Date of submission: 01.09.2016

BI Norwegian Business School – Thesis

Portfolio Optimization: On Risk Measures and Estimation Techniques

Written by Johannes Andreas Barstad,
MSc in Business and Economics (Major in Finance)
Olve Heitmann
MSc in Business and Economics (Major in Finance)

Supervised by Johann Reindl

Abstract

This thesis focuses on the risk measure in the Markowitz algorithm. We discuss why assuming normality is unrealistic, and why the unconditional sample covariance matrix is an inappropriate input for the algorithm. We compare the minimum variance portfolio of Markowitz to the minimum CVaR portfolio, and discuss how the use of GARCH and Copula models can improve upon the precision of the risk estimate. We compare these techniques in two real data applications. Our results suggest that GARCH, and Copula and GARCH in combination outperform the sample estimates if sample correlation is low, and that minimizing variance or CVaR gives very similar results.

Keywords: GARCH, Copula, Portfolio Optimization, Modern Portfolio Theory, Risk Measures, Coherent Risk Measures, Conditional Value-at-Risk, Risk Management.

Acknowledgements

We want to thank Johann Reindl for his guidance on finding the topic, and supervision during the process of writing the thesis.
We also want to thank family and friends for support.
The writing process has been challenging and highly educational, and we believe that the skillsets developed will be useful in the future.

Contents

List of Figures

List of Tables

Introduction

In 1952 Howard Markowitz pioneered the application of normative decision theory for constructing optimal portfolios with his Modern Portfolio Theory (MPT). The theory put the tradeoff between the expected return of the portfolio and the portfolios risk in the spotlight. In particular, MPT's message was that for any given level of expected portfolio return, the optimal portfolio were the one with the lowest variance, or equivalently, for any level of portfolio variance the optimal portfolio were the one with the highest expected return. Plotting these various combinations of portfolio variance (or standard deviation) and expected return, one can create what has become known as the "efficient frontier".

MPT's problems and research questions

While MPT's idea of an efficient frontier seem both sound and desirable, estimating the inputs for the algorithm (namely the expected return vector and the covariance matrix) is hard, and small deviations have big effects on the suggested portfolio weights. In practice, MPT is therefore often implemented along with a number of different "fixes", including shrinkage of the estimates, imposing a factor structure on the covariance matrix, estimating the expected return vector from an asset pricing model (e.g. the CAPM, Ross' APT, various index models) and constraining the portfolio weights. In this thesis we focus on the second input of the algorithm, i.e. the covariance matrix, or to put it in more general: the risk measure. We discuss the limits of standard deviation as a risk metric for a financial portfolio, what alternatives exist, and how one should go about estimating them. In particular, we discuss how assuming normality and i.i.d. doesn't hold up to the stylized facts of the empirical distribution of financial returns.

The research are centered around if it makes a difference if we optimize the portfolio for CVaR or portfolio variance, and if we are able better take into account these stylized facts using more advanced estimation procedures than the sample estimates. We compare minimizing portfolio variance to minimizing portfolio CVaR of an all US equity portfolio and a full commodity portfolio. We compare univariate models that take into account the leverage effect to those that

don’t, constant correlation to dynamic, and optimizing based on the assumption of normality versus assuming that returns follow the student t distribution. Since the focus is on risk measures, the natural evaluation criteria are those that focus on dispersion, accordingly we have chosen to look at the out of sample values of common risk metrics as sample variance, semi-variance, VaR and CVaR.

Thesis structure and writing style

While we aim to keep the style of writing on a level that should be understandable to readers who have taken basic statistics, some of the material that is covered is more advanced in nature, making this hard. To alleviate this, we have included a short list of some statistical and econometrical concepts and definitions in the start of Appendix A. This section does not aim at being exhaustive, and is not a must read per se, but might be used as a mini-encyclopedia while reading the thesis. In addition several excellent text books on the subject are listed in our bibliography at the end of the paper.

The rest of the paper is laid out as following in two main parts: Section 1-6 containing general theory on and related to risk measures for portfolio optimization, and Section 7-9 where said theory is applied and interpreted on real data. In particular; in section 1 we review the original Markowitz paper from 1952, and expand further on the problems of implementing it. Desirable properties of a risk measure are discussed in section 2, and the empirical distribution of financial asset returns are discussed in section 3. Section 4 reviews the risk measures Value-at-Risk (VaR) and the closely related Conditional-Value-at-Risk (CVaR), portfolio optimization with CVaR as a risk measure, and why it’s likely a good idea. In section 5 we review the GARCH framework for modeling volatility over time. Section 6 deals with copulas and measures of dependence between assets. In section 7 we describe the methodology used to conduct the research, while section 8 deals with the data used, and the results we get. Section 9 summarizes the results from section 8, discuss limitations and make suggestions for future work.

Section 1: Modern Portfolio Theory

This section reviews Howard Markowitz' pioneering paper on portfolio selection from 1952. We lay out his framework for constructing an optimal portfolio, the assumptions he relied on, and summarize the biggest critique of the paper.
The main references for this section are: Markowitz (1952, 1959)

1.1 Markowitz 1952

The foundation for modern portfolio theory as a mathematical and statistical problem was laid out by 1989 Nobel Prize laureate Howard Markowitz in 1952.

Markowitz argues that simply maximizing discounted, expected returns as an investment rule is rejected both as a hypothesis to explain historic investor behavior, and as a maxim to guide optimal investment. Instead he considers a rule that expected return is a desirable thing while variability of return is undesirable. The general principle here, risk aversion, had been known long before Markowitz. Daniel Bernoulli initiated this hypothesis in 1738 (Bodie et al [2014], 199), while Morgenstern and von Neumann 1944 showed that maximizing expected utility was rational behavior, consistent with Markowitz proposed rule. What however was new was the concept of efficient frontiers, i.e. the idea that for every level of expected portfolio return there is a portfolio that is the most efficient in terms of the lowest level of variability of return. If one then plots many efficient portfolios, the efficient frontier develops, and we get a graphical representation of the risk-reward tradeoff.

In his framework, Markowitz argues that the return of a security can be modelled as a random variable. Then, the return of a portfolio is a weighted average of the returns of the individual assets included in the portfolio, i.e. also a random variable.

$$R_P = \sum_{i=1}^{n} W_i R_i \qquad \textbf{(1)}$$

where

- R_i is the return of security i
- W_i is the relative weight allocated to security i in the portfolio, P.
- R_P is the return of the portfolio

Further Markowitz measures the variability of the return with the statistical concept variance. This implies a demand for diversification, not solely by reducing variance through increasing numbers of different securities held, but also choosing assets with low covariance.

1.1.1 The Efficient Frontier

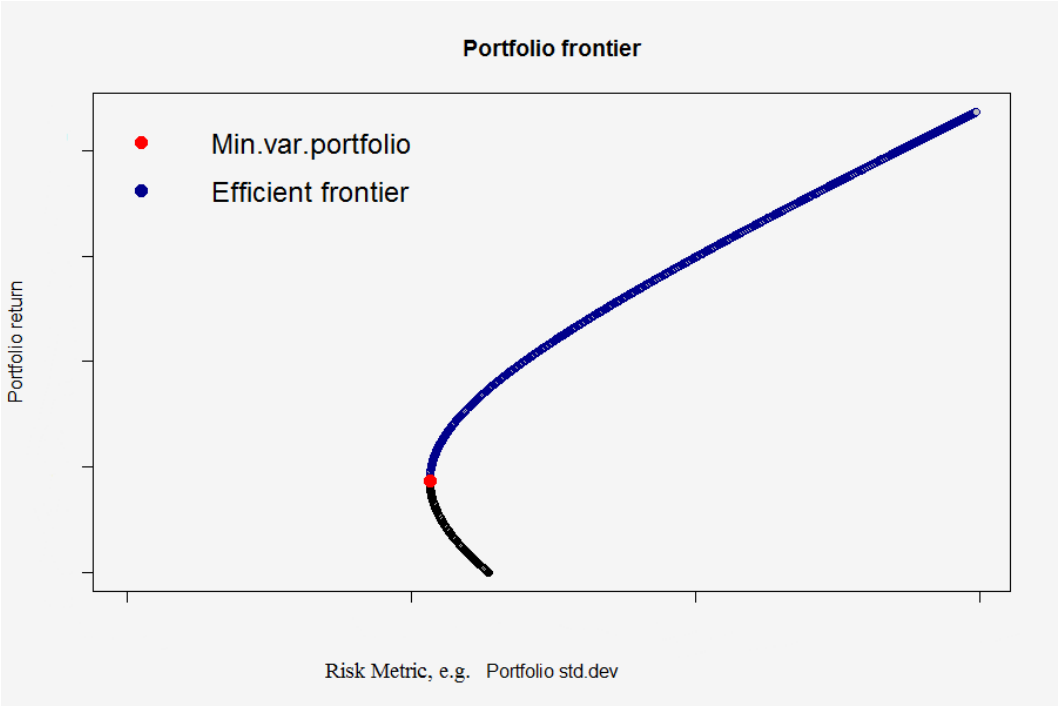


Figure 1 - The Efficient Frontier

The efficient frontier represents the set of portfolios that generate the highest expected portfolio return for a given level of risk, in Markowitz (1952) measured by the portfolios standard deviation. The concept of the efficient frontier was new in Markowitz (1952), and was a welcomed addition as it allows for a visual representation of the tradeoff between risk and return. . The mathematical mapping of the frontier is discussed in Appendix A3.

1.1.1 The Minimum Variance Portfolio

Of particular interest is the red point furthest to the left on the efficient frontier, i.e. the minimum variance portfolio. This portfolio is interesting as it is calculated without taking into account the expected return vector, i.e.:

$$Min\ \sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \sigma_{ij} \quad \textbf{(2)}$$

Subject to

$$\sum_{i=1}^{n} Wi = 1 \quad \textbf{(3)}$$

This makes the minimum variance portfolio a useful benchmark in theses like this where the subject we study is measuring risk. Instead of optimizing portfolios for some arbitrary utility function where parameters have to be estimated or assumed, we can rather focus on minimizing risk and compare portfolios by looking at measures of dispersion.

To see just how important the dependence structure of the assets is to forming this portfolio, it can be useful to look at the exact expression for the weight vector $\boldsymbol{W}$ that minimizes variance;

$$\boldsymbol{W_{MV}} = \frac{\boldsymbol{\Sigma^{-1}1}}{\boldsymbol{1'\Sigma^{-1}1}} \quad \textbf{(4)}$$

As we can see, both the denominator and the nominator depend solely on vectors of 1 and the covariance matrix. This expression is solved explicitly in Appendix A3.

1.2 Critique of Markowitz' assumptions

Markowitz argues that procedures for finding expected returns and correlations should involve statistical techniques and the judgment of practical knowledge. As portfolio optimization is a task with the future in mind, one needs to look at the forecasted expected return, as well as the forecasted covariance matrix. To estimate these, sample statistics are suggested, but also remarks that better methods can be found. In particular, Markowitz argues that investors need only use the first two moments of the probability distribution of returns to optimize their portfolio, implying that asset returns is assumed to follow the normal distribution. As we will see in the Section 3, this is an unrealistic assumption. In the real world, the distribution of financial returns tends to be left (negatively) skewed, leptokurtic and time-varying. Due to this, sample variance and covariance as measures of risk does not adequately capture the real distribution of returns, leading us to underestimating the potential losses, and the weights of the portfolio not offering the optimal risk-return tradeoff.

If we accept Markowitz' proposition of mean-dispersion per se being undesirable, as well as accepting the assumption of normality of financial returns and "stable probability beliefs"(Markowitz 1952, 4) to be reasonable, then there is simply not much room for improvement of the original algorithm. Both the sample mean and the sample covariance matrix should be reasonable inputs, and these are fairly simple to estimate. As we have already pointed out, the assumptions made on the distribution of financial returns are unrealistic. However, the notion that mean-dispersion is undesirable is also questionable. Variance as a risk measure equally punishes positive and negative deviations from the mean, while for risk averse investors the emphasis should be placed on the risk of large losses. Measures such as value at risk (VaR) and expected shortfall/conditional value at risk (CVaR) better takes this into account. In his book from 1959, Markowitz proposes semi-variance as an alternative to variance, where semi variance is defined as:

$$\Sigma_p^2 = E\left(\left[R_p - \mu_p\right]^{-2}\right) = E(\left[Min\left(R_p - \mu_p, 0\right)\right]^2) \quad \textbf{(5)}$$

Here only negative deviations are considered, making it an improvement over regular variance.

Section 2: On the properties of Risk Measures

This section reviews what the literature deems desirable properties of a risk measure. The main references for this section are: Alexander (2008c), Artzner et al (1999) and Krause (2002).

2.1 What is risk?

Most people have an intuition about what risk is. However, to define more precisely what risk is, it can be useful trying to understand the absence of risk. A risk-free asset is an asset which has a certain future value. If such an asset truly exists is debatable (often government bonds are used as a proxy for it). A risky asset must then be an asset whose future value is uncertain, and thus risk is clearly related to uncertainty. However, risk is normally not thought of simply as uncertainty in and of itself, but rather the probability and magnitude of outcomes that leaves us in a worse than expected or "average" state.

While there are many types of risk, we focus on *market risk*, i.e. the risk of unexpected changes in prices. Modeling this risk is important because investors require a premium to take on risk (i.e. they are risk averse). This thesis covers the statistical approach to modeling risk in the context of portfolio optimization.

2.2 Coherent risk measures

To manage the risk of a portfolio, we need appropriate risk measures. Artzner et al. (1999) presents the concept of *coherent risk measures*, arguing that any risk measure to be used to effectively regulate or manage risk should follow a set of axioms, making them *coherent*.

Let X and Y be two risky assets and $\rho(\cdot)$ the risk measure we are studying. We also assume that we have access to a riskless asset providing a fixed outcome of $RF > 1$ for each unit invested. The amount invested in the risk free asset is denoted by k.

We can then compare two investment strategies;

1. Invest in one of the risky assets and the risk free asset
2. Invest only in the risky asset (same asset as in strategy 1).

The riskless asset provides a certain profit, and thus reduces the potential losses arising from the risky asset by exactly the amount invested into it (i.e. *k*). This should be reflected in a risk measure and leaves us with the first axiom:

Translation invariance (Axiom T):

$$\rho(X + kRF) = \rho(X) - k$$

When adding quantity k to the risk free asset, the risk is reduced by the same amount.

Secondly, we know that risk can be reduced by diversification, and a risk measure should reflect that;

Sub-additivity (Axiom S):

$$\rho(X + Y) \leq \rho(X) + \rho(Y)$$

The risk of the combined portfolio is less than or equal to the sum of the risk of the individual assets.

Thirdly, the risk measure speaks only of uncertainty regarding the object at study, not of the risk attitude of the investor. Thus a risk metric should reflect that our risk is proportional to our bet;

Positive homogeneity (Axiom PH):

$$\rho(kX) = k\rho(X)$$

By increasing the amount invested in the asset by factor k, the measured risk is increased by the same factor.

Lastly, risk is typically thought of as bad deviations from our expectations, and not positive. If one investment stochastically dominates another, it's not intuitive for the dominating investment to have a higher measured risk (This is discussed in detail in Yamai and Yoshiba [2002]). ;

Monotonicity (Axiom M)

For all X and Y with $X \geq Y,\ \rho(X) \leq \rho(Y)$

That is, if X (weakly) stochastically dominates Y, then X should be judges as less or equally risky as Y.

2.2.1 Potential consequences of using an incoherent risk measure

Portfolios should be managed relating to their risk measure, i.e. the measure should have an impact on decisions, it shouldn't simply be something we monitor and then "leave it at that". If the risk metric we manage our portfolio under for instance breaks the sub-additivity axiom, we will heavily undervalue the benefits of diversification, one of the biggest advantages to having portfolios rather than single assets in the first place. In fact, if the risk measure both exhibits sub-additivity and positive homogeneity, then the risk measure must also exhibit convexity. This is desirable as it ensures that if there exists a local minimum, it must also be the global minimum, making mathematical optimization a much more powerful tool (This is discussed in further detail in Rockafellar and Uryasev [2000]). On the contrary, if our risk measure doesn't exhibit both sub-additivity and positive homogeneity, we run into the risk of possibly believing we have optimized our portfolio, while in reality we have only found one of many local minimums. Similarly, if our risk measure doesn't comply with monotonicity, for instance if we manage our portfolio by standard deviation, we could lose out on "free lunches" of stochastically dominant assets (how realistic free lunches are is another debate, but the point remains true).

2.2.2 On the coherency of volatility

The standard deviation (or volatility) of investment outcomes has been a common risk measure since Markowitz 1952. From the section above we see that standard deviation is not a coherent risk measure, as it fails both axiom T and axiom M. Only measures expressed in absolute terms can fulfill axiom T, while volatility is typically expressed in relative terms. More problematic is perhaps axiom M failing. Consider two normally distributed independent investments. Investment 1 has an expected return of 10%, and a standard deviation of 5%. Investment 2 has an expected return of 1%, and a standard deviation of 4%. Ranking the riskiness of the investments based on standard deviation alone implies that investment 2 is less risky than 1 – clearly this doesn't make sense.

Section 3: The Stylized facts

The stylized empirical facts of financial time series is a collection of empirical observations and inferences of statistical properties common across a wide range of instruments, markets and time periods.
The main references for this section are: Black (1976), Christie (1982), Cont (2001), Erb (1994), Goetzmann (2005), Ledoit et al (2003), Longin and Solnik (1995), McNeil et al (2005) and Mandelbrot (1963).

The stylized facts are potentially very useful in determining how we should model financial risk, as they are able to give us guidelines to which properties our models should exhibit. Cont (2001) argues that in order to let the data speak for itself as much as possible the facts should be formulated as qualitative assumptions, which statistical models then can be fitted to, rather than assume that the data belongs to any pre specified parametric family.

Much has been written on stylized facts, and the exact formulation of each stylized fact varies from author to author, and some authors include "facts" that others don't. With that being said, most of the stylized facts are reoccurring in most of the literature on the topic. The list we present are based on what seems to be the most reoccurring facts, with formulations inspired by McNeil et al (2005) and Cont (2001).

(1) Linear autocorrelations of asset returns are often insignificant. The exception to this is typically for small intraday timescales (~ 20 minutes) for which market microstructure effects enter (Cont 2001). This stylized fact is often cited as support for weak market efficiency, as significant autocorrelations would imply that previous returns could be used to predict future returns (and thus "statistical arbitrage"

(2) Volatility appears to cluster and vary over time, and in a somewhat predictable manner. It is observed that series of absolute or squared returns show profound serial correlation. A commonly used metric to measure volatility clustering is the autocorrelation function of the squared returns;

$$AC_2 = corr(|r_{t+\tau,\Delta t}|^2,\ |r_{t,\Delta t}|^2) \quad \textbf{(6)}$$

(3) Asymmetric relationship between gains and losses. One can typically observe large drawdowns in stock prices but not equally large upward movements. As the mean return tends to be positive or close to zero, this implies a skewed distribution.

(4) Slow decay of autocorrelation in absolute returns. The autocorrelation function of absolute returns decays slowly, often modelled by a power law with exponent $\beta \in [0.2, 0.4]$ (Cont 2001). This can be interpreted as a sign of long-range dependence.

(5) Return series are leptokurtic, i.e. heavy-tailed. The unconditional distribution of returns has excess kurtosis relative to that off the normal distribution (> 3). This effect is often still present even after correcting returns (e.g. via GARCH-type models), but reduced compared to that of the unconditional distribution.

(6) Leverage effects. Most measures of volatility of an asset are negatively correlated with the returns of that asset, e.g. $L_\tau = corr(|r_{t+\tau,\Delta t}|^2, r_{t,\Delta t})$ start from a negative value and decays to zero, suggesting that negative returns leads to increased volatility.
Black (1976) suggested that this could be attributed to the fact that bad news drives down the stock price, increasing the debt to equity ratio (i.e. the leverage) and thus causing the stock to be more volatile(risky).

In addition to these stylized facts (that reach across different asset classes and financial instruments) a lot of research has also been done on the individual asset classes. Of particular relevance to this thesis is the research stream surrounding the dynamic nature of equity correlations. Goetzmann et al. found that correlations between equity returns vary substantially over time, and peak during periods of highly integrated financial markets (as one would expect). Longin and Solnik found evidence for rejection of the hypothesis of constant correlations among international stock markets, while Ledoit et al. and Erb et al. show time-varying (dynamic) correlations tend to be higher during periods of recession. The latter observation is particularly interesting (or worrisome) as it would imply that if we model financial risk in “normal” or “boom” periods, our correlations would

be understated and should a recession come, our risk measures would be understated in the time we needed them the most.

Even if stylized facts can be a useful tool, the gain in generality across financial instruments, markets and time do come at the cost of precision of the statements that can be made about asset returns (this of course holds true in general to statistical models). Nevertheless, these stylized facts present properties that are regarded very constraining for a model to exhibit, even as an ad hoc stochastic process (Cont 2001). A question which should be noted in this regard is whether a stylized fact is relevant for the economic task at hand. If deemed not, it should not be a constraint to the model we are seeking either.

3.1 The normal distribution, i.i.d. assumption and the stylized facts

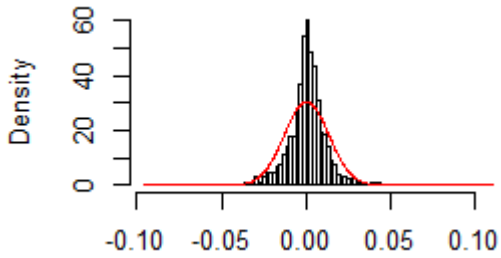


Figure 2 - The S&P 500 vs the normal distribution

The inappropriateness of modeling the marginal distribution of asset returns with the normal distribution was pointed out as early as 1963 by B. Mandelbrot. The properties of the normal distribution simply doesn't reconcile with the stylized facts.

We can characterize the needs for a parametric model to be able to successfully reproduce the observed empirical features with it having at least four parameters; a location parameter (e.g. mean), a scale parameter (e.g. standard deviation), a parameter describing the tail decay and eventually an asymmetry parameter allowing different behavior in each of the tails. The normal distribution only meets two of these requirements.

3.2 The random walk hypothesis

The insignificance of autocorrelations in return gave support for the random walk hypothesis of prices, where returns are considered independent random variables. However, the lack of linear dependence doesn't imply independence: one also have to consider nonlinear functions of return. As we know from stylized fact 2, there is nonlinear dependence (which is exactly what is drawn from in order to create GARCH models for one). Log prices are therefore not properly modelled by random walks. Traditional tools of signal processing such as ARMA modeling and autocovariance analysis, can't distinguish between asset returns and white noise. This points out the need for nonlinear dependence measures (e.g. GARCH modeling) to properly measure the dependence of asset returns (Cont 2001).

3.3 Assuming elliptical distributions in general

Much has been written on the validity of assuming that financial asset returns follow a normal distribution. A highly related and interesting discussion is that of the validity of assuming financial returns follow elliptical distributions in general. For instance Owen and Rabinovitch (1983) take the position that non-normal elliptical distributions such as the student t can be useful as it allows for describing tail decay through the degrees of freedom parameter, despite that the asymmetry parameter is still lacking (While this paper was written before high impact statistical methods such as ARCH rose to popularity in finance, we are of the opinion that the arguments are still valid). Chicheportiche and Bouchaud (2012) argue that elliptical distributions might be a fair assumption when assets are highly correlated, but also argue that it is very unrealistic when correlations are low. From our point of view, the only real consensus seems to be that these assumptions need to be assessed case-by-case. We'd also like to note that while some form of asymmetry parameter probably is desirable for most financial asset classes, it is absolutely necessary when modeling a joint distribution including non-linear assets such as options, almost regardless of the underlying asset. These assets are not considered in the real data application of this thesis.

Section 4: Tail-based Risk Measures

Tail-based risk measures focus on the returns that fall within a certain quantile of the return distribution. In this section we review quantile risk measures at the portfolio level, without mapping out the portfolios individual risk factors.
The main references for this section are: Alexander (2008c), McNeil et al (2005), Rockafellar and Uryasev (2000, 2002) and Yamai and Yoshiba (2002).

4.1 Value at Risk (VaR)

The α-Value at Risk is defined as the maximum expected level of loss l given a confidence level of α, over a specified risk horizon h. Defining the significance level as $1-\alpha$, we can also say that the significance level states the probability of losses at or exceeding the Value at Risk for the specified risk horizon. The term became widely used in finance in the mid 1990's with JPMorgan's publishing of RiskMetrics. As we don't know the future, we can only get a number for the Value at Risk if we make some assumptions about the underlying probability distribution. Thus Value at Risk is sometimes defined in a less assertive manner, i.e. the potential loss under "normal conditions" (implying that the assumptions made holds under the normal conditions).

To progress further we need a way to define a loss. In some works on Value at Risk and related measures (e.g. Alexander [2008c], McNeil et al [2005, 38]), the loss is defined as the amount one fall short of some benchmark, or "mean-VaR". For instance, if the benchmark is the S&P 500, and the S&P 500 has a return of 10% while our portfolio only has a return of 5%, the loss relative to the benchmark would be $10\% - 5\% = 5\%$. In this thesis we take the more conventional approach of viewing the loss function of a portfolio as the negative of the return of the portfolio:

$$L(w,r) = -(w_1 r_1 + \cdots + w_n r_n) = -w^T r \tag{7}$$

Where r is the returns of the assets, and w is the portfolio weights. The probability of the loss $L(w,r)$ not exceeding l is defined as:

$$F(w,l) = \int_{L(w,r) \leq l} p(r)dr \quad \textbf{(8)}$$

Where $p(r)$ is the joint density function of returns and $F(w,l)$ is the cumulative distribution function for losses. The Value at Risk is then given by:

$$VaR_\alpha(L) = l_\alpha(w) = \min(l: F(w,l) \geq \alpha) \quad \textbf{(9)}$$

Here $l_\alpha(w)$ is the VaR and the left endpoint of the interval so that $F(w,l) = \alpha$. This follows as $F(w,l)$ is continuous per assumption and non-decreasing with respect to l.

The risk horizon h should ideally reflect the period over which we are committed to holding the period. The length of this period is affected by contractual and legal constraints, and liquidity considerations. The latter imply that it will likely vary across markets, and the investment policy of the institution holding the portfolio.

4.1.1 Estimating Value at Risk

While the concept of Value at Risk is intuitive, obtaining a good estimate of Value at Risk isn't easy. The approaches for estimating Value at Risk is typically divided into three categories:

1. Analytical computation by making assumptions about the return distributions.
2. Estimates based on the histogram of past returns.
3. Estimates based on simulation techniques.

All of these approaches can have merit depending on which assets are modeled and what the use is. It is however important to keep in mind the stylized facts while making estimates. For instance, usually past returns are used to some extent in the estimating process, and these past returns are likely to exhibit ARCH effects. If one then makes the naïve assumption of equally weighting each day of the historic sample without taking into account these ARCH effects (e.g. with a GARCH model), then the estimate is more likely to be higher (if volatility is trending down towards the end of the historic sample) or lower (if volatility is

trending upwards towards the end of the historic sample) than the actual risk taken. Hence, all three of these approaches could potentially benefit from econometric techniques. We elaborate further on the three approaches in Appendix A4.

4.1.2 Pros and Cons of Value at Risk

Unfortunately, Value at Risk has some large drawbacks.

Firstly, it does not measure losses exceeding the VaR, giving us no real information about the possible consequences when things *really* don't go our way.

Secondly, VaR is often given a very literal interpretation, which can be misleading and dangerous. The estimate of the loss distribution is subject to estimation error and model risk (misspecified models or unrealistic assumptions). Additionally, the literal interpretation of VaR neglects any issues related to liquidity. If trades have a large impact on prices, or we are simply not able to trade due to no counterparty being willing to take the other side of the transaction, the literate interpretation of VaR becomes inaccurate. This latter problem was brought to the attention finance academics by Lawrence and Robinson (1995), and is discussed in detail there.

Thirdly, as proved by Artzner et al (1999), the measure is incoherent as it does not fulfill the non-subadditivity axiom, meaning that it is not by itself support diversification. Good, practical examples demonstrating this can be found in e.g. McNeil et al (2005) [example 6.7 page 241, 6.12 page 246]. Further, this implies that the convexity of Value at Risk is not guaranteed, making it potentially difficult to optimize as it allows for multiple local extrema.

Even if Value-at-Risk isn't perfect, it has been (and is) widely used in practice due to some desirable features. Value-at-Risk focuses on potential losses and not simply variability in returns, and is thus true to the intuitive perception of risk. Secondly, when explained it is easily interpretable for people without a background in finance or statistics, and thus can be useful when dealing with clients.

Luckily, there exists a measure that maintains the positives of Value at Risk, while avoiding some of the negatives: Conditional Value at Risk. This measure is discussed in more detail in the following subsection.

4.2 Conditional Value at Risk (CVaR)

While Value at Risk concerns itself with what the maximum loss over risk horizon h given a confidence level of α is, Conditional Value at Risk measures the expected loss the worst $(1-\alpha)100\%$ of times. That is;

$$\alpha\text{-}CVaR = E[L(w,r)|L(w,r) > \alpha\text{-}VaR] \quad \textbf{(10)}$$

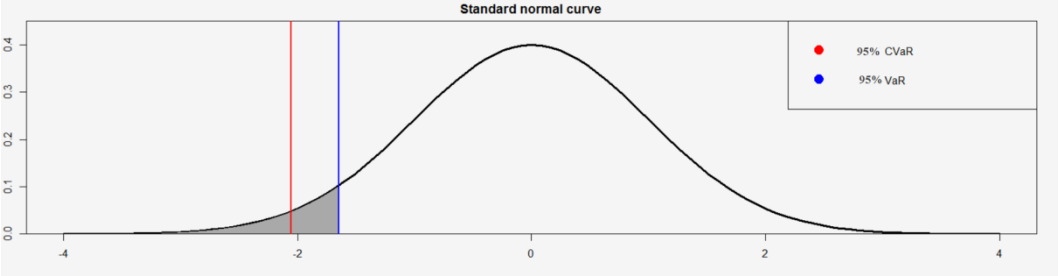


Figure 3 - Graphical representation of CVaR and VaR

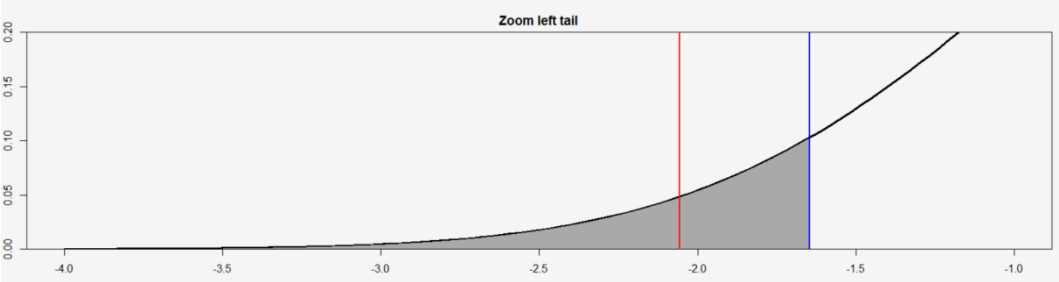


Figure 4- Graphical representation of CVaR and VaR, tail zoom

Acerbi and Tasche (2002a, 2002b) prove that CVaR is a coherent risk measure, while Yamai and Yoshiba (2002) show that CVaR is consistent with maximizing expected utility under way more general conditions (utility functions) than the standard deviation (requiring a quadratic utility function) or even the unconditional Value at Risk. Additionally, CVaR as a risk measure shares the desirable properties of VaR in focusing on potential losses and being interpretable to the average investor.

However, as the definition of CVaR ensure that CVaR for any given level of confidence and risk horizon is higher than the corresponding VaR, minimizing CVaR ensures a low VaR as well.

4.2.1 Mean-CVaR optimization

In 2000 Rockafellar and Uryasev presented the approach of minimizing Conditional Value-at-Risk (CVaR) rather than minimizing the potentially non-convex and non-coherent Value-at-Risk. The new approach can be combined with analytical or simulation-based methods to optimize portfolios.

Following the notation from the Value at Risk section, Rockafeller and Uryasev give the following expressions for α-VaR and α-CVaR;

$$\alpha\text{-VaR}(w) = l_\alpha(w) = \min\{l \in \mathbb{R}: F(w,r) \geq \alpha\} \quad \textbf{(11)}$$

$$\alpha\text{-CVaR}(w) = \phi_\alpha(w) = (1-\alpha)^{-1} \int_{L(w,r)\geq l_\alpha(w)} L(w,r)p(r)dr \quad \textbf{(12)}$$

Where $P[L(w,r) \geq l_\alpha(w)] = 1 - \alpha$ by the definition of VaR.

Rockafellar and Uraysev then define the following function F_α on $W \times \mathbb{R}$

$$F_\alpha(w,l) = l + (1-\alpha)^{-1} \int_{r\epsilon R^n} (-w^T r - l)^+ p(r)dr \quad \textbf{(13)}$$

$$= l + (1-\alpha)^{-1} \int_{r\epsilon R^n} [L(w,r) - l]^+ p(r)dr \quad \textbf{(14)}$$

Rockafellar and Uryasev (2002) prove that F_α is convex and continuously differentiable, and that the α-CVaR then can be computed by

$$\phi_\alpha(w) = \min_{\alpha \in R} F_\alpha(w,l) \quad \textbf{(15)}$$

The main contribution of Rockafellar and Uryasev (2000) is exactly this, continuously differentiable convex function are easy to minimize numerically. Additionally, α-CVaR can be calculated without first calculating α-VaR, which we know is troublesome from section 4.1.2.

As the joint distribution (and thus analytical expression for $p[r]$) is not known, we instead have to use either historical values of returns or simulated returns.

Rockafellar and Uryasev (2000) proposes to approximate this integral by sampling the probability distribution of r according to density $p(r)$. If we take the approach of simulated returns (scenarios), then the corresponding approximation to $F_\alpha(w,l)$ is

$$\hat{F}_\alpha(w,l) = l + (S - S\alpha)^{-1} \sum_{s=1}^{S} [L(w,r_s) - l]^+ \quad \textbf{(16)}$$

Minimization of $\hat{F}_\alpha(w,l)$ can then be reduced to convex programming by replacing $[L(w,r_s) - l]^+$ with the dummy variables Z_s for $= 1, \dots, r$:

And then minimizing the linear expression

$$l + (S - S\alpha)^{-1} \sum_{s=1}^{S} Z_s \quad \textbf{(17)}$$

Subject to

1. $w_i \geq 0$ for $i = 1, \dots, n$ with $\sum_{i=1}^{n} w_i = 1$
2. $w^T E(r) \geq R$ if we want to guarantee a certain return, else this constraint is dropped
3. $Z_s \geq 0$, for $s = 1, \dots, r$
4. $w^T r + l + Z_s \geq 0$ for $s = 1, \dots, r$

Where

- $s = 1, \dots, S$ denotes simulated scenarios
- r_s denotes the simulated returns vector for scenario s.

This expression is also convex and is piecewise linear with respect to l. It is not differentiable with respect to l, but more importantly it can be minimized through linear programming.

4.2.1.1 Mean-CVaR optimization vs. Mean-Variance optimization

Rockafeller and Uryasev (2000) show that Mean-Variance optimization and Mean-CVaR optimization generates equivalent efficient frontiers if the loss function follows the normal distribution. More interestingly, they show that differences arise when the loss distribution is non-normal, i.e. when it exhibits fat tails and asymmetry. Krokhmal et al (2002) shows that Mean-CVaR optimization yields a higher standard deviation than Mean-Variance optimization for a given level of return, with differences increasing with the confidence level $1 - \alpha$. Both these observations are in thread with what one would expect based on the stylized facts, i.e. negative skew and fat tails.

Section 5: Autoregressive Conditional Heteroscedasticity

The ARCH model is the foundation for the popular GARCH models. We briefly review the original ARCH model from 1982 in Appendix A5.

The rest of this section will discuss using the GARCH-framework to model the empirical distribution of financial returns in order to get good estimates of risk. The main references for this section are: Bollerslev (1986, 1990), Brooks (2008), Engle (1982, 2002), Francq and Zakoïan (2014), Nelson (1991), Silvennoinen and Teräsvirta (2009) and Zivot (2009).

5.1 Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

We will start off by looking at the univariate GARCH model. The GARCH model originated as an extension of Engle's (1982) ARCH model, and was first introduced by Bollerslev in 1986 to "allow for both a longer memory and more flexible lag structure" (Bollerslev 1986, page 2). Brooks (2008, p. 393) argues that this makes the model more parsimonious than the ARCH model, as well as helping to avoid overfitting. While in the regular ARCH model, the decay rate of the unconditional autocorrelation of ϵ_t is too rapid for what is typically observed in finance unless you include many lags, the GARCH model allows for a slower, albeit still exponential decay. This in turn makes the model less likely to breach non-negativity constraints, compared to that of the ARCH model.

To illustrate for our application; let r_t be a stochastic process of daily log returns, then a univariate symmetric normal GARCH(p,q) process is formally defined as

$$r_t = \mu_t + \epsilon_t \tag{18}$$

$$\epsilon_t = {h_t}^{1/2} z_t \tag{19}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \tag{20}$$

Where

- r_t is the log return of a security at time *t*
- μ_t is the expected value of the conditional log return of a security at time *t*
- ϵ_t is the mean corrected return of a security at time *t*
- h_t is the square of the conditional volatility, i.e. the conditional variance at time *t*
- h_t is the conditional volatility at time *t*
- z_t is a sequence of standardized i.i.d. random variables
- $\alpha_i \forall$ i are parameters of the model
- $\beta_j \forall$ i are parameters of the model

As with the ARCH model, the parameters of the model can be constrained to be positive to ensure positive and finite conditional variance. Some have argued that one should avoid imposing constraints such as these on the parameter estimation routine, as if such constraints indeed are necessary it is indicative of the specification being improper (Alexander 2008b, 136).

It is important to note that the GARCH(*p*,*q*) process is weakly stationary if and only if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \qquad \textbf{(21)}$$

5.1.1 Conditional Mean specification

The expected value of the conditional log return (μ_t) (i.e. the conditional mean) is sometimes modelled as an ARMA-process, but also commonly modelled simply as a constant. The ARMA-process modeling gets used in order to capture autocorrelation caused by market microstructure effects such as the bid-ask bounce(the phenomenon that transaction prices may take place either close to the ask or close to the bid price and then tend to bounce between these two prices), or non-trading effects (Zivot 2009). If extreme market events happened during the sample period, one can also opt to remove these effects with dummy variables, but caution and a thorough understanding of why the events were extreme

(unusual, unlikely to happen again and thus "noise") are needed. When no ARMA effects are found we often default to the constant. Putting all this into an equation, the typical conditional mean specification takes the form of:

$$E_{t-1}[y_t] = c + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \sum_{l=0}^{L} \beta_l' X_{t-l} + \epsilon_t \quad \textbf{(22)}$$

Where X_t is a $k \times 1$ vector of exogenous explanatory variables, and the rest of the variables are an intercept and conventional ARMA parameters reviewed in the appendix.

5.1.2 The unconditional volatility of GARCH models

If there were no market shocks, the GARCH variance h_t would eventually settle down to a steady state value, $\bar{h}$ so that $h_t = \bar{h}$ for all *t*. This is called the unconditional variance of the GARCH model, and is not to be confused for the unconditional variance in a moving average volatility (which is based on the i.i.d. returns assumption). This steady state value varies based on the GARCH-specification we choose.

E.g. for a "vanilla GARCH (1,1)" (symmetric normal) we can calculate the value by substituting $h_t = h_{t-1} = \bar{h}$, and then use the fact that $\mathbb{E}[{\epsilon_{t-1}}^2] = h_{t-1} = \bar{h}$ to finally obtain $\bar{h} = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$

For the general GARCH(*p*,*q*) model, we get

$$\bar{h} = \frac{\alpha_0}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j)} \quad \textbf{(23)}$$

Consensus on the GARCH model is that it should be useful for forecasting short-term (and perhaps medium-term) volatility, but the long term volatility is more questionable (Alexander 2008b, 144) . This can be somewhat alleviated by constraining the value of the constant in the GARCH estimation ($\propto_0$) to some

level we deem realistic for long term variance, and only let the remaining parameters be estimated by the data.

5.1.3 The Exponential GARCH

A critique of the GARCH (1,1) model is that it assumes the response of the conditional variance to negative shocks to be exactly the same as the response of a positive shock of the same magnitude. Since we know that leverage effects are present in most financial data, this assumption is not very good. The asymmetric volatility response can be captured by assuming a skewed error distribution, or altering the volatility equation. Nelson (1991) proposed the following E-GARCH model as a solution of the latter type;

$$\ln(\sigma_t^2) = h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j h_{t-j} \qquad \textbf{(24)}$$

We demonstrate how this specification can capture asymmetric responses in the appendix, and discuss some other properties of the specification.

5.1.4 Non-Gaussian Error Distributions

The original presentations of the ARCH and GARCH models assumed a normal error distribution. However, based on the stylized facts of financial time series, it's intuitive to use an error distribution that can capture fat tails better than the normal distribution. Many fat-tailed error distributions have been proposed, with Bollerslev (1987)'s proposal of the Student's t distribution being among the most popular. The Student's t density has a symmetric bell shape similar to that of the normal distribution. The density function of the general t distribution is given by

$$f_v(z) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi s v}}\left(1 + \frac{z^2}{sv}\right)^{-\left(\frac{v+1}{2}\right)} \qquad \textbf{(25)}$$

Where $\Gamma(\cdot)$ is the gamma function, v the degree of freedoms controlling the thickness of the tails and s the scale parameter (Alexander 2008a, 97-98).

The first two moments of the general t distribution are given by

$$E(Z) = 0, v > 0 \qquad \textbf{(26)}$$

$$Var(Z) = \frac{sv}{v-2}, \quad v > 2 \tag{27}$$

As z_t in the GARCH model is standardized (variance = 1), the scale parameter must have the value $\frac{v-2}{v}$ to ensure unit variance.

5.1.5 Estimation of the parameters

Estimation of the parameters in the GARCH model is usually done using Maximum Likelihood Estimation (MLE), meaning that we choose values for the parameters that maximize the likelihood for getting the data we have. The data chosen are typically daily data, as this is the "nicest" data for GARCH models. If one uses less frequent data, the volatility clustering effects are likely to disappear (Alexander 2008b, 137), while if intraday data is used one have to account for "seasonality"-effect, i.e. that more trading happens during the start and end of a day than during.

MLE for GARCH can be done in two ways: Quasi-MLE and Full MLE. Quasi-MLE are used when the focus of the study are the actual GARCH-parameters. We assume that the correct specification is chosen (e.g. a GARCH [1,1]), but that the errors are Gaussian. Essentially, the Gaussian likelihood is treated as the objective function to be maximized rather than a proper likelihood. This method is proven to give good estimates of α and β under appropriate assumptions on the true innovation function (McNeil et al 152).
Full-MLE uses the density of the true distribution in the likelihood function. This method gives more information, but we need to know the true distribution of the data.

For a more technical piece on estimating GARCH models by Quasi-MLE, see Francq and Zakoïan 2010, 141-179.

5.2 Multivariate GARCH

In portfolio optimization, dependence between the assets is arguably even more important than the variance of the individual stocks themselves. Therefore, an extension to a multivariate framework is appropriate.
Multivariate ARCH models appeared almost at the same time as the univariate models, and the extension to GARCH followed shortly after by Bollerslev et al. (1988). The first models had too many parameters to be useful for modeling much more than two assets at a time, and this "dimensionality curse" is still prevalent today with more parsimonious models and much faster computers.

Today, the most popular multivariate GARCH specifications seem to be the Constant Conditional Correlations (CCC) models by Bollerslev (1990) Jeantheau (1998), the BEKK model by Baba et al. (1995), and the Dynamic Conditional Correlations (DCC) models proposed by Tse and Tsui (2002) and Engle (2002) (Francq and Zakoïan [2014]).

A multivariate GARCH process is defined as

$$r_t = \mu_t + \epsilon_t \quad \textbf{(28)}$$

$$\epsilon_t = H_t^{\frac{1}{2}} z_t \quad \textbf{(29)}$$

Where

- r_t is a $n \times 1$ vector of the log returns of *n* securities at time *t*
- μ_t is a $n \times 1$vector of the expected value of the conditional log return of *n* securities at time *t*
- ϵ_t is a $n \times 1$vector of the mean corrected return of *n* securities at time *t*
- H_t is a $n \times n$ matrix of the conditional covariance's of ϵ_t at time *t*
- $H_t^{1/2}$ is a $n \times n$ matrix so that $H_t^{1/2}(H_t^{1/2})' = H_t$ is the conditional covariance matrix of ϵ_t. $H_t^{1/2}$ can be obtained by Cholesky decomposition of H_t.
- z_t is a $n \times 1$ vector of standardized i.i.d. random variables

Note that $H_t^{1/2}$ is *not* the square root of the conditional covariance matrix H_t (negative covariances would then be impossible).

Remaining is the specification of the matrix process H_t. Similarly to ensuring that the conditional variance is positive in the univariate case, we here have to ensure that the specification imposes positive definiteness (as the covariance matrix per definition is positive definite). Models for specifying H_t are often divided into four categories;

1. Models where H_t is modelled directly, such as the VEC and BEKK models.
2. Factor models motivated by parsimony, where r_t is assumed to be generated by a number of unobserved heteroskedastic factors.
3. Decomposition models where H_t is modeled through first modeling the conditional variances and conditional correlations and then combining them to construct H_t
4. Semi- and nonparametric approaches.

In this thesis we focus on and apply models from the third category, namely CCC and DCC type models. The conditional correlations in these models are modeled sequentially using the standardized residuals $u_t = D_t^{-1}\epsilon_t$. A second requirement also arises from decomposing H_t, namely that the values of the conditional correlation matrix has to be larger than or equal to minus one and less than or equal to plus one, i.e. $-1 \leq \rho_{ij} \leq 1$ for $i = 1, \dots, n$ and $j = 1, \dots, n$.

5.2.1 Constant Conditional Correlation (CCC) Models

The perhaps simplest multivariate conditional correlation model is the CCC model of Bollerslev (1990). As the name suggests, the correlation structure P is time-invariant in this model, leaving H_t to be expressed as follows:

$$H_t = D_t P D_t \tag{30}$$

Where

D_t is $diag = (h_{1t}^{1/2}, \dots, h_{nt}^{1/2})$. That is, D_t is the diagonal matrix containing the conditional standard deviations $h_{it}^{1/2}$;

$$D_t = \begin{pmatrix} h_{1_t}^{1/2} & 0 & \dots & 0 \\ 0 & h_{2_t}^{1/2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_{n_t}^{1/2} \end{pmatrix}$$

P is the positive definite correlation matrix, $P = [\rho_{i,j}]$ so that $\rho_{i,i} = 1$ for $i = 1, \dots, n$;

$$P = \begin{pmatrix} 1 & \rho_{1,2} & \dots & \rho_{1,n} \\ \rho_{2,1} & 1 & \dots & \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \dots & 1 \end{pmatrix}$$

The off-diagonal elements of H_t will then be defined as

$[H_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}, \;\; i \neq j$ so that

$$H_t = \begin{pmatrix} h_{1_t} & h_{1_t}^{1/2} h_{2_t}^{1/2} \rho_{1,2} & \dots & h_{1_t}^{1/2} h_{n_t}^{1/2} \rho_{1,n} \\ h_{2_t}^{1/2} h_{1_t}^{1/2} \rho_{2,1} & h_{2_t} & \dots & h_{2_t}^{1/2} h_{n_t}^{1/2} \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_t}^{1/2} h_{1_t}^{1/2} \rho_{n,1} & h_{n_t}^{1/2} h_{2_t}^{1/2} \rho_{n,2} & \dots & h_{n_t} \end{pmatrix}$$

If the conditional variances are modelled with a standard univariate GARCH model (we could also use another specification, e.g. the E-GARCH), the conditional variances can be written in vector form;

$$h_t = \omega + \sum_{j=1}^{q} A_j r_{t-j}^2 + \sum_{j=1}^{p} B_j h_{t-j} \quad \textbf{(31)}$$

Where

- ω has the form of a $n \times 1$ vector
- A_j and B_j are diagonal $n \times n$ matrices
- $r_t^2 = r_t \odot r_t$ i.e. the element-wise (Hadamard) product (defined in Appendix A1).

Positive definiteness of H_t is then guaranteed if P is positive definite and the elements of ω and the diagonal elements of A_j and B_j are positive. The positivity A_j and B_j's diagonal elements are not necessary unless $p = q = 1$ (Silvennoinen and Teräsvirta (2009)), but this case is extremely common in finance.

5.2.2 Dynamic Conditional Correlation models

Although the CCC model is attractive in terms of parameterization, the stylized facts of financial returns suggest that correlations should be modeled as dynamic. To alleviate this restriction, the dynamic correlation models also allows the correlation matrix P to be time-varying, i.e. P_t isn't necessarily equal P_{t+z} for any z but $z = 0$. Hence,

$$P_t = \begin{pmatrix} 1 & \rho_{1,2_t} & \dots & \rho_{1,n_t} \\ \rho_{2,1_t} & 1 & \dots & \rho_{2,n_t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1_t} & \rho_{n,2_t} & \dots & 1 \end{pmatrix}$$

$$H_t = D_t P_t D_t = \begin{pmatrix} h_{1_t} & h_{1_t}^{1/2} h_{2_t}^{1/2} \rho_{1,2_t} & \dots & h_{1_t}^{1/2} h_{n_t}^{1/2} \rho_{1,n_t} \\ h_{2_t}^{1/2} h_{1_t}^{1/2} \rho_{2,1_t} & h_{2_t} & \dots & h_{2_t}^{1/2} h_{n_t}^{1/2} \rho_{2,n_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_t}^{1/2} h_{1_t}^{1/2} \rho_{n,1_t} & h_{n_t}^{1/2} h_{2_t}^{1/2} \rho_{n,2_t} & \dots & h_{n_t} \end{pmatrix}$$

This makes the dynamic correlation models harder to estimate, as the correlation matrix has to be inverted for each time step t. There exists a vast number of methods for specifying P_t in the literature, we will focus on Engle (2002)'s Dynamic Conditional Correlation model. As this is a thesis on portfolio optimization and not econometrics, we will focus on the intuition and not dive too deep into the technicalities of the model. For a more rigorous discussion, see e.g. Engle and Sheppard (2001) and Francq and Zakoïan (2010, 2014).

Following notation of Silvennoinen and Teräsvirta (2009), we consider a dynamic matrix process

$$Q_t = S(1 - \alpha - \beta) + \alpha u_{t-1} u'_{t-1} + \beta Q_{t-1} \quad \textbf{(32)}$$

Where

- $u_t = D_t^{-1}\epsilon_t$, i.e. the standardized errors.
- S is the unconditional correlation matrix of the standardized errors u_t.
- α is a positive scalar parameter.
- β is a non-negative scalar parameter.
- $\alpha + \beta < 1$
- Q_0 is positive definite

This ensures positive definiteness, but not necessarily valid correlation matrices. To get that Q_t has to be rescaled;

$$P_t = diag\{Q_t\}^{-1/2}\, Q_t\, diag\{Q_t\}^{-1/2} = (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2} \quad \textbf{(33)}$$

Section 6: Copulas

The main references for this subsection are: Alexander (2008b), McNeil et al (2005), Nelsen (1999), Schmidt (2006)

Copulas are a statistical tool for capturing the dependence structure of a joint distribution, independent of the properties of the marginal distributions (see e.g. Nelsen, R. 1999). The term *copula* was first used by Sklar (1959), and is derived from *copulare,* which is latin for "to connect", or "join". The purpose of copulas is exactly that: to connect the marginal distributions of random variables into a joint distribution. This makes copulas an interesting alternative (or supplement) to multivariate GARCH models for combining univariate volatility models into multivariate models of asset returns. As with multivariate GARCH specifications, copulas can be made to be either constant or time-varying (dynamic).

6.1 Pitfalls of covariance and correlation

It can be proved mathematically that correlation is only a good measure for elliptical joint distributions. An obvious example would be that $Corr(X, X^2) = 0$ even though X and X^2 clearly have perfectly quadratic dependence. A famous paper by Embrechts et al. (as reviewed by Alexander (2008b)) from 2002 identifies and illustrates several major problems associated with Pearson's correlation, such as

- Correlation is not invariant under transformation of variables, e.g. $Corr(X, Y) \neq Corr(\ln[X], \ln[Y])$
- Feasible values for correlation depends strongly on the marginal distributions, e.g. if X and Y are lognormal then certain correlations (inside of the normal range of values, i.e. $[-1, 1]$) is impossible! This confuses conventional interpretation of the correlation coefficient drastically, as 1 is no longer necessarily perfect positive dependence, and similar for -1.

To put it briefly, these problems arise because correlation is really only a proper dependence measure when the joint distribution is elliptical. We formalize the

desired properties of a dependence measure in the part on Concordant Metrics and Tail dependence in the appendix.

6.2 Definition of a copula and Sklar's theorem

An n-dimensional copula is a distribution function on $[0,1]^n$ with uniformly distributed margins on $[0,1]$.

According to Sklar 1959 (as reviewed by Alexander [2008b] and McNeil et al [2005]), there exists a function *C*, a copula, mapping $[0,1]^n$ into [0,1] so that

$$\boldsymbol{F}(y_1, \dots, y_n) = C[F_1(y_1), \dots, F_n(y_n)] \quad \textbf{(34)}$$

Where

- $\boldsymbol{F}(y_1, \dots, y_n)$ is an *n*-variate cumulative distribution function
- $F_i(y_i),\ i = 1, \dots, n$ is a set of univariate margins

If $F_i(y_i)$ are continuous for all *i*, then C is unique. Otherwise, C is uniquely determined on Ran $F_1 \times$ Ran $F_2 \times \dots \times$ Ran F_n, where Ran F_i denotes the range of F_i. Conversely, if *C* is a copula and $F_1, \dots, F_n$ are univariate distribution functions, then $\boldsymbol{F}$ must be a joint distribution function with margins $F_1, \dots, F_n$. For a full proof of this theorem, see Nelsen (1999, 18).

Denoting $y_i = F_i^{-1}(u_i)$ for $i = 1, \dots, n$ where $F_i^{-1}(u_i)$ is the inverse marginal distribution and u_i are uniformly distributed [0,1], it follows that $u_i = F_i(y_i)$. In turn inserting this into Sklar's equation, we get the following;

$$\boldsymbol{F}[F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)] = C[F_1(F_1^{-1}[u_1]), \dots, F_n(F_n^{-1}[u_n])]$$

$$= C(u_1, \dots, u_n) \quad \textbf{(35)}$$

The importance of this theorem is that it shows that all multivariate distribution functions contain copulas, and that these copulas may be used in conjunction with univariate distribution functions to construct the joint distributions. It also shows that we can think of C as the joint distribution of the uniformly distributed variables $u_i\ ,\quad i = 1, \dots, n$. Thus, the copula represents the dependence structure separated from the marginal distributions.

The joint density function is then obtained by differentiating once using the chain rule,

$$f(y_1,..,y_n) = c[F_1(y_1),\dots,F_n(y_n)]f_1(y_1) \times f_2(y_2) \times \dots \times f_n(y_n) \quad \textbf{(36)}$$

6.3 Bivariate Copulas

To develop intuition on how copulas work, it is useful to consider the bivariate case (i.e. two random variables);

Two random variables, X_1 and X_2 are fully described by their respective cumulative distribution functions (cdf) (i.e. the marginals),
for X_1: $F_1(x_1) = P(X_1 \leq x_1)$
and similarly for X_2: $F_2(x_2) = P(X_2 \leq x_2)$

The marginal do however give us no information about the joint behavior of X_1 and X_2. From basic probability rules, we know that $P(A\ and\ B) = P(A) \times P(B)$ if A and B are independent events. Applying it to our random variables (and assuming independence) we get $P(X_1 \leq x_1, X_2 \leq x_2) = F(x_1) \times F(x_2)$. Here we have expressed the joint distribution through three ingredients; each of the marginals and a description of the type of relation between the variables (here independence i.e. independent copula). The beauty of copulas is that it allows us to do this when the variables are dependent of each other, and we can do so with different types of marginals.

6.4 Tail Dependence

A formal definition of tail dependence is given in Appendix A6; here we focus on the intuition. Tail dependence looks at the dependence (association) in the tails of the joint distribution. The reason for looking at tail dependence is that it provides a measure of the strength of dependence between extreme tail events. Of particular interest for risk management is what is called the lower tail dependence, which can be interpreted as the probability that asset 1 gets a very low return given that asset 2 has a very low return. Modeling this is important in finance, as

we know that downfalls in one market often unleash a "domino-effect" over other markets.

6.5 Elliptical Copulas

The Elliptical copulas are part of the implicit copula category (Implicit copula are explained in Appendix A6. The important thing to know is that implicit copulas use the copula density rather than the distribution function). We will study the two most popular elliptical copulas, namely the Gaussian (i.e. normal) copula and the Student t copula.

6.5.1 Gaussian Copulas

The Gaussian copula is derived from the n-dimensional multivariate and univariate standard normal distribution functions, i.e. it's defined by

$$\boldsymbol{C}(u_1, \dots, u_n; P) = \Phi_m[\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)] \quad \textbf{(37)}$$

Where

- Φ_m is the *n*-dimensional multivariate standard normal distribution function.
- Φ is the univariate standard normal distribution function.
- P is the correlation matrix.

6.5.1.1 Tail dependence of the Gaussian copula

It can be shown that the Gaussian tail dependence approaches zero when one goes far enough into the tail (see e.g. McNeil et al 2005, 210-211). This is sometimes referred to as "asymptotic independence", and means that regardless of how high the correlation between two assets are, if we just go far enough into the tails, extreme events appear to occur independently. This is very problematic when seen in relation to the stylized facts, as we know that when we are the furthest into the left tail for one asset, is exactly when we expect "bad things" to happen the most often for other assets. To put it differently, the Gaussian copula's tail dependence imply the exact opposite.

6.5.2 Student t Copulas

Another example of an elliptical copula is the multivariate Student t copula, defined by

$$\boldsymbol{C}(u_1, \dots, u_n; P) = mt_v[t_v^{-1}(u_1), \dots, t_v^{-1}(u_n)] \quad \textbf{(38)}$$

Where

- mt_v is the *n*-dimensional multivariate Student t distribution function with v degrees of freedom.
- t_v is the univariate Student t distribution function with v degrees of freedom.
- P is the correlation matrix.

6.5.2.2 Tail dependence of the Student t copula

Contrary to the Gaussian copula, it can be shown that the Student t copula is asymptotically dependent in both the upper and lower tail (see McNeil et al 2005, 211). Even for zero or negative correlations there is still some tail dependence. This gives the Student t copula an edge over the Gaussian copula when it comes to modelling most financial assets. As with all elliptical copulas, the tail dependence is however symmetric. This is not necessarily what we want when modeling in finance. An obvious example would be if the marginals are belonging to different stocks. While an extreme downfall in one stock is often associated with a high probability of other stocks also taking a huge fall (i.e. the market "collapses"), the opposite is not necessarily true. A large upswing in one stock's price tends to be associated with positive news, often firm-specific. In the context of pure risk management, the lower tail is what typically is considered, but this is likely to matter more when optimizing for both a high return and a low risk.

6.6 Copulas for Portfolio Optimization

Multivariate copulas can be used in portfolio optimization to specify the dependence between assets. If we use the normal or student t copula, we make the assumption that the return distribution is elliptical, and the optimization will be

based on a correlation matrix. However, we are now free to specify the asset marginals to have their individual empirical distribution, or a parametric distribution we believe can describe the asset well.

One way to apply copulas for portfolio optimization is to combine the marginals in a simulation approach for estimating Conditional Value at Risk. The copulas then offer a great advantage in terms of flexibility. We apply this on real data through CVaR optimization in section 8.

6.6.1 Simulation using the Gaussian Copula

1. Simulate $x \sim \mathcal{N}_d(0, P)$
   Simulating n rows from the multivariate normal distribution gives us an $n \times d$ matrix x of standard normal realizations with correlation given by P.
2. Set $u = \Phi(x)$
   We transform each column vector using probability transformation. That is, each realization is put through the standard normal marginal density function to get uniform (0,1) variables.
3. We are now free to transform each vector d in our (n x d) matrix of uniform variables with the appropriate inverse marginal density function. Thus, we can get d column vectors of different marginal distributions with length n of, which still possess the correlation structure given by P.

6.6.2 Simulation using the Student t copula

For the Student t Copula the steps are equivalent, except that we use the multivariate t distribution in step 1, and student t marginals in step 2. All with v degrees of freedom.

1. Simulate $x \sim t_d(0, P, v)$
   Simulating n rows from the multivariate t distribution.
2. Set $u = t_v(x)$
   Transform each column vector using probability transformation.
3. Transform columns d to suitable marginal distributions using inverse density functions.

Section 7: Methodology

As the emphasis of this thesis is purely on the risk measure input of the Markowitz algorithm and not the expected return vector (that is to the extent that they are separable), we have chosen to conduct the comparison by minimizing portfolio risk, rather than search for an optimal risk-reward tradeoff. This in turn means that we don't have to specify an expected return vector, and we believe this will reduce any unnecessary noise from the comparison. To put it differently, we simply choose the minimum risk portfolio on efficient frontier.

All the empirical work done in this thesis is limited by the common mathematical assumptions of finance listed in Appendix A2.

7.1 Strategies

To compare the risk measures and estimation techniques we create a separate strategy for each combination of estimation technique and risk measure considered. The strategies are rebalanced weekly, i.e. every fifth trading day. This reduces the impact of assuming no transaction costs compared to that of daily rebalancing. For a more thorough discussion on rebalancing, see "Suggestions for future work" in section 9.

For all portfolios, the weight per asset is constrained to a maximum of 40%, to ensure that the portfolios exhibit some diversification across assets.

7.1.1 The naïve strategy

The naïve strategy is included for benchmarking purposes, and is simply to invest an equal share in each available asset, i.e.
$W_i = \frac{1}{n}$, for all i, where n is the number of available assets. As both datasets contain 10 assets, $Wi = \frac{1}{10}$ for all i.

7.1.2 Regular specification strategies

The regular specification strategies are calculated using the simplest estimation techniques.

7.1.2.1 Regular Markowitz MV, expanding window.

The simplest MV-strategy we consider is the Markowitz minimum variance portfolio, assuming that returns are i.i.d. This strategy serves as a benchmark and is calculated by estimating the expanding window sample covariance matrix, i.e. for every trading day that goes by the input for the sample covariance matrix is expanded by one observation.

7.1.2.2 Regular Min CVaR, expanding window.

The simplest Min CVaR-strategy we consider is minimizing the CVaR based on the historical return distribution. This strategy serves as a benchmark and is calculated by minimizing CVaR for the expanding window i.e. for every trading day that goes by the input is expanded by one observation.

7.1.3 MV-GARCH specification strategies

The MV-GARCH specification strategies are minimum variance strategies using the GARCH framework to model variance. We only consider order (1,1) univariate models. For the portfolio optimization procedure, we have calculated a rolling estimation of forecasted covariance matrices from the DCC and CCC models. I.e. are doing a 5-day forecast of the covariance matrix each 5th day, with estimation of the model done with data up to that day. We use each of these matrices as input in the Markowitz minimum variance optimization algorithm, and thus we rebalance each 5th day with a new estimate of future volatility and covariance given by the multivariate GARCH models.

7.1.3.1 CCC S-GARCH

The CCC S-GARCH strategy models the univariate volatility with a standard GARCH (1,1) model, using a normal error distribution. The dependence structure is modeled with the CCC specification.

7.1.3.2 DCC S-GARCH

The DCC S-GARCH strategy models the univariate volatility with a standard GARCH (1,1) model, using a normal error distribution. The dependence structure is modeled with the DCC specification.

7.1.3.3 DCC E-GARCH

The DCC E-GARCH strategy models the univariate volatility with the E-GARCH (1,1) model, using a Student t error distribution. The dependence structure is modeled with the DCC specification.

7.1.4 Copula-GARCH CVaR specifications

The Copula-GARCH CVaR specifications minimize $95\% - CVaR_{1day}$ on a simulated sample given by the algorithm in section 7.1.5.

7.1.4.1 Normal Copula, DCC S-GARCH

This strategy uses the GARCH (1,1) model with normal error distribution as the univariate specification. Returns are simulated based on the normal copula and the simulation algorithm described in 7.1.5. The correlation structure implied by the fitted DCC model.

7.1.4.2 Normal Copula, DCC E-GARCH

This strategy uses the E-GARCH (1,1) model with student t error distribution as the univariate specification. Returns are simulated based on the normal copula and the simulation algorithm described in 7.1.5. The correlation structure implied by the fitted DCC model.

7.1.4.3 Student t Copula, DCC S-GARCH

This strategy uses the GARCH (1,1) model with normal error distribution as the univariate specification. Returns are simulated based on the student t copula and the simulation algorithm described in 7.1.5. The correlation structure implied by the fitted DCC model.

7.1.4.4 Student t Copula, DCC E-GARCH

This strategy uses the E-GARCH (1,1) model with student t error distribution as the univariate specification. Returns are simulated based on the student t copula and the simulation algorithm described in 7.1.5. The correlation structure implied by the fitted DCC model.

7.1.5 Simulation algorithm

For the simulation process we use the copula- and GARCH-framework in combination. The elliptical copulas, normal copula and t-copula, both use the correlation matrix as the measure for the dependence structure between assets. The DCC framework forecast a correlation matrix based on past information. In the out of sample testing period, we do a rolling estimation and forecasting of correlation matrices. That is, we start at the beginning of the out of sample testing period, i.e. at time 0 of that sample, and fit the model to the training data. Thus we get parameters for the model, and can forecast the matrix. For every 5^{th} time step, we forecast 5 days ahead, and rebalance and refit every 5 days. Thus, after 5 days, we have 5 more data point in our sample. We use the estimated parameters on the new data set (the old plus the 5 new observations) and forecast a correlation matrix for the next 5 days. This is done successively every 5 days in the out of sample testing period, and we thus get a number of matrices equal to the number of data points in our out of sample data divided by 5, i.e. 100 in this case.

Now that the matrices are available, we impose on the copulas that the dependence structure is given by the correlation outputs from the DCC model. We use student t marginals for all assets. We use the diagonal of the covariance matrix outputs from the DCC-GARCH model as our estimates of variances for each of the marginals. The mean is estimated from the training data, and the degrees of freedom parameter for the student t-marginals is estimated by MLE for

all assets. From section 6.6, when given a correlation matrix of dimension d, we can simulate $n \times d$ uniform random variables in the [0,1] space which possess the dependence structure from the copula. We can then fit the suitable distributions, in this case the marginal student t distributions, to each of the d simulated uniform vectors of length n, using the appropriate quantile function for each margin.

For each time step, we simulate a new dataset with the parameters estimated as explained above. Thus, for each correlation matrix, we get a simulated dataset with dependence structure given by the DCC-GARCH and the respective copulas. This is useful when we want to estimate a portfolio using CVaR risk criterion, as we get to implement the time varying dependence structure along with the time varying second moment of the marginals, and in addition we can simulate thousands of data points, making α-CVaR optimization more robust (keep in mind, only $(1-\alpha)100\%$ of the points are actually used). The drawbacks of this procedure are the reliance on the assumptions we make on the marginal distributions and that the DCC-GARCH framework is able to adequately reflect the true dependence structure in conjunction with the copula.

To summarize the procedure:

1. Calculate a rolling forecast of DCC-GARCH models, to obtain the forecasted correlation and covariance matrices with data as if we were in real time.
2. Choose either a normal or student t copula to model the dependence, and impose the correlations from step 1 to be able to simulate data with that dependence.
3. Simulate data at each time step *t* using the estimated correlation matrix P_t, and transform the uniform output from the copula with the quantile functions of the suitable marginals. Each marginal is using second moment input from the DCC-covariance matrix at the same time step t.

7.2 Ranking the strategies

To rank the measures (strategies) we will take a similar approach to that of Yilmaz (2010), i.e. compare the realized risk of the strategies with sample statistics. In addition to sample volatility, we will look at sample Value at Risk, sample Conditional Value at Risk and Downside deviation (i.e. the square root of semi-variance).

To assert the significance of the differences in variance we apply the Brown-Forsythe test (Brown and Forsythe 1974). The Brown-Forsythe test (also referred to as modified Levene's test or simply Levene's test) is a modified, more robust version of the more well-known Levene's test for homogeneity of variances. Rather than using the mean, the B-F test uses the median, a more robust measure of central tendency when the distribution is skewed.
As far as the authors know, there is no "good way" to do the same with Value at Risk and Conditional Value at Risk. An intuitive way of getting an indication of the significance of differences in VaR (or CVaR) is the approach taken in Hendricks (1996). The issue with this approach is that it requires a relatively high amount of samples to give any accuracy, so it wasn't really feasible for this paper (some of the estimations take a couple of hours to complete in R). Thus, we merely note the differences in VaR and CVaR, as the statistical power of these measures (as presented) are low.

Section 8: Empirical application

In this section we will apply and compare the risk measures and estimation techniques discussed in the previous sections.

8.1 Data

In the empirical application of the thesis we apply two data sets: One all equity dataset, and one dataset of commodity indices. Both of the samples are split into a "training sample" and a "testing" sample (also referred to as the "out of sample" period). The testing sample is chosen to be the last 500 observations (trading days) for both samples.

8.1.1 Equity dataset

The first dataset is a 10-dimensional dataset from the US large cap equity market. The dataset contains a total of 2531 observations (trading days), spanning from the 31st of July 2006 until the 17th of August 2016. The US large cap equity market is highly efficient and liquid, making it a great candidate for quantitative approaches. However, equities are generally highly correlated, making diversification benefits potentially harder to detect. Initially we wanted to make use of the entire S&P500 for comparing the risk measures, but we quickly realized that this simply wasn't feasible – while both computational power and statistical techniques have made leaps forward in the recent years, the dreaded "dimensionality curse" is still very much alive. Accordingly, we had to find a way to limit the amounts of assets chosen and we thought the best way was to use indices of stocks rather than the individual stocks themselves.

S&P Dow Jones present a total of 11 "Select Sector Indices" on their web site, with the constituents all being members of the S&P 500, and each constituent being assigned to at least one index. Unfortunately, two of the indices, Real Estate and Financial Services, only have data from 2015 and beyond, and our models require quite a bit of data for both estimation and out of sample testing. As a

substitute for the Real Estate index we added the “Select Industry” index Home builders. The Home builders’ index isn’t constructed purely by securities from the S&P 500, but there is a requirement for both market capitalization and trading liquidity. As for the Financial Services index we assume that it is highly correlated with at least the Financials index, and that we don’t lose much of the “total equity picture” by omitting it. Thus we end up with the following 10 indices;

- S&P 500 Consumer Discretionary Select Sector Index
- S&P 500 Consumer Staples Select Sector Index
- S&P 500 Energy Select Sector Index
- S&P 500 Financials Select Sector Index
- S&P 500 Health Care Select Sector Index
- S&P Home builders Select Industry Index
- S&P 500 Industrials Select Sector Index
- S&P 500 Materials Services Select Sector Index
- S&P 500 Technology Select Sector Index
- S&P 500 Utilities Select Sector Index

A detailed description of how the indices are constructed is available on the S&P Dow Jones web site (see the Bibliography, S&P Dow Jones Indices. 2016a).

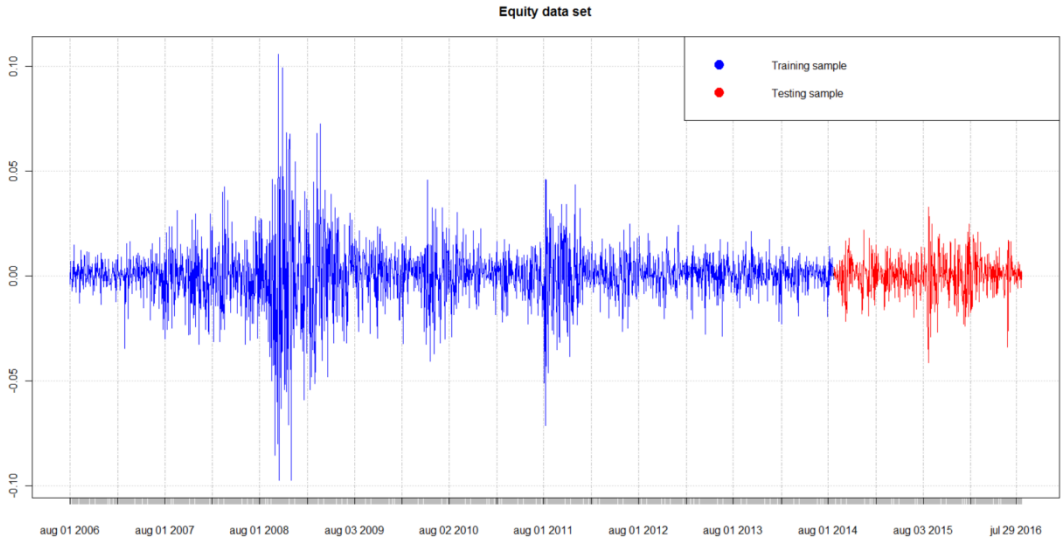


Figure 5 - Graphical representation of the Equity dataset through the naive portfolio

8.1.2 Commodity dataset

The second dataset is also 10-dimensional, containing data from the commodity market. For this dataset we handpicked commodities we believed to be lowly correlated, and used the S&P GSCI indices as a measure of their performance over time. The S&P GSCI indices are designed to benchmarks for investing in the commodity markets, and is also designed to be tradable (e.g. there exists ETFs aiming to track the indices). This dataset contains a total of 2537 observations (trading days), spanning from the 31st of July 2006 until the 22nd of August 2016.

The indices we chose were:

- S&P GCSI Brent Crude
- S&P GCSI Cocoa
- S&P GCSI Coffee
- S&P GCSI Corn
- S&P GCSI Cotton
- S&P GCSI Feeder Cattle
- S&P GCSI Gold
- S&P GCSI Sugar
- S&P GCSI Zinc

A detailed description of how the indices are constructed is available on the S&P Dow Jones web site (see the Bibliography, S&P Dow Jones Indices. 2016d).

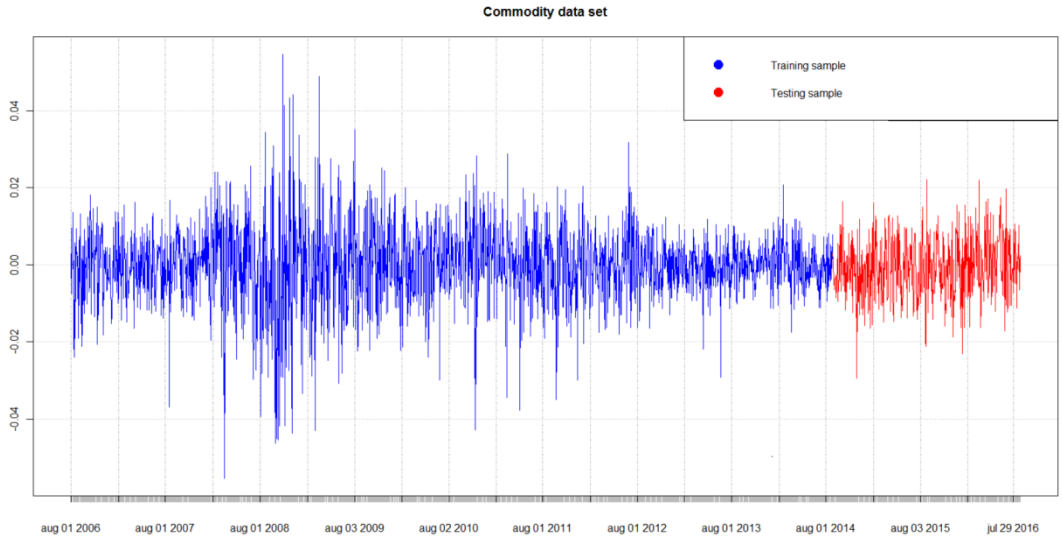


Figure 6 - Graphical representation of the Commodity dataset through the naive portfolio

8.2 Empirical results

8.2.1 Equity dataset

8.2.1.1 Initial examination

First we examine the marginal distributions to check for pronounced deviations from normality. The QQ plots examine the quantiles of two probability distributions (here the theoretical normal distribution on the X axis and the sample distribution on the y axis). If the two distributions compared are similar, the plot will approximate the line $y = x$. If the quantiles are linearly related, the plot will approximate a line, but not necessarily $y = x$.

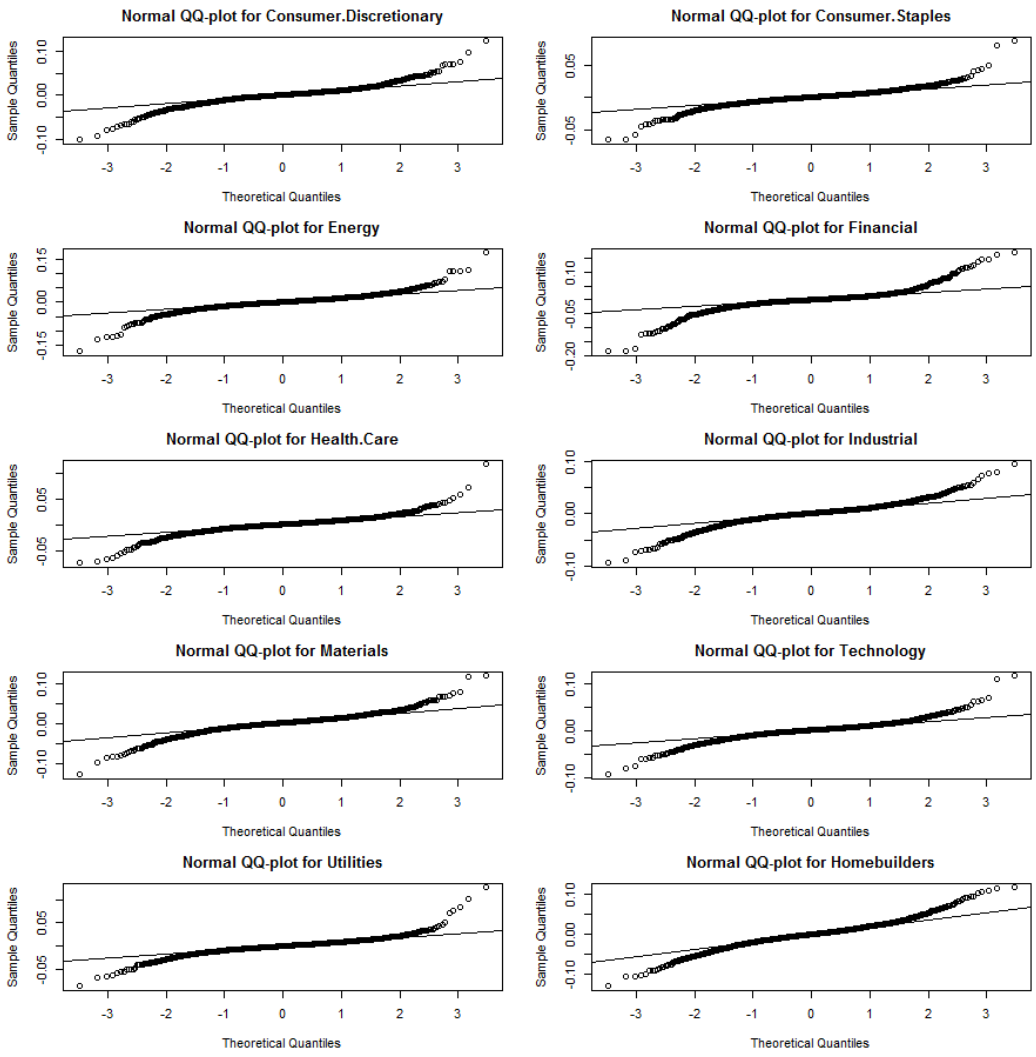


Figure 7 - Normal QQ plot, Equity

As expected, we see clear evidence of deviations in the tails for all assets. In particular, we see that the sample quantiles have more extreme values than the normal distribution in the tails. Thus, we can conclude the assets have heavier tails than the normal distribution (Note: We also perform a formal test in the form of the Jarque-Bera statistic, which can be found in Appendix B1. As expected, the hypothesis that the assets are normally distributed are rejected at all conventional confidence levels.)

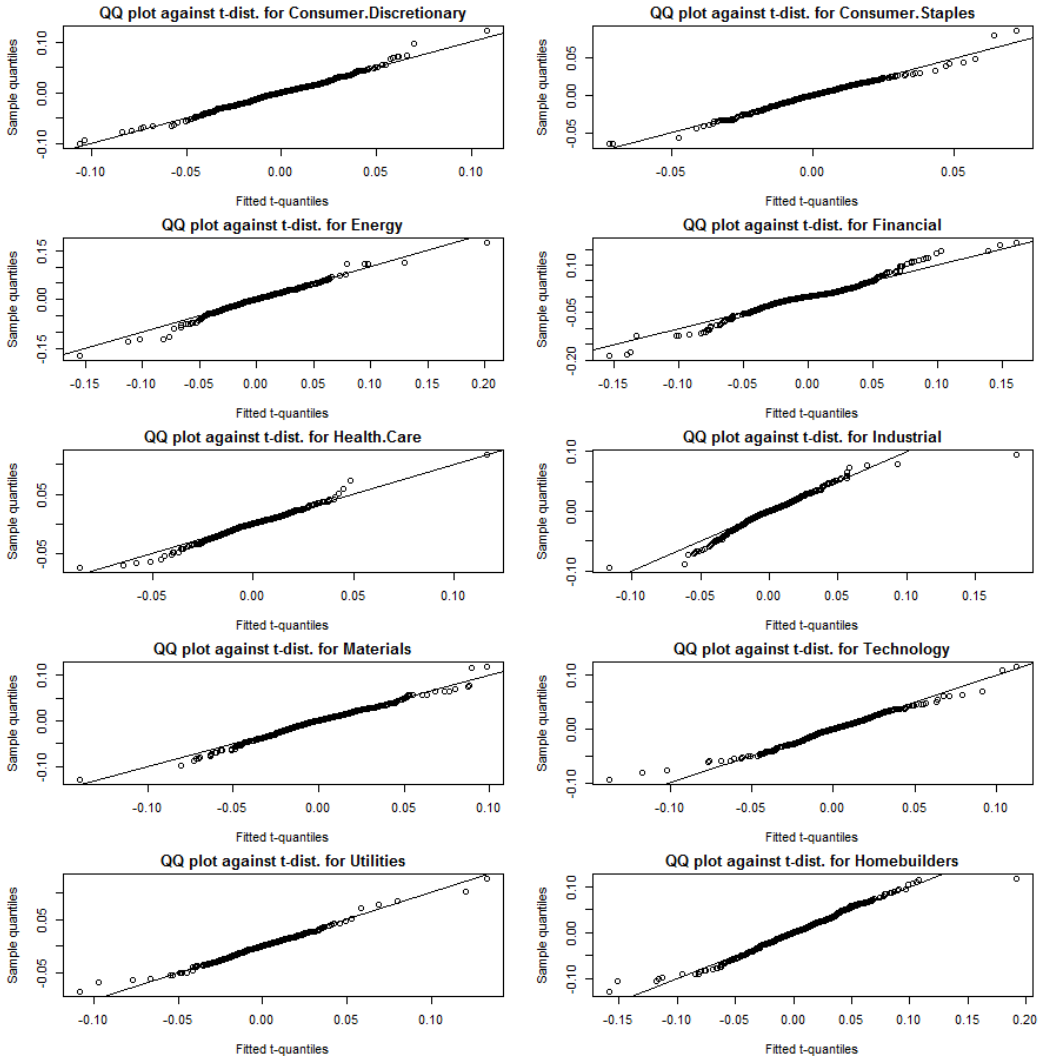


Figure 8 - Student t QQ plots, Equity

We see a clear improvement in approximation of the sample distributions by the straightness of the plots in the tails, indicating that a t-distribution is better at capturing the extreme tail events.

We also examined the ACFs of the distributions. As one would expect, the linear ACF was non-significant while the squared ACF was significant. These plots can be found in Appendix B1.

8.2.1.2 Fitting the univariate GARCH models

We proceed by fitting the univariate GARCH models. We fit one standard GARCH model with assumed normally distributed errors, and we fit an exponential GARCH with assumed t-distributed errors.

The ARMA parameters in the conditional mean equations at times turn up mostly significant. As for the S-GARCH parameters, all the alphas and betas are significant, while the omegas are typically insignificant. For the E-GARCH all parameters are significant, with the exception of the gamma parameter of Homebuilders.

Table 1 - Univariate S-GARCH(1,1) fit, Equity

S-GARCH Model Fit

Optimal Parameters:

	Consumer.Discretionary	Consumer.Staples	Energy	Financial	Health.Care	Industrial	Materials	Technology	Utilities	Homebuilders
mu	0.00091	0.00070	0.00084	0.00074	0.00079	0.00086	0.00082	0.00092	0.00057	0.00064
ar1	0.76325	0.61151	0.71910	0.66072	0.63352	0.88454	0.90999	0.72725	0.95785	-0.01130
ma1	-0.79819	-0.68676	-0.76427	-0.73002	-0.68610	-0.90995	-0.94509	-0.77465	-0.97706	0.05252
omega	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
alpha1	0.08749	0.12327	0.08256	0.11832	0.11488	0.08879	0.09497	0.08919	0.10807	0.07567
beta1	0.89922	0.84776	0.91200	0.88068	0.85650	0.90095	0.89795	0.89544	0.87949	0.91679

Table 2 - Univariate E-GARCH(1,1) fit, Equity

e-GARCH Model Fit

Optimal Parameters:

	Consumer.Discretionary	Consumer.Staples	Energy	Financial	Health.Care	Industrial	Materials	Technology	Utilities	Homebuilders
mu	0.00088	0.00059	0.00081	0.00068	0.00071	0.00078	0.00076	0.00087	0.00051	0.00029
ar1	0.73023	0.61315	-0.44037	0.63905	0.60028	0.74933	0.71433	0.36660	0.73277	0.01307
ma1	-0.75568	-0.68193	0.38837	-0.70131	-0.64437	-0.77233	-0.75633	-0.40988	-0.77350	0.02515
omega	-0.11216	-0.25031	-0.10248	-0.07239	-0.28039	-0.11532	-0.08661	-0.16565	-0.16254	-0.06696
alpha1	-0.10402	-0.11093	-0.07945	-0.09195	-0.12483	-0.11140	-0.09499	-0.12900	-0.05946	-0.05862
beta1	0.98789	0.97471	0.98818	0.99183	0.97059	0.98740	0.99032	0.98221	0.98268	0.99161
gamma1	0.14507	0.16610	0.14787	0.18400	0.16648	0.12811	0.12471	0.12772	0.18359	0.13260

We see from the parameters that all series are stationary as $alpha1 + beta1 < 1$ for the standard GARCH, and $beta1 < 1$ for the E-GARCH. From the large beta values we see that prior volatility has a high impact on forecasted volatility. This is true for both models.

As for the other E-GARCH parameters we get some indications that the model might be inappropriate. The alpha parameters are negative, which doesn't make sense, a shock should increase volatility increase, not decrease it. As for the gamma-parameters, they as expected have the opposite sign of the alphas, indicating the presence of leverage effects in the equity markets.

We also inspect the ACF's of the squared standardized residuals. If the GARCH specifications work as intended, we should now see little to no significant autocorrelation.

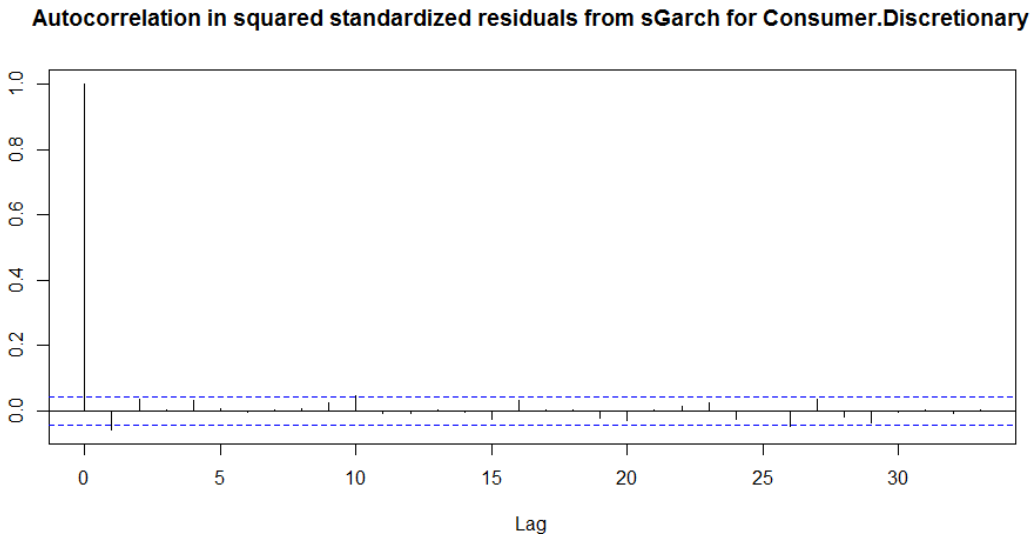


Figure 9 - ACF for the squared GARCH(1,1) standardized residuals, Consumer.Discretionary

We note that the autocorrelations are largely reduced, and this is similar for the other indices (Appendix B1).

8.2.1.3 Fitting the multivariate GARCH models

To fit the CCC models we use the unconditional sample correlation, in combination with the stand GARCH (1,1) models.

sGarch DCC for equities				
[Joint]dcca1	0.022309	0.001414	1.5773e+01	0.000000
[Joint]dccb1	0.960437	0.003027	3.1724e+02	0.000000
eGarch DCC for equities				
[Joint]dcca1	0.020279	0.001386	1.4634e+01	0.000000
[Joint]dccb1	0.962812	0.003013	3.1960e+02	0.000000

Figure 10 - DCC parameters, Equity

Here a1 and b1 is the α and β in equation **(32)** from section 5.2.2. The high b1 indicates that recent correlation has a large effect on the future correlation, while the low a1 indicates that shocks in correlation have less of an effect.

8.2.1.4 Simulation diagnostics

We proceed to simulate data for the Copula-GARCH CVaR optimization. This is done according to the algorithm provided in section 7.1.5. To see if the simulation algorithm is working as intended we test it on the training data. We do some informal tests in the form of visual diagnostics, to see if the characteristics of the simulated data are similar to the characteristics of the training data. For this we use the sample correlation and covariance matrix as inputs. We want the simulated data to have similar frequency of tail events as the training data set. We also want the t-copula to have tail dependence similar to that of the training data. At last, we want the copula to simulate data with a similar correlation structure to the training data.

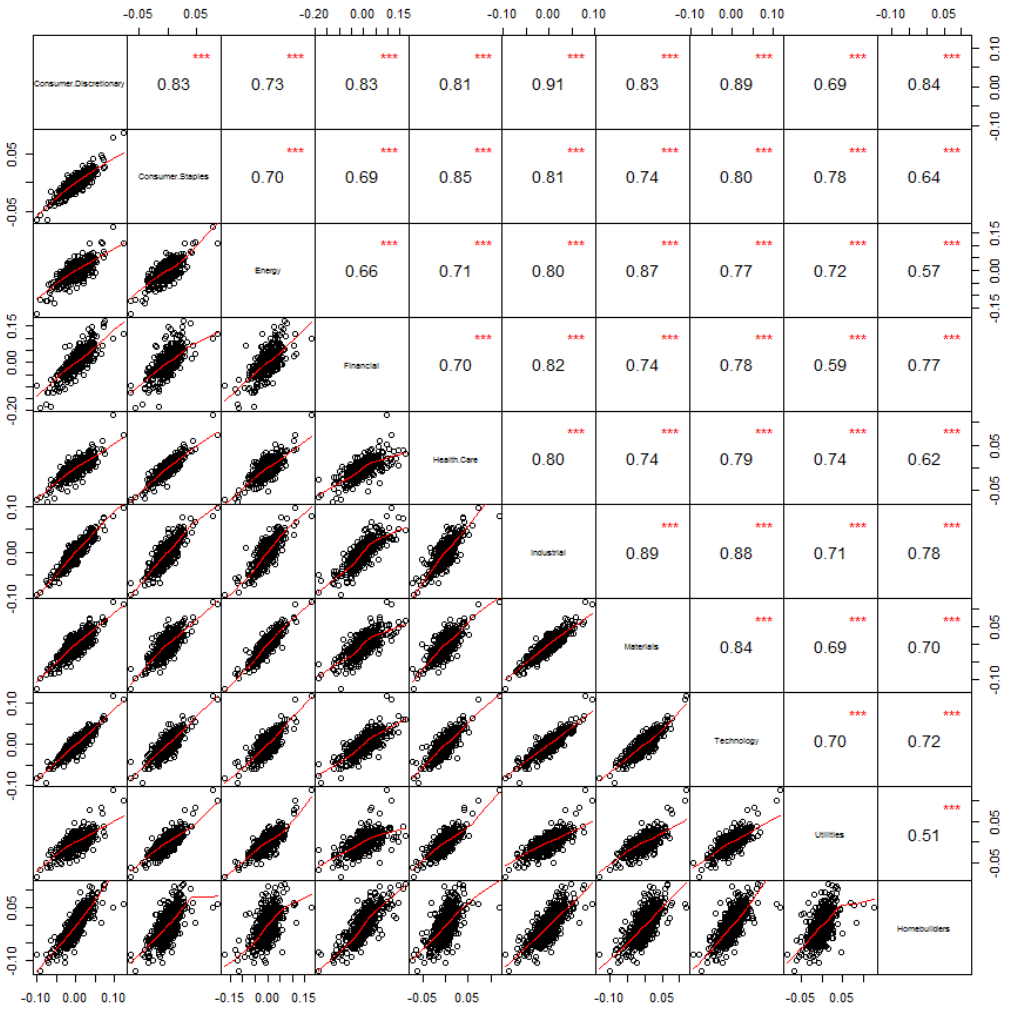


Figure 11 - Pairs plot training data, Equity

The Pairs plots are divided in two parts. On the lower triangular, there are pairwise scatterplots of the data. These illustrate the linear dependence between the assets. On the upper triangular we have the correlation coefficients. On the diagonal we have the asset names. We observe mainly lower tail dependence from the fact that the plots narrows as we go further into the tails, and that the extreme events seem to pair up in the lower left corner.

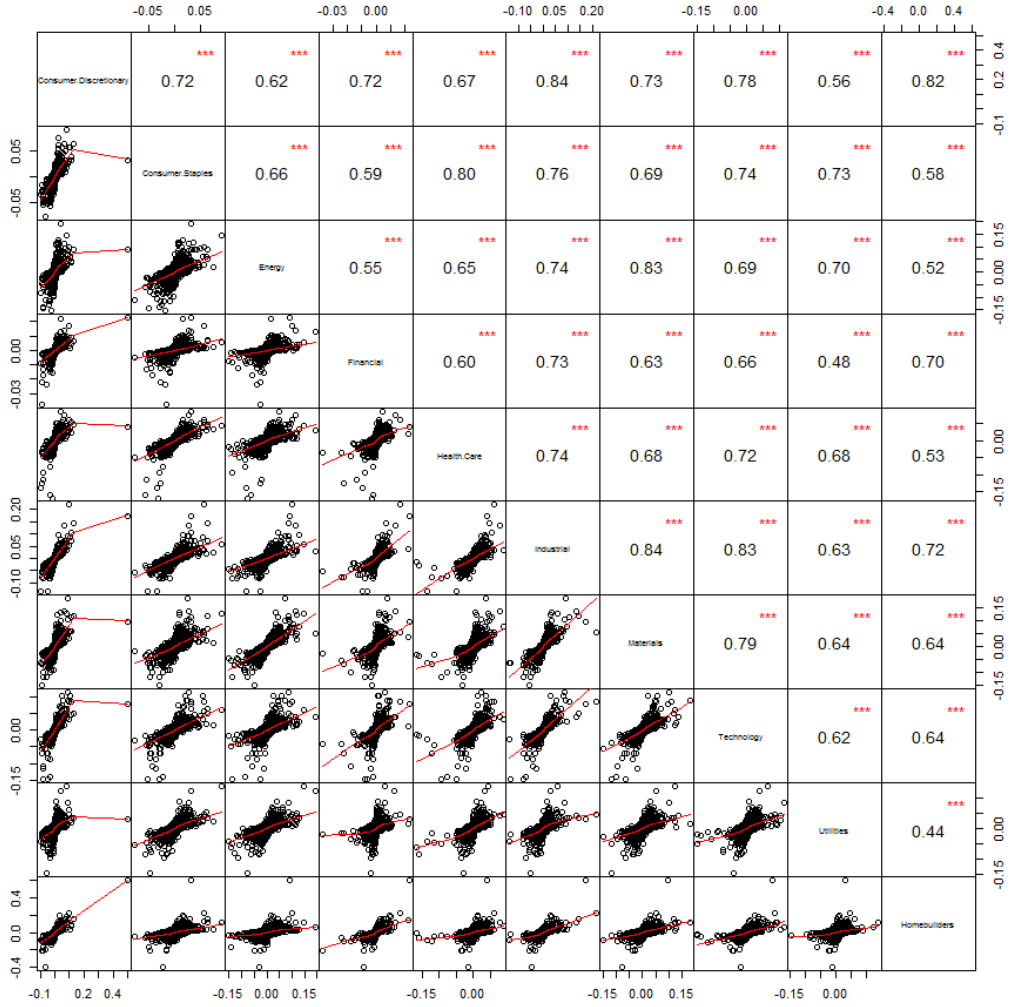


Figure 12 - Pairs plot G-SIM, Equity

The above picture is the Pairs plots from the simulations from the Gaussian copula. We see that the correlation structure is intact, but that the tail dependence as expected is not modelled. We see this from the absence of the data points in the lower left corner. Rather, when one asset reaches extreme values, they clump close to one axis, and stay in "the middle" of the other, indicating the extremely low probability of two tail events at the same time.

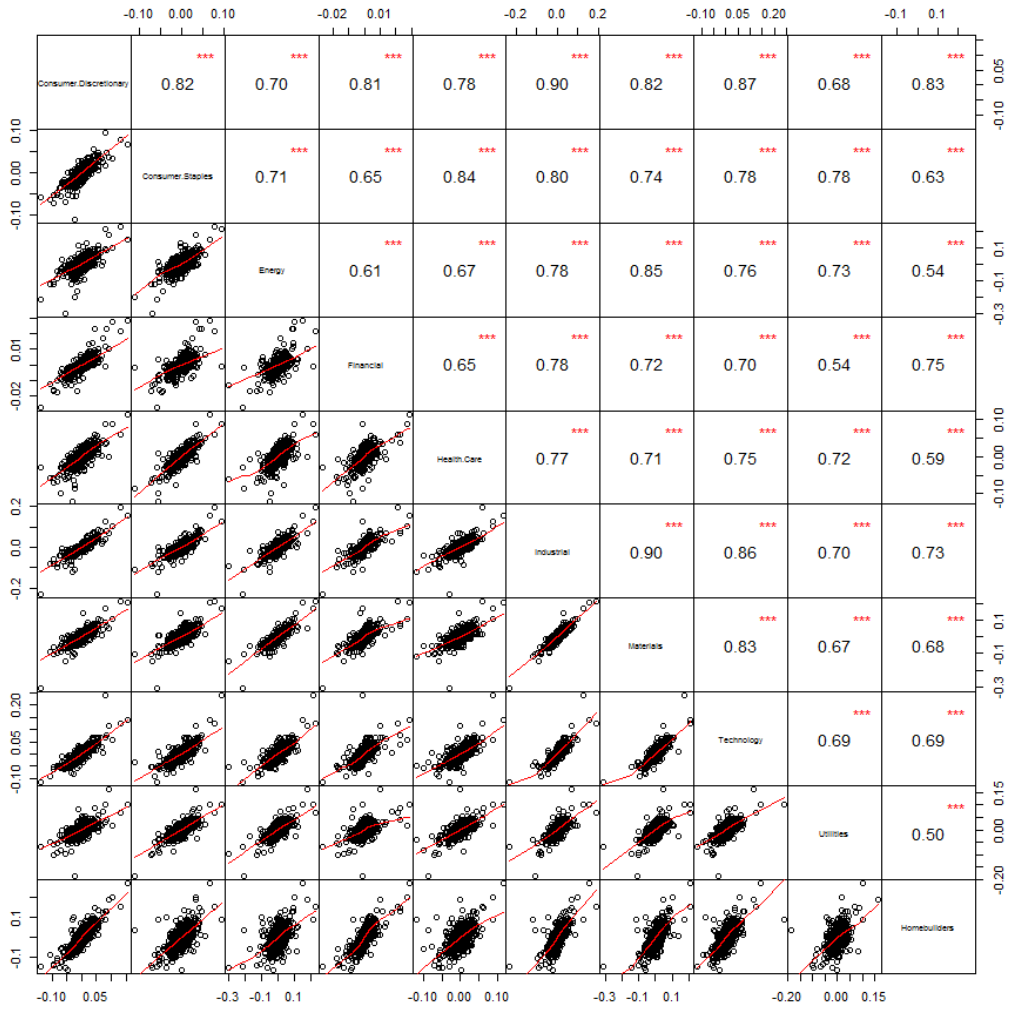


Figure 13 - Pairs plot t-SIM, Equity

Here we see the Pairs plot from the t-copula simulated data. We see that the tail dependence is present, but symmetric as extreme points are located in both the lower left and upper right corner. This is expected for a t-copula simulation, but is not a perfect approximation of the real world. Also the correlation structure is modelled well compared to the training data, i.e. the values in the upper right triangle are similar to that of the training data in Figure 11.

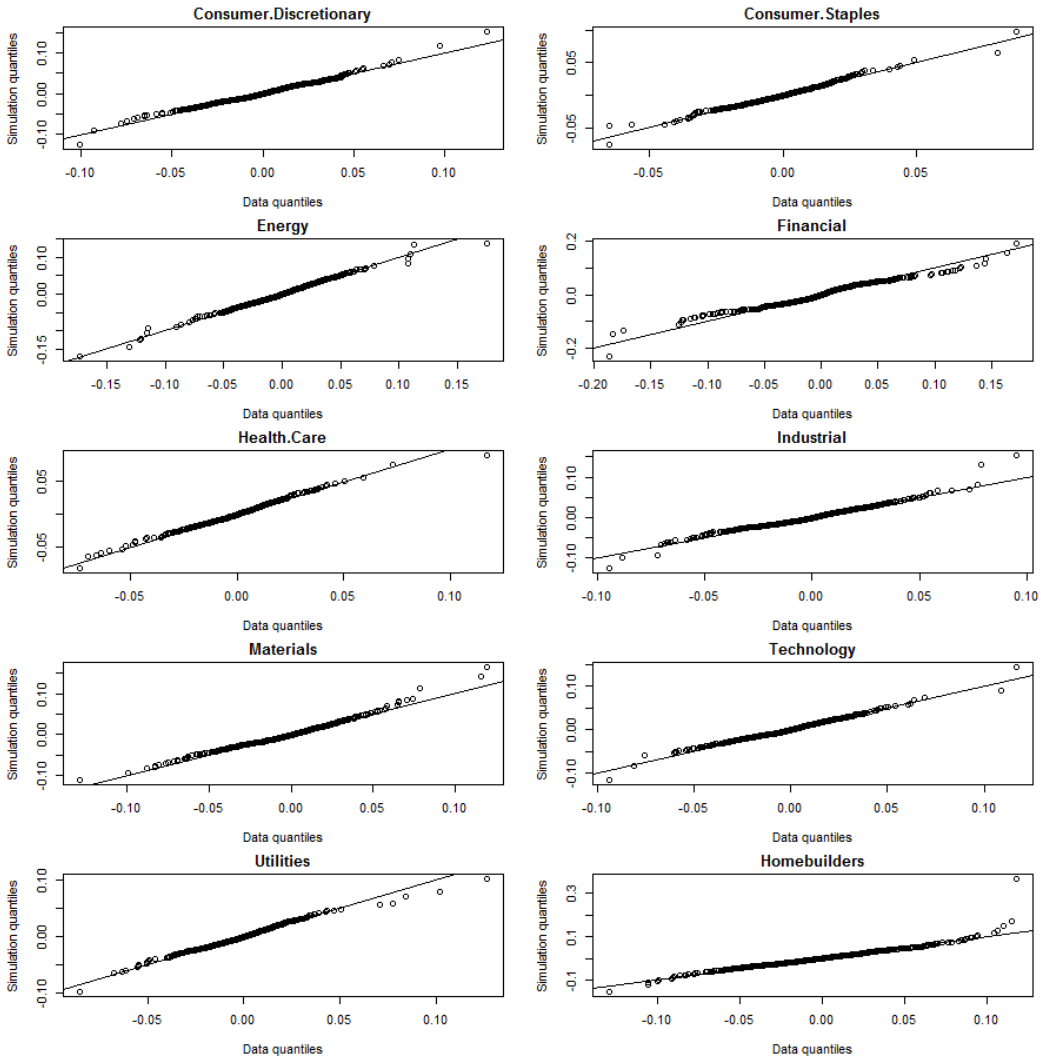


Figure 14 - Simulated vs training data QQ plot, Equity

Lastly, we look at QQ-plots comparing the simulated data against the training data. We see that the lines are fairly straight for all the assets, and conclude that the simulation algorithm is working as intended. This plot is from the t-copula, the Gaussian looks very similar, as we use the same marginal specification for both copulas. Therefore only one QQ-plot is needed for visualization.

We simulate data from both copulas, for both univariate GARCH specifications, and are therefore left with four sets of simulation data to do CVaR optimization on.

8.2.1.5 Comparing the strategies

Regular specification portfolios

Table 3 - Regular specification moments, Equity

	mean	sd	skew	ex. kurtosis
Equal Weight	2e-04	0.0092	-0.296	1.498
Markowitz	5e-04	0.008	-0.1682	1.3677
CVaR	5e-04	0.008	-0.1766	1.3618

As expected we see that both risk minimized portfolios have lower standard deviation. There are however no difference between the two methods of optimization for this metric.

Table 4 - Regular specification downside measures, Equity

	Equal Weight	Markowitz	CVaR
Downside Deviation (0%)	0.0066	0.0055	0.0055
Historical VaR (95%)	-0.0158	-0.0131	-0.0133
Historical ES (95%)	-0.0213	-0.0174	-0.0173

The downside risk measures are unsurprisingly all in favor of the risk minimized portfolios. We see however again very little difference between the two optimization procedures.

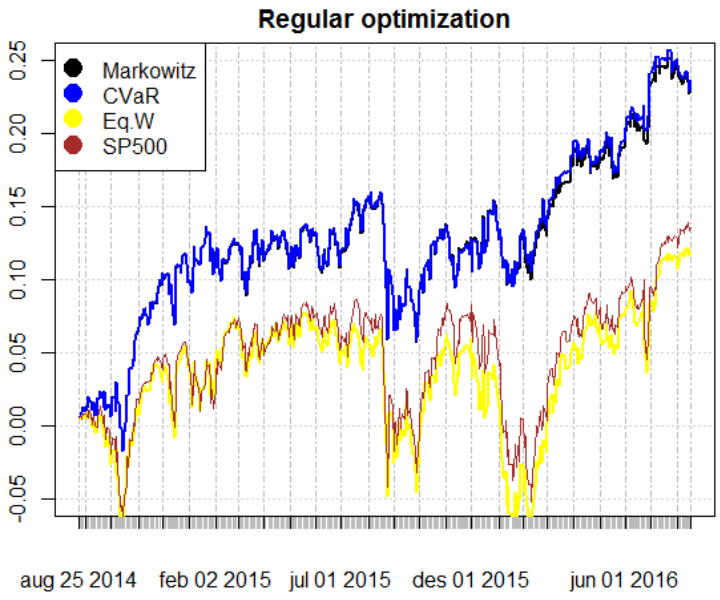


Figure 15 - Regular specification cum. log return, Equity

The cumulative returns graph displays the benefit of the risk optimal portfolios compared to the equal weight and S&P 500 benchmark. We see that the drops in cumulative returns for the optimized portfolios are much smaller overall, leading also to a higher overall return in the span of the testing sample.

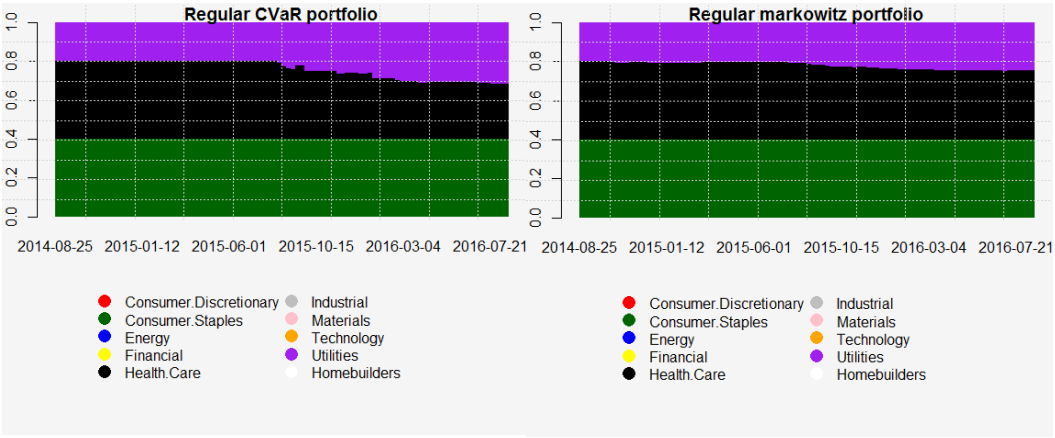


Figure 16 - Regular specification weighting, Equity

These plots display the weight development in the portfolio as time passes. The assets are given a color representation, and the weights are stacked at each time step, so that the sum of the weights is 1 on the y-axis. We see that for the optimal portfolios of both CVaR and Markowitz minimum variance, only three assets are represented through the entire testing sample. We note that the maximum allocation restriction is restricting Consumer Staples through the entire sample, and it is also restricting Health Care for much of the sample. However, the weight allocation is very similar for both procedures.

MV-GARCH portfolios

Table 5 - MV GARCH moments, Equity

	mean	sd	skew	ex. kurtosis
M.DCC sGarch port.	4e-04	0.0079	-0.2963	1.3922
M.DCC eGarch port.	3e-04	0.0078	-0.2743	1.4306
M.CCC-sGarch port	4e-04	0.008	-0.3148	1.3774

We see that MV-GARCH portfolio moments are similar to the moments for the regular optimization in Table 3. The standard deviation for the DCC portfolios are slightly lower than the CCC portfolio.

Table 6 - MV GARCH downside measures, Equity

	M.DCC sGarch port.	M.DCC eGarch port.	M.CCC-sGarch port
Downside Deviation (0%)	0.0056	0.0055	0.0057
Historical VaR (95%)	-0.0138	-0.013	-0.0142
Historical ES (95%)	-0.0178	-0.0178	-0.0181

The tail risk measures of these portfolios are similar to those of the portfolios optimized using the regular specifications, i.e. table 4. We see that the DCC portfolios are slightly less risky than the CCC portfolios for this sample, but by a very small margin.

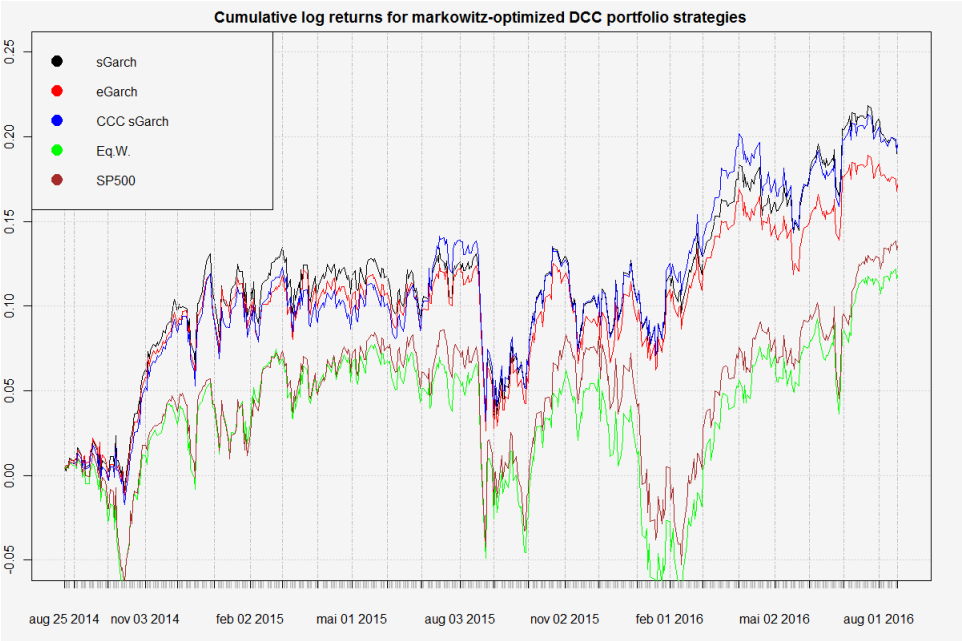


Figure 17 - MV-GARCH cum. log return, Equity

The plot for the cumulative returns looks similar to the corresponding plot for the regular specification, i.e. Figure 15. We see that the large drops from the S&P and the equal weight portfolio are avoided to some extent, but no large improvement from the regular optimization is spotted.

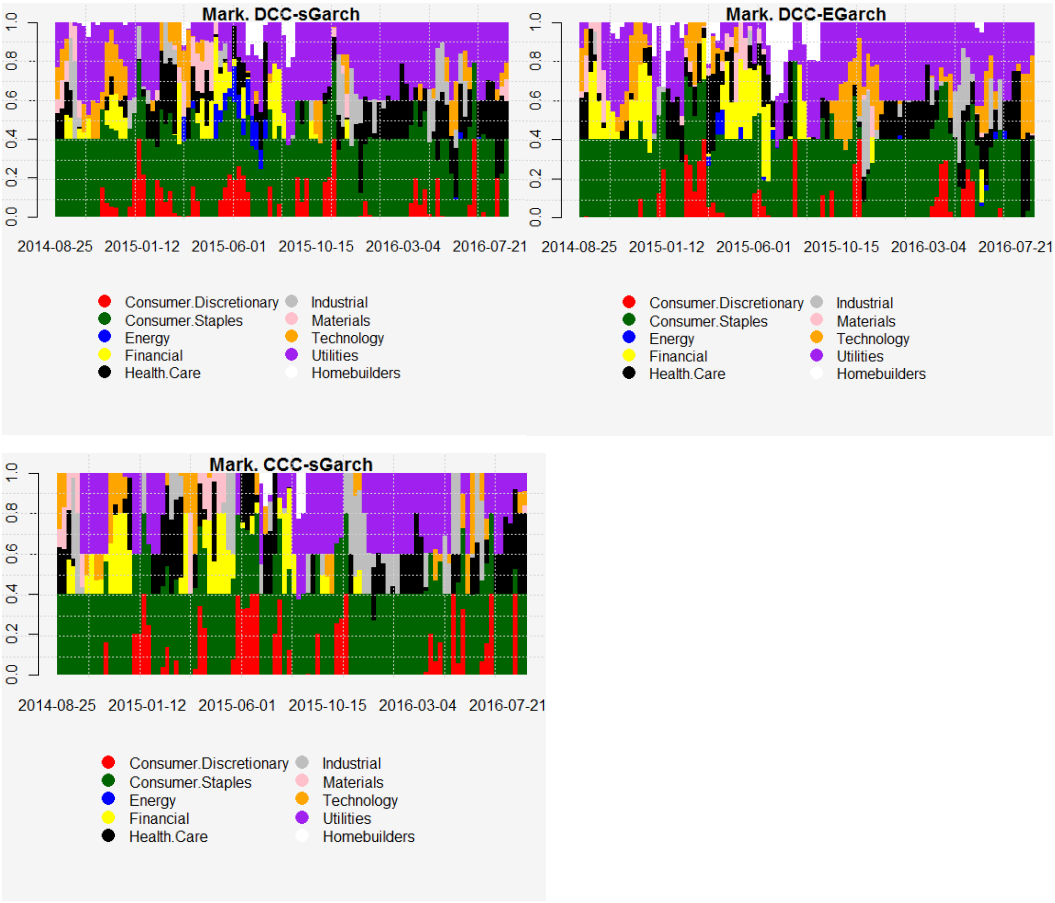


Figure 18 - MV-GARCH weighting, Equity

The weight plots for the MV-GARCH strategies suggest a more diversified portfolio than the regular specification, in the sense that more assets are represented. But this also requires more rebalancing to obtain compared to the regular portfolios, which has fairly stable weights. We see a large emphasis on Consumer Staples also in these optimizations.

Copula-GARCH portfolios

Table 7 - Copula-GARCH moments, Equity

	mean	sd	skew	ex. kurtosis
Normal copula sGarch	4e-04	0.0079	-0.3047	1.3949
t-copula sGarch	4e-04	0.0079	-0.253	1.3244
Normal copula eGarch	4e-04	0.0078	-0.307	1.42
t-copula eGarch	3e-04	0.0079	-0.2772	1.3578

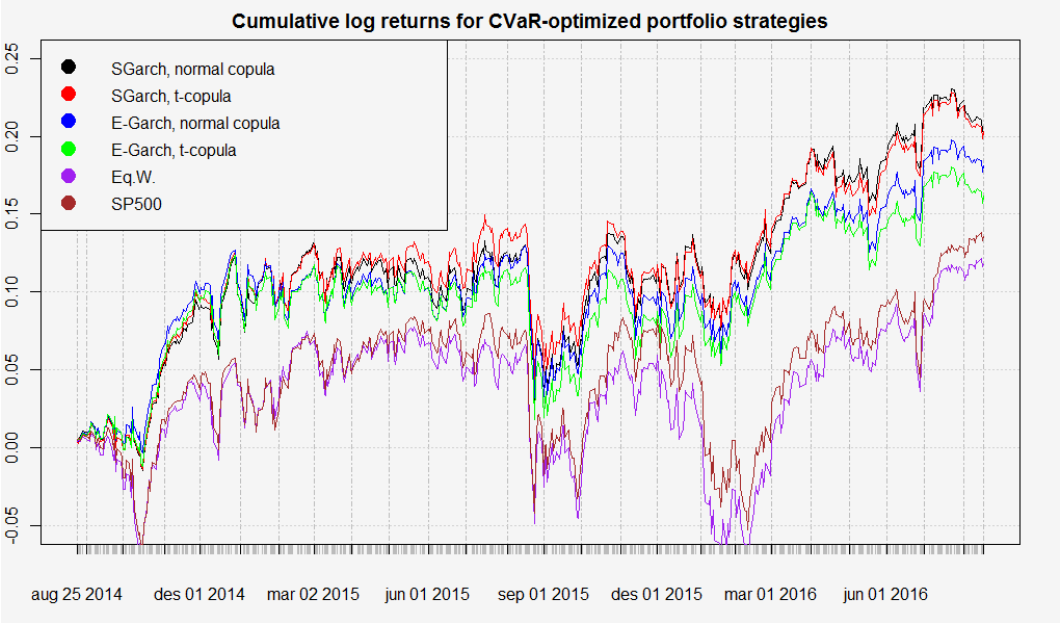
Again, we see similar results to the regular optimization in terms of standard deviation. Little difference between the t and Gaussian (Normal) copula are present in the moments.

Table 8 - Copula-GARCH downside measures, Equity

	Normal copula sGarch	t-copula sGarch	Normal copula eGarch	t-copula eGarch
Downside Deviation (0%)	0.0055	0.0055	0.0055	0.0055
Historical VaR (95%)	-0.0137	-0.0138	-0.0134	-0.0133
Historical ES (95%)	-0.0177	-0.0176	-0.0178	-0.0178

The tail risk metrics of these portfolios are again similar to that of the regular portfolio optimization procedure. As for the t-copula vs Gaussian copula, we cannot based on these findings suggest that one has an advantage over the other. They both perform in line with the regular portfolio optimizations done at the start of this section.

Table 9 - Copula-GARCH cum. log returns, equity



We see from the cumulative returns plot, that also these portfolios have an advantage over the naïve portfolio, and are similar in development to the regular portfolios.

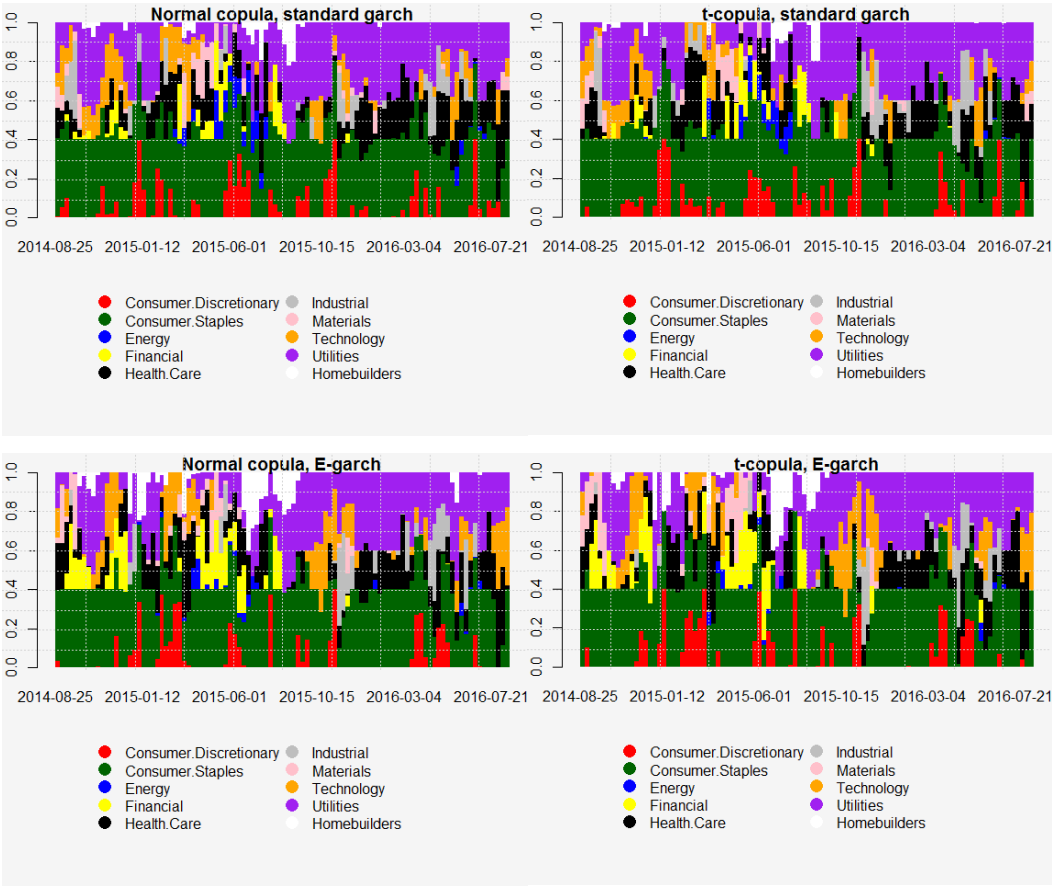


Figure 19 - Copula-GARCH weighting, Equity

As for the weight development they look similar to the MV-GARCH strategies. These portfolios have a higher degree of diversification, but unlike the regular portfolios they require much rebalancing in order to achieve the suggested weights.

Table 10 - Levene's (B-F) p values, Equity

	Regular CVaR	Regular Markowitz
CVaR DCC optim.		
sGarch, G-cop	0.760104861296857	0.711028923520854
sGarch, T-cop	0.860470183534243	0.809359507218278
eGarch, G-cop	0.667251484465646	0.620610366819076
eGarch, T-cop	0.72478440358063	0.676496812431844
Markowitz DCC optim		
sGarch-DCC	0.847688367426817	0.796804078644067
eGarch-DCC	0.655009714264267	0.608829295222316
sGarch-CCC	0.841218489986821	0.791135506620887

This table displays the p-values for the Brown–Forsythe test. The test compares the variance of two different distributions, where the null hypothesis is that the variances are equal. A low p-value would indicate that the variances are significantly different. The table compares each of the dynamic GARCH type strategies with both of the regular strategies. We see that for all the GARCH type strategies, the p-value is high. This indicates an insignificant difference in the standard deviation of the portfolio returns.

Using this sample, we have little evidence to suggest that the copula and GARCH frameworks improve upon portfolio optimization compared to the regular framework. We have located a plausible reason why these proposed improvements has little practical benefit. Our sample is in nature highly correlated as all the assets are part of the large cap index in the U.S. Additional unconditional correlation is present in the training sample due to the presence of the financial crisis. When all the correlations are this high, the diversification benefit from the portfolio optimizations is small, and the weights will be extremely tilted towards the assets with low volatility. This is mathematically sound, because when the correlation is this large, holding the asset with the lowest standard deviation is almost equivalent of holding the entire portfolio in terms of risk minimization. When the correlations between all assets are 1, the efficient frontier will be a straight line, and no diversification benefits are available. As correlations decrease, the efficient frontier will be shaped as a parabola, increasing in peakedness as correlation decrease.

When the optimizers allocate this high weights to a few assets, comparing the portfolios might not give much insight, as they are all very similar.

8.2.2 Commodity dataset

8.2.2.1 Initial examination

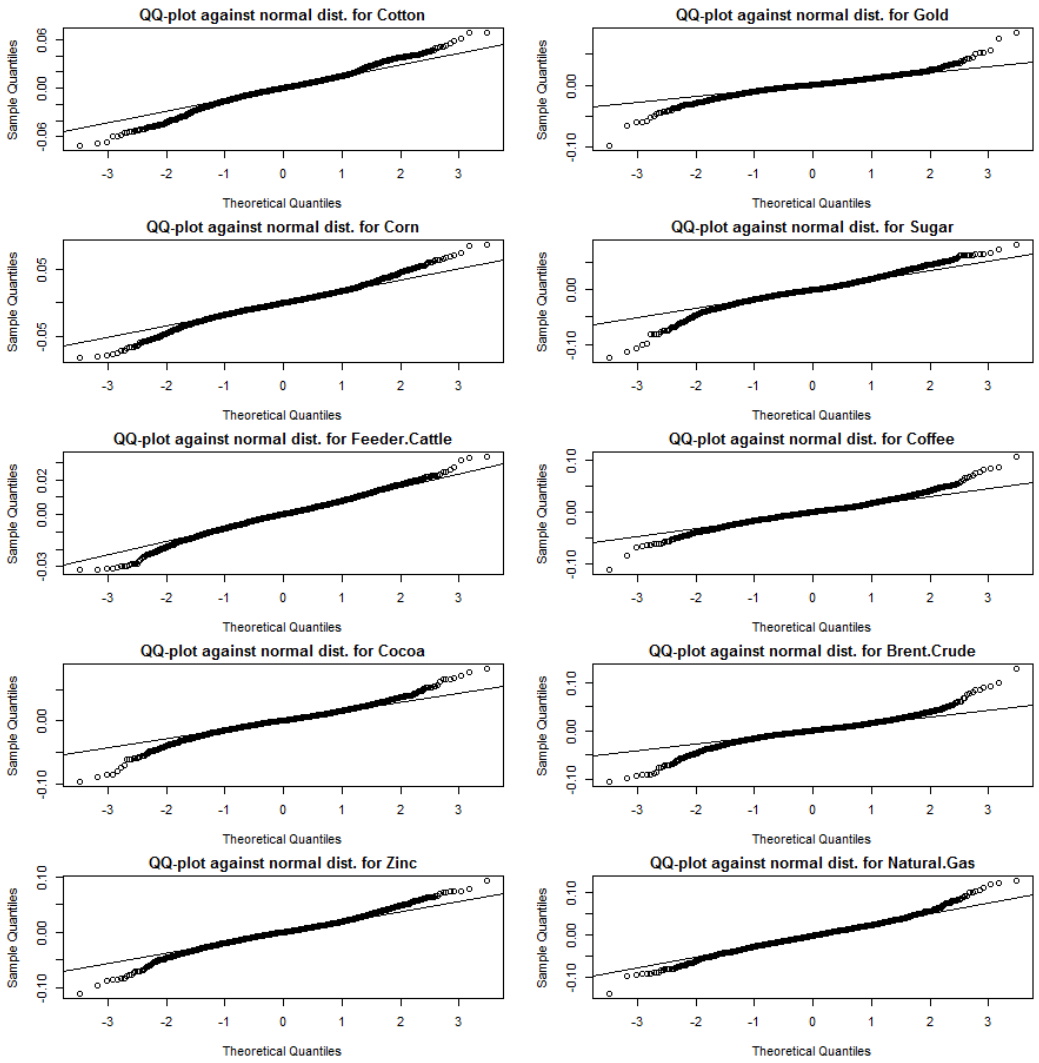


Figure 20 - Normal QQ plots, Commodities

These plots show similar characteristics as the equities, i.e. there is heavier tails in the data than what would be expected from a normally distributed variable.

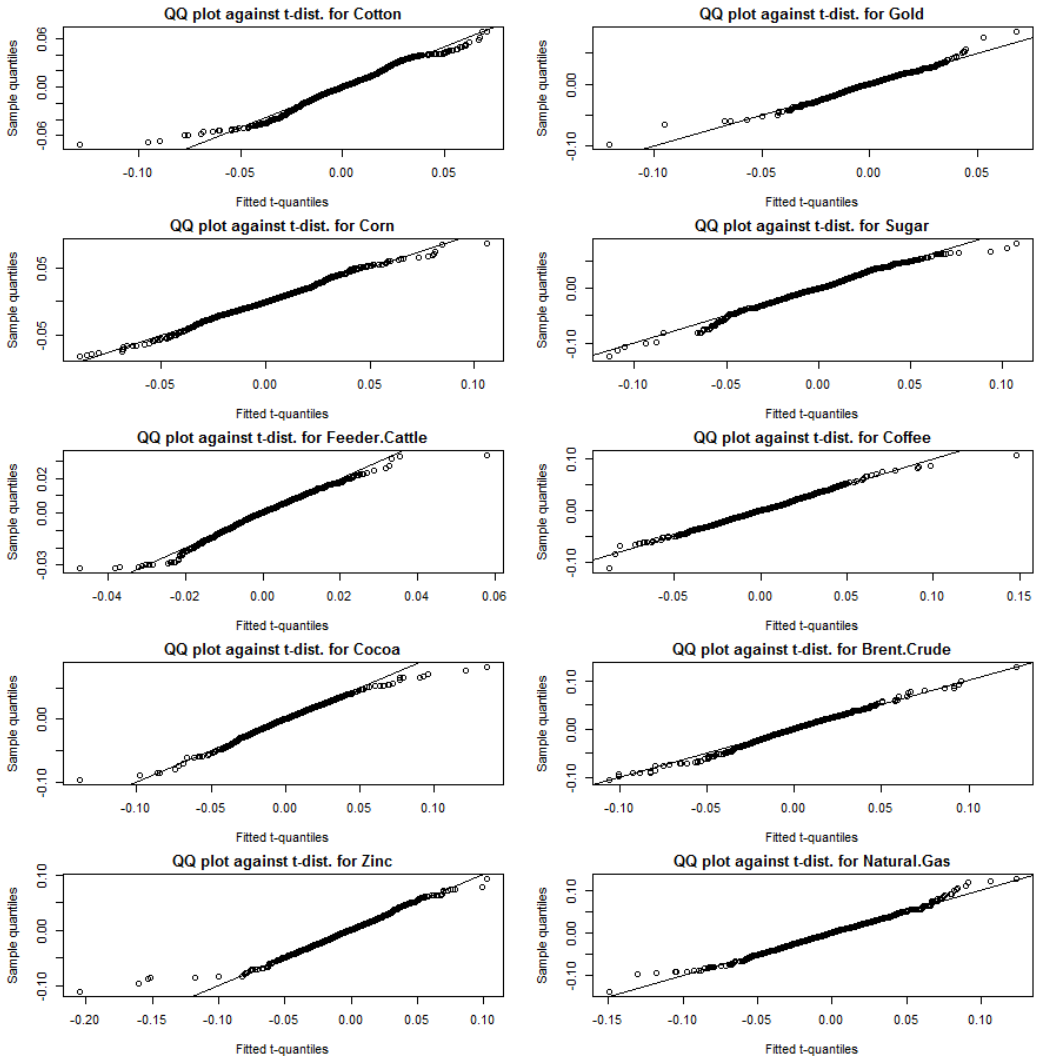


Figure 21 - Student t QQ plots, Commodities

The tail events are better captured by the t-distribution in general, as the dotted line follows the straight line more closely in the tails.

We again examined the ACFs of the distributions. As one would expect, the linear ACF was non-significant while the squared ACF was significant. These plots can be found in Appendix B2.

8.2.2.2 Fitting univariate GARCH the models

Two univariate GARCH specifications are fitted to this data set as well.
The ARMA parameters in the conditional mean equations turn up at times insignificant. However, as using ARMA to model the residuals are likely to make the residuals closer to being i.i.d. by removing any traces of linear autocorrelation, they are kept in the model. As for the S-GARCH parameters, all the alphas and betas are significant, while the omegas are typically insignificant. For the E-GARCH all parameters are significant, with the exception of the gamma parameter of Homebuilders.

Table 11 - Univariate GARCH(1,1) fit, Commodities

S-GARCH Model Fit

Optimal Parameters:

	Cotton	Gold	Corn	Sugar	Feeder.Cattle	Coffee	Cocoa	Brent.Crude	Zinc	Natural.Gas
mu	-0.00013	0.00041	0.00006	-0.00051	0.00017	-0.00030	0.00038	0.00031	0.00005	-0.00138
ar1	-0.09592	-0.93068	-0.79463	-0.68157	-0.22531	0.07706	0.30753	-0.40921	-0.80831	-0.58160
ma1	0.14998	0.94441	0.82729	0.70740	0.33057	-0.12387	-0.30308	0.37626	0.79149	0.52634
omega	0.00000	0.00000	0.00001	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00001
alpha1	0.05681	0.05168	0.05588	0.04477	0.03280	0.03604	0.02772	0.04677	0.02093	0.05684
beta1	0.93349	0.93420	0.92624	0.95263	0.95881	0.94527	0.97128	0.95029	0.97807	0.93016

Table 12 - Univariate E-GARCH(1,1) fit, Commodities

e-GARCH Model Fit

Optimal Parameters:

	Cotton	Gold	Corn	Sugar	Feeder.Cattle	Coffee	Cocoa	Brent.Crude	Zinc	Natural.Gas
mu	0.00008	0.00081	-0.00003	-0.00055	0.00011	-0.00006	0.00064	0.00033	-0.00006	-0.00148
ar1	-0.13024	0.70464	-0.82480	0.12317	-0.19774	-0.09747	-0.27884	-0.27334	-0.76275	-0.58662
ma1	0.17724	-0.73357	0.85491	-0.14025	0.29948	0.03296	0.26983	0.23120	0.74290	0.53274
omega	-0.07980	-0.05702	-0.11180	-0.03781	-0.16904	-0.12732	-0.00137	-0.03393	-0.00682	-0.11227
alpha1	0.00460	0.01097	-0.02416	0.01159	-0.04271	0.04024	0.00264	-0.03882	-0.01845	-0.00324
beta1	0.99023	0.99352	0.98588	0.99519	0.98239	0.98383	0.99987	0.99599	0.99926	0.98467
gamma1	0.13125	0.09917	0.13468	0.10010	0.07612	0.07508	0.06024	0.10223	0.04428	0.12409

We see from the parameters that all series are stationary as $alpha1 + beta1 < 1$ for the standard GARCH, and $beta1 < 1$ for the E-GARCH. From the large beta values we see that prior volatility has a high impact on forecasted volatility. This is true for both models.

As for the other E-GARCH parameters we again get some indications that the model might be inappropriate. Some of the alpha parameters are negative, which doesn't make sense, a shock should increase volatility increase, not decrease it. As for the gamma-parameters, they often have the same sign of the alphas. While the leverage effects per se can't exist in commodity returns (they can't take on debt), we still would expect the gammas to have the opposite sign of the alphas based on the stylized facts.

The volatility is less jumpy than that of equities. This can be seen from the lower alpha, and higher beta parameters.

We also here inspect the ACF's of the squared standardized residuals. If the GARCH specifications work as intended, we should now see little to no significant autocorrelation.

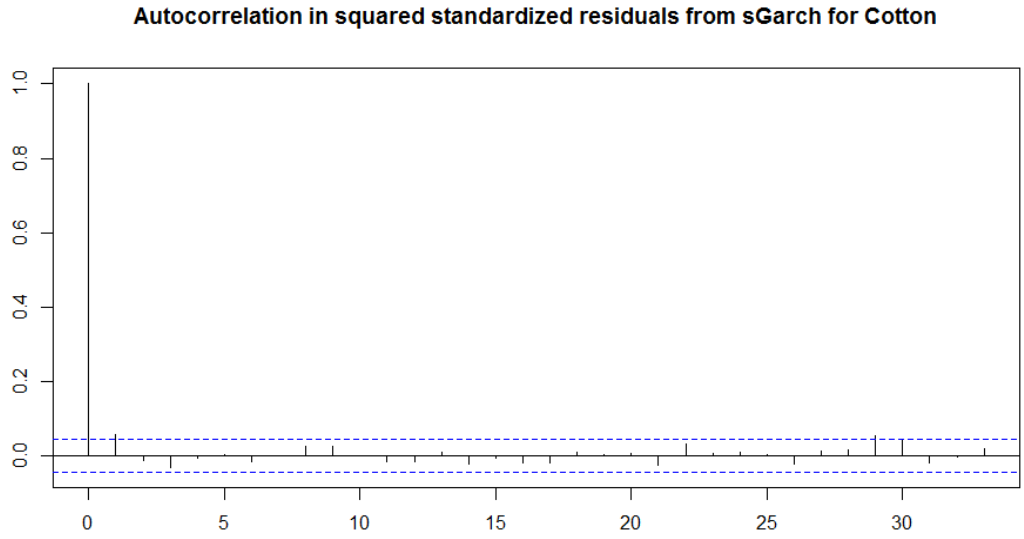


Figure 22 - ACF of squared standardized residuals for cotton.

The squared ACF's are largely improved upon by standardizing after GARCH, here illustrated by Cotton using the GARCH(1,1) specification.

8.2.2.3 Fitting the multivariate GARCH models

To fit the CCC models we use the unconditional sample correlation, in combination with the stand GARCH (1,1) models.

sGarch DCC for commodities

[Joint]dcca1	0.007034	0.001207	5.825965	0.000000
[Joint]dccb1	0.978917	0.005245	186.622440	0.000000

eGarch DCC for commodities

[Joint]dcca1	0.006805	0.001053	6.4611e+00	0.000000
[Joint]dccb1	0.981449	0.004057	2.4193e+02	0.000000

Figure 23 - DCC fit, Commodities

Here a1 and b1 is the α and β from section 5.2.2. The high b1 indicates that recent correlation has a large effect on the future correlation, while the low a1 indicates that shocks in correlation have less of an effect. Compared to the equity dataset we observe that commodities seem less "jumpy" or responsive to shocks in that the a1is even lower here.

8.2.2.4 Simulation diagnostics

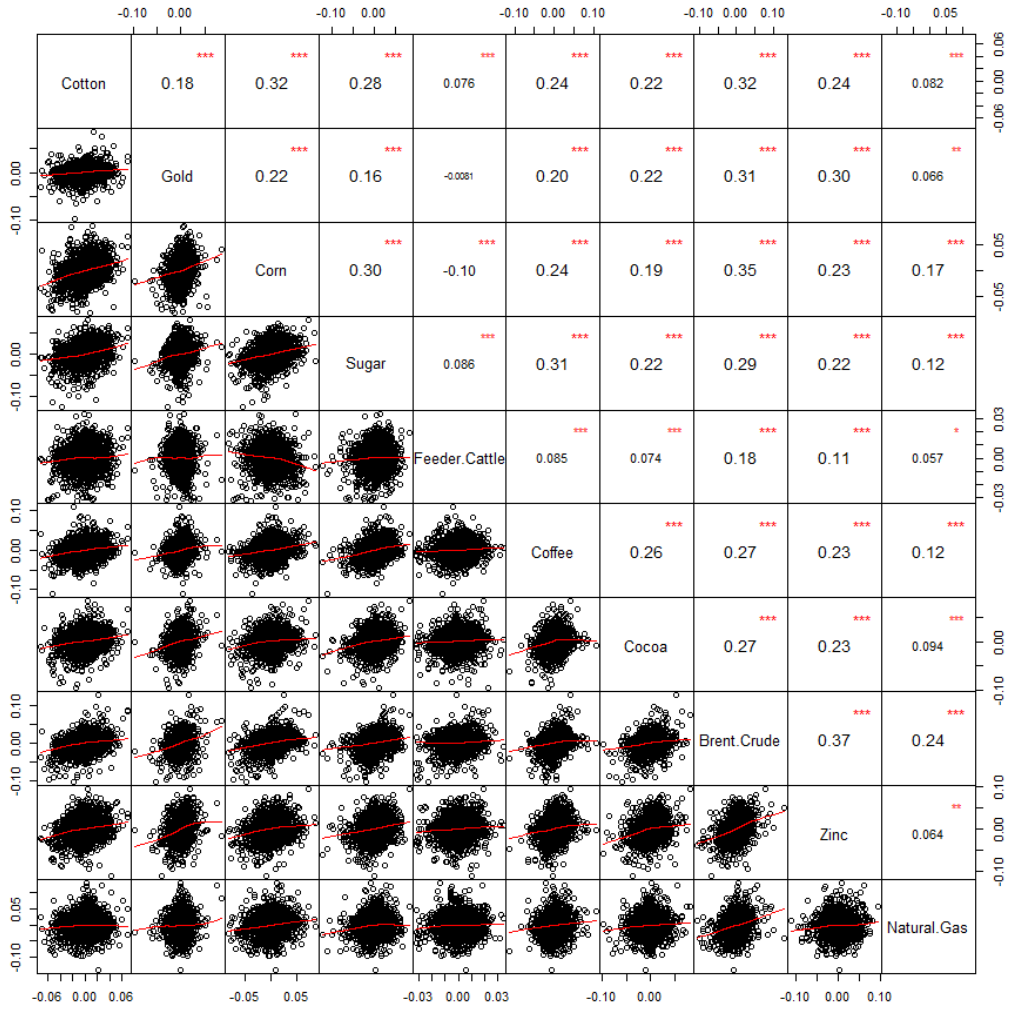


Figure 24 - Pairs plot training data, Commodities

From the pairs plot we see that this data set has much lower correlation than that of the equity data. We also see less tail dependence in this data set, from the more circular shape of the Pairs plots above, with a lower peak in the tails.

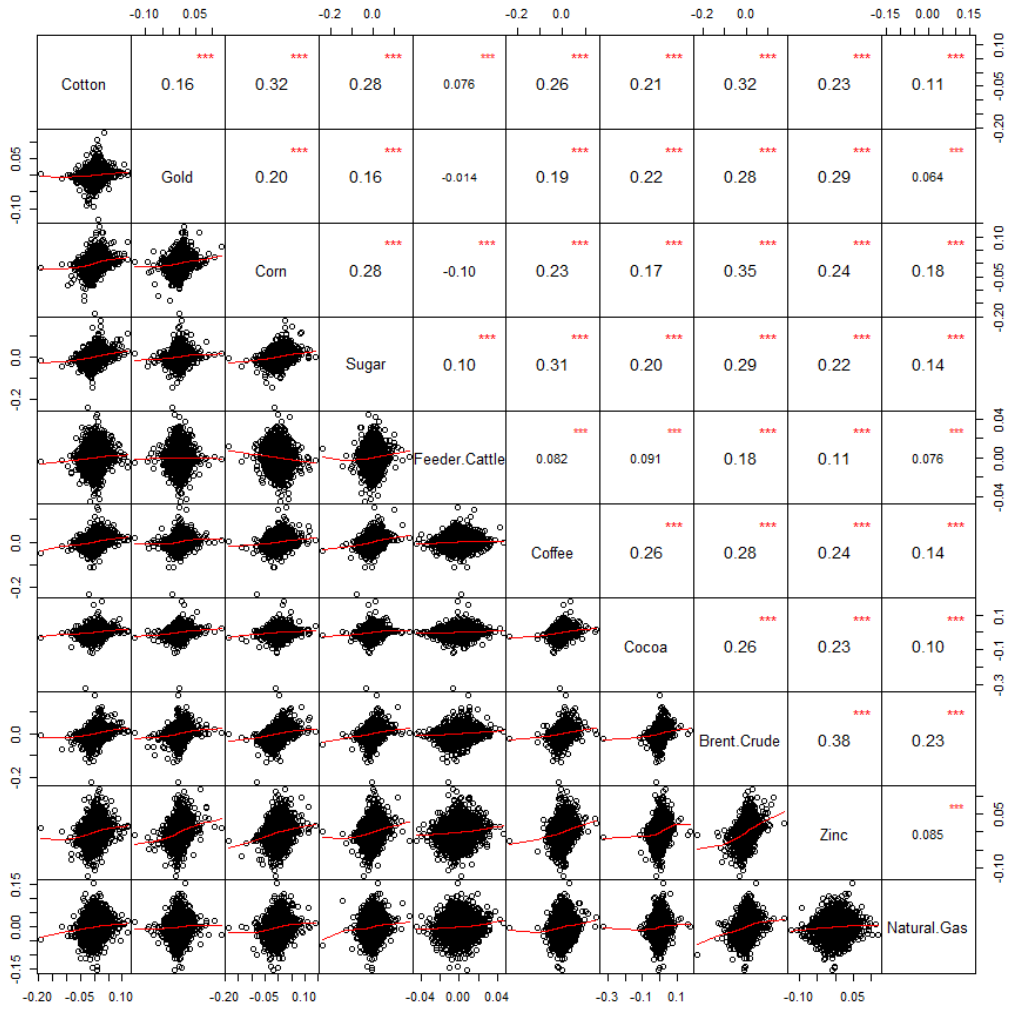


Figure 25 - Pairs plot G-SIM, Commodities

The above pairs plot shows the simulated data from the Gaussian copula fitted to the training sample for comparison purposes. We see that the fit is better than for the equities, but as expected, no tail dependence is present. This can be seen from the extreme data points tend to be positioned parallel to the axes instead of in the corners.

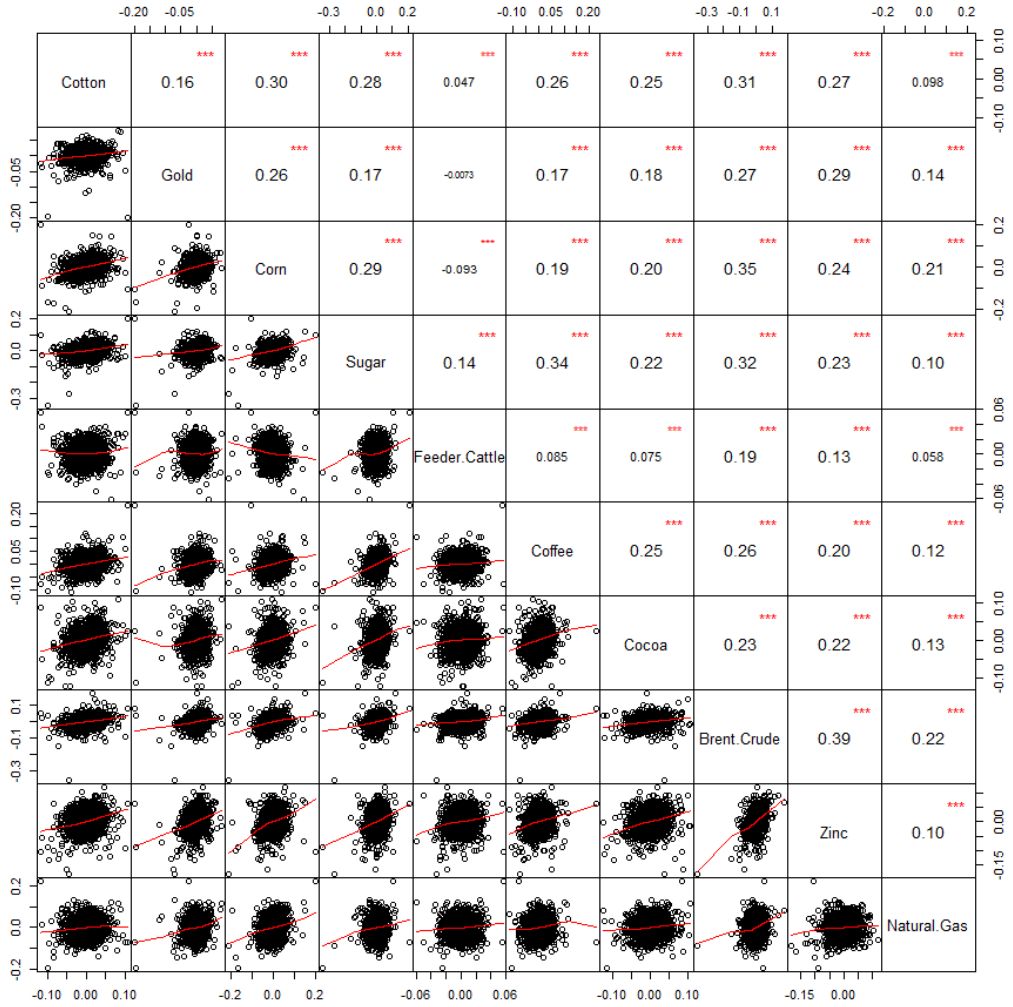


Figure 26 - Pairs plots t-SIM, Commodities

To get tail dependence, we simulate using the t-copula. The shapes of the plots are circular, indicating the lower levels of correlation. We see that the shape of the pairs look similar, but more concentrated than the data. The tail dependence is less pronounced than for the equity sample, but we do see a tendency of extreme events to be located in the bottom left and top right corners of the plot.

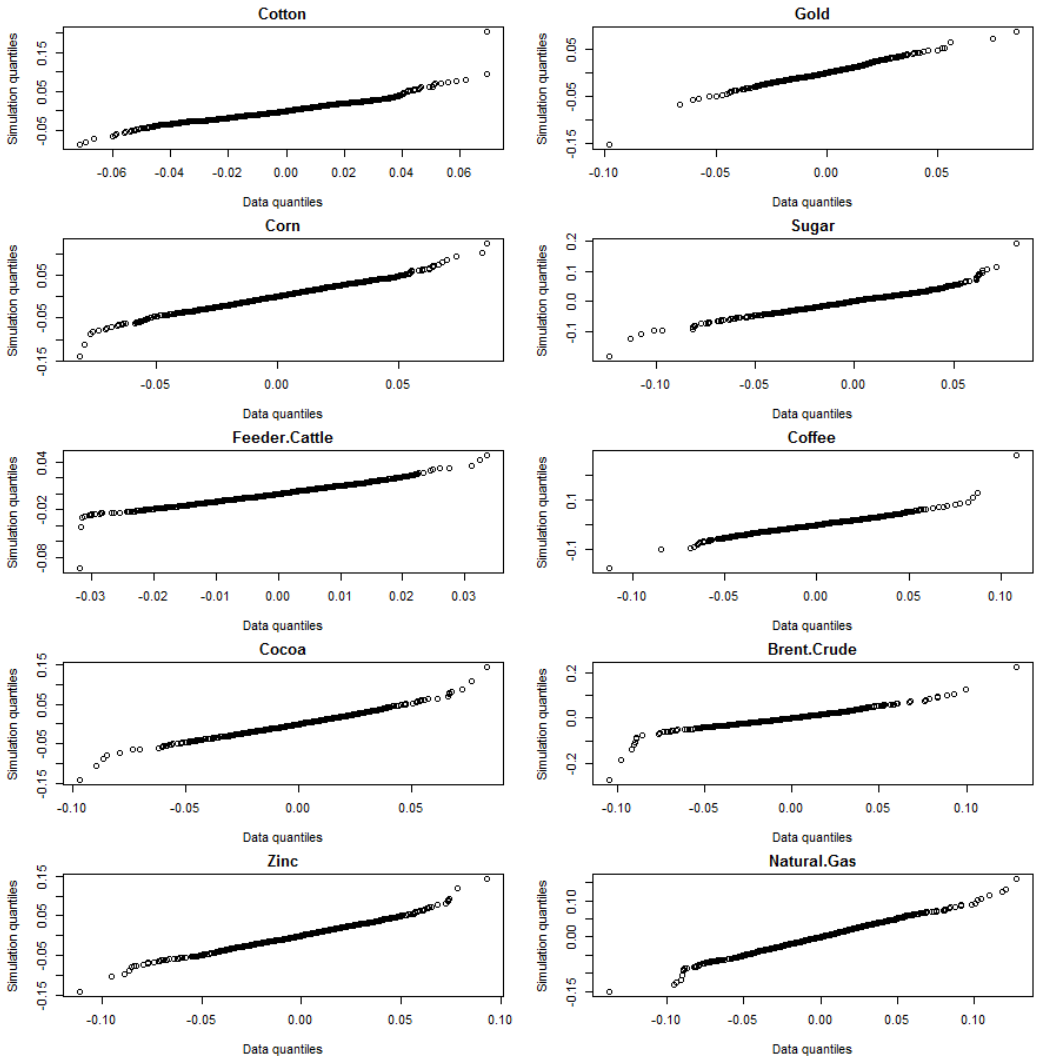


Figure 27 - Simulation QQ plot, Commodities

To check the marginal distribution fit, we look at QQ-plots of the simulated data compared to the training data. There seems to be a slight tendency for the simulations to have more extreme returns than the training sample, but only slight. We conclude that the simulation is reasonably effective at capturing the desired characteristics of the data.

8.2.2.5 Comparing the strategies

Regular specification portfolios

Table 13 - Regular specifications moments, Commodities

	mean	sd	skew	ex. kurtosis
Equal Weight	-6e-04	0.0078	0.0126	0.0416
Markowitz	-3e-04	0.0062	0.0736	0.0552
CVaR	-3e-04	0.0064	0.0971	0.0788

The standard deviation from the regular portfolio optimizations indicates a lower risk than that of the equally weighted portfolio, as would be expected.

Table 14 - Regular specifications downside measures, Commodities

	Equal Weight	Markowitz	CVaR
Downside Deviation (0%)	0.0058	0.0046	0.0047
Historical VaR (95%)	-0.0127	-0.0104	-0.0108
Historical ES (95%)	-0.0165	-0.0133	-0.0136

The same is present in the downside risk of the regular portfolios vs the equally weighted.

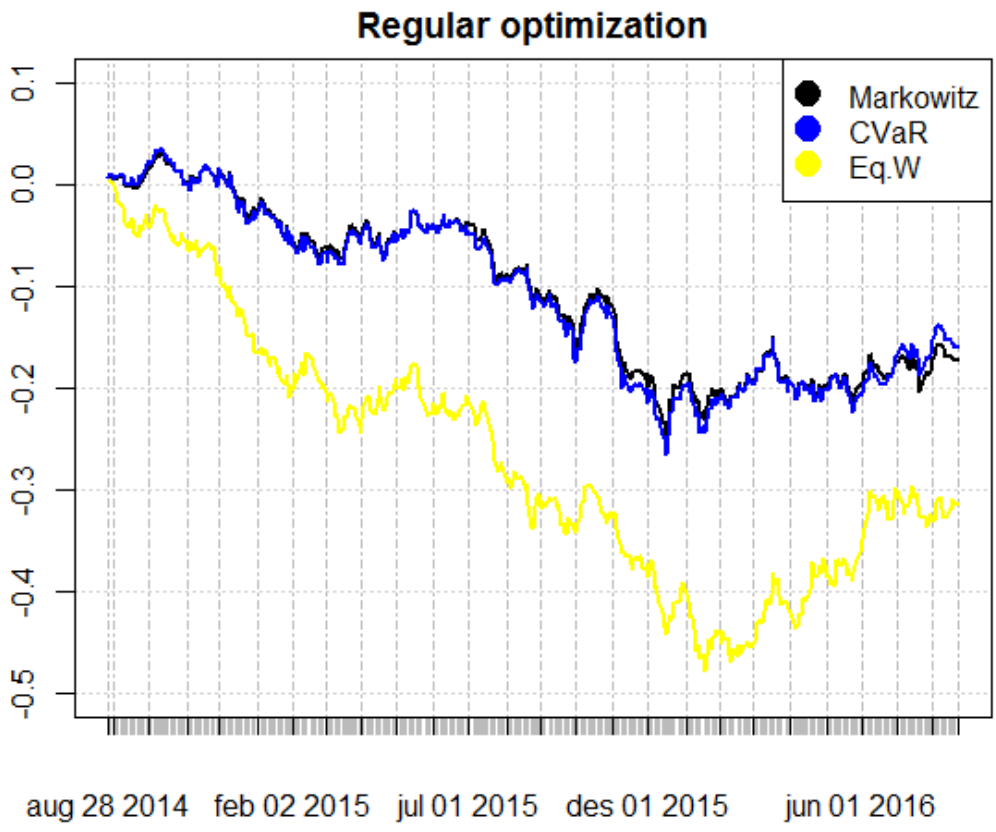


Figure 28 - Regular specifications cum. log return, Commodities

We see from the cumulative returns plot that both risk optimized portfolios has lower losses than the equally weighted. It is also evident that the two optimization procedures are very similar for this dataset as well, as they follow each other closely.

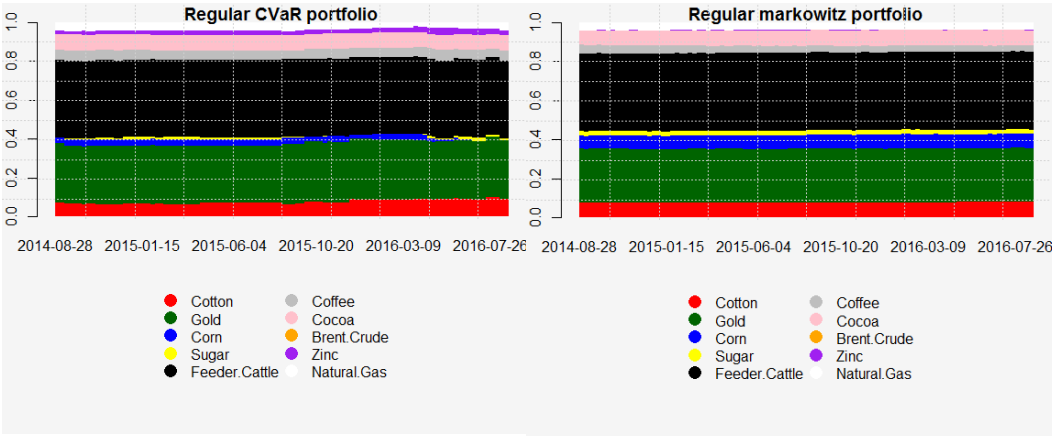


Figure 29 - Regular specifications weighting, Commodities

From the weight development plots, we see that both portfolios are dominated by the same assets. We see more diversification in these plots in comparison with the equities for the same optimization procedures. This is expected, as the correlation between these assets are a lot lower, and thus provide more diversification benefit.

MV-GARCH portfolios

	mean	sd	skew	ex. kurtosis
M.DCC sGarch port.	-3e-04	0.0056	-0.0484	-0.2029
M.DCC eGarch port.	-4e-04	0.0056	-0.0378	-0.1954
M.CCC-sGarch port	-3e-04	0.0059	-5e-04	-0.0363

Figure 30- MV-GARCH moments, Commodities

We see that the standard deviation from the different MV-GARCH portfolio optimization strategies seems to be improved compared to the regular portfolios in table 13. This is a promising result, and is further investigated by the Brown–Forsythe test later.

	M.DCC sGarch port.	M.DCC eGarch port.	M.CCC-sGarch port
Downside Deviation (0%)	0.0041	0.0042	0.0043
Historical VaR (95%)	-0.0091	-0.0093	-0.0099
Historical ES (95%)	-0.012	-0.0119	-0.0127

Figure 31 - MV-GARCH downside measures, Commodities

The downside risk metrics for the DCC portfolio looks low compared to both the CCC-portfolio and the regular portfolios. All metrics favor the DCC portfolios,

but no conclusive results are present for the different univariate GARCH specifications. S-GARCH and E-GARCH perform equally well in combination with the DCC framework.

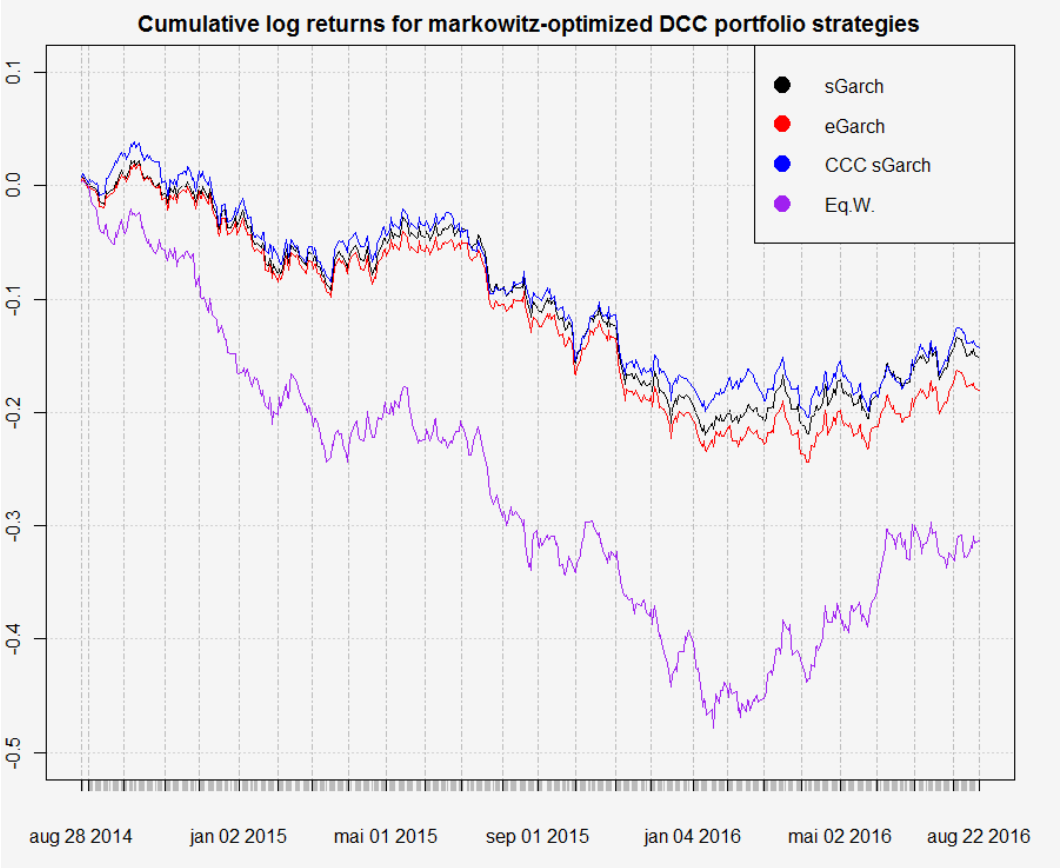


Figure 32 - MV-GARCH cum. log return, Commodities

The cumulative returns plot for all the strategies looks well in terms of loss-avoidance compared to the equally weighted portfolio. As for the different strategies, they look similar in performance and move closely together.

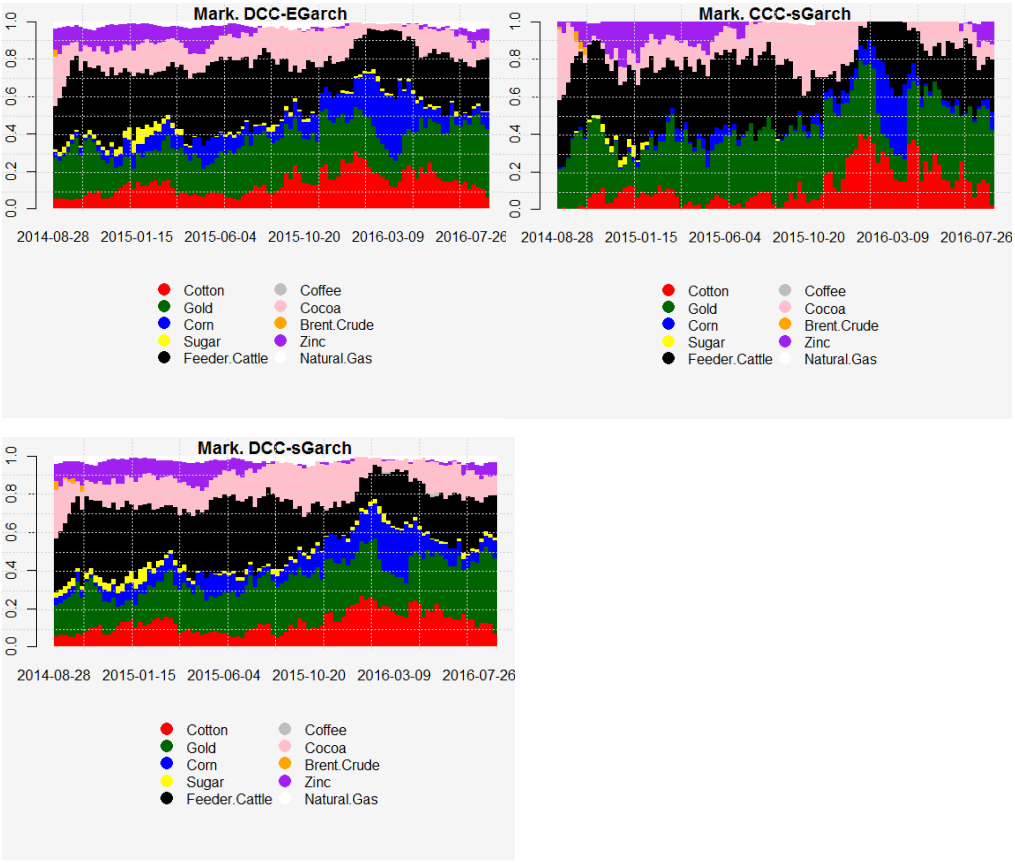


Figure 33 - MV-GARCH weighting, Commodities

The weight development for the different strategies look similar. The same assets are in prioritized, as would be expected, and the weighs are fairly evenly distributed among the active assets. This is good for diversification. As for rebalancing purposes, the weights are fairly stable. This is beneficial for potential rebalancing costs.

Copula-GARCH portfolios

Table 15 - Copula-GARCH moments, commodities

	mean	sd	skew	ex. kurtosis
Normal copula sGarch	-3e-04	0.0056	-0.0189	-0.0774
t-copula sGarch	-3e-04	0.0057	-0.0748	-0.1902
Normal copula eGarch	-4e-04	0.0056	-0.0276	-0.1579
t-copula eGarch	-4e-04	0.0057	-0.0081	-0.0403

CVaR optimization in combination with copula simulated data show similar standard deviations to the MV-GARCH procedures presented above. They too, are low compared to the regular portfolio optimizations. As for difference between the copulas, it is miniscule and does not give any clear indications of one outperforming the other.

Table 16 - Copula-GARCH downside measures, Commodities

	Normal copula sGarch	t-copula sGarch	Normal copula eGarch	t-copula eGarch
Downside Deviation (0%)	0.0041	0.0042	0.0042	0.0042
Historical VaR (95%)	-0.0092	-0.0095	-0.0093	-0.0095
Historical ES (95%)	-0.012	-0.0122	-0.0119	-0.0122

The tail risk measures are also similar to that of the MV-GARCH portfolios presented previously. Neither here is there any indication of benefit of using the t-copula over the normal copula. The Normal copula is actually performing better than the t-copula on both standard deviation and downside risk. However, this difference is very small, and cannot be viewed as a clear indication of outperformance. As for the GARCH specifications, S-GARCH and E-GARCH have close to equal performance on all metrics.

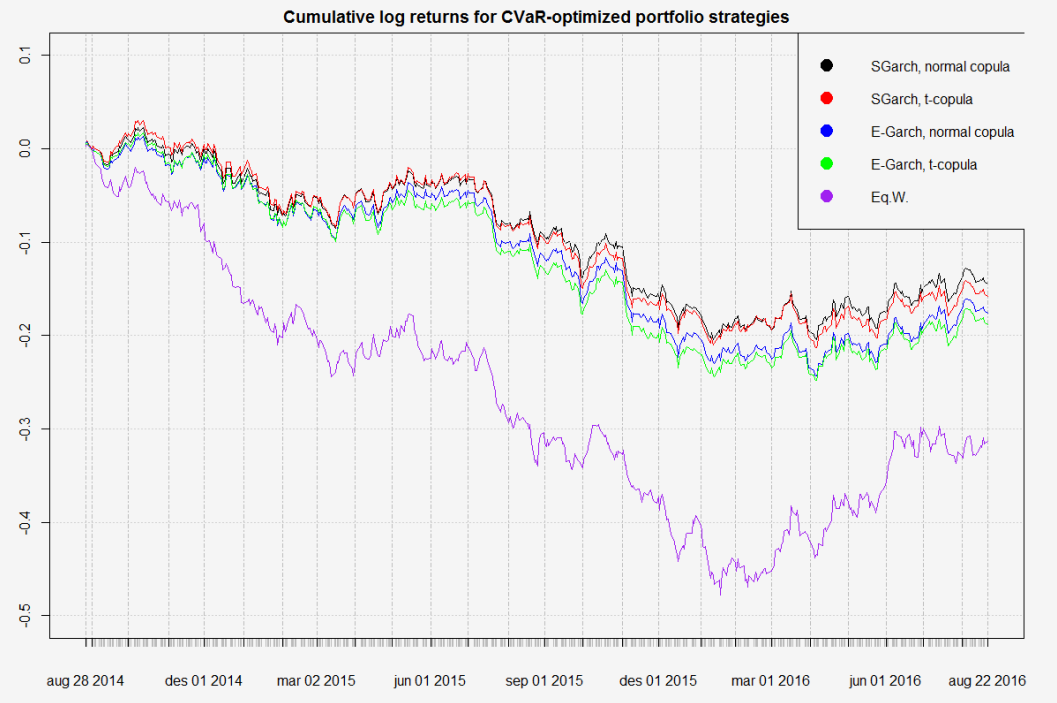


Figure 34 - Copula-GARCH cum. log return, Commodities

The cumulative plot looks similar to that of the MV-GARCH optimization, with no clear advantage to either strategy; although we see that both the S-GARCH-specifications are above the others for most of the period. The difference is very small from the between the MV-GARCH strategies, and they follow each other closely.

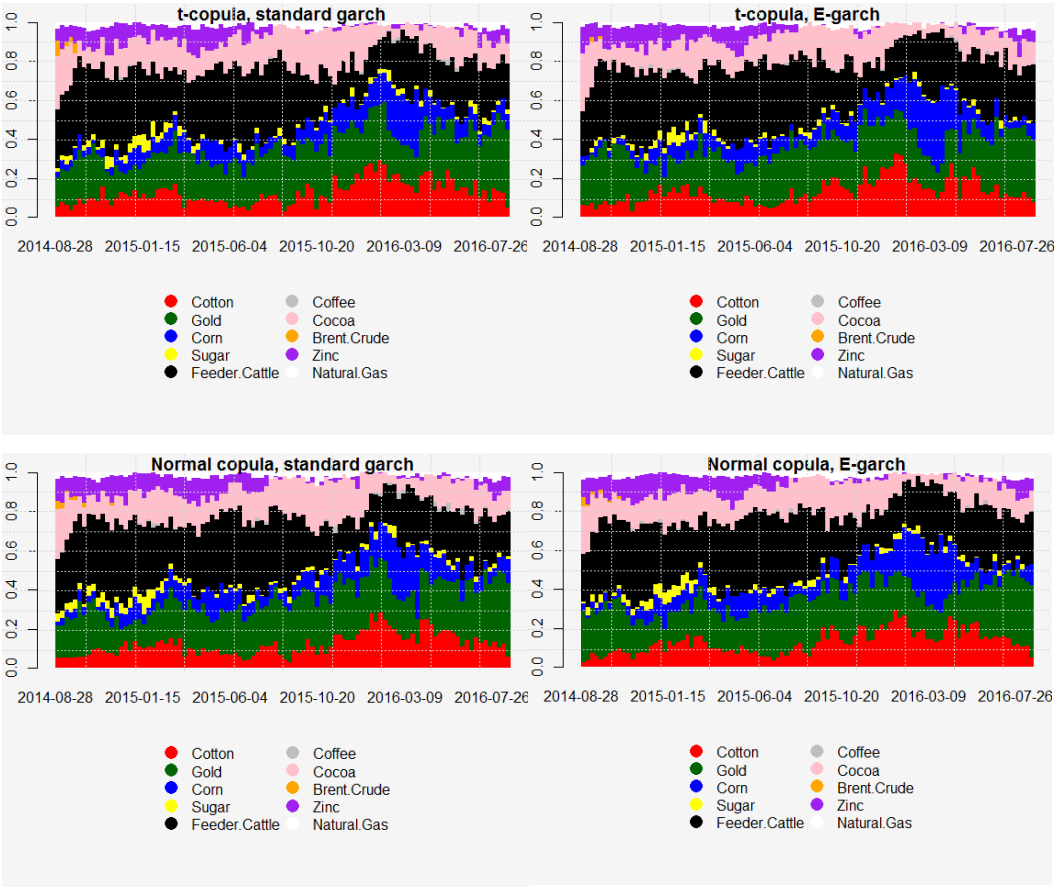


Figure 35 - Copula-GARCH weighting, Commodities

As for the weight development, the plots looks very similar to each other, and to the MV-GARCH portfolios presented previously. We see the same assets dominating the portfolios, and we also see a larger degree of diversification here, compared to the regular portfolios. Comparing it to the equity-portfolios, we see a larger degree of diversification here, and the restriction on asset weights are less constraining, as the natural diversification benefit is larger for this dataset.

We want to examine the standard deviation of the portfolios more closely using the Brown–Forsythe test. The p-values are presented in the following table.

Table 17 - Levene's (B-F) p values, Commodities

	Regular CVaR	Regular Markowitz
CVaR DCC optim.		
sGarch, G-cop	0.0167730568701707	0.0459345726009979
sGarch, T-cop	0.0425091501440481	0.104607348042707
eGarch, G-cop	0.0160391208295015	0.0444757108856151
eGarch, T-cop	0.0253052452445013	0.0661013664133644
Markowitz DCC optim		
sGarch-DCC	0.0189980334404685	0.0517862296062437
eGarch-DCC	0.0169747410496942	0.0468956584399867
sGarch-CCC	0.11123783806985	0.233817238849495

The test compares two sample variances. A low p-value indicates that the difference between the variances is large. In the table, we have compared the regular portfolios to each of the Copula-GARCH and MV-GARCH portfolios to see if there is some additional benefit in terms of reduced variance. We see that the differences for the commodities sample are a lot more significant than that of equities. Comparing to the regular CVaR optimization, we see that the only portfolio that is not significantly different at 95% confidence is the S-GARCH-CCC. As for the regular Markowitz optimization, two of the portfolio strategies do not have significantly different variance at the 90% confidence level, and three strategies have significantly different variance at the 95% confidence level. These values are indicative of improved risk minimization for the more advanced models, compared to the regular portfolio optimization.

Section 9: Concluding Remarks

In this thesis we have reviewed and discussed what constitutes a proper risk measure for portfolio optimization. We have reviewed the stylized facts of the empirical distribution of financial returns, and discussed the failure of the normal distribution and i.i.d. assumption in modelling this distribution.

We examine the use of GARCH, and Copula-GARCH models for providing estimates for the Markowitz and Min-CVaR algorithm, respectively. We find empirical indications suggesting that when in an investment universe with low correlation (such as a commodity portfolio); they are able to significantly outperform the sample counterparts. Additionally, it seems like the DCC-GARCH outperforms the CCC-GARCH in terms of low risk for this sample. We find no evidence suggesting that using the Student t copula outperforms the Gaussian for CVaR simulation, nor that the E-GARCH with t-distributed errors outperforms the standard GARCH with normally distributed errors in forecasting accuracy.

The equity sample does not significantly favor any of the GARCH or Copula-GARCH model estimates over the sample counterparts. A plausible reason for this is that the sample consists of highly correlated assets, which gives little freedom to the optimizers in portfolio selection. Thus, the portfolios are generally severely dominated by the low-risk alternatives, providing little chance for diversification.

We find that it doesn't make an economically meaningful difference whether we optimize with respect to Conditional Value-at-Risk or Variance. This is surprising for various reasons;
According to theory this should make a difference when returns are non-normal, and as we have demonstrated, both our samples are far from normal.
Additionally, results presented in related papers (e.g. Krokhmal et al 2002) find that minimizing CVaR results in a portfolio with lower CVaR and higher variance than when minimizing Variance, as one would expect.

We have tried to come up with some possible explanations as for why:

1. The difference may come down to sample specifics. If some assets in the dataset dominate the others in terms of low risk-levels, they will be

heavily favored in the asset allocation regardless of risk measure used for optimization. This may delude the interpretable results, as the differences between the models become less significant. We see some evidence that this is improved upon in the commodity sample over the equity sample.

2. The CVaR approaches considered in this thesis uses elliptical copulas, so simulated tail dependence is symmetrical. Differences between CVaR and Variance optimization should occur when tails are fat or when the distribution is skewed, our simulation only consider fat tails.
3. Krokhmal et al (2002) minimizes CVaR and Variance when the expected return is constrained to a certain goal. It is not obvious that this should make a difference, but it is a difference in methodology compared to our thesis.
4. Model specifications, such as choice of alpha-level for the CVaR, and copula parameters for the simulation algorithm may have impact on the result.

9.1 Suggestions for future work

There are many possible directions for future work. For the DCC-type models one might consider an ADCC (Asymmetric Dynamic Conditional Correlation) specification, allowing the model to take into consideration the empirical observation that equity correlations increase in times of recession (see section 2).

For applications to larger datasets (e.g. 100 assets rather than 10) the cDCC specification might be considered, as there is evidence suggesting that the cDCC specification outperforms the regular DCC specification when n get high (Aielli 2013). Alternatively (or additionally!) the CVaR and VaR measurements might involve risk mapping. While this thesis takes the approach of estimating CVaR and VaR by looking at the multivariate joint distribution, risk mapping allows for each risk factor (asset) to be studied alone, and then combined. This is generally necessary when the number of assets get very high. For an excellent introduction, see Alexander (2008c).

As for the copula specifications, one might consider vine copulas, and copulas from the Archimedean family. Vine copulas are considered more flexible than the

copula specifications used in this thesis, while copulas from the Archimedean family might be better suited for capturing asymmetry.

In addition more sophisticated univariate models might give better results. A word of caution is however needed here, theoretically superior models often fail to beat the simple GARCH (1,1) in out of sample forecasts, as demonstrated by Hansen and Lunde (2005) and illustrated by the disappointing results of the E-GARCH in this thesis. Specifications considering volatility spillover effects might be considered if one deal with a multi-market dataset, e.g. the US equity market and Norwegian equity market. One can also consider other error distributions than the normal and the Student's t distribution. For instance, a skewed Student's t distribution should in theory allow for both modeling fat tails and asymmetry, without needing to specify any measures of asymmetry in the conditional volatility equation.

A final suggestion is altering the methodology for (perhaps) more practically relevant results. For instance the thesis ignores transaction costs, and assumes that rebalancing weekly is reasonable. Krokhmal et al (2002) discuss how one can incorporate transaction costs as an additional constraint in the optimization problem, while Mendes and Marques discuss rebalancing strategies more generally. An idea could be to only rebalance assets if the new weights differ significantly from before. If this results in more infrequent balancing, correct multistep predictions for the dynamic conditional covariance matrix become increasingly more important (Hlouskova et al [2009]). One should also probably consider two-component GARCH-type models, as these are more likely to handle the persistence better for volatility predictions long into the future (see section 7 and 8 of Zivot [2009]).

Bibliography

Acerbi, C. and D. Tasche. 2002a. "On the coherence of Expected Shortfall." *Journal of Banking & Finance* 26(7): 1487-1504.

---------. 2002b. "Expected Shortfall: A natural coherent Alternative to Value at Risk." *Economic Notes* 31(2): 379-388.

Aielli, G.P. 2013. "Dynamic Conditional Correlation: On Properties and Estimation." *Journal of Business & Economic Statistics* 31(3): 282-299.

Alexander C. 2008a. *Market Risk Analysis Volume I: Quantitative Methods in Finance*. West Sussex, England: John Wiley & Sons Ltd.

---------. 2008b. *Market Risk Analysis Volume II: Practical Financial Econometrics*. West Sussex, England: John Wiley & Sons Ltd.

---------. 2008c. *Market Risk Analysis Volume IV: Value-at-Risk Models*. West Sussex, England: John Wiley & Sons Ltd.

Artzner P., F. Delbaen, J.M. Eber, D. Heath. 1998. "Coherent Measures of Risk." *Mathematical Finance* 9(3): 203-228.

Berkelaar, M. and others. 2015. "lpSolve: Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs. R package version 5.6.13."
https://CRAN.R-project.org/package=lpSolve

Black, F. 1976. "Studies of stock price volatility changes, Proceedings of the 1976 Meetings of the American Statistical Association." *Business and Economics: Statistics Section* 177-181.

Bodie, Z., A. Kane and A. J. Marcus. 2014. *Investments*. 10th ed. NY: McGraw-Hill Education.

Bollerslev, T. 1987. "A Conditional Heteroskedastic Time Series Model for Speculative Prices Rates of Return." *Review of Economics and Statistics* 69(3): 542-547.

---------. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31(3): 307-327.

---------. 1990. "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model." *Review of Economics and Statistics* 72(3): 498–505.

Bollerslev, T., R.F. Engle. and J.M. Wooldridge. 1988. "A capital asset pricing

model with time-varying covariances."
The Journal of Political Economy 96(1): 119–138.

Brooks, C. 2008. *Introductory Econometrics for Finance*. 2nd ed.
NY:Cambridge University Press.

Brown, M.B. and A.B. Forsythe. 1974. "Robust tests for the Equality of Variances." *Journal of the American Statistical Association*. 69(346): 364-367.

Chicheportiche, R. and J.P. Bouchaud. 2012. "The joint distribution of stock returns is not elliptical." *International Journal of Theoretical and Applied Finance* 15(3): 1250019-1-1250019-23.

Christie, A. A. 1982. "The stochastic behavior of common stock variances: Value, leverage, and interest rate effects."
Journal of Financial Economics 10(4): 407-432.

Cont R. 2001. "Empirical properties of asset returns: stylized facts and statistical issues." *Quantitative Finance* 1(2): 223-236.

Delignette-Muller, M.L. and C. Dutang. 2015. fitdistrplus: "An R Package for Fitting Distributions." *Journal of Statistical Software* 64(4): 1-34.
http://www.jstatsoft.org/v64/i04/

Embrechts, P., A.J. McNeil, D. Straumann. 2002 "Correlation and dependence in Risk Management: Properties and pitfalls" in *Risk Management: Value at Risk and Beyond*, edited by M. Dempster.
Cambridge: Cambridge University Press.

Engle, R.F. 1982. "Autoregressive conditional hetroscedasticity with estimates of the variance of United Kingdom Inflation." *Econometria* 50(4): 987-1007.

---------. 2002. "Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models."
Journal of Business and Economic Statistics 20(3): 339-350.

Engle, R.F. and Sheppard, K. 2001. "Theoretical and empirical properties of dynamic conditional multivariate GARCH."
University of California San Diego, discussion paper.

Erb, C., Harvey, C. and Viskanta, T. 1994. "Forecasting international equity correlations" *Financial Analyst Journal* 50(6): 32-45.

Focardi, S. M., F.J. Fabozzi and T.G. Bali. 2013. *Mathematical Methods for Finance*. NJ: John Wiley & Sons, Inc.

Fox, J. and S.Weisberg. 2011. *An {R} Companion to Applied Regression.* 2nd ed. Thousand Oaks CA: Sage.

http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Francq, C and J-M. Zakoïan. 2014. "Estimating multivariate GARCH and Stochastic Correlation models equation by equation." CREST and Université de Lille (EQUIPPE). Working paper retrieved 2016 from:

http://perso.univ-lille3.fr/~cfrancq/Christian-Francq/statistics-econometrics-papers/EbEE.pdf

---------. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. West Sussex, United Kingdom: John Wiley & Sons Ltd.

Genz, A. and F. Bretz. 2009. *Computation of Multivariate Normal and t Probabilities, Lecture Notes in Statistics Vol. 195.* Heidelberg, Germany: Springer-Verlag, Heidelberg.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, T. Hothorn. 2016. "mvtnorm: Multivariate Normal and t Distributions. R package version 1.05."

http://CRAN.R-project.org/package=mvtnorm

Ghalanos, A. 2015a. "rmgarch: Multivariate GARCH models. R package version 1.3-0."

---------. 2015b. "rugarch: Univariate GARCH models. R package version 1.3-6."

Goetzmann, W. N., L. Li and K.G. Rouwenhorst. 2005."Long-term global market correlations." *Journal of Business* 78(1): 1-38.

Hansen, P.R. and A. Lunde. 2005. "A forecast comparison of volatility models: does anything beat a GARCH(1, 1)?" *Journal of Applied Econometrics* 20(7): 873–889.

Harter, R., K. Hornik and S. Theussl. 2016. "Rsymphony: SYMPHONY in R. R package version 0.1-22."

https://CRAN.R-project.org/package=Rsymphony

Hendricks, D. 1996. "Evaluation of value-at-risk models using historical data." *Economic Policy Review* 2(1): 39-69.

Hlouskova, J., K. Schmidheiny and M. Wagner. 2009. "Multistep predictions for multivariate GARCH models: Closed form solution and the value for portfolio management." *Journal of Empirical Finance* 16(2): 330-336.

Hofert, M., I. Kojadinovic, M. Maechler and J. Yan. 2015. “copula: Multivariate Dependence with Copulas. R package version 0.999-14.”
http://CRAN.R-project.org/package=copula

Hofert, M., M. Maechler. 2011. “Nested Archimedean Copulas Meet R: The nacopula Package.” *Journal of Statistical Software* 39(9): 1-20.
http://www.jstatsoft.org/v39/i09/

Kojadinovic, I., J. Yan. 2010. “Modeling Multivariate Distributions with Continuous Margins Using the copula R Package.” *Journal of Statistical Software* 34(9): 1-20.
http://www.jstatsoft.org/v34/i09/

Komsta, L. and F. Novomestky. 2015. “moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. “
https://CRAN.R-project.org/package=moments

Krause, A. 2002. “Coherent risk measurement: an introduction.” *Balance Sheet.* 10(4): 13 – 17.

Krokhmal, P., J. Palmsquist and S. Uryasev. 2002. “Portfolio optimization with conditional value-at-risk objective and constraints.” *Journal of Risk* 4(2): 11-27.

Lawrence, C. and G. Robinson. 1995. “Liquid measures.” *Risk* 8(7): 52–55.

Ledoit, O., Santa-Clara, P. and Wolf, M. 2003 “Flexible multivariate GARCH modeling with an application to international stock markets.” *Review of Economics and Statistics* 85(3): 735-747.

Longin, F. and B. Solnik 1995 “Is the international correlation of equity returns constant: 1960-1990?” *Journal of International Money and Finance* 14(1): 3-26.

Luethi, D. and W. Breymann. 2013. ghyp: A package on the generalized hyperbolic distribution and its special cases. R package version 1.5.6.
https://CRAN.R-project.org/package=ghyp

Mandelbrot B. 1963. “The variation of certain speculative prices.” *Journal of Business* 36(4): 394-419.

Markowitz, H. 1952. “Portfolio selection.” *Journal of Finance* 7(1): 77-91.

---------. 1959. *Portfolio Selection: Efficient Diversification of Investments*. NY:John Wiley & Sons, Inc.

McNeil, A.J., R. Frey and P Embrechts. 2005. *Quantitative Risk Management*. NJ:Princeton University Press.

Mendes B. V. M. and D.S. Marques. 2012. “Choosing an optimal investment strategy: The role of robust pair-copulas based portfolios.” *Emerging Markets Review* 13(4): 449-464.

Nelsen, R. (1999). *An Introduction to Copulas, Lecture Notes in Statistics*. New York: Springer-Verlag New York, Inc.

Nelson, D. B. 1991. “Conditional Heteroskedasticity in Asset Returns: a New Approach.” *Econometrica* 59(2): 347-370.

Owen, J. and R. Rabinovitch. 1983. “On the class of elliptical distributions and their applications to the theory of portfolio choice.” *Journal of Finance* 38(3): 745-752.

Peterson, B.G. and P. Carl. 2014. “PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.4.3541.”
https://CRAN.R-project.org/package=PerformanceAnalytics

---------. 2015. “PortfolioAnalytics: Portfolio Analysis, Including Numerical Methods for Optimization of Portfolios. R package version 1.0.3636.”
https://CRAN.R-project.org/package=PortfolioAnalytics

Rmetrics Core Team, D. Wuertz and T. Setz. 2014. “fMultivar: Rmetrics-Analysing and Modeling Multivariate Financial Return Distributions. R package version 3011.78.”
https://CRAN.R-project.org/package=fMultivar

Rmetrics Core Team, D. Wuertz, T. Setz and Y. Chalabi. 2014a. “fBasics: Rmetrics - Markets and Basic Statistics. R package version 3011.87.”
https://CRAN.R-project.org/package=fBasics

---------. 2014b.” fPortfolio: Rmetrics - Portfolio Selection and Optimization. R package version 3011.81. “
https://CRAN.R-project.org/package=fPortfolio

Rockafellar R.T. and S. Uryasev. 2002. “Conditional value-at-risk for general loss distributions.” *Journal of Banking & Finance* 26(7): 1443-1471.

---------. 2000. “Optimization of Conditional Value-at-Risk” *Journal of Risk* 2(3): 21-41.

Ryan, J.A. 2015. “quantmod: Quantitative Financial Modelling Framework. R package version 0.4-5.”
https://CRAN.R-project.org/package=quantmod

Ryan, J.A. and J. M. Ulrich. 2014. "xts: eXtensible Time Series. R package version 0.9-7."
https://CRAN.R-project.org/package=xts

S&P Dow Jones Indices. 2016a. "S&P Select Industry Indices *Methodology*" Retrieved 17.08.2016.
http://us.spindices.com/documents/methodologies/methodology-sp-us-indices.pdf?force_download=true

---------. 2016b. "U.S. EQUITY INDUSTRY INDEX RETURNS." Retrieved 17.08.2016.
http://us.spindices.com/additional-reports/all-returns/index.dot?parentIdentifier=b3a06701-8feb-4ca1-ac99-ce0d61acf409&sourceIdentifier=index-family-specialization&additionalFilterCondition=

---------. 2016c. "Commodities: S&P GSCI." Retrieved 24.08.2016.
http://us.spindices.com/index-family/commodities/sp-gsci

---------. 2016d. "S&P GSCI Methodology" Retrieved 24.08.2016.
http://us.spindices.com/documents/methodologies/methodology-sp-gsci.pdf?force_download=true

Schmidt, T. 2006 «Coping with copulas» Forthcoming in Risk Books *Copulas From Theory to Applications in Finance*

Silvennoinen, A. and T. Teräsvirta. 2009. "Multivariate GARCH Models" in *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J-P. Kreiß and T. Mikosch, 208-232. Berlin, Germany: Springer-Verlag Berlin Heidelberg.

Skintzi, V.D., G. Skaidoupoulous and A.P.N. Refenes. 2005. "The Effect of Mis-Estimating Correlation on Value at Risk." *Journal of Alternative Investments* 7(4): 66-82.

Sklar, A. 1959. "Fonctions de répartition à n dimensions e leurs marges." *Publications de l'Institut de Statistique de l'Univiversité de Paris* 8: 229 - 231.

Trapletti, A. and K. Hornik. 2016. tseries: "Time Series Analysis and Computational Finance. R package version 0.10-35"

Turlach, B. A. R port by Weingessel, A. 2013. quadprog: Functions to solve Quadratic Programming Problems. R package version 1.5-5. https://CRAN.R-project.org/package=quadprog

Von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press

Wuertz, D., Y. Chalabi with contribution from M. Miklovic, C. Boudt, P. Chausse and others. 2013. "fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. R package version 3010.82." https://CRAN.R-project.org/package=fGarch

Wuertz, D., Y. Chalabi, W. Chen and A. Ellis. 2009. *Portfolio Optimization with R/Rmetrics*. Zürich, Germany: Finance Online

Yan, J. 2007. "Enjoy the Joy of Copulas: With a Package copula." *Journal of Statistical Software* 21(4): 1-21. http://www.jstatsoft.org/v21/i04/

Yamai, Y and T. Yoshiba. 2002. "Comparative analyses of expected shortfall and value-at-risk: Expected Utility Maximization and Tail Risk." *Monetary and Economic Studies* 20(2): 95-115.

Yilmaz, T. 2010. "Improving Portfolio Optimization by DCC and DECO GARCH: Evidence from Istanbul Stock Exchange." Ph.D. candidate paper, Yeditepe University. Retrieved 2016 from http://mpra.ub.uni-muenchen.de/27314/

Zeileis, A. and G. Grothendieck. 2005. "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14(6): 1-27. doi:10.18637/jss.v014.i06

Zivot, E. 2009. "Practical Issues in the Analysis of Univariate GARCH Models" in *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J-P. Kreiß and T. Mikosch, 113-155. Berlin, Germany: Springer Verlag Berlin Heidelberg.

Appendix A: General Theory

A1: Basic statistics and econometrics

This section briefly reviews some basic statistical and econometrical concepts and definitions that the thesis draws upon, but isn't discussed in depth.
The main references for this section are: Alexander (2008a), Francq and Zakoian (2010) and McNeil, Frey and Embrechts (2005).

Financial returns as a random variable

It is common to model financial returns as a continuous random variable. If the returns of financial assets were *not random*, i.e. predictable, it would be possible to systematically earn positive returns while avoiding market downfalls. This would contradict the market efficiency hypothesis. As for the continuous part of the assumption, financial asset prices typically move in "ticks" (increments), so this is only an approximation. However it is an approximation that makes the modeling and mathematical part of finance a lot more convenient than the discrete alternative, at a very low cost.
In addition, what is considered is usually what is referred to as the log returns, i.e.

$$r_{t,\Delta t} = X_{t+\Delta t} - X_t, \qquad where\ X_t = \ln S_t\,, \qquad and$$
$$S_t\ denotes\ the\ price\ of\ a\ financial\ asset$$

Log returns have the advantage over regular returns that accumulation can be done by addition.

When referring to returns throughout the thesis log returns is implied, unless otherwise explicitly stated.

Expected Value and the expectations operator

The first moment of a probability distribution is called the expected value. Throughout the thesis we will denote the expectation of a random variable, say X as $E(X)$

Standardized random variable

When we refer to a standardized random variable, it simply means a random variable that is scaled so that it exhibit a zero mean and unit variance

Independent and identically distributed random variables

A common assumption for a stationary process is that the random variables are *independent and identically distributed (i.i.d.).* This implies that there is no autocorrelation, and that the statistical moments is the same for all of the random variables. In particular the variance is assumed the same for all time periods, i.e. the process is *homoscedastic*, in contrast to a process where variance is varying, that is a *heteroscedastic* process.

The i.i.d. assumption is generally used to ease the task of statistical inference, but as shown in section, is often not suitable for finance.

Definition – Variance

Variance represents the dispersion about the mean of the density,

$$\sigma^2 = V(X) = E([X - E(X)]^2)$$

Standard deviation

The square root of variance is called the standard deviation, or sometimes in econometrics, volatility.

Definition – Covariance

Covariance is the first central moment of the joint density function of X and Y,

$$Cov(X,Y) = E\big[(X - \mu_x)\big(Y - \mu_y\big)\big], \quad \mu_x = E(X), \quad \mu_y = E(Y)$$

Definition – Correlation

Covariance is determined not only by the degree of dependence between X and Y, but also their size and the size of their deviations. For this reason it's preferable to work with a "more scaled" parameter, correlation. Throughout the thesis we will focus Pearson's correlation unless otherwise stated;

$$Corr = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

Auto correlation (serial correlation): Correlation with previous observations (lags) of itself over a given time interval.

Stationary Processes

Stationarity is important in financial time series analysis, as it sort of replaces the i.i.d. assumption in standard statistics (Francq, C. and J-M. Zakoïan 2010, 1) and allows us to make statistical inference.

If we consider a sequence of real random variables (X_t), $t \in \mathbb{Z}$ defined on the same probability space (i.e. a stochastic time series), then strict stationarity is defined as:

Definition – Strict Stationarity

$The\ process\ X_t\ is\ said\ to\ be\ strictly\ stationary\ if\ the\ vectors\ (X_1, \dots, X_k)'$ $and\ (X_{1+h}, \dots, X_{k+h})'\ have\ the\ same\ joint\ distribution\ for\ any\ k \in \mathbb{N}\ and$ $any\ h \in \mathbb{Z}$

We can also have weak stationarity, which as the name implies, is often less demanding in that it only constrains the first two statistical moments (I.e. the mean and variance/autocovariance. The moments do however have to exist.)

Definition – Weak Stationarity (Second-order stationarity)

$$The\ process\ X_t\ is\ said\ to\ be\ weakly\ stationary\ if:$$

1) $\mathbb{E}\, X_t^2 < \infty, \quad \forall\, t \in \mathbb{Z}$

2) $\mathbb{E}\, X_t = m, \quad \forall\, t \in \mathbb{Z}$

3) $Cov(X_t, X_{t+h}) = \gamma x(h), \quad \forall\, t\,, \quad h \in \mathbb{Z}$

$$Where\ \gamma x(\cdot)\left(\rho X(\cdot) := \frac{\gamma x(\cdot)}{\gamma x(0)}\right)$$

$$is\ the\ autocovariance\ (autocorrelation) function\ of\ X_t$$

An important example of a weakly stationary process is a white noise process;

Definition – Weak white noise

$$The\ process\ X_t\ is\ said\ to\ be\ weak\ white\ noise\ if:$$

1) $\mathbb{E}\, \varepsilon_t^2 = \sigma^2, \quad \forall\, t \in \mathbb{Z}$

2) $\mathbb{E}\, \varepsilon_t = 0, \quad \forall\, t \in \mathbb{Z}$

3) $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0, \quad \forall\, t, \quad h \in \mathbb{Z}\,, \quad h \neq 0$

Strong white noise differs from weak in that instead of assuming no autocorrelation, we assume independence. I.e. hypothesis (3) gets replaced by the more constraining

$$3')\ \ \varepsilon_t\ and\ \ \varepsilon_{t+h}\ are\ i.i.d.$$

ARMA Models

ARMA (autoregressive moving average) models are the most widely used model type for the prediction of weakly stationary processes (Francq, C. and J-M. Zakoïan 2010, 4). ARMA models are often preferred to MA for parsimony reasons, as they in general require fewer parameters to be estimated.

Definition – AutoRegressive Moving Average (ARMA(*p, q*)) process

$$A\ weakly\ stationary\ process\ X_t\ is\ called\ ARMA(p,q)\ where\ p\ and\ q\ are\ integers, if\ there\ exist\ real\ coefficients\ c, a_1, \dots, a_p, b_1, \dots, b_q\ such\ that$$

$$\forall\, t \in \mathbb{Z}, \quad X_t + \sum_{i=1}^{p} a_i X_{t-i} = c + \varepsilon_t + \sum_{j=1}^{q} b_j\, \varepsilon_{t-j}$$

$$where\ \varepsilon_t\ is\ the\ linear\ innovation\ process\ of\ X_t$$

(A formal mathematical definition of a linear innovation process can be found in most advanced books on econometrics, for instance Francq & Zakoian 2010. An adequate understanding/interpretation for the rest of the material covered in this thesis would be to interpret ε_t as the latest of many shocks to X_t)

Elliptical distributions

A family of distributions where the *level sets*, or contours, of the bivariate distribution's density function forms ellipses. Examples are the multivariate normal distribution and the multivariate Student t-distribution. In the bivariate form of these distributions there is a single parameter, ϱ, the correlation between the two variables X and Y.

Hadamard product

Throughout the thesis $A \odot B$ will denote the Hadamard product between matrix A and B, i.e. the operation where every cell in matrix A is multiplied with the correspondent cell in matrix B;

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$A \odot B = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}$$

A2: Assumptions of mathematical finance

When trying to model the financial world with mathematics, assumptions are often needed. This subsection review general assumptions of mathematical finance used in the empirical application (Focardi et al [2013]).

Not moving the market

We assume that our actions do not affect the market price. In free markets this is not true, increasing demand (buying securities) increases the price, while increasing supply (selling securities) lower the price. If we are trading in small quantities the effects will be negligible, while they can have an impact if we are buying or selling large amounts of small cap stocks. As we deal with the large cap US equity market, we don't see this as a big problem.

Market liquidity

Closely related to the first is the assumption of market liquidity. We assume that we can buy or sell as much as we want to at the market price at any time. Again, we deal with the large cap US equity market, so this is unlikely to be a problem.

Shorting

Shorting tends to be restricted for most investors, with hedge funds being one of the few exceptions. We assume that shorting isn't allowed, this tends to negatively affect the amount of diversification we can get compared to that of an investor who can short.

Fractional quantities

Financial models and algorithms (such as those applied in portfolio selection) tend to seek out the optimal quantities through mathematical processes, often leaving us with recommendations of purchasing fractional quantities of assets. This clearly isn't possible in the markets, but if we assume that our investment is big, we are often able to come close to the relative proportions allocated by the model regardless.

No transaction costs

Selling and buying securities come at a price, and typically one also has to face the extra costs implied by the bid-ask spread. We ignore these costs.

A3: Markowitz Calculations

Minimum Variance derivation

If we denoted the covariance matrix by $\boldsymbol{\Sigma}$, we can write $\sigma_p^2 = \boldsymbol{W'\Sigma\, W}$

$$\sigma_p^2 = (W_1 \quad W_2 \quad \dots \quad W_n)\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \dots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_n^2 \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}$$

The Lagrangian for the minimum variance portfolio in matrix notation is then simply

$$L(\boldsymbol{W}, \lambda) = \boldsymbol{W'\Sigma\, W} + \lambda(\boldsymbol{W'1} - 1)$$

Where $\mathbf{1}$ is an $n \times 1$ vector of 1's.

The first order conditions (FOC's) are then the linear equations

#1 $\frac{\partial L(\boldsymbol{W},\lambda)}{\partial \boldsymbol{W}} = 2\boldsymbol{\Sigma W} + \lambda\mathbf{1} = 0,$

#2 $\frac{\partial L(\boldsymbol{W},\lambda)}{\partial \lambda} = \boldsymbol{W'1} - 1 = 0.$

Solving #1 for $\boldsymbol{W}$ we get

$$\boldsymbol{W} = -\frac{1}{2}\lambda\boldsymbol{\Sigma^{-1}1}$$

Multiplying both sides by $\mathbf{1}'$ we get

$$\boldsymbol{1'W} = -\frac{1}{2}\lambda\boldsymbol{1'\Sigma^{-1}1}$$

As #2 implies that $\boldsymbol{1'W} = 1$, we can solve for λ:

$$1 = -\frac{1}{2}\,\lambda\boldsymbol{1'\Sigma^{-1}1}$$

$$\lambda = -\frac{2}{\boldsymbol{1'\Sigma^{-1}1}}$$

Inserting λ into #1 again we get:

$$2\boldsymbol{\Sigma W} - \frac{2}{\mathbf{1'\Sigma^{-1}1}}\mathbf{1} = 0$$

$$\boldsymbol{\Sigma W} = \frac{\mathbf{1}}{\mathbf{1'\Sigma^{-1}1}}$$

Multiplying both sides by $\boldsymbol{\Sigma^{-1}}$:

$$\boldsymbol{W} = \frac{\boldsymbol{\Sigma^{-1}1}}{\mathbf{1'\Sigma^{-1}1}}$$

Finally, recalling that the exact solution for $\boldsymbol{W}$ that minimizes variance is noted $\boldsymbol{W_{MV}}$, we end up with:

$$\boldsymbol{W_{MV}} = \frac{\boldsymbol{\Sigma^{-1}1}}{\mathbf{1'\Sigma^{-1}1}}$$

Mapping of the Efficient Frontier

Markowitz argues that we can map out the efficient frontier by minimizing portfolio return for a given level of portfolio return, i.e.

$$Min\ \sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \sigma_{ij}$$

Subject to

$$1. \sum_{i=1}^{n} Wi\ E(R_i) = E^*$$

where E^* is the targeted expected portfolio return and

$$2. \sum_{i=1}^{n} Wi = 1$$

secures that the weights of the portfolio sum up to 100%.

Alternatively, giving an equivalent frontier (when plotted for enough target portfolio returns or variances);

$$Max \sum_{i=1}^{n} Wi\ E(R_i)$$

Subject to

$$1.\ \sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \sigma_{ij} = \sigma_P^{2^*}$$

where $\sigma_P^{2^*}$ is the targeted portfolio variance and

$$2. \sum_{i=1}^{n} Wi = 1$$

This problem can for instance be solved with the Lagrangian method and matrix algebra.

For our first formulation (i.e. minimize variance for a given expected portfolio return) we can formulate the following objective function:

$$Min\ L = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \sigma_{ij} + \lambda_1 \sum_{i=1}^{n} [W_i\ E(R_i) - E^*] + \lambda_2 \left(\sum_{i=1}^{n} W_i - 1\right)$$

We can then take the partial derivatives of this function with respect to each of the variables, $W_1, W_2, \dots, W_n,\ \lambda_1$ and λ_2 and set the resulting equations equal to zero and solve for the portfolio weights. Once we have solved the algorithm for enough combinations of return and risk, we are able to map out the frontier, and the investor can pick the portfolio most in line with his or her preferences (e.g. via a utility function).

A4: On VaR estimation

Estimating VaR through analytical computation

The most common example of a method from the first category would be to assume that returns are normal and i.i.d., so that we can use the sample mean vector and sample covariance matrix as inputs. Value at Risk of the portfolio is then given by $VaR_\alpha = \Phi^{-1}(\alpha)\sigma_h - \mu_h$

where

- Φ^{-1} is the inverse of the standard normal cumulative density function
- α the confidence level, so that $1 - \alpha$ is the quantile return
- σ_h and μ_h are the h −day portfolio standard deviation and mean return calculated from the sample statistics, in a similar fashion to that described in section 1 (Note: The value we choose for h will then also be our chosen risk-horizon).

The main weaknesses of the analytical approach are that the estimate is very sensitive to the distribution we assume and inputs we enter into the inverse density. As we have already seen in section 2, the assumptions made in the above example are unrealistic, and unfortunately there are no really good alternatives, especially for the density function. Even if we are able to specify densities that match the empirical distribution better, these are often very difficult to work with and estimate inputs for.

Estimating VaR through historical returns

We can also estimate Value at Risk by plotting the histogram of a sample of returns we believe to be representative of the joint return distribution. The actual math for getting the estimate of the VaR then is very simple, we only have to look for the value given by the $(1 - \alpha)$-quantile of the distribution. The issue here is that we again have to rely heavily on our assumptions of the past being a good predictor of the future, when there is good reasons to believe it might not be.

When we estimate tail risk from histograms, good estimates of the tails of a distribution require many observations, as we in reality only use $(1-\alpha)x\ 100\%$ of the sample. Not all assets have a long history to draw data from, and even for those that do have this can be problematic. We know from the stylized facts that the empirical distribution is hardly stable (e.g. there is heteroscedasticity), so using a large sample to plot the histogram might leave us with a lot of irrelevant data. On the other hand, if we use a more recent, smaller sample of the joint return distribution to create the histogram, the estimate of VaR might be very imprecise and have large confidence bands.

Estimating VaR by scenario simulation

Lastly we can estimate Value at Risk by simulation. For example, if we use x simulations, we have then created x simulated portfolio returns at the risk horizon in h days. The VaR can then simply be obtained by ranking the returns from the lowest to the highest, and plotting these in a frequency diagram. The VaR is then given by the $-(1-\alpha)$ quantile of this distribution. The list of possible simulation types are almost endless, as we can make almost any assumptions we want to. In the simplest form the simulated Value at Risk uses the assumptions that returns are i.i.d. and multivariate normal, but we can also for instance use a copula to model the dependence, or draw upon aspects of the historical distribution. However, the power of the simulation approach is not without downsides. The risk of simulation errors are always there and the approach is computationally intensive. There is also evidence of correlation error being even more detrimental to simulation approaches than the other Value at Risk approaches, see Skintzi et al (2006).

A5: Autoregressive Conditional Heteroskedasticity

Conditional and unconditional variance

Unconditional i.i.d. variance is simply what is most often referred to as just variance, defined as:

$$Var(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

The unconditional variance is assumed to be constant over the entire data period we consider, and can be thought of as the average variance for the respective period. In this manner the measure is *static*. However, unconditional variance also comes in other forms when what we wish to model is not assumed to be i.i.d., e.g. in GARCH-modeling.

Conditional variance is conditional on the previous knowledge (history) we have of that variable. We can say that we account for the *dynamic* properties of what we are trying to model (e.g. financial returns) by modeling their distribution as time dependent. For instance if we are interested in the conditional variance h_t of the error term u_t in a regression, we can write it as

$$h_t = var(u_t|\mathcal{F}_{t-1}) = \mathrm{E}[(u_t - \mathrm{E}[u_t])^2|\mathcal{F}_{t-1}]$$

Where $\mathcal{F}_{t-1}$ is the information set of previous values of u_t ($e.g.\ u_{t-1}, u_{t-2}$)

(theoretically speaking $\mathcal{F}_{t-1}$ is the σ-field generated by $u_{t-j}, j \geq 1$)

Further, it's common to assume $\mathrm{E}[u_t] = 0$, so

$$h_t = var(u_t|\mathcal{F}_{t-1}) = \mathrm{E}[u_t^2|\mathcal{F}_{t-1}]$$

Stating that the conditional variance of a zero mean, normally distributed random variable equals the conditional expected value of the squared random variable (Brooks 2008, 387-388). The intuition is that the conditional variance is allowed to vary over time as a function of previous error terms (residuals), and thus changes at every point in time.

The ARCH model is the predecessor to the popular GARCH model. The ARCH model was the first model of conditional heteroscedasticity, and was presented by Engle in 1982. The model is intuitively appealing in that since volatility in financial time series acts in clusters, i.e. variance expresses autocorrelation, a model conditional on the history of previous levels of variance should have prediction power.

ARCH-model formal definition

The ARCH(*p*) model is formally defined as:

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$$

Where ϵ_t may be a directly observable variable, or more widely used, the error term of a regression. In the latter case we also need a "mean equation" for a complete model, i.e. a regression model of the variable we are interested in. Often the variable is tested for indications of following an ARMA process (A more thorough discussion of modeling the conditional mean is found in the univariate GARCH section below).

Engle assumed that ϵ_t could be decomposed as $\epsilon_t = z_t h_t^{1/2}$, where z_t is an sequence of i.i.d. standardized random variables.
$\alpha_i \; \forall \;\; i$ are the parameters of the model, and are constrained to be positive in the regression in order to ensure positive conditional variance.

On the properties of the E-GARCH

$$\ln(\sigma_t^2) = h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j h_{t-j}$$

To see how the model is able to capture asymmetric responses, assume we estimate a (1,q) E-GARCH. When ϵ_{t-1} is positive (i.e. good news), the total effect of ϵ_{t-1} on the log variance is $\alpha_1(1+\gamma_1)|\epsilon_{t-1}|$, while if ϵ_{t-1} is negative (i.e. bad news), the effect is $\alpha_1(1-\gamma_1)|\epsilon_{t-1}|$. As bad news are expected to have a larger impact on volatility than good news, we anticipate $\gamma_i$to have the opposite sign of α_i for all *i*.

Additionally, the E-GARCH model has an advantage over the standard GARCH model in that the conditional variance σ_t^2 is guaranteed to be positive regardless of the sign of the estimated coefficients due to the logarithm of σ_t^2 being modelled over σ_t^2.

Conditions for the covariance stationarity of the E-GARCH model can be found in Nelson (1991). For practical purposes we require $\sum_{j=1}^{q} \beta_j < 1$

A6: Copulas and dependence measures

Implicit and Explicit Copulas

Implicit copulas are one of the two main categories of copulas, the other being explicit copulas. The implicit copulas do not possess a closed form, i.e. they can only be expressed as an integral and therefore it is easier and common to work with the copulas density function rather than the distribution function itself. The name implicit copulas stems from the fact that these copulas are derived from a standard multivariate distribution, and the dependence is isolated and *implied* by the chosen distribution. In this thesis we focus solely on implicit copulas

Concordance Metrics

As we have discussed in the thesis, interpretation of Pearson's linear correlation is problematic when the underlying distributions aren't elliptical. To formalize this, we consider concordance metrics;

Two pairs of observations on the continuous random variables X and Y, $(x1, y1)$ and $(x2, y2)$ are said to be *concordant* if $x1 - x2$ has the same sign as $y1 - y2$, and *discordant* if not.

A concordance metric $m(X, Y)$ is then defined as a numerical metric of dependence between X and Y so that:

- $m(X, Y) \in [-1,1]$ and its value within this range depends on the joint distribution of X and Y
- $m(X, X) = 1$ and $m(X, -X) = -1$
- $m(X, Y) = m(Y, X)$ and $m(X, -Y) = -m(X, Y)$
- If X and Y are independent then $m(X, Y) = 0$
- Given two possible joint distributions $F(X, Y)$, $G(X, Y)$, with $m_F(X, Y)$ and $m_G(X, Y)$ denoting the respective concordance metrics. Then if $F(X, Y) \geq G(X, Y)$, $m_F(X, Y) \geq m_G(X, Y)$

The problem with Pearson's linear correlation is that it is not a concordance metric, except for the strict condition that X and Y have elliptical distributions. Even when this is the case, correlation of 0 doesn't imply independence if the bivariate distribution is Student t, confusing the interpretation further (Alexander

2008b, 273-275). If there is strong evidence suggesting that the distributions indeed are not elliptical, one should consider rank correlations (see e.g. McNeil et al 2005, 206).

Tail dependence

There are many definitions of tail-dependence measures in the literature, here we use the one presented in McNeil et al (2005).

For the upper tail dependence, we measure the probability that X exceeds its q-quantile, given that Y exceeds its q-quantile, and the consider the limit value as $q \to 100\%$. Additionally, X and Y are interchangeable, i.e. the tail dependence between X and Y is the same as the tail dependence between Y and X.

Formally, the coefficient of upper tail dependence is given by

$$\lambda_u(X,Y) = \lim_{q \to 1^-} P[Y > F_Y^{-1}(q) | X > F_X^{-1}(q)]$$

Provided a limit of $\lambda_u(X,Y) \in [0,1]$ exist.

Analogously, the coefficient of lower tail dependence is given by

$$\lambda_l(X,Y) = \lim_{q \to 0^+} P[Y \leq F_Y^{-1}(q) | X \leq F_X^{-1}(q)]$$

Again, provided a limit of $\lambda_l(X,Y) \in [0,1]$ exist.

Gaussian Copula density

The Gaussian copula is derived from the n-dimensional multivariate and univariate standard normal distribution functions, i.e. it's defined by

$$\boldsymbol{C}(u_1, \dots, u_n; P) = \Phi_m[\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)]$$

Where

- Φ_m is the *n*-dimensional multivariate standard normal distribution function.
- Φ is the univariate standard normal distribution function.

- P is the correlation matrix.

Differentiating the copula function yields the copula's density (Alexander 2008b, 267);

$$c(u_1, \dots, u_n; P) = |P|^{-1/2} e^{-\xi'(P^{-1}-I)\xi}$$

Where

- $|P|$ denotes the determinant of P.
- I is the identity matrix.
- ξ is a vector of realized normal variables, $\xi = [\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)]$, i.e. a vector of uniform variables put through the inverse univariate standard normal distribution.

In the special case where $n = 2$, the normal copula distribution is

$$C(u_1, u_2; \rho_{1,2}) = \Phi_m[\Phi^{-1}(u_1), \Phi^{-1}(u_2)]$$

Where Φ_m is the 2-dimensional multivariate standard normal distribution, i.e. the bivariate standard normal distribution function.

The bivariate normal copula density is then given by

$$c(u_1, u_2; \rho_{1,2}) = \left(1 - \rho_{1,2}^2\right)^{-\frac{1}{2}} \times e^{-\left[\frac{\xi_1^2 - 2\rho_{1,2}\xi_1\xi_2 + \xi_2^2}{2(1-\rho_{1,2}^2)}\right]}$$

Where

- $\xi_1 = \Phi^{-1}(u_1)$
- $\xi_2 = \Phi^{-1}(u_2)$

Student t Copula density

The multivariate Student t copula is defined by

$$\boldsymbol{C}(u_1, \dots, u_n; P) = mt_v[t_v^{-1}(u_1), \dots, t_v^{-1}(u_n)]$$

Where

- mt_v is the *n*-dimensional multivariate Student t distribution function with v degrees of freedom.
- t_v is the univariate Student t distribution function with v degrees of freedom.
- P is the correlation matrix.

Using the definition of the multivariate Student t density (Alexander 2008a, 117)

$$f(x) = \Gamma\left(\frac{v}{2}\right)^{-1} \Gamma\left(\frac{v+n}{2}\right)(v\pi)^{-\frac{n}{2}}|P|^{-\frac{1}{2}}(1 + v^{-1}x'\Sigma^{-1}x)^{-\frac{(v+n)}{2}}$$

where

- Γ denotes the gamma function.
- $|P|$ denotes the determinant of P.

we can express the Student t copula distribution, and differentiate this expression to obtain the Student t copula density (for a more detailed derivation see Alexander 2008b, 268);

$$c_v(u_1, \dots, u_n; P) = \Gamma\left(\frac{v}{2}\right)^{n-1} \Gamma\left(\frac{v+1}{2}\right)^{-n} \Gamma\left(\frac{v+n}{2}\right)|\Sigma|^{-\frac{1}{2}}(1 + v^{-1}\xi' P^{-1}\xi)^{-\frac{(v+n)}{2}} \prod_{i=1}^{n}(1 + v^{-1}\xi_i^2)^{\frac{(v+1)}{2}}$$

where $\xi = [t_v^{-1}(u_1), \dots, t_v^{-1}(u_n)]$ is a vector of realized Student t variables, i.e. a vector of uniform variables put through the inverse univariate Student's t distribution.

If we have two marginals, i.e. $n = 2$, we have a symmetric bivariate t copula density of

$$c(u_1, u_2; \rho_{1,2}) =$$

$$\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{v+1}{2}\right)^{-2}\Gamma\left(\frac{v+2}{2}\right)\left(1 - \rho_{1,2}^2\right)^{-\frac{1}{2}} \times$$

$$\left[1 + v^{-1}\left(1 - \rho_{1,2}^2\right)^{-1}\left(\xi_1^2 - 2\rho_{1,2}\xi_1\xi_2 + \xi_2^2\right)\right]^{-\frac{(v+2)}{2}} \times [(1 + v^{-1}\xi_1^2)(1 + v^{-1}\xi_2^2)]^{\frac{(v+1)}{2}}$$

$$= \frac{\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{v+2}{2}\right)}{\sqrt{\left(1-\rho_{1,2}^2\right)}\times\Gamma\left(\frac{v+1}{2}\right)^2} \times \frac{\left[\left(1+v^{-1}\xi_1^2\right)\left(1+v^{-1}\xi_2^2\right)\right]^{\frac{(v+1)}{2}}}{\left[1+v^{-1}\left(1-\rho_{1,2}^2\right)^{-1}\left(\xi_1^2-2\rho_{1,2}\xi_1\xi_2+\xi_2^2\right)\right]^{\frac{(v+2)}{2}}}$$

Where

- $\xi_1 = t_v^{-1}(u_1)$
- $\xi_2 = t_v^{-1}(u_2)$

Appendix B: Additional plots and tables

B1: Equity dataset

	statistic	parameter	p.value	method
Consumer.Discretionary	4448.71	2	0	"Jarque Bera Test"
Consumer.Staples	10013.06	2	0	"Jarque Bera Test"
Energy	11702.49	2	0	"Jarque Bera Test"
Financial	11241.39	2	0	"Jarque Bera Test"
Health.Care	11380.23	2	0	"Jarque Bera Test"
Industrial	2482.445	2	0	"Jarque Bera Test"
Materials	3445.796	2	0	"Jarque Bera Test"
Technology	5566.195	2	0	"Jarque Bera Test"
Utilities	14759.77	2	0	"Jarque Bera Test"
Homebuilders	760.1258	2	0	"Jarque Bera Test"

Figure 36 - Jarque Bera test, Equity indices

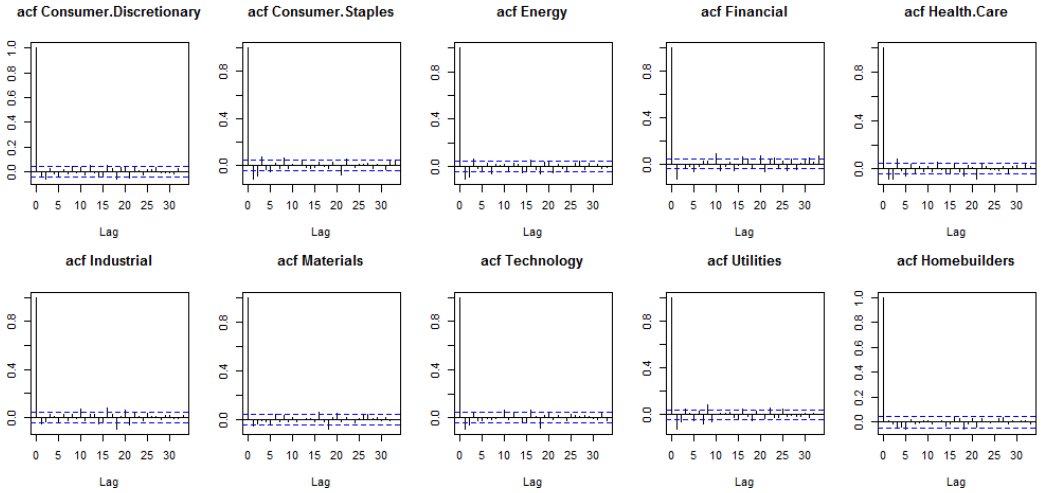


Figure 37 - Linear ACFs, Equity indices

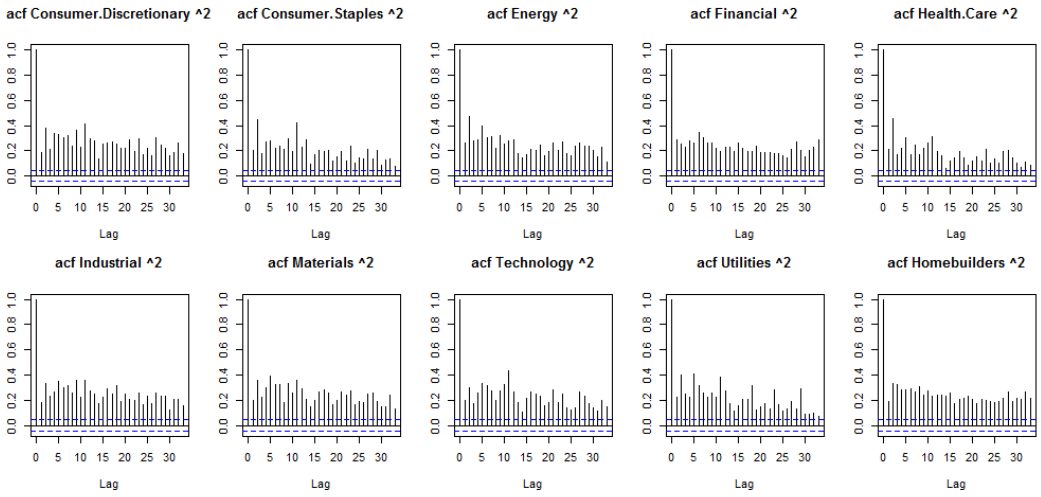


Figure 38 - Squared ACFs, Equity indices

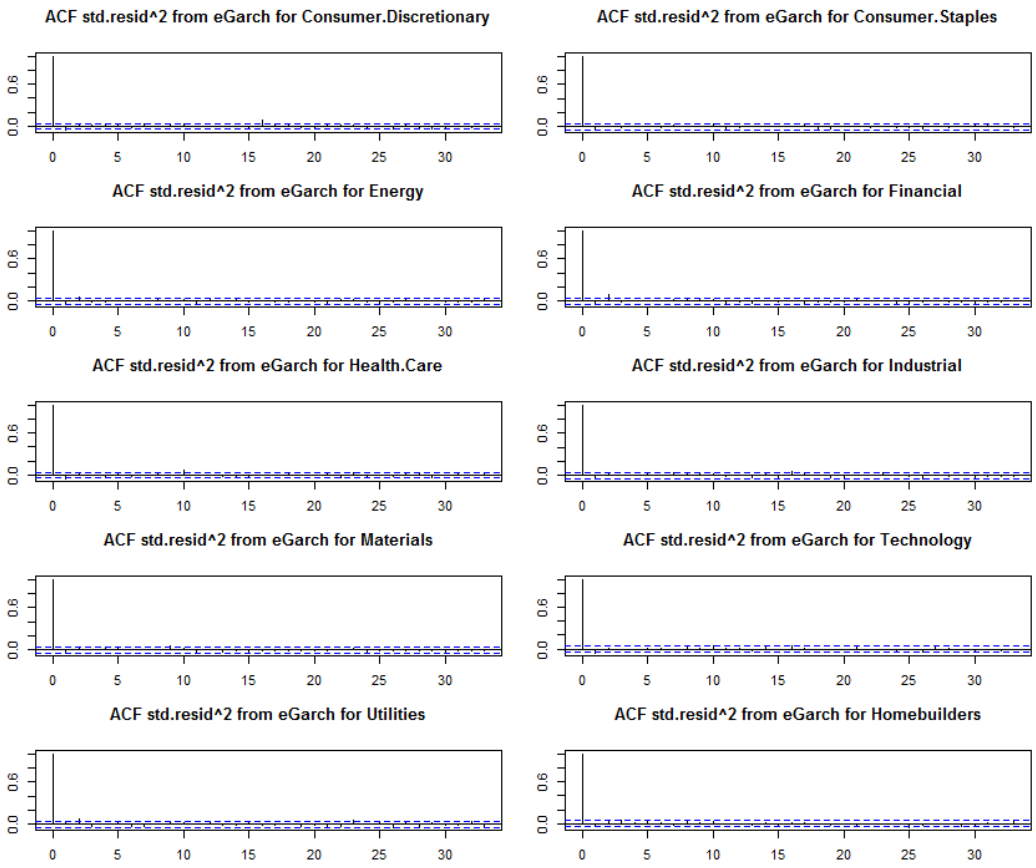


Figure 39 - Squared ACFs for E-GARCH (1,1) standardized residuals

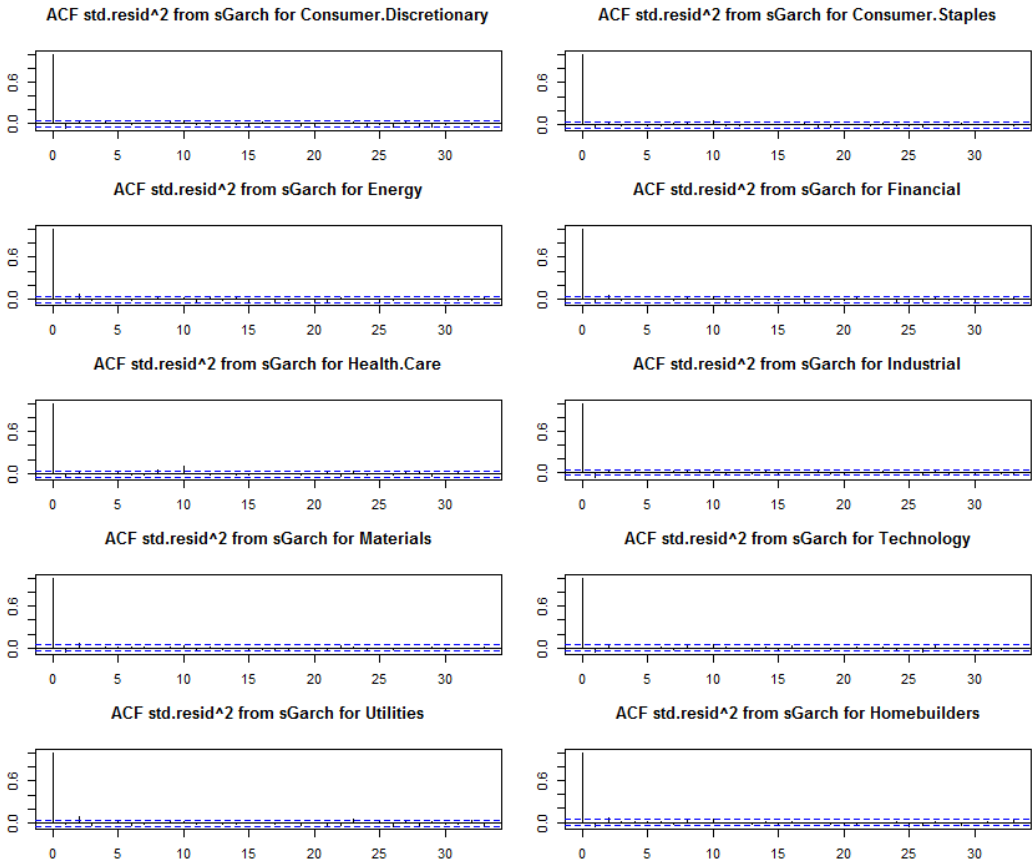


Figure 40 - Squared ACFs for GARCH (1,1) standardized residuals

Table 18 - Univariate S-GARCH(1,1) parameter p values, Equity

	Consumer Discretionary	Consumer Staples	Energy	Financial	Health Care	Industrial	Materials	Technology	Utilities	Homebuilders
mu	0.000002	0	0.000365	0.000277	0	0	0	0	0	0.099598
ar1	0	0.000087	0.018846	0	0.006983	0.00004	0	0.000001	0	0.969839
ma1	0	0.000002	0.007178	0	0.001878	0.000003	0	0	0	0.860042
omega	0.072945	0.036985	0.329081	0.113119	0.033079	0.120358	0.18012	0.067759	0.16055	0.145766
alpha1	0	0	0.000016	0	0	0	0	0	0	0
beta1	0	0	0	0	0	0	0	0	0	0

Table 19 - Univariate E-GARCH(1,1) parameter p values, Equity

	Consumer Discretionary	Consumer Staples	Energy	Financial	Health Care	Industrial	Materials	Technology	Utilities	Homebuilders
mu	0.000001	0	0.000364	0.000176	0.000007	0.000001	0.000937	0.000002	0.000636	0.004464
ar1	0	0	0	0	0	0	0	0.000004	0	0.974119
ma1	0	0	0	0	0	0	0	0	0	0.950125
omega	0	0	0	0	0	0	0	0	0	0.036479
alpha1	0	0	0	0	0	0	0	0	0.000066	0.000002
beta1	0	0	0	0	0	0	0	0	0	0
gamma1	0	0	0	0	0	0	0	0	0	0.076849

B2: Commodity dataset

	statistic	parameter	p.value	method
Cotton	102.3317	2	0	"Jarque Bera Test"
Gold	2100.094	2	0	"Jarque Bera Test"
Corn	185.0201	2	0	"Jarque Bera Test"
Sugar	533.7416	2	0	"Jarque Bera Test"
Feeder.Cattle	90.89546	2	0	"Jarque Bera Test"
Coffee	459.2053	2	0	"Jarque Bera Test"
Cocoa	568.3735	2	0	"Jarque Bera Test"
Brent.Crude	1620.602	2	0	"Jarque Bera Test"
Zinc	245.0847	2	0	"Jarque Bera Test"
Natural.Gas	183.4358	2	0	"Jarque Bera Test"

Figure 41 - Jarque Bera test, Commodity indices

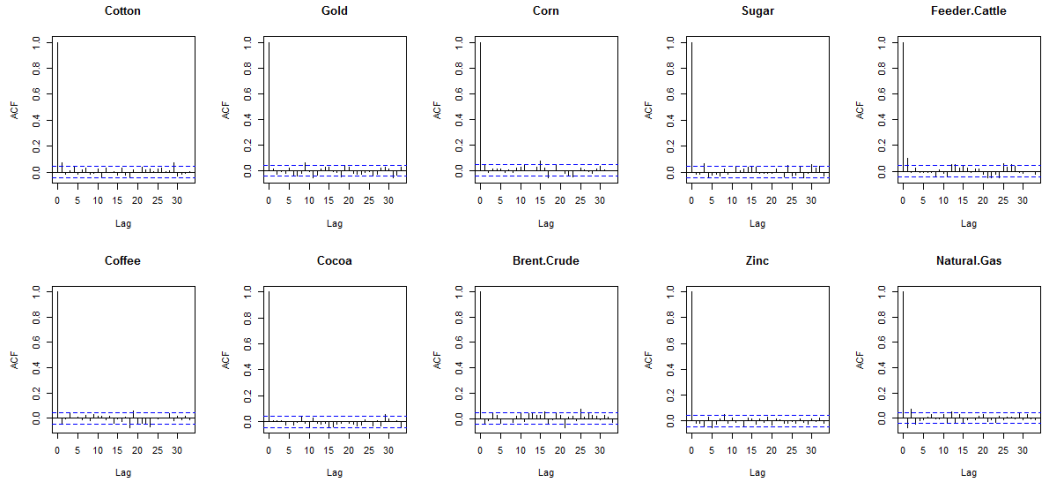


Figure 42 - Linear ACF's Commodity indices

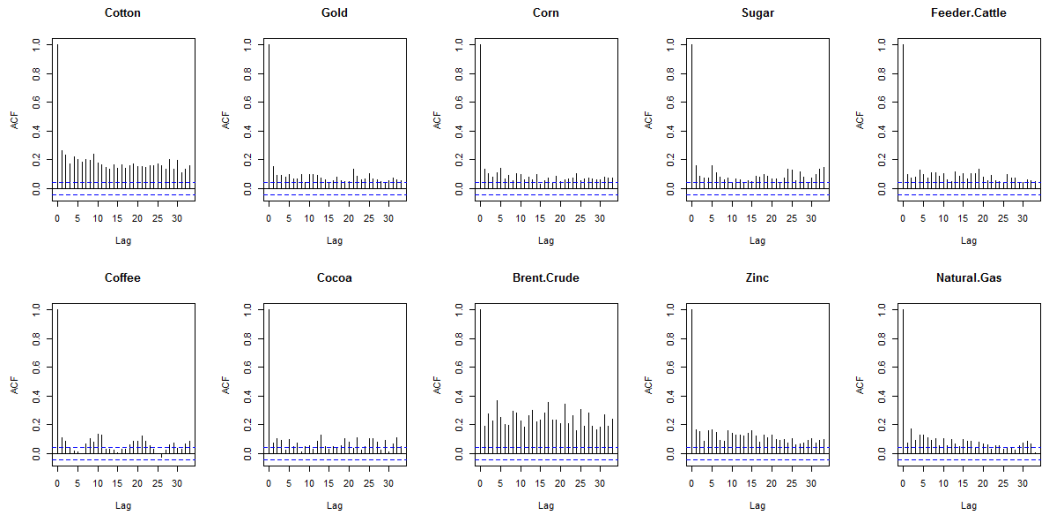


Figure 43 - Squared ACF's Commodity indices

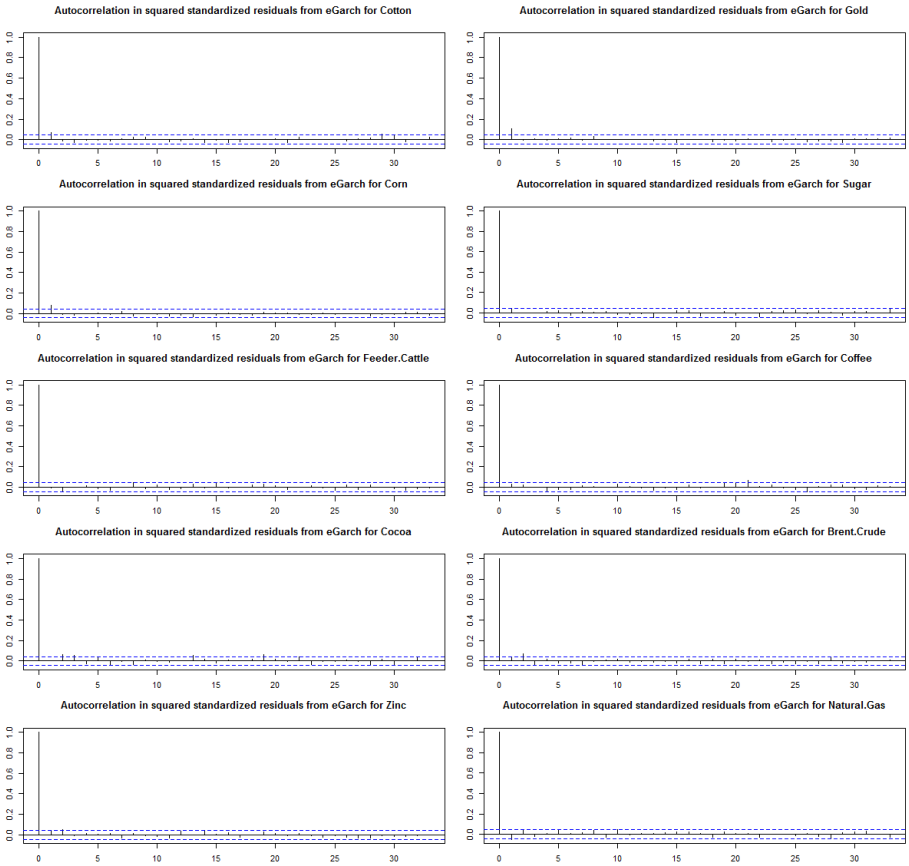


Figure 44 - Squared ACFs for E-GARCH (1,1) standardized residuals

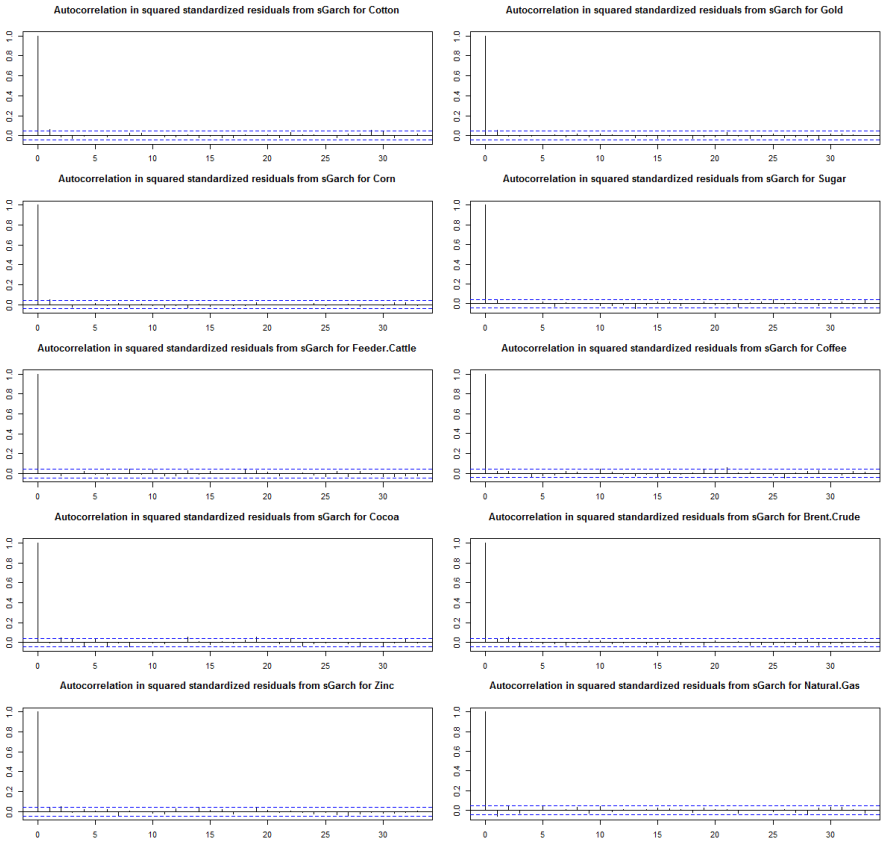


Figure 45 - Squared ACFs for GARCH (1,1) standardized residuals

Table 20 - Univariate S-GARCH(1,1) parameter p values, Commodities

	Cotton	Gold	Corn	Sugar	Feeder Cattle	Coffee	Cocoa	Brent Crude	Zinc	Natural Gas
mu	0.69757	0.103132	0.881941	0.186348	0.383079	0.43078	0.26722	0.313102	0.897282	0.009565
ar1	0.782763	0	0.000008	0.000076	0.245496	0.8011	0.7712	0.680346	0.000001	0.008882
ma1	0.664182	0	0.000001	0.00002	0.079001	0.68376	0.77472	0.709505	0.000002	0.023606
omega	0.211583	0.141847	0.000008	0.381504	0.295511	0	0.67724	0.730668	0.78203	0.001365
alpha1	0.000006	0.000003	0	0	0	0	0	0.028201	0	0
beta1	0	0	0	0	0	0	0	0	0	0

Table 21 - Univariate E-GARCH(1,1) parameter p values, Commodities

	Cotton	Gold	Corn	Sugar	Feeder Cattle	Coffee	Cocoa	Brent Crude	Zinc	Natural Gas
mu	0.834622	0.000052	0.946084	0.047791	0.538818	0.859381	0.045442	0.36659	0.870108	0.001071
ar1	0.000001	0	0	0.000014	0	0.325297	0	0.00261	0.000003	0
ma1	0	0	0	0.000001	0	0.727488	0	0.011674	0.00001	0
omega	0.004338	0	0	0	0	0	0.052878	0.019705	0	0
alpha1	0.651712	0.306217	0.057352	0.232194	0.000007	0.000341	0.779778	0.000192	0.007015	0.77748
beta1	0	0	0	0	0	0	0	0	0	0
gamma1	0.017472	0	0	0.000001	0	0	0	0.007398	0	0

Appendix C: R-code for empirical estimation

The R code was developed using the following packages and resources: Berkelaar and others (2015), Delignette-Muller and Dutang (2015), Fox and Weisberg (2011), Genz and Bretz (2009), Genz et al (2016), Ghalanos, A. (2015a, 2015b), Harter et al (2016), Hofert et al (2015), Hofert and Maechler (2011), Kojadinovic and Yan (2010), Komsta and Novomestky (2015), Luethi and Breymann (2013), Peterson and Carl. (2014, 2015), Rmetrics Core Team, Wuertz and Setz (2014), Rmetrics Core Team, Wuertz et al (2014a, 2014b), Ryan (2015), Ryan and Ulrich (2014), Trapletti and Hornik (2016), Turlach and Weingessel (2013), Wuertz et al (2009, 2013), Yan (2007) and Zeileis and Grothendieck (2005).

As the total code for the thesis is long, we have only included the most important steps.

Excerpt from R script

```
# Dat = training sample
# Dat.all = full sample
k = ncol(Dat)
########## Univariate garch models specified ##########
# standard garch, normal spec
xspec = ugarchspec(mean.model = list(armaOrder = c(1,1)), variance.model =
list(garchOrder = c(1,1), model = "sGARCH"), distribution.model = "norm")
# e-garch spec, t dist
espec = ugarchspec(mean.model = list(armaOrder = c(1,1)), variance.model =
list(garchOrder = c(1,1), model = "eGARCH"), distribution.model = "std")

# Standard multispec
mspec = multispec(replicate(n = k, xspec))
# exponential garch multispec
emspec = multispec(replicate(n = k, espec))

# Fitting the univariate models
# standard garch with normal
multf = multifit(mspec, Dat, solver = "hybrid")
# e-garch with t-dist
emultf = multifit(emspec, Dat, solver = "hybrid")

########## DCC model specified and fitted ##########
## Specifying the dcc models for different ugarchspecs.
# standard garch multispec with normal marginals and multivariate normal
spec1 = dccspec(uspec= mspec, dccOrder = c(1,1), model = "DCC", distribution =
"mvnorm")
# E-garch with t margins and multivariate t dist
e.dccspec = dccspec(uspec= emspec, dccOrder = c(1,1), model = "DCC",
distribution = "mvt")

## Fitting the DCC models with the fitted univariate garch models.
# standard garch multispec with normal marginals and multivariate normal
```

```
fit1 = dccfit(spec = spec1, data = Dat, fit.control = list(eval.se = T), fit = multf)
# E-garch with t margins and multivariate t dist
e.dccfit = dccfit(spec = e.dccspec, data = Dat, fit.control = list(eval.se = T), fit =
emultf)

########## Rolling forecast of the covariance matrix ##########
L.ofs = 500
# 5 Day rolling forecast standard Garch and normal, multi.normal dist.
roll.forecast = function(data){
  multf = multifit(mspec, data, solver = "hybrid")
  fit1 = dccfit(spec = spec1, data = data, fit.control = list(eval.se = T), fit = multf)
  dcc.forecast1 = dccforecast(fit1, n.ahead = 5)
  cor1 = rcor(dcc.forecast1, type="R")
  cov1 = rcov(dcc.forecast1)
  return(list(cor=cor1[[1]][,,5], cov=cov1[[1]][,,5]))
}
# 5 day rolling forecast, EGarch, std mstd.
et.roll.forecast = function(data){
  emultf = multifit(emspec, data, solver = "hybrid")
  e.dccfit = dccfit(spec = e.dccspec, data = data, fit.control = list(eval.se = T), fit =
emultf)
  e.dcc.forecast = dccforecast(e.dccfit, n.ahead = 5)
  ecor = rcor(e.dcc.forecast, type="R")
  ecov = rcov(e.dcc.forecast)
  return(list(cor=ecor[[1]][,,5], cov=ecov[[1]][,,5]))
}

# 5 day ugarch rolling forcast
# sGarch spec
roll.forcast = lapply(seq(0,L.ofs-5,5), function(x)
roll.forecast(Dat.all[1:nrow(Dat)+x]))
# eGarch spec
e.roll.forcast = lapply(seq(0,L.ofs-5,5), function(x)
et.roll.forecast(Dat.all[1:nrow(Dat)+x]))
```

```
# Function for getting a correlation matrix for each out of sample point in an
array. Same for covariance.
cor.roller = function(matrices){
 roll.matrix = array(dim=c(k,k,500))
 loop=1
 loop1=5
 for(i in 1:100){
  roll.matrix[,,loop:loop1] = matrices[[i]]$cor
  loop=loop+5
  loop1=loop1+5
 }
 return(roll.matrix)
}
cov.roller = function(matrices){
 roll.matrix = array(dim=c(k,k,500))
 loop=1
 loop1=5
 for(i in 1:100){
  roll.matrix[,,loop:loop1] = matrices[[i]]$cov
  loop=loop+5
  loop1=loop1+5
 }
 return(roll.matrix)
}

# 500 matrixes for each spec, ie each of the 100 5-day forecasts repeated 5 times
every fifth day.
for.c.cov = cov.roller(roll.forcast)
for.c.cor = cor.roller(roll.forcast)
e.for.c.cov = cov.roller(e.roll.forcast)
e.for.c.cor = cor.roller(e.roll.forcast)

############ Function for simulating with copula with specified correlation and
covariance ##########
# Simulating with normal copula and t marginals
```

```
norm.cop.sim = function(data=Dat, S, P){
  c.var = diag(S)
  para = P[lower.tri(P)]
  G.cop = normalCopula(param=para, dim=k, dispstr='un')
  margin.par = lapply(1:k, function(x) list(mean=mean(data[,x]), sd=sqrt(c.var[x]),
nu=nu[x]))
  G.cop = normalCopula(param=para, dim=k, dispstr='un')
  g.mvdc = mvdc(copula = G.cop,
                margins=c(rep("std", k)),
                paramMargins = margin.par)
  G.sim = rMvdc(mvdc = g.mvdc, n = n)
  return(G.sim)
}
# Simulating with t-copula and t marginals
t.cop.sim =function(data=Dat, S, P){
  c.var = diag(S)
  para = P[lower.tri(P)]
  T.cop = tCopula(param=para, dim=k, dispstr='un', df=3.4)
  margin.par = lapply(1:k, function(x) list(mean=mean(data[,x]), sd=sqrt(c.var[x]),
nu=nu[x]))
  t.mvdc = mvdc(copula = T.cop,
                margins=c(rep("std", k)),
                paramMargins = margin.par)
  T.sim = rMvdc(mvdc = t.mvdc, n = n)
  return(T.sim)
}

########## Simulating data using copula-simulation function for each estimated
DCC matrix ####
## Simulation with the normal copula
cop.dcc.sim.NN = lapply(1:100, function(x) norm.cop.sim(data =
Dat.all[1:(nrow(Dat)+(5*(x-1))),], S=roll.forcast[[x]]$cov,
P=roll.forcast[[x]]$cor))
```

```
cop.dcc.sim.ET = lapply(1:100, function(x) norm.cop.sim(data =
Dat.all[1:(nrow(Dat)+(5*(x-1))),], S=e.roll.forcast[[x]]$cov,
P=e.roll.forcast[[x]]$cor))

## Simulation with student t copula
t.cop.dcc.sim.NN = lapply(1:100, function(x) t.cop.sim(data =
Dat.all[1:(nrow(Dat)+(5*(x-1))),], S=roll.forcast[[x]]$cov,
P=roll.forcast[[x]]$cor))
t.cop.dcc.sim.ET = lapply(1:100, function(x) t.cop.sim(data =
Dat.all[1:(nrow(Dat)+(5*(x-1))),], S=e.roll.forcast[[x]]$cov,
P=e.roll.forcast[[x]]$cor))

########## Regular CVaR and markowitz optimization and equal weights
portfolio ##########

##Equal weights
eq.w = as.xts(matrix(rep(rep(1/ncol(Dat), ncol(Dat)), L.ofs),ncol=ncol(Dat),
nrow=L.ofs), order.by = OOS.index)
colnames(eq.w)=id
eq.w.ret = eq.w*OOS.Dat
eq.w.port.ret = apply(eq.w.ret, 1, sum)
eq.w.cum.ret = matrix(ncol=1, nrow=L.ofs)
for(i in 1:L.ofs){
  eq.w.cum.ret[i] = sum(eq.w.port.ret[1:i])
}
eq.w.cum.ret=as.xts(eq.w.cum.ret, order.by=OOS.index)
weight.plot(eq.w, ma="Equal weights portfolio")

##CVAR
# Weights through time
CVaR.port.weights = t(sapply(1:L.ofs, function(x) mincvar(data =
Dat.all[1:nrow(Dat)+x,],constraints = boxConstraints)))
# Rebalancing
Cvar.port.weights.reb = lapply(seq(1,L.ofs,reb) , function(x)
matrix(rep(CVaR.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
```

```
Cvar.port.weights.reb = do.call(rbind, Cvar.port.weights.reb)
# Return
Cvar.weight.ret = Cvar.port.weights.reb*OOS.Dat
Cvar.por.ret = apply(Cvar.weight.ret,1,sum)
# Cumulative return
cvar.port.cum.ret = matrix(ncol=1, nrow=L.ofs)
for(i in 1:L.ofs){
  cvar.port.cum.ret[i] = sum(Cvar.por.ret[1:i])
}
cvar.port.cum.ret=as.xts(cvar.port.cum.ret, order.by=OOS.index)

##Markowitz
M.port.weights = t(sapply(1:L.ofs, function(x) Min.var.port(Covar = NULL,
data=Dat.all[1:(nrow(Dat)+x),], max.all = max.Alloc)))
# Rebalancing
M.port.weights.reb = lapply(seq(from = 1, to = 500, by = reb), function(x)
matrix(rep(M.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
M.port.weights.reb = do.call(rbind, M.port.weights.reb)
# Return
M.weight.ret = M.port.weights.reb*OOS.Dat
M.port.ret = apply(M.weight.ret,1,sum)
# Cumulative return
M.port.cum.ret = matrix(ncol=1, nrow=L.ofs)
for(i in 1:L.ofs){
  M.port.cum.ret[i] = sum(M.port.ret[1:i])
}
M.port.cum.ret=as.xts(M.port.cum.ret, order.by=OOS.index)

#### standard garch, normal margins, mvnorm. DCC portfolio ####

sn.port.w = sapply(1:L.ofs, function(x) Min.var.port(Covar = for.c.cov[,,x], data =
Dat.all[1:(nrow(Dat)+x),], max.all = max.Alloc))
sn.port.w = t(sn.port.w)
dim(sn.port.w) # 500x10
# Returns
```

```
sn.w.ret = OOS.Dat*sn.port.w
colnames(sn.w.ret) = id
index(sn.w.ret) = OOS.index
sn.port.ret = as.xts(apply(sn.w.ret, 1, sum))
colnames(sn.port.ret) = "sGarch dcc min.var ret."
#### e-garch, t margins, mvt. DCC portfolio ####
ett.port.w = sapply(1:L.ofs, function(x) Min.var.port(Covar = e.for.c.cov[,,x], data
= Dat.all[1:(nrow(Dat)+x),], max.all = max.Alloc))
ett.port.w = t(ett.port.w)
dim(ett.port.w) # 500x10
# Returns
ett.w.ret = OOS.Dat*ett.port.w
colnames(ett.w.ret) = id
index(ett.w.ret) = OOS.index
ett.port.ret = as.xts(apply(ett.w.ret, 1, sum))
colnames(ett.port.ret) = "eGarch dcc min.var ret"

#### standard garch, normal margin, mvnorm, CCC portfolio ####
CCC.cor = cor(Dat)
# Calculating rolling sigma forecasts
sgarchsigma = lapply(1:100, function(x) matrix(rep(diag(roll.forcast[[x]]$cov),5),
ncol=k, nrow=5, byrow=T))
sgarchsigma = do.call(rbind, sgarchsigma)
# Covariance-matrixes
CCC.cov = lapply(1:500, function(x)
(as.vector(sgarchsigma[x,]))%*%t(as.vector(sgarchsigma[x,]))*CCC.cor)
# Portfolio weights
CCC.port.w = sapply(1:L.ofs, function(x) Min.var.port(Covar = CCC.cov[[x]],
max.all = max.Alloc))
CCC.port.w = t(CCC.port.w)
dim(CCC.port.w) # 500x10
head(CCC.port.w)
# Returns
CCC.w.ret = OOS.Dat*CCC.port.w
colnames(CCC.w.ret) = id
```

```
index(CCC.w.ret) = OOS.index
CCC.port.ret = as.xts(apply(CCC.w.ret, 1, sum))
colnames(CCC.port.ret) = "CCC port.ret"

########## CVAR optimization based on the copula-simulated data ##########

## optimal weights based on the normal copula sgarch-normal-normal simulated
data.
Ncop.NN.port.weights = t(sapply(1:100, function(x) mincvar(data =
cop.dcc.sim.NN[[x]],constraints = boxConstraints)))
#rebalanced
Ncop.NN.port.weights.reb = lapply(1:100, function(x)
matrix(rep(Ncop.NN.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
Ncop.NN.port.weights.reb = do.call(rbind, Ncop.NN.port.weights.reb)
# Returns
Ncop.NN.weights.ret = Ncop.NN.port.weights.reb*OOS.Dat
Ncop.NN.port.ret = apply(Ncop.NN.weights.ret, 1, sum)

## optimal weights based on the t copula sgarch-normal-normal simulated data.
Tcop.NN.port.weights = t(sapply(1:100, function(x) mincvar(data =
t.cop.dcc.sim.NN[[x]],constraints = boxConstraints)))
# rebalanced
Tcop.NN.port.weights.reb = lapply(1:100, function(x)
matrix(rep(Tcop.NN.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
Tcop.NN.port.weights.reb = do.call(rbind, Tcop.NN.port.weights.reb)
# Returns
Tcop.NN.weights.ret = Tcop.NN.port.weights.reb*OOS.Dat
Tcop.NN.port.ret = apply(Tcop.NN.weights.ret, 1, sum)

# optimal weights based on the normal copula egarch-t-t simulated data.
Ncop.ET.port.weights = t(sapply(1:100, function(x) mincvar(data =
cop.dcc.sim.ET[[x]],constraints = boxConstraints)))
# rebalanced
Ncop.ET.port.weights.reb = lapply(1:100, function(x)
matrix(rep(Ncop.ET.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
```

```
Ncop.ET.port.weights.reb = do.call(rbind, Ncop.ET.port.weights.reb)
# Returns
Ncop.ET.weights.ret = Ncop.ET.port.weights.reb*OOS.Dat
Ncop.ET.port.ret = apply(Ncop.ET.weights.ret, 1, sum)

# optimal weights based on the t- copula egarch-t-t simulated data.
Tcop.ET.port.weights = t(sapply(1:100, function(x) mincvar(data =
t.cop.dcc.sim.ET[[x]],constraints = boxConstraints)))
# rebalanced
Tcop.ET.port.weights.reb = lapply(1:100, function(x)
matrix(rep(Tcop.ET.port.weights[x,], reb), ncol=k, nrow=reb, byrow=T))
Tcop.ET.port.weights.reb = do.call(rbind, Tcop.ET.port.weights.reb)
# Returns
Tcop.ET.weights.ret = Tcop.ET.port.weights.reb*OOS.Dat
Tcop.ET.port.ret = apply(Tcop.ET.weights.ret, 1, sum)
```

***** mincvar and Min.var.port are functions for minimizing CVaR and variance of a portfolio respectively, each returning a vector of optimal weights. This excerpt from the full script will not run on insertion into R, and are included for illustrational purposes.

Appendix D: Preliminary Thesis

BI Norwegian Business School – GRA 19003 Preliminary Thesis Report

Master of Science in Business with Major in Finance

Portfolio Optimization: A Comparison of Risk Measures

By Johannes Andreas Barstad 0934937
Olve Heitmann 0930663

Supervised by Johann Reindl

15.02.2015

Introduction

In this thesis we seek to compare various risk measures in the context of portfolio optimization and risk diversification. We are interested in measures and approaches that can be, or are being, used by practitioners in the finance industry. Theoretical superiority is thus not the sole criterion for choosing measures to test, but also feasibility. To make the thesis as relevant to the real world as possible, we intend to use Value at Risk (VaR) and/or Conditional Value at Risk (CVaR, also known as expected shortfall/ES) constraints to the portfolio. The risk measures we are currently considering is the most popular multivariate GARCH models (see methodology section) and copula based approaches. The research questions we seek to answer is then:

- Which of these approaches seem to perform the best in the long run?
- Are some of the measures performances affected more by a certain type of financial climate (boom/recession) than the others?
- Do certain measures perform better in certain sectors (i.e. the BEKK model in the oil service provider sector), or does one measure trump the others regardless of sector?
- Other questions that arise during the process.

The foundation for modern portfolio theory as a mathematical and statistical problem was laid out by Markowitz in 1952. Markowitz' argues that simply maximizing discounted, expected returns as an investment rule is rejected both as a hypothesis to explain historic investor behavior, and as a maxim to guide optimal investment. Instead he considers a rule that expected return is a desirable thing while variance of return is undesirable. This idea implies a demand for diversification, not solely by reducing variance through increasing numbers of different securities held, but also choosing assets with low covariance.

Modelling volatility (measured by the standard deviation) and covariance in financial time series is crucial for good decision-making. Volatility is well documented in the literature to be time-varying, as well as moving in "clusters" of high and low volatility. Much has been written on this topic, with variants of the *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) models

being among the most popular.
Multivariate GARCH models however have a major drawback. The number of parameters can become very large as the cross section of stocks increases, generating a “dimensionality curse”. Existing approaches work around this problem by either relying on constraints to the structure of the model, or using alternative estimation criterion to the usual quasi likelihood criterion.
In this thesis we seek to employ and Equation by Equation (EbE) framework originally proposed by Engle and Sheppard (2001), and Engle (2002). This approach will allow us to avoid putting constraints to the structure of the model, while using quasi likelihood estimation.

Literature review

Portfolio optimization

From Markowitz’ paper on portfolio optimization, an optimal portfolio is determined based on expected return on the assets, and variance as the measure of risk. By minimizing the variance-covariance of the assets in the portfolio subject to a return requirement, we get a set of optimal portfolios. These different combinations map out the efficient frontier, which is the optimal relationship between expected return and risk for different portfolio compositions, i.e. the lowest risk/reward relationship.

The return of security i, r_i, is a random variable. The return of a portfolio, r_p, is a weighted average of the returns of the individual assets included in the portfolio, i.e. also a random variable.

$$r_p = \sum_{i=1}^{n} w_i r_i$$

Subject to the constraint:

$$\sum_{i=1}^{n} w_i = 1$$

The setup of Markowitz has some major drawbacks, however. By using variance as a measure of risk, it is assumed that the returns are jointly normally distributed.

This is in fact not the case in the real world, as the distribution of returns generally are left skewed and leptokurtic. Another drawback is that the variance-covariance is constant over time, while in the real world the variance and covariance tend to change with unexpected events. Due to this, sample variance and covariance as measure of risk does not adequately capture the real distribution of returns. This will lead us to underestimate the potential losses, and the weights of the portfolio is not optimal.

Another drawback of the original model is that it equally weighs upside and downside moves of the return distribution, while for risk averse investors the emphasis should be placed on the downside risk of the portfolio, i.e. the risk of large losses. Measures such as value at risk (VaR) and expected shortfall/conditional value at risk (CVaR) takes this into account.

Risk

We can define risk as the uncertainty regarding a future event, in this case asset returns. To measure the risk of a process, we need an appropriate risk measure. Artzner *et al.* (1999) presents the concept of coherent risk measures. For a risk measure to be coherent, it needs to satisfy a set of axioms:

If $r_1$and r_2 are random variables, and $\rho(\cdot)$ is the risk measure, then the axioms states:

1. **Translation invariance (Axiom T):**

   $$\rho(r_1 + k) = \rho(r_1) - k$$

   When adding quantity k to the asset, the risk is reduced by the same amount.
2. **Subadditivity (Axiom S):**

   $$\rho(r_1 + r_2) \leq \rho(r_1) + \rho(r_2)$$

   Risk can be reduced by diversification. The risk of the combined portfolio is less or equal to the sum of the risk for the individual assets.
3. **Positive homogeneity (Axiom PH):**

   $$\rho(kr_1) = k\rho(r_1)$$

   By increasing the amount invested in the asset by factor k, the risk is increased by the same factor.
4. **Monotonicity (Axiom M)**

$$\rho(r_1) \leq \rho(r_2), r_1 \leq r_2$$

IF the value of r1 is greater than r2 then the risk of r1 is less than that of r2.

Risk measures

Value at risk:

Value at risk is the maximum level of expected losses given a confidence level, over a specified time interval. In other words, the confidence level chosen states the probability of losses exceeding the VaR in the specified time horizon. We can view the loss function of a portfolio as the negative of the return of the portfolio:

$$L(\boldsymbol{w}, \boldsymbol{r}) = -(w_1 r_1 + \cdots + w_n r_n) = -\boldsymbol{w}^T \boldsymbol{r}$$

Where $\boldsymbol{r}$ is the returns of the assets, and $\boldsymbol{w}$ is the portfolio weights. The probability of the loss $L(\boldsymbol{w}, \boldsymbol{r})$ not exceeding l is defined as:

$$F(\boldsymbol{w}, l) = \int_{L(\boldsymbol{w},\boldsymbol{r}) \leq l} f(\boldsymbol{r}) d\boldsymbol{r}$$

Where $f(\boldsymbol{r})$ is the joint density function of returns and $F(\boldsymbol{w}, l)$ is the cumulative distribution function for losses. The VaR is given by:

$$VaR_\alpha(\boldsymbol{w}) = \min(l : F(\boldsymbol{w}, l) \geq \alpha)$$

Where 1- α is the confidence level chosen.

VaR has some large drawbacks. First, it does not take into account the actual extent of the losses exceeding the limit. Second, it does not fulfill the non-subadditivity axiom which means that it is not by itself supporting diversification. Another drawback is that the regular value at risk measure is assuming normality, while return series in reality is generally leptokurtic. This means that VaR is underestimating the frequency of large losses when assuming normality.

Copula

We are considering a copula based approach for estimating the dependence structure between assets. This is due to the criticism of covariance as a measure of dependence, due to multivariate normality assumption of returns. In the copula

framework, the individual distributions can be joined through the copula function, allowing us to have non-normality in the individual distributions.

Data

We intend to collect data for the thesis through DataStream. For our data sample we intend to use time series data from the US stock market, starting January 2006, ending December 2015. This data sample includes events such as the financial crisis of 2008, allowing us to monitor the models performance in relation to volatility clustering and leverage effects. We will study the daily logarithmic relative differences of market prices, known as *log-returns.* Using logarithmic scaling is the standard in financial time series studies, and has several advantages;

- Compounded log returns can conveniently be computed by summation
- Returns can be interpreted as continuously compounded returns, so that the compounding frequency become irrelevant and returns across assets can be compared. This matters in regards to benchmarking.

Methodology

To answer the research questions from the introduction, we plan to build several portfolio optimization models. Each measure we decide to test should result in one separate model. The models will require input in term of price information from the market, and should give us both a starting portfolio, as well as signals of when and by how much weights should be rebalanced.

Rebalancing

We are not fully decided upon how to model rebalancing as of yet. The options we consider are:

- Continuous rebalancing, ignoring transaction costs. This means that every time new information is added, the model should rebalance.
- Modelling transaction costs, and rebalance only when the model says its “optimal” considering said costs
- Ignoring transactions costs, but only allow for rebalancing if the model projects optimal move is to change the weights by above a certain

minimum. E.g. if the minimum change in portfolio weight is 10%, and the current weight of stock A is 15%, we would need new optimal weight to be below 13,5% or above 16,5%.

The goal of the thesis is to compare the risk measures, yet to do so in a realistic manner. Option 1 is the by far easiest to implement, option 2 the most realistic and option 3 somewhere in between the two.

For the discussion of the methodology we intend to use, the concepts of conditional and unconditional variance is needed. Unconditional variance is simply what is most often referred to as just variance, defined as:

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Conditional variance on the other hand, is conditional on the previous knowledge (history) we have of that variable. If we are interested in the conditional variance of the error term in a regression, u_t, denoted as ${\sigma_t}^2$, we can write it as

$${\sigma_t}^2 = var(u_t|u_{t-1}, u_{t-2}, \dots) = \mathbb{E}[(u_t - \mathbb{E}[u_t])^2|u_{t-1}, u_{t-2}, \dots]$$

Further, it's common to assume $\mathbb{E}[u_t] = 0$, so

$${\sigma_t}^2 = var(u_t|u_{t-1}, u_{t-2}, \dots) = \mathbb{E}[{u_t}^2|u_{t-1}, u_{t-2}, \dots]$$

Stating that the conditional variance of a zero mean normally distributed random variable equals the conditional expected value of the squared random variable (Brooks 2008, p.387-388).

The conditional variance is then allowed to vary over time as a function of previous error terms (residuals).

GARCH

The GARCH model originated as an extension of Engle's (1982) ARCH model, and was first introduced by Bollerslev in 1986 to "allow for both a longer memory and more flexible lag structure" (Bollerslev 1986, page 2).
To illustrate for our application; let y_t be a stochastic process of daily returns, with mean equation AR(s)

$$y_t = \mu + \sum_{i=1}^{s} \propto_i y_{t-i} + u_t$$

Then a GARCH(p,q) process is formally defined as

$$u_t = \sigma_t z_t$$

$$\sigma_t{}^2 = \propto_0 + \sum_{i=1}^{p} \propto_i u_{t-i}{}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}{}^2$$

This is the univariate case. In portfolio optimization, covariances between the securities are arguably even more important than the variance of the individual stocks themselves.

The last 20 years of financial time series research a significant portion has been devoted to multivariate extensions of GARCH. The most popular specifications seem to be the Constant Conditional Correlations (CCC) model by Bollerslev (1990) Jeantheau (1998), the BEKK model by Baba *et al.* (1995), and the Dynamic Conditional Correlations (DCC) models proposed by Tse and Tsui (2002) and Engle (2002).

The use of MGARCH models in applied works has suffered to their complexity. In portfolio management, cross-sections of hundreds of stocks are not uncommon. When the dimension of the cross section increases, the number of parameters to be estimated can become very large in MGARCH models. This is the case for most multivariate time series, but maybe even more so in GARCH models as the parameters in the conditional variance matrix has to be inverted in Gaussian likelihood-based estimation methods. Existing approaches work around this problem by either relying on constraints to the structure of the model, or using alternative estimation criterion to the usual quasi likelihood criterion. Workarounds of the first type has been presented in Engle *et al's* Factor ARCH models in 1990, van der Weide's Generalized Orthogonal GARCH model (2002), and Lanne and Saikkonen's Generalized Orthogonal Factor GARCH model (2008). The second type approach was presented by Engle *et al.* in 2008, using a composite likelihood instead of the usual quasi-likelihood. Engle and Kelly also presented a workaround combining the two concepts in the Dynamic Equicorrelation (DECO) model (2012).
The solution we apply in this thesis is an equation by equation (EbE) framework

initially proposed by Engle and Sheppard (2001) and Engle (2002) in the context of DCC models, and later by Pelletier (2006) for regime-switching dynamic correlation models. The EbE approach alleviates the dimensionality curse in two steps;

1. Estimate univariate GARCH models for each individual series (equation by equation)

2. Standardized residuals from the individual stages are used to estimate parameters of the dynamic correlation matrix.

From Engle and Sheppard's paper we know that the standard errors of the first step remain consistent, and only the standard errors for the correlation matrix needs to be modified.

The statistical properties of this approach (involving "two-step estimators") had not been established until recently by Francq and Zakoïan in a working paper from late 2014. That paper establishes strong consistency and asymptotic normality (CAN) in a very general framework, including DCC and CCC models.

Ranking the measures

To assess the performance of the risk measures and in turn rank them, we will perform in and out of sample testing. If the portfolios then is constrained by having the same VaR and/or CVaR, return should be the only relevant criteria, and we should be able to rank the measures. We are also considering adding in other types of performance measures.

Biblography

Markowitz, H. 1952. “Portfolio selection”. *Journal of Finance* 7 (1): 77-91

Francq, C and Zakoïan, J-M. 2014. “Estimating multivariate GARCH and Stochastic Correlation models equation by equation” [Online], CREST and Université de Lille (EQUIPPE), working paper available at: http://perso.univ-lille3.fr/~cfrancq/Christian-Francq/statistics-econometrics-papers/EbEE.pdf

Engle, R.F. and Shepppard, K. 2001. “Theoretical and empirical properties of dynamic conditional multivariate GARCH”. University of California San Diego, discussion paper.

Engle, R.F. 2002. “Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models”. *Journal of Business and Economic Statistics* 20 (3): 339-350

Chamberlain, G. 1983. “A characterization of the distributions that imply mean-variance utility functions”. *Journal of Economic Theory* 29: 185-201

Owen, J. and Rabinovitch, R. 1983. “On the class of elliptical distributions and their applications to the theory of portfolio choice” *Journal of Finance* 38(3): 745-752

Chicheportiche, R. and Bouchaud, J.P. 2012. “The joint distribution of stock returns is not elliptical” *International Journal of Theoretical and Applied Finance* 1-23

Artzner P., Delbaen F., Eber J.M., Heath D. 1999. “Coherent measures of risk” *Mathematical finance* 9:203-228

Brooks, C. 2008. Introductory Econometrics for Finance, Second edition. Cambridge University Press: New York

Bollerslev, T. 1986. “Generalized autoregressive conditional heteroskedasticity” *Journal of Econometrics* 31: 307-327

Engle, R.F. 1982. “Autoregressive conditional hetroscedasticity with estimates of the variance” *Econometria* 50: 987-1007

Herwartz H. and Raters F.H.C., 2015, “Copula-MGARCH with continuous covariance decomposition”, *Economics Letters* 133: 73-76

Jondeau, E. and Rockinger M. 2006. “The Copula-GARCH model of conditional dependencies: An international stock market application”, *Journal of International Money and Finance* 25: 827-853

Rockafellar R.T. and Uryasev S. 2002. “Conditional value-at-risk for general loss distributions”, *Journal of Banking & Finance* 26: 1443-1471