ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

TITLE
“Portfolio Optimization using Coherent Risk Measures: The case of Conditional Value at Risk”

By

Eleni P. Perdikouri

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece

May 2011

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΤΙΤΛΟΣ

«Βελτιστοποίηση Χαρτοφυλακίου με την χρήση Συνεπών Μέτρων Κινδύνου: Η Περίπτωση του Conditional Value at Risk»

Ελένη Π. Περδικούρη

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Μάιος 2011

ACKNOWLEDGEMENTS

I would like to thank Professor A. Yannacopoulos for supervising the completion of this dissertation and I am grateful for his assistance and his comments concerning my thesis.

Furthermore, I am grateful to the faculty and staff of the program of Master of Science in Statistics of the Athens University of Economics and Business.

My special thanks to my family, I am deeply indebted for their support in everything I do. I also would like to thank my dearest friends J.Sagriotis, M. Darivianaki and Z. Papanikolaou for their encouragement and their support.

VITA

I was born in Athens on the 8th of January 1983. I graduated from High School Arsakeio and after taking the National Examinations I entered on the Department of Mathematics of the University of Athens. In 2009 I graduated and, at the same year, I entered in the Master of Science of Statistics of the Athens University of Economics and Business.

ABSTRACT

Eleni P. Perdikouri

“Portfolio Optimization using Coherent Risk Measures: The case of Conditional Value at Risk”

May 2011

This thesis analyzes the notion of Value at Risk and Conditional Value at Risk as risk measures. The aim of this work is to describe the idea of Value at Risk and Conditional Value at Risk, its properties and its ways of computing. Furthermore, this study points out the advantages and disadvantages of both risk measures. The central point of this work is to present the way that VaR and CVaR can be used to optimize a portfolio, whether the portfolio is linear or not, with main objective to reduce as possible the potential risk.

Keywords: Value at Risk, Conditional Value at Risk, Optimization of the portfolio, Coherent Risk Measures

ΠΕΡΙΛΗΨΗ

Ελένη Π. Περδικούρη

«Βελτιστοποίηση Χαρτοφυλακίου με την χρήση Συνεπών Μέτρων Κινδύνου: Η Περίπτωση του Conditional Value at Risk»

Μάιος 2011

Η διπλωματική αυτή εργασία αναλύει την έννοια του Value at Risk και Conditional Value at Risk σαν μέτρα κινδύνου. Σκοπός της εργασίας είναι να περιγράψει την ιδέα των Value at Risk και Conditional Value at Risk, τις ιδιότητές τους και τους τρόπους υπολογισμού τους. Επίσης, περιγράφονται τα προτερήματα και τα ελαττώματα των δύο αυτών μέτρων κινδύνου. Κύριος άξονας της εργασίας είναι να παρουσιάσει τον τρόπο που το VaR και το CVaR μπορούν να χρησιμοποιηθούν για να βελτιστοποιήσουν ένα πιθανό χαρτοφυλάκιο, είτε αν το χαρτοφυλάκιο είναι γραμμικό είτε μη γραμμικό, έτσι ώστε να μειωθεί όσο το δυνατόν περισσότερο το ρίσκο και η πιθανότητα ζημίας.

Keywords: Value at Risk, Conditional Value at Risk, Βελτιστοποίηση γραμμικού και μη γραμμικού χαρτοφυλακίου, Συνεπή Μέτρα Κινδύνου

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The main concern of an investor is to have a specific answer to the question “What is the worst possible loss from my investment?”. In other words, every investor would like to know, a priori if possible, how much risk he/she is going to take, applying a chosen strategy.

There is not a precise definition of the term *risk* in the literature because different investors are able to have different conceptions of risk and so different approaches, or else different strategies, in order to reduce the possibility of losses. In some sense, risk is a ***subjective*** concept, which probably is the main characteristic of risk. For that reason, even if someone can identify some desirable features of a risk measure or a precise strategy, probably there is not a unique one to solve an investor’s problem.

Looking back at the history of risk, the optimal investment decision traditionally refers to the solution of an expected utility maximization problem. Though, despite the fact that risk is a subjective decision, there is a possibility to state some common risk characteristics in order to select the optimal choices of investors. That’s why risk measures were created.

Risk measures are statistical tools, which are used by investors in finance, to make predictions about investment risk. Risk measures are part of the portfolio theory, which relies on the use of standardized tools and predictions to make decisions about how and where to invest. This was made in order to find a way to limit risks and, in the same time, to maximize returns. For example, standard deviation is a very common measure of risk, which measures how much the return of an investment varies comparing with the expectations, which are relying in historical performance and the data.

The main assumption for constructing an optimal portfolio was that the variance of the return distribution was the only tool needed for characterizing the risk of an investment. However, a whole new philosophy was created with the introduction of the notion of *Value at Risk (VaR)* by RiskMetrics (1996). The main idea behind the use of VaR is the fact that a single number is able to encapsulate all the information needed about the possible portfolio losses. The losses are implied by the left hand side tail of the return distribution. VaR became quickly a useful tool for computing the potential losses, such that the managers or the investors were taking their decisions based on VaR.

However, in 1999 Artzner et al. laid out a paper with the desirable mathematical properties that a risk measure must satisfy in order to be useful and precise. Artzner et al. in the paper "Coherent measures of Risk" defined the term of *"coherent risk measures"* and their properties. VaR did not satisfy all of them, though. As an alternative to VaR, a new risk measure was created called *Conditional Value at Risk (CVaR)*. The CVaR is an extension of the notion of VaR, providing the investors with more conservative results and, additionally, is a coherent risk measure. The first introduction of this measure was made in 2000 by Rockafellar and Uryasev in their paper "Optimization of Conditional Value at Risk". In that paper, Uryasev and Rockafellar used this measure for portfolio optimization in conjunction with linear programming algorithm. This combination made CVaR popular in academic community. Furthermore, in the last years there is a tremendous growth of the computing power which has made the computation of CVaR extremely ease.

The organization of this thesis is as follows: in section 2 I introduce the most common and widespread risk measure *Value at Risk* along with the idea of VaR and ways of computing VaR. In section 3 I describe the coherent risk measures and the reasons why Value at Risk is not a coherent risk measure. In section 4 I present an alternative risk measure, *Conditional Value* *at Risk* which has better properties than Value at Risk and is able to product more precise results. I analyze the properties of Conditional Value at Risk have and the way of computing it. In section 5 I present ways for computing VaR and CVaR n several cases, such as for linear portfolios with elliptic distribution and for non linear portfolios.

CHAPTER 2

VALUE AT RISK

2.1 Risk measures

The first time that the term Value at Risk was initiated was in the late 1980s. The event that caused the initiation was the stock market crash of 1987. This was the first major financial crash which pointed out that the standard statistical models used until then, were insufficient since none of them came even close to predict the crash.

At the same time, the commercial banks and the trading portfolios were becoming not only larger, but also more volatile. So the need to create a trustworthy risk measure was imperative. In the early 1900s, the financial events that occurred found many firms in trouble because they were not capable to cover their potential damage.

VaR was developed as a systematic way of quantifying extreme events, occurring in long-term history and broad market events from everyday price movements. Development was most extensive at J.P. Morgan, which not only published a methodology for the utility and application of VaR, but also gave free access to its data base. That was the very first time that the use of VaR surpassed a small group of analysts and was extended to a large scientific, and not only, community. Two years later, the methodology was integrated into an independent for- profit business, now part of RiskMetrics Group.

In 1997, the Securities and Exchange Commission (SEC) ruled that all public corporations must give out quantitative information concerning their activity in derivatives. That rule in combination with the vast usage of VaR, made some banks to choose to implement the rule by including VaR information to their financial statements.

Though, the greater adoption of Value at Risk as a risk measure was invoked by Basel II, beginning in 1999 until nowadays. According to paragraph 178 of Basel II *"As an alternative to the use of standard or own-estimate haircuts, banks may be permitted to use a VaR models approach to reflect the price volatility of the exposure and collateral for repo-style transactions, taking into account correlations effects between security positions. This approach would apply to repo-style transactions covered by bilateral netting agreements on a counterparty-by-counterparty basis. In addition, other similar transactions (like prime brokerage), that meet the requirements for repo-style transactions are also eligible to use the VaR models. The VaR models approach is available to banks that have received supervisory recognition for an internal market risk model under the Market Risk Amendment. Banks which have not received supervisory recognition for use of models under the Market Risk Amendment can separately apply for supervisory recognition to use their interval VaR models for calculation of potential price volatility for repo-style transactions. Interval models will only be accepted when a bank can prove the quality of its model to the supervisor through the back testing of its output using one year of historical data. "*

As we can see from above, Basel II has given the possibility to the banks to use VaR models to compute their volatility of the portfolios officially.

2.2 Value at Risk

Value at Risk (VaR) is a measure of risk in finance theory and as we already mentioned, probably the most popular.

Generally, Value at Risk measures the potential loss in value of a risky asset or portfolio over a defined period for a given confidence interval. In the context of finance, Value at Risk is an estimate, with a given degree of confidence, of how much one can lose from a portfolio over a given time horizon. Thus, if the VaR of an asset is 50 million euro at a one-week, 95% confidence level, there is only 5% chance that the value of the asset will drop more than 50 million € over any given week.

Let X be a random variable, with cumulative distribution $F_X(z) = P(X \leq z)$. Generally, X may have the meaning of loss or gain, though in our study with X we denote the potential losses. The distribution F_X is called the loss distribution. To define VaR, first we must set:

a) The time horizon (the period) Δ, in which we intend to keep our assets.
b) The confidence level at which we want to estimate our losses.
c) The monetary unit we use.
d) The probability distribution of F.

Definition 2.1: Value at Risk[31]

The VaR of X with confidence level α∈[0,1] is:

$$\boxed{\mathrm{VaR}_\alpha(X) = -\min\{z|F_X(z) \geq \alpha\}} \tag{2.1a}$$

From definition (2.1a) we can clearly see that $\mathrm{VaR}_\alpha(X)$ is the lower α-percentile of the variable X, i.e. is the smallest value such that the probability that losses exceed or equal this value is greater or equal to α . Note that if X follows a continuous distribution then $F_X(z) = \alpha$.

An alternative way of presenting VaR is the followed:

$$Prob\left[\Delta\widetilde{P}(\Delta t, \Delta\tilde{x}) > -VaR\right] = 1 - \alpha \qquad (2.1b)$$

where $\Delta\widetilde{P}(\Delta t, \Delta\tilde{x})$ denotes the change in the market value of our portfolio in the period Δt. $\Delta\tilde{x}$ is the vector of changes of the variable $\tilde{x}$.

Typical values for α are α=0.95 or α=0.99. Usually, in market risk management we use a time horizon Δ of 1 or 10 days, in credit risk management and operational risk management Δ is usually one year. At this point, we have to note that, by its definition, VaR at confidence level α does not give any information about the severity of losses which occur with a probability less than 1-α [26], which is clearly a drawback of VaR as a risk measure.

In Figure 2.1, we depict graphically the notion of VaR. The VaR of a portfolio at the confidence level α is given by the smallest number z, such that the probability that the loss X will exceed z is no larger than 1-α. At this point, we must make clear that VaR represents the losses. Thus, it usually is a positive number. If we find a negative VaR that means that we have a good possibility of gaining.

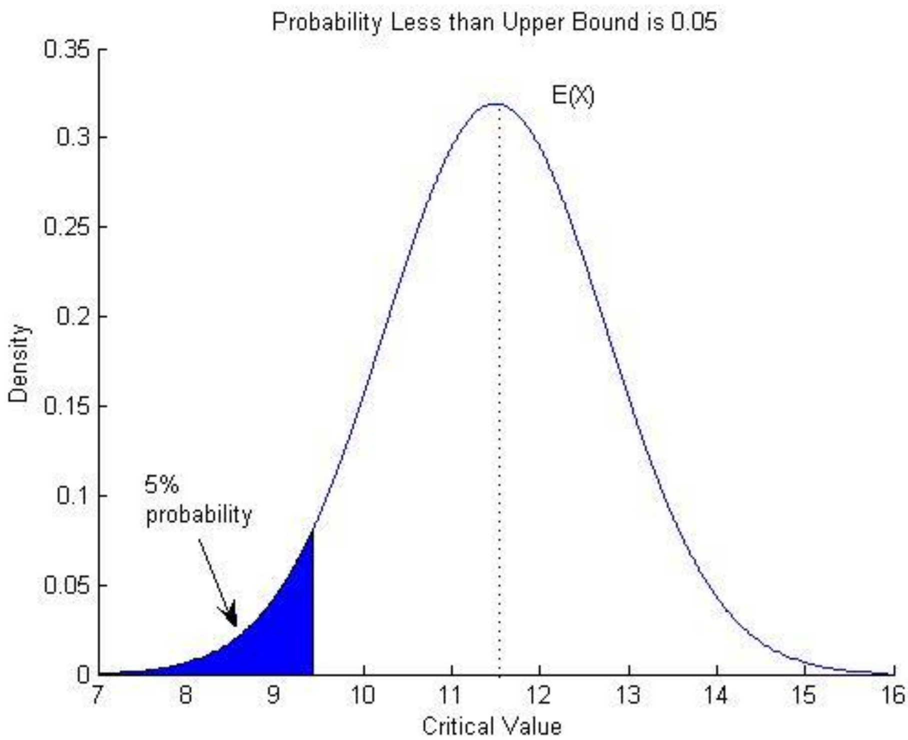

Figure 2.1: The Value at Risk (VaR)

In the case that the returns follow the normal distribution, which is symmetric, then the figure above can provide us with the same results:

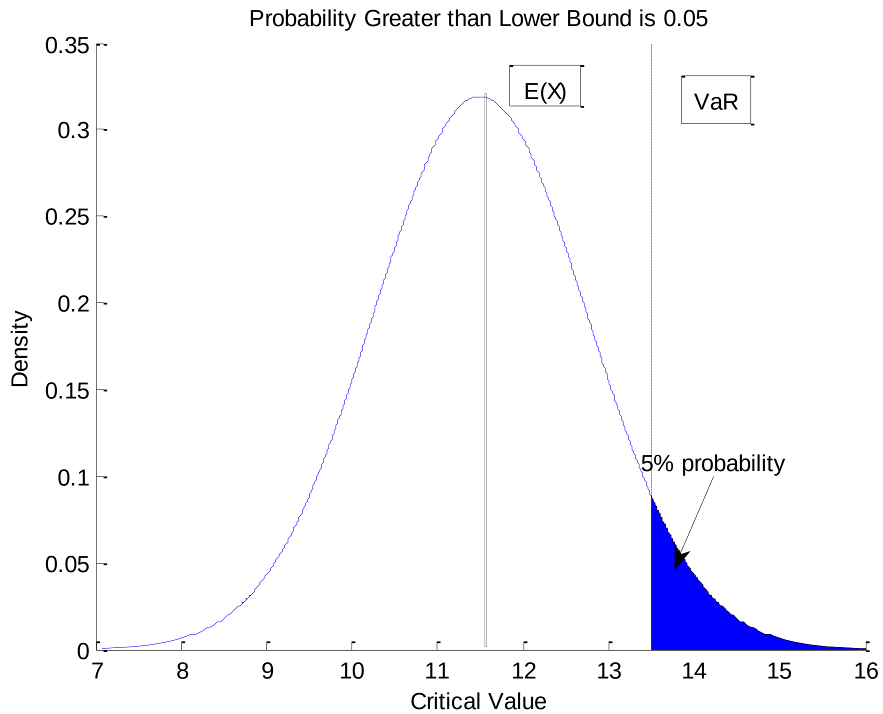

Figure 2.2: The Value at Risk (VaR)

Lemma 2.1:

A point $x_0 \in \mathbb{R}$ is the α- quantile of some density function F if an only f the following two conditions are satisfied:

$$F(x) \geq \alpha \text{ and } F(x) < \alpha \text{ for all } x < x_0$$

■

In that point, for our better understanding most in the computation part of Vale at Risk, we must say that practically Value at Risk is the inverse cumulative distribution of F, i. e.

$$\mathrm{VaR} = \mathrm{F}^{-1}(\alpha) \qquad (2.1c)$$

Mostly, in our study, we will use (2.1c) I the computing part of VaR and not in the theoretical examination of this measure.

Example 2.1: Var for Normal and t loss distributions [26]

For normally distributed random variables, Value at Risk is proportional to the standard deviation. Suppose that the loss distribution F_X is normally distributed with mean μ and standard deviation σ^2 (i.e. $X \sim N(\mu,\sigma^2)$) . We fix $\alpha \in (0,1)$. Then we have:

$$\mathrm{VaR}_{\alpha} = -(\mu + \sigma\Phi^{-1}(\alpha))$$

where Φ denotes the standard normal density function and $\Phi^{-1}(\alpha)$ is the α-quantile of Φ.

Proof:

F_X is strictly increasing. So, using the lemma 1, we only have to prove that $F_X(\mathrm{VaR}_\alpha) = \alpha$. Now,

$P(X \leq -\mathrm{VaR}_\alpha) = P(\frac{X-\mu}{\sigma} \leq \Phi^{-1}(\alpha)) = \Phi\left(\Phi^{-1}(\alpha)\right) = \alpha.$

This result is commonly used in variance- covariance approach, which we are going to analyze later.

The above result is used in every location- scale family, such as Student t loss distribution. Suppose that our loss X is of the form $\frac{X-\mu}{\sigma}$ and has a standard t distribution with ν degrees of freedom, i.e. $X \sim t(\nu, \mu, \sigma^2)$. We note that the moments in that case are given by:

$E(L) = \mu$ *and* $var(X) = \nu\sigma^2/(\nu - 2)$, when ν>2.

We get:

$$\mathrm{VaR}_{\alpha} = -(\mu + \sigma t_{\nu}{}^{-1}(\alpha))$$

where t_ν is the df of standard t.

■

2.3 Interpretation of Value at Risk

Value at Risk (VaR) is a central concept in risk management. As we said earlier, Value at Risk is a real number which measures the potential portfolio loss. Losses which are greater than VaR are suffered only with a specific small probability. In that point, we must point out that Value at Risk refers only to “Normal Market”[1].

The simplicity of Value at Risk lies on the fact that a single number can include all the possible risks of the portfolio. Furthermore, this single number can be used in every report referring to the portfolio and it can be easily understood by everyone. So, if someone “forgets” the use of statistics as a way to compute VaR, the fact that VaR refers to monetary values is straightforward to understand.

For our better understanding of the concept of VaR, consider an example included in Linsmeier and Pearson’s paper [24]. This example involves an FX forward contract entered into by a U.S. company at some point in the past. Suppose that the current date is 20 May1996 and the forward contract has 91 days remaining until the delivery date of 19 August. The 3-month US dollar (USD) and British pound (GBP) interest rate are r_{USd} = 5.469% and r_{GBP} = 6.063%, respectively, and the spot exchange rate is 1.5335 \$/£ . On the delivery date the U.S. Company will deliver \$15 million and receive £1 million. The US dollar mark-to-market value of the forward contract can be computed using the interest and exchange rates prevailing on 20 May. Specifically,

$$\begin{aligned} \text{USD}_{\text{mark-to-market value}} &= \left[\left(\text{exchange rate in}\frac{\text{USD}}{\text{GBP}}\right) \times \frac{\text{GBP 10 million}}{1+r_{\text{GBP}}(91/360)}\right] - \frac{\text{USD 15 million}}{1+r_{\text{USD}}(91/360)} \\ &= \left[(1.5335\text{USD/GBP})\times \frac{GBP\ 10\ million}{1+0.06063(\frac{91}{360})}\right] - \frac{\text{USD 15 million}}{1+0.05463(\frac{91}{360})} \\ &= \text{USB } 327.771 \end{aligned}$$

In this calculation we use that fact that one leg of the forward contract is equivalent to a pound- denominated 91-day zero coupon bond and the other leg is equivalent to a dollar-denominated 91-day zero coupon bond.

On the next day, 21 May, it is likely that interest rates, exchange rates, and thus the value of the forward contract have all changed. Moreover, suppose that the probability that the loss will exceed \$130,000 is 2%, the

[1] We define Normal Market as follows “For non-interest rate futures, a situation in which the distant months are at a premium to the nearby months. For interest rate futures, a situation in which the nearby months are at a premium to the distant months.”

probability that the loss will be between $110,000 and $130,000 is 1% and the probability that the loss will be between $90,000 and $110000 is 2%. Summing these probabilities, there is a 5% probability that the loss will exceed approximately $90,000. If we deem a loss that is suffered less than 5% of the time to be a loss due to unusual or “abnormal” market movements, then $90,000 divides the losses due to “abnormal” market movements from the normal ones. If we use this 5% probability as a cutoff to define a loss due to normal market movements, then $90,000 is the (approximate) Value at Risk. The above example is based on the assumption that losses follow the Normal distribution.

■

The probability used as the reference confidence level needs not be 5%, but rather is chosen by the either the user or the provider of the Value at risk number: perhaps the risk manager, risk management committee or designer of the system used to compute the Value at Risk. If instead the probability was chosen to be 2%, the Value at Risk would be $130,000, because the loss is predicted to exceed $130,000 only 2% of the time.

Suppose we use a probability of *x* percent and a holding period of *t* days. From the example above and the definition of Value at risk we gave in the previous section, the Value at Risk is the loss that is expected to be exceeded with a probability of only *x* percent during the next *t*-day holding period. In other words, <u>it is the loss that is expected to be exceeded during x percent of the t-day holding period.</u>

Unfortunately, there is no methodology which will guide us for the optimal choice of x. This is up to the designer’s discretion and his/hers subjective opinion on the probability of losses. For example, JP Morgan’s RiskMetrics system uses 5%, while Mobil Oil’s annual report indicates that it uses 0.3% [4]. Additionally, it is really important to keep in mind the holding time period t and the probability distribution of x. Two companies holding identical portfolios will come up with totally different values of VaR if the time period or the probability or both of them are different. The loss that is suffered with probability 1% is larger than the one suffered with probability

5%. Using the Variance- Covariance approach for the calculation of VaR, mentioned in the section 2.4.1, it had been found that it is 1.41 times as large.

2.4 Methods of measuring Value at Risk

In the following section we quote the standard methods used in finance for the computation of VaR. There are three basic ways of measuring Value at Risk, though there are several variations of each method. The three basic ways are:

1) The Variance-Covariance method
2) The historical data approach
3) The Monte- Carlo simulation

2.4.1 ***Variance- Covariance Method:***

As we have seen previously, Value at Risk computes the probability that the value of an asset will drop below a specific value in a particular time horizon. So if we had the option to derive a probability distribution of potential values of risk, it would be even simpler to compute VaR. That is in simple words what Variance- Covariance method does.

General Description:

This method is based on the assumption that the underlying market factors are normally distributed. We denote X_{t+1} the risk-factors which follow the multivariate normal distribution, i.e. $X_{t+1} \sim N_d(\mu, \Sigma)$ where μ is the mean vector and Σ is the variance- covariance matrix of the distribution. Using this assumption it is easy to determine the distribution of our portfolio profits and losses, which is also Normal. After that, due to the mathematical properties of the Normal distribution, is easy enough to determine the loss which is equaled or exceeded a specific x per cent of the time.

For example, assume we want to evaluate the Value at Risk for a single asset, where the potential values are, as we mentioned earlier, normally distributed with mean \$100 million and standard deviation \$20 million. With 95% confidence interval, which is translated to 1.96 standard deviations on both either side of the mean, we can assess that the value of this asset will not fall below \$60 million or rise above \$140 million, i.e.

$$VaR = 1.96 \times (\text{standard deviation}) = 1.96 \times 20 = 39.2 \cong 40$$

Even though this method is simple enough to understand, it is difficult when we have portfolios with a large number of assets. The difficulty lies in the fact that we must compute not only the variances, but also the covariances of every pair of assets. The number of computations is usually a large one. For instance, if we have a portfolio with 100 assets, we must estimate 49.500 covariances, in addition to 100 variances [34].

Clearly, this method is not very practical when we have large portfolios. For this reason, we want to simplify this method. To proceed, we map the risk in the individual investments in the portfolio to more general market risks, when we compute the VaR, and then we estimate the measure based on this market risk exposures. For the mapping we follow four basic steps:

<u>Step 1:</u>

In the first step we try to map every financial asset into a set of instruments representing the underlying market. We have to identify the basic market factors and the standardized positions related to market factors of course to do the mapping. In this step, we have a subjective opinion of the risk analyst in the decision of how important a risk factor is and how much it influences the market. The subjective choice in that point, can give to the analyst the flexibility in setting up the mapping.

This step of mapping is essential in this method. Not only can you reduce extremely the number of risk factors, but also instead of estimating the variances and covariances of hundreds of individual assets, you just estimate the risk instruments that those assets are exposed to.

<u>Step 2:</u>

In the second step, each financial asset is stated as a set of positions in the standardized market instruments.

Step 3:

In the third step, we estimate the variances and covariances of each instrument and across the instruments, respectively. At this point, the variance and covariance method captures the variability and comovement of the market factors: standard deviation (or else variance) captures the variability and correlation coefficients capture the comovements. Usually, this step i.e. the estimations of variances and covariances, are obtained by historical data.

Step 4:

In the final forth step, we compute VaR for the examined portfolio using weights on the standardized instruments of step 2 and on variances and covariances of step 3.

This method is called risk mapping.

For example, assume we want to compute a portfolio of a three risky assets. The initial value of that portfolio is $100 million. The three risky assets are X_1, X_2 and X_3. Assume that the parameters are known. We have μ_1, μ_2 and μ_3 the means and

$$S = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

the variance-covariance matrix of the returns, which is symmetric.

We have σ_{ij} the covariance of the returns between assets i and j, and σ_{ii} the variances of the return of the asset i. Assume that $30 million are invested in asset 1, $25 million in asset 2 and $45 million in asset 3. Then the return distribution of the portfolio is given by:

$$\text{mean return} = \mathrm{x}\mu = \mathrm{x}_1\mu_1 + \mathrm{x}_2\mu_2 + x_3\mu_3$$

$$\begin{aligned} \text{variance of return} &= \mathrm{XSS^T} \\ &= x_1^2\sigma_{11}^2 + x_2^2\sigma_{22}^2 + x_3^2\sigma_{33}^2 + 2x_1x_2\rho_{12}\sigma_1\sigma_2 + 2x_1x_3\rho_{13}\sigma_1\sigma_3 \\ &+ 2x_2x_3\rho_{23}\sigma_2\sigma_3 \end{aligned}$$

where $X=(x_1, x_2, x_3)=(0.3\ 0.25\ 0.45)$ is the vector of the three risky assets and S^T is the inverted matrix of S and

$\mu=(0.1\ 0.12\ 0.13)$

$$S = \begin{bmatrix} 0.1 & 0.04 & 0.03 \\ 0.04 & 0.2 & -0.04 \\ 0.03 & -0.04 & 0.6 \end{bmatrix}$$

We have, mean return=0.11185 and variance of return=0.0384838. So,

Portfolio mean=111.85

Portfolio variance=38.4838

So the VaR at 1% level is 177.677 [5].

■

Comments:

1) We must point out that the first assumption of this method is that returns on individual risk factors are normally distributed. Furthermore, despite the fact that returns themselves may not be normally distributed and large outliers are very common in finance (because commonly the distributions are fat tailed), the assumption is that the standardized return[2] follows the Normal distribution.
2) When we analyze the results, we must not focus on the size of the returns, but on size of the returns related to the standard deviation. For example, a large return in a period of high volatility may result in a low standardized return, whereas the same return in a period of low volatility may give an abnormal high standardized return.

The basic advantage of this approach is its flexibility, its simplicity, a combination which made the Variance- Covariance method very popular and widespread used. Furthermore, this method enables the addition of specific scenarios and enables the analysis of the sensitivity of the results with respect to the parameters. However, there are some drawbacks lies of this method when the analyst goes to the estimation process.

[2] We note that $standardized\ return = \frac{\text{return}}{\text{forecasted standard deviation}}$

- Wrong distributional assumption: If the conditional returns are not normally distributed, the computed VaR will underestimate the true VaR. That means that if there are more outliers in the actual return distribution than would be expected given the normal distribution, the actual VaR will be much higher than the computed one.
- Input errors: As we mentioned earlier, the estimation of variances and covariances is made using historical data. Even accepting the assumption of normality, the estimations have already standard errors. In other words, the variance-covariance matrix that is an input to the VaR measure is a collection of estimates, some of which may have large error terms.
- Non-stationarity: Another problem occurs when the variances and covariances change over time. That's a very common issue due to the fact that the markets change every day and a slight change may have a large effect in the computations. For example, the correlation between the U.S. dollar and European euro may change if a financial crisis takes place in a country which uses euro. That incident may lead to a breakdown in the computed VaR.

Modifications:

Since the disadvantages of the computation of VaR using this method are known, efforts have been made to revitalize the approach. First, researchers tried to compute VaR disregarding the assumption of normality. Hull and White [17] underlined that the most commonly used model of calculating VaR assumes that the probability distribution of the daily changes in each market variable is normal distribution. Though, they point out that this assumption is far from the truth.

The daily changes in many variables, particularly exchange rates, present an amount of positive kurtosis. That means that the probability distribution of daily changes is fat tailed, so that extreme outcomes may happen much more frequently than the normal distribution predicts. Duffie and Pan identify jumps and stochastic volatility as possible causes of kurtosis.

Hull and White suggested ways of calculating VaR in models where the variables are not normally distributed. This model allows the analysts to assume a probability for the variables, such as daily changes, different than normal. Then, they transform these variables into new ones, which must follow the normal distribution. We assume that the new variables are multivariate normal. Those transformations are a way of handling the correlations between the variables.

For the second problem we mentioned earlier, which is the input errors, researchers tried to find new estimating techniques to overcome this problem. Some researchers suggested that the ways of calculation the estimates just need some changes. Other authors proposed that conventional estimates are based on the assumption that the standard deviation in returns does not change over time, i.e. homoskedasticity. Engle [12] suggested that we can have better estimators if we use model that explicit allow the standard deviation to change over time, i.e. eteroskedasticity. He suggested Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) as two variants that provide better forecasts [12].

Last but not least, the final issue for which this approach is criticized is the fact that the Variance- Covariance method is based on the fact that there is a linear relationship between risk and portfolio positions. That is not always true, since the relationship between risk and payoffs in a portfolio which includes options for example is not linear. To deal with this kind of problem, researchers proposed *Quadratic Value at Risk.* Quadratic measures give the researchers the opportunity to estimate VaR for complicated portfolios that includes options, but there is a cost: firstly the mathematics involved in the calculations of deriving the VaR is very complicated and secondly some of the intuition is getting lost along the way.

2.4.2 __Historical Simulation__

The historical simulation is probably the simplest way of estimating Value at Risk and the major parameters, such as means, standard deviations, and correlations. The key in this method is the fact that it uses past data, which is collected over a specific horizon of the past, to create possible future scenarios. This allows the analysts to compute the changes that would occurred in each period. These scenarios are thought to be a representative collection of all possibilities that could take place between a time period which is being specified from the start by the analysts.

In addition, in historical data simulation there is no assumption of a complex structure of markets, fact that simplifies a lot the calculations. Furthermore, in HS we make no assumptions about the distribution that the returns follow, so it is actually a non parametric method of calculating VaR. That happens because this method uses the empirical distribution of the portfolio’s return. However, the main disadvantage of HS is the assumption we make that all the returns are iid. Furthermore, each day of the historical simulation comes with the some weight, so the analyst cannot evaluate the potential volatility. Lastly, the method is based on the assumption of history repeating itself.

General description:

The way historical simulation works is simple enough: as in variance-covariance method, we have derived our portfolio in market risk factors. For each factor we run a time series using past data. The difference lies in the fact that we do not use past data to estimate the variances and covariances to be, since the changes in the portfolio over time yields all the information we need to compute the Value at Risk.

Analytically, with HS we go back in time and we generate scenarios by sampling historical returns. Each return is associated with each risk factor of the portfolio. The overall value, of all linear and derivative position, is the produced portfolio. We repeat this procedure as many times as needed. The returns we use to create our data, may be drawn with or without replacement. In fact, we use a methodology called *bootstrapping* to create, using the actual distribution of the historical data which we have weighted, to create “new” data.

In other worlds, we could say that we weight our time series of historical asset returns. We denote with [4]:

Θ: the set of historical returns

$e^* = (e_1^*, e_2^*, \dots., e_T^*)$: element with $e^* \in \Theta$, $\quad i = 1,2, \dots., T$

$Y_{T+1}^* = Y_T + Y_T^* * e^*$ *is the simulated price for asset Y*

We repeat the process of finding the Y_{T+i}^* as many times needed to create the simulated price series Y^*. The scenario of the risk factor Y is formed of the simulated prices for day T+1, T+2 … T+N.

Comments:

The historical simulation approach is without a doubt the easiest and the most popular method of calculation VaR. The great advantage of this method is, apart of its easiness, the fact that is extremely easy to implement and is fully understood even from someone who is not a statistician. Furthermore, HS uses the empirical distribution of the returns for the calculations so there is no need to assume a distribution. Moreover, HS can be applied to all kind of portfolios, even in nonlinear ones.

However, it has some drawbacks. Firstly, as we already mentioned, the major disadvantage of this method is the assumption that the distribution of returns is iid. If the returns are in fact iid and we know the moments of the distribution, then any result about the portfolio will be accurate and they will not change during time. Independence indicates that the size of price movement in one period does not play any role to every other period, i.e. each period's changes does not influence any other period. Stationarity makes us sure that the probability that a specific loss will be occurred is the same every day. Those two abilities of iid plus the assumption of normality can make the VaR calculation very simply, for longer or shorter periods. However, if the distribution of returns is not stationary and we use for our calculations the same constant volatility, then we will be driven in misleading results[3].

The above disadvantage can be combined with the fact that HS does not give any space to volatility changes. Large price changes usually are followed by even larger changes. That create tendency in the data, known as *volatility clustering*, which must be taken into account in our calculations.

[3] In days where we have greater volatility we will have greater losses.

Secondly, HS depends on the motto that “History repeats itself”. This method uses only the past data to calculate VaR, assuming that what happened to the past will happen to the future too. There is no possibility of subjective information, as in Monte Carlo, neither of distributional assumptions, as in Variance-Covariance method. That is happening due to the entirely computation based in historical price changes.

Last but not least, HS does not give us the possibility to insert in our analysis new assets or market risks to optimize our measurement. This is logical of course if we think that our data is based in the past. However, as the time passes, new factors must be taken into account. For example, a financial crisis which may be occurred is a regulatory and extremely serious factor which can change our results in a heartbeat.

Modifications:

Since the problems of this approach are known, it is expected that researchers have tried to find ways of eliminating them.

a) Weighting the recent past: A logical thought can be made here, that recent past must be taken into more account for the future’s calculations of VaR than the distant one. Boudoukh, Richardson and Whitelaw worked exactly on that road. They suggested that the probabilities must be weighted based on their recency [6]. In other words, if we assume that decay factor is 0.9, the most recent probability weight p, the observation prior to it will be 0.9p, the one before 0.81p and so on [34].
b) Combining historical simulation with time series models: Cabedo and Moya in their paper “Estimating oil price “Value at Risk” using the historical simulation approach” suggested another way of computing VaR using time series model. They prove that fitting our historical data to a time series model and by estimating the parameters of that model we will have better and more accurate results. They fit an autoregressive moving average model (ARMA) to their oil data and use this model to forecast the returns with 99% confidence interval. Using their methodology, they did not directly used the distribution of past

returns, but the distribution of forecasting errors derived from an estimated ARMA model.

Firstly, the calculated the past portfolio returns in absolute value and then they analyzed the produced autocorrelation. They pointed out, though, that autocorrelation may provide negative results, which means that the time series may present a non-stationary behavior or they may have a non- statistically significant autocorrelation level. In the first case, the analyst must transform the series using the lowest differentiation for creating a stationary behavior. In the second case, the methodology is equivalent to standard historical simulation approach. We can continue to the second phase only when a statistically significant autocorrelation has been determined.

Secondly, we have the ARMA model estimation. Using past returns, we estimate a model for oil price behavior. In this phase, we must remove the autocorrelation, so we must determine the number of necessary to do so.

In the third stage, we have our preliminary forecasts. Those forecasts are made using the coefficients estimated in the previous step. In that stage we also calculate the percentile associated with the desirable likelihood level.

In the final stage, we calculate the future returns, which are corrected with the use of the percentiles calculated in stage three, by using the model of forecast estimated in the second stage. These forecasts are used to calculate the desirable VaR. Figure 2.3 shows the differences:

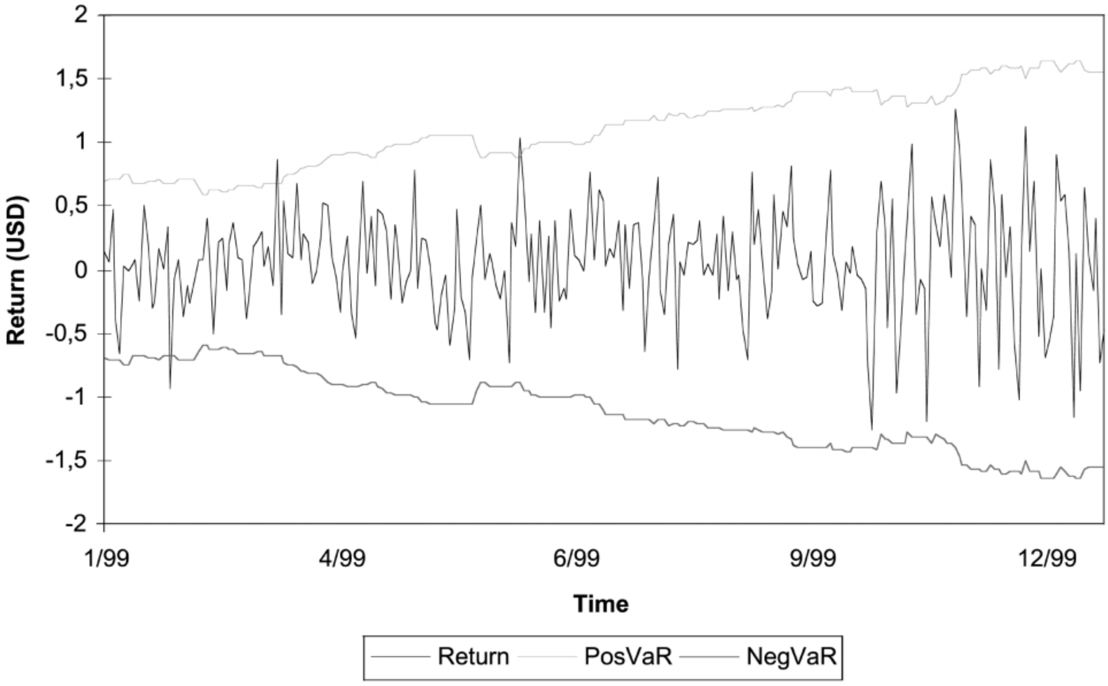

Figure 2.3: Value at Risk estimates from time series model [26]

The figure 2.3 illustrates that the actual oil price returns in 1999 fall within the predicted bounds 98.8% of the time, in contrast to 97.7% of the time they do with the unadjusted historical simulation [7]. The innovation in that method is that use time series which are sensitive in the changes of variance unlike historical simulation.

c) Volatility updating: There is possibility that the recent volatility of an asset to be higher than the one of the historical data. To skip this problem, Hull and White [18] suggested that the historical data must be adjusted to reflect any possible changes. For example, if the standard deviation now is 0.8% and the standard deviation in our historical data was 0.6% 20 days ago, Hull and White recommended scaling that number to reflect the change in volatility. For 1% return on that day we have:

$\frac{0.8}{0.6} * 1\% = 1.33\%\ return$ [18]. Their approach requires day-specific estimates of variances that change over the historical time period, which they obtained by using the GARCH models.

Other versions based on historical simulation:

A useful version of historical simulation is when the procedure of risk mapping gives us the opportunity to define the price of a whole portfolio as a deterministic function of the market parameters P(p), where P is the pricing function and p is the vector of all significant market parameters. When we examine today's price, say today is day t, we have $P(p_t)$ price. Using the notation above, if we examine the parameters at some day i, we have the parameters vector p_i and in day i+1 we have p_{i+1}. As the basic goal to do the above is to model the possible changes in today's market, we use the following ways: either we multiply each market parameter with the ratio of the same parameter at the day i+1 and day i, nor we add to today's value the difference between the values at day i+1 and day i for each parameter. The second method is recommended when increase in volatility is being observed with the level of parameter, so this method is very useful for stock indexes, exchange rates etc [32].

2.4.3 Monte Carlo simulation

Monte Carlo simulation is generally a widespread way of generating values and computing parameters. In our case, our focus is on calculating Value at Risk. This is both a parametric and non parametric way of calculating VaR depending on our assumptions. However, in the case of Monte Carlo simulation, we are not interested in computing the entire distribution of losses. We are interested only in the probabilities of losses exceeding a specific value.

Monte Carlo simulation has a lot of similarities with the historical data approach. The difference between them is that in the case of historical simulation we carry out the simulation using the observed changes in the market factors over the last N periods to generate N hypothetical portfolios, though in the case of Monte Carlo simulation we have the possibility to choose a distribution that is believed to adequately capture or approximate the possible changes in the market factor [24].

General description:

For the Monte Carlo simulation, we follow the same two first steps as in Variance- Covariance method. One should build a joint distribution of these factors based on one of the following: historical data, data implicitly implied by observed prices or data based on specific economic scenarios. The assumed distribution is not obligatory to be the multivariate Normal, despite the fact that the natural interpretation of its parameters, such as mean, standard deviation and correlations, and the ease with which these parameters can be estimated is for its favor. The analysts are free to choose any distribution that could reasonably fit their data and it could be capable of describing possible future changes in the market factors.

The difference between those two methods begins in stage 3. Instead of computing the variances and covariances of the market risks, we generate values, in other words we stimulate, by specifying probabilities for each market risk. Also, we specify how those market risks move together.

Then, the simulation is being performed for a large number of possible scenarios. After a repeated series of runs, usually greater than 1.000 and maybe greater 10.000, we will have a distribution of portfolio values that can be used to assess Value at Risk. The profit and losses at the end of the period are ordered from the largest profit to the largest lost for each scenario and the x% quantile of the worst results is the VaR estimate.

For example, we assume that we run a series of 10.000 simulations and derive corresponding values for the portfolio. These values can be ranked from highest to lowest and the 95%percentile VaR will correspond to the 500^{th} lowest value of the 99^{th} percentile to the 100^{th} lowest value.

In this method no assumption for the form of distributions is being made in advance. The analyst assumes a probability that in his/her belief is the one that the parameters follow. It is clear that if we assume the normally distributed market risks, our life becomes extremely easy. However, the power of Monte Carlo simulation lies on the fact that we have the freedom to choose any probability other than normal, which we believe it is more suitable for our data. In other words, this method is the only method among the three that which is better suited to the analyst with the possibility of *subjective judgments.*

Assessment:

This method, without a doubt, has several important advantages, and as every other method, some also important disadvantages. First, and maybe the most important, is the fact that this method does not assumes a priori neither a specific model nor a specific probability for the risk markets. So it can be easily adjusted to economic forecasts. Also, we have the possibility of improving our result by taking a larger number f simulated scenarios.

In addition, Monte Carlo simulation is not affected by the nonlinear relationship between the parameters, so we can include nonlinear instruments such as options. Last but not least, an analyst can track path- dependence because the whole market process is simulated rather than the final result alone.

However, disadvantages exist too. First of all the simulation will work well and provide us with good results only if the assumption we have made for the probability distribution of the inputs is correct. In other words, if we assume a probability that is far from the real one, then our results will be extremely misleading and we may be driven to wrong conclusions.

In addition, as the number of market risk factors increases making the comovements more complex, Monte Carlo simulations become more difficult to run for two reasons. First, we now have to estimate the probability distributions for hundreds of market risk variables than just the handful as in a single project or asset. This is also true in cases in which the variables are highly correlated. Second, the number of simulations that someone needs to run for having a good estimation of VaR will have to increase substantially. That leads us to the probably most important disadvantage of Monte Carlo simulation: the convergence. Any Monte Carlo simulation converges to the true value with rate $\frac{1}{\sqrt{N}}$, where N is the total number of simulated trajectories. So, if we want to increase the precision of our estimates, let’s say 10 times, we must run 100 times more simulations.

However, despite the fact that this maybe the most serious disadvantage of this method, in many cases there are well developed techniques of variance reduction, which are based on known properties of the portfolio such as correlations, or known analytical approximations to options and fixed income instruments. Another way of speeding the standard Monte Carlo approach is portfolio compression. In that case, we represent a large portfolio of similar instruments as a single instrument with risk characteristics similar to the original, requiring though similar risk characteristic to all the instruments we want to unify.

Undeniably, Monte Carlo approach has a lot of strengths, which are more obvious trough the comparison of the three approaches. Unlike the Variance- covariance method, we are not obligated to make the assumption of normality in returns, which most of the times is wrong. On the contrary, we have the ability to assume a probability which in our beliefs fits better the data. Also, compared to the historical data approach, we still start with historical data, but we also have the possibility to enter in our model subjective judgments and other information which can improve our forecasting. Finally, Monte Carlo simulation gives us the possibility to include in our data even nonlinear instruments and by that possibility, to be more flexible.

Modifications:

As with all the other approaches, modifications have been made to eliminate the disadvantages of Monte Carlo simulation too, which are mostly focused on the computational part. If we have a large number of instruments, a large number of simulations is required, for example say we have a model with 15 key rates and 4 possible values for each. Then, it is required 4^{15} simulations to be completed, i.e. 1.073.741.824 simulations. The modifications are directed to reduce the required number of simulations.

a) Scenario Simulation: We have the possibility to reduce the number of simulations required by doing the analysis over a number of discrete scenarios. Jamshidan and Zhu [19] suggested a way to do that so called scenario simulations. According to them, we can use principal components analysis to reduce the number of the market risks which affect our portfolio. Rather than allowing every factor to take part in every scenario, they are looking for possible combinations of these variables to conclude at scenarios. Then, they compute across these scenarios to arrive at the simulation results.

b) Monte Carlo simulations with Variance- Covariance method modification: In that modification we combine the speed of variance-covariance method and the flexibility of Monte Carlo simulation. The strong part in variance- covariance method is its speed and leaving aside the assumption of normality, it is capable to calculate VaR in a heartbeat. On the other hand, Monte Carlo gives the opportunity to the analyst to assume any distribution he believes it fits better to the data. "Glasserman, Heidelberger and Shahabuddin [12] use approximations from the variance covariance method to guide the sampling process in Monte Carlo simulations and report a substantial savings in time and resources, without any appreciable loss of precision [34]. "

2.5 **Comparing Approaches**

Questions arise concerning which method provides best estimations. Variance-covariance method is very simple to compute and is quickly enough. However, this method requires strong assumptions about the return distribution of standardized assets. On the other hand, historical simulation is the simplest of all the methods with no assumptions about the distribution of the returns, but assumes that the past will repeat itself so the data used in simulation is straightforward. Last but not least, Monte Carlo simulation provides good estimates and is flexible enough in the point you must assume a distribution for the future returns, but it can be very complicated in the computing part and slow especially when a large number of market factors come in the model.

But the query still remains: how different are the estimates of VaR produced by each method. If we use, for instance, the historical data approach and the historical data returns are normally distributed then this method will give the same results as the variance- covariance method, if that data is used to estimate the variance covariance matrix. Furthermore, the variance covariance method will give the same result as the Monte Carlo simulation if we assume that the data which we use is normally distributed with consistent means and variances. Moreover, the historical data approach will give the same results as the Monte Carlo simulation if the distributions we use in the latter are entirely based on historical data.

Also, another query comes in mind: if we do find different values of VaR, which of the three is the best estimate? That actually depends on what data we want to examine and on what approach we use. In very method mentioned above, we gave not only the advantages and disadvantages but also modifications that have been made to eliminate the problems in each approach.

So, in short, if we want to find which method is more appropriate and gives the best estimates, maybe we must look what kind of data we have in our hands. If we want to assess the Value at Risk for portfolios that do not include options over short period of time (a day or a week), the variance-covariance approach does a very good job regardless the assumption of normality. Moreover, if our risk source, for which we want to compute Value at Risk, is stable and where there is substantial historical data (for example commodity prices) then historical simulation gives good estimates. Finally, if we have non-linear portfolios, which may for example includes options, over large time periods, where the historical data is volatile and non- stationary and the assumption of normal is questionable, then Monte Carlo simulation is better.

CHAPTER 3

COHERENT RISK MEASURES

INTRODUCTION

In section 2.1 we introduced the idea of risk measure and then we analyzed one of the most common risk measure Value at Risk. Though, despite the fact that anyone can, intuitively, understand what a financial risk may be, it is most of the times hard to give a precise assessment of financial risk, unless of course we have a good risk measure. It is like the temperature. Anyone can feel the cold, but we need the thermometer to calculate exactly the temperature.

But a question arises after that. Are there some criteria that must be followed if we want to have a, so called, good risk measure? Properties which can guarantee the efficiency of our risk measure?

The answer is positive. A list of desirable properties for financial risk measures was proposed by Artzner et al. in 1999 in their paper called "Coherent Measures of Risk" [2]. This paper actually made the researchers to understand that the gap between the market practice and theoretical progress had widened enormously.

In their paper, they proposed a number of axioms that any so- called *coherent* risk measure should satisfy. Also, they studied the coherence properties of the most common used risk measures such as Value at Risk and gave a characterization of all coherent risk measures in terms of generalized scenarios. Furthermore, this specific paper made risk management a "science" with its own theory and framework.

3.1 PROPERTIES OF COHERENT RISK MEASURES

In this section we will introduce the properties that a risk measure should satisfy in order to be called coherent. Furthermore, we will comment every property and its significance in our study.

First of all we will give a strict definition of a risk measure.

We denote with:

(Ω, F, P): the probability space

Δ: the time horizon (which will leave unspecified and it will be entered in specific problems)

$X^0 (\Omega, F, P)$: the set of all variables on (Ω, F) which are considered to be almost surely finite.

G: is a set of all risks and it is real valued function on Ω. Also, we assume that G is convex cone[4].

ρ: the risk measure which is real- valued function and $\rho: G \rightarrow R$

X and Y: are random variables which denote the future value for each instrument.

Furthermore, when we say $\rho(X)$ is the minimum amount of capital that should be added to a risky position X, so that the position becomes acceptable to an external or internal risk controller [26]. If it is negative, i.e. $-\rho(X)$ capital may be even withdrawn. So we have:

[4] Convex cone's definition: if $X \in G$ and $Y \in G$ then $X+Y \in G$ and $\lambda X \in G$ for every $\lambda>0$.

$$\rho(X) = \begin{cases} \rho(X) \leq 0, acceptable\ without\ addition\ of capital \\ \rho(X) > 0, capital\ may\ be\ even\ withdrawn \end{cases}$$

The definition of ρ(X) illustrates that if ρ(X)>0 is positive, then a positive amount of money must be added to make the position acceptable or else the money must be withdrawn. If ρ(X) <0 is negative then we already have an acceptable position. In that case, we can even withdraw money from our position (amount of money which equals the absolute value of ρ(X)) and still have an acceptable position.

Now, we represent the four properties that a risk measure must follow in order to be called coherent.

Property 1: Translation invariance:

For all $X \in G$ and all real numbers α, we have $\rho(X+\alpha\tau)=\rho(X)-\alpha$

Property 2: Subadditivity

For all X_1 and $X_2 \in G$, we have $\rho(X_1+X_2) \leq \rho(X_1) + \rho(X_2)$

Property 3: Positive homogeneity:

For all $\lambda \geq 0$ and $X \in G$, we have $\rho(\lambda X)=\lambda\rho(X)$

Property 4: Monotonicity

For all X and Y $\in$ G with $X \leq Y$, we have $\rho(X) \leq \rho(Y)$.

Comments:

Property 1 is very useful in finance. It ensures that by adding an amount α in the initial position τ, the risk decreases by α. That is something logical actually. If we invest more money in a position, we do decrease the risk compared to the initial model.

Property 2 means that the capital required for two positions X and Y combined will be less than the capital required for the risk treated separately. In other words, a portfolio which is made up of sub-portfolios will risk an amount of money less or equal than the sum of the risks of each sub-portfolios. This property is necessary and we expect to be satisfied by a risk measure, because it actually says that if we aggregate individual risks then we will not increase the overall risk, on the contrary. That property is very important because:

a) If a company was forced to meet a requirement of extra capital, without satisfying this property, it could be capable of breaking into two separate incorporated affiliates, in order to reduce its regulatory capital requirements.
b) Consider an example of two trading desks, D_1 and D_2. Denote with L_1 and L_2 the possible risks of X and X_2, respectively. Assume that the analyst wants to ensure ρ (Z), where Z= D_1+ D_2 is the overall loss, is smaller than M. If the risk measure ρ is subadditive, then the only thing he has to do is to choose bounds M_1 and M_2, such that $M_1+M_2\leq M$ and ρ $(X_i) \leq M_i$. So, it is ensured that $\rho(Z)\leq M_1+M_2\leq M$.

Property 3 means that the required capital is independent of the currency in which the risk is measured. This property is very reasonable and emerges from property 2. Subadditivity implies for $\lambda\in \mathbb{N}$,

$$\rho(\lambda X) = \rho(X + X + \cdots X) \leq \lambda\rho(X)$$

Last but not least, property 4 simply means that if one risk includes greater losses than another, then it will require greater amount of capital. In other words, and in financial terms, if we have a position with higher losses in every state of the world will require more risk capital. Or else, if Y has a greater value than X, then Y should have lower risk, i.e. less amount of money should be added in Y than in X to make it acceptable (the amount to be added is the risk measure).

As we already mentioned, those four properties will define the coherent risk measure. We have the definition:

Definition 3.1: Coherent risk measure

A risk measure satisfying the four properties of translation invariance, subadditivity, positive homogeneity and Monotonicity, is called coherent [2].

There is an argument against property 3, the property of positive homogeneity. Some analysts believe that if λ is a large number then somehow it must being seen the concentration of risk and the fact that liquidity problems will be present. There for, the inequality $\rho(\lambda X) > \lambda\rho(X)$ should stands. Though, this is impossible for the subadditive risk measures. All this conflict led to the study of another family of risk measures, known as *convex risk measures*. In this family of risk measures, we have the properties 2 and 3(i.e. the subadditivity and the positive homogeneity) relaxed and anyone just asks for the weaker property of convexity, i.e. for all X, Y $\in$ G

$$\rho(tX + (1-t)Y) \leq t\rho(X) + (1-t)\rho(Y) \tag{3.1}$$

The fact is that these four properties guarantee that the risk function is convex, which actually in turn corresponds to risk aversion. Economically speaking, the relationship (3.1), gives the idea that diversification can reduce the potential risk.

Moreover, the fact that coherent risk measures are also convex can be helpful in another aspect too. The risk surface of any coherent risk measure is convex. This means that any line drawn between two coherent risk measures lies above the coherent risk surface. This is very important in minimization routines because it can insure the uniqueness of a global risk minimum. However, if a risk measure is not a convex one, then we may face the problem of multiple local minima, and it may be difficult to establish which of those minimums is the global.

3.2 THEORETICAL PROBLEMS IN THE APPLICATION OF VALUE AT RISK

3.2.1 THE VIOLATION OF SUBADDITIVITY OF VALUE AT RISK:

In chapter 2, we introduced one of the most widespread risk measures, Value at Risk. It is a fact that VaR is very ease in its application and computation among risk measures, along with its interpretation. That’s way Value at Risk was adopted as the best risk measure by essentially all banks and regulators. However, Value at Risk is not a coherent risk measure, because it contradicts with property 2, the property of subadditivity, although it does follow the other three properties. In fact, Value at Risk is subadditive in the case where the loss distribution is an elliptical one. In the real world though, losses are not always elliptically distributed.

In order to understand the mathematical problem as far the subadditivity is concerned, we quote some examples.

Example 3.1:

(Artzner et al. 1999: Coherent Measures of Risk [2] page 13)

« In this example we will show that Value at Risk follows properties 1, 2, 4 but fails to satisfy property 3. Furthermore we will show that Value at Risk is not convex.

In this example, we have two digital options[5] on a stock, having the same exercise date T, the end of the folding period. The first option denoted by A and initial price u, pays:

$$p_A = \begin{cases} 1000, \text{if the value of the stock at time T is more than a given U} \\ 0, \text{otherwise} \end{cases}$$

where U is given. The second option denoted with B and initial price ℓ, pays:

$$p_B = \begin{cases} 1000, \text{if the value of the stock at T is less than a given L} \\ 0, \text{otherwise} \end{cases}$$

Choosing L and U such that:

$$P\{S_T < L\} = P\{S_T > U\} = 0.008$$

We look for the 1% values of the future net works of positions taken by two traders writing respectively 2 options A and 2 options B. They are -2u and -2ℓ respectively, supposing that r=1. The positive number 1000- ℓ-u is the 1% value at risk of the future net worth of the position taken by a trader writing A+B. This implied that the set *is not convex.* »

In other words, if an agent sells both options at the same time, since those options are written on the same underlying asset, they are negatively correlated. So, the probability that one of those two options will be exercised is 1.6% and the 1% VaR is 1000- ℓ-u, a positive value. Then, we have:

[5] DEFINITION:A digital option is an option whose payout is fixed after the underlying stock exceeds the predetermined threshold or strike price

$$\text{VaR}(-A-B) = 1000 - u - \ell > -u - \ell = \text{VaR}(-A) + \text{VaR}(-B)$$

The relationship above, clearly contradicts with the property of subadditivity.

Example 3.2:

(Artzner et al. 1999: Coherent Measures of Risk [2] page 14)

«In this example we will show that Value at Risk follows properties 1, 2, 4 but fails to satisfy property 3.

For this example, we allow an infinite set Ω and consider two independent identically random variables X_1 and X_2, having the same density

$$f = \begin{cases} 0.90, \text{on the interval } [0,1] \\ 0.05, \text{on the interval } [-2,0] \end{cases}$$

Assume that each of them represents a future random net worth with positive expected value, which is a possibly interesting risk. Furthermore, in terms of quantile, the 10% Values at Risk of X_1 and X_2 being equal to 0, whereas an easy calculation showing that the 10% Value at risk of X_1+X_2 is certainly larger than 0. So we conclude that the individual controls of these risks do not allow directly a control of their sum, if we were to use the 10% Value at Risk »

Example 3.3:

(Alexander J. McNeil, Rudiger Frey, Paul Embrechts: Quantitative Risk Management [26] page 241)

«Consider a portfolio of d=10 defaultable corporate bonds. We assume that defaults are independent; the default probability is identical for all bonds and is equal to 2%. The current price of the bonds is 100. If there is no default, a bond pays t+1 (one year from now, say) an amount of 105, otherwise there is no repayment. Hence, L_i, the loss of bond i, is :

$$L_i = \begin{cases} 100, \text{when the bond defaults} \\ -5, \text{otherwise} \end{cases}$$

Denote by Y_i the default indicator of firm i, i.e.

$$Y_i = \begin{cases} 1, \text{if bond i defaults in } [t, t+1] \\ 0, \text{otherwise} \end{cases}$$

We get:

$L_i = 100Y_i - 5(1 - Y_i) = 105Y_i - 5.$

Hence, the L_i form a sequence of iid rvs with $P(L_i = -5) = 0.98$ and

$P(\mathrm{L_i} = 100) = 0.02.$

We compare the two portfolios, both with current value equal to 10.000. Portfolio A is fully concentrated and consists of 100 units of bond one. Portfolio B is completely diversified; it consists of one unit of each of the bonds. Economic intuition suggests that portfolio B is less risky than portfolio A and hence should have less VaR.

If we compute VaR at a confidence level of 95% for both portfolios, we will have:

Portfolio A:

portfolio loss is $\mathrm{L_A} = 100\mathrm{L_1}$ and $\mathrm{VaR_{0.95}(L_A)} = 100VaR_{0.95}(L_1)$

Now,

$P(\mathrm{L_1} \leq -5) = 0.98 \geq 0.95$ and $P(\mathrm{L_1} \leq \ell) = 0 < 0.95$ for ℓ<-5.

Hence, $\mathrm{VaR_{0.95}(L_A)} = -500.$

This means that even after a withdrawal of a risk capital of 500 the portfolio is still acceptable to a risk controller working with VaR at the 95% level.

Portfolio B:

Portfolio loss $\mathrm{L_B} = \sum_{i=1}^{100} \mathrm{L_i} = 105 \sum_{i=1}^{100} \mathrm{Y_i} - 500$ and $\mathrm{VaR_a(L_B)} = 105\mathrm{q}_\alpha(\sum_{i=1}^{100} \mathrm{Y_i}) - 500$

The sum $M \coloneqq \sum_{i=1}^{100} \mathrm{Y_i}$ has a binomial distribution, i.e. M~B(100,0.02). We get by inspection that:

P (M≤5) ≈0.984≥0.95 and

P (M≤4) ≈0.949< 0.95. So:

$$\mathrm{q_{0.95}(M)} = 5$$

Hence,

$$\mathrm{VaR_{0.95}(L_B)} = 525 - 500 = 25$$

In this case a bank would need an additional risk capital of 25 to satisfy a regulator working with VaR at the 95% level. Clearly the risk capital required for portfolio B is higher than for portfolio A ».

From example 3, we have two facts we have to point out:

1) Despite the fact that our intuition leads us to believe that portfolio B should be less risky, we just proved the exact opposite. So we conclude that VaR may give nonsensical results.
2) We proved that VaR is not subadditive.

3.2.2 COMMENTS:

Need of Subadditivity:

Generally, we have seen the violation of the subadditivity. This property, in short, expresses that, if we have a portfolio made by sub-portfolios, then the amount we risk when we use the whole portfolio as one, will be less or equal than the sum of the separate amounts risked by its sub-portfolios. The equality holds only when concurrent events take place, i.e. when all the sources of these risks can act altogether. In every other case, we have a strict inequality, i.e. the amount of risk of the whole portfolio is strictly less than the sum of the risks of the separate sub-portfolios.

Subadditivity allows risk reduction after diversification of portfolio. On the contrary, if a measure contradicts this property, it allows the possibility that diversification will lead to an increase of the value of risk, even if the partial risks are influenced by concurrent events. Moreover, subadditivity is necessary for capital adequacy in banking supervision. Think of a bank which is made of several branches and assume that there is a diversification of the whole capital into each branch. Subadditivity can guarantee that the overall bank capital will be adequate. On the other hand, if the employed risk measure violates subadditivity, then the risk of the whole bank could be much higher than the sum of the risks of the branches.

Another essential use of subadditivity is in portfolio optimization problems. This is due to convexity, which actually follows from subadditivity and positive homogeneity. Convexity assures the existence of a unique absolute minimum, rather than the local one in case of non-convexity. So the minimization comes with only a unique, well- diversified solution.

Where does the fact that Value at Risk is not subadditive lie? Usually it is caused by the fact that the assets that make the portfolio have very skewed loss distributions. In other words, the reason for the non subadditivity of Value at Risk is the fact that the underlying random variables may be independent but they are very heavy-tailed

3.3 TAIL ANALYSIS

When there is the special case of normality of returns, i.e. when all risks are jointly normally distributed, Value at Risk is known to be coherent below the mean. That happens due to the fact that quantiles satisfy the subadditivity as long as probabilities of exceedence are smaller than 0.5 [2].

However, that is extremely unlike to happen. Since Mandelbrot [25] in 1963 and Fama [13] in 1965, we know that usually returns are fat tailed. Thus, it has not been generally known when subadditivity is violated. What we do know is that Value at Risk is subadditive for the tails of all fat tailed distributions provided that the tails are not extremely fat. An example for that, is some assets whose, due to the fact that are so fat tailed, first moment is not defined, such as those who follow the Cauchy distribution.

The standard measure to calculate tail-fatness is kurtosis, which is the expected forth power, i.e. $E(S_i^4)$. A distribution has fat tails when kurtosis is present. Actually, when we have large returns usually kurtosis estimates are highly sensitive.

When a distribution has higher than normal kurtosis we expect a major concentration of the mass in the center of the distribution, leaving thin tails. Kurtosis can be higher than normal if either the tails of the cdf are heavier than the normal or, as we already told, if there is great mass in the center of the distribution or both. If we want to be more formal, we define a fat tail distribution requiring to be regularly varying. A function is regularly varying if it has a Pareto distribution-like power expansion at infinity. Then, to measure the thickness of the tails we use the tail index, which for financial assets is between three and five [8].

It is very common to observe super fat tails for assets whose prices are very unlike to change (either the prices have very little change or they do not change at all.) Furthermore, fat tailed are likely to be found in short term assets. In all of those cases, the violation of subadditivity can create a variation of problems.

3.4 STABLE AND ELLIPTICAL DISTRIBUTIONS

3.4.1 Stable distributions

As we mentioned before, if the distribution of returns is normal, then subadditivity is not violated. However, Value at Risk estimation under the assumption of normality tends to give bad results, even at the conventional 1% and 5% levels. Usually, it tends to overestimate Value at risk at the high probability levels and underestimate it at the lower probability levels. These bad results may lead the risk manager to underestimate the true risk or to miss part of the risk.

Furthermore, Value at Risk does not pay attention to the shape of the distribution beyond the VaR point. If a risk manager wants to compare two portfolios he/she must to be sure that the distributions aroused from the portfolios are similar. If, however, he/she wants to increase returns by selling derivatives which have high risk, then the VaR will not be greatly affected. However, that move has changed the distribution of losses. But, as we said, Value at Risk does not take into account what happens to the loss distribution after the point of VaR. That is a major drawback of Value at Risk systems.

A generalization of normal distribution, which can over pass the problem of the shape beyond VaR point, is the *stable distribution*, or else a-stable distribution. Usually the assumption of normality is based on the Central Limit Theorem, which is actually used in the case of a-stable distribution too. The differences from the normal distribution are that a-stable allows bigger concentration of mass around the mean, more extreme values and bigger skewness.

Definition 3.2: α- Stable distribution [26]

A random variable X is said to follow a α-stable distribution if for any positive number a and b, there is a positive number c and a real number d such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d \tag{3.2}$$

where X_1 and X_2 are independent copies of X and $\stackrel{d}{=}$ denotes equality in distribution.

The stable distributions have a characteristic function, defined as:

$$Eexp(i\theta X) = \exp\left\{-\sigma^{a}|\theta|^{a}\left[1 - i\beta(\sin ng)\tan\frac{\Pi\alpha}{2}\right] + i\mu\theta\right\}, 0 \leq \alpha \leq 2 \tag{3.3}$$

The parameters σ, β and μ are unique. Furthermore, μ is a location parameter and σ is a scale parameter. We use the notation:

$$X \sim S_a[\sigma,\beta,\mu] \Leftrightarrow \frac{X-\mu}{\sigma} \sim S_{a}[1,\beta,0]$$

The distribution $S_a(\sigma,\beta,0)$ is called:

$$\begin{cases} \text{Skewed to the right if } \beta > 0 \\ totally\ skewed\ if\ \beta = 1 \\ \text{Skewed to the left if } \beta < 0 \\ totally\ skewed\ if\ \beta = -1 \end{cases}$$

The number α s called *index of stability*, such that: $c^a = a^a + b^a$ [15].

The α -stable distribution depends on 4 parameters a, b, c and d. For those parameters we have the following [14]:

- α, with $0<a\leq 2$ is the stability parameter which determines the weight in the tails. The smaller the value of a, the greater the frequency and size of the extreme events.
- b is a skewness parameter with $-1\leq b\leq 1$. If b=0 we have a symmetric distribution. If b has positive or negative value then the distribution is skewed to the right or to the left, respectively.
- The parameter c is positive and it measures the dispersion. It is similar to the variance of a normal distribution.
- Finally d is a real number which can be thought as a location measure, it is similar to the mean of a normal distribution.

For the case of a normal distribution, a=2 and b=0, i.e. there is no skewness. Also, a normal distribution has a variance equal to $2c^2$. When a=1 and b=0, we then have the Cauchy distribution.

Stable distributions have a very useful property, from which the term "stable" actually derives. Let X_1, X_2.... X_n be mutually independent random variables with a common distribution R and let $S_n=X_1+X_2+.....X_n$. Then the distribution R is stable if for each n there are constants c_n and d_n such that:

$$S_n \stackrel{d}{=} c_n X + d_n$$

That means that the density of a sum of independent identically distributed a-stable random variables is up to a scale and location parameter the same as the distribution of the initial variables, i.e. stable. We have to point out that this is a property that only the a-stable distributions have.

α-stable distributions have yet another useful property. A generalization of the Central Limit Theorem says that: "If X_1, X_2 ..... X_n are independent, identically distributed (iid) random variables with finite variance, then the sum $S_n = X_1 + X_2 + \ldots\ldots + X_n$ will asymptotically follow the normal distribution. " These assumptions may be weaked considerably. If we keep the assumption of independence along with the fact that no individual variable has a significant effect on the mean, but drop the assumptions of identically distributed and finite variance then the Central Limit Theorem continues to hold. Furthermore, if we keep the independence requirement but allow for cases the individual measurements to have an effect on the mean, then if the sum converges, it converges to an a-stable distribution. So, each member of the a-stable distribution is actually an asymptotic limit for some set of independent identically distributed random variables.

Using the wider class of stable distributions, we do have a way to over pass the violation of subadditivity through proper conditioning, which are valid particularly for stable distributions. In that point, we must point out that Normal distribution itself is subadditive and it is not heavy tailed. A larger family of distributions, which includes the Normal distribution, is stable distribution. Stable distributions are heavy tailed but they hold all the nice properties of the Normal distribution. This is an essential property. When we have normally distributed losses we do have subadditivity but we do not have a good model due to the heavy tails. This problem is solved by using the stable distributions because I have both subadditivity and all the features from the Normal distribution and from tail analysis.

According to Fama [13] and Mandelbrot [25], the stable distributions present heavy- tailed which are well suited for modeling financial data. That is the reason they produce measures of risk based on distribution tails, like Value at Risk, which are reliable. Though, due to the properties we mentioned above, we can see a stable distribution with fat tails, as a mixture of stable distributions with less fat tails, or even as a mixture of normal distributions. So, as a result, the violation of subadditivity to the aggregate distribution can be over passed if we consider the whole distribution as a sum of separate stable distributions, in which there is no violation of the subadditivity.

Proposition 3.1 [9]:

Suppose that X_1 and X_2 are two asset returns with jointly regularly varying non-degenerate tails with tail index $\alpha>1$. Then VaR is subadditive in the tail region.

■

From the above proposition we are sure that the VaR of a portfolio, at sufficient enough low probability levels, is indeed lower than the sum of the VaRs of the individual positions, if the return distribution exhibits fat tails. In other words, subadditivity is not violated for fat tailed data, regardless of its dependency structure. An example of this application is the multivariate t-Student distribution with degrees of freedom larger than 1. To show the power of proposition 3.1, we illustrate the following example, considering zero colleration but where portfolios may be dependent.

Example 3.4 [9]:

(Danielsson et al.: Fat Tails, VaR and Subadditivity, page 10)

Consider the case of a bank and a market neutral hedge fund that both have exposures to two company returns from the same sector. The respective random returns are X_1 and X_2. For simplicity these are assumed to be independent and identically distributed. By nature, a bank is long in the economy and hence in the two companies. The bank's exposure is therefore $X_1 + X_2$. A market neutral hedge fund is short in the first firm and long in the second. Its exposure is $X_1 - X_2$. Assume alternatively that the returns are standard normally distributed, or Student-t, with $\alpha > 2$ degrees of freedom. It is immediate that $E\,[(X_1 + X_2)(X_1 - X_2)] = 0$, and hence the correlation is zero. Consider the following conditional probability

$$\lim_{s\to\infty} \Pr\,(X_1 + X_2 > s | X_1 - X_2 > s)$$

as a measure for dependence in the tails.

For the case of the normal distribution, since the colleration is zero, the two portfolios are independent and hence

$$\lim_{s\to\infty} \Pr(X_1 + X_2 > s | X_1 - X_2 > s) = \lim_{s\to\infty} \frac{\Pr(X_1 + X_2 > s | X_1 - X_2 > s)}{\Pr(X_1 - X_2 > s)}$$

$$= \lim_{s\to\infty} \frac{\Pr(X_1 + X_2 > s)\Pr(X_1 - X_2 > s)}{\Pr(X_1 - X_2 > s)}$$

$$= \lim\nolimits_{s\to\infty} \Pr(X_1 + X_2 > s) = 0$$

This is different under the assumption of Student-t distributed returns:

$$\lim_{s\to\infty} \Pr(X_1 + X_2 > s | X_1 - X_2 > s) = \lim_{s\to\infty} \frac{\Pr(X_1 + X_2 > s | X_1 - X_2 > s)}{\Pr(X_1 - X_2 > s)}$$

$$= \lim_{s\to\infty} \frac{s^{-\alpha}}{2s^{-\alpha}} = \frac{1}{2}$$

Thus for the Student returns there is clear evidence of dependence (in the tail area).

The implication for the VaR is as follows. For the normal case, below the mean the VaR is known to be subadditive. For the non-linear dependent case with the Student-t risk drivers, one can calculate the VaR sufficiently deep into the tail area by using Feller's convolution theorem. Since for large s

$$p = \Pr(X_1 + X_2 > s) = \Pr(X_1 - X_2 > s) \cong 2s^{-\alpha}$$

upon inversion, the univariate VaRs are $s \cong \left(\frac{2}{p}\right)^{1/\alpha}$. The VaR of the combination of the portfolios, i.e. when the bank is integrated with the hedge fund, is obtained from

$$p = \Pr(X_1 + X_2 + X_1 - X_2 > s) = \Pr(2X_1 > s) \cong 2^{\alpha} s^{-\alpha}$$

upon inversion $s \cong 2\left(\frac{1}{p}\right)^{1/\alpha}$. It follows immediately that this VaR is smaller than the sum of the individual VaRs $2\left(\frac{2}{p}\right)^{1/\alpha}$.

■

Finishing this section, we will add some characteristics that Student-t distribution has. Student distribution is a very widespread, especially in finance. One of the reasons is that it allows heavy tails. Another reason is that the degrees of freedom equal the tail index.

Furthermore, if we simulate data and then analyze the results, we will see that if the degrees of freedom are higher than 1, then the first moment is well defined and the violation of the subadditivity is almost zero. However, if the degrees of freedom equal to 1, then the number of times when subadditivity fails is very high and the first moment is not well defined.

3.4.2 Elliptic Distributions

Another extremely useful family of distributions is the *elliptical distribution family,* which actually is a generalization of the multivariate normal distribution. That is the reason why elliptical distributions have, like Normal distributions, attractable and interesting properties. The elliptic family is a family of symmetric distributions which includes the multivariate Normal and the multivariate t-distributions, as long as multivariate Cauchy, multivariate Logistic and multivariate Stable.

The elliptic distributions have a constant density on ellipsoids. Actually, in a two- dimension of space, the contour lines of the density surface are ellipses. A good example is the t- distribution. Due to the fact that this distribution has standard univariate t marginal distributions which are heavy-tailed, is often used when we have to deal with asset returns.

In the past few years, elliptical distributions have gained prominence as an effective and attractive tool for multivariate modeling in risk management. They were introduced by Kelker in 1970 and further discussed by Fang et al. at their paper “Symmetric Multivariate and Related Distributions” in 1987. Elliptical distributions are actually a generalization of multivariate normal family, with the difference that elliptical distributions allow both the presence of heavy tails and asymptotic tail dependence and the existence of short tails. Thus, elliptical distributions are more flexible than just the normal distribution. The fact that this family is able to control the problem of kurtosis is very important in financial risk management where the analyst observes empirical distributions of losses, which usually appear with heavier tails than the normal.

Elliptical distributions have some very convenient properties, which are the reason for the wide use of that family as a risk management tool. The properties are three:

1. It is very easy to simulate elliptical distributions, even in high dimensions, like the normal distribution. The fact that they are so easy to simulate from, is extremely useful for Monte Carlo simulation of dependent risks.
2. All linear transformations of elliptical distributed vectors also belong to elliptical distribution family.
3. The covariance matrix, exactly like in normal distributions, “carries” all the information about the potential dependency among the variables, even if covariances do not exist. Therefore, when we deal with elliptical distributed risks, the problems which arise from the examination of the variability and dependence in terms of the covariance matrix disappear.

In order to proceed with the formal definition of elliptical distributions, we denote with Ψ_n a class of functions $\psi(t)\colon [0,\infty) \to \mathbb{R}$ such that the function $\psi(\sum_{i=1}^{n} t_i^2)$ is a n-dimensional characteristic function. Then we have: $\Psi_n \subset \Psi_{n-1} \ldots . \subset \Psi_1$

Definition 3.4a: Elliptical distribution [23]

Consider an n-dimensional random vector $X = (X_1, X_2, \ldots X_n)^T$. The random vector X has a multivariate elliptical distribution, denoted by $X \sim E_n(\mu, \Sigma, \psi)$ if its characteristic function can be expressed as:

$$\varphi_X(t) = \exp(it^T\mu)\,\psi\left(\frac{1}{2}t^T\Sigma t\right) \tag{3.4}$$

for some column-vector μ, n x n positive- definite matrix Σ and for some function ψ (t)∈ Ψ_n, which is called the *characteristic generator*.

The probability density function, if it exists, has the form:

$$f_x(x) = \frac{c_n}{\sqrt{|\Sigma|}} g_n\left[\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \tag{3.5}$$

for some function $g_n(\cdot)$, which is called *density generator*.

From (3.4) we have that $X \sim E_n(\mu, \Sigma, \psi)$. As we have already mentioned, elliptical distribution have the property of linear transformation. Therefore, if we denote with A a mxn dimensional matrix of rank m ≤ n and b a m dimensional vector, then:

$$AX + b \sim E_n(A\mu + b, A\Sigma A^T, \psi) \quad (3.6)$$

So any linear combination of elliptical distributions is in fact another elliptical distribution with the same density generator function.

Another way to define elliptical distributions is the following:

Definition 3.4b: Elliptical distributions [26]

X has an elliptical distribution if

$$X \stackrel{d}{=} \mu + AY \quad (3.7)$$

Where $Y \sim S_k(\psi)$ and $A \in \mathbb{R}^{dxk}$ and $\mu \in \mathbb{R}^d$ are a matrix and vector of constants, respectively[6,7].

So, elliptical distributions are obtained by multivariate affine transformations of spherical distributions.

Properties of elliptic distributions [26]:

In this section, we will further examine the properties of elliptical distributions and we will compare them with the properties of multivariate normal distributions. The similarities between elliptical and multivariate normal distributions can make easier to state the assumption that the risk-

[6] Definition: Spherical distributions
A random vector $X=(X_1,\ldots X_d)'$ has a spherical distribution if, for every orthogonal map $U \in \mathbb{R}^{dxd}$

$$UX \stackrel{d}{=} X$$

[7] Theorem 2: The following are equivalent.

(1) X is spherical (A random vector $X=(X_1,\ldots,X_d)'$ has a spherical distribution if for every orthogonal map $U \in \mathbb{R}^{dxd}$ i.e. maps satisfying $UU'=U'U=I_d$) $UX \stackrel{d}{=} X$)>
(2) There exists a function ψ of a scalar variable such that, for all $t \in \mathbb{R}^d$, $\varphi_X(t) = E(e^{it'X}) = \psi(t't) = \psi(t_1^2 + \cdots + t_d^2)$
(3) For every $\alpha \in \mathbb{R}^d$, $\alpha' X \stackrel{d}{=} ||\alpha|| X_1$, where $||\alpha||^2 = a'a = a_1^2 + \cdots + a_d^2$

factors changes have an approximately elliptical distribution, rather than multivariate normal, which is usually false. We say that $X \sim M_d(\mu, \Sigma, \widehat{H})$, so X follows the multivariate normal distribution with μ the mean, Σ the variance-covariance matrix and $\widehat{H} = \int_0^\infty e^{-\theta u} dH(u)$ is the Laplace- Stieltjes transform of the df H of W, where W is a positive mixing variable.

Linear combinations:

Let $X \sim M_d(\mu, \Sigma, \widehat{H})$ and Y=BX+ b, where $B \in \mathbb{R}^{kxd}$ and $b \in \mathbb{R}^k$, then

$Y \sim M_k(B\mu + b, B\Sigma B', \widehat{H})$. So Y remains to be multivariate normal. The exact the same happens when X follows an elliptical distribution. If we take a linear combination of elliptical distribution, then these remain elliptical with the same characteristic generator ψ. Let $X \sim E_d(\mu, \Sigma, \psi)$ and take $B \in \mathbb{R}^{kxd}$ and $b \in \mathbb{R}^d$. Then we have

$$BX + b \sim E_k(B\mu + b, B\Sigma B', \psi)$$

Marginal distributions:

The marginal distribution of X must be elliptical distributions with the same characteristic generator. Let $X = (X_1, X_2)'$ then

$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

Conditional distributions:

In the case of multivariate normal, the conditional distribution of X_2 given X_1 is multivariate normal with the same characteristic generator ψ. In the case of elliptical distributions, the conditional distribution remains elliptical distribution with a different characteristic generator $\widetilde{\psi}$.

Convolutions:

The convolution of two independent elliptical vectors with the same dispersion matrix Σ is also elliptical. If X and Y are independent d-dimensional random vectors satisfying $X \sim E_d(\mu, \Sigma, \psi)$ and $Y \sim E_d(\bar{\mu}, \Sigma, \bar{\psi})$ then we have:

$$X + Y \sim E_d(\mu + \bar{\mu}, \Sigma, \bar{\bar{\psi}})$$

Where $\tilde{\psi}(u) = \psi(u)\bar{\psi}(u)$

Though, if the dispersion matrices of X and Y differ more than a constant factor, then the convolution may not be elliptical distributed, even if the two generators are identical.

Furthermore, we denote with X_{sum} the sum of X_i with i=1,….n variables which follow the elliptic distribution, i.e. $X_{sum} = X_1 + X_2 + \cdots + X_n = \overline{1}^T X$. Then, by using (3.6) it follows that:

$$X_{sum} \sim E_1(\overline{1}^T \mu, \overline{1}^T \Sigma \overline{1}, \psi)$$

Similarly, if we denote with $X_{weighted}$the weighted sum of X_i variables which follow the elliptic distribution with i=1,…n, then we have:

$$X_{weighted} \sim E_1\left(\theta^T \mu, \theta^T \Sigma\right)\theta, \psi)$$

Those properties are the tools which help us handling the subadditivity problem of VaR. Due to the fact that elliptical distributions are a generalization of multivariate normal distributions; they hold all the essential properties of normality along with the fact that elliptical distributions allow heavy tails and kurtosis in general. So these distributions have a large range of applications, especially in financial problems where normality hardly applies.

In risk management, elliptical distributions are amenable to the standard approaches, a fact that makes them extremely useful. In the elliptical world, Value at Risk is a coherent risk measure, holding all the properties of coherence, including subadditivity. So, computing the Value at Risk of variables which follow the elliptical distribution is just a routine and we do not have to be afraid about the violation of the subadditivity.

Theorem 3.1: Subadditivity of VaR for elliptical risk factors

(Quantitative Risk Management [26], McNeil, Freyand and Embrechts, p.242-243)

Suppose that $X \sim E_d(\mu, \Sigma, \psi)$ and define the set $\mathcal{M}$ of linearized portfolio losses of the form

$$\mathcal{M} = \left\{ L: L = \lambda_0 + \sum_{i=1}^{d} \lambda_i X_i, \lambda_i \in \mathbb{R} \right\}$$

Then for any two losses $L_1, L_2 \in \mathcal{M}$ and $0.5 \leq \alpha < 1$,

$$\mathrm{VaR}_\alpha(\mathrm{L}_1+\mathrm{L}_2) \le \mathrm{VaR}_\alpha(\mathrm{L}_1)+\mathrm{VaR}_\alpha(\mathrm{L}_2)$$

Proof:

Without any loss of generality, we assume that $\lambda_0=0$. For any $L \in \mathcal{M}$ it follows from the definition of elliptical distribution that we can write $L=\lambda' X \stackrel{d}{=} \lambda' AY+\lambda'\mu$ for a spherical random vector $Y \sim S_k(\psi)$, a matrix $A \in \mathbb{R}^{dxk}$ and a constant vector $\mu \in \mathbb{R}^d$. By part (3) of Theorem 2 we have

$$L \stackrel{d}{=} ||\lambda' A|| Y_1+\lambda'\mu$$

showing that every $L \in \mathcal{M}$ is a random variable of the same type. Moreover, the translation invariance and homogeneity of VaR imply that $L=\lambda' X$,

$$\mathrm{VaR}_\alpha(\mathrm{L}) = ||\lambda' A|| \, VaR_\alpha(Y_1)+\lambda'\mu$$

Now, set $\mathrm{L}_1=\lambda_1'\mathrm{X}$ and $\mathrm{L}_2=\lambda_2'\mathrm{X}$. Since $||(\lambda_1+\lambda_2)'\mathrm{A}|| \le ||\lambda_1'\mathrm{A}|| + ||\lambda_2'\mathrm{A}||$ and since $\mathrm{VaR}_\alpha(\mathrm{Y}_1) \ge 0$ for $\alpha \ge 0.5$, the results follows.

■

Theorem 3.1 guarantees the subadditivity of Value at Risk for elliptically distributed losses. In theorem 3.2 it can be seen that when we deal with elliptical distributions, the Markowitz[8] variance-minimizing portfolio does minimize risk measures like Value at Risk, which is coherent in the elliptical world.

Theorem 3.2 [11]:

Suppose $X \sim \mathrm{E}_n(\mu,\Sigma,\psi)$ with $\sigma^2[\mathrm{X_i}] < \infty$ for all i. Let

$$\mathrm{P} = \left\{ \mathrm{Z} = \sum_{i=1}^{n} \lambda_i \mathrm{X_i} \,|\, \lambda_i \in \mathbb{R} \right\}$$

be the set of all linear portfolios. Then the following are true.

1. (Subadditivity of VaR) For any two portfolios $\mathrm{Z}_1,\mathrm{Z}_2 \in \mathrm{P}$ and $0.5 \le \alpha \le 1$,

[8] When we use the Markowitz variance approach, we develop a mean-variance analysis for selecting a portfolio of common stocks. Someone can find more information in the paper "Asset Allocation Models Using the Markowitz Approach", written by Paul D. Kaplan, in January 1998.

$$\mathrm{VaR}_\alpha(Z_1+Z_2) \le \mathrm{VaR}_\alpha(Z_1)+\mathrm{VaR}_\alpha(Z_2)$$

2. (Equivalence of variance and positive homogenous risk measurement). Let ϱ be a real-valued risk measure on the space of real-valued random variables which depends only on the distribution of a random variable X. Suppose this measure satisfies the property of positive homogeneity. Then for $Z_1, Z_2 \in P$:

$$\varrho(Z_1-\mathbb{E}[Z_1]) \le \varrho(Z_2-\mathbb{E}[Z_2]) \Leftrightarrow \sigma^2[Z_1] \le \sigma^2[Z_2]$$

3. (Markowitz risk- minimizing portfolio). Let ϱ be as in 2, and assume that the property of translation invariance is also satisfied. Let

$$\mathcal{E} = \left\{ Z = \sum_{i=1}^{n} \lambda_i X_i \middle| \lambda_i \in \mathbb{R}, \sum_{i=1}^{n} \lambda_i = 1, \mathbb{E}[Z] = r \right\}$$

Be the subset of portfolios giving expected return r. Then

$$\mathrm{argmin}_{Z\in\mathcal{E}}\varrho(Z) = \mathrm{argmin}_{Z\in\mathcal{E}}\sigma^2[Z]$$

<u>Proof:</u>

1. Let q_α be the α-quantile of the standardized distribution of this type. Then,

$$\mathrm{VaR}_\alpha(Z_1) = \mathbb{E}[Z_1] + \sigma[Z_1]q_\alpha$$
$$VaR_\alpha(Z_2) = \mathbb{E}[Z_2] + \sigma[Z_2]q_\alpha$$
$$VaR_\alpha(Z_1+Z_2) = \mathbb{E}[Z_1+Z_2] + \sigma[Z_1+Z_2]q_\alpha$$

Since $\sigma[Z_1+Z_2] \le \sigma[Z_1]+\sigma[Z_2]$ and $q_\alpha \ge 0$ the results follows.

2. Since Z_1 *and* Z_2 are random variables of the same type, there exists an $\alpha > 0$ such that $Z_1 - \mathbb{E}[Z_1] = \alpha(Z_2-\mathbb{E}[Z_2])$. It follows that

$$\varrho(Z_1-\mathbb{E}[Z_1]) \le \varrho(Z_2-\mathbb{E}[Z_2]) \Leftrightarrow \alpha \le 1 \Leftrightarrow \sigma^2[Z_1] \le \sigma^2[Z_2]$$

3. Follows from 2 and the fact that we optimize over portfolios with identical expectation.

■

<u>Remark:</u>

We will follow our study by introducing the term of *mean-VaR*. We denote by μ the mean of the loss distribution. Sometimes, instead of VaR, we use the statistic

$$\mathrm{VaR}_{\alpha}^{\mathrm{mean}} = \mathrm{VaR}_{\alpha} - \mu \tag{3.8}$$

We suppose to have a time horizon equals to a day. Then the mean-VaR refers as *daily earnings at risk*. When we refer to market risk management, the distinction between VaR and $\mathrm{VaR}_{\alpha}^{\mathrm{mean}}$ does not play an exceptional role, due to the fact that time horizon is usually short. However, in credit market, where the time horizon is longer or in asset management risk, $\mathrm{VaR}_{\alpha}^{\mathrm{mean}}$ does play an exceptional role, and is more frequently used than VaR.

Example 3.5 [11]:

Suppose that $X = (\mathrm{X}_1, \ldots\ldots \mathrm{X}_{\mathrm{n}})^{\mathrm{t}}$ represents n risks with an elliptical distribution. Also, we have linear portfolios of such risks, which are

$$\left\{\mathrm{Z} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_{\mathrm{i}} \mathrm{X}_{\mathrm{i}} | \lambda_{\mathrm{i}} \in \mathbb{R}\right\}$$

with distribution F_{Z}.Then, the Value at Risk of portfolio Z at probability level α is given by:

$$\mathrm{VaR}_{\alpha}(\mathrm{Z}) = \mathrm{F}_{\mathrm{Z}}^{-1}(\alpha) = \inf\{\mathrm{z} \in \mathbb{R}: \mathrm{F}_{\mathrm{Z}}(\mathrm{z}) \geq \alpha\}$$

■

SPECIAL CASES OF ELLIPTIC DSTRIBUTIONS

Examples [22]:

(Landsman and Valdez: Tail Conditional Expectations for Elliptical Distributions)

3.6 Multivariate Student-t Family (page 59)

An elliptically distributed vector X has a multivariate student-t distribution if its density generator can be expressed as:

$$g_n(u) = \left(1 + \frac{u}{k_p}\right)^{-p}$$

where the parameter $p > \frac{n}{2}$ and k_p is some constant that may depend on p. We write $X \sim t_n(\mu, \Sigma, p)$ if X belongs to this family. Its joint density therefore has the form:

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}}\left[1 + \frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2k_p}\right]^{-p}$$

Using equation:

$$c_n = \frac{\Gamma(\frac{n}{2})}{(2\pi)^{n/2}} \left[\int_0^\infty x^{\frac{n}{2}-1} g_n(x)dx\right]^{-1} \qquad (*)$$

it can be shown that the normalizing constant is:

$$c_n = \frac{\Gamma(p)}{\Gamma(p-n/2)}(2\pi k_p)^{-n/2}$$

Here, we introduce the multivariate student-t distribution in its most general form. Taking for example, $p = \frac{(n+m)}{2}$, where n and m are integers and $k_p = \frac{m}{2}$ we get the traditional form of the multivariate student-t distribution with density

$$f_X(x) = \frac{\Gamma[(n+m)/2]}{(\pi m)^{n/2}\Gamma(m/2)\sqrt{|\Sigma|}}\left[1 + \frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{m}\right]^{-(n+m)/2} \qquad (3.9)$$

In the univariate case where n=1, Bian and Tiku (1997) and MacDonald (1996) suggested putting $k_p = (2p-3)/2$ if p>3/2 to get the so-called generalized student-t (GST) univariate distribution with density. This normalization leads to the important property that Var(X) = σ^2. Extending this to the multivariate case, we suggest keeping $k_p = (2p-3)/2$ if p>3/2, then this multivariate GST has the advantage that Cov(X)= Σ

In particular, for p= (n+ m) /2 we suggest, instead of equation (3.9), considering

$$f_X(x) = \frac{\Gamma[(n+m)/2]}{[\pi(n+m-3)]^{n/2}\Gamma(m/2)\sqrt{|\Sigma|}}\left[1+\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{n+m-3}\right]^{-(n+m)/2} \quad (3.10)$$

Because it also has the property that the covariance is Cov (X) =Σ. If $1/2<p\leq 3/2$, the variance does not exist and we have a heavy-tailed multivariate distribution. If $1/2<p\leq 1$, even the expectation does not exist. In the case where p=1, we have the *multivariate Cauchy distribution* with density:

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2}\sqrt{|\Sigma|}}[1+(x-\mu)^T\Sigma^{-1}(x-\mu)]^{-(n+1)/2} \quad (3.11)$$

3.7 Multivariate Logistic Family (page 59)

An elliptical vector X belongs to the family of multivariate logistic distributions if its density generator has the form:

$$g_n(u) = \frac{e^{-u}}{(1+e^{-u})^2} \quad (3.12)$$

Its joint density has the form:

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}}\frac{\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]}{\left\{1+\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]\right\}^2} \quad (3.13)$$

Where the normalizing constant can be evaluated using equation $(*)$ as follows[9]:

$$c_n = (2\pi)^{-n/2}\left[\sum_{j=1}^{\infty}(-1)^{j-1}j^{1-n/2}\right]^{-1} \quad (3.14)$$

If X belongs to the family of multivariate logistic distributions, we shall write $X \sim ML_n(\mu,\Sigma)$

---

[9] $\left(\frac{e^{-x}}{(1+e^{-x})^2}\right) = \sum_{j=1}^{\infty}(-1)^{j-1}je^{jx}$

3.8 Multivariate Exponential Power Family (page 60)

An elliptical vector X belongs to the family of multivariate exponential power distributions if its density generator has the form:

$$g(u) = e^{-ru^s} \quad for\; r, s > 0 \tag{3.15}$$

The jointly density of X can be expressed in the form:

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}} \exp\left\{-\frac{r}{2^s}[(x-\mu)^T\Sigma^{-1}(x-\mu)]^s\right\} \tag{3.16}$$

Where the normalizing constant is given by:

$$c_n = \frac{\Gamma(\frac{n}{2})}{(2\pi)^{n/2}}\left(\int_0^\infty x^{\frac{n}{2}-1}e^{-rx^s}dx\right)^{-1}$$

$$= \frac{s\Gamma(n/2)}{(2\pi)^{\frac{n}{2}}\Gamma(\frac{n}{2s})} r^{n/(2s)}$$

When r=s=1, this family of distributions clearly reduces to the multivariate normal family. When s=1 alone, this family reduces to the original Kotz multivariate distribution suggested by Kotz in 1975. If s=1/2 and r=$\sqrt{2}$, we have the family of double exponential or Laplace distributions.

■

The next graph illustrates the differences between the normal and the student distribution with mean 0 and the same variance which is equal to 1. From this graph, the differences in the tail become clear.

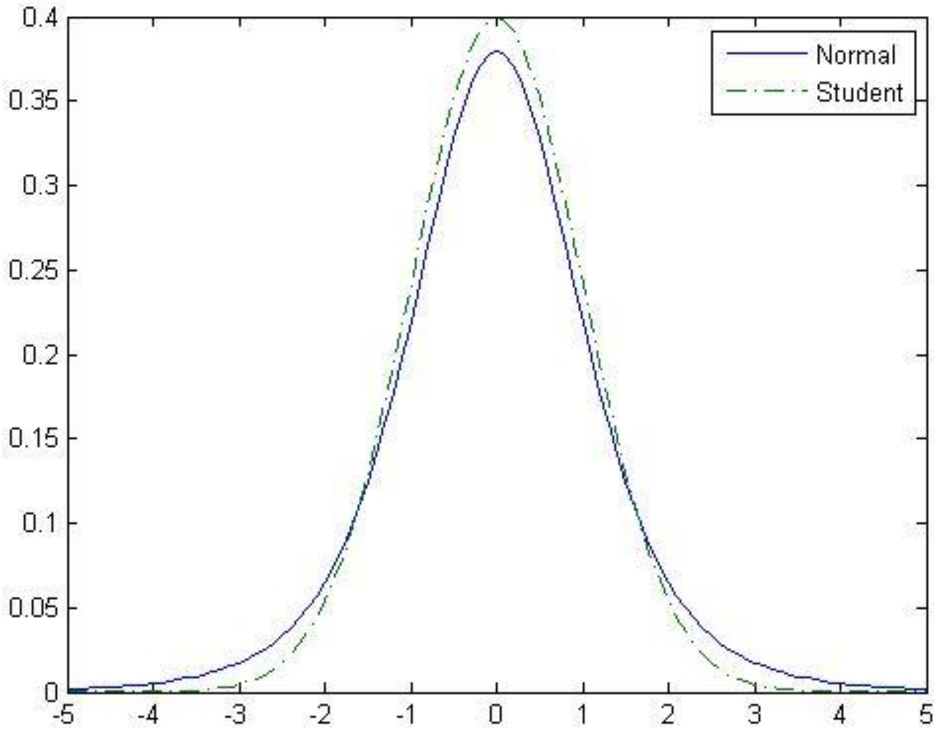

Figure 3.1: Comparison of Normal and Student distribution

Generally, VaR has been criticized for its lack of subadditivity, especially in the case where fat tails are present. Fat tails are extremely common in finance. However, in chapter 3 we proved that there are cases where subadditivity stands, making VaR a good risk measure. Of course, analysts want a risk measure which will not be standing only in some cases but generally. That is why Conditional Value at Risk was made.

CHAPTER 4

CONDITIONAL VALUE AT RISK

4.1 Conditional Value at Risk

As we examined in the previous chapters, Value at Risk is a widespread risk measure which is easy to use and implement but with serious drawbacks.

Value at Risk is a risk measure which can be applied to every type of portfolio and allows the comparison among the risks factors of different portfolios. Furthermore, Value at Risk gives us the opportunity to aggregate the risks, taking into account the internal correlations among all risk factors. Additional, Value at Risk is able to assess on a complete portfolio and not only on individual positions in it, which is a major difference comparing to other risk measures. Moreover, due to the fact that Value at Risk is a probabilistic risk measure, gives to the analyst all the necessary information on the probabilities which are associated with specific loss amounts. Last but not least, Value at Risk is extremely easy to implementation.

However, Value at Risk is extremely untrustworthy when the loss distribution is not normal but tends to be fat tailed, a fact which is very common in real data. Furthermore, Value at Risk ignores and does not take into account what happens after the exceeding point, or else the threshold, that this measure has indicated itself. Value at Risk only tells us what we can lose in “good” states, where a tail event does not occur, but not in “bad” states, when a tail event does occur, leading the analysts to wrong results. So, that measure can be thought of as a “biased one towards optimism instead of the conservatism that ought to prevail in risk management”, as Rockafellar and Uryasev point out in their paper “Conditional Value at Risk for General Loss Distributions” [29]. In addition, the fact that Value at Risk does not take into account what happens in “bad” states can encourage traders to promote their own interests at the expense of the institutions they work in. Last but not least, Value at Risk is not a coherent risk measure, based on the definition of Artzner, because it does not satisfy the property of subadditivity unless in the case of standard normal distribution where VaR is proportional to standard deviation, so VaR is coherent.

Given these problems, the need of a new alternative risk measure which will be able to over pass these issues is vital. This alternative risk measure should be able to keep all the beneficial properties of Value at Risk along with the ability to solve the problems arising from it. For these reasons, Artzner et al proposed the Conditional Value at Risk.

However, when we want to define Conditional Value at Risk (CVaR), we have to make a distinction between the random variables which follow a continuous distribution and the random variables which follow a distribution which are discrete or with jumps.

When the underlying factors follow a continuous distribution, then the CVaR is the conditional expected loss given that the loss is greater than or equal to the value of VaR.

Definition 4.1: Conditional Value at Risk

Suppose X is a continuous random variable denoting the loss of a given portfolio and $VaR_\alpha(X)$ is the VaR at the 100(1-α) percent confidence level, with α∈ [0, 1]. The CVaR of X is the mean of the generalized α-tail distribution:

$$\boxed{\mathrm{CVaR}_{\alpha}(\mathrm{X}) = \mathrm{E}[\mathrm{X}|\mathrm{X} \geq \mathrm{VaR}_{\alpha}(\mathrm{X})]} \qquad (4.1)$$

Conditional Value at Risk actually measures the amount of money someone can lose on average in states beyond Value at Risk level. That holds for continuous distributions. Generally, <u>for both continuous and discontinuous distributions</u> CVaR can be expressed as followed:

$$\boxed{\mathrm{CVaR}_{\alpha}(\mathrm{X}) = \int_{-\infty}^{\infty} \mathrm{z}\mathrm{dF}_{\mathrm{X}}^{\alpha}(\mathrm{z})} \qquad (4.2)$$

where :

$$\mathrm{F}_{\mathrm{X}}^{\alpha}(\mathrm{z}) = \begin{cases} 0, \ \text{when z} < \mathrm{VaR}_{\alpha}(\mathrm{X}) \\ \dfrac{\mathrm{F}_{\mathrm{X}}(\mathrm{z}) - \alpha}{1-\alpha}, \text{when z} \geq \mathrm{VaR}_{\alpha}(\mathrm{X}) \end{cases}$$

In the case of a continuous distribution CVaR can be also called Mean Excess Loss, Expected Shortfall, Mean Shortfall or Tail Value at Risk.

However, when we have to deal with discrete distributions or with distributions which may have discontinuities, which is very common in everyday life, we must be more careful. To define CVaR in that case, we must first introduce two new quantities, $\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X})$ and $\mathrm{CVaR}_{\alpha}^{-}(\mathrm{X})$. The $\mathbf{CVaR}_{\boldsymbol{\alpha}}^{+}(\mathbf{X})$ is called **Expected Shortfall (ES)** or else “upper CVaR” and $\mathbf{CVaR}_{\boldsymbol{\alpha}}^{-}(\mathbf{X})$ is called **Tail VaR** or else “lower CVaR”, where:

$$\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X}) = \mathrm{E}[\mathrm{X}|\mathrm{X} > \mathrm{VaR}_{\alpha}(\mathrm{X})] \qquad (4.3)$$

and

$$\mathrm{CVaR}_{\alpha}^{-}(\mathrm{X}) = \mathrm{E}[\mathrm{X}|\mathrm{X} \geq \mathrm{VaR}_{\alpha}(\mathrm{X})] \qquad (4.4)$$

Generally, we have: $\mathrm{VaR} \leq \mathrm{CVaR}^{-} \leq \mathrm{CVaR} \leq \mathrm{CVaR}^{+}$. The equality stands only in the case where there are no jumps at the VaR threshold. Otherwise, both the inequalities can be strict. Thus, when we have to deal with continuous distributions CVaR and ES coincides, but when the distributions are discrete, CVaR may differ from ES.

Using the definition (4.3) we can define alternative CVaR by weighting both $\mathrm{VaR}_{\alpha}(\mathrm{X})$ and $\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X})$:

$$\boxed{\mathrm{CVaR}_{\alpha}(\mathrm{X}) = \lambda_{\alpha}(\mathrm{X}) VaR_{\alpha}(\mathrm{X}) + \big(1 - \lambda_{\alpha}(\mathrm{X})\big)\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X})} \tag{4.5}$$

where:

$$\lambda_{\alpha}(X) = \frac{F_X(VaR_{\alpha}(X) - \alpha)}{1 - \alpha} \tag{4.6}$$

The definition (4.5) makes the relation between CVaR and VaR even clearer. When $F_X\big(VaR_{\alpha}(X)\big) = 1$, where $\mathrm{VaR}_{\alpha}(\mathrm{X})$ is the biggest loss that our portfolio can occur in the portfolio, then $\mathrm{CVaR}_{\alpha}(\mathrm{x}) = \mathrm{VaR}_{\alpha}(\mathrm{x})$.So, the relation (4.5) can be written as follows:

$$\mathrm{CVaR}_{\alpha}(\mathrm{X}) = \begin{cases} \lambda_{\alpha}(\mathrm{X}) VaR_{\alpha}(\mathrm{X}) + \big(1 - \lambda_{\alpha}(\mathrm{X})\big)\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X}), & \text{if } \mathrm{F_X}\big(\mathrm{VaR}_{\alpha}(\mathrm{X})\big) < 1 \\ \mathrm{VaR}_{\alpha}(\mathrm{X}), & \text{if } \mathrm{F_X}\big(\mathrm{VaR}_{\alpha}(\mathrm{X})\big) = 1 \end{cases} \tag{4.7}$$

The weights in the relationship (4.7) arise from the way that CVaR splits the atom of probability at the VaR value, a situation which we will analyze further below. The use and the importance of relationship (4.7) is major. Despite the fact that neither VaR nor $\mathrm{CVaR}_{\alpha}^{+}(\mathrm{X})$ behave well in general, because a vast number of cases have a loss distribution which is discontinuous, CVaR manages to hold important properties: it is both continuous with respect to α and is jointly convex in (X,α).

When the distribution of the random variables is continuous, we have that $CVaR = \mathrm{CVaR}^{-}$. So, in the case of a continuous distribution of random variables CVaR and ES coincide.

In the case of a discrete loss distribution, where y is the stochastic parameter and can take the values $y_1, y_2, \dots y_J$ with probabilities θ_j, with j=1,….,J. Then CVaR is defined as [20]:

$$\mathrm{CVaR}_\alpha(\mathrm{x}) = \frac{1}{1-\alpha}\left[\left(\sum_{\mathrm{j}=1}^{\mathrm{j}_\alpha} \theta_j - \alpha\right) \mathrm{f}(\mathrm{x}, \mathrm{y}_{\mathrm{j}_\alpha}) + \sum_{\mathrm{j}=\mathrm{j}_\alpha+1}^{\mathrm{J}} \theta_j \mathrm{f}(\mathrm{x}, \mathrm{y}_\mathrm{j})\right] \tag{4.8}$$

Where j_α satisfies:

$$\sum_{j=1}^{\mathrm{j}_\alpha - 1} \theta_j < \alpha \leq \sum_{j=1}^{j_\alpha} \theta_j$$

The big innovation of the definition of CVaR is that for discrete distributions, CVaR can split the atom. If $\mathrm{F_X(x)}$ presents a vertical gap, then there is an interval of confidence level α where we have the same value of VaR. This interval has the endpoints:

$$[\alpha^-, \alpha^+] = [F_\mathrm{X}(\mathrm{VaR}_\alpha^-(\mathrm{X})), F_\mathrm{X}(\mathrm{VaR}_\alpha(\mathrm{X}))] = [\mathrm{P}\{\mathrm{X} < \mathrm{VaR}_\alpha(\mathrm{X})\}, F_\mathrm{X}(\mathrm{VaR}_\alpha(\mathrm{X}))]$$

When $F_\mathrm{X}(\mathrm{VaR}_\alpha^-(\mathrm{X})) < \alpha < F_\mathrm{X}(\mathrm{VaR}_\alpha(\mathrm{X})) < 1$, the atom $\mathrm{VaR}_\alpha(\mathrm{X})$, which has total probability $\alpha^+ - \alpha^-$, is split by the confidence level α n two pieces with probabilities $\alpha^+ - \alpha$ and $\alpha - \alpha^-$[31]. To explain better the idea of splitting we illustrate the following examples, which can be found in the "Value at Risk vs. Conditional Value at Risk in Risk Management and Optimization" written by Sarykalin, Serraino and Uryasev:

"Suppose we have six equally likely scenarios with losses $\mathrm{f}_1, \dots, \mathrm{f}_6$. Let: $\alpha = \frac{2}{3} = \frac{4}{6}$. In this case α does not split the atom

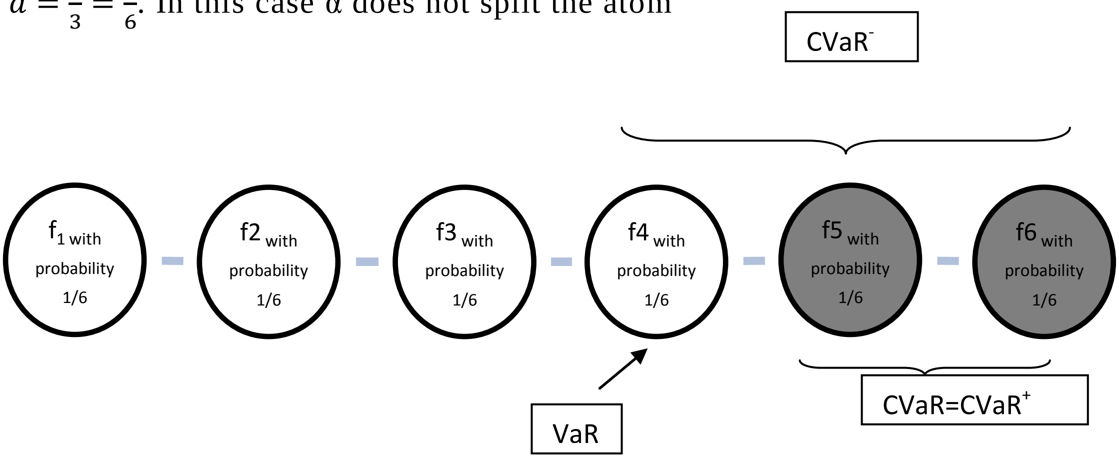

Figure 4.1: Computation of CVaR when α does not split the atom

Then $\mathrm{VaR}_\alpha(\mathrm{X}) < \mathrm{CVaR}_\alpha^-(\mathrm{X}) < \mathrm{CVaR}_\alpha(\mathrm{X}) = \mathrm{CVaR}_\alpha^+(\mathrm{X})$ and

$\lambda_\alpha(X) = \frac{\mathrm{F_X}(\mathrm{VaR}_\alpha(X)-\alpha}{1-\alpha} = 0$ and

$$\mathrm{CVaR}_\alpha(\mathrm{X}) = \frac{1}{5}\mathrm{VaR}_\alpha(\mathrm{X}) + \frac{4}{5}\mathrm{CVaR}_\alpha^+(\mathrm{X}) = \frac{1}{5}\mathrm{f}_5 + \frac{1}{5}\mathrm{f}_6$$

Now, let $\alpha = \frac{7}{12}$

In that case α does split the $\mathrm{VaR}_\alpha(\mathrm{X})$ atom, and:

$\lambda_\alpha(\mathrm{X}) = \frac{F_{\mathrm{X}}(\mathrm{VaR}_\alpha(\mathrm{X})-\alpha)}{1-\alpha} > 0$ and

$\mathrm{CVaR}_\alpha(\mathrm{X}) = \frac{1}{5}\mathrm{VaR}_\alpha(\mathrm{X}) + \frac{4}{5}\mathrm{CVaR}_\alpha^+(\mathrm{X}) = \frac{1}{5}\mathrm{f}_4 + \frac{2}{5}\mathrm{f}_5 + \frac{2}{5}\mathrm{f}_6$

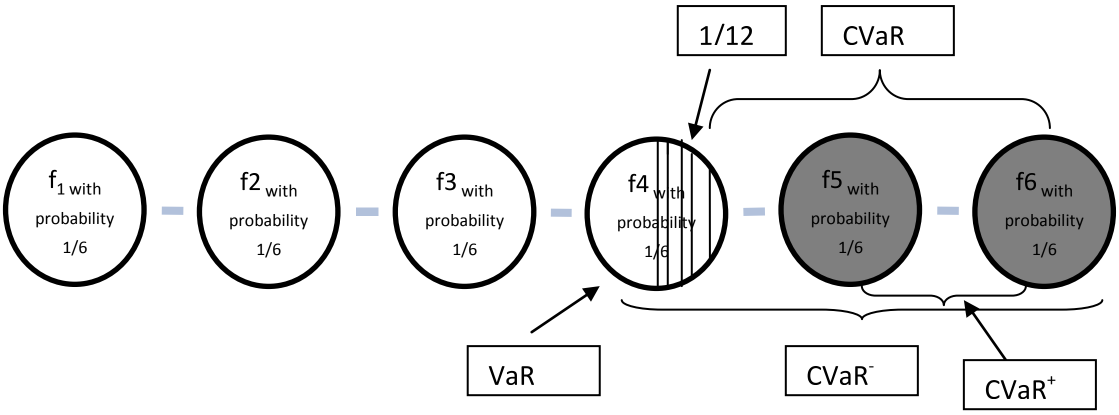

Figure 4.2: Computation of CVaR when α does split the atom

The biggest advantage of CVaR, as we mentioned before is the fact that CVaR is a coherent risk measure, i.e. it follows all the four properties of the coherence, including subadditivity. The first, third and forth property of the coherence are easy to be proven. We are going to prove the second property, i.e. the property of subadditivity.

Subadditivity of CVaR [26]

We denote with, using the notation from the book “Quantitative Risk Management” by McNeil, Frey and Embrechts [26]:

$\mathbf{L_1}, \ldots, \mathbf{L_n}$**:** is a generic sequence of random variables with associated order statistics $L_{1,n} \geq \ldots \geq L_{n,n}$ and note that for arbitrary *m* satisfying $1 \leq m \leq n$ we have:

$$\sum_{i=1}^{m} L_{i,n} = \sup\{L_{i_1} + \cdots + L_{i_m}: 1 \le i_1 < \cdots < i_m \le m\}$$

Now consider two random variables L and $\tilde{L}$ with joint df F and a sequence of iid bivariate random vectors $(L_1, \tilde{L}_1)$... $(L_n, \tilde{L}_n)$ with the same df F. Writing $(L + \tilde{L})_i \coloneqq L_i + \tilde{L}_i$ and $(L + \widetilde{L})_{i,n}$ for an order statistic of $(L + \tilde{L})_1 \ldots . (L + \tilde{L})_n$, we observe that we must have:

$$\sum_{i=1}^{m} (L + \tilde{L})_{i,n} = sup\left\{(L + \tilde{L})_{i_1} + \cdots + (L + \tilde{L})_{i_m}: 1 \le i_1 < \cdots < i_m \le m\right\}$$

$$\le sup\{L_{i,1} + \cdots + L_{i,m}: 1 \le i_1 < \cdots < i_m \le m\}$$
$$+ sup\{\tilde{L}_{i_1} + \cdots + \tilde{L}_{i_1}: 1 \le i_1 < \cdots < i_m \le m\}$$

$$= \sum_{i=1}^{m} L_{1,n} + \sum_{i=1}^{m} \tilde{L}_{i,n}$$

By setting m= [n(1-α)] and letting n→∞ we have that

$$\mathrm{CVaR}_{\alpha}(L + \tilde{L}) \le \mathrm{CVaR}_{\alpha}(L) + \mathrm{CVaR}_{\alpha}(\tilde{L})"$$

■

CVaR has some other properties too. As we can see in the book of Pflug [27] *"Some remarks on VaR and CVaR"*, CVaR satisfies the following properties:

i. CVaR_{α} is *translation- equivariant*, i.e.

$$\mathrm{CVaR}_{\alpha}(Y + c) = \mathrm{CVaR}_{\alpha}(Y) + c$$

ii. CVaR_{α} is *positively homogenous*, i.e.

$$\mathrm{CVaR}_{\alpha}(cY) = cCVaR(Y), if\ c > 0$$

iii. If Y has a density,

$$\mathbb{E}(Y) = (1 - \alpha)\mathrm{CVaR}_{\alpha}(Y) - \alpha CVaR_{(1-\alpha)}(-Y)$$

iv. CVaR_{α} is *convex* in the following sense: For arbitrary (possibly dependent) random variables Y_1 and Y_2 and 0<λ<1,

$$\mathrm{CVaR}_{\alpha}(\lambda Y_1 + (1 - \lambda)Y_2) \le \lambda CVaR_{\alpha}(Y_1) + (1 - \lambda)CVaR_{\alpha}(Y_2)$$

v. CVaR_α is *monotonic* w.r.t. SD(2)[10] (and a fortiori w.r.t. SD(1)), i.e. if $Y_1 \prec_{SD(2)} Y_2$, then

$$\mathrm{CVaR}_\alpha(Y_1) \leq \mathrm{CVaR}_\alpha(Y_2)$$

vi. CVaR_α is *monotonic* w.r.t. MD(2)[11], i.e. if $Y_1 \prec_{MD(2)} Y_2$ then

$$\mathrm{CVaR}_\alpha(Y_1) \leq \mathrm{CVaR}_\alpha(Y_2)$$

Properties (i) and (ii) are obvious from the definition of CVaR. In that point we will prove property (iii) and (iv).

Proof (iii) [27]:

$$\begin{aligned}\mathrm{CVaR}_{(1-1)}(-Y) &= \mathbb{E}\big(-Y| - Y \geq \mathrm{VaR}_{(1-\alpha)}(-Y)\big)\\ &= \mathbb{E}\big(-Y\big|-Y \geq -\mathrm{VaR}_\alpha(Y)\big)\\ &= -\mathbb{E}(Y|Y \leq \mathrm{VaR}_\alpha(Y))\end{aligned}$$

One sees that:

$$\begin{aligned}\mathbb{E}(Y) &= \alpha\mathbb{E}\big(Y\big|Y \leq \mathrm{VaR}_\alpha(Y)\big) + (1-\alpha)\mathbb{E}(Y|Y \geq VaR_\alpha(Y))\\ &= \alpha \mathrm{CVaR}_{(1-\alpha)}(-Y) + (1-\alpha)\mathrm{CVaR}_\alpha(Y)\end{aligned}$$

Proof (iv) [27]:

Let α_i be such that:

$$\mathrm{CVaR}_\alpha(Y_i) = \alpha_i + \frac{1}{1-\alpha}\mathbb{E}[Y_i - \alpha_i]^+$$

.

Since $y \rightarrow [y-\alpha]^+$ is convex, we have:

$$\mathrm{CVaR}_\alpha(\lambda Y_1 + (1-\lambda Y_2)$$

$$\leq \lambda\alpha_1 + (1-\lambda)\alpha_2 + \tfrac{1}{1-\alpha}\mathbb{E}[\lambda Y_1 + (1-\lambda) - \lambda\alpha_1 + (1-\lambda)\alpha_2]^+$$

$$\leq \lambda\alpha_1 + (1-\lambda)\alpha_2 + \tfrac{1}{1-\alpha}\mathbb{E}[Y_1 - \alpha_1]^+ + \tfrac{1-\lambda}{1-\alpha}\mathbb{E}[Y_2 - \alpha_2]^+$$

---

[10] SD= Stochastic dominance is a term which refers to a set of relations that may hold between a pair of distributions. We say that SD (2) is stochastic dominance of order 2 when the relation $Y_1 \prec_{SD(2)} Y_2$ holds if and only if $\mathbb{E}[\psi(Y_1)] \leq \mathbb{E}[\psi(Y_2)]$ for all integrable concave, monotonic functions ψ.

[11] MD= Monotonic dominance. We say that MD (2) is monotonic dominance of order 2 when the relation $Y_1 \prec_{MD(2)} Y_2$ holds if and only if $\mathbb{E}[\psi(Y_1)] \leq \mathbb{E}[\psi(Y_2)]$ for all integrable concave functions ψ.

$$\leq CVaR_{\alpha}(Y_1) + (1-\lambda)CVaR_{\alpha}(Y_2)$$

The properties (v) and (vi) follow from the fact that y→$[y-\alpha]^+$ is monotone and convex.

■

The fact that CVaR is a convex function of portfolio positions is substantial. Additionally, the set of the minimum points is convex too on a convex set, a fact that makes the minimization process of CVaR easier.

We illustrate some examples for our better understanding of the concept of CVaR.

Example 4.1: CVaR for Normal loss distribution [26]:

"Suppose that the loss distribution F_L is normal with mean μ and variance σ^2. Fix α∈ (0, 1). Then:

$$\mathrm{CVaR}_{\alpha} = \mu + \sigma\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}$$

where φ is the density of the standard normal distribution.

First note that

$$\mathrm{CVaR}_{\alpha} = \mu + \sigma E\left(\left\{\frac{L-\mu}{\sigma}\middle|\frac{L-\mu}{\sigma} \geq q_{\alpha}\left(\frac{L-\mu}{\sigma}\right)\right\}\right)$$

Hence it suffices to compute the Conditional Value at Risk for the standard normal random variable $\tilde{L} \coloneqq \frac{L-\mu}{\sigma}$. Here we get:

$$\mathrm{CVaR}_{\alpha}(\tilde{L}) = \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} l\varphi(l)dl = \frac{1}{1-\alpha}[-\varphi(l)]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\varphi(\Phi^{-1}(\alpha))}{1-a}$$

■

Example 4.2: CVaR for Student t loss distribution [26]:

Suppose the loss L is such that $\tilde{L} = \frac{L-\mu}{\sigma}$ has a standard t distribution with ν degrees of freedom. Suppose further that ν>1. By the reasoning of example 3.1, which applies to any location- scale family, we have

$$\mathrm{CVaR}_{\alpha} = \mu + \sigma CVaR_{\alpha}(\tilde{L})$$

The Conditional Value at Risk of the standard t distribution is easily calculated by direct integration to be

$$\mathrm{CVaR}_{\alpha}(\tilde{\mathrm{L}}) = \frac{\mathrm{g}_{\nu}(\mathrm{t}_{\nu}^{-1}(\alpha))}{\nu - 1}\left(\frac{\nu + (\mathrm{t}_{\nu}^{-1}(\alpha))^{2}}{\nu - 1}\right)$$

Where t_{ν} denotes the df and g_{ν} the density of standard t.

■

4.2 INTERPRETATION

VaR is actually, as we have already mentioned the quantile. If for example, we want to calculate VaR with α=5%, then the result indicates that 5% of the cases will overpass the 5% quantile. However, the big disadvantage of VaR is that it does not give us any information of how much the loss will be.

That is the great difference from CVaR. CVaR is the expected, or else, the average value of the loss in the tail. CVaR is taking into account only the area of the distribution which is exceeding the value of VaR and it calculates the averages loss in that precise area. So CVaR is the average of the tail. However, there is no insurance that the loss will not be further to the right. So graphically, if we assume that we have a standard normal distribution and we want to calculate the 5% VaR we have:

$$\mathrm{CVaR}_{\alpha}(\mathrm{X}) = \mathrm{E}(\mathrm{X}|\mathrm{X} > q) = \int_{\alpha}^{1} \frac{1}{1-a} q_{\mathrm{p}} \mathrm{dp}$$

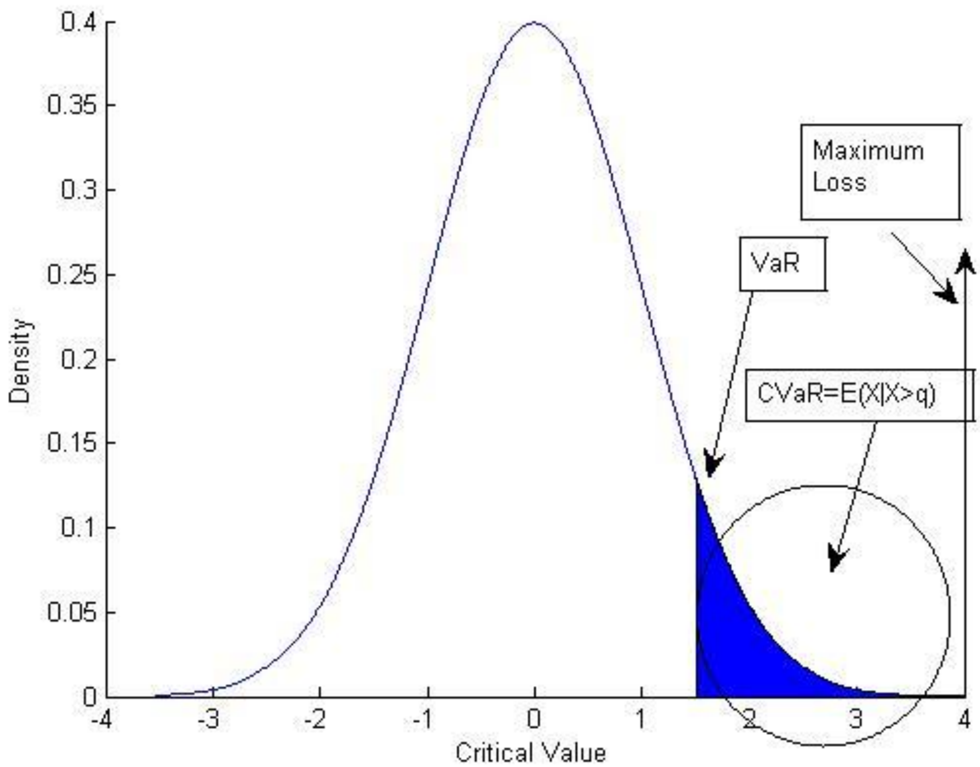

Figure 4.3: Loss Distribution, VaR, CVaR and Maximum Loss

For the same value of α we have that CVaR ≥VaR, which was actually the reason why CVaR was "created". CVaR is stricter than VaR. Due to that characteristic; CVaR is able to "secure" our portfolio from potential losses, by not letting it exposed to extreme danger.

When the value of the confidence level α is close to 1, then CVaR coincides with Maximum Loss.

4.3 OPTIMIZATION OF CVaR

CVaR as we have already said is an alternative risk measure with better properties than VaR. In this chapter, we will study CVaR's optimization.

As a tool of optimization, CVaR can be easily treaded due to the fact that it is able to keep all the good properties as a risk measure. When we have to deal with normal or maybe elliptical distributions of loss, then CVaR, Var and even the minimum variance are equivalent and easy to be calculated.

However, those distributions are not very common in everyday life, where we mostly have to deal with fat tailed distributions. Nevertheless, in those cases CVaR can be expressed by an easy minimization formula. Generally, this formula can be used with respect to $x \in X$ to minimize CVaR so as to optimize the portfolio by minimizing the risk.

Those computational advantages of CVaR over VaR have made CVaR more flexible and easier to be used. Until now, there are not formulas or algorithms which can be used for optimizing VaR when we have to deal with high- dimensional instruments. An optimization formula for CVaR has first been introduced by Rockafellar and Uryasev in 2000. In this approach, along with the calculation and the optimization of CVaR, we obtain also the computation of VaR. Furthermore, the minimization of CVaR also leads close to optimal solutions for VaR, so minimizing CVaR is very closely related with minimizing VaR. That is quite logical however, because VaR never exceeds CVaR, so the minimization of CVaR operates as a barrier for VaR. That is the reason why portfolios with low CVaR have low VaR either.

When we optimize a portfolio, we actually solve a stochastic optimization problem, for which a numerous algorithms have been made. These algorithms combine the mathematical features of the portfolios along with simulation-based methods. One of the great contributions of the approach introduced by Rockafellar and Uryasev is that linear programming techniques can be used for optimization of the conditional Value at Risk. That technique is analyzed in their paper: "Conditional Value at Risk: Optimization Algorithms and applications". The linearity has be proven by Rockafellar and Uryasev for distribution which is given by a fixed number of scenarios and the loss function is linear, then the CVaR function can be replaced by a linear function and an additional set of linear constraints.

The problem of optimization of CVaR can be reduced in a linear programming problem, which is easier to be performed. Furthermore, the fact that CVaR can be minimized with the use of linear programming techniques allows the handling of portfolios with large number of instruments and scenarios.

<u>Description of the approach [28]:</u>

First of all, we will notate the quantities we will use as they are used in the paper of Rockafellar and Uryasev “Optimization of Conditional Value at Risk”.

Let **f(x, y)** be the loss associated with the decision vector x. The vector **x∈ X**, with X being a certain subset of $\mathbb{R}^n$ and the random vector y in $\mathbb{R}^m$. The vector **x** can be thought as the representing portfolio, and **X** as the set of all the available portfolios, which are subject to various constraints. The vector **y** represents the uncertainties which affect the current examined portfolio and by extension the loss. In the case where we find a negative loss, we conclude that we have a gain.

For each x, the loss f(x, y) is a random variable which follows a distribution in $\mathbb{R}$, induced by the distribution of y. For our convenience, we assume that y has the density p(y), but we will see later that there is no need an analytical expression of that density for our implementation. We only need a proper algorithm which generates random samples from p(y). To obtain an analytical expression of p(y) or to perform a Monte-Carlo simulation, we follow the next two steps:

(1) We first model the risk factors in $\mathbb{R}^{m_1}$, with $m_1 < m$.

(2) “based on the characteristics of instrument I (i=1,2…,n), the distribution p(y) can be derived or code transforming random samples of risk factors to the random samples from density p(y) can be constructed.”

The probability of f(x, y) is given:

$$\Psi(\mathbf{x},\alpha) = \int_{f(x,y)\leq\alpha} p(y)dy \tag{4.9}$$

The relation (4.9) as a function of α for fixed x is the cumulative distribution for the loss associated with x. The importance of Ψ(x, α) is significant. First of all, it determines the behavior of the random variable plus it has a substantial role in the computation of VaR and CVaR. Furthermore, Ψ(x, α) is a <u>nondecreasing</u> function <u>with respect to α</u> and is <u>continuous from the right.</u> There is no need of the assumption that Ψ(x, α) must be continuous from the left too, because there is a possibility of jumps. However, for our simplicity, we will assume that $\Psi(x, \alpha)$ is everywhere continuous with respect to α, i.e. there is no possibility of jumps. If we do not make that assumption, the only problem is just a complication in mathematical properties, but those complications can be over passed.

We have:

$$\mathrm{VaR}_{\beta}(\mathrm{x}) = \min\{\alpha \in \mathbb{R}: \Psi(x, \alpha) \geq \beta\} \tag{4.10}$$

and:

$$\mathrm{CVaR}_{\beta}(\mathrm{x}) = \frac{1}{1-\beta} \int_{\mathrm{f(x,y)} \geq \mathrm{VaR}_{\beta}(\mathrm{x})} \mathrm{f(x, y)p(y)dy} \tag{4.11}$$

Due to the fact that $\Psi(x, \alpha)$ is nondecreasing and continuous with respect to α, we conclude that $\mathrm{VaR}_{\beta}(\mathrm{x})$ is the <u>left endpoint of the nonempty interval consisting of the values α such that $\Psi(x, \alpha) = \beta$.</u> From the relation (4.11), we have that: $P\left(f(x, y) \geq \mathrm{VaR}_{\beta}(\mathrm{x})\right) = 1$. Thus, $\mathrm{CVaR}_{\beta}(\mathrm{x})$ is actually the conditional expected value of the loss f(x, y) associated with x, related to that the loss is greater or equal to $\mathrm{VaR}_{\beta}(\mathrm{x})$.

As Rockafellar and Uryasev point out, "the key to this approach is the characterization of $\mathrm{VaR}_{\beta}(\mathrm{x})$ and $\mathrm{CVaR}_{\beta}(\mathrm{x})$ in terms of the function F_{β} on $X \times \mathbb{R}$, defined as follows:

$$\mathrm{F}_{\beta}(\mathrm{x}, \alpha) = \alpha + \frac{1}{1-\beta} \int_{y \in \mathbb{R}^m} [f(x, y) - \alpha]^{+} p(y) dy \tag{4.12}$$

where:

$$[\mathrm{t}]^{+} = \begin{cases} \mathrm{t, when\ t} > 0 \\ 0, \mathrm{when\ t} \leq 0 \end{cases}$$

The important characteristic of F_{β} is the property of convexity. The property of convexity ensures that the local minimum is a global minimum as well.

THEOREM 4.1 [28]:

As a function of α, $F_\beta(x,\alpha)$ is convex and continuously differentiable. The β-CVaR of the loss associated with any x∈X can be determined from the formula:

$$CVaR_\beta(x) = \min_{\alpha\in\mathbb{R}} F_\beta(x,\alpha) \tag{4.13}$$

In this formula, the set consisting of the values of α for which the minimum is attained, namely

$$A_\alpha(x) = argmin_{\alpha\in\mathbb{R}} F_\beta(x,\alpha) \tag{4.14}$$

is a nonempty closed bounded interval (perhaps reducing to a single point) and the β-VaR of the loss is given by

$$VaR_\beta(x) = left\ endpoint\ of\ A_\beta(x) \tag{4.15}$$

In particular, one always has

$$VaR_\beta(x) \in argmin_{\alpha\in\mathbb{R}} F_\beta(x,\alpha)\ and \quad CVaR_\beta(x) = F_\beta\left(x, VaR_\beta(x)\right) \tag{4.16}$$

■

To prove Theorem 4.1, we rely on the assumption that Ψ(x,α) is continuous with respect to α, i.e. regardless of the x, if f(x,y)=α

When 1-β is quite small, someone can minimize, instead of the quantity $F_\beta(x,\alpha)$, the quantity (1-β)$F_\beta(x,\alpha)$. This, can be done for computational purposes because there will be no need to divide the integral by (1-β).

The power of theorem 4.1 is that:

(1) The minimization process becomes easier numerically due to the fact that the functions are continuous, differentiable and convex.
(2) There is no need to calculate VaR separately, but the value of VaR derives from the calculation of the endpoint of the function $A_\beta(x)$. Also, if the VaR is not needed, then the part of the calculation of VaR can e omitted.
(3) Additionally to the second point, we can calculate CVaR without first having calculated VaR (do not forget that CVaR depends by definition directly from VaR), which sometimes can be very complicated.

In addition to theorem 4.1, there is theorem 4.2.

The integral of $F_{\beta}(x,\alpha)$ can be approximated in several ways. For example, we can sample from the probability distribution of y according to its density p(y). Then the sampling will generate a collection of vectors $y_1, \dots y_n$ and the approximation of $F_{\beta}(x,\alpha)$ will be:

$$\widetilde{F_{\beta}}(x,\alpha) = \alpha + \frac{1}{q(1-\beta)}\sum_{k=1}^{q}[f(x,y_k) - \alpha]^{+} \tag{4.17}$$

We have the following corollaries:

COROLLARY 4.1 [30]: Convexity of CVaR

If f(x, y) is convex with respect to x, then $CVaR_{\alpha}(x)$ is convex with respect to x as well. Indeed, in this case $F_{\beta}(x,a)$ is jointly convex in (x,α).

Likewise, if f(x, y) is sublinear with respect to x, then $CVaR_{\alpha}(x)$ is sublinear with respect to x. Then too, $F_{\beta}(x,a)$ is jointly sublinear in (x,α).

Proof:

The joint convexity of $F_{\beta}(x,a)$ in (x,α) is an elementary consequence of the definition of this function in (4.12) and the convexity of the function (x,α)⟼$[f(x,y) - \alpha]^{+}$ when f(x, y) s convex in x. The convexity of $CVaR_{\alpha}(x)$ in x follows immediately then from the minimization formula (3.12)[12].

The argument for sublinearity is entirely parallel to the argument just given. Only the additional feature of positive homogeneity needs attention, according to the remark about sublinearity above.

■

A case where the sublinearity which is presented in Corollary 4.1 is obvious is the one where the function f(x, y) is linear with respect to x:

$$f(x,y) = x_1 f_1(y) + \cdots + x_n f_n(y) \tag{4.18}$$

[12] In convex analysis, when a convex function of two vector variables is minimized with respect to one of them, the residual is a convex function of the other.

COROLLARY 4.2 [30]: Coherence of CVaR

On the basis of definition (4.1), α-CVaR is a coherent risk measure: when f(x,y) is linear with respect to x, not only is $CVaR_\alpha(x)$ *sublinear with respect to x, but furthermore it satisfies*

$$CVaR_\alpha(x) = c \; when \; f(x,y) \equiv c \qquad (A)$$

And it obeys the monotonicity rule that

$$CVaR_\alpha(x) \leq CVaR_\alpha(x') \;\; when \; f(x,y) \leq f(x',y) \qquad (B)\;.$$

Proof:

In terms of z=f(x, y) and $z' = f(x,y)$ in the context of the linearity in the relation (3.17), these properties come out easily. The sublinearity of CVaR, in the case of (3.17) has already been noted as ensured in the Corollary 4.1. Like that, the additional properties (A) and (B) too can be seen as simple consequences of the fundamental minimization formula for CVaR_α.

■

PROPOSITION 4.1 [30]: Stability of CVaR

The value $CVaR_\alpha(x)$ *behaves continuously with respect to the choice of α∈(0,1) and even has left and right derivatives, given by*

$$\frac{\partial^-}{\partial\alpha} CVaR_\alpha(x) = \frac{1}{(1-\alpha)^2} E\{[f(x,y) - \alpha^-(x)]^+\}$$

$$\frac{\partial^+}{\partial\alpha} CVaR_\alpha(x) = \frac{1}{(1-\alpha)^2} E\{[f(x,y) - \alpha^+(x)]^+\}$$

■

THEOREM 4.2 [28]:

Minimizing the β-CVaR of the loss associated with x over all x∈ X is equivalent to minimizing $F_\beta(x,\alpha)$ *over all (x,α)∈X×ℝ, in the sense that*

$$\min_{x\in X} CVaR_\beta(x) = \min_{(x,\alpha)\in X\times\mathbb{R}} F_\beta(x,\alpha) \qquad (4.19)$$

where, moreover, a pair (x^*, α^*) achieves the second minimum if and only if x^* achieves the first minimum and $\alpha^* \in A_\beta(x^*)$. In particular, therefore, in circumstances where the interval $A_\beta(x^*)$ reduces to a single point (as is typical), the minimization of F(x, α) over (x,α)∈X×ℝ produces a pair (x^*, α^*), not necessarily unique, such that x^*minimizes the β-CVaR and α^*gives the corresponding β-VaR.
Furthermore, $F_\beta(x, \alpha)$ is convex with respect to (x,α) and $CVaR_\beta(x)$ is convex with respect to x, when f(x, y) is convex with respect to x, in which case, if the constraints are such that X is a convex set, the joint minimization is an instance of convex programming.

■

The power of Theorem 4.2 is the fact that is not necessary to work directly with the CVaR_β. So we eliminate by this some difficulties like the fact that the computation of CVaR_β is directly associated with the value of VaR. However, the value of VaR, as we have already mentioned, can be hard to be computed, a fact that makes difficult the calculation of CVaR. Moreover, there is no need to define a specific x which minimizes the CVaR. Instead, we can use the $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ which is convex with respect to α and even sometimes with respect to (x, α).

A last comment for this approach is that the optimization process of $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ is actually a stochastic approximation problem, because the definition of $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ includes an expectation.

To sum up, the authors in that paper proved that:

1) $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ is convex with respect to α
2) $VaR_a(x)$ is a minimum point of $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ with respect to
3) Minimizing $\mathrm{F}_\beta(\mathrm{x}, \alpha)$ with respect to α gives $\mathrm{CVaR}_\alpha(\mathrm{x}) = \min_\alpha \mathrm{F}_\beta(\mathrm{x}, \alpha)$

4.4 APPLICATION OF THE METHOD [28]

Let $x = (\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n})$ be a portfolio of financial instruments with x_j being the position in instrument j and

$$x_j \geq 0 \; for \; j = 2, \ldots, n \; with \sum_{j=1}^{n} x_j = 1 \tag{4.20}$$

We have the random vector $y = (y_1, \ldots., y_n)$, where y_i is the return on instrument i. The distribution of y has density p(y) and is a joint distribution of the several returns and is independent of x.

The return on a portfolio x is the sum of the returns on the individual instruments in the portfolio weighted by the proportions x_j. Because we have to deal with a quantity with negative meaning, the loss function is negative and is given by:

$$f(x, y) = -[x_1 y_1 + \cdots + x_n y_n] = -x^T y \tag{4.21}$$

The continuity of p(y) with respect to y insures the continuity of the cumulative distribution function for the loss associated with x.

Usually we measure VaR and CVaR in terms of monetary value. However in this application we will measure CVaR and VaR in terms of percentage returns, so as to make the comparison of minimum CVaR and the minimum variance approach consist.

The connection between VaR and CVaR is being made by the relation:

$$F_\beta(x, \alpha) = \alpha + \frac{1}{1-\beta} \int_{y \in \mathbb{R}^n} [-x^T y - \alpha]^+ p(y) dy \tag{4.22}$$

In this point, we must point out that $F_\beta(x, \alpha)$ is a convex function of both x and α. Also, often is differentiable in these two variables. Those two properties make, by mathematical aspect, easier the application of the approach.

For the loss associated with x we have the mean denoted by μ(x) and the variance denoted by σ(x). For the variable y the mean is m and the variance V. We have the relations:

$$\mu(x) = -x^T m \;\text{ and }\; \sigma^2(x) = x^T V x \tag{4.23}$$

From the relation (4.23) it can be seen that the relation between μ and m is linear and the relation between σ and V is quadratic, both functions of x. We have the quantity R, which is the least amount that a portfolio is expected to return. So when we set the constraints we will accept only the portfolios for which stands the inequality:

$$\mu(x) \leq -R \tag{4.24}$$

So we take the portfolios:

$$X = \{\mathrm{x}: \mathrm{x}\ \text{satisfies}\ (4.19)\ \text{and}\ (4.24)\} \tag{4.25}$$

We have that the set X is a convex one, and due to theorem 4.2 the problem of minimizing F_β over X×ℝ can be dealt as a problem of convex programming. Now, if we take a sample from the probability distribution of y, then we have an approximation of F_β like the form of the relation (4.17). So if we have the sample $y_1, \dots, y_q$ then, based on the relation (4.17), we have the approximation:

$$\widetilde{F_\beta}(\mathrm{x}, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{\mathrm{k}=1}^{\mathrm{q}} [-\mathrm{x}^{\mathrm{T}} \mathrm{y}_{\mathrm{k}} - \alpha]^+ \tag{4.26}$$

Our goal in this part of the approach is to get an approximate solution to the minimization of F_β over X×ℝ by the minimization of $\widetilde{F_\beta}$ over X×ℝ. This problem can be easily reduced to convex programming. Furthermore, despite the fact that the expression (4.26) is not differentiable with respect to α, it can "be minimized either by line search techniques or by representation in terms of an elementary linear programming problem" as Rockafellar and Uryasev point out in their paper "Optimization of Conditional Value at Risk" . In order to express this step with real variables like w_k with k=1,…r for example, we have the minimization of the linear expression:

$$\alpha + \frac{1}{(1-\beta)q} \sum_{\mathrm{k}=1}^{\mathrm{r}} \mathrm{w}_{\mathrm{k}}$$

Subject to the linear constraints:

1) $\mathrm{x}_{\mathrm{j}} \geq 0\ for\ j = 2, \dots, n\ with\ \sum_{j=1}^{n} x_j = 1$

2) $\mu(x) \leq -R$ and

3) $\mathrm{w}_{\mathrm{k}} \geq 0$ and $\mathrm{x}^{\mathrm{T}} \mathrm{y}_{\mathrm{k}} + \alpha + w_k \geq 0\ for\ k = 1, \dots., r$

A useful note in that point is that the above reduction to linear programming does not depend on a special assumption about the distribution of y, i.e. we do not assume a normal distribution, but it works for a nonnormal distribution too.

Since this part, we have not mentioned at all a way of finding a portfolio that minimizes β-VaR, i.e.:

$$P(1) \qquad \text{minimize} VaR_{\beta}(x) \text{ over } x \in X$$

Since now our approach was based on Theorem 2 and on finding a way to minimize F_{β} and by extension β-CVaR, i.e.:

$$P(2) \qquad \text{minimize} CVaR_{\beta}(x) \text{over } x \in X$$

However, when we find a way of minimizing CVaR and due to the fact that $CVaR_{\beta}(x) \geq VaR_{\beta}(x)$, the solutions of the problem 2 can be good for problem 1 as well. In other words, CVaR, due to the fact that measures the losses exceeding the value of VaR, acts as an upper bound for VaR.

Additionally, we can compare the problems P (1) and P (2) with a third problem, which is very popular because it minimizes maybe the simplest measure of the three: the variance. So we have the problem:

$$P(3) \qquad \text{minimize} \sigma^2(x) \ \text{ over } x \in X$$

The attractiveness of P (3) is that it reduces to a quadratic problem. However, it is not, like P (1), very suitable for all kind of portfolios.

Both P (1) and P (2) negotiate to find the optimal portfolio which is going to minimize both VaR and CVaR. We present proposition 3.1:

PROPOSITION 4.2 [28]:

Suppose that the loss associated with each x is normally distributed, as holds when y is normally distributed. If β≥0.5 and the constraint (4.23) is active at solutions to any two of the problems P (1), P(2) and P(3), then the solutions *to those two problems are the same; a common portfolio x^* is optimal by both criteria.*

■

Using Proposition 4.2 we have the opportunity to use quadratic programming solutions as the base of testing if the reduction of (4.25) into a linear programming problem is appropriate.

To conclude, we showed a way of minimizing the CVaR of a portfolio and by extension, despite the fact it is not clear by our approach, to lower VaR which stands due to the inequality $\mathrm{CVaR}_\beta(\mathrm{x}) \geq \mathrm{VaR}_\beta(\mathrm{x})$.

4.5 COMPARATIVE ANALUSIS OF VaR AND CVaR

VaR is without a doubt a risk measure which is very popular and, as Rockafellar and Uryasev pointed out in their paper "Conditional Value at Risk for General Loss distributions" [29], "VaR has achieves the high status of being written into industry regulations". However, VaR has some drawbacks, which the use of CVaR instead can be overcome.

CVaR has superior mathematical properties than VaR as we have already mentioned, as it follows all the necessary, to be a coherent risk measure, properties defined by Artzner. Plus, from those four properties we conclude that CVaR is also a convex measure, making the optimization procedure much easier that VaR's.

Furthermore, VaR, by definition, cannot control scenarios which exceed the VaR level, and that is a huge disadvantage of this measure. VaR is not able to distinguish the differences among situations where there is a large possibility of losses and situations where the possibility of losses is astounding. On the other hand, CVaR does account the losses which exceed the VaR levels, providing the analysts with better and more accurate results.

The CVaR was created to estimate the tails of the distribution of loss, so as an estimator of potential losses, can provide the analysts with results more stable and most of the times more accurate than VaR. That is the reason why risk management using CVaR can be done very efficient. The CVaR is a stable statistical estimator, but it must not in no way be compared with VaR. It is a huge mistake to compare the results of CVaR and VaR assuming that they are compared with the same value of confidence level α. The problem of that comparison lays in the fact that VaR and CVaR measure different parts of the distribution.

However, there is a possibility that VaR and CVaR to be equal. Actually, if we have two confidence levels α_1 *and* α_2, then the comparison of VaR and CVaR can be found from the equation:

$$\mathrm{VaR}_{\alpha 1} = \mathrm{CVaR}_{\alpha 2}$$

In the paper "Risk Return Optimization with different Risk Aggregation Strategies" written by Serraino, Theiler and Uryasev, is illustrated a credit risk example, in which they find that CVaR with confidence level α=0.95 is equal to VaR with confidence level α=0.99.

Furthermore, Yamai and Yoshiba, in their paper "Comparative Analysis of Expected Shortfall and Value at Risk, Their Estimation, Error, Decomposition and Optimization" [33], they examine the estimations of Var and CVaR of the parametrical family of stable distributions.

"When a random variable X follows the stable distribution, there exist α and γ_n such that:

$$\mathrm{S_n} \overset{\mathrm{d}}{=} \mathrm{n}^{1/\alpha}\mathrm{X} + \gamma_{\mathrm{n}}$$

$\mathrm{S_n}$ is the sum of independently and identically distributed n copies of X and α is the index of stability. The smaller α is, the heavier the tail of the distribution[13].

They ran 10.000 simulations of size 1.000 and they compared the standard deviations of VaR and CVaR estimators. The conclusions were enlightening. VaR estimators were more stable than the ones of CVaR with the same confidence level. However, when we have to deal with fat-tailed distributions the differences are obvious compared with distributions which are closed to normal. Additionally, as the sample size increases at the same time increases the accuracy of CVaR.

The explanation for the difference between fat tailed distributions and normal distributions is simple enough. When we have to deal with fat tailed

[13] When α=2 we have the normal distribution and when α=1 we have the Cauchy distribution.

distributions, the possibility of an infrequent and large loss is high; so the CVaR is affected since it considers the right tail of the distribution. Contrary, VaR is not affected of that kind of loss since VaR disregards the loss beyond the VaR level. That is the reason why CVaR estimations errors in fat tailed distributions are larger than the estimation errors of VaR.

In the same paper, Yamai and Yoshiba [33] examined the influence of the sample size in the estimation errors of CVaR. They ran 10.000 sets of Monte Carlo simulation with sample sizes of 1.000, 10.000 and 100.000, under the assumption that the underlying loss distributions are stable with α= 2.0, 1.5, 1.1. As the sample size increases, the relative standard deviations of the CVaR were reducing. So they conclude that a way of reducing the estimation errors of CVaR is the increase of the sample size.

Another difference between VaR and CVaR is the fact that CVaR can be easier decomposed according to risk factors than VaR. Say that X_i ae the loss of individual risk factors and w_i are the sensitivities to the risk factors with i=1,….,n. We then have the portfolio:

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

Also, we have the following decompositions:

$$VaR_\alpha(X) = \sum_{i=1}^{n} \frac{\partial VaR_\alpha(X)}{\partial w_i} w_i = E[\{X_i | X = VaR_\alpha(X)\}] w_i$$

And

$$CVaR_\alpha(X) = \sum_{i=1}^{n} \frac{\partial CVaR_\alpha(X)}{\partial w_i} w_i = E[\{X_i | X \geq VaR_\alpha(X)\}] w_i$$

It is easier to calculate quantities $E[\{X_i | X \geq VaR_\alpha(X)\}]$ in the CVaR decomposition than quantities $E[\{X_i | X = VaR_\alpha(X)\}]$ in the VaR decomposition. That is due to the fact that the estimators of the fraction $\frac{\partial VaR_\alpha(X)}{\partial w_i}$ are not as stable as the estimators of the fraction $\frac{\partial CVaR_\alpha(X)}{\partial w_i}$.

To sum up we have:

Value at Risk:

Advantages: VaR is a simple risk measure, which measures *"how much you may lose with a certain confidence level"*. That information is summarized in a single number, so it is really easy to be understood. Furthermore, using VaR is easy to compare two different distributions in the same confidence level. If we calculate VaR for all the confidence levels, we will have a full image of the distribution we study. Moreover, due to the fact that VaR focuses on a specific part of the loss distribution is more stable comparing to the standard deviation, because it is not affected by high tail losses, when we study a fat tailed distribution.

Disadvantages: The fact that VaR disregards what happens beyond the VaR level, it may lead to an unintentional high risk. So, if someone wants to overcome this possibility, he/she must calculate several VaRs with different confidence levels. Furthermore, VaR may increase a lot with a slight change of α. Also, when we have to deal with distributions that present skewness, VaR calculations may lead to wrong results. Last but not least, VaR is a nonconvex measure at risk and is discontinuous function for discrete distributions. Those facts make the problem of optimization of VaR very complicated. However, there are codes that overcome these problems, like PSG [1].

Conditional Value at Risk:

Advantages: The implementation of CVaR is clear: what is the possibility for a worst case scenario to be held. For example, if $\bar{L}$ is the worst potential losses and L is the loss, then the restriction $\mathrm{CVaR}_{\alpha}(\mathrm{L}) \leq \bar{\mathrm{L}}$ ensures that the average of (1-α) % highest losses will not exceed $\bar{L}$. Furthermore, just like VaR, CVaR for all the confidence levels α specifies fully the distribution, so CVaR is also a better measure than standard deviation. Moreover, CVaR is a convex risk measure and is continuous with respect to α. So the optimization process is not as difficult as with VaR. Additionally, in the case where the loss distribution is normal, the CVaR optimization approach coincides with mean-variance approach, making our work very easy. However, that does not mean that optimization of CVaR without the normality is a difficult issue. On the contrary, is easy to control and to achieve the optimization of CVaR for non- normal loss distributions. Also, CVaR can be used for linear programming, even when the number of the instruments is big. That is a big difference between VaR and CVaR, where VaR may crack when the number of instruments is big.

Disadvantages: The accuracy of CVaR is totally dependent of the assumption of the model we have already done. If the model we assume is not good for the tail of the distribution (which is the part that CVaR calculates), then the results will be misleading. For example, the historical simulation method does not give us enough information about the tail of the distribution. Thus, we must assume a distribution to fit in our tails. If the assumption is a little far from the truth, then the results will be extremely wrong.

Despite the advantages and the disadvantages of VaR and CVaR, the question which of those two risk measures is more appropriate to be used providing better results still remains. Probably the answer is related to who actually uses the risk measure. A trader for example will probably prefer to use VaR for various reasons. First of all, VaR is not as restricted as CVaR for the same confidence level. Yet, VaR is easier to be calculated. In case of extreme losses, the trader will not pay the price of his option because his s just an employee. On the other hand, the board of the company will prefer to use CVaR as a risk measure. The results of CVaR will cover the case of large losses, so it is better as risk measure for them.

Additionally, as we already mentioned before, a bad assumption in the distribution of losses or the ignoring of a possible fat tail distribution will not "damage" VaR, since VaR disregards the tails. That does not happen in the case of CVaR. CVaR may not perform well in that case, providing the analysts with wrong results. However, if the assumption of the model is correct, then without a doubt CVaR will give extremely accurate estimations, so it is better to be used.

To sum up, the choice as to which of the two measures, VaR or CVaR we will use is neither easy nor straightforward. The choice of the suitable risk measure is being affected by the differences in the mathematical properties, the simplicity of the optimization processes of those two measures, the stability of the providing estimators and finally by the acceptance of the regulators.

CHAPTER 5

WAYS OF COMPUTING VaR AND CVaR

INTRODUCTION

As we analyzed in the second chapter, there are several different ways of computing Value at Risk. In this chapter, first of all we will present ways of computing VaR and CVaR for linear and nonlinear portfolios and we will compute Value at Risk and Conditional Value at Risk for data which follow either Normal Distribution or t- Student distribution.

5.1 Computation of Value at Risk

As we have already introduced in chapter 2, Value at Risk can be expressed as followed:

$$Prob\left[\Delta\widetilde{\mathrm{P}}(\Delta t, \Delta\tilde{x}) > -VaR\right] = 1 - \alpha \tag{5.1}$$

as we have said in (2.1b). The definition of VaR has an extremely ease interpretation: the value of the tested portfolio will not decline more than VaR α% of the time, over a specific number of trading days. The time horizon can be a day, a week or more and is an exogenous factor to our model. However it must be determined from the start and plays a crucial role to our calculation and to our interpretation. Furthermore, the value of the chosen confidence level is often between 1% and 10%. The confidence level is a subjective parameter determined by the analyst and only by him. This is the only subjective factor in our model. The choice of α depends o the risk tolerance of management.

A major role for the computation of VaR has the knowledge or not of the distribution of the returns. If the returns are normally distributed (or log-normally distributed) then the task is ease. However, difficulties are present when the returns follow a different distribution than normal, which may present kurtosis are skewness. Furthermore, another factor that is very important is the linearity or not of the portfolio. If the portfolio is linear there are straightforward ways to compute everything. Otherwise, we follow the rout of historical or Monte-Carlo simulation along with specific methodologies to calculate VaR and CVaR.

5.1.1 Computation of VaR for a linear portfolio with elliptic distribution [30]

We are going to present the computation of Value at Risk assuming that the pricing function of the portfolio is linear in the risk factors. This method is parametric and is capable of providing fast answers as long as the linearity holds.

We will use the following notations:

The row vectors: $x=(x_1, \dots, x_n)$ and

$y=(y_1, \dots, y_n)$

The matrices: $A=(A_{ij})_{i,j}$ and

B

Furthermore we have the Euclidean inner product $x * x = x * x^t = x_1^2 + \cdots + x_n^2$

We will call a portfolio linear for a specific time t and a value Π (t) if its profit and loss function $\Delta\Pi(t)=\Pi(t)-\Pi(0)$ over a time window [0,t] is a linear function of the returns $X_1(t), \dots, X_n(t)$, i.e.

$$\Delta\Pi(t) = \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_n X_n \tag{5.2}$$

The (5.2) holds not only if we use percentage returns, but also hold to good approximation with log-returns, which are commonly used in everyday market, as long as the time window is small.

We will keep the time t fixed, so we will leave it out from our notation for simplicity. So, we will introduce now

$$X=(X_1,\ldots,X_n)$$

so that $\Delta\Pi=\delta X=\delta X^t$.

We assume that X_j are elliptically distributed with mean μ and a correlation matrix $\Sigma=AA^t$, i.e. $(X_1,\ldots,X_n)\sim N(\mu,\Sigma,\varphi)$. This leads us to the following pdf:

$$f_X(x) = |\Sigma|^{-\frac{1}{2}} g\big((x-\mu)\Sigma^{-1}(x-\mu)^t\big) \tag{5.3}$$

Where $g:\mathbb{R}_{>0} \to 0$ is such that the Fourier transform of $g(|x|)^2)$ as a generalization function of $\mathbb{R}^n$, is equal to $\varphi(|\xi|^2)^{-1}$. Now, we will assume that g is a non zero everywhere and a continuous function, the Value at Risk for a confidence level 1-α is:

$$\mathrm{Prob}\{\Delta\Pi(t) < -\mathrm{VaR}_\alpha\} = \alpha \tag{5.4}$$

Since our parameters are elliptic distributed, we must solve the equation

$$\alpha = |\Sigma|^{-1/2} \int_{\{\delta x \le -VaR_\alpha\}} g((x-\mu)\Sigma^{-1}(x-\mu^t))dx \tag{5.5}$$

We change variables to $y = (x-\mu)A^{-1}$, $dy=|A|dx$, where we use the Cholesky decomposition of A and we have $\Sigma= A^tA$, the (4.4) now becomes:

$$\alpha = \int_{\{\delta Ay \le -\delta\mu - VaR_\alpha\}} g(|y|^2)dy \tag{5.6}$$

Let R be a notation which sends $\delta A \to (|\delta A|, 0, \ldots, 0)$. If we change variables one more time to $y=zR$, we have the equation:

$$\alpha = \int_{\{|\delta A| z_1 \leq -\delta\mu - VaR_\alpha\}} \mathrm{g}(|\mathrm{z}|^2)\mathrm{dz} \tag{5.7}$$

If $|z|^2 = z_1^2 + |z'|^2$ with $z' \in \mathbb{R}^{n-1}$ then we have shown that:

$$\alpha = \mathrm{Prob}\{\delta X < -\mathrm{VaR}_\alpha\} = \int_{\mathbb{R}^{n-1}} [\int_{-\infty}^{\frac{-\delta\mu - VaR_\alpha}{|\delta A|}} g(z_1^2 + |z'|^2) dz_1] dz' \tag{5.8}$$

By using spherical variables, we have the next transformations:

$$z' = r\xi, with\ \xi \in S_{n-2}$$

$$dz' = r^{n-2} s\sigma(\xi) dr$$

we see that for solving VaR, we have to solve the equation[14]:

$$\alpha = |S_{n-2}| \int_0^{+\infty} r^{n-2} \left[\int_{-\infty}^{\frac{-\delta\mu^t - VaR_\alpha}{|\delta A|}} g(z_1^2 + r^2) dz_1 \right] dr \tag{5.9}$$

We introduce the function[15]:

$$G(s) = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{-s} \left[\int_0^{+\infty} r^{n-2} g(z_1^2 + r^2) dr \right] dz_1$$

$$= \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \int_s^{+\infty} \left[\int_{z_1^2}^{+\infty} (u - z_1^2)^{\frac{n-2}{2}} g(u) du \right] dz_1$$

We have proved the following result:

Theorem 5.1 [30]***:*** *Suppose that the portfolio's Profit and Loss function over the time window of interest is, to good approximation, given by*

---

[14] The $|S_{n-2}|$ is the surface measure of the unit-sphere in $\mathbb{R}^{n-1}$, i.e. $|S_{n-2}| = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}$

[15] In the second line we have made the transformation $u = r^2 + z_1^2$ and we have replaced z_1 with $-z_1$.

$$\Delta\Pi = \delta_1 X_1 + \cdots + \delta_n X_n \ ,$$

with constant portfolio weights δ_j. Suppose moreover that the random vector $X = (X_1, \dots, X_n)$ of underlying risk factors follows a continuous elliptic distribution, with probability density given by

$$f_X(x) = |\Sigma|^{-\frac{1}{2}} \mathrm{g}((x-\mu)\Sigma^{-1}(\mathrm{x}-\mu)^{\mathrm{t}})$$

where μ is the vector mean and Σ is the variance- covariance matrix, and where we suppose that $g(s^2)$ is integrable over $\mathbb{R}$, continuous and nowhere 0. Then the portfolio's Delta-elliptic VaR_αat confidence level 1-α is given by

$$\boldsymbol{VaR_\alpha = -\delta\mu + q^g_{\alpha,n}\sqrt{\delta\Sigma\delta^t}}$$

where s=$q^g_{\alpha,n}$ is the unique positive solution of the transcendental equation α=G(s).

■

Remarks:

1. The financial interpretation of |δA| is simply the portfolio's volatility with |δA|=$\sqrt{\delta\Sigma\delta^t}$.
2. If there is a Short-term Risk management, one can easily assume that μ≅0. Specifically in that case, we have

   $$\mathrm{VaR}_\alpha = \sqrt{\delta\Sigma\delta^t} \cdot \mathrm{q}^{\mathrm{g}}_{\alpha,n}$$

   The above result is analogous with the results found in the case of linear portfolios with the risk factors to follow the Normal distribution. The only difference is, for example in the case where α=0.05 the normal 5% quantile is 1.65, but now is replaced by the g-dependent constant $\mathrm{q}^{\mathrm{g}}_{0.05}$.
3. For the case of t- Student distribution we have the following theorem:

   Theorem 5.2 [30]: *Assuming that $\Delta\Pi \cong \delta_1 X_1 + \cdots + \delta_n X_n$ with a multivariate Student-t random vector $(X_1, \dots, X_n)$ with vector mean μ, and variance- covariance matrix Σ, the linear Value at Risk at confidence level 1-α is given by the following formula:*

   $$\boldsymbol{VaR_\alpha = -\delta \cdot \mu + q^t_{\alpha,v}\sqrt{\delta\Sigma\delta^t}}$$

where now $s = q^t_{\alpha,\nu}$ *is the unique positive solution of the transcendental equation* $G^t_\nu(s) = \alpha$.[16]

■

In the case of Student-t distribution $q^t_{\alpha,\nu}$ *does not depend on n.*

5.1.2 Computation of CVaR for a linear portfolio with elliptic distribution [30]

In the third chapter we introduced an alternative risk measure, the Conditional Value at Risk which has better properties than the most popular risk measure, Value at Risk. CVaR actually describes how large losses are on average when they exceed the VaR level. CVaR is a subadditive risk statistic which is stricter than VaR and which indicates the size of the extreme losses when we have VaR as threshold. Despite the fact that CVaR has more attractive properties than VaR, has not still become a standard risk tool in the financial area, though, it is gaining in the insurance industry.

Since the definition of CVaR is strongly related to VaR, the computation of CVaR leads to close optimal solutions for VaR too. Also, we must not forget that when the distribution of returns is normal then CVaR and VaR those two measures are equivalent, which means that they give the same optimum portfolio. However, when we have to deal with skewed distributions the optimal portfolios which take from VaR and CVaR can and will differ a lot. Rockafellar and Uryasev in 2000 in their paper “Optimization for Conditional Value at Risk” [28] proved that linear programming techniques can be used for optimization of the CVaR.

In this section, we will represent a way of computing CVaR for a linear portfolio assuming that the risk factors follow an elliptical distribution. From (5.1), we have that:

$$CVaR = \mathbb{E}(-\Delta\Pi \mid -\Delta\Pi > VaR) \tag{5.10}$$

[16] We have: $C(v,n) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma(n/2)\sqrt{(\nu\pi)^n}}$ with v>2 and $G^t_\nu(s) = \frac{1}{\nu\sqrt{\pi}}\left(\frac{\nu}{s^2}\right)^{\nu/2}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\,{}_2F_1\left(\frac{1+\nu}{2},\frac{\nu}{2};1+\frac{\nu}{2};-\frac{\nu}{s^2}\right)$

Due to the fact that we are dealing with a continuous distribution, we have that Conditional Value at Risk coincides with Expected Shortfall (ES). Assuming that the risk factors are elliptically distributed, we have the density as t is in (5.3). So, the CVaR, or else the ES, fat the confidence level 1-α is given by:

$$\begin{aligned}
-\mathrm{CVaR}_{\alpha} = -\mathrm{ES}_{\alpha} &= \mathbb{E}(\mathit{\Delta\Pi} \mid \mathit{\Delta\Pi} > VaR_{\alpha}) = \\
&= \frac{1}{\alpha}\mathbb{E}\big(\mathit{\Delta\Pi}\cdot 1_{\{\mathit{\Delta\Pi}\leq -VaR_{\alpha}\}}\big) \\
&= \frac{1}{a}\int_{\{\delta x^{t}\leq -VaR_{a}\}} \delta x^{t} f(x)dx \\
&= \frac{|\Sigma|^{-1/2}}{\alpha}\int_{\{\delta x^{t}\leq -VaR_{a}\}} \delta x^{t} g((x-\mu)\Sigma^{-1}(x-\mu)^{t})dx
\end{aligned}$$

We will leave $\Sigma = A^{t}A$ and we will do the same linear transformations that we did in the previous section, when we wanted to calculate Value at Risk. Doing all that, we have the following result:

$$\begin{aligned}
-\mathrm{CVaR}_{\alpha} = -\mathrm{ES}_{\alpha} &= \frac{1}{\alpha}\int_{\{|\delta A|z_1\leq -\delta\mu - VaR_{a}\}} (|\delta A|z_1 + \delta\mu)g(\|z\|^2)dz \\
&= \frac{1}{\alpha}\int_{\{|\delta A|z_1\leq -\delta\mu - VaR_{a}\}} (|\delta A|z_1)\cdot g(\|z\|^2)dz + \delta\mu
\end{aligned}$$

If we make exactly the same transformations as before, i.e. the transformation $\|z\|^2 = z_1^2 + \|z'\|^2$ and by using again spherical coordinates

$$z' = r\xi, with\ \xi \in S_{n-2}$$

we have:

$$-\mathrm{CVaR}_{a} = \delta\mu + \frac{|S_{n-2}|}{\alpha}\int_{0}^{\infty} r^{n-2}\left[\int_{-\infty}^{\frac{-\delta\mu^{t}-VaR_{\alpha}}{|\delta A|}} |\delta A| z_1 g(z_1^2 + r^2)dz_1\right]dr \qquad (5.11)$$

Now, we first do the change z_1 to $-z_1$ and then we introduce the quantity $u = r^2 + z_1^2$, as we did before. From theorem (5.1) combined with the remark 1, we have that:

$$q_{\alpha,n}^{g} = \frac{\delta \cdot \mu + VaR_\alpha}{|\delta A|}$$

We denote $q_{\alpha,n}^{g} = q_\alpha$ and we have:

$$\mathrm{CVaR_a} = -\delta\mu + |\delta A| \frac{|S_{n-2}|}{\alpha} \int_{q_\alpha} \int_{z_1^2}^{\infty} z_1 (u - z_1^2)^{\frac{n-3}{2}} g(u) du dz_1$$

$$= -\delta\mu + |\delta A| \frac{|S_{n-2}|}{\alpha} \int_{q_\alpha^2}^{\infty} \frac{1}{n-1} (u - q_\alpha^2)^{\frac{n-1}{2}} g(u) dud \qquad (5.12)$$

After using the functional equation for the Γ-function Γ(x+1)=xΓ(x) and the formula for $|S_{n-2}|$, we have the next theorem:

Theorem 5.3* [1]*: *Suppose that the portfolio is linear in the risk factors X=($X_1, \ldots, X_2$), ΔΠ=δ·X and that X~N(μ,Σ,φ), with pdf*

$$f_X(x) = |\Sigma|^{-\frac{1}{2}} \mathrm{g}((x-\mu)\Sigma^{-1}(\mathrm{x}-\mu)^{\mathrm{t}})$$

If we write $q_{\alpha,n}^{g} = \frac{\delta \cdot \mu + VaR_\alpha}{(\delta \Sigma \delta)^{1/2}}$*, then the Expected Shortfall at level α is given as:*

$$\boldsymbol{ES_\alpha = -\delta\mu + K_{ES}^{g} |\delta \Sigma \delta^t|^{1/2}} \qquad (5.13)$$

where the constant

$$K_{ES}^{g} = \frac{\pi^{\frac{n-1}{2}}}{\alpha \Gamma(\frac{n+1}{2})} \cdot \int_{(q_{\alpha,n}^{g})^2}^{\infty} \left(u - (q_{\alpha,n}^{g})^2\right)^{\frac{n-1}{2}} g(u) du$$

■

Let us see now the application of the above methodology when we have a multivariate t-Student distribution. In that case, we have that

$g(u) = C(\nu, n)(1 + \frac{u}{\nu})^{-\frac{(\nu+n)}{2}}$, where C(ν+n) is the same with the one that we introduced in the previous section. For our simplicity, we will write q instead of $q_{\alpha,n}^{g}$. We can evaluate the integral of the (5.13) and we have[17]:

$$\int_{q^2}^{\infty} (u - q)^{\frac{n-1}{2}} \left(1 + \frac{u}{\nu}\right)^{-\frac{n+\nu}{2}} du = \nu^{\frac{n+\nu}{2}} (q^2 + \nu)^{-\frac{\nu-1}{2}} B(\frac{\nu-1}{2}, \frac{n+1}{2})$$

By doing some computation, which is irrelevant for our study, we find the following result:

Theorem 5.4 ***[30]:*** *The Expected Shortfall at confidence level 1-α for a multivariate Student distributed linear portfolio δ·X, with*

$$X \sim \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \cdot \sqrt{|\Sigma|(\nu\pi)^n}} \left(1 + \frac{(x-\mu)\Sigma^{-1}(x-\mu)}{\nu}\right)^{-\frac{n+\nu}{2}}$$

Is given by:

$$ES_{\alpha,\nu}^{t} = -\delta\mu + |\delta\Sigma\delta^t|^{\frac{1}{2}} \cdot \frac{1}{a\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \nu^{\frac{\nu}{2}} ((q_{\alpha,\nu}^{t})^2 + \nu)^{-\frac{\nu+1}{2}} =$$

$$= -\delta\mu + |\delta\Sigma\delta^t|^{\frac{1}{2}} \cdot \frac{1}{a\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \nu^{\frac{\nu}{2}} \left(\left(\frac{\delta \cdot \mu + VaR_\alpha}{(\delta\Sigma\delta)^{\frac{1}{2}}}\right)^2 + \nu\right)^{-\frac{\nu+1}{2}}$$

$$= -\delta\mu + es_{a,\nu}|\delta\Sigma\delta^t|^{1/2}.$$

■

[17] To do the evaluation of the integral we use the following lemma: "If $\left|\arg\left(\frac{u}{\beta}\right)\right| < \pi$ and Re(ν_1)>Re(μ)>0, then $\int_{q^2}^{\infty} (u - q)^{\frac{n-1}{2}} \left(1 + \frac{u}{\nu}\right)^{-\frac{n+\nu}{2}} du = \nu^{\frac{n+\nu}{2}} (q^2 + \nu)^{-\frac{\nu-1}{2}} B(\frac{\nu-1}{2}, \frac{n+1}{2})$ Where B is the Euler Beta function with $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

5.1.3 Computation of VaR for non linear portfolios [3]

In the previous section, we described a way of computing Value at Risk for linear portfolios. Unfortunately, in the real life not all portfolios are linear. On the contrary, we often have to deal with nonlinearities, especially when we refer to option position, which present nonlinearities caused by the portfolio's payoff structure[18]. In those cases, Value at Risk cannot be calculated using the risk factor distribution. Instead, the risk factor distribution must first be converted into a profit and loss distribution for the portfolio and then, Value at Risk is computed through the profit and loss distribution.

There have been proposed several ways to overpass the problem of nonlinearity for computing Value at Risk. There are the parametric models, such as the *delta-normal*, which, for computing VaR they use statistical parameters such the mean and standard deviation of the risk factor distribution. Using the delta of the position and these parameters, they calculate directly from the risk factor distribution the VaR[19]. Another way is the *Delta- Gamma Value at Risk technique* which is quite simple. Briefly, that technique has the following steps:

1. We collect all the information about our portfolio, like the types of instruments in the portfolio and all the data we need.
2. We calculate the covariance matrix of the risk factors: to do so, we evaluate the returns of the time series data. Then, we normalize the time series. Furthermore, we calculate the correlation matrix and then the covariance matrix.
3. We use finite difference method to calculate the Delta Matrix of first derivatives of portfolio value with respect to the risk factors.
4. We use finite difference method to calculate the Hessian Matrix of second and cross derivatives of Portfolio Value with respect to the risk factors.
5. We perform a Cholesky decomposition of the covariance matrix. This step is useful for achieving to a lower triangular matrix.

[18] Generally, a non linear portfolio can be caused by options, flexibility or other price dependent elements in the portfolio.
[19] The delta is actually used as an approximation for the conversion from the risk factor distribution to the profit-loss distribution.

6. We calculate the adjusted delta matrix to arrive at zero Eigen and non-zero Eigen vectors.
7. We calculate the cumulative distribution with the help of the above matrices.
8. Finally, we take the 1-α percentile cut off from the inverse cdf.

Non parametric models are simulation or historical models. The simulation approaches can be separated in two categories: *the full valuation* and *the partial valuation*. In the full simulation, we create a number of possible scenarios for the risk factors and in each scenario we perform a complete revaluation of the portfolio. By that, we get the profit and loss distribution of the portfolio. The difference in the partial simulation is that we indeed use simulations to create the distribution of the risk factors, but we do not fully revalue the portfolio. Instead of revaluating the portfolio, in this method, we use delta gamma approximations to calculate the portfolio value. The next figure shows graphically the different ways to calculate models.

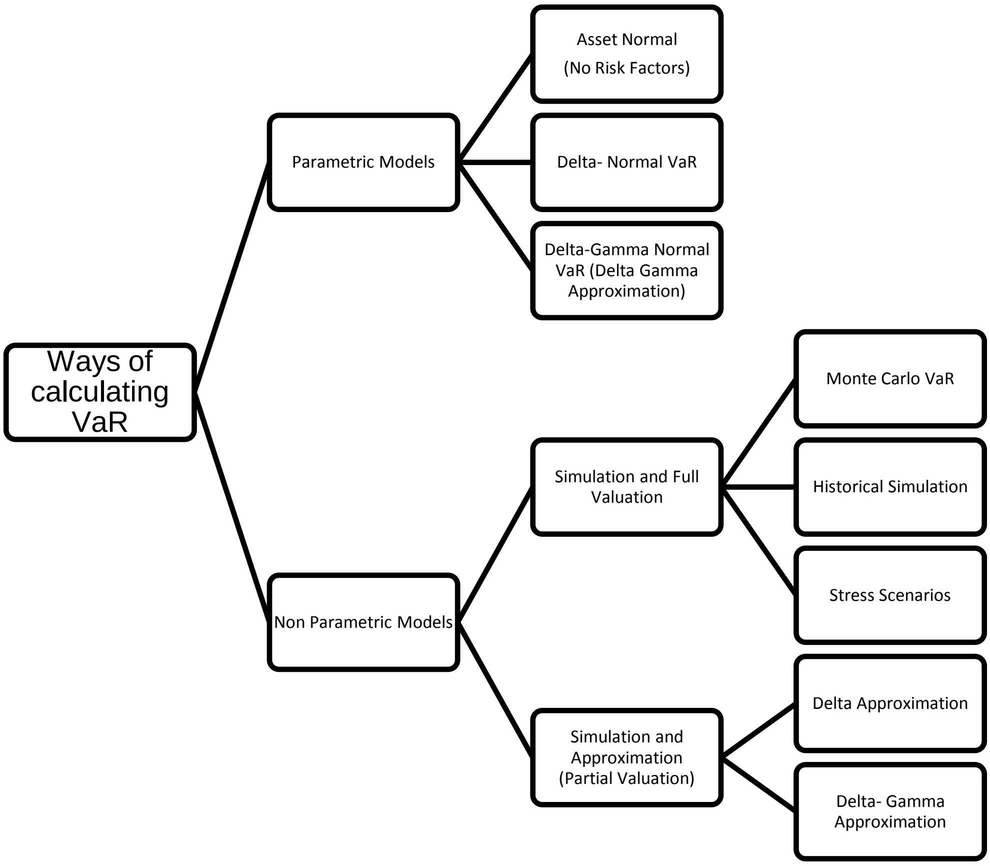

Figure 5.1: Ways of calculating Value at Risk for non linear portfolios

The most common way of dealing with portfolios which have nonlinear returns, or portfolios with non normal distributed assets, is the Monte Carlo methods. However, despite the fact that Monte Carlo methodology has the advantage of being universally applicable, has also a big drawback: is much slower and resource consuming compared with parametric methods when the latter are available. Additionally, the accuracy of the computation using Monte Carlo is usually limited to order $1/\sqrt{n}$, where n is the number of the performed trials.

5.1.4 Asset- Normal and Delta Normal VaR

The basic way of calculating VaR is the Asset-Normal VaR. To use this method, we assume normality for the respective values of the position. Then, the VaR is:

$$\mathrm{VaR}_{\alpha}(\mathrm{t},\mathrm{T}) = \mathrm{z}_{\alpha}\sqrt{\mathrm{w}'\Sigma w} \cdot \sqrt{\mathrm{T}-\mathrm{t}} \cdot PF(t)$$

where Σ is the variance- covariance matrix, w is the vector of the portfolio weights, z_{α} is the α- quantile derived from the standard normal distribution, (T-t) is the time window on which we do are calculations and PF is the real value of the portfolio.

Now, if we want to reduce the dimension of the problem, we use the *Delta-Normal method*. With this method, "the model is based on a risk factor representation of the individual positions. In other words, every position in the portfolio is modeled exclusively on the basis of market risk factors." This method has the following assumptions:

1. There is a linear dependent relationship between the changes of in the value of the portfolio and the respective changes in the value of the risk factors. Due to the fact that a Taylor approximation of first order of the change of the value of the portfolio is used, the computation of VaR when we have to deal with linear instruments (such as forward contracts) is exact. However, when we have to deal with non linear instruments (such as options) the calculation is VaR is a local approximation.

2. The distribution of the changes in the value of the risk factors assumed to be the joint normal distribution.
3. Last but not least, the composition of the portfolio is thought to be constant over the time.

Then, the VaR is given by:

$$\mathrm{VaR}_{\alpha}(\mathrm{t},\mathrm{T}) = \mathrm{z}_{\alpha}\sqrt{\mathrm{D}'\Sigma D} \cdot \sqrt{\mathrm{T}-\mathrm{t}}$$

with Σ is the variance covariance matrix of the market risk factors, i.e.

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \sigma_{\mathrm{M},1} & \cdots & \sigma_{\mathrm{M},\mathrm{M}} \end{pmatrix}$$

and D the (M×1) matrix of modified portfolio sensitivities relative to the changes of the market risk factors, i.e.

$$\mathrm{D} = \begin{pmatrix} d_1 \\ \vdots \\ d_{\mathrm{M}} \end{pmatrix} = \begin{pmatrix} \delta_1 \cdot \mathrm{S}_1(\mathrm{t}) \\ \vdots \\ \delta_{\mathrm{M}} \cdot \mathrm{S}_{\mathrm{M}}(\mathrm{t}) \end{pmatrix}$$

with δ_{i} the i-th portfolio sensitivity and $\mathrm{S}_{\mathrm{i}}(\mathrm{t})$ the value of the ith market risk factor at the time t.

However, we have to point out, that this model has some drawbacks which are:

- It does not take into account the presence of a leptokurtic or a skewed distribution.
- It assumes the knowledge of the distribution of the market risk factors and the fact that the structure of the variance covariance matrix is deterministic.
- In the definition of the VaR, the use of the $\sqrt{\mathrm{T}-\mathrm{t}}$ lies in the assumption that the volatility of the changes in the risk factors is constant and that there is no correlation among them, facts that are often violated in practice.

Despite the drawbacks, this method is ease to interpret and to understanding, has low cost and is fast enough. In the case of leptokurtic risk factors' distribution, where the assumption of normality does no longer exists, Albanese, Levin and Chao suggested a method which uses a stochastic variance-covariance matrix than a deterministic one with known probability distribution and unknown parameters. The parameters are been estimated with a Bayesian approach.

5.1.5 <u>Monte-Carlo and Partial Monte-Carlo VaR</u>

In the Monte- Carlo method what we do is to simulate the values of the changes of the market risk factors and then revaluating the entire portfolio for each simulated trial. As it is inevitable, there are assumptions in that method too.

First of all, we assume a Geometric Brownian motion for the stock price process. To calculate S(t) we take the stochastic differential equation:

$$\frac{\mathrm{dS}}{\mathrm{S}} = \mu(t)dt + \sigma(t)dz$$

with $dz = \varepsilon_{\mathrm{i}}\sqrt{\mathrm{dt}}$ to be the infinitesimal increment of the Brownian motion (or Weiner process) and ε_{i} a standard normal variable. If we assume a time T, the solution of the above equation for the asset price is given by:

$$S_{\mathrm{j}}(\mathrm{T}) = S_{\mathrm{j}}(0) \cdot \exp\left[\left(\mu_{\mathrm{j}} - \frac{\sigma_{\mathrm{j}}^{2}}{2}\right) \cdot \mathrm{T} + \sigma_{\mathrm{j}} \cdot \varepsilon_{\mathrm{i}}\sqrt{\mathrm{T}}\right]$$

In the above equation, we denote with μ_{j} the annual expected value of the relative change, σ_{j} the annual standard deviation of the relative change and T the time horizon in years.

In that point, we must point out that the use of the Brownian motion takes for granted that the distribution of the price of the underlying is lognormal and the continuously compounded return follows the normal distribution. Furthermore, due to the short horizon in which we do our calculations for VaR we assume that the expected change in the price of a market risk factor over the time period is zero. This assumption is reasonable enough if we consider that the expected change in the price of a market risk is small compared with the volatility of the respective changes.

Since we have specified the use of a stochastic process we can, now, generate future scenarios for the different market risk factors. However, we must first generate random normal variables which are going to be correlated between them. These correlations are given from the estimated variance-covariance matrix. To generate those variables we make use of the Cholesky factorization. We follow the next steps:

1. We estimate the historical correlation matrix.
2. We decompose the correlation matrix to get the Cholesky matrix, which is a lower triangular matrix.
3. We create a vector which includes uncorrelated random normal variables.
4. Finally, we multiply the vector from the third step with the lower triangular matrix and we get a vector which contains normal random variables which are correlated according to the estimated correlation matrix.
5. For each trial we revaluate the single positions and the whole portfolio.

Then, the VaR we want to calculate is the α-quantile of the change of the portfolio. As expected, the higher the number of trials, the better the estimation we get.

Monte Carlo method has plenty of advantages. First of all, we are able to assume an arbitrary process for the underlying asset. Furthermore, is easy enough to generate different correlated scenarios. When we make Monte Carlo simulation with fully valuation, the portfolio is revalued for each simulation trial, giving the analysts more accurate results for portfolios with option components, which are difficult to handle otherwise. Also, if the number of the risk factors is high, then it might be difficult to generate variables since the matrix Σ needs first to be Cholesky factorized. However, despite of the advantages of this method, there is a huge disadvantage: Monte-Carlo simulation combined with full valuation needs high technical requirements and the calculations can be slow because the portfolio in every trial is been revaluated.

A way to overpass the problem of low calculations is to use Partial Monte-Carlo VaR, which is slightly different from the regular Monte- Carlo VaR. However, the drawback of this method is that "*it does not improve approximations over parametric models if it applies the same approximation methods and assumes the same processes for the market risk factors*."

Last but not least, we would like to point out that when we have to deal with only one risk factor, then using the Monte- Carlo simulation is simple enough. However, when we have to deal with options, where there are more factors that must be acknowledged such as volatility risk or the interest rate risk, then this method may be complicated. Additionally, the assumption that the risk factors are normally distributed is, as we have already pointed out before, not always real. To be more realistic, someone could run Monte-Carlo simulation using the third or the fourth moments.

However, sometimes is even better to use the *Delta-Gamma approximation.* With this method, we first take a Taylor series approximation in the returns for the value of each of the assets in the portfolio. These component approximations are then summed over all assets in the portfolio. By this, we obtain the Taylor approximation for the overall portfolio. Due to the fact that these components are quit small, we keep only the linear and the quadratic terms which often give us sufficient enough results. The linear terms are called deltas and the quadratic ones gammas. That is why the second-order approximation is called a *delta-gamma approximation.* Even with this method though, the Monte-Carlo approach can present computational difficulties.

5.2 Efficient frontier [21]

Efficient frontier was first introduced as a definition by Markowitz in 1952 in his paper "Portfolio Selection". By efficient frontier we can, considering a universe of risky investments, explore what may be the optimal portfolio based on those investments.

In our case, if we want to work with Conditional Value at Risk, if we consider the minimum of the expected returns we can minimize the value of CVaR. Actually, we consider different expected returns and by that we generate an efficient frontier. Another way of working is to maximize the expected returns along with the fact that we do not allow risk. To be more precise, when the optimal portfolio is defined in one of two ways:

1. If we consider that all the portfolios have the same volatility level, we say that the optimal portfolio is the one with the highest return.

2. If we consider that all the portfolios have the same expected return, we say that the optimal portfolio is the one with the lowest volatility.

So, if we follow the first way of efficient frontier we produce an optimal portfolio for each possible level of risk. On the other hand, if we follow the second way, we produce an optimal portfolio for each expected return. However, we must point out that these two ways of producing an optimal portfolio are equivalent because the set of optimal portfolios produced by the one definition or the other is exactly the same. That precise set of optimal portfolios is called *efficient frontier*.

There are three formulations for the optimization problem with the use of CVaR. These formulations are equivalent because they produce the same efficient frontier. We have the following theorem:

***Theorem 5.5* [21]:** *Let us consider the functions φ(x) and R(x) dependent on the decision vector x, and the following three problems:*

$(P1)\qquad \min_x \varphi(x) - \mu_1 R(x), \qquad x \in X, \qquad \mu_1 \geq 0$

$(P2)\qquad \min_x \varphi(x), \qquad R(x) \geq 0, \qquad x \in X$

$(P3)\qquad \min_x -R(x), \qquad \varphi(x) \leq \omega \qquad x \in X$

Suppose that constraints R(x)≥ρ, φ(x)≤ω have interval points. Varying the parameters μ_1*, ρ and ω traces the efficient frontiers for the problems (P1)-(P3), accordingly. If φ(x) is convex, R(x) is concave and the set **X** is convex, then the three problems, (P1)-(P3), generate the same efficient frontier.*

■

The equivalence we mentioned before is known as "mean-variance" and "mean-regret" efficient frontiers. It is actually holds for every concave and convex risk function, as the Conditional Value at Risk is, with convex constraints.

5.3 Example:

In that section, we will illustrate all the above theory in an example. We assume that we have three assets, which construct a portfolio x=(x_1, x_2, x_3) of financial instruments with x_i being the position in instrument i. We must not forget that the sum of the weights must be 1, i.e. $x_1 + x_2 + x_3$=1. Furthermore, we have that $x_3 = 1 - x_1 - x_2$.

We have the returns r_i and we denote with r the random vector of the returns y_i, i.e. r=(r_1, r_2, r_3). We take two cases: first that the distribution of y is normal and second that the distribution of y is t-Student.

The return of the portfolio x is, as we have defined in (3.20), the following:

$$f(x, y) = -[x_1 r_1 + x_2 r_2 + x_3 r_3]$$

Value at Risk and Conditional Value at Risk will be calculated with Matlab programming.

We have the following 3D-plots:

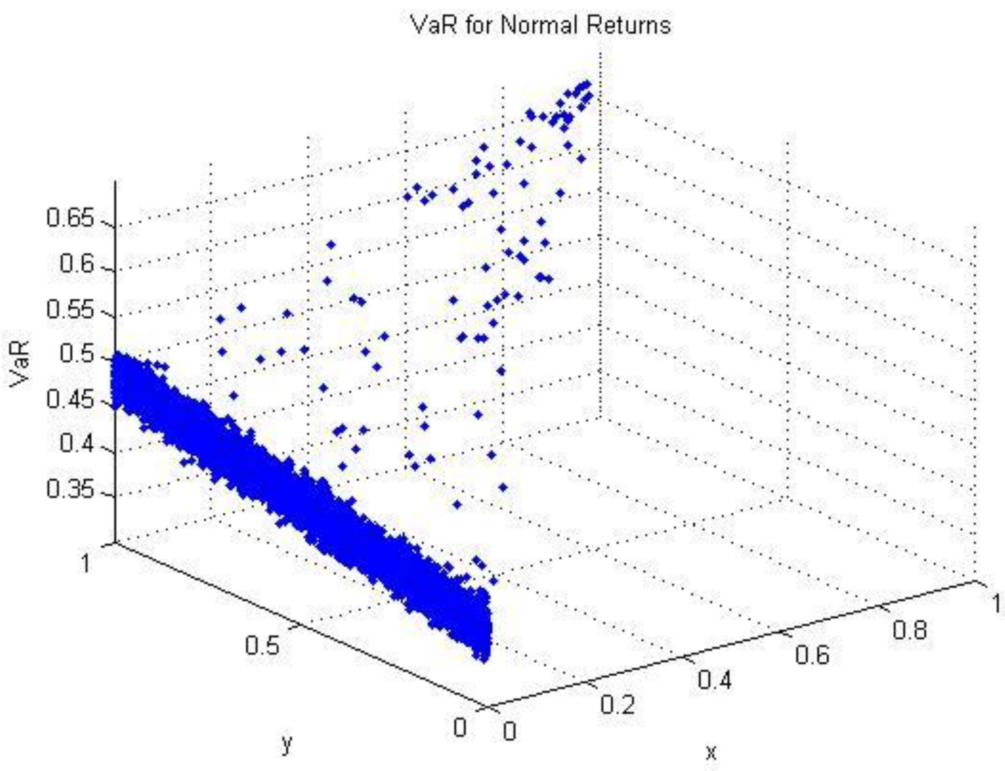

Figure 5.2

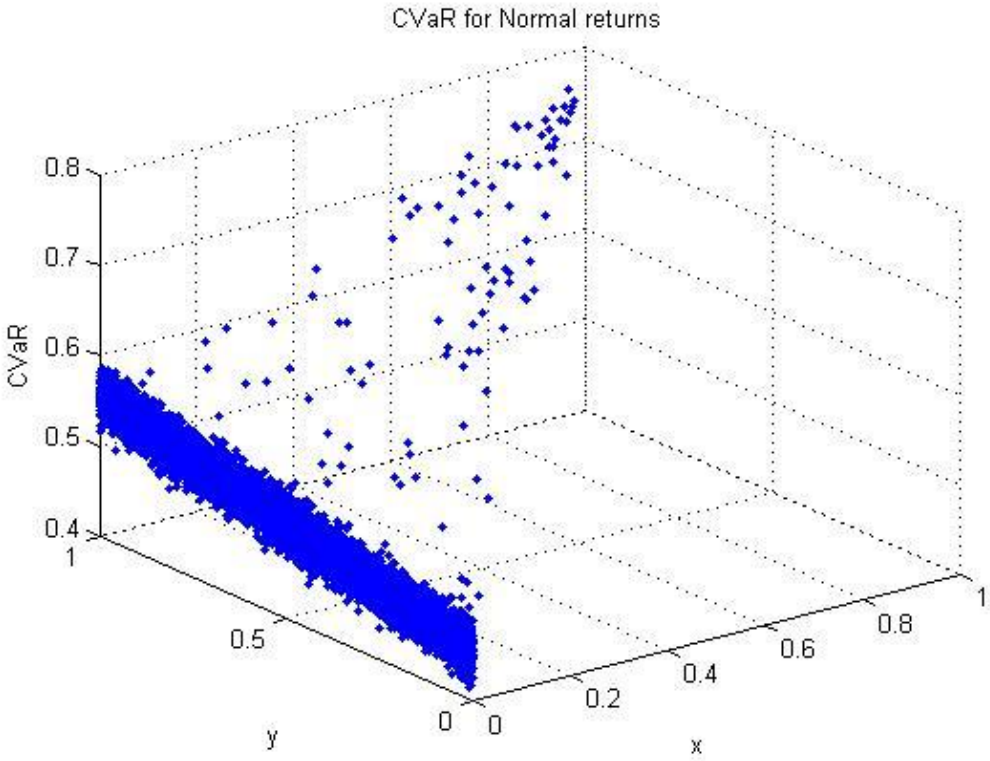

Figure 5.3

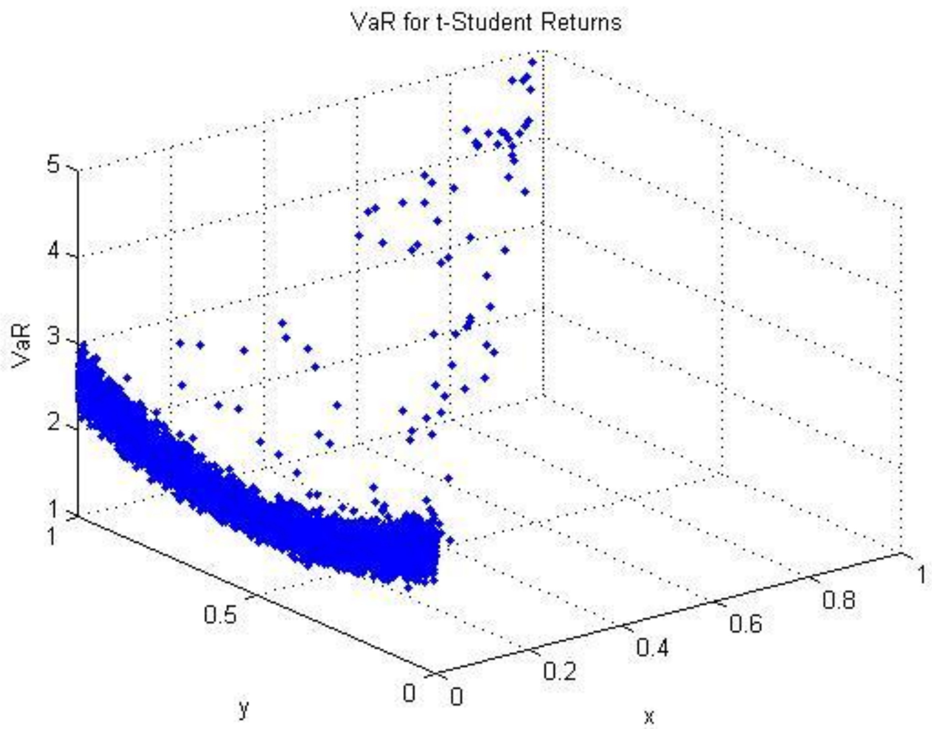

Figure 5.4

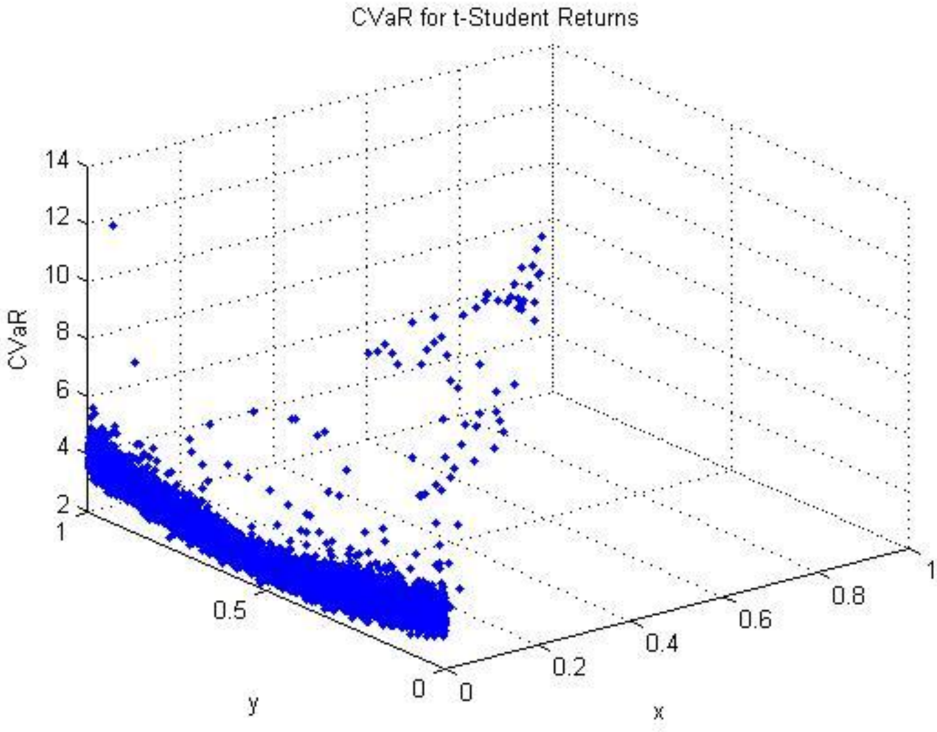

Figure 5.5

We can also plot the expected return of the portfolio with the Value at Risk:

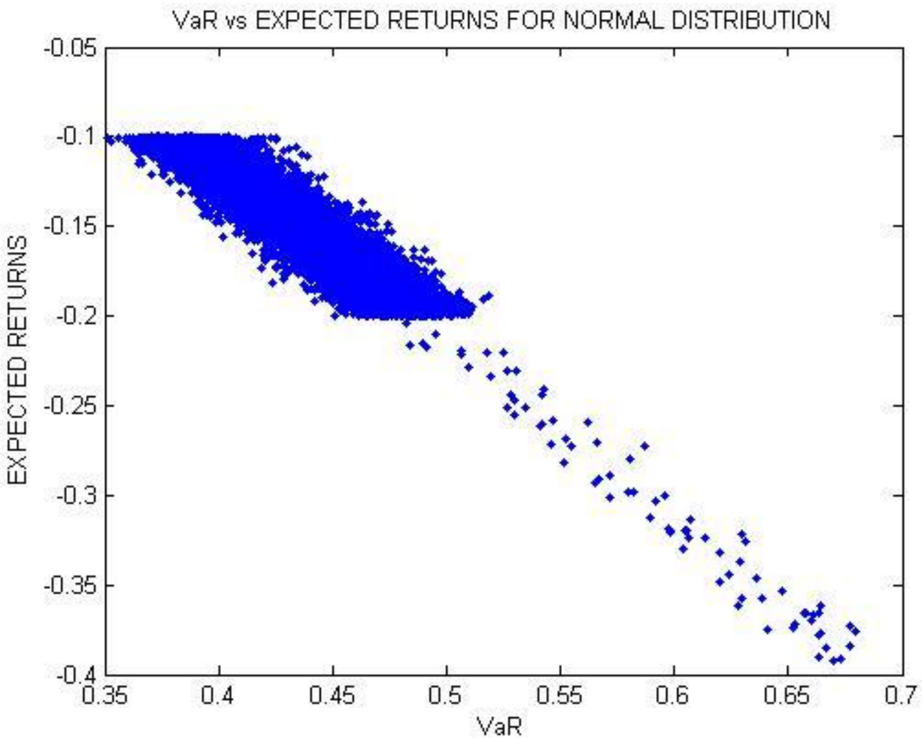

Figure 5.6

We can see a linear relationship between the expected returns of the portfolio and the Value at Risk.

We can also plot the expected returns with the Conditional Value at Risk:

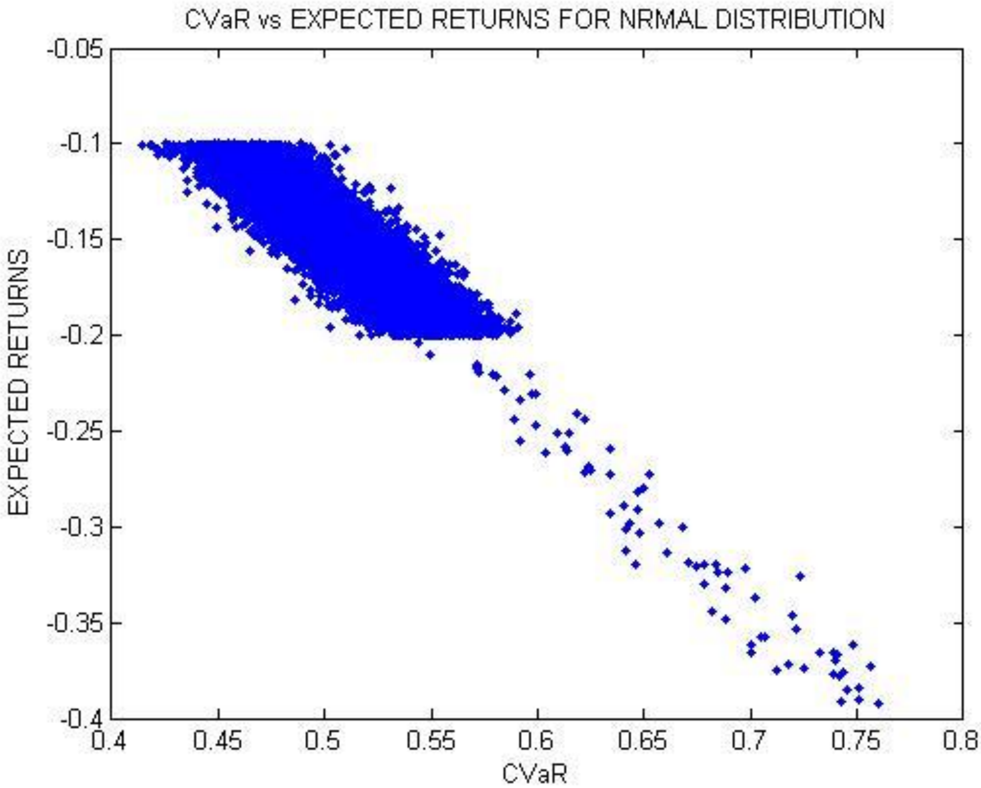

Figure 5.7

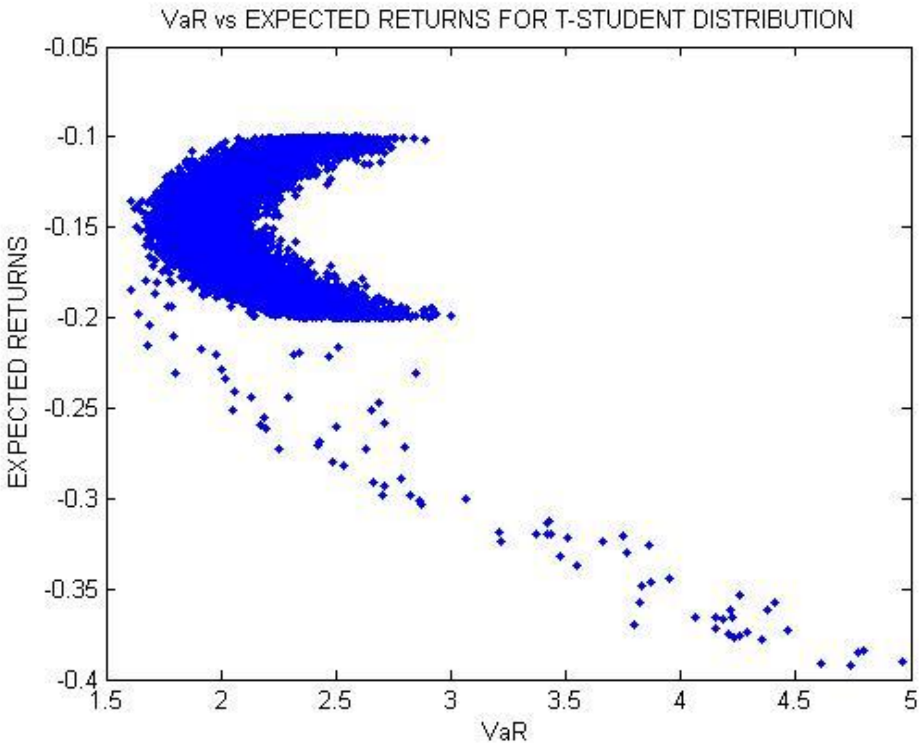

Figure 5.8

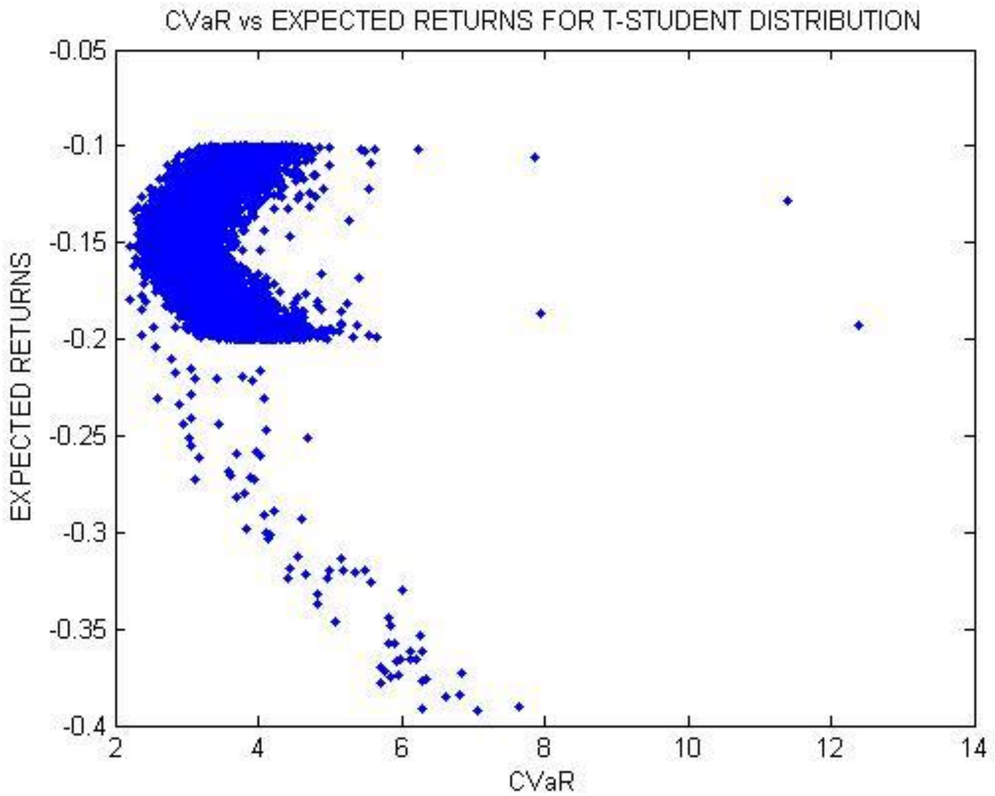

Figure 5.9

References

[1]: **Acerbi and Tasche (2002)**:"On the coherence of Expected Shortfall"

[2]: **Artzner et al. (1998)** "Coherent Measures of Risk", Carnegie Mellon University, Pittsburgh, Pennsylvania

[3]: **Ammann and Reich**: "Value at Risk for Nonlinear Financial Instruments, Linear Approximation of full Monte-Carlo?"

[4]: **Barone Adesi Giovanni and Kostas Gianopoulos (2000)** "Non Parametric VaR Techniques. Myths and Realities

[5]: **Benninga Simon and Zvi Wiener**: "Value-at-Risk (VaR)"

[6]: **Boudoukh, Richardson, Whitelaw**: "The Best of Both Worlds", *Risk, v11, 64-67*

[7]: **Cabedo and Moya (2003)** "Estimating oil price "Value at Risk" using the historical simulation approach", *energy Economics*

[8]: **Danielson et al. (2005)** "Subadditivity Re- examined: The case for Value at Risk"

[9]: **Danielson et al. (2010)** "Fat Tails, VaR and Subadditivity" *(first version of the paper in 2005 was called:"* Subadditivity Re- examined: The case for Value at Risk*")*

[10]: **Dowd and Blake** "After VaR: The Theory, Estimation and Insurance Applications of Quantile- Based Risk Measures"

[11]: **Embrechts, McNeil, Straumann** "Correlation and Dependence in Risk Management: Properties and Pitfalls"

[12]: **Engle Robert** "Garch 101: The use of ARCH/GARCH models in Applied Econometrics."

[13]: **Fama (1965)** "The Behavior of Stock Market Prices", *The Journal of Business, Vol 38, No.1*

[14]: **Frain (2008)** "Value at Risk and the a-stable distribution", *Trinity Economics papers*

[15]: **Garcia, Renault and Tsafack (2003)** "Proper Conditioning for Coherent VaR in Portfolio management"

[16]: **Glasserman, Heiderberger and Shahabuddin (2000)** "Efficient Monte Carlo Methods for VaR",*working paper, Columbia University.*

[17]: **Hull and White (1997)** "Value at Risk when daily changes in market variables are not normally distributed"

[18]: **Hull and White** "Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk", Journal of Risk, v1 5-19

[19]: **Jamshidian, Farshid and Yu Zhu** "Scenario Simulation: Theory and Methodology", *Finance Stochastics, vol. 1, p. 43-67*

[20]: **Krokhmal, Uryasev and Zrazhevsky** "Numerical Comparison of CVaR and CDaR Approaches: Application to Hedge Funds"

[21]: **Krokhmal, Palmquist and Uryasev** " Portfolio Optimization with Conditional Value at Risk, Objective and Constraints"

[22]: **Landsman and Valdez** Tail Conditional Expectations for Elliptical Distribution

[23]: **Landsman and Tsanakas (2004)** Elliptical Distributions and Stochastic Orders

[24]: **Linsmeier J. Thomas and Neil D. Pearson (1996)** "Risk measurement: An introduction to Vale at Risk"

[25]: **Mandelbrot (1963)** "The Variation of Certain Speculative Prices", *The Journal of Business Vol. 36, Issue 4*

[26]: **J. McNeil Alexander, Rudiger Frey and Paul Embrechts** " Quantitative risk Management"

[27]: **Pflug**: "Some remarks on Value at Risk and Conditional Value at Risk"

[28]: **Rockafellar and Uryasev (2000)** "Optimization of conditional Value at Risk"

[29]: **Rockafellar and Uryasev "**Conditional Value at Risk for general loss distributions", *Journal of Banking and Finance*

[30]: **Sadefo Kamdem** "Value at Risk and Expected Shortfall for Linear Portfolios with Elliptically distributed Risk Factors"

[31]: **Sarykalin Sergey, Gaia Serraino and Stan Uryasev (2008)** " Value at Risk vs Conditional Value at Risk in Risk Management and Optimization",

[32]: **Zvi Wiener (1997)** "Introduction to VaR, Risk management and regulation in Banking"

[33]: **Yamai and Yoshiba (2002)**: "Comparative Analyses of Expected Shortfall and Value at Risk: Their Estimation, Error, Decomposition and Optimization" *Monetary and Economic Studies*

[34]: www.stern.nyu.edu/~adamodar/pdfiles/papers/**VAR**.pdf