The Influence of the Investment Horizon on the Asset Allocation

Joël Meili and Fabio Bührer

June 7, 2019

Abstract

Portfolio optimization usually involves minimizing the risk of a portfolio for a given level of expected return. This requires a definition of risk. In Markovitz portfolio optimization, risk is defined as the standard deviation of returns. As the standard deviation makes no reference to a client's investment horizon, this yields the same optimal portfolio for long-term investors as for short-term investors, which seems counter-intuitive. To address this issue, we use value-at-risk and expected shortfall over a given investment horizon as risk measures, instead of the standard deviation of returns. We include a review and comparison of these different risk measures, their advantages and their shortcomings. In addition, we account for the fact that the distributions of market returns are not exactly normal, but exhibit different degrees of fat tails. Based on analyzing 35 years of daily market returns, we model them by univariate and multivariate skewed Student's t-distributions, whose numbers of degrees of freedom range from 3 to 10. Fewer degrees of freedom, i.e., fatter tails, correspond to shorter-term returns. We include a review of normal distributions, symmetric Student's t-Distributions, skewed Student's t-distributions and their generalization, the generalized hyperbolic distributions. We then first use numerical techniques to minimize value-at-risk over a given investment horizon for a portfolio consisting of assets that are described by symmetric Student's t distributions with different degrees of freedom. We find that the optimal portfolio now depends on the investment horizon. In particular, the appetite for heavy-tailed assets grows with longer investment horizons, while for shorter-term horizons the assets with more normally-distributed returns are overweighted. In a second step, we analytically minimize Expected Shortfall over various levels, corresponding to different investment horizons, for a portfolio consisting of assets that are described by the multivariate skewed Student's t distribution. The results confirm again that longer-term investors should take more tail risk, while short-term investors should avoid it. Implications for institutional portfolio management are briefly discussed.

Table of Contents

1 Introduction

1.1 Objectives

The goal of this paper is to model the "fat tails" of the return distributions of financial assets and, based on this, to understand how the length of the investment horizon impacts portfolio allocation among various assets for different levels of risk. Addressing this problem requires essentially to (1) establish a measure of financial risk and (2) to build a model for representing the evolution of each asset's value and the effects of financial risk.

The notion of financial risk plays a crucial role in defining an investment strategy. In the case of this paper, a constant rebalancing strategy is followed, where the portfolio return is defined as the weighted average of the single asset returns. The shorter the time horizon of the investment strategy, the more the investor needs to consider the likelihood and the effects of large economic and financial downturns on the portfolio's value. Growing globalization and interconnectedness of financial systems increase the chances of acute, but local issues to spread to other geographic areas and to start chain reactions affecting different markets and actors. Another characteristic of crashes is their potential to undo long periods of profitability in a very short time. The 2007-2008 financial crisis is a prime example of these concepts. The crisis started with the failures of subprime loans in the US. mortgage market and spread globally to affect virtually every financial market and asset class; as the crisis developed, even long-standing and often successful financial firms were forced into bankruptcy, restructuring, acquisition, or bailout programs after suffering large losses. To reflect these events quantitatively requires establishing a financial risk metric which is intuitive, computable, and gives a concrete, realistic indication of the loss a portfolio would incur in relation to an occurrence, the latter being specified by its assumed likelihood.

In the case of this paper, modeling the value of an asset means to analyze its historical returns and to determine a statistical distribution which approximates its behavior as closely as possible. This step entails not oversimplifying the mathematical approximation for the sake of the model's elegance, since a slightly different statistical distribution could cause market events with significant bearing on an asset to be under- or overrepresented. It is important that the historical data include as many instances of important market movements as possible to better reflect their likelihood and impact. The reference data must also be correctly interpreted and, where appropriate, transformed, to ensure the same property is being quantified and used to model all instruments, e.g. in the case where an asset's price uses an otherwise unusual notation.

1.2 Background

This paper continues and builds upon the investigation of a preliminary paper, some key points of which are summarized here:

1. We go beyond volatility as a risk measure. While at the core of important theories in the field of mathematical finance, risk measurement in terms of volatility has several drawbacks. One disadvantage of volatility from a conceptual standpoint is that it equally weighs positive and negative deviations from the mean. From a quantitative point of view, a common critique to volatility-based risk management is that it does not give a concrete indication of the size of possible losses. Value-at-risk (VaR) is used to measure the minimal potential loss over a defined time frame τ with a stated confidence level $1-\alpha$, α being the significance level.
2. We go beyond the use of the normal (or Gaussian) distribution. While its flexibility and tame mathematical nature make it enticing, it carries substantial disadvantages. The heavy-tailed distributions of returns on various financial instruments are in stark contrast with their normal approximations. The higher the frequency at which the returns on an asset are being measured, the more evident extreme movements are. For instance, approximating daily returns on a stock market index with a normal distribution underrepresents the proportion of tail-events that actually occurred in the index' lifetime. The Student's t-distribution allows, by changing the degrees of freedom ν, to adjust how heavy its tails are, and is used to model returns generated by eight instruments from four different asset classes.

3. Data analysis indicates that riskier investments, e.g. stock indices, display heavier tails than assets considered safer, such as government bonds. Also, extreme, negative downturns are more frequent in risky instruments, whereas their safer counterparts are more often subject to market surges. Further, by analyzing correlations between the assets, it is noticed that the safer assets negatively correlate to risky investments. These facts seem to confirm the notion of safe haven assets being bought in times of distress or uncertainty in the financial markets.

4. Logarithmic returns are used for ease of aggregation to compute returns over longer time horizons and to test the assumption of returns being log-normally distributed, i.e. that log returns are normally distributed. By performing the Anderson-Darling normality test, it is established that log returns are generally not normally distributed. However, the result changes according to the timespan and the instrument being considered: the longer the investment period and the less volatile the asset, the more the normality assumption is plausible.

5. Empirical validation of modern portfolio theory for optimizing a portfolio demonstrates that the framework works (1) with Student's t-distributed instruments and variance as a risk metric and (2) with normally distributed instruments and VaR as a risk metric. However, the optimal weights differ from the mean-variance optimized results when one asset is normally distributed, the other is Student's t-distributed, and simultaneously VaR is used as the risk metric. For this reason, further weight optimizations are done numerically.

6. To make the optimal portfolio weights for different investment horizons τ comparable, the significance level of the VaR is calculated as a function of τ over the duration of the total period of data being considered: $\alpha = \frac{\tau}{Duration}$. This modified version of VaR accounts for the risk-reducing effect of longer investment horizons. The optimization indicates a stronger appetite for instruments with lower degrees of freedom, i.e. riskier investments, with growing investment horizon.

In this context, the secondary goal is the improvement of both non-normal and non-volatility numerical results obtained in the preliminary paper. Compared to the preliminary paper, this paper introduces new approaches for quantifying financial risk, modeling assets behavior, and determining optimal model parameters.

2 Review

2.1 Statistical measurements

2.1.1 Log-Returns

In finance, returns measure the amount of profit or loss one has gained by investing in an asset, such as stocks or government bonds. Using logarithmized returns r_t allows to aggregate returns over periods as the product collapses to a sum.

Mathematical definition:

$$
\begin{aligned}
R_t &= \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1 \\
\Rightarrow R_{i,j} &= \prod_{t=i\geq 1}^{j} (1 + R_t) = \prod_{t=i\geq 1}^{j} \left(\frac{P_t}{P_{t-1}}\right) \\
\ln(R_{i,j}) &= \sum_{t=i\geq 1}^{j} \underbrace{\ln\left(\frac{P_t}{P_{t-1}}\right)}_{\overset{\|}{\because} \, r_t}
\end{aligned}
$$

Where R_t and P_t are respectively the return on an asset and its price at time t, and $R_{i,j}$ is the return on an asset in the period between $i, \ldots, j$.

2.1.2 Risk Measures

The ability to identify, analyze, quantify, and act on risks is essential for the long-term profitability of financial services and individual investors. Also, the need for more comprehensible, more accurate, and possibly more conservative risk assessment techniques became clearer after the fallout from the 2007-2008 global financial crisis. The importance of financial risk management is not limited to reasons of earnings or corporate longevity; in a bid to improve the resilience of financial systems, regulators and¬ other authorities tend to impose stricter capital requirements and harsher consequences for companies – and their managements – who do not abide by the rules. Nowadays it is not uncommon for large financial institutions to split financial risk into multiple subclasses, each with its dedicated personnel, e.g. credit risk, liquidity risk, equity risk, and currency risk.

In the context of this paper, the distinction between risk metric and risk measure is not important, since each risk concept presented henceforth is both a different abstraction, i.e. risk metric, and a different mathematical construct, i.e. risk measure. There is, however, a set of formal requirements for a method to qualify as a coherent risk measure ρ (Philippe Artzner and Heath (1998)):

Translation invariance:

Adding a sure initial amount α to a portfolio X and investing it in an instrument with a strictly positive price r decreases ρ by α:

$$\rho(X + r) = \rho(X) - \alpha$$

Subadditivity:

Merging portfolios X_1 and X_2 into one portfolio does not increase the associated ρ:

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$$

Positive homogeneity:

The risk of a portfolio X is proportional to the portfolio's size:

$$\rho(\lambda X) = \lambda \rho(X)$$

Where $\lambda \geq 0$.

Monotonicity:

If portfolio Y has always better values than portfolio X, then the risk of Y is lower than that of X:

$$\forall\, Y \geq X : \rho(Y) \leq \rho(X)$$

These axioms form part of the argumentations for the preference of one risk metric over another in this paper.

2.1.2.1 Volatility

Volatility measures the dispersion of the returns on an asset and is calculated as the standard deviation of the returns. Higher volatility coincides with larger movements of the returns in any direction, and thus can be a proxy for its riskiness.

Mathematical Definition:
Sample standard deviation:

$$\sigma \;=\; \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i \;-\; \mu)^2}$$

Where N is the number of observations and $$ is the sample mean

$$\mu \;=\; \frac{1}{N}\;\sum_{i=1}^{N}\; x_i$$

2.1.2.2 Value-at-Risk

Value-at-risk (VaR) measures the value that the potential loss will not exceed at a specific confidence level over a stated time horizon. For example, a one-day 95%-VaR of 1 million USD means that losses should not exceed 1 million USD with a certainty of 95% (Alexander F. McNeil and Embrechts (2015)).

Mathematical Definition:
Value-at-risk:

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(1-\alpha)$$

Where α is the significance level.

It has been established that VaR does not always qualify as a coherent risk measure because, in general, it does not conform to the subadditivity axiom (Philippe Artzner and Heath (1998)). This means that the sum of the VaR of multiple, single assets can be less than the VaR of the single, diversified portfolio composed of the same instruments. This phenomenon violates the notion of risk reduction through diversification, whereby the risk of a portfolio of assets with imperfectly correlated behaviors is less than the sum of each instrument's risk. At most, the level of risk of a diversified portfolio should equal the sum of the individual asset risks, and only in the case where all instruments are perfectly correlated.

However, VaR adheres to subadditivity when asset returns follow an elliptical distribution, e.g. any distribution that belongs to the set of normal variance mixture distributions X:

$$\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z}$$

Where $\boldsymbol{\mu}$ is the location vector, $W \geq 0$ is a random variable, A is a real-valued matrix, and $\boldsymbol{Z} \sim N(\boldsymbol{0}, I)$

Compared to variance, value-at-risk at various levels of confidence is a more fine-grained measure of risk, since it states the extent of potential losses. However, VaR can be viewed as an optimistic view of the loss scenario; contrary to the principle of conservatism, the quantile given by VaR is the least bad outcome that satisfies the conditions of a specific timespan and a given level of confidence. Despite it being a more specific risk metric than variance, VaR fails to quantify the real loss an instrument is expected to make when the downside event takes place. Moreover, comparing VaRs for assets with different distributions can be misleading. An assets can be riskier than another for the same VaR if the left tail of the distribution of its returns is comparatively heavier. Still, if accounted for its limitations, VaR can serve as an indicator and warning signal for managing risks related to financial instruments. Notwithstanding its technical and conceptual shortcomings, presently VaR plays a significant role in financial risk management and is included as a measure of risk in regulatory frameworks such as Basel II.

2.1.2.3 Expected Shortfall

Expected shortfall (ES) measures the expected loss suffered from an event occurring in the worst $\alpha\%$ of cases. It is the mean quantile of the portion of the distribution being considered, i.e. the center of mass of the tail from $-\infty$ to the $\alpha\%$-quantile, the latter having the same meaning as the $\mathrm{VaR}_\alpha(X)$. Because of this relationship, it follows that $|ES_\alpha(X)| \geq |\mathrm{VaR}_\alpha(X)|$. Expected shortfall is also known as conditional VaR by virtue of it being the expected outcome given that the loss occurs at or below the indicated VaR (Alexander F. McNeil and Embrechts (2015)).

Mathematical Definition:
Expected shortfall:

$$ES_q(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(X)\mathrm{d}u$$

Expected shortfall satisfies all axioms of a coherent risk measure. Unlike value-at-risk, ES always fosters risk reduction where it can be achieved through portfolio diversification with imperfectly correlated assets, regardless of the assets' returns distributions. Compared to VaR, expected shortfall also gives a more realistic indication of the magnitude of losses, since it considers the mean of the tail and not its most optimistic value, capturing the tail's inherent risk, and leading to more conservative exposures. Another advantage over VaR is that ES is less likely to be misleading in the comparison of assets with different distribution characteristics. For two assets with the same VaR, the instrument with the heavier tail will have a higher absolute expected shortfall. The standards published in 2016 by the Basel Committee on Banking Supervision also cite ES as a more prudent alternative to VaR ((“Minimum Capital Requirements for Market Risk” 2016)).

2.1.3 Heavy Tails

Heavy-tailed distributions can be specified by means of mathematical methods. However, there is no universally agreed upon theory for the technical classification of distributions by tail behavior, nor there is an academically accepted convention for the use of terms to describe it Schuster (1984). Some authors (Asmussen (2003)) use precise nomenclature to differentiate among subsets of what they regard as the wider heavy tail distribution class. Nonetheless, in this paper the words “fat” and “heavy” are interchangeable. Dacorogna et al. (Michel M. Dacorogna (2001)) use a classification system of three categories of distribution behavior mainly based on how the tails of the cumulative distribution functions decline when $x \to \pm\infty$:

1. Thin-tailed: the tails decline in an exponential fashion, i.e. $O(e^{-\lambda|x|}), \lambda \in \mathbb{R}_{>0}$
2. Fat-tailed: the tails decline like a power law, i.e. $O\left(|x|^{-\alpha}\right), \alpha \in \mathbb{R}_{>0}$
3. Bounded distributions: absence of tails caused by domain finiteness

Under this classification system, the normal distribution falls into the thin-tailed category, whereas the distributions of daily asset returns plausibly fall into the fat-tailed category. The discrepancy between the normal distribution and financial data due to heavy tails is well documented (Mandelbrot (1963)), and although numerous authors agree on the need for models with heavy tails, the scientific literature does not converge on the type of distribution to use.

2.1.4 Skewness

Skewness is a property of probability distributions characterizing the degree of asymmetry. The value of the skewness of a distribution includes the sign, which signals the direction of the asymmetry, and the numerical value, which measures of pronounced the asymmetry is. Intuitively, a unimodal distribution, as is the case for the returns on most financial instruments, is negatively (or left-) skewed if its left tail is longer and its mass is concentrated on the right tail, and positively (or right-) skewed if the opposite is true. Rule of thumbs exist involving the values of different location measures to determine the skewness, e.g. comparing mean and median; these approximations can fail if the distribution is multimodal or if one tail is short and thick, while the other is long and thin. Despite the variety of alternative definitions of measures of skewness, the current convention is that the skewness γ_1 is the third standardized central moment of a distribution.

Mathematical Definition:

$$\gamma_1 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

This parameter is important when analyzing the behavior of financial assets and can be a measure of how often extreme market movements in one direction happen compared to events of opposite sign. Assets deemed safe often exhibit positive skewness in virtue of price surges during market meltdowns, whereas riskier investments tend to display negative skewness owing to prices precipitating in times of crisis.

2.2 Distributions

2.2.1 Normal Distribution

The normal (or Gaussian) distribution is a symmetric, unimodal distribution with tails growing thinner from its center, where mean, median and mode are located. The distance from the center is often expressed in standard deviations, notated σ; this is a compact method for expressing the likelihood of – assumedly – normally-distributed data.

Mathematical Definition:
Probability density function:

$$f_X\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Cumulative distribution function:

$$F_X\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x$$

Expected value:

$$\mathrm{E}\left(X\right) = \mu$$

Variance:

$$\mathrm{Var}\left(X\right) = \sigma^2$$

Standard (or z-) score:

$$z = \frac{x-\mu}{\sigma}$$

The portion of data within one, two and three σ from the population mean μ in both the left and right tail is summarized below:

$$\begin{aligned}
P(\mu-\sigma \leq x \leq \mu+\sigma) &\approx 68.27\% \\
P(\mu-2\sigma \leq x \leq \mu+2\sigma) &\approx 95.45\% \\
P(\mu-3\sigma \leq x \leq \mu+3\sigma) &\approx 99.73\%
\end{aligned}$$

Where x is a realization of the random variable X.

From the formula for the probability density function it can be gleaned that the tails of the Gaussian distribution decline like $O(e^{-\lambda|x|^2})$, i.e. they are extremely thin, making this distribution unsuitable for modeling phenomena with heavy tails.

2.2.2 Student's t-Distribution

The Student's t-distribution is a symmetric, unimodal distribution with declining, unbounded tails on both sides of the center. It originates from the estimation of the mean of a normally distributed population, where the population variance is unknown and the sample size is small. The Student's t-distribution has one main feature that distinguishes it from the normal distribution: in general its tails are heavier, meaning they fall off like a power of $|x|$. The decay of the tails is governed by the only parameter ν, the number of degrees of freedom. The similarity between the Student's t- and standard normal distribution grows with ν; if $\nu = \infty$, the Student's t-distribution and the normal distribution are identical (Walck (2007)).

Mathematical Definition:
Probability density function:

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

Where $\Gamma(\cdot)$ and $\mathrm{B}(c,d) = \frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)}$ are respectively the gamma and the complete beta functions, needed to normalize the distribution.

Cumulative distribution function:

$$F_X(x;\nu) = \begin{cases} \frac{1}{2} I_{\frac{\nu}{\nu+x^2}}\left(\frac{\nu}{2},\frac{1}{2}\right) & \text{for } x \leq 0 \\ 1-\frac{1}{2} I_{\frac{\nu}{\nu+x^2}}\left(\frac{\nu}{2},\frac{1}{2}\right) & \text{for } x > 0 \end{cases}$$

Where $I_y\left(c,d\right) = \frac{\mathrm{B}(y;c,d)}{\mathrm{B}(c,d)}$ is the regularized incomplete beta function with $\mathrm{B}(y;c,d)$ being the incomplete beta function.

Expected value:

$$\mathrm{E}(X) = \begin{cases} \text{undefined} & \text{for } \nu \leq 1 \\ 0 & \text{for } \nu > 1 \end{cases}$$

Variance:

$$\mathrm{Var}(X) = \begin{cases} \text{undefined} & \text{for } \nu \leq 1 \\ \infty & \text{for } 1 < \nu \leq 2 \\ \frac{\nu}{\nu-2} & \text{for } \nu > 2 \end{cases}$$

The above formula for the variance implies that when estimating ν for a model, it necessary to set a lower bound b, $2 < b \leq \tilde{\nu}$ for the variance of the model to be finite.

2.2.3 Generalized Inverse Gaussian Distribution

The generalized inverse Gaussian (GIG) distribution is a continuous probability distribution family which includes, among others, the Gamma distribution. Its relevance in this paper derives from it being the distribution of the mixing variable W in the variance-mean mixture generating the generalized hyperbolic distribution (Alexander F. McNeil and Embrechts (2015)).

Mathematical Definition:
Probability density function:

$$f_W\left(w;\lambda,\chi,\psi\right) = \frac{\chi^{-\lambda}\left(\chi\psi\right)^{\frac{\lambda}{2}}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} w^{\lambda-1} e^{-\frac{1}{2}\left(\frac{\chi}{w}+\psi w\right)}$$

Where $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ. The parameters χ and ψ are subject to:

$$\begin{cases} \chi > 0, \psi \geq 0 & \text{if } \lambda < 0 \\ \chi > 0, \psi > 0 & \text{if } \lambda = 0 \\ \chi \geq 0, \psi > 0 & \text{if } \lambda > 0 \end{cases}$$

Expected value:

$$\mathrm{E}(W) = \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}$$

Variance:

$$\mathrm{Var}(W) = \frac{\chi}{\psi}\left[\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)} - \left(\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\right)^2\right]$$

The formulas for all the moments are valid only if χ and ψ are strictly positive.

2.2.4 Generalized Hyperbolic Distribution

The generalized hyperbolic (GH) distribution is a continuous probability distribution superclass which includes several classes of distributions, e.g. the Student's t-distribution. Introduced by Barndorff-Nielsen during his studies of aeolian processes, it possesses the same closure properties as the Gaussian distribution, making it relatively easy to handle analytically. Furthermore, it allows to include both asymmetric and heavy tails, i.e. tails that decline like a power law, which makes it suitable for modeling the dynamics in financial markets. In its multivariate version, the parameters are: (1) λ: real number; (2) χ: real number; (3) ψ: real number; (4) $\boldsymbol{\mu}$: location vector; (5) Σ: dispersion matrix; (6) $\boldsymbol{\gamma}$: asymmetry vector. The parameters $\boldsymbol{\mu}$ and Σ are not, in general, related to respectively the mean vector and the covariance matrix of the GH distribution because of the normal variance-mean mixture explained below (Alexander F. McNeil and Embrechts (2015)).

Mathematical Definition:

$\boldsymbol{X} \sim GH(\boldsymbol{x};\lambda,\chi,\psi,\boldsymbol{\mu},\Sigma,\boldsymbol{\gamma})$ is a normal variance-mean mixture distribution, meaning:

$$\boldsymbol{X} = \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\boldsymbol{Z}$$

Where W is a random variable that follows a generalized inverse Gaussian distribution $f_W(w; \lambda, \chi, \psi)$, A is a real-valued matrix subject to $AA^\top = \Sigma$, and $\boldsymbol{Z} \sim (\boldsymbol{0}, I)$.

Probability density function:

$$\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}; \lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) &= \int \frac{w^{\frac{-d}{2}} e^{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2w} - w\frac{\boldsymbol{\gamma}^\top\Sigma^{-1}\boldsymbol{\gamma}}{2}\right)} f_W(w; \lambda, \chi, \psi)\,\mathrm{d}w \\
&= c\frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi + (\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\,(\psi + \boldsymbol{\gamma}^\top\Sigma^{-1}\boldsymbol{\gamma})}\right) \mathrm{e}^{(\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\chi + (\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\,(\psi + \boldsymbol{\gamma}^\top\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}-\lambda}}
\end{aligned}$$

Where c is the normalizing constant

$$c = \frac{(\chi\psi)^{-\frac{\lambda}{2}}\psi^\lambda(\psi + \boldsymbol{\gamma}^\top\Sigma^{-1}\boldsymbol{\gamma})^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_\lambda(\sqrt{\chi\psi})}$$

d is the number of dimensions (i.e. variables), and $K_i(\cdot)$ is the modified Bessel function of the third kind with index i.

Expected value:

$$\mathrm{E}(\boldsymbol{X}) \overset{(1)}{=} \boldsymbol{\mu} + \mathrm{E}(W)\boldsymbol{\gamma} \overset{(2)}{=} \boldsymbol{\mu} + \sqrt{\frac{\chi}{\psi}}\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\boldsymbol{\gamma}$$

Covariance:

$$\mathrm{Cov}(\boldsymbol{X}) \overset{(1)}{=} \mathrm{E}(W)\Sigma + \mathrm{Var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}^\top \overset{(2)}{=} \sqrt{\frac{\chi}{\psi}}\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\Sigma + \frac{\chi}{\psi}\left[\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)} - \left(\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\right)^2\right]\boldsymbol{\gamma}\boldsymbol{\gamma}^\top$$

From the first equality (1) in each of the above two formulas it can be inferred why in general $\boldsymbol{\mu}$ and Σ are not related to respectively the mean vector and the covariance matrix of $\boldsymbol{X}$. For $\boldsymbol{\mu}$ to be the mean vector, the joint distribution has to be symmetric, i.e. $\boldsymbol{\gamma} = \boldsymbol{0}$; additionally, for Σ to correspond to the covariance matrix, $\mathrm{E}(W) = 1$ has to hold. The last step (2) from the general to the specific form of the moment formulas is valid only if the first two moments of W can be calculated analytically, i.e. if χ and ψ are strictly positive. It should be noted that, all else being equal, for $\boldsymbol{\gamma} = \boldsymbol{0}$ the GH distribution is symmetric because the normal variance-mean mixture above simplifies to a normal mean mixture $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z}$. It is the mixture of variances that enables the GH distribution to have tails with different weights.

2.2.5 Skewed Student's t-Distribution

The skewed Student's t-distribution (SST) is a limiting case of the GH distribution that occurs when $\lambda = -\frac{\nu}{2}$, $\chi = \nu$, and $\psi = 0$, ν being the degrees of freedom of one tail. By calibrating the parameters, the tails of the SST distribution can be made to have different weights, so that it may be used to model phenomena which display asymmetric behavior and exhibit extreme movements in one direction more often than in the other, e.g. financial instruments with large losses more frequent than windfall profits, a characteristic of many equities. This means that in the multivariate case, for $\gamma \neq 0$ the tails behave as follows (Kjersti Aas (2006)):

$$\gamma < 0 : \begin{cases} \text{The left tail decays like} & O\left(|x|^{-\left(\frac{\nu}{2}+1\right)}\right) \\ \text{The right tail decays like} & O\left(|x|^{-\left(\frac{\nu}{2}+1\right)} e^{-2|\gamma||x|}\right) \end{cases}$$
$$\gamma > 0 : \begin{cases} \text{The left tail decays like} & O\left(|x|^{-\left(\frac{\nu}{2}+1\right)} e^{-2|\gamma||x|}\right) \\ \text{The right tail decays like} & O\left(|x|^{-\left(\frac{\nu}{2}+1\right)}\right) \end{cases}$$

Mathematical Definition:
Probability density function:

$$f_X(\boldsymbol{x}; \nu, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) = c \frac{K_{\frac{\nu+d}{2}}\left(\sqrt{(\nu + (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\, \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma}}\right) e^{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} \boldsymbol{\gamma}}}{\left(\left(\nu + (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right) \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma}\right)^{-\frac{\nu+2}{4}} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{\frac{\nu+d}{2}}}$$

Where c is the normalizing constant

$$c = \frac{2^{\left(1-\frac{\nu+d}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}$$

Expected value:

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} + \frac{\nu}{\nu-2} \Sigma^{-1} \boldsymbol{\gamma}$$

Covariance:

$$\mathrm{Cov}(\boldsymbol{X}) = \begin{cases} \infty & \text{for } \nu \leq 4 \\ \frac{2\nu^2}{(\nu-2)^2(\nu-4)} \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma} + \frac{\nu}{\nu-2} I & \text{for } \nu > 4 \end{cases}$$

Because of the covariance formula, the estimated $\tilde{\nu}$ for an asymmetric model delivers a distribution with finite covariance only if $\tilde{\nu} > 4$. Being a special case of the GH distribution, the SST distribution becomes a (non-central) symmetric Student's t-distribution when $\boldsymbol{\gamma} \to \mathbf{0}$ (Alexander F. McNeil and Embrechts (2015), Kjersti Aas (2006)).

2.2.6 Copulas

A copula C is a multivariate cumulative distribution function on the d-dimensional unit cube $[0,1]^d$ with standard uniform marginal distributions $U_n \sim \mathrm{U}(0,1)$, i.e. a mapping $C : [0,1]^d \to [0,1]$. Copulas are omnipresent and of foremost importance in any field employing multivariate statistics. A copula describes the interdependence among random variables of a statistical model and can be used to produce multivariate distributions with specific dependence structures. The wide use of copulas is justified by Sklar's theorem: it proves that every multivariate cumulative distribution function (CDF) has a copula, and that copulas can be used to join univariate CDFs to create a specific multivariate distribution. Sklar's theorem can be summarized as follows (Alexander F. McNeil and Embrechts (2015)):

1. There exists a copula C such that, $\forall(x_1, \dots, x_d) \in \mathbb{R}$:

$$F(x_1, \dots, x_d) = C\left(F_1(x_1), \dots, F_d(x_d)\right)$$

Where $F(\cdot)$ is a joint CDF with marginal CDFs $F_1(\cdot), \dots, F_d(\cdot)$.

This means that all multivariate distributions have a copula which couples together the marginal CDFs.

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

Where u_i denotes the i-th component of the d-dimensional random vector $\boldsymbol{U} \sim \mathrm{U}(0,1)$ and $F_i^{-1}(\cdot)$ is the inverse CDF $F_i^{-1}(u_i) = \inf\{x_i : F_d(x_i) \geq u_i\}$.

This shows how a copula can be extracted from a joint CDF, and that the dependence is in terms of u_i-quantiles.

Mathematical Definition:

For C to be a copula, it has to fulfil three properties:

1. $C(u_1, \dots, u_d) = 0$ if $u_i = 0$ for any $i \in [1, d]$
2. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i \; \forall i \in \{1, \dots, d\}, u_i \in [0, 1]$

This property ensures that marginal distributions are uniform.

3. $\forall(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0,1]^d, a_i \leq b_i : \displaystyle\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0$

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

This inequality ensures that for a random vector $(U_1, \dots, U_d)^\top$ with CDF C, the probability $P(a_1 \leq U_1 \leq b_1, \dots, a_d \leq U_d \leq b_d)$ is non-negative.

In the context of copulas, it is important to delve into the topic of dependence. In statistics, dependence is a concept used to describe how variables interact and to try to establish possible cause-and-effect relationships. In science, this notion of dependence often becomes a concrete description of the mechanism in a phenomenon, commonly referred to as law, e.g. Newton's laws of motion. However, the dynamics of financial assets follow a very limited and porous set of laws, if any, and most approaches are mere attempts at modeling the behavior of instruments. Nonetheless, even in mathematical finance some facts can be established by analyzing historical data; very often, one of these stylized facts is the interdependence of extreme movements in the returns on a set of instrument. From a statistical perspective, this means that the (absolute) correlations between the variables increase with the distance from the center of the distribution. However, Pearson's linear correlation can be a misleading dependence measure. One can generate multivariate sets of data with identical correlation matrices, but with observable differences in their behaviors, even when by starting from identical marginal distributions; this fallacy is due to linear correlation not being exclusively a function of the copula of the joint distribution. Still, as explained by McNeil et al. (Alexander F. McNeil and Embrechts (2015)), linear correlation is a reliable dependence measure for elliptical distributions, which can be explained by the fact that these distributions are a fully described by their mean vector μ, covariance matrix Σ and characteristic generator. Mean and variance being the sole parameters influencing the marginal CDFs, it follows that the copula of the related elliptical distribution solely depends on the correlation matrix and characteristic generator. Furthermore, even copulas of elliptical distribution can exhibit interesting traits. The t copula, based on the elliptical Student's t-distribution, depends on both the correlation matrix and the degrees of freedom ν; from a practical standpoint, ν allows to introduce tail dependence. This entails that, as demonstrated by Demarta and McNeil (Demarta and McNeil (2005)), as long as $\nu < \infty$ there is tail dependence in the multivariate distribution, an indication that extreme outcomes in each direction are dependent on one another. In contrast, joint distributions based on the Gaussian copula, which depends solely on the correlation matrix, are independent in the tails if the correlations are zero. In this paper, modeling financial instruments is approached with a skew Student's t-distribution based on the GH distribution. The copula of the SST has a third parameter $\boldsymbol{\gamma}$, which introduces the sought skewness in the dependence structure. Figure 2 below shows how the t copula still grows towards the quadrants' extremities even when $\rho = 0$, as opposed to its Gaussian counterpart.

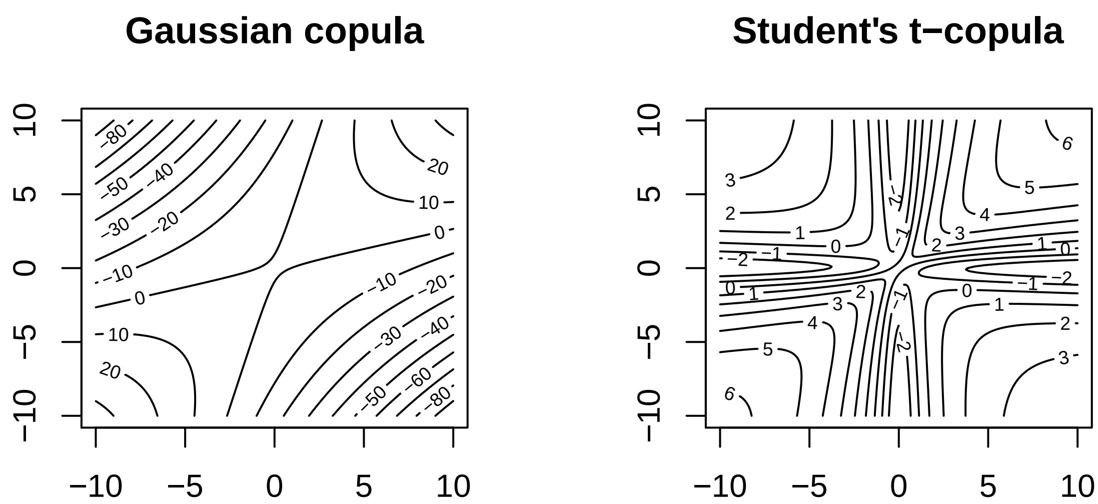

Figure 1: Log-density of Gaussian copula vs. Student's t-copula with $\rho = 0.5, \nu = 2.5$

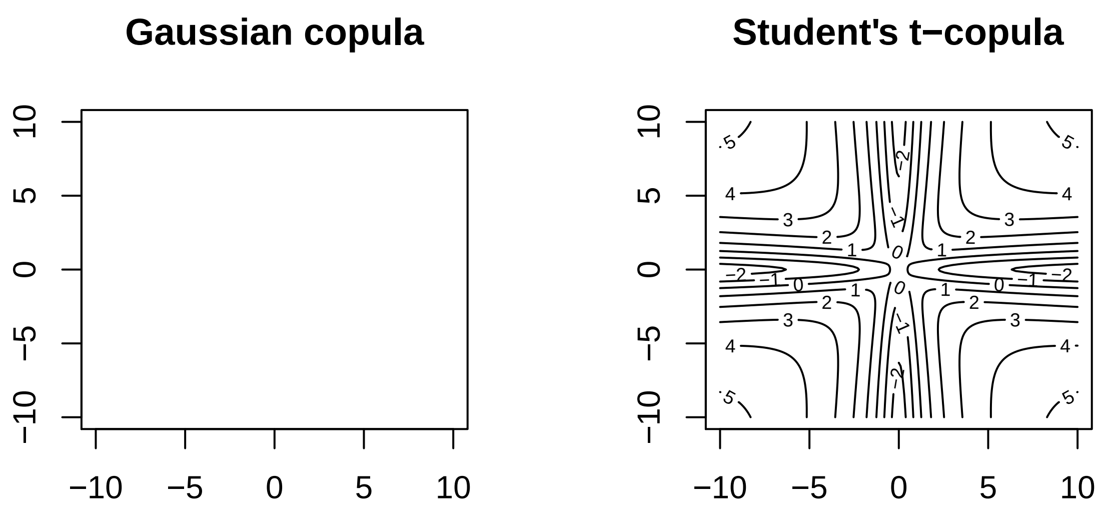

Figure 2: Log-density of Gaussian copula vs. Student's t-copula with $\rho = 0, \nu = 2.5$

2.3 Portfolio Optimization (Mean-Variance Analysis)

Mean-variance analysis (MVA), also known as modern portfolio theory, is a mathematical approach to the problem of asset allocation in a portfolio. It was introduced by Harry M. Markowitz in an article published in the "Journal of Finance" in March 1952. Markowitz divided the process of portfolio selection into two separate stages: the first stage is concerned with gathering current knowledge and beliefs about the future performance of the assets, the second stage focuses on portfolio selection basing on relevant data collected in the first stage. Mean-variance analysis is the mathematical framework used to execute the second stage. Markowitz dismissed the idea of return maximization as an adequate criterion to allocate a portfolio and instead pointed to the risk-reducing property of diversification, whereby selecting assets with uncorrelated price behavior reduces the variance of the portfolio. In the context of MVA, variance is the proxy for risk and increased variance is the premium for selecting assets with higher expected returns. Mean-variance analysis focuses on maximizing the expected return on a portfolio for a set level of variance by finding a solution on the efficient frontier in the risk-return space (Luenberger (1997)).

Mathematical Definition:
Variance of a portfolio:

$$\sigma^2 = \boldsymbol{w}^\top \Sigma \boldsymbol{w}$$

Where Σ is the covariance matrix of the expected returns on the N selected assets and $\boldsymbol{w}$ is the unknown vector expressing their weights $w_1, \ldots, w_N$ in the portfolio.

The vector $\boldsymbol{w}$ is determined such that a predetermined expected return on the portfolio r_{pf} is reached while minimizing σ^2. The quadratic matrix equation can be formulated as follows:

$$\arg\min_{\boldsymbol{w}} \frac{\sigma^2}{2} = \arg\min_{\boldsymbol{w}} \frac{1}{2} \boldsymbol{w}^\top \Sigma \boldsymbol{w}$$

Where $\arg\min_{\boldsymbol{w}}$ minimizes the expression through the best $\boldsymbol{w}$.

The above equation is subject to the constraints:

$$\begin{cases} \boldsymbol{w}^\top \boldsymbol{r} = r_{\text{pf}} & (1) \\ \boldsymbol{w}^\top \mathbf{1} = 1 & (2) \end{cases}$$

The first equation minimizes the portfolio's half variance $\frac{\sigma^2}{2}$; $\frac{\sigma^2}{2}$ being a linear transformation of σ^2, the solution set to the first expression also minimizes the portfolio's full variance σ^2. This specific formulation further eases the quadratic matrix equation, which turns into a linear expression when using the method of Lagrange multipliers. Condition (1) requires that the portfolio with expected returns on single assets $\boldsymbol{r}^\top = r_1, \ldots, r_n$ generates the investor's chosen aggregated expected return r_{pf} and condition (2) ensures that the weights of the portfolio's assets sum up to one. This constrained minimization problem can be expressed by the Lagrange function:

$$\mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2) = \frac{1}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w} - \lambda_1(\boldsymbol{w}^\top \boldsymbol{r} - r_{\text{pf}}) - \lambda_2(\boldsymbol{w}^\top \mathbf{1} - 1)$$

Where λ_1 and λ_2 are the Lagrange multipliers.

The solution $\boldsymbol{w}$ is found by solving the system of partial differential equations:

$$\boldsymbol{\nabla}_{\boldsymbol{w}, \lambda_1, \lambda_2} \mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2) = \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}, \frac{\partial \mathcal{L}}{\partial \lambda_1}, \frac{\partial \mathcal{L}}{\partial \lambda_2}\right) = (\mathbf{0}, 0, 0)$$

Because of the first order partial derivatives, the quadratic expression of the $\frac{\sigma^2}{2}$-minimization part in $\mathcal{L}$ becomes linear, leaving a system of $N + 2$ linear equations with $N + 2$ variables:

$$\begin{cases} \Sigma \boldsymbol{w} - \lambda_1 \boldsymbol{r} - \lambda_2 \mathbf{1} = \mathbf{0} \\ \boldsymbol{w}^\top \boldsymbol{r} - r_{\text{pf}} = 0 \\ \boldsymbol{w}^\top \mathbf{1} - 1 = 0 \end{cases}$$

This system of linear equations has the explicit, unique solution $\boldsymbol{w}^* = \Sigma^{-1}(\lambda_1 \boldsymbol{r} + \lambda_2 \mathbf{1})$, where

$$\lambda_1 = \frac{(\mathbf{1}^\top \Sigma^{-1} \mathbf{1}) r_{\text{pf}} - \boldsymbol{r}^\top \Sigma^{-1} \mathbf{1}}{(\boldsymbol{r}^\top \Sigma^{-1} \boldsymbol{r})(\mathbf{1}^\top \Sigma^{-1} \mathbf{1}) - (\boldsymbol{r}^\top \Sigma^{-1} \mathbf{1})^2}$$

$$\lambda_2 = \frac{\boldsymbol{r}^\top \Sigma^{-1} \boldsymbol{r} - (\boldsymbol{r}^\top \Sigma^{-1} \mathbf{1}) r_{\text{pf}}}{(\boldsymbol{r}^\top \Sigma^{-1} \boldsymbol{r})(\mathbf{1}^\top \Sigma^{-1} \mathbf{1}) - (\boldsymbol{r}^\top \Sigma^{-1} \mathbf{1})^2}$$

The above process can be reformulated as a r_{pf}-maximization problem for a given level of σ^2. If the chosen σ^2 is the variance of the portfolio found for the σ^2-minimization problem, all else being equal the r_{pf}-maximization problem leads to the same solution $\boldsymbol{w}^*$.

MVA in its original version has the main disadvantage of having variance as its risk metric, and thus it is prone to the same pitfalls that can be incurred when basing financial risk management methods on volatility. However, MVA can be modified to implement another risk metric, provided the latter is coherent with the (expected) returns distribution followed by the assets.

2.4 R - Statistical Programming Language

The entirety of the practical part of the research is conducted using the programming language R and the RStudio interface. The other candidate for the choice of programming language is Python, over which R is selected for various reasons. In contrast to Python, which is a general-purpose programming language, R is specifically developed for statistical computing; in the case of this paper, the specialized nature of R means that the software includes many useful functions without prior installation and loading of additional packages. The abundance of built-in statistical functions in R also makes the code more readable and concise in the case of statistical computing. We also find that R integrates computing and graphing capabilities more intuitively, which is important for visualizing and analyzing results. A further reason for selecting R is the fact that it is better understood by the researchers and the academic advisor, thus allowing better cooperation and feedback. RStudio is chosen as the integrated developed environment because its features are specifically intended to for use with the R language; this characteristic and the user-friendly graphics device enable efficient coding and data visualization.

2.5 GARCH Model

The generalized autoregressive conditional heteroskedasticity (GARCH) statistical model is used for modeling time series, in particular to model data where the error terms ϵ_t, e.g. the difference between the logarithms of the t-th observation and the previous one $\ln(x_t) - \ln(x_{t-1})$, are affected by alternating periods of lower and higher variance, the latter known in finance as volatility clusters. GARCH models are generalized autoregressive methods because they generate forecasts for the instantaneous error variance σ_t^2 based on an autoregressive moving average (ARMA) process, as opposed to the autoregressive (AR) process used in simpler ARCH models; conditional heteroskedasticity means that σ_t^2 depends on past data, allowing it to change over time.

Mathematical Definition:

A GARCH(n, m) model for ϵ_t has:

$$\epsilon_t = \sigma_t u_t$$

$$\sigma_t^2 = c\sigma^2 + \sum_{j=1}^{n} \alpha_j \sigma_{t-j}^2 + \sum_{k=1}^{m} \beta_k \epsilon_{t-k}^2$$

Where ϵ_t is the t-th error term, u_t is a white noise process with independent and identically distributed random variables having $\mathrm{E}(u_t) = 0$ and $\mathrm{Var}(u_t) = 1$, σ_t^2 is the conditional variance $\mathrm{Var}(\epsilon_t|\sigma_{t-1}, \ldots, \sigma_{t-n}, \epsilon_{t-1}, \ldots, \epsilon_{t-m})$, c is the fixed weight of the unconditional variance $\sigma^2 = \mathrm{Var}(\epsilon_t)$, n and m are the specified lag lengths, and α_j and β_k are the weights of past data points for each respective given lag j and k; the quantities c, β_k, and α_j have to be estimated (Wildi (2018)).

In this paper, the GARCH model is used to confirm the existence of volatility clusters in the data and thus justify methods that do not assume normally distributed logarithmic returns.

2.6 MCECM Algorithm

The expectation-maximization (EM) algorithm is a numerical method for the estimation of statistical parameters. The EM procedure is formalized as a problem where both unknown parameters $\boldsymbol{\theta}$ and missing values $\boldsymbol{Z}$ have to be estimated simultaneously, the observed data being $\boldsymbol{X}$. The EM algorithm alternates between the E-step and the M-step to solve this problem.

Mathematical definition:

The steps of the EM method at the t-th iteration are the following (Neath (2013)):

1. Random initialization of the estimates $\tilde{\boldsymbol{\theta}}_t$
2. E-step: Define $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}_t) = \mathrm{E}_{\boldsymbol{Z}|\boldsymbol{X},\tilde{\boldsymbol{\theta}}_t}\left[\ln\left(L\left(\boldsymbol{\theta};\boldsymbol{X},\boldsymbol{Z}\right)\right)\right]$

Where $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}_t)$ is the expected value of the log-likelihood $\ln\left(L\left(\boldsymbol{\theta};\boldsymbol{X},\boldsymbol{Z}\right)\right)$ with complete data, given $\boldsymbol{X}$ and the estimates $\tilde{\boldsymbol{\theta}}_t$.

3. M-step: Find $\tilde{\boldsymbol{\theta}}_{t+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}_t)$

Where $\tilde{\boldsymbol{\theta}}_{t+1}$ is the estimate which maximizes $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}_t)$.

4. Iterate the E- and M-steps until a chosen degree of convergence is reached

The multicycle expectation conditional maximization (MCECM) algorithm is a variant of the EM algorithm used in this paper. The MCECM algorithm replaces the M-step with a CM-step, i.e. a sequence of conditional maximizations (CMs) of $\mathrm{E}_{\boldsymbol{Z}|\boldsymbol{X},\tilde{\boldsymbol{\theta}}_t}\left[\ln\left(L\left(\boldsymbol{\theta};\boldsymbol{X},\boldsymbol{Z}\right)\right)\right]$, and performs multiple E-steps during the same iteration, e.g. before each conditional maximization (Andrew Gelman (2014)). In this paper, the estimation of the parameters of the multivariate skewed Student's t-distribution is carried out by way of the MCECM algorithm implemented in the R ghyp package (Luethi and Breymann (2016)).

3 Analysis

3.1 Modelling Market Data

The first step of our research project is to quantify the:

- heavy tails (kurtosis)
- asymmetry (skewness)
- tail correlation
- volatility clustering

of the distributions of log-returns of the following financial assets:

Table 1: Description of data

Asset.Name	Class	Description
DIJA	Equities	Stock market index of of 30 large, publicly owned companies in the United States.
S&P500	Equities	Stock market index of 500 large-capitalization companies
US 10-Year Treasury Note	Fixed income (bonds)	US government debt issued by the Department of the Treasury
JPY/USD Currency Pair	Foreign exchange	Exchange rate of US dollars against one Japanese yen
Gold	Commodities	

3.1.1 Heavy Tails and Asymmetry

In figure 3 it is shown that the standardized S&P500 log-returns are not normally distributed as the observations in the tails, especially the lower tail, are outside the confidence band of the standard normal distribution. Additionally, it is shown that the observations in the lower tail are more pronounced than in the upper tail, which contrasts with the characteristic of asymmetry as under the normality assumption the observations would be evenly distributed. In contrast to the S&P500 index, the JPY/USD exchange rate shows an inverted asymmetry, which can be explained by considering what happens in the minds of investors during an economic collapse such as the 2008 financial crisis. When markets crash investors tend to panic sell risky assets to minimize their losses and invest in more secure ones. This behavior leads to even further losses in the riskier assets, yet higher than usual gains in the more secure ones. These assets are considered “safe haven” instruments as they promise higher returns during market turmoil e.g. the Japanese Yen, the Swiss Franc and government bonds. In table 2 correlations between assets are illustrated to highlight the concept of safe haven instruments. It is shown that stock indices have a negative relationship with the Japanese Yen, which coincides with the concept of safe haven instruments during market turmoil. Further a t-test for the correlation was carried out to evaluate which correlations are statistically significant (Plackett (1962)). In table 3 the significant t-values relating to the values in table 2 have been highlighted.

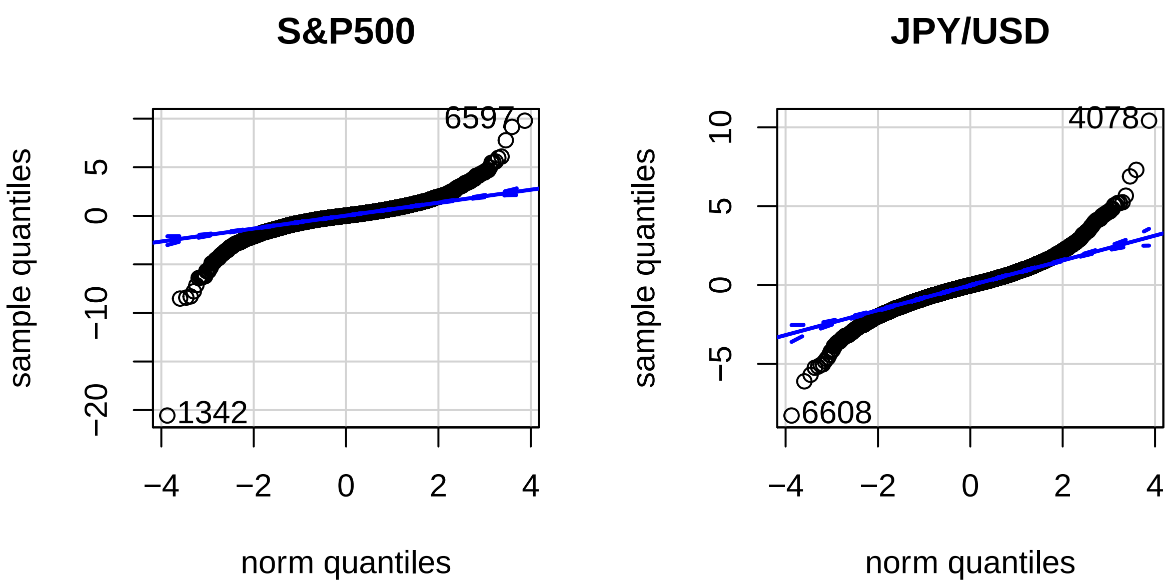

Figure 3: Quantile-Quantile plots of daily, standardized S&P500 and JPY/USD log-returns

Table 2: Correlation of daily log-returns of various assets

	DIJA	S&P500	10Y US Bond Futures	JPY/USD	Gold
DIJA	1.00	0.97	-0.01	-0.17	-0.04
S&P500	0.97	1.00	-0.02	-0.17	-0.02
10Y US Bond Futures	-0.01	-0.02	1.00	0.16	0.06
JPY/USD	-0.17	-0.17	0.16	1.00	0.23
Gold	-0.04	-0.02	0.06	0.23	1.00

Table 3: T-statistics of two-sided t-test to the 95%-significance level for the correlations (significant values are highlighted bold)

	DIJA	S&P500	10Y US Bond Futures	JPY/USD	Gold
DIJA	**Inf**	**383.07**	-0.96	**-16.56**	**-3.84**
S&P500	**383.07**	**Inf**	-1.92	**-16.56**	-1.92
10Y US Bond Futures	-0.96	-1.92	**Inf**	**15.56**	**5.77**
JPY/USD	**-16.56**	**-16.56**	**15.56**	**Inf**	**22.69**
Gold	**-3.84**	-1.92	**5.77**	**22.69**	**Inf**

3.1.2 Volatility Clustering

As demonstrated in figure 4, the absolute S&P500 log-returns are heavily auto-correlated, which indicates that there is volatility clustering. Volatility clusters imply that large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes Mandelbrot (1967), thus the variance itself is not constant and time-dependent. The normal distribution on the other hand assumes constant variance, thus the observations would disperse evenly around its mean. In figure 4 it is also highlighted that the volatility clustering is the most prominent during times of market collapse e.g. 2008 financial crisis. The time-dependent volatility can be estimated by using a GARCH model that assumes volatility to be a latent variable instead of a constant. In figure 5 it is shown that the estimated volatility for the normally distributed sample is dispersing evenly around its true volatility, whereas it varies drastically for the S&P500 log-returns. During market turmoil the estimated volatility is almost six times bigger than what the empirical volatility implies, which could be lead back to the likelihood of tail events. As the occurrence of tail events is usually very unlikely, it seems that they arise more often than the normal distribution assumes and especially during market turbulence.

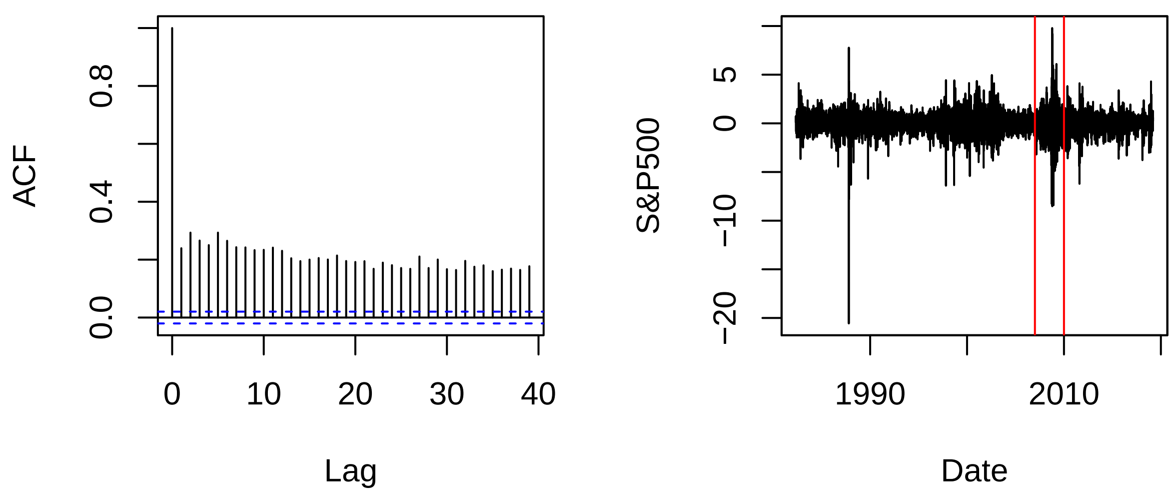

Figure 4: Auto-correlation plot of daily, absolute standardized S&P500 log-returns and time series plot of daily, standardized S&P500 log-returns highlighting the 2008 financial crisis from 01.01.2007 to 01.01.2010

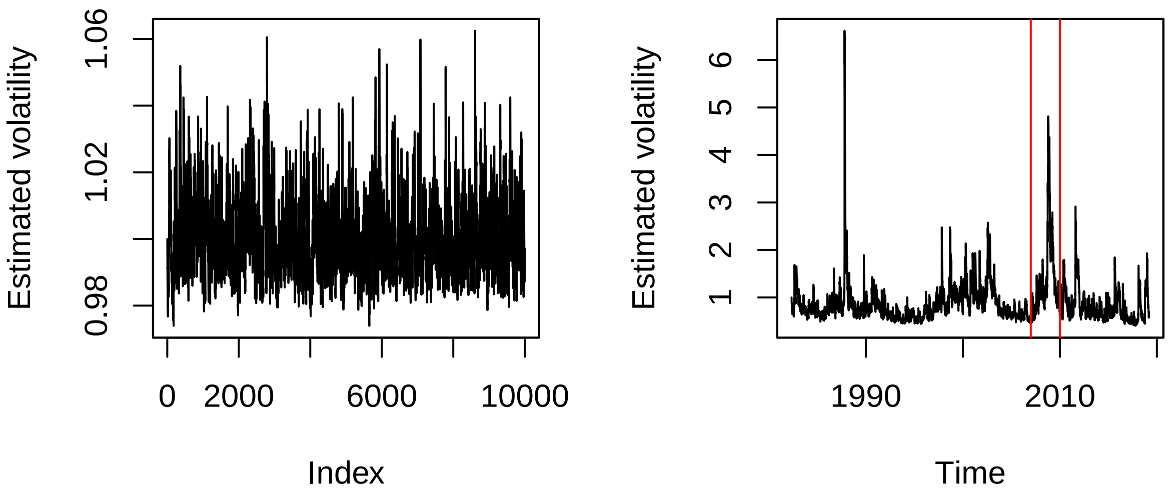

Figure 5: Volatility clustering in normally distributed sample vs. S&P500 log-returns highlighting the 2008 financial crisis from 01.01.2007 to 01.01.2010

3.1.3 Tail Correlation

The scatterplot of SP500 index data and ten-year US bonds futures in figure 6 displays outliers in all four quadrants, as opposed to no outliers in the corners of the scatterplot of two independent normally distributed samples. The key fact is that the relationship between assets cannot be adequately represented by correlation alone as shown in figure 6, since the normal distribution is incapable of replicating the outliers.

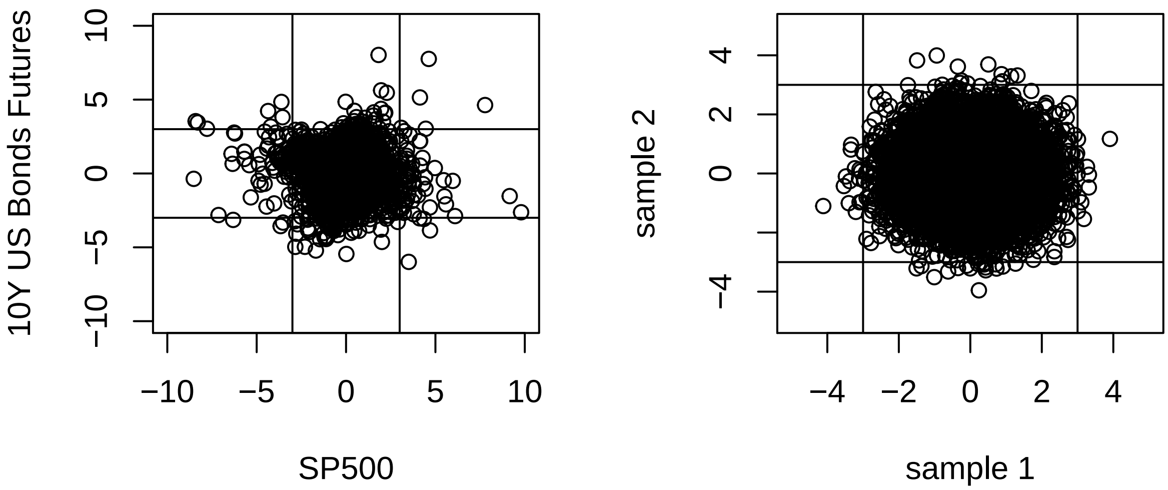

Figure 6: Tail correlations of S&P500 vs. two normally distributed samples

3.1.4 Gauss Optimization

Another issue with normally distributed assets is that a portfolio does not deviate for whichever risk measure used to construct it. The normal distribution does not distinguish between different risk profiles as for example the value-at-risk optimized portfolios are the same for all levels of α and equal to the volatility optimized portfolio as illustrated in the following equations:

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_1 w_2 \mathrm{Cov}(X,Y)$$

$$\frac{\partial \sigma^2}{\partial w_1} = 2w_1\sigma_1^2 + w_2 \mathrm{Cov}(X,Y)$$

$$\frac{\partial \sigma^2}{\partial w_2} = 2w_2\sigma_2^2 + w_1 \mathrm{Cov}(X,Y)$$

$$w_1 = \frac{w_2\sigma_2^2 - w_2\mathrm{Cov}(X,Y)}{\sigma_1^2 - \mathrm{Cov}(X,Y}$$

$$q_\alpha = \mu_{\mathrm{pf}} + \sigma_{\mathrm{pf}} q_{\alpha,\mathrm{norm}}$$

$$q_\alpha = \mu + \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_1 w_2 \mathrm{Cov}(X,Y)}\; q_{\alpha,norm}$$

$$\frac{\partial q_\alpha}{\partial w_1} = \frac{\partial q_\alpha}{\partial w_2}$$

$$w_1 = \frac{w_2\sigma_2^2 - w_2\mathrm{Cov}(X,Y)}{\sigma_1^2 - \mathrm{Cov}(X,Y}$$

Where $q_{\alpha,\mathrm{norm}}$ is the α-quantile of the standard normal distribution.

3.1.5 Aggregated Log-returns

Although the normal distribution is not suitable for modeling daily log-returns, it still has its purposes when the frequency is increased e.g. yearly log-returns. In figure 7 it is shown that the effect of kurtosis and skewness decrease after aggregating the log-returns and the observations are within the confidence interval. Thus, the normal distribution approximates the characteristics of log-returns better when the horizon is increased. Therefore, the horizon must play an important role for the evaluation of a suitable distribution model. Although, the effect of kurtosis and skewness can be reduced, the effect of tail correlation persists.

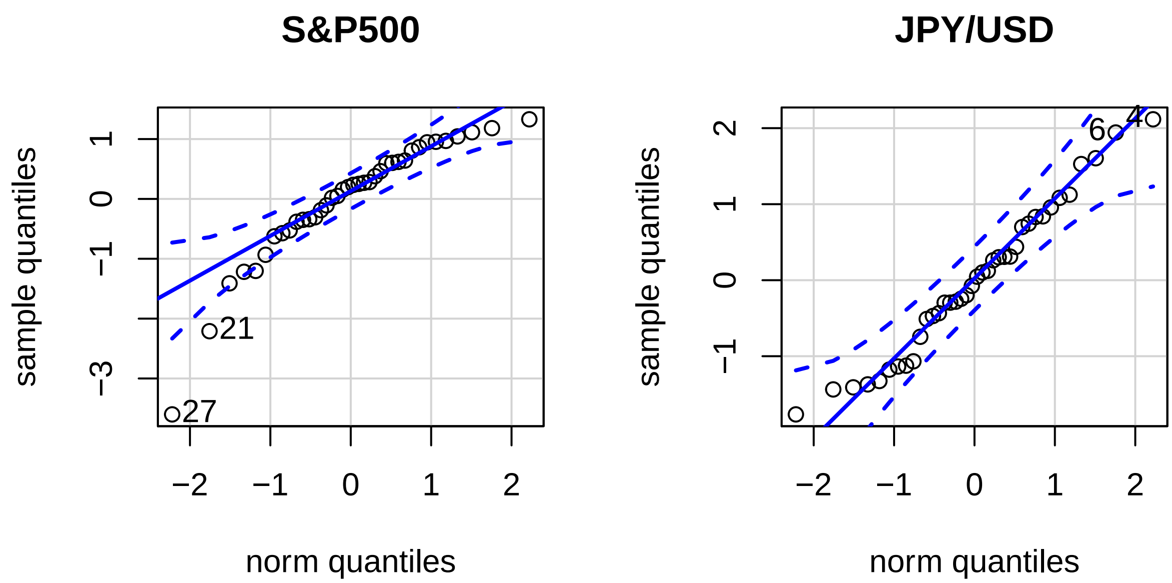

Figure 7: Quantile-Quantile plots of yearly, standardized S&P500 and JPY/USD log-returns

3.2 Distribution Fitting

After the qualitative observations of previous subsections, we now quantify the parameters that describe the fat tails and skewness of the return distributions. As evaluated before the normal distribution is not a suitable choice to model daily log-returns. A more adequate option could be the Student's t-distribution as it exhibits fat tails that differ for the degrees of freedom parameter as shown in figure 8. The Student's t-distribution converges towards the normal distribution as the degrees of freedom parameter is increased. In figure 9 it is demonstrated that the Student's t-distribution approximates the log-returns well as almost all observations lie within the confidence interval. It was assumed that the S&P500 follows a Student's t-distribution with three degrees of freedom respectively four for the JPY/USD exchange rate.

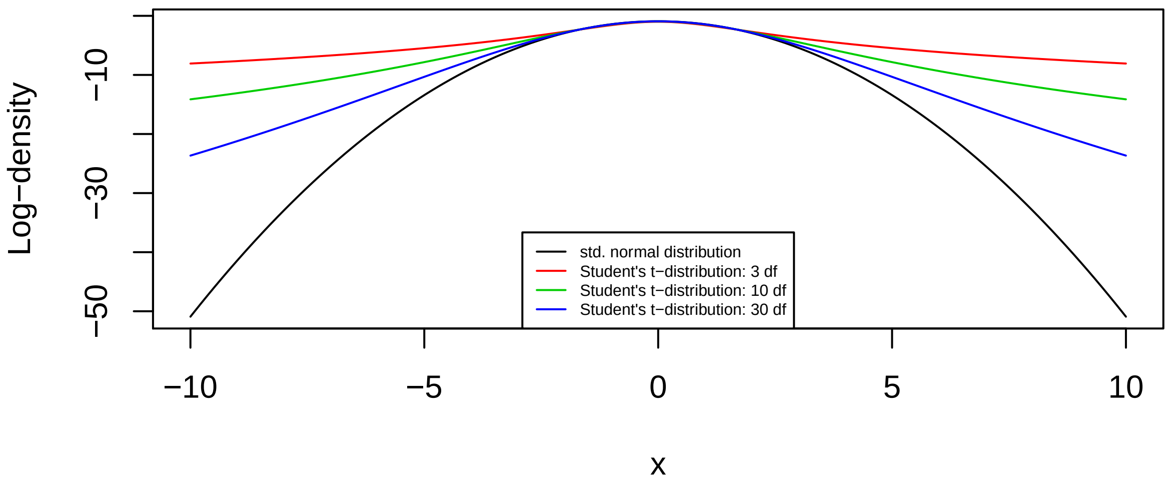

Figure 8: Comparing the log-density for different distributions

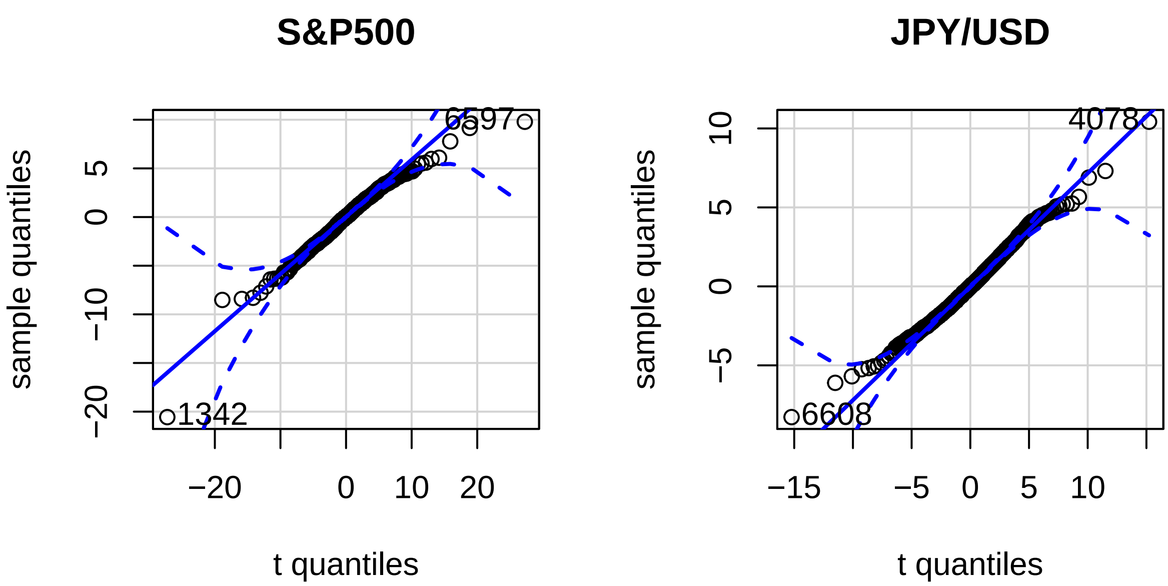

Figure 9: Quantile-Quantile plots of daily, standardized S&P500 and JPY/USD log-returns

Although, the Student t-distribution incorporates kurtosis, it neglects asymmetry as it is a symmetric distribution. Thus, the skewed Student's t-distribution was a logical continuation for the modeling process. At first the data was fitted to univariate skewed Student's t-distributions to evaluate whether the additional skewing parameter was of significance. The fitting process included the usage of the ghyp-package, which was implemented by Wolfgang Breymann and David Luethi. The package focuses on the generalized hyperbolic distributions and especially their contributions to finance and risk management. The optimizer uses a maximum likelihood algorithm to fit a univariate, generalized hyperbolic distribution to provided sample data. In figures 10-11 it is shown that the tails were better approximated by the skewed Student's t-distribution compared to both the normal distribution and the symmetric Student's t-distribution. As the normal distribution underestimates the frequency of extreme events in both the lower and upper tail, the symmetric Student's t-distribution overestimates the upper tail for the S&P500 and overestimates the lower tail for the JPY/USD exchange rate. Further the significance of the skewing parameter was examined by applying a log-likelihood ratio test that tests whether the simpler model is the true underlying model (King (1989)). The simpler model in this case is the symmetric Student's t-distribution as it incorporates one less parameter as the skewed Student's t-distribution. In table 4 it is shown that the p-values of the log-likelihood ratio test are both significant and thus the null hypothesis, which states that the parameter $\gamma = 0$ explains the data, was rejected. As a second measure of goodness of fit a value-at-risk forecast backtest was concluded to evaluate whether violations against the value-at-risk forecast occur more often than suggested. First, the number of violations against the value-at-risk forecast were counted for the normal distribution, the symmetric Student's t-distribution and the skewed Student's t-distribution. Second, a binomial test was conducted to test whether the ratio of violations is higher than suggested by an arbitrary significance level. As shown in tables 5-8 the skewed Student's t-distribution does not violate the value-at-risk forecast for any of the significance levels, thus making it the best available choice to model log-returns.

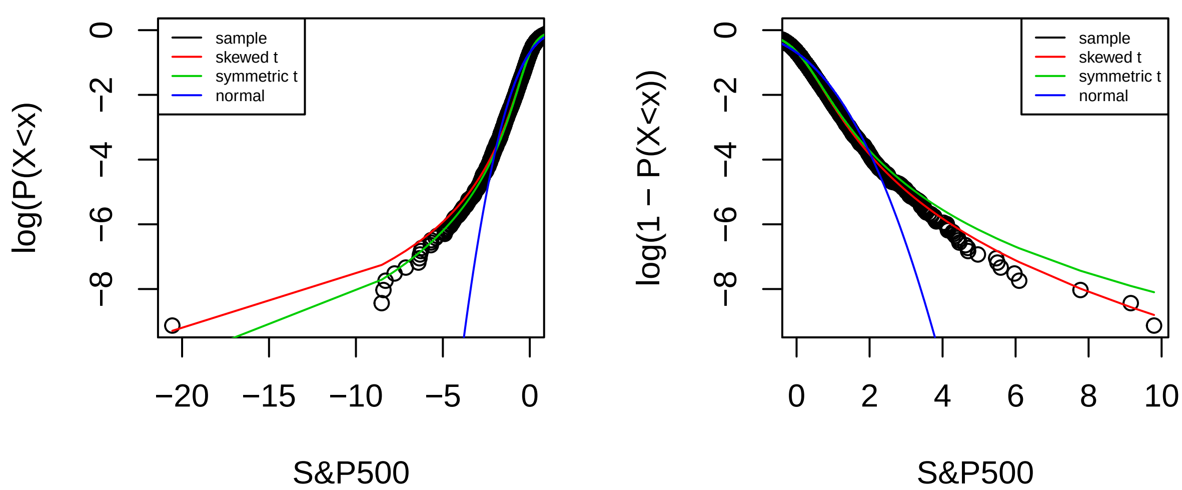

Figure 10: Graphical analysis of the lower and the upper tail regarding the fitted parameters of the univariate skewed Student's t-distribution for the S&P500

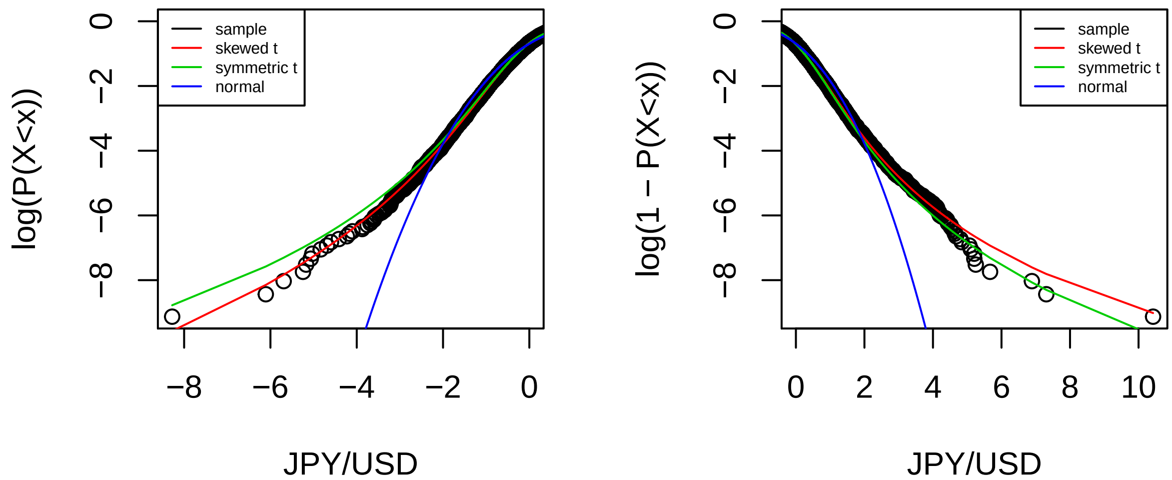

Figure 11: Graphical analysis of the lower and the upper tail regarding the fitted parameters of the univariate skewed Student's t-distribution for the JPY/USD

Table 4: P-values of the log-likelihood ratio test for the SP500 and the JPY/USD

	S&P500	JPY/USD
P-value	0.0037031	0.0011639

Table 5: Observed number of violations of SP500 log-returns against 1-day value-at-risk forecast (lower tail)

	0.001	0.005	0.01
Gaussian	65	115	151
Symmetric Student's t	6	43	89
Skewed Student's t	5	35	72

Table 6: P-values of binomial test of SP500 log-returns against 1-day value-at-risk forecast (lower tail)

	0.001	0.005	0.01
Gaussian	**0**	**0**	**0**
Symmetric Student's t	0.897	0.696	0.644
Skewed Student's t	0.952	0.961	0.987

Table 7: Observed number of violations of JPY/USD log-returns against 1-day value-at-risk forecast (upper tail)

	0.999	0.995	0.99
Gaussian	71	122	169
Symmetric Student's t	8	60	109
Skewed Student's t	4	47	93

Table 8: P-values of binomial test of JPY/USD log-returns against 1-day value-at-risk forecast (upper tail)

	0.999	0.995	0.99
Gaussian	**0**	**0**	**0**
Symmetric Student's t	0.701	**0.028**	**0.047**
Skewed Student's t	0.982	0.466	0.479

With the information obtained by assessing the characteristics of log-returns and the accurate approximation of the univariate skewed Student's t-distribution, the multivariate skewed Student's t-distribution was chosen to model the behavior of log-returns. The multivariate skewed Student's t-distribution is a suitable choice as it accounts both for extreme values in the tails and also their skewed appearance. Additionally, it incorporates the tail correlation structure that was examined in the chapter before, which is crucial when modeling non-linear dependence. It also has computing advantages as linear combinations of the marginal distributions can be described as a univariate skewed Student's -t distribution as explained in the following equations:

$$\overline{\mu} = \boldsymbol{w}^\top \boldsymbol{\mu}$$

$$\overline{\sigma} = \sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}$$

$$\overline{\gamma} = \boldsymbol{w}^\top \boldsymbol{\gamma}$$

The estimation method used to estimate the multivariate skewed Student's t-distribution uses a multi-cycle, expectation, conditional estimation (MCECM) algorithm to identify the unknown parameters. It adopts an iterative approach by updating the starting values for the parameters after each iteration. In table 9 the fitted parameters are summarized, the first parameter refers to nu whereas the rest of the parameters are the individual skewing parameters.

Table 9: Bootstrap estimated parameters for a multivariate skewed Student's t-distribution (N = 100)

ν	γ_{DIJA}	γ_{SP500}	$\gamma_{10YUSBonds}$	$\gamma_{JPY/USD}$	γ_{Gold}
4.18±0.107	-0.065±0.017	-0.077±0.018	-0.021±0.017	0.07±0.019	-0.025±0.017

3.3 Portfolio Optimization and the Role of the Investment Horizon

3.3.1 Brief Introduction

In the previous sub-section, we have modelled the returns of various asset classes (equities, bonds, FX rates and commodities) by symmetric and skewed Student's t-distributions. In this sub-section, we discuss optimal asset allocations for these asset classes. As risk measures, we use value-at-risk and expected shortfall.

To define value-at-risk and expected shortfall, we must specify two parameters: the time horizon T ($T = 1, 5, 22, 250$ for daily, weekly, monthy, yearly returns) and the level of confidence α. If α were chosen independently of T, the optimization results for different horizons would not be comparable. E.g., $\alpha = 99\%$ value-at-risk of yearly returns vs. daily returns measures, roughly, the worst annual loss expected in 100 years vs. the worst daily loss expected in 100 days, or approximately 5 months. However, to cover the same time period of 100 years, i.e. 25,000 days, we should instead compare the 99% value-at-risk of yearly returns with the $1 - \frac{1}{25000} = 99.996\%$ value-at-risk of daily returns, and 99% expected shortfall of yearly returns with 99.996% expected shortfall of daily returns.

To illustrate this point, suppose we sell insurance against a major earthquake in Zürich against an insurance premium. Suppose such an earthquake occurs, on average, once in a hundred years and causes a loss of CHF 1 billion within one day. In this case, 99% 1-year expected shortfall for this "investment" is CHF 1 billion, which is the same as the 99.996% 1-day expected shortfall. By comparison, the 99% 1-day expected shortfall would only be CHF $\frac{\text{CHF1billion}}{250} = \text{CHF4million}$.

In the following, we let α depend on τ as follows: $\alpha = 1 - \frac{\tau}{30000}$. Thus, our risk measure is sensitive to losses that occur once in up to 30,000 days, or 117 years (we choose 30,000 days, because $1 - \frac{1}{30000} = 99.997\%$ VaR corresponds to a 4-standard-deviation event for normal distributions).

In a first step, we model the asset classes by symmetric t-distributions in sub-section 3.3.2, choosing a variety of degrees of freedom within the same portfolio. In this case, we solve the optimization problem numerically, using Monte Carlo simulations. As a risk measure, we use value-at-risk (the results for expected shortfall would be very similar in this case). We find that long-term investors can tolerate more tail-risk than short term investors. I.e., the optimal weights of assets with fat left tails are small for short-term investors, and larger for longer-term investors.

While symmetric t-distributions capture the fat tails of real market return distributions, the analysis of sub-section 3.3.2 has at least three shortcomings:

a. It does not account for skewness

b. It does not account for tail correlations between the assets

c. It assumes that the returns at different days are independent of each other, thereby ignoring the autocorrelation of absolute returns described above

We therefore supplement the analysis of sub-section 3.3.2 by a second, more realistic analysis in sub-section 3.3.3. There, we use the multivariate student-t distribution with 3.5 degrees of freedom to model market returns, thereby resolving issues (a) and (b). We then use analytical methods to optimize expected shortfall.

As for issue (c), the autocorrelation of absolute returns indicates that aggregating daily returns to T-day returns does not lead to normal return distributions as quickly as expected by the central limit theorem. I.e., the number of deegrees of freedom of the t-distribution does not go to infinity as quickly as if the returns on different days were completely independent of each other. We do not have enough historical data to measure reliably how the number of degrees of freedom grows with the investment horizon. We therefore make the conservative assumption that the aggregation of daily returns to T-day returns does not change the number of degrees of freedom at all, but that it merely changes the dispersion.

Based on this, in subsection 3.3.3, we optimize expected shortfall for the same multivariate skewed t-distribution as for daily returns, but at different levels of confidence, with higher levels again corresponding to shorter investment horizons. Despite our conservative assumption, the results are again consistent with those of subsection 3.3.2: the optimal asset allocation for short-term investors gives smaller weights to assets with heavy left tails, and higher weights to assets with thin left tails.

3.3.2 Univariate symmetric Student's t-distribution

In a first attempt a range of assets were simulated using univariate symmetric Student's t-distribution with different degrees of freedom and allocated to the following investment horizons using value-at-risk as risk measure:

- daily returns
- weekly returns
- monthly returns
- yearly returns

For the process of asset allocation we used a specific value-at-risk that is dependent on the length of the investment horizon:

To define value-at-risk, we must specify two parameters: the time horizon τ and the level of confidence α. If α were independent of τ, the optimization results for different horizons would not be comparable. E.g., $\alpha = 99\%$ value-at-risk of yearly returns vs. daily returns measures, roughly, the worst annual loss expected in 100 years vs. the worst daily loss expected in 100 days $\approx$ 5 months. E.g., suppose we sell insurance against a major earthquake in Zurich against an insurance premium. Suppose such an earthquake occurs, on average, once in a hundred years and causes a loss of CHF 1 million within one day. In this case, 99% 1-year value-at-risk for this "investment" is CHF 1 million, but 99% 1-day value-at-risk is almost zero. To "see" the earthquake in the daily value-at-risk, we must consider it at a level of at least $1 - \frac{1}{26000}$ (assuming that there are 260 trading days per year). To cover the same time period (we use 30000 days) for each investment horizon τ , we therefore let α depend on τ as follows:

$$\mathrm{VaR}_{\alpha} = F^{-1}\left(1 - \frac{\tau}{26000}\right)$$

In Figure 12 it shown that the weight of the normally distributed asset shrinks when the investment horizon is increased. Said phenomenon corresponds with the knowledge obtained earlier, as the heavy tails feature diminishes when returns are aggregated. This means short-term investors can take less tail risk Thus, the weights are more evenly distributed throughout the range of assets. Although this first attempt illustrates the relationship between short term and long term risks it also has several drawbacks:

- Independence copula (no tail correlation)
- Symmetric distribution model (no skewness)

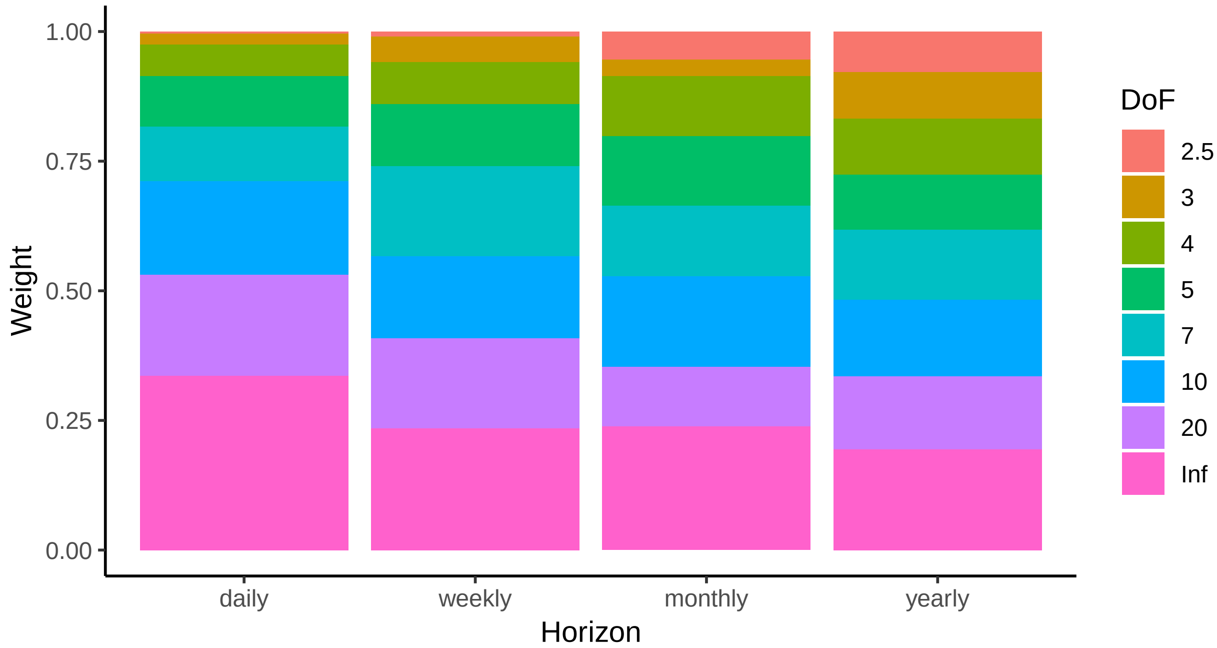

Figure 12: Weight allocation in a portfolio of Student's t-distributed assets

3.3.3 Multivariate skewed Student's t-distribution

The next step of the portfolio optimization process was conducted using an arbitrary multivariate skewed Student's t-distribution. In this stage, we make the following assumptions:

- The absolute log-returns are not autocorrelated
- The degrees of freedom of the single assets remain constant across all investment horizons

The following parameters were used for the multivariate model:

$$\nu = 3.5$$

$$\boldsymbol{\mu} = \begin{bmatrix} -0.0625 \\ -0.0125 \\ 0.0125 \\ -0.0375 \end{bmatrix}$$

$$\boldsymbol{\gamma} = 0.025 + 2\boldsymbol{\mu}$$

$$\boldsymbol{\gamma} = \begin{bmatrix} -0.1 \\ 0 \\ 0.05 \\ -0.05 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Where ν denotes the degrees of freedom parameters, $\boldsymbol{\gamma}$ the skewing parameter and $\boldsymbol{\mu}$, Σ the location-scale parameters. The parameters were chosen that $\boldsymbol{\gamma}$ is a linear combination from $\boldsymbol{\mu}$, which allows to model an efficient frontier that is independent of the risk measure. This approach has computing advantages as the portfolio optimization calculations can be done using the mean-variance method, which is less complex than optimizing value-at-risk or expected shortfall and does not require an iterative optimization algorithm e.g. Nelder-Mead.

Mathematical proof:

$$\boldsymbol{\gamma} = a + b\boldsymbol{\mu}$$

$$\overline{R} = \boldsymbol{w}^\top(\boldsymbol{\mu} + \boldsymbol{\gamma})$$

$$\boldsymbol{w}^\top\boldsymbol{\gamma} = a + b(\overline{R} - \boldsymbol{w}^\top\boldsymbol{\gamma})$$

$$\boldsymbol{w}^\top\boldsymbol{\gamma} = \frac{a + b\overline{R}}{1 + b}$$

Where $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$ are the location and skewness parameters of a constructed portfolio and $\overline{R}$ is the expected portfolio return. Thus, the portfolio parameters are only dependent upon three constants: the linear transformation constants a, b and the expected return. Therefore, the efficient frontier is the same for any risk measure.

In figure 13 it is shown that the optimal portfolio allocation for an investor that considers expected shortfall, changes significantly as the investment horizon increases. Especially equities are weighted differently as they are the riskiest of the bunch and thus can be tolerated more the longer the investment horizon. It is also demonstrated that volatility is not a suitable risk measure as there is no way to include the investment horizon in the calculations. However, the changes in the weights happen rather fast as the x-axis had to be log-scaled to highlight the impact of the investment duration.

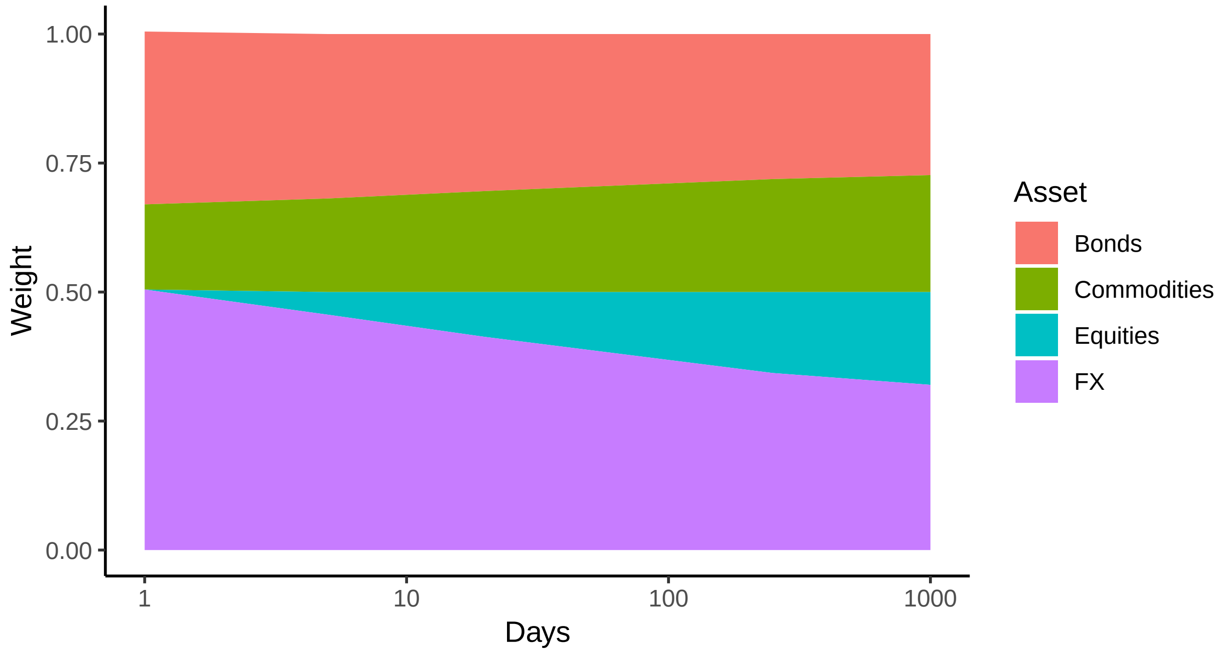

Figure 13: Graphical analysis of the portfolio optimization

4 Discussion

This study exposes several results. Firstly, the assumption of normality for returns of financial instruments is not valid, notably for short-term investment horizons. Therefore, especially for short time scales, portfolio management methodologies should put more emphasis on models based on heavy-tailed distributions.

Secondly, the skewness in the returns of various asset classes is confirmed and dealt with by introducing asymmetry in the respective distributions. Skewness represents an additional parameter to be considered by investors looking to optimize profits. Generally, hedgers will tend to prefer assets with positively skewed return distributions, while speculators will tend to prefer assets with negative skewness, if this is compensated for by additional return.

We found that the multivariate skewed Student's t-distribution is a good proxy for different asset classes as it incorporates heavy tails and skewness, and the related t copula provides the marginal distributions with the empirically observed tail correlation, all features the Gaussian distribution and its copula are unable to replicate. We find a range of 3-5 for the number of degrees of freedom of these distributions.

Lastly, results from the portfolio optimization show that the optimal asset allocation depends on the investor's investment horizon. Assets with heavy left tails are strongly under-weighted for short-term investors. On the other hand, long-term investors can allocate a greater weight of their portfolios to assets with heavier left tails, as they can sit out occasional sharp losses, provided that they are over-compensated in the long run by higher long-term returns.

As an implication, we find that long-term investors such as pension funds should not be too afraid of tail risk. Rather, they should be willing to leverage their long-term investment horizon to benefit from the long-term premium that is expected to accompany tail risk.

Appendix

Source Code

Data preparation

```
# @title: Source code for data preparation @author:
# joel.meili

# - load packages
library(dplyr)

# - prepare data
path <- "~/Dropbox/BA18/data/"
files <- c("DOW", "SP500", "10YUS_BOND_FUTURES", "JPY_USD", "GOLD")

data <- lapply(files, FUN = function(x) {
    read.csv(paste0(path, x, ".csv"), header = T, stringsAsFactors = F)
})

for (i in 1:length(data)) {
    data[[i]][, "Date"] <- as.Date(data[[i]][, "Date"], format = "%m/%d/%Y")
    data[[i]] <- data[[i]][order(data[[i]][, "Date"]), ]
    rownames(data[[i]]) <- NULL
    data[[i]][, files[i]] <- c(0, diff(log(data[[i]][, "Last.Price"])))
    data[[i]] <- data[[i]][-1, ]
    data[[i]] <- data[[i]][, c("Date", files[i])]
}

assets <- data[[1]]

for (i in 2:length(files)) {
    assets <- assets %>% inner_join(data[[i]], by = "Date")
}

save(assets, file = paste0(path, "log.return.assets.Rdata"))
```

Normality assumption

```
# @title: Source code for assessment of the normality
# assumption @author: joel.meili

# - load packages
library(ghyp)

# - contour plots of copulas in review section
x <- seq(-10, 10, by = 0.1)
y <- x
n <- gauss(mu = rep(0, 2), sigma = matrix(c(1, 0.5, 0.5, 1),
```

```r
    ncol = 2, byrow = TRUE))
t <- student.t(nu = 2.5, mu = rep(0, 2), sigma = matrix(c(1,
    0.5, 0.5, 1), ncol = 2, byrow = TRUE), gamma = rep(0, 2))
norm <- function(x, y) {
    dghyp(cbind(x, y), object = n, logvalue = TRUE) - dghyp(x,
        object = n[1], logvalue = TRUE) - dghyp(y, object = n[2],
        logvalue = TRUE)
}
tn <- function(x, y) {
    dghyp(cbind(x, y), object = t, logvalue = TRUE) - dghyp(x,
        object = t[1], logvalue = TRUE) - dghyp(y, object = t[2],
        logvalue = TRUE)
}
z.norm <- outer(x, y, norm)
z.t <- outer(x, y, tn)
par(mfrow = c(1, 2))
contour(x, y, z.norm, main = "Gaussian copula")
contour(x, y, z.t, main = "Student's t-copula")

# - quantile-quantile plots for daily S&P500 and JPY/USD
# log-returns
par(mfrow = c(1, 2))
qqPlot(assets$SP500 %>% scale(), ylab = "sample quantiles", main = "S&P500")
qqPlot(assets$JPY_USD %>% scale(), ylab = "sample quantiles",
    main = "JPY/USD")

# - correlation of log-returns
z <- round(cor(assets[, 2:ncol(assets)]), digits = 2)
colnames(z) <- c("DIJA", "S&P500", "10Y US Bond Futures", "JPY/USD",
    "Gold")
rownames(z) <- colnames(z)

# - volatility clustering
par(mfrow = c(1, 2))
acf(assets$SP500 %>% scale() %>% abs(), main = "")
z <- zoo(assets$SP500 %>% scale(), order.by = assets$Date)
plot(z, xlab = "Date", ylab = "S&P500")
abline(v = as.Date("2007-01-01"), col = 2)
abline(v = as.Date("2010-01-01"), col = 2)

# - GARCH modelling for volatiliy clustering
set.seed(100)
par(mfrow = c(1, 2))
garch.norm <- garchFit(data = rnorm(10000) %>% scale())
garch.sp500 <- garchFit(data = assets$SP500 %>% scale())
plot(zoo(garch.norm@sigma.t), ylab = "Estimated volatility")
plot(zoo(garch.sp500@sigma.t, order.by = assets$Date), xlab = "Time",
    ylab = "Estimated volatility")
abline(v = as.Date(c("2007-01-01", "2010-01-01")), col = 2)

# - tail correlation in daily log-returns
set.seed(100)
par(mfrow = c(1, 2))
```

```r
temp <- assets[, c(3, 4)] %>% scale()
plot(temp, xlim = c(-10, 10), ylim = c(-10, 10), ylab = "10Y US Bonds Futures")
abline(lm(temp[, 2] ~ temp[, 1]), col = 2)
lines(lowess(temp[, 2] ~ temp[, 1]), col = 3)
abline(h = c(3, -3), v = c(3, -3))
legend("bottomright", legend = c("linear regression", "local regression"),
    col = 2:3, lty = 1, cex = 0.5, bg = "white")
temp <- rghyp(10000, gauss(mu = rep(0, 2))) %>% scale()
plot(temp, xlim = c(-5, 5), ylim = c(-5, 5), xlab = "sample 1",
    ylab = "sample 2")
abline(lm(temp[, 2] ~ temp[, 1]), col = 2)
abline(h = c(3, -3), v = c(3, -3))
lines(lowess(temp[, 2] ~ temp[, 1]), col = 3)
legend("bottomright", legend = c("linear regression", "local regression"),
    col = 2:3, lty = 1, cex = 0.5, bg = "white")

# - quantile-quantile plots for yearly log-returns
assets.yearly <- assets %>% group_by(Year = year(Date)) %>% summarise(SP500 = sum(SP500),
    JPY_USD = sum(JPY_USD))
par(mfrow = c(1, 2))
qqPlot(assets.yearly$SP500 %>% scale(), ylab = "sample quantiles",
    main = "S&P500")
qqPlot(assets.yearly$JPY_USD %>% scale(), ylab = "sample quantiles",
    main = "JPY/USD")

# - tail correlation in yearly log-returns
temp <- assets.yearly[, c(2, 3)] %>% scale()
plot(temp, ylab = "10Y US Bonds Futures")
abline(lm(temp[, 2] ~ temp[, 1]), col = 2)
lines(lowess(temp[, 2] ~ temp[, 1]), col = 3)
legend("bottomright", legend = c("linear regression", "local regression"),
    col = 2:3, lty = 1, cex = 0.5, bg = "white")
```

Multivariate distribution fitting

```r
# @title: Source code for distribution fitting @author:
# joel.meili

# - load packages
library(dplyr)
library(ghyp)

# - load data
path <- "~/Dropbox/BA18/data/"
load(paste0(path, "log.return.assets.Rdata"))

# - demonstrating tail differences in Gaussian and Student's
# t-distribution
q <- seq(-10, 10, by = 0.01)
plot(q, dnorm(q, log = TRUE), type = "l", xlab = "x", ylab = "Log-density")
lines(q, dt(q, df = 3, ncp = 0, log = TRUE), col = 2)
```

```r
lines(q, dt(q, df = 10, ncp = 0, log = TRUE), col = 3)
lines(q, dt(q, df = 30, ncp = 0, log = TRUE), col = 4)
legend("bottom", legend = c("std. normal distribution", "Student's t-distribution: 3 df",
    "Student's t-distribution: 10 df", "Student's t-distribution: 30 df"),
    col = 1:4, lty = 1, cex = 0.5, bg = "white")

# - eye testing symmetric univariate Student's t-distribution
# for S&P500 and JPY/USD samples
par(mfrow = c(1, 2))
qqPlot(assets$SP500 %>% scale(), "t", df = 3, ylab = "sample quantiles",
    main = "S&P500")
qqPlot(assets$JPY_USD %>% scale(), "t", df = 4, ylab = "sample quantiles",
    main = "JPY/USD")

# - valdiating univariate skewed Student's t-distribution for
# S&P500 and JPY/USD samples
temp <- apply(assets[, 2:ncol(assets)], 2, scale)
temp <- apply(temp, 2, FUN = function(x) x[order(x)])
N <- nrow(temp)
fit.sp500 <- fit.tuv(temp[, 2], silent = TRUE)
fit.sp500.symm <- fit.tuv(temp[, 2], symmetric = TRUE, silent = TRUE)
fit.jpy <- fit.tuv(temp[, 4], silent = TRUE)
fit.jpy.symm <- fit.tuv(temp[, 4], symmetric = TRUE, silent = TRUE)
par(mfrow = c(1, 2))
plot(log(1:N/(N + 1)) ~ temp[, 2], xlim = c(min(temp[, 2]), 0),
    xlab = "S&P500", ylab = "log(P(X<x)")
lines(log(pghyp(temp[, 2], object = fit.sp500)) ~ temp[, 2],
    col = 2)
lines(log(pghyp(temp[, 2], object = fit.sp500.symm)) ~ temp[,
    2], col = 3)
lines(log(pnorm(temp[, 2], mean = mean(temp[, 2]), sd = sd(temp[,
    2]))) ~ temp[, 2], col = 4)
legend("topleft", legend = c("sample", "skewed t", "symmetric t",
    "normal"), col = 1:4, lty = 1, cex = 0.5)
plot(log((N + 1 - 1:N)/(N + 1)) ~ temp[, 2], xlim = c(0, max(temp[,
    2])), xlab = "S&P500", ylab = "log(1 - P(X<x))")
lines(log(1 - pghyp(temp[, 2], object = fit.sp500)) ~ temp[,
    2], col = 2)
lines(log(1 - pghyp(temp[, 2], object = fit.sp500.symm)) ~ temp[,
    2], col = 3)
lines(log(1 - pnorm(temp[, 2], mean = mean(temp[, 2]), sd = sd(temp[,
    2]))) ~ temp[, 2], col = 4)
legend("topright", legend = c("sample", "skewed t", "symmetric t",
    "normal"), col = 1:4, lty = 1, cex = 0.5)

# -
par(mfrow = c(1, 2))
plot(log(1:N/(N + 1)) ~ temp[, 4], xlim = c(min(temp[, 4]), 0),
    xlab = "JPY/USD", ylab = "log(P(X<x))")
lines(log(pghyp(temp[, 4], object = fit.jpy)) ~ temp[, 4], col = 2)
lines(log(pghyp(temp[, 4], object = fit.jpy.symm)) ~ temp[, 4],
    col = 3)
lines(log(pnorm(temp[, 4], mean = mean(temp[, 4]), sd = sd(temp[,
```

```
    4]))) ~ temp[, 4], col = 4)
legend("topleft", legend = c("sample", "skewed t", "symmetric t",
    "normal"), col = 1:4, lty = 1, cex = 0.5)
plot(log((N + 1 - 1:N)/(N + 1)) ~ temp[, 4], xlim = c(0, max(temp[,
    4])), xlab = "JPY/USD", ylab = "log(1 - P(X<x))")
lines(log(1 - pghyp(temp[, 4], object = fit.jpy)) ~ temp[, 4],
    col = 2)
lines(log(1 - pghyp(temp[, 4], object = fit.jpy.symm)) ~ temp[,
    4], col = 3)
lines(log(1 - pnorm(temp[, 4], mean = mean(temp[, 4]), sd = sd(temp[,
    4]))) ~ temp[, 4], col = 4)
legend("topright", legend = c("sample", "skewed t", "symmetric t",
    "normal"), col = 1:4, lty = 1, cex = 0.5)

# - log-likelihood-ratio test to validate whether symmetric
# Student's t-distribution would be sufficient
log.lik.sp500 <- lik.ratio.test(fit.sp500, fit.sp500.symm)
log.lik.jpy <- lik.ratio.test(fit.jpy, fit.jpy.symm)
log.lik.mat <- matrix(c(log.lik.sp500$p.value, log.lik.jpy$p.value),
    ncol = 2, byrow = TRUE)
rownames(log.lik.mat) <- "P-value"
colnames(log.lik.mat) <- c("S&P500", "JPY/USD")

# - value-at-risk backtesting and one-sided binomial test for
# S&P500 and JPY/USD estimates
alphas <- c(0.001, 0.005, 0.01)
viols <- matrix(NA, nrow = 3, ncol = length(alphas))
rownames(viols) <- c("Gaussian", "Symmetric Student's t", "Skewed Student's t")
colnames(viols) <- alphas
k <- 1
temp <- assets$SP500 %>% scale()
for (alpha in alphas) {
    viols.t <- temp < qghyp(alpha, object = fit.sp500)
    viols.symm <- temp < qghyp(alpha, object = fit.sp500.symm)
    viols.norm <- temp < qnorm(alpha)
    viols[1, k] <- sum(viols.norm)
    viols[2, k] <- sum(viols.symm)
    viols[3, k] <- sum(viols.t)
    k <- k + 1
}
viols <- as.data.frame(viols)
rownames(viols) <- c("Gaussian", "Symmetric Student's t", "Skewed Student's t")

p.values <- viols
for (i in 1:nrow(p.values)) {
    for (j in 1:ncol(p.values)) {
        p.values[i, j] <- binom.test(viols[i, j], nrow(assets),
            alphas[j], alternative = "greater")$p.value
    }
}
p.values <- round(p.values, digits = 3)
p.values <- as.data.frame(p.values) %>% mutate_all(~cell_spec(.x,
    bold = ifelse(.x < 0.05, TRUE, FALSE)))
```

```r
rownames(p.values) <- rownames(viols)

alphas <- c(0.001, 0.005, 0.01)
viols <- matrix(NA, nrow = 3, ncol = length(alphas))
rownames(viols) <- c("Gaussian", "Symmetric Student's t", "Skewed Student's t")
colnames(viols) <- 1 - alphas
k <- 1
temp <- assets$JPY_USD %>% scale()
for (alpha in alphas) {
    viols.t <- temp > qghyp(1 - alpha, object = fit.jpy)
    viols.symm <- temp > qghyp(1 - alpha, object = fit.jpy.symm)
    viols.norm <- temp > qnorm(1 - alpha)
    viols[1, k] <- sum(viols.norm)
    viols[2, k] <- sum(viols.symm)
    viols[3, k] <- sum(viols.t)
    k <- k + 1
}
viols <- as.data.frame(viols)

p.values <- viols
for (i in 1:nrow(p.values)) {
    for (j in 1:ncol(p.values)) {
        p.values[i, j] <- binom.test(viols[i, j], nrow(assets),
            alphas[j], alternative = "greater")$p.value
    }
}
p.values <- round(p.values, digits = 3)
p.values <- as.data.frame(p.values) %>% mutate_all(~cell_spec(.x,
    bold = ifelse(.x < 0.05, TRUE, FALSE)))
rownames(p.values) <- rownames(viols)

# fit univariate skewed Student's-t distribution to daily
# DIJA log-returns
uv.fit <- fit.tuv(assets$DOW %>% scale(), silent = T)

# - fit multivariate skewed Student's-t distribution to daily
# log-returns
assets.use <- assets[, 2:ncol(assets)] %>% mutate_each(scale)
mv.fit <- fit.tmv(assets.use, silent = T)

# - bootstrap estimated multivariate skewed Student's
# t-distribution parameters
R <- 100
boot.params <- matrix(NA, nrow = ncol(assets.use) + 1, ncol = R)
rownames(boot.params) <- c("nu", paste0("gamma", ".", colnames(assets.use)))

for (i in 1:R) {
    print(paste("Iteration", i, "of", R))
    idx <- sample(1:nrow(assets), replace = TRUE)
    mv.temp <- fit.tmv(assets.use[idx, ], silent = T)
    boot.params[1, i] <- mv.temp@lambda * -2
    boot.params[2:nrow(boot.params), i] <- mv.temp@gamma
}
```

```r
save(boot.params, file = paste0(path, "fitted.parameters.Rdata"))
```

Asset allocation

```r
# @title: Source code for asset allocation @author:
# joel.meili

# - load packages
library(ghyp)
library(dplyr)
library(tidyr)

# - create multivariate skewed Student's-t distribution
path <- "~/Dropbox/BA18/data/"
skew <- c(-0.1, 0, 0.05, -0.05)
mv <- student.t(nu = 3.5, mu = -2 * skew + 0.025, sigma = diag(4),
    gamma = skew)

# - asset allocation
skew <- c(-0.1, 0, 0.05, -0.05)
mv <- student.t(nu = 3.5, mu = (skew - 0.025)/2, gamma = skew)
horizons <- c(1, 5, 20, 250, 1000)
opt.weights <- sapply(horizons, FUN = function(x) portfolio.optimize(mv,
    risk.measure = "expected.shortfall", level = 1 - x/26000,
    distr = "return", silent = TRUE)$opt.weights)
opt.weights <- as.data.frame(t(opt.weights))
colnames(opt.weights) <- c("Equities", "Bonds", "FX", "Commodities")
opt.weights$Days <- horizons
opt.weights <- opt.weights %>% gather(Asset, Weight, 1:4)
opt.weights$Weight[opt.weights$Asset == "Equities" & opt.weights$Days ==
    1] <- 0
save(opt.weights, file = paste0(path, "opt.weights.Rdata"))
```

List of Figures

List of Tables

References

Alexander F. McNeil, Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Andrew Gelman, Hal S. Stern, John B. Carlin. 2014. *Bayesian Data Analysis*. CRC Press.

Asmussen, Søren. 2003. *Applied Probability and Queues*.

Demarta, Stefano, and Alexander J. McNeil. 2005. "The T Copula and Related Copulas." *International Statistical Review / Revue Internationale de Statistique* 73 (1). International Statistical Institute: 111–29.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambdridge University Press.

Kjersti Aas, Ingrid Hobæk Haff. 2006. "The Generalized Hyperbolic Skew Student's T-Distribution." *Journal of Financial Econometrics* 4 (2). Oxford University Press: 275–309.

Luenberger, David G. 1997. *Investment Science*. Oxford University Press.

Luethi, David, and Wolfgang Breymann. 2016. *Ghyp: A Package on Generalized Hyperbolic Distribution and Its Special Cases*. https://CRAN.R-project.org/package=ghyp.

Mandelbrot, Benoit. 1963. "The Variation of Certain Speculative Prices." *The Journal of Business* 36 (4). University of Chicago Press: 394–419. http://www.jstor.org/stable/2350970.

———. 1967. "The Variation of Some Other Speculative Prices." *The Journal of Business* 40 (4). University of Chicago Press: 393–413. http://www.jstor.org/stable/2351623.

Michel M. Dacorogna, Ulrich A. Müller, Ramazan Gençay. 2001. *An Introduction to High-Frequency Finance.* Academic Press.

"Minimum Capital Requirements for Market Risk." 2016. Bank for International Settlements.

Mohd Razali, Nornadiah, and Yap Bee Wah. 2011. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests." *Journal of Statistical Modeling and Analytics* 2 (1): 21–33.

Neath, Ronald C. 2013. "On Convergence Properties of the Montecarlo Em Algorithm." *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton* 10. Institute of Mathematical Statistics: 43–62.

Philippe Artzner, Jean-Marc Eber, Freddy Delbaen, and David Heath. 1998. "Coherent Measures of Risk." https://people.math.ethz.ch/~delbaen/ftp/preprints/CoherentMF.pdf.

Plackett, R. L. 1962. *Journal of the Royal Statistical Society. Series A (General)* 125 (2). [Royal Statistical Society, Wiley]: 284–86. http://www.jstor.org/stable/2982331.

Schuster, Eugene F. 1984. "Classification of Probability Laws by Tail Behavior." *Journal of the American Statistical Association* 79 (388). [American Statistical Association, Taylor & Francis, Ltd.]: 936–39. http://www.jstor.org/stable/2288727.

Stephens, M. A. 1974. "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association* 69 (347). American Statistical Association, Taylor & Francis, Ltd.: 730–37. http://www.jstor.org/stable/2286009.

Walck, Christian. 2007. *Hand-Book on Statistical Distribution for Experimentalists.* http://www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf.

Wildi, Marc. 2018. "Econometrics Iii: Conditional Heteroscedasticity Models."