Derivatives of Value at Risk and Expected Shortfall

November 2003

Hans Rau-Bredow

hans.rau-bredow@mail.uni-wuerzburg.de

Dr. Hans Rau-Bredow
University of Cologne; University of Würzburg
Leo Weismantel Str. 4
D-97074 Würzburg
Germany
phone.: +49(0)178-8218853

Derivatives of Value at Risk and Expected Shortfall

Abstract

This paper analyses derivatives of Value at Risk (*VaR*) and Expected Shortfall (*ES*). First, an elementary result is stated for continuous probability distributions by which derivatives of *VaR* and *ES* of arbitrary order can be derived through recursive application. The case of discrete distributions with only a finite number of possible values is also considered. In this case, expressions for the first derivatives of *VaR* and *ES* can be developed if an exception is made for certain discontinuity points.

JEL classification: D 81, G 11

Keywords : Risk Management, Value at Risk, Expected Shortfall, Marginal Risk Contributions, Decomposing VaR and ES

1. Introduction

Value at Risk (*VaR*) and Expected Shortfall (*ES*) are two closely related and widely used risk measures. An important issue are the derivatives of these risk measures: If a new position is added to the portfolio, how does the risk of the entire portfolio change? In practice, it is often assumed that marginal risk contributions are proportional to the contributions to the portfolio standard deviation, but this is only justified under certain restrictions on the probability distribution: For example, if returns are normal distributed, *VaR* is always a multiple of the standard deviation (e.g., $VaR_{99\%}$ is 2.33 standard deviations). But it is wellknown that the assumption of a normal distribution does not fit with reality, in particular because financial markets returns are fat tailed and because credit risk is asymmetric and skewed. The question arises if general expressions for the derivatives of *VaR* and *ES* exist which do not rely on specific assumptions about the probability distribution.

A general and surprisingly simple result for the derivative of *VaR* is due to *Tasche (1999), (2000), Lemus (1999)* and *Gourieroux/Laurent/Scaillet (2000).* These authors have shown that under certain continuity assumptions, a first order approximation of *VaR*-contributions is given by the conditional expectation of the marginal risk, on condition that the portfolio value is exactly equal to *VaR*. Similarly, Tasche (1999) has shown that the first derivative of *ES* is also given as a conditional expectation, now on condition that losses are equal to or greater than *VaR*. A result for the second derivative of *VaR* was given by *Gourieroux/Laurent/Scaillet (2000)*[1].

For a simple intuition of these results, consider a Monte Carlo simulation with e.g. 1000 random scenarios, which are ranked from the worst loss to the highest gain. $VaR_{99\%}$ is then the outcome in the 10^{th} worst scenario (losses do not exceed this amount in 99% of all cases) and *ES* is the average loss in the 10 worst scenarios. The simple fact that the value of the entire portfolio is always given as the sum of its constituents then suggests a decomposition of *VaR* and *ES* where the result of a particular position in the 10^{th} worst scenario is considered as its contribution to *VaR* and the average loss of this position in the 10 worst scenarios is considered as its contribution to *ES*. Because of that, it is plausible that marginal contributions to *VaR* or *ES* are both given as conditional expectations, in case of *VaR* on condition that losses are exactly equal to *VaR*[2] and in case of *ES* on condition that losses are equal to or greater than *VaR*.

However, it is not easy to complete such simple considerations to a rigorous proof. An illustration of the problems that will arise can be given with reference to the work of *Hallerbach (2002)* (notations translated): Assume, through varying the exposure ε to *Y*, a small change in the allocation of a portfolio $Z = X + \varepsilon Y$, where *X* and *Y* are arbitrary random variables. Then, in writing

$$VaR(Z) = E(X + \varepsilon Y | Z = VaR(Z)) \tag{1}$$

Hallerbach (2002) wants to exploit the linearity of the expectation and concludes that

$$\frac{\partial VaR(Z)}{\partial \varepsilon} = E(Y | Z = VaR(Z)) \tag{2}$$

However, this neglects that, because of $Z = X + \varepsilon Y$, the portfolio allocation has also an influence on the condition $Z = VaR(Z)$. For (2) to be a correct result it must also be the case that

$$\left.\frac{\partial E(X + \varepsilon Y | X + sY = VaR(X + sY))}{\partial s}\right|_{s = \varepsilon} = 0 \tag{3}$$

Of course, because it has been proven in the literature that the derivative of *VaR* is given as the conditional expectation and (2) is indeed correct, (3) must also be true. But (3) in itself is not a trivial result which is obvious without a mathematical proof. Therefore, although *Hallerbach (2002)* arives at a correct result, his reasoning is not completely convincing and cannot replace a rigorous proof.

---

[1] A different derivation for the first and secnond derivative of *VaR*, which uses the Laplacwe transform, was given by Martin/Wilde (2002).

This paper has two aims. First, an elementary result about the expectation of a so called indicator function is derived, which follows from the law of iterated expectations. Derivatives of *VaR* and *ES* of arbitrary order can then be derived through recursive application of this result. This will be explicitly shown for the first and second derivatives. The second aim of the paper is to investigate the derivatives of *VaR* and *ES* in case of a discrete probability distribution. Until now, the case of a discrete dirtribution has not been studied in the literature. It will be shown by a counter-example that the derivative of *VaR* is not always given by the conditional mean in the discrete case. However, it turns out that such counter-examples are related to certain discontinuity points of the conditional mean. If an exception is made for such discontinuities, the results for the derivatives of *VaR* and *ES* can be generalized to the discrete case.

The paper is organized as follows: The next chapter gives the formal definition of *VaR* and *ES* for continuous and discrete distributions. Thereafter, a basic result concernig the expectations of the indicator function is developed and applied in order to calculate derivatives of *VaR* and *ES* in the continuous case. Subsequently, the case of a discrete probability distributionis is discussed. A few concluding remarks follow.

[2] But note that for a continuous distribution, the probability that the portfolio value will be exactly equal to a given VaR-number is always zero. The conditional expectation must then be interpretated in a non-elementary sense.

2. Definition of Value at Risk and Expected Shortfall

Following Tasche (2002), the formal definition of *VaR* with confidence level *p* is as follows:

$$VaR_p(X) = inf\{t \in R | Prob(-X \leq t) \geq p\} \quad (4)$$

If the random variable *X* describes gains (positive values) and losses (negative values) of a bank portfolio, *VaR* according to this definition would be the minimal amount of economic capital required in order to preserve solvency (⇔ *X* + *economic capital* ≥ *0*) with a probability of at least *p*.

If the inverse F_X^{-1} of the cumulative distribution function of *X* exists, the definition simplifies to:

$$VaR_p(X) = -F_X^{-1}(1-p)\,. \quad (5)$$

VaR does not differentiate between small and very large violation of the *VaR*-threshold. Indeed, VaR is not a so-called "coherent" risk measure in the sense of Artzner et al. (1999). For this reason, Expected shortfall (*ES*) has been proposed as an alternative to *VaR*. *ES* is defined as the average loss on condition that losses are greater or equal than *VaR*[3]. This definition can be motivated by the fact that not only the probability that the bank will collapse is of interest to the depositors of the bank, but also if they will loss everything or only a small amount of money in case that the bank actually collapses. *ES* then expresses the expected loss in case of a bank collapse[4].

Again following Tasche (2002), the formal definition of *ES* is as follows:

---

[3] The definition of *ES* is similar to that of other downside risk measures like lower partial moments *LPM. LPM* are defined through $LPM_k = E(max(t-X,0)^k)$. The difference between *LPM* and *ES* is that in case of *ES* the target *t* = *VaR* = *VaRp(X)* is not fixed at the outset but also depends on the portfolio allocation. Lower partial moments were discussed in *Fishburn (1977).*

$$ES_p(X) = \frac{E(-X1_{\{X \leq -VaR_p(X)\}}) - VaR_p(X)(p - Prob(X > -VaR_p(X)))}{1-p} \tag{6}$$

Here, 1_A is the indicator function which has value *1* if *A* is true and zero otherwise.

First consider a continuous distribution. Then it is always the case that $p = Prob(X > -VaR_p(X))$ and (6) reduces to:

$$ES_p(X) = \frac{E(-X1_{\{X \leq -VaR_p(X)\}})}{1-p} = E(-X \mid X \leq -VaR_p(X)) \tag{7}$$

The conditional expectation in (7) is sometimes also called Tail Value-at-Risk (*TailVaR*)[5]. *ES* and *TailVaR* are therefore the same for a continuous probability distribution. From the substitution $F_X(t) = 1 - s \Leftrightarrow t = -VaR_s(X)$ one sees that *ES* or *TailVaR* can then also be written as the average *VaR* for all confidence levels greater than *p*:

$$ES_p(X) = \frac{\int_{-\infty}^{-VaR_p(X)} -t\,dF_X(t)}{1-p} = \frac{\int_p^1 VaR_s(X)\,ds}{1-p} \tag{8}$$

Conversely, $VaR_p(X)$ is the negative derivative with respect to the confidence level *p* of $ES_p(X)$ times $1 - p$. Because of (8), $ES_p(X)$ is given in figure 1a by the hatched area times $1/(1-p)$ $(= 100$ if $p=99\%)$.

Now consider a discrete distribution. If there are only a finite number of possible realizations, then the probability that losses are less than *VaR* could differ from the confidence level *p*. In this case, a correction has to be made in order to ensure that *ES* is the average loss in the worst *(1-p)%* cases and not the average loss in all cases where losses are not less than *VaR*. In (6), this correction is given by the second term of the numerator. In particular if

[4] However, this implicitly assumes that depositors are risk-neutral or that a pseudo risk-neutral probability distribution for the valuation of state-dependent pay-offs has been used.
[5] See for example Artzner et al. (1999) p.223.

$$X_1 < X_2 < \ldots < X_n \tag{9}$$

and $\lceil (1-p)\, n \rceil = i^*$, where the ceiling function $\lceil x \rceil$ denotes the least integer greater than or equal to x, then $VaR_p(X) = -X_{i^*}$ and

$$\begin{aligned} ES_p(X) &= \frac{-\sum_{i=1}^{i=i^*} X_i p_i - X_{i^*}(p - \sum_{i=i^*+1}^{i=n} p_i)}{1-p} \\ &= -\frac{\sum_{i=1}^{i=i^*-1} X_i p_i + X_{i^*}[1 - p - \sum_{i=1}^{i=i^*-1} p_i]}{1-p} \end{aligned} \tag{10}$$

with the notation $p_i = Prob(X = X_i)$. Obviously, $ES_p(X)$ is the weighted sum of $X_1,\ldots,X_{i^*}$, and the weight of X_{i^*} is such that the weights sum up to 1. The difference between the continuous and the discrete case is also tried to illustrate by figure 1b.

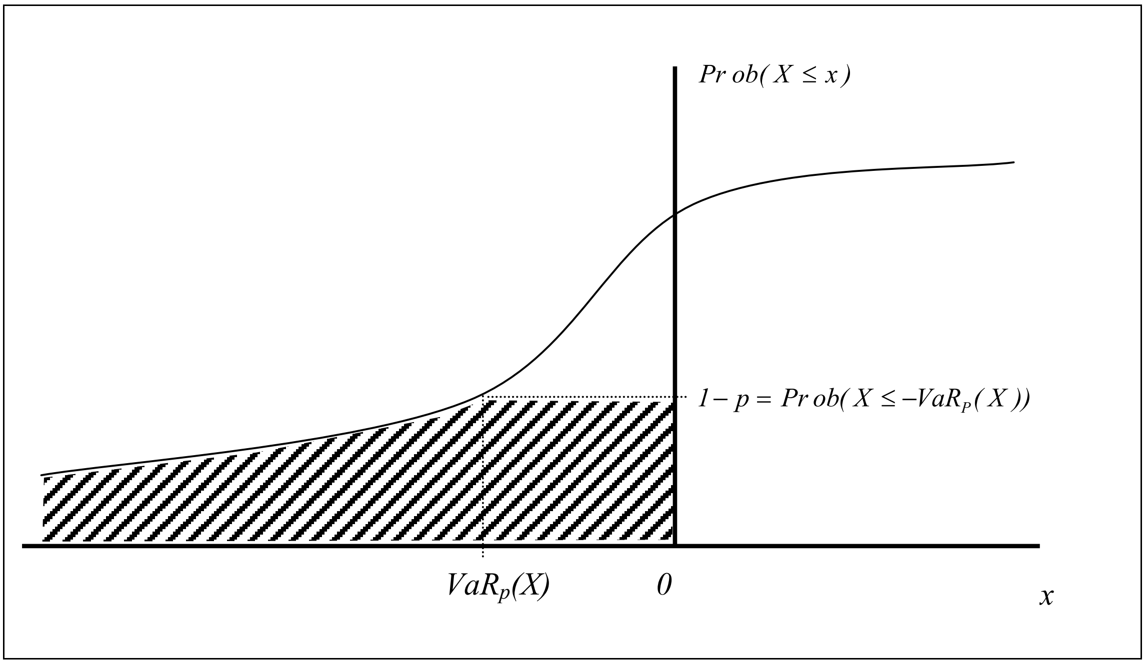

Figure 1a: *ES* for a continuous distribution (hatched area times *1/(1-p)*)

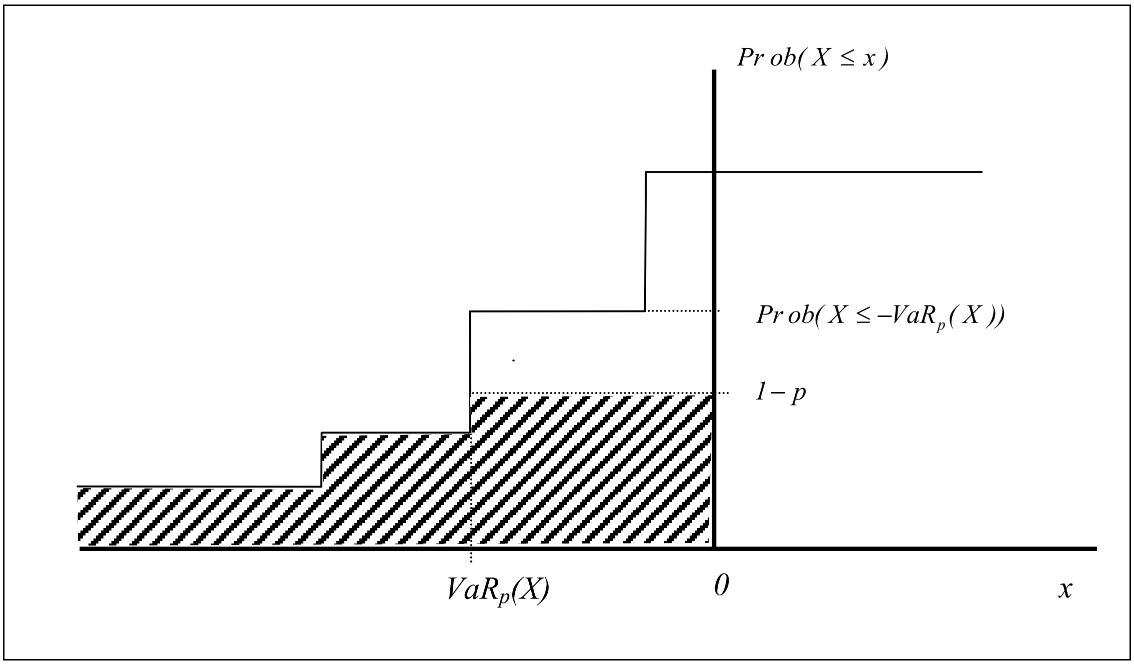

Figure 1b: *ES* for a discrete distribution (hatched area times *1/(1-p)*)

3. Derivatives of Value at Risk and Expected Shortfall

3.1 Continuous probability distributions

In this section it will be shown that the derivatives of *VaR* and *ES* can be derived by a recursive application of an elementary result about the expectation of an indicator function (theorem 1). For this purpose, the indicator function $1_{\{X \le t\}}$ is considered which is defined such that its value is one if $X \le 1$ is true and zero otherwise. The relationsship to the conditional expectations is given by the following two equations:

$$E(Z|X \le t) = \frac{\int_{-\infty}^{t}\int_{-\infty}^{\infty} zf(x,z)dzdx}{\int_{-\infty}^{t}\int_{-\infty}^{\infty} f(x,z)dzdx} = \frac{E(Z1_{\{X \le t\}})}{E(1_{\{X \le t\}})} \tag{11}$$

$$E(Z|X = t) = \frac{\int_{-\infty}^{\infty} zf(t,z)dz}{\int_{-\infty}^{\infty} f(t,z)dz} = \frac{\frac{\partial}{\partial t}E(Z1_{\{X \le t\}})}{\frac{\partial}{\partial t}E(1_{\{X \le t\}})} \tag{12}$$

Also note that

$$\frac{\partial}{\partial t}E(1_{\{X \le t\}}) = \frac{\partial}{\partial t}Prob(X \le t) = f_X(t) \tag{13}$$

where $f_X(t)$ is the density of *X*.

The following theorem 1 analyses the impact if a small perturbation εY is added to *X*. It states that the derivative with respect to the exposure ε to *Y* can be replaced, after multiplying $Z1_{\{X+\varepsilon Y \le t\}}$ by $-Y$, by the derivative with respect to the upper threshold *t*:

Theorem 1:

If X, Y and Z are continuously distributed random variables, then:

$$\frac{\partial}{\partial\varepsilon}E(Z1_{\{X+\varepsilon Y\le t\}})=\frac{\partial}{\partial t}E(-YZ1_{\{X+\varepsilon Y\le t\}})$$

Proof:

First note that for any fixed values $Y=y, Z=z$:

$$\frac{\partial E(z1_{\{X+\varepsilon y\le t\}})}{\partial\varepsilon}=\frac{\partial E(z1_{\{X\le t-\varepsilon y\}})}{\partial\varepsilon}=\frac{\partial(t-\varepsilon y)}{\partial\varepsilon}\frac{\partial E(z1_{\{X\le s\}})}{\partial s}\Big|_{s=t-\varepsilon y}=-y\frac{\partial E(z1_{\{X\le t-\varepsilon y\}})}{\partial t}$$

Then, assuming that derivative and expectations can be interchanged, by the law of iterated expectations[6]:

$$\frac{\partial}{\partial\varepsilon}E(Z1_{\{X+\varepsilon Y\le t\}})=\frac{\partial}{\partial\varepsilon}E(E(Z1_{\{X+\varepsilon Y\le t\}}|Y,Z))=E(\frac{\partial}{\partial\varepsilon}E(Z1_{\{X+\varepsilon Y\le t\}}|Y,Z))$$

$$=E(\frac{\partial}{\partial t}E(-YZ1_{\{X+\varepsilon Y\le t\}}|Y,Z))=\frac{\partial}{\partial t}E(-YZ1_{\{X+\varepsilon Y\le t\}})$$

■

Results for higher order derivatives can be get by a recursive application of theorem 1[7]:

$$\frac{\partial^n}{\partial\varepsilon^n}E(Z1_{\{X+\varepsilon Y\le t\}})=\frac{\partial^n}{\partial t^n}E((-Y)^n Z1_{\{X+\varepsilon Y\le t\}}) \qquad (14)$$

---

[6] Define $g(y,z)=E(X|Y=y,Z=z)$. Consider the random variable $g(Y,Z)$ given as a function of the random variables Y and Z. The law of iterated expectations then states: $E(g(Y,Z))=E(E(X|Y,Z))=E(X)$.

[7] For a proof, interchange the differentation with respect to ε and t and replace Z by $(-Y)^n Z$ in each step.

As an example, the first derivative of *VaR* follows from theorem 1 in the following way: First note that $VaR = VaR_p(X + \varepsilon Y)$ is implicitly defined as a function of ε by:

$$Prob(X + \varepsilon Y \le -VaR) = E(1_{\{X+\varepsilon Y \le -VaR\}}) = p = const. \tag{15}$$

Differentiating this with respect to ε and taking into account that *VaR* depends on ε yields

$$\begin{aligned} 0 &= \frac{\partial E(1_{\{X+\varepsilon Y \le t\}})}{\partial \varepsilon}\Big|_{t=-VaR} + \frac{\partial E(1_{\{X+\varepsilon Y \le t\}})}{\partial t}\Big|_{t=-VaR} \frac{\partial VaR}{\partial \varepsilon} \\ &= \frac{\partial E(-Y 1_{\{X+\varepsilon Y \le t\}})}{\partial t}\Big|_{t=-VaR} + \frac{\partial E(1_{\{X+\varepsilon Y \le t\}})}{\partial t}\Big|_{t=-VaR} \frac{\partial VaR}{\partial \varepsilon} \end{aligned} \tag{16}$$

and finally because of (12):

$$\frac{\partial VaR}{\partial \varepsilon} = \frac{\frac{\partial}{\partial t} E(-Y 1_{\{X+\varepsilon Y \le t\}})\Big|_{t=-VaR}}{\frac{\partial}{\partial t} E(1_{\{X+\varepsilon Y \le t\}})\Big|_{t=-VaR}} = E(-Y \mid X + \varepsilon Y = -VaR) \tag{17}$$

Other derivatives of *VaR* and *ES* can be calculated more or less in the same way. Theorem 2 collects the results for the first and second derivatives of *VaR* and *ES*.

Theorem 2 :

First and second derivatives of $VaR_p(X + \varepsilon Y) = VaR$ and $ES_p(X + \varepsilon Y) = ES$ are given as follows:

(i) $\dfrac{\partial VaR}{\partial \varepsilon} = E(-Y \mid X + \varepsilon Y = -VaR)$

(ii) $\dfrac{\partial^2 VaR}{\partial \varepsilon^2} = \dfrac{1}{f_{X+\varepsilon Y}(-VaR)} \dfrac{\partial(\sigma^2(Y \mid X + \varepsilon Y = t) f_{X+\varepsilon Y}(t))}{\partial t}\Big|_{t=-VaR}$

(iii) $\frac{\partial ES}{\partial \varepsilon} = E(-Y \vert X + \varepsilon Y \leq -VaR)$

(iv) $\frac{\partial^2 ES}{\partial \varepsilon^2} = \frac{\sigma^2(Y \vert X + \varepsilon Y = -VaR) f_{X+\varepsilon Y}(-VaR)}{1-p}$

To proof (ii)-(iv), it is useful to first show the following generalization of (iv):

(v) $\frac{\partial E(Z \vert X + \varepsilon Y \leq -VaR)}{\partial \varepsilon} = -\frac{cov(Y, Z \vert X + \varepsilon Y = -VaR) f_{X+\varepsilon Y}(-VaR)}{1-p}$

Proof:

ad(v): From equations (11), (12), (13) and (17) it follows that:

$$\frac{\partial E(Z \vert X + \varepsilon Y \leq -VaR)}{\partial \varepsilon}$$

$$= \frac{\partial}{\partial \varepsilon} \frac{E(Z 1_{\{X+\varepsilon Y \leq -VaR\}})}{1-p}$$

$$= \frac{1}{1-p} \left(\frac{\partial E(-YZ 1_{\{X+\varepsilon Y \leq t\}})}{\partial t} - \frac{\partial E(Z 1_{\{X+\varepsilon Y \leq t\}})}{\partial t} \frac{\partial VaR}{\partial \varepsilon} \right) \Big|_{t=-VaR}$$

$$= \frac{1}{1-p} \left\{ \left[\frac{\frac{\partial}{\partial t} E(-YZ 1_{\{X+\varepsilon Y \leq t\}})}{\frac{\partial}{\partial t} E(1_{\{X+\varepsilon Y \leq t\}})} - \frac{\frac{\partial}{\partial t} E(Z 1_{\{X+\varepsilon Y \leq t\}})}{\frac{\partial}{\partial t} E(1_{\{X+\varepsilon Y \leq t\}})} E(-Y \vert X + \varepsilon Y = t) \right] \frac{\partial E(1_{\{X+\varepsilon Y \leq t\}})}{\partial t} \right\} \Big|_{t=-VaR}$$

$$= -\frac{[E(YZ \vert X + \varepsilon Y = t) - E(Y \vert X + \varepsilon Y = t) E(Z \vert X + \varepsilon Y = t)] f_{X+\varepsilon Y}(t)}{1-p} \Big|_{t=-VaR}$$

$$= -\frac{cov(Y, Z \vert X + \varepsilon Y = -VaR) f_{X+\varepsilon Y}(-VaR)}{1-p}$$

ad (iii):

$$\frac{\partial ES}{\partial \varepsilon} = \frac{\partial E(-X-\varepsilon Y | X+\varepsilon Y \le -VaR)}{\partial \varepsilon}$$

$$= \frac{\partial E(-X-sY | X+\varepsilon Y \le -VaR)}{\partial s}\Big|_{s=\varepsilon} + \frac{\partial E(-X-\varepsilon Y | X+sY \le -VaR)}{\partial s}\Big|_{s=\varepsilon}$$

$$= E(-Y | X+\varepsilon Y \le -VaR) + \frac{cov(Y,-X-\varepsilon Y | X+\varepsilon Y = -VaR) f_{X+\varepsilon Y}(-VaR)}{1-p}$$

$$= E(-Y | X+\varepsilon Y \le -VaR) + 0$$

where the last but one steps follows from (v) with $Z = -X - \varepsilon Y$

ad (iv): (iv) is a special case of (v) with $Z = -Y$.

ad(iii): It follows from (8) that VaR is the negative derivative with respect to the confidence level p of ES times $1-p$. Interchanging derivatives and applying (iv) then yields:

$$\frac{\partial^2}{\partial \varepsilon^2} VaR = \frac{\partial^2}{\partial \varepsilon^2} \frac{\partial(-(1-p)ES)}{\partial p}$$

$$= -\frac{\partial[\sigma^2(Y | X+\varepsilon Y = -VaR) f_{X+\varepsilon Y}(-VaR)]}{\partial p}$$

$$= \frac{1}{f_{X+\varepsilon Y}(-VaR)} \frac{\partial[\sigma^2(Y | X+\varepsilon Y = t) f_{X+\varepsilon Y}(t)]}{\partial t}\Big|_{t=-VaR}$$

The last equation follows because of (5) and

$$\frac{\partial}{\partial p}(-VaR) = \frac{\partial}{\partial p}(F_X^{-1}(1-p)) = \frac{-1}{f_{X+\varepsilon Y}(-VaR)}$$

■

Derivatives of *VaR* and *ES* of a degree higher than 2 can be get by further recursive applications of theorem 1. This requires that the results of theorem 2, in particular the conditional expectations and the density $f_{X+\varepsilon Y}(t)$, can be expressed with the use of the indicator function, which is possible because of (11), (12) and (13).

3.2 Discrete probability distributions

In this section it will be studied whether the previous results are also valid for a discrete probability distribution. First, a counter-example will be developed which shows that the derivative of *VaR* not always coincides with the conditional expectation. Consider the following case:

$$X = \begin{cases} -100 & 2{,}2\% \\ 100 & 97{,}8\% \end{cases}$$

$$Y = \begin{cases} -1 & 50\% \\ 1 & 50\% \end{cases} \tag{18}$$

Assume that X, Y are stochastically independent. Then:

$$X + \varepsilon Y = \begin{cases} -100 - \varepsilon & 1{,}1\% \\ -100 + \varepsilon & 1{,}1\% \\ 100 - \varepsilon & 48{,}9\% \\ 100 + \varepsilon & 48{,}9\% \end{cases} \tag{19}$$

It follows from definition (4) that *VaR* of $X + \varepsilon Y$ with confidence level $p=99\%$ for sufficiently small ε is given by:

$$VaR_{99\%}(X + \varepsilon Y) = 100 + \varepsilon \tag{20}$$

The minimal economic capital is $VaR = 100 + \varepsilon$, because with less economic capital the bank would collapse with probability *1,1% > 1%.* Then for the derivative of *VaR*:

$$\frac{\partial}{\partial\varepsilon} VaR_{99\%}(X + \varepsilon Y) = 1 \tag{21}$$

On the other side, because X and Y are assumed to be stochastically independent, one gets for the conditional expectation on condition that the loss is equal to *VaR* at $\varepsilon = 0$:

$$E(-Y \,|\, X + \varepsilon Y = -VaR)\big|_{\varepsilon=0} = E(-Y \,|\, X = -10) = E(-Y) = 0 \tag{22}$$

Obviously, this conditional expectation differs from the derivative of *VaR*. This is a contradiction to theorem 2, result (i).

However, if ε different from, but very close to zero, and with the distributions of X and Y as presumed in the example above, the only possible solution of $X + \varepsilon Y = -VaR \Leftrightarrow X + \varepsilon Y = -100 - \varepsilon$ is $X = -100$ and $Y = -1$. Therefore, for a sufficiently small $\varepsilon \neq 0$, it is always $E(-Y \,|\, X + \varepsilon Y = -VaR) = 1$.

It seems that it is only the discontinuity of the conditional expectation at $\varepsilon = 0$ that leads to a contradiction to theorem 2. In the neighbourhood of $\varepsilon = 0$, also in our discrete example the derivative of *VaR* always coincides with the conditional expectation. Theorem 3 shows that this observation can be generalized:

Theorem 3:

Consider two random variables X and Y with only a finite number $(X_1,...X_n)$ and $(Y_1,...Y_m)$ of possible realizations. Denote $VaR_p(X + \varepsilon Y) = VaR$ and $ES_p(X + \varepsilon Y) = ES$. Then for all sufficiently small $\varepsilon \neq 0$ in a certain neighborhood of $\varepsilon = 0$:

(i) $\frac{\partial}{\partial\varepsilon} VaR = E(-Y \,|\, X + \varepsilon Y = -VaR)$

(ii) $$\frac{\partial}{\partial \varepsilon} ES = \frac{E(-Y1_{\{X+\varepsilon Y \le -VaR\}}) - \frac{\partial VaR}{\partial \varepsilon}(p - Prob(X + \varepsilon Y > -VaR))}{1-p}$$

Proof:

First note that if the two random variables X and Y have a finite number of n resp. m possible realizations, then there is a strict ranking order of the $n \cdot m$ possible realizations of $X + \varepsilon Y$ and this ranking order will be the same for all sufficiently small $\varepsilon \neq 0$ in a certain neighborhood of $\varepsilon = 0$:

$$X_{k(1)} + \varepsilon Y_{l(1)} < X_{k(2)} + \varepsilon Y_{l(2)} < \ldots < X_{k(nm)} + \varepsilon Y_{l(nm)}$$

Necessary conditions for this are $X_{k(i)} \le X_{k(i+1)}$ for all i and $Y_{l(i)} < Y_{l(i+1)}$ if $X_{l(i)} = X_{l(i+1)}$.

With this notation, the proof can be stated as follows:

ad(i) VaR with confidence level p is the minimal realization of $X + \varepsilon Y$ on condition that less than $(1-p)\,n\,m$ realizations of $X + \varepsilon Y$ are below $-VaR$. If the ceiling function $\lceil (1-p)\,n\,m \rceil = i^*$ denotes the least integer greater than or equal to $(1-p)\,n\,m$, then:

$$VaR = -X_{k(i^*)} - \varepsilon Y_{l(i^*)} \;\Rightarrow\; \frac{\partial}{\partial \varepsilon} VaR = -Y_{l(i^*)}$$

On the other side, if the perturbation εY of X is sufficiently small, and on condition that $\varepsilon \neq 0$, there is only one solution of $X + \varepsilon Y = -VaR$ <=> $X + \varepsilon Y = X_{k(i^*)} + \varepsilon Y_{l(i^*)}$, namely $X = X_{k(i^*)}$ and $Y = Y_{l(i^*)}$. Therefore:

$$E(-Y \mid X + \varepsilon Y = -VaR) = E(-Y \mid \; X = X_{k(i^*)};\; Y = Y_{l(i^*)}) = -Y_{l(i^*)} = \frac{\partial}{\partial \varepsilon} VaR$$

ad(ii): With notation $Prob(X+\varepsilon Y = X_{k(i)} + \varepsilon Y_{l(i)}) = p_i$, and recording that small changes of ε have no effect on the ranking order, it can be stated that

$$\frac{\partial}{\partial\varepsilon}E((-X-\varepsilon Y)1_{\{X+\varepsilon Y\le -VaR_p(X+\varepsilon Y)\}}) = \frac{\partial}{\partial\varepsilon}\sum_{i=1}^{i=i^*}[(-X_{k(i)}-\varepsilon Y_{l(i)})p_i]$$

$$=\sum_{i=1}^{i=i^*}(-Y_{l(i)}p_i)$$

$$=E(-Y1_{\{X+\varepsilon Y\le -VaR_p(X+\varepsilon Y)\}})$$

It is then also the case that

$$\frac{\partial}{\partial\varepsilon}(p-Prob(-X-\varepsilon Y<VaR_p(-X-\varepsilon Y)))=0$$

Applying this to the definition of *ES* (equation (6)) immediately yields:

$$\frac{\partial}{\partial\varepsilon}ES_p(X+\varepsilon Y)=\frac{\partial}{\partial\varepsilon}\Big(\frac{E((-X-\varepsilon Y)1_{\{X+\varepsilon Y\le -VaR_p(-X-\varepsilon Y)\}})}{1-p} - \frac{VaR_p(X+\varepsilon Y)(p-Prob(X+\varepsilon Y>-VaR_p(-X-\varepsilon Y)))}{1-p}\Big)$$

$$=\frac{E(-Y1_{\{X+\varepsilon Y\le -VaR\}})-\frac{\partial VaR}{\partial\varepsilon}(p-Prob(X+\varepsilon Y>-VaR))}{1-p}$$

■

It is clear that the that the above result for the first derivative of *ES* is a generalization of the result of theorem 2 for the continuous case. Note that in the discrete case, higher order derivatives of *VaR* and *ES* are always zero, because *VaR* and *ES* are then locally linear functions of ε. This can also be seen as in accordance with the results in the continuous case, because if the condition $X+\varepsilon Y=-VaR$ determines that $Y=Y_{k(i^*)}$, then Y becomes non-stochastic and the conditional variance $\sigma^2(Y|X+\varepsilon Y=-VaR)$, which appears in theorem 2 results (ii) and (iv), will be zero.

4. Concluding remarks

It has been shown in the existing literature that general expressions for marginal contributions to *VaR* and *ES* are both given by conditional expectations, in case of *VaR* on condition that losses are exactly equal to *VaR* and in case of *ES* on condition that losses are equal to or greater than *VaR*. These results do not rely on any specific assumptions about the probability distribution. In this paper, a more elementary and general relationship (theorem 1) has been derived from which derivatives of arbitrary order can be get by recurrent application. This relationship may also serve for a deeper understanding of the mathematics behind the result that marginal contributions to *VaR* and *ES* are in fact given by conditional expectations.

It has also been studied in this paper whether these results are valid for random variables with discrete probability distributions. Although counter-examples in particular for the first derivative of *VaR* can then be constructed, it is obvious that such counter-examples are caused by certain discontinuity points which could emerge if the conditional expectation is considered as a function of the portfolio allocation. It has been shown that it is possible to calculate first order derivatives of *VaR* and *ES* also in the discrete case provided an exception is made for such discontinuity points. However, higher order derivatives are always zero in the discrete case, because *VaR* and *ES* are then locally linear functions of the portfolio weights[8].

References

Artzner, P., Delbaen F., Eber J.-M. and Heath, D. (1999): Coherent Measures of Risk, in: Mathematical Finance 9 3, pp. 203-228.

Fishburn, P.C. (1977): Mean Risk Analysis with Risk Associated with below Target Returns, American Economic Review, 67, pp.116-126

Gourieroux C., Laurent J.P., Scaillet O. (2000): Sensitivity Analysis of Values at Risk, in: Journal of Empirical Finance Vol. 7 (3-4) pp. 225-245.

[8] It could be added that practitioners are not only interested in theoretical results for the derivatives of *VaR* and *ES*, but also how these derivatives can be estimated in practise. This paper has not focused on this estimation problem, but see e.g. Gourieroux (2000), Lemus (1999), Hallerbach (2002), Tasche/Tibiletti (2001).

Hallerbach W. (2002): Decomposing Portfolio Value-at-Risk, A General Analysis, in: The Journal of Risk 5/2, vol. 5/2, Winter 2002, pp.1-18

Lemus-Rodriguez, G. (1999): Portfolio optimization with quantile-based risk measures, PhD thesis, Massachusetts Institute of Technology.

Martin, R., Wilde, T. (2002): Unsystematic credit risk, in: Risk, pp. 123-128

Tasche, D. (1999): Risk Contributions and Performance Measurement, working paper.

Tasche, D. (2000): Conditional expectation as quantile derivative, working paper.

Tasche, D., Tibiletti, L. (2001): Approximations for the Value-at-Risk approach to risk-return analysis, working paper.