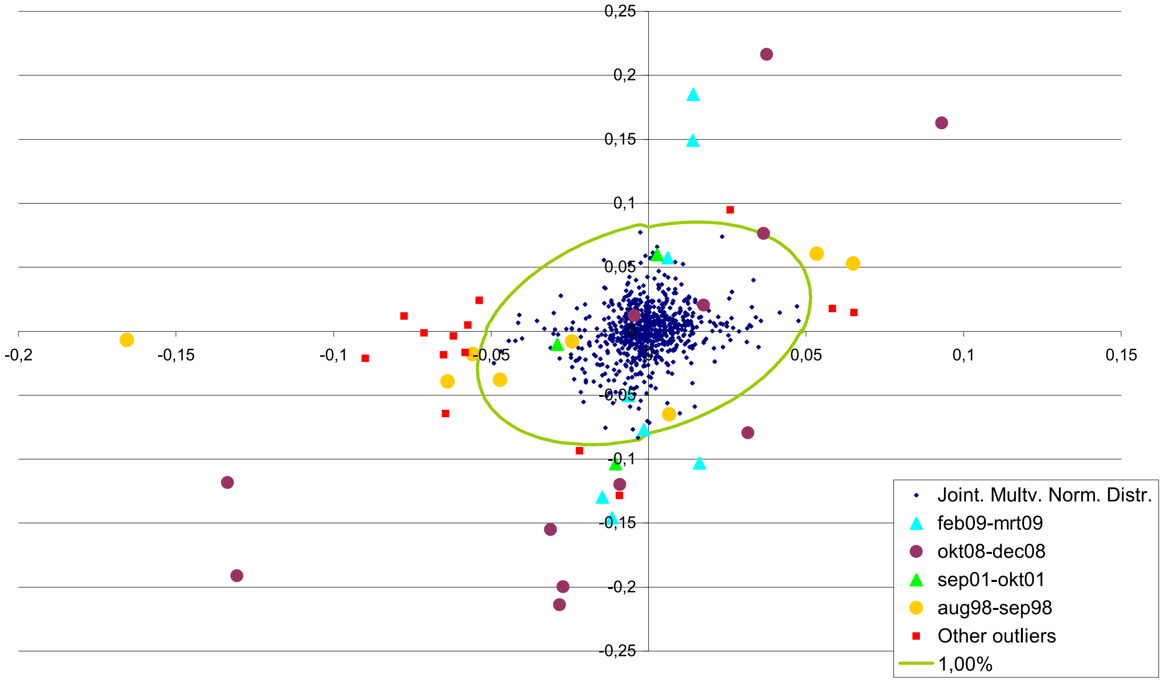


Figure 11: Scatterplot of Treasury vs High Yield

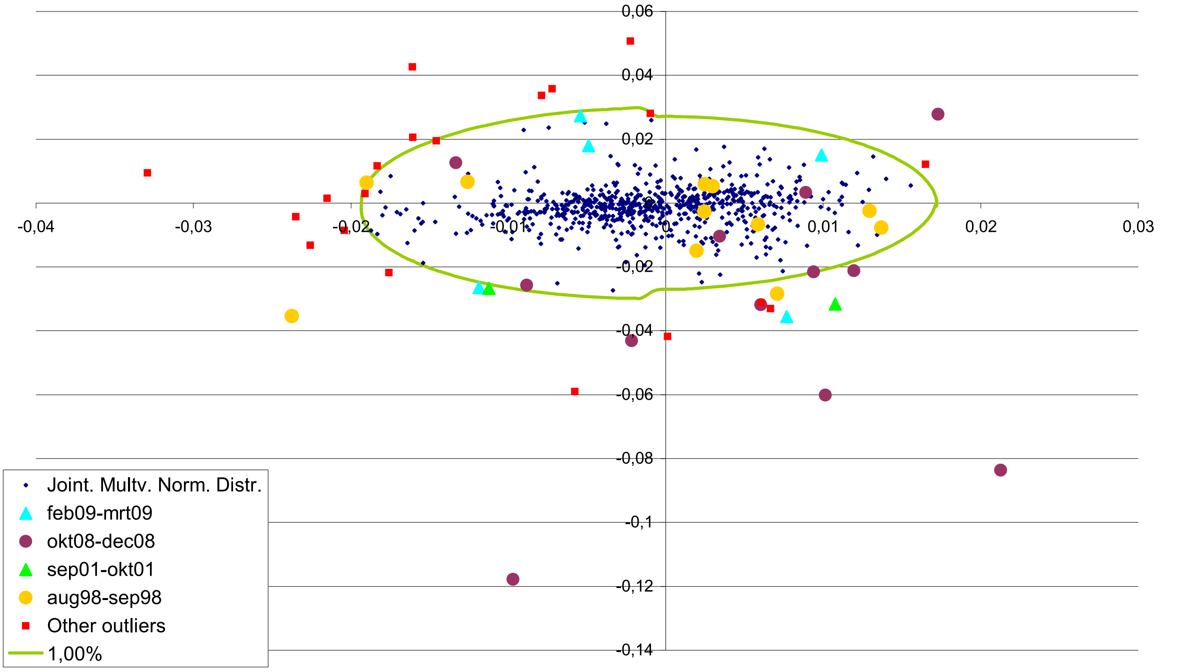


Figure 12: Histogram of the data series

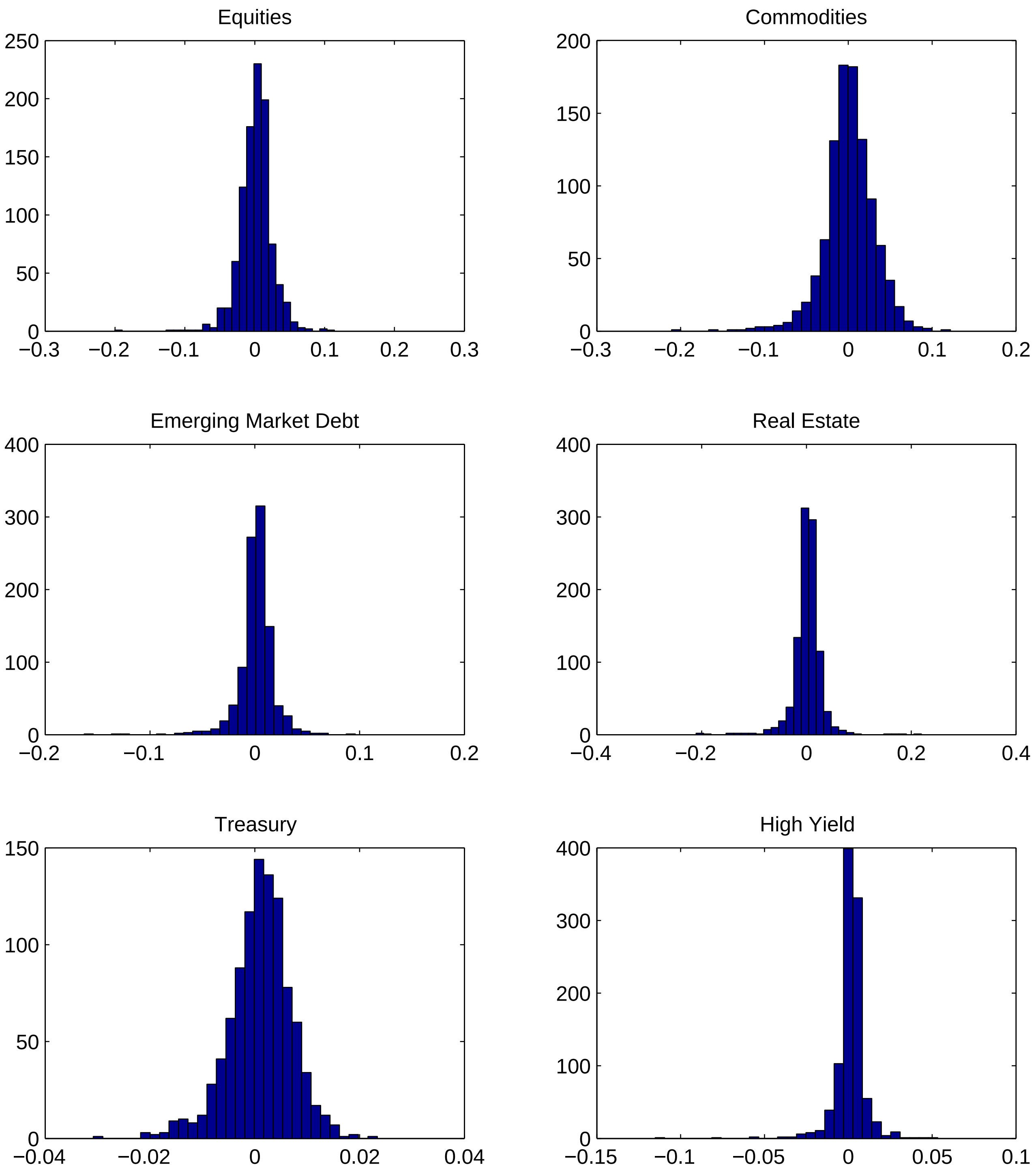


Figure 13: Histogram of fitted marginals with optimal mixture of Student's t or Skewed t components. The histograms of Equities, Commodities, Treasury and High Yield all show the combined mixture models of Student's t components (the total distribution is an accumulation of these different components). EMD and Real Estate are both fitted using a single Skewed t distribution

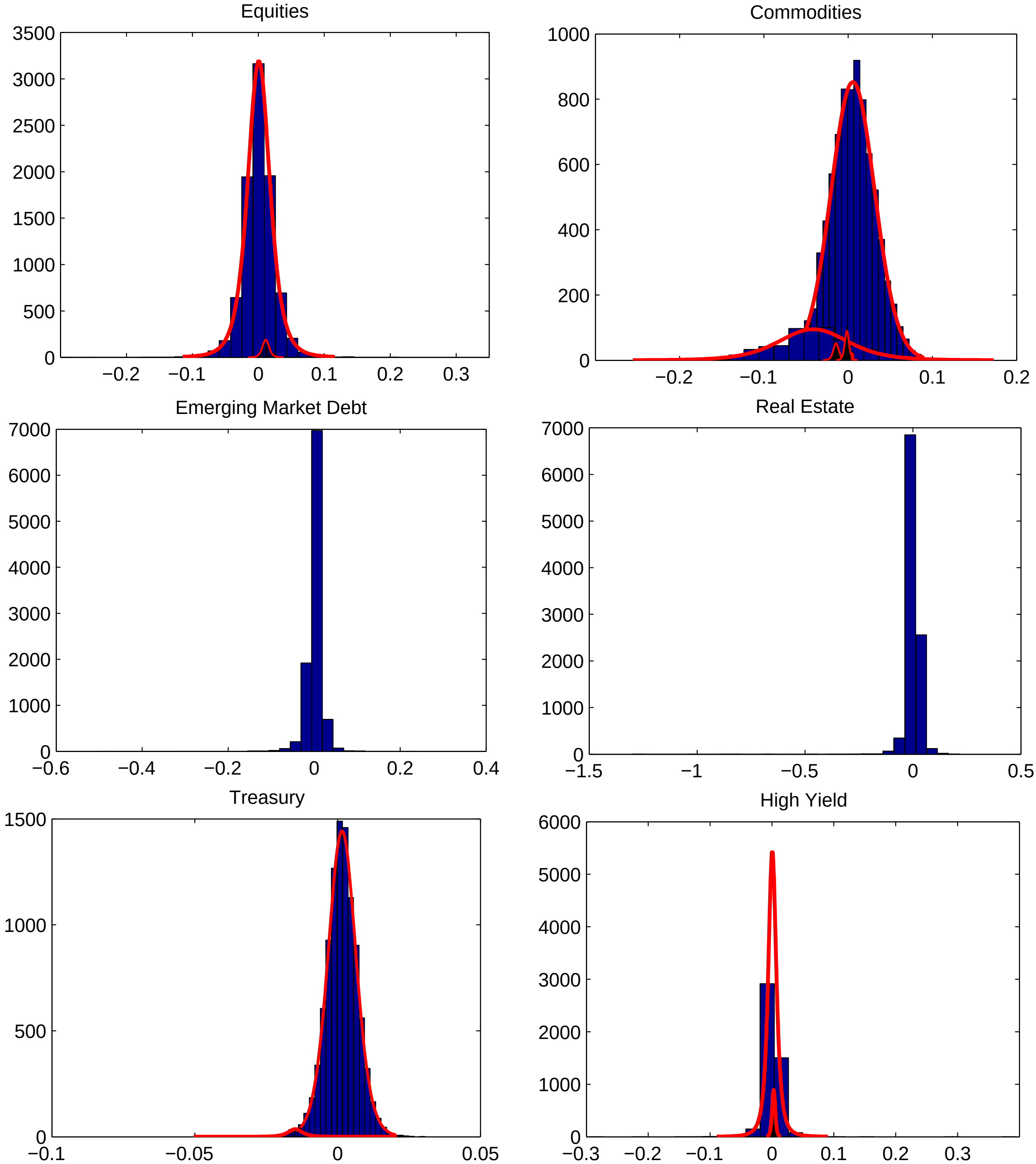


Figure 14: Histogram of fitted marginals with optimal mixture of normal components. The histograms show the combined mixture models of normal distributions (the total distribution is an accumulation of these different components).

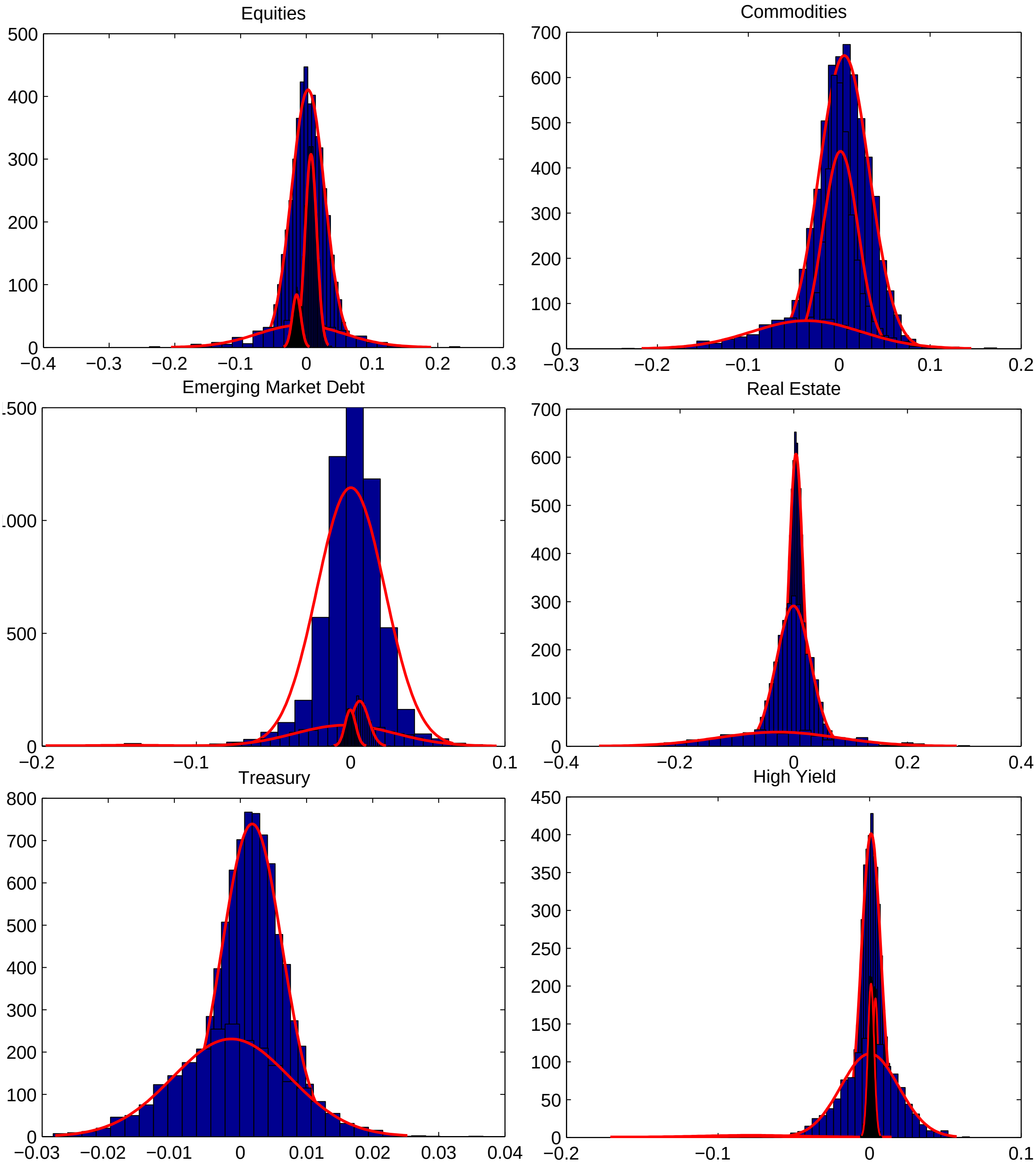


Figure 15: Scatter diagram of empirical data vs multivariate normal distribution. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

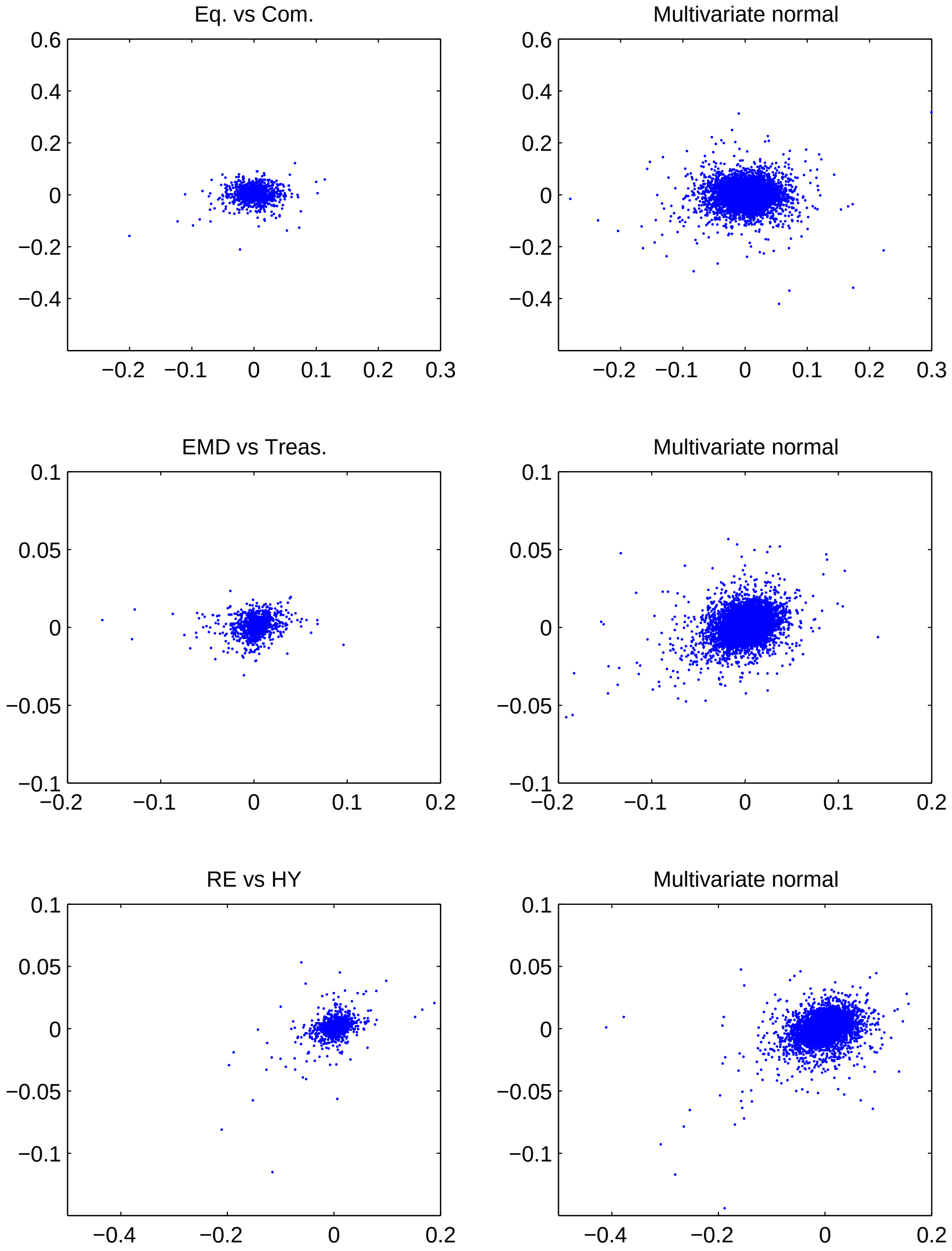


Figure 16: Scatter diagram of empirical data vs multivariate Student's t distribution. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

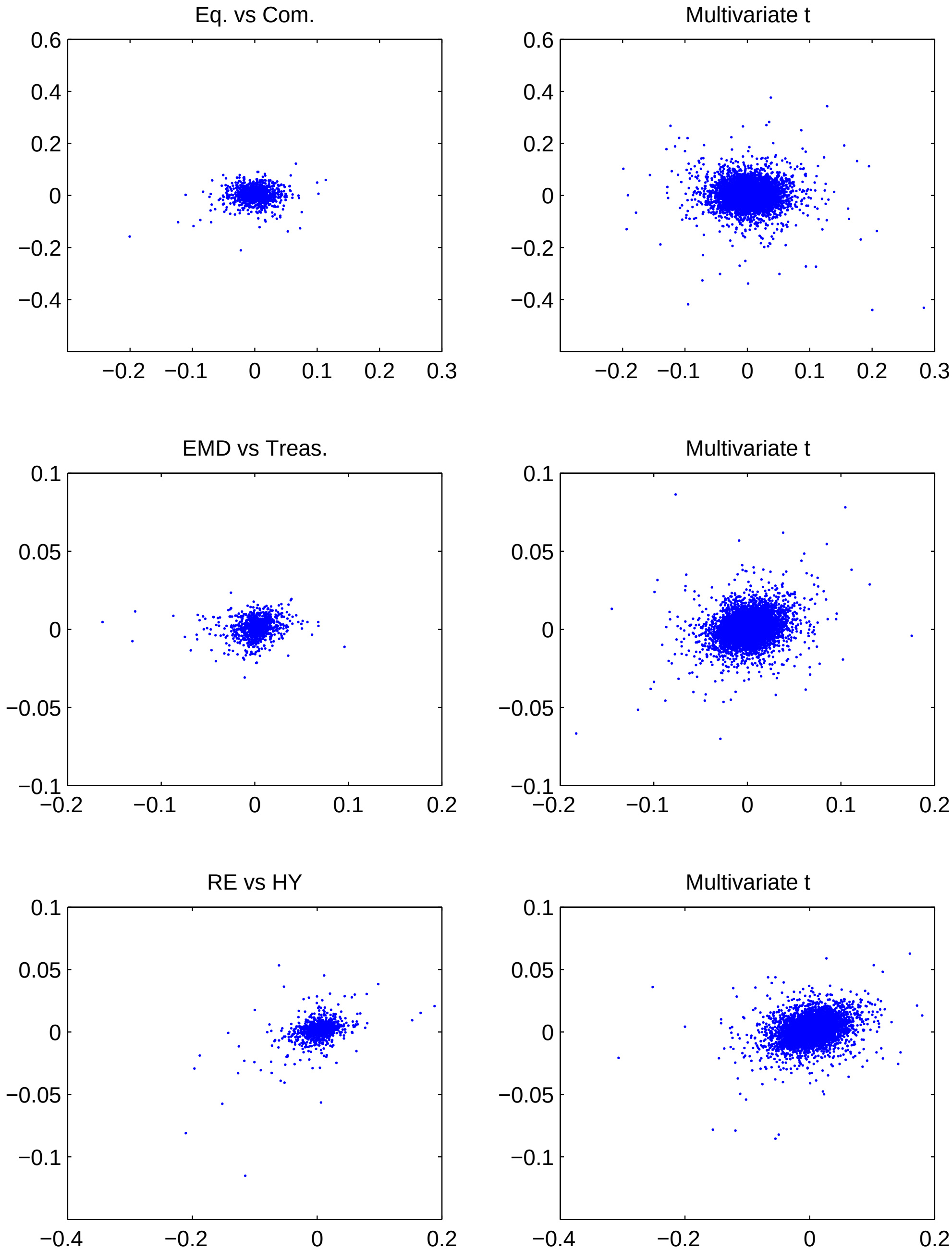


Figure 17: Scatter diagram of empirical data vs multivariate Skewed t distribution. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

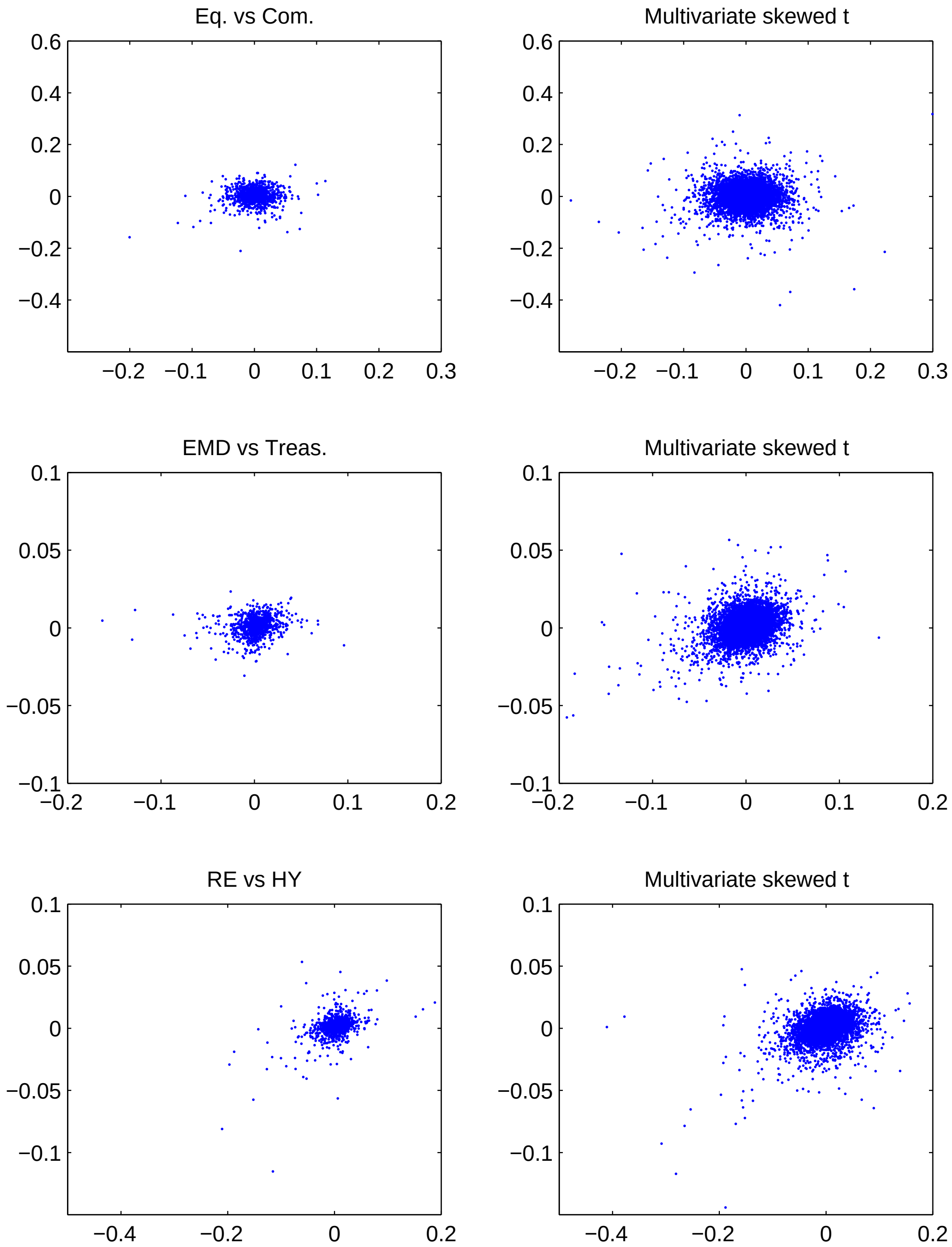


Figure 18: Scatter diagram of empirical data vs normal copula with mixture of normal components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

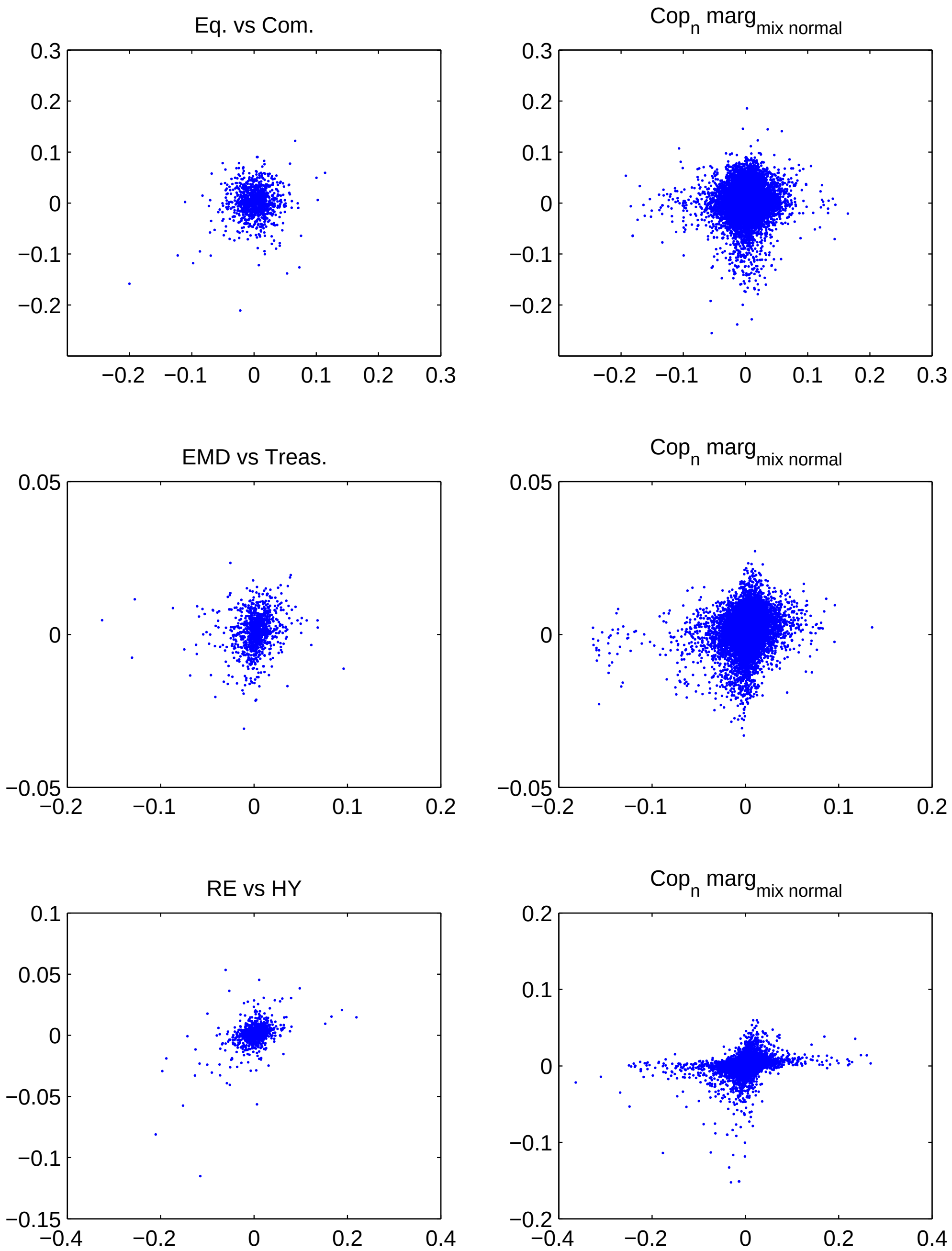


Figure 19: Scatter diagram of empirical data vs Student's t copula with mixture of normal components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

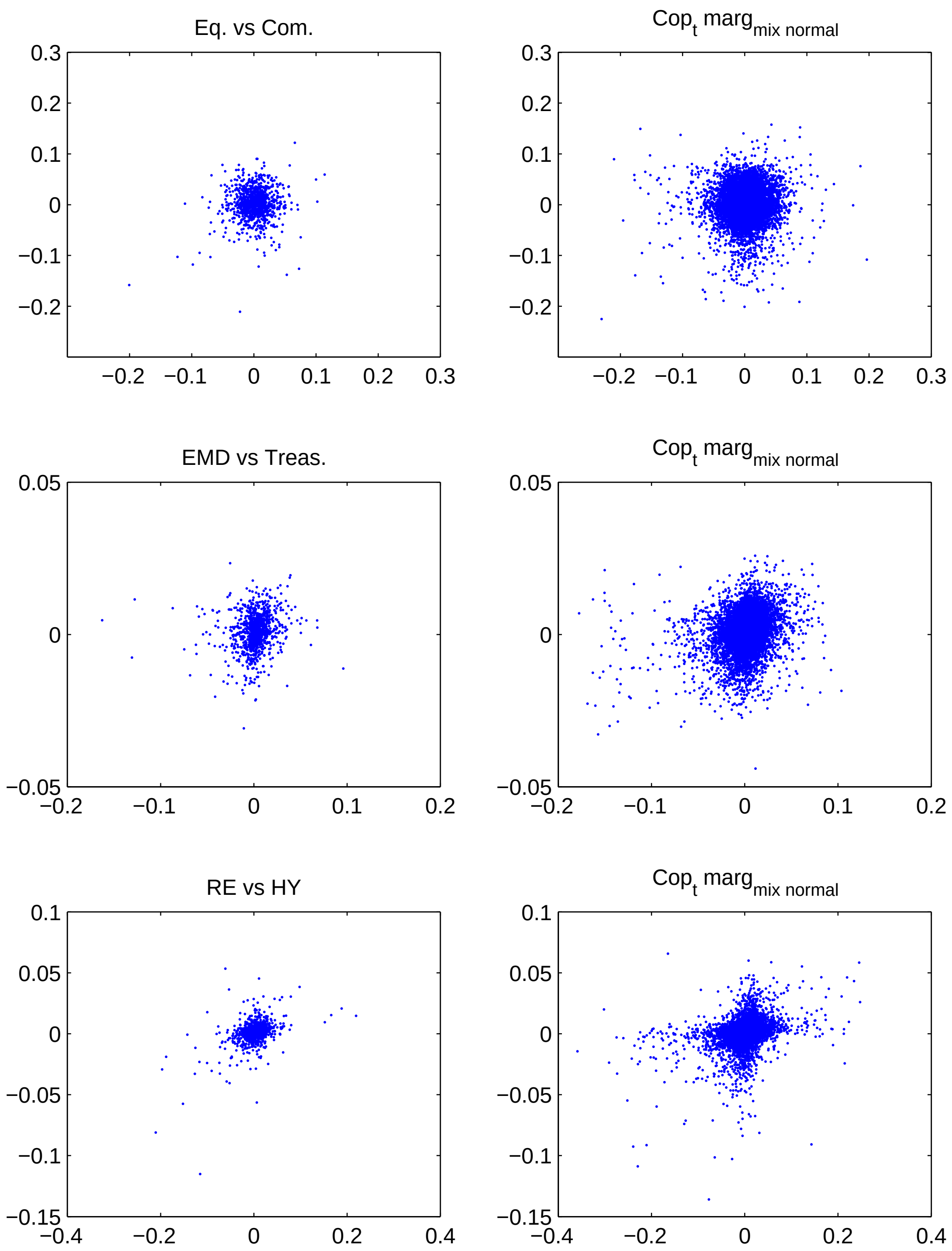


Figure 20: Scatter diagram of empirical data vs Skewed t copula with mixture of normal components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

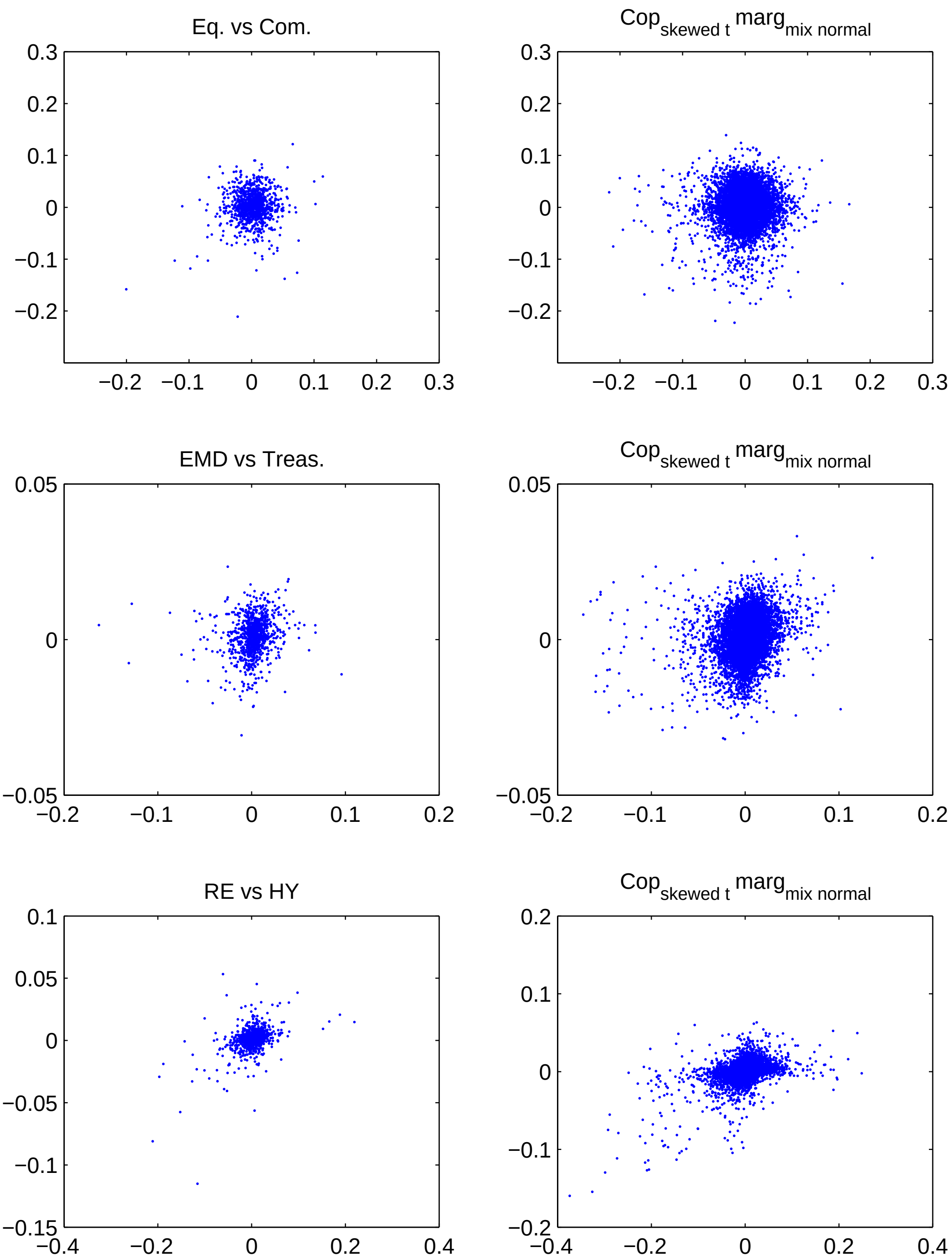


Figure 21: Scatter diagram of empirical data vs normal copula with mixture of Student's t or Skewed t components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

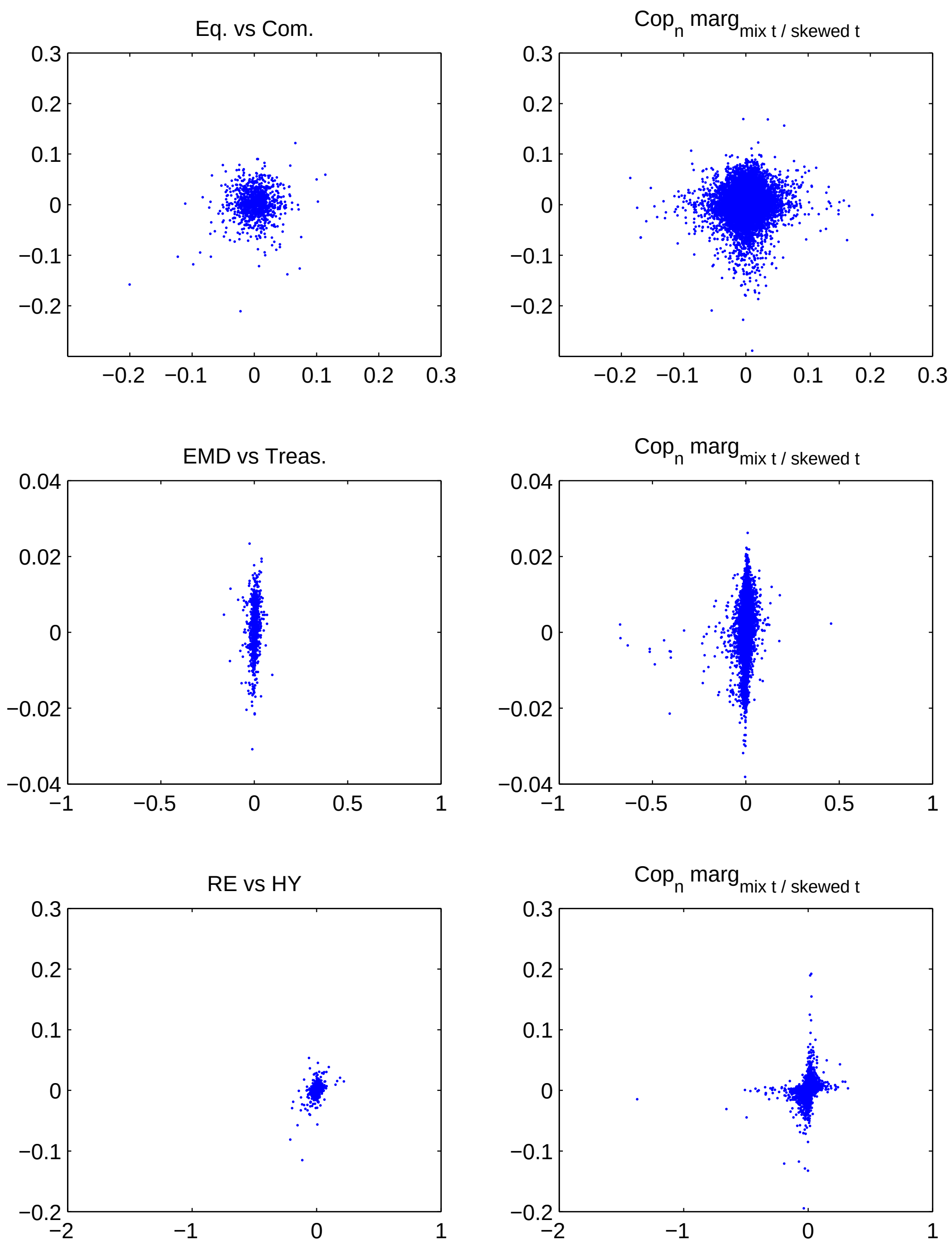


Figure 22: Scatter diagram of empirical data vs Skewed t copula with mixture of Student's t or Skewed t components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.

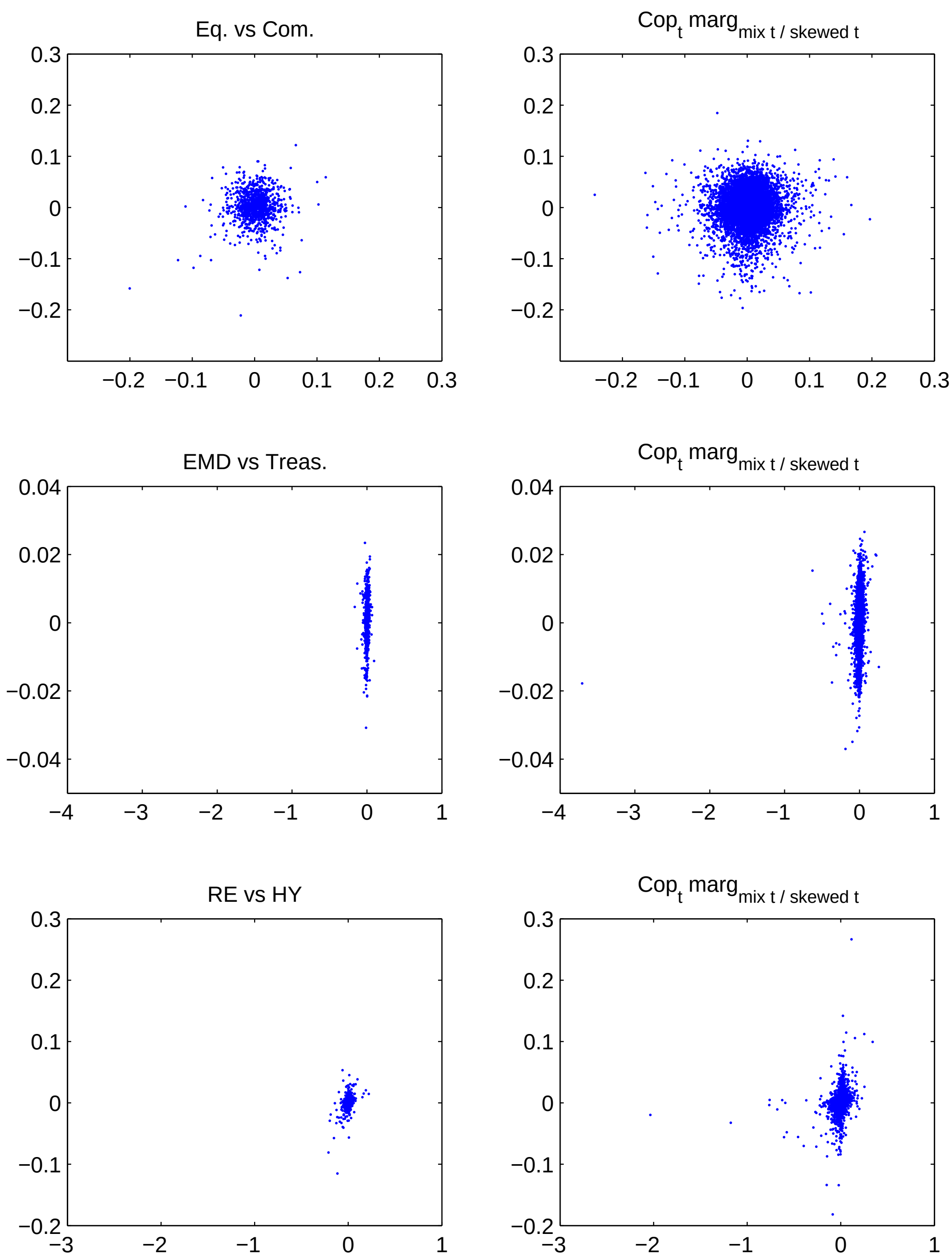
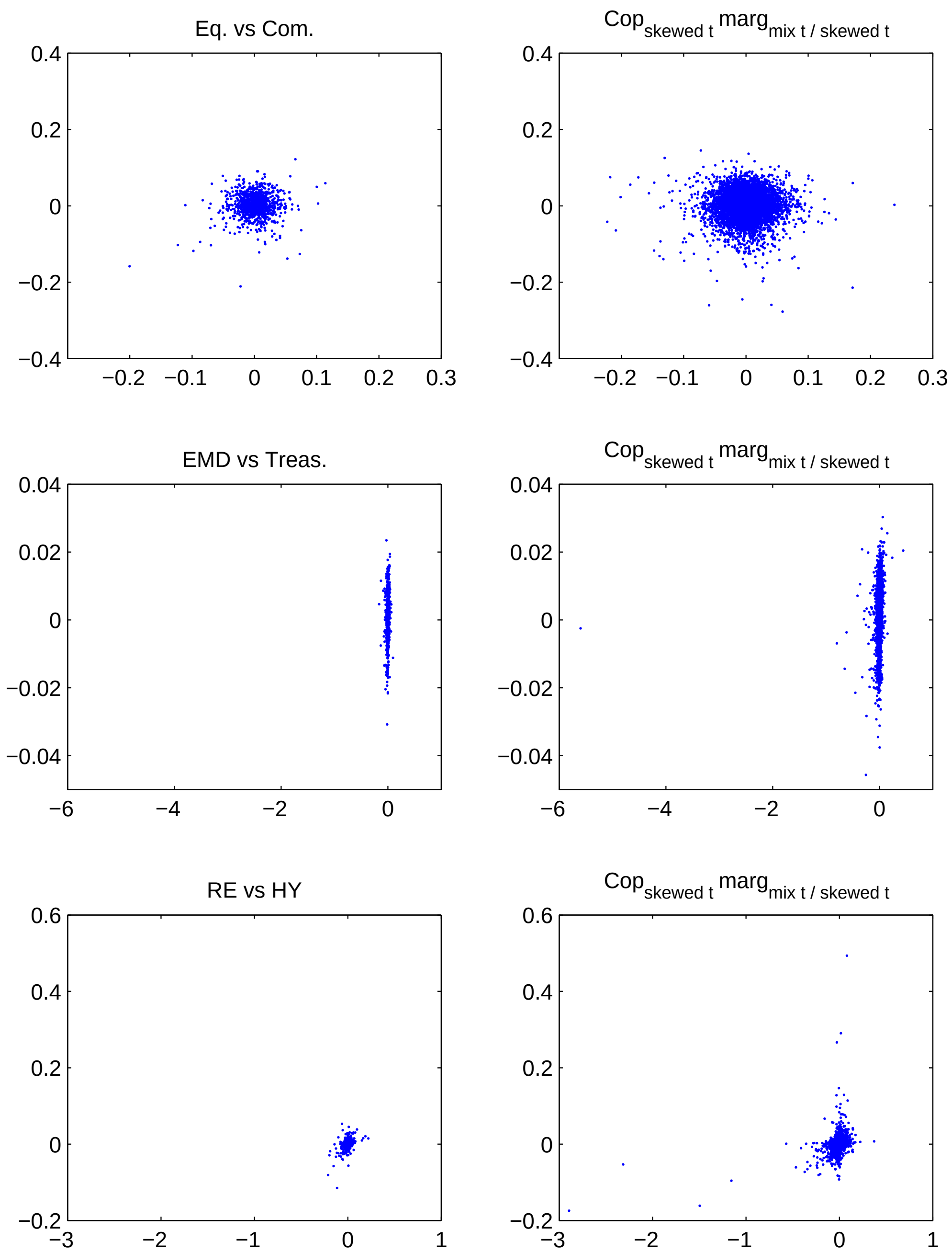


Figure 23: Scatter diagram of empirical data vs Skewed t copula with mixture of Student's t or Skewed t components. The empirical sample contains 1.000 observations while the simulated samples contain 10.000 observations.



E Optimization figures

Figure 24: Optimal portfolio weights of the assets over the different risk measures measures

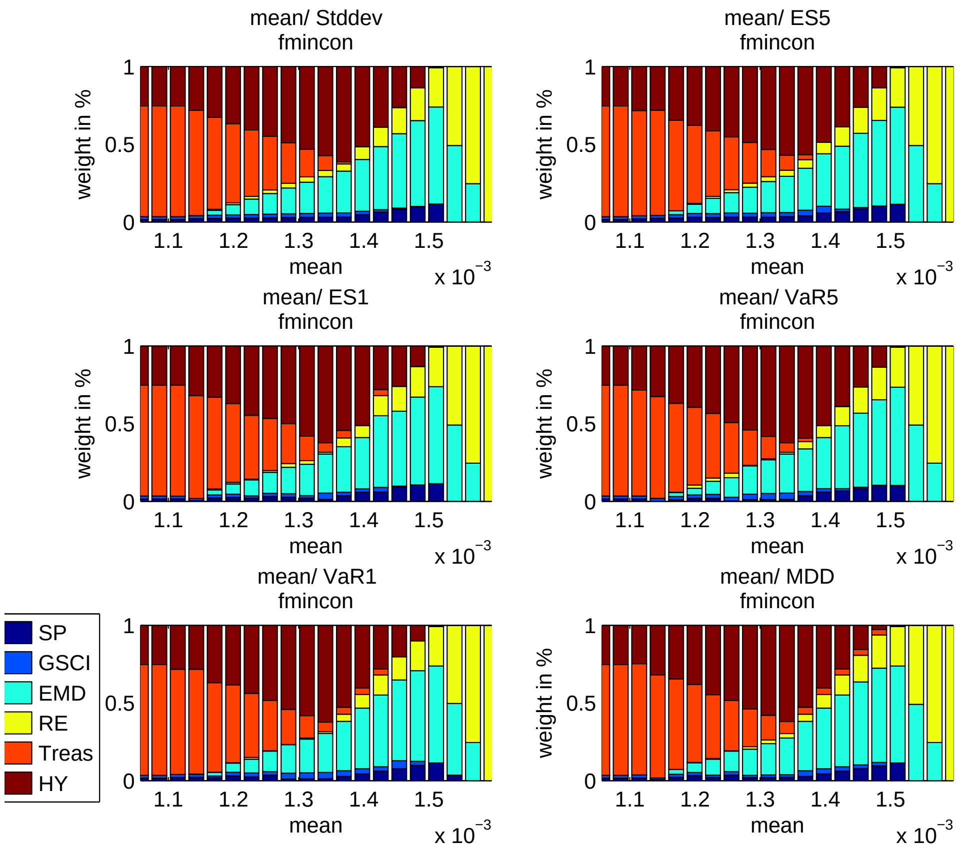


Figure 25: Optimal portfolio weights per model at minimized standard deviation

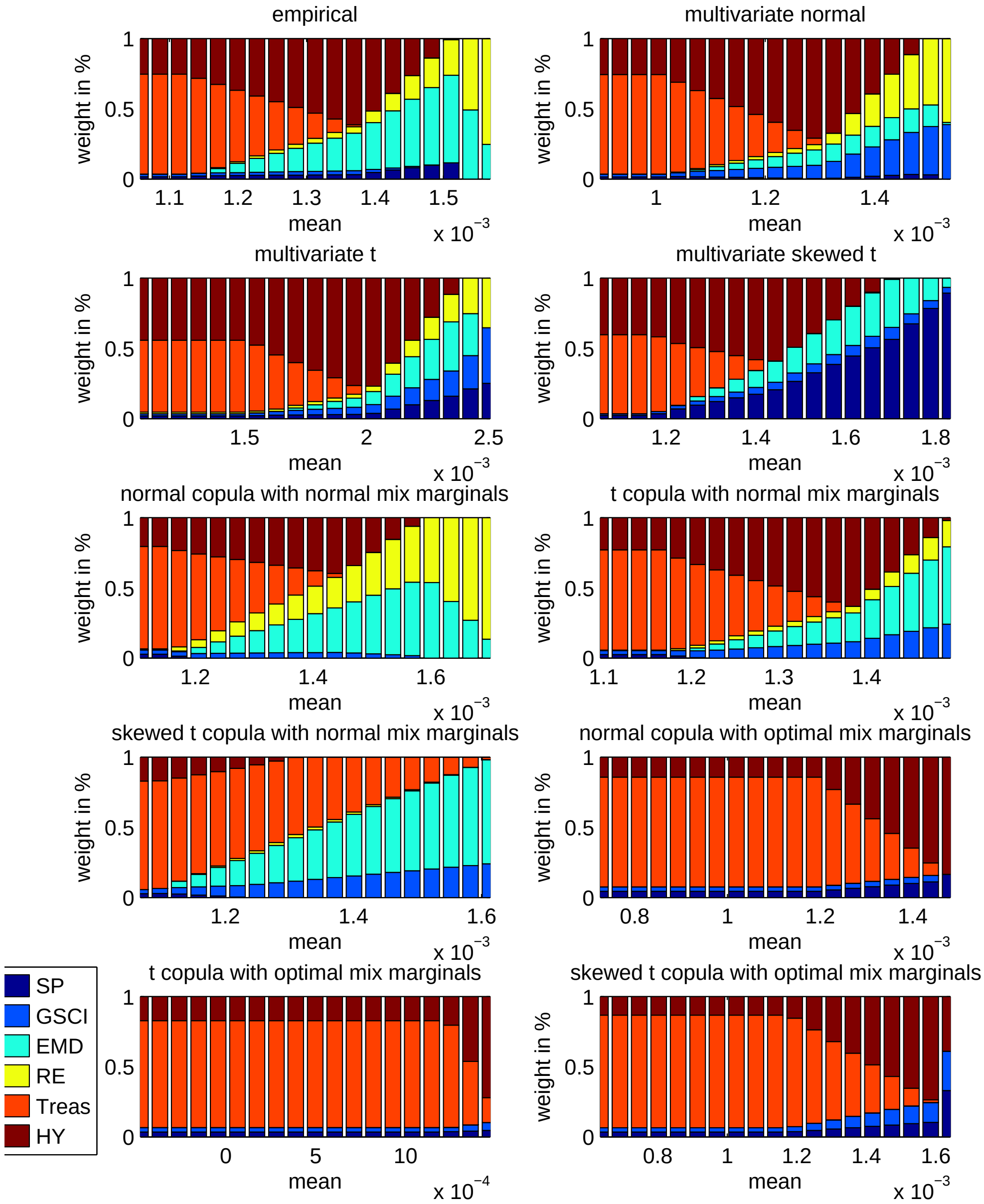


Figure 26: Optimal portfolio weights per model at minimized 5%VaR

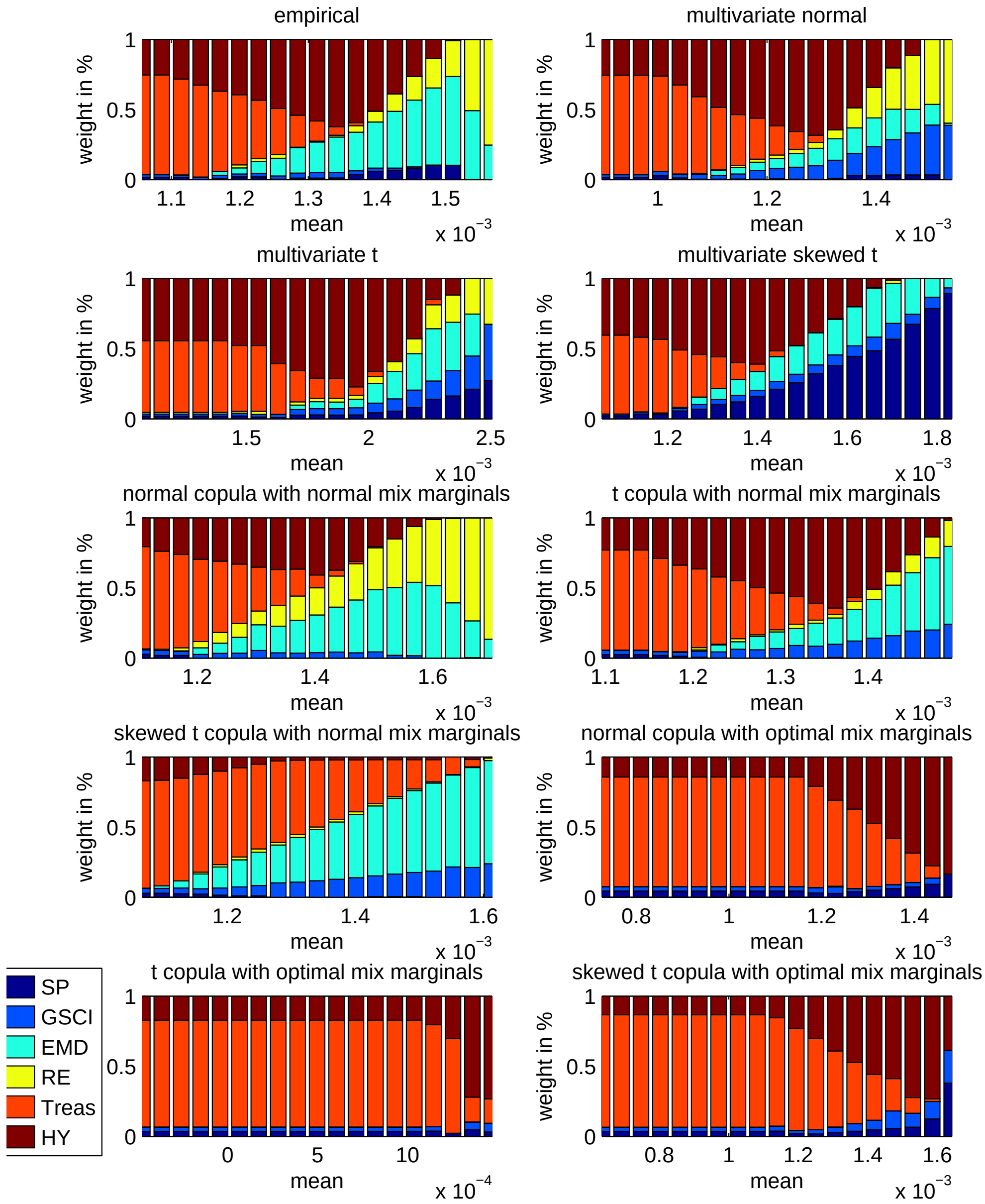


Figure 27: Optimal portfolio weights per model at minimized 1%VaR

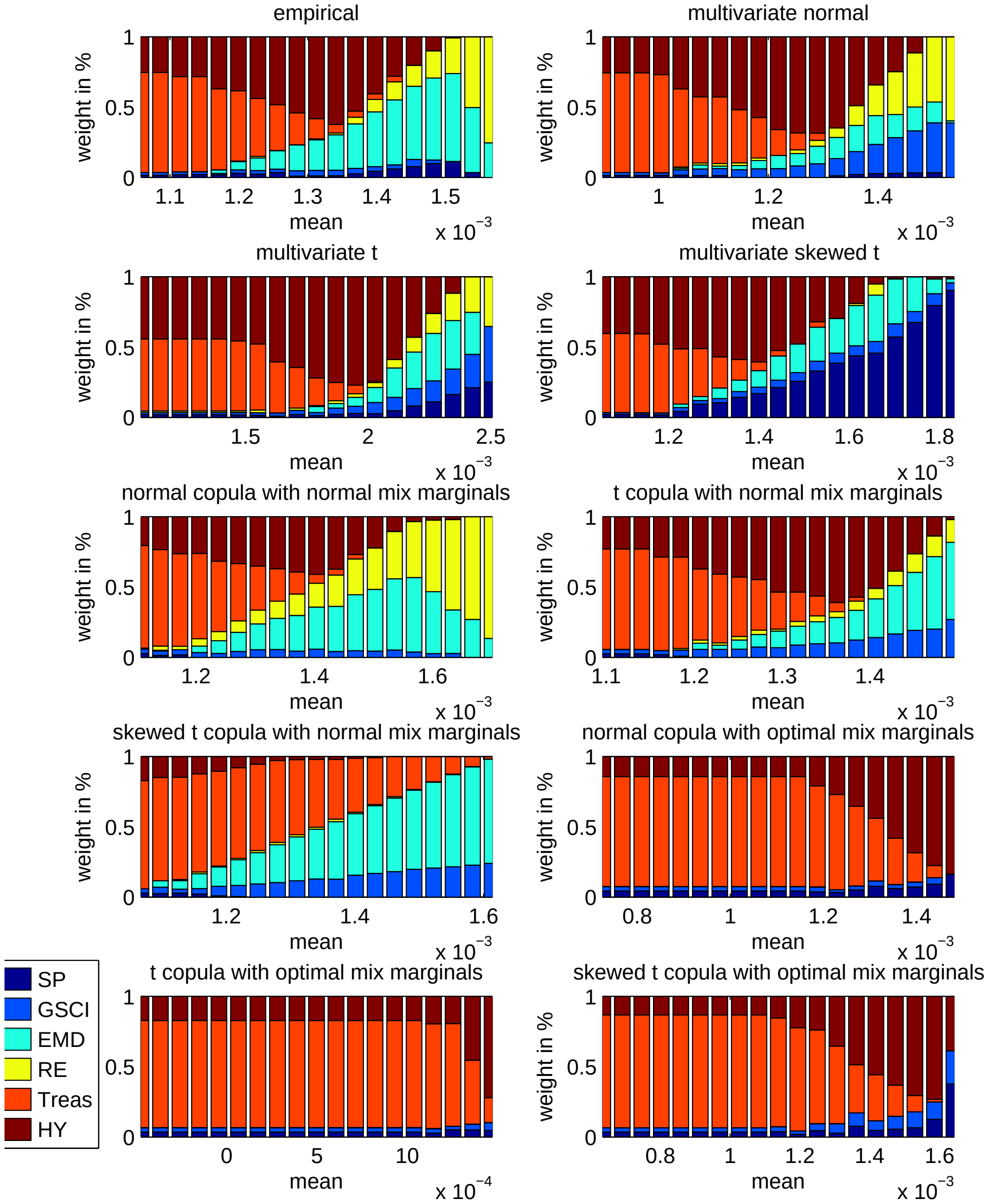


Figure 28: Optimal portfolio weights per model at minimized 5%CVaR

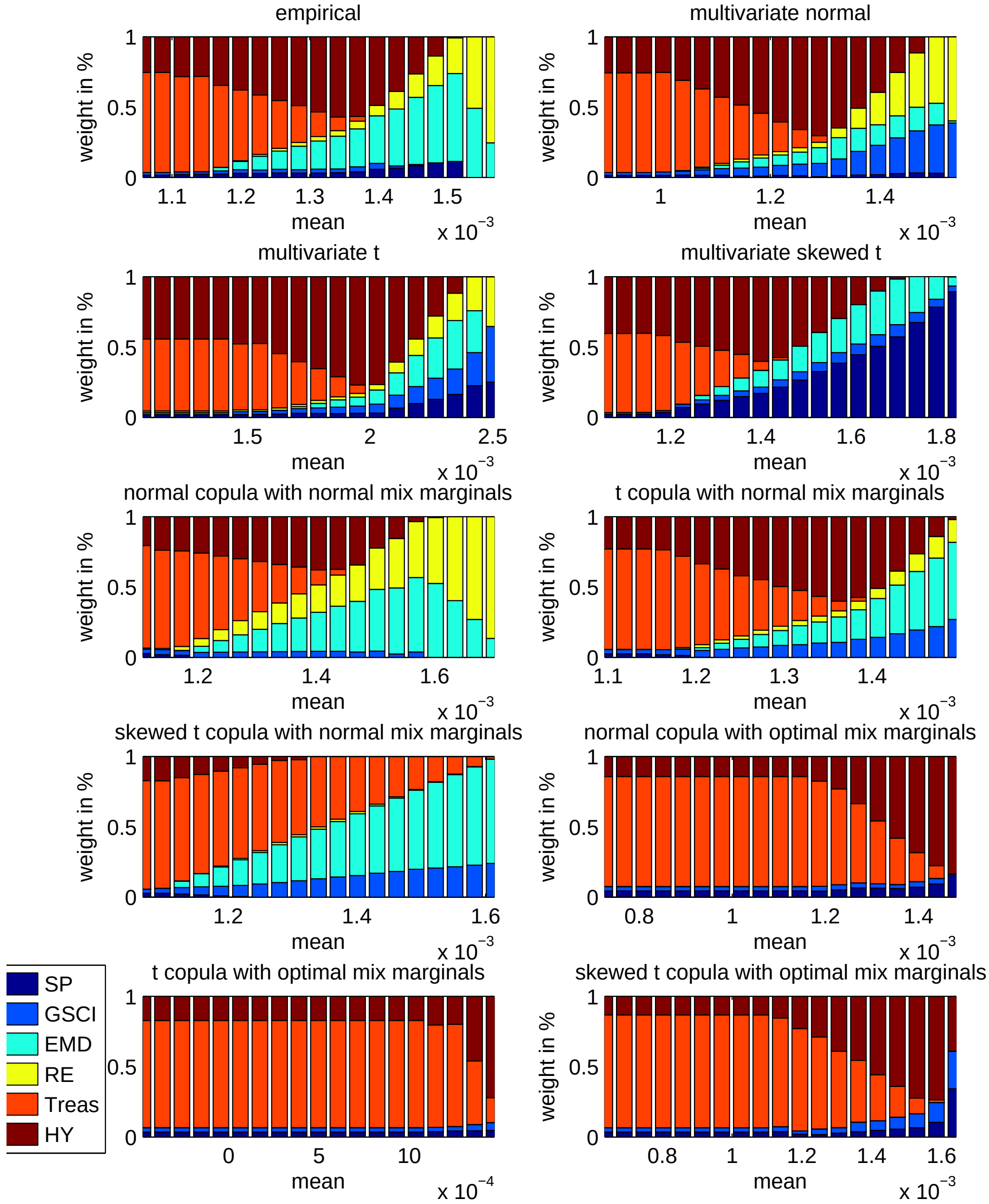


Figure 29: Optimal portfolio weights per model at minimized 1%CVaR

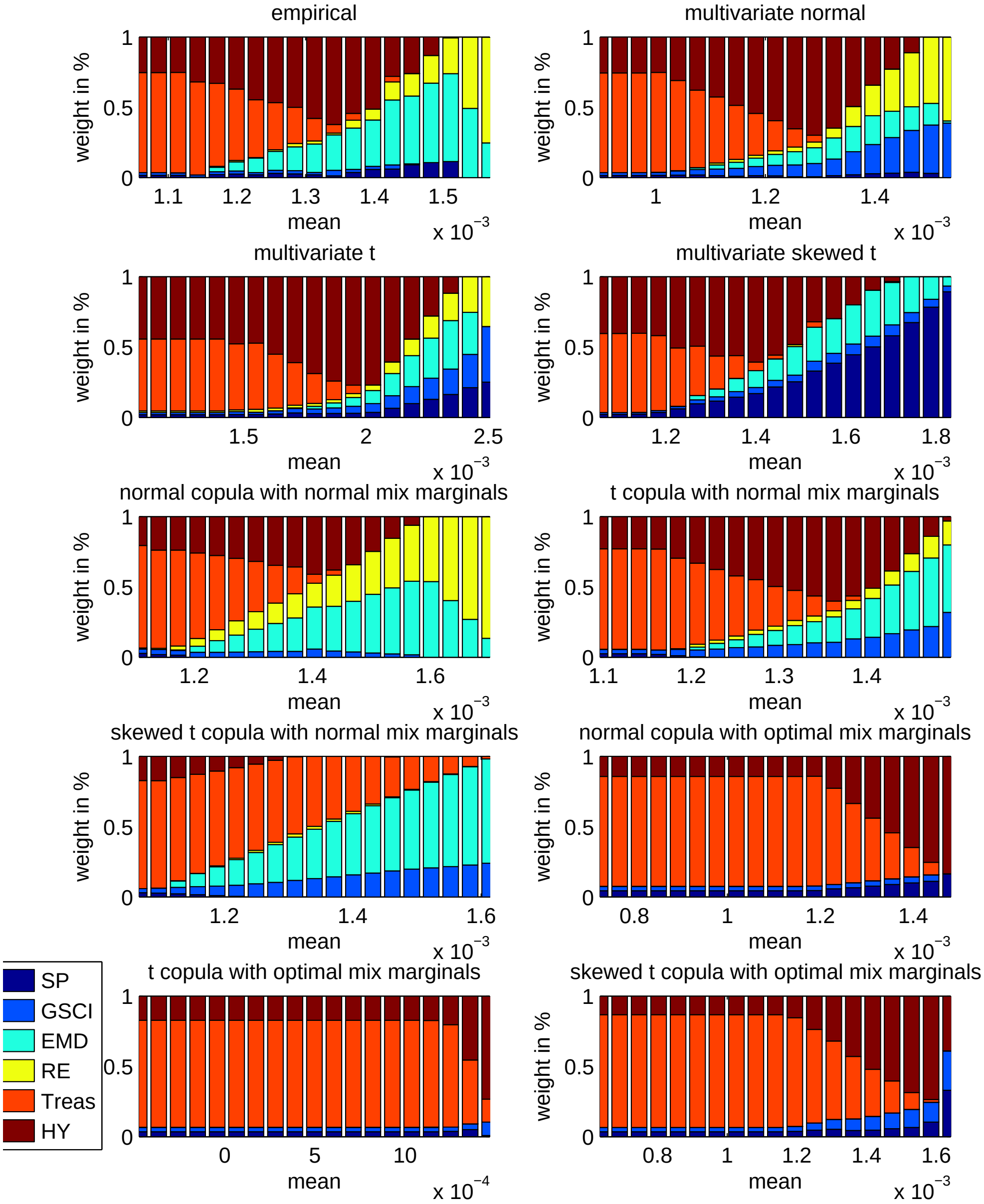


Figure 30: Optimal portfolio weights per model at minimized MDD

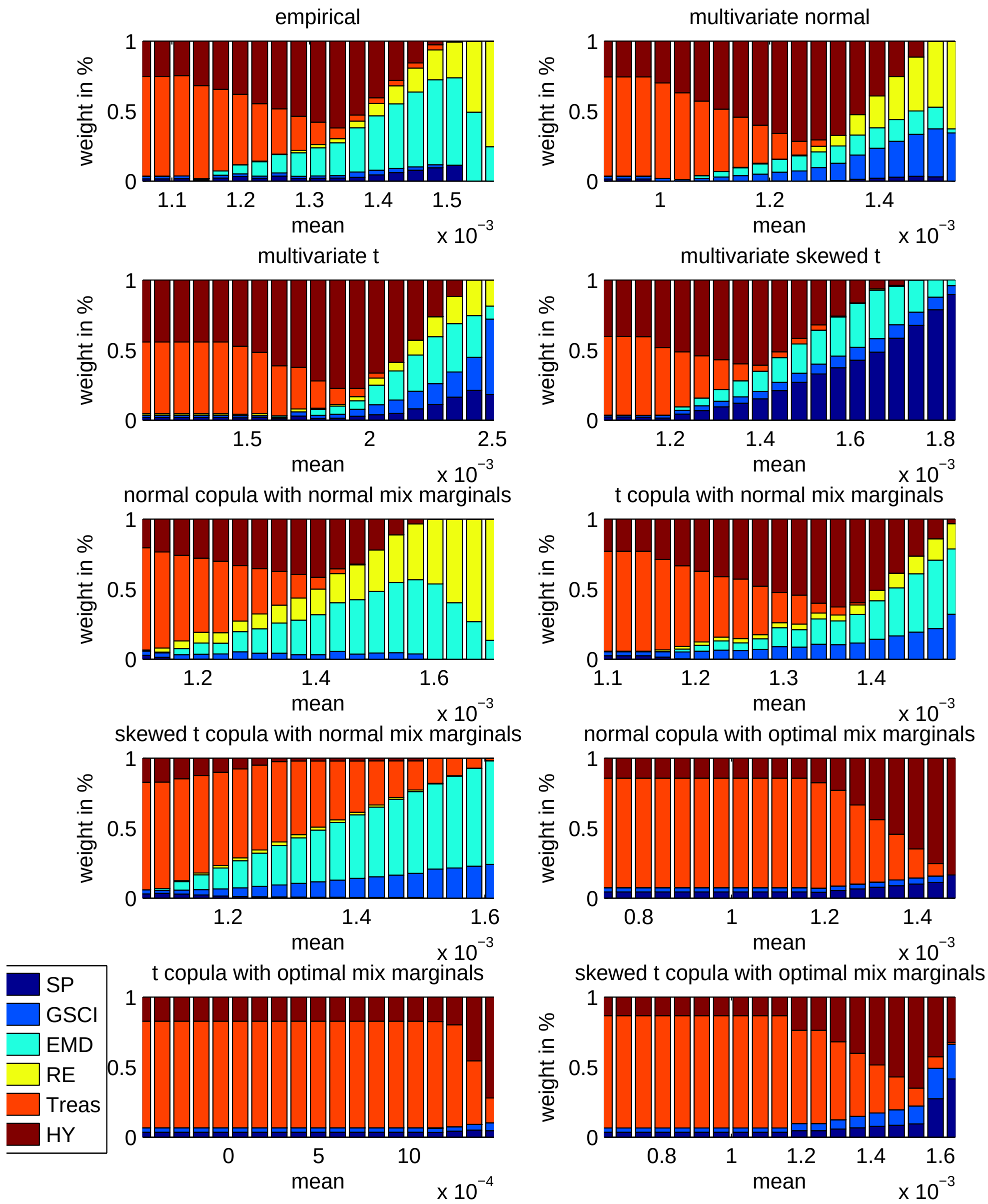


Figure 31: Mean / Std optimization

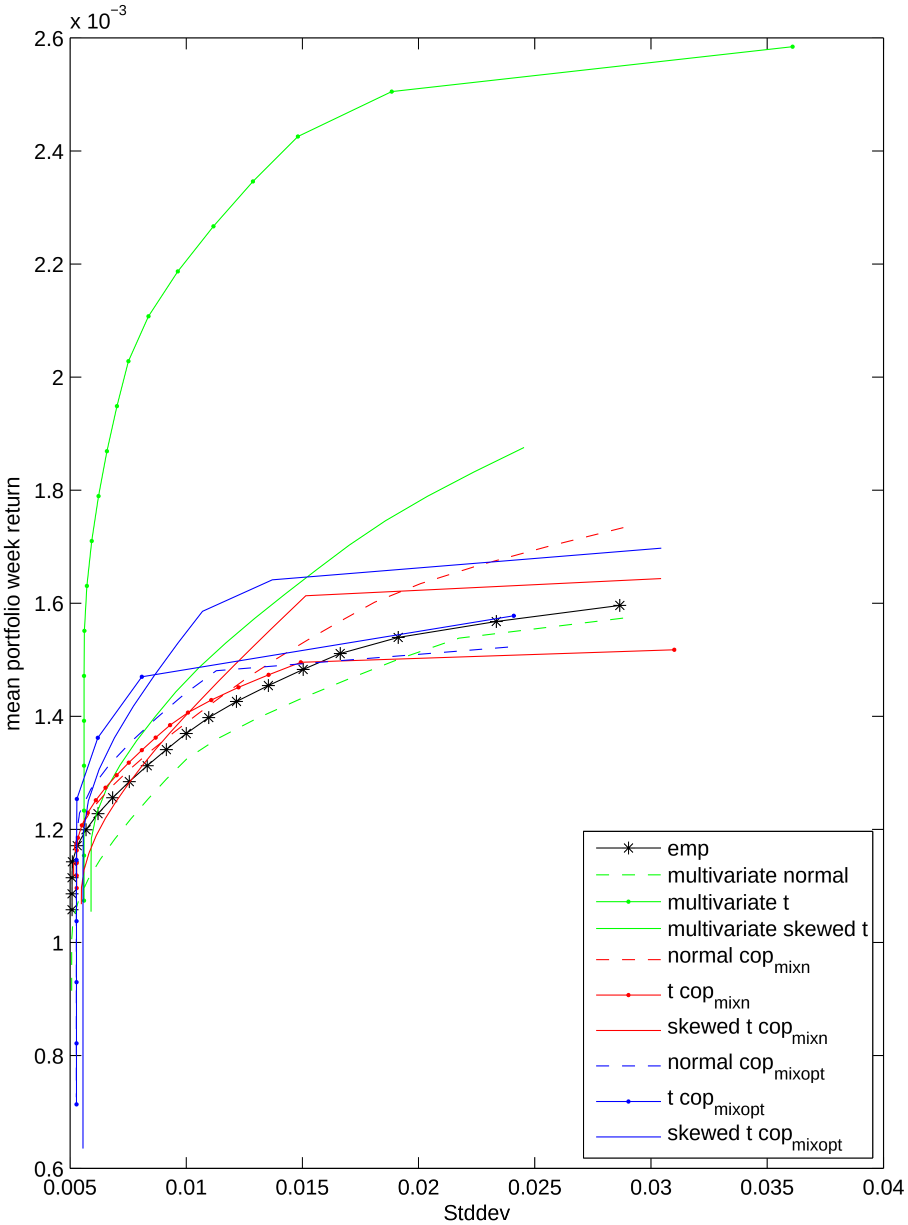


Figure 32: Mean / 5%VaR optimization

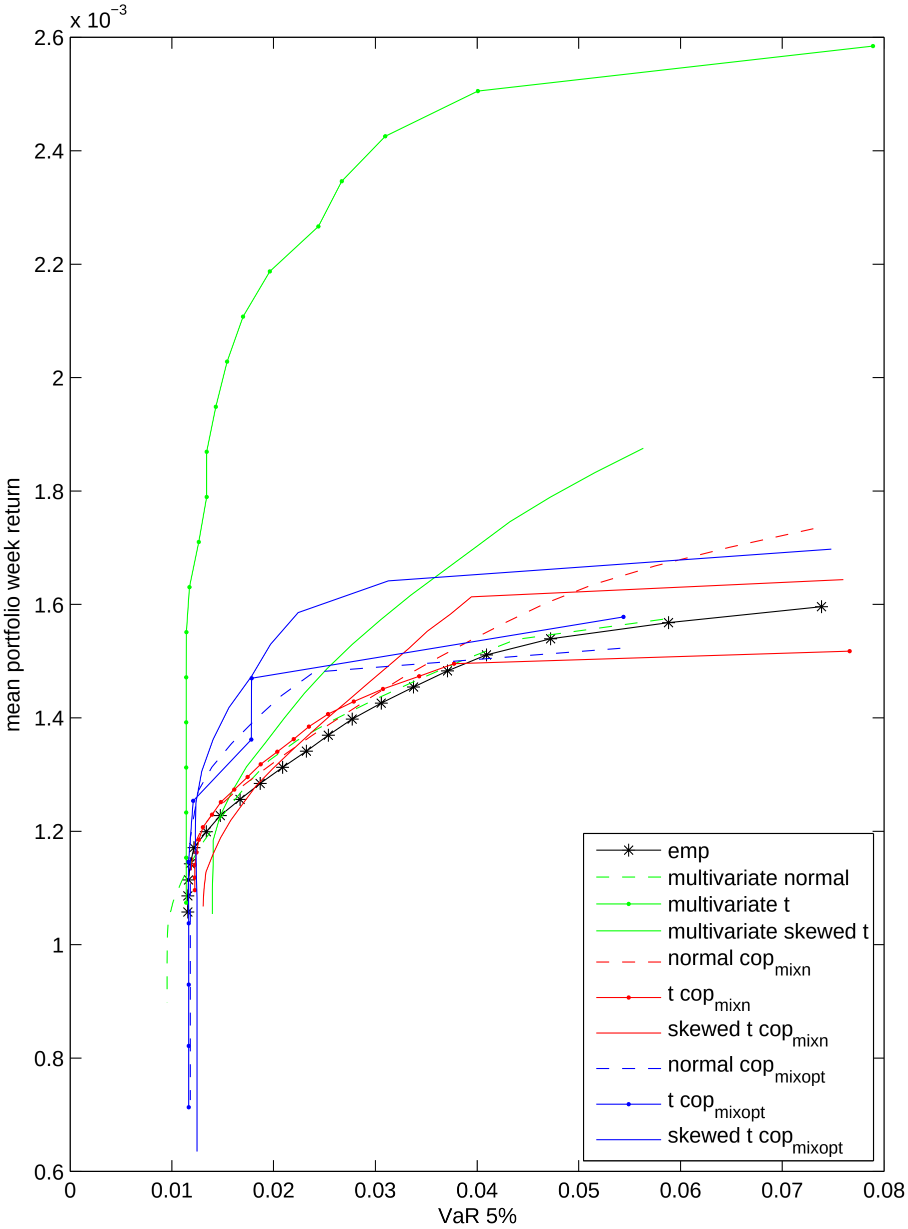


Figure 33: Mean / 1%VaR optimization

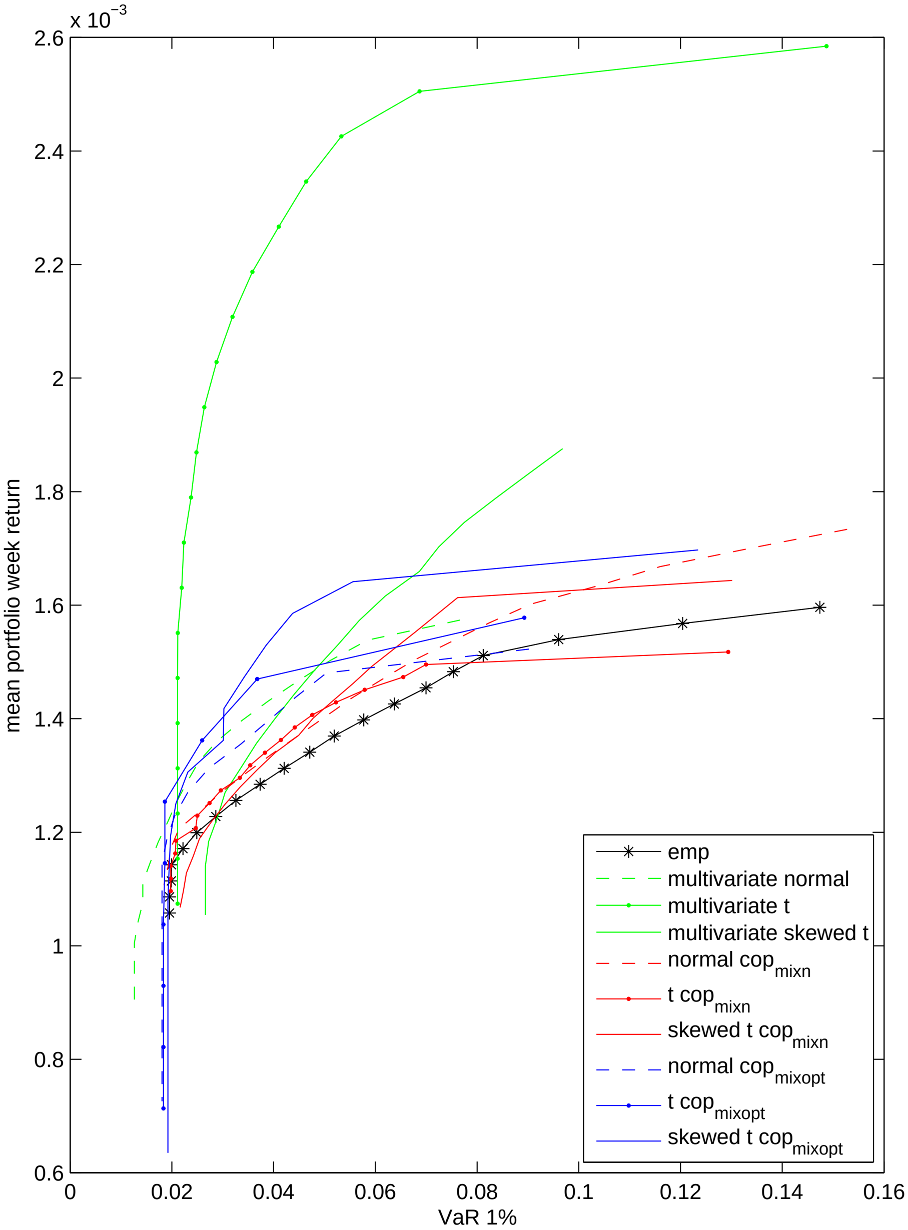


Figure 34: Mean / 5%ES optimization

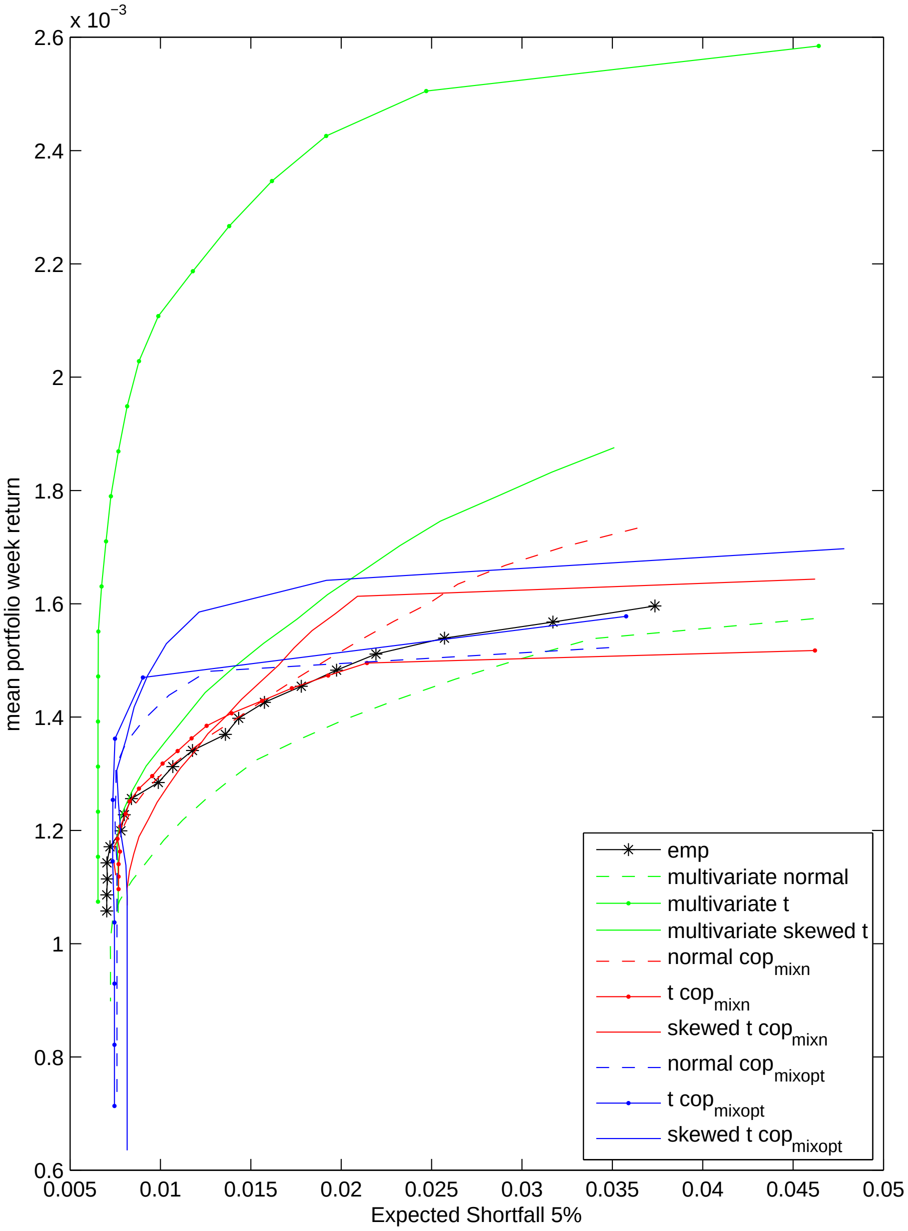


Figure 35: Mean / 1%ES optimization

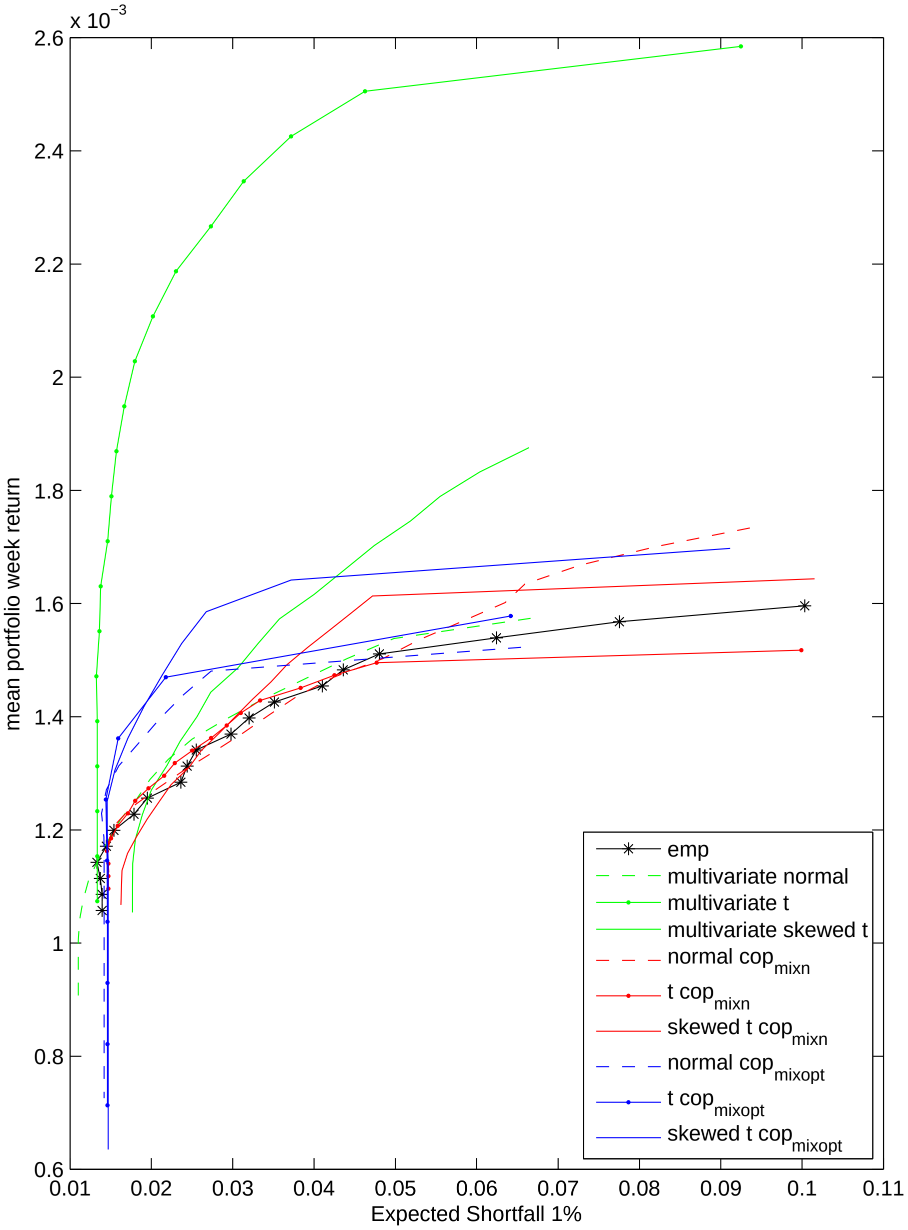


Figure 36: Mean / MDD optimization

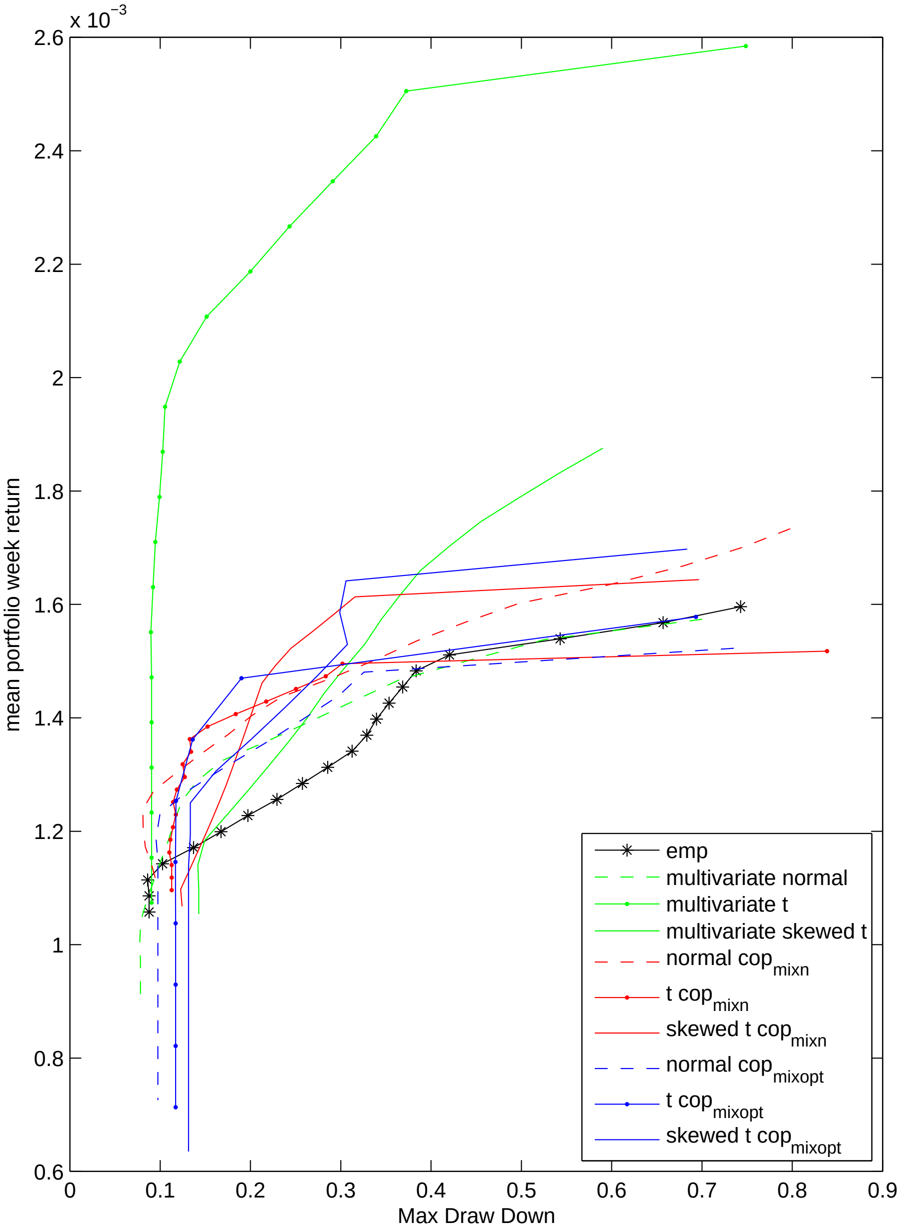


Figure 37: Mean / Std optimization

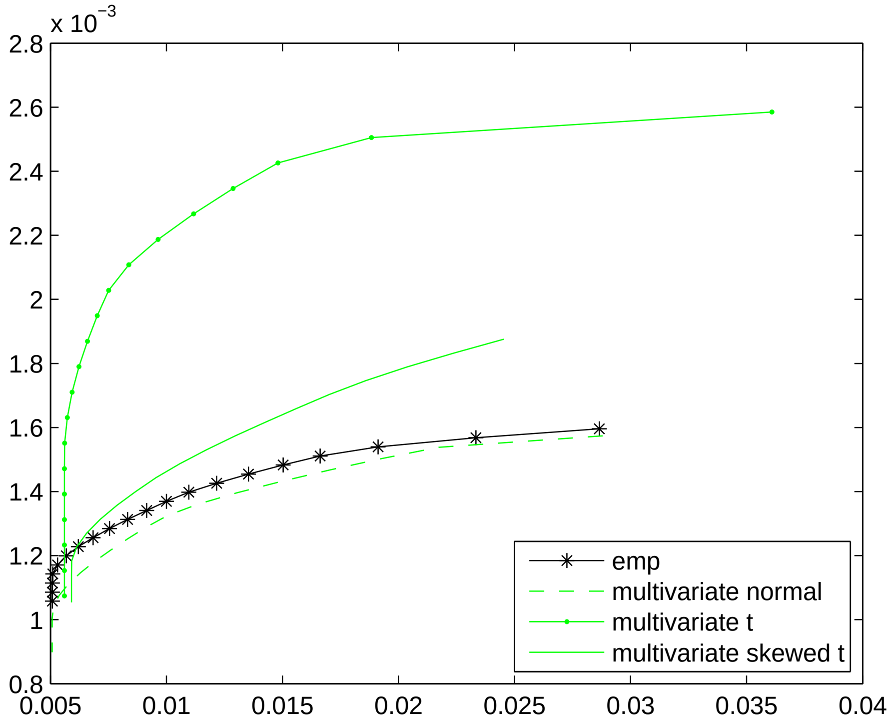


Figure 38: Mean / Std optimization

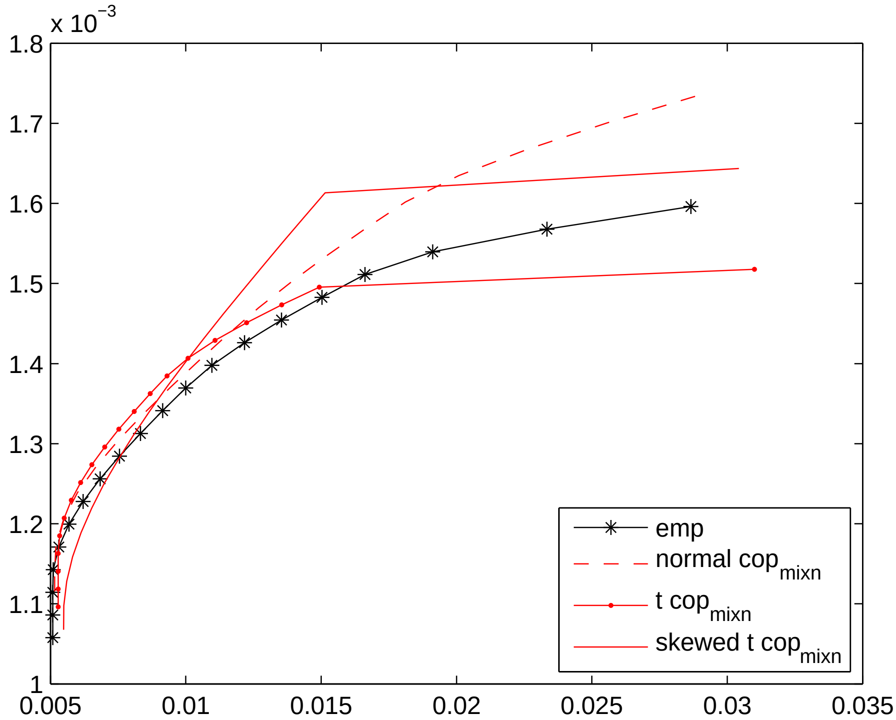


Figure 39: Mean / Std optimization

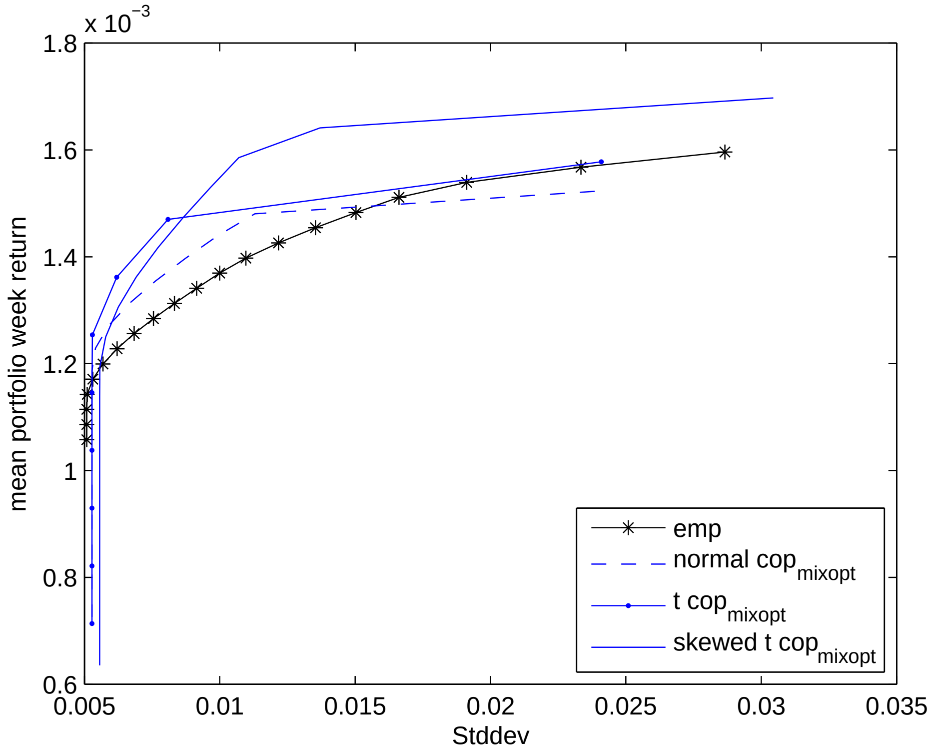


Figure 40: Mean / 5% VaR optimization

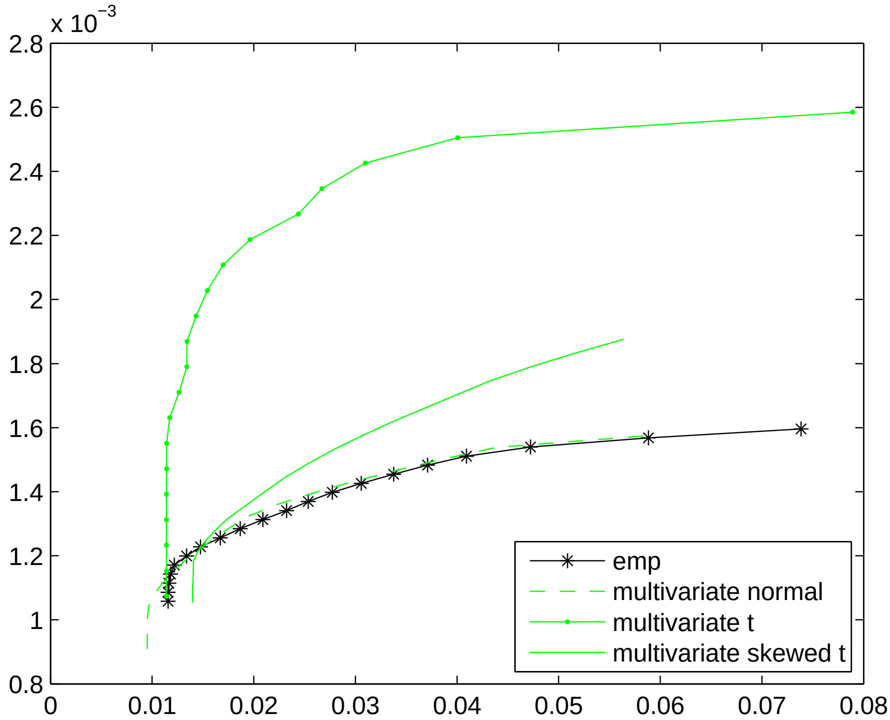


Figure 41: Mean / 5% VaR optimization

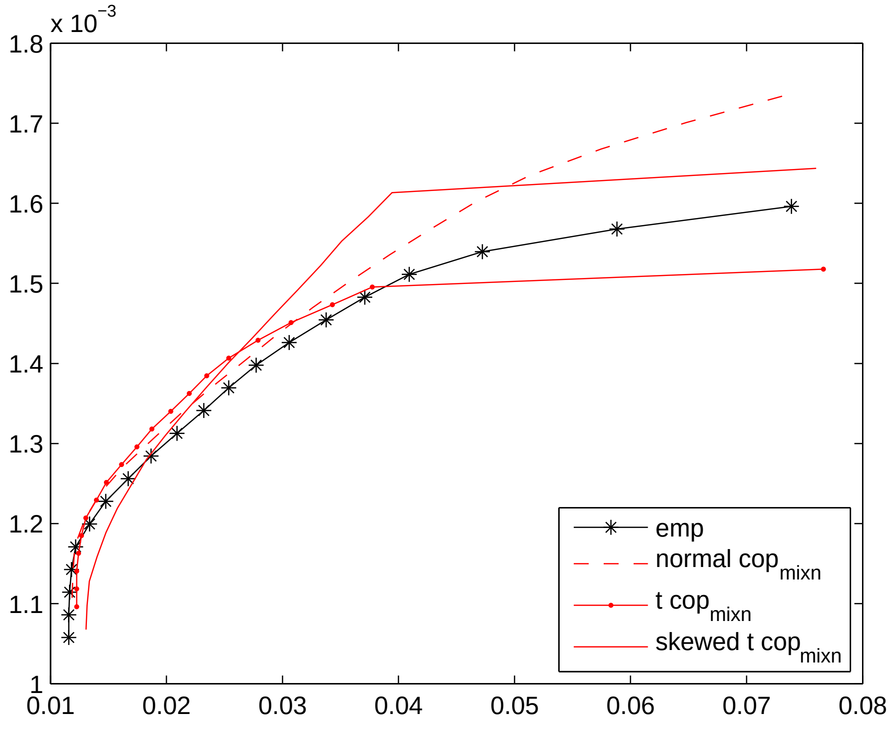


Figure 42: Mean / 5% VaR optimization

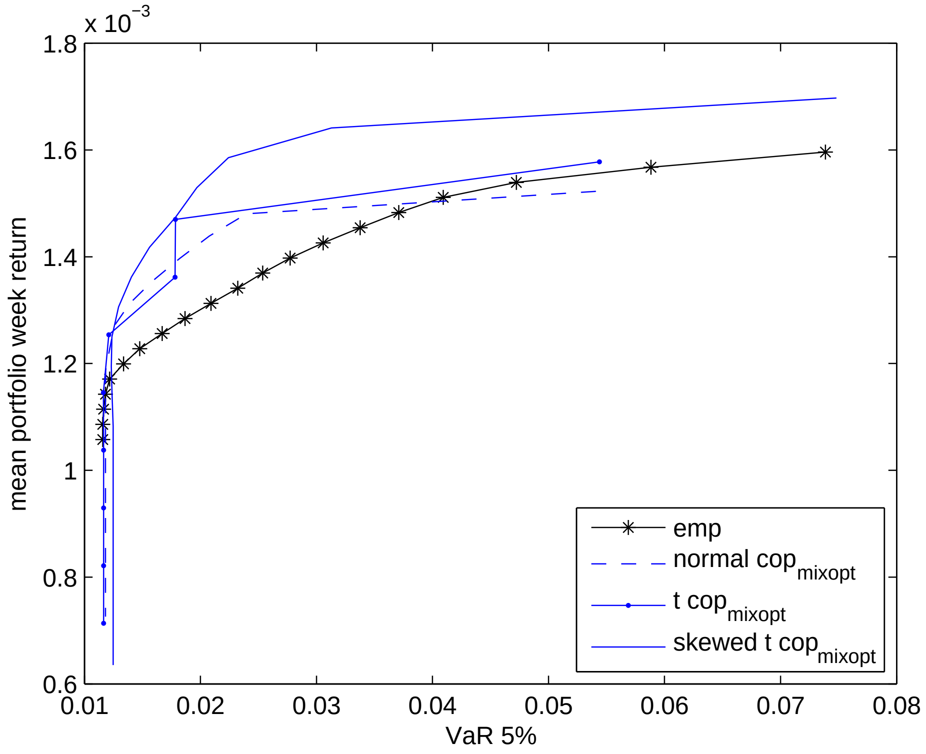


Figure 43: Mean / 1% VaR optimization

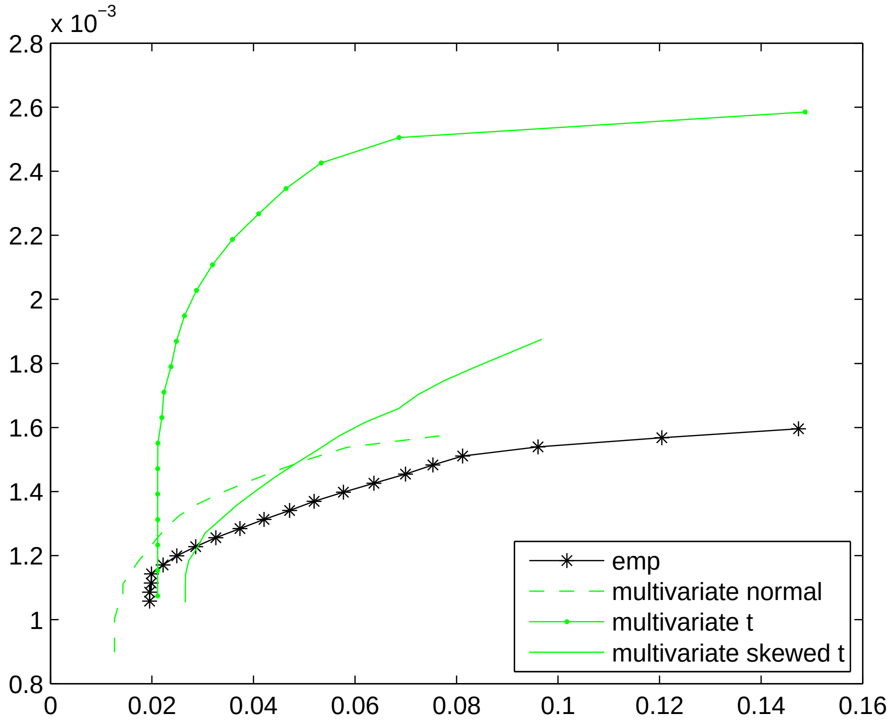


Figure 44: Mean / 1% VaR optimization

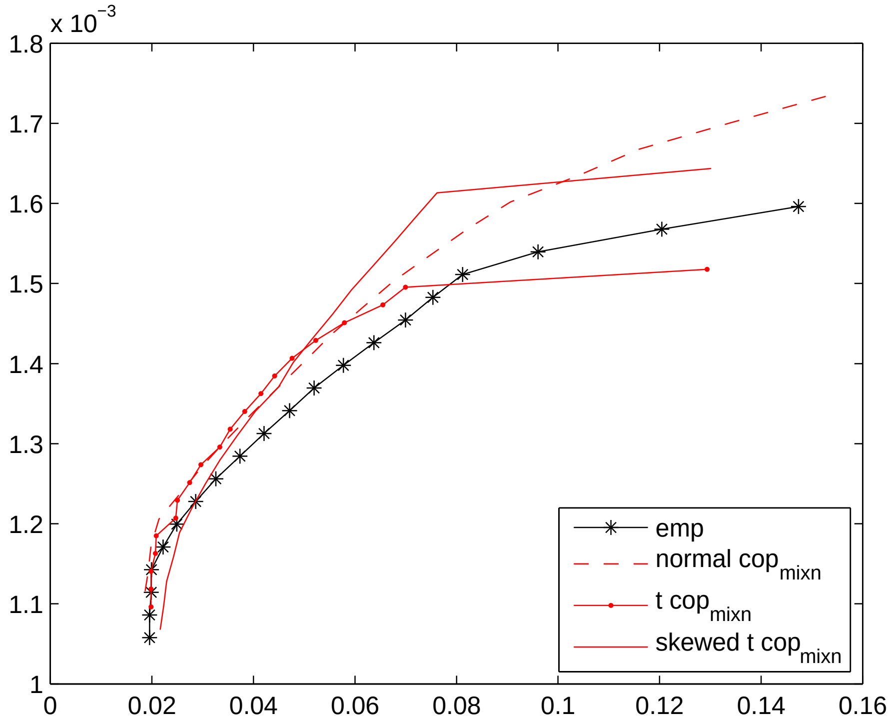


Figure 45: Mean / 1% VaR optimization

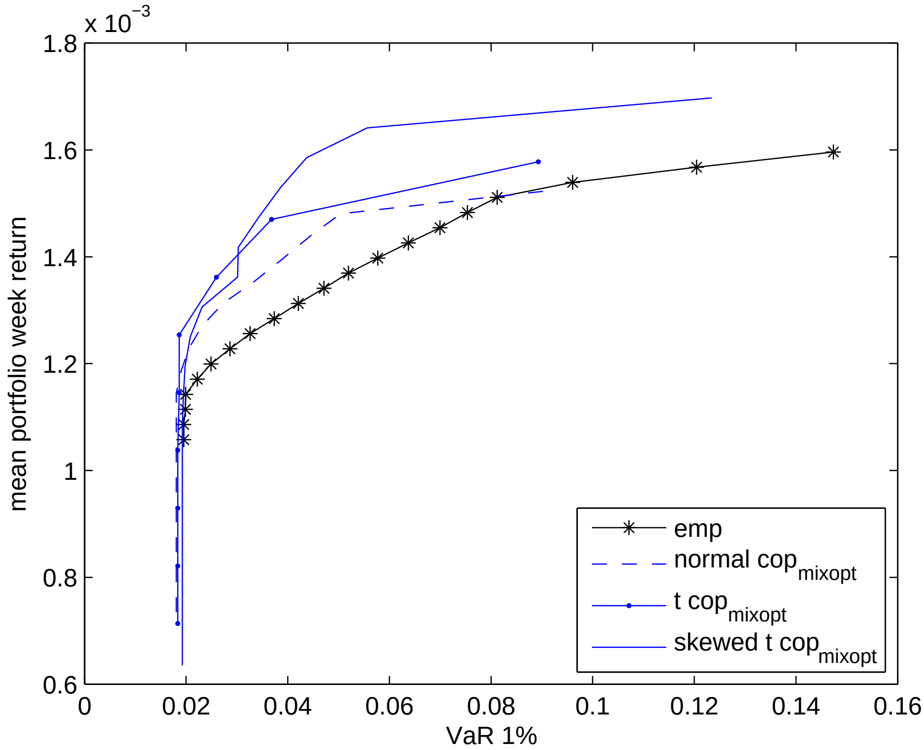


Figure 46: Mean / 5% ES optimization

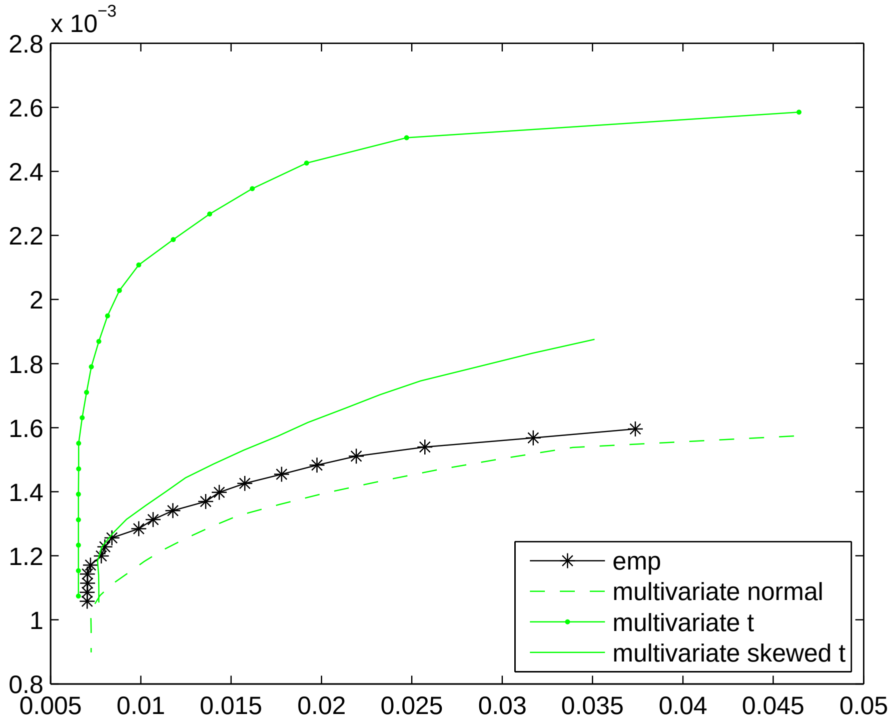


Figure 47: Mean / 5% ES optimization

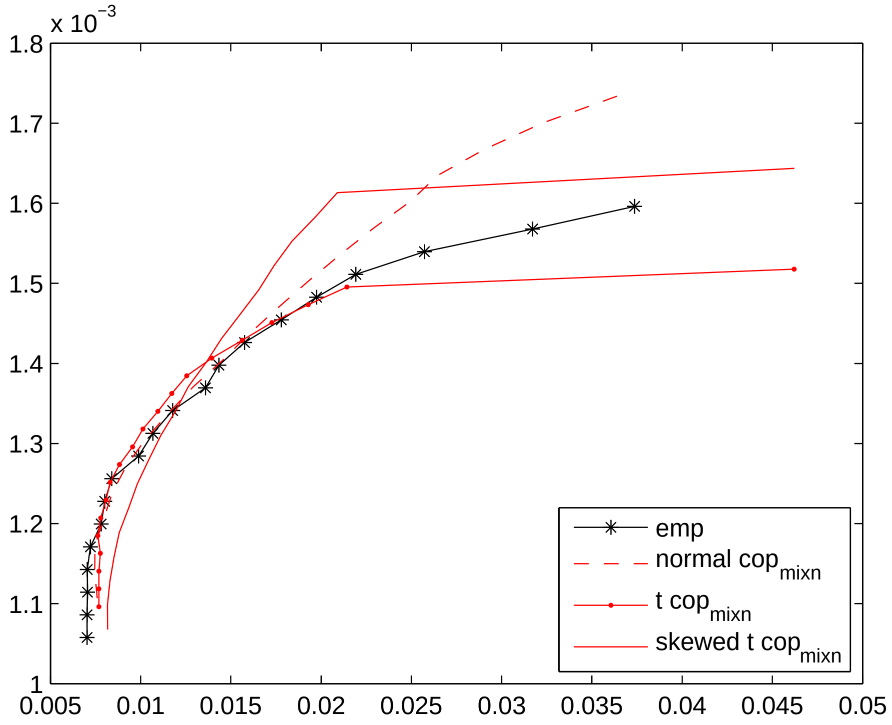


Figure 48: Mean / 5% ES optimization

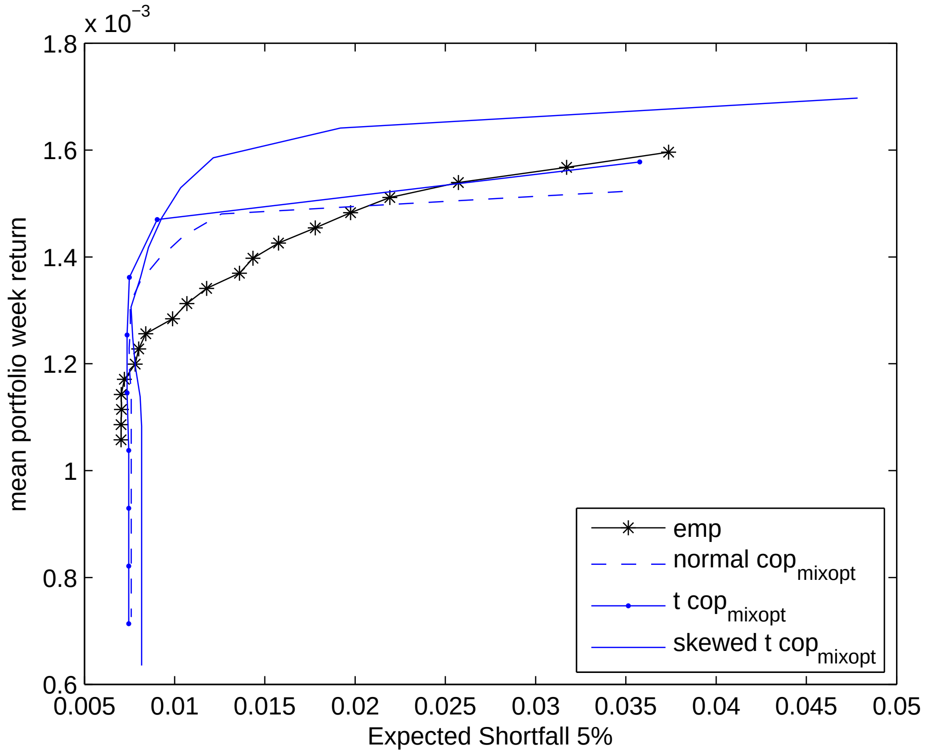


Figure 49: Mean / 1% ES optimization

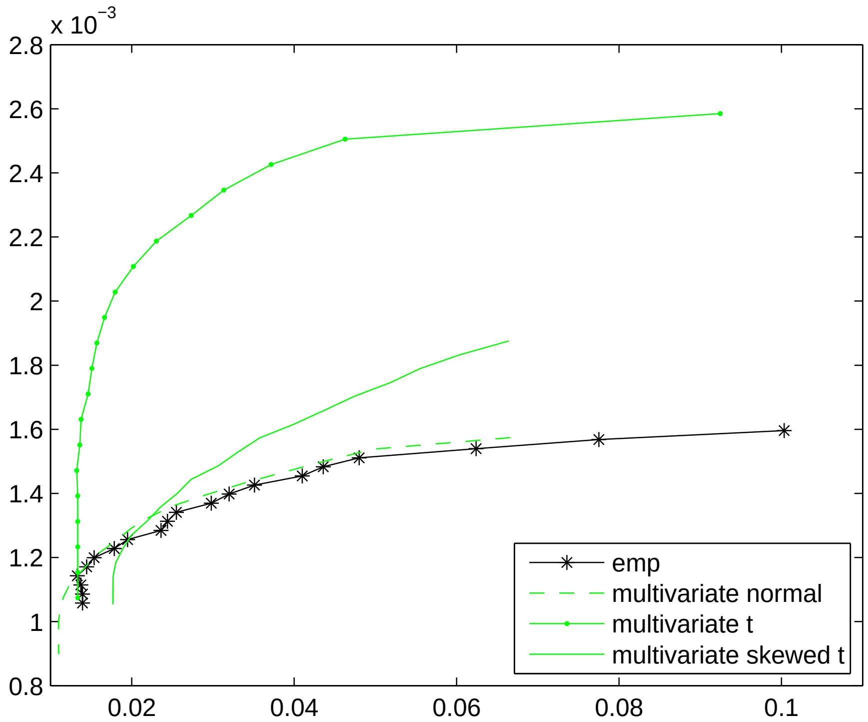


Figure 50: Mean / 1% ES optimization

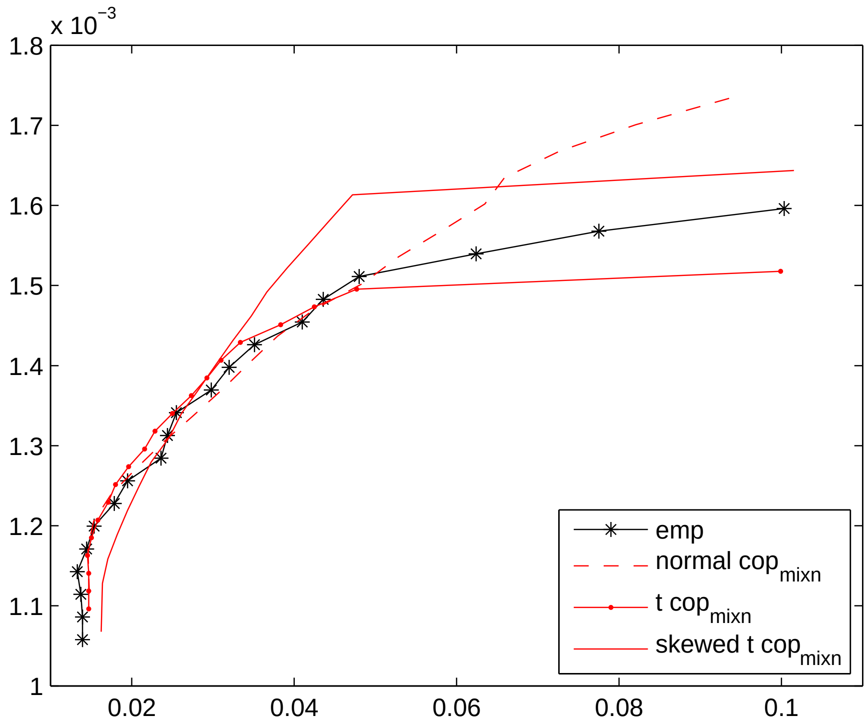


Figure 51: Mean / 1% ES optimization

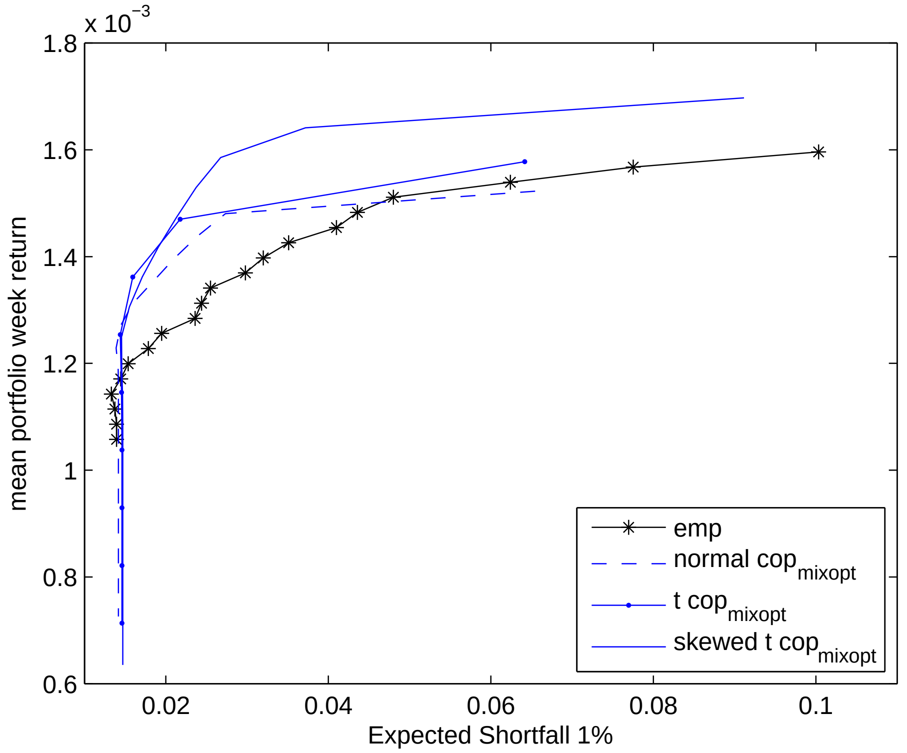


Figure 52: Mean / MDD optimization

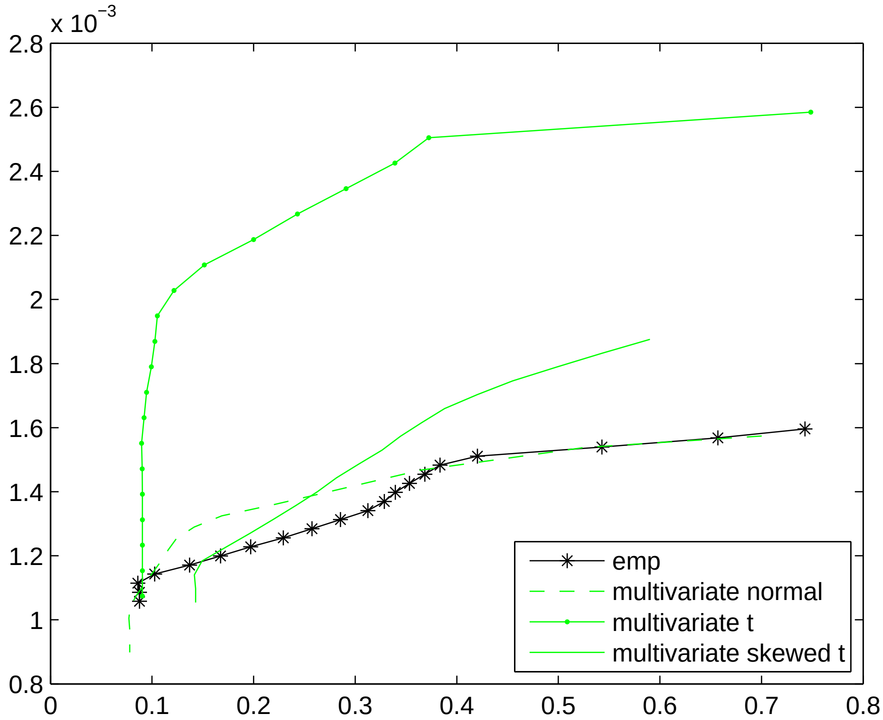


Figure 53: Mean / MDD optimization

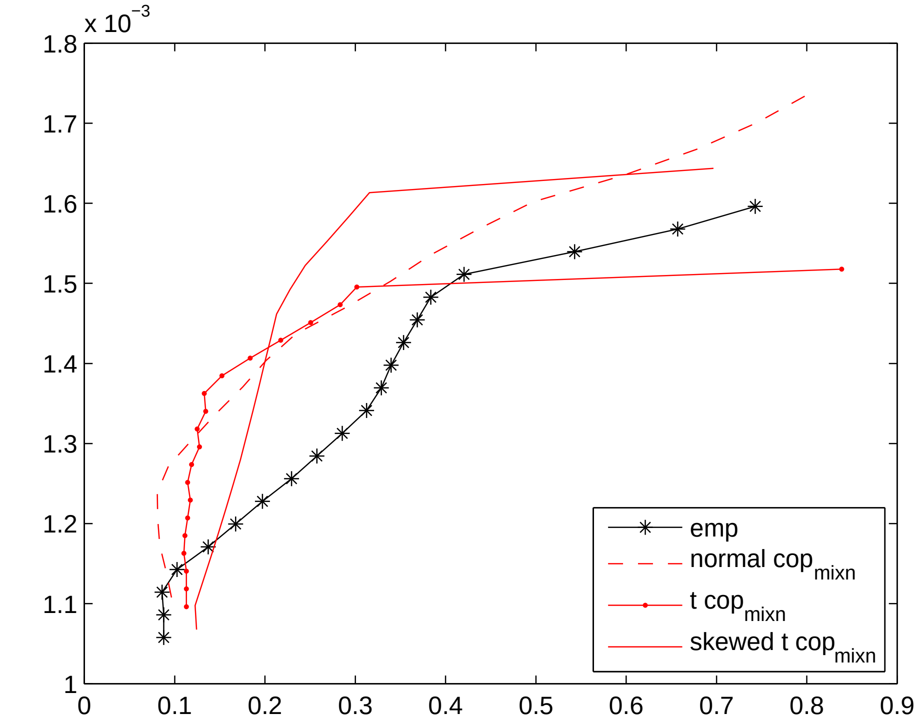


Figure 54: Mean / MDD optimization