An Integrated System for Market Risk, Credit Risk and Portfolio Optimization Based on Heavy-Tailed Models and Downside Risk Measures

Dissertation an der Fakultät für Mathematik, Informatik und Statistik der Ludwig–Maximilians–Universität München

Vorgelegt am

Borjana Racheva-Iotova

Erstgutachter	Prof. Stefan Mittnik, PhD
Zweitgutachter	Prof. Dr. Marc Paolella
Tag des Rigorosums	22.09.2010

Zusammenfassung

Gängige Theorien der mathematischen Finanzwirtschaft wie zum Beispiel der Mean-Variance-Ansatz zur Portfolio-Selektion oder Modelle zur Bewertung von Wertpapieren basieren alle auf der Annahme, dass Renditen im Zeitablauf unabhngig und identisch verteilt sind und einer Normalverteilung folgen. Empirische Untersuchungen liefern jedoch signifikante Hinweise dahingehend, dass diese Annahme für wichtige Anlageklassen unzutreffend ist. Stattdessen sind Wertpapierrenditen zeitabhängige Volatilitäten, Heavy Tails (schwere Verteilungsränder), Tail Dependence (Extremwertabhängigkeit) sowie Schiefe gekennzeichnet. Diese Eigenschaften haben Auswirkungen sowohl auf die theoretische als auch praktische Modellierung in der Finanzwirtschaft. Nach der Präsentation des theoretischen Hintergrundes spricht die Arbeit die Modellierungsprobleme an, die sich aus diesen häufig beobachteten Phänomenen ergeben. Speziell werden Fragen bezüglich der Modellierung von Markt- und Kreditrisiken volatiler Märkte behandelt als auch Probleme bei der Portfolio-Optimierung unter Verwendung alternativer Risikomae und Zielfunktionen. Fragen der praktischen Implementierung wird dabei besondere Aufmerksamkeit gewidmet.

Abstract

The cornerstone theories in finance, such as mean-variance model for portfolio selection and asset pricing models, that have been developed rest upon the assumption that asset returns follow an iid Gaussian distribution. There is, however, strong empirical evidence that this the assumption does not hold for most relevant asset classes. Financial return series typically exhibit volatility clustering, heavy-tailedness, tail dependence, and skewness. These properties have implications for both theoretical and practical modeling in finance. After providing some theoretical background, this thesis addresses modeling issues arising from these commonly observed phenomena. Specifically, questions pertaining assessing and modeling market- and credit-risk in volatile markets and portfolio-optimization problems under use of alternative risk measures and objective functions are investigated. Practical implementation is a concern throughout the analysis.

Contents

Preface

The recent crash in September-October 2008 demonstrated there is a clear need for a unified system based on realistic assumptions for the behavior of financial variables. Most of the existing systems are based on the mutivariate Gaussian distribution which empirical work has shown in numerous studies it fails to describe the returns of financial variables. Not only does it fail on a stand-alone level but also on a multivariate level, where a failure to describe dependence between variables can be greatly misleading about diversification opportunities.

Evidently, there is a need for the decision making process in financial institutions to be supported by a more sophisticated tool. It should be a unified framework for market and credit risk estimation and portfolio optimization based on heavy-tailed, skewed distributions with realistic downside risk measures and flexible copula functions to capture properly dependence. In this thesis, we present such a conceptual approach working with the class of stable Paretian and Skewed Student's t distributions and the Expected tail loss risk measure. We present numerical examples supporting our approach. The system described here has been implemented successfully in the Cognity risk management platform.

I would like to thank Prof Stefan Mittnik for his patience and encouragement over the years and the members of the examination committee. I would also like to thank my husband Todor Iotov for his constant support.

January 20, 2009 Borjana Racheva-Iotova

Chapter 1

Probabilistic Models for Assets Returns

1.1 Introduction

The cornerstone theories in finance such as mean-variance model for portfolio selection and asset pricing models that have been developed rest upon the assumption that asset returns follow a normal distribution. Yet, there is little, if any, credible empirical evidence that supports this assumption for financial assets traded in most markets throughout the world. Moreover, the evidence is clear that financial return series are heavy-tailed and, possibly, skewed. Fortunately, several papers have analyzed the consequences of relaxing the normality assumption and developed generalizations of prevalent concepts in financial theory that can accommodate heavy-tailed returns (see Rachev and Mittnik (2000) and Rachev (2003) and references therein).

Mandelbrot (1963) strongly rejected normality as a distributional model for asset returns, conjecturing that financial return processes behave like non-Gaussian stable processes. To distinguish between Gaussian and non-Gaussian stable distributions. The latter are commonly referred to as "stable Paretian" distributions or "Levy stable" distributions [1].

While there have been several studies in the 1960s that have extended Mandelbrot's investigation of financial return processes, probably, the most notable is Fama (1963) and Fama (1965). His work and others led to a consolidation of the stable Paretian hypothesis. In the 1970s, however, closer empirical scrutiny of the "stability" of fitted stable Paretian distributions also produced evidence that was not consistent with the stable Paretian hypothesis. Specifically, it was often reported that fitted characteristic expo-

[1] Stable Paretian is used to emphasize that the tails of the non-Gaussian stable density have Pareto power-type decay "Levy stable" is used in recognition of the seminal work of Paul Levy's introduction and characterization of the class of non-Gaussian stable laws.

nents (or tail-indices) did not remain constant under temporal aggregation[2]. Partly in response to these empirical "inconsistencies", various alternatives to the stable law were proposed in the literature, including fat-tailed distributions being only in the domain of attraction of a stable Paretian law, finite mixtures of normal distributions, the Student t-distribution, and the hyperbolic distribution.

A major drawback of all these alternative models is their lack of stability. As has been stressed by Mandelbrot and argued by Rachev and Mittnik (2000), among others, the stability property is highly desirable for asset returns. This is particularly evident in the context of portfolio analysis and risk management. Only for stable distributed returns does one obtain the property that linear combinations of different return series (e.g., portfolios) follow again a stable distribution. Indeed, the Gaussian law shares this feature, but it is only one particular member of a large and flexible class of distributions, which also allows for skewness and heavy-tailedness.

Recent attacks on Mandelbrot's stable Paretian hypothesis focus on the claim that empirical asset return distributions are not as heavy-tailed as the non-Gaussian stable law suggests. Studies that come to such conclusions are typically based on tail-index estimates obtained with the Hill estimator. Because sample sizes beyond 100,000 are required to obtain reasonably accurate estimates, the Hill estimator is highly unreliable for testing the stable hypothesis. More importantly, the Mandelbrot's stable Paretian hypothesis is interpreted too narrowly, if one focuses solely on the *marginal* distribution of return processes. The hypothesis involves more than simply fitting marginal asset return distributions. Stable Paretian laws describe the fundamental "building blocks" (e.g., innovations) that drive asset return processes. In addition to describing these "building blocks", a complete model should be rich enough to encompass relevant stylized facts, such as

1. non-Gaussian, heavy-tailed and skewed distributions
2. volatility clustering (ARCH-effects)
3. temporal dependence of the tail behavior
4. short- and long-range dependence

An attractive feature of stable models — not shared by other distributional models — is that they allow us to generalize Gaussian-based financial theories and, thus, to build a coherent and more general framework for financial modeling. The generalizations are only possible because of specific probabilistic properties that are unique to (Gaussian and non-Gaussian) stable laws, namely, the stability property, the Central Limit Theorem and the

[2]For a more recent study, see Akgiray and Booth (1988) and Akgiray and Lamoureux (1989).

Invariance Principle for stable processes [3]. Concerning short- and long-range dependence, see Racheva-Iotova and Samorodnitsky (2003). For an empirical study of mortgage pass-through securities, see Fabozzi et al. (2006).

1.2 One-dimensional unconditional models

As pointed out in Rachev and Mittnik (2000), stable distributions are attractive because they have an important desirable property — domains of attraction. Loosely speaking, according to this property, if a distribution is in the domain of attraction of a stable law, it has properties which are close to those of the specified stable law. The domain of attraction is completely determined by the tail behavior of the distribution. As a result, it is reasonable to adopt the stable law as the "idealized" model if the true distribution has the appropriate tail behavior.

Another attractive feature is the stability property. Stable laws have an important shape parameter which governs the properties of the distribution. It is called the index of stability and is denoted by α. Because of the significance of the index of stability, stable distributions are also called α-stable. According to the stability property, appropriately centralized and normalized sums of independent identically distributed (i.i.d.) α-stable random variables is again α-stable.

A well-known property of stable non-Gaussian distributions is that, due to the power decay of the tails, they do not possess a finite second moment. Certainly the application of infinite-variance distributions as theoretical models of bounded variables, such as financial assets returns, seems inappropriate. Moreover any empirical distribution has a finite variance, hence it may seem that infinite variance distributions are inapplicable in any context. Nevertheless there is ample empirical evidence that the probability of large deviations of the changes in stock market prices is so great that any statistical theory based on finite-variance distributions is impossible to predict accurately. As it is remarked in Mandelbrot (1963) and Fama (1965), the sum of a large number of these variables is often dominated by one of the summands which is a theoretical property of infinite variance distributions. Hence an infinite-variance distribution may be an appropriate probabilistic model.

The problem of parameter estimation of stable distributions was first tackled by Mandelbrot then by Fama and Roll, Fama and Roll (1968) and Fama and Roll (1971). This is a non-trivial task because, with a few exceptions, there are no closed-form expressions for the probability density functions (p.d.f.) and cumulative distribution functions (c.d.f.)[4]. For example

[3] Detailed accounts of properties of stable distributed random variables can be found in Samorodnitsky and Taqqu (1994) and Janicki and Weron (1994).

[4] A description of the various methods to approximate the stable p.d.f. and c.d.f., see

the classical maximum likelihood method, in this case, depends on numerical approximations of the density and could be extremely time-consuming. Moreover standard estimation techniques based on asymptotic results which rely on a finite second moment are irrelevant.

Mathematical models with application of stable laws in finance, economics and other areas can be found in Adler et al. (1998), Embrechts et al. (1997) and Rachev and Mittnik (2000). We continue with a rigorous definition of the family of α-stable distributions.

1.2.1 Definition and basic properties of stable distributions

Definition and parametrizations

There are several equivalent ways to define the class of α-stable distributions. The first definition identifies the stability property.

Definition 1. *A random variable X is said to have stable distribution if for any $n \geq 2$, there is a positive number C_n and a real number D_n such that*

$$X_1 + X_2 + \ldots + X_n \stackrel{d}{=} C_n X + D_n$$

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

The second definition states that stable distributions are the only distributions that can be obtained as limits of properly normalized sums of i.i.d. random variables.

Definition 2. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e. if there is a sequence of i.i.d. random variables $Y_1, Y_2, \ldots$ and sequences of positive numbers $\{d_n\}$ and real numbers $\{a_n\}$ such that*

$$\frac{Y_1 + Y_2 + \ldots + Y_n}{d_n} + a_n \stackrel{d}{\to} X$$

where $\stackrel{d}{\to}$ denotes convergence in distribution.

The third definition specifies the characteristic function of stable laws.

Definition 3. *A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma > 0$, $-1 \leq \beta \leq 1$, $\mu \in \mathbb{R}$ such that its characteristic function (ch.f.) has the following form*

$$\varphi(t) = Ee^{itX} = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha (1 - i\beta \frac{t}{|t|} \tan(\frac{\pi\alpha}{2})) + i\mu t\}, & \alpha \neq 1 \\ \exp\{-\sigma|t|(1 + i\beta \frac{2}{\pi} \frac{t}{|t|} \ln(|t|)) + i\mu t\}, & \alpha = 1 \end{cases} \tag{1.1}$$

Stoyanov and Racheva-Iotova (2004*b*)

where $\frac{t}{|t|} = 0$ *if* $t = 0$.

Proofs of the equivalence between the three definitions can be found in Zolotarev (1986).

The parameter α is the index of stability, β is a skewness parameter, σ is a scale parameter and μ is a location parameter. Since stable distributions are uniquely determined by the four parameters, the common notation is $S_\alpha(\sigma, \beta, \mu)$. If X is said to have stable distribution, we write $X \sim S_\alpha(\sigma, \beta, \mu)$. One can easily notice that if $\beta = 0$ and $\mu = 0$, the ch.f. becomes real-valued, hence the random variable is symmetric. If X belongs to the class of symmetric α-stable distributions, we write $X \sim S\alpha S$. If $X \sim S\alpha S$ then the ch.f. of X has the simple form

$$\varphi(t) = \exp\{-\sigma^\alpha |t|^\alpha\}$$

If $\alpha = 2$, we arrive at the ch.f. of the Gaussian distribution, that is if $X \sim S_2(\sigma, \beta, \mu)$, then

$$\varphi(t) = Ee^{itX} = \exp\{-\sigma^2 |t|^2 + i\mu t\}$$

Hence X has the Gaussian distribution with mean equal to μ and variance equal to $2\sigma^2$: $X \sim N(\mu, 2\sigma^2)$. Note that in this case β is irrelevant. Nevertheless the Gaussian distribution is usually associated with $\beta = 0$.

The parametrization in Definition 3 is one possible way to define the characteristic function. It has the advantage that the parameters are easy to interpret in terms of shape and location. However there is a serious disadvantage when it comes to numerical or statistical work — it is discontinuous at $\alpha = 1$ and $\beta \neq 0$, i.e. if $\alpha \to 1$, $\beta \to \beta^* \neq 0$, $\sigma \to \sigma^*$ and $\mu \to \mu^*$ then the limit function is not the ch.f. of the stable random variable $S_1(\sigma^*, \beta^*, \mu^*)$. This drawback is not an inherent property of the class of α-stable laws and appears because of the special form of the classical ch.f. given in Definition 3. As noted in Samorodnitsky and Taqqu (1994), it is possible to change the parametrization in order to have convergence in distribution when $\alpha \to \alpha^*$, $\beta \to \beta^*$, $\sigma \to \sigma^*$ and $\mu \to \mu^*$. An alternative parametric representation of the ch.f. equipped with this property is the following

$$\varphi(t) = \begin{cases} \exp\{-|\sigma t|^\alpha + i\sigma t\beta(|\sigma t|^{\alpha-1} - 1)\tan(\frac{\pi\alpha}{2}) + i\mu_1 t\}, & \alpha \neq 1 \\ \exp\{-|\sigma t| + i\sigma t\beta\frac{2}{\pi}\ln|\sigma t| + i\mu_1 t\}, & \alpha = 1 \end{cases} \tag{1.2}$$

where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma > 0$ and $\mu_1 \in R$.

Let us denote the parametrization defined in equation (1.1) as P_0 and the continuous one defined in (1.2) as P_1. The relation between P_0 and P_1 is given in terms of the parameters μ and μ_1

$$\mu_1 = \begin{cases} \mu + \beta\sigma \tan\frac{\pi\alpha}{2}, & \alpha \neq 1 \\ \mu, & \alpha = 1 \end{cases} \tag{1.3}$$

Obviously P_0 is different from P_1 only when $\beta \neq 0$. An attractive feature of the continuous parametrization is that it is not necessary to consider the case $\alpha = 1$ separately, i.e. it can be defined by means of the limit $\alpha \to 1$. Convergence in distribution follows because of the one-to-one relationship between the cumulative distribution functions (c.d.f.s) and the ch.f.s. Moreover it is preferable to have a continuous parametrization in statistical and numerical work.

The two parametrizations defined in this section do not comprise all possible parametric representations of the ch.f. of stable laws. In Zolotarev (1986) there are more examples which appear to be appropriate in different situations.

Basic properties

The basic properties we shall consider are easier to establish when working with P_0. Most of them follow directly from the particular form of equation (1.1). The proofs can be found in Samorodnitsky and Taqqu (1994).

Property 1. *Let X_1 and X_2 be independent random variables such that $X_1 \in S_\alpha(\sigma_1, \beta_1, \mu_1)$ and $X_2 \in S_\alpha(\sigma_2, \beta_2, \mu_2)$. Then $X_1 + X_1 \in S_\alpha(\sigma, \beta, \mu)$, with*

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{\frac{1}{\alpha}}, \quad \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2$$

Property 2. *Let $X \in S_\alpha(\sigma, \beta, \mu)$ and $a \in \mathbb{R}$. Then $X + a \in S_\alpha(\sigma, \beta, \mu + a)$*

Property 3. *Let $X \in S_\alpha(\sigma, \beta, \mu)$ and $a \in \mathbb{R}$, $a \neq 0$. Then*

$$aX \in S_\alpha(|a|\sigma, sign(a)\beta, a\mu), \quad \alpha \neq 1$$

$$aX \in S_1(|a|\sigma, sign(a)\beta, a\mu - \frac{2}{\pi}(\ln(|a|)\sigma\beta)), \quad \alpha = 1$$

The first three properties identify σ and μ as a scale and a shift parameter respectively.

Property 4. *For any $0 < \alpha < 2$, if $X \in S_\alpha(\sigma, \beta, 0)$, then $-X \in S_\alpha(\sigma, -\beta, 0)$*

We shall use the standard notation for the cumulative distribution function and probability density function: $P(X < x) = F(x; \alpha, \beta)$ specifies the c.d.f. and $f(x; \alpha, \beta) = F'(x; \alpha, \beta)$ denotes the p.d.f. of a random variable $X \sim S_\alpha(1, \beta, 0)$. The fact that we consider only standardized random variables is not limiting. By Properties 2 and 3 it follows that if $X \in S_\alpha(\sigma, \beta, \mu)$,

then $(X-\mu)/\sigma \in S_\alpha(1,\beta,0)$. Furthermore because of the symmetry introduced by Property 4, it is sufficient to examine only the c.d.f. and the p.d.f. of $X \in S_\alpha(1,\beta,0)$, with $\beta \geq 0$ and then take advantage of the expressions:

- if $\beta < 0$:

$$f(x;\alpha,\beta) = f(-x;\alpha,-\beta), \quad F(x;\alpha,\beta) = 1 - F(-x;\alpha,-\beta) \tag{1.4}$$

- if $\sigma \neq 1$ and $\mu \neq 0$:

$$f(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma};\alpha,\beta\right), \quad F(x) = F\left(\frac{x-\mu}{\sigma};\alpha,\beta\right) \tag{1.5}$$

  where $f(x)$ and $F(x)$ are the p.d.f. and the c.d.f. of $X \in S_\alpha(\sigma,\beta,\mu)$ respectively.

The parameter β is a skewness parameter because of the next property.

Property 5. *$X \in S_\alpha(\sigma,\beta,\mu)$ is symmetric if and only if $\beta = 0$ and $\mu = 0$. It is symmetric about μ if and only if $\beta = 0$.*

The distribution is said to be skewed to the right if $\beta > 0$ and to the left if $\beta < 0$. It is said to be totally skewed to the right if $\beta = 1$ and totally skewed to the left if $\beta = -1$.

Stable distributions can be used as a theoretical model when empirical data is heavy-tailed. As we have mentioned in the introduction, empirical studies confirm that financial time series possess this property. The application of stable laws in this aspect is motivated by the fact that the tail of the stable law approaches zero as a power function. This is what is called "Pareto-like" behavior of the tail because of the same power decay of the tail of the Pareto distribution. The next property provides a rigorous description of the tail behavior.

Property 6. *Let $X \in S_\alpha(\sigma,\beta,\mu)$ $0 < \alpha < 2$. Then*

$$\lim_{\lambda\to\infty} \lambda^\alpha P(X > \lambda) = C_\alpha \frac{1+\beta}{2}\sigma^\alpha$$

$$\lim_{\lambda\to\infty} \lambda^\alpha P(X < -\lambda) = C_\alpha \frac{1-\beta}{2}\sigma^\alpha$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin(x)dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, & \alpha \neq 1 \\ 2/\pi, & \alpha = 1 \end{cases}$$

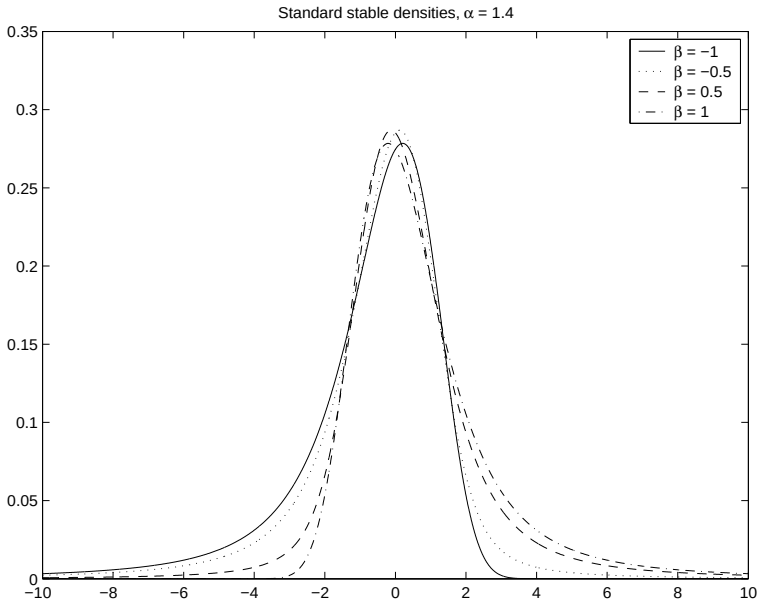

Figure 1.1: Stable densities.

The tail behavior of the Gaussian distribution as $\lambda \to \infty$ is specified by

$$P(X < -\lambda) = P(X > \lambda) \sim \frac{1}{2\sqrt{\pi}\sigma\lambda} e^{-\frac{\lambda^2}{4\sigma^2}}$$

The difference in the asymptotic behavior of the tails of α-stable distributions, with $\alpha < 2$ and the Gaussian distribution motivates the distinction between Gaussian and non-Gaussian stable distributions. The latter are also called Pareto stable or Lévy stable as mentioned in the introduction.

The power decay of the tail of Pareto stable distributions implies that they do not possess a finite second moment. More exactly

Property 7. *Let $X \in S_\alpha(\sigma, \beta, \mu)$ and $0 < \alpha < 2$. Then*

$$E|X|^p < \infty, \quad 0 < p < \alpha$$

$$E|X|^p = \infty, \quad \alpha \leq p$$

As a consequence, if $\alpha \leq 1$, then the corresponding α-stable distribution does not have a finite first absolute moment. Therefore statistical techniques valid for the Gaussian distribution are not applicable for the stable Paretian distributions[5].

The absolute moments considered in Property 7 can be computed.

[5]For a review of the different methods of estimation of the stable distribution parameters, see Stoyanov and Racheva-Iotova (2004*c*)

Property 8. *Let $X \in S_\alpha(\sigma, \beta, \mu)$, $0 < \alpha < 2$ and $\beta = 0$ if $\alpha = 1$. Then for any $0 < p < \alpha$ there exists a constant $c_{\alpha,\beta}(p)$, such that:*

$$(E|X - \mu|^p)^{1/p} = c_{\alpha,\beta}(p)\sigma, \tag{1.6}$$

where $c_{\alpha,\beta}(p) = (E|X_0|^p)^{1/p}$, $X_0 \in S_\alpha(1, \beta, 0)$

The constant $c_{\alpha,\beta}(p)$ can be explicitly calculated, see Samorodnitsky and Taqqu (1994)

$$\begin{aligned}(c_{\alpha,\beta}(p))^p = &\frac{2^{p-1}\Gamma\left(1-\frac{p}{\alpha}\right)}{p\int_0^\infty u^{-p-1}\sin^2 udu}\cdot \\ &\cdot\left(1+\beta^2\tan^2\frac{\pi\alpha}{2}\right)^{p/2\alpha}\cos\left(\frac{p}{\alpha}\arctan\left(\beta\tan\frac{\alpha\pi}{2}\right)\right) \\ &= (c_{\alpha,0}(p))^p\cdot \\ &\cdot\left(1+\beta^2\tan^2\frac{\pi\alpha}{2}\right)^{p/2\alpha}\cos\left(\frac{p}{\alpha}\arctan\left(\beta\tan\frac{\alpha\pi}{2}\right)\right)\end{aligned} \tag{1.7}$$

The shift parameter μ has the following nice property if $\alpha > 1$:

Property 9. *If $1 < \alpha \leq 2$, the location parameter μ equals the mathematical expectation of $X \in S_\alpha(\sigma, \beta, \mu)$.*

Property 10. *Let $X \in S_{\alpha'}(\sigma, 0, 0)$ and let $0 < \alpha < \alpha'$. Let Y be an α/α'-stable random variable, totally skewed to the right*

$$Y \in S_{\alpha/\alpha'}\left(\left(\cos\frac{\pi\alpha}{2\alpha'}\right)^{\alpha'/\alpha}, 1, 0\right)$$

and assume that X and Y are independent. Then

$$Z = Y^{1/\alpha'}X \in S_\alpha(\sigma, 0, 0)$$

This property[6] implies that if X is zero mean Gaussian random variable and if Y is a positive $\alpha/2$-stable random variable independent of X, then

$$Z = Y^{1/2}X$$

is symmetric α-stable. This shows that every symmetric α-stable random variable is conditionally Gaussian.

[6]For a proof, see Samorodnitsky and Taqqu (1994). This property has an important implication in the multivariate sub-Gaussian model which finds application in portfolio theory, see Section 1.5 and Ortobelli et al. (2004) and the references therein.

1.2.2 Definition of skewed Student's t distribution

The skewed Student's t distribution is defined through the following stochastic representation:[7]

$$X := \mu + \gamma W + Z\sqrt{W} \tag{1.8}$$

where $W \in IG(\nu/2, \nu/2)$ and $Z \in N(0, \sigma^2)$, Z is independent of W, $\gamma \in \mathbb{R}$ is a parameter accounting for the skewness, $\mu \in R$ is location parameter vector and ν stands for the degrees of freedom. The notation $IG(\nu/2, \nu/2)$ stands for the inverse gamma distribution with parameters $\nu/2$.

The skewed Student's t distribution allows for closed-form expression of its density,

$$f_X(x) = \frac{aK_{(\nu+1)/2}\left(\sqrt{(\nu + (x-\mu)^2/\sigma^2)\gamma^2/\sigma^2}\right)\exp((x-\mu)\gamma/\sigma^2)}{(\nu + (x-\mu)^2/\sigma^2)\gamma^2/\sigma^2)^{-\frac{\nu+n}{4}}\left(1 + \frac{(x-\mu)^2/\sigma^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

where $x \in \mathbb{R}$, K is the modified Bessel function of the third kind and

$$a = \frac{2^{\frac{2-\nu-1}{2}}}{\Gamma(\nu/2)(\pi\nu)^{1/2}\sqrt{|\Sigma|}}.$$

The skewed Student's t distribution can be used to model heavy-tails and skewness. Figure 1.2 shows the stable and skewed Student's t fits to a sample of observed returns. Both models are clearly better than the standard normal model. In different market conditions, the two skewed and heavy-tailed models may perform differently. Therefore, a practical risk management system should have them both as possible choices for modeling one-dimensional returns.

1.3 Parameter estimation of stable distributions

Generally speaking, parameter estimation techniques for the class of stable laws come into three categories - quantile methods, characteristic function based methods and maximum likelihood. The approaches from the first type use predetermined empirical quantiles to estimate stable parameters. For example the method of Fama and Roll (1968), Fama and Roll (1971) for symmetric α-stable distributions and its modified version of McCulloch (1986) for the skewed case belong to this group.

Chf based methods include the method of moments approach suggested by Press (1972) and regression-type procedures proposed by Koutrouvelis

[7] For more information, see Section 12.7 in Rachev and Mittnik (2000).

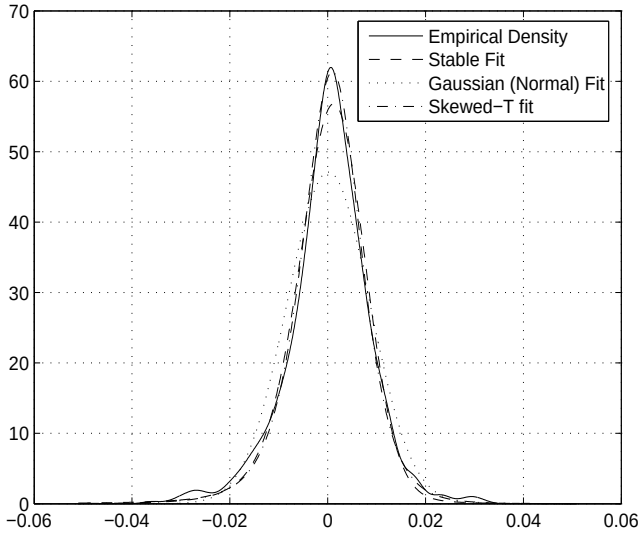

Figure 1.2: Stable, skewed Student's t, and Gaussian distributions fitted to real data.

(1981) and Kogon and Williams in Adler et al. (1998). Simulation studies available in the literature (Adler et al. (1998), Akgiray and Lamoureux (1989)), show the superiority of the regression-type estimation over the quantile methods.

The validity of maximum likelihood estimation (MLE) theory was demonstrated by DuMouchel (1973). The comparison studies between MLE and the quantile method of McCulloch in Rachev and Mittnik (2000) recommend the maximum likelihood estimator.

In this section we shall review and compare McCulloch's quantile method, the method of moments, the regression-type estimator of Kogon and Williams and MLE.[8]

1.3.1 Quantile method of McCulloch

The estimation procedure proposed by McCulloch (1986) is a generalization of the quantile method in Fama and Roll (1971) for the symmetric case. The estimates of stable parameters in parametrization P_0 are consistent and asymptotically normal if $0.6 < \alpha \leq 2$. We shall adopt the standard notation for theoretical and empirical quantiles, namely x_p is the p-th quantile if $F(x_p) = p$, where $F(x)$ is the cdf of a random variable and given a sample

[8]For additional information, see Stoyanov and Racheva-Iotova (2004*a*).

of observations $x_1, x_2, \ldots, x_n$, then $\hat{x}_p$ is the sample quantile if $F_n(\hat{x}_p) = p$, where $F_n(x)$ is the sample cdf.

According to McCulloch (1986), let us define two functions of theoretical quantiles :

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}$$

$$v_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.50}}{x_{0.95} - x_{0.05}}$$

The functions v_α and v_β have this special form because by expression (1.5) it appears that they do not depend on the scale and the location parameter, i. e.

$$\left| \begin{array}{l} v_\alpha = \phi_1(\alpha, \beta) \\ v_\beta = \phi_2(\alpha, \beta) \end{array} \right. \tag{1.9}$$

Employing equation (1.4), we have that $F(-x_p; \alpha, -\beta) = F(x_{1-p}; \alpha, \beta)$ and therefore we have the relations:

$$\begin{array}{l} \phi_1(\alpha, \beta) = \phi_1(\alpha, -\beta) \\ \phi_2(\alpha, \beta) = -\phi_2(\alpha, -\beta) \end{array} \tag{1.10}$$

The system of equations (1.9) can be inverted and the parameters α and β can be expressed as functions of the quantities v_α and v_β:

$$\left| \begin{array}{l} \alpha = \psi_1(v_\alpha, v_\beta) \\ \beta = \psi_2(v_\alpha, v_\beta) \end{array} \right. \tag{1.11}$$

Replacing v_α and v_β in equations (1.11) with their sample counterparts $\hat{v}_\alpha$ and $\hat{v}_\beta$:

$$\hat{v}_\alpha = \frac{\hat{x}_{0.95} - \hat{x}_{0.05}}{\hat{x}_{0.75} - \hat{x}_{0.25}}$$

$$\hat{v}_\beta = \frac{\hat{x}_{0.95} + \hat{x}_{0.05} - 2\hat{x}_{0.50}}{\hat{x}_{0.95} - \hat{x}_{0.05}}$$

yields estimators $\hat{\alpha}$ and $\hat{\beta}$:

$$\left| \begin{array}{l} \hat{\alpha} = \psi_1(\hat{v}_\alpha, \hat{v}_\beta) \\ \hat{\beta} = \psi_2(\hat{v}_\alpha, \hat{v}_\beta) \end{array} \right. \tag{1.12}$$

The functions $\psi_1(.)$ and $\psi_2(.)$ are tabulated in Tables 1.1 and 1.2. It should be noted that because of property (1.10), we have that:

		v_β						
		0	0.1	0.2	0.3	0.5	0.7	1
v_α	2.439	2	2	2	2	2	2	2
	2.5	1.916	1.924	1.924	1.924	1.924	1.924	1.924
	2.6	1.808	1.813	1.829	1.829	1.829	1.829	1.829
	2.7	1.729	1.73	1.737	1.745	1.745	1.745	1.745
	2.8	1.664	1.663	1.663	1.668	1.676	1.676	1.676
	3	1.563	1.56	1.553	1.548	1.547	1.547	1.547
	3.2	1.484	1.48	1.471	1.46	1.448	1.438	1.438
	3.5	1.391	1.386	1.378	1.364	1.337	1.318	1.318
	4	1.279	1.273	1.266	1.25	1.21	1.184	1.15
	5	1.128	1.121	1.114	1.101	1.067	1.027	0.973
	6	1.029	1.021	1.014	1.004	0.974	0.935	0.874
	8	0.896	0.892	0.887	0.883	0.855	0.823	0.769
	10	0.818	0.812	0.806	0.801	0.78	0.756	0.691
	15	0.698	0.695	0.692	0.689	0.676	0.656	0.595
	25	0.593	0.59	0.588	0.586	0.579	0.563	0.513

Table 1.1: $\alpha = \psi_1(v_\alpha, v_\beta) = \psi_1(v_\alpha, -v_\beta)$

		v_β						
		0	0.1	0.2	0.3	0.5	0.7	1
v_α	2.439	0	2.16	1	1	1	1	1
	2.5	0	1.592	3.39	1	1	1	1
	2.6	0	0.759	1.8	1	1	1	1
	2.7	0	0.482	1.048	1.694	1	1	1
	2.8	0	0.36	0.76	1.232	2.229	1	1
	3	0	0.253	0.518	0.823	1.575	1	1
	3.2	0	0.203	0.41	0.632	1.244	1.906	1
	3.5	0	0.165	0.332	0.499	0.943	1.56	1
	4	0	0.136	0.271	0.404	0.689	1.23	2.195
	5	0	0.109	0.216	0.323	0.539	0.827	1.917
	6	0	0.096	0.19	0.284	0.472	0.693	1.759
	8	0	0.082	0.163	0.243	0.412	0.601	1.596
	10	0	0.074	0.174	0.22	0.377	0.546	1.482
	15	0	0.064	0.128	0.191	0.33	0.478	1.362
	25	0	0.056	0.112	0.167	0.285	0.428	1.274

Table 1.2: $\beta = \psi_2(v_\alpha, v_\beta) = -\psi_2(v_\alpha, -v_\beta)$

		β				
		0	0.25	0.5	0.75	1
	0.5	2.588	3.073	4.534	6.636	9.144
	0.6	2.337	2.635	3.542	4.808	6.247
	0.7	2.189	2.392	3.004	3.844	4.775
	0.8	2.098	2.244	2.676	3.265	3.912
	0.9	2.04	2.149	2.461	2.886	3.356
	1	2	2.085	2.311	2.624	2.973
	1.1	1.98	2.04	2.205	2.435	2.696
	1.2	1.965	2.007	2.125	2.294	2.491
α	1.3	1.955	1.984	2.067	2.188	2.333
	1.4	1.946	1.967	2.022	2.106	2.211
	1.5	1.939	1.952	1.988	2.045	2.116
	1.6	1.933	1.94	1.962	1.997	2.043
	1.7	1.927	1.93	1.943	1.961	1.987
	1.8	1.921	1.922	1.927	1.936	1.947
	1.9	1.914	1.915	1.916	1.918	1.921
	2	1.908	1.908	1.908	1.908	1.908

Table 1.3: $v_\sigma = \phi_3(\alpha, \beta) = \phi_3(\alpha, -\beta)$

$$\psi_1(v_\alpha, v_\beta) = \psi_1(v_\alpha, -v_\beta)$$
$$\psi_2(v_\alpha, v_\beta) = -\psi_2(v_\alpha, -v_\beta)$$

In other words, the sign of $\hat{v}_\beta$ determines the sign of β.

Since there are no closed-form expressions for the functions $\psi_1(.)$ and $\psi_2(.)$, we compute estimates of α and β from the statistics $\hat{v}_\alpha$ and $\hat{v}_\beta$ using the values in Tables 1.1 and 1.2 and linear interpolation for intermediate values. If it happens that $\hat{v}_\alpha$ is below 2.439, $\hat{\alpha}$ should be set equal to 2 and $\hat{\beta}$ equal to zero. Table 1.2 contains values larger than 1 for more precise interpolation. If $\hat{\beta} > 1$, it should be reduced to 1.

McCulloch provides estimator for the scale parameter σ which is very similar to the estimator given by Fama and Roll. Let us first define v_σ as:

$$v_\sigma = \frac{x_{0.75} - x_{0.25}}{\sigma} = \phi_3(\alpha, \beta)$$

The function $\psi_3(\alpha, \beta)$ is given in Table 1.3. Employing the same arguments that led us to equations (1.10) yields the relation $\phi_3(\alpha, \beta) = \phi_3(\alpha, -\beta)$. The estimator $\hat{\sigma}$ is received after replacing α and β with the estimates found according to equations (1.12):

$$\hat{\sigma} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\phi_3(\hat{\alpha}, \hat{\beta})}$$

Estimation of the location parameter μ is a more involved affair because of the discontinuity of the parametric representation of the chf P_0 when

		β				
		0	0.25	0.5	0.75	1
α	0.5	0	-0.061	-0.279	-0.659	-1.198
	0.6	0	-0.078	-0.272	-0.581	-0.997
	0.7	0	-0.089	-0.262	-0.52	-0.853
	0.8	0	-0.096	-0.25	-0.469	-0.742
	0.9	0	-0.099	-0.237	-0.424	-0.652
	1	0	-0.098	-0.223	-0.383	-0.576
	1.1	0	-0.095	-0.208	-0.346	-0.508
	1.2	0	-0.09	-0.192	-0.31	-0.447
	1.3	0	-0.084	-0.173	-0.276	-0.39
	1.4	0	-0.075	-0.154	-0.241	-0.335
	1.5	0	-0.066	-0.134	-0.206	-0.283
	1.6	0	-0.056	-0.111	-0.17	-0.232
	1.7	0	-0.043	-0.088	-0.132	-0.179
	1.8	0	-0.03	-0.061	-0.092	-0.123
	1.9	0	-0.017	-0.032	-0.049	-0.064
	2	0	0	0	0	0

Table 1.4: $\phi_4(\alpha, \beta) = \phi_4(\alpha, -\beta)$

$\alpha \to 1$ and $\beta \neq 0$. First we estimate the shifted location parameter ζ defined by:

$$\zeta = x_{0.50} + \sigma \text{sign}(\beta)\phi_4(\alpha, \beta) \tag{1.13}$$

where $\phi_4(\alpha, \beta)$ is tabulated in Table 1.4 and has the property $\phi_4(\alpha, \beta) = \phi_4(\alpha, -\beta)$. The location parameter μ is related to ζ according to:

$$\mu = \begin{cases} \zeta - \beta\sigma \tan \frac{\pi\alpha}{2}, & \alpha \neq 1 \\ \zeta, & \alpha = 1 \end{cases} \tag{1.14}$$

Replacing the parameters in equations (1.13) and (1.14) with their sample counterparts yields the estimator $\hat{\mu}$:

$$\hat{\zeta} = \hat{x}_{0.50} + \hat{\sigma} sign(\hat{\beta})\phi_4(\hat{\alpha}, \hat{\beta})$$

and

$$\hat{\mu} = \begin{cases} \hat{\zeta} - \hat{\beta}\hat{\sigma} \tan \frac{\pi\hat{\alpha}}{2}, & \hat{\alpha} \neq 1 \\ \hat{\zeta}, & \hat{\alpha} = 1 \end{cases}$$

It should be observed that a significant advantage of the method considered is the lack of heavy computations. On the personal homepage of McCulloch (http://www.econ.ohio-state.edu/jhm/jhm.html) a FORTRAN implementation of the algorithm is publicly available.

1.3.2 Chf based methods

The characteristic function based methods rely on the sample chf for parameter estimation. The sample chf is defined as:

$$\hat{\varphi}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itx_j}, \quad t \in \mathbb{R} \tag{1.15}$$

where $x_1, x_2, \ldots, x_n$ is a sample of independent, identically distributed (iid) observations on a random variable X. Since $|\hat{\varphi}(t)| \leq 1$, all moments of the random variable $\hat{\varphi}(t)$ are finite and, according to equation (1.15), for any t it is the sample mean of the iid random variables e^{itx_j}. As a consequence, from the law of large numbers, it can be inferred that the sample chf is a consistent estimator of the chf $\varphi_X(t) = \mathbb{E}e^{itX}$, $t \in \mathbb{R}$ of a random variable X.

The method of moments

Press (1972) suggested a simple and straightforward approach to estimation of parameters of stable laws which was called the method of moments. His approach is based on certain transformations of the chf in parametrization P_0. From the parametric representation (1.1) it follows that

$$|\varphi(t)| = \exp(-\sigma^\alpha |t|^\alpha), \quad t \in \mathbb{R} \tag{1.16}$$

and therefore $-\ln|\varphi(t)| = \sigma^\alpha |t|^\alpha$ for any real t.

Case $\alpha \neq 1$. If we choose t_1 and t_2 such that $t_1 \neq t_2 \neq 0$, we have the following system of two equations:

$$\left| \begin{array}{l} -\ln|\varphi(t_1)| = \sigma^\alpha |t_1|^\alpha \\ -\ln|\varphi(t_2)| = \sigma^\alpha |t_2|^\alpha \end{array} \right.$$

which can be solved for α and σ. Replacing the chf for its sample equivalent $\hat{\varphi}(t)$ yields the estimators $\hat{\alpha}$ and $\hat{\sigma}$:

$$\hat{\alpha} = \frac{\ln \frac{\ln|\hat{\varphi}(t_1)|}{\ln|\hat{\varphi}(t_2)|}}{\ln\left|\frac{t_1}{t_2}\right|} \tag{1.17}$$

and

$$\ln \hat{\sigma} = \frac{\ln|t_1| \ln(-\ln|\hat{\varphi}(t_2)|) - \ln|t_2| \ln(-\ln|\hat{\varphi}(t_1)|)}{\ln\left|\frac{\hat{\varphi}(t_1)}{\hat{\varphi}(t_2)}\right|} \tag{1.18}$$

Estimation of the skewness and the location parameter requires more efforts. Let us first denote the imaginary part of the logarithm of the chf in P_0 as $u(t)$:

$$u(t) = \Im(\ln \varphi(t)) = \mu t + \sigma^\alpha |t|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2}$$

Then if we choose two non-zero values t_3 and t_4 such that $t_3 \neq t_4$ we can write a system of two equations:

$$\begin{cases} \frac{u(t_3)}{t_3} = \mu + \sigma^\alpha |t_3|^{\alpha-1} \beta \tan \frac{\pi\alpha}{2} \\ \frac{u(t_4)}{t_4} = \mu + \sigma^\alpha |t_4|^{\alpha-1} \beta \tan \frac{\pi\alpha}{2} \end{cases}$$

It is possible to solve the system for β and μ and again replacing α, σ and $u(t)$ with their sample counterparts yields the required estimators. Since

$$\hat{\varphi}_\xi(t) = \left(\frac{1}{n} \sum_{j=1}^{n} \cos t x_j \right) + i \left(\frac{1}{n} \sum_{j=1}^{n} \sin t x_j \right)$$

and taking advantage of the properties of complex numbers we achieve the estimator $\hat{u}(t)$:

$$\tan \hat{u}(t) = \frac{\sum_{j=1}^{n} \sin t x_j}{\sum_{j=1}^{n} \cos t x_j}$$

Finally for $\hat{\beta}$ and $\hat{\mu}$ we have:

$$\hat{\beta} = \frac{\frac{\hat{u}(t_4)}{t_4} - \frac{\hat{u}(t_3)}{t_3}}{\left[|t_4|^{\hat{\alpha}-1} - |t_3|^{\hat{\alpha}-1} \right] \hat{\sigma}^{\hat{\alpha}} \tan \frac{\pi\hat{\alpha}}{2}} \tag{1.19}$$

and

$$\hat{\mu} = \frac{|t_4|^{\hat{\alpha}-1} \frac{\hat{u}(t_3)}{t_3} - |t_3|^{\hat{\alpha}-1} \frac{\hat{u}(t_4)}{t_4}}{|t_4|^{\hat{\alpha}-1} - |t_3|^{\hat{\alpha}-1}} \tag{1.20}$$

Case $\alpha = 1$. If $\alpha = 1$, equation (1.16) allows us to construct the estimator $\hat{\sigma}$ directly:

$$\hat{\sigma} = -\frac{\ln |\varphi(t_1)|}{t_1}$$

where $t_1 \neq 0$. Similar arguments as in the case $\alpha \neq 1$ lead us to:

$$\hat{\beta} = \frac{\frac{\hat{u}(t_3)}{t_3} - \frac{\hat{u}(t_4)}{t_4}}{\frac{2}{\pi} \hat{\sigma} \ln |\frac{t_4}{t_3}|}$$

$$\hat{\mu} = \frac{\ln |t_4| \frac{\hat{u}(t_3)}{t_3} - \ln |t_3| \frac{\hat{u}(t_4)}{t_4}}{\ln |t_4| - \ln |t_3|}$$

where $t_3 \neq t_4$ and both are non-zero.

The estimators of stable parameters are consistent since they are based on $\hat{\varphi}(t)$, $\Re\hat{\varphi}(t)$ and $\Im\hat{\varphi}(t)$ which are consistent estimators of $\varphi(t)$, $\Re\varphi(t)$ and $\Im\varphi(t)$ by the law of large numbers. The question which still remains is the best way to choose $t_1, \ldots, t_4$, since obviously the derived estimators are not invariant of their choice. Koutrouvelis in his simulation studies in Koutrouvelis (1980) uses the values $t_1 = 0.2$, $t_2 = 0.8$, $t_3 = 0.1$ and $t_4 = 0.4$, which are selected for the normalized case ($\sigma = 1, \mu = 0$). Because of the following property of the chf of an arbitrary random variable X:

$$\varphi_{\sigma X+\mu}(t) = e^{it\mu}\varphi_X(\sigma t)$$

it is clear that for different σ and μ we shall have to choose different values for $t_1, \ldots, t_4$ to achieve equal performance, i.e. the values determined for the normalized case will not be equally "good" for a non-normalized case. For this reason, if we aim at estimation of stable parameters by the method of moments, we need first to find initial estimates of the scale and the location parameter and to normalize the sample. Without incurring significant additional computational burden, initial estimates could be computed with the help of a quantile method. For such purposes Koutrouvelis (1980) uses the method of Fama and Roll (1968), Fama and Roll (1971) despite the bias in the estimate of σ even in the symmetric case. We shall adopt his approach in our computations.

To summarize, the algorithm for estimating stable parameters by the method of moments, given a sample of iid observations $x_1, x_2, \ldots, x_n$, is as follows:

1. Compute initial estimates $\hat{\sigma}_0$ and $\hat{\mu}_0$ of σ and μ respectively, according to:

   $$\hat{\sigma}_0 = \frac{\hat{x}_{0.72} - \hat{x}_{0.28}}{1.654}$$

   and $\hat{\mu}_0$ equals the 50% truncated sample average - the mean of the middle 50% of the ordered observations.

2. Normalize the sample with the initial estimates:

   $$x'_k = (x_k - \hat{\mu}_0)/\hat{\sigma}_0, \quad k = 1, 2, \ldots n$$

3. Using the normalized sample $x'_1, x'_2, \ldots, x'_n$, calculate $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_1$ and $\hat{\mu}_1$ according to equations (1.17), (1.19), (1.18) and (1.20) respectively.

4. Compute the final estimates $\hat{\sigma}$ and $\hat{\mu}$:

   $$\hat{\sigma} = \hat{\sigma}_0\hat{\sigma}_1, \qquad \hat{\mu} = \hat{\sigma}_0\hat{\mu}_1 + \hat{\mu}_0$$

The regression-type estimator of Kogon-Williams

Regression-type estimators are also based on the sample chf. It is possible to derive simple expressions, linear with respect to stable parameters, and construct estimators using the least squares technique. Kogon and Williams suggest such a procedure in Adler et al. (1998) with the chf being parametrized according to the continuous parametrization P_1 defined in equation (1.2). Their approach is similar to the that of Koutrouvelis (1980) and Koutrouvelis (1981).

The linear equations follow directly from the convenient form of the logarithm of the chf:

$$\ln[-\Re(\ln\varphi(t))] = \alpha\ln\sigma + \alpha\ln|t| \tag{1.21}$$

$$\Im(\ln\varphi(t)) = \mu_1 t + \beta\sigma t(|\sigma t|^{\alpha-1} - 1)\tan\frac{\pi\alpha}{2} \tag{1.22}$$

Estimators of the stable parameters can be constructed using the method of least squares after replacing the chf for the sample chf. Certainly here we face the same problem as in the method of moments - the sample chf should be evaluated for certain values of the argument. Koutrouvelis gives tables in Koutrouvelis (1980) and Koutrouvelis (1981) which relate the values of the sample chf argument to the value of the index of stability α and the sample size. The major advantage of the procedure in Adler et al. (1998) is that the provided values of the sample chf argument are invariant of any other parameters. Having conducted numerous experiments, Kogon and Williams report in Adler et al. (1998) that the most suitable choice is $t_k = \{0.1 + 0.1k, \quad k = 0, 1, \ldots, 9\}$ - 10 equally spaced points in the interval $[0.1, 1]$. Undoubtedly the sample should be normalized before applying the method of least squares, otherwise the optimal selection of the sample chf arguments would depend on the scale and the modified location parameter. For preliminary estimation of σ and μ, it is suggested to use the quantile method of McCulloch.

The algorithm is as follows:

1. Given a sample of iid observations $x_1, x_2, \ldots, x_n$ first we find preliminary estimates σ_0 and μ_{01} utilizing the quantile method of McCulloch and we normalize the observations:

$$x'_j = \frac{x_j - \hat{\mu}_{01}}{\hat{\sigma}_0}, \quad j = 1, 2, \ldots n$$

2. Next we consider the regression equation constructed from equation (1.21):

$$y_k = b + \alpha w_k + \epsilon_k, \quad k = 0, 1, \ldots 9$$

where $y_k = \ln[-\Re(\ln\hat{\varphi}(t_k))]$, $w_k = \ln|t_k|$, $t_k = \{0.1 + 0.1k, \quad k = 0, 1, \ldots, 9\}$ and ϵ_k denotes the error term. We find $\hat{\alpha}$ and $\hat{b}$ according to the method of least squares using the normalized sample $x'_1, x'_2, \ldots, x'_n$. The estimator $\hat{\sigma}_1$ of the scale parameter of the normalized sample is:

$$\hat{\sigma}_1 = \exp\left(\frac{\hat{b}}{\hat{\alpha}}\right)$$

3. Estimators $\hat{\beta}$ and $\hat{\mu}_{11}$ of the skewness parameter and the modified location parameter respectively are derived from the second regression equation based on (1.22):

$$z_k = \mu_{11}t_k + \beta v_k + \eta_k, \quad k = 0, 1, \ldots 9$$

where $z_k = \Im(\ln\hat{\varphi}(t_k))$, $v_k = \hat{\sigma}_1 t_k(|\hat{\sigma}_1 t_k|^{\hat{\alpha}-1} - 1)\tan\frac{\pi\hat{\alpha}}{2}$, $t_k = \{0.1 + 0.1k, \quad k = 0, 1, \ldots, 9\}$ and η_k is the error term.

4. The final estimators $\hat{\sigma}$ and $\hat{\mu}_1$ proceed from:

$$\hat{\sigma} = \hat{\sigma}_0\hat{\sigma}_1, \quad \hat{\mu}_1 = \hat{\mu}_{01} + \hat{\sigma}_0\hat{\mu}_{11}$$

If we aim at estimating the location parameter μ, we need to take advantage of the connection between the two parametric forms P_0 and P_1:

$$\hat{\mu} = \hat{\mu}_1 - \hat{\beta}\hat{\sigma}\tan\frac{\pi\hat{\alpha}}{2}$$

In Adler et al. (1998) there is a huge Monte Carlo study in which the method of Kogon-Williams is compared to the approach in Koutrouvelis (1980), Koutrouvelis (1981). The result is that from computational viewpoint the former is more efficient. It is definitely superior to the latter when α is close to zero and $\beta \neq 0$. The approach of Koutrouvelis outperforms that of Kogon-Williams only in the estimation of β.

1.3.3 Maximum likelihood

The method of maximum likelihood is very attractive because of the good asymptotic properties of the estimates, provided that the likelihood function obeys certain general conditions. The likelihood function is defined as:

$$L(x_1, x_2, \ldots x_n|\theta) = \prod_{k=1}^{n} f(x_k|\theta)$$

where $x_1, x_2, \ldots x_n$ is a sample of iid observations of a random variable X, $f(x|\theta)$ is the pdf of X, and θ is a vector of parameters. In the case of stable distributions, $\theta = (\alpha, \beta, \sigma, \mu)$. Maximum likelihood estimates are found by

searching for that parameter values which maximize the likelihood function, or equivalently, the log-likelihood function:

$$\hat{\theta}_n = \arg\max_{\theta} \log(L(x_1, x_2, \dots x_n|\theta)) \tag{1.23}$$

DuMouchel studied the applicability of maximum likelihood theory in the case of α-stable distributions in DuMouchel (1973) by verifying whether the likelihood function complies with a set of conditions that guarantee the validity of the theory. The theorem proved in the paper and adapted to parametrization P_0 is the following

Theorem 1. *When sampling from a stable distribution, $\hat{\theta}_n$, the maximum likelihood estimate for $\theta = (\alpha, \beta, \sigma, \mu)$ based on the first n observations, restricted so that $\hat{\alpha}_n$, the estimate of α, satisfies $\hat{\alpha}_n > \epsilon$, ϵ arbitrarily small and positive, is consistent and asymptotically normal as long as θ_0, the true value of θ is in the interior of the parameter space (that is the cases $\alpha_0 \leq \epsilon$, $\alpha_0 = 2$ and $|\beta| = 1$ are excluded) and the additional case $(\alpha_0 = 1, \beta_0 \neq 0)$ is excluded.*

Clearly if we intend to derive expressions for MLE analytically, we need to have closed-form expressions for the pdfs of stable laws. Such expressions are not known to exist in the general case and the problem of MLE of stable parameters should be attacked numerically, i.e. we have to numerically search for the solution of problem (1.23) in which the pdf is approximated. Approaches to numerical approximation of the pdfs and the cdfs of stable laws are considered in Stoyanov and Racheva-Iotova (2004*c*). For practical calculations, we use the FFT-method combined with a Bergström series expansion for tail approximation.

1.3.4 Choice of an estimation method

A risk system based on stable distributions have to be flexible enough to support a number of fitting methods. In terms of accuracy, the MLE is superior but is slow while the characteristic function based method is a lot faster but may have a difficulty in estimating the skewness. The quantile method is fastest but is least accurate. Thus, the choice which method to use does not have a definite answer since sometimes practitioners prefer speed to accuracy, especially when portfolios are large.

1.4 Conditional one-dimensional models

The analysis of empirical behavior of stock returns data has shown that the conditional first and second moments are time-dependent, see for example Fama and French (1988) and French et al. (1989). The temporal dependence is different with respect to whether the returns are measured over shorter

or longer horizons. In a model that explains these phenomena, information of past market movements is incorporated and thus a *conditional* distribution model is more suitable than a *unconditional* one. We say that the model is *conditional* if the distribution of asset returns is conditional on the information of past market movements or a more general information set.

One natural class of models is the auto-regressive moving average (ARMA) processes. They have the property that the conditional distribution is *homoskedastic*, i.e. the conditional volatility is constant. If we assume the ARMA(p,q) process of auto-regressive order p and moving average order q to model the return time series r_t of an asset, then

$$r_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} \tag{1.24}$$

where $\{\epsilon_t\}$ is a white-noise process[9], also called the innovations process, $\{a_i\}_{i=1}^{p}$ and $\{b_j\}_{j=1}^{q}$ are parameters that are estimated from historical data. The order (p, q) of the process can be determined by following the standard Box-Jenkins identification methodology[10].

The assumption of conditional homoskedasticity often does not hold in financial data. An empirically observed phenomenon is the volatility clustering — a large return in absolute value is usually followed by relatively large returns in absolute value and their magnitude gradually subsides. Models that can capture such behavior are the conditional *heteroskedastic* models. The first proposed model belonging to this class is the autoregressive conditional heteroskedastic model (ARCH) and was described in Engle (1982). It was further generalized by Bollerslev's GARCH model in Bollerslev (1986). If we assume a GARCH(p,q) model for r_t, then

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \end{aligned} \tag{1.25}$$

where $\{\epsilon_t\}$ is a white-noise process, $\{a_i\}_{i=1}^{p}$ and $\{b_j\}_{j=1}^{q}$ are parameters that are estimated from historical data. If $q = 0$, then (1.25) reduces to an ARCH(p) model. A common assumption for the white-noise process is that $\{\epsilon_t\}$ are independent, zero-mean random variables distributed according to the Gaussian law. Even though the tails of the Gaussian distribution decay exponentially, the unconditional distribution of r_t has heavier tails and certain moments could be infinite. Nevertheless recent studies suggest that

[9]The process $\{\epsilon_t\}$ is white-noise process if the random variables $\{\epsilon_t\}$ are uncorrelated, have zero mean and finite variance.

[10]See for example, Box and Jenkins (1976), Hamilton (1994) and Brokwell and Davis (1991).

GARCH-filtered residuals have heavier tails than the Gaussian distribution and thus, a heavy-tailed assumption for the innovations could be more adequate. The stable GARCH model was proposed in Panorska et al. (1995) and Mittnik et al. (1996). Necessary and sufficient conditions for existence and uniqueness of a stationary stable GARCH process were derived. In Mittnik et al. (2002), the stable power-GARCH model, $S_{\alpha,\beta,\delta}$GARCH(p,q), is studied. If $r_t \in S_{\alpha,\beta,\delta}$GARCH(p,q), then

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^\delta &= a_0 + \sum_{i=1}^{p} a_i |r_{t-i}|^\delta + \sum_{j=1}^{q} b_j \sigma_{t-j}^\delta \end{aligned} \tag{1.26}$$

where $\{\epsilon_t\}$ are i.i.d. and $\epsilon_t \in S_\alpha(1, \beta, 0)$ and the power parameter δ satisfies $0 < \delta < \alpha$. The GARCH(p,q) model with Gaussian innovations is a special case of the stable power-GARCH model when $\alpha = \delta = 2$.

Many other autoregressive models have been proposed in literature such as Integrated GARCH, Exponential GARCH, etc. and also ARMA-GARCH generalizations that combine processes of type (1.24) and (1.26).

1.5 Multivariate models

For the purposes of portfolio risk estimation, beside assuming a proper probabilistic model for the distribution of the return of an asset, yet another thing is important. It is the proper modeling of the dependence between assets returns. A well-known fact is that in market crashes, the prices of many assets fall down, i.e. many assets have negative returns at one moment of time and thus the idea that assets returns might be independent breaks down. Certainly it is not only during market crashes that the proper modeling of dependence matters.

The most simple notion of dependence between two random variables is the correlation between them. Unfortunately it is a measure of linear dependence and is the right notion of dependence if the multivariate distribution of assets return is Gaussian. It is certainly inappropriate if assets returns are heavy-tailed, as empirical studies have shown, because of infinite second moments.

Many ways exist to model dependence and many measures have been introduced[11]. One way to deal with the problem is to presume a multivariate distribution for the vector of assets returns. Knowing the multivariate c.d.f., it is possible in theory to compute all kinds of joint probabilities. There are some popular multivariate models like the multivariate Gaussian or the multivariate Student t, etc. that belong to the class of the multivariate elliptical

[11] For more details, see Embrechts et al. (2002)

distributions. This class is fairly flexible and contains representatives with heavy-tails.

Another example is the class of the multivariate stable distributions. It should be remarked that both classes are not mutually exclusive — the sub-Gaussian laws are both elliptical and multivariate stable. Except for the multivariate Gaussian distribution, all multivariate stable laws are heavy-tailed. They are a suitable choice for a theoretical model also because all linear combinations of the marginals, and thus the marginals themselves, are one-dimensional α-stable laws. There is one peculiarity, all linear combinations have one and the same index of stability which is the same as the index of stability of the multivariate law. In effect, using this class as a probabilistic model means that we assume one and the same tail behavior for the assets returns.

This restriction is not present in the class of the operator stable laws. As a matter of fact, multivariate stable laws appear as a special case, i.e. operator stable laws generalize the family of the multivariate stable distributions. Operator stable laws arise from generalization of the central limit theorem and they have domains of attraction. All marginals have different indices of stability that depend on the characteristic exponent of the multivariate law, which is a matrix in this case.

There is yet another way to arrive at a multivariate model. It is to specify separately the marginal distributions and the dependence structure. This could be done through the use of copulas.

1.5.1 Elliptical distributions

The class of the elliptical distributions provides a rich source of multivariate distributions that share many of the tractable properties of the multivariate Gaussian distribution. We begin with a definition and characterization of the sub-class of the spherical distributions. Then we proceed with the elliptical distributions.

Definition 4. *Let X be a n-dimensional random vector. X is said to be spherically distributed, or simply spherical, if and only if $X \stackrel{d}{=} OX$ for every $n \times n$ orthonormal matrix O and $\stackrel{d}{=}$ denotes equality in distribution.*

The class of the spherical distributions is rotationally symmetric according to the definition. This is because orthonormal transforms of vectors do not change their norms but just their orientation. Let U be uniformly distributed on the unit sphere $\mathbb{S}^n$, $\mathbb{S}^n := \{x \in \mathbb{R}^n : ||x|| = 1\}$ where $||x||$ denotes the Euclidean norm. Then every random vector X which can be represented as $X \stackrel{d}{=} RU$ where R is a non-negative random variable independent of U, is rotationally symmetric and thus spherical. The statement also appears correct if considered in reverse direction, any spherical random vector is necessarily representable by RU and R is called the *generating random variate*

of X.

It is possible to characterize the spherical distributions in terms of their characteristic function $\varphi_X(t) = E\exp(i\langle t, X\rangle)$ where $t \in \mathbb{R}^n$ and $\langle t, X\rangle = \sum_{k=1}^n t_k X_k$ denotes the dot product.

Theorem 2. *A random vector X is spherically distributed if and only if its characteristic function has the form $\varphi_X(t) = \phi_X(t^T t)$.*

The function $\phi_X(\cdot)$ is called *characteristic generator* of X. This representation implies a simple form of the characteristic function of any affine transformation of a spherical distribution.

Proposition 1. *Let X be a k-dimensional spherically distributed random vector with characteristic generator $\phi_X(\cdot)$. Further, let $\Lambda \in \mathbb{R}^{n\times k}$ be an arbitrary matrix and $\mu \in \mathbb{R}^n$. Then the characteristic function of $Y = \mu + \Lambda X$ has the form*

$$\varphi_Y(t) = \exp(it^T\mu)\phi_X(t^T\Sigma t), \qquad t \in \mathbb{R}^n$$

where $\Sigma = \Lambda\Lambda^T$.

Proof. The statement is verified directly using the special representation of the characteristic function of spherical random vectors. □

Definition 5. *If X is a n-dimensional random vector and, for some $\mu \in \mathbb{R}^n$ and some $n \times n$ nonnegative definite, symmetric matrix Σ, the characteristic function $\varphi_{X-\mu}(t)$ of $X - \mu$ is a function of the quadratic form $t^T\Sigma t$, $\varphi_{X-\mu}(t) = \phi(t^T\Sigma t)$, we say that X has an elliptical distribution with parameters μ, Σ and ϕ, and we write $X \in E_n(\mu, \Sigma, \phi)$.*

When $n = 1$, the class of elliptical distributions coincides with the class of one-dimensional symmetric distributions. If $X \in E_n(\mu, I, \phi)$ where I is the identity matrix, then X is spherically distributed. Also, because of Proposition 1, every affinely transformed spherical random vector is elliptically distributed. The following stochastic representation shows that the converse is true if the transformation matrix has full rank.

Theorem 3. *$X \in E_n(\mu, \Sigma, \phi)$ with rank$(\Sigma) = k$ if and only if there exist a random variable $R \geq 0$ independent of U, a k-dimensional random vector uniformly distributed on the unit sphere $\mathbb{S}^k$, and a $n \times k$ matrix Λ with $\Lambda\Lambda^T = \Sigma$, such that*

$$X \stackrel{d}{=} \mu + R\Lambda U$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Proof. For the proof, see Fang et al. (1994). □

If $X \in E_n(\mu, \Sigma, \phi)$, where Σ is a diagonal matrix, then X has uncorrelated components if the variance of the components of X is finite. If X has independent components, then $X \in N(\mu, \Sigma)$. It should be remarked that the multivariate normal distribution is the only one among the elliptical distributions where uncorrelated components imply independent components. A random vector $X \in E_n(\mu, \Sigma, \phi)$ does not necessarily have density. If X has a density function $f(x)$, then it has a special form.

Theorem 4. *Let $X \in E_n(\mu, \Sigma, \phi)$ where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite. Then X can be represented stochastically by $X \stackrel{d}{=} \mu + R\Lambda U$ with $\Lambda\Lambda^T = \Sigma$ according to Theorem 3. Further let the c.d.f. of R be absolutely continuous. Then the p.d.f. of X is given by*

$$f_X(x) = \sqrt{\det(\Sigma^{-1})} \cdot g_R\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right), \qquad x \neq \mu, \tag{1.27}$$

where

$$g_R(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{n/2}} t^{-\frac{n-1}{2}} \cdot f_R\left(\sqrt{t}\right), \qquad t > 0 \tag{1.28}$$

and $f_R(t)$ is the p.d.f. of R.

Proof. A proof and a more general result for semi-definite Σ can be found in Frahm (2004). □

The function $g_R(\cdot)$ is called *density generator*. Given the density of the generating variate R, one can compute the density generator of the corresponding elliptical distribution. Note that the contour lines of the density function form ellipsoids in $\mathbb{R}^n$. For this reason the elliptical distributions are often called *elliptically contoured* distributions.

Given the distribution of X, the representation $E_n(\mu, \Sigma, \phi)$ is non-unique. It uniquely determines μ but Σ and $\phi(\cdot)$ are determined up to a positive constant. More precisely, if $X \in E_n(\mu_1, \Sigma_1, \phi_1)$ and $X \in E_n(\mu_2, \Sigma_2, \phi_2)$, then

$$\mu_1 = \mu_2 \qquad \Sigma_1 = c\Sigma_2 \qquad \phi_1(\cdot) = \phi_2(\cdot/c)$$

for some constant $c > 0$. It comes out that it is possible to choose the characteristic generator ϕ such that $\mathrm{cov}(X) = \Sigma$ if covariances are defined, see Embrechts et al. (2003) for an example.

Affine transformations of elliptical random vectors have also elliptical distribution

Theorem 5. *Let $X \in E_n(\mu, \Sigma, \phi)$ and B be a $q \times n$ matrix and $b \in \mathbb{R}^q$. Then*

$$b + BX \in E_n(b + B\mu, B^T \Sigma B, \phi)$$

Proof. For a proof, see Embrechts et al. (2003). □

If we partition X, μ and Σ into

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where X_1 and μ_1 are $r \times 1$ vectors and Σ_{11} is a $r \times r$ matrix. As a corollary from the above theorem if follows that

$$X_1 \in E_r(\mu_1, \Sigma_{11}, \phi) \qquad \text{and} \qquad X_2 \in E_{n-r}(\mu_2, \Sigma_{22}, \phi)$$

Therefore the marginal distributions of the elliptical distributions are elliptical and with the same characteristic generator. The next result states that the conditional distribution of X_1 given X_2 is elliptical but in general not with the same characteristic generator.

Theorem 6. *Let $X \in E_n(\mu, \Sigma, \phi)$ with Σ strictly positive definite. Then the conditional distribution of X_1 given that $X_2 = x$*

$$X_1 | X_2 = x \in E_r(\tilde{\mu}, \tilde{\Sigma}, \tilde{\phi})$$

where $\tilde{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x - \mu_2)$ and $\tilde{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Moreover $\tilde{\phi} = \phi$ if and only if $X \in N(\mu, \Sigma)$.

Proof. For the proof, see Fang et al. (1994). □

The next result states that linear combinations of independent, elliptically distributed random vectors with the same dispersion matrix Σ up to a positive constant remain elliptical.

Theorem 7. *Let $X \in E_n(\mu, \Sigma, \phi)$ and $\widetilde{X} \in E_n(\widetilde{\mu}, c\Sigma, \widetilde{\phi})$ with $c > 0$ be independent. Then for $a, b \in \mathbb{R}$,*

$$aX + b\widetilde{X} \in E_n(a\mu + b\widetilde{\mu}, \Sigma, \phi^*), \quad \textit{where} \quad \phi^*(u) = \phi(a^2 u)\widetilde{\phi}(b^2 c u)$$

Proof. The proof uses the partficular form of the characteristic function from the definition, see Embrechts et al. (2003). □

A practical problem with the elliptical distributions in multivariate risk modeling is that all marginals are of the same type, which may not be very realistic. In the remaining part of the section we give as examples some of the most widely used representatives of the class.

Example 1. ***(Multivariate Gaussian distribution)*** *Let $\mu \in \mathbb{R}^n$ and $\Lambda \in \mathbb{R}^{n \times n}$ such that $\Sigma := \Lambda\Lambda^T \in \mathbb{R}^{n \times n}$ is positive definite. The random vector $X \in N(\mu, \Sigma)$ is elliptically distributed since it is representable as*

$$X \stackrel{d}{=} \mu + \sqrt{\chi_n^2}\Lambda U$$

where U is a n-dimensional random vector with uniform distribution on the unit sphere $\mathbb{S}^n$ and χ^2_n is a χ^2-distributed random variable with n degrees of freedom, independent of U, see Frahm (2004) and the references therein. The random variable $\sqrt{\chi^2_k}$ is the generating variate of X. This is easily seen by considering the standard normal distribution which is the underlying spherical distribution with characteristic generator $\phi_X(s) = \exp(-s/2)$. The density generator can be readily obtained because $R = \sqrt{\chi^2_n}$,

$$g_R(t) = \frac{1}{(2\pi)^{n/2}} \cdot \exp(-t/2)$$

and in line with Theorem 4, we obtain the multivariate Gaussian p.d.f.

$$f_X(x) = \frac{\sqrt{\det \Sigma^{-1}}}{(2\pi)^{n/2}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right), \qquad x \in \mathbb{R}^n \qquad (1.29)$$

Example 2. *(**Multivariate t-distribution**) Consider the random vector*

$$Y \overset{d}{=} \mu + \frac{X}{\sqrt{\frac{\chi^2_\nu}{\nu}}}, \qquad \nu \in \mathbb{N}$$

where $\mu \in \mathbb{R}^n$ and $X \in N(0, \Sigma)$. Then Y is said to be multivariate t-distributed with ν degrees of freedom, location vector μ and dispersion matrix Σ, $Y \in t_n(\mu, \Sigma, \nu)$. The random vector allows for the stochastic representation

$$Y \overset{d}{=} \mu + \frac{1}{\sqrt{\frac{\chi^2_\nu}{\nu}}} \cdot \sqrt{\chi^2_n}\Lambda U$$

where χ^2_ν and χ^2_n are independent and have χ^2 distribution, U is uniformly distributed on the unit sphere $\mathbb{S}^n$ provided that Λ has full rank, and is independent of χ^2_ν and χ^2_n, $\Sigma = \Lambda\Lambda^T$. Hence for the generating variate we have

$$R \overset{d}{=} \sqrt{\frac{\chi^2_n}{\frac{\chi^2_\nu}{\nu}}} \overset{d}{=} \sqrt{nF_{n,\nu}}$$

where $F_{n,\nu}$ denotes a F-distributed random variable. The density generator is

$$g_R(t) = \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{(\nu\pi)^{n/2}} \cdot \left(1 + \frac{t}{\nu}\right)^{-\frac{n+\nu}{2}}$$

and in line with Theorem 4, we obtain the multivariate t-distribution p.d.f.

$$f_X(x) = \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\sqrt{\det \Sigma^{-1}}}{(\nu\pi)^{n/2}} \cdot \left(1 + \frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{\nu}\right)^{-\frac{n+\nu}{2}}, \quad x \in \mathbb{R}^n \tag{1.30}$$

1.5.2 Multivariate stable distributions

Multivariate stable laws can be introduced by extending Definition 1 of one-dimensional stable distributions. The multivariate Gaussian distribution is a particular representative. Gaussian vectors possess the property that any linear combination of its components is a one-dimensional Gaussian distribution. This property is shared by multivariate stable laws — linear combinations of vector components have univariate stable distribution.

As in the one-dimensional case, stable random vectors can be characterized in terms of their characteristic function. Its definition involves a finite measure on the unit sphere in $\mathbb{R}^n$ called *spectral measure* and a vector μ that can be interpreted as a shift parameter. The spectral measure describes the skewness and the scale of the distribution. If certain conditions are imposed on it, the multivariate distribution can become symmetric, that is X and $-X$ have the same distribution.

The sub-Gaussian distributions are a sub-class of symmetric multivariate stable laws and are often considered in the applications because of their simple structure. They can be viewed as Gaussian random vectors with randomized covariance matrix. They also belong to the class of the elliptical distributions and estimators of their dispersion matrix can be readily constructed.

Definition and basic properties

The stability property in $\mathbb{R}^n$ is introduced as in $\mathbb{R}^1$.

Definition 6. *A random vector X is said to be a stable random vector in $\mathbb{R}^n$ if for any positive numbers A and B there is a positive number C and a vector $D \in \mathbb{R}^n$ such that*

$$AX^{(1)} + BX^{(2)} \stackrel{d}{=} CX + D$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of X.

The vector X is called strictly stable if Definition 6 holds with $D = 0$ for any $A > 0$ and $B > 0$. The vector is called symmetric stable if it is stable and satisfies the additional relation $X \stackrel{d}{=} -X$. Concerning the marginal distributions and linear combinations of vector components, we have the next

Theorem 8. *Let X be a stable (respectively, strictly stable, symmetric stable) vector in $\mathbb{R}^n$. Then there is a constant $\alpha \in (0, 2]$ such that in Definition 6, $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Moreover, any linear combination of the components of X of the type $Y = \sum_{k=1}^n b_k X_k = b^T X$ is an α-stable (respectively, strictly α-stable, symmetric α-stable) random variable.*

Proof. A proof is given in Samorodnitsky and Taqqu (1994). □

It is straightforward to obtain the next

Corollary 1. *A random vector is stable if and only if for any $k \geq 2$, there is an $\alpha \in (0, 2]$ and a vector D_k such that*

$$X^{(1)} + X^{(2)} + \ldots + X^{(k)} \stackrel{d}{=} k^{1/\alpha} X + D_k$$

where $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ are independent copies of X.

Multivariate stable vectors have domains of attraction — they are weak limits of normalized sums of i.i.d. random vectors. The result contained in Corollary 1 is a special case of this property. The random vector Y is said to be in the normal domain of attraction of the α-stable vector X if for any sequence of i.i.d. copies $Y^{(1)}, Y^{(2)}, \ldots, Y^{(k)}$..., there is a sequence of vectors A_k such that

$$k^{-1/\alpha}(Y^{(1)} + Y^{(2)} + \ldots + Y^{(k)} - A_k) \stackrel{d}{\to} X \tag{1.31}$$

where the notation $\stackrel{d}{\to}$ denotes convergence in distribution. As a matter of fact, Corollary 1 states that α-stable random vectors belong to their own normal domains of attraction.

The parameter α is called the index of stability or characteristic exponent. According to Theorem 8, all linear combinations $b^T X$ of a stable random vector are α-stable random variables. The converse is generally not true.

Theorem 9. *Let X be a random vector in $\mathbb{R}^n$.*

1. *If all linear combinations $b^T X$ have strictly stable distributions, then X is a strictly stable random vector.*

2. *If all linear combinations are symmetric stable, then X is a symmetric stable random vector.*

3. *If all linear combinations are stable with index of stability $\alpha \geq 1$, then X is a stable random vector.*

Proof. A proof is given in Samorodnitsky and Taqqu (1994). □

As we have remarked, stable random vectors can be described in terms of their characteristic function $\varphi_X(t) = \exp(i\langle t, X\rangle)$.

Theorem 10. *Let $0 < \alpha < 2$. Then X is an α-stable random vector in $\mathbb{R}^n$ if and only if there exists a finite measure Γ on the unit sphere $\mathbb{S}^n$ and a vector $\mu \in \mathbb{R}^n$ such that*

a) if $\alpha \neq 1$,

$$\varphi_X(t) = \exp\left\{-\int_{\mathbb{S}^n} |\langle t,s\rangle|^\alpha \left(1 - i \cdot sign(\langle t,s\rangle) \tan\frac{\pi\alpha}{2}\right)\Gamma(ds) + i\langle t,s\rangle\right\}$$

b) if $\alpha = 1$,

$$\varphi_X(t) = \exp\left\{-\int_{\mathbb{S}^n} |\langle t,s\rangle| \left(1 - i\frac{2}{\pi} sign(\langle t,s\rangle) \log|\langle t,s\rangle|\right)\Gamma(ds) + i\langle t,s\rangle\right\}$$

The pair (Γ, μ) is unique and is called spectral decomposition.

Example 3. *(**One-dimesional stable distributions**) Suppose that $n = 1$. Then $\mathbb{S}^1$ consists of the two points $\{-1\}$ and $\{1\}$ and the spectral measure is concentrated on them. Writing the characteristic function for $\alpha \neq 1$ as given in Theorem 10 and comparing it to the one in Definition 3, we obtain the following relations between the univariate stable parameters and the spectral measure:*

$$\sigma = (\Gamma(\{1\}) + \Gamma(\{-1\}))^{1/\alpha}, \qquad \beta = \frac{\Gamma(\{1\}) - \Gamma(\{-1\})}{\Gamma(\{1\}) + \Gamma(\{-1\})}$$

The skewness parameter β is zero only if the measure is symmetric.

Example 4. *(**Linear combinations**) Let X be a stable random vector with characteristic function given in Theorem 10. We know that any linear combination $Y_b = \langle b, X\rangle$ has an α-stable distribution $S_\alpha(\sigma_b, \beta_b, \mu_b)$. To determine the parameters σ_b, β_b and μ_b, let γ be any real number and set $t = \gamma b$ in the characteristic function given in Theorem 10. The resulting function of γ is the characteristic function of the random variable Y_b. We receive*

$$\sigma_b = \left(\int_{\mathbb{S}^n} |\langle b,s\rangle|^\alpha \Gamma(ds)\right)^{1/\alpha}, \qquad \beta_b = \frac{1}{\sigma_b^\alpha}\int_{\mathbb{S}^n} |\langle b,s\rangle|^\alpha sign(\langle b,s\rangle)\Gamma(ds)$$

and

$$\mu_b = \begin{cases} \langle b,\mu\rangle & \text{if} \quad \alpha \neq 1 \\ \langle b,\mu\rangle - \dfrac{2}{\pi}\displaystyle\int_{\mathbb{S}^n} \langle b,s\rangle \log|\langle b,s\rangle|\Gamma(ds) & \text{if} \quad \alpha = 1 \end{cases}$$

By choosing suitable vectors b, one can obtain the marginal distributions.

A necessary and sufficient condition for a random variable to be symmetric is its characteristic function to be real-valued. The next result describes the characteristic function of symmetric α-stable ($S\alpha S$) vectors.

Theorem 11. *X is a symmetric α-stable vector in $\mathbb{R}^n$ with $0 < \alpha < 2$ if and only if there exists a unique symmetric finite measure on the unit sphere $\mathbb{S}^n$ such that*

$$\varphi_X(t) = \exp\left\{ - \int_{\mathbb{S}^n} |\langle t, s\rangle|^\alpha \Gamma(ds) \right\} \tag{1.32}$$

Γ is the spectral measure of the symmetric α-stable random vector X.

Proof. For a proof, see Samorodnitsky and Taqqu (1994). □

Equation (1.32) holds also in the Gaussian case, but then the spectral measure Γ is not unique and thus the spectral decomposition is not a useful concept[12].

The covariance function is an extremely powerful tool in the study of Gaussian random elements but it is not defined when $\alpha < 2$. The *covariation* is designed to replace the covariance when $1 < \alpha < 2$ but unfortunately it lacks some of the nice properties of the covariance. Nevertheless it is a useful concept. We start with a definition of *signed power*. Let a and p be real numbers. The signed power $a^{<p>}$ is introduced by $a^{<p>} = |a|^p \text{sign}(a)$.

Definition 7. *Let X_1 and X_2 be jointly $S\alpha S$ with $\alpha > 1$ and let Γ be the spectral measure of the random vector (X_1, X_2). The covariation of X_1 on[13] X_2 is the real number*

$$[X_1, X_2]_\alpha = \int_{\mathbb{S}^2} s_1 s_2^{<\alpha-1>} \Gamma(ds)$$

If $\alpha = 2$, it is possible to show that for a bi-variate random vector,

$$\int_{\mathbb{S}^2} s_k^2 \Gamma(ds) = \frac{1}{2} DX_k, \quad k = 1, 2, \qquad \text{and} \qquad [X_1, X_2]_2 = \frac{1}{2}\text{cov}(X_1, X_2)$$

and therefore covariation reduces to covariance.

In the more general case, when we consider n-dimensional random vectors, the following result holds. It shows how to compute the covariation of two components of a random vector in terms of the spectral measure of the vector.

[12] See Samorodnitsky and Taqqu (1994) for several examples.

[13] It appears that covariation is not symmetric in its arguments. Nevertheless, $[X_1, X_2]_\alpha$ is often called *the covariation of X_1 and X_2*, see Samorodnitsky and Taqqu (1994) for details.

Proposition 2. *Let $X = (X_1, X_2, \ldots, X_n)$ be n-dimensional $S\alpha S$ random vector with $\alpha > 1$ and spectral measure Γ_X. Then*

$$[X_i, X_j]_\alpha = \int_{\mathbb{S}^n} s_i s_j^{<\alpha-1>} \Gamma_X(ds)$$

and

$$[X_i, X_i]_\alpha = \int_{\mathbb{S}^n} |s_i|^{<\alpha>} \Gamma_X(ds) = \sigma_{X_i}^\alpha$$

where σ_{X_i} denotes the scale parameter of the vector element X_i.

Proof. For a proof, see Samorodnitsky and Taqqu (1994). □

The covariation is related to the joint moment $EXY^{<p-1>}$.

Proposition 3. *Let (X, Y) be jointly $S\alpha S$ with $\alpha > 1$. Then for all $1 < p < \alpha$,*

$$\frac{EXY^{<p-1>}}{E|Y|^p} = \frac{[X, Y]_\alpha}{||Y||_\alpha^\alpha}$$

where $||Y||_\alpha$ denotes the scale parameter of Y.

In effect, using the expression for the p-th absolute moment in Property 7, we can obtain an estimator for the covariation.

Example 5. *(**Sub-Gaussian distributions**) Let $X \in N(0, Q)$ and Y be a positive random variable such that*

$$Y \in S_{\alpha/2}\left(\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 1, 0\right)$$

Then the random vector $Z = \sqrt{Y}X$ has a $S\alpha S$ distribution since any linear combination of its components is a symmetric stable random variable, see Theorem 9 and Property 10. The random variable Z is said to be a sub-Gaussian symmetric α-stable vector. The characteristic function of Z has a simple form

$$\varphi_Z(t) = \exp\left(-\left(\frac{1}{2}t^T Q t\right)^{\alpha/2}\right) \tag{1.33}$$

see Samorodnitsky and Taqqu (1994) for a proof. Now it is easy to see that the sub-Gaussian stable random vectors have elliptical distributions. Let $\mu \in \mathbb{R}^n$. According to Definition 5, it follows that the random vector $\mu + \sqrt{Y}X \in E_n(\mu, Q, \phi)$, with characteristic generator $\phi(s) = \exp(-(s/2)^{\alpha/2})$.

The covariation between the components of Z can be explicitly computed. Let us denote the elements of Q by q_{ij}, $i, j = 1, 2, \ldots, n$. Thus $q_{ij} = EX_iX_j$

is the covariance between the elements of the Gaussian vector. The covariation between Z_i and Z_j equals[14]

$$[Z_i, Z_j]_\alpha = 2^{-\alpha/2} q_{ij} q_{jj}^{(\alpha-2)/2} = \frac{q_{ij}}{2} \cdot \left(\frac{q_{jj}}{2}\right)^{(\alpha-2)/2} \tag{1.34}$$

and according to Proposition 2, the covariation between Z_j and Z_j equals

$$[Z_j, Z_j]_\alpha = \left(\frac{q_{jj}}{2}\right)^{\alpha/2} = \sigma_{Z_j}^\alpha \tag{1.35}$$

Taking advantage of Proposition 3, Property 7 and Lemma **??**, *it is possible to construct moment-type estimators of the dispersion matrix. Transforming equation (1.35), we obtain*

$$\frac{q_{jj}}{2} = \left(\frac{\cos(p_1\pi/2)\Gamma(1-p_1)}{\Gamma(1-p_1/\alpha)} E|Z_j|^{p_1}\right)^{2/p_1}$$

where $j = 1, 2, \ldots, n$ and from equation (1.34),

$$\frac{q_{ij}}{2} = \left(\frac{q_{jj}}{2}\right)^{\frac{2-p}{2}} \frac{\cos(p\pi/2)\Gamma(1-p)}{\Gamma(1-p/\alpha)} E Z_i Z_j^{<p-1>}$$

where $i, j = 1, 2, \ldots, n$, $i \neq j$, $0 < p_1 < \alpha$ and $1 < p < \alpha$. Certainly the diagonal elements are estimated separately and thus p_1 should not be necessarily equal to p. Suppose that we have a sample of i.i.d. observations $Z^{(1)}, Z^{(2)}, \ldots, Z^{(N)}$ of the random vector Z. Then the moment-type estimators corresponding to the equations above are

$$\frac{\widehat{q}_{jj}}{2} = \left(\frac{\cos(p\pi/2)\Gamma(1-p)}{\Gamma(1-p/\alpha)} \frac{1}{N} \sum_{k=1}^{N} \left|Z_j^{(k)}\right|^p\right)^{2/p}$$

where $j = 1, 2, \ldots, n$ and

$$\frac{\widehat{q}_{ij}}{2} = \left(\frac{\widehat{q}_{jj}}{2}\right)^{\frac{2-p}{2}} \frac{\cos(p\pi/2)\Gamma(1-p)}{\Gamma(1-p/\alpha)} \frac{1}{N} \sum_{k=1}^{N} Z_i^{(k)} \left(Z_j^{(k)}\right)^{<p-1>}$$

where $i, j = 1, 2, \ldots, n, \quad i \neq j$.

The multivariate Gaussian distribution appears as a special case of the sub-Gaussian family when $\alpha \to 2$. This is easily seen from the characteristic function in equation 1.33. Therefore when $p \to \alpha \to 2$, the dispersion matrix Q transforms into the variance-covariance matrix.

[14] See Samorodnitsky and Taqqu (1994) for a proof.

1.5.3 Operator stable distributions

Theorem 8 implies that if we assume the multivariate stable distribution as a theoretical model for assets returns, then all marginals, i.e. asset returns, have one and the same index of stability, hence one and the same tail behavior. Moreover, any portfolio has the same index of stability as any of its components. This may not be very realistic as empirical studies suggest that the tail behavior varies with asset classes, see Rachev and Mittnik (2000). Operator stable distributions generalize the multivariate α-stable distributions. They arise from the generalized central limit theorem with matrix scaling. Matrix scaling is natural when we are dealing with random vectors and the limit distribution allows for components with different tail behavior.

Definition and basic properties

In this section we give the formal definition and some basic properties.

Definition 8. *A random vector X is said to be an operator stable random vector in $\mathbb{R}^n$ if there exists a matrix $E \in \mathbb{R}^{n\times n}$ and a vector a_k such that*

$$k^{-E}(X^{(1)} + X^{(2)} + \ldots + X^{(k)} - a_k) \stackrel{d}{=} X \tag{1.36}$$

where $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ are independent copies of X.

In Definition 8, k^{-E} means an integer number raised to a matrix power. This is defined through the power series expansion of the exponent

$$t^A = \exp(\log(t)A) = I + \sum_{m=1}^{\infty} \frac{(\log(t)A)^m}{m!} \tag{1.37}$$

where t is a positive number, A is an arbitrary square matrix and I is the identity matrix. Certainly equation (1.37) makes sense for linear operators defined in abstract spaces but we shall confine our considerations in $\mathbb{R}^n$ where the linear space of linear operators is represented by the linear space of square matrices. The matrix E is called an *exponent* of the operator stable random vector X.

It is straightforward to find the inverse and the transpose of the scaling matrix from (1.37). Actually

$$\begin{aligned} t^A \cdot t^{-A} &= I \\ (t^A)^T &= t^{A^T} \end{aligned} \tag{1.38}$$

and therefore $(k^{-E})^{-1} = k^E$ and $(k^{-E})^T = k^{-E^T}$. Using the definition and (1.38), we can arrive at an expression for the characteristic function of operator stable laws. Rewriting equation (1.36), we obtain

$$X^{(1)} + X^{(2)} + \ldots + X^{(k)} \stackrel{d}{=} k^E X + a_k$$

The characteristic function of the right hand-side equals

$$\begin{aligned}\varphi_{k^E X + a_k}(t) &= E\exp(i\langle t, k^E X + a_k\rangle) = \exp(i\langle t, a_k\rangle) E\exp(i\langle t, k^E X\rangle) \\ &= \exp(i\langle t, a_k\rangle) E\exp\left(i\langle (k^E)^T t, X\rangle\right) \\ &= \exp(i\langle t, a_k\rangle)\varphi_X\left((k^E)^T t\right) \\ &= \exp(i\langle t, a_k\rangle)\varphi_X\left(k^{E^T} t\right)\end{aligned}$$

The characteristic function of the left hand-side is the k-th power of $\varphi_X(t)$ because the vectors in the sum are i.i.d. and the equality in distribution implies[15]

$$(\varphi_X(t))^k = e^{i\langle t, a_k\rangle}\varphi_X\left(k^{E^T} t\right)$$

Operator stable vectors have domains of attraction.

Definition 9. *We say that the random vector Y belongs to the generalized domain of attraction of some operator stable random vector X with exponent E if there is a sequence of vectors a_k*

$$k^{-E}(Y^{(1)} + Y^{(2)} + \ldots + Y^{(k)} - a_k) \stackrel{d}{\to} X$$

where $Y^{(1)}, Y^{(2)}, \ldots, Y^{(k)}$ are independent copies of Y.

Thus the operator stable laws belong to their own domain of attraction.

The tail behavior of the operator stable random vector is determined by the eigenvalues of its exponent. It is possible to represent[16] every $n \times n$ exponent by a unique spectral decomposition based on the real part of the eigenvalues which allows us to write $E = PBP^{-1}$, where P is a change of coordinates matrix and B is block-diagonal with

$$B = \begin{pmatrix} B_1 & 0 & \ldots & 0 \\ 0 & B_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & B_p \end{pmatrix}$$

where B_i is a $n_i \times n_i$ matrix. Every eigenvalue of B_i has real part equal to a_i, $1/2 \leq a_1 < a_2 < \ldots < a_p$ and $n_1 + n_2 + \ldots + n_p = n$. Since B is block-diagonal and consists of p blocks, it follows that we arrive at

[15] See (Rachev and Mittnik, 2000, 339) for the general case with linear operators, Meerschaert and Scheffler (2003), Meerschaert and Scheffler (2001) and the references therein for more information on operator stable distributions.

[16] See Meerschaert and Scheffler (2003) and Meerschaert and Scheffler (2001).

a spectral decomposition of the entire space $\mathbb{R}^n$ such that any $x \in \mathbb{R}^n$ is representable as $x = x_1 + \ldots + x_p$ where x_i belong to the sub-space spanned by the eigenvectors corresponding to the i-th eigenvalue. Moreover, these sub-spaces are E-invariant, that is if y belongs to the sub-space spanned by the eigenvectors corresponding to the i-th eigenvalue, then the vector Ey belongs to the same sub-space.

Given a non-zero vector $b \in \mathbb{R}^n$, we can write its spectral decomposition as $b = b_1 + \ldots + b_p$ where b_i is in the corresponding E-invariant sub-space. Let us define

$$\alpha(b) = \min_i \left(\frac{1}{a_i} : b_i \neq 0 \right)$$

that is the number $\alpha(b)$ is the smallest $1/a_i$ from those eigenvalues that correspond to non-zero components in the spectral decomposition of the vector b. In Meerschaert and Scheffler (2001) it is shown that the next property holds

Proposition 4. *Let X be an operator stable vector in $\mathbb{R}^n$ with exponent E and let $b \in \mathbb{R}^n$. Then for any small $\delta > 0$ we have*

$$\lambda^{-\alpha(b)-\delta} < P(|\langle b, X\rangle| > \lambda) < \lambda^{-\alpha(b)+\delta} \tag{1.39}$$

for all $\lambda > 0$ sufficiently large.

In other words, the tail behavior of the linear combination $\langle b, X\rangle$ is dominated by the component with the heaviest tail. This also means that $E|\langle b, X\rangle|^s$ exists for $0 < s < \alpha(b)$ and diverges for $s > \alpha(b)$.

If we write the spectral decomposition of the random vector $X = X_1 + \ldots + X_p$, projecting it on the sub-spaces spanned by the corresponding eigenvectors, we see that each X_i from the spectral decomposition is operator stable with some exponent E_i that has the same real part a_i. We say that X_i is spectrally simple with index $\alpha_i = 1/a_i$ and it has the tail behavior (1.39) with $\alpha(b) = \alpha_i$.

Example 6. *(**Multivariate α-stable distributions**) If X is operator stable with the simple exponent $E = aI$ where $a \geq 1/2$ is a real number and I is the identity matrix, then Definition 8 shows that X has a multivariate α-stable distribution. In this case the change of coordinates matrix P in the spectral decomposition is the identity matrix and $B = E$. There is only one spectral component and the tail behavior is the same in every radial direction by Proposition 4.*

Example 7. *(**Operator stable distributions with a diagonal exponent matrix**) Let X be a n-dimensional operator stable vector with exponent*

$$E = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix}$$

where $1/2 \leq a_1 < a_2 < \ldots < a_n$*. Then the spectral decomposition is such that* $p = n$*,* $P = I$ *and* $B = E$*, that is the sub-spaces spanned by the eigenvectors are the coordinate axes. Projecting the vector* X *onto the i-th coordinate axis, shows that* X_i *is* α_i*-stable with* $\alpha_i = 1/a_i$*.*

As a matter of fact, when the exponent is a diagonal matrix, it is possible to calculate the rescaling matrix using the definition (1.37) and noticing that the m-th power of a diagonal matrix translates into raising the diagonal elements to the m-th power,

$$k^{-E} = \begin{pmatrix} k^{-1/\alpha_1} & 0 & \dots & 0 \\ 0 & k^{-1/\alpha_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & k^{-1/\alpha_n} \end{pmatrix}$$

For any non-zero vector b*, the tail of the linear combination* $\langle b, X \rangle$ *decays like the power function* $\lambda^{-\alpha(b)}$*, where* $\alpha(b) = \min(\alpha_i : b_i \neq 0)$*. Thus, if we have a portfolio of assets with multivariate distribution as* X *in this example, the tail behavior of the portfolio returns will be governed by the asset with the heaviest tail in which we have invested a non-zero amount.*

For more examples, see Meerschaert and Scheffler (2003).

1.5.4 Multivariate skewed Student's t distribution

Another multivariate model with applications in finance is a skewed version of Student's t distribution. As the name suggests, it allows for skewed and heavy-tailed representatives. In the literature, there are many ways to introduce a skewed version of Student's t distribution. We focus on one particular model which has a suitable stochastic representation and belongs to the class of subordinated models.

Definition

The form of the multivariate skewed Student's t distribution we are using is defined through the following stochastic representation:[17]

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

[17] For more information, see Section 12.7 in Rachev and Mittnik (2000).

where $W \in IG(\nu/2, \nu/2)$, and $Z \in N(0, \Sigma)$, Z is independent of W, $\gamma = (\gamma_1, \ldots, \gamma_n)$ is a n-dimensional vector accounting for the skewness, $\mu = (\mu_1, \ldots, \mu_n)$ is n-dimensional location parameter vector and ν stands for the degrees of freedom. We denote this distribution by $X \in t_n(\nu, \mu, \Sigma, \gamma)$. The notation $IG(\nu/2, \nu/2)$ stands for the inverse gamma distribution with parameters $\nu/2$. Thus, W is a one-dimensional random variable and Z is a random vector having a zero-mean multivariate normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{pmatrix}.$$

The multivariate skewed Student's t distribution allows for closed-form expression of its density,

$$f_X(x) = \frac{aK_{(\nu+n)/2}\left(\sqrt{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma}\right)\exp((x-\mu)'\Sigma^{-1}\gamma)}{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma)^{-\frac{\nu+n}{4}}\left(1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\nu}\right)^{\frac{\nu+n}{2}}}$$

where $x \in \mathbb{R}^n$, and K denotes the modified Bessel function of the third kind and

$$a = \frac{2^{\frac{2-\nu-n}{2}}}{\Gamma(\nu/2)(\pi\nu)^{n/2}\sqrt{|\Sigma|}}.$$

1.5.5 Generalized subordinated models

In section 1.5.3, we discussed operator stable distributions. We mentioned they are a class of multivariate distributions allowing for different tail behavior of the marginals. In practice however, they are too complicated to be used. Thus, we employ other multivariate models which, in a way similar to operator stable distributions, allow for different tail behavior of the marginals. We call these models *generalized subordinated models* because the construct is based on the idea of subordination. Their specific feature is that we use different subordinators for the different marginals. Thus, we have the flexibility of choosing the subordinators to be a vector of dependent or independent random variables. Since the subordinators control the extreme events, the corresponding models can be called *dependent* or *independent* tail models. From a practical viewpoint, the dependent tail models make more sense because in market crashes extreme negative returns occur jointly. Thus, we will describe the dependent tail models only.

The generalized sub-Gaussian stable model

The generalized sub-Gaussian stable model, as the name suggests, is based on the idea of the sub-Gaussian stable distribution we considered in Example 5. The multivariate distribution is defined through the following stochastic representation

$$Z = \begin{pmatrix} \sqrt{Y_1}X_1 \\ \sqrt{Y_2}X_2 \\ \dots \\ \sqrt{Y_n}X_n \end{pmatrix} \tag{1.40}$$

where $X = (X_1, \dots, X_n) \in N(0, \Sigma)$ and $Y = (Y_1, \dots, Y_n)$ is a vector of subordinators such that

$$Y_i \in S_{\alpha_i/2}\left(\left(\cos\frac{\pi\alpha_i}{4}\right)^{2/\alpha_i}, 1, 0\right), \quad i = 1, n$$

and all components of Y are functionally dependent. That is, all components of Y can be represented as $F_i^{-1}(U)$, $i = 1, n$, where U is one and the same random variable for all components with a uniform distribution. The random vector Z is said to be a generalized sub-Gaussian symmetric α-stable vector.

Apparently, the tail behavior of each component is governed by the tail index α_i which means that the model allows for different tail behavior. Furthermore, the sub-Gaussian distribution appears as a special case if $\alpha_1 = \ldots = \alpha_n$.

The generalized skewed stable model

The generalized skewed stable model can be viewed as an extension to the generalized sub-Gaussian stable model in which an additional term is added to account for skewness. The stochastic representation is as follows,

$$W = \begin{pmatrix} 2^{-1/\alpha_1}Z_1 + V_1 \\ 2^{-1/\alpha_2}Z_2 + V_2 \\ \dots \\ 2^{-1/\alpha_n}Z_n + V_n \end{pmatrix}$$

where the vector $Z = (Z_1, \dots, Z_n)$ has the generalized sub-Gaussian stable distribution with the stochastic representation given in (1.40) and the random variables $V_1, \dots, V_n$ are independent and V_i distributed according to a stable law with parameters:

$$\alpha = \alpha_i, \quad \beta = 2\beta_i, \quad \sigma = 2^{-1/\alpha_2}\sigma_i, \quad \mu = \mu_i$$

In effect, the random variable W_i, which is the i-th component of the vector W, has a skewed stable distribution $W_i \in S_{\alpha_i}(\sigma_i, \beta_i, \mu_i)$.

The generalized skewed Student's t model

The generalized skewed Student's t model is constructed in a similar way to the generalized sub-Gaussian stable model. The multivariate distribution is introduced through the following stochastic representation,

$$Z = \begin{pmatrix} \mu_1 + \gamma_1 W_1 + \sqrt{W_1} Z_1 \\ \mu_2 + \gamma_2 W_2 + \sqrt{W_2} Z_2 \\ \dots \\ \mu_n + \gamma_n W_n + \sqrt{Y_n} Z_n \end{pmatrix} \tag{1.41}$$

where $Z \in N(0, \Sigma)$, $W_i \in IG(\nu/2, \nu/2)$, $i = 1, \dots, n$ and W_i are independent of Z but W_i are functionally dependent among themselves, i.e. $W_i = F_i^{-1}(U)$ where U is one and the same uniformly distributed random variable. The random vector Z is said to be a generalized skewed Student's t vector.

1.5.6 Constructing multivariate distributions with copulas

It is possible to construct a multivariate model by merging together one-dimensional distributional assumptions for the marginals and one special multivariate function responsible for the modeling of the dependence between them. This special function is called a *copula*. The word *copula* is coined to emphasize that this is a function which "joins togethe" one-dimensional distribution functions to form multivariate distribution functions. Technically a function is a copula if it is a cumulative distribution function on the unit cube $[0, 1]^n$ with uniformly distributed marginals. In this section, we briefly describe this technique without delving into the details of copula properties or exhaust all available methods for copula construction. Throughout this section, for a function $f(\cdot)$, we denote by $Domf$ and $Ranf$ the domain and the range of $f(\cdot)$, i.e. $f : Domf \to Ranf$.

Definition and basic properties

We give the definition[18] and some basic properties.

Definition 10. *An n-dimensional copula is a function $C(\cdot)$ with domain $[0, 1]^n$ such that $C(\cdot)$ is a cumulative distribution function and the marginals $C_k(\cdot)$, $k = 1, 2 \dots, n$ satisfy $C_k(x) = x$, $x \in [0, 1]$.*

The most important result regarding copulas is the following theorem known as Sklar's theorem.

Theorem 12. *Let $H(\cdot)$ be an n-dimensional c.d.f. with marginals $F_1(\cdot)$, $F_2(\cdot)$, $\dots$, $F_n(\cdot)$. Then there exists a copula $C(\cdot)$ such that for all $x \in \mathbb{R}^n$,*

[18] For a more detailed introduction into the properties of the copula functions and their application, see Embrechts et al. (2003) and Nelsen (1999).

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

If $F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot)$ are all continuous, then $C(\cdot)$ is unique; otherwise the copula is uniquely determined on $RanF_1 \times \ldots \times RanF_n$. Conversely, if $C(\cdot)$ is a copula and $F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot)$ are distribution functions, then the function $H(\cdot)$ is an n-dimensional c.d.f. with marginals $F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot)$.

Proof. For a proof, see Sklar (1996). □

From Sklar's theorem, we see that for continuous multivariate c.d.f., the univariate marginals and the multivariate dependence structure can be separated and the dependence structure can be represented by a copula.

Next we define the generalized inverse of a c.d.f.

Definition 11. *Let $F(\cdot)$ be a univariate distribution function. We define the generalized inverse of $F(\cdot)$ as*

$$F^{-1}(t) = \inf\{x \in \mathbb{R} | F(x) \geq t\}$$

for all $t \in [0, 1]$ using the convention that $\inf\{\emptyset\} = -\infty$.

Corollary 2. *Let $H(\cdot)$ be an n-dimensional c.d.f. with continuous marginals $F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot)$ and a copula $C(\cdot)$. Then for any $u = (u_1, \ldots, u_n) \in [0, 1]^n$,*

$$C(u_1, \ldots, u_n) = H(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n))$$

The corollary can be used to calculate the copula of a given multivariate c.d.f. In the following examples we illustrate that.

Example 8. *(**The Gaussian copula**) Let $\Phi(\cdot)$ denote the standard one-dimensional normal distribution function and $\Phi_\Sigma(\cdot)$ denote the standard multivariate normal distribution function with correlation matrix Σ. Then*

$$\begin{aligned} C(u_1, \ldots, u_n) &= \Phi_\Sigma\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \ldots \int_{-\infty}^{\Phi^{-1}(u_n)} f(x_1, \ldots, x_n) dx_1 \ldots dx_n \end{aligned} \tag{1.42}$$

where $f(x_1, \ldots, x_n)$ is the multivariate Gaussian p.d.f. given in (1.29), represents the Gaussian copula because we know that all marginals of the standard multivariate normal distribution are one-dimensional standard normal distributions. Now it is possible to construct a multivariate model $G(\cdot)$ that has a Gaussian copula and arbitrary continuous marginals by considering

$$G(x_1, \ldots, x_n) = \Phi_\Sigma\left(\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_n(x_n))\right)$$

where $F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot)$ are the desired continuous c.d.f. of the marginals.

Example 9. *(**The t-distribution copula**) Let $t_\nu(\cdot)$ denote the one-dimensional t-distribution c.d.f. with ν degrees of freedom and $t_{\nu,\Sigma}(\cdot)$ denote the multivariate t-distribution c.d.f. with ν degrees of freedom and dispersion matrix Σ. The copula of the t-distribution*[19] *is given by*

$$\begin{aligned} C(u_1,\ldots,u_n) &= t_{\nu,\Sigma}\left(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_n)\right) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \ldots \int_{-\infty}^{t_\nu^{-1}(u_n)} f(x_1,\ldots,x_n)dx_1\ldots dx_n \end{aligned} \tag{1.43}$$

where $f(x_1,\ldots,x_n)$ is the multivariate t-distribution p.d.f. given in equation (1.30). Again, as in the previous example, one can construct a multivariate model with the t-distribution copula and arbitrary continuous marginals.

Example 10. *(**The skewed t-distribution copula**) The skewed Student's t copula is defined as the copula of the multivariate distribution of X. Therefore, the copula function is*

$$C(u_1,\ldots,u_n) = F_X(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n))$$

where F_X is the multivariate distribution function of X and $F_k^{-1}(u_k)$, $k = 1,n$ is the inverse c.d.f of the k-th marginal of X. That is, $F_X(x)$ has the density $f_X(x)$ defined above and the density function $f_k(x)$ of each marginal is

$$f_k(x) = \frac{aK_{(\nu+1)/2}\left(\sqrt{\left(\nu+\frac{(x-\mu_k)^2}{\sigma_{kk}}\right)\frac{\gamma_k^2}{\sigma_{kk}}}\right)\exp\left((x-\mu_k)\frac{\gamma_k}{\sigma_{kk}}\right)}{\left(\left(\nu+\frac{(x-\mu_k)^2}{\sigma_{kk}}\right)\frac{\gamma_k^2}{\sigma_{kk}}\right)^{-\frac{\nu+1}{4}}\left(1+\frac{(x-\mu_k)^2}{\nu\sigma_{kk}}\right)^{\frac{\nu+1}{2}}}, \tag{1.44}$$

where $x \in R$, σ_{kk} is the k-th diagonal element in the matrix Σ.

A copula can also be specified explicitly, without a reference to some known multivariate distribution.

Example 11. *(**The bivariate Gumbel copula**) Consider the bivariate family of functions*

$$C_\theta(u,v) = \exp\left(-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right) \tag{1.45}$$

where $(u,v) \in [0,1]^2$ and $\theta \geq 1$. It is simple to verify that $C_\theta(u,v)$ is a bivariate c.d.f. in the unit cube. Equation (1.45) is known as the bivariate Gumbel family. It is a class of copulas, which is a sub-class of the Archimedean copulas, see Embrechts et al. (2003) and Nelsen (1999).

[19] The copula of the t-distribution is usually called just t-copula.

Let $X_1, X_2, \ldots, X_n$ be random variables with csontinuous c.d.f. $F_1(\cdot)$, $F_2(\cdot), \ldots, F_n(\cdot)$ and joint distribution function $H(\cdot)$. Then there is a unique copula describing the dependence of the vector $(X_1, X_2, \ldots, X_n)^T$,

$$H(x_1, \ldots, x_n) = P(X_1 < x_1, \ldots, X_n < x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

The random variables are independent if and only if

$$H(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_k(x_k)$$

and we can arrive at a particular expression for the copula of a vector with independent components.

Theorem 13. *Let $(X_1, X_2, \ldots, X_n)^T$ be a vector of continuous random variables with copula $C(\cdot)$. $X_1, X_2, \ldots, Xn$ are independent if and only if $C(u_1, \ldots, u_n) = u_1 u_2 \ldots u_n$.*

One nice property of copulas is that for strictly monotone transformations of the random variables, copulas are either invariant, or change in certain simple ways. Note that if a random variable is continuous, a strictly monotone transformation of it is also a continuous random variable.

Theorem 14. *Let $(X_1, X_2, \ldots, X_n)^T$ be a vector of continuous random variables with copula $C(\cdot)$. If for $k = 1, 2, \ldots, n$, $g_k(\cdot)$ is strictly increasing, then the vector $(g_1(X_1), g_2(X_2), \ldots, g_n(X_n))^T$ has the same copula $C(\cdot)$.*

Proof. For a proof, see Embrechts et al. (2003). □

The general case, in which $g_k(\cdot)$ are strictly monotone, is discussed for example in Embrechts et al. (2003).

The reasoning in examples 8 and 9 shows that in order to construct a multivariate model, it is sufficient to have one-dimensional assumptions for the marginals and a different assumption for the dependence structure through the copula. Thus choosing an appropriate copula is related to presuming a proper dependence model. As intuition suggests, some dependence concepts are indeed copula properties. We shall consider only the concept of tail dependence, for a more thorough description see Embrechts et al. (2003). The concept of tail dependence relates to the amount of dependence in the upper-right-quadrant tail or the lower-left-quadrant tail and is appropriate for studying the dependence between extreme events. Since it appears that the tail dependence between two random variables X_1 and X_2 is a copula property, it is invariant under strictly increasing transformation of X_1 and X_2.

Definition 12. *Let $(X_1, X_2)^T$ be a vector of continuous random variables with marginal distribution functions $F_1(\cdot)$ and $F_2(\cdot)$, and a copula $C(\cdot)$.*

a) The coefficient of upper tail dependence λ_U *of* $(X_1, X_2)^T$ *is defined as*

$$\lambda_U = \lim_{u \to 1^-} P\left(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)\right)$$

or equivalently as

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

provided that the limit $\lambda_U \in [0, 1]$ *exists. If* $\lambda_U \in (0, 1]$, X_1 *and* X_2 *are said to be asymptotically dependent in the upper tail, or the copula is said to have upper tail dependence. If* $\lambda_U = 0$, X_1 *and* X_2 *are said to be asymptotically independent in the upper tail, or the copula is said to have upper tail independence.*

b) The coefficient of lower tail dependence λ_L *of* $(X_1, X_2)^T$ *is defined as*

$$\lambda_U = \lim_{u \to 0^+} P\left(X_2 < F_2^{-1}(u) | X_1 < F_1^{-1}(u)\right)$$

or equivalently as

$$\lambda_U = \lim_{u \to 0^+} \frac{C(u, u)}{u}$$

provided that the limit $\lambda_L \in [0, 1]$ *exists. If* $\lambda_L \in (0, 1]$, X_1 *and* X_2 *are said to be asymptotically dependent in the lower tail, or the copula is said to have lower tail dependence. If* $\lambda_L = 0$, X_1 *and* X_2 *are said to be asymptotically independent in the lower tail, or the copula is said to have lower tail independence.*

It turns out that the Gaussian copula is upper and lower tail independent on condition that the correlation coefficient is below one. The t-distribution copula is upper and lower tail dependent which is influenced by the degrees of freedom and the correlation coefficient, a table of values for λ_U is given in Embrechts et al. (2003). The tail dependence is radially symmetric in these two examples because they are particular cases of the family of the elliptical distributions. In the general case, symmetry is not necessarily present.

1.6 The Monte Carlo method

Numerical methods that are known as Monte Carlo methods can be loosely described as statistical simulation techniques. Statistical simulation is defined, in quite general terms, to be any method that utilizes sequences of random numbers to perform simulation. This technique derives its name

from the casinos in Monte Carlo — a Monte Carlo simulation uses random numbers to model a process.

In the field of mathematical finance, Monte Carlo methods are used to value and analyze basic financial models as well as complex instruments, portfolios and investments by simulating various sources of uncertainty affecting their value. Afterwards, on the basis of the generated scenarios, we determine a number of statistics. The advantage of Monte Carlo methods over other techniques become more evident as the dimensions (sources of uncertainty) of the problem increase.

Traditionally, Monte Carlo techniques depend on a number generation method that mimics randomness as well as possible. The general principle of random number generation is as follows. Given the current value of one or more (usually internally stored) state variables, apply an iterative mathematical algorithm to obtain a new set of values for the state variables, and use a specific formula to obtain a new uniform (0, 1) variate from the current values of all the state variables. The generated numbers can never be truly random, only *pseudo*-random as they are generated according to some deterministic algorithm and after a (large) number of random number generations the sequence will start repeating itself. The number of iterations before replication starts is a measure of the quality of a random number generator and usually is called the *period* of the random number generator.

The most commonly used pseudo-random number generation methods are known as *congruential generators* first proposed by Lehmer (1951). The basic idea is to produce integer values m_n on a given interval $[0, M-1]$, and to return a uniform $(0, 1)$ variate u_n by rescaling[20]. The next integer variate is calculated by the mod[21] operation

$$m_{n+1} = (am_n + c) \mod M \tag{1.46}$$

where M, a, and c are called the modulus, the multiplier, and the increment, respectively.

Note that equation (1.46) is piecewise affine with the same multiplier over all pieces. Thus, it preserves the volume of any given subinterval of $[0, M-1]$, which is why it is called a congruential generator. Frequently, the constant c in equation (1.46) is chosen to be zero, whence we commonly encounter the name *linear congruential generator*. There is a lot of literature on good choices for a and M, and not all of it is trustworthy. One reasonable choice, proposed by Park and Miller (1988) is: $a = 16807$ and $M = 2^{31} - 1$.

[20] The common method for rescaling is to set $u_n = m_n/M$. Since m_n can, however, take on the value 0, which we usually want to avoid, it is recommended to rescale according to $u_n = (m_n + 1)/(M + 1)$.

[21] *mod* operation finds the remainder of division of one number by another. Given two numbers, a (the dividend) and n (the divisor), a **modulo** n (abbreviated as, and sometimes red as "a mod n") is the remainder, on division of a by n.

The period of this generator is equal to $2^{31}-2$. These constants are normally set in the code and are never user-defined.

Another random number generator technique that has become increasingly popular recently is the *Mersenne Twister*. The name is to indicate that the period of the sequence is a Mersenne number, i.e. a prime number that can be written in the form $2^n - 1$ for some $n \in \mathbb{N}$. The period of the Mersenne twister as published by Matsumoto and Nishimura (1998) is $2^{19937} - 1$. Clearly, for all practical purposes, this number generator can be assumed to have infinite periodicity. The Cognity scenario generation engine uses the Mersenne twister generator.

Most simulations entail sampling random variables or random vectors from distributions other than the uniform. A typical simulation procedure uses methods for transforming samples from the uniform distribution to samples from other distributions. There is a large literature on both general purpose methods and specialized algorithms for specific cases. In the appendix to this chapter, we provide details about two of most widely used methods: the *inverse transform method* and the *acceptance-rejection method*.

1.6.1 Generation of scenarios from a stable distribution

For generation of random numbers from a stable distribution, we take advantage of the Chambers-Mallows-Stuck generator which is described below.

Property 11. ***Chambers-Mallows-Stuck generator.*** *The following algorithm generates random numbers distributed according to* $S_\alpha(\sigma, \beta, \mu)$.

1. *Generate two independent random numbers* $u_1 \in Exp(1)$ *and* $u_2 \in U(-\pi/2, \pi/2)$.

   (a) *If* $\alpha \neq 1$,

   $$z_i = s_{\alpha,\beta} \times \frac{\sin \alpha(u_2 + b_{\alpha,\beta}))}{(\cos u_2)^{1/\alpha}} \times \left(\frac{\cos(u_2 - \alpha(u_2 + b_{\alpha,\beta}))}{u_1} \right)^{(1-\alpha)/\alpha}$$

   where

   $$s_{\alpha,\beta} = \left[1 + \beta^2 \tan^2 \frac{\pi\alpha}{2} \right]^{1/(2\alpha)}$$

   $$b_{\alpha,\beta} = \frac{\arctan(\beta \tan \frac{\pi\alpha}{2})}{\alpha},$$

   (b) *If* $\alpha = 1$,

   $$z_i = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta u_2 \right) \tan u_2 - \beta \log \left(\frac{u_1 \cos u_2}{\frac{\pi}{2} + \beta u_2} \right) \right]$$

2. *The random variable z has normalized stable distribution, i.e. the scale parameter is equal to 1 and the location parameter is equal to 0. For arbitrary values of σ and μ, use the following transform*

$$s = \sigma z + \mu$$

The distribution of the random variable z is $S_\alpha(\sigma, \beta, \mu)$.

For additional information on numerical work with stable distributions, see for example Janicki and Weron (1994).

1.6.2 Generation of scenarios from a multivariate normal distribution

The first step in the algorithm for sampling from a multivariate normal law is the decomposition of the covariance matrix. In a risk system, it makes sense to have a number of methods for covariance matrix decomposition. The two traditional algorithms are — the Cholesky decomposition and the eigenvalue decomposition. The former is computationally more efficient but works only with positive definite matrices. The latter is more general and works with positive semi-definite matrices. Before any method for decomposition is executed, the loaded covariance matrix can be verified for consistency. We use a necessary condition, which is the Cauchy-Schwartz inequality:

$$cov(r_1, r_2) \leq \sqrt{Var(r_1)}\sqrt{Var(r_2)}$$

or in terms of the covariance matrix elements:

$$|\sigma_{ij}| \leq \sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}, \quad \text{where} \quad \Sigma = \{\sigma_{ij}\}_{i,i=1}^{n}$$

This inequality may signal which elements in the covariance matrix are problematic if the covariance matrix is improper. This question arises when the covariance matrix is manipulated after being estimated if, for example, we would like to include stress-tests on the correlations between some variables.

Cholesky decomposition The assumption here is that the estimated matrix Σ is positive definite. Find a matrix C, such that:

$$C^T C = \Sigma$$

The matrix C is a triangular one. The result from this decomposition, which the next step uses, is the transposed matrix C, i.e. the result is $A = C^T$.

Eigenvalue decomposition The assumption here is that we work with positive semi-definite matrices. Every covariance matrix is positive semi definite. Decompose the estimated covariance matrix into a product of three matrices:

$$\Sigma = B\Lambda B^T$$

where Λ is a diagonal matrix with the eigenvalues λ_i on the main diagonal, B is a matrix with the eigenvectors corresponding to the eigenvalues in Λ.

This decomposition is always possible because the covariance matrix is real and symmetric; moreover the eigenvalues are non-negative (this fact is used for a second verification for covariance matrix consistency). There are standard algorithms to find the eigenvalues and the corresponding eigenvectors. Having obtained the matrix B and the matrix Λ , we receive the matrix A:

$$A = B\sqrt{\Lambda}B^T$$

where with $\sqrt{\Lambda}$ we denote a matrix, such that $\sqrt{\Lambda}\sqrt{\Lambda} = \Lambda$. Therefore the matrix $\sqrt{\Lambda}$ is diagonal and the elements on the main diagonal are $\sqrt{\lambda_i}$. It can be easily verified that the resulting matrix A in this case is symmetric.

Generation of multivariate normal vectors This stage can be subdivided into three parts:

- Generate n independent random numbers from standard Gaussian distribution $N(0, 1)$ with mean 0 and variance 1. We have a n-dimensional vector $X = (X_1, X_2, \ldots, X_n)$ with independent components, i.e. the multivariate distribution is Gaussian with zero mean and the covariance matrix is the identity matrix. The algorithm for generation of Gaussian random numbers follows the Box-Muller approach:

  1. Generate two independent uniformly distributed random numbers $U_1, U_2 \sim U(0, 1)$,
  2. Compute X_1 and X_2 using the following transform:

  $$X_1 = \sqrt{-2\ln U_1}\cos 2\pi U_2$$

  $$X_2 = \sqrt{-2\ln U_1}\sin 2\pi U_2$$

  3. X_1 and X_2 are independent, $N(0, 1)$ distributed random numbers. Repeat previous steps until we have the desired number of random variates.

- Using the matrix A, computed in the previous step, calculate

  $$Y = AX$$

  As a result, Y has multivariate Gaussian distribution with covariance matrix equal to the estimated matrix Σ.

- Repeat previous two steps N number of times. In effect we have random vectors $(Y_k, k = 1, 2, \ldots, N)$ drawn from a n-dimensional Gaussian distribution with desired parameters, or, to phrase it differently, N states of the world.

$$\begin{pmatrix} Y_{11} & Y_{12} & \ldots & Y_{1n} \\ Y_{21} & Y_{22} & \ldots & Y_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ Y_{N1} & Y_{N2} & \ldots & Y_{nN} \end{pmatrix}$$

In a practical risk system, the algorithm starts with the Cholesky decomposition and if the algorithm cannot produce the decomposed matrix, it switches to the more general eigenvalue decomposition. This is necessary for greater efficiency. In case the matrix is positive definite, there is no need to resort to the slower algorithm even if it is more general.

1.6.3 Generation of scenarios from the Generalized sub-Gaussian stable model

The Generalized sub-Gaussian stable model is a practical application of the stochastic representation. In this section, we assume that the stable parameters of all marginal distributions have been already estimated. We denote them with $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i, \hat{\mu}_i)$, for $i = 1, \ldots, n$. The estimation can be performed according to any of the methods discussed in this Chapter.

1. **Generation of multivariate Gaussian random vectors.** Simulate i.i.d. n-dimensional random vectors y_k, $k = 1, 2, \ldots, N$ drawn from a multivariate Gaussian law with covariance matrix equal to the estimated matrix Σ.

2. **Generation of subordinators.** Simulate i.i.d. n-dimensional random vectors x_k, $k = 1, 2, \ldots, N$ with components distributed according to a stable law with parameters

$$\alpha_i = \frac{\widehat{\alpha}_i}{2}, \quad \beta_i = 0, \quad \sigma_i = \frac{\widehat{\sigma}_i^2}{\mathrm{cov}(i,i)/2} \left[\cos \frac{\pi \widehat{\alpha}_i}{4}\right]^{2/\widehat{\alpha}_i}, \quad \mu_i = 0 \quad (1.47)$$

   where $i = 1, 2, \ldots, n$. The generation is performed by fixing the random numbers u_1 and u_2 in the Chambers-Mallows-Stuck algorithm for all risk variables, see Property 11. In this manner, we obtain functionally dependent subordinators.

3. **Generation of final simulations.**

   Using the generated vectors from the previous steps, compute:

$$z_k = y_k\sqrt{x_k}, \qquad k = 1, 2, \ldots, N$$

where by multiplication of vectors we mean element by element multiplication, i.e. z_k is an n-dimensional random vector, the i-th component of the vector z_k is distributed according to a symmetric stable law with parameters $\widehat{\alpha}_i$ and $\widehat{\sigma}_i$.

1.6.4 Generation of scenarios from the Generalized skewed stable model

One additional step is added to the algorithm for the Generalized Symmetric Stable Model that is responsible for the asymmetry.

1. **Generation of multivariate Gaussian random vectors.** The same as the corresponding step in Section 1.6.3.

2. **Generation of subordinators.** The same as the corresponding step in Section 1.6.3.

3. **Generation of final simulations.** The same as the corresponding step in Section 1.6.3.

4. **Generation of skewed stable variates.** Simulate i.i.d. vectors v_k, $k = 1, 2, \ldots, N$ with independent components, each component distributed according to a stable law with parameters:

$$\alpha_i = \widehat{\alpha}_i, \quad \beta_i = 2\widehat{\beta}_i, \quad \sigma_i = 2^{1/\widehat{\alpha}_i}\widehat{\sigma}_i, \quad \mu_i = 0$$

Then using the random vectors z_k generated in the previous step, compute

$$w_k = 2^{-2/\widehat{\alpha}_i}\left(2^{1/\widehat{\alpha}_i} z_k + v_k\right)$$

Chapter 2

Market Risk

2.1 Introduction

The concept of investment risk has become central in portfolio theory and many definitions have been proposed in the literature. In a recent survey, Holton (2004), risk is defined as "an exposure to a proposition of which one is uncertain". It can be argued that risk is a subjective phenomenon, it is related to an investor's perception of exposure and uncertainty. Even if a true operational definition is impossible to obtain, still we would like to define some aspects of risk that would allow to quantify it.

Risk is commonly characterized as a subjective, relative and multi - dimensional phenomenon.[1] It is relative and multi-dimensional in the sense that we might want to measure it with respect to a benchmark, or more generally, with respect to multiple benchmarks. Risk is also *asymmetric* because it is related to a potential loss; that is, downside movement of prices. In this aspect, risk is not identical with uncertainty, thought the two concepts are related.

Attempts to quantify risk have led to the notion of a *risk measure*. It is a functional that assigns a numerical value to a random variable which is interpreted as loss. Since risk is subjective, risk measures are strongly related to "utility functions". From historical point of view, the optimal investment decision always corresponds to the solution of an "expected utility maximization problem". In particular, the link between expected utility theory and the risk of some admissible investments is generally represented by the consistency of the risk measure with a stochastic order. As a consequence of the consistency, the best investments of a given category of investors (non-satiable, risk-averse, non-satiable and risk-averse) are among the less risky ones. But the converse is generally not true; that is, we cannot guarantee that all the less risky choices are the best ones even if the risk measure is consistent with some stochastic order. In fact, any risk measure associates

[1] See the Rachev, Fabozzi and Menn (2005)

only a real number to a random variable, while the stochastic orders compare cumulative distribution functions. Then, intuitively, a single number cannot summarize the information derived from a portfolio distribution function. This is the principal reason every risk measure is in a sense incomplete.

From historical perspective, Markowitz (1952) was the first to recognize the relationship between risk and reward and introduced standard deviation as a proxy for risk. Thus standard deviation was recognized as a measure of risk. Unfortunately risk is an asymmetric phenomenon and thus standard deviation is not a good risk measure because it symmetrically penalizes potential loss as well as potential profit. Also, it is incapable of describing the risk of low-probability events as the default risk. Finally, it is not consistent with second-order stochastic dominance and thus with the expected utility approach for portfolio selection[2]. The deficiencies of the standard deviation as a risk measure were acknowledged by Markowitz who was the first to suggest the semi-standard deviation as a substitute, Markowitz (1959). Ogryczak and Ruszczynski (2001) also proposed semi-variance models and showed the consistency with second-order stochastic dominance. Other risk measures have also been proposed, for example Value-at-Risk and Expected tail loss. The former is widely used in practice but does not satisfy the very natural property of sub-additivity[3]. The latter is consistent with second order stochastic dominance under some weak conditions.

It is only recently that a systematic approach towards risk measures has been undertaken. Artzner et al. (1998), concerned with banking regulations, proposed an axiomatic approach to the definition of a risk measure. They presented a set of four properties for measures of risk and they called measures satisfying these properties, *coherent risk measures*. Unfortunately, coherent risk measures are generally not consistent with second order stochastic dominance, see Pflug (1998). Consistency with second order stochastic dominance of the more general family of the *convex risk measures* was studied in DeGiorgi (2005).

2.2 Risk measures

In this section we give a formal definition of a risk measure and also examples of some classes of risk measures[4]. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G}$ be a set of real-valued random variables which are $\mathcal{F}$-measurable.

Definition 13. *A risk measure is a mapping* $\rho : \mathcal{G} \to \mathbb{R}$.

Every risk measure induces a class of random variables

[2]This is illustrated by the (μ, σ) paradox.
[3]See the next section.
[4]More examples can be found in Pflug (1998)

$$\mathcal{A}_\rho := \{X \in \mathcal{G} | \rho(X) \leq 0\}$$

which are *acceptable* in the sense that they carry no positive risk. The class $\mathcal{A}_\rho$ is called the *acceptance set* of the risk measure $\rho(\cdot)$.

2.2.1 Coherent risk measures

Artzner et al. (1998) propose several axioms that a risk measure should satisfy, otherwise it may lead to undesirable conclusions. The risk measure is defined on the future value of a position.

Definition 14. *A functional $\rho(\cdot)$ on the space of real-valued random variables is called a coherent risk measure if it is:*

1. *monotonous:* $X,\ Y \in \mathcal{G};\ Y \geq X \quad \Longrightarrow \quad \rho(Y) \leq \rho(X)$
2. *sub-additive:* $X,\ Y,\ X+Y \in \mathcal{G} \quad \Longrightarrow \quad \rho(X+Y) \leq \rho(X) + \rho(Y)$
3. *positively homogeneous:* $X \in \mathcal{G},\ h > 0,\ hX \in \mathcal{G} \quad \Longrightarrow \quad \rho(hX) = h\rho(X)$
4. *translation invariant:* $X \in \mathcal{G},\ a \in \mathbb{R} \quad \Longrightarrow \quad \rho(X+a) = \rho(X) - a$

Monotonicity implies that the risk measure should give a warning when the financial asset has a sure loss (if $Y = 0$ then $\rho(X) > 0$). Sub-additivity property ensures a diversification effect. Positive homogeneity makes sense because of liquidity concerns — when all positions are increased by a multiple, risk is increased by the same multiple because it is getting more difficult to liquidate larger positions. Translation invariance implies that the risk-free asset should reduce the amount of risk by exactly the worth of the risk-free asset.

These axioms are desirable for the purpose of risk management. It is possible to find a representation of coherent risk measures in terms of a functional defined on classes of probability measures, for details see Artzner et al. (1998).

2.2.2 Convex risk measures

If we relax the positive homogeneity assumption, we obtain the more general class of *convex risk measures*.

Definition 15. *A functional $\rho(\cdot)$ on the space of real-valued random variables is called a convex risk measure if it is:*

1. *monotonous:* $X,\ Y \in \mathcal{G};\ Y \geq X \quad \Longrightarrow \quad \rho(Y) \leq \rho(X)$
2. *convex:* $X,\ Y, \in \mathcal{G},\ \lambda \in [0,1] \quad \Longrightarrow \quad \rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$

3. translation invariant: $X \in \mathcal{G},\ a \in \mathbb{R} \quad \Longrightarrow \quad \rho(X + a) = \rho(X) - a$

Representation theorems for convex risk measures can be found in Föllmer and Schied (2002*a*), and for a more extensive treatment, see Föllmer and Schied (2002*b*).

2.3 Particular representatives

In this section we give particular examples of risk measures and briefly comment on their advantages and disadvantages.

2.3.1 Value-at-Risk

Value-at-risk, or VaR, has been widely accepted as a risk measure in the last decade and has been incorporated into industrial regulations (see Khindanova and Rachev (2000) for an overview). The main reason is because it is conceptually easy. It is defined as the minimum level of loss at a confidence level of solvency of $1 - \alpha$. That is VaR can be interpreted as the minimum amount of capital needed as reserve in order to prevent insolvency which happens with probability α.

Definition 16. *The VaR at confidence level* $1 - \alpha$ *is defined as the negative of the lower* α*-quantile of the gain/loss distribution, where* $\alpha \in (0, 1)$*,*

$$VaR_\alpha(X) = -\inf_x \{x | P(X \leq x) \geq \alpha\}$$

While VaR is intuitive, generally it fails to satisfy the sub-additivity property of coherent risk measures which makes it a bad choice for portfolio optimization problems. Yet another pitfall of VaR is that it only provides a minimum bound for losses and ignores any huge potential loss beyond that level.

2.3.2 Expected tail loss

While VaR has gained a lot of attention during the late nineties and early this century, that fact that it is not a coherent risk measure casts doubt on any application of VaR. As a coherent alternative to VaR, conditional value-at-risk (CVaR), expected shortfall or expected tail loss (ETL) was introduced. Similar concepts were also introduced: mean excess loss, mean shortfall, worse conditional expectation, tail conditional expectation or tail VaR. The definition varies across different authors. Acerbi and Tasche (2002) clarify the ambiguity in the definitions of VaR and CVaR and show the equivalence of the CVaR and the expected shortfall. At the same time independently, Rockafellar and Uryasev (2002) also show the equivalence and generalize

their finding to general loss distributions, which incorporate discontinuity. For a discussion and comparison between ETL and VaR, see Yamai and Yoshiba (2002*b*) and Yamai and Yoshiba (2002*a*).

Definition 17. *The expected tail loss at level α is defined as*

$$ETL_\alpha(X) = E(-X| - X > VaR_\alpha(X))$$

where $VaR_\alpha(X)$ is the value-at-risk measure.

The ETL can be interpreted as the average loss beyond VaR. For a proof that ETL is coherent, see for example Zhang and Rachev (2004).

2.4 Closed-form expressions of ETL

If we assume a particular distribution of X in Definition 17, then it could be possible to arrive at particular expressions[5] for $ETL_\epsilon(X)$. In this section first we give some simple examples and then we proceed with the case $X \in S_\alpha(\sigma, \beta, \mu)$ which is much more involved.

Example 12. ***(The Normal distribution)*** *Let us assume that $X \in N(\mu, \sigma^2)$. Then $(X - \mu)/\sigma \in N(0, 1)$ and since ETL is coherent,*

$$ETL_\epsilon(X) = \sigma ETL_\epsilon\left(\frac{X-\mu}{\sigma}\right) + \mu \tag{2.1}$$

The expected tail loss of the standardized random variable is easier to compute.

$$\begin{aligned} ETL_\epsilon\left(\frac{X-\mu}{\sigma}\right) &=: ETL_\epsilon(Y) \\ &= -\frac{1}{\epsilon}\int_{-\infty}^{-VaR_\epsilon(Y)} \frac{x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{(VaR_\epsilon(Y))^2}{2}\right) \end{aligned}$$

Therefore

$$ETL_\epsilon(X) = \frac{\sigma}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{(VaR_\epsilon(Y))^2}{2}\right) + \mu$$

where $Y = (X - \mu)/\sigma$.

[5]Throughout this section, there is a minor change in notation. To avoid ambiguity, the ETL level is denoted by ϵ, while the letter α is reserved for the tail index of stable distributions.

Example 13. ***(The t-distribution)*** *Let us assume that X has t-distribution with ν degrees of freedom, $X \in t(\nu)$. Then*

$$\begin{aligned} ETL_\epsilon(X) &= -\frac{1}{\epsilon}\int_{-\infty}^{-VaR_\epsilon(X)} x\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx \\ &= -\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\frac{\nu}{2\epsilon}\int_{-\infty}^{-VaR_\epsilon(X)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}d\left(1+\frac{x^2}{\nu}\right) \\ &= \begin{cases} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{\sqrt{\nu}}{(\nu-1)\epsilon\sqrt{\pi}}\left(1+\frac{(VaR_\epsilon(X))^2}{\nu}\right)^{\frac{1-\nu}{2}} & , \quad \nu > 1 \\ \infty & , \quad \nu = 1 \end{cases} \end{aligned}$$

If $\nu = 1$, then X follows a Cauchy distribution which has unbounded mathematical expectation and the infinite ETL in this case is not surprising.

As the above example indicates, the ETL is not always finite. The Cauchy distribution is a special case of the class of α-stable distributions and we can expect that for some members of the class the ETL will be finite, for others it will diverge. The condition for the absolute convergence of the corresponding integral is $E(\max(-X, 0)) < \infty$. Therefore from Property 7 we can see that for $\alpha > 1$, the ETL is finite and is unbounded otherwise.

Example 14. ***(The stable distribution)*** *The expression will be stated for the standardized case, i.e. $X \in S_\alpha(1, \beta, 0)$. As a matter of fact, using the properties of translation invariance and positive homogeneity of the ETL, given that $\sigma > 0$ and $\mu \in \mathbb{R}$ are a scale and a location parameter respectively, it follows*

$$ETL_\epsilon(\sigma X + \mu) = \sigma ETL_\epsilon(X) - \mu$$

where $\sigma X + \mu \in S_\alpha(\sigma, \beta, \mu)$; the strategy is the same as in Example 12.

Let $X \in S_\alpha(1, \beta, 0)$ with $\alpha > 1$. If $VaR_\epsilon(X) \neq 0$, then the conditional VaR of X at significance level ϵ admits the following integral representation

$$ETL_\epsilon(X) = \frac{\alpha}{1-\alpha}\frac{|VaR_\epsilon(X)|}{\pi\epsilon}\int_{-\overline{\theta}_0}^{\pi/2} g(\theta)\exp\left(-|VaR_\epsilon(X)|^{\frac{\alpha}{\alpha-1}}v(\theta)\right)d\theta \tag{2.2}$$

where

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

$$v(\theta) = \left(\cos\alpha\overline{\theta}_0\right)^{\frac{1}{\alpha-1}}\left(\frac{\cos\theta}{\sin\alpha(\overline{\theta}_0+\theta)}\right)^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha\overline{\theta}_0+(\alpha-1)\theta)}{\cos\theta},$$

$\overline{\theta}_0 = \frac{1}{\alpha}\arctan\left(\overline{\beta}\tan\frac{\pi\alpha}{2}\right)$ *and* $\overline{\beta} = -sign(VaR_\epsilon(X))\beta$.

Furthermore, if $VaR_\epsilon(X) = 0$, *then*

$$ETL_\epsilon(X) = \frac{2\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{(\pi - 2\theta_0)}\frac{\cos\theta_0}{(\cos\alpha\theta_0)^{1/\alpha}}. \tag{2.3}$$

Example 15. ***(The skewed Student's t distribution)***

It is possible to compute the VaR and ETL of the skewed Student's t distribution defined in (1.8) *even though the formulae are involved and the complexity is similar to the case of stable distributions. We first provide the expressions for VaR and then proceed with the ETL.*

The VaR formula for skewed-T random variable, that is, the value of $x_0 = VaR_\alpha$ *is coming as the unique zero of the following equation (provided* $\gamma > 0$*)*

$$g(x_0) = -\alpha + \frac{2C\sqrt{\pi}}{\gamma}\int_0^\infty t^{-(\nu+2)/2}e^{-\frac{\nu\gamma^2}{4t}}\Phi\left(\frac{\gamma x_0}{\sqrt{2t}} - \sqrt{2t}\right)dt = 0.$$

For negative skewness the value of $x_0 = VaR_\alpha$ *is coming as the unique zero of the next equation (i.e., provided* $\gamma < 0$*)*

$$g(x_0) = 1 - \alpha + \frac{2C\sqrt{\pi}}{\gamma}\int_0^\infty t^{-(\nu+2)/2}e^{-\frac{\nu\gamma^2}{4t}}\Phi\left(\frac{\gamma x_0}{\sqrt{2t}} - \sqrt{2t}\right)dt = 0,$$

In both cases the zero, x_0 *of* g, *is sought on the interval* $(-\infty, 0]$ *provided* $\alpha < \alpha_0$ *or, on the interval* $[0, +\infty)$ *provided* $\alpha > \alpha_0$.

In order to arrive at an expression for the ETL, we consider four different cases depending on the skewness parameter and the VaR level:

Case 1: $x_0 \geq 0, \gamma > 0$,

$$ETL_1 = E[X|X > x_0] = \frac{1}{P(X > x_0)}\int_{x_0}^\infty xf(x)dx$$

Case 2: $x_0 \leq 0, \gamma > 0$,

$$ETL_2 = E[X|X < x_0] = \frac{1}{P(X < x_0)}\int_{-\infty}^{x_0} xf(x)dx$$

Case 3: $x_0 \geq 0, \gamma < 0$,

$$ETL_3 = E[X|X > x_0] = \frac{1}{P(X > x_0)}\int_{x_0}^\infty xf(x)dx$$

Case 4: $x_0 \leq 0, \gamma < 0$,

$$ETL_4 = E[X|X < x_0] = \frac{1}{P(X < x_0)}\int_{-\infty}^{x_0} xf(x)dx$$

It is possible to compute the ETL in these four cases:

$$
\begin{aligned}
ETL_1 &= \frac{\gamma\nu}{(1-\alpha)(\nu-2)} + \frac{4C}{(1-\alpha)\gamma^2}KI \\
ETL_2 &= \frac{-4C}{\alpha\gamma^2}KI \\
ETL_3 &= \frac{4C}{(1-\alpha)\gamma^2}KI \\
ETL_4 &= \frac{\gamma\nu}{\alpha(\nu-2)} + \frac{-4C}{\alpha\gamma^2}KI
\end{aligned}
$$

where

$$
KI = \left[K_{\lambda-1}(\beta) \left(\frac{2}{\beta}\right)^{\lambda-1} e^{\gamma x_0} - \sqrt{\pi} \int_0^\infty t^{-\lambda+1/2} e^{-\frac{\nu\gamma^2}{4t}} \Phi(h_0) dt \right]
$$

in which

$$
\beta = \sqrt{\nu\gamma^2 + (\gamma x_0)^2} \quad \textit{and} \quad h_0 = \frac{\gamma x_0}{\sqrt{2t}} - \sqrt{2t}.
$$

For poofs and additional details, see Dokov et al. (2008).

2.5 ETL estimation from a sample

Suppose that we have a sample of observed portfolio returns and we are not aware of their distribution. Provided that we do not impose any distributional model, the ETL of portfolio return can be estimated from the sample of observed portfolio returns. Denote the observed portfolio returns by $r_1, r_2, \ldots, r_n$ at time instants $t_1, t_2, \ldots, t_n$. The numbers in the sample are given in order of observation. Denote the sorted sample by $r_{(1)} \leq r_{(2)} \leq, \ldots, \leq r_{(n)}$. Thus, $r_{(1)}$ equals the smallest observed portfolio return and $r_{(n)}$ is the largest. The ETL of portfolio returns at tail probability ϵ is estimated according to the formula[6]

$$
\widehat{ETL}_\epsilon(r) = -\frac{1}{\epsilon}\left(\frac{1}{n} \sum_{k=1}^{\lceil n\epsilon \rceil - 1} r_{(k)} + \left(\epsilon - \frac{\lceil n\epsilon \rceil - 1}{n} \right) r_{(\lceil n\epsilon \rceil)} \right) \tag{2.4}
$$

where the notation $\lceil x \rceil$ stands for the smallest integer larger than x.[7] The "hat" above ETL denotes that the number calculated by equation (2.4) is

[6] This formula is a simple consequence of the definition of ETL for discrete distributions, see the appendix to this chapter. A detailed derivation is provided by Rockafellar and Uryasev (2002).

[7] For example, $\lceil 3.1 \rceil = \lceil 3.8 \rceil = 4$.

an estimate of the true value because it is based on a sample. This is a standard notation in statistics.

We demonstrate how equation (2.4) is applied in the following example. Suppose that the sorted sample of portfolio returns is -1.37%, -0.98%, -0.38%, -0.26%, 0.19%, 0.31%, 1.91% and our goal is to calculate the portfolio ETL at 30% tail probability. In this case, the sample contains 7 observations and $\lceil n\epsilon \rceil = \lceil 7 \times 0.3 \rceil = 3$. According to equation (2.4), we calculate

$$\begin{aligned}\widehat{ETL}_{0.3}(r) &= -\frac{1}{0.3}\left(\frac{1}{7}(-1.37\% - 0.98\%) + (0.3 - 2/7)(-0.38\%)\right) \\ &= 1.137\%.\end{aligned}$$

Formula (2.4) can be applied not only to a sample of empirical observations. We may want to work with a statistical model for which no closed-form expressions for ETL are known. Then we can simply sample from the distribution and apply formula (2.4) to the generated simulations.

Besides formula (2.4), there is another method for calculation of ETL. It is based on the minimization formula

$$ETL_\epsilon(X) = \min_{\theta \in (R)} \left(\theta + \frac{1}{\epsilon} E(-X - \theta)_+\right) \tag{2.5}$$

in which we replace the mathematical expectation by the sample average,

$$\widehat{ETL}_\epsilon(r) = \min_{\theta \in (R)} \left(\theta + \frac{1}{n\epsilon} \sum_{i=1}^{n} \max(-r_i - \theta, 0)\right). \tag{2.6}$$

Even though it is not obvious, equations (2.4) and (2.6) are completely equivalent.

The minimization formula in equation (2.6) is appealing because it can be calculated through the methods of linear programming. It can be restated as a linear optimization problem by introducing auxiliary variables $d_1, \ldots, d_n$, one for each observation in the sample,

$$\begin{aligned} \min_{\theta, d} \quad & \theta + \frac{1}{n\epsilon} \sum_{k=1}^{n} d_k \\ \text{subject to} \quad & -r_k - \theta \leq d_k,\ k = 1, n \\ & d_k \geq 0,\ k = 1, n \\ & \theta \in \mathbb{R}. \end{aligned} \tag{2.7}$$

The linear problem (2.7) is obtained from (2.6) through standard methods in mathematical programming. We briefly demonstrate the equivalence between them. Let us fix the value of θ to θ^*. Then the following choice of the auxiliary variables yields the minimum in (2.7). If $-r_k - \theta^* < 0$, then $d_k = 0$. Conversely, if it turns out that $-r_k - \theta^* \geq 0$, then $-r_k - \theta^* = d_k$.

In this way, the sum in the objective function becomes equal to the sum of maxima in equation (2.6).

Applying (2.7) to the sample in the example above, we obtain the optimization problem,

$$\begin{aligned} \min_{\theta,d} \quad & \theta + \frac{1}{7 \times 0.3} \sum_{k=1}^{7} d_k \\ \text{subject to} \quad & 0.98\% - \theta \leq d_1 \\ & -0.31\% - \theta \leq d_2 \\ & -1.91\% - \theta \leq d_3 \\ & 1.37\% - \theta \leq d_4 \\ & 0.38\% - \theta \leq d_5 \\ & 0.26\% - \theta \leq d_6 \\ & -0.19\% - \theta \leq d_7 \\ & d_k \geq 0,\ k = 1, 7 \\ & \theta \in \mathbb{R}. \end{aligned}$$

The solution to this optimization problem is the number 1.137% which is attained for $\theta = 0.38\%$. In fact, this value of θ coincides with the VaR at 30% tail probability and this is not by chance but a feature of the problem which is demonstrated in the appendix to this chapter. We verify that the solution of the problem is indeed the number 1.137% by calculating the objective in equation (2.6) for $\theta = 0.38\%$,

$$ETL_\epsilon(r) = 0.38\% + \frac{0.98\% - 0.38\% + 1.37\% - 0.38\%}{7 \times 0.3} = 1.137\%.$$

Thus, we obtain the number calculated through equation (2.4).

2.6 Computing portfolio ETL in practice

We assume that there are n common stocks with random returns described by the random variables $X_1, \ldots, X_n$. Thus, the portfolio return is represented by

$$r_p = w_1 X_1 + \ldots + w_n X_n$$

where $w_1, \ldots, w_n$ are the weights of the common stocks in the portfolio.

2.6.1 The multivariate normal assumption

If the stock returns are assumed to have a multivariate normal distribution, then the portfolio return has a normal distribution with variance $w'\Sigma w$, where w is the vector of weights and Σ is the covariance matrix between stock returns. The mean of the normal distribution is

$$Er_p = \sum_{k=1}^{n} w_k EX_k$$

where E stands for the mathematical expectation. Thus, under this assumption the ETL of portfolio return at tail probability ϵ can be expressed in closed-form through the closed-form expression of the normal distribution,

$$\begin{aligned} ETL_\epsilon(r_p) &= \frac{\sqrt{w'\Sigma w}}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{(VaR_\epsilon(Y))^2}{2}\right) - Er_p \\ &= C_\epsilon\sqrt{w'\Sigma w} - Er_p \end{aligned} \tag{2.8}$$

where C_ϵ is a constant independent of the portfolio composition and can be calculated in advance. In effect, due to the limitations of the multivariate normal assumption, the portfolio ETL appears symmetric and is representable as the difference between the properly scaled standard deviation of the random portfolio return and portfolio expected return.

2.6.2 The Historical Method

Generally, the historical method is not related to any distributional assumptions. We use the historically observed portfolio returns as a model for the future returns and apply formula (2.4) or (2.6).

While the historical method seems to be more general as it is free of any distributional hypotheses, it has a number of major drawbacks.

1. It assumes that the past trends will continue in the future. This is not a realistic assumption because we may experience extreme events in the future, for instance, which have not happened in the past.

2. It treats the observations as independent and identically distributed (i.i.d.) which is not realistic. The daily returns data exhibits clustering of the volatility phenomenon, autocorrelations and so on, which are sometimes a significant deviation from the i.i.d. assumption.

3. It is not reliable for estimation of ETL at very high confidence levels. A sample of one year of daily data contains 250 observations which is a rather small sample for the purpose of the 99% ETL estimation.

We emphasize that it is very inaccurate for low tail probabilities, e.g. 1% or 5%. Even with one year of daily returns which amounts to 250 observations, in order to estimate the ETL at 1% probability, we have to use the 3 smallest observations which is quite insufficient. What makes the estimation problem even worse is that these observations are in the tail

of the distribution; that is, they are the *smallest* ones in the sample. The implication is that when the sample changes, the estimated ETL may change a lot because the smallest observations tend to fluctuate a lot.

2.6.3 The Hybrid Method

The *hybrid method* is a modification of the historical method in which the observations are not regarded as i.i.d. but certain weights are assigned to them depending on how close they are to the present. The weights are determined using the *exponential smoothing* algorithm. The exponential smoothing accentuates the most recent observations and seeks to take into account time-varying volatility phenomenon.

The algorithm of the hybrid approach consists of the following steps.

1. Exponentially declining weights are attached to historical returns, starting from the current time and going back in time. Let $r_{t-k+1}, \ldots, r_{t-1}, r_t$ be a sequence of k observed returns on a given asset, where t is the current time. The i-th observation is assigned a weight

$$\theta_i = c^* \lambda^{t-i},$$

   where $0 < \lambda < 1$, and $c = \frac{1-\lambda}{1-\lambda^k}$ is a constant chosen such that the sum of all weights is equal to one, $\sum \theta_i = 1$.

2. Similarly to the historical simulation method, the hypothetical future returns are obtained from the past returns and sorted in increasing order.

3. The VaR measure is computed from the empirical c.d.f. in which each observation has probability equal to the weight θ_i.

Generally, the hybrid approach is appropriate for VaR and ETL estimation of heavy-tailed time series. It overcomes, to some degree, the first and the second deficiency of the historical method but it is also not reliable for VaR estimation of very high confidence levels.

According to the hybrid method, different weights are assigned to the observations by which the more recent observations get a higher weight. The rationale is that the observations far back in the past have less impact on the portfolio risk at the present time.

The hybrid method can be adapted for ETL estimation. The weights assigned to the observations are interpreted as probabilities and, thus, the portfolio ETL can be estimated from the resulting discrete distribution according to the formula

$$\widehat{ETL}_{\epsilon}(r) = -\frac{1}{\epsilon}\left(\sum_{j=1}^{k_\epsilon} p_j r_{(j)} + \left(\epsilon - \sum_{j=1}^{k_\epsilon} p_j\right) r_{(k_\epsilon+1)}\right) \tag{2.9}$$

where $r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(k_m)}$ denotes the sorted sample of portfolio returns or payoffs and $p_1, p_2, \ldots, p_{k_m}$ stand for the probabilities of the sorted observations; that is, p_1 is the probability of $r_{(1)}$. The number k_ϵ in equation (2.9) is an integer satisfying the inequalities,

$$\sum_{j=1}^{k_\epsilon} p_j \leq \epsilon < \sum_{j=1}^{k_\epsilon+1} p_j.$$

Equation (2.9) follows directly from the definition of ETL[8] under the assumption that the underlying distribution is discrete without the additional simplification that the outcomes are equally probable.

2.6.4 The Monte Carlo Method

In practice, computing portfolio ETL is done through the Monte Carlo method. We hypothesize a parametric model for the multivariate distribution of financial returns, we fit the model, and then we generate a large number of scenarios. From the generated scenarios, we compute scenarios for portfolio return. Employing formula (2.4), we calculate portfolio ETL at a specified tail probability ϵ.

We can regard the generated scenarios as a sample from the fitted model and thus the computed ETL in the end appears as an estimate of the true ETL. The larger the sample, the closer the estimated ETL is to the true value. If we regenerate the scenarios, the portfolio ETL number will change and it will fluctuate around the true value. Figure 2.1 illustrates this phenomenon for the standard normal distribution with $\epsilon = 0.01$, for which the true value $ETL_{0.01}(X) = 2.665$. This stochastic variability is an issue inherent in the Monte Carlo method and cannot be avoided. In this context, the Monte Carlo method can be viewed as a numerical method of computing portfolio ETL when the hypothesized multivariate model does not allow portfolio ETL to be computed analytically. In this section, we discuss the asymptotic distribution of the estimator in (2.4) which we can use to determine approximately the variance of (2.4) when the number of scenarios is large.

Before proceeding to a more formal result, let us check what intuition may suggest. If we look at equation (2.4), we notice that the leading term

[8] A formal proof can be found in Rockafellar and Uryasev (2002). The reasoning in Rockafellar and Uryasev (2002) is based on the assumption that the random variable describes losses while in equation (2.9), the random variable describes the portfolio return or payoff.

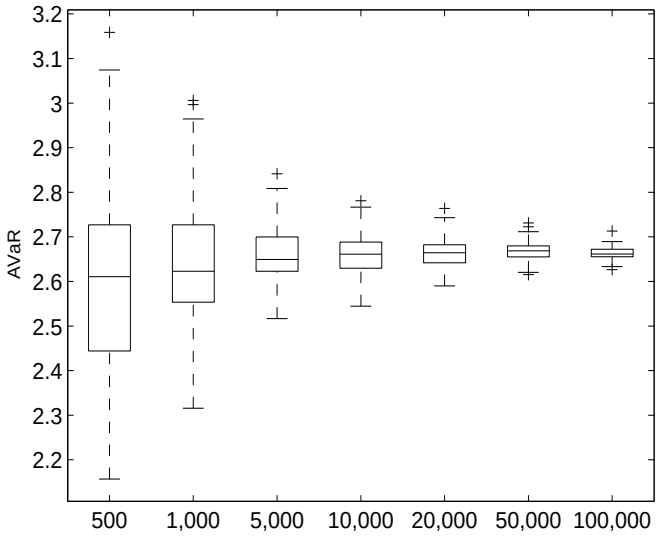

Figure 2.1: Boxplot diagrams of the fluctuation of the ETL at $\epsilon = 1\%$ of the standard normal distribution based on 100 independent samples. The x-axis represents number of scenarios. (Reproduced from Figure 7.4 in Rachev, Stoyanov and Fabozzi (2008))

is the average of the smallest observations in the sample. The fact that we average observations reminds of the central limit theorem (CLT) and the fact that we average the smallest observations in the sample suggests that the variability should be influenced by the behavior of the left tail of the portfolio return distribution. Basically, a result based on the CLT would state that the distribution of the ETL estimator becomes more and more normal as we increase the sample size. Applicability of the CLT however depends on certain conditions such as finite variance which guarantee certain regularity of the random numbers. If this regularity is not present, the smallest numbers in a sample may vary quite a lot as they are not naturally bounded in any respect. Therefore, for heavy-tailed distributions we can expect that the CLT may not hold and the distribution of the estimator in such cases may not be normal at all.

The formal result in Stoyanov and Rachev (2007*b*) confirms these observations. Taking advantage of the generalized CLT, we can demonstrate that

Theorem 15. *Suppose that X is random variable with distribution function $F(x)$ which satisfies the following conditions*

- *$x^{\alpha}F(x) = L(x)$ is slowly varying at infinity, i.e. $\lim_{x\to\infty} L(tx)/L(x) = 1$, $\forall t > 0$.*
- *$\int_{-\infty}^{0} x dF(x) < \infty$*
- *$F(x)$ is differentiable at $x = q_\epsilon$ where q_ϵ is the ϵ-quantile of X.*

Then, there exist c_n $n = 1, 2, ...,$ such that for any $0 < \epsilon < 1$,

$$c_n^{-1}\left(\widehat{ETL}_\epsilon(X) - ETL_\epsilon(X)\right) \xrightarrow{w} S_{\alpha^*}(1, 1, 0) \tag{2.10}$$

in which $\xrightarrow{w}$ denotes weak limit, $1 < \alpha^* = \min(\alpha, 2)$, and $c_n = n^{1/\alpha^*} L_0(n)/\epsilon$ where L_0 is a function slowly varying at infinity and $\widehat{ETL}_\epsilon(X)$ is computed from a sample of independent copies of X according to equation (2.4).

This theorem implies that the limit distribution of the ETL estimator in (2.4) is necessarily a stable distribution totally skewed to the left. In the context of the theorem, we can think of X as a random variable describing portfolio return. If the index α governing the left tail of X is $\alpha \geq 2$, then the above result reduces to the classical CLT as in this case $\alpha^* = 2$ and the limit distribution is normal. This case is considered in detail in Stoyanov and Rachev (2007*a*).

Stable distributed returns

In this section, we consider the case in which portfolio return distribution is a stable law with parameter $1 < \alpha < 2$. Under this assumption, $\alpha^* =$

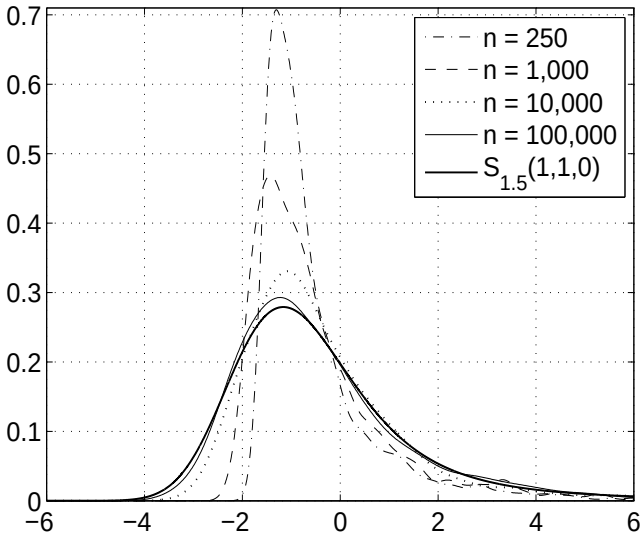

Figure 2.2: The density of the sample ETL as the number of scenarios increases together with the limit stable law, $X \in S_{1.5}(1, 0.7, 0)$. (Reproduced from Figure 2 in Stoyanov and Rachev (2007*b*))

α and thus the limit distribution of the ETL estimator is also a stable law with parameter α. That is, we are not in the case of the classical CLT. We carry out a Monte Carlo study in order to see for how many scenarios the limit distribution Theorem 15 is sufficiently close to the real distribution of the estimator. We choose $X \in S_{1.5}(1, 0.70)$ and we generate 2,000 samples from the corresponding distribution the size of which equals $n = 250;\ 1,000;\ 10,000;$ and $100,000$. Figure 2.2 shows the density of the random variable in the left part of equation (2.10) and how it approaches the limit distribution, i.e. the right part of (2.10), as the number of scenarios increases.

Student's *t* distribution

If we assume that portfolio return has a Student's t distribution, then the degree of freedom parameter ν determines the tail thickness. In the notation of Theorem 15, $\alpha = \nu$ and, therefore, if $\nu \geq 2$, then the asymptotic distribution of the ETL estimator is normal. Even though it is normal, the tail thickness influences the convergence rate which means that the acceptance of the limit distribution as an approximate model decreases as ν decreases. The rationale is that the higher ν is, the more regular the random variable is.

	No tail truncation		With tail truncation	
ν	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.05$
3	70000	17000	12000	4000
4	60000	9000	11500	3600
5	50000	7000	11000	3300
6	23000	4500	11000	3200
7	14000	4200	10500	3100
8	13000	4100	10000	3000
9	12000	4000	10000	3000
10	12000	3900	10000	3000
15	11000	3850	10000	2950
25	10000	3800	10000	2900
50	10000	3750	10000	2900
∞	10000	3300	10000	2900

Table 2.1: The number of observations sufficient to accept the normal distribution as an approximate model when X has Student's t distribution for different values of ν and ϵ (Reproduced from Table 1 and Table 4 in Stoyanov and Rachev (2007b).)

The first two columns of Table 2.6.4 show how the number of scenarios changes when $\epsilon = 0.01$ and $\epsilon = 0.05$ which correspond to the two standard choices of 95% and 99% for the confidence level of the ETL. The numbers in the table are calculated by generating 2,000 samples of a given size which we use to approximate the distribution of the left part of equation (2.10). We use the Kolmogorov distance in order to check if the hypothesis of a normal distribution can be accepted. We notice that the minimum number of scenarios increases from about 10,000 when $\nu = 25$ to 70,000 when $\nu = 3$ for $\epsilon = 0.01$. For additional details, see Stoyanov and Rachev (2007a).

The effect of tail truncation

In the introduction, we noted that one way to deal with the issue that a heavy-tailed model may have an infinite volatility is to apply a tail truncation. This means that we truncate the tails of the distribution very far away from the center, e.g. at 0.1% and 99.9% quantiles. This is not so artificial as it may seem. Stock exchanges usually have regulations according to which trading stops if a given market index drops too much. Essentially, this regulation does not allow panick in the market to result in arbitrarily large losses. If the truncation is done far away from the center of a distribution, the general shape of the distribution is preserved and the descriptive power of the model does not deteriorate.

The implementation of a tail truncation method increases dramatically

the stochastic stability of the ETL estimator in two ways. First, if the portfolio return distribution is such that the limit law of the ETL estimator is stable, after tail truncation the limit law becomes normal. However, the convergence rate to the normal distribution depends on how far away from the center we truncate — the deeper we go into the tails, the slower the convergence rate. Second, if the limit law of the ETL estimator is normal, tail truncation increases the convergence rate. The last two columns of Table 2.6.4 illustrate this observation when portfolio returns are assumed to have Student's t distribution for a truncation threshold equal to the 0.1% quantile. We notice that the minimum number of scenarios drops from 70,000 to 12,000 when $\nu = 3$ and $\epsilon = 0.01$. Stoyanov and Rachev (2007*b*) provide additional examples and details.

For an additional discussion of ETL including geometric interpretations, see Rachev, Stoyanov and Fabozzi (2008) and Stoyanov et al. (2008).

2.7 Risk budgeting with ETL

The concept of ETL allows for scenario-based risk decomposition which is a concept similar to the standard deviation based percentage contribution to risk (PCTR). The practical issue is to identify the contribution of each position to portfolio risk and since ETL is a tail risk measure, percentage contribution to ETL allows one to build a framework for tail risk budgeting. The approach largely depends on one of the coherence axioms given Artzner et al. (1998), which is the positive homogeneity property

$$ETL_\epsilon(aX) = aETL_\epsilon(X), \quad a > 0.$$

Euler's formula is valid for such functions. According to it, the risk measure can be expressed in terms of a weighted average of the partial derivatives with respect to portfolio weights,

$$ETL_\epsilon(w'X) = \sum_i w_i \frac{\partial ETL_\epsilon(w'X)}{\partial w_i}$$

where w is a vector of weights, X is a random vector describing the multivariate return of all financial instruments in the portfolio, and $w'X$ is the portfolio return. The left hand-side of the equation equals total portfolio risk and if we divide both sides by it, we obtain the tail risk decomposition,

$$\begin{aligned} 1 &= \sum_i \frac{w_i}{ETL_\epsilon(w'X)} \frac{\partial ETL_\epsilon(w'X)}{\partial w_i} \\ &= \sum_i p_i. \end{aligned} \tag{2.11}$$

In order to compute the percentage contribution to risk of the i-th position, the i-th summand p_i in (2.11), we have to calculate first the partial derivative. It turns out that the derivative can be expressed as a conditional expectation,

$$\frac{\partial ETL_\epsilon(w'X)}{\partial w_i} = -E(X_i | w'X < -VaR_\epsilon(w'X)).$$

when X is an absolutely continuous random variable, see Zhang and Rachev (2004) and the references therein. The conditional expectation can be computed through the Monte Carlo method.

2.7.1 Identifying risk diversifiers and risk contributors

Exploiting a link between percentage contribution to ETL and the global minimum ETL portfolio, we can arrive at a rule for identifying ETL diversifiers and ETL contributors in long-only portfolios.

Consider the global minimum portfolio ETL problem,

$$\begin{aligned} &\min_w ETL_\epsilon(w'X) \\ &\text{s.t.} \\ &w'e = 1. \end{aligned}$$

where e is a vector of ones and the condition $w'e = 1$ means that the sum of all weights should be equal to 1. Due to the convexity property of ETL, this problem has a unique minimum which can be obtained through the standard first-order optimality conditions for constrained optimization problems. The solution of the optimization problem is the global minimum ETL portfolio which should satisfy the conditions,

$$\nabla ETL_\epsilon(w'X) = \lambda e, \tag{2.12}$$

where $\nabla ETL_\epsilon(w'X)$ is the gradient of ETL computed at the optimal solution and λ is the Lagrange multiplier. The Lagrange multiplier can be explicitly computed,

$$\lambda = w' \nabla ETL_\epsilon(w'X) = ETL_\epsilon(w'X),$$

in which we make use of Euler's formula. Equation (2.12) implies that the partial derivatives of the global minimum ETL portfolio are all equal and the percentage contribution to ETL of the i-th position equals the weight of this position. On the basis of this observation, a simple rule can be designed identifying risk diversifiers and risk contributors for long-only portfolios:

- If $p_i > w_i$, then the i-th position is a risk contributor.

- If $p_i < w_i$, then the i-th position is a risk diversifier.
- If $p_i = w_i$, then the i-th position is neither a contributor nor a diversifier.

The rationale is that if we increase marginally the weights of the diversifiers and decrease marginally the weights of the contributors, making sure that the weights sum up to one, the ETL of the new portfolio is marginally improved. In case $p_i = w_i$ for all i, then we are holding the global minimum ETL portfolio and the risk cannot be marginally reduced.

The analysis presented above is valid for long-only portfolios because negative weights change the sign of the percentage contribution statistics p_i and identifying contributors and diversifiers by comparing to w_i may be problematic. For arbitrary portfolios however, we can use another rule for marginal rebalancing which is based solely on the partial derivatives of ETL,

- Compute the partial derivatives of ETL for a given portfolio w.
- Sort the portfolio positions by the derivatives in a decreasing order.
- The position on top is a risk contributor and the position at the bottom is a risk diversifier.
- In order to improve marginally portfolio risk, decrease the weight of the position on top by a small amount and increase the weight of the position at the bottom making sure all weights sum up to one.

If the partial derivative of the position on top equals the partial derivative of the position at the bottom, then we are holding the global minimum ETL portfolio and no marginal improvement of risk is possible.

2.8 Computing the market risk of financial portfolios

From the discussion so far, we can draw the conclusion that the Monte Carlo method is general enough to be applied in reality and to become the framework of building a risk management system. ETL can be reliably computed using scenarios which can also be used to compute marginal and incremental risk statistics. Therefore, a practical and powerful risk management system can be based upon this concept.

Such a risk management system should have two basic components — a scenario generation engine and a risk calculation engine. FinAnalytica's Cognity risk management system is an industry first on-line portfolio risk management system based on demonstrably realistic asset return distributions and which is built on the basis of the Monte Carlo engine.

2.8.1 Scenario generation engine

The scenario generation engine itself can be organized to contain a variety of models. The most simple workflow is through the discussed generalized multivariate stable and skewed Student's t models. Under this assumption, the marginal distributions are assumed to be either skewed stable or skewed Student's t and the dependence structure is introduced through the covariance matrix of the Gaussian component in the stochastic representation. Thus, the marginal distribution parameters and the covariance matrix can be estimated separately and then scenarios can be produced by means of the stochastic representation.

Alternatively, one can directly work with a copula model such as the skewed Student's t copula. In this case, the marginal distributions can be any combination between stable, skewed Student's t, and normal distributions. Similarly, the marginal distribution parameters can be estimated separately from the copula model parameters and after that scenarios can be produced from the multivariate model.

We can generalize the two cases into one three-step procedure.

Algorithm SG 1

1. Estimate the marginal distribution parameters for the variables which need to be simulated.

2. Estimate the parameters of the dependence model.

3. Produce simulations for the variables from the fitted multivariate model.

This procedure can be directly applied for equity portfolios, for example. The variables we are interested in are only the equity returns. For more complex portfolios however, some variables have to be modeled in a more specific way. For instance, if there are bonds in the portfolio, then we have to model the corresponding yield curve and the scenario generation engine should provide scenarios for it as well. One simple model for the yield curve is through a principal component analysis (PCA) which means that we assume a linear model for the interest rates comprising the yield curve.

A linear model can be required by other considerations as well. For example, we may want to explain equity returns through certain factors and then check the portfolio risk coming from a part or all of these factors. In such a case, we would construct a regression model in which equity returns are the dependent variable and the factor returns are the explanatory variables,

$$r = \alpha + \sum_{k=1}^{n} \beta_k f_k + \epsilon,$$

where r is the return of the dependent variable, f_k is the return of the k-th independent variable and ϵ is the residual. Then we would simulate all dependent variables and the residuals and produce scenarios for them. The scenarios for the equity returns are then produced through the fitted factor model. It is apparent that the general structure of the algorithm of scenarios generation for the dependent variable is invariant of the probabilistic model. Therefore, we can develop an algorithm according to the following steps.

Algorithm SG 2

1. Fit the presumed linear model using the available historical data. Store the regression coefficients (all α and β_k, $k = 1, n$ in the equation above) and the calculated residuals.

2. Estimate the distribution parameters of the factor returns and the residuals.

3. Estimate the parameters of the assumed dependence model between factor returns and residuals.

4. Generate scenarios for the factor returns and the residuals.

5. Compute the scenarios for the dependent variables through the fitted linear model making use of the stored regression coefficients.

Clearly, Algorithm SG 1 is a special case of Algorithm SG 2 when all coefficients in the linear model are equal to zero.

In this discussion, we skip questions related to estimation of linear models and assume that some algorithm is employed for this purpose. Our goal is to demonstrate that the all situations arising in practice can be fitted into the current framework of modeling financial variables with skewed and heavy-tailed distributions with a copula model for the dependence structure.

2.8.2 Risk calculation engine

Having generated scenarios for the relevant variables, it remains to carry out the calculation of portfolio risk. Depending on the portfolio composition, this part of the process may turn out to be the most computationally intensive one. For example, availability of scenarios for equity returns is sufficient for risk analysis of purely equity portfolios. However, if there are derivatives in the portfolio, they need to be evaluated.

In the Monte Carlo framework, the derivatives can be evaluated in each state of the world; that is, for each vector of scenarios. Only then we can compute the positions value and, finally, the portfolio value in each state of the world. The very final step is to compute the ETL and the corresponding marginal and incremental risk statistics using the available scenarios on

position and portfolio level. Therefore, the key steps in the risk calculation engine are the following ones.

Algorithm RC

1. Evaluate the positions values in each state of the world. This step may demand time consuming derivatives pricing routines.
2. Compute the portfolio value in each state of the world.
3. Calculate risk and additional risk statistics on portfolio and position level.

As we noted, sometimes it is important to see the risk on portfolio and position level coming from one factor or group of factors. For example, this factor could be a market index, a currency, or a given yield curve. In such a case, we can assume that all other variables retain their present values in the scenarios and we simulate only the variables we are interested in. Thus, the final scenarios are driven only by the changes of the simulated variables and, therefore, the portfolio risk statistics represent the impact on the portfolio coming from the corresponding variables. Evidently, this does not require reworking Algorithm RC but only the scenarios input.

For additional information, see Rachev, Racheva-Iotova, Stoyanov and Fabozzi (2008) and also Rachev, Martin, Racheva-Iotova and Stoyanov (2008).

Chapter 3

Credit Risk

3.1 Introduction

Cognity CreditRisk System is an integrated risk framework based on heavy-tailed risk models and downside risk measures. The methodology is applicable to all financial instruments worldwide with inherent credit risk and provides a full picture of portfolio credit risks incorporating extreme loss dependencies which can signal over concentration and indicate actions to benefit from diversification in a mark-to-market framework.[1] The system includes all necessary components for active risk management which help determine investment decisions, actions to reduce portfolio risk, consistent risk-based credit limits and rational economic capital. For more information on the recent advances in credit risk modeling, see Cowell and ans S. Trueck (2008).

Cognity CreditRisk System comprises two models for credit risk evaluation for complex portfolios of instruments with inherent credit risk – Asset Value Approach (Model 1) and Stochastic Default Rate Model (Model 2). The choice of the model depends on several factors:

- *The nature of the analyzed portfolio.* Asset Value model (AVM) will work best for portfolios with a great number of market driven instruments and mainly corporate obligors. On the opposite, the accuracy of the Stochastic Default Rate Model (DRM) is not influenced by the types of obligors (corporate, retails, etc.).

- *Goals of the analysis.* Assessing portfolio risk driven by changes in debt value caused by changes in obligor credit quality (including default) is achievable using the first model. The second model considers only default as a credit event.

[1] See D'Sourza et al. (2002) for comparison between stable non-Gaussian and Gaussian approach for credit default swap valuation.

- *Availability of data.* The two models have significant differences in the type of input data requirements.
- *IT Capacity.* The second model requires is more computationally intensive.

3.2 Asset Value Model

There are four key steps in the Monte Carlo approach to credit risk modeling in the Asset-Value model:

Step 1. Modeling the dependence structure between market risk factors and the credit risk drivers.

Step 2. Scenario generation — each scenario corresponds to a possible "state of the world" at the end of our risk horizon. For the purposes of credit risk modeling, the "state of the world" is just the credit rating of each of the obligors in our portfolio and the corresponding values of the market risk factors affecting the portfolio.

Step 3. Portfolio valuation — for each scenario, we evaluate the portfolio to reflect the new credit ratings and the values of the market risk factors. This step offers us a large number of possible future portfolio values.

Step 4. Summarize results — having the scenarios generated in the previous steps, we get an estimate for the distribution of the portfolio value. One may then choose to report any number of descriptive statistics for this distribution.

We provide a schema in Figure 3.1 which summarizes the calculation process and then we explain in detail the key steps outlined above. The general methodology described below is valid for every Monte Carlo approach to credit risk modeling in the Asset-Value model. We describe the improvements we have introduced in the models.

3.2.1 Model dependence structure

The general assumption of the model is that the driver of credit events is the asset value of a company. This means one should be able to model the dependence structure between asset values of the counterparties. As in CreditMetrics, at this point we assume that the dependence structure between asset values of two firms can be approximated by the dependence structure between the stock prices of those firms. This fact offers a very natural solution to the problem: if we are successful in modeling dependence structure between the stock prices and all relevant market risk factors (interest rates, exchange rates, etc.), then we accomplish simultaneously two goals:

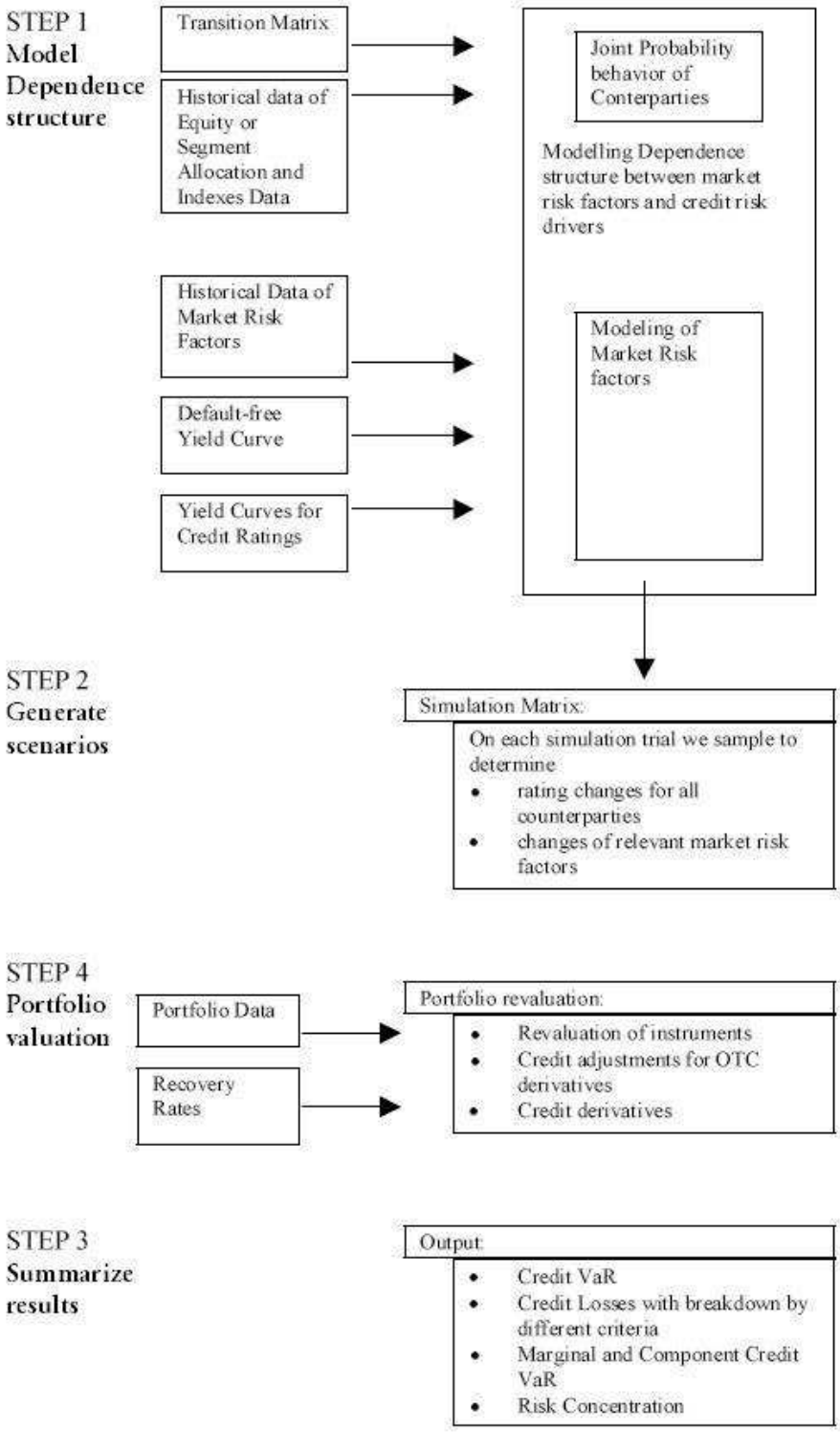

Figure 3.1: The schema summarizes the calculation process by showing the data required for each step and highlighting the key components of the model.

- We construct dependency between credit risk events of our obligors
- We model dependency between market risk factors and the credit risk drivers

If one uses as a measure of dependence the correlation between risk factors (as in CreditMetrics), the above task is trivial — all one needs is to estimate the correlation matrix for the stock prices and the relevant market risk factors.

The correlation is a widespread concept in modern finance and insurance and stands for a measure of dependence between random variates. However, as we noted in Section 1.5.6, this term is very often incorrectly used to mean *any* notion of dependence. Actually correlation is one particular measure of dependence among many. Of course, in the world of multivariate normal distribution and, more generally in the world of spherical and elliptical distributions, it is the accepted measure, see Chapter 1 for more information regarding spherical and elliptical distributions. Yet empirical research shows that real data seldom seems to have been generated from a distribution belonging to this class.

There are at least three major drawbacks of the correlation method. Let us consider the case of two real-valued random variables X and Y.

1. The variances of X and Y must be finite or the correlation is not defined. This assumption causes problems when working with heavy-tailed data. For instance the variances of the components of a bivariate $t(n)$ distributed random vector for $n \leq 2$ are infinite, hence the correlation between them is not defined.

2. Independence of two random variables implies correlation equal to zero, the opposite, generally speaking, is not correct — zero correlation does not imply independence. A simple example is the following: Let $X \sim N(0,1)$ and $Y = X^2$. Since the third moment of the standard normal distribution is zero, the correlation between X and Y is zero despite the fact that Y is a function of X which means that they are dependent. Only in the case of multivariate normal distribution zero correlation and independence are interchangeable notions. This statement is not valid if only the marginal distributions are normal and the joint distribution is non-normal. The example on Figure 3.2 illustrates this fact

3. The correlation is not invariant under non-linear strictly increasing transformations T: $\mathbb{R} \to \mathbb{R}$ which is a serious disadvantage. In general $\text{corr}(T(X), T(Y)) \neq \text{corr}(X, Y)$.

A more prevalent approach is to model dependency using copulas, see Section 1.5.6 for more details on copulas. The use of copulas offers the following advantages:

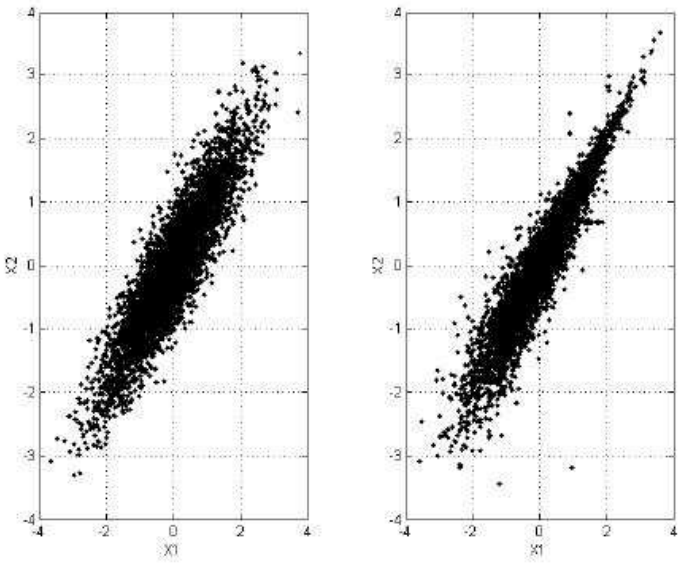

Figure 3.2: 5000 simulated data from two bivariate distributions with normal marginals and identical correlation of 0.9 but different dependence structures.

- The nature of dependency that can be modeled is more general. In comparison, only linear dependence can be explained by the correlation.

- Dependence of extreme events can be modeled.

- Copulas are indifferent to continuously increasing transformations (not only linear as it is true for correlations): If $(X_1, \ldots, X_n)^t$ has copula C and $T_1, \ldots, T_n$ are increasing continuous functions, then the vector $(T_1(X_1), \ldots, T_n(X_n))^t$ also has copula C.

This is extremely important in Asset-Value models for credit risk, because this property postulates that the asset values of two firms shall have exactly the same copula as the stock prices of these two companies. The latter is true if we consider the stock price of a company as a call option on its assets and if the option pricing function giving the stock price is continuously increasing with respect to the asset values. Cognity Credit Risk Module supplies both models for describing dependence structure:

- The simplified approach using correlations like a measure for dependency

- The copula approach. Our model accounts for the joint extreme movements at the same time not incurring significant computational burden. This is done by different treatment of the extremes and the body of the distribution trying to avoid in this way the curse of dimensionality.

As a conclusion to this part of the discussion it is worth saying that in case there is no information about stock prices for a given obligor we employ the idea of segmentation described in CreditMetrics. The essence of this approach is that the user determines the percentage of the allocation of obligor volatility among the volatilities of certain market indexes and explains the dependence between obligors by the dependence of the market indexes that drive obligor volatility.

3.2.2 Scenario generation

In this section, we discuss how to generate scenarios of future credit ratings for the obligors in our portfolio and simultaneously for the changes of the market risk factors. Each set of future credit ratings and values of the market risk factors corresponds to a possible "state of the world" at the end of our risk horizon.

We shall rely heavily on the asset value model. The steps to scenario generation are as follows:

1. Establish asset-return thresholds for the obligors in the portfolio.

2. Generate scenarios of asset returns and the market risk factors using appropriate distribution — this is an assumption to be imposed.

3. Map the asset return scenarios to credit rating scenarios.

If we are using a multivariate normal distribution as a probability model for the log-returns of asset values and market risk factors, generating scenarios is a simple matter of generating correlated, normally distributed variables. There is a well known algorithm, see Section 1.6.2 in Chapter 1.

However, it is a widely accepted critique of the normal distribution that it fails to explain certain properties of financial variables — fat tails and excess kurtosis. The family of Stable distributions that we are proposing contains as a special case the Gaussian (Normal) distribution. However, non-Gaussian Stable models do not possess the limitations of the normal one and all share a similar feature that differentiates them from the Gaussian one – heavy probability tails. In addition they are completely described by four parameters which control, in addition to the variability and mean, the degrees of heavy tails and skewness. Thus they can model greater variety of empirical distributions including skewed ones. For more information on

stable distributions and how multivariate models can be constructed and simulated, see Chapter 2.

When using stable distributions the following sub-steps are required:

1. Estimate the parameters of the stable distribution for each factor (asset values or market risk factor) using historical data. There are several approaches; maximum likelihood estimation provides the most accurate results.

2. This first sub-step models the marginal distributions of our risk drivers.

3. Employ the dependence structure model in Section 1 and construct a multivariate distribution the marginals of which are stable with estimated parameters from step 2.

4. Generate scenarios sampled from the multivariate probability model developed in step 3.

5. Transform the generated scenarios to stable marginals by using the universe of the fitted stable cdf in step 1.

Once we have scenarios for the asset values, we only need to assign credit ratings for each scenario. This is done by comparing the asset value in each scenario to the rating thresholds.

3.2.3 Portfolio valuation

For non-default scenarios, the portfolio valuation step consists in applying a valuation model for each particular position within the portfolio over each scenario. The yield curve corresponding to the credit rating of the obligor for this particular scenario should be used.

For default scenarios, the situation is slightly different. There are two approaches dealing with the recovery rate required for default scenarios:

- Assume constant recovery rates — then the value of a position in case of a default scenario is simply the nominal amount times recovery rate.

- Allow the recovery rate to be a random variable.

As discussed in many empirical analysis recovery rates are not deterministic quantities but rather display a large amount of variation. This variation of value in the case of default is a significant contributor to risk. Recovery rates are modelled using a Beta distribution with a specified mean and standard deviation. In this case, for each default scenario for a given obligor, we should generate a random recovery rate for each particular transaction with defaulted obligor. The value of a given position in case a particular default scenario is realized will be different.

If we have derivatives in the portfolio, it is possible to apply non-standard derivative pricing under the stable assumption. For this purpose we model the volatility in the stochastic process of the underlying as random. The distribution of the random volatility is a maximally skewed stable law called a subordinator and is in line with a more general assumption about the price process of the underlying, called a subordinated process. For more details, refer to Rachev and Mittnik (2000). In the case of non-standard derivative pricing, we have the following steps:

1. Estimate the parameters of the subordinator from the returns series of the underlying.

2. In each state of the world generate a random volatility from the fitted distribution in step 1.

3. In each state of the world estimate the value of the derivative using the available realizations of the random volatility and the value of the underlying.

3.2.4 Summarize results

Having the portfolio value scenarios generated in the previous steps, we obtain an estimate for the distribution of the portfolio values. We may then choose to report any number of descriptive statistics for this distribution. The calculation of statistics is one and the same for both models. For example, mean and standard deviation of future portfolio value can be obtained from the simulated portfolio values using sample statistics. Because of the skewed nature of the portfolio distribution, the mean and standard deviation may not be good measures of risk. Since the distribution of values is not normal, we cannot infer percentile levels from the standard deviation. Given the simulated portfolio values, we can compute better measures, for example empirical quantiles, or mean shortfall.

To this point, we have considered only statistics, which describe the portfolio distribution. We would also like to consider individual assets and to ascertain how much risk each asset contributes to the portfolio. To this end, we can use marginal and incremental risk statistics:

- **Marginal ETL**. Marginal ETL stands for the derivative of ETL with respect to the position weight. The resulting number indicates how much portfolio risk changes provided that we increase by a small amount the weight of the corresponding position. In Section 2.7 we consider the more general idea of risk budgeting with ETL which can be directly applied.

- **Incremental ETL**. Incremental ETL is simply the difference between portfolio ETL and the ETL of the same portfolio but with this position

removed. Thus, if the position is a risk diversifier, the incremental ETL will be a negative number because the ETL of the portfolio with the position removed will be larger. This analysis concerns adding or removing the entire position.

3.3 Default-rate model

There are five key steps in the Monte Carlo Approach to Credit Risk Modeling based on Stochastic Modeling of Default Rate:

Step 1. Build econometric models for default rates and for the explanatory variables. An econometric model is evaluated for the default probability of a segment based on explanatory variables (macro-factors, indexes, etc.) using historical data for default frequencies in a given segment and historical time series for the explanatory variables.

Step 2. Generate scenarios. Each scenario corresponds to a possible "state of the world" at the end of our risk horizon. Here, the "state of the world" is a set of values for the market variables and for the explanatory variable defined in Step 1.

Step 3. Estimate default probabilities under each scenario for each of the segments using the scenario values of the explanatory variables and the model estimated in Step 1. *Simulate sub-scenarios for the status of each obligor* (default/non-default) based on the estimated default probability.

Step 4. Portfolio valuation. For each scenario, revalue the portfolio to reflect the new credit status of the obligor and the values of the market risk factors. This step generates a large number of possible future portfolio values.

Step 5. Summarize results. Having the scenarios generated in the previous steps, we possess an estimate for the distribution of portfolio values. We may then choose to report any descriptive statistics for this distribution.

We provide a schema in Figure 3.3 describing the calculation process and then we continue with a detailed description of the key steps outlined above.

3.3.1 Build the econometric models

This first step is in fact the most challenging and critical task of the model. Two crucial models should be defined and estimated:

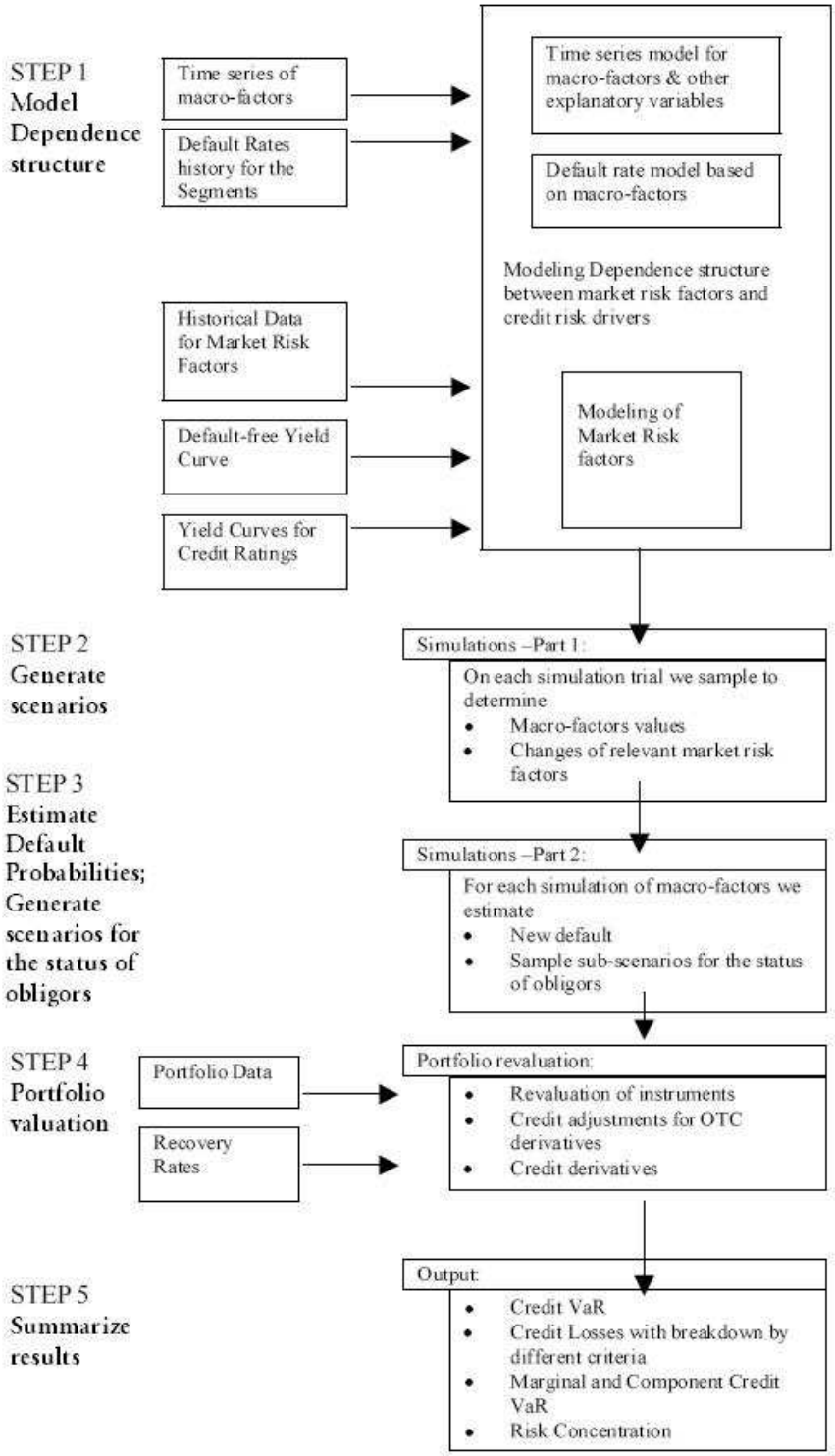

Figure 3.3: The schema summarizes the process showing the data required for each step and highlighting the key components of the model.

- An econometric model for default probability of a segment based on explanatory variables like macro-factors, indexes, etc.
- Time series model for explanatory variables.

Default probability models should be evaluated for each user-defined segment. The segment definitions can be flexible enough based on the following criteria

- credit rating
- industry
- region
- size of the company, provided the time series of default rates are available for each of the segments.

The explanatory variables that might be suitable to represent the systematic risk of the default rates in the chosen country-industry-segments depend on the nature of the portfolio and might be

- industry indices
- macro variables (GDP, unemployment)
- long-term interest rates, exchange rates, etc.

The first task is to define a model for the default probability of a segment based on explanatory variables (macro-factors, indexes, etc.) using historical data for default frequencies in a given segment and historical time series for the explanatory variables. In other words we choose a function f and estimate its parameters such that

$$DF_{s,t} = f(X_{1,t}, \ldots, X_{N,t}) \tag{3.1}$$

where $DF_{s,t}$ is the default frequency in the segment s for the time period t, $X_{i,t}$ is the value of the i-th explanatory variable at time $t, i = 1 \ldots N$

The second model is a time-series model for explanatory variables. The usual way to model dependent variables (as also suggested in CreditPortfolioView) is to employ some kind of ARMA (p, q) model. That is the same as assuming that

$$X_t = a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} b_j e_{t-i} + e_t, \text{ where } e_t \text{ is } N(0, \sigma^2).$$

It is important to note that the proper modeling of the default rate will depend very much on the proper modeling of the dependent variables.

Another crucial component of the model is the residual distribution modeling. We can observe that the real distribution of residuals deviates form the assumption of the model — residuals are not normal. They are

- skewed
- with fatter tails
- with higher peak around the center of the distribution
- and there is a volatility clustering

Thus the improper use of normal residuals will end up with "incorrect" scenarios (simulations) for the possible default rates.

For the modeling of macro-factors, we propose the following more general model – stable vector AR(1)-ARCH – type model. Under this model $X_t = A_1 X_{t-1} + E_t$, where $X_t = (X_1, \ldots, X_n)'$ is the vector of explanatory variables, A_1 is n by n-matrix, $E_t = (e_{1,t}, \ldots, e_{n,t})'$ is the vector of residuals which are modeled by multivariate stable ARCH model, see also Racheva-Iotova et al. (2003). Employing stable residuals results in:

- fatter tails of the residuals
- higher variability of default rates

The ARCH component of the model takes care of volatility clustering. Moreover, since the model is a vector autoregressive model we will eventually succeed in modeling *joint* behavior of macro-factors.

Note 1: Relevant market variables (like interest rates) are also included in the model.

3.3.2 Generate scenarios

Each scenario corresponds to a possible "state of the world" at the end of our risk horizon. Here, the "state of the world" is a set of values for the market variables and for the explanatory variable defined in Step 1.

The scenarios are simulated according to the vector-autoregressive model discussed in Step 1. The important feature of the model is that the macro-factors and market risk factors are modeled by a *joint* probability distribution (See note 1 above). The latter means that the realizations of the market risk factors depend on the realizations of the macro-factors.

3.3.3 Estimate default probabilities

In the third step, the default probabilities are estimated under each scenario for each of the segments using the simulated values of the explanatory variables and the model estimated in Step 1. Then we simulate sub-scenarios for the status of each obligor (default/non-default) based on the estimated default probability. For each scenario set j in Step 2, we estimate the probability of default for each segment based on the model estimated in Step 1

$$P_{s,j} = f_s(X_{1,j}, \ldots, X_{N,j})$$

where $P_{s,j}$ is the default probability for the segment s under j-th simulation.

Now, for each scenario j, we generate independent sub-scenarios for each counterparty state (default or no-default) based on its probability of default $P_{s,j}$.

At this point, we have obtained the full set of scenarios describing the possible "states of the world" in terms of market risk factor values and obligor states (default or non-default).

3.3.4 Portfolio valuation

For each scenario, we revalue the portfolio to reflect the new credit status of the obligor and the values of the market risk factors. The result is a large number of possible future portfolio values. This step is similar to the corresponding step in the AVM described in the previous section. The only difference is that in the current model there is a simplification –it considers only default or non-default status for an obligor.

3.3.5 Summary of the results

At this point, we have created a number of possible future portfolio values. The final task is then to synthesize this information into meaningful risk estimates. We can use any of the descriptive statistics from the previous section.

Chapter 4

Optimal Portfolios

4.1 Introduction

There are two basic approaches to the problem of portfolio selection under uncertainty. One of them is the stochastic dominance approach, based on the axiomatic model of risk-averse preferences. Unfortunately, the optimization problems that arise are not easy to solve in practice. The other is the reward-risk analysis. According to it, the portfolio choice is made with respect to two criteria — the expected portfolio return and portfolio risk. A portfolio is preferred to another one if it has higher expected return and lower risk. There are convenient computational recipes and geometric interpretations of the trade-off between the two criteria. A disadvantage of the latter approach is that it cannot capture the richness of the former. As a matter of fact, the relationship between the two approaches is still a research topic (see Ogryczak and Ruszczynski (2001) and the references therein).

Related to the reward-risk analysis is the reward-risk ratio optimization. Since the publication of the Sharpe ratio, see Sharpe (1966), which is based on the mean-variance analysis, some new performance measures like the STARR ratio, the Minimax measure, Sortino-Satchell ratio, Farinelli-Tibiletti ratio and most recently the R-ratio[1] and the GR-ratio[2] have been proposed (for an empirical comparison, see Biglova et al. (2004), Rachev, Jasic, Biglova and Fabozzi (2005) and the references therein). The new ratios take into account empirically observed phenomena, that assets returns distributions are fat-tailed and skewed, by incorporating proper reward and risk measures.

Another area, where the reward-risk ratio analysis is applicable, is the construction of momentum strategies. In the last decade, it was discovered that asset returns exhibit persistent momentum behavior in intermediate horizons — simple strategies of ranking and selecting stocks as winners and

[1]The R-ratio is an abbreviation for the Rachev-ratio

[2]The GR-ratio is an abbreviat0ion for the Generalized Rachev ratio

losers are able to generate significant profit. It has been ascertained that these effects are present in the US markets, Europe and emerging markets and also in samples from different periods of time. Yet there is not an accepted explanation of the existence of momentum strategies by any conventional or non-conventional theory, see Rachev, Jasic, Biglova and Fabozzi (2005).

In this chapter, first we briefly describe the the Markowitz problem and the related Sharpe ratio optimization. Next we continue with a generalization, introducing the stochastic dominance analysis and the consequences when different multivariate assumptions for the assets return distribution are imposed. In section 4.4 we describe the optimal portfolio problem with ETL as a risk measure, and comment on the numerical difficulties. Finally, we describe how a portfolio optimization module can be integrated into the scenarios based risk management discussed in the previous chapters.[3] For additional information on the framework, see also Racheva-Iotova and Stoyanov (2008).

4.2 The mean-variance analysis and the Sharpe ratio

The classical mean-variance framework introduced by Markowitz in the 1950's (Markowitz (1952)) is the first proposed model of the second type and we shall briefly describe it. Suppose that at time $t_0 = 0$ we have an investor who can choose to invest among a universe of n assets. Having made the decision, he keeps the allocation unchanged until the moment t_1 when he can make another investment decision based on the new information accumulated up to t_1. The vector of assets returns $r = (r_1, r_2, \ldots, r_n)^T$ is stochastic with expected value $Er = (Er_1, Er_2, \ldots, Er_n)^T$. The result of the investment decision is a portfolio with composition $w = (w_1, w_2, \ldots, w_n)^T$ where w_i is the portfolio weight corresponding to the i-th item, i.e. the share of the initial endowment invested in the i-th asset. We require that the weights of all portfolio items sum up to 1, $w^T e = \sum_{i=1}^{n} w_i = 1$ where $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$. The expected portfolio return, expressed in terms of the individual items, equals

$$\mu_p = \sum_{i=1}^{n} w_i Er_i = w^T Er$$

A key point in Markowitz's approach is that the standard deviation of portfolio return σ_p is assumed to be the measure of risk. If we denote

[3] Concerning application of heavy-tailed distributions in portfolio choice theory, see Ortobelli et al. (2005).

$\Sigma = \{cov(r_i, r_j)\}_{i,j=1}^{n}$ to be the covariance matrix of the portfolio items, then

$$\sigma_p^2 = \sum_{i,j} w_i w_j \mathrm{cov}(r_i, r_j) = w^T \Sigma w$$

Sometimes the investor faces certain exogenous constraints. For instance, a certain subset of the assets is not allowed to constitute more than a given fraction of total portfolio value. A portfolio that satisfies all constraints in the selection problem will be called *admissible* or *feasible*. Where appropriate, we shall denote the set of all feasible portfolios by X.

The main principle behind the mean-variance analysis can be summarized briefly in two ways:

1. From all feasible portfolios with a given upper bound on σ_p, find the ones that have maximum expected return μ_p;
2. From all feasible portfolios with a given lower bound on μ_p, find the ones that have minimum risk σ_p;

Behind the two formulations of the principle, we can find two optimization problems:

$$\begin{array}{rl} \max\limits_{w} & w^T Er \\ \text{subject to} & w^T e = 1 \\ & w^T \Sigma w \leq R^* \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.1}$$

and

$$\begin{array}{rl} \min\limits_{w} & w^T \Sigma w \\ \text{subject to} & w^T e = 1 \\ & w^T Er \geq R_* \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.2}$$

where R^* is the upper bound on portfolio risk, R_* is the lower bound on portfolio return, $A \in \mathbb{R}^{k \times n}$ is a matrix, $Lb \in \mathbb{R}^k$ is a vector of lower bounds and $Ub \in \mathbb{R}^k$ is a vector of upper bounds. The set of k double linear inequalities $Lb \leq Aw \leq Ub$ generalizes all exogenous constraints. The solution of Problem (4.1) or Problem (4.2) represents the optimal portfolio or the portfolio that is most preferable among the set of all feasible portfolios. As a matter of fact, the originally proposed problem by Markowitz in his seminal work is Problem (4.2) and it is known as the Markowitz problem[4]. The optimal portfolio w^o found in this way is a function of the imposed bounds

[4]For the general dualty theorems of the type of Problems (4.1) and (4.2), see Rachev and Rüschendorf (1998)

R^* or R_* depending on whether we consider Problem (4.1) or Problem (4.2). Let us choose Problem (4.2) for the sake of being unambiguous. Then as we have explained $w^o = w^o(R_*)$. Changing the parameter R_* we obtain the set of all optimal portfolios, or the *mean-variance efficient set*. We denote it by $\mathcal{E}_o$. The curve $(w^{oT}Er, w^{oT}\Sigma w^o)$ where $w^o \in \mathcal{E}_o$ is called the *efficient frontier*.

We need to remark that there is a third way to arrive at the mean-variance efficient set. It is by considering the optimization problem:

$$\begin{array}{rl} \max\limits_{w} & w^T Er - \lambda w^T \Sigma w \\ \text{subject to} & w^T e = 1 \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.3}$$

where $\lambda > 0$ is a parameter. In this representation, the objective function $w^T Er - \lambda w^T \Sigma w$ is interpreted as a utility function and λ is called *the risk-aversion* parameter. Since it is possible to show that the three problems are equivalent (see, for example, Rockafellar and Uryasev (2002) and Palmquist et al. (2002)), the mean-variance efficient set can be obtained by the problem above via varying the risk-aversion parameter.

Suppose that we have received the portfolios from the mean-variance efficient set and that we can compare and choose among all of them. Are we indifferent towards all these portfolios? We can compare them in terms of their expected return for a unit of risk, that is we can compare the ratios

$$SR(w^o) = \frac{w^{oT}Er}{\sqrt{w^{oT}\Sigma w^o}} \tag{4.4}$$

for all portfolios $w^o \in \mathcal{E}_o$. We would prefer the portfolio with the highest ratio as it provides the highest expected return for a unit of risk. That is we solve the problem

$$\max_{w \in \mathcal{E}_o} \quad \frac{w^T Er}{\sqrt{w^T \Sigma w}} \tag{4.5}$$

Geometrically, the point on the Efficient frontier that corresponds to the solution of Problem (4.5) is where a straight line passing through the origin is tangent to the Efficient frontier (see Figure 4.1). The optimal portfolio received is called the *tangent portfolio*[5]. The ratio defined in equation (4.4) is a version of the reward-to-variability ratio called the Sharpe ratio, hence the notation. It was first introduced to measure the performance of mutual funds and was originally proposed as the ratio between the expected excess return (the expected return of the fund above a benchmark portfolio return) and the standard deviation of the returns of the fund, see Sharpe (1966) and Sharpe (1994).

[5]The optimal portfolio is known as the Markowitz market portfolio, or can also be called *tangent portfolio with zero risk-free rate*.

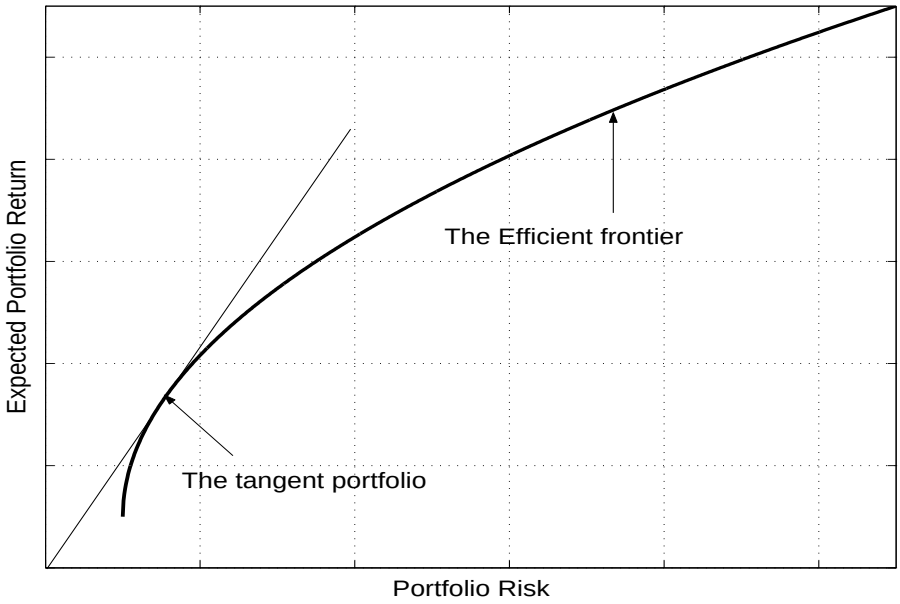

Figure 4.1: The efficient frontier and the tangent portfolio

4.3 Generalization of the mean-variance analysis

An accepted theory of asset choice under uncertainty is the stochastic dominance approach, which is connected to the expected utility theory. Expected utility representation of agent's preferences is present if there exists a function $u(\cdot)$ such that a random consumption X is preferred to a random consumption Y if and only if $E(u(X)) \geq E(u(Y))$, where $E(\cdot)$ is the expectation under the agent's probability belief. The formal approach was first introduced by Neumann and Morgenstern (1953).

Usually two assumptions are made for a utility function:

1. $u(\cdot)$ is a non-decreasing and a non-constant function, meaning that a rational decision maker would prefer more than less. Such decision makers are called *non-satiable*. Let us denote the set of those utility functions by $\mathcal{U}_1$

2. $u(\cdot)$ is assumed concave, i.e. the decision maker is *risk averse*. Let us denote by $\mathcal{U}_2$ the set of all $u(\cdot) \in \mathcal{U}_1$ that are concave.

In the stochastic dominance approach, random variables are compared by pointwise comparison of some functions constructed from their distribution function, see for example Ortobelli and Rachev (2001) and the numerous references therein. In the context of portfolio selection, the random variables are the random portfolio returns. Let w and v are two feasible

portfolio compositions. Then the corresponding portfolio returns are $w^T r$ and $v^T r$, where r is the random vector of portfolio items returns. The preference relation is defined on the random portfolio returns and it is naturally translated into the space of portfolio compositions. We say that portfolio w dominates portfolio v in the sense of the first degree stochastic dominance (FSD), $w^T r \succeq_{FSD} v^T r$, if all investors having utility functions in $\mathcal{U}_1$ prefer $w^T r$ to $v^T r$. The following equivalence holds:

$$w^T r \succeq_{FSD} v^T r \quad \Longleftrightarrow \quad F_{w^T r}(t) \leq F_{v^T r}(t), \quad \forall t \in \mathbb{R} \tag{4.6}$$

where $F_X(t)$ is the cumulative distribution function of the random variable X. FSD is a general relation, according to which $w^T r \succeq_{FSD} v^T r$ within all non-satiable investors, no matter how risk-averse or risk-seeking they are.

For decision making under risk, another concept is more important. Portfolio w dominates portfolio v in the sense of the second degree stochastic dominance (SSD), $w^T r \succeq_{SSD} v^T r$, if all investors having utility functions in $\mathcal{U}_2$ prefer $w^T r$ to $v^T r$. We have the next equivalence relation:

$$w^T r \succeq_{SSD} v^T r \quad \Longleftrightarrow \quad \int_{-\infty}^{t} F_{w^T r}(u)du \leq \int_{-\infty}^{t} F_{v^T r}(u)du, \quad \forall t \in \mathbb{R} \tag{4.7}$$

If $w^T r \succeq_{SSD} v^T r$, then $w^T r$ is preferred among all non-satiable investors, who are risk-averse.

An additional relation is constructed by Rothschild and Stiglitz (1970). Portfolio w dominates portfolio v in the sense of Rothschild-Stiglitz dominance (RSD), $w^T r \succeq_{RSD} v^T r$, if all investors with concave utility functions prefer $w^T r$ to $v^T r$. Again, we have an equivalence relation in terms of some functions of the corresponding cumulative distribution functions:

$$w^T r \succeq_{RSD} v^T r \quad \Longleftrightarrow \quad \begin{cases} w^T Er = v^T Er, \\ \displaystyle\int_{-\infty}^{t} F_{w^T r}(u)du \leq \int_{-\infty}^{t} F_{v^T r}(u)du, \quad \forall t \in \mathbb{R} \end{cases} \tag{4.8}$$

Further on, it is possible to show that the next implications hold: $FSD \Longrightarrow SSD$ and $RSD \Longrightarrow SSD$. The relationship between stochastic dominance and the mean-risk models, see Ogryczak and Ruszczynski (2001) and DeGiorgi (2005).

4.3.1 The portfolio choice problem

Assume that the market is frictionless, there are no arbitrage opportunities and all investors are price takers. An investor with a utility function $u(\cdot)$ would like to know which feasible portfolio composition w maximizes the expected value of the utility function $u(\cdot)$. In effect, the investor would

be interested in the following portfolio composition sets, which are called *efficient sets*, for more details see Rachev and Mittnik (2000):

$$W_i = \left\{ w \in X \;\middle|\; \begin{array}{l} w \in \arg\sup\limits_{x \in X} E(u(x^T r)), \quad u(\cdot) \in \mathcal{U}_i, \quad \text{and} \\ \nexists y \in \arg\sup\limits_{x \in X} E(u(x^T r)) : E(v(y^T r)) \geq E(v(w^T r)), \\ \text{for all} \quad v(\cdot) \in \mathcal{U}_i \quad \text{and strictly for some} \quad v(\cdot), \quad i = 1, 2 \end{array} \right\}$$

where X is the set of all admissible portfolios. W_1 represents the set of all feasible portfolio compositions which are not dominated according to FSD, i.e. which maximize the expected utility of non-satiable investors. W_2 represents all feasible portfolios which are not dominated according to SSD — portfolios that maximize the expected utility of non-satiable, risk-averse investors. In a similar way, we define the efficient set for the risk-averse investors:

$$\widetilde{W} = \left\{ w \in X \;\middle|\; \begin{array}{l} w \in \arg\sup\limits_{x \in X} E(u(x^T r)), \quad u(\cdot) \quad \text{is concave, and} \\ \nexists y \in \arg\sup\limits_{x \in X} E(u(x^T r)) : E(v(y^T r)) \geq E(v(w^T r)), \\ \text{for all concave} \quad v(\cdot) \quad \text{and strictly for some} \quad v(\cdot) \end{array} \right\}$$

$\widetilde{W}$ represents all feasible portfolio compositions that are not dominated according to RSD.

In order to obtain the above definitions, we implicitly fix the investor with utility function $u(\cdot)$. Also we can consider the next group of efficient sets:

$$T_i = \left\{ w \in X \;\middle|\; \begin{array}{l} \nexists y \in X : \quad E(v(y^T r)) \geq E(v(w^T r)), \\ \text{for all} \quad v(\cdot) \in \mathcal{U}_i \quad \text{and strictly for some} \quad v(\cdot), \quad i = 1, 2 \end{array} \right\}$$

and

$$\widetilde{T} = \left\{ w \in X \;\middle|\; \begin{array}{l} \nexists y \in X : E(v(y^T r)) \geq E(v(w^T r)), \\ \text{for all concave} \quad v(\cdot) \quad \text{and strictly for some} \quad v(\cdot) \end{array} \right\}$$

As a consequence of these definitions and the relations among various types of stochastic dominance rules, we have that $T_i \subseteq W_i$, $i = 1, 2$, $\widetilde{T} \subseteq \widetilde{W}$, $\widetilde{T} \subseteq T_2$ and $\widetilde{W} \subseteq W_2$.

Clearly, the optimization problems that generate the efficient sets T_i, $i = 1, 2$ and $\widetilde{T}$ on the basis of relations (4.6), (4.7) and (4.8) are continuum dimensional. It appears that if we assume a particular distributional hypothesis for the random vector r, sometimes it is possible to formulate an

optimization problem that is more simple and is consistent with a stochastic dominance order. For example, if r follows the multivariate Gaussian distribution with mean Er and covariance matrix Σ, $r \in N(Er, \Sigma)$, problem (4.2) is consistent with RSD, and therefore with SSD.

The class of the elliptical distributions

Owen and Rabinovitch (1983) extend the classical results with the normal distribution for the more general class of the elliptical distributions, $E(\mu, Q, \phi)$. There are representatives of the elliptical family which have infinite variance and can be used to model heavy-tailed series. For a description of the basic properties, see Section 1.5.1. The next dominance rules can be verified under the assumption of the elliptical distribution for portfolio items returns, see (Rachev and Mittnik, 2000, 433-435).

1. *Elliptical stochastic order for non-satiable investors.* Suppose that portfolio $v^T r$ has the same dispersion as portfolio $w^T r$, i.e. $w^T Q w = v^T Q v$, and also $v^T Er < w^T Er$. Then for every $t \in \mathbb{R}$,

$$\frac{t - w^T r}{w^T Q w} \leq \frac{t - v^T r}{v^T Q v} \quad \text{and} \quad F_{w^T r}(t) \leq F_{v^T r}(t)$$

   Therefore $w^T r \succeq_{FSD} v^T r$.

2. *Elliptical stochastic order for risk-averse investors.* Suppose that $w^T Er = v^T Er = m$ and we have the dispersion relation $w^T Q w < v^T Q v$. Then for every $t \in \mathbb{R}$,

$$\int_{-\infty}^{t} F_{w^T r}(u) du \leq \int_{-\infty}^{t} F_{v^T r}(u) du$$

   Therefore $w^T r \succeq_{RSD} v^T r$.

3. *Elliptical stochastic order for non-satiable, risk-averse investors.* Give two portfolios w and v with mean relation $w^T Er \geq v^T Er$ and dispersion relation $w^T Q w \leq v^T Q v$, where at least one inequality is strict, it follows that $w^T r \succeq_{SSD} v^T r$.

Markowitz (1959) and Tobin (1958) define the efficient frontier for non-satiable, risk-averse investors whose portfolio distributions are unambiguously characterized by mean and variance. The same can be expressed by the following mean-dispersion dominance rule in the elliptical world. We say that the portfolios in the market are ordered by the mean-dispersion dominance rule if for every couple of portfolios $w^T r$ and $v^T r$ with $w^T Er \geq v^T Er$ and $w^T Q w \leq v^T Q v$ where at least one inequality is strict, $w^T r$ is preferable

to $v^T r$ and, in effect, dominates it. It is clear that the elliptical stochastic order 3) justifies the Markowitz-Tobin rule. By applying this rule to all admissible portfolios, we obtain the *Markowitz-Tobin efficient frontier* — these are the portfolios which are not dominated in the sense of SSD.

Under the assumption of the elliptical distribution, the portfolio dispersion is a measure of risk. Hence solving a problem of type (4.2) with $\Sigma = Q$ provides optimal portfolios which are not dominated in the sense of RSD and belong to the classic extended Markowitz-Tobin mean-dispersion frontier. If we do not restrict the portfolio positions to be long only and allow unlimited short selling, it is possible to find a closed-form solution.

Theorem 16. *Suppose that there are $n \geq 2$ risky assets traded in a frictionless economy where unlimited short selling is allowed and suppose that portfolio returns belong to the same elliptical family with non-singular dispersion matrix Q. It is also assumed that the random gross returns on any asset cannot be expressed as a linear combination of the gross returns on other assets. Then the solutions of the constrained problem:*

$$\begin{array}{rl} \min\limits_{w} & w^T Q w \\ \textit{subject to} & \\ & w^T e = 1 \\ & w^T E r = m \end{array} \tag{4.9}$$

are all the portfolio compositions w, satisfying the following analytic relation:

$$w = \frac{(CQ^{-1}\mu - BQ^{-1}e)m + AQ^{-1}e - BQ^{-1}\mu}{AC - B^2} \tag{4.10}$$

Equivalently, in the mean-dispersion plane we have the relationship

$$\sigma^2(AC - B^2) - m^2 C + 2mB - A = 0$$

where $\mu = Er$, $m = w^T\mu$, $e = (1, 1, \ldots, 1)^T$, $A = \mu^T Q^{-1}\mu$, $B = e^T Q^{-1}\mu$, $C = e^T Q^{-1} e$ and $\sigma^2 = w^T Q w$.

Proof. See (Rachev and Mittnik, 2000, 435-437). □

A risk-free asset can be included in the optimal portfolio problem, see (Rachev and Mittnik, 2000, 436).

All portfolio compositions from (4.10) are not dominated in the sense of RSD. As a matter of fact, by the elliptical stochastic order rule for risk averse investors, every portfolio not dominated in the sense of RSD must have minimum dispersion for a fixed mean. Therefore the portfolio compositions from (4.10) belong to $\widetilde{T}$, which is also called *Markowitz-Tobin frontier*. If we apply the mean-dispersion dominance rule, we obtain the Markowitz-Tobin efficient frontier. It consists of all portfolios that satisfy:

$$w = \frac{(CQ^{-1}\mu - BQ^{-1}e)m + AQ^{-1}e - BQ^{-1}\mu}{AC - B^2}, \quad \text{with} \quad m \geq \frac{B}{C}$$

It is possible to show that if portfolio items returns belong to the elliptical family with non-singular dispersion matrix, the mean-dispersion dominance rule is equivalent to SSD, see (Rachev and Mittnik, 2000, 438-439) and the references therein.

Translation and scale invariant classes of distributions

Some of the results obtained for the family of the elliptical distributions can be generalized for any location and scale invariant class of distributions, for more details see Ortobelli et al. (2003) and Ortobelli et al. (2004). We say that a class of distributions is location and scale invariant, we use the notation $\sigma\tau_k(a)$, if it has the following characteristics:

1. Every distribution $F_X \in \sigma\tau_k(a)$ is described by a vector of k parameters
$$(m_X, \sigma_X, a_{1,X}, \ldots, a_{k-2,X}) \in \mathbb{R}^k$$
where m_X is the mean of X and σ_X is a positive scale parameter of X. It is assumed that the class is weakly determined by its parameters. That is the equality
$$(m_X, \sigma_X, a_{1,X}, \ldots, a_{k-2,X}) = (m_Y, \sigma_Y, a_{1,Y}, \ldots, a_{k-2,Y})$$
implies that $X \stackrel{d}{=} Y$ and the converse is not necessarily true.

2. If $F_X \in \sigma\tau_k(a)$, then for every admissible real t the translated distribution F_{X+t} belongs to the same class with parameters
$$(m_X + t, \sigma_X, a_{1,X}, \ldots, a_{k-2,X})$$

3. If $F_X \in \sigma\tau_k(a)$, then for every admissible positive λ the re-scaled distribution $F_{\lambda X}$ belongs to the same class with parameters
$$(\lambda m_X, \lambda\sigma_X, a_{1,X}, \ldots, a_{k-2,X})$$

On condition that the distribution of the random portfolio returns belong to a translation and scale invariant class, we can identify a stochastic dominance relation.

Theorem 17. *Suppose that the distribution functions of all random portfolio returns belong to one and the same class $\sigma\tau_k(a)$. Let $w^T r$ and $v^T r$ be two portfolios which are unbounded from below. Their distribution functions are described by the parameters*

$$(m_{w^T r}, \sigma_{w^T r}, a_1, \ldots, a_{k-2}) \quad \textit{and} \quad (m_{v^T r}, \sigma_{v^T r}, a_1, \ldots, a_{k-2})$$

respectively. Then the following statements are equivalent:

1. $m_{w^T r} \geq m_{v^T r}$ *and* $\sigma_{w^T r} \leq \sigma_{v^T r}$ *with at least one equality strict.*
2. $w^T r \succeq_{SSD} v^T r$

Proof. For the proof, see Ortobelli et al. (2003). □

As a corollary, it follows that we can formulate an optimization problem the solutions of which are consistent with SSD. The optimization problem is given in the next

Corollary 3. *Under the conditions of Theorem 17, let us consider the problem*

$$\begin{array}{rl} \min\limits_{w} & \sigma_{w^T r} \\ \textit{subject to} & \\ & w^T e = 1 \\ & w^T E r = m \\ & a_{1,w^T r} = a_1^* \\ & a_{2,w^T r} = a_2^* \\ & \ldots \\ & a_{k,w^T r} = a_k^* \end{array} \tag{4.11}$$

where $a_1^*, a_2^*, \ldots, a_k^*$ are predefined values. The solutions of problem (4.11) are consistent with SSD.

The optimization problem (4.11) can be more or less difficult, depending on the relationship between the scale of the portfolio distribution, $\sigma_{w^T r}$, and portfolio weights w on one hand, and the functions $a_{i,w^T r} = f(w)$, $i = 1, \ldots, k$ on the other. For example, problem (4.9) is a version of (4.11) for the translation and scale invariant class of the elliptical distributions. Because of the specific properties of the elliptical family, the portfolio scale is a quadratic function of portfolio weights which makes (4.9) suitable to handle computationally. Unfortunately this is not always the case — there are classes of distributions for which $\sigma_{w^T r}$ is a more complicated function of w and the additional parameters of the location and scale invariant family that can make the optimization problem very difficult.

Let us have a closer look at the conditions of Theorem 17 and problem (4.11). Taking advantage of the generic properties of the location and scale invariant class, we obtain the following result.

Theorem 18. *Suppose that the distribution functions of all random portfolio returns belong to one and the same class $\sigma\tau_k(a)$ and that $\rho(\cdot)$ is a positive risk measure with the property $\rho(ax) = a\rho(x)$, $a > 0$. The following statements hold.*

1. *Let us have two portfolios w and v such that their distributions are described by*

$$(m, \sigma_{w^T r}, a_1, \ldots, a_{k-2}) \quad \textit{and} \quad (m, \sigma_{v^T r}, a_1, \ldots, a_{k-2})$$

respectively. Then

$$v^T r \preceq_{RSD} w^T r \quad \Longleftrightarrow \quad \rho(w^T r) \leq \rho(v^T r)$$

2. *Let $X = \{w \in D \subset \mathbb{R}^n | w^T E r = m, a_{j,w^T r} = a_j^*, j = 1, \ldots, k\}$, then*

$$\arg\min_{w \in X} \rho(w^T r) = \arg\min_{w \in X} \sigma_{w^T r}$$

Proof. For a proof, see Stoyanov (2005). □

Since, as we have mentioned, the class of the elliptical distributions is a particular example of a location and scale invariant class, the result of Theorem 18 holds for it as well. Both results in Theorem 18 do not hold if we allow some of the additional parameters a_j, $j = 1, \ldots, k$ to depend on the portfolio composition. In this case, the choice of different classes of risk measures will generate different optimal solutions.

The class of the multivariate α-stable distributions

Empirical research in the field of finance has shown that the Gaussian distribution cannot account for the observed properties of the distribution of assets returns — heavy tails, asymmetry and excess kurtosis. Mandelbrot and Fama, in their pioneering work in the 1960s, reject the assumption of normality and suggest a more general family of distributions — the class of stable laws. For more recent studies on the application of stable distributions in finance, see Rachev and Mittnik (2000), Rachev (2003) and Rachev (2004).

A natural candidate for a multivariate distributional assumption is the class of the multivariate stable laws. If we assume that the vector of assets returns is in the domain of attraction of a multivariate stable distribution then we can take the limit distribution as an approximate probabilistic model. That is, suppose $r \in S_\alpha(\Gamma, \mu)$, where $\Gamma(\cdot)$ is the spectral measure and μ is a vector location parameter, then according to the properties of the multivariate stable distributions, portfolio returns follow a univariate stable distribution with the same index of stability, $w^T r \in S_\alpha(\sigma_w, \beta_w, \mu_w)$ where

$$\sigma_w = \left(\int_{S^n} |w^T s|^\alpha \Gamma(ds) \right)^{1/\alpha}, \quad \beta_w = \frac{1}{(\sigma_{w^T r})^\alpha} \int_{S^n} |w^T s|^\alpha sign(w^T s) \Gamma(ds),$$
$$\text{and} \quad \mu_w = w^T \mu$$

An important assumption is $\alpha > 1$, because in this case the mathematical expectation of portfolio returns is finite and $\mu_w = w^T Er$. Actually this is a consequence of the more general result about the distribution of linear combinations of the components of multivariate α-stable vectors, see Section 1.5.2 for more details about their properties.

Clearly portfolio returns belong to one and the same distribution family. Moreover σ_w and μ_w are a scale and a location parameters respectively. Hence we have a location and scale invariant class with four parameters, $\sigma\tau_4(a)$, and all results from the previous paragraph can be applied (Theorems 17 and 18). We can formulate an optimization problem, equivalent to problem (4.11), which is consistent with SSD:

$$\begin{array}{rl} \min\limits_{w} & \sigma_w \\ \text{subject to} & \\ & w^T e = 1 \\ & w^T Er = m \\ & \beta_w = \beta^* \end{array} \tag{4.12}$$

Clearly, in this case, the scale parameter σ_w and the additional parameter β_w are complicated functions of w. For alternative optimization problems, in which σ_w and β_w are replaced by functions of some moments, see Ortobelli et al. (2003).

Another difficulty with problem (4.12), and generally with the class of multivariate stable distributions, is that the spectral measure $\Gamma(\cdot)$ is hard to estimate from empirical data. The portfolio choice problem is significantly simplified if we consider a sub-class of the multivariate stable family, the class of the sub-Gaussian distributions where the parameter estimation is a more simple task. The sub-Gaussian stable laws are shifted symmetric multivariate α-stable laws, see Section 1.5.2 for the definition and some properties. For this multivariate distributional assumption, portfolio returns follow the symmetric stable law $w^T r \in S_\alpha(\sigma_w, 0, \mu_w)$ with

$$\sigma_w = \sqrt{w^T Q w}, \quad \text{and} \quad \mu_w = w^T \mu$$

where Q is a positive definite matrix called the dispersion matrix and $\mu = Er$ is the vector of expected returns (recall that we assume $\alpha > 1$). The sub-Gaussian stable laws form a special case of the more general class of the elliptical distributions and for this reason problem (4.12) reduces to problem (4.9) and has a closed-form solution provided that unlimited short selling

is allowed, see Theorem 16. For more details about this kind of elliptical problem and empirical studies of the optimization problem, see Ortobelli et al. (2003), Ortobelli et al. (2004) and Ortobelli et al. (2005).

The class of the operator stable distributions

A property of the class of the multivariate stable distributions is that all marginals have one and the same index of stability. Empirical research has shown that the indices of stability vary with asset types. Thus a more realistic distributional assumption would be one that allows for different tail behaviour of the marginal distributions. The class of operator stable distributions has this advantage. This class also arises from limit theorems and the multivariate stable laws appear as a special case; employing it, we do not lose the theoretical appeal that the limit theorems provide. See Meerschaert and Scheffler (2003) and the references therein and Kozubowski et al. (2003) for other multivariate models. Also, see Section 1.5.3 for more details about the class of operator stable distributions.

If we assume that the vector of assets returns is in the domain of attraction of an operator stable distribution, then we can use the limit distribution as an approximate mathematical model. Suppose that r is operator stable with exponent

$$E = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix}$$

where $1/2 \leq a_i < 1$, $i = 1, \dots, n$ then each marginal is α-stable with exponent $\alpha_i = 1/a_i \in (1, 2]$, $i = 1, \dots, n$. It comes out that if we have a portfolio $w = (w_1, \dots, w_n)$, then the portfolio return $w^T r$ is a random variable with tail behavior $P(|w^T r| > t) \sim Ct^{\alpha_w}$ where C is some constant and $a_w = \min_i(\alpha_i : w_i \neq 0)$, which means that the heaviest tail among those of the items with non-zero weight will dominate. That is, the random variable $w^T r$ is in the domain of attraction of some α-stable distribution $S_{\alpha_w}(\sigma_w, \beta_w, \mu_w)$. In case $\alpha_w > 2$, then the portfolio return is in the domain of attraction of the Gaussian distribution because it has a finite second moment. That is why we can write that $\alpha_w = \min(\alpha_w, 2)$. Since σ_w and μ_w are a scale and a location parameter, we can again pose the problem in terms of a location and scale invariant class $\sigma\tau_4(a)$ and use the corresponding results in Theorems 17 and 18. The optimization problem corresponding to (4.11), and hence consistent with SSD, is

$$\begin{aligned} \min_{w} \quad & \sigma_w \\ \text{subject to} \quad & \\ & w^T e = 1 \\ & w^T Er = m \\ & \alpha_w = \alpha^* \\ & \beta_w = \beta^* \end{aligned} \tag{4.13}$$

As in the case of the multivariate stable assumption, σ_w, β_w and α_w can be replaced by functions of some moments, see Ortobelli et al. (2003). The constraint $\alpha_w = \alpha^*$ can also be regarded as being equivalent to excluding certain assets from the optimal portfolio. Since $a_w = \min_i(\alpha_i : w_i \neq 0)$, then if α_k, $k \in K \subset \{1, \ldots, n\}$ is below the target α^*, then the assets in the index set K will not be included in the optimal portfolio. Due to the existence of some estimation error, one could leave that to the optimization problem.

4.4 The ETL portfolio optimization model

In recent years, significant efforts have been dedicated to building extensions to the classical mean-variance analysis. The principal reason is that it leads to correct decisions only when the vector of assets returns follows the multivariate normal distribution, i.e. $r \in N(Er, \Sigma)$, and, as we have noted, there is ample empirical evidence against that assumption. The extensions involve including different risk measures in the optimization problems. In the next sections we describe the extension of the modified equivalents to (4.1) and (4.2) and the related optimal reward-risk ratio problems.

The mean-variance principle can be extended for a general risk measure $\rho(\cdot)$, not necessarily coherent, and the corresponding optimization problems can be re-stated. For example, Problem (4.2) becomes:

$$\begin{aligned} \min_{w} \quad & \rho(w^T r) \\ \text{subject to} \quad & w^T e = 1 \\ & w^T Er \geq R_* \\ & Lb \leq Aw \leq Ub \end{aligned} \tag{4.14}$$

The difficulty of solving in practice problem (4.14) depends on the particular choice of the risk measure $\rho(\cdot)$. Generally, if the risk measure is a convex function of w, then the optimization problem belongs to the class of the convex problems and can be solved using the methods of convex optimization. The assumption of convexity is very natural because it allows diversification. If additionally $\rho(\cdot)$ is differentiable, then one can take advantage of a method for numerical optimization of smooth functions, for example gradient methods, or sequential quadratic programming methods.

Sometimes, when the risk measure is more specific, the problem can be further simplified. We have already discussed the case of the standard deviation being a risk-measure or, more generally, the dispersion in the elliptical family, in which case the (4.14) reduces to a quadratic programing problem. There are examples in which (4.14) reduces to a linear programing problem — when the risk measure can be linearized. For instance this is the case when $\rho(\cdot)$ is the ETL function or a spectral measure of risk, see Rockafellar and Uryasev (2002) and Acerbi and Simonetti (2002).

Certainly very often more than one approach is applicable for solving (4.14). For a linearizable risk measure, it is possible to use either the underlying linear programming problem or the more general convex problem. Each approach has its advantages and one should be able to assess the trade-off and choose the most suitable for the corresponding application. We are not going to discuss (4.14) in its most general form. Such general discussions can be found in Rachev, Stoyanov and Fabozzi (2008). We consider the special case when the risk measure is the ETL.

In the the beginning of this chapter, we noted that the mean-variance framework employs variance as a proxy for risk. However, variance is not a true risk measure because it penalizes symmetrically profit and loss. Therefore, a reasonable question is if we can extend the reasoning in the previous section with a quantity which performs better at measuring risk. In the Cognity system, we have adopted ETL and, thus, we have extended the mean-variance framework into the mean-ETL framework in which we measure risk by means of ETL.

Practically, the mean-variance principles remain valid after replacing variance with ETL. We can re-define the three basic types of problems we discussed in the previous section:

- minimize ETL subject to a constraint on portfolio expected return;
- maximize portfolio expected return subject to a constraint on portfolio ETL;
- maximize a utility function composed of portfolio expected return and portfolio ETL;

In order to illustrate the new framework, we provide an example with the utility function approach. The utility function is composed of two terms — one responsible for portfolio expected return and another for portfolio ETL

$$\max_{w} \quad [Er_p - \lambda \times ETL_\epsilon(r_p)] \tag{4.15}$$

where w is a vector containing portfolio weights, r_p stands for portfolio return $r_p = w'r$, and λ is a positive risk-aversion coefficient. The magnitude

of the risk-aversion coefficient represents portfolio manager risk preferences, i.e., higher λ value stipulates higher risk-averse attitude. Note that unlike the mean-variance framework, the optimal solution depends on one additional parameter which defines the risk measure — the tail probability of the ETL. Varying the tail probability while holding the risk-aversion parameter fixed will lead to different optimal allocations.

4.4.1 Model statement

By definition, ETL at tail probability $1-\epsilon$, denoted by $ETL_\epsilon(r_p)$, is the average of the value-at-risk (VaR) numbers larger than the VaR of the portfolio loss $(-r_p)$[6] at tail probability $1-\epsilon$. The complement of the tail probability, i.e., the value of ϵ, is a user-specified quantity. It should be chosen in such a way that it best suits investment and risk-averse preferences of the investors.

From a mathematical viewpoint, the optimization problems involving ETL belong to the general class of convex programming problems. However, the structure of the ETL allows for a significant simplification of the optimization problem when we consider scenarios. Thus, at the cost of introducing additional variables, the optimization problem can be formulated as a linear programming problem which, as far as the structure is concerned, is simpler than quadratic problems. We say that a problem is linear when the objective function is a linear function also the set of constraints is defined by means of linear equations. In the sequel, we define the model by means of the linearization technique developed in Rockafellar and Uryasev (2002).

The scenarios utilized in the portfolio optimization module are generated on the basis of the assumptions specified in the scenario generation settings. Generally, the model relies on a set of available scenarios which can be generated from any parametric model. Thus, we can generate scenarios based on stable distributions, or a more general model involving factor analysis, and a dependence model such as the skewed Student's t copula. Suppose that for a given number of n funds, we simulate k returns scenarios per fund. Thus we obtain a $k \times n$ matrix, H, containing those scenarios in rows (and columns per fund)

$$H = \begin{pmatrix} r_1^1 & r_2^1 & \dots & r_n^1 \\ r_1^2 & r_2^2 & \dots & r_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^k & r_2^k & \dots & r_n^k \end{pmatrix}, \tag{4.16}$$

where r_j^i stands for the i-th return scenario of the j-th fund. The matrix H can be constructed from historical observations[7] as well. Employing this

[6] Note, portfolio return r_p implies portfolio loss $(-r_p)$.

[7] In such case the time is modeled by the index i for historical dates ranging from time t_1 to time t_k.

matrix notation let us state the linear programming formulation of the minimum ETL problem for portfolio w and funds returns scenarios H

$$\begin{array}{rcll} ETL_\epsilon(Hw) & = & \min\limits_{w,\theta,d} & \theta + \dfrac{1}{k\epsilon} d'e \\ & & \text{s.t.} & -Hw - \theta e \leq d \\ & & & w'e = 1 \\ & & & d \geq 0,\, \theta \in \mathbb{R} \end{array} \tag{4.17}$$

where $d' = (d_1, \ldots, d_k)$ is a vector of auxiliary variables, $e = (1, \ldots, 1) \in \mathbb{R}^k$ is a vector of ones, and $\theta \in \mathbb{R}$ is an additional variable coming from the linearization technique and yielding the VaR quantile at tail probability ϵ. For additional information about the optimization problem, see Rockafellar and Uryasev (2002). The first inequality in (4.17) involves vectors and must be interpreted in a component-wise manner (i.e., one inequality for each vector entry, in other words, one inequality for each scenario). Mathematically, this is expressed by the following equivalent relation

$$-Hw - \theta e \leq d \qquad \Longleftrightarrow \qquad \left| \begin{array}{l} -r^1 w - \theta \leq d_1 \\ -r^2 w - \theta \leq d_2 \\ \ldots \\ -r^k w - \theta \leq d_k \end{array} \right.$$

where r^i is the i-th row in H for i ranging from 1 to k, that is, the vector r^i contains the i-th scenario for each fund in our portfolio.

The portfolio model in (4.17) optimizes the ETL of a portfolio by finding a composition w which minimizes ETL. The model may seem somewhat complicated because of the vector-matrix notations, but in fact, it has a very simple structure. The objective function is a linear function and all constraints are linear equalities and inequalities. There are very efficient algorithms for solving such types of so-called *linear programming problems*.

The minimum ETL problem (4.17) can be extended with other linear constraints such as the expected returns constraint

$$Er_p \equiv w'\mu \geq R_*.$$

All relevant types of such constraints are described in the next section.

In a similar fashion, we formulate the maximum expected portfolio return problem subject to an upper bound R^* on the ETL quantity. For this reason, we include the objective function of (4.17) in the constraint set as follows

$$\begin{array}{rl} \max\limits_{w,\theta,d} & w'\mu \\ \text{s.t.} & -Hw - \theta e \leq d \\ & w'e = 1 \\ & \theta + \dfrac{1}{k\epsilon} d'e \leq R^* \\ & d \geq 0,\, \theta \in \mathbb{R}. \end{array} \tag{4.18}$$

The structure of the resulting formulation (4.18) is again a linear programming problem.

The objective function in the last two optimization models (4.17) and (4.18) can have a more composite form such as the one shown in (4.15) combining risk and return. We still can transform it in easy-to-optimize linear function

$$\max_{w,\theta,d} \quad \left[w'\mu - \lambda \times \left(\theta + \frac{1}{k\epsilon} d'e \right) \right]. \tag{4.19}$$

In this utility-type objective function we might additionally include a turnover expression without breaking the nice structure.

4.4.2 The ETL efficient frontier

By varying λ and solving problem (4.19), we arrive at the set of *mean-ETL efficient portfolios*. These portfolios can be also obtained either through solving (4.17) or (4.18) with an appropriate lower or upper bound constraint, R_* or R^*, on the expected return or on the ETL quantity, respectively. When we plot the expected return and the ETL of the efficient portfolios in the mean-ETL plane we arrive at the mean-risk efficient frontier (EF). The EF shows the trade-off between risk (namely ETL) and expected return of the mean-ETL efficient portfolios.

From a mathematical viewpoint, the most robust way to obtain the EF is through the minimum ETL optimization given in (4.17). In an EF optimization, the corresponding model is repeatedly solved subject to a series of lower bounds on the portfolio expected return.

The EF graph shows the trade-off between risk and return of the reward-risk efficient portfolios. he EF is built in the following way. First, we solve for the so-called global minimum risk portfolio. That is, we minimize the portfolio risk (measured through variance or ETL) leaving the return unconstrained. Second, we solve for the maximum portfolio expected return leaving the portfolio risk unconstrained. These two portfolio optimization models yield the end points of the EF graph. Then, a uniform grid is constructed for the upper bounds on the expected return, and a series of portfolio optimization problems is solved — minimizing the risk for each level of portfolio return on the uniform grid. The resulting risk-return pairs are plotted as shown on top part of Figure 4.2.

The bottom part of Figure 4.2 illustrates how portfolio weights change along the efficient frontier. An important observation from the plots on these figures concerns the so-called portfolio diversification. That is, portfolios corresponding to lower risk and return values (EF points on the left side of the graph) are more diversified portfolios. In other words, more funds are being included in the optimal portfolio allocation. So that the portfolio risk is being diversified among more funds. The left side of the two figures

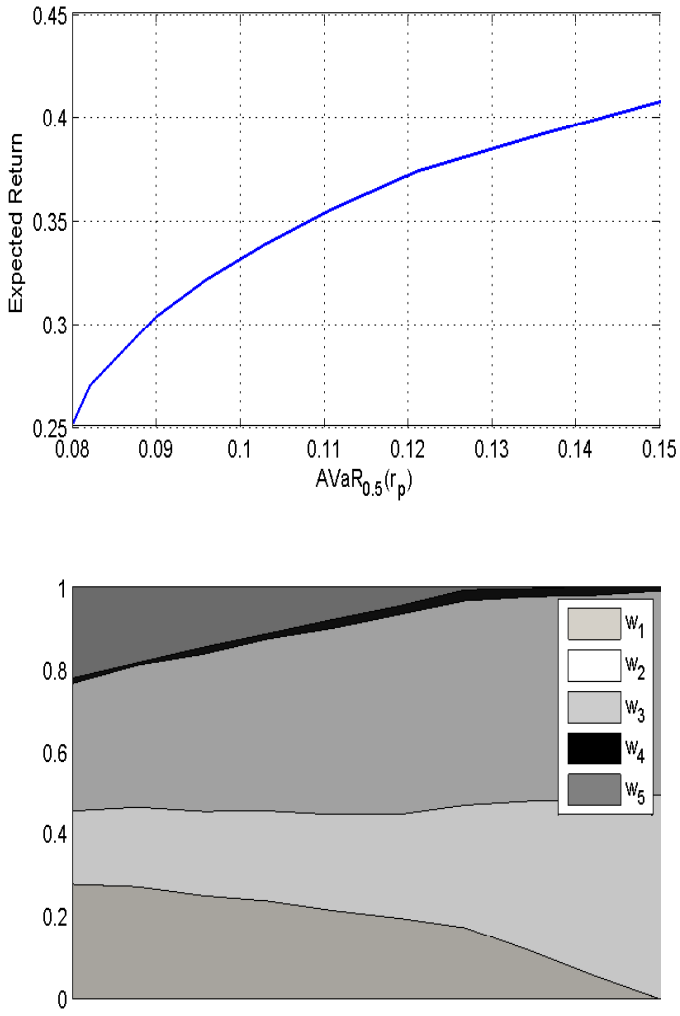

Figure 4.2: The top plot shows the efficient frontier in the mean-ETL plane. The bottom plot shows the compositions of the optimal portfolios along the efficient frontier. Both plots have the same horizontal axis. The shaded areas correspond to the values of the portfolio weights w_1 through w_5 located from the bottom to the top on the graph.

shows significant percent holdings allocated among all weights w's. On the other hand, the right side of the two figures shows percent holdings allocated only between two w's weights which indicates that these portfolios are more concentrated.

This observation is related to portfolio diversification is actually the cornerstone of the highly spelled-out *trade-off* between risk and return. Diversified portfolios are less risky but they also have smaller expected return. Higher return portfolios are more risky and, i.e. less diversified.

An important property of the EF plot is that if the expected return is viewed as a function of the ETL or standard deviation, this function is a concave one. This property is basically due to the diversification effect, for a mathematical proof, see Stoyanov (2005). If a fund manager constructs all possible portfolios from the funds included, then all these portfolios will appear below the EF line. Portfolios which are not placed on the EF line are called suboptimal portfolios. When we solve a minimum risk or maximum return portfolio optimization problem, then we always obtain a portfolio on the EF line.

Concerning the efficient portfolios along the EF line, they can also be compared in terms of their risk-adjusted return. Thus, among the EF portfolios the most valuable one is the portfolio which yields the maximum return to risk ratio. Provided the portfolio risk is identified with the portfolio standard deviation then the resulting return-to-risk ratio is called the *Sharpe Ratio*. Similarly, if the portfolio risk is measured by the portfolio ETL then the resulting ratio is called the *STARR Ratio*.

4.5 STARR ratio optimization

A key step in the investment management process is measurement of portfolio performance. A formula which quantifies the portfolio performance is called a performance measure. A widely used measure of performance is the Sharpe ratio. In essence, it is the ratio between the average active portfolio return and the standard deviation of portfolio return. In this way, it is a reward-to-variability ratio in which the variability is computed by means of the standard deviation.

As pointed out in the previous chapter, standard deviation penalizes both the upside and the downside potential of portfolio return. Therefore, it is not a very appropriate choice as a measure of performance. This deficiency was recognized and many alternatives to the Sharpe ratio have been proposed in the literature. Some of them are reward-to-variability ratios in which a downside dispersion measure is used in the denominator. One example is the Sortino ratio, in which the downside semi-standard deviation is used as a measure of variability.

Other types of performance measures are reward-to-risk ratios. In con-

trast to the reward-to-variability ratios, they calculate the risk-adjusted active reward of the portfolio. For example, the Sortino-Satchell ratio calculates the average active return divided by a lower partial moment of the portfolio return distribution. And, the STARR ratio calculates the average active return divided by ETL at a given tail probability. Another ratio commonly referred as the MAR ratio is the Calmar ratio, see Rachev, Fabozzi and Menn (2005) and Rachev, Stoyanov and Fabozzi (2008). It is a performance measure that is used in comparing Commodity Trading Advisors or alternative asset managers. The Calmar ratio is the absolute value of the ratio of the annual compounded return divided by the largest drawdown[8] incurred to date. Such ratios with values close to 1 are very rare in real world trading for an extended period of time. For example, if we are striving for a compounded annual return of 20% than we can expect our largest drawdown to be at least minus twenty percent.

There are examples in which a reward measure is used instead of the average active return. For instance, the Farinelli-Tibiletti ratio is essentially a ratio between an upside and a downside partial moment of the portfolio return distribution. And, the Rachev ratio (R-ratio) is a ratio between the average of upper quantiles of the portfolio return distribution and the portfolio ETL which is, essentially, an average of the lower quantiles.

In this section we focus on solving the maximum STARR ratio portfolio optimization problem. This is a reward-to-risk ratio where the risk is measured by portfolio ETL. The abbreviation STARR stands for *stable tail-adjusted return ratio*. The abbreviation initially comes from the assumption that funds returns follow the stable distribution. However, the concept behind STARR can be translated to any distributional assumption. Formally, STARR is defined as

$$STARR_{\epsilon}(w) = \frac{E(r_p - r_b)}{ETL_{\epsilon}(r_p - r_b)}$$

where r_p is funds portfolio return, r_b is a target benchmark return, and ETL_{ϵ} is the average of the lower $\epsilon\%$ of the portfolio active returns. If r_b is a constant benchmark return, then STARR equals

$$STARR_{\epsilon}(w) = \frac{w'\mu - r_b}{ETL_{\epsilon}(r_p) + r_b}$$

where $w'\mu = Er_p$ is the fund portfolio expected return. The target benchmark return can be set equal to zero if we are not interested in an active optimization problem.

The allocation of the optimal STARR ratio problem is found by solving the following optimization problem

[8]The drawdown is the measure of the decline from a historical peak in the cumulative profit of a financial trading strategy.

$$\begin{array}{ll} \max\limits_{w} & \dfrac{E(r_p - r_b)}{ETL_\epsilon(r_p - r_b)} \\ \text{s.t.} & w'e = 1 \end{array} \tag{4.20}$$

where additional constraints can also be added. Mathematically, by introducing additional variables and constraints (4.20) can be reduced to a problem with a simpler structure, see Rachev, Stoyanov and Fabozzi (2008) for more information. As a result, it can be solved with a solver for linear programming problems.

The solution of the *max STARR* ratio optimization problem yields a portfolio which is represented by the point on the mean-ETL efficient frontier studied in the previous section — this is the point which delivers the highest mean return to ETL ratio.

4.6 Types of constraints

The optimization models in the mean-variance and the mean-ETL portfolio optimization frameworks described in the last two sections can be solved subject to a variety of constraints. That is, each objective function such as minimum variance, maximum return, minimum ETL, or the relevant utility-type function combinations can be optimized subject to different user-specified constraints. Generally, the constraints materialize certain investment goals which portfolio managers should meet. Among those goals we may list some fund-specific, geography, or strategy preferences as well as standard risk, return, and turnover requirements. All of these cases can be constructed in the Cognity system.

In this section, we present in a more technical way four different types of portfolio constraints that can be modeled in the system. These general types of constraints, and their relevant subtypes, are:

- weight constraints
  - box-type
  - grouped by strategy, industry, geography
- expected return constraints
  - absolute
  - relative to a benchmark
- turnover constraints
- ETL constraints
  - absolute

- tracking error type
- difference

From a methodology viewpoint, any constraint which is either linear or is equivalently representable by a set of linear constraints, can be incorporated in the optimization problem.

Weight constraints

The box-type weight constraints allow bounding the portfolio weight of a particular fund. If we denote the weight of the i-th fund by w_i, then it can be constrained, for example, between three and eight percent by the following linear expression

$$a \leq w_i \leq b.$$

The lower and upper bounds[9] for w_i are user-supplied in a box-type interface in the Cognity system. We can use this constraint to specify a particular investment amount rather than lower and upper bounds. If we want to invest an exact amount of capital in a certain fund then we can specify an equality $w_i = a = b$ instead of a two-sided inequality. In this case, regardless of the type of optimization performed in Cognity, we will end up with an optimal portfolio which satisfies our investment wish for that specific fund. In the system, the lower and upper bounds, and the comparison sign are specified in the weight constraint preferences.

Weight constraints applied on groupings of funds are also supported. They have the general form

$$a \leq w_{i_1} + w_{i_2} + ... + w_{i_s} \leq b$$

where all s funds with corresponding weights $w_{i_1}, w_{i_2}, ..., w_{i_s}$ belong to a certain group, i.e., belong to some strategy, industry, geography, etc. In this way, we can bound the invested amount in a given preselected group of funds.

Expected return constraints

The expected return constraint is modeled as explained already in the previous two sections. In particular, the *absolute* expected return constraint has the form

$$Er_p \equiv w'\mu \geq R_* \tag{4.21}$$

[9]Throughout Section 4.6, we assume that the lower and upper bound notations a and b stand for some percent numbers, for example, $a = 3\%$ and $b = 8\%$.

The selection of the R_* number is a responsibility of the portfolio manager. The choice of R_* should be made in a reasonable way because an inappropriate choice for the lower bound R_* can make the corresponding model infeasible — that is, the optimization solver in Cognity might not find any allocation of portfolio weights w_i's that achieve an unrealistically high goal R_* for the portfolio expected return.

The expected return constraint may also be relative to the expected return of a benchmark portfolio. The model formulation of such a constraint is as follows

$$w'\mu - Er_b \geq R_*$$

where r_b denotes the return of the benchmark portfolio. Since in this case the constraint is relative to the expected performance of a benchmark portfolio, we call this constraint *relative*. The relative expected return constraint is an *ex-ante* formulation of the goal of a portfolio manager to "beat" the expected return of a benchmark portfolio, Er_b, by a certain amount R_*.

Equivalently, the relative expected return constraint can be written as

$$E\left(r_p - r_b\right) \geq R_*.$$

The last formulation of the expected return benchmark-type constraint inspires the formulation of a similar constraint for the ETL risk measure presented later in this section. Because of the ETL specifics, it turns out that in that case we can distinguish two separate benchmark-type constraints.

Turnover constraints

Basically, the turnover constraint allows bounding from above the trading volume incurred in portfolio rebalancing. The optimization goal can be any of the goals already described in the chapter.

In the Cognity system, this feature is modeled through portfolio weights w_i's and the initial portfolio weights denoted here by w_i^0's. That is, we assume we currently hold the open positions represented by w_i^0's portfolio weights. Based on the new information we gather, we look for a new optimal allocation represented by w_i's weights. Mathematically, the turnover quantity is modeled by the absolute value of the difference between the initial and the (new) optimal holdings. That is, regardless of the fact whether we will buy some or sell some shares the generated trading volume is given by the following quantity

$$i\text{-th fund turnover} = |w_i - w_i^0|.$$

This holds because we get the money added or deducted from the position by multiplying the initial and the final weights by the portfolio value. Summing

up the turnover for all positions, we get the turnover for the whole portfolio, which we can bound from above by a maximal admissible turnover,

$$\sum_{i=1}^{n} |w_i - w_i^0| \leq TO\%$$

where TO denotes the upper bound. We can convert all numbers in this constraint in dollar values by multiplying both sides of the inequality by portfolio value. In effect, we buy and sell fund holdings for less than TO percents of our invested capital.

From a mathematical viewpoint, the turnover constraint is not linear but it can be equivalently transformed to a set of linear constraints by introducing a set of auxiliary variables δ_i^+ and δ_i^-. While these auxiliary variables are "hidden" from the Cognity user, they can be easily interpreted. The first variable δ_i^+ denotes the money added to the i-th position as percentage from portfolio value and the variable δ_i^- denotes money deducted from the i-th position as percentage from portfolio value. Thus, both variables are non-negative and one of them should be always equal to zero because it is always suboptimal to simultaneously buy and sell one and the same fund. As a result, the difference between the initial and the optimal weights can be represented in terms of the new variables,

$$w_i - w_i^0 = \delta_i^+ - \delta_i^-.$$

The non-negativity of the new variables and the fact that one of them is always zero implies that the turnover can be expressed as

$$|w_i - w_i^0| = \delta_i^+ + \delta_i^-.$$

Finally, the constraint on the portfolio turnover can be replaced by the following set of linear constraints,

$$\begin{aligned} w_i - w_i^0 &= \delta_i^+ - \delta_i^- \\ \sum_{i=1}^{n} (\delta_i^+ + \delta_i^-) &\leq TO \end{aligned}$$

where $\delta_i^+ \geq 0, \delta_i^- \geq 0$ and the condition $\delta_i^+ \times \delta_i^- = 0$ for $i = 1, n$ is satisfied due to non-optimality of the converse.

ETL constraints

The ETL constraints provide a way for bounding the ETL on a portfolio or sub-portfolio level from above. That is, generally speaking this constraint ensures that the risk of the optimal portfolio will not exceed a certain limit. A typical ETL constraint is

$$\begin{aligned} &\theta + \frac{1}{k\epsilon} d'e \ \leq R^* \\ &-Hw - \theta e \leq d \\ &d \geq 0,\ \theta \in \mathbb{R}, \end{aligned} \tag{4.22}$$

where we bound the portfolio ETL quantity by an upper bound of R^*. In this formulation, we utilize the linearization technique and a short-hand notation for the linear representation (4.22) of the ETL constraint is

$$ETL_\epsilon(r_p) \equiv ETL_\epsilon(Hw) \leq R^*.$$

where r_p denotes portfolio return. Concerning the linearized version, we have to keep in mind that the additional variables d and θ are variables in the optimization problem, i.e. we optimize with respect to them as well.

The constraint in (4.22), we call *absolute* ETL constraint. The quantities ϵ, H, and w correspond respectively to the tail probability, the set of scenarios for the funds returns, and to the portfolio allocation, respectively. Apart from the absolute ETL constraint, we distinguish between two other types which concern the *relative* optimization problem, i.e. when the risk is measured relative to a benchmark portfolio. These two types are the *tracking-error* type and the *difference* type.

The *tracking error* ETL constraint is defined in a way similar to the tracking-error functional in the benchmark tracking problem. Basically, the tracking-error is defined as the standard deviation of the active return which is the difference $r_p - r_b$. Similarly, the tracking-error ETL is introduced as the ETL of the active return, $ETL(r_p - r_b)$, at some tail probability level. As a result, the tracking-error ETL constraint is simply an upper bound on this quantity,

$$ETL_\epsilon(r_p - r_b) \leq R^* \tag{4.23}$$

where r_b denotes the return of a benchmark portfolio.

In contrast, the difference type, as the name suggests, is essentially an upper bound on the difference between the portfolio ETL and the benchmark ETL,

$$ETL_\epsilon(r_p) - ETL_\epsilon(r_b) \leq R^* \tag{4.24}$$

which bounds the ETL of our portfolio by the ETL of the benchmark portfolio $ETL_\epsilon(r_b)$ plus a certain constant R^* which is user-defined in Cognity.

4.7 A schematic description of a portfolio optimization module

From a general viewpoint, the main input to a portfolio optimization problem based on ETL is the matrix of scenarios. They are used to approximate

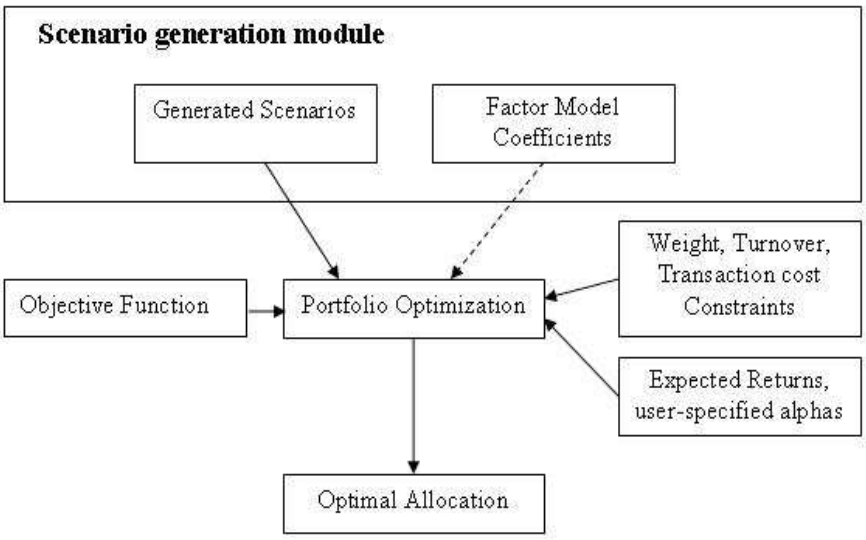

Figure 4.3: A schematic description of an ETL-based portfolio optimization module and its relationship with the scenario generation module.

portfolio risk and, as we noted, with their help the optimization problem can be transformed to a linear programing problem. These generic properties can be used to construct a portfolio optimization module which is based on heavy-tailed scenarios produced from a scenario generation module. The settings in the scenario generation module include preferences for marginal distributions and dependence structure — the two components of the multivariate model describing the joint distribution of stock returns. Therefore, the input to the portfolio optimization problem can be divided into two groups:

- future scenarios describing possible outcomes for the vector of stock returns
- optimization problem specific — constraints, objective function, expected returns (or user-specified alpha forecasts)

We discussed already there are a number of possible constraints. Some of them do not require any additional information except upper or lower bounds. These are easiest to specify directly in a user interface. There are other, more involved constraints, such as the ETL constraint. It requires the generated scenarios from the scenario generation module. The beta constraint requires additional information from a factor model.

Figure 4.3 describes the types of input. From the viewpoint of the portfolio optimization problem, the model we employ to generate scenarios from is completely irrelevant. The information needed for the computation of portfolio risk is extracted from the generated scenarios and, thus, as long as there are scenarios, the actual distributional assumptions are not needed. In contrast, in the Markowitz framework the optimization problem is consistent with a multivariate normal distribution by design. In the ETL-based framework, the scenarios can be generated through a factor model or through complicated combinations of time-series models and factor models. For this reason, in case a factor model is defined, the factor model coefficients needed for the beta constraint are provided by the scenario generation module.

4.8 Strategy construction based on ETL minimization

In general, there are several types of strategies depending on whether the risk and/or the expected return of the optimal portfolio are compared to those of a benchmark with the goal of beating the benchmark. If so, then the strategy is said to be active as opposed to the passive strategy of replicating a benchmark. In the classical Markovitz framework, an example of a benchmark tracking problem is the tracking error problem. In it, the standard deviation of the excess return, which is called tracking error, is minimized subject to constraints. The tracking error in this case is used to measure the deviation of the optimal portfolio return relative to the benchmark. Certainly, this goal can be achieved through a better measure based on the ETL rather than the standard deviation.

Other types of strategies can be distinguished. If short positions are not allowed, then we have long only strategies, if short positions are allowed we have long-short and as an extreme case of these appear the zero-dollar strategies in which the dollar value of the long part exactly equals the dollar value of the short part. The zero-dollar strategies generally give rise to more complicated problems from mathematical viewpoint. One of the points in having short positions in the portfolio is to gain return in both bearish and bullish markets.

All problems considered can be extended with transaction cost or turnover constraints. We do not consider them for brevity of exposition and clarity of the optimization problem formulations. Also, all strategies are considered self-financing in the sense that inflows and outflows of capital are not allowed.

4.8.1 Long-only active strategy

The traditional equity portfolio is constructed and managed relative to an underlying benchmark and aims at achieving active return while minimizing the active risk. Typically the portfolio construction process in such cases suffers from two main disadvantages - employment of tracking error as a measure of the active risk which penalizes upside deviations from the market and the lack of consideration of absolute risk which can play dramatic effect in case of market crashes. For example, in periods when the risk of the benchmark increases, the risk of the optimal solution will increase as well since there is no absolute measure of it in the problem.

The traditional definition of the problem is:

$$\begin{array}{rl} \min\limits_{w} & \text{var}(w'r - r_b) \\ \text{subject to} & w'e = 1 \\ & w'Er - Er_b \geq R \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.25}$$

where var stands for variance; i.e. in the objective there is the active variance. If we change the active variance for the tracking error, the optimal solution will not change. Problem (4.25) is easier to solve because it reduces to a quadratic problem for which there are efficient solvers. In (4.25), and in all problems considered below, the double inequalities $Lb \leq Aw \leq Ub$ in matrix form are linear constraints on the weights and generalize all possible linear constraints that can be imposed, such as constraints by industry, or simple box-type constraints either in absolute terms or relative to the weight of the corresponding stocks in the benchmark. Actually, the relative expected return constraint and the constraint that all weights sum up to 1, $w'e = 1$, can also be generalized in this fashion but we give them separately as they have special meaning. The relative expected return constraint is of practical importance and the constraint $w'e = 1$ guarantees that the strategy is self-financing; i.e. in absence of transaction costs, the optimal portfolio present value equals the present value of the initial portfolio.

A modification of (4.25) based on the ETL can be suggested in which an additional absolute risk constraint is added. One possible version is:

$$\begin{array}{rl} \min\limits_{w} & ETL_{\beta}(w'r - r_b) \\ \text{subject to} & w'e = 1 \\ & ETL_{\alpha}(w'r) \leq ETL^{*} \\ & w'Er - Er_b \geq R \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.26}$$

where the ETL in the objective is at the confidence level $1 - \beta$ which is relative to the benchmark portfolio returns r_b. The ETL in the constraint set is in absolute terms and its confidence level is $1 - \alpha$ which may differ

form $1-\beta$. The upper bound ETL^* is selected by the portfolio manager. It has a direct interpretation due to the ETL being the average loss beyond the corresponding VaR level, see the ETL definition in the previous chapter.

Suppose that all the constraints and the parameters in an optimization problem have been selected, including the risk measure, in this case the ETL and its confidence level, the additional weight constraints and, finally, the multivariate model. Then the next step in the analysis is to run back-testing calculations in order to see what we would have gained in terms of risk-adjusted return, had we followed the strategy in the previous 1, 3 or 5 years for instance. In this way, we can compare several strategies and decide on the best one. Problem (4.26) has the disadvantage that it may become infeasible for some periods in a back-testing calculation because of the absolute risk constraint or the relative expected return constraint. One reason is that for some periods the upper bound ETL^* may be too low and there may not be a single portfolio satisfying it. An alternative formulation of (4.26) can be more advantageous:

$$\begin{array}{rl} \min\limits_{w} & \lambda_1 ETL_\beta(w'r - r_b) + \lambda_2 ETL_\alpha(w'r) - w'Er \\ \text{subject to} & w'e = 1 \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.27}$$

In (4.27), the objective has a more complicated form. It can be viewed as a utility function composed of three components with two positive risk-aversion coefficients, λ_1 and λ_2, which signify the relative importance of the three components. Note that, due to the linearity of the expectation, the expected return of the benchmark can be ignored in (4.27). The two aversion coefficients are free parameters in the problem and can be chosen by the portfolio manager. Alternatively, they can be calibrated by a sequence of back-testing calculations.

The main advantages of (4.26) or (4.27) to the traditional problem (4.25) are:

- they allow for a non-normal distributional assumption
- the risk measure used penalizes only the downside deviations from the market
- the absolute risk of the optimal portfolio can be controlled

In practice, (4.26) and (4.27) can be solved using the linearization procedure described in ... after we have produced scenarios from the fitted parametric model. The type of the parametric model does not influence the structure of the optimization problem; the technique is one and the same on condition that scenarios are provided.

4.8.2 Long-short strategy

Long/short hedge funds focus on security selection and aim at achieving absolute returns keeping small market risk exposure by offsetting short and long positions. The increased flexibility introduced by short positions allows reducing dependence with the market and allows taking advantage of overvalued as well as undervalued securities.

Suppose that the portfolio manager knows which stocks will be long (winners) and which will be short (losers). The only remaining problem is their relative proportions. In this case, both problems (4.26) and (4.27) can be used without modification. The set $Lb \leq Aw \leq Ub$ can be defined to account for portfolio manager's decision. For example, the box constraints of all short stocks are of the type $a\% \leq w_i \leq 0$ ad those of the long stocks, $0 \leq w_i \leq b\%$. The constraint $w'e = 1$ means that the weights are interpreted as a percentage of the optimal portfolio present value, which, in absence of transaction costs, equals the present value of the initial portfolio as the strategy is by construction self-financing. Therefore this problem does not control the ratio of the exposure of the long and the short legs. A constraint of this type can easily be implemented by including

$$\sum_{i \in S} w_i \leq \sum_{i \in L} w_i$$

where L and S are the sets of the long and short stocks respectively.

If the manager does not know which stocks to short, then the problem becomes more involved. Basically, there are two options. On one hand, the manager can screen the stocks prior to the optimization and decide on the winners and the losers and after that proceed to the optimization. The screening can be done by adopting a relevant criterion which could be a performance measure. For example, the STARR ratio and the R-ratio are completely compatible with problems (4.26) or (4.27) as they involve the ETL as a risk measure; see the previous section for more details. In summary, we suggest a two-step procedure:

a) rank the stocks according to a performance measure and choose the winners and the losers

b) solve the optimization problem to get the optimal weights

On the other hand, the portfolio manager may decide to use a one-step process. In this case, binary variables can be employed since the sign of the weights is not known beforehand. The optimization problem becomes numerically much more involved than the two-step alternative but provides the true optimal solution. Such problems are solved with mixed-integer programming (MIP) methods.

In all problems considered above, there is a numerical stability issue. When the long/short ratio approaches one, i.e. the zero-dollar problem, the optimization problem becomes numerically unstable because the weights are defined as a percentage of the present value which becomes zero at the limit. Therefore, the zero-dollar problem is a singularity. To avoid this, the weights can be redefined as a percentage of the total exposure instead; that is the value of the long leg minus the value of the short leg. In effect, the constraint $w'e = 1$ should be replaced.

4.8.3 Zero-dollar strategy

The zero-dollar strategy is a limit case of the long/short strategy in which the value of the long leg equals in absolute terms the value of the short leg, or, alternatively, the total exposure of the portfolio is split in halves between the short and the long part. The optimization problem becomes more involved because the intuitive formulation in (4.26) or (4.27) should be modified since the weights cannot be defined as a percentage of the present value which is equal to zero. The alternative of (4.27) is, for example:

$$\begin{array}{rl} \min\limits_{w} & \lambda_1 ETL_\beta(w'r - r_b) + \lambda_2 ETL_\alpha(w'r) - w'Er \\ \text{subject to} & \sum\limits_{i \in L} w = 1 \\ & \sum_{i \in S} w = -1 \\ & Lb \leq Aw \leq Ub \end{array} \tag{4.28}$$

where S is the set of the short stocks and L is the set of the long stocks. The two new weight constraints mean that they are defined as a percentage of the value of the long part. Thus, the sum of all weights is zero in contrast to the formulation in (4.27). Here we face the same problems as in the long/short case. Either we can use the two-step procedure outlined there, or a one-step process involving MIP.

The same trick of changing the weight definition works with (4.25) and (4.26) as well. Actually, if the weights are defined as a percentage of the total exposure, then the sum of the positive weights equals $1/2$ and the sum of the negative weights to $-1/2$.

4.8.4 Other aspects

Apart from being long or long/short, there are other important features of a strategy. One aspect is if it is correlated with an industry represented by an index. It is very often the case that the portfolio manager demands that the optimal portfolio should not be correlated with one or several indexes. Similar requirements can be implemented in the optimization problems considered in the paper by including additional linear weight constraints.

Suppose that the returns of the stocks can be represented by a multifactor model in which the explanatory variables are the returns of the indexes:

$$r_i = a_i + \sum_{j=1}^{K} b_j^i r_j^I + \epsilon_i, \quad i = 1, \ldots, N$$

where r_j^I are the returns of the corresponding indexes, a_i and b_j^i are the regression coefficients and ϵ_i are the residuals. The returns of a portfolio of these stocks can be expressed in a similar way,

$$w'r = w'a + \sum_{j=1}^{K} w'b_j r_j^I + e_i.$$

Thus, in the optimization problem, we have to include additional constraints involving the coefficients in the above equation since an optimal portfolio with $w'b_j = 0$ will be uncorrelated with the index j.

Certainly the reasoning relies on whether the multifactor model is realistic. If it is not, or the regression coefficients are not properly estimated, then the optimal solution in reality will not be uncorrelated with the corresponding indexes.

4.9 Optimization Example

In this section, we provide a back-testing example of a long-only optimal portfolio strategy using the Russell 2000 universe. The back-testing time period is ten years — from December 1993 to December 2004 with monthly frequency. In the optimization algorithm, we use the proprietary stable model in Cognity Risk & Portfolio Optimization System. In the strategies, the Russell 2000 index is used as the benchmark; that is r_b is the return of Russell 2000.

The optimization constraints are the following.

- 0% to 3% limit on single stock
- $\pm$3% industry exposure with respect to the benchmark; the industries being defined by Ford Equity Research
- The active return is strictly positive
- The two-way turnover is below 15% per month. This constraint is used as a soft constraint; i.e. may be exceeded at times. Also, no limit is imposed in July because the benchmark is adjusted in July.

The back-testing is performed in the following way. We use 450 stocks as initial universe. One year of daily data is used to calibration the model

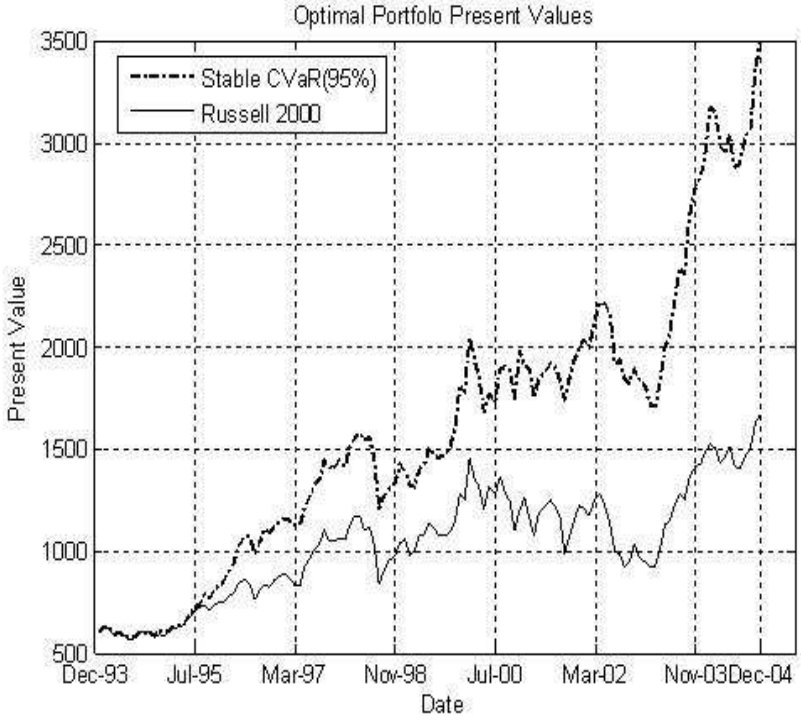

Figure 4.4: The time evolution of the present values of the stable ETL portfolio compared to the Russell 2000 index.

and monthly scenarios are produced by it. Then a version of the optimal portfolio problem (8) is solved in which a tail probability of 5% is selected for the ETL. At the end of the month, the portfolio PV is calculated. The process is repeated next month. Figure 4.4 shows the stable ETL portfolio PV compared to the Russell 2000 index.

Additional information is given in Tables 1 and 2. The average monthly turnover is defined as the dollar weighted purchases plus the dollar weighted sales. Tables 3 and 4 provide details on return-risk ratios. The information ratio is the active return per unit of tracking error.

On the basis of the charts and the information in the tables, it is obvious that both the optimization problem type and the multivariate distributional assumption are vital for the optimal portfolio performance and should be considered.

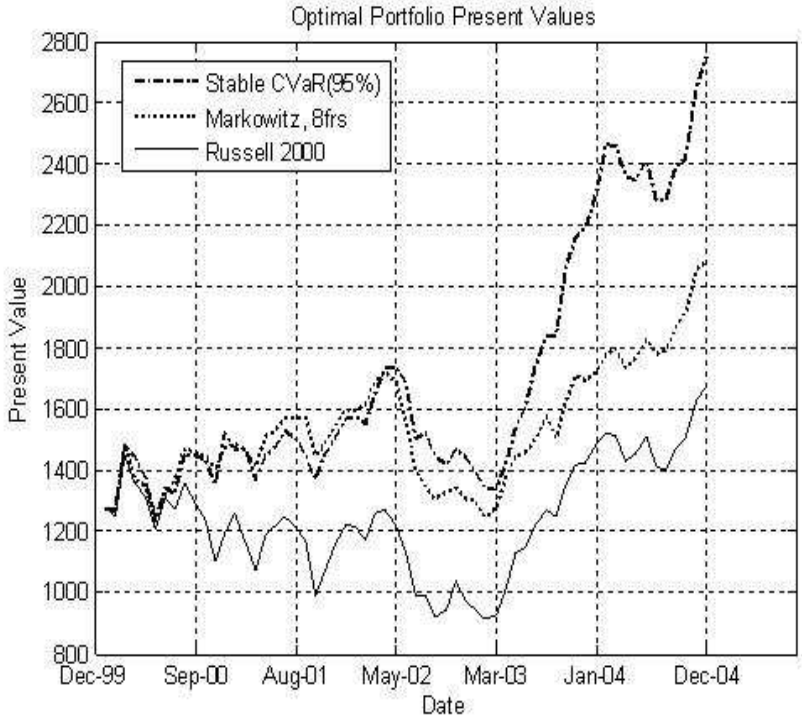

Figure 4.5: The time evolution of the present values of the Markowitz and the stable ETL (scaled) portfolios compared to the Russell 2000 index

	Stable ETL	Markowitz
10 year	112	
5 year	105	137
3 year	102	110
2 year	100	104
1 year	104	100

Table 4.1: Average holdings count

	Stable ETL	Markowitz
11 months	16%	18%
July	163%	85%
All months	27%	24%

Table 4.2: Average monthly turnover

	Stable ETL	Markowitz
11 months	16%	18%
July	163%	85%
All months	27%	24%

Table 4.3: Average monthly turnover

	Stable ETL	Markowitz
10 year	0.74	
5 year	0.71	0.29
3 year	0.93	-0.24
2 year	0.74	-0.57
1 year	1.22	1.03

Table 4.4: Annualized information ratio

	Stable ETL	Markowitz	Russell 2000
10 year	1.01		0.42
5 year	0.92	0.68	0.36
3 year	1.22	0.71	0.58
2 year	2.13	1.99	1.82
1 year	1.66	2.16	1.19

Table 4.5: Sharpe ratios

Return	Stable ETL	Markowitz	Russell 2000
10 year	17.90%		11.20%
5 year	18.10%	11.10%	7.90%
3 year	20.30%	9.90%	21.10%
2 year	34.00%	24.30%	28.60%
1 year	24.10%	21.50%	17.10%
Volatility	Stable ETL	Markowitz	Russell 2000
10 year	17.70%		26.40%
5 year	19.60%	16.50%	21.80%
3 year	16.50%	14.00%	21.10%
2 year	15.90%	12.20%	15.70%
1 year	14.50%	10.00%	14.40%

Table 4.6: Annualized return and volatility

Active Return	Stable ETL	Markowitz
10 year	6.70%	
5 year	7.80%	3.30%
3 year	8.10%	-2.20%
2 year	5.30%	-4.30%
1 year	7.00%	6.40%
Tracking-error	Stable ETL	Markowitz
10 year	9.00%	
5 year	11.10%	11.40%
3 year	8.70%	9.50%
2 year	7.20%	7.70%
1 year	5.70%	6.30%

Table 4.7: Annualized active return and annualized tracking-error

Conclusions

In this thesis, we presented a unified Monte-Carlo based framework for market and credit risk estimation and portfolio optimization. The framework is based on heavy-tailed, skewed distributions for modeling risk factor returns and downside risk measures. As far as modeling stand-alone variables is concerned, we considered the class of stable distributions and the class of Skewed Student's t distributions. For capturing the dependence structure between variables, we considered a copula approach.

The system can work with or without a predefined factor model. When the risk variables are highly correlated, it is more desirable to capture the dependencies first via a factor model and then apply the framework on the factor returns and the residuals. As a simple example, we can consider a yield curve. The interest rates in one yield curve are extremely dependent and in this case, it is desirable to take advantage of some factor model such as the statistical factor model behind the principal components analysis.

The described system is a forward looking tool as it is based on the Monte Carlo method. All risk statistics are computed on the basis of the generated scenarios from the fitted multivariate model. The downside risk measure we considered is the Expected tail loss, which apart from the straightforward practical interpretation has appealing theoretical properties. As a consequence, we can incorporate it with the heavy-tailed framework and build risk budgeting tools and forward-looking portfolio optimization tools.

The recent turbulent events in September and October 2008 represent a clear evidence that measuring and managing properly risk is a complicated task. This task has to be performed looking at the portfolio from different angles with an arsenal of different tools. Any myopic, one-sided approach is doomed. For this purpose, practitioners need an integrated system based on a common framework built upon realistic assumptions. Only in such environment can practitioners make their decisions better informed, achieving an improved protection against adverse market movements.

The empirical examples presented here and published in a number of papers support this concept. A more comprehensive study is forthcoming in a book to be published by Wiley, Racheva-Iotova et al. (2009).

List of Figures

List of Tables

Bibliography

Acerbi, C. and D. Tasche (2002), 'On the coherence of expected shortfall', *Journal of Banking and Finance* **26, (7)**, 1487–1503.

Acerbi, C. and P. Simonetti (2002), 'Portfolio optimization with spectral measures of risk', *working paper, Abaxbank, Milano* .

Adler, R. J., R. Feldman and M. Taqqu (1998), *A practical guide to heavy tails: Statistical techniques and applications*, Birkhäuser.

Akgiray, A. and G. Booth (1988), 'The stable-law model of stock returns', *Journal of Business and Economic Statistics* **6**, 51–57.

Akgiray, A. and G. Lamoureux (1989), 'Estimation of stable-law parameters: A comparative study', *Journal of Business and Economic Statistics* **7**, 85–93.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1998), 'Coherent measures of risk', *Math. Fin.* **6**, 203–228.

Biglova, A., S. Ortobelli, S. Rachev and S. Stoyanov (2004), 'Different approaches to risk estimation in portfolio theory', *The Journal of Portfolio Management* **Fall 2004**, 103–112.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedacity', *Journal of Econometrics* **31**, 307–327.

Box, G. and G. Jenkins (1976), *Time series analysis: forecasting and control, 2ed*, San Francisco: Holden-Day.

Brokwell, P. and R. Davis (1991), *Time series: theory and methods, 2ed*, New York: Springer.

Cowell, F. and B. Racheva-Iotova ans S. Trueck (2008), 'Recent advances in credit risk management', *Risk Assessment: Decisions in Banking and Finance, G. Bol et al.(edt.), Physika, Springer, to appear* .

DeGiorgi, E. (2005), 'Reward-risk portfolio selection and stochastic dominance', *Journal of Banking and Finance* **29, (4)**, 895–926.

Dokov, S., S. Stoyanov and S. Rachev (2008), 'Computing var and avar of skewed-t distribution', *Journal of Applied Functional Analysis* **3, (2)**, 189–207.

D'Sourza, D., K. Amir-Afeti and B. Racheva-Iotova (2002), 'Valuation of a credit default swap: The stable non-gaussian versus the gaussian approach', *Proceedings of the 8-th Econometric Conference on Credit Risk Management, March 13-15-th* .

DuMouchel, W.H. (1973), 'On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution', *Annals of Statistics* **1**, 948–957.

Embrechts, P., A. McNeil and D. Straumann (2002), 'Correlation and dependence in risk management: properties and pitfalls', *In Risk management: Value at risk and beyond, edited by M. Dempster, published by Cambridge University Press, Cambridge* **10, (3)**, 341–352.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997), *Modeling extremal events for insurance and finance*, Springer.

Embrechts, P., F. Lindskog and A. McNeil (2003), 'Modelling dependence with copulas and applications to risk management', *In: Handbook of heavy-tailed distributions in finance, ed. S. Rachev* pp. 329–384.

Engle, R. (1982), 'Autoregressive conditional heteroskedacity with estimates of the variance of u.k. inflation', *Econometrica* **50**, 987–1008.

Fabozzi, F., S. Stoyanov and B. Racheva-Iotova (2006), 'An empirical examination of the return distribution characteristics of agency mortgage pass-through oprimization', *Applied Financial Economics* **16(15)**, 1085–1094.

Fama, E. (1963), 'Mandelbrot and the stable paretian hypothesis', *Journal of Business* **36**, 420–429.

Fama, E. (1965), 'The behavior of stock market prices', *Journal of Business* **38**, 34–105.

Fama, E. and K. French (1988), 'Permanent and temporary components of stock prices', *Journal of Political Economy* **96**, 246–273.

Fama, E. and R. Roll (1968), 'Some properties of symmetric stable distributions', *Journal of the American Statistical Association* **63**, 817–836.

Fama, E. and R. Roll (1971), 'Parameter estimates for symmetric stable distributions', *Journal of the American Statistical Association* **66**, 331–338.

Fang, K.-T., S. Kotz and K.-W. Ng (1994), *Symmetric multivariate and related distributions*, Marcel Dekker, New York.

Föllmer, H. and A. Schied (2002*a*), 'Robust preferences and convex risk measures', *Advances in Finance and Stochastics, Essays in Honor of Dieter Sondermann, Eds. K. Sandmann, Ph. Schonbucher. Berlin, Heidelberg, New York: Springer.* pp. 39–56.

Föllmer, H. and A. Schied (2002*b*), *Stochastic finance: An introduction in discrete time, second revised and extended revision*, Walter de Gruyter, Berlin.

Frahm, G. (2004), *Generalized elliptical distributions: Theory and applications*, PhD thesis, University of Köln.

French, K., G. Schwert and R. Stambaugh (1989), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 3–29.

Hamilton, J. (1994), *Time series analysis*, Princeton University Press.

Holton, Glyn A. (2004), 'Defining risk', *Financial Analysts Journal* **60, (6)**, 19–25.

Janicki, A. and A. Weron (1994), *Simulation and chaotic behavior of alpha-stable stochastic processes*, Marcel Dekker, New York.

Khindanova, I. and S. Rachev (2000), 'Value-at-risk: recent advances', *Journal of Risk Analysis* **2**, 45–76.

Koutrouvelis, I.A. (1980), 'Regression-type estimation of the parameters of stable laws', *Journal of the American Statistical Association* **7**, 918–928.

Koutrouvelis, I.A. (1981), 'An iterative procedure for the estimation of the parameters of stable laws', *Communications in Statistics. Simulation and Computation* **10**, 17–28.

Kozubowski, T., A. Panorska and S. Rachev (2003), 'Statistical issues in modeling multivariate stable portfolios', *In: Handbook of heavy-tailed distributions in finance, ed. S. Rachev* pp. 131–167.

Lehmer, D.H. (1951), 'Mathematical methods in large-scale computing units', *Proceedings of the Second Symposium on Large Scale Digital Computing Machinery* pp. 141–146.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *Journal of Business* **26**, 394–419.

Markowitz, H. M. (1952), 'Portfolio selection', *Journal of Finance* **7, (1)**, 77–91.

Markowitz, H. M. (1959), *Portfolio selection: Efficient diversification of investments*, John Wiley & Sons, New York.

Matsumoto, M. and T. Nishimura (1998), 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator', *ACM Transactions on Modeling and Computer Simulation* **8(1)**, 3–30.

McCulloch, J.H. (1986), 'Simple consistent estimators of stable distribution parameters', *Communications in Statistics. Simulation and Computation* **15**, 1109–1136.

Meerschaert, M. and H. Scheffler (2001), *Limit distributions for sums of independent random vectors: Heavy tails in theory and practice*, Wiley, New York.

Meerschaert, M. and H. Scheffler (2003), 'Portfolio modeling with heavy-tailed random vectors', *In: Handbook of heavy-tailed distributions in finance, ed. S. Rachev* pp. 595–640.

Mittnik, S., M. Paolella and S. Rachev (2002), 'Stationarity of stable power-garch processes', *Journal of Econometrics* **106**, 97–107.

Mittnik, S., S. Rachev and M. Paolella (1996), 'Integrated stable garch processes', *Technical report, Institute of Statistics and Econometrics, Christian Albrechts University at Kiel* .

Nelsen, R. (1999), *An introduction to copulas*, Springer, New York.

Neumann, J. Von and O. Morgenstern (1953), *Theory of games and economic behavior*, Princeton University Press, Princeton, N.J.

Ogryczak, W. and A. Ruszczynski (2001), 'On consistency of stochastic dominance and mean-semideviation models', *Mathematical Programming* **89**, 217–232.

Ortobelli, S., A. Biglova, I. Huber, B. Racheva-Iotova and S. Stoyanov (2005), 'Portfolio choice with heavy-tailed distributions', *Journal of Concrete and Applicable mathematics* **3, (3)**, 353–376.

Ortobelli, S., I. Huber and S. Rachev (2003), 'Portfolio choice theory with non-gaussian distributed returns', *In: Handbook of heavy-tailed distributions in finance, ed. S. Rachev* pp. 547–594.

Ortobelli, S. and S. Rachev (2001), 'Safety-first analysis and stable paretian approach to portfolio choice theory', *Mathematical and Computer Modelling* **34**, 1037–1072.

Ortobelli, S., S. Rachev, I. Huber and A. Biglova (2004), 'Optimal portfolio selection and risk management: A comparison between the stable paretian approach and the gaussian one', *In: Handbook of computational and numerical methods in finance, ed. S. Rachev* pp. 197–252.

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

Palmquist, J., S. Uryasev and P. Krokhmal (2002), 'Portfolio optimization with conditional value-at-risk objective and constraints', *Journal of Risk* **4, (2)**, 43–68.

Panorska, A., S. Mittnik and S. Rachev (1995), 'Stable garch models for financial time series', *Applied Mathematics Letters* **815**, 33–37.

Park, S. K. and K. W. Miller (1988), 'Random number generators: Good ones are hard to find', *Communications of the ACM* **31(10)**, 1192–1201.

Pflug, C. (1998), 'How to measure risk?', *working paper* .

Press, S.J. (1972), 'Estimation in univariate and multivariate stable distribution', *Journal of the American Statistical Association* **67**, 842–846.

Rachev, S. (2003), *Handbook of heavy-tailed distributions in finance*, editor, Rachev, Elsevier/North-Holland.

Rachev, S. (2004), *Handbook of computational and numerical methods in finance*, editor, Rachev, Birkhäuser.

Rachev, S., B. Racheva-Iotova, S. Stoyanov and F. Fabozzi (2008), 'Risk management and portfolio optimization for volatile markets', *to appear in a Springer volume dedicated to Harry Markowitz* .

Rachev, S., D. Martin, B. Racheva-Iotova and S. Stoyanov (2008), 'Stable etl optimal portfolios and extreme risk management', *to appear in Risk Assessment: Decisions in Banking and Finance, G.BoL (edt.) Physika, Springer* .

Rachev, S., F. Fabozzi and C. Menn (2005), *Fat-tails and skewed asset return distributions*, Wiley, Finance.

Rachev, S. and L. Rüschendorf (1998), *Mass transportation problems, V1, V2*, Springer-Verlag, NY.

Rachev, S., T. Jasic, A. Biglova and F. Fabozzi (2005), 'Risk and return in momentum strategies: Profitability from portfolios based on risk-adjusted stock ranking criteria', *Technical report, Department of probability and applied statistics, University of California, Santa Barbara.* .

Rachev, S. T., Stoyan V. Stoyanov and F. J. Fabozzi (2008), *Advanced stochastic models, risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures*, Wiley, Finance.

Rachev, S.T. and S. Mittnik (2000), *Stable Paretian Models in Finance*, John Wiley & Sons, Series in Financial Economics.

Racheva-Iotova, B. and G. Samorodnitsky (2003), 'Long range dependence in heavy tailed stochastic processes', *Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance, chapter 16, Elsevier/North Holland* **1**, 641–662).

Racheva-Iotova, B., S. Rachev, S. Stoyanov and F. Fabozzi (2009), *Advanced Risk Management and Portfolio Optimization*, to be published by John Wiley & Sons.

Racheva-Iotova, B. and S. Stoyanov (2008), 'Post-modern approaches for portfolio optimization', *Handbook on Information Technology in Finance, D. Seese at al. Editors, Springer* pp. 613–633.

Racheva-Iotova, B., S. Stoyanov and S. Rachev (2003), 'Stable non-gaussian credit risk model; the cognity approach', *Bol et al (Eds), Credit Risk – Measurement, Evaluation and Management* pp. 175–193.

Rockafellar, R. T. and S. Uryasev (2002), 'Conditional value-at-risk for general loss distributions', *Journal of Banking and Finance* **26, (7)**, 1443–1471.

Rothschild, M. and J. E. Stiglitz (1970), 'Increasing risk: A definition', *Journal of Economic Theory* **2**, 225–243.

Samorodnitsky, G. and M.S. Taqqu (1994), *Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance*, Chapman & Hall, New York, London.

Sharpe, W. F. (1966), 'Mutual funds performance', *Journal of Business* **January**, 119–138.

Sharpe, W. F. (1994), 'The sharpe ratio', *Journal of Portfolio Management* **Fall**, 45–58.

Sklar, A. (1996), 'Random variables, distribution functions and copulas — a personal look backward and forward.', *In: Rüschendorf, L., Schweizer, B., Taylor, M. D. (Eds.), Distributions with fixed marginals and related topics. Institute of Mathematical Statistics, Hayward, CA* pp. 1–14.

Stoyanov, S. (2005), *Optimal financial portfolios in highly volatile markets*, PhD Thesis, University of Karlsruhe.

Stoyanov, S. and B. Racheva-Iotova (2004*a*), 'Numerical methods for stable modeling in financial risk management', *Handbook of Computational and Numerical Methods in Finance, ed. S. Rachev, Birkhauser* pp. 299–329.

Stoyanov, S. and B. Racheva-Iotova (2004*b*), 'Univariate stable laws in the field of finance — approximations of density and distribution functions', *Journal of Concrete and Applicable Mathematics* **2, (1)**, 38–57.

Stoyanov, S. and B. Racheva-Iotova (2004*c*), 'Univariate stable laws in the field of finance — parameter estimation', *Journal of Concrete and Applicable Mathematics* **2, (4)**, 24–49.

Stoyanov, S., B. Racheva-Iotova, S. Rachev, S. Ortobelli and F. Fabozzi (2008), 'Stochastic models for risk estimation in volatile markets: A survey', *forthcoming in Annals of Operations Research* .

Stoyanov, S. and S. Rachev (2007*a*), 'Asymptotic distribution of the sample average value-at-risk', *forthcoming in Journal of Computational Analysis and Applications* .

Stoyanov, S. and S. Rachev (2007*b*), 'Asymptotic distribution of the sample average value-at-risk in the case of heavy-tailed returns', *forthcoming in Journal of Applied Functional Analysis* .

Tobin, J. (1958), 'Liquidity preference as behavior toward risk', *Review of Economic Studies* **25**, 65–86.

Yamai, Y. and T. Yoshiba (2002*a*), 'Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition and optimization', *Monetary and Economic Studies, Institute for Monetary and Economic Studies, Bank of Japan* **20, (1)**, 87–122.

Yamai, Y. and T. Yoshiba (2002*b*), 'On the validity of value-at-risk: comparative analyses with expected shortfall', *Monetary and Economic Studies, Institute for Monetary and Economic Studies, Bank of Japan* **20, (1)**, 57–86.

Zhang, Y. and S. Rachev (2004), 'Risk attribution and portfolio performance measurement — an overview', *Technical report, Department of Probability and Applied Statistics, University of California, Santa Barbara, USA* .

Zolotarev, V. M. (1986), *One-dimensional stable distributions (Translation of mathematical monographs, Vol 65)*, American Mathematical Society.

Index