A Polynomial Algorithm for a NP-hard to Solve Optimization Problem

Dissertation der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
Stefan Eberle
aus Kaufbeuren

eingreicht am 17. Oktober 2008

1. Gutachter: Prof. Dr. W. Richert

2. Gutachter: Prof. Dr. A. Schenzle

Tag der mündlichen Prüfung: 29. Januar 2009

Abstract

Since Markowitz in 1952 described an efficient and practical way of finding the optimal portfolio allocation in the normal distributed case, a lot of progress in several directions has been made. The main objective of this thesis is to replace the original risk measure of the Markowitz setting by a more suitable one, Value-at-Risk. In adressing the optimal allocation problem in a slightly more general setting, thereby still allowing for a large number of different asset classes, an efficient algorithm is developed for finding the exact solution in the case of specially distributed losses. Applying this algorithm to even more general loss distributions results in a not necessarily exact matching of the VaR optimum. However, in this case, upper bounds for the euclidean distance between the exact optimum and the output of the proposed algorithm are given. An investigation of these upper bounds shows, that in general the algorithm results in quite good approximations to the VaR optimum. Finally, an application of a stochastic branch & bound algorithm to the current problem is discussed.

Contents

Danksagung

An dieser Stelle möchte ich Herrn Prof. Dr. Richert ganz herzlich für die Unterstützung bei meiner Dissertation danken. Ein weiterer Dank gilt allen Freunden und Kollegen bei der Münchener Rück, welche meine Arbeit durch anregende Diskussionen unterstützten. Der finanziellen Unterstützung durch die Münchener Rück in Form eines Promotionsstipendiums sei ebenfalls gedankt.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit der Lösung einer sehr allgemeinen Problemstellung, bei der für n gegebene Zufallsvariablen

$$X_1, ..., X_n \in \mathbb{R}$$

die gewichtete "Mischung"

$$X(x) := x_1 \cdot X_1 + ... + x_n \cdot X_n \tag{0.1}$$

für $x_1, ..., x_n \in \mathbb{R}$ näher untersucht werden soll. Hierbei wird unterstellt, dass die Abhängigkeitsstruktur der Zufallsvariablen $X_1, ..., X_n$ als bekannt angenommen werden kann und als explizite Szenarienvektoren

$$X_i = \begin{pmatrix} \widehat{X}_i^1 \\ \vdots \\ \widehat{X}_i^k \end{pmatrix}, \widehat{X}_i^j \in \mathbb{R}$$

der Länge k vorliegen. In dieser Schreibweise ist die Abhängigkeitsstruktur durch die Ausprägungen von $\widehat{X}_i^j$ für $i = 1, ..., n$ implizit gegeben[1]. Weiter sollen die Gewichte x_i für $i = 1, ..., n$ unter Einhaltung einer festgelegten erwarteten Ausprägung $\mathbb{E}(X(x))$ so gewählt werden, dass $X(x)$ eine möglichst geringe Varianz aufweist.

Darüber hinaus wurde der Fall untersucht, bei welchem einseitige positive Abweichungen vom Erwartungswert keiner Beschränkung unterliegen, während das Unterschreiten eines vorgegebenen Schwellenwertes als Abweichung von der Erwartung mit möglichst hoher Konfidenz ausgeschlossen werden kann.

[1] Durch die Verwendung einer geeigneten Anzahl an Szenarien k können über diesen Ansatz auch als stetig vorausgesetzte Zufallsvariable hinreichend approximiert werden.

Mit anderen Worten soll die optimierte Mischung der X_i für zulässige Werte x_i bestimmt werden, so dass zu gegebener Konfidenz das Unterschreiten eines Schwellenwertes durch die Zufallsvariable $X(x)$ ausgeschlossen werden kann.

Auch wenn der oben beschriebene allgemeine Fall an einem konkreten Beispiel erarbeitet wird, so ist hervorzuheben, dass an keiner Stelle der Analysen die Allgemeinheit der Aussage beschränkende Annahmen einfließen. Wird also das vorliegende Problem an einem konkreten Praxisbeispiel erläutert, so dient dies einzig und allein der Anschaulichkeit. Unbenommen hiervon handelt es sich bei dem vorgestellten Algorithmus um eine sehr allgemeine "Quantils-Optimierung" einer beliebigen Mischung von Zufallsvariablen, wobei der Erwartungswert der untersuchten Verteilung als bekannt vorausgesetzt werden darf. Bei dem beschriebenen Lösungsalgorithmus handelt es sich um eine Vorgehensweise, welche ohne weiteren Anpassungsbedarf auf allgemeine mathematisch-statistische Problemstellungen angewendet werden kann.

Wie bereits erwähnt wurde, wird die oben beschriebene allgemeine Problemstellung anhand einer konkreten Problemstellung eingehend analysiert. In der Tat beschäftigen wir uns im vorliegenden Fall mit einer Erweiterung des klassischen Portfoliooptimierungsproblems, welches erstmals von Markowitz in seiner wegweisenden Arbeit [34] im Detail untersucht wurde. Die bereits beschriebenen Zufallsvariablen X_i werden in diesem Zusammenhang als Verlustverteilungen einzelner Assetklassen interpretiert und in der Markowitzschen Analyse als normalverteilt unterstellt. Basierend auf dieser Annahme wird das optimale Portfolio als eine Positionierung zwischen Risiko und erwartetem Verlust formuliert. Über den ursprünglich von Markowitz gewählten Ansatz hinaus wollen wir uns jedoch nicht auf den Fall normalverteilter Zufallsvariablen beschränken, sondern die Verlustverteilungen, gegeben als die Zufallsvariablen $X_i, i = 1, ..., n$ mit größtmöglichem Freiheitsgrad wählbar erlauben. Interpretieren wir $x \in \mathbb{R}^n$ als Portfolioallokation der n verschiedenen Assetklassen, so ergibt sich die entsprechende Verlustverteilung $X(x)$ mittels Gleichung 0.1.

Eine sehr verbreitete Vorgehensweise zur Risikomessung bei nicht normalverteilten Verlustverteilungen stellt die Verwendung des Value-at-Risks zum Konfidenzniveau η (VaR_η oder kurz VaR) dar. Dieses Risikomaß mißt den kleinsten zu erwartenden Verlust, so dass die Wahrscheinlichkeit eines den VaR übersteigenden Verlust nicht höher als $(1-\eta) \cdot 100\%$ ist. Obwohl dieses sehr bedeutende Risikomaß weite Verbreitung gefunden hat, fehlen ihm einige sehr wichtige Eigenschaften, welche Risikomaße im Allgemeinen aufweisen sollten (vgl. [4]). Neben der Tatsache, dass es sich beim VaR um ein im Allgemeinen äußerst instabiles und in der Optimierung als komplex zu klas-

sifizierendes Risikomaß handelt, wird die Ausprägung der den VaR übersteigenden Verluste nicht berücksichtigt. In diesem Sinne wird bei gleicher Wahrscheinlichkeit des den VaR übersteigenden Verlusts die Situation, in welcher eine Konzentration von Verlusten den VaR deutlich übersteigt nicht unterschieden von einer (oftmals favorisierten) Verlustverteilung, bei welcher die Verluste geglättet über ein weites Spektrum unterschiedlicher Verlustausprägungen auftreten.

Es gibt eine Vielzahl an Bemühungen, Optimierungsprobleme der vorliegenden Art effizient zu lösen. Ein Grund dafür, dass es noch immer keinen Algorithmus zur effizienten Lösung großdimensionierter Problemstellungen gibt, ist hauptsächlich in der fehlenden Subadditivitätseigenschaft des VaR begründet. Eine direkte Konsequenz ist im Allgemeinen das Auftreten zahlreicher lokaler Optima, welche die Verwendung klassischer Lösungsansätze von vornherein ausschließen. In der Tat können Yang et al. in ([62]) zeigen, dass die Komplexität des Optimierungsproblems mit dem VaR als Zielfunktion in die Klasse NP-schwieriger Problemstellungen fällt. Die Behauptung von Yang et al. wurde erneut aufgegriffen und in verallgemeinertem Kontext mit neuer Beweisführung nachvollzogen. Aufgrund der Komplexität des Problems gibt es in der Forschung derzeit verschiedene Richtungen, um das Problem fehlender Subadditivität zu umgehen.

In einem ersten Schritt kann eine deutliche Reduktion der Komplexität durch die Beschränkung auf bestimmte Verlustverteilungen erzielt werden. Neben den bereits erwähnten, normalverteilten Verlustverteilungen zählen auch elliptische Verteilungsannahmen zu jenen Verteilungen, für welche das korrespondierende Allokationsproblem mittels geeigneter Algorithmen effizient gelöst werden kann. Darüber hinaus konnte in dieser Arbeit erstmals gezeigt werden, dass unter der Annahme α-stabil verteilter Verluste und bestimmter (wenig restriktiver) Beschränkungen an das zu untersuchende Konfidenzniveau η das ursprüngliche VaR Optimierungsproblem ebenfalls deutlich in seiner Komplexität reduziert werden kann.

Will man auf die Annahme möglichst allgemeiner Verlustverteilungen nicht verzichten, so kann eine Komplexitätsreduktion durch die Verwendung angepasster und in der Regel subadditiver Risikomaße erzielt werden. Neben dem "Smooth Value-at-Risk" ($SVaR$) oder dem "Worst Conditional Expectation" ist an dieser Stelle vor allem der "Conditional Value-at-Risk" ($CVaR$) für diese Dissertation von zentraler Bedeutung. Dass es sich bei letztgenanntem Risikomaß um ein kohärentes (und damit subadditives) Risikomaß handelt, konnte von den Autoren Uryasev und Rockafellar ([52], [53]) für eine weite Klasse von Verlustverteilungen gezeigt werden. In der vorliegen-

den Dissertation konnte dieses Resultat dazu verwendet werden, um unter wenig beschränkenden Annahmen an die Verlustverteilungen und für zahlreiche Konfidenzniveaus η stets ein affines $CVaR$-Konfidenzniveau anzugeben, so dass die korrespondierende Lösung des $CVaR$ Optimierungsproblems eine oftmals ausreichende Approximation des ursprünglichen VaR Problems darstellt. Die Subadditivität der affinen Problemstellung erlaubt es mit Hilfe von Optimierungsverfahren zur Lösung hochdimensionaler linearer Problemstellungen auch das ursprüngliche Optimierungsproblem für eine äußerst große Anzahl an Freiheitsgraden x_i effizient zu lösen. Eine Beschreibung aller hierfür notwendigen Annahmen als auch Aussagen über die Güte der erzielten Approximation stellt den zentralen Bestandteil dieser Arbeit dar.

Sollen einerseits möglichst allgemeine Verlustverteilungen abgebildet werden und andererseits die Verwendung des VaR als Zielfunktion aufrecht erhalten bleiben, so existieren auch hier unterschiedliche Ansätze, welche im Allgemeinen allerdings nur für eine sehr begrenzte Zahl an zu optimierenden Freiheitsgraden x_i effiziente Anwendung finden. Eine hierunter fallende Klasse von Lösungsalgorithmen mit herausragender Bedeutung stellen heuristische Algorithmen (Tabu Search, Simulated Annealing, Threshold Accepting, etc.) dar. Nach dem Abschluss der Analyse zu vorgeschlagener approximativer Vorgehensweise wird ein speziell auf die Gegebenheiten der VaR Optimierung angepasster stochastischer Branch & Bound Algorithmus vorgestellt und dessen Performance anhand eines Beispiels numerisch untersucht. Als ein Resultat zeigt sich, dass dieser Algorithmus interessante Fortschritte bezüglich der Effizienz stochastischer Algorithmen aufweisen kann. Allerdings zeigt sich auch an diesem Beispiel die Notwendigkeit alternativer Vorgehensweisen bei der Handhabung großdimensionierter Problemstellungen. Die im Hauptteil der Dissertation beschriebene approximative Vorgehensweise erfährt auch hier eine erneute Rechtfertigung, da nur über diesen Weg auch Optimierungsprobleme höchster Dimensionalität in voller Allgemeinheit effizient lösbar erscheinen.

Chapter 1

Introduction

Within this thesis we want to address the solution of a very general problem setting, where for n given random variables

$$X_1, ..., X_n \in \mathbb{R}$$

the weighted composition

$$X(x) := x_1 \cdot X_1 + ... + x_n \cdot X_n \tag{1.1}$$

for $x_1, ..., x_n \in \mathbb{R}$ is analyzed in detail. Here, we presume the dependency structure among the random variables $X_1, ..., X_n$ to be known and given as scenario vectors of length k

$$X_i = \begin{pmatrix} \widehat{X}_i^1 \\ \vdots \\ \widehat{X}_i^k \end{pmatrix}, \widehat{X}_i^j \in \mathbb{R}$$

Using this notation the dependency structure is given implicitly using realizations of the $\widehat{X}_i^j, i = 1, ..., n$[1]. Furthermore, the weights x_i for $i = 1, ..., n$ are to be chosen in such a way that the expected value $\mathbb{E}(X(x))$ admits a preassigned value and such that $X(x)$ admits the lowest variance possible.

[1] Using an appropriate number of scenarios k, also continuous random variables can be approximated sufficiently.

Moreover, the case, where there are no bounds on the one-sided positive deviations from the mean, whereas the shortfall of a given threshold can be excluded with high confidence, is analyzed. In other words, the optimized composition of $X(x)$ for feasible x_i is to be determined in such a way that for some given confidence the shortfall of a threshold can be excluded.

While the general problem setting described above is worked out using a very special example, it is important to point out that at no point of the analysis the generality of the results restricting assumptions are made. When the problem at hand is explained using a concrete practical example the only reason for this is clearness. Besides this, the developed algorithm deals with the optimization of quantiles of a general weighted composition of random variables under a predetermined value of its expected value $\mathbb{E}(X(x))$. Without any further efforts, the algorithm can therefore easily be applied to general mathematical and statistical problem settings.

As already mentioned above the more general problem setting at hand is analyzed using some concrete example. In fact, we want to address the classical portfolio optimization problem at first rigorously investigated by Markowitz's pioneering work [34] on portfolio selection. There, Markowitz allows for the different asset classes' returns to be normally distributed and based on this forecast he constructs the optimal portfolio as the right trade-off between risk and expected return. In our setting we will also act on the assumption that the underlying universe of asset classes has been predefined by well-known distribution functions. However, we will not restrict to the case of normal distributions but accept almost any kind of distributions as given by random variables $X_i, i = 1, ..., n$. Writing $x \in \mathbb{R}^n$ and interpreting it as some portfolio allocation, the corresponding loss distribution of portfolio x is then given by equation 1.1.

One possible way of accurately measuring risk in the context of non-normal distributed asset returns is using the Value-at-Risk (VaR), which measures the smallest loss value, s.t. the probability of a loss exceeding VaR is not higher than $(1-\eta)\cdot 100\%$. Although this very important risk measure actually found its way to be written into industry regulations (see for example [28]) it lacks some very important properties risk measures in general should have ([4]). Apart from being unstable to work with numerically, VaR does not look at the extent of losses higher than VaR itself. In this sense, as long as there is the same probability of losses exceeding VaR, the measure does not distinguish between the situation where there is a cumulate of losses clearly exceeding VaR and the more preferable one where the tail-losses are more smoothed out.

There have been many efforts of efficiently solving optimization problems where the VaR appears either in the objective function or is part of the constraints. The reason that there is still no solver at hand which can handle large scale problems is mainly due to the fact that the VaR risk measure lacks subadditivity. A direct consequence is (in general) the existence of numerous local optima that exclude the usage of classical solvers such as the steepest descent method. In fact, Yang et al. show in ([62]) that the plain vanilla VaR optimization as considered in this thesis is NP hard if accounting for general loss distributions. Due to this result there are now several different directions current research addresses the solution of this problem.

The first possible reduction in complexity can be achieved by restricting oneself to a special class of distribution functions. It is widely known that assuming normally distributed asset returns results in a VaR optimization problem that is equivalent to the problem of minimizing standard deviation. A similar structured problem is the wider class of elliptically distributed returns, not only since this class again inherits the property of being closed under taking the sum of elliptically distributed returns. Another class of distribution function which inherits this property is the class of α-stable distributed returns. We will show that under some minor assumptions on the confidence level the complexity of the corresponding VaR optimization problem will only be of polynomial type.

On the contrary, if one wants to use arbitrary return distributions to match investment returns, another way of getting around the problem of increased complexity is the usage of similar but in some sense adapted risk measures. One example is the further usage of value at risk while smoothing out the irregular local behavior. Especially dealing with a large number of scenarios, the non-subadditivity and hence the existence of local optima becomes only important on a very small scale that can well be smoothed out by using the so called $SVaR$ as further described in ([21]). Away from explicitly approximating VaR, some authors also investigated new coherent (in particular subadditive) risk measures hence inheriting all the theoretical and optimizational desired properties. Among them is the so called "worst conditional expectation", proposed by Artzner et al. ([4]). Another in the context of this thesis very important invention of risk measure is the usage of "tail value at risk", "expected shortfall" or simply "conditional value at risk" ($CVaR$) which will be seen to have in many aspects a much more favorable behavior while in its definition being as simple as VaR. Whereas the usage of the term $CVaR$ mainly refers to the case of discrete distributions, all the other notations implicitly account for continuous distribution functions to be used. In order to avoid any further confusion, we will refer

to this risk measure as $CVaR$, regardless of the distribution function being continuous or discrete. That this measure of risk is in fact a coherent one for an almost arbitrary choice of distribution functions was shown only recently by the authors Uryasev and Rockafellar ([52] and [53]). The mean absolute deviation ([30]), the mean regret ([13]) and the maximum deviation ([63]) are some further examples of risk measures which, dealt as the objective of the corresponding portfolio optimization problem, can be stated as a linear optimization problem. Moreover, Cheklov et al. ([9]) develop a risk measure called conditional drawdown-at-risk ($CDaR$) closely related to the concept of $CVaR$. The subadditivity property of $CDaR$ is also shown to hold true.

Another possible approach in reducing complexity is the restriction to higher moments such as skewness and kurtosis besides mean value and standard deviation as in the Markowitz approach. This proceeding is based on the fact that the value at risk under some assumptions on the distribution function can be developed into a series of its central moments. This is done in an analogous way to the Taylor series expansion for differentiable functions and depends on different techniques such as the Cornish-Fisher or Gram-Charlie expansion ([25]). The author in [48] chooses a Cornish-Fisher expansion accounting for an additional skewness parameter in the objective function. However, this skewness parameter is set to be constant, hence it is not varying with the different portfolio allocations.

If either the most generality of distribution functions used and VaR as the objective should be attained, several algorithms were developed. On the one hand there are some techniques based on solving linear programming problems and isolating the corresponding VaR minimum using branch and bound methods. Pang and Leyffer ([46]) develop some good overview on this strategy. Pflug in [47] also develops some interesting reformulation of the VaR. In principle, classical methods could be applied to solve this restated objective function. Besides classical methods of optimization also heuristics such as Genetic Algorithms, Tabu Search, Simulated Annealing or Threshold Accepting (for a good overview compare e.g. [61]) among others were adjusted to match the particularities of the VaR objective. However, all the proceedings of this paragraph have in common not to be able to address problems of higher (i.e. higher than 15) dimensions efficiently. After concentrating on the development of a good approximating algorithm to handle the VaR optimization problem we will also present a stochastic branch & bound algorithm at the end of this thesis. Using a slight reformulation of the original problem it is possible to apply stochastic optimization tools which appear to obtain interesting features regarding the efficiency of the corresponding solvers.

Investigation of the strong relationship between $CVaR$ and VaR being the central topic of this thesis, it is worth to mention again the very promising approach of using $CVaR$ instead of the VaR objective in portfolio optimization. Whereas from a practitioner'ss point of view $CVaR$ can be shown to be always the more conservative measure of risk Uryasev and Rockafellar succeed in proving the coherence of this measure thereby showing superiority when compared to VaR since it is exactly this property which allows for optimization problems to be easily solved. Dealing with a relatively easy to solve measure of risk on the other hand enables numerous other authors to solve for more advanced settings. For example Krokhmal et al. in [31] extend the setting to problems with $CVaR$ being part of the constraints. Moreover, they investigate the efficient line within the $CVaR$ framework.

We now want to give a short overview on the different meander towards our final result as stated in section 3.3. Before using g-and-h distributions to cope the matching of arbitrary distributions, we tried to investigate several other methods of approximation. First, as already mentioned before, we used different series such as Edgeworth-, Gram-Charlier- or Cornish-Fisher-Series which all approximate the respective probability functions in terms of their cumulants, hence can be used to approximately express both VaR and $CVaR$. However, convergence of the corresponding series is not always guaranteed and the error of truncating the series after some finite cumulant to our knowledge cannot be estimated within the portfolio context. Moreover, it is not entirely clear how to handle the additional cumulants in the portfolio optimization process.

Next we tried to use the approximate result for stable distributed returns already at hand to extend it to arbitrary distributions by approximating them through α-stable distributions. It is important to note, that using stable distributions to approximate the original VaR-problem directly, one has to find an appropriate function on the n-dimensional sphere, since such a function can be shown to give a dependence structure for multidimensional stable distributed random vectors (see e.g. [2], [42] or [39]). In order to significantly reduce complexity of the related optimization problem, we have to guarantee convexity of $CVaR_\zeta(x)$ as a function of portfolio composition x. This can only be done by giving an overall dependence structure, as can be seen by following the proofs in [52]. However, the adaptation of a general dependence structure, as the search for an appropriate function on the n-sphere, to arbitrary distributions seemed too inflexible and too expensive.

Our third approach was driven by the observation, that in general the VaR as well as the $CVaR$ can quite well be approximated by using the first mo-

ments of a distribution. Again using the result for stable distributed asset returns, we tried to make usage of truncated stable distributions in order to firstly approximate the original distributions via the first three moments. In a second step we tried to use small truncation parameters $\lambda > 0$ (see e.g. [35] for the work with truncated distributions) of the truncated stable distribution to best approximate a corresponding $VaR/CVaR$-optimization problem characterized by the parameter values $(\mu(x), \sigma(x), \beta(x))$ which again correspond to the expected value, standard deviation and skewness, respectively. However, this proceeding also has several disadvantages. On the one hand, approximation by only the first three moments may not be enough to yield appropriate results. Moreover, to match the stable distribution by the corresponding truncated stable distribution would lead to the usage of very small values λ. This on the other hand leads to very small values $\sigma(x)$, $\beta(x)$ with problems for the corresponding optimization problem. Besides these disadvantages, convexity of the resulting $CVaR$-optimization problem cannot be guaranteed.

We firstly develop the general notations used dealing with the VaR and $CVaR$ objective in Chapter 1. After introducing the main definitions and properties of these two risk measures we next have a closer look on the complexity of the general VaR optimization problem. Moreover, in this introductory chapter, Section 2.5 gives a rough overview on the two major families of distribution functions used in this thesis, stable distributions and g-and-h distributions.

Chapter 2 starts with solving the $CVaR$ optimization problem as transforming it to a linear programming problem. In a second step Benders decomposition is used to reduce the dimensionality of the resulting linear problem. Having recognized the $CVaR$ optimization problem to be of polynomial complexity, we analyse polynomial VaR optimization problems for special classes of distribution functions in Section 3.2. In the subseeding sections we firstly use g-and-h distributions to match arbitrary distribution functions in order to find approximate statements for the euclidean distance between the VaR optimum and the $CVaR$ optimum for some suitably chosen confidence level $\zeta_0 \in (0, 1)$. The necessary conditions for this estimation are considered numerically. In a next step, this approximate result can be shown to hold true for almost any kind of distribution function. This is part of Section 3.3.

Within Chapter 3 we will develop the adaption of stochastic branch & bound techniques to directly address the VaR optimization problem. The implementation of such an algorithm allows for the solution of moderately (i.e. up to 10) dimensioned optimization problems efficiently. Chapter 4 summarizes

our results and gives some directions of further research.

Chapter 2

Problem Statement and Conjecture

2.1 General Notations and Definitions

Throughout this thesis, random variables are denoted by capital letters Y, Z, $X, X_1, ...$ and are assumed to be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution of the random variable $X_i, i = 1, ..., n$ can be interpreted as a loss distribution, giving the probability of losses for the respective insecure investments. Clearly, using this setting, $\mathbb{E}(X_i)$ is to be interpreted as the mean value of the loss distribution, hence the negative of the corresponding return distribution. We will denote by $x = (x_1, ..., x_n)^t \in \mathbb{R}^n$ the concrete portfolio compositions given as the percentage allocations to the investment possibilities, i.e. x_i denotes the percentage exposure to asset class $X_i, i = 1, ..., n$. The set of permissible portfolio allocations is denoted by $\mathbb{X} \subset \mathbb{R}^n$ and is restricted to be a subspace of $\mathbb{R}^n$ given, for example, in standard form

$$\mathbb{X} = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Moreover, we will always require to be fully invested without the possibility of going short, i.e. the minimum requirement on $\mathbb{X}$ will be

$$\mathbb{X} = \left\{ x = (x_1, ..., x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1, x_i \geq 0, i = 1, ..., n \right\}. \tag{2.1}$$

We are now in the position to define the two main objectives of this paper,

VaR and $CVaR$.

Definition 2.1.1 *For $x \in \mathbb{X}$, $\eta \in (0,1)$ the $VaR_\eta(x)$ is implicitly defined by*

$$VaR_\eta(x) := \inf\{z \in \mathbb{R} : \mathbb{P}\{X(x) \geq z\} \leq \eta\}. \tag{2.2}$$

Moreover, for $x \in \mathbb{X}, \zeta \in (0,1)$, the expected shortfall $CVaR_\zeta(x)$ is defined by

$$CVaR_\zeta(x) := \mathbb{E}[X(x) \mid X(x) > VaR_\zeta(x)]. \tag{2.3}$$

Remark 2.1.1 *It is important to note that Definition 2.2 is suitable for continuous distribution functions as well as for discrete ones. However, in the case of discrete distribution functions (as in any real world example characterized by scenario representations) more subtle definitions are required for rigorously defining expected shortfall, c.f. [53]. The same authors show that by dealing with the discrete case one can define $CVaR$ as*[1]

$$CVaR_\zeta(x) := \inf_{a \in \mathbb{R}} \left\{ a + \frac{1}{1-\zeta}\mathbb{E}[X(x) - a]^+ \right\}, \tag{2.4}$$

where $[z]^+ = \max(z, 0)$. It is shown that for smooth distribution functions both definitions are equivalent (compare Theorem 10 in [53]). Since the expected shortfall is nothing else than a conditional expectation, the aforementioned definitions are not the only possible way of defining $CVaR$. If $X(x)$ admits a probability density function $f_x(t)$, this can be used to define $CVaR$ as

$$CVaR_\zeta(x) = \frac{1}{1-\zeta}\int_{VaR_\zeta(x)}^{\infty} t \cdot f_x(t)dt.$$

If the distribution function F_x under consideration is smooth, it is easy to see that Definition 2.2 is equivalent to

$$VaR_\eta(x) = F_x^{-1}(\alpha).$$

Remark 2.1.2 *Comparing the two definitions for VaR and $CVaR$, it becomes obvious that while VaR is not handling the extend of the losses beyond*

[1] For a motivation of this definition compare the proof to theorem 2.4.1.

some threshold, taking the conditional mean in the definition of $CVaR$ allows for the recognition of the loss structure beyond this threshold. Whilst only providing some lower bound for the losses in the tail, VaR is the more optimistic risk measure if compared to the more conservative usage of $CVaR$. Compare the discussion in ([53], p. 1444).

Definition 2.1.2 *During the whole paper, we will mainly focus on the following two optimization problems*

$$\min_{x\in\mathbb{X}} VaR_\eta(x) \tag{2.5}$$

and

$$\min_{x\in\mathbb{X}} CVaR_\zeta(x). \tag{2.6}$$

for various confidence levels $\eta, \zeta \in (0,1)$. However, in order to show some equivalence results of these two optimization problems we will also solve the problem

$$\min_{x\in\mathbb{X}} CVaR_{\zeta_0,c}(x), \tag{2.7}$$

where we denote by

$$CVaR_{\zeta_0,c}(x) := c\sigma(x)\cdot\overline{CVaR}_{\zeta_0}(x) + \mu(x)$$

the $CVaR$ objective with transformed scaling parameter $c\sigma(x)$. $\overline{CVaR}_{\zeta_0}(x)$ is thereby defined to be

$$\overline{CVaR}_{\zeta_0}(x) := CVaR_{\zeta_0}\left(\frac{x-\mu(x)}{\sigma(x)}\right).$$

Remark 2.1.3 *Without explicitly referring to it, the usage of a transformed scaling parameter $c\sigma(x)$ will always assume the existence of an appropriate scaling parameter $\sigma(x)$.*

As noted earlier, the set $\mathbb{X}$ can also account for different linear restrictions to the optimization problem. In particular, restrictions on the minimum achievable return of the optimal portfolio allocation, denoted by x_η^* respectively x_ζ^*, can be imposed. This gives us the possibility to generate efficient lines in an environment, where risk is measured by VaR respectively $CVaR$. However, for the purpose of proving our main results we will always make usage of the

Karush-Kuhn-Tucker conditions. To guarantee nondegeneratedness of these necessary conditions for a local minimum x^*, a number of regularity conditions, such as the Linear Independence Constraint Qualification (LICQ) were developed (see e.g. [37]). The LICQ states that the optimization problem is regular in x^* if the gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at x^*. $\mathbb{X}$ as defined in equation 2.1 clearly satisfies LICQ since at least one of the constraints is not active. The inclusion of additional constraints, however, must be in a way such that we are dealing with x^* being regular.

Besides the plain vanilla optimization problems as stated in (2.5) and (2.6) and its extensions to the generation of efficient lines, there exist various interesting generalizations that account for more complex settings. Among all these possible extensions we would like to emphasize the incorporation of liabilities in the present context rather than restricting oneself to the analysis of an asset only portfolio. The optimal matching of stochastic liabilities that cannot be directly influenced by changing some portfolio allocation is known as Asset Liability Management (ALM) and attains increased significance not only in the insurance industry but also within corporate companies and investment banks. Considering stochastic liabilities as an additional input parameter to the optimization problem can have a tremendous effect on the optimal allocation x^*. Denoting by ξ some arbitrarily distributed random variable, those liabilities ξ could easily be incorporated by setting

$$X := X(x) := (X_1, ..., X_n) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \xi,$$

where we used the same notation as before. Kuenzi-Bay and Mayer ([29]) in their description of numerically minimizing $CVaR$ allow for such a setting and show that the additional presence of liabilities do not disturb the nice property of $CVaR$ being a coherent measure of risk. However, it is not entirely clear how VaR behaves with additional liabilities changing the overall loss distributions. In particular, it is not known in how far liabilities would destroy the properties of the algorithm presented in the next chapter. Future research on this topic may be worth to perform.

2.2 Formulation of the Algorithm

To describe the fundamental algorithm of this thesis, we will firstly have to introduce a set of striking portfolio compositions. Based on the optimization problem (2.7) we therefore define

$$\mathbb{X}^* := \left\{ x \in \mathbb{X} \,|x \text{ solution of (2.7)}, \zeta \in (0,1), c \in \mathbb{R}^+ \right\}.$$

and denote by $x^*_{\zeta,c}$ the respective elements of $\mathbb{X}^*$. Furthermore, by x^*_ζ we will shorthand refer to the portfolio $x^*_{\zeta,1}$. We will then consider the following optimization problem

$$\min_{x \in \mathbb{X}^*} VaR_\eta(x) \tag{2.8}$$

and show that the resulting optimal portfolio is either equal to the solution of the original problem (2.5) in the case of specially distributed asset returns. For those cases, where we are not able to show exact matching of optimal portfolios, we will state an upper bound for the euclidean distance of the two portfolio compositions as given by the arguments of (2.5) and (2.8). Hence, in searching for the global minimum of the original (maybe huge dimensional) problem one gets reasonable results in restricting oneself to the set $\mathbb{X}^*$ and therefore reducing to a two dimensional search which generally is much more efficient to solve. Moreover, some of the results of this thesis choose c in the definition of the set $\mathbb{X}^*$ to be equal to one hence resulting in a one dimensional search equivalent. It is interesting to note, that in those cases, where we allow for an arbitrary positive value c, we found that the optimal value for c in the sense of (2.8) is also close to one.

Therefore, the main complexity in order to solve (2.8) lies in the determination of the set $\mathbb{X}^*$ which itself heavily depends on the effective solution of the affine $CVaR$ optimization problem for suitable confidence levels $\zeta \in (0,1)$. There we can use Rockafellar's and Uryasev's result on efficiently optimizing $CVaR$ for some fixed confidence level ζ. In practice, for a suitable choice c_l, c_u we will take some sufficiently close meshed grid on $(0,1) \times [c_l, c_u]$ and for every element (ζ, c) on this grid we solve the corresponding $CVaR$ optimization problem. That this procedure in fact will produce a helpful approximation to the set $\mathbb{X}^*$ is the result of the next lemma which states that optimization (2.8) generally is well behaved.

Lemma 2.2.1 *Let $VaR_\eta(x)$ be a continuous function of x and suppose that for every confidence level $\zeta \in (0,1)$ there exists a unique solution to the*

$CVaR$ optimization problem (2.6). Then for $c > 0$ the function

$$\begin{aligned} h : (0,1) &\longrightarrow \mathbb{R} \\ \zeta &\longmapsto VaR_\eta\left(x^*_{\zeta,c}\right) \end{aligned} \tag{2.9}$$

is continuous.

Proof: W.l.o.g. we can restrict to the case $c = 1$. Other values of c do not change the qualitative statement. We prove continuity in an arbitrarily chosen confidence level $\zeta_0 \in (0,1)$.

Step 1: *$CVaR_\zeta(x)$ is a continuous function of the confidence level ζ.* This result was proven for arbitrary distribution functions in ([53]).

Step 2: *The mapping*

$$\zeta \longmapsto CVaR_\zeta(x^*_\zeta)$$

is continuous. For each $a \in \mathbb{R}$ consider the function of $\gamma \in \mathbb{R}, x \in \mathbb{X}$ defined by

$$\theta_{a,x}(\gamma) = a + \gamma\mathbb{E}\left[(X(x) - a)^+\right],$$

and let

$$\theta(\gamma) = \min_{x\in\mathbb{X}, a\in\mathbb{R}} \theta_{a,x}(\gamma). \tag{2.10}$$

Hence by Theorem 10 and 14 of ([53])

$$\begin{aligned} CVaR_\zeta(x^*_\zeta) &\equiv \min_{x\in\mathbb{X}} CVaR_\zeta(x) \\ &= \min_{x\in\mathbb{X}, a\in\mathbb{R}} \theta_{a,x}(\gamma) \\ &= \theta(\gamma) \text{ for } \gamma = 1/(1-\zeta), \end{aligned} \tag{2.11}$$

with the minimum in (2.10) being attained when a belongs to the interval $[VaR_\zeta(x^*_\zeta), VaR^+_\zeta(x^*_\zeta)]$.

According to (2.10), θ is the pointwise minimum of the collection of functions $\theta_{a,x}$. Those functions are affine, hence θ is concave (c.f. [51], Theorem 5.5). A finite, concave function on $\mathbb{R}^n$ is necessarily continuous (c.f. [51], Theorem 10.1). The result follows through 2.11 by considering the function

$\zeta \mapsto CVaR_\zeta(x^*_\zeta)$ as the composition of θ with $\zeta \mapsto 1/(1-\zeta)$ and invoking the chain rule.

Step 3: *The mapping*

$$\zeta \longmapsto x^*_\zeta \tag{2.12}$$

is continuous. Using continuity of Step 1, for any $\varepsilon > 0$ there exists a $\delta_1 > 0$, s.t.

$$|\zeta - \zeta_0| < \delta_1 \Rightarrow \left|CVaR_\zeta(x^*_{\zeta_0}) - CVaR_{\zeta_0}(x^*_{\zeta_0})\right| < \frac{\varepsilon}{2}$$

Moreover, by Step 2 there exists $\delta_2 > 0$, s.t.

$$|\zeta - \zeta_0| < \delta_2 \Rightarrow \left|CVaR_\zeta(x^*_\zeta) - CVaR_{\zeta_0}(x^*_{\zeta_0})\right| < \frac{\varepsilon}{2}$$

Defining $\delta := \min\{\delta_1, \delta_2\}$ and for all confidence levels ζ satisfying $|\zeta - \zeta_0| < \delta$ we can write

$$\begin{aligned}\left|CVaR_\zeta(x^*_\zeta) - CVaR_\zeta(x^*_{\zeta_0})\right| \leq \quad & \left|CVaR_\zeta(x^*_\zeta) - CVaR_{\zeta_0}(x^*_{\zeta_0})\right| \\ + \quad & \left|CVaR_{\zeta_0}(x^*_{\zeta_0}) - CVaR_\zeta(x^*_{\zeta_0})\right| \\ < \quad & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

Hence by the convexity of $CVaR$ and the assumed uniqueness of its optimum for any sufficiently small $\widetilde{\delta} > 0$, $\varepsilon > 0$ can be chosen, s.t.

$$\left\|x^*_\zeta - x^*_{\zeta_0}\right\| < \widetilde{\delta}.$$

Since $VaR_\eta(x)$ is a continuous function of the portfolio composition x, also the composition with the continuous function as defined in (2.12) is continuous which is the claim of the lemma. □

2.3 Complexity of the VaR-Optimization Problem

The authors in [62] investigate the optimization problem 2.5 and show that the problem of finding the optimal allocation x^*_η is generally Non-deterministic Polynomial (NP) -hard.

Although being formally correct, in their proof to NP-hardness, the authors are using some predefined distribution function of the individual asset classes. Moreover, the number of used scenarios is not to be arbitrarily chosen and the proof only allows for the confidence level $\eta = \frac{1}{2}$. Hence, it is not entirely clear if the optimization problem looses its property of being NP-hard when we are allowing for an arbitrary number of scenarios and/or if we are only interested in some confidence level close to one. Below, we are stating an alternative (and in our sense more intuitive) proof of NP-hardness, thereby accounting for the above mentioned remedies.

Theorem 2.3.1 *The scenario-based VaR_η optimization problem (2.5) is NP-hard.*

Proof: Suppose there exists some joint continuous density function $p(y)$ for the losses denoted by y. Before reducing the NP-hard "Subset Sum" problem to the current one, let us implicitly define VaR_η using the joint density function

$$\int\limits_{x \cdot y \geq VaR_\eta} p(y)dy = \eta.$$

The main complexity in solving for the optimal portfolio allocation in (2.5) lies in the decision whether or not there exists a feasible solution which accounts for some predefined value VaR_η (c.f. [62]). To be more rigorous, denoting by V the minimum achievable value of the original problem,

$$V := \min_{x \in \mathbb{X}} VaR_\eta(x),$$

for some given value V^t, problem (2.5) is in complexity equivalent to the decision whether there exists $x \in \mathbb{X}$ with

$$\int\limits_{x \cdot y \geq V^t} p(y)dy = \eta. \tag{2.13}$$

Discretising Equation 2.13 yields[2]

$$\min \left\{ \sum_{x \cdot y_i \geq V^t} p(y_i) | \sum_{x \cdot y_i \geq V^t} p(y_i) \geq \eta \right\} \tag{2.14}$$

[2] Compare the definition of VaR for discrete distributions.

for loss values y_i. Suppose now there exists a polynomial algorithm that decides for any joint density function and any choice of loss values $y_i, i = 1, ..., n$ whether there exists a corresponding $x \in \mathbb{X}$, s.t. (2.14) holds. Defining $\frac{a_i}{\sum_k a_k} := p(y_i)\ \forall i = 1, ..., n$ and

$$\begin{aligned} S_1 &: = \left\{ a_i \mid xy_i \geq V^t \right\} \\ S_2 &: = \left\{ a_i \mid xy_i < V^t \right\} \end{aligned}$$

the problem is equivalent to the one of deciding whether there exists a subset S_1 of $S := \{a_1, ..., a_n\}$, s.t.

$$\sum_{a_i \in S_1} a_i = \eta. \tag{2.15}$$

which by assumption would also be solvable within polynomial time. This would be in contradiction to the well known fact that the so called Subset Sum problem (2.15) is NP-hard (compare [56]). □

Note that the existence of a joint density does not generally imply the corresponding VaR problem to be hard to solve. For example, the case of normally distributed asset returns with joint distribution function is known to be easily treatable. Moreover, J. Danielsson et al. (c.f. [11]) give some further examples where either the confidence level or the special structure of the distribution functions are chosen in a way that the resulting optimization problem is of polynomial type. All these examples correspond to some simple structured Subset Sum problem as in the proof of the theorem. However, following the lines of the theorem, one attains any kind of Subset Sum problem by a suitable choice of the joint density function.

Although we cannot give a rigorous proof, the experience of the calculations to this thesis as well as the results given below teaches most of the undesired behavior to be caused by discretising the originally assumed continuous or even differentiable distribution functions. Discretising the random variables to work with numerically seems to produce a large number of local extremas that would not be observed solving the problem analytically. Using the advantage of convex $CVaR$ optimization in the proposed algorithm for VaR optimization surpasses the problem connected with the discretization of the original distribution functions.

2.4 Properties of VaR and CVaR

Within this section we want to present the major properties of the two risk measures under consideration. It is not our intention to give a full review of all the VaR and $CVaR$ inherent properties but to restrict to those properties that are directly used in our approach of minimizing VaR. In particular, we will concentrate on the coherence of VaR and $CVaR$ as well as stating some important relations between those two risk measures. The subadditivity property will be seen to be the crucial difference between the two risk measures; its absence is the reason for the increased complexity in optimizing VaR.

2.4.1 Coherent Measures of Risk

Artzner et al. [4] show that there exists a set of axioms that risk measures should fulfill. The authors also show that these axioms are complete in the sense that if a measure does not satisfy all the axioms, this may lead to undesirable conclusions. Considering a set V of real valued random variables with ρ describing some arbitrary risk measure, these axioms can be stated as

(a) (translation invariance) $Y \in V, c \in \mathbb{R} \Rightarrow \rho(Y + c) = \rho(Y) + c$

(b) (positive homogeneity) $Y \in V, c > 0, cY \in V \Rightarrow \rho(cY) = c\rho(Y)$

(c) (sub-additivity) $Y_1, Y_2, Y_1 + Y_2 \in V \Rightarrow \rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$

(d) (monotonicity) $Y_1, Y_2 \in V, Y_1 \prec Y_2 \Rightarrow \rho(Y_1) \leq \rho(Y_2)$

where in (d) we used the relation $Y_1 \prec Y_2$ iff $\mathbb{E}[\psi(Y_1)] \leq \mathbb{E}[\psi(Y_2)]$ for all integrable monotonic functions ψ.

Without any further assumptions it can easily be seen (compare e.g. [47]) that both VaR and $CVaR$ as defined above satisfy translation invariance, positive homogeneity as well as monotonicity. The property of being subadditive is always fulfilled using $\rho = CVaR$, hence by positive homogeneity $CVaR$ admits the property of being convex in the sense that for $0 < \lambda < 1$

$$CVaR_\zeta(\lambda Y_1 + (1-\lambda)Y_2) \leq \lambda CVaR_\zeta(Y_1) + (1-\lambda)CVaR_\zeta(Y_2).$$

We will have a closer look at this property in Subsection 2.4.3. Since VaR is not always subadditive (i.e. the property depends strongly on the confidence level and the set V used) it does not inherit convexity in general. Apart from the above described fact that VaR tends to be a too optimistic measure of risk, the lack of being subadditive is one of the major drawbacks of using VaR in risk management. Beyond accounting for nice mathematical properties, in practice the usage of convex risk measures has the advantage that risk measurement on a subsidiary level guarantees the overall risk not to exceed the sum of the individual risks. Using VaR this conclusion might be false and result in a wrong risk allocation. We will have a closer look on this topic in Subsection 2.5.6.

2.4.2 Relations between VaR and CVaR

VaR is a quantile whereas $CVaR$ measures the conditional tail expectation. Hence these two risk measures obviously coincide only if the tail is cut off and the cut is in $VaR_\eta(x)$ itself. Nevertheless, within the next theorem we will state two very important relations between VaR and $CVaR$.

Theorem 2.4.1 **(i)** *Considering the same confidence level ζ for VaR and $CVaR$, $CVaR_\zeta(Y)$ always succeeds $VaR_\zeta(Y)$,*

$$VaR_\zeta(Y) \leq CVaR_\zeta(Y),$$

for any random variable Y.

(ii) *Writing F_x for the distribution function of some random variable suppose ζ is in the range of F, i.e. $F(F^{-1}(\alpha)) = \alpha$, then*

$$CVaR_\zeta(Y) = \frac{1}{1-\zeta}\int_\zeta^1 VaR_\varepsilon(Y)d\varepsilon. \tag{2.16}$$

Proof: Using the definitions of VaR and $CVaR$ (i) is an obvious result. Under the assumptions of the theorem, equation (2.16) is again nothing else than definition (2.3) and writing the conditional expectation in integral form. Starting with the definition for $CVaR$ as given in (2.4), the following relation gives an interesting insight to the motivation of defining $CVaR$ by (2.4). Let

b be chosen such that $F(b) = \zeta$. Then $\mathbb{P}\{Y > b\} = 1 - \zeta$ and

$$\begin{aligned}\mathbb{E}(Y \mid Y > b) &= \frac{\mathbb{E}\left(Y\mathbb{I}_{\{Y>b\}}\right)}{\mathbb{P}\{Y > b\}} \\ &= \frac{\mathbb{E}\left(b\mathbb{I}_{\{Y>b\}} + [Y-b]^+\right)}{\mathbb{P}\{Y > b\}} \\ &= b + \frac{1}{1-\zeta}\mathbb{E}\left([Y-b]^+\right).\end{aligned}$$

This shows, that under the assumptions of the theorem equation (2.16) holds also when using the alternative definition of $CVaR$ for discrete distribution functions. □

2.4.3 Subadditivity of CVaR

The main purpose of this paper is to solve the VaR optimization problem by a sequence of affine $CVaR$ optimization problems. It is therefore not surprising that the aspect of efficiently finding the solution of the $CVaR$ objective is one of our main focuses. Since in practice we are always confronted with a finite set of scenarios representing some (theoretical) distribution function, we want to show that the above mentioned subadditivity property is not only true for sufficiently smooth distribution functions, but also holds for those cases where the distribution function is a discrete step function. In financial applications this is the only case of interest since any optimization has to be performed on a finite set of scenarios. Moreover, it is the intention of this subsection to formulate the corresponding $CVaR$ optimization problem in a way that allows for a simple implementation. Hence this subsection can be interpreted to provide the necessary background for further developments of efficient $CVaR$ solvers as described in Section 3.1. All the results of this section can be found in the joint work of Rockafellar and Uryasev ([53]). Although the authors are dealing in a more general context, we are concentrating on the pure portfolio allocation problem hence stating their results in a more specialized framework.

As the authors in [53] show, using discrete distribution functions instead of continuous ones makes it necessary to distinguish between the definition of $VaR_\eta(x)$ as given in (2.2) and

$$VaR^+_\eta(x) := \inf\{z \in \mathbb{R} : \mathbb{P}\{X(x) \geq z\} > \eta\}.$$

However, this distinction only becomes relevant for a small number of scenarios. If we are dealing with thousands of scenarios, the difference between the definition of $VaR_\eta^+(x)$ and $VaR_\eta(x)$ will only be marginal. For a small number of scenarios an appropriate definition of analogous values $CVaR_\zeta^+(x)$ and $CVaR_\zeta^-(x)$ to the existing definition of $CVaR_\zeta(x)$ becomes necessary. Since in this paper we are assuming to deal with a sufficiently large number of scenarios and since we do not need this distinction in the further course of this thesis, we will omit the corresponding definitions. For further details consult the above mentioned paper by Rockafellar and Uryasev.

We are now in the position to state one of the main advantages of using $CVaR$ over VaR as a risk measure.

Theorem 2.4.2 *Considering the definition of $CVaR_\zeta(x)$ as given in 2.4 define*

$$G_\zeta(x,a) := a + \frac{1}{1-\zeta}\mathbb{E}\left[X(x)-a\right]^+,$$

with $[t]^+ := \max\{0,t\}$.

(i) *Since $X(x)$ is a linear function of x, $G_\zeta(x,a)$ is a jointly convex function in x and a. In particular, $CVaR_\zeta(x)$ is a convex function of the portfolio composition x.*

(ii) *Minimizing $CVaR_\zeta(x)$ as a function of $x \in \mathbb{X}$ is equivalent to minimizing $G_\zeta(x,a)$ over all $(x,a) \in \mathbb{X}\times\mathbb{R}$ in the sense that*

$$\min_{x\in\mathbb{X}} CVaR_\zeta(x) = \min_{(x,a)\in\mathbb{X}\times\mathbb{R}} G_\zeta(x,a).$$

Moreover,

$$(x^*,\zeta^*) \in \arg\min_{(x,a)\in\mathbb{X}\times\mathbb{R}} G_\zeta(x,a) \Longleftrightarrow \begin{cases} x^* \in \arg\min_{x\in\mathbb{X}} CVaR_\zeta(x), \\ \zeta^* \in \arg\min_{a\in\mathbb{R}} G_\zeta(x^*,a) \end{cases}$$

Proof: For a full proof of the statements of the theorem see ([53]). □

The statement of the theorem can now be used to implement the corresponding $CVaR_\zeta(x)$ optimization problem in an efficient way. In fact, we will see that the special structure of the objective function can be used to state the

minimization problem as a linear one. Further simplifications using Benders decomposition can be applied in order to increase efficiency. We will focus on these developments in Section 3.1 below.

After having stated the subadditivity property of $CVaR_\zeta(x)$, we want to show that this property is stable under some minor change of the objective. Below, we will show that the VaR_η optimization problem is in some sense and under several assumptions on the distribution functions equivalent to the detection of the $CVaR_{\zeta_0,c}$ minimizing portfolio for some given $(\zeta_0, c) \in (0,1) \times \mathbb{R}^+$. Here, the objective function is given by

$$CVaR_{\zeta_0,c}(x) := c \cdot \sigma(x)\overline{CVaR}_{\zeta_0}(x) + \mu(x),$$

again using $\overline{CVaR}$ to denote the $CVaR$ of the standardized distribution.

The next lemma shows that this objective function is also a convex function of the portfolio composition $x \in \mathbb{X}$.

Lemma 2.4.3 *For all $c \in \mathbb{R}^+$, $\zeta_0 \in (0,1)$ and under the assumptions of Theorem 2.4.2 $CVaR_{\zeta_0,c}(x)$ is a convex function of x.*

Proof: Theorem 2.4.2 gives us the convexity of $CVaR_{\zeta_0,1}(x)$. Hence, $c \cdot CVaR_{\zeta_0,1}(x)$ is a convex function and we can write

$$\begin{aligned} c \cdot CVaR_{\zeta_0,1}(x) &= c \cdot \sigma(x)\overline{CVaR}_{\zeta_0}(x) + c \cdot \mu(x) \\ &=: \widetilde{CVaR}_{\zeta_0,c}(x), \end{aligned}$$

Here $\widetilde{CVaR}_{\zeta_0,c}(x)$ denotes the expected shortfall, where for all $i = 1, ..., n$ the original asset classes X_i are replaced by linearly transformed $\widetilde{X}_i$ with $\mu(\widetilde{X}_i) = c\mu(X_i)$. It is important to note, that by the linearity of the expected value $\mu(x)$, $\widetilde{CVaR}_{\zeta_0,c}(x)$ is convex iff $CVaR_{\zeta_0,c}(x)$ is convex. □

2.4.4 Differentiability of VaR and CVaR

Our results on the equivalence of the VaR_η and the $CVaR_\zeta$ optimization problem heavily rely on the differentiability of the respective objective function in the portfolio context, i.e. differentiation with respect to the portfolio

composition x. First developments into this direction can be found in [24] and [23]. Although these papers already provide readily interpretable formulae for the case of linear combinations of random variables, the authors did not yet attend the question under whichsoever general conditions their formulae in fact hold true, hence addressing the question of differentiability. Addressing both the VaR and the $CVaR$ risk measure, Tasche in ([59], Sec. 5.2 and 5.3), develops some certain conditions under which differentiability is fulfilled. There one also finds some helpful discussion on distribution functions more frequently used in financial applications. For completeness, we will state those conditions summarized in Assumption (S). Compare ([59], p. 16) for the original context.

Assumption (S) For fixed $\eta \in (0,1)$, we say that an $\mathbb{R}^n$ valued random vector $(X_1, ..., X_n)$ satisfies Assumption (S) if $n \geq 2$ and the conditional distribution of X_1 given $(X_2, ..., X_n)$ has a density

$$\begin{aligned} \Phi : \mathbb{R} \times \mathbb{R}^{n-1} &\longrightarrow [0, \infty) \\ (t, x_2, ..., x_n) &\longmapsto \Phi(t, y_2, ..., y_n) \end{aligned}$$

which satisfies the following four conditions

(i) For fixed $y_2, ..., y_n$ the function $t \longmapsto \Phi(t, y_2, ..., y_n)$ is continuous in t.

(ii) The mapping

$$\begin{aligned} &\mathbb{R} \times \mathbb{R}\backslash\{0\} \times \mathbb{R}^{n-1} \longrightarrow [0, \infty) \\ (t, x) \longmapsto \mathbb{E}&\left[\Phi\left(x_1^{-1}\left(t - \sum_{j=2}^{n} x_j X_j\right), X_2, ..., X_n\right)\right] \end{aligned}$$

is finite-valued and continuous.

(iii) For each $x \in \mathbb{R}\backslash\{0\} \times \mathbb{R}^{n-1}$

$$0 < \mathbb{E}\left[\Phi\left(x_1^{-1}\left(VaR_\eta(x) - \sum_{j=2}^{n} x_j X_j\right), X_2, ..., X_n\right)\right].$$

(iv) For each $i = 2, ..., d$ the mapping

$$\begin{aligned} &\mathbb{R} \times \mathbb{R}\backslash\{0\} \times \mathbb{R}^{n-1} \longrightarrow \mathbb{R} \\ (t, x) \longmapsto \mathbb{E}&\left[X_i \Phi\left(x_1^{-1}\left(t - \sum_{j=2}^{n} x_j X_j\right), X_2, ..., X_n\right)\right] \end{aligned}$$

is finite-valued and continuous.

Under these assumptions, Tasche shows that $VaR_\eta(x)$ is a differentiable function of x. If the mean value of $X(x)$ is finite, he also shows differentiability of $CVaR_\zeta(x)$. Moreover, he again develops the formulae first published in [24] and [23] expressing the partial derivatives of $VaR_\eta(x)$ and $CVaR_\zeta(x)$ in the following descriptive way

$$\frac{\partial}{\partial x_i} VaR_\eta(x) = \mathbb{E}\left\{X_i : X(x) = VaR_\eta(x)\right\}, i = 1, ..., n. \tag{2.17}$$

similarly, the partial derivative of $CVaR_\zeta(x)$ yields the expressive form

$$\frac{\partial}{\partial x_i} CVaR_\zeta(x) = \mathbb{E}\left\{X_i : X(x) \geq VaR_\zeta(x)\right\}, i = 1, ..., n. \tag{2.18}$$

Scaillet in [55] derives similar formulas and discusses its application to the normal distributed case. Formulas (2.17) and (2.18) show the main difference between the two risk measures. Whereas under some assumptions on the distribution function VaR appears to be partially differentiable, its derivative becomes in some sense unstable, indicating that generally only very restrictive assumptions on the distribution function will yield derivatives of higher order. The aforementioned condition of being unstable becomes apparent when numerically evaluating expression (2.17). Having vectors $\widehat{X}_i$ representing the random variable $X_i, i = 1, ..., n$, the corresponding partial derivative can be approximated by taking the entry of $\widehat{X}_i$ which corresponds to the entry of $\widehat{X}^t \cdot x$ which again represents $VaR_\eta(x)$. Obviously, such evaluations can result in very different partial derivatives.

On the other hand, formula (2.18) can be used in a similar way by first sorting the vector $\widehat{X}_i$ with respect to the sorted entries of $\widehat{X}^t \cdot x$ and taking the mean over the $(1 - \zeta) \cdot 100\%$ of the last entries. Taking the mean makes the partial derivative much more stable. Expressions for the VaR and the $CVaR$ derivatives of arbitrary order are given in [50]. However, actual differentiability conditions for higher order derivatives in an arbitrary context are (to our best knowledge) not yet developed.

At some stage we are not only interested in the differentiability with respect to the portfolio composition but also in the differentiability with respect to the confidence level η. For distributions with continuous distribution function, the result for VaR immediately follows from

$$\frac{\partial}{\partial \eta} VaR_\eta(x) = \frac{\partial}{\partial \eta} F_x^{-1}(\eta)$$

and the differentiability of the inverse function. Relation (2.16) gives the equivalent result for $CVaR$. For more general distribution functions the result might not hold true for the VaR risk measure. Using $CVaR$ as a risk measure, Rockafellar and Uryasev show ([53], p. 1458) that even for discrete distributions $CVaR_\zeta(x)$ is a continuous function of ζ which exhibits left and right derivatives.

2.5 Some Comments on Stable / G-and-H Distributed Asset Returns

Some of our results are heavily dependant on the usage of stable respectively g-and-h distributions. Whereas stable distributions admit the favorable property of being closed under taking the sum, hence allowing for explicit representations in terms of the defining parameters, g-and-h distributions exhibit tremendous flexibility in matching arbitrary distributions. Moreover, the VaR for g-and-h distributed asset returns appears to be given by an explicit formula. Because of the importance to our results, we will give a short overview on the properties of these two classes of distribution functions.

2.5.1 Properties of Stable Distributed Asset Returns

Within this paragraph we want to consider the class of stable probability distributions which can be interpreted as a generalization of the normal law allowing for skewness and heavy tails. This class was originally introduced by Paul Levy in the 1920s in his study of sums of i.i.d. terms. Since there is empirical evidence for skewness and/or kurtosis within finance and economic modelling this class of distributions is nowadays often used in the modelling of financial data. For a list of research on this topic we refer to the citations in [41]. Another reason for the usage of stable distributions is the Generalized Central Limit Theorem which states that the only possible non-trivial limit of normalized sums of i.i.d. distributed terms is again stable. Since it is argued that e.g. the price of stocks is the sum of many small terms consequently a stable model should be used to model this data.

Recently, there was a huge development in the area of making stable distributions more accessible to practitioners. As an example, several methods of estimating stable parameters, such as maximum likelihood, quantiles, empirical characteristic functions and fractional moments were developed. Also

a reason for the widespread distribution of this special kind of distribution function is justified by the development of reliable and fast methods to compute stable densities, distribution functions and quantiles. Moreover, as we will see in the next paragraph there were developed fast methods to generate stable random variables.

The class of stable probability distributions encloses some very nice theoretical properties. However it is not our intention to give a full reference to all these properties. We will therefore restrict ourselves to those selected properties that are directly related to portfolio allocation and to our desire to find the optimal solution of the VaR optimization problem.

Every stable distributed random variable is uniquely defined by four parameters: an index of stability or characteristic exponent $\alpha \in (0, 2]$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$ and a location parameter $\delta \in \mathbb{R}$. Although there are several different representations used to describe stable distributions we are concentrating on the two most common ones. Moreover, within this thesis we will focus on the 1-parametrization (as in the notation of Nolan in [41]) for the reason that using this representation the location parameter δ corresponds to the mean value (for $\alpha > 1$). In order to implement efficient VaR and $CVaR$ calculators we mainly use the representations of the distribution function as given in [38] where the 0-parametrization is used. To compare the results with the 1-parametrization we will have to adjust the values for δ accordingly. The respective transformation functions are given below.

Definition 2.5.1 *For $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$ and $\delta \in \mathbb{R}$ a stable random variable X in the 0-parametrization, $X \in S_\alpha(\beta, \gamma, \delta; 0)$, is given by the characteristic function*

$$\mathbb{E}\exp(iuX) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 + i\beta \left(\tan\frac{\pi\alpha}{2}\right)(sign\ u)\left(|\gamma u|^{1-\alpha} - 1\right)\right] + i\delta u\right) & \text{if } \alpha \neq 1 \\ \exp\left(-\gamma |u| \left[1 + i\beta\frac{2}{\pi}(sign\ u)\log(\gamma |u|)\right] + i\delta u\right) & \text{if } \alpha = 1. \end{cases}$$

Remark 2.5.1 *Within this paper we will only consider distributions that allow for the existence of a mean value. For stable distributions this is only the case for the index of stability being greater than one. However, for completeness we will give the definitions in more general terms by allowing α to be an element of the interval $(0, 2]$. It is interesting to note that some well known distribution functions can be seen to be stable. For example, it can be shown that the case $\alpha = 2$ corresponds to the normal distribution, $\alpha = 1$ results in*

some Cauchy distribution whereas the choice $\alpha = \frac{1}{2}$ *represents Levy stable distributions.*

We often handle standardized distributions, i.e. $\gamma = 1$ and $\delta = 0$. In this case we will use $S_\alpha(\beta; 0)$ as an abbreviation of $S_\alpha(\beta, 1, 0; 0)$. Further on we will also use the notation

$$VaR_\eta(\beta(x))$$

to describe the VaR of a standardized random variable $X \backsim S_\alpha(\beta; 0)$. Similar notations are used for $CVaR$ as well as for the 1-parametrization as described next.

Definition 2.5.2 *For* $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$ *and* $\delta \in \mathbb{R}$ *a stable random variable* X *in the* 1*-parametrization,* $X \in S_\alpha(\beta, \gamma, \delta; 1)$*, is given by the characteristic function*

$$\mathbb{E}\exp(iuX) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 - i\beta\left(\tan\frac{\pi\alpha}{2}\right)(sign\ u)\right] + i\delta u\right) & \text{if } \alpha \neq 1 \\ \exp\left(-\gamma |u| \left[1 + i\beta\frac{2}{\pi}(sign\ u)\log(|u|)\right] + i\delta u\right) & \text{if } \alpha = 1. \end{cases}$$

If not explicitly mentioned the 1-parametrization will be used denoted by $S_\alpha(\beta, \gamma, \delta)$ and $S_\alpha(\beta)$ for arbitrary and standardized stable distributions, respectively. Within the two parametrizations described above, α, β and the scale γ are always the same; the location parameter however admits different values for the same stable distributed random variable $X \backsim S_\alpha(\beta, \gamma, \delta_k; k)$ for $k = 0, 1$. Writing δ_0 and δ_1 for the location parameter of the respective parametrization, one easily sees that

$$\delta_0 = \begin{cases} \delta_1 + \beta\gamma\tan\frac{\pi\alpha}{2}, \text{ for } \alpha \neq 1 \\ \delta_1 + \beta\frac{2}{\pi}\gamma\log\gamma, \text{ for } \alpha = 1 \end{cases} \qquad \delta_1 = \begin{cases} \delta_0 - \beta\gamma\tan\frac{\pi\alpha}{2}, \text{ for } \alpha \neq 1 \\ \delta_0 - \beta\frac{2}{\pi}\gamma\log\gamma, \text{ for } \alpha = 1. \end{cases}$$

Therefore it is quite easy to transform between the two parametrizations. One of the most important properties of stable distributed returns is the fact that any linear combination of α-stable distributed r.v. is again α-stable distributed. Hence for some arbitrary portfolio composition $x \in \mathbb{X}$, the resulting distribution function can be exactly described and analyzed by the behavior of the four parameters. The next theorem gives us some first insight to the behavior of these parameters as functions of the portfolio composition x.

Theorem 2.5.1 *Let $X_i \backsim S_\alpha(\beta_i, \gamma_i, \delta_i; k)$, $k = 0, 1$, be independently distributed random variables and let $x = (x_1, ..., x_n)^t \in \mathbb{X}$ describe some feasible allocation. Then*

$$X := x_1 \cdot X_1 + ... + x_n \cdot X_n \backsim S_\alpha(\beta(x), \gamma(x), \delta(x); k).$$

Moreover, the parameters β, γ and δ are explicitly given by

$$\begin{aligned}
\gamma(x)^\alpha = & \quad \sum_{i=1}^n |x_i\gamma_i|^\alpha \\
\beta(x) = & \quad \frac{\sum_{i=1}^n \beta_i (sign x_i) \, |x_i\gamma_i|^\alpha}{\gamma^\alpha} \\
\delta(x) = & \quad \begin{cases}
\sum_i x_i\delta_i + \tan\frac{\pi\alpha}{2}\left(\beta\gamma - \sum_i \beta_i x_i \gamma_i\right) & \text{if } k = 0,\ \alpha \neq 1 \\
\sum_i x_i\delta_i + \frac{2}{\pi}\left(\beta\gamma\log\gamma - \sum_i \beta_i x_i\gamma_i \log|x_i\gamma_i|\right) & \text{if } k = 0,\ \alpha = 1 \\
\sum_i x_i\delta_i & \text{if } k = 1,\ \alpha \neq 1 \\
\sum_i x_i\delta_i - \frac{2}{\pi}\sum_i \beta_i x_i\gamma_i \log|x_i| & \text{if } k = 1,\ \alpha = 1.
\end{cases}
\end{aligned} \tag{2.19}$$

Proof: For a proof of this result we refer the interested reader to [54], [41]. □

In order to introduce some dependence structure among the different asset allocations it is practical to consider multivariate stable distributions. If we denote by

$$\mathbb{S} := \{u \in \mathbb{R}^n : \|u\| = 1\} \subset \mathbb{R}^n$$

the unit sphere in $\mathbb{R}^n$, then Feldheim (in [19]) showed that any stable random vector can be characterized by a spectral measure Λ (a finite measure on $\mathbb{S}$) and a shift vector $\delta \in \mathbb{R}^n$. Any stable random vector $\mathbf{X} \backsim S_\alpha(\Lambda, \delta)$ can therefore be represented as

$$\mathbb{E}\exp\left(i\left\langle u, \mathbf{X}\right\rangle\right) = \begin{cases}
\exp\left(\begin{array}{c} -\int_{\mathbb{S}}\left[|\langle u,s\rangle|^\alpha - i(\operatorname{sign}\langle u,s\rangle)\tan\frac{\pi\alpha}{2}|\langle u,s\rangle|^\alpha\right]\cdot \\ \Lambda(ds) + i\langle u,\delta\rangle \end{array}\right) & \text{if } \alpha \neq 1 \\
\exp\left(\begin{array}{c} -\int_{\mathbb{S}}\left[|\langle u,s\rangle|^\alpha - i(\operatorname{sign}\langle u,s\rangle)\tan\frac{\pi\alpha}{2}|\langle u,s\rangle|^\alpha\right]\cdot \\ \Lambda(ds) + i\langle u,\delta\rangle \end{array}\right) & \text{if } \alpha = 1
\end{cases}$$

Alike in the univariate case of stable distributions an equivalent way of representing any stable distributed random vector is given by using projections. This uses the fact that for any vector x, the projection $\langle x, \mathbf{X}\rangle$ is univariate

α-stable with some skewness $\beta(x)$, some scale $\gamma(x)$ and some shift $\delta(x)$.We will write $\mathbf{X} \backsim S_\alpha(\beta(.), \gamma(.), \delta(.))$ if $\mathbf{X}$ is stable with

$$\langle x, \mathbf{X} \rangle \backsim S_\alpha(\beta(x), \gamma(x), \delta(x))$$

for every $x \in \mathbb{R}^n$ (again implicitly using the 1-parametrization). Note that the spectral measure determines the projection parameter functions by

$$\begin{aligned} \gamma^\alpha(x) &= \int_{\mathbb{S}} |\langle x, s \rangle|^\alpha \, \Lambda(ds) \\ \beta(x) &= \frac{\int_{\mathbb{S}} |\langle x, s \rangle|^\alpha \operatorname{sign}(\langle x, s \rangle) \, \Lambda(ds)}{\gamma(x)^\alpha} \\ \delta(x) &= \begin{cases} \langle x, \delta \rangle & \alpha \neq 1 \\ \langle x, \delta \rangle - \frac{2}{\pi} \int_{\mathbb{S}} \langle x, s \rangle \log |\langle x, s \rangle| \, \Lambda(ds) & \alpha = 1 \end{cases} \end{aligned} \tag{2.20}$$

There exists no explicit formula for the density function nor the cumulative distribution function of an α-stable random vector. However, it can be shown that any nondegenerate (i.e. $\gamma \neq 0$) stable distribution is a continuous distribution with an infinitely differentiable density function (compare [41]). In the univariate as well as in the multivariate case the authors Abdul-Hamid and Nolan ([1]) give integral representations of the density functions.

In the further proceeding we will implicitly use the differentiability of VaR and $CVaR$ as a function of the portfolio allocation. That this is indeed well-defined for stable distributions is the content of the next remark.

Remark 2.5.2 *Suppose $(X_1, ..., X_n)^t$ is a stable distributed random vector. Then, by [1] there exists a multidimensional continuous density function*

$$p : \mathbb{R}^n \to [0, \infty[.$$

In particular, p is a bounded function, such that for fixed $x_2, ..., x_n$ the function

$$\begin{aligned} \widetilde{p} : \mathbb{R} &\to [0, \infty[\\ t &\longmapsto p(t, x_1, ..., x_n) \end{aligned}$$

is continuous and for every $u \in \mathbb{R} \backslash \{0\} \times \mathbb{R}^{n-1}$

$$\mathbb{E}\left[p\left(u_1^{-1} \cdot \left(VaR_\eta(u) - \sum_{i=2}^n u_i X_i \right), X_2, ..., X_n \right) \right] > 0.$$

Moreover, by restricting oneself to α-stable random vectors with $\alpha > 1$, it is well known, that for $i = 1, ..., n$

$$\mathbb{E}\left[|X_i|\right] < \infty.$$

Hence, Assumption S is satisfied for α -stable random vectors ($\alpha > 1$) (cf. [59], p.17 following the discussion on special situation where Assumption S is satisfied) and thus (cf. chapter 2.4.4) $VaR_\eta(x)$ and $CVaR_\zeta(x)$ are partially differentiable functions of x.

2.5.2 Generation of Stable Distributed Random Variables

If we want to investigate either the VaR or the $CVaR$ structure of stable distributed portfolio allocations it is often helpful to generate α-stable distributed random variables. By generating a large set of corresponding scenarios it is then relatively easy to get a rough idea of the VaR respectively the $CVaR$ behavior for different allocations. Although it does not seem obvious how to generate stable distributed random variables, Chambers et al. give an easy way of generating those r.v. based on a nonlinear transformation of two independent uniform respectively exponential distributed random variables.

Theorem 2.5.2 *Let Θ and W be independent with Θ uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and W exponentially distributed with $\mathbb{E}(W) = 1$ and let $0 < \alpha \leq 2$.*

(a) The symmetric random variable

$$Z = \begin{cases} \frac{\sin \alpha\Theta}{(\cos)^{1/\alpha}} \left[\frac{\cos(\alpha-1)\Theta}{W}\right]^{(1-\alpha)/\alpha} & \alpha \neq 1 \\ \tan\Theta & \alpha = 1 \end{cases} \tag{2.21}$$

has a $S_\alpha(0;0) = S_\alpha(0;1)$ distribution.

(b) In the nonsymmetric case, for any $-1 \leq \beta \leq 1$, define

$$\Theta_0 = \arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)/\alpha$$

when $\alpha \neq 1$. Then

$$Z = \begin{cases} \frac{\sin\alpha(\Theta_0+\Theta)}{(\cos\alpha\Theta_0\cos\Theta)^{1/\alpha}} \left[\frac{\cos(\alpha\Theta_0+(1-\alpha)\Theta)}{W}\right]^{(1-\alpha)/\alpha} & \alpha \neq 1 \\ \frac{2}{\pi}\left[\left(\frac{\pi}{2}+\beta\Theta\right)\tan\Theta - \beta\log\left(\frac{\frac{\pi}{2}W\cos\Theta}{\frac{\pi}{2}+\beta\Theta}\right)\right] & \alpha = 1 \end{cases} \tag{2.22}$$

has a $S_\alpha(\beta;1)$ *distribution.*

Proof: A proof of this result can be found in [8]. □

To simulate stable random variables X with arbitrary scale and location parameters γ, δ one has to apply the following transformations

$$X = \begin{cases} \gamma\left(Z - \beta\tan\frac{\pi\alpha}{2}\right) + \delta & \alpha \neq 1 \\ \gamma Z + \delta & \alpha = 1 \end{cases} \tag{2.23}$$

in the 0-parametrization, whereas in the 1-parametrization the necessary transformations look like

$$X = \begin{cases} \gamma Z + \delta & \alpha \neq 1 \\ \gamma Z + \left(\delta + \beta\frac{2}{\pi}\gamma\log\gamma\right) & \alpha = 1. \end{cases} \tag{2.24}$$

2.5.3 VaR Computation in the α-stable Case

In principle, to perform the VaR calculation for an arbitrary confidence level, it is possible to use equations (2.21), (2.22) to generate a set of random variables, sort the corresponding vector and take the entry that corresponds to the α-quantile. However, if we are dealing with very heavy tailed distributions as in the case of α near 1, especially for higher confidence levels the number of scenarios generated to yield an accurate value for the VaR estimate will have to be very high. To compass these difficulties with modeling the heavy tail directly, the so called technique of Importance Sampling was developed (see [57], [7]). However, in our context it will be a more efficient approach to calculate the corresponding distribution function and use this to determine the VaR. Nolan in [38] gives formulas for the density and distribution function of stable random variables that only involve the numerical evaluation of one integral over a finite interval. Since in the present context we are only interested in the formulas for the distribution function, we will not state the corresponding formulas for the density functions. For a numerical implementation one should take into account the discussion on numerical considerations in [38], page 766.

The below stated formulas only account for normalized distribution functions. To transform the respective distribution function to the case of arbitrary scale and location parameters equations (2.23) or (2.24) should be applied.

Before we can state the exact formulas we will have to introduce some further notations. Define

$$\begin{aligned}
\zeta &= \zeta(\alpha,\beta) = \begin{cases} -\beta\tan\frac{\pi\alpha}{2} & \alpha \neq 1 \\ 0 & \alpha = 1 \end{cases} \\
\theta_0 &= \theta_0(\alpha,\beta) = \begin{cases} \frac{1}{\alpha}\arctan\left(\beta\tan\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \frac{\pi}{2} & \alpha = 1 \end{cases} \\
c_1(\alpha,\beta) &= \begin{cases} \frac{1}{\pi}\left(\frac{\pi}{2}-\theta_0\right) & \alpha < 1 \\ 0 & \alpha = 1 \\ 1 & \alpha > 1 \end{cases} \\
V(\theta;\alpha,\beta) &= \begin{cases} (\cos\alpha\theta_0)^{\frac{1}{\alpha-1}}\left(\frac{\cos\theta}{\sin\alpha(\theta_0+\theta)}\right)^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha\theta_0+(\alpha-1)\theta)}{\cos\theta} & \alpha \neq 1 \\ \frac{2}{\pi}\left(\frac{\frac{\pi}{2}+\beta\theta}{\cos\theta}\right)\exp\left(\frac{1}{\beta}\left(\frac{\pi}{2}+\beta\theta\right)\tan\theta\right) & \alpha = 1, \beta \neq 0. \end{cases}
\end{aligned}$$

Theorem 2.5.3 *Let X be stable distributed with characteristic function in the 0- or the 1-parametrization with $\gamma = 1$, $\delta = 0$. The distribution function of X is then given by:*

(a) When $\alpha \neq 1$ and $t > \zeta$,

$$F(t;\alpha,\beta) = c_1(\alpha,\beta) + \frac{sign(1-\alpha)}{\pi}\int_{-\theta_0}^{\frac{\pi}{2}} \exp\left(-(t-\zeta)^{\frac{\alpha}{\alpha-1}}V(\theta;\alpha,\beta)\right)d\theta.$$

(b) When $\alpha \neq 1$ and $t = \zeta$,

$$F(\zeta;\alpha,\beta) = \frac{1}{\pi}\left(\frac{\pi}{2}-\theta_0\right).$$

(c) When $\alpha \neq 1$ and $t < \zeta$,

$$F(t;\alpha,\beta) = 1 - F(-t;\alpha,-\beta).$$

(d) When $\alpha = 1$,

$$F(t;1,\beta) = \begin{cases} \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp\left(-e^{-\frac{\pi x}{2\beta}}V(\theta;1,\beta)\right)d\theta & \beta > 0 \\ \frac{1}{2}+\frac{1}{\pi}\arctan t & \beta = 0 \\ 1 - F(t;\alpha,-\beta) & \beta < 0. \end{cases}$$

Proof: A proof of these formulas is given in [38]. □

Remark 2.5.3 *Using the explicit form for the distribution function of stable distribution functions $F(t;\alpha,\beta)$ in the theorem above, it is easy to see, that $F(t;\alpha,\beta)$ is continuously differentiable with respect to t. By the well-known formula for the derivative of the inverse function and the property*

$$VaR_\eta(\alpha,\beta) = F^{-1}_{\alpha,\beta}(\eta),$$

we also know about the continuous differentiability of the function $VaR_\eta(\alpha,\beta)$ with respect to η. Note, that this property also holds for $CVaR_\zeta(\alpha,\beta)$ and all $\zeta \in (0,1)$. We will use this fact to analyse Assumption A^{ST} in Chapter 3.2.3 numerically.

2.5.4 CVaR Computation in the α-stable Case

In the last section we suggested to generate some large number of α-stable random variables in order to estimate the corresponding VaR. Whereas this proceeding only yields acceptable results for α sufficiently high or an appropriate number of random variables, one could try to do a similar calculation to also estimate $CVaR$. Putting all the generated random variables into one vector, sorting and taking the mean over the $(1-\zeta)\%$ highest entries would by the Law of Large Numbers directly result in $CVaR_\zeta$. However, due to the heavy tailedness of α-stable distributed r.v. an accurate estimation of the $CVaR$ would afford an even higher number of scenario generations than compared to the VaR calculation. In fact, this phenomenon is one of the drawbacks of using $CVaR$ as a risk measure in finance. Since we are taking the mean value over heavy tailed securities the modelling of the tails becomes crucial in estimating the $CVaR$. One way out of this dilemma could be to assume polynomially decreasing tails followed by estimating the parameters of this polynomial. For example, in Extreme Value Theory (EVT), a generalized Pareto model is used to match the tail. If we take this procedure to be a sufficient one it is however not entirely clear how this tail dependence moves with varying portfolio allocations.

Another way of getting around the above mentioned disadvantages of evaluating $CVaR$ accurately could be the usage of the following connection (or similar ones involving the density function of stable distributions) between VaR and $CVaR$

$$CVaR_\zeta(Y) = \frac{1}{1-\zeta}\int_\zeta^1 VaR_\varepsilon(Y)d\varepsilon.$$

This again would imply the repeated numerical calculation of the VaR for different confidence levels. All in all this would result in numerically evaluating a double integral, which is not very efficient. Stoyanov et al. in [58] develop an integral representation that only involves one integral and that can be used to efficiently evaluate the $CVaR_\zeta$ for high confidence levels $\zeta \in (0,1)$. The exact formula will be stated within the next theorem.

As in the proceeding cases, the following theorem only states its result for standardized α-stable distributed r.v. Y in the 1-parametrization. To get the respective expression for r.v. with arbitrary scale and location parameters γ and δ one simply uses

$$CVaR_\zeta(\gamma Y + \delta) = \gamma CVaR_\zeta(Y) + \delta.$$

Theorem 2.5.4 *Let $Y \in S_\alpha(\beta, 1, 0)$ with $\alpha > 1$.*

(a) If $VaR_\zeta(Y) \neq 0$, then the expected shortfall of Y at confidence level ζ admits the following integral representation

$$CVaR_\zeta(Y) = \frac{\alpha}{1-\alpha}\frac{|VaR_\zeta(Y)|}{\pi(1-\zeta)}\int_{-\overline{\theta}_0}^{\frac{\pi}{2}} g(\theta)\exp\left(-|VaR_\zeta(Y)|^{\frac{\alpha}{\alpha-1}}v(\theta)\right)d\theta$$

where

$$g(\theta) = \frac{\sin\left(\alpha\left(\overline{\theta}_0+\theta\right)-2\theta\right)}{\sin\alpha\left(\overline{\theta}_0+\theta\right)} - \frac{\alpha\cos^2\theta}{\sin^2\alpha\left(\overline{\theta}_0+\theta\right)},$$

$\overline{\theta}_0 := \theta_0(\alpha, \overline{\beta})$, $\overline{\beta} := -sign(VaR_\zeta(Y))\beta$ and $V(\theta;\alpha,\beta)$ have the same meaning as in the relay to theorem 2.5.3.

(b) If $VaR_\zeta(Y) = 0$, then

$$CVaR_\zeta(Y) = \frac{2\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{(\pi - 2\theta_0)}\frac{\cos\theta_0}{(\cos\alpha\theta_0)^{1/\alpha}}.$$

Proof: A proof of this result and a discussion on some special cases can be found in [58]. □

2.5.5 Properties of the G-and-H Distribution

The g-, h- and g-and-h distributions are defined in terms of quantiles and were first introduced by J. W. Tuckey in [60] (c.f. also [26] and [27]). Also a good overview and several applications are given in [14] and [12]. Badrinath and Chatterjee ([5], [6]) and Mills ([36]) used the g-and-h distribution to model the return on an equity index as well as the return on equity in various markets. Dutta and Babbel ([16], [17]) used the g-and-h distribution to model interest rates and interest rate options. For a multivariate extension of this flexible class of distributions compare [20].

Although already being described in 1960 and despite their practical attractiveness of simplicity and flexibility in handling skewness as well as kurtosis, the various g-and-h distributions did not find their way through to financial modelling. In fact, until recently, the g-and-h distributions received only little interest in financial literature, which might in part stem out of the fact that fat tailedness is not of such interest in traditional areas of finance as it is in risk measurement.

We first give the formal definition of a g-and-h distribution.

Definition 2.5.3 *Let $Z \sim \mathcal{N}(0,1)$ be a standard normal random variable. A random variable X is said to be g-and-h distributed, denoted $X \in GH$, with parameters $a, b, g, h \in \mathbb{R}$, if X satisfies*

$$X \sim a + b \cdot Y, \tag{2.25}$$

with

$$Y \sim \frac{e^{gZ} - 1}{g} e^{hZ^2/2}. \tag{2.26}$$

a, b are the location and scale parameters, respectively.

Within Definition 2.26, the case $g = 0$ is naturally interpreted as

$$Y \sim Z e^{hZ^2/2}$$

and called the h-distribution. In analogy, the case $h = 0$ will be referred to as the g-distribution. Moreover, X following a g-distribution can be interpreted to be a scaled lognormal random variable. Although there are generalizations

of Definition 2.26 to the case, where both g and h are polynomials in Z (c.f. [15]), we will firstly restrict ourselves to the case of constant parameter values g and h. However, in order to perform the matching of both VaR and $CVaR$, we will define g and h to be differentiable functions of the portfolio allocation $x \in \mathbb{X}$.

Definition 2.26 suggests the class of g-and-h distributions to be an extraordinary flexible class in the sense of being a good match to other families of distribution functions used in the financial industry. In fact, Martinez and Iglewicz showed the g-and-h distribution to properly match such a wide variety of distribution functions as the uniform, exponential or Student-t distributions, to mention only a few. This is mainly due to the fact that the location and scale parameters in (2.25) can easily be used to match expected value and standard deviation of some arbitrary distribution. Moreover, it is easy to see, that parameters g and h can be used to account for skewness and kurtosis, respectively. To be more precise, for $g > 0$, increased g results in greater skewness to the right, whereas $g < 0$ results in skewness to the left. Since, by definition the case $g = h = 0$ reduces to a normal distributed random variable and successively incorporates heavier tails as h is increased (for fixed $g = 0$), also the g-and-h distribution (for both g and $h > 0$) has heavier tails than the normal.

Another appealing feature of a g-and-h distributed random variable X is the fact, that the VaR can be given in explicit terms as a function of the standard normal quantile. This property is stated in the next theorem.

Theorem 2.5.5 *Suppose $h > 0$, $\eta \in (0,1)$ and Φ denotes the standard normal distribution function. Then*

$$VaR_\eta(Y) = \frac{e^{g\Phi^{-1}(\eta)} - 1}{g} e^{h(\Phi^{-1}(\eta))^2/2}. \tag{2.27}$$

Proof: For $h > 0$, the function

$$k(x) = \frac{e^{gx} - 1}{g} e^{hx^2/2}$$

is strictly increasing. Therefore, the distribution function $F(x)$ of a g-and-h random variable can be written as

$$F(x) = \Phi(k^{-1}(x)).$$

The result of the theorem follows immediately. □

Remark 2.5.4 *There are mainly two reasons for restricting ourselves to the parameter space $h > 0$. The first reason is driven by the desire to match financial motivated distribution functions. In general, these can be seen to inhere leptocurtic behavior and we have noted above that it is exactly $h > 0$, that gives us the possibility to match heavy tailed distribution functions. On the other hand, there is also a rigorous mathematical reason for choosing h to be positive since otherwise the function $k(x)$ defined in the proof to the theorem can be shown not to be one to one but two to one. As a consequence, $VaR_\eta(Y)$ cannot be stated in an analogous explicit way as in (2.27). More subtle numerical methods will have to be applied (c.f. [60]).*

In Chapter 3.2.4, we will use the location and scale parameters to match normalized random variables. In order to do so, we first state the expressions of the centralized moments of an arbitrary g-and-h distributed random variable X.

Theorem 2.5.6 *Suppose $h < \frac{1}{n}$. The n-th central moment of a g-and-h distributed random variable Y is then given by*

$$\mathbb{E}(Y^n) = \frac{1}{g^n(1-nh)^{1/2}} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \exp\left[\frac{((n-i)\,g)^2}{(2\,(1-nh))}\right], \text{ if } g \neq 0,$$

$$\mathbb{E}(Y^n) = \begin{cases} 0 & \text{for } n \text{ odd} \\ \frac{n!}{2^{n/2}(\frac{n}{2})!}(1-nh)^{-(n+1)/2} & \text{for } n \text{ even} \end{cases}, \text{ if } g = 0. \tag{2.28}$$

Proof: *For a proof of this result see [60].* □

Corollary 2.5.7 *Using the result of the theorem, the expected value and the variance of a g-and-h distributed random variable can be seen to have the expressions*

$$\mathbb{E}(Y) = \frac{1}{g(1-h)^{1/2}}\left[\exp\left(\frac{g^2}{2(1-h)}\right) - 1\right], \text{ for } h < 1,$$

$$Var(Y) = \frac{1}{g^2(1-2h)^{1/2}}\left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1\right] - \mathbb{E}(Y),$$
$$\text{for } h < \tfrac{1}{2}.$$

Remark 2.5.5 *In the notation of Definition 2.5.3, in what follows, we mostly want the g-and-h distribution X to be normalized, i.e. $\mathbb{E}(X) = 0$ and $\sigma(X) = 1$ for various values g, h. This can be attained by using*

$$\begin{aligned} a &= -\frac{\mathbb{E}(Y)}{\sigma(Y)} \\ b &= \frac{1}{\sigma(Y)}, \end{aligned}$$

where Y denotes the above stated g-and-h distributed random variable.

To get a first impression of what the g-and-h density function for selected parameter values g, h looks like, we plot in figure 2.1 the density functions of the standardized normal distribution (solid line), the g-and-h density function to the parameter values given in the title of the respective subplot (dashed line) as well as the normalized g-and-h density function (dotted line) as described above. Using these plots it is interesting to notice the influence of the parameter g, which in both cases (the original g-and-h transformed density function as well as the normalized one) operates as skewness parameter. Notice also the symmetry by changing the sign of the skewness parameter. Higher values for h result in density functions with heavier tails. Normalizing mean and standard deviation of the respective g-and-h distribution does not change the values for skewness and kurtosis. Since higher values for h result in higher standard deviations (with the limiting case $\sigma \to \infty$ as $h \to 0.5$), the normalized density functions exhibit much more peaked behavior as can especially be seen in the lower subplots of Figure 2.1.

2.5.6 Subadditivity of VaR

Having stated the subadditivity property of $CVaR$ for continuous as well as for discrete distribution functions in Section 2.4.3 we will have a closer look on the subadditivity property of VaR for selected classes of distribution functions. In the ongoing sections, we will see that the class of elliptical distributions is subadditive for $\eta > \frac{1}{2}$ which is mainly due to its symmetry property. Moreover, for stable distributions with constant skewness parameter β Garcia et al. (see [22], Corollary 3.2) find sufficient conditions on the VaR confidence level such that the subadditivity property is fulfilled. We will derive this result within an even simpler setting. Suppose the two independently stable distributed r.v. X_1, X_2 admit the same skewness parameter

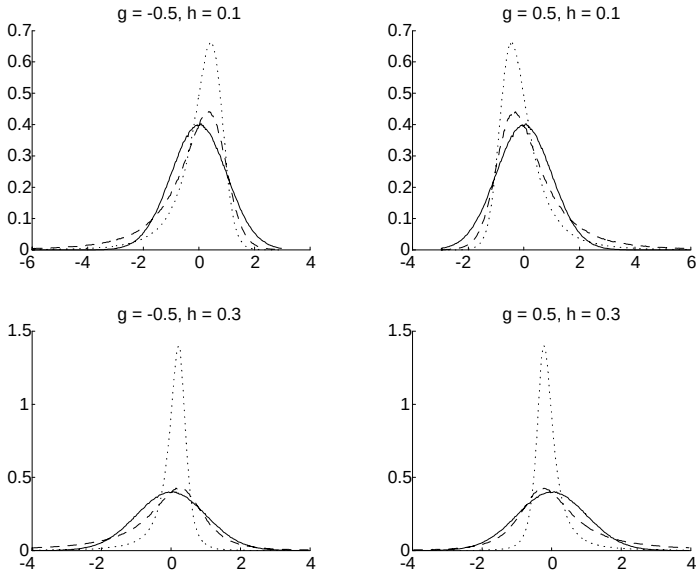

Figure 2.1: Density functions of the standard normal distribution (solid line), the g-and-h density function (dashed line) and the normalized g-and-h density function (dotted line) for several parameter values.

β. Then Theorem 2.5.1 shows that $X_1 + X_2$ is again stable distributed with skewness β and the claimed subadditivity follows from

$$\begin{aligned} VaR_\eta(X_1 + X_2) &= (\mu_1 + \mu_2) + (\gamma_1^\alpha + \gamma_2^\alpha)^{\frac{1}{\alpha}} \cdot VaR_\eta(\beta) \\ VaR_\eta(X_1) + VaR_\eta(X_2) &= (\mu_1 + \mu_2) + (\gamma_1 + \gamma_2) \cdot VaR_\eta(\beta). \end{aligned}$$

where we used μ_1, μ_2 for the respective mean and γ_1, γ_2 for the respective scale parameter. Hence, by

$$(\gamma_1 + \gamma_2) > (\gamma_1^\alpha + \gamma_2^\alpha)^{\frac{1}{\alpha}}$$

for $\alpha > 1$ we have

$$VaR_\eta(X_1 + X_2) \begin{cases} < VaR_\eta(X_1) + VaR_\eta(X_2) & \text{if } VaR_\eta(\beta) > 0 \\ = VaR_\eta(X_1) + VaR_\eta(X_2) & \text{if } VaR_\eta(\beta) = 0 \\ > VaR_\eta(X_1) + VaR_\eta(X_2) & \text{if } VaR_\eta(\beta) < 0 \end{cases}.$$

To get an impression of those confidence levels, for that VaR_η is subadditive we plot in Figure 2.2 the function

$$\begin{aligned} h_\alpha : (0,1) &\to \mathbb{R} \\ \beta &\mapsto \eta(\beta), \end{aligned}$$

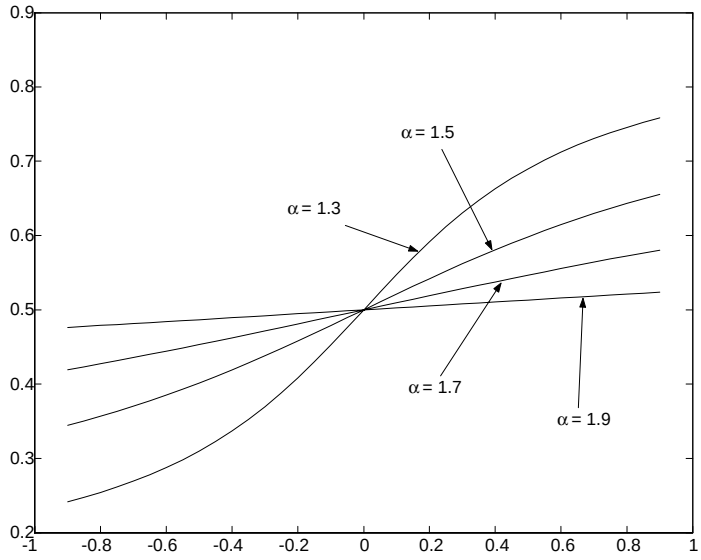

Figure 2.2: Confidence levels η, for which $VaR_\eta(\beta)$ is subadditiv. Considered is the case of α-stable and independantly distributed random variables with the same skewness parameter β. Those confidence levels η, that lie above the corresponding graph are guaranteeing VaR to be subadditive.

$\eta(\beta)$ implicitly defined by the expression $VaR_\eta(\beta) = 0$ and for several values $\alpha \in (1, 2)$. There are several observations we can make. Firstly, as we are increasing α towards $\alpha = 2$ (corresponding to the normal distribution), the graph $\eta(\beta)$ also flattens towards the constant function $\eta(\beta) = \frac{1}{2}$, the corresponding borderline for normal distributions. It is also interesting to note, that by introducing of heavy tails via the stability parameter α, the set of confidence levels, for that the subadditivity property holds is strongly dependant on the skewness parameter $\beta \in (-1, 1)$: for some fixed α, higher skewness parameters also imply the set of "subadditivity levels" to be cut. Lower values of β result in a widening of the respective set of confidence levels.

Clearly, all plots have the point $(0, \frac{1}{2})$ in common. This is due to the fact, that for $\beta = 0$ the resulting distribution is symmetric and has expected value equal to zero, which implies $VaR_{\frac{1}{2}}(0) = 0$.

If dealing with stable distributions in practice, we will usually not deal with the case of constant skewness parameters. For X_1, X_2 admitting selected skewness and scaling parameters, we give in Appendix B those confidence

levels that are the solution of the equation

$$\gamma_{12} VaR_\eta(\beta_{12}) - \gamma_1 VaR_\eta(\beta_1) - \gamma_2 VaR_\eta(\beta_2) = 0.$$

Here, γ_{12} and β_{12} denote the scale respectively the skew of the r.v. $X_1 + X_2$. Hence, for a suitable combination of parameters the table shows the lowest confidence levels, s.t. for all higher levels the corresponding r.v. X_1 and X_2 admit the desired subadditivity property.

As we are also interested in g-and-h distributed random variables one might also ask the question of subadditivity in this special case. Degen et al. ([12]) address this problem and can show that also for g-and-h distributed r.v. the subadditivity property for VaR holds true for sufficiently high confidence levels. Again, those confidence levels strongly depend on the skewness and kurtosis of the corresponding distributions. Moreover, the authors investigate dependence on the dependance structure. For further details see [12].

Chapter 3

VaR and CVaR Optimization

3.1 CVaR Optimization

$CVaR$ does not only have the desirable property of being a coherent measure of risk but also has some nice behavior in its direct numerical implementation. At a first step, using scenarios to represent the random variables involved, the $CVaR$ optimization problem can be easily restated as a linear one. Hence, as we will see, in the present case the problem reduces to the finding of the solution of a large scale linear optimization problem. There are many developments in this direction and recently there are also implementations available, that easily handle 100.000 variables or even more. For the large scale optimization results of this thesis MOSEK, a very powerful optimization toolbox for MATLAB, was used.

Although there are suitable solvers at hand to address the problem directly, in the succeeding chapters we will discuss an application of stochastic programming which allows for a reduction in dimensionality by successively solving growing systems of linear equalities. We start with stating the $CVaR$ optimization as a linear problem.

3.1.1 Linearity of CVaR Optimization

If the analytical form of the portfolio instruments' losses is not available (and in practical applications this is usually the case nor would it be of any help for the implementation), the price of these instruments could be represented

by historical price observations or by Monte Carlo simulations. Using the definition of $CVaR$ as in (2.4) we define

$$F_\zeta(a,x) := a + \frac{1}{1-\zeta}\mathbb{E}\left[X(x)-a\right]^+ \tag{3.1}$$

which can be approximated by the expression

$$\widetilde{F}_\zeta(a,x) := a + \frac{1}{(1-\zeta)J}\sum_{j=1}^{J}\left[\widehat{X}^j(x)-a\right]^+,$$

$\widehat{X}^j(x)$ representing the j-th realization of the random variable $X(x)$ which itself represents the losses of portfolio x. As already discussed in the proceeding chapters, $\widetilde{F}_\zeta(a,x)$ is a convex, nonsmooth function w.r.t. the vector (a,x). Moreover, introducing auxiliary variables z_j, the minimization over the last expression in the notation of (2.6) can be written as

$$\begin{aligned} & \min_{x,a,z} a + \frac{1}{(1-\zeta)J}\sum_{j=1}^{J} z_j \\ \text{s.t.} \quad & x \in \mathbb{X} \\ & z_j \geq 0 \\ & z_j \geq \widehat{X}^j(x) - a \end{aligned} \tag{3.2}$$

In fact, since $\mathbb{X}$ is a convex set, methods of linear programming can be used to solve problem (3.2). It is clear that the dimension of the LP problem increases with both, the number of instruments under consideration as well as the number of scenarios used to match the instruments' behavior. However, besides our own experiences with the implementation of (3.2) several case studies (see e.g. [52], [3]) have this representation identified to be a numerically stable technique on tackling the $CVaR$ optimization problem.

Albeit being a useful and easy to solve formulation, there exists an even more efficient way by using stochastic programming methods. We will introduce this very useful technique in the following chapter.

3.1.2 Reduction of Dimensionality using Benders Decomposition

The description of the following algorithm follows the lines of the original work performed by Kuenzi-Bay and Mayer, where we restrict to the present

setting. For more details on the mathematical background especially on the embedding of the present problem into the more general class of two-stage recourse problems, follow the discussion in ([29]) and the citations therein. There, one also finds some more general setting also accounting for the presence of liabilities.

At a first step, the authors in ([29]) show the equivalence of optimization problem $\min_{x\in\mathbb{X},a\in\mathbb{R}} F_\zeta(a,x)$ as defined in equation (3.1) with

$$\begin{aligned} &\min_{x,a} a + \mathbb{E}\left[Q_c(x,a)\right] \\ &\text{s.t. } x \in \mathbb{X} \end{aligned} \tag{3.3}$$

with the recourse subproblem

$$\begin{aligned} Q_c(x,a) := \frac{1}{1-\zeta} \min_y \quad & y \\ \text{s.t.} \quad & y \geq X(x) - a \\ & y \geq 0 \end{aligned} \tag{3.4}$$

and its LP dual

$$\begin{aligned} Q_c(x,a) = \frac{1}{1-\zeta} \quad & \max_u \left(X(x) - a\right) u \\ & \text{s.t. } 0 \leq u \leq 1, \end{aligned} \tag{3.5}$$

which admits its optimal solution for either $u = 0$ or $u = 1$, depending on the sign of $X(x) - a$. Hence the optimal objective value is $\frac{1}{1-\zeta}\left(X(x) - a\right)^+$. In what follows, we again restrict ourselves to the case where there are concrete realizations $\widehat{X}^j, j = 1, ..., J$, each with the same probability of occurrence, of the random variable X available. Applying the L-shaped method of stochastic programming for the current two stage recourse problem, one makes usage of the solutions u^j of the dual recourse sub-problems corresponding to the realizations $\widehat{X}^j, j = 1, ..., J$. Now the extremely simple structure of the dual problem (3.5) (the solutions are either $u^j = 0$ or $u^j = 1$) can be used to construct the aggregate cuts as appearing in the L-shaped method based on the J-dimensional vector $(u^1, ..., u^J)^t$. Here, this vector is a binary one, which can be identified in a one-to-one manner with a subset of the index-set $\mathcal{J} = \{1, ..., J\}$, where we are choosing those indices j as elements for which $u^j = 1$ holds. Therefore, Künzi-Bay and Mayer obtain the following representation of (3.3), which they call the *full master problem*

$$\min_{x,a,w} a + \frac{1}{1-\zeta} w$$

$$\begin{aligned} &\text{s.t. } w \geq \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \widehat{X}^k(x) - a, \mathcal{K} \subset \mathcal{J} \\ &x \in \mathbb{X}, \end{aligned} \tag{3.6}$$

where the case $\mathcal{K} = \emptyset$ corresponds to w being nonnegative. It is clear that already for a moderate number J of realizations entering the optimization process, the full master problem comprises a huge number of constraints. However, in general, many of these constraints are redundant since the constraint $x \in \mathbb{X}$ forces not all J-dimensional binary vectors to be part of the underlying duality consideration. Now, in the L-shaped method, this fact is used in order to solve a sequence of affine optimization problems, which starting from an almost unconstrained problem successively includes a growing number of constraints. Under a more general framework this approach is known as the Benders Decomposition; the convergence of the corresponding optimization problems within the current setting is shown in ([29]). The successive optimization problems are called *relaxed master problems* and can be stated as follows

$$\begin{aligned} \underline{G}_\nu &: = \min_{x,a} a + \frac{1}{1-\zeta} w \\ \text{s.t. } w &\geq \frac{1}{|\mathcal{K}_i|}\sum_{k\in\mathcal{K}_i} \widehat{X}^k(x) - a, i = 1, ..., \nu \\ w &\geq 0 \\ x &\in \mathbb{X}. \end{aligned} \tag{3.7}$$

Here, ν is the number of constraints generated so far, $\mathcal{K}_i \subset \mathcal{J}$ for all i and $\mathcal{K}_i \neq \mathcal{K}_l$ for $i \neq l$. Based on our assumptions, problem (3.7) admits a solution which is the basis for the next constraint added.

Now we are in the position to formulate the formal algorithm, which resolves the original $CVaR$ optimization problem (2.6) in a finite number of steps. This follows directly from the fact that the algorithm is nothing else than the Benders decomposition method applied to (3.3).

0) Let $\mathcal{K}_1 := \mathcal{J}$ and set $\nu = 1$. The single inequality constraint in the relaxed master problem will be $w \geq \frac{1}{J}\sum_{k\in\mathcal{J}} \widehat{X}^k(x) - a$.

1) Solve the current relaxed master problem (3.7). Let (x^*, a^*, w^*) be an optimal solution, and let

$$\mathcal{K}^* := \left\{ k \mid 1 \leq k \leq J, \widehat{X}^k(x^*) - a^* > 0 \right\} \text{ and}$$

$$w_+^* := \frac{1}{|\mathcal{K}^*|} \sum_{k \in \mathcal{K}^*} \left(\widehat{X}^k(x^*) - a^* \right).$$

2) If $w_+^* \leq w^*$, then x^* is an optimal solution of (2.6). Otherwise continue with the next step.

3) Set $\nu := \nu + 1, \mathcal{K}_\nu = \mathcal{K}^*$ and append the corresponding cut to the set of constraints in the relaxed master problem (3.7). Continue with Step 1.

A discussion on the particularities of implementing this algorithm, in particular the usage of a refined stopping criterion at *Step* 2, can also be found in the aforementioned citation. We implemented the algorithm using Matlab and thereby making the experience that especially for a huge set of scenario representations this proceeding generally gives much more efficient results than using the linear representation of the last section. However, it is not generally clear how many of the relaxed master problems will have to be solved in order to get an optimal solution in the sense of the stopping criterion in *Step* 2. Therefore, it is possible that the algorithm gets stuck in a finitely but huge number of successively to solve optimization problems.

3.2 VaR Optimization for Certain Classes of Asset Returns

Within this chapter we will focus on some selected classes of distribution functions, explicitly showing the exact coincidence of the solution to (2.8) and the "real" VaR_η optimum of the original problem (2.5) to solve. We start considering the case of normally distributed asset returns.

3.2.1 Spherical and Elliptical Distributions

One of the reasons for the widespread use of normally distributed asset returns is its inherent simplicity in describing the complete behavior by two parameter values, mean and standard deviation. Moreover, if we expect the original asset classes to be normally distributed, every linear combination of these normally distributed returns are normally distributed as well. Considering the VaR risk measure in the case of normally distributed asset returns

another nice property of this class of distributions becomes obvious. By standardizing the portfolio distribution in the sense that for every $X(x)$ we write

$$Y = \frac{X(x) - \mu(x)}{\sigma(x)},$$

Y appears to be standard normal distributed with mean zero and standard deviation $\sigma = 1$. Then by the linearity property of the VaR we can write

$$VaR_\eta(X(x)) := VaR_\eta(x) = \mu(x) + \sigma(x) \cdot VaR_\eta(Y).$$

Hence, the standardized VaR is no longer dependant on the portfolio composition. A similar approach gives us an equivalent result for the $CVaR$ risk measure

$$CVaR_\zeta(x) = \mu(x) + \sigma(x) \cdot CVaR_\zeta(Y).$$

Thus using the $CVaR$ confidence level for some fixed VaR confidence level η implicitly defined by

$$VaR_\eta(Y) = CVaR_{\zeta_0}(Y)$$

it is easy to see that the $VaR_\eta(x)$ and the $CVaR_{\zeta_0}(x)$ optimization problems yield the same optimal portfolio. However, this proceeding crucially depends on the existence of ζ_0 as described above. In the case of normally distributed returns the existence of a corresponding ζ_0 is guaranteed for all confidence levels $\eta > \frac{1}{2}$. This is because $VaR_\eta(x) > 0$ for $\eta > \frac{1}{2}$ and

$$\begin{aligned} \lim_{\zeta \to 0} CVaR_\zeta(Y) &= 0 \\ \lim_{\zeta \to 1} CVaR_\zeta(Y) &= \infty. \end{aligned}$$

The existence of a suitable confidence level follows by the continuity of $CVaR_\zeta(Y)$ as a function of ζ.

Within this paragraph we will see that these nice properties can be maintained by expanding this class of distributions to the more general class of elliptically distributed random variables. As an important consequence this very general class of distributions will be seen to also inherit the property of the VaR optimization reducing to some affine $CVaR_\zeta$ optimization.

In order to define elliptically distributed random vectors we will firstly introduce spherical distributions.

Definition 3.2.1 *A n-dimensional random vector Y is said to have a spherical distribution $S_n(\phi)$ if its characteristic function $\psi_Y(s)$ satisfies*

$$\psi_Y(s) = \phi(s^t s)$$

for some scalar function $\phi_n(.)$ which is called the characteristic generator of the spherical distribution $S_n(\phi)$.

Theorem 3.2.1 *Let Y be spherically distributed with characteristic generator $\phi_n(.)$. Then Y can be written as*

$$Y \stackrel{d}{=} R_n U^n,$$

where the random variable $R_n \geq 0$ is independent from the n-dimensional random vector U^n,which is uniformly distributed on the n-dimensional unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$.

A proof of this relationship can be found e.g. in [18]. We are now in the position to define elliptically distributed random vectors.

Definition 3.2.2 *A n-dimensional random vector $\widetilde{Y}$ is said to have an elliptical distribution with parameters $\mu \in \mathbb{R}^{n,1}$ and $\Sigma \in \mathbb{R}^{n.n}$ if $\widetilde{Y}$ has the same distribution as $\mu + \mathcal{A}^t Y$, where $Y \in S_n(\phi)$ and $\mathcal{A} \in \mathbb{R}^{k,n}$ is such that $\mathcal{A}^t\mathcal{A} = \Sigma$ with $rank(\Sigma) = k$. The set of all n-dimensional elliptically distributed random vectors with parameters μ and Σ and characteristic generator ϕ will be denoted by $E_n(\Sigma, \phi, \mu)$.*

As in the case of spherical distributions every random vector $Y \in E_n(\Sigma, \phi, \mu)$ can be represented as

$$Y \stackrel{d}{=} \mu + R_k \mathcal{A}^t U^k,$$

where $\mathcal{A}^t\mathcal{A} = \Sigma$ and the random variable $R_k \geq 0$ is independent from the uniformly distributed k-dimensional r.v. U^k on the unit sphere $S^{k-1} \subset \mathbb{R}^k$. Using this representation of an elliptically distributed random vector it is easy to see that the linear combination of elliptically distributed random variables is again elliptically distributed.

Theorem 3.2.2 *Let $X \backsim E_n(\Sigma, \phi, \mu)$ with $rank(\Sigma) = k$ and $x \in \mathbb{R}^n$ represent some vector of portfolio allocations. Then also*

$$X^t x \backsim E_n(x^t \Sigma x, \phi, \mu^t x).$$

Remark 3.2.1 *A direct consequence of the last theorem is that the $VaR_\eta(Y)$ and the $CVaR_\zeta(Y)$ for any random variable $Y \in E_1(\sigma^2, \phi, \mu)$ can be standardized as in the case of normally distributed random variables. Hence in the notation of the theorem*

$$VaR_\eta(x^t X) = x^t \mu + x^t \Sigma x \cdot VaR_\eta(\widetilde{Y}),$$
$$CVaR_\zeta(x^t X) = x^t \mu + x^t \Sigma x \cdot CVaR_\zeta(\widetilde{Y})$$

where $\widetilde{Y} \in E_1(1, \phi, 0)$. As in the case of normally distributed asset returns we have for $\eta > \frac{1}{2}$, $VaR_\eta(Y) > 0$ wherefore there exists $\zeta_0 \in (0, 1)$, s.t.

$$VaR_\eta(\widetilde{Y}) = CVaR_{\zeta_0}(\widetilde{Y})$$

and the corresponding optimization problems are seen to be equivalent.

Remark 3.2.2 *The family of elliptically distributed random variables are a very rich extension of the family of normally distributed random variables. It also accounts for the set of Student-t distributed random variables with m degrees of freedom. Another important subclass of elliptical distributions is the family of sub-gaussian distributions.*

3.2.2 Stable Distributions with Constant Skewness Parameter

Having described the set of elliptically distributions in the foregoing paragraph we will now partially leave the class of symmetrically distributed random variables in order to investigate constantly skewed stable r.v. in the sense that the main behavior of the r.v. is still given by location and scale parameters whereas the skewness parameter is not necessarily restricted to be zero. Apart of some interesting properties such as incorporating for heavy tailed behavior, one of the crucial properties of stable distributions is the fact that any linear combination of stable distributed r.v. again is stable

distributed. We will firstly restrict ourselves to the case where the constant skewness parameter is of the form $\beta = 0$, i.e. the case of symmetric α-stable ($S\alpha S$) distributions. In order to guarantee the closedness of this property under taking the linear combinations of $S\alpha S$ distributed r.v., one has to impose further restrictions on the dependence structure given by the spectral measure Λ or by assuming independency of the r.v. under consideration (as can easily be seen by theorem 2.19). If we can guarantee that all the allowed linear portfolio combinations are again symmetric this enables us to write

$$\begin{aligned} VaR_\eta(x) &= \mu(x) + \gamma(x) \cdot VaR_\eta(\widetilde{Y}), \\ CVaR_\zeta(x) &= \mu(x) + \gamma(x) \cdot CVaR_\zeta(\widetilde{Y}) \end{aligned} \tag{3.8}$$

with $\widetilde{Y} \in S_\alpha(1, 0, 0)$. Hence for $\eta > \frac{1}{2}$ there exists some value $\zeta_0 \in (0, 1)$ such that the corresponding optimization problems are equivalent.

Remark 3.2.3 *A very important subset of the family of $S\alpha S$ distributions is the family of sub-Gaussian stable distributed random variables. These are characterized as inheriting the property of being stable distributed as well as being elliptically contoured. Thus a sub-Gaussian distributed random variable Y is described by its characteristic function having the form*

$$\mathbb{E}(\exp(i \langle x, X \rangle)) = \exp(-(x^t \Sigma x)^{\alpha/2} + i \langle x, \delta \rangle)$$

for some positive definite matrix Σ and shift vector $\delta \in \mathbb{R}^n$. X describing an arbitrary n-dimensional stable vector. The corresponding projection parameter functions are then given by

$$\gamma(x) = (x^t \Sigma x)^{\frac{1}{2}}, \quad \beta(x) = 0, \quad \delta(x) = \langle x, \delta \rangle \ .$$

A closer description of this family of distributions as well as a justification of its name is e.g. given in [40].

Clearly, we can extend the foregoing result to the case where we are dealing with portfolio compositions that impose a constant skewness parameter β not necessarily being equal to zero. Thus the equivalent to equation 3.8 holds with an $\widetilde{Y} \in S_\alpha(1, \beta, 0)$ for a fixed $\beta \in (-1, 1)$. However, the case of nonzero but constant $\beta \in (0, 1)$ is quite unrealistic to hold for every portfolio composition since even for independent distributed assets with equal skewness parameter β the linear combination will generally result in a portfolio

whose β is unequal to the original stand alone skewness (compare theorem 2.5.1).

Although for nonzero skewness parameters the corresponding $VaR_\eta(x)$ optimization problem seems more complex to solve due to the term $VaR_\eta(\widetilde{Y})$ not being constant in x, within the next paragraph we will develop sufficient conditions that again allow the reduction to some convex $CVaR_{\zeta_0}(x)$ optimization problem.

3.2.3 Stable Distributions

In [44] the authors investigate the problem of optimizing a portfolio with stable distributed asset returns. However, the authors restrict themselves to the two asset classes case, where the investor can only choose between the allocation to a risk-free asset and a stable distributed one. Moreover, risk measurement in this context is given by the expression

$$\mathbb{E}(\,|W - \mathbb{E}(W)|^r\,)$$

for some positive real number r and $W := \lambda X_0 + (1-\lambda)X_1, \lambda \in [0,1]$. Under these assumptions, the authors find the best allocation λ in a mean-risk context and compare the results to the case of normally distributed asset returns.

Within this chapter, we will also consider the optimal portfolio allocation problem for the special case, where the $X_i, i = 1, ..., n$ are stable distributed.

A Polynomial Algorithm for Stable Distributed Asset Returns

We will see, that under the following assumption on the stable distributed asset returns the complexity of the VaR optimization problem can significantly be reduced.

Assumption $\mathbf{A}^{ST}$ For $\eta \in (0,1)$ and $\beta^* \in (-1,1)$ the skewness parameter of the optimal $VaR_\eta(x)$ solution, there exists $\zeta_0 \in (0,1)$, such that

$$\frac{VaR_\eta(\beta^*)}{CVaR_{\zeta_0}(\beta^*)} = \frac{\frac{\partial}{\partial\beta}VaR_\eta(\beta^*)}{\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta^*)} =: c \tag{3.9}$$

with $\frac{\partial}{\partial\beta}VaR_\eta(\beta^*) \neq 0 \neq \frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta^*)$.

Remark 3.2.4 *(i) In Assumption A^{ST} above, we suppressed the subindex c in $CVaR_{\zeta_0,c}(\beta^*)$, since we implicitly set $c = 1$, i.e. we are considering the CVaR optimization problem with normalized $\mu(x) = 0$ and $\sigma(x) = 1$ for all $x \in \mathbb{X}$.*

(ii) It must be secured that all the denominators appearing within 3.2.3 are not zero. However, we will see that slight perturbations will force all denominators to be nonzero.

Theorem 3.2.3 *Suppose Assumption A^{ST} holds and all allocations $x \in \mathbb{X}$ with $\mathbb{X}$ as defined in (2.1) are supposed to be stable distributed. Let*

$$x^*_{\zeta_0,c} = \arg\min_{x\in\mathbb{X}} CVaR_{\zeta_0,c}(x)$$
$$x^*_\eta = \arg\min_{x\in\mathbb{X}} VaR_\eta(x)$$

denote the optimal solutions of the corresponding $CVaR_{\zeta_0,c}$ and VaR_η optimization, then

$$x^*_\eta = x^*_{\zeta_0,c}$$

for (ζ_0, c) as given by Assumption A^{ST}.

Proof: Since x^*_η denotes an optimal solution of the VaR_η optimization problem, it follows by the differentiability of the VaR

$$\begin{aligned}\nabla VaR_\eta(x^*_\eta) &= VaR_\eta(\beta^*)\cdot\nabla\sigma(x^*_\eta) + \sigma(x^*_\eta)\cdot\nabla VaR_\eta(\beta^*) + \nabla\mu(x^*_\eta)\\ &= VaR_\eta(\beta^*)\cdot\nabla\sigma(x^*_\eta) + \sigma(x^*_\eta)\cdot\nabla\beta(x^*_\eta)\frac{\partial}{\partial\beta}VaR_\eta(\beta^*) + \nabla\mu(x^*_\eta)\\ &\stackrel{!}{=} \lambda\cdot\mathbf{1} + \sum_{i=1}^n \mu_i e_i \qquad (3.10)\end{aligned}$$

and

$$\mu_i \cdot x^*_\eta(i) = 0 \ \forall i = 1, ..., n \qquad (3.11)$$

for $\lambda \in \mathbb{R}$, $\mu_i \geq 0$. On the other hand, for the $CVaR$ we can write

$$\begin{aligned}
\nabla CVaR_{\zeta_0,c}(x_\eta^*) &= CVaR_{\zeta_0}(\beta^*) \cdot c \cdot \nabla\sigma(x_\eta^*) \\
&\quad + c \cdot \sigma(x_\eta^*) \cdot \nabla CVaR_{\zeta_0}(\beta^*) + \nabla\mu(x_\eta^*) \\
&= CVaR_{\zeta_0}(\beta^*) \cdot c \cdot \nabla\sigma(x_\eta^*) \\
&\quad + c \cdot \sigma(x_\eta^*) \cdot \nabla\beta(x_\eta^*) \frac{\partial}{\partial\beta} CVaR_\eta(\beta^*) + \nabla\mu(x_\eta^*) \\
&= VaR_\eta(\beta^*) \frac{\frac{\partial}{\partial\beta} CVaR_{\zeta_0}(\beta^*)}{\frac{\partial}{\partial\beta} VaR_\eta(\beta^*)} \cdot c \cdot \nabla\sigma(x_\eta^*) \\
&\quad + c \cdot \sigma(x_\eta^*) \cdot \frac{\partial}{\partial\beta} VaR_\eta(\beta^*) \frac{CVaR_{\zeta_0}(\beta^*)}{VaR_\eta(\beta^*)} \cdot \nabla\beta(x_\eta^*) \\
&\quad + \nabla\mu(x_\eta^*) \\
&= VaR_\eta(\beta^*) \cdot \nabla\sigma(x_\eta^*) + \sigma(x_\eta^*) \cdot \frac{\partial}{\partial\beta} VaR_\eta(\beta^*) \cdot \nabla\beta(x_\eta^*) \\
&\quad + \nabla\mu(x_\eta^*) \\
&= \lambda \cdot \mathbf{1} + \sum_{i=1}^{n} \mu_i e_i, \qquad (3.12)
\end{aligned}$$

and

$$\mu_i \cdot x_\eta^*(i) = 0 \ \forall i = 1, ..., n$$

with the same parameters λ, $\mu_i, i = 1, ..., n$ as given in equations 3.10 and 3.11. We again suppressed the index c in $CVaR(\beta)$, since here $c = 1$ and therefore the notation is redundant. Using convexity of $CVaR_{\zeta_0,c}(x)$ (cf. Lemma 2.4.3) according to the Karush-Kuhn-Tucker conditions, equality (3.12) is sufficient for x_η^* being the global minimum of $CVaR_{\zeta_0,c}(x)$, as required. □

Numerical Evaluations of Assumption A^{ST}

In what follows, we want to consider Assumption A^{ST} numerically. Lemma 3.2.4 is quite helpful in the numerical analysis on Assumption A^{ST}.

Lemma 3.2.4 *Let $VaR_\eta(\beta)$ denote the Value-at-Risk of a normalized stable random variable X, i.e. $X \in S_\alpha(\beta, 1, 0)$. Then for fixed $1 < \alpha \leq 2$*

$$h_\alpha : (-1, 1) \to \mathbb{R}$$

$$\beta \mapsto VaR_\eta(\beta)$$

is continuously differentiable with respect to β. Moreover, in the notation of theorem 2.5.3 (1-parametrization), the derivative is given by

$$\begin{aligned}\frac{\partial}{\partial\beta}VaR_\eta(\beta) &= \frac{\frac{\partial}{\partial\beta}F(VaR_\eta(\beta);\alpha,\beta)}{f(VaR_\eta(\beta);\alpha,\beta)} \\ &\equiv \frac{G(VaR_\eta(\beta);\alpha,\beta)}{f(VaR_\eta(\beta);\alpha,\beta)},\end{aligned}$$

where

$$\begin{aligned}G(t;\alpha,\beta) := \exp\left[-(t)^{\frac{\alpha}{\alpha-1}}\cdot V(\theta_0;\alpha,\beta)\right]\cdot\frac{\tan\left(\frac{\pi\alpha}{2}\right)}{\alpha(1+\beta^2\tan^2(\frac{\pi\alpha}{2}))} \\ -\frac{1}{\pi}\int_{-\theta_0}^{\frac{\pi}{2}}\frac{\partial}{\partial\beta}V(\theta;\alpha,\beta)\exp(-(t)^{\frac{\alpha}{\alpha-1}}\cdot V(\theta;\alpha,\beta))d\theta\end{aligned}$$

and as before

$$\theta_0 := \theta_0(\alpha,\beta) := \frac{1}{\alpha}\arctan(\beta\tan(\frac{\pi\alpha}{2}))$$

$$V(\theta,\alpha,\beta) := (\cos(\alpha\theta_0))^{\frac{1}{\alpha-1}}\left(\frac{\cos(\theta)}{(\sin\alpha)(\theta_0+\theta)}\right)^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha\theta_0+(\alpha-1)\theta)}{\cos\theta}.$$

Proof: Using the explicit form of $F(t;\alpha,\beta)$ as stated in Theorem 2.5.3, we see that the distribution function is continuously differentiable with respect to t and β. Moreover, we know that

$$F(VaR_\eta(\beta);\alpha,\beta) = \eta$$

holds for all $\beta \in (-1,1)$. With

$$\frac{\partial}{\partial t}F(t;\alpha,\beta) = f(t;\alpha,\beta)$$

we are given the density function of a normalized, stable distributed random variable. In particular, it holds

$$\frac{\partial}{\partial t}F(t;\alpha,\beta) > 0 \ \forall t \in \mathbb{R}.$$

Hence, the assumptions of the implicit functions theorem are satisfied and $VaR_\eta(\beta)$ is continuously differentiable with respect to β with derivative

$$\frac{\partial}{\partial\beta}VaR_\eta(\beta) = \frac{\frac{\partial}{\partial\beta}F(VaR_\eta(\beta);\alpha,\beta)}{f(VaR_\eta(\beta);\alpha,\beta)} \ \forall\beta \in (-1,1).$$

□

In order to analyse Equality 3.9, we firstly write it in the form

$$\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta^*)\cdot VaR_\eta(\beta^*) = \frac{\partial}{\partial\beta}VaR_\eta(\beta^*)\cdot CVaR_{\zeta_0}(\beta^*) \tag{3.13}$$

This expression has the advantage that we do not have to care about denominators becoming equal to zero. Using (3.13), we can easily find those values for the VaR confidence levels η, for which there exists an ζ, such that (3.13) is fulfilled. Our results are summarized in Figure 3.1. For four selected values of α, we plot the corresponding borderlines for which we know that for an explicit value $\beta \in (-0.9, 0.9)$[1], all the values η above this line yield a corresponding value ζ, such that (3.13) holds. To be more precise, for numerical reasons, we plot the borderline that represents the set of η that corresponds to the values $\zeta = 0.2$ (in the analysis of $\alpha = 1.3$ and $\alpha = 1.5$) respectively $\zeta = 0.1$ ($\alpha = 1.7$ and $\alpha = 1.9$). Since there also exist values η corresponding to $\zeta < 0.2$ respectively $\zeta < 0.1$, the given plots can be interpreted as an upper bound for the set of all η fulfilling (3.13)[2].

Equations 3.9 and 3.13 are not exactly the same, since Assumption A^{ST} implicitly imposes the additional condition, that both $\frac{\partial}{\partial\beta}VaR_\eta(\beta)$ and $\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta)$ are not equal to zero. Although the condition of these two expressions being equal to zero is not stable in the sense that a slight perturbation will drive the expressions unequal to zero, we depict in Figure 3.2 the set of values η for which there are no zeros of $\frac{\partial}{\partial\beta}VaR_\eta(\beta)$. The plots are the results of finding the greatest zeros in $(0,1)$ of $\frac{\partial}{\partial\beta}VaR_\eta(\beta)$. Hence, all values η lying above the respective graph for $\alpha = 1.3, 1.5, 1.7$ and 1.9 are guaranteed to result in a denominator which is unequal to zero.

[1] We restrict ourselves to the interval $(-0.9, 0.9)$ for practical reasons. There is no difficulty to perform an equivalent calculation for higher respectively lower values than given by this interval.

[2] By numerical observation we claim that for higher levels than η given in the boarderlines, there also exist values ζ, such that Assumption A^{ST} holds. However, to be rigorous, this will have to be considered analytically.

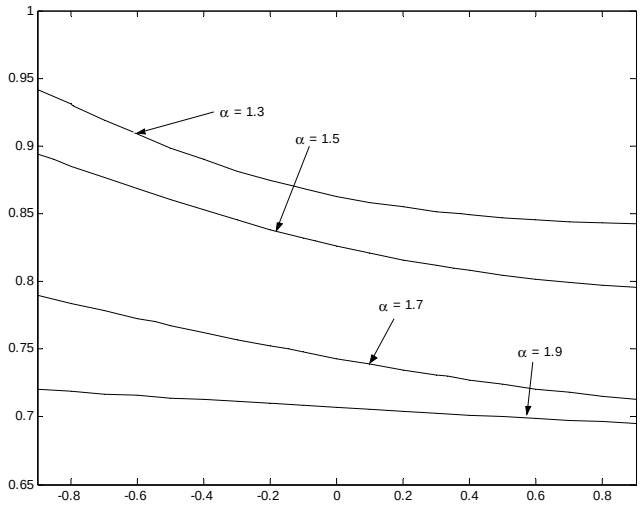

Figure 3.1: β versus η. The figure sketches those confidence levels, for which Assumption A^{ST} is fulfilled. Above the line of the corresponding stability parameter α, Equation 3.9 is satisfied for some suitable ζ_0.

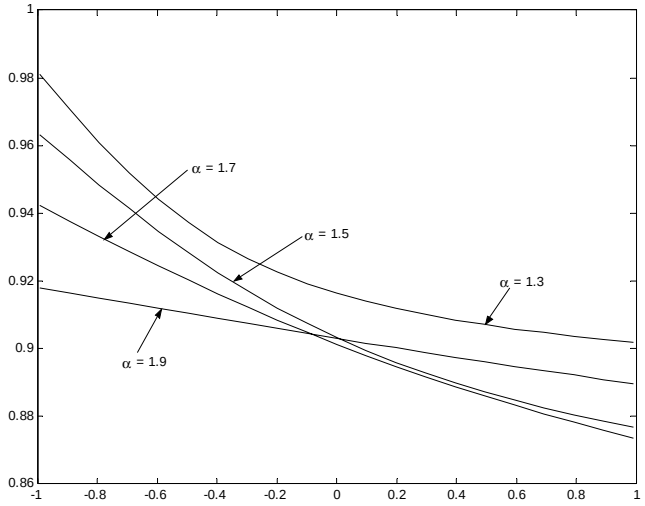

Figure 3.2: β versus η. There are no zeros of the function $\frac{\partial}{\partial\beta}VaR_\eta(\beta)$ for values η above the sketched lines for the respective stability parameter α and depending on the skewness parameter β.

In Figure 3.3 we analyse the analogous result for the expression $\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta)$ not getting equal to zero. It is intuitively clear that this results in much lower (greatest) zeros in the interval $(0,1)$. This is due to the fact

$$\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta) = \frac{1}{1-\zeta_0}\int_{\zeta_0}^{1}\frac{\partial}{\partial\beta}VaR_{\varepsilon}(\beta)d\varepsilon.$$

Since $\frac{\partial}{\partial\beta}VaR_{\varepsilon}(\beta)$ for sufficiently high levels ε admits the same sign, at least for those levels ε, $\frac{\partial}{\partial\beta}CVaR_{\varepsilon}(\beta)$ also admits the same sign. Figure 3.3 also shows that guaranteeing $\frac{\partial}{\partial\beta}CVaR_{\zeta_0}(\beta) \neq 0$ in (3.9) gives us another restriction within the analysis performed in Figure 3.1.

Regarding Figure 3.3 it is conspicuous that independently of the stability parameter α, the corresponding lines have one point in common. Hence by our numerical observation for any $\alpha \in (1,2]$ we have

$$\frac{\partial}{\partial\beta}CVaR_{\frac{1}{2}}(0) = 0.$$

Our numerical analysis also shows another observation. For $\alpha \in [1.5,2]$ we observe that by increasing the parameter α, the corresponding lines are also successively increasing for negative skewness parameters. For positive skewness parameters, the converse is true; the corresponding lines are decreasing by increasing the values for α. It is now interesting to point out that for values $\alpha \in (1,1.5]$ exactly the converse of the described behavior is true. As a summarizing result, we can state that for all values $\alpha \in (1,2]$ and $\beta \in (-1,1)$ the resulting zeros of

$$f_{\alpha,\beta} : \zeta \mapsto \frac{\partial}{\partial\beta}CVaR_{\zeta}(\beta)$$

are included in the set

$$\{\zeta : \zeta \text{ is a zero of } f_{1.5,\beta} \text{ for all } \beta \in (-1,1)\},$$

which again using Figure 3.3 can be approximated to be a subset of $(0.55, 0.45)$.

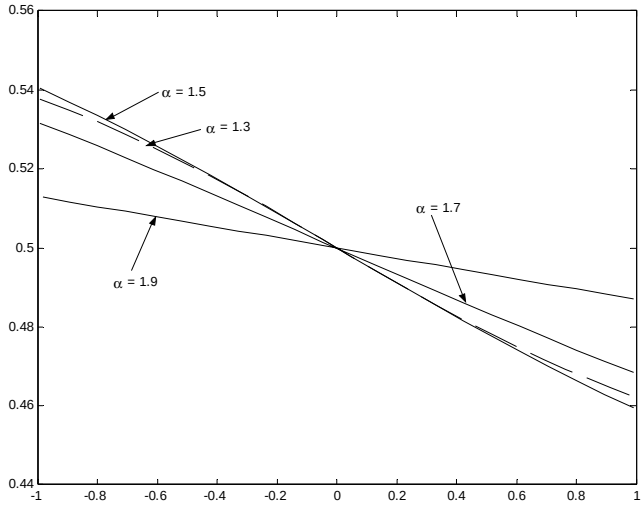

Figure 3.3: β versus ζ. There are no zeros of the function $\frac{\partial}{\partial\beta}CVaR_\zeta(\beta)$ for values ζ above the sketched lines for the respective stability parameter α and depending on the skewness parameter β.

3.2.4 First Estimations for G-and-H distributed asset returns

Within this chapter we will suppose that every portfolio allocation can be represented by a g-and-h distribution, i.e.

$$X(x) \sim X(\mu(x), \sigma(x), g(x), h(x)) \in GH$$

for all permissible allocations x. For the Value-at-Risk respectively the expected shortfall we write

$$\begin{aligned} VaR_\eta^{g,h}(x) &= \sigma(x)\overline{VaR^{g,h}}_\eta(g(x), h(x)) + \mu(x) \\ CVaR_\eta^{g,h}(x) &= \sigma(x)\overline{CVaR}_\eta^{g,h}(g(x), h(x)) + \mu(x) \end{aligned}$$

$\mu(x)$ and $\sigma(x)$ are shorthand for the expected value and the standard deviation of the loss distribution, respectively[3]. In order to state the result of

[3] For the equalities above to be well defined $\mu(x)$, $\sigma(x)$ will have to be finite. Additional restrictions on the parameter values g, h will have to be imposed.

the theorem we introduce Assumption $\mathrm{A}^{g,h}$. This is done in analogy to the above Assumption A^{ST} for stable distributions.

Assumption $\mathbf{A}^{g,h}$ Let $g^* = g(x_\eta^*)$ and $h^* = h(x_\eta^*)$ denote the parameter values for the VaR optimal allocation x_η^* as given by the solution to (2.5) and for g-and-h distributed asset returns. Then there exists $\zeta_0 \in (0,1)$, s.t.

$$\frac{\overline{VaR}_\eta^{g,h}(g^*,h^*)}{\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)} = \frac{\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g^*,h^*)}{\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)} := c. \tag{3.14}$$

Moreover, define

$$\frac{\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)}{\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g^*,h^*)} - \frac{\frac{\partial}{\partial h}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)}{\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g^*,h^*)} := \varepsilon.$$

Remark 3.2.5 *(i) In Assumption $A^{g,h}$ above we suppressed the subindex c in $\overline{CVaR}_{\zeta_0,c}^{g,h}(g^*,h^*)$, since we implicitly set $c=1$, i.e. we are considering the CVaR optimization problem with normalized $\mu(x)=0$ and $\sigma(x)=1$ for all $x \in \mathbb{X}$.*

(ii) It must be secured that all the denominators appearing within Assumption $A^{g,h}$ are not zero. However, as in the case of stable distributed asset returns, it is clear, that slight perturbations will force all denominators to be nonzero.

Theorem 3.2.5 *Suppose all allocations $x \in \mathbb{X}$ are g-and-h distributed, s.t. Assumption $A^{g,h}$ holds true, the determining parameter functions ($\mu(x)$, $\sigma(x)$, $g(x)$ and $h(x)$) are all differentiable with respect to x and $CVaR_\zeta^{g,h}(g(x),h(x))$ is a convex function of x for any confidence level $\zeta \in (0,1)$. Let*

$$x_{\zeta_0,c}^* \in \arg\min_{x\in\mathbb{X}} CVaR_{\zeta_0,c}^{g,h}(x)$$
$$x_\eta^* \in \arg\min_{x\in\mathbb{X}} VaR_\eta^{g,h}(x)$$

denote the optimal solutions of the corresponding $CVaR_{\zeta_0}^{g,h}$ and $VaR_\eta^{g,h}$ optimization, then

$$||x_\eta^* - x_{\zeta_0}^*|| \leq R^{-1} \cdot |\varepsilon| \cdot c \cdot \sigma(x_\eta^*) \cdot \left|\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g^*,h^*)\right| \cdot ||\nabla h(x_\eta^*)|| \tag{3.15}$$

holds. We wrote R for the Rayleigh quotient of $\overline{G}^2$*, the matrix* $\overline{G}$ *being defined by*

$$\overline{G} := \left(\int_0^1 \nabla^2 CVaR_{\zeta_0,c}(x^*_{\zeta_0} + \tau(x^*_\eta - x^*_{\zeta_0}))d\tau \right).$$

Proof: Since we are dealing with smoothly transformed normal distributions, twice continuous differentiability of $CVaR^{g,h}$ is fulfilled and the usage of Rayleigh quotients is well defined. For convenience, we will suppress the indices g, h in the notation of $VaR^{g,h}$ and $CVaR^{g,h}$. x^*_η stands for the optimal asset allocation of the VaR optimization problem. A necessary condition for x^*_η to be optimal is

$$\begin{aligned}
\nabla VaR_\eta(x^*_\eta) &= \overline{VaR}_\eta(g^*,h^*)\cdot\nabla\sigma(x^*_\eta) + \sigma(x^*_\eta)\cdot\nabla\overline{VaR}_\eta(g^*,h^*) + \nabla\mu(x^*_\eta) \\
&= \overline{VaR}_\eta(g^*,h^*)\cdot\nabla\sigma(x^*_\eta) + \sigma(x^*_\eta)\cdot\frac{\partial}{\partial g}\overline{VaR}_\eta(g^*,h^*)\cdot\nabla g(x^*_\eta) \\
&\quad +\sigma(x^*_\eta)\cdot\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\cdot\nabla h(x^*_\eta) + \nabla\mu(x^*_\eta) \\
&= \lambda\cdot\mathbf{1} + \sum_{i=1}^n \mu_i e_i
\end{aligned}$$

and

$$\mu_i x^*_\eta(i) = 0, \ \forall\ i = 1, ..., n \tag{3.16}$$

for $\lambda \in \mathbb{R}$, $\mu_i \geq 0$, $\forall i = 1, ..., n$. In analogy, we can now write for $CVaR_{\zeta_0}(x)$

$$\begin{aligned}
\nabla CVaR_{\zeta_0,c}(x^*_\eta) &= \overline{CVaR}_{\zeta_0}(g^*,h^*)\nabla(c\sigma(x^*_\eta)) + c\sigma(x^*_\eta)\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}(g^*,h^*)\nabla g(x^*_\eta) \\
&\quad +c\sigma(x^*_\eta)\frac{\partial}{\partial h}\overline{CVaR}_{\zeta_0}(g^*,h^*)\nabla h(x^*_\eta) + \nabla\mu(x^*_\eta) \\
&= \overline{VaR}_\eta(g^*,h^*)\frac{\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}(g^*,h^*)}{\frac{\partial}{\partial g}\overline{VaR}_\eta(g^*,h^*)}\nabla(c\sigma(x^*_\eta)) \\
&\quad +c\sigma(x^*_\eta)\frac{\partial}{\partial g}\overline{VaR}_\eta(g^*,h^*)\frac{\overline{CVaR}_{\zeta_0}(g^*,h^*)}{\overline{VaR}_\eta(g^*,h^*)}\nabla g(x^*_\eta) \\
&\quad +c\sigma(x^*_\eta)\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\left(\frac{\overline{CVaR}_{\zeta_0}(g^*,h^*)}{\overline{VaR}_\eta(g^*,h^*)}\right) - \varepsilon\nabla h(x^*_\eta) \\
&\quad +\nabla\mu(x^*_\eta)
\end{aligned}$$

$$
\begin{aligned}
&= \overline{VaR}_\eta(g^*,h^*)\nabla\sigma(x_\eta^*) + \sigma(x_\eta^*)\frac{\partial}{\partial g}\overline{VaR}_\eta(g^*,h^*)\nabla g(x_\eta^*) \\
&\quad +\sigma(x_\eta^*)\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\nabla h(x_\eta^*) \\
&\quad -\varepsilon\cdot c\cdot\sigma(x_\eta^*)\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\nabla h(x_\eta^*) + \nabla\mu(x_\eta^*) \\
&= \lambda\cdot\mathbf{1} + \sum_{i=1}^{n}\mu_i e_i - \varepsilon\cdot c\cdot\sigma(x_\eta^*)\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\nabla h(x_\eta^*). \qquad (3.17)
\end{aligned}
$$

On the other hand, using Taylor's theorem we can write for the gradient of a twice continuously differentiable function $f(x)$ with a (local) extremum in x^*

$$\nabla f(x) = \nabla f(x^*) + \overline{G}(x-x^*)$$

with

$$\overline{G} = \int_0^1 \nabla^2 f(x+\tau(x^*-x))d\tau.$$

Since $\nabla f(x^*) = 0$ we can therefore write

$$||\nabla f(x)||^2 = R^2(x)\cdot||x-x^*||^2 \qquad (3.18)$$

where

$$R^2(x) = \frac{(x-x^*)^t\cdot\overline{G}^2\cdot(x-x^*)}{||x-x^*||^2}$$

describes the Rayleigh quotient of $\overline{G}^2$ in the direction of $x-x^*$. If we set

$$f(x) \equiv CVaR_{\zeta_0,c}(x) - \lambda\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu_i x_i \qquad (3.19)$$

and using equations (3.17), (3.18) and (3.16), according to the Karush-Kuhn-Tucker conditions for optimality we get

$$
\begin{aligned}
||x_\eta^* - x_{\zeta_0}^*||^2 &\leq R^2(x_\eta^*)\left\|\nabla(CVaR_{\zeta_0,c}(x) - \lambda\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu_i x_i)\right\|^2 \\
&= R(x_\eta^*)^{-2}\left(\varepsilon\cdot c\cdot\sigma(x_\eta^*)\cdot\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*)\right)^2\cdot\left\|\nabla h(x_\eta^*)\right\|^2
\end{aligned}
$$

since $\nabla f(x_{\zeta_0}^*) = 0$. The above stated result follows. □

Remark 3.2.6 *(i) Denoting by* $\lambda_{\min}$ *and* $\lambda_{\max}$ *the smallest respectively the highest eigenvalue of the matrix* $\overline{G}^2$*, it can easily be seen that*

$$\lambda_{\min} \leq R \leq \lambda_{\max},$$

hence $\lambda_{\min}$ *can be used instead of* R *to achieve a similar estimation as 3.15. However, numerical analysis suggests (compare section 3.3.3) that in general the respective eigenvalue is very small.*

(ii) It can be shown that the following inequality between the minimal eigenvalues of Hermitian $(n \times n)$ *matrices* $A, B \in H_n$ *is true*

$$\lambda_{\min}(A) + \lambda_{\min}(B) \leq \lambda_{\min}(A + B). \tag{3.20}$$

If the functions $g(x), h(x)$ *can be chosen, s.t.* $CVaR^{g,h}_{\zeta_0}(g(x), h(x))$ *is a convex function of* x*, the corresponding hessian matrix is hermitian and inequality (3.20) gives us a lower bound for* R *:*

$$R \geq \int_0^1 \lambda_{\min}(x^*_{\zeta_0} + \tau(x^*_\eta - x^*_{\zeta_0}))d\tau, \tag{3.21}$$

where we again denote by $\lambda_{\min}(x)$ *the smallest (positive) eigenvalue of the Hermitian matrix* $\nabla^2 CVaR^{g,h}_{\zeta_0}(x)$*. The value of* R *is therefore mainly a function of the curvature of* $CVaR^{g,h}_{\zeta_0}(x)$*. Hence, for those portfolio optimization problems where the standard deviation* $\sigma(x)$ *makes up a large portion of the resulting* $CVaR^{g,h}_{\zeta_0}(x)$ *(as e.g. in the case of elliptically distributed asset returns), or where the curvature of the portfolio standard deviation in the direction* $\overline{x^*_\eta x^*_{\zeta_0}}$ *is "high", the resulting values of* R *can be guaranteed not to be too small. Moreover, it is worth mentioning the dependency of* R *from* ζ_0 *respectively* η*. This enables us in the following to show that for sufficiently high values* $\eta \in (0, 1)$*, the proposed* $CVaR^{g,h}_{\zeta_0,c}(x)$*-optimization will give good approximations to the optimal allocation* x^*_η*.*

(iii) We are mainly interested in finding an upper bound for $\left\| x^*_\eta - x^*_{\zeta_0} \right\|$*. However, the approach of the theorem can easily be seen to also give a lower bound on the difference of the two minimizing portfolio allocations*

$$\left\| x^*_\eta - x^*_{\zeta_0} \right\| \geq \lambda^{-1}_{\max} \cdot |\varepsilon| \cdot c \cdot \sigma(x^*_\eta) \cdot \left| \frac{\partial}{\partial h} \overline{VaR}^{g,h}_\eta(g^*, h^*) \right| \cdot ||\nabla h(x^*_\eta)||,$$

where $\lambda_{\max}$ *denotes the maximum eigenvalue of the matrix* $\overline{G}^2$*. Moreover, a similar inequality as the one in (3.21) can be found*

$$R \leq \int_0^1 \lambda_{\max}(x^*_{\zeta_0} + \tau(x^*_\eta - x^*_{\zeta_0}))d\tau.$$

Moreover, we are always working with a fixed $\zeta_0 \in (0,1)$ *given by Assumption* $A^{g,h}$. *Hence it is not clear whether or not this choice is in fact the* $VaR^{g,h}$ *minimizing one among all possible* x^*_ζ *in the sense that*

$$\min_{(\zeta,c)\in(0,1)\times\mathbb{R}^+} VaR^{g,h}_\eta(x^*_{\zeta,c}) = VaR^{g,h}_\eta(x^*_{\zeta_0}).$$

Hence, by searching the two dimensional space of solutions given by (2.7) one might even find better approximations to x^*_η *than the one analyzed in the theorem above.*

3.3 Approximative Results for the VaR Optimum

In what follows, we want to consider a four parametric family of distributions to approximate the original Value-at-Risk and Expected Shortfall, respectively. More precisely, we will focus on a special class of distributions known as the g-and-h distribution. The reason for choosing this distribution was the major flexibility of this class in the sense, that a broad class of financially relevant distributions can be replicated using the g-and-h distribution. For example, the class of g-and-h distributions includes normal and log-normal as well as heavy skewed distributions with possibly high kurtosis. Moreover, the class of g-and-h distributions inherits the attracting feature to have an explicitly known formula for the quantile, i.e. VaR-calculations appear to be more or less simple.

Besides the enormous flexibility of the proposed approach, we will see that the main results are referring to special restrictions. Using numerical evaluations, we will focus on the restrictiveness of these additional conditions within this chapter.

3.3.1 Using G-and-H Distributions to Match Real-World Optimization Problems

As stated in Theorem 3.2.5, optimization problems with g-and-h distributions have quite nice properties when used as approximates for the actual distributions. However, it is not at first glance clear how to define the functions $g(x), h(x)$ in a way that the resulting function $CVaR^{g,h}_{\zeta_0,c}(g(x),h(x))$ is

a convex function of x. Since only a convex $CVaR$ optimization will lead to the desired reduction in complexity of the affine VaR optimization problem, for practical and theoretical (uniqueness of the $CVaR$ minimum) reasons $CVaR_{\zeta_0,c}(g(x),h(x))$ will have to be chosen as a convex function. Moreover, as stated in the remark to the theorem, for a convex choice of $CVaR$ the value R in the corresponding estimation (3.15) can be shown to give better estimations.

In the following theorem we will use the implicit function theorem in order to guarantee convexity of $CVaR$ by choosing the functions $g(x), h(x)$ in an appropriate way. Moreover, we will see that under some additional (not too restrictive) conditions, these functions can be chosen, s.t. the resulting $VaR_\eta^{g,h}(x)$ matches the $VaR_\eta(x)$ values of the original optimization problem. In order to guarantee convexity of the function $CVaR_{\zeta_0}^{g,h}(x)$, we will also match it with the original (convex) $CVaR_{\zeta_0}(x)$. All these allocations satisfying the additional constraints are collected in the following definition

$$AX \;:=\; \Big\{ x \in \mathbb{X} : \exists (g,h) \in (-1,0)\times(0,\frac{1}{2}), \overline{VaR}_\eta^{g,h}(g,h) = \overline{VaR}_\eta(x) \\ \overline{CVaR}_{\zeta_0}^{g,h}(g,h) = \overline{CVaR}_{\zeta_0}(x) \Big\}.$$

Within this definition and within Theorem 3.3.1, we restrict ourselves to $h^* \in (0,\frac{1}{2})$, since otherwise the corresponding standard deviation $\sigma(g,h^*)$ does not exist for any value g. Moreover, in order to guarantee the explicit form of the VaR as given in equation 2.27 we restrict ourselves to the case where h^* is truly positive. Note that in the definition of AX, ζ_0 is defined by the minimum value x_η^* and Assumption $A^{g,h}$, therefore independently of the choice of g and h. This fact is used in our numerical considerations below.

Theorem 3.3.1 *Suppose x_η^*, $x_{\zeta_0}^*$ are both elements of the same connected component of AX. Moreover, suppose that $\overline{VaR}_\eta(x)$ is once and $\overline{CVaR}_{\zeta_0}(x)$ is twice continuously differentiable in x and Assumption $A^{g,h}$ holds. Then there exist continuously differentiable functions $g(x), h(x)$, s.t.*

$$\left\| x_\eta^* - x_{\zeta_0}^* \right\| \le R^{-1}\cdot\sigma(x_\eta^*)\cdot\left\| c\cdot\nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla\overline{VaR}_\eta(x_\eta^*)\right\|. \qquad (3.22)$$

As in the previous results, R denotes the Rayleigh quotient of the matrix $\overline{G}^2$.

Proof: For any $x^* \in AX$, choose open sets $U_1 \subset AX$, $U_2 \subset (-1,0)$ and

$U_3 \subset (0, \frac{1}{2})$ of x^* with corresponding values g^*, h^* and define the function

$$\begin{aligned} F : U_1 \times U_2 \times U_3 &\rightarrow \mathbb{R}^2 \\ (x, g, h) &\longmapsto \begin{pmatrix} \overline{VaR}_\eta^{g,h}(g,h) - \overline{VaR}_\eta(x) \\ \overline{CVaR}_{\zeta_0}^{g,h}(g,h) - \overline{CVaR}_{\zeta_0}(x) \end{pmatrix} \end{aligned}$$

Hence, under the assumptions of the theorem we know

$$F(x^*, g^*, h^*) = 0.$$

If we set

$$C := \begin{pmatrix} \frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g^*,h^*) & \frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g^*,h^*) \\ \frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*) & \frac{\partial}{\partial h}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*) \end{pmatrix} \tag{3.23}$$

then numerical examinations (compare the numerical part below) suggest that

$$\det(C) \not= 0,$$

for all parameter values $(g, h) \in (-1, 0) \times (0, \frac{1}{2})$. Applying the theorem of implicit functions yields the existence of unique functions $g(x), h(x)$ defined on appropriate subsets of AX. Since the choice of $x^* \in AX$ was arbitrary, by uniqueness $g(x), h(x)$ are well-defined on the whole of AX; the respective gradients in the VaR optimal allocation x_η^* are

$$\begin{aligned} \nabla g(x_\eta^*) &= \frac{1}{\det(C)} \cdot \begin{bmatrix} \frac{\partial}{\partial h}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*) \cdot \nabla\overline{VaR}_\eta(x_\eta^*) \\ -\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g^*,h^*) \cdot \nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) \end{bmatrix} \\ \nabla h(x_\eta^*) &= \frac{1}{\det(C)} \cdot \begin{bmatrix} \frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g^*,h^*) \cdot \nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) \\ -\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*) \cdot \nabla\overline{VaR}_\eta(x_\eta^*) \end{bmatrix} \end{aligned}$$

Now, by the construction of the functions g and h and the relation

$$\begin{aligned} CVaR_{\zeta_0}^{g,h}(x) &\equiv \mu(x) + \sigma(x)\overline{CVaR}_{\zeta_0}^{g,h}(x) \\ &= \mu(x) + \sigma(x)\overline{CVaR}_{\zeta_0}(x) \equiv CVaR_{\zeta_0}(x) \end{aligned}$$

$CVaR_{\zeta_0}^{g,h}(x)$ is a convex function of x whenever $CVaR_{\zeta_0}(x)$ is convex. Hence, the assumptions of theorem 3.2.5 are fulfilled and we can write

$$\begin{aligned}\left\|x_\eta^* - x_{\zeta_0}^*\right\| &\leq R^{-1} \cdot |\varepsilon| \cdot c \cdot \sigma(x_\eta^*) \cdot \left|\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*, h^*)\right| \cdot ||\nabla h(x_\eta^*)|| \\ &= R^{-1} \cdot \sigma(x_\eta^*) \cdot \left\|c \cdot \nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla\overline{VaR}_\eta(x_\eta^*)\right\|,\end{aligned}$$

which is the claim of the theorem. □

Remark 3.3.1 *It is interesting to note that by the use of g-and-h-distributions to fit either VaR or CVaR, equation (3.22) is independent of the matching error as defined in Assumption $A^{g,h}$.*

Numerical evaluations of the conditions for the approximate approach

Within the foregoing analysis we used several partly implicitly given assumptions in order to get inequality 3.22. It is the purpose of this chapter to shed some light on the restrictiveness of these assumptions. We will firstly analyse Assumption $A^{g,h}$ with its inherent assumption of all the expressions being unequal to zero. After having a closer look on the determinant of matrix C as given in Equation 3.23, we will analyse the most restrictive assumption, x_η^* and $x_{\zeta_0}^*$ both being in the same connected component of the set AX. Although it seems not possible to catch the complexity of Assumption $A^{g,h}$ in full generality, we will give some hint on its fulfillment. Again, in all the numerical analysis that follows, we restrict ourselves to the case of positive parameters $h < \frac{1}{2}$, s.t. the corresponding distribution admits finite second moments. Since for higher values h the corresponding distribution already admits exorbitant high standard deviations, for practical reasons we restrict to the case $h < 0.4$. Further research could avoid these numerical disadvantages and extent the analysis to the case of higher values h.

Moreover, since most of the poor behavior (at least for confidence levels being high enough) takes place for parameter values $g > 0$, one could restrict oneself to the case $g \leq 0$. However, as we will see, the set AX strongly depends on positive parameters g. Hence, we extend our analysis to the region of those values g, h which imply a strictly positive determinant C. Again caused by numerical considerations we restrict the region of admissible values g to the case $|g| < 1$, thereby always respecting all the other active constraints as

mentioned above. Further research will enable to extend this region in order to improve the set AX.

One advantage of the following analysis is the fact that both the $VaR^{g,h}$ for any confidence level as well as the respective partial derivatives with respect to g and h are explicit functions of the quantiles of a standard normal distribution. Moreover, explicit expressions for the mean and the standard deviation are known and can be used to give explicit forms of the VaR of a standardized g-and-h distributed random variable. However, these calculations are quite lengthy and are the best performed by using symbolic computation systems such as Maple. This proceeding has the advantage that we can do the numerical calculations without referring to finite approximations of the derivatives. Analogous results hold for $CVaR$ and its partial derivatives with respect to g and h.

Assumption $\mathbf{A}^{g,h}$

In order to properly analyse Assumption $A^{g,h}$, we first write equation 3.14 in the form

$$\overline{VaR}_{\eta}^{g,h}(g^*,h^*)\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*) = \overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)\frac{\partial}{\partial g}\overline{VaR}_{\eta}^{g,h}(g^*,h^*)$$

to get rid of the denominators. These are further analyzed below. For several values $\eta \in (0,1)$ we depict in Figure 3.4 the respective lines for which the regions to the left impose values g, h, s.t. Assumption $A^{g,h}$ is satisfied with a ζ_0 larger than 0.1. Hence, for confidence levels $\eta > 0.94$ Assumption $A^{g,h}$ is satisfied in the whole region under consideration. Moreover, extending the analysis to values $g < -1$ seem not to impose additional restrictions.

In what follows, we want to consider the nonzeroness of the denominators appearing in the definition of Assumption $A^{g,h}$. Although there are points, where the respective denominators vanish, a slight perturbation in g^*, h^* (or, alternatively in η respectively ζ_0) will result in nonzero denominators by not influencing the overall result too much. This is due to the fact that the requirement of nonzero denominators only has to be fulfilled in one point (g^*, h^*).

Assumption $\frac{\partial}{\partial g}\overline{VaR}_{\eta}^{g,h}(g^*,h^*) \neq 0$

For different confidence levels η, Figure 3.5 shows the respective zeros of the expression

$$\frac{\partial}{\partial g}\overline{VaR}_{\eta}^{g,h}(g,h)$$

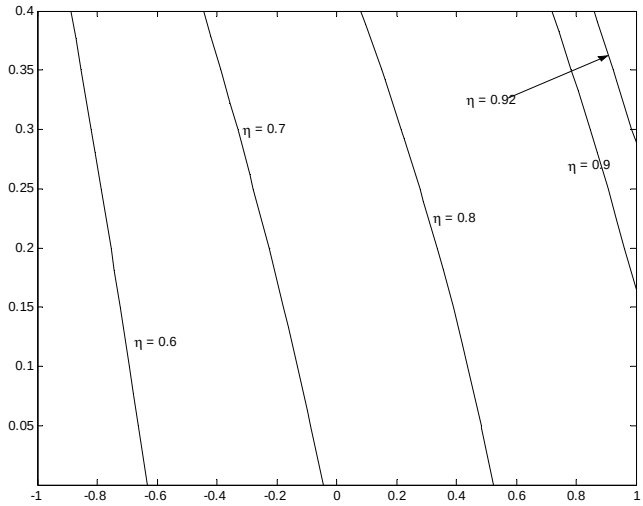

Figure 3.4: g versus h. For all pairs g, h in the region left to the respective line for some $\eta \in (0,1)$ assumption $A^{g,h}$ is fulfilled. Depicted are the lines where there exists $\zeta_0 > 0.1$, s.t. assumption $A^{g,h}$ holds.

within the region $(-1,1) \times (0, 0.4)$ (dotted lines). Moreover, we depict those regions around the zeros where the absolute value of the expression under consideration is less than 0.05. This calculation was performed in order to get an impression of the sensitivity of these zeros. It can be observed that for the analyzed confidence levels, all the zeros only appear for positive values g. In addition, the higher we are choosing the confidence level η, the greater will be the region left to the respective line of zeros and the more sensitive this zero moves away from zero by a slight perturbation in (g^*, h^*).

Assumption $\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*, h^*) \neq 0$

We could also do a similar analysis as performed in the preceding section. However, by analyzing the expression

$$\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g, h)$$

for every value $(g, h) \in (-1,1) \times (0, 0.4)$ we would have to calculate the corresponding value ζ_0 satisfying Assumption $A^{g,h}$. In fact, this analysis becomes as easy as just referring to the analysis of the preceding section. This apparently becomes clear by realizing that for sufficiently high confidence

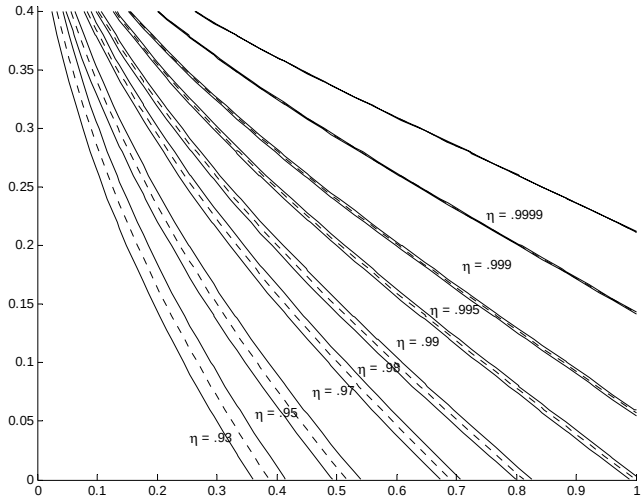

Figure 3.5: g versus h. Zeros of the function $\frac{\partial}{\partial g}\overline{VaR}_\eta(g,h)$ (dotted lines) for different values η. In order to evaluate sensitivity of the zeroes, we include the level sets to -0.05 respectively 0.05 (solid lines) for every confidence level η.

levels $\eta \in (0,1)$ the corresponding value $\overline{VaR}_\eta^{g,h}(g,h)$ is strictly positive, hence by the property $\overline{CVaR}_{\zeta_0}^{g,h}(g,h) > 0$ for any choice of parameters g^*, h^* we conclude

$$\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g^*,h^*) = c \cdot \frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g^*,h^*)$$

for some positive value c as a direct consequence of Assumption $A^{g,h}$. We would like to point out that the confidence levels in Figure 3.5 are sufficiently high to guarantee $\overline{VaR}_\eta^{g,h}(g,h) > 0$ for any combination (g,h) in the region of interest. To summarize, the zeros of the expression exactly coincide with the zeros given in Figure 3.5.

Assumption $\frac{\partial}{\partial h}\overline{VaR}_\eta(g^*,h^*) \neq 0$

As Figure 3.6 shows, depending on the confidence level η of the $VaR_\eta(x)$ optimization problem there are in fact lines representing zeros of the expression

$$\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g,h)$$

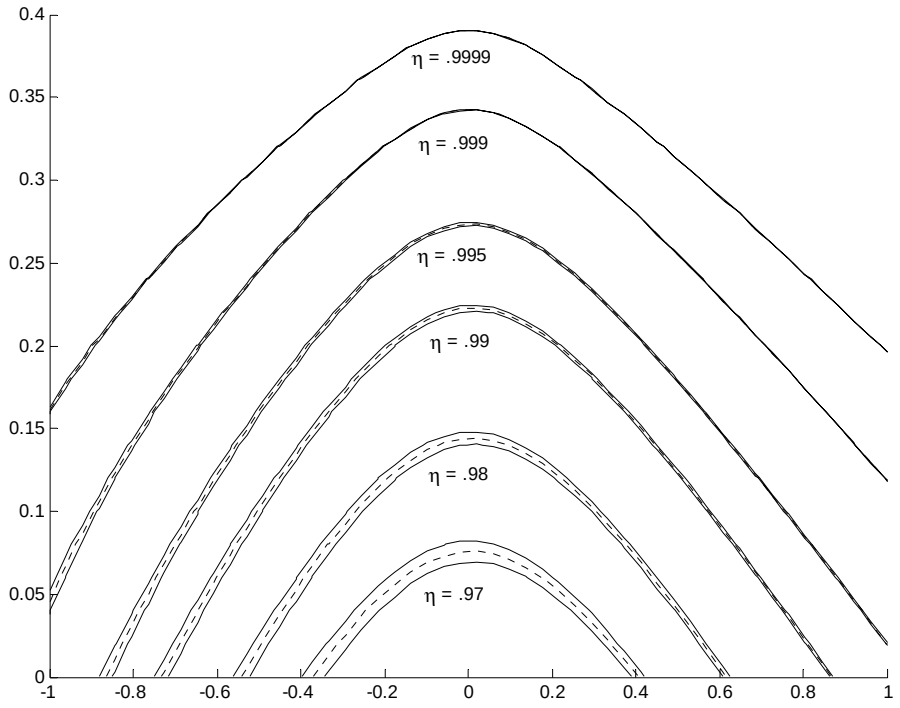

Figure 3.6: g versus h. Zeros of the function $\frac{\partial}{\partial h} VaR_\eta^{g,h}(g,h)$ (dotted lines) for different values η. In order to evaluate sensitivity of these zeroes, we include the level sets to -0.05 respectively 0.05 (solid lines) for every confidence level η.

within the region of interest. As in the preceding analysis on $\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g,h)$ in addition to the actual zeros (dotted lines) we also plot the regions where $\left|\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g,h)\right| < 0.05$ holds. Again this is done in order to get an impression of the sensitivity to possible perturbations. As before, such a perturbation could be performed either in the parameter values g, h as well as in the confidence level η in order to achieve a nonzero value of $\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g,h)$ in the optimal parameter value (g^*, h^*) as defined above.

To achieve the results shown in Figure 3.6, we just calculated the value of $\frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g,h)$ over a fine mesh of the region $(-1,1) \times (0, 0.4)$. Shown are the respective level sets of these evaluations.

Assumption of a nonzero determinant C

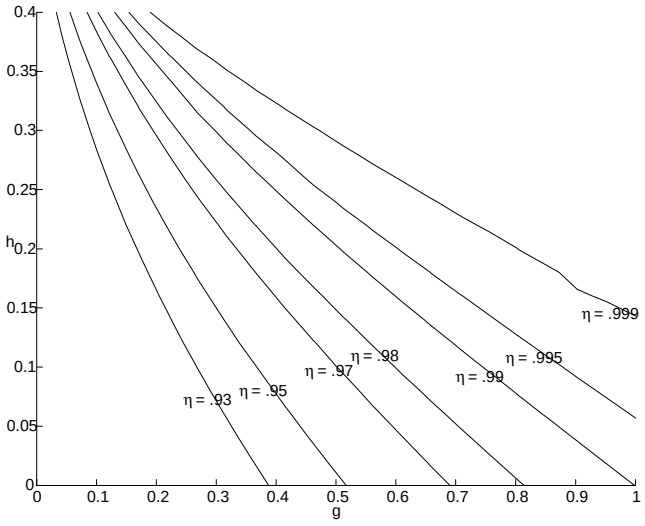

Figure 3.7: g^* versus h^*.Zeros of the determinant of $C(\eta, g, h)$ for the selected values $\eta \in \{.93, .95, .97, .98, .99, .995, .999\}$

The determinant of matrix C as defined in the proof of theorem 3.3.1 is strongly related to the defining equation of assumption $A^{g,h}$. Therefore it is not too surprising, that the set of parameters g, h, s.t.

$$\det(C) = 0$$

again coincides with the set of zeros given by figure 3.5. It is clear that this set must be a subset of the set of parameter values g, h for that $\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g,h)$ vanishes. This is due to the fact that

$$\begin{aligned}\det(C) &= \frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g,h) \cdot \frac{\partial}{\partial h}\overline{CVaR}_{\zeta_0}^{g,h}(g,h) \\ &\quad -\frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g,h) \cdot \frac{\partial}{\partial h}\overline{VaR}_\eta^{g,h}(g,h)\end{aligned}$$

and

$$\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g,h) = 0 \Leftrightarrow \frac{\partial}{\partial g}\overline{CVaR}_{\zeta_0}^{g,h}(g,h) = 0.$$

In Figure 3.7, we did a full analysis of the zeros over the region of interest. The picture shows, that for the choice of confidence levels there are in fact no more zeros of determinant C than given by the equality

$$\frac{\partial}{\partial g}\overline{VaR}_\eta^{g,h}(g,h) = 0.$$

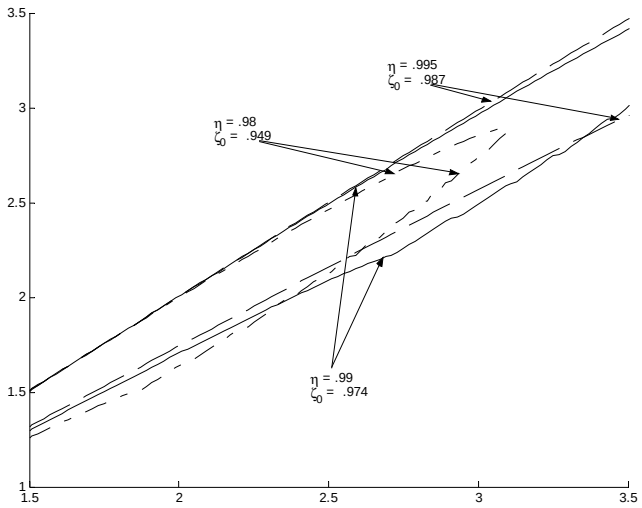

Figure 3.8: $CVaR_{\zeta_0}(x)$ versus $VaR_\eta(x)$. For fixed confidence levels (η, ζ_0) we sketch the boundaries given as the maximum and minimum value $VaR_\eta(x)$, s.t. there exists a corresponding value (g, h) in a suitable subset of $(-1, 1) \times (0, 0.4)$ that satisfies the condition given in the definition of AX. η and ζ_0 are chosen s.t. they correspond to the pair satisfying Assumption $A^{g,h}$ for normal distributed random variables.

A characterization of the set AX

Whereas the foregoing assumptions can be seen to be satisfied for a wide range of parameter values and can even be turned into the case of fully satisfying all the assumptions by a slight perturbation, the following condition will be the one that truly puts a quite restrictive requirement on our overall desire to match the original VaR_η optimization problem by a suitable optimization of $CVaR_{\zeta_0}$.

In order to achieve the results depicted in Figure 3.8 we firstly calculated for some fixed combination (η, ζ_0) (which we chose in Figure 3.8 in such a way that ζ_0 satisfies Assumption $A^{g,h}$ for normal distributions) the corresponding values $CVaR_{\zeta_0}(g, h)$ for all parameter values $(g, h) \in (-1, 1) \times (0, 0.4)$ that satisfy the above analyzed assumptions. After calculating the corresponding level sets we determined the respective reachable values $VaR_\eta(g, h)$ for every level set. Figure 3.8 depicts for any level set with $CVaR_{\zeta_0} \in [1.5, 3.5]$ the corresponding maximum and minimum value of VaR_η.

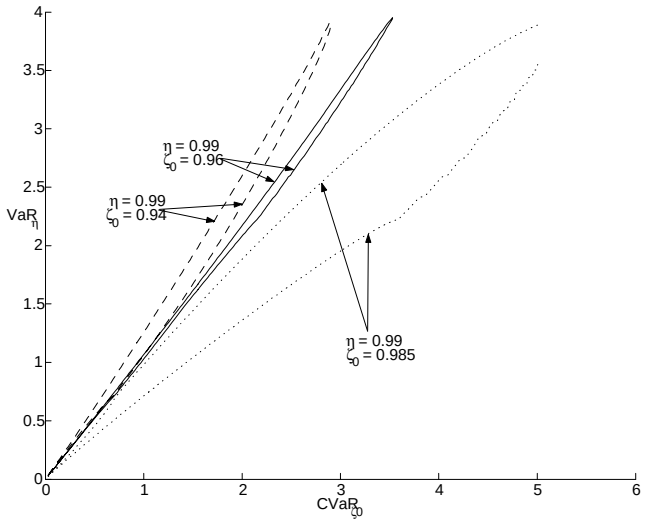

Figure 3.9: $CVaR_{\zeta_0}$ vs. VaR_η. For three combinations (η, ζ_0) the corresponding sets AX are shown. Several other assumptions are incorporated, s.t. the result of Theorem 3.3.1 is fulfilled (compare text).

After determining the optimum value ζ_0 that corresponds to its equivalent in the statement of Assumption $A^{g,h}$ and the value of $CVaR_{\zeta_0}(x)$, Figure 3.8 can be used to determine the maximum allowed variation of $VaR_\eta(x)$, s.t. the assumptions of theorem 3.3.1 are fulfilled. We see that with increasing confidence levels η the allowed variations in $VaR_\eta(x)$ for fixed $CVaR_{\zeta_0}(x)$ become larger. It is also interesting to note that all the regions are lying below or on the bisecting line. However, this observation changes if we allow the value ζ_0 to deviate from the optimum value ζ_0 of normal distributed asset returns. Figure 3.9 shows that for $CVaR$ confidence levels ζ_0 lower than the corresponding value for normal distributions in tendency lies above the bisecting line. On the contrary values for ζ_0 that are larger than the corresponding value for normal distributions result in regions that lie below the bisecting line. Moreover, the overall size of the inclosing regions of allowed variations depend heavily on the respective values ζ_0. Starting with a relatively large value ζ_0 near η the enclosed area of reachable variation $VaR_\eta(x)$ is shrinking down as ζ_0 is decreased. After reaching some value ζ_0' (approximately value $\zeta_0 = 0.96$ within Figure 3.9) where the enclosed region reaches its minimum, a further decrease in ζ_0 again results in larger possible variations of $VaR_\eta(x)$ (compare solid line within Figure 3.9).

3.3.2 VaR-optimization for general distributions

Within the next section we will derive a similar estimation of the distance $\left\|x_\eta^* - x_{\zeta_0}^*\right\|$ as formulated in Theorem 3.3.1 using approximations by g-and-h distributions. Those similar results will be derived for arbitrary (sufficiently smooth) distributions and by setting $c = 1$. This proceeding obviously has the advantage of using the proposed algorithm for finding good approximations to the VaR optimum, thereby not needing to check the various assumptions necessary in the foregoing analyses.. Moreover, Equation 3.24 is constructed to yield results for the case $c = 1$, hence showing that the main proposed algorithm of this thesis also yields appropriate results when restricting the (original) two dimensional search as formulated in 2.2 to a one dimensional one.

Estimations for arbitrary distributions

Theorem 3.3.2 *Suppose $\overline{VaR}_\eta(x_\eta^*) > 0$ and using the usual definition of $\mathbb{X}$ as given in 2.1. Then in the notational conventions of theorem 3.3.1, the following inequality holds*

$$\left\|x_\eta^* - x_{\zeta_0}^*\right\| \leq R^{-1} \cdot \sigma(x_\eta^*) \cdot \left\|\nabla \overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla \overline{VaR}_\eta(x_\eta^*)\right\|. \tag{3.24}$$

Proof: Since we are always dealing with continuous distribution functions it is clear that the mapping

$$\begin{array}{rcl} (0,1) & \to & \mathbb{R} \\ \zeta & \longmapsto & \overline{CVaR}_\zeta(x) \end{array}$$

is continuous for any choice of $x \in \mathbb{X}$. Moreover,

$$\lim_{\zeta \to 0} \overline{CVaR}_\zeta(x) = \mu(x) = 0$$

and

$$\lim_{\zeta \to 1} \overline{CVaR}_\zeta(x) = \infty.$$

Since $\overline{VaR}_\eta(x_\eta^*) > 0$, there exists a confidence level $\zeta_0 \in (0,1)$, s.t.

$$\overline{CVaR}_{\zeta_0}(x_\eta^*) = \overline{VaR}_\eta(x_\eta^*). \tag{3.25}$$

Using the definition of ζ_0 we can therefore write

$$\begin{aligned}\nabla CVaR_{\zeta_0}(x_\eta^*) &= \nabla\mu(x_\eta^*) + \nabla\sigma(x_\eta^*)\cdot\overline{CVaR}_{\zeta_0}(x_\eta^*) + \sigma(x_\eta^*)\cdot\nabla\overline{CVaR}_{\zeta_0}(x_\eta^*)\\ &= \nabla\mu(x_\eta^*) + \nabla\sigma(x_\eta^*)\cdot\overline{VaR}_\eta(x_\eta^*) + \sigma(x_\eta^*)\cdot\nabla\overline{VaR}_\eta(x_\eta^*)\\ &\quad +\sigma(x_\eta^*)\cdot\left[\nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla\overline{VaR}_\eta(x_\eta^*)\right]\\ &= \lambda\cdot\mathbf{1} + \sum_{i=1}^n \mu_i e_i + \sigma(x_\eta^*)\cdot\left[\nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla\overline{VaR}_\eta(x_\eta^*)\right]\end{aligned}$$

for $\lambda\in\mathbb{R}$, $\mu_i\geq 0$ defined via the Karush-Kuhn-Tucker condition for optimality

$$\nabla VaR_\eta(x) = \lambda\cdot\mathbf{1} + \sum_{i=1}^n \mu_i e_i.$$

Following the lines of the proof of Theorem 3.3.1, one sees that the norm of the gradient of the related Lagrange function can be estimated. This again can be used to find an estimate of the distance $||x_\eta^* - x_{\zeta_0}^*||$,

$$\begin{aligned}||x_\eta^* - x_{\zeta_0}^*|| &\leq R^{-1}\left\|\nabla(CVaR_{\zeta_0}(x) - \lambda\sum_{i=1}^n x_i - \sum_{i=1}^n \mu_i x_i)\right\| \qquad (3.26)\\ &= R^{-1}\cdot\sigma(x_\eta^*)\cdot\left\|\nabla\overline{CVaR}_{\zeta_0}(x_\eta^*) - \nabla\overline{VaR}_\eta(x_\eta^*)\right\|\end{aligned}$$

which is the result of Theorem 3.3.1 for $c=1$, but without referring to the same restrictive assumptions. □

Remark 3.3.2 *(i) Suppose the assumptions of Theorem 3.3.2 are fulfilled for some η_0. Then they are also fulfilled for any $\eta > \eta_0$.*

(ii) By the ζ_0 defining Equation 3.25 it is apparently clear that as η approaches 1, ζ_0 tends to 1 as well. Consequently, referring to the investigation as given in 3.3.3 for increasing η sufficiently high, one expects R^{-1} to decrease, therefore resulting in better estimates.

Error estimations using best linear and quadratic predictors

Let us briefly recall the notion of a best linear predictor. In chapter 2.4.4 under certain assumptions on the distribution function, we stated the partial

derivatives of $VaR_\eta(x)$ as a function of x. In this context it is important to note, that the conditional expectation of X_i given $X(x)$ may be interpreted as the best predictor of X_i by elements of the space

$$M := \{f(X(x)) : f : \mathbb{R} \to \mathbb{R} \text{ measurable}\}$$

(c.f. [59], S.19). Let us denote the minimizing element of M by

$$f_i^*(X(x)) = \frac{\partial}{\partial x_i} VaR_\eta(x) = \mathbb{E}\left\{X_i : X(x) = VaR_\eta(x)\right\}.$$

Thereby, distance is measured by the expression

$$\mathbb{E}\left\{(X_i - f_i(X(x)))^2\right\}.$$

It is easy to see (c.f. [45], Chapter 3.3) that for elliptically distributed random vectors $(X_1, ..., X_n)$ the best predictor can be chosen to be linear in the sense that the best predictor is an element of the space $\{m \cdot X(x) : m \in \mathbb{R}\}$. In this special case it can be shown that the minimizing value m^* is given by

$$m^* = \frac{cov(X_i, X(x))}{var(X(x))}.$$

For more general than elliptically distributions the best linear predictor may considerably differ from the best predictor over all measurable functions. In the next lemma, we will therefore assume that the best *quadratic predictor* will sufficiently approximate $f_i^*(x)$.

Lemma 3.3.3 *Suppose $f_i^*(x)$ can sufficiently be approximated by a quadratic predictor*

$$f_{i,2}^*(X(x)) = \overline{a_i} \cdot [X(x)]^2 + \overline{b_i} \cdot [X(x)] + \overline{c_i}$$

and $\mathbb{E}(X(x_\eta^*)^4)$ exists. Under the assumptions of theorem 3.3.1 the error estimation reduces to

$$\left\| x_\eta^* - x_{\zeta_0}^* \right\| \leq R^{-1} \cdot \sigma(x_\eta^*) \cdot \|a\| \cdot \left[\frac{c}{1-\zeta_0} \cdot \int_{\zeta_0}^1 \overline{VaR}_\varepsilon^2(x_\eta^*) d\varepsilon - \overline{VaR}_\eta^2(x_\eta^*) \right] \tag{3.27}$$

with $a := (\overline{a}_1, ..., \overline{a}_n)$ and

$$\overline{a}_i := \frac{\mathbb{E}(X^2) \cdot cov(X_i X^2) - \mathbb{E}(X^3) cov(X_i X)}{-\mathbb{E}(X^2)\mathbb{E}(X^4) + (\mathbb{E}(X^2))^3 + (\mathbb{E}(X^3))^2}.$$

where in the last equation we wrote X shorthand for $X(x_\eta^*)$.

Remark 3.3.3 *(i) For distributions with high kurtosis, for all $i = 1, ..., n$ the value $\overline{a}_i$ is mainly driven by the term $\mathbb{E}(X^4)$. Hence $\|a\|$ will tend to be small and consequently inequality (3.27) is (at least for sufficiently high confidence levels $\eta \in (0,1)$) quite strict. Moreover, in order to find better approximations to the "real" best predictor, this will involve higher moments, that on the one hand make the corresponding coefficients less important in an analogous estimation as (3.27). On the other hand, from a practical point of view, it seems not desirable to include moments of higher than the fourths order. For an accurate evaluation, this would afford a huge number of scenario evaluations, that for practical considerations cannot be provided.*

(ii) A straight forward estimate for $\sigma(x_\eta^*)$ in (3.27) could be

$$\sigma_{\max} := \max_{i=1,...,n} \{\sigma(X_i)\},$$

since portfolio diversification will force $\sigma_{\max}$ to be an upper bound for $\sigma(x_\eta^*)$.

(iii) Using Equation 3.24 instead of 3.22 yields the same result for $c = 1$ and without referring to the very restrictive assumptions of Theorem 3.3.1.

(iv) Using equation 3.24 an analogous result with $c = 1$ in 3.27 can be achieved.

Proof: Using the assumptions of the lemma we can write for any $\varepsilon \in (0,1)$

$$\frac{\partial}{\partial x_i} \overline{VaR}_\varepsilon(x_\eta^*) = \mathbb{E}\left[X_i : X(x_\eta^*) = \overline{VaR}_\varepsilon(x_\eta^*)\right],$$

which itself can be approximated by

$$\overline{a}_i \cdot X^2 + \overline{b}_i \cdot X + \overline{c}_i = \overline{a}_i \cdot \left(\overline{VaR}_\varepsilon(x_\eta^*)\right)^2 + \overline{b}_i \cdot \overline{VaR}_\varepsilon(x_\eta^*) + \overline{c}_i.$$

The coefficients will have to be chosen in a way that $\overline{a}_i \cdot X^2 + \overline{b}_i \cdot X + \overline{c}_i$ represents the best quadratic predictor, i.e. minimizes the expression

$$\mathbb{E}\left[\left(X_i - \overline{a}_i \cdot X^2 - \overline{b}_i \cdot X - \overline{c}_i\right)^2\right],$$

which by expansion, differentiation with respect to $\overline{a}_i, \overline{b}_i, \overline{c}_i$ respectively and by setting the resulting equations equal to zero can easily be shown to be

minimized for the values

$$\begin{aligned}
\overline{a}_i &= \frac{\mathbb{E}(X^2)\cdot\mathbb{E}(X_iX^2)-\mathbb{E}(X^3)\mathbb{E}(X_iX)}{-\mathbb{E}(X^2)[\mathbb{E}(X^4)-\mathbb{E}(X^2)^2]+(\mathbb{E}(X^3))^2} \\
\overline{b}_i &= \frac{-\mathbb{E}(X_iX)[\mathbb{E}(X^4)-\mathbb{E}(X^2)^2]-\mathbb{E}(X^3)\mathbb{E}(X_iX^2)}{-\mathbb{E}(X^2)[\mathbb{E}(X^4)-\mathbb{E}(X^2)^2]+(\mathbb{E}(X^3))^2} \\
\overline{c}_i &= \frac{\mathbb{E}(X^2)[\mathbb{E}(X^2)\mathbb{E}(X_iX^2)-\mathbb{E}(X^3)\mathbb{E}(X_iX)]}{-\mathbb{E}(X^2)[\mathbb{E}(X^4)-\mathbb{E}(X^2)^2]+(\mathbb{E}(X^3))^2}
\end{aligned}$$

Here, we used $\mathbb{E}(X_i) = \mathbb{E}(X) = 0$ since without loss of generality all random variables can be chosen to have zero mean. Now, using the connection of VaR and $CVaR$ as described in section 2.4 we know

$$\begin{aligned}
\frac{\partial}{\partial x_i}\overline{CVaR}_{\zeta_0}(x_\eta^*) &= \mathbb{E}\left[X_i : X(x_\eta^*) \geq \overline{VaR}_{\zeta_0}(x_\eta^*)\right] \\
&= \frac{1}{1-\zeta_0}\int_{\zeta_0}^1 \frac{\partial}{\partial x_i}\overline{VaR}_\varepsilon(x_\eta^*)d\varepsilon \\
&\approx \frac{1}{1-\zeta_0}\int_{\zeta_0}^1 \overline{a}_i\cdot\left(\overline{VaR}_\varepsilon(x_\eta^*)\right)^2 + \overline{b}_i\cdot\overline{VaR}_\varepsilon(x_\eta^*) + \overline{c}_i\; d\varepsilon
\end{aligned}$$

Inserting these results into Equation (3.22) will yield

$$\begin{aligned}
c\cdot\frac{\partial}{\partial x_i}\overline{CVaR} &\quad \zeta_0(x_\eta^*) - \frac{\partial}{\partial x_i}\overline{VaR}_\eta(x_\eta^*) = \\
&= \frac{c\overline{a}_i}{1-\zeta_0}\int_{\zeta_0}^1 \left(\overline{VaR}_\varepsilon(x_\eta^*)\right)^2 d\varepsilon - \overline{a}_i\left(\overline{VaR}_\eta(x_\eta^*)\right)^2 \\
&\quad + \underbrace{c\overline{b}_i\overline{CVaR}_{\zeta_0}(x_\eta^*)}_{\overline{b}_i\overline{VaR}_\eta(x_\eta^*)} - \overline{b}_i\overline{VaR}_\eta(x_\eta^*) + \overline{c}_i - \overline{c}_i \\
&= \overline{a}_i\left[\frac{c}{1-\zeta_0}\int_{\zeta_0}^1 \overline{VaR}_\varepsilon^2(x_\eta^*)d\varepsilon - \overline{VaR}_\eta^2(x_\eta^*)\right]
\end{aligned}$$

which is the statement of the lemma. □

3.3.3 An illustrative example

In the preceding sections we showed that for α-stable distributed asset returns in principle and under some conditions on the confidence level, the

optimization of VaR and $CVaR$ are the same. This was mainly due to the fact that standardizing the involved random variables resulted in expressions that only depend on the one additional skewness parameter β. We want to use equation (3.26) to show that the equivalence of VaR and $CVaR$ optimization roughly holds true for the special case of independent h-distributed asset returns.

Suppose the assets $X_i, i = 1, ..., n$ are h-distributed and independent. It is a well known fact that the sum of two independent, symmetric and univariate random variables is again symmetric and univariate. Although there are some statements in the literature that the sum of h-distributed random variables is again h-distributed, we will at first show that this will generally fail to be true. Using equation (2.28), one easily sees that for any h-distributed random variable $Y \backsim Z \exp\left(\frac{hZ^2}{2}\right)$,

$$\begin{array}{ll} \mathbb{E}(Y) = 0 & \text{Skewness}(Y) = 0 \\ \sigma^2(Y) = \frac{1}{(1-2h)^{\frac{3}{2}}} & \text{Kurtosis}(Y) = 3\frac{(1-2h)^3}{(1-4h)^{5/2}} \end{array}$$

Given two h-distributed random variables Y_0, Y_1 (for simplicity we take in both cases the same value h), suppose the linear combination $Y_\lambda := \lambda Y_1 + (1-\lambda)Y_0$, $\lambda \in (0,1)$ is again h-distributed for some arbitrary value h. Considering the case $h = 0$, one clearly derives the case of standard normal distributions, hence admit the property of being closed under addition. Hence, in order to disprove closedness under taking the sum, we do not only have to allow for the parameters g, h but also the location parameter a and the scale parameter b to vary with the portfolio allocation. In the current setting, a represents some kind of mean value, which is equal to zero for all allocations λ. b can be chosen to match standard deviation (existing for $h < \frac{1}{2}$), thus resulting in a standardized random variable Y, which for the case $h = 0$ reduces to be truly deterministic. Allowing for $h \neq 0$, skewness is equal to zero by construction and the corresponding value h_λ can be constructed in order to match the kurtosis of each allocation λ given by the equations above. Assuming closedness under addition, the value h_λ and portfolio's standard deviation would completely determine the 80% -quantile, hence giving the same results as directly evaluating $VaR_{0.8}(\lambda)$ for every $\lambda \in (0,1)$ via Monte-Carlo simulations. Within Figure 3.10 we performed the necessary calculations and plotted the resulting 80%-quantiles using the different approaches. Comparing the two graphs, this obviously shows that the sum of h-distributed random variables is in general not h-distributed.

Figure 3.10 depicts the situation where both "corner" portfolios are h distrib-

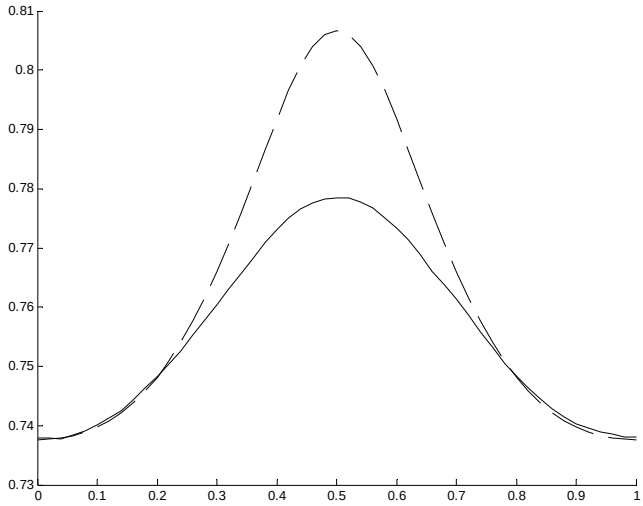

Figure 3.10: λ versus $VaR_{0.8}(\lambda)$. The solid line represents the standardized $VaR_{0.8}(\lambda)$ for different allocations λ using Monte-Carlo techniques with 3 million simulations. The dotted line firstly calculates h_λ of the standardized allocation and uses the analytic expression to express $VaR_{0.8}(\lambda)$.

uted with $h = 0.1$. Although we were using a huge number of Monte-Carlo simulations (3 million) to evaluate the 80% quantile, there is still a slight error in the evaluation, as can be seen in the difference between the two graphs at the endpoints $\lambda = 0$ and $\lambda = 1$. It is also interesting to see, that the 80% confidence level results in riskier allocations as we are diversifying the portfolio. This is due to the fact that we are considering standardized random variables. Although diversification results in lower kurtosis, this will only yield a lower VaR_η, if the confidence level η is sufficiently high. In the present situation for computational reasons, the confidence level was chosen to be relatively low, hence resulting in the qualitative behavior of figure 3.10.

Retaining the current setting, we will now have a closer look on the behavior of the respective Rayleigh quotient and its estimation via the smallest eigenvalue as used in the valuations above. Since the random variables Y_0, Y_1 are independent, it is clear that for diversificational reasons the corresponding VaR_η and $CVaR_\zeta$ problems always yield $x^*_\eta = x^*_\zeta = (\frac{1}{2}, \frac{1}{2})^t$ for any confidence levels $\eta, \zeta \in (0, 1)$. Hence the two optimizational problems are trivially equivalent. Evaluating the Rayleigh quotient as well as the smallest eigenvalues of $\overline{G}$ as defined in the proof to theorem (3.2.5) in an interval around the $CVaR$ optimal portfolio composition $x^*_\zeta = (\frac{1}{2}, \frac{1}{2})$ allows to get some hint on

the general behavior of the Rayleigh quotient. Moreover, we can estimate the overall error we make for only accounting for the smallest eigenvalues in order to approximate the real Rayleigh quotient.

To perform the above described approach, it is necessary to accurately evaluate second derivatives of the conditional value-at-risk. In order to do so, there are mainly two procedures, each of them having its own disadvantage. The first one takes the explicit form of the $CVaR$ derivative as stated in equation (2.18) and numerically evaluates the second derivative based on finite differences. Disregarding either the question of differentiability or the existence of a suitable density function, the second approach uses the following representation

$$\frac{\partial^2 CVaR_\zeta(x)}{\partial X_i^2} = \frac{1}{1-\zeta}\left(\sigma^2\left(X_i | X(x) = VaR_\zeta(x)\right) \cdot f_{X(x)}(VaR_\zeta(x))\right), \quad (3.28)$$

as developed in [49]. Although not developed in the setting of the cited literature, this representation could probably be extended to account for mixed derivatives as well. However, formulation (3.28) as in the case of the first derivative of VaR admits the problem of evaluating the standard deviation of the random variable X_i conditioned on

$$X(x) = VaR_\zeta(x).$$

Hence this approach has the numerical disadvantage of being unstable, since two different evaluations can result in quite different values for the corresponding standard deviation. Due to numerical stability, we used the approach of approximating the second derivative via finite differences. For some value x_η^* near x_ζ^*, we integrated second derivatives as in the definition of $\overline{G}$. From the resulting matrix we easily derived the corresponding Rayleigh quotient. The results are plotted in figure 3.11.

In Figure 3.11 we plot the Rayleigh coefficient as a function of the optimal $CVaR$ confidence level ζ_0^* as used in the various estimations of this thesis. We used some fixed optimal portfolio allocations as defined by $x_\eta^* = [0.45, 0.55]$ and $x_\zeta^* = [0.5, 0.5]$ as we made the experience that varying these values only has a minor effect on the qualitative statement. We made the calculations for different values $h \in \{0, 0.05, 0.1, 0.15\}$, using simulation sizes of 0.5 mil, 1 mil, 2 mil and 3 mil g-and-h distributed scenarios respectively. The reason for the successively increased number of simulations lies in the increased importance of tail modelling. Although using a huge number of simulations, this inaccuracy in the tail modelling of the respective distributions can still

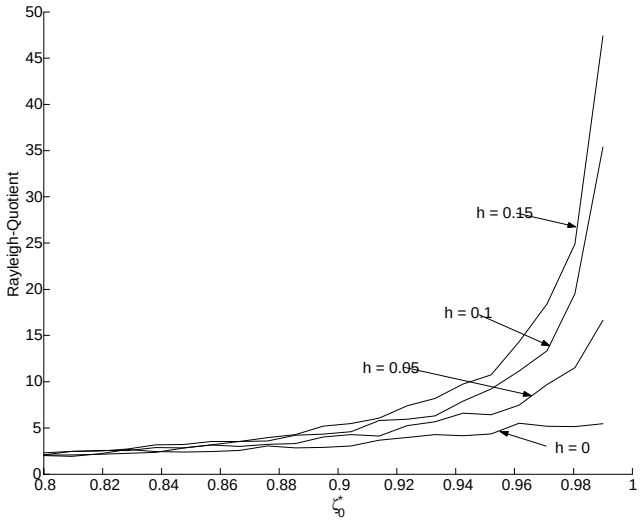

Figure 3.11: ς_0 versus the Rayleigh-Quotient as a function of ς_0 in the case of the linear combination of two i.i.d distributed r.v. Y_1, Y_2.

be observed in the nonsmooth behavior of the different graphs as depicted in Figure 3.11. There are actually two very important observations regarding the current investigation. Note, that in all the estimations above we were using the inverse of the Rayleigh coefficient in order to estimate the accuracy of our proposed algorithm. Hence, a larger value for the Rayleigh coefficient also yields better estimating results. As a higher value η for the VaR confidence level generally results in higher confidence levels of the affine $CVaR_{\zeta_0^*}$ problem (as e.g. can be seen by 3.25) we see that for increased values η we also get better estimations. Moreover, one observes that enlarging kurtosis of the VaR optimal portfolio generally results in better estimating results. All in all, as for practical considerations confidence levels η near one are of importance, our estimations will generally give quite good upper bounds.

We would like to stress again that the plots as given in Figure 3.11 are to be considered only in a qualitative way. Although we were using a huge number of simulations, a recalculation based on a different random vector of same size might result in quite different evaluations. Moreover, using such a number of scenarios, also the accuracy in evaluating the respective integrals has to be lowered. Otherwise computation time appears unreasonable long.

As we have seen in the proceeding examples it is numerically quite expensive to evaluate Rayleigh quotients since this involves the integration of sec-

ond derivatives of $CVaR$, hence for accurate evaluations a huge number of Monte-Carlo simulations has to be generated. We will now restrict to the case $h = 0$, since then we are dealing with normally distributed allocations for which there are explicit expressions for the $CVaR$ gradient. Therefore, the necessary calculations can be performed without reference to Monte-Carlo simulations. The next theorem summarizes the main formulas for the gradient expression of $CVaR$.

Theorem 3.3.4 *Under the usual notational conventions used in this paper and denoting by $\mu = \mathbb{E}([X_1,...,X_n]^t)$ and Ω the mean vector respectively the covariance matrix of normally distributed random variables $X_i, i = 1,...,n$, the following analytic representations of the $CVaR$ gradient hold true*

$$\begin{aligned} \nabla CVaR_\zeta(x) &= -\mu + \frac{\Omega x}{(x'\Omega x)^{1/2}} \frac{\varphi(z(\zeta))}{1-\zeta} \\ &= -\mu + \frac{\Omega x}{x'\Omega x} \left(CVaR_\zeta(x) + \mu^t x\right) \\ &= -\mu + \frac{\Omega x}{(x'\Omega x)^{1/2}} \frac{\varphi\left(\frac{VaR_\zeta(x)+\mu^t x}{(x'\Omega x)^{1/2}}\right)}{(1-\zeta)}. \end{aligned} \tag{3.29}$$

Here, φ denotes the normal probability density function, whereas $z(\zeta)$ describes the ζ-quantile of a standard normal distributed random variable.

Proof: The stated results can be found in ([55]). □

We used equation (3.29) to evaluate the corresponding Rayleigh quotient in the above outlined context, i.e. in the two asset spanning portfolio optimization problem, where the two assets X_1, X_2 are i.i.d. standard normal distributed. However, we also allowed for different covariance matrices in depicting our results in Figure (3.12). More concrete, within Figure (3.12), we depict the Rayleigh quotient as a function of the value ζ_0^*, the optimal $CVaR$ confidence level satisfying some additional property. Within this plot, the solid line represents covariance matrix $\Omega_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, whereas $\Omega_1 = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \Omega_2 = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix}$ are represented by the dotted respectively the dashed line. During our numerical analysis we found that for any confidence level ζ_0^* the corresponding Rayleigh coefficient approaches 0 as the off diagonal of Ω tends to 1. Another important feature is the

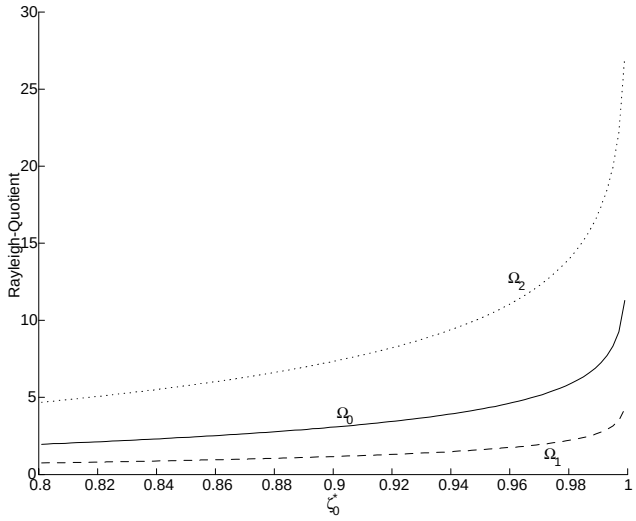

Figure 3.12: ζ_0^* versus the Rayleigh quotient R. The figure plots R as a function of the optimal confidence level ζ_0^* for the situation $x_\eta^* = (0.45, 0.55)$ and $x_\zeta^* = (0.5, 0.5)$ and for different covariance matrices.

strongly increasing Rayleigh coefficient for ζ_0^* getting closer to one, resulting in a successively decreasing value for R^{-1}, hence a more and more restricting inequality (3.24) as ζ_0^* increases.

Although we know that under the present assumptions on the distribution functions $x_\eta^* = x_\zeta^* = [0.5, 0.5]$ for explanatory reasons we will vary x_η^* to achieve the depicted results with $x_\eta^* = [0.45, 0.55]$ and $x_\zeta^* = [0.5, 0.5]$. Further variations only result in marginal differences which do not change the qualitative behavior of the depicted calculations. Therefore, we passed on sketching those results.

Earlier in this work, we used the smallest eigenvalue to approximate the corresponding Rayleigh quotient. However, during the analysis of the present setting of this section, the smallest eigenvalue appears to be a poor estimator since it is either zero or close to zero. Taking the inverse results in an almost unbounded RHS in equation (3.24).

Taking the knowledge about the Rayleigh coefficient in the normal distributed case as a projection of the behavior for arbitrary distributions in an arbitrary sized setting, we conclude that for sufficiently high confidence levels η, R^{-1} generally will attain values that give good upper bounds on the distance $\left\|x_\eta^* - x_\zeta^*\right\|$. Allowing for some kind of dependence structure among

the different asset classes, negative correlation will generally result in better approximations whereas highly correlated assets only allow for a slightly worse approximation. However, an adequate choice of η will continue to hold an favored value R^{-1}.

Chapter 4

Stochastic Branch & Bound

4.1 General Method

In contrast to the procedure of the last chapter we want to describe an application of a truly stochastic algorithm specially adapted to the case of optimizing VaR. Not only in the stochastic but also in the deterministic case, effectively implementing a branch and bound algorithm depends crucially on three problem related properties. These are the structure of the current objective, the choice of partitioning the set of feasible solutions $\mathbb{Q}$ into smaller subsets Z and an effectively enumerable set of upper and lower bounds for the removal of suboptimal subsets. In order to solve a stochastic optimization problem, we will have to impose certain restrictions to achieve meaningful results. In what follows, we are mainly referring to the work of Norkin et al. ([43]).

Considering the objective function within the stochastic branch and bound algorithm, we demand it to be of the special form

$$\min_{x \in \mathbb{Q} \cap D} \mathbb{E}\left[g(x, X(\omega))\right], \tag{4.1}$$

again writing $X(\omega)$ for some n-dimensional random vector and $\mathbb{Q}$ defined to be the unit cube $[0,1]^n \subset \mathbb{R}^n$. In contrast to the foregoing analysis we allow further restrictions on the set of feasible solutions to be described by a closed subset of $\mathbb{R}^n$ denoted by D. Hence, at least the requirement of being fully invested will have to be incorporated using the set D. The function $g : \mathbb{Q} \times D \to \mathbb{R}$ is assumed to be continuous in the first argument and measurable

in the second one and we want the expectation operator $\mathbb{E}$ to be well defined. Restricting to the form (4.1), one might expect this kind of objective function only to address some minor problems. However, the authors in [43] give some interesting applications to real world problems, including the stochastic tackling of a multiperiodic portfolio allocation problem under the additional treatment of transaction costs. Moreover, as we will see in the next section, it is exactly this form that allows for the efficient (in the context of branch and bound algorithms) solution of the VaR optimization problem as described in Chapter 1.

The second very important characterization is the branching rule of successively dividing the original compact set $\mathbb{Q}$ into smaller subsets $Z \subset \mathbb{Q}$ generating a partition $\mathcal{P}$ of $\mathbb{Q}$, s.t. $\cup_{Z \in \mathcal{P}} Z = \mathbb{Q}$, hence resulting in subsequently refined optimization problems (4.1), $\mathbb{Q}$ being replaced by the corresponding compact set Z. A specialized branching method in the context of VaR optimization will be considered within the next section. For more details on the subdivision of the compact set $\mathbb{Q}$ consult any book on deterministic branch and bound algorithms (compare e.g. [33]).

The third characterization of any branch & bound algorithm is where the stochastic of the whole process comes in. In order to effectively exclude designated sets Z in the bounding process the right choice of lower and upper bounds becomes crucial. In the next section we will develop such stochastic bounds based on the interchange relaxation, hence satisfying all the assumptions on yielding meaningful upper and lower stochastic bounds for any subset Z of the partition $\mathcal{P}$. These bounds are supposed to be random variables $\overline{U}(Z, X(\omega))$ respectively $\overline{L}(Z, X(\omega))$ defined on the collection of all compact $Z \subseteq \mathbb{Q}$ with $Z \cap D \neq \emptyset$ and such that

$$\begin{aligned} L(Z) &: = \mathbb{E}\left[\overline{L}(Z, X(\omega))\right] \\ U(Z) &: = \mathbb{E}\left[\overline{U}(Z, X(\omega))\right] \end{aligned}$$

are real-valued functions, again well-defined on the same domain as $\overline{U}$, $\overline{L}$.

4.1.1 The Algorithm

Depending on the use of different branching techniques, in the theory of deterministic branch and bound optimization one distincts between different graph search algorithms such as e.g. the Depth-First Search, the Breadth-First Search or the Best-First Search (compare e.g. [10]). These approaches highly

differ in the successive partitioning process, hence resulting in quite different numerical costs of finding the optimal solution. However, increased numerical efficiency unfortunately comes along with reduced significance about the quality of the achieved solution. Moreover, besides the drawback of yielding a possibly bad estimate to the true solution, the implementation of those efficient search algorithms does not give any hint on the sensitivity of the proposed solution. Although in [43] the authors already choose a very broad way of branching the region, within the following algorithm we are describing the broadest possible way of branching any region of interest. At every stage of the algorithm we are partitioning every active set, only disregarding subsets if they either do not belong to the set of feasible solutions or the corresponding lower bound increases the minimum upper bound taken over all feasible subsets.

Initialization. Form an initial partition $\mathcal{P}_1 = \{\mathbb{Q}\}$. Observe independent lower and upper bounds, $\overline{L}_1(\mathbb{Q}, X(\omega))$ respectively $\overline{U}_1(\mathbb{Q}, X(\omega))$. Set $k = 1$. Before iteration k we have partition $\mathcal{P}_k$ and bound estimates $\overline{L}_k(Z, X(\omega))$, $\overline{U}_k(Z, X(\omega))$, $Z \in \mathcal{P}_k$.

Partitioning. For every $Z \in \mathcal{P}_k$ construct a partition

$$\Psi_k(Z) = \{Y_i,\ i = 1, 2, ...\}$$

such that $Z = \cup_i Y_i$. Define a new full partition

$$\mathcal{P}_k' = \{\Psi_k(Z) : Z \in \mathcal{P}_k\}$$

as well as approximate solutions $x_k \in Y_k$,

$$Y_k \in \arg\min \left\{\overline{U}_k(Z, X(\omega)) : Z \in \mathcal{P}_k\right\}.$$

Deletion. For some positive value c define

$$\mathcal{S} := \left\{Z \in \mathcal{P}_k' : \overline{L}_k(Z, X(\omega)) > \overline{U}_k(Y_k, X(\omega)) + c\right\}$$

and clean partition $\mathcal{P}_k'$ of suboptimal and non-feasible subsets, defining

$$\mathcal{P}_{k+1} := \mathcal{P}_k' \setminus \left\{\mathcal{S} \cup \left\{Z \in \mathcal{P}_k' : Z \cap D = \emptyset\right\}\right\}.$$

Bound estimation. For all $Z \in \mathcal{P}_{k+1}$ observe random variables $\overline{L}_{k+1}(Z, X(\omega))$, independently observe $\overline{U}_{k+1}(Z, X(\omega))$ and set $k := k + 1$.

Go to **Partitioning.**

On some additional assumptions on the stochastic bounds (which can be shown to hold true for the bounds presented in the next section), Norkin et al. achieve not only on showing the convergence of the described algorithm but also give an estimation of goodness as stated in the following lemma.

Lemma 4.1.1 *Suppose a uniform bound σ^2 is known for the variances of all random variables $\overline{L}_k(Z, X(\omega))$, $\overline{U}_k(Z, X(\omega))$, $Z \in \cup_k \mathcal{P}_k$, $k = 1, 2, ...$ and let x^* be a solution to (4.1). Then*

$$\mathbb{P}\{x^* \text{ is lost in the final deletion}\} \leq 2\frac{\sigma^2}{c^2}.$$

4.2 Application to VaR Optimization

Adoption of the Objective Function. As already formulated in the previous section, the proposed algorithm heavily depends on a special structure of the objective function. To adapt the current algorithm to the case of VaR optimization we firstly want to recall the definition of $CVaR$ as given by the authors Uryasev and Rockafellar. Not only restricting to the case

$$f(x, X) = X^t \cdot x =: X(x)$$

(here f describes the loss function) but allowing for any function $f(x, X)$ that is continuous in the first and measurable in the second argument with $\mathbb{E}[|f(x, X)|] < \infty$. Allowing for such an arbitrary behavior of the loss function the $CVaR$ can be shown to be alternatively defined by some minimization problem (compare Equation 2.4). Based on this result we want to define

$$CVaR_\zeta^m(x) := \inf_{a \in \mathbb{R}} \left\{ a + \frac{1}{1-\zeta}\mathbb{E}\left[\min\left\{X^t x; m\right\} - a\right]^+ \right\},$$

and thus allowing the losses to be given by

$$f_m(x, X(\omega)) := \min\left\{X^t x; m\right\}. \tag{4.2}$$

Using this definition of losses, we are not interested in those losses exceeding some level m; all the occurring loss distributions admit a cut at level m, thereby not changing the original behavior for losses less or equal to m. Figure 4.1 shows how this approach is acting on $VaR_\alpha(x)$ as a function of

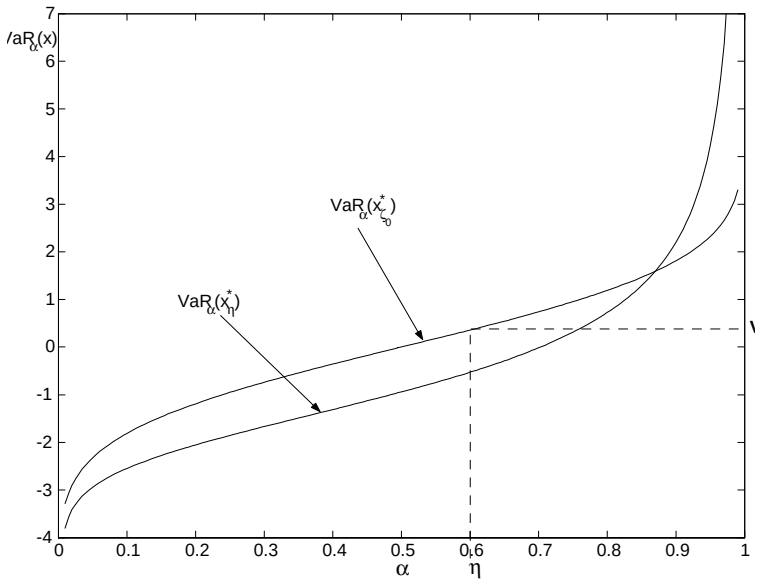

Figure 4.1: α versus VaR_α. Cutting VaR_α in order to apply the stochastic branch and bound algorithm as described in the text.

α. Since there are no losses exceeding level m, the corresponding cut results in a VaR graph that always exhibits a constant part for sufficiently high confidence levels.

In principle, m can be chosen to admit any real value. However, in what follows, it is of practical interest to choose the value m to be close to the optimal value of the solution to the corresponding VaR problem. In particular, m could be chosen to match the VaR value given by the output of some foregoing heuristic algorithm. Using the notation of the previous chapters we could therefore choose

$$m := VaR_\eta(x^*_{\zeta_0}).$$

It is clear, that by the definition of m, this results in the following estimations for $CVaR^m_\eta(x)$

$$CVaR^m_\eta(x) \begin{cases} < m & \text{if } VaR_\eta(x) < VaR_\eta(x^*_{\zeta_0}) \\ = m & \text{if } VaR_\eta(x) \geq VaR_\eta(x^*_{\zeta_0}) \end{cases}.$$

Based on some (heuristic) guess for the optimal VaR_η portfolio, the further stochastic optimization of the $CVaR^m_\eta(x)$ objective either results in an overall

constant function always being equal to m (in the case, where the heuristic already found an optimal allocation) or it identifies more promising regions among the set of feasible solutions.

Defining

$$\begin{aligned} g_m : \mathbb{Q} \times \mathbb{R} \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x, a, X) &\longmapsto a + \frac{1}{1-\eta} \left[\min\left\{X^t x; m\right\} - a\right]^+, \end{aligned}$$

our objective is to solve the minimization problem

$$\min_{x \in \mathbb{Q} \cap D, a \in \mathbb{R}} \mathbb{E}\left[g_m(x, a, X(\omega))\right],$$

which already has got the form treatable by the stochastic branch and bound algorithm described in the last section.

Remark 4.2.1 *Cutting every loss function according to the definition in (4.2) allows to interpret $CVaR^m$ as the weighted integral over the VaR as a function of its confidence level. However, $CVaR^m$ appears to loose the convexity property of $CVaR$ thereby resulting in a non-convex, hence generally not easy to solve optimization problem. In this sense, the complexity of solving $CVaR^m$ lies between the one of VaR and $CVaR$. Using the expectation operator in the definition of the objective function gives some smoother behavior compared to VaR but not enough to hold up convexity.*

Stochastic Bounds. Let us now consider the question of using stochastic upper and lower bounds such that the different assumptions as stated in [43] are fulfilled. As an upper stochastic estimate of

$$F^*(Z \cap D) := \min_{x \in Z \cap D, a \in \mathbb{R}} \mathbb{E}\left[g_m(x, a, X(\omega))\right]$$

one can use Monte-Carlo estimates of the objective function at some feasible point $x \in Z \cap D$.

The definition of lower stochastic bounds appears to be more sophisticated but in general there could be used any deterministic bounding technique from global optimization. However, in the context of the optimized branching method of the next paragraph we will describe a very natural lower stochastic

bound. Interchanging the minimization and the expectation operator in the definition of $F^*(Z \cap D)$ yields

$$\begin{aligned} F^*(Z \cap D) &= \min_{x \in Z \cap D, a \in \mathbb{R}} \mathbb{E}\left[g_m(x, a, X(\omega))\right] \\ &\geq \mathbb{E} \min_{x \in Z \cap D, a \in \mathbb{R}} g_m(x, a, X(\omega)). \end{aligned}$$

Thus defining the deterministic lower bound

$$L(Z) := \mathbb{E} \min_{x \in Z \cap D, a \in \mathbb{R}} g_m(x, a, X(\omega))$$

and its stochastic analogon based on Monte-Carlo estimates $\widehat{X}^j(\omega)$

$$\overline{L}(Z, \widehat{X}^j) := \frac{1}{J} \sum_{j=1}^{J} \min_{x \in Z \cap D, a \in \mathbb{R}} g_m(x, a, \widehat{X}^j).$$

The next lemma shows that the minimization problems in the definition of $\overline{L}$ are quite efficiently solvable.

Lemma 4.2.1 *The joint minimization of*

$$G_m^*(Z \cap D) := \min_{x \in Z \cap D, a \in \mathbb{R}} g_m(x, a, \widehat{X}^j)$$

in (x, a) can be performed by firstly minimizing over $x \in \mathbb{Q}$, followed by minimizing over $a \in \mathbb{R}$

$$\begin{aligned} G_m^*(Z \cap D) &= \min_{x \in Z \cap D, a \in \mathbb{R}} \left\{ a + \frac{1}{1-\eta} \left[f_m(x, \widehat{X}^j) - a \right]^+ \right\} \\ &= \min_{a \in \mathbb{R}} \left\{ a + \frac{1}{1-\eta} \left[\min_{x \in Z \cap D} f_m(x, \widehat{X}^j) - a \right]^+ \right\} \\ &= \mathbb{E}\left[\min_{x \in Z \cap D} f_m(x, \widehat{X}^j) \mid \min_{x \in Z \cap D} f_m(x, \widehat{X}^j) \geq m \right] \end{aligned}$$

Proof: We first note that

$$h_y(a) := a + \frac{1}{1-\eta} \left[y - a \right]^+$$

attains its minimum for $y = a$ for any fixed $y \in \mathbb{R}$ [$h_a(a+\varepsilon) = a + \frac{1}{1-\eta}\left[a - a - \varepsilon\right]^+ \geq a = h_a(a)$ and $h_a(a-\varepsilon) = a + \frac{1}{1-\eta}\left[a - a + \varepsilon\right]^+ \geq a = h_a(a)$]. Thus minimizing $h_y(a)$ in both y and a reduces to the finding of a minimum feasible point $y \in \mathbb{R}$ with the consequence that

$$\min_{a,y}\left\{a + \frac{1}{1-\eta}\left[y-a\right]^+\right\} = \min_{a}\left\{a + \frac{1}{1-\eta}\left[\left\{\min_{y} y\right\} - a\right]^+\right\}.$$

Substituting y by some arbitrary function $f(x, X(\omega))$ with values in $\mathbb{R}$ yields the claim of the lemma. The last equality follows from the $CVaR$ representation as given in ([53]). The only conditions on

$$\min_{x \in Z \cap D} f_m(x, \widehat{X}^j)$$

are the continuity in x and the measurability in $X(\omega)$, which are both trivially satisfied. □

The corresponding minimization problem which we will have to solve (for a possibly large number of scenarios) appears to be of linear type, hence easy to solve. By the special choice of subsets $Z \subset \mathbb{Q}$ as described in the next paragraph the numerical complexity of evaluating these minimization problems can considerably be reduced.

A Special Branching Method. The most common optimization problem in portfolio theory uses a setting where the most rudimentary set of feasible allocations consists of the compact set $\mathbb{X}$ as defined in (2.1). This restriction, together with more sophisticated restrictions such as the demand for some minimum expected return are collected in the closed set D and do not have to be of further analysis. In what follows, we are restricting to the branching process of the very basic set given by $\mathbb{Q}$.

In fact, the proposed branching method is nothing special at all. As can be seen in Figure 4.2, the partition $\Psi_k(Z)$ as in the description of the branch and bound algorithm is determined by successively refined n-dimensional cubes which result from successively dividing every edge into two pieces of equal length, thus resulting in 2^n new cubes for every partitioned cube. The great advantage of this procedure stems out of the ease of handling the huge number of linear optimization problems as described in the last section. Using some available solver for linear optimization one firstly searches for the optimal vertex within the original set $\mathbb{Q} = [0,1]^n$ by solving

$$\min_{x \in \mathbb{Q}} \left(\widehat{X}^j\right)^t \cdot x \tag{4.3}$$

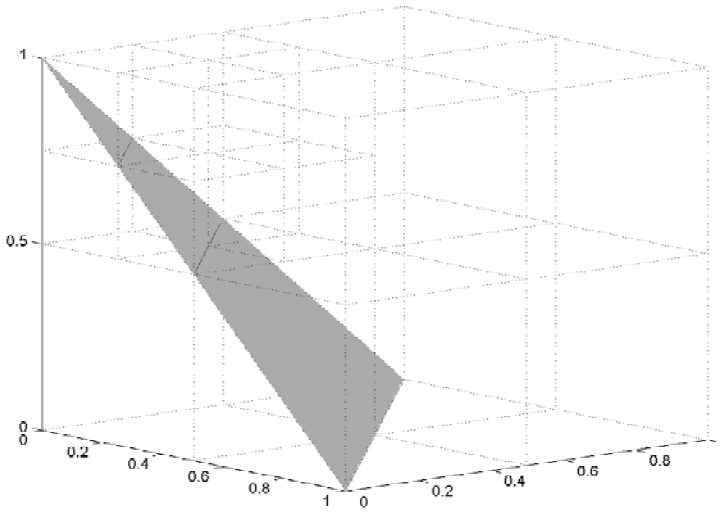

Figure 4.2: How to partition the search space $\mathbb{Q}$.

for any representation $\widehat{X}^j, j = 1, ..., J$ of $X(\omega)$ in order to yield

$$\min_{x \in \mathbb{Q}} f_m(x, \widehat{X}^j).$$

As the proposed branching method results in suitable scaled and/or transformed cubes of the original set $\mathbb{Q}$, the respective optimal vertex of

$$\min_{x \in Z} f_m(x, \widehat{X}^j)$$

can easily be attained by evaluating the corresponding vertex as given by the initializing optimization problem 4.3. By successively excluding all the non-feasible sets, the algorithm can be efficiently implemented for a possibly high number of scenarios. However, for a higher number of investment alternatives (n greater than 10, largely depending on the computer power used), the resulting algorithm also becomes inefficient due to the full evaluation of the whole region as proposed in the algorithm above. By restricting the search to the most promising regions of interest one might result in more efficient algorithms, thereby allowing for an even higher number of asset classes to solve for.

4.3 An Analysis of the Algorithm's Efficiency

Using a setting of normally and independently distributed asset classes it is relatively easy to find the true VaR minimum with the help of quadratic optimization techniques. On the other hand, applying the proposed algorithm to this setting gives us the possibility to analyse the algorithm's performance in finding the (theoretically) true minimum. To be more precise, in our analysis we are considering different n-dimensioned problem settings, the respective means always given as a subset of the first n values out of the set $\{0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$. Moreover, the respective standard deviations are chosen to be equal to the securities' mean, therefore accounting for the fact that a rise in expected returns generally comes along with an increase in the asset's inherent risk. The parameter setting was chosen arbitrarily; its economic relevancy is of no interest for the results of the analysis, which are given in the next table.

Dim	Depth	Time	Active Set	MaxDist	MinVaR	MaxVaR	OptVaR
3	3	5s	0.0566	0.4467	0.2635	0.3874	0.3597
	4	5s	0.0164	0.3233	0.2991	0.3824	0.3597
	5	7s	0.0048	0.2859	0.3177	0.3748	0.3597
	6	8s	0.0016	0.2505	0.3243	0.3698	0.3597
	7	14s	$5.79\ 10^{-4}$	0.2240	0.3313	0.3650	0.3597
	8	30s	$2.45\ 10^{-4}$	0.2236	0.3321	0.3627	0.3597
4	3	10s	0.0244	0.4771	0.1929	0.3943	0.3001
	4	12s	0.0047	0.3697	0.2444	0.3372	0.3001
	5	23s	$7.12\ 10^{-4}$	0.3288	0.2676	0.3211	0.3001
	6	50s	$1.18\ 10^{-4}$	0.2378	0.2791	0.3128	0.3001
	7	128s	$2.44\ 10^{-5}$	0.2094	0.2872	0.3091	0.3001
5	3	12s	0.01	0.5266	0.1509	0.3914	0.2540
	4	45s	0.0019	0.3971	0.1855	0.3129	0.2540
	5	240s	$2.79\ 10^{-4}$	0.3302	0.2110	0.2943	0.2540
	6	934s	$4.71\ 10^{-5}$	0.2956	0.2228	0.2891	0.2540
6	3	28s	0.0032	0.5159	0.1337	0.3914	0.2154
	4	247s	$6.29\ 10^{-4}$	0.4178	0.1461	0.3074	0.2154
	5	2791s	$9.12\ 10^{-5}$	0.3565	0.1651	0.2727	0.2154
7	3	80s	$8.48\ 10^{-4}$	0.5229	0.1103	0.3943	0.1811
	4	1243s	$1.27\ 10^{-4}$	0.4062	0.1139	0.3000	0.1811
8	3	241s	$2.61\ 10^{-4}$	0.5367	0.0789	0.4128	0.1496
	4	7618s	$3.46\ 10^{-5}$	0.4344	0.0751	0.3000	0.1496
9	3	882s	$7.00\ 10^{-5}$	0.5835	0.0789	0.4864	0.1199

VaR calculations are based on the 99% quantile using a scenario representation of 2000 paths. The first column describes the dimensionality of the corresponding Markowitz optimization problem, followed by the number of consecutive decompositions of the original set of feasible solutions into smaller subsets. Column three shows the elapsed time for searching the space by successively dividing into smaller subsets. Clearly, the computational time for doing so increases in both, the dimension of the original problem setting and the search depth. The numbers as stated in columns four up to seven give some hint on the accuracy of the search-space reduction and how far we are lying off from the true VaR optimum as given in column eight.

The values given in the "Active Set"-column are defined as the proportion of the combined cube volume of those cubes not yet discarded within n-dimensional space. It is intuitively clear that this number strongly decreases for an increased number of asset classes.. Moreover, increasing the number of subdivisions as given by column "Depth" results in much more adequate volume reductions.

By calculating the figures in columns five through seven we always calculated the respective figure at a predefined fixed corner of the corresponding subcube thereby underestimating the true "MinVaR" and overestimating "-MaxVaR". This becomes especially apparent for those rows which correspond to a relatively low degree of search depth.

Although we propose a very efficient way of solving the resulting subproblems, table **4.3** shows that the use of the described algorithm for solving the VaR optimization at hand is only practical for a moderate number of asset classes. However, it seems reasonable to combine the foregoing approach with the use of some meta heuristic optimization procedure in order to achieve faster approximative results. Another way of increasing the number of realistically solvable dimensions could be in the algorithm's implementation on several computers.

It is important to note another reason for the relatively high execution times. In the implementation of the current algorithm a subset of a former active set was rejected if and only if the corresponding lower bound equals the "cutting parameter" $m := OptVaR$. However, since we only used 2000 scenarios to represent the current setting, the corresponding value of the VaR optimum could significantly deviate from the true value $OptVaR$. Thus an increase of the used scenarios to simulate the optimization problem could result in a more efficient rejection policy. However, increasing the number of scenarios also results in an increase of computational complexity.

Chapter 5

Conclusion and Further Investigations

In the course of this thesis we tried to examine the approximation of the original NP-hard VaR optimization problem via a relatively easy to solve affine problem, namely of finding the optimal solution to the optimization problem where the VaR objective is replaced by the $CVaR$ risk measure. We saw that there are very general classes of distribution functions, where the described procedure results in an exact identification of the VaR optimum. However, for general loss distributions the quality of the approximation is highly dependent on the interplay of the loss distributions' moments. If the mean as well as the standard deviation are the main drivers of the overall behavior, i.e. if the behavior of the standardized $\overline{VaR}$ (as a function of portfolio composition) can be neglected, our proposed approach will automatically result in good outcomes.

Another driver for the accurateness of the proposed proceeding seems to be the problem's inherent dependency structure. Generally speaking, a good diversification between asset classes will also result in good approximations to the VaR optimum through an equivalent $CVaR$ optimum. This mainly stems out of the fact that for a setting with good diversification effects, the Hermitian matrix of $CVaR_{\varsigma_0}(x)$ in the optimal portfolio allocation $x^*_{\varsigma_0}$ admits some non-negligible Rayleigh-Quotient.

However, besides proving the result for special classes of distribution functions, we were also able to provide upper and lower bounds on the euclidean distance of the preferred element of $\mathbb{X}^*$ to the true VaR optimum x^*_η in a most general setting. Independent of the problem's nature, this enables us

to show that for sufficiently high confidence levels the approximation is also quite good. In particular, when dealing with distribution functions that do not heavily deviate from the normal distribution (as in the case of a well diversified portfolio incorporating a large number of asset classes, where the optimal portfolio composition is also expected to consist of a large number of asset classes) our approach to the VaR optimization problem will yield sufficient good results.

Besides the justification of approximating the VaR optimum by a sequence of affine $CVaR$ optimizations, in Chapter 3 we gave a short overview of a specially tailored stochastic branch and bound algorithm. While also giving good approximations to the VaR optimization problem, this approach is for computational reasons restricted to a moderate number of asset classes.

Besides the results of this thesis, there are some important extensions for further investigation. At first, it would be of general interest not to only look at the stand alone portfolio allocation problem, but to also allow for liabilities being part of the investor's portfolio (compare [29]). Within the $CVaR$ objective setting, this problem was already successively solved. However, it is not entirely clear if the additional consideration of liabilities disturbs the nice approximation results of this thesis.

Another development of interest would be the extension of the original VaR objective function to the one, where the investor is interested in optimizing the linear combination of $VaRs$ for different confidence levels. This would be done to represent some kind of spectral risk and would also involve relatively low confidence levels (in the sense of this thesis). Also a very interesting extension of the present setting would be the application to multi periodic optimization settings.

Besides these more general developments in extending the present problem setting, further research on the results of this thesis are imaginable. Not to mention all, it would be of some interest to investigate the exact relationship between the set of $\eta \in (0,1)$ for which Assumption $\mathrm{A}^{ST}/\mathrm{A}^{g,h}$ is true and the set of confidence levels, for that VaR_η proves to be subadditive. Moreover, further insights to the behavior of the proposed algorithm could be achieved by a deeper understanding of the behavior of the two integrals I_1 and I_2 in (3.27).

Appendix A

Abbreviations and Notations

x	Vector of portfolio compositions as percentages of the whole portfolio
$\mathbb{X}$	Set of possible portfolio allocations
$\mathbb{Q}$	$:= [0,1]^n$.
$\mu(x)$	Expected value of the loss distribution of portfolio composition x
$\sigma(x)$	Standard deviation of the loss distribution of portfolio composition x
$VaR_\eta(x)$	Value-at-Risk of portfolio composition x for convidence level η
$CVaR_\zeta(x)$	Expected shortfall of portfolio composition x for convidence level ζ
$VaR_\eta^{g,h}(x)$	Value-at-Risk for g-and-h distributed losses
$CVaR_\zeta^{g,h}(x)$	Expected shortfall for g-and-h distributed losses
$\overline{VaR}_\eta(x)$	$:= \sigma^{-1}(x) \cdot (VaR_\eta(x) - \mu(x))$
$\overline{CVaR}_\zeta(x)$	$:= \sigma^{-1}(x) \cdot (CVaR_{\zeta_0}(x) - \mu(x))$
x_η^*	Globally optimal portfolio for the $VaR_\eta(x)$-optimization
$x_{\zeta_0}^*$	Globally optimal portfolio for the $CVaR_{\zeta_0}(x)$-optimization
$X(x)$	$:= X^t \cdot x$, random number representing portfolio allocation x
X	$:= (X_1, ..., X_n)^t$, random vector
X_i	Random number, representing asset i
$\widehat{X}^j$	j-th realization of random vector X
$\widehat{X}^j(x)$	$:= \left(\widehat{X}^j\right)^t \cdot x$, j-th realization of random vector $X(x)$
$S_\alpha(\beta, \gamma, \delta)$	Set of α-stable distributed random variables with parameters β, γ, δ
$f_0(t; \alpha, \beta)$	Density fct. of an α-stable r.v. with parameter β using the 0-param.
$F_0(t; \alpha, \beta)$	Distribution fct. of an α-stable r.v. with parameter β using the 0-param.
$f_0(t; \alpha, \beta)$	Density fct. of an α-stable r.v. with parameter β using the 1-param.
$F_1(t; \alpha, \beta)$	Distribution fct. of an α-stable r.v. with parameter β using the 1-param.
$\mathbb{R}^+$	Set of positive reell numbers
e_i	Vector with the i-th entry equal to 1, otherwise zeros
GH	Set of g-and-h distributions
$[t]^+$	$:= \max\{t; 0\}$

Appendix B

Subadditivity of VaR for Stable Distributions

$\alpha = 1.5$	γ_1	0.1				0.3			0.5		0.7
	γ_2	0.1	0.3	0.5	0.7	0.3	0.5	0.7	0.5	0.7	0.7
β_1	β_2										
−0.8	−0.8	.357	.357	.357	.357	.357	.357	.357	.357	.357	.357
	−0.6	.371	.365	.363	.361	.371	.368	.366	.371	.369	.371
	−0.4	.388	.374	.369	.367	.388	.381	.377	.388	.383	.388
	−0.2	.408	.385	.377	.373	.408	.396	.389	.408	.400	.408
	0.2	.450	.408	.394	.387	.450	.429	.416	.450	.436	.450
	0.4	.470	.418	.402	.393	.470	.444	.429	.470	.452	.470
	0.6	.486	.428	.409	.399	.486	.457	.439	.486	.467	.486
	0.8	.500	.435	.414	.403	.500	.467	.448	.500	.478	.500
−0.6	−0.6	.385	.385	.385	.385	.385	.385	.385	.385	.385	.385
	−0.4	.402	.395	.392	.391	.402	.398	.396	.402	.400	.402
	−0.2	.422	.405	.400	.397	.422	.413	.408	.422	.416	.422
	0.2	.464	.428	.417	.411	.464	.446	.435	.464	.452	.464
	0.4	.483	.439	.424	.417	.483	.461	.448	.483	.468	.483
	0.6	.500	.448	.431	.423	.500	.474	.458	.500	.482	.500
	0.8	.514	.455	.437	.427	.514	.484	.467	.514	.494	.514
−0.4	−0.4	.419	.419	.419	.419	.419	.419	.419	.419	.419	.419
	−0.2	.439	.430	.427	.426	.439	.434	.432	.439	.436	.439
	0.2	.481	.453	.444	.439	.481	.467	.458	.481	.471	.481
	0.4	.500	.463	.452	.446	.500	.481	.471	.500	.488	.500
	0.6	.517	.472	.458	.451	.517	.494	.481	.517	.502	.517
	0.8	.530	.480	.464	.456	.530	.505	.490	.530	.513	.530
−0.2	−0.2	.458	.458	.458	.458	.458	.458	.458	.458	.458	.458
	0.2	.500	.481	.475	.472	.500	.490	.485	.500	.494	.500
	0.4	.519	.492	.483	.478	.519	.505	.497	.519	.510	.519
	0.6	.536	.501	.489	.484	.536	.518	.508	.536	.524	.536
	0.8	.550	.508	.495	.488	.550	.529	.516	.550	.536	.550
0.2	0.2	.542	.542	.542	.542	.542	.542	.542	.542	.542	.542
	0.4	.561	.552	.550	.548	.561	.557	.554	.561	.558	.561
	0.6	.578	.562	.556	.554	.578	.570	.565	.578	.573	.578
	0.8	.592	.569	.562	.558	.592	.581	.574	.592	.584	.592
0.4	0.4	.581	.581	.581	.581	.581	.581	.581	.581	.581	.581
	0.6	.598	.590	.588	.586	.598	.594	.591	.598	.595	.598
	0.8	.612	.598	.593	.591	.612	.605	.600	.612	.607	.612
0.6	0.6	.615	.615	.615	.615	.615	.615	.615	.615	.615	.615
	0.8	.629	.622	.620	.619	.629	.626	.624	.629	.627	.629
0.8	0.8	.643	.643	.643	.643	.643	.643	.643	.643	.643	.643

$\alpha = 1.7$	γ_1	0.1				0.3			0.5		0.7
	γ_2	0.1	0.3	0.5	0.7	0.3	0.5	0.7	0.5	0.7	0.7
β_1	β_2										
-0.8	-0.8	.428	.428	.428	.428	.428	.428	.428	.428	.428	.428
	-0.6	.436	.432	.430	.430	.436	.434	.432	.436	.434	.436
	-0.4	.445	.436	.433	.432	.445	.440	.437	.445	.442	.445
	-0.2	.454	.440	.436	.434	.454	.447	.443	.454	.450	.454
	0.2	.474	.449	.442	.439	.474	.461	.454	.474	.465	.474
	0.4	.483	.453	.445	.441	.483	.468	.459	.483	.473	.483
	0.6	.492	.458	.448	.443	.492	.474	.464	.492	.480	.492
	0.8	.500	.461	.450	.445	.500	.480	.469	.500	.486	.500
-0.6	-0.6	.444	.444	.444	.444	.444	.444	.444	.444	.444	.444
	-0.4	.453	.449	.447	.447	.453	.451	.450	.453	.452	.453
	-0.2	.463	.453	.450	.449	.463	.458	.455	.463	.459	.463
	0.2	.482	.462	.456	.453	.482	.472	.466	.482	.475	.482
	0.4	.491	.466	.459	.456	.491	.478	.471	.491	.483	.491
	0.6	.500	.470	.462	.458	.500	.485	.476	.500	.490	.500
	0.8	.508	.474	.464	.460	.508	.491	.481	.508	.496	.508
-0.4	-0.4	.462	.462	.462	.462	.462	.462	.462	.462	.462	.462
	-0.2	.472	.467	.465	.465	.472	.469	.468	.472	.470	.472
	0.2	.491	.476	.471	.469	.491	.483	.478	.491	.485	.491
	0.4	.500	.480	.474	.471	.500	.490	.484	.500	.493	.500
	0.6	.509	.484	.477	.474	.509	.496	.489	.509	.500	.509
	0.8	.517	.488	.480	.476	.517	.502	.493	.517	.507	.517
-0.2	-0.2	.481	.481	.481	.481	.481	.481	.481	.481	.481	.481
	0.2	.500	.490	.487	.486	.500	.495	.492	.500	.496	.500
	0.4	.509	.494	.490	.488	.509	.501	.497	.509	.504	.509
	0.6	.518	.498	.493	.490	.518	.508	.502	.518	.511	.518
	0.8	.526	.502	.495	.492	.526	.514	.507	.526	.518	.526
0.2	0.2	.519	.519	.519	.519	.519	.519	.519	.519	.519	.519
	0.4	.528	.523	.522	..521	.528	.526	.524	.528	.527	.528
	0.6	.537	.528	.525	.523	.537	.532	.529	.537	.534	.537
	0.8	.546	.531	.527	.525	.546	.538	.534	.546	.541	.546
0.4	0.4	.538	.538	.538	.538	.538	.538	.538	.538	.538	.538
	0.6	.547	.542	.540	.540	.547	.544	.543	.547	.545	.547
	0.8	.555	.546	.543	.542	.555	.550	.547	.555	.552	.555
0.6	0.6	.556	.556	.556	.556	.556	.556	.556	.556	.556	.556
	0.8	.546	.559	.558	.558	.564	.562	.560	.564	.562	.564
0.8	0.8	.572	.572	.572	.572	.572	.572	.572	.572	.572	.572

Appendix C

Karush-Kuhn-Tucker Conditions

The Karush-Kuhn-Tucker conditions are necessary conditions of a feasible solution yielding optimality in a quite arbitrary setting. By restricting the objective as well as the constraint functions to be convex, respectively linear functions, the necessary conditions on some feasible solution can even be shown to be sufficient. To make these observations more rigorous, let us consider the following minimization problem

$$\begin{aligned} &\min_x f(x) \\ &g_i(x) \leq 0 \\ &h_j(x) = 0 \end{aligned} \tag{C.1}$$

with objective $f(x)$, nonequality constraints $g_i(x)$, $i = 1, ..., m$ and equality constraints $h_j(x)$, $j = 1, ..., l$.

Theorem C.0.1 *(i) Suppose that all the functions*

$$f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$$

occurring in the problem setting (C.1) are continuously differentiable at a local minimum x^*. Then there exist constants $\lambda \geq 0$, $\mu_i \geq 0$, $(i = 1, ..., m)$ and ν_j, $(j = 1, ..., l)$, s.t.

$$\lambda + \sum_{i=1}^{m} \mu_i + \sum_{j=1}^{l} |\nu_j| > 0$$

$$\lambda \nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \nu_j \nabla h_j(x^*) = 0 \tag{C.2}$$

$$\mu_i g_i(x^*) = 0, \ i = 1, ..., m \tag{C.3}$$

(ii) Let the objective function f and the constraint functions g_i be convex functions and h_j affine functions with an arbitrary feasible point x^*. If there exist constants $\mu_i \geq 0$, $(i = 1, ..., m)$ and ν_j, $(j = 1, ..., l)$, satisfying conditions (C.2) and (C.3) for some $\lambda \neq 0$, then x^* is a global minimum.

Proof: The proof of the theorem can be found in [32] □

Bibliography

[1] H. Abdul-Hamid and J. Nolan. Multivariate stable densities as functions of one dimensional projections. *Journal of Multivariate Analysis*, 67:80–89, 1998.

[2] H. Abdul-Hamid and J. P. Nolan. Multivariate stable densities as functions of one dimensional projections. *Journal of Multivariate Analysis*, 67:80–89, 1998.

[3] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming*, (89):273–291, 2001.

[4] P. Artzner, F. Delbean, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[5] S. G. Badrinath and S. Chatterjee. On measuring skewness and elongation in common stock return distributions: The case of the market index. *The Journal of Business*, 61(4):451–472, 1988.

[6] S. G. Badrinath and S. Chatterjee. A data-analytic look at skewness and elongation in common-stock-return distributions. *The Journal of Business and Economic Statistics*, 9(9):223–233, 1991.

[7] J. A. Bucklew. *Introduction to rare event simulation*. Springer-Verlag, New York, 2004.

[8] J. Chambers, C. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, June 1976.

[9] A. Chekhlov, S. Uryasev, and M. Zabarankin. Portfolio optimization with drawdown constraints. Technical report, Risk Management and Financial Engineering Lab, Center for Applied Optimization.

[10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press and McGraw-Hill, second edition, 2001.

[11] D. Danielsson, B. N. Jorgensen, C. De Vries, and X. Yang. Optimal portfolio allocation under a probabilistic risk constraint and the incentives for financial innovation. Technical report, London School Economics, Harvard Business School, Erasmus University, Chinese Acadamy of Sciences, 2001.

[12] M. Degen, P. Embrechts, and D. D. Lambrigger. The quantitative modeling of operational risk: Between g-and-h and evt. Technical report, Department of Mathematics, ETH Zürich, 2006.

[13] R. S. Dembo and A. J. King. Tracking models and the optimal regret distribution in asset allocation. *Applied Stochastic Models and Data Analysis*, 8:151–157, 1992.

[14] N. Doowoo. *Value-at-Risk: A Quantile Based Distribution Approach for Incorporating Skewness and Fat-Tailedness*. PhD thesis, University of Alabama, 2001.

[15] K. Dutta and J. Perry. A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. Technical Report 6-13, Federal Reserve Bank of Boston, 2006.

[16] K. K. Dutta and D. F. Babbel. On measuring skewness and kurtosis in short rate distributions: The case of the us dollar london inter bank offer rates. Technical report, Wharton Financial Institutions Center, 2002.

[17] K. K. Dutta and D. F. Babbel. Extracting probabilistic information from the prices of interest rate options: Tests of distributional assumptions. *The Journal of Business*, 78(3):841–870, 2005.

[18] K. T. Fang, S. Kotz, and K. Ng. *Symmetric Multivariate and Related Distributions*, volume 36 of *Monographs of Statistics and Applied Probability*. Chapman and Hall, London.

[19] E. Feldheim. *Etude de la stabilite des lois de probabilite*. PhD dissertation, Faculte des Sciences de Paris, Paris, France, 1937.

[20] C. Field and M. G. Genton. The multivariate g-and-h distribution. *Technometrics*, 48(1):104–111, Feb. 2006.

[21] A. A. Gaivoronski and G. Pflug. Value-at-risk in portfolio optimization: Properties and computational approach.

[22] R. Garcia, E. Renault, and G. Tsafack. Proper conditioning for coherent var in portfolio management. Technical report, University of of Montreal, University of North Carolina at Chapel Hill, 2000.

[23] Gouriéroux. Sensitivity analysis of values at risk. Technical report, 1999.

[24] W. Hallerbach. Decomposing portfolio value-at-risk: A general analysis. Technical report, Tinbergen Institute Rotterdam, 1999.

[25] G. Hill and A. Davis. Generalized asymptotic expansions of cornish-fisher type. *The Annals of Mathematical Statistics*, 1968.

[26] D. C. Hoaglin. g-and-h-distributions. In S. Kotz, L. J. Norman, and C. B. Read, editors, *Encyclopedia of Statistical Sciences*, number 3, pages 298–301. Wiley and Sons, New York, 1983.

[27] D. C. Hoaglin. Summarizing shape numerically: The g-and-h distributions. In D. C. Hoaglin, F. Mosteller, and J. W. Tuckeys, editors, *Exploring Data Tables, Trends and Shapes*, pages 461–513. John Wiley and Sons, New York, 1985.

[28] P. H. Jorion. *Value-at-Risk: A new Benchmark for Measuring Derivatives Risk.* Irwin Professional Publishers, New York, 1996.

[29] A. Künzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. Technical report, Institute for Operations Research, University of Zürich, 2005.

[30] H. Konno and H. Yamazaki. Mean absolute deviation portfolio optimization model and its application to tokyo stock market. *Management Science*, 37:519–531, 1991.

[31] P. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. Technical report, University of Florida, 2001.

[32] H. W. Kuhn and A. W. Tucker. *Nonlinear Programming.* Proceedings of 2nd Berkeley Symposium. University of California Press, Berkely, 1951.

[33] A. H. Land and A. G. Doig. An automatic method for solving discrete programming problems. *Econometrica*, 28:497–520, 1960.

[34] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[35] C. Menn and S. T. Rachev. Smoothly truncated stable distributions, garch-models, and option pricing. Technical report, Cornell University, 2005.

[36] C. T. Mills. Modelling skewness and kurtosis in the london stock exchange ftse-index return distributions. *The Statistician*, 44(3):323–332, 1995.

[37] A. Mordecai. *Nonlinear Programming: Analysis and Methods*. Dover Publishing, 2003.

[38] J. Nolan. Numerical calculation of stable densities and distribution functions. *Commun. Statist.-Stochastic Models*, 13(4):759–774, 1997.

[39] J. Nolan. Multivariate stable distributions: Approximation, estimation, simulation and identification. pages 509–525, 1998.

[40] J. Nolan. Multivariate stable densities and distribution functions: General and elliptical case. *Deutsche Bundesbank's 2005 Annual Fall Converence*, 2005.

[41] J. Nolan. *Stable Distributions*. Birkhauser, 2007.

[42] J. P. Nolan, A. K. Panorska, and J. H. McCulloch. Estimation of stable spectral measures. *Mathematical and Computer Modelling*, 34:1113–1122, 2001.

[43] V. I. Norkin, G. C. Pflug, and A. Ruszczynski. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.

[44] S. Ortobelli, S. T. Rachev, and E. Schwartz. The problem of optimal asset allocation with stable distributed returns. Technical report, Anderson Graduate School of Management, 2000.

[45] A. M. P. Embrechts and D. Straumann. Correlation and dependency in risk management: Properties and pitfalls. *ETHZ Zurich*, 1998.

[46] J.-S. Pang and S. Leyffer. On the global minimization of the value-at-risk. *eprint arXiv:math/0401063*, 2003.

[47] G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. Technical report, Department of Statistics and Decision Support System, University of Vienna.

[48] M. Radulesco. Optimal portfolio selection: Mean-variance versus mean-var. Technical report, Institute of Mathematical Statistics and Applied Mathematics, Bukarest.

[49] H. Rau-Bredow. Value at risk, expected shortfall, and marginal risk contribution. Technical report, University of Würzburg, 2002.

[50] H. Rau-Bredow. Derivatives of value at risk and expected shortfall. Technical report, University of Würzburg, 2003.

[51] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[52] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

[53] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[54] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, NY, 1994.

[55] O. Scaillet. Nonparametric estimation and sensitivity analysis of expected shortfall. Technical report, Catholic University of Louvain, 2000.

[56] N. Soma and P. Toth. An exact algorithm for the subset sum problem. *European Journal of Operational Research*, (136):57–66.

[57] R. Srinivasan. *Importance sampling - Applications in communications and detection*. Springer-Verlag, Berlin, 2002.

[58] S. Stoyanov, G. Samorodnitsky, S. T. Rachev, and S. Ortobelli. Computing the portfolio conditional value-at-risk in the α -stable case. *working paper*, 2004.

[59] D. Tasche. Risk contributions and performance measurement. *working paper*, 1999.

[60] J. W. Tuckey. The practical relationship between the common transformations of percentages of counts and of amounts. Technical Report 36, Princeton University Statistical Techniques Research Group, Princeton, NJ, 1960.

[61] P. Winker. *Optimization Heuristics in Econometrics, Apllications of Threshold Accepting*. Wiley Series in Probability and Statistics. Wiley and Sons, 2001.

[62] X. Yang, S. Tao, R. Liu, and M. Cai. Complexity of scenario-based portfolio optimization problem with var objective. *International Journal of Computer Science*, 13(5):671–679, 2002.

[63] M. R. Young. A minimax portfolio selection rule with linear programming solution. *Management Science*, 44(5):673–683, 1998.