Measuring the Risk of Large Losses

Kay Giesecke[*] Thorsten Schmidt[†] Stefan Weber[‡]

Stanford University *Universität Leipzig* *Cornell University*

August 5, 2005; this version January 15, 2008[§]

Abstract

Risk management is an important component of the investment process. It requires quantitative measures of risk that provide a metric for the comparison of financial positions. In this expository note we give an overview of risk measures. In particular, we contrast different risk measures with respect to their sensitivity to potentially large losses due to market wide shocks. The industry standard value at risk exhibits many deficiencies. It does not account for the size of the losses and may penalize diversification. We compare value at risk to alternative risk measures including average value at risk and the less well know but superior utility-based shortfall risk.

Key words: Distribution-invariant risk measures, utility-based shortfall risk, average value at risk, value at risk, event risk, extreme events, extreme value distributions

[*]Management Science and Engineering, Terman Engineering Center, 380 Panama Way, Stanford University, Stanford, California 94305-4026, USA, phone: +1 650-723-9265, email giesecke@stanford.edu.

[†]Department of Mathematics, Universität Leipzig, 04081 Leipzig, Germany, phone: +49 341 9732141, email Thorsten.Schmidt@math.uni-leipzig.de.

[‡]School of Operations Research and Information Engineering, Cornell University, 279 Rhodes Hall, Ithaca, NY 14853-3801, USA, phone: +1 607 254-4825, email sweber@orie.cornell.edu.

[§]We are extremely grateful for useful comments of an anonymous referee.

1 Introduction

Risk management is an important component of the investment process. It requires quantitative measures of risk that provide a metric for the comparison of financial positions. In this expository note we give an overview of risk measures. In particular, we contrast different risk measures with respect to their sensitivity to potentially large losses due to market wide shocks. Sources of such shocks include severe currency devaluations, credit default clusters, liquidity crises, market crashes, natural disasters and terrorist attacks.

The distribution of profit and loss is very complex. It contains an infinite amount of information by specifying the probability of a profit or loss of any given size. A meaningful comparison of distributions requires therefore to focus on specific properties of distributions, e.g. their risk. While distributions are sophisticated objects, risk measures, in contrast, constitute an attempt to summarize the risk of a profit and loss distribution in a single number. The goal is to reduce the complexity of the problem of risk management. Apparently, a single number provides only very limited information about a profit and loss distribution. Thus, risk measures need to be designed in such a way that they capture the relevant features of distributions. The choice of risk measures with desirable properties will be discussed in this article.

Risk management concerns the *lower tail* of this distribution – events that cause excessive losses. Therefore, we are led to focus on risk measures that summarize the features of the lower tail of the profit and loss distribution. Besides being sensitive to excessive losses, a good risk measure should have some more virtues. First, it should measure risk on a monetary scale: the notion of risk entails the amount of capital we need to set aside in order to make a position acceptable from a risk management perspective. Second, a risk measure should penalize concentrations and encourage diversification. Third, a risk measure should support dynamically consistent risk measurements over multiple horizons.

While these requirements are very natural, many risk measures that are implemented throughout the industry and sanctioned by the supervisory authorities fall short of some of them. A case in point is the ubiquitous value at risk, which does not account for the size of the losses exceeding the value at risk. Equally bothersome, value at risk may even penalize diversification. Average value at risk, which is also known as expected shortfall, does somewhat better – however not perfectly. Less well known but superior is utility-based shortfall risk.

Utility-based shortfall risk is specified by a loss function ℓ and a loss threshold. If X is the value of a position at a future horizon, then utility-based shortfall risk is the smallest amount of cash m such that the expected value of the weighted shortfall $\ell(-(X+m))$ does not exceed the loss threshold. Compare this definition with value at risk, which is the smallest amount of cash m such that the probability of $X+m$ falling below 0 does not exceed a threshold. Both utility-based shortfall risk and value at risk can be seen as a capital requirement. The well-known methods for estimating and simulating value at risk can also be applied to utility-based shortfall risk.

The remainder of this note is organized as follows. Section 2 reviews the definition of basic risk measures. Section 3 considers utility-based shortfall risk. It shows that utility-based shortfall risk can be very sensitive to event risk, depending on the choice of the loss function ℓ. Section 4 reviews some extreme value distributions. Section 5 numerically compares value at risk, average value at risk and utility-based shortfall risk when positions follow extreme value distributions.

2 Risk Measures

Consider a financial position such as a bond, a stock or a portfolio of securities. The value of the position at a future horizon, or the change of this value relative to the current value, is uncertain and described by a random variable X.[1] A risk measure assigns to X a number that expresses the risk of the position. *Value at risk* is one such measure. It is the smallest amount of cash that must be added to X such that the probability of a loss does not exceed $\lambda \in (0,1)$. In mathematical terms,

$$\mathrm{VaR}_\lambda(X) = \inf \big\{ m \in \mathbb{R} : \; P[m + X < 0] \le \lambda \big\}.$$

Unfortunately, value at risk has several deficiencies. It does not take account of the size of the losses that exceed the value at risk. Further, it does not always encourage diversification. The following examples illustrate these points.

Example 2.1. Consider two positions X_1 and X_2 given by

$$X_1 = \begin{cases} 1 & \text{with probability } 99\% \\ -1 & \text{with probability } 1\% \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{with probability } 99\% \\ -10^{10} & \text{with probability } 1\% \end{cases}$$

Both positions have a 1% value at risk equal to -1. While X_2 has much higher downside risk than X_1, the value at risk does not distinguish these two positions.

Example 2.2. Consider the two positions X_1 and X_2 given by

$$X_i = \begin{cases} 1 & \text{with probability } 50\% \\ -1 & \text{with probability } 50\% \end{cases}$$

Here $\mathrm{VaR}_{50\%}(X_i) = -1$. If X_1 and X_2 are independent, then we can calculate the value at risk of the diversified position X that is given by

$$X = \frac{X_1 + X_2}{2} = \begin{cases} 1 & \text{with probability } 25\% \\ 0 & \text{with probability } 50\% \\ -1 & \text{with probability } 25\% \end{cases}$$

[1]We fix an atomless probability space $(\Omega, \mathcal{F}, P)$ that describes the uncertainty of investors.

We get $\text{VaR}_{50\%}(X) = 0$. Thus, the value at risk suggests the diversified position X is riskier than any individual position. In other words, value at risk does not account for diversification effects.

The examples point to several properties that a reasonable risk measure should have: it should quantify risk on a monetary scale, be sensitive to large losses, and encourage diversification. This normative reasoning leads to an axiomatic approach developed by Artzner, Delbaen, Eber & Heath (1999), in which a risk measure is defined as a function on a space D of positions that has certain properties.[2]

Definition 2.3. A real-valued function M on the position space D is called a distribution-invariant *risk measure* if it satisfies the following conditions for all $X_1, X_2 \in D$:

- *Inverse Monotonicity:* If $X_1 \leq X_2$, then $M(X_1) \geq M(X_2)$.
- *Translation Property:* If $m \in \mathbb{R}$, then $M(X_1 + m) = M(X_1) - m$.
- *Distribution Invariance:* If X_1 and X_2 have the same probability distribution, then $M(X_1) = M(X_2)$.

The first property says that the risk of a position decreases if its value increases. The second property requires that risk is measured on a monetary scale: if an amount m of cash is added to a position, then the risk of the combined position is reduced by m. The third property says that the risk of a position depends only on the distribution of its value. This property is illustrated in the following example.

Example 2.4. Suppose that the state of the economy is described by the random variable

$$S = \begin{cases} g & \text{with probability } 50\% \\ b & \text{with probability } 50\% \end{cases}$$

g denotes the good state of the economy, and b the bad state.

Consider the two positions X_p and X_a defined by

$$\begin{array}{lll} X_p = 1, & X_a = -1 & \text{if} \quad S = g \\ X_p = -1, & X_a = 1 & \text{if} \quad S = b, \end{array}$$

so X_p behaves procyclical and X_a behaves anticyclical. While X_p and X_a pay different amounts in any given state, they have the same distribution. If M is distribution-invariant, then $M(X_p) = M(X_a)$. In this case risk depends on individual distributions but not on the dependence structure of the position and the rest of the world: individual risk measurements give no information on this dependence.

[2]We assume that D is a vector space of integrable random variables that contains the constants. Examples include the space of bounded positions L^∞ and the space of positions with finite variance L^2.

Value at risk is a distribution-invariant risk measure. However, it does not always encourage diversification as we have seen in Example 2.1 above. In order to discuss alternative risk measures, we consider a set of additional properties.

- *Convexity:* If two positions $X_1, X_2 \in D$ are combined, then the risk of the diversified position $\alpha X_1 + (1-\alpha)X_2$ does not exceed the weighted sum of the individual risks:

  $$M(\alpha X_1 + (1-\alpha)X_2) \leq \alpha M(X_1) + (1-\alpha)M(X_2), \quad \alpha \in [0,1].$$

  This property formalizes the idea that diversification reduces risk.

- *Invariance under randomization:* If two positions $X_1, X_2 \in D$ are acceptable, i.e. if $M(X_1) \leq 0$ and $M(X_2) \leq 0$, then the randomized position X given by[3]

  $$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1-\alpha \end{cases} \qquad \alpha \in (0,1)$$

  is also acceptable. From a normative perspective, the uncertainty associated with randomization should not matter. After tossing a coin, an investor gets either the acceptable X_1 or the acceptable X_2. Thus X should also be accepted. Similarly, if the individual positions are not acceptable with respect to M, if i.e. $M(X_1) > 0$ and $M(X_2) > 0$, then X should also not be acceptable.

- *Positive homogeneity:* If a position $X \in D$ is increased by a positive factor, then the risk increases by the same factor:

  $$M(\lambda X) = \lambda M(X), \quad \lambda \geq 0.$$

  This property is economically less meaningful since it neglects the asymmetry between gains and losses. Increasing the size of a position by a factor λ may increase the risk by a factor larger than λ if the costs of bearing losses grow faster than their size. For example, the larger a position, the more costly it typically becomes to liquidate it. From an investor's perspective, these additional costs should be measured appropriately. From a regulatory perspective, high costs of large losses may lead to instability of the financial system by triggering losses at other institutions.

Example 2.5.

- The main idea behind the notion of *convexity* was already discussed in Example 2.2.

- The intuition behind *invariance under randomization* can be illustrated as follows. Suppose a family plans to buy a car within a week. The family agrees on a specific model, but is indifferent about the color of the car, say wether to buy a red or a blue car. A dealer provides them with the information that he can get the model at

[3]The choice should be made independently of X_1 and X_2.

a good price within five days, but does unfortunately not know which color will be available. In such a situation it makes sense to assume that the family does not care and will find the proposed deal acceptable. Conversely, if the family does neither wish to buy a red nor a blue car, they would not accept to buy a car which is either red or blue but whose color is not known in advance.

- *Positive homogeneity* implies that if a position is acceptable then also any multiple of the position is acceptable. Here, acceptability means that a position has non positive risk. Let us demonstrate that this assumption is usually not reasonable. Suppose that a manager would like to invest in portfolio which might lose $\$1,000,000$ with a probability of 5%. Since the upside is very promising, he is willing to accept this risk. Although unpleasant, a loss of $\$1,000,000$ would not cause a serious problem. Suppose now that the manger considers to invest 1000 times as much in the same investment opportunity. If we assume that the manager's actions do not influence the price or the chances of the outcome, his potential gains would also multiply by the large factor of 1000. However, he might loose \$1 billion with a probability of 5% which might be prohibitively large. A positively homogeneous risk measure would nevertheless classify the position as acceptable.

Value at risk is invariant under randomization and positively homogeneous, but not convex in general.[4] An alternative to value at risk is *average value at risk*, which is also known as expected shortfall. For a level $\lambda \in (0,1)$, it is defined as

$$\mathrm{AVaR}_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda \mathrm{VaR}_\alpha(X)\,d\alpha.$$

Average value at risk can be used as a building block for a much larger class of distribution-invariant risk measures via robust representation, see Föllmer & Schied (2004), Kusuoka (2001) and Kunze (2003). Average value at risk is a distribution-invariant risk measure that is convex and positively homogeneous. However, it is not invariant under randomization. In Section 3 below, we discuss *utility-based shortfall risk*, a less well-known alternative to (average) value at risk. Utility-based shortfall risk has many desirable properties, including convexity and invariance under randomization. If chosen appropriately, it is more sensitive to the risk of large losses than average value at risk as we show in Section 5.

[4]Convexity relies on the domain of the risk measure. If the domain contains the space of bounded positions, then value at risk is *never* convex. If we consider instead a vector space of positions of the *same type*, then value at risk is a convex risk measure on this space, see Embrechts, McNeil & Strautman (2002). Examples include vector spaces of Gaussian, or more general, elliptical distributions. However, it is well-known that these families often fail to adequately describe the profit and loss distribution of financial positions.

3 Utility-based Shortfall Risk

To define utility-based shortfall risk, consider a convex loss function $\ell : \mathbb{R} \to \mathbb{R}$, i.e. a function that is increasing and not constant. Let z be a point in the interior of the range of ℓ. The space of positions D is chosen such that for $X \in D$ the expectation $E[\ell(-X)]$ is well-defined and finite. A position is acceptable if the expected value of $\ell(X)$ does not exceed z. The corresponding set A of positions with non positive risk, or acceptance set, is given by[5]

$$A = \left\{X \in D : \; E[\ell(-X)] \leq z\right\}.$$

The acceptance set A induces the risk measure *utility-based shortfall risk* M as the smallest amount of cash that must be added to the position X to make it acceptable:

$$M(X) = \inf\left\{m \in \mathbb{R} : \; X + m \in A\right\}.$$

We summarize the properties of utility-based shortfall risk.

(1) It is convex and therefore encourages diversification.

(2) It is invariant under randomization. The same is true for value at risk, but *not* for average value at risk, which is *not* a utility-based risk measure. More generally, it can be shown that utility-based shortfall risk measures are essentially the only distribution-invariant convex risk measures that are invariant under randomization. Thus, utility-based shortfall risk measures are the only distribution-invariant convex risk measures that should be used for the dynamic measurement of risk over time, see Weber (2006). This is an important advantage of utility-based shortfall risk.

(3) It is positively homogeneous if and only if the loss function is of the form

$$\ell(x) = z - \alpha x^- + \beta x^+, \qquad \beta \geq \alpha \geq 0$$

where x^- denotes the negative part of x and x^+ denotes the positive part.

The loss function controls the sensitivity of utility-based shortfall risk to large losses.

Example 3.1 (Entropic risk measure)**.** For $\ell(x) = \exp(\alpha x)$ with $\alpha > 0$ we get[6]

$$M(X) = \frac{1}{\alpha}(\log E[\exp(-\alpha X)] - \log z).$$

Example 3.2 (One sided loss function)**.** Examples include $\ell(x) = x^\alpha \cdot \mathbf{1}_{[0,\infty)}(x)$ for $\alpha \geq 1$ and $\ell(x) = \exp(\alpha x) \cdot \mathbf{1}_{[0,\infty)}(x)$ for $\alpha > 0$. Here, $\mathbf{1}_H$ denotes the indicator function of the event H. The associated risk measures focus on downside risk only and thus neglect tradeoffs between gains and losses.

[5] An alternative definition of A starts with the Bernoulli utility function u given by $u(x) = -\ell(-x)$. Then $U(X) = E[u(X)]$ defines a von Neumann-Morgenstern utility. A position X is acceptable if $U(X) \geq -z$. This explains the terminology "utility-based."

[6] In the context of extreme value distributions exponential moments may not exist, see Remark 4.6.

Consider the positions X_1 and X_2 analyzed in Example 2.1. The 1% value at risk does not detect that X_2 has a significantly lower downside than X_1. If M is USBR with loss function $\ell(x) = x^2 \cdot \mathbf{1}_{[0,\infty)}(x)$ and loss threshold $z = 1\%$, we obtain $M(X_1) = 0$ and $M(X_2) = 10^{10} - 1$. This shows that utility-based shortfall risk is sensitive to extreme risks for a suitable loss functions. This property is further investigated in Section 5 below.

Example 3.3. The following example demonstrates that UBSR often provides more useful information about large losses of a portfolio than VaR and AVaR. Consider

$$0 < x_1 < x_2,$$

and let

$$X \quad = \quad \begin{cases} 1 & \text{with probability } 99\% \\ -x_1 & \text{with probability } 0.005\% \\ -x_2 & \text{with probability } 0.005\% \end{cases}$$

If $x_1 = x_2 = 1$, we recover X_1 from Example 2.1. As in Example 2.1,

$$\mathrm{VaR}_{0.01}(X) = -1$$

for any choice of x_1 and x_2. Moreover, it can be shown that

$$\mathrm{AVaR}_{0.01}(X) = \frac{x_1 + x_2}{2}.$$

Average value at risk is thus insensitive to changes of the values of the variables x_1 and x_2 as long as the sum of these remains constant. For example, the average value at risk is the same in either case $x_1 = x_2 = 1$ or $x_1 = 0.3$, $x_2 = 1.7$. The extreme downside risk of the second position is significantly higher than the downside risk of the first position, but this is neither measured by VaR nor AVaR.

The situation is completely different for UBSR. Consider for example $l_1(x) = c \cdot x\mathbf{1}_{\{x \geq 0\}}$ and $l_2(x) = c \cdot x^2 \mathbf{1}_{\{x \geq 0\}}$ as in Example 3.2 with $c = 0.01^{-1}$. For the first loss function we obtain

$$L_1(0) := E(l_1(-X)) = \frac{x_1 + x_2}{2}$$

which is exactly equal to AVaR. The second loss function leads to

$$L_2(0) := E(l_2(-X)) = \frac{x_1^2 + x_2^2}{2}.$$

Note that for $x_1 = x_2 = 1$ we calculate $L_2(0) = 1$, while for $x_1 = 0.3$ and $x_2 = 1.7$ we obtain $L_2(0) = 2.98$ which clearly reflects the larger downside risk.

Apart from these desirable properties, utility-based shortfall risk also has advantages in the implementation. We can adapt the well-known methods for the estimation and simulation of value at risk, see Glasserman (2004). These methods rely on the fact that the $\mathrm{VaR}_\lambda(X)$ of a position $X \in D$ is the λ-*quantile* of the random variable X. The quantile is an inverse to the cumulative distribution function

$$F(x) = P[X \leq x] \qquad (x \in \mathbb{R})$$

of X. Since F may not be injective, we define a right-continuous inverse by the upper quantile function

$$q(\alpha) = \inf\{x : \ P(X \leq x) > \alpha\} = \sup\{x : \ P[X < x] \leq \alpha\} \qquad (\alpha \in (0,1)).$$

The value at risk at level α is given by

$$\mathrm{VaR}_\alpha(X) = -q(\alpha). \tag{1}$$

The quantile function q is essentially the inverse of the distribution function F. A good strategy for the estimation of $\mathrm{VaR}_\alpha(X)$ is to determine the value of the distribution function $F(x_i) = P[X \leq x_i] = E[\mathbf{1}_{\{X \leq x_i\}}]$ at a sequence of suitable points $x_1, x_2, x_3, \dots \in \mathbb{R}$ in a neighborhood of the (unknown) quantile $q(\alpha)$. Choosing these points recursively may accelerate the computation. The calculation of the values $F(x_i) = E[\mathbf{1}_{\{X \leq x_i\}}]$ usually requires Monte Carlo simulation. Variance reduction techniques can be used to reduce computation time.

Value at risk is essentially given by the inverse of the distribution function. Utility-based shortfall risk admits a similar representation. Let ℓ be a convex loss function, z be an interior point of its range, and $\mathrm{UBSR}_z^\ell(X)$ be the associated risk measure. In this context, the distribution function is replaced by the function L given by

$$L(x) = E[\ell(-X - x)] \qquad (x \in \mathbb{R}),$$

which is decreasing in x. Since ℓ is convex and increasing, the interior of its range equals (a, ∞) for some $a \in \mathbb{R} \cup \{-\infty\}$. For any $y \in (a, \infty)$, the equation $L(x) = y$ has a unique solution. This defines the inverse L^{-1} of L on (a, ∞). Then

$$\mathrm{UBSR}_z^\ell(X) = L^{-1}(z) = \inf\{x : E[\ell(-X - x)] \leq z\}.$$

This shows that for a given loss function ℓ and threshold z, the $\mathrm{UBSR}_z^\ell(X)$ of the position X is the smallest level x such that the expected, ℓ-weighted shortfall $-X - x$ is less than z. For the estimation of $\mathrm{UBSR}_z^\ell(X)$ we consider the value of the function

$$L(x_i) = E[\ell(-X - x_i)] \tag{2}$$

for a suitable, recursively chosen sequence $x_1, x_2, x_3, \dots \in \mathbb{R}$. The standard variance reduction techniques known from the estimation of value at risk can also be used in this context. This allows a reduction in computing time in the Monte Carlo simulation of (2), see Dunkel & Weber (2007).

Finally we show that L can be expressed in terms of AVaR in the case of a quadratic loss function ℓ. In the following we will always write X for the financial position itself, while we use $Y = -X$ for the financial loss.

Example 3.4. Suppose the loss $Y = -X$ of a position X has a continuous distribution function $F = P(Y \leq x)$. With a loss function of the form $\ell(x) = f(x)\mathbf{1}_{\{x>0\}}$ we obtain that $L(x) = E[f(Y-x)\mathbf{1}_{\{Y>x\}}]$. If a quadratic loss function is considered, i.e. $f(x) = x^2$, an interesting connection to AVaR appears. First,

$$L(x) = E(Y^2\mathbf{1}_{\{Y>x\}}) - 2xE(Y\mathbf{1}_{\{Y>x\}}) + x^2(1-F(x)).$$

The second term is related to average value at risk. Note that $\mathrm{VaR}_\alpha(X) = F^{-1}(1-\alpha)$ since F is the distribution function of $Y = -X$. Second,

$$E(Y\mathbf{1}_{\{Y>x\}}) = \int_{F(x)}^{1} F^{-1}(z)dz = \int_0^{\bar{F}(x)} \mathrm{VaR}_\alpha(X)d\alpha = \bar{F}(x)\mathrm{AVaR}_{\bar{F}(x)}(X),$$

where we set $\bar{F}(x) := 1 - F(x)$. Summarizing,

$$L(x) = E[X^2\mathbf{1}_{\{X<-x\}}] - 2x\bar{F}(x)\,\mathrm{AVaR}_{\bar{F}(x)}(X) + x^2\bar{F}(x),$$

which incorporates the average value at risk and also the expected second moment of X subject to shortfall. In special cases the above expressions can be computed explicitly. For example, if F is an extreme value distribution the above expressions can be computed explicitly, compare Examples 4.3 to 4.5. Extreme value distributions will be discussed in the next section.

4 Extreme Value Distributions

We would like to compare the sensitivity of the risk measures discussed above with respect to extreme fluctuations in the value of a financial position. This calls for profit and loss distributions that are realistic models for extreme events. Such distributions are often *heavy tailed*: intuitively, they assign a relatively high probability to excessive losses.

In Figure 1 we compare the densities of heavy and light tailed distributions. The log-plot in the right panel highlights the difference in the tails. In Section 5 below we compare the sensitivity of risk measures for heavy tailed profit and loss distributions. The purpose of the current section is to review some of the most important distributions.

The analysis of distributions that account for excessive fluctuations is at the core of *extreme value theory*. These distributions are typically heavy-tailed.[7] A systematic discussion is in Embrechts, Klüppelberg & Mikosch (1997). McNeil et al. (2005) focus on risk management applications.

[7]Heavy tailed distributions also arise through *mixing*. For example, the distribution of a light tailed normal random variable whose variance is a suitable independent random variable rather than being a constant, is heavy tailed. The Student t and the generalized hyperbolic distributions can be obtained this way. There are numerous other examples, see McNeil, Frey & Embrechts (2005, Section 7.3.3).

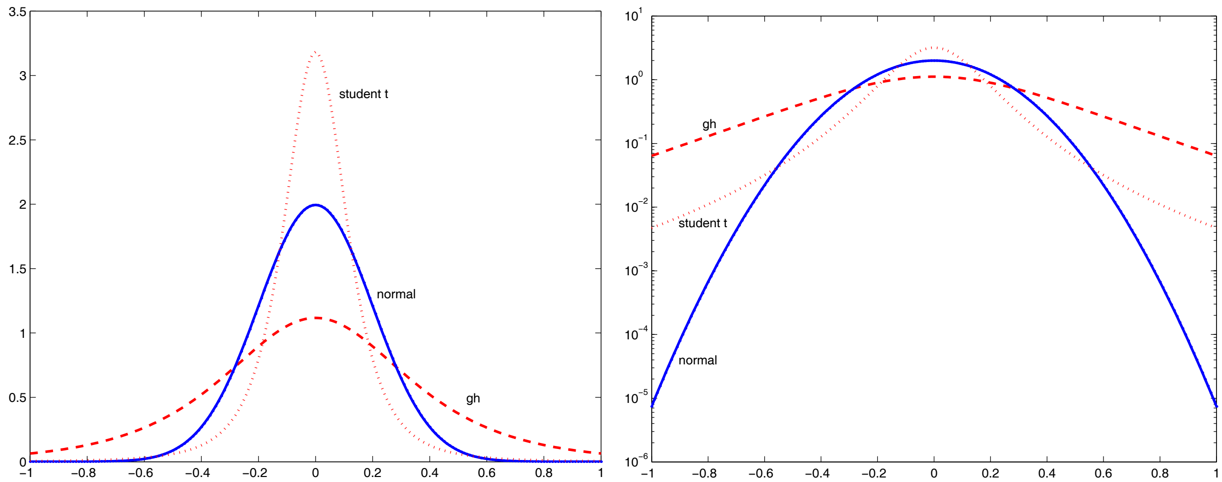

Figure 1: The densities of normal, Student t and generalized hyperbolic distributions with equal mean and variance. The right plot is on a log-scale and shows the differences in the tails.

Generalized Extreme Value Distributions. While the normal distribution arises in connection with sums of random variables, a large class of extreme value distributions arises in connection with their maxima. Let (Y_i) be a sequence of random variables and consider the running maximum

$$M_n = \max_{i \le n} Y_i.$$

We are interested in the distribution of M_n if n becomes large. If the Y_i's are independent and identically distributed with cumulative distribution function $F(x) = P[Y_1 \le x]$, then the distribution function of M_n is the n-fold product F^n:

$$P[M_n \le x] = P[Y_1 \le x, \dots, Y_n \le x] = \big(F(x)\big)^n. \tag{3}$$

This is hardly a useful observation on large samples since $\lim_{n\to\infty} F^n(x) = 0$ for any $x \in \mathbb{R}$ with $F(x) < 1$. Instead of considering the distribution of M_n directly, it is more insightful to investigate a rescaled running maximum. This is illustrated by the following example.

Example 4.1. Let the (Y_i) be independent, $\mathrm{Exp}(\lambda)$ distributed random variables. Consider the distribution of the rescaled running maximum $M_n - \log n$. We get

$$P[M_n - \log n \le x] = \big(F(x + \log n)\big)^n = \big(1 - \frac{1}{n} e^{-x}\big)^n \longrightarrow \exp\big(-e^{-x}\big),$$

for $n \to \infty$. In other words, if F is exponential, then the asymptotic distribution of the rescaled maximum is a Gumbel distribution.

This example suggests to consider the *rescaled* running maximum. The Fisher-Tippett theorem states that for suitable constants $c_n > 0$ and $d_n \in \mathbb{R}$, the distribution of the rescaled maximum $c_n^{-1}(M_n - d_n)$ of the iid sequence (Y_i) converges to one of three standard

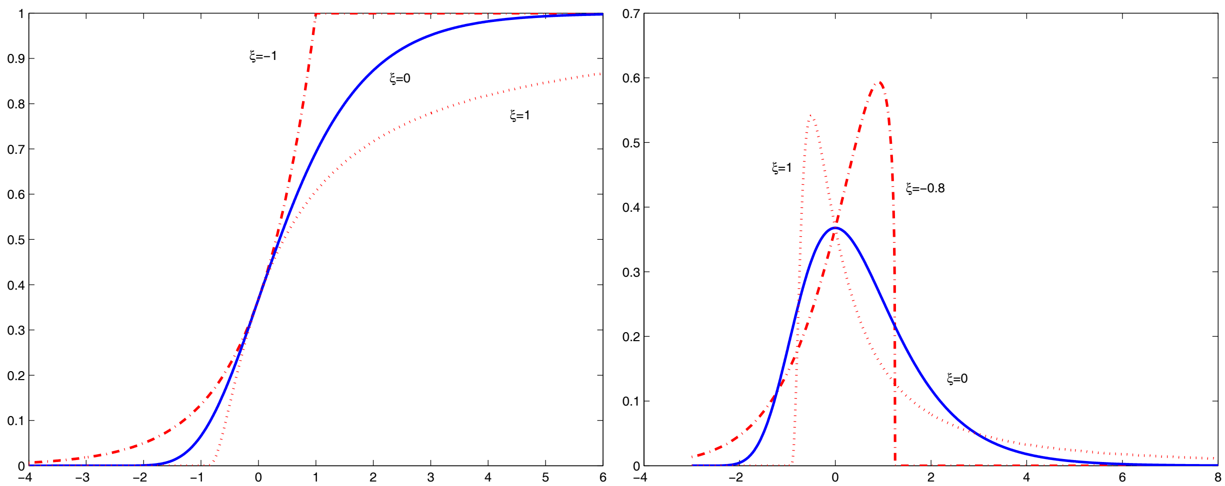

Figure 2: Plot of the distribution function (left) and density (right) of the generalized extreme value distribution (GEV). For $\xi < 0, \xi = 0$ and $\xi > 0$ the GEV distribution is the Weibull, Gumbel and Fréchet distribution, respectively. The parameters are $\xi = -1, 0, 1$ (left) and $\xi = -0.8, 0, 1$ (right).

extreme value distributions as $n \to \infty$. These are the Weibull, Gumbel and Fréchet distributions; they are special cases of the *generalized extreme value distribution* (GEV) with cumulative distribution function H_ξ given by

$$H_\xi(x) = \begin{cases} \exp\big(-(1+\xi x)^{-1/\xi}\big), & \xi \neq 0 \\ \exp\big(-e^{-x}\big), & \xi = 0 \end{cases}$$

for $x \in \mathbb{R}$ and shape parameter $\xi \in \mathbb{R}$ such that $1 + \xi x > 0$.

If the Y_i's have common distribution function F and the limit distribution of the rescaled maximum is H_ξ, then we say that F lies in the *maximum domain of attraction* of H_ξ. We consider the standard extreme value distributions; see Figure 2 for graphs.

(1) For $\xi < 0$ we obtain the *Weibull distribution*. Since it has a fixed right endpoint, it is less relevant for our purposes.

(2) For $\xi = 0$ we obtain the *Gumbel distribution*. Its domain of attraction includes the exponential, normal, lognormal, hyperbolic and generalized hyperbolic distributions.

(3) For $\xi > 0$ we obtain the *Fréchet distribution*. Its domain of attraction includes the Pareto, inverse gamma, Student t, loggamma and the F distributions. These distributions have a heavier, so-called power tail. For $m \geq 1/\xi$, their mth moment does not exist.

Example 4.2. The inverse of the GEV takes an easy form which leads to immediate expression for VaR. Assume that the loss Y of a financial position $X = -X$ has a GEV with parameter $\xi \neq 0$, i.e. $F(x) = P(Y \leq x) = H_\xi(x)$. Hence, we obtain by (1) that

$$\mathrm{VaR}_\alpha(X) = F^{-1}(1-\alpha) = \frac{1}{\xi}\Big[\big(-\log(1-\alpha)\big)^{-\xi} - 1\Big].$$

For the average value at risk this function has to be integrated, which is more complicated. If we consider exceedences over a large threshold instead of maxima the calculations become much simpler, as shown in the following.

Generalized Pareto Distributions. Another important class of distributions in extreme value theory captures the peaks that exceed a certain large threshold level. Let Y be a random variable. We fix a threshold u and consider the excess variable $Y_u = (Y-u)^+$. We are interested in the distribution of Y_u conditional on the event that Y itself exceeds the threshold level. The conditional distribution of Y_u is approximately a *generalized Pareto distribution* (GPD) when the threshold u is raised. This result is actually very closely connected to the Fisher-Tippet theorem. If the distribution of Y lies in the maximum domain of attraction of H_ξ for some shape parameter $\xi \in \mathbb{R}$, then ξ is also the shape parameter in the limiting GPD. The Pickands-Balkema-de Haan theorem states this relation.[8]

The class of generalized Pareto distributions constitutes a two parametric family. Letting $\xi \in \mathbb{R}$ and $\beta > 0$, the generalized Pareto distribution function is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} & \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \xi = 0, \end{cases}$$

where $x \in [0, \infty)$ for $\xi \geq 0$ and $x \in [0, -\beta/\xi]$ for $\xi < 0$.

As shown in Figure 3, the parameters ξ and β control the shape and the scale of the distribution:

(1) For $\xi = 0$ the GPD is the *exponential distribution*.

(2) For $\xi > 0$ it is the *Pareto distribution*. As in the case of a Fréchet distribution, the mth moment does not exist for $m \geq 1/\xi$.

(3) For $\xi < 0$ it is the *Pareto Type II distribution*. The latter is less significant for financial applications since it has a fixed right endpoint.

We will now consider risk measures of generalized Pareto distributions. For the GPD both value at risk and average value at risk are given explicitly in terms of the parameters.

Example 4.3. Assume the loss Y_1 of a financial position $X_1 = -Y_1$ is distributed according to the distribution function $G_{\xi,\beta}$, $\xi \neq 0$. The distribution function is easily inverted and we get, again using (1),

$$\mathrm{VaR}_\alpha(X_1) = \frac{\beta}{\xi}\left(\alpha^{-\xi} - 1\right).$$

[8] The Pickands-Balkema-de Haan theorem states that there exists a function $\beta(u) > 0$ s.t.

$$\lim_{u \to x_F} \sup_{0 \leq x \leq x_F - u} \left| P(Y_u \leq x | Y > u) - G_{\xi,\beta(u)}(x) \right| = 0,$$

if and only if the distribution of Y is in the maximum domain of attraction of H_ξ. Here $Y_u = (Y-u)^+$ and x_F is the right endpoint of F (possibly $+\infty$). See, for example, McNeil et al. (2005).

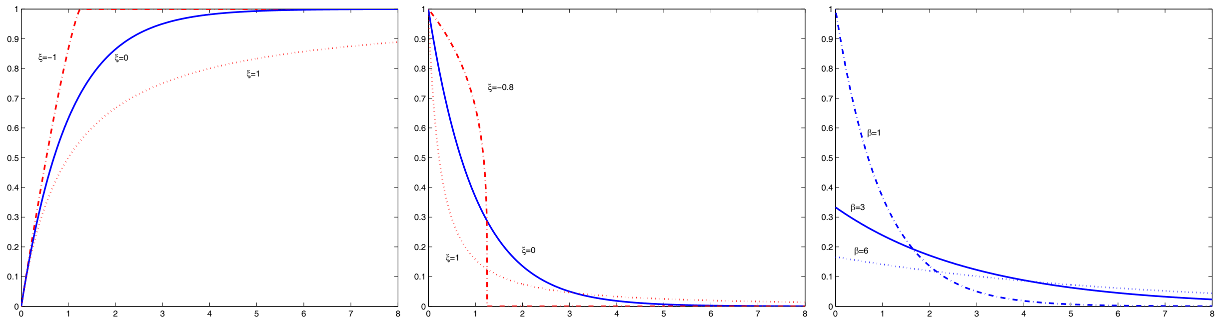

Figure 3: Plot of the distribution function (left) and densities (mid, right) of the generalized Pareto distribution (GPD). For $\xi < 0, \xi = 0$ and $\xi > 0$ this distribution is the Pareto, exponential and Pareto type II distribution, respectively. The parameters are $\xi = -1, 0, 1, \beta = 1$ (left) $\xi = -0.8, 0, 0.5, \beta = 1$ (mid) and $\xi = 0, \beta = 1, 3, 6$ (right).

Next, consider the loss $Y_2 = -X_2$. Assume the probability that Y_2 exceeds a high threshold u is positive. Suppose $P[Y_2 > u + x | Y_2 > u] = G_{\xi,\beta}(x)$ for $x \geq 0$. For $\lambda < P[Y_2 > u]$ we get

$$\mathrm{VaR}_\alpha(X_2) = u + \frac{\beta}{\xi}\Big(\Big(\frac{\alpha}{P[Y_2 > u]}\Big)^{-\xi} - 1\Big).$$

Example 4.4. Again, consider the loss $Y_1 \sim G_{\xi,\beta}$, $\xi \neq 0$ and assume that $\xi < 1$, such that the 1st moment exists. The AVaR of $X_1 = -Y_1$ is given by

$$\mathrm{AVaR}_\alpha(X_1) = \frac{\mathrm{VaR}_\alpha(X_1)}{1-\xi} + \frac{\beta}{1-\xi}.$$

If Y_2 is distributed as in the previous example with $\lambda < P[Y_2 > u]$, we obtain a similar formula, where β is replaced by $\beta - \xi u$, i.e.

$$\mathrm{AVaR}_\lambda(X_2) = \frac{\mathrm{VaR}_\lambda(Y_2)}{1-\xi} + \frac{\beta - \xi u}{1-\xi}.$$

Example 4.5. We can easily compute the UBSR for a quadratic loss function as introduced in Example 3.4. Again consider $Y_1 \sim G_{\xi,\beta}$, $\xi \neq 0$; to guarantee existence of 2nd moments assume $\xi < \frac{1}{2}$. With $c := \beta/\xi$,

$$\begin{aligned} E[X_1^2 \mathbf{1}_{\{X_1 < -x\}}] &= \int_0^{\bar{F}(x)} \big(F^{-1}(1-z)\big) dz \\ &= \int_0^{\bar{F}(x)} \Big[c(z^{-\xi} - 1)\Big]^2 dz \\ &= \frac{c^2 \bar{F}(x)^{1-2\xi}}{1-2\xi} - 2c\bar{F}(x)\mathrm{AVaR}_{\bar{F}(x)}(X_1) - c^2\bar{F}(x), \end{aligned}$$

where we used $\bar{F}(x) = 1 - F(x)$. With the formula for $L(x)$ from Example 3.4 we finally arrive at

$$L(x) = \frac{c^2 \bar{F}(x)^{1-2\xi}}{(1-2\xi)} - 2(c+x)\bar{F}(x)\mathrm{AVaR}_{\bar{F}(x)}(X_1) + \bar{F}(x)\big(x^2 - c^2\big).$$

Remark 4.6. (Existence of moments) As seen in the above examples, UBSR relies on the existence of several moments. For $\ell(x) = x^\alpha \cdot \mathbf{1}_{[0,\infty)}(x)$ the existence of UBSR in the case of a GEV, respectively GPD, distribution is always guaranteed if $\xi \leq 0$. For $\xi > 0$, however, we need that $\xi \leq \frac{1}{\alpha}$.

In the case of the entropic risk measure, i.e. $\ell(x) = e^{\alpha x}$, all moments need to exist, such that this measure does not exist for $\xi > 0$.

Also for a Student t-distribution with ν degrees of freedom, the mth moments do not exist for $m \geq \nu$. That is, the entropic risk measure will never exist, just for the limiting case $\nu = \infty$, which is the normal distribution. For the loss function $\ell(x) = x^\alpha \cdot \mathbf{1}_{[0,\infty)}(x)$ existence follows for $\alpha < \nu$.

5 Sensitivity of Risk Measures

We compare the sensitivity of the convex risk measures average value at risk and utility-based shortfall risk with respect to extreme events in the context of the heavy tailed distributions discussed above.

We construct a benchmark distribution for the loss Y of a position by mixing a Student t distribution with a normal distribution.

Benchmark distribution. Let Y_1 be a Student t distributed scaled by $\frac{1}{2}$ and 2 degrees of freedom and Y_2 be normally distributed with mean μ and variance 0.4. Let W be a random variable independent of Y_1 and Y_2 with $P(W = 1) = w$ and $P(W = 0) = 1 - w$. Set

$$Y = (1 - W)Y_1 + WY_2.$$

μ is chosen to be very large: we use the normal random variable to model a rare extreme event. For $w = 0.04$ this means e.g. that with probability 96 % we are in the Student t case and with probability 4 % some extreme event occurs, scattered around μ.

The distribution of the financial loss Y models event risk through a hump in the tail, whose location is governed by the mean μ of the normal distribution; see the left panel of Figure 4. The higher μ or σ, the more mass is allocated to the tail of the distribution[9] and the more excessive is the fluctuation of Y, holding the other parameters fixed.

[9] More precisely: If we increase only σ, more mass is allocated to the right of $\mu + \delta$ for some $\delta > 0$.

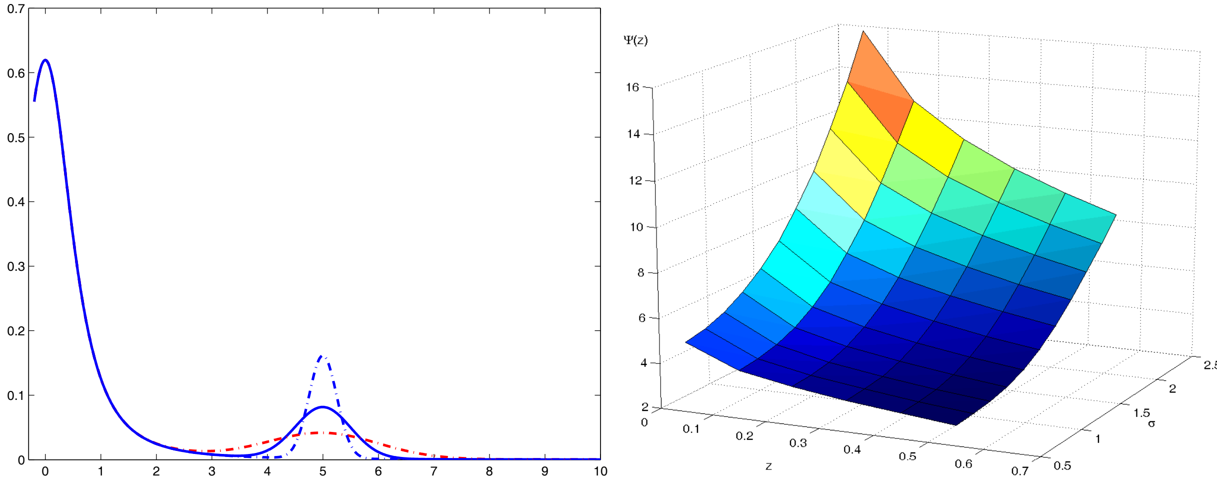

Figure 4: Utility-based shortfall risk with respect to a two-hump distribution. Left panel: density of a convex combination of Student t and normal distributions ($\mu = 5, \sigma = 0.25, 0.5, 1$). Right panel: utility-based shortfall risk as a function of the threshold z for varying standard deviations $\sigma = 0.5, \ldots, 2.5$. Note that these distributions all have similar $\mathrm{VaR}_{0.05}$ (5.1) and $\mathrm{AVaR}_{0.05}$ (5.6 to 6.2).

Sensitivity of risk measures. We consider now VaR, AVaR and UBSR for varying values of μ to illustrate their sensitivity with respect to extreme events. Remember that a larger μ corresponds to additional event risk. In our numerical case study, we choose the loss function of utility-based shortfall risk as $\ell(x) = x^\alpha \mathbf{1}_{[0,\infty)}(x)$. In the left panel of Figure 5 we plot the risk measures for varying exponents $\alpha \in \{1, \frac{3}{2}, 2\}$ and a threshold $z = 0.3$ as functions of μ. The value at risk is mainly governed by the t-distribution, while all other risk measures start to depend linearly on μ after μ reaches a certain level. As expected, except for value at risk all risk measures are able to capture this event risk.

As shown in right panel of Figure 5, the conclusion is different if we fix μ and vary σ, the volatility of the peak. From a risk management perspective, this is a more dangerous situation. The increasing variance indicates that the values will be substantially larger. We consider $w = 0.04$. Note that in this case, value at risk even decreases. Only utility-based shortfall risk captures this extreme risk properly. For different exponents α utility-based shortfall risk behaves qualitatively different. For $\alpha = 1$ it is similar to average value at risk. In general, the larger α the more sensitive is utility-based shortfall risk with respect to extreme risks. For varying threshold level z and fixed exponent α, the right panel of Figure 4 shows utility-based shortfall risk as a function of σ.

Since the loss function can be freely chosen, utility-based shortfall risk offers a great deal of flexibility. Both loss function and threshold level z can be tailored to the specific needs of any financial institution or regulating authority. The example above illustrated that utility-based shortfall risk is a powerful tool for the measurement of event risk.

This virtue of value at risk is indeed not limited to the specific example we considered. We will finally investigate the measurement of extreme events for further profit and loss

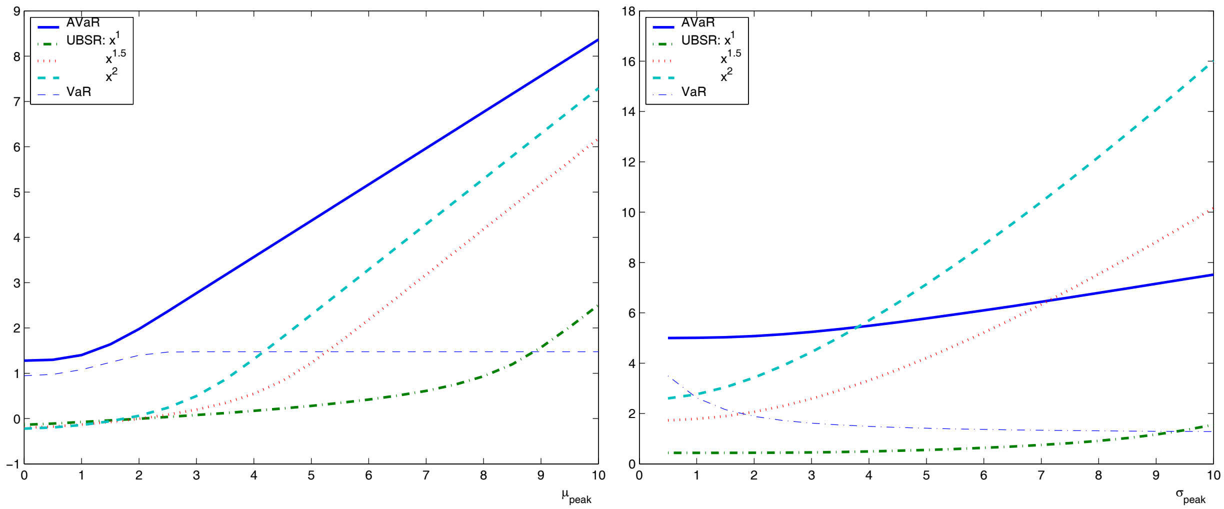

Figure 5: Left: $\mathrm{VaR}_{0.05}$,$\mathrm{AVaR}_{0.05}$ and utility-based shortfall risk ($\alpha \in \{1, \frac{3}{2}, 2\}$, $z = 0.3$) as functions of μ when the distribution of loss Y is given by a convex combination ($w = 0.04$) of a Student t and normal distribution with mean μ. Right: The same risk measures as functions of σ when Y is normally distributed with mean 0 and variance σ.

distributions.

Example 5.1. In the left panel of Figure 6, we consider the risk measures with respect to a GPD, where $\beta = 1$ and we vary ξ. The tail becomes heavier with increasing ξ. Value at risk and average value at risk hardly show the rapidly increasing probability of large values, while utility-based shortfall risk is clearly able to detect this. The exponent α expresses a measure of the sensitivity of utility-based shortfall risk. For $\alpha = 1$ it is similar to average value at risk.

An analysis of risk measures evaluated for GEV distributions shows similar results. The reason is that both distributions, GEV and GPD, have indeed the same tail behavior. If we vary ξ and consider value at risk, average value at risk and utility-based shortfall risk of a GEV distribution H_ξ, again utility-based shortfall risk is capable to detect the risk stemming from the heavy tails, while this is not the case for value at risk and average value at risk. This resembles the results for a GPD $G_{1,\xi}$, since ξ is the common shape parameter for both distributions H_ξ and $G_{1,\xi}$.

Example 5.2. To shed a different light on the sensitivity to heavy tails, we compare all risk measures with respect to a t-distribution with n degrees of freedom. The t-distribution has heavy tails for small n and converges to a normal distribution with light tails for $n \to \infty$. To focus on the tail effects, we rescale the t-distribution such that the value at risk remains constant for different n. For utility-based shortfall risk we use the loss function $\ell(x) = cx^\alpha \mathbf{1}_{\{x>0\}}$ and, to make the results comparable, we chose c such that it matches the value at risk in a benchmark situation, i.e. a t_6-distribution. The results are shown in the right panel of Figure 6. Again, value at risk does not and average value at

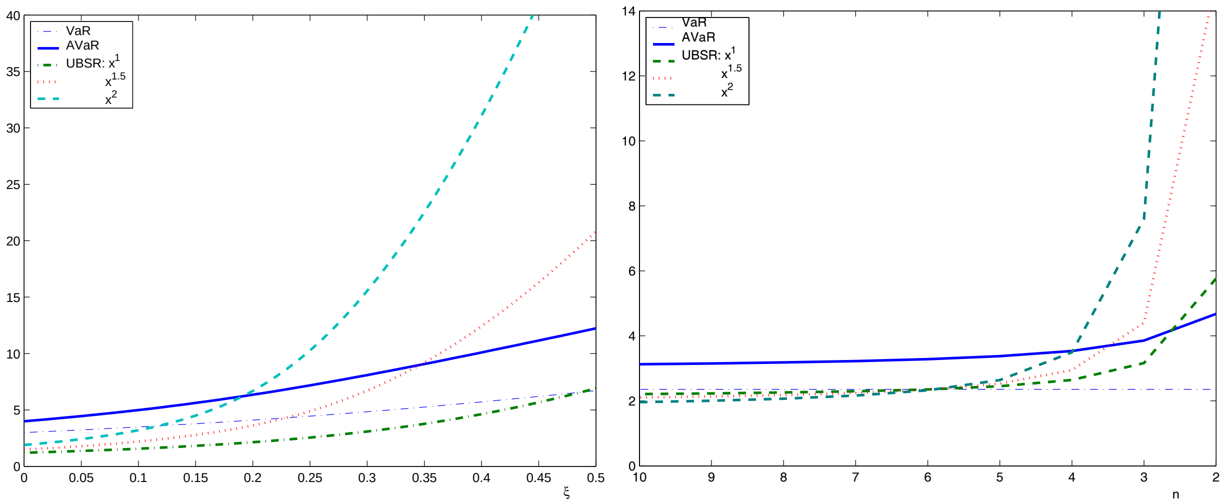

Figure 6: Left: $\mathrm{VaR}_{0.05}$, $\mathrm{AVaR}_{0.05}$ and utility-based shortfall risk for varying exponents $\alpha \in \{1, 1.5, 2\}$ as functions of ξ, where the loss Y has a GPD-distribution with $\beta = 1$ and ξ. The threshold for utility-based shortfall risk is $z = 0.3$. Right: As functions of n, where Y has a t-distribution with n degrees of freedom. The t-distribution is rescaled for each n such that the VaR remains constant. For utility-based shortfall risk, $z = 0.3$ and $\ell(x) = cx^\alpha \mathbf{1}_{\{x>0\}}$, where c is chosen such that for $n = 6$ the utility-based shortfall risk equals the VaR.

risk does hardly indicate the heaviness of the tail. In contrast, utility-based shortfall risk is lower for light tails but increases sharply if the tails get heavier.

All numerical examples indicate that utility-based shortfall risk takes adequate account of event risk if the loss function is of the form $\ell(x) = cx^\alpha \mathbf{1}_{\{x>0\}}$ with $\alpha > 1$. In our examples $\alpha = 1.5$ seems indeed to be a good choice, since in this case the utility-based shortfall risk is less conservative than value at risk and average value at risk in the light-tailed case and still sufficiently sensitive to heavy tails.

References

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber & David Heath (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Dunkel, Jörn & Stefan Weber (2007), 'Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models', *WSC 2007 Proceedings* . To appear.

Embrechts, Paul, Alexander McNeil & Daniel Strautman (2002), Correlation and dependency in risk management: Properties and pitfalls, *in* M.Dempster, ed., 'Risk Management: Value at Risk and Beyond', Cambridge University Press.

Embrechts, Paul, Claudia Klüppelberg & Thomas Mikosch (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Föllmer, Hans & Alexander Schied (2004), *Stochastic Finance - An Introduction in Discrete Time*, Walter de Gruyter, Berlin. 2nd edition.

Glasserman, Paul (2004), *Monte Carlo Methods in Financial Engineering*, Springer, New York.

Kunze, Mathias (2003), Verteilungsinvarinate konvexe Risikomaße. Diplomarbeit, Humboldt-Universität zu Berlin.

Kusuoka, Shigeo (2001), 'On law invariant coherent risk measures', *Advances in Mathematical Economics* **3**, 83–95.

McNeil, Alexander, Rüdiger Frey & Paul Embrechts (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press.

Weber, Stefan (2006), 'Distribution-invariant risk measures, information, and dynamic consistency', *Mathematical Finance* **16**, 419–442.