Lecture 7: Average value-at-risk

Prof. Dr. Svetlozar Rachev

Institute for Statistics and Mathematical Economics
University of Karlsruhe

Portfolio and Asset Liability Management

Summer Semester 2008

Copyright

The material is based on the text-book:

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi

Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures

John Wiley, Finance, 2007

Prof. Svetlozar (Zari) T. Rachev
Chair of Econometrics, Statistics
and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Kollegium am Schloss, Bau II, 20.12, R210
Postfach 6980, D-76128, Karlsruhe, Germany
Tel. +49-721-608-7535, +49-721-608-2042(s)
Fax: +49-721-608-3811
http://www.statistik.uni-karslruhe.de

Introduction

- The average value-at-risk (AVaR) is a risk measure which is a superior alternative to VaR.
- There are convenient ways for computing and estimating AVaR which allows its application in optimal portfolio problems.
- It satisfies all axioms of coherent risk measures and it is consistent with the preference relations of risk-averse investors.
- AVaR is a special case of spectral risk measures.

Average value-at-risk

A disadvantage of VaR is that it does not give any information about the severity of losses beyond the VaR level. *Consider the example.*

- Suppose that X and Y describe the random returns of two financial instruments with densities and distribution functions such as the ones in *Figures 1,2*.
- The expected returns are 3% and 1%, respectively. The standard deviations of X and Y are equal to 10%. The cumulative distribution functions (c.d.f.s) $F_X(x)$ and $F_Y(x)$ cross at $x = -0.15$ and $F_X(-0.15) = F_Y(-0.15) = 0.05$.
- The 95% VaRs of both X and Y are equal to 15%. That is, the two financial instruments lose more than 15% of their present values with probability of 5%.

$\Rightarrow$ We may conclude that their risks are equal because their 95% VaRs are equal.

Average value-at-risk

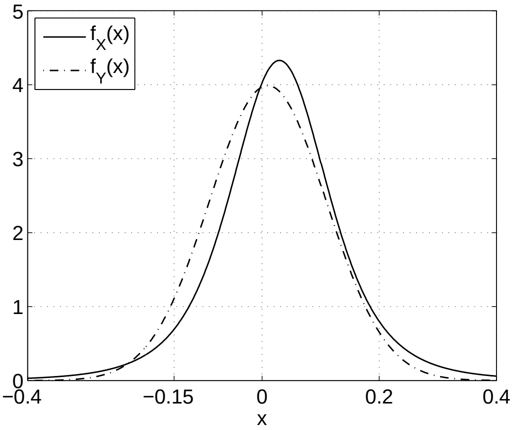

Figure 1. The plot shows the densities of X and Y. The 95% VaRs of X and Y are equal to 0.15 but X has a thicker tail and is more risky.

Average value-at-risk

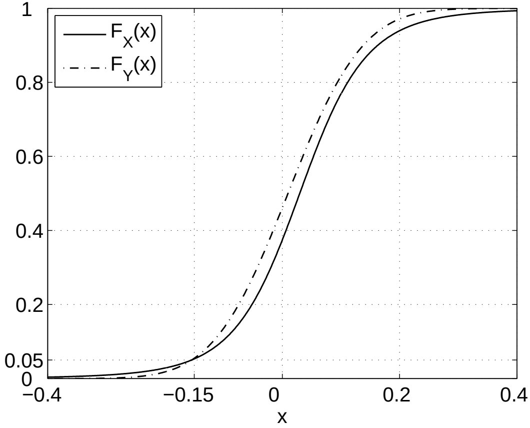

Figure 2. The bottom plot shows their c.d.f.s. The 95% VaRs of X and Y are equal to 0.15 but X has a thicker tail and is more risky.

Average value-at-risk

- The conclusion is wrong because we pay no attention to the losses which are larger than the 95% VaR level.
- It is visible in *Figure 1* that the left tail of X is heavier than the left tail of Y. It is more likely that the losses of X will be larger than the losses of Y, on condition that they are larger than 15%.
- Looking only at the losses occurring with probability smaller than 5%, the random return X is riskier than Y.
- If we base the analysis on the standard deviation and the expected return, we would conclude that both X and Y have equal standard deviations and X is actually preferable because of the higher expected return.
- In fact, we realize that it is exactly the opposite which shows how important it is to ground the reasoning on a proper risk measure.

- The disadvantage of VaR, that it is not informative about the magnitude of the losses larger than the VaR level, is not present in the risk measure known as average value-at-risk.
- In the literature, it is also called conditional value-at-risk or expected shortfall but we will use average value-at-risk (AVaR) as it best describes the quantity it refers to.

Average value-at-risk

- The AVaR at tail probability ϵ is defined as the average of the VaRs which are larger than the VaR at tail probability ϵ.
- The AVaR is focused on the losses in the tail which are larger than the corresponding VaR level. The average of the VaRs is computed through the integral

$$AVaR_\epsilon(X) := \frac{1}{\epsilon}\int_0^\epsilon VaR_p(X)dp \qquad (1)$$

  where $VaR_p(X)$ is defined in equation (5) of Lecture 6.
- The AVaR is well-defined only for random variables with finite mean; that is $AVaR_\epsilon(X) < \infty$ if $E|X| < \infty$.
- For example, random variable, used for a model of stock returns, is assumed to have finite expected return.

$\Rightarrow$ Random variables with infinite mathematical expectation have limited application in the field of finance.

Average value-at-risk

- The AVaR satisfies all the axioms of coherent risk measures. It is convex for all possible portfolios which means that it always accounts for the diversification effect.
- A geometric interpretation of the definition in equation (1) is provided in *Figure 3*, where the inverse c.d.f. of a r.v. X is plotted.
- The shaded area is closed between the graph of $F_X^{-1}(t)$ and the horizontal axis for $t \in [0, \epsilon]$ where ϵ denotes the selected tail probability.
- $AVaR_\epsilon(X)$ is the value for which the area of the drawn rectangle, equal to $\epsilon \times AVaR_\epsilon(X)$, coincides with the shaded area which is computed by the integral in equation (1).
- The $VaR_\epsilon(X)$ value is always smaller than $AVaR_\epsilon(X)$.

Average value-at-risk

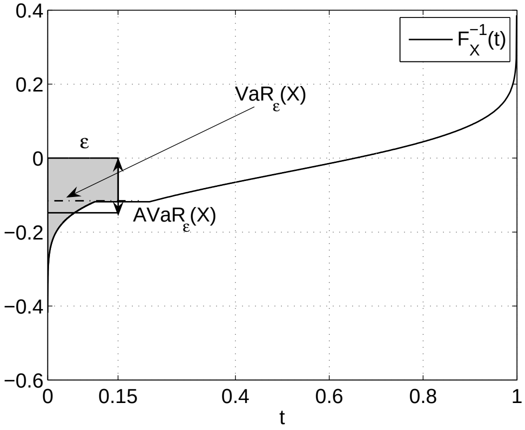

Figure 3. Geometrically, $AVaR_\epsilon(X)$ is the height for which the area of the drawn rectangle equals the shaded area closed between the graph of the inverse c.d.f. and the horizontal axis for $t \in [0, \epsilon]$. The $VaR_\epsilon(X)$ value is shown by a dash-dotted line.

Average value-at-risk

- Recall the example with the VaRs at 5% tail probability, where we saw that both random variables are equal. X is riskier than Y because the left tail of X is heavier than the left tail of Y; that is, the distribution of X is more likely to produce larger losses than the distribution of Y on condition that the losses are beyond the VaR at the 5% tail probability.
- We apply the geometric interpretation illustrated in *Figure 3* to this example.
- First, notice that the shaded area which concerns the graph of the inverse of the c.d.f. can also be identified through the graph of the c.d.f. (See the illustration in *Figure 4*).

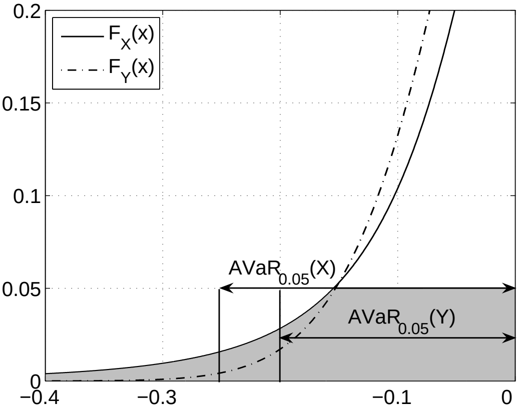

Figure 4. The plot shows a magnified section of the left tails of the c.d.f.s plotted in *Figure 1*. Even though the 95% VaRs are equal, the AVaRs at 5% tail probability differ, $AVaR_{0.05}(X) > AVaR_{0.05}(Y)$.

- In *Figure 4*, the shaded area appears as the intersection of the area closed below the graph of the distribution function and the horizontal axis, and the area below a horizontal line shifted at the tail probability above the horizontal axis.
- The area for $F_Y(x)$ at 5% tail probability is smaller because $F_Y(x) \leq F_X(x)$ to the left of the crossing point of the two c.d.f.s which is exactly at 5% tail probability.
- The $AVaR_{0.05}(X)$ is a number, such that if we draw a rectangle with height 0.05 and width equal to $AVaR_{0.05}(X)$, the area of the rectangle ($0.05 \times AVaR_{0.05}(X)$) equals the shaded area.
- The same exercise for $AVaR_{0.05}(Y)$ shows that $AVaR_{0.05}(Y) < AVaR_{0.05}(X)$ because the corresponding shaded area is smaller and both rectangles share a common height of 0.05

- Besides the definition in equation (1), AVaR can be represented through a minimization formula,

$$AVaR_{\epsilon}(X) = \min_{\theta \in \mathbb{R}} \left(\theta + \frac{1}{\epsilon} E(-X - \theta)_{+} \right) \tag{2}$$

where $(x)_{+}$ denotes the maximum between x and zero, $(x)_{+} = \max(x, 0)$ and X describes the portfolio return distribution.

- This formula has an important application in optimal portfolio problems based on AVaR as a risk measure[1].

[1] See the appendix to the lecture for details

Average value-at-risk

How can we compute the AVaR for a given return distribution?

- We assumed that the return distribution function is a continuous function, i.e. there are no point masses.
- Under this condition, using the fact that VaR is the negative of a certain quantile, the AVaR can be represented in terms of a conditional expectation,

$$AVaR_\epsilon(X) = -\frac{1}{\epsilon}\int_0^\epsilon F_X^{-1}(t)dt \\ = -E(X|X < -VaR_\epsilon(X)), \tag{3}$$

  which is called expected tail loss (ETL), denoted by $ETL_\epsilon(X)$.
- The conditional expectation implies that the AVaR equals the average loss provided that the loss is larger than the VaR level. In fact, the average of VaRs in equation (1) equals the average of losses in equation (3) only if the c.d.f. of X is continuous at $x = VaR_\epsilon(X)$.

Average value-at-risk

- Equation (3) implies that AVaR is related to the conditional loss distribution.
- If there is a discontinuity, or a point mass, the relationship is more involved.
- Under certain conditions, it is the mathematical expectation of the conditional loss distribution, which represents only one characteristic of it.
- See the appendix to this lecture for several sets of characteristics of the conditional loss distribution, which provide a more complete picture of it.

For some continuous distributions, it is possible to calculate explicitly the AVaR through equation (3).

1. The Normal distribution
   Suppose that X is distributed according to a normal distribution with standard deviation σ_X and mathematical expectation EX. The AVaR of X at tail probability ϵ equals

$$AVaR_\epsilon(X) = \frac{\sigma_X}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{(VaR_\epsilon(Y))^2}{2}\right) - EX \qquad (4)$$

   where Y has the standard normal distribution, $Y \in N(0,1)$.

2. The Student's t distribution

Suppose that X has Student's t distribution with ν degrees of freedom, $X \in t(\nu)$. The AVaR of X at tail probability ϵ equals

$$AVaR_\epsilon(X) = \begin{cases} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\sqrt{\nu}}{(\nu-1)\epsilon\sqrt{\pi}} \left(1 + \frac{(VaR_\epsilon(X))^2}{\nu}\right)^{\frac{1-\nu}{2}} & , \nu > 1 \\ \infty & , \nu = 1 \end{cases}$$

where the notation $\Gamma(x)$ stands for the gamma function.

It is not surprising that for $\nu = 1$ the AVaR explodes because the Student's t distribution with one degree of freedom, also known as the *Cauchy distribution*, has infinite mathematical expectation.

Average value-at-risk

- The equation (4) can be represented in a more compact way,

$$AVaR_\epsilon(X) = \sigma_X C_\epsilon - EX, \tag{5}$$

  where C_ϵ is a constant which depends only on the tail probability ϵ.
- Therefore, the AVaR of the normal distribution has the same structure as the normal VaR — the difference between the properly scaled standard deviation and the mathematical expectation. Also, the normal AVaR properties are dictated by the standard deviation.
- Even though AVaR is focused on the extreme losses only, due to the limitations of the normal assumption, it is symmetric.
- Exactly the same conclusion holds for the AVaR of Student's t distribution. The true merits of AVaR become apparent if the underlying distributional model is skewed.

AVaR estimation from a sample

- Suppose that we have a sample of observed portfolio returns and we are not aware of their distribution.
- Then the AVaR of portfolio return can be estimated from the sample of observed portfolio returns.
- Denote the observed portfolio returns by $r_1, r_2, \ldots, r_n$ in order of observation at time instants $t_1, t_2, \ldots, t_n$. Denote the sorted sample by $r_{(1)} \leq r_{(2)} \leq, \ldots, \leq r_{(n)}$. $r_{(1)}$ equals the smallest observed portfolio return and $r_{(n)}$ is the largest.
- The AVaR of portfolio returns at tail probability ϵ is estimated according to the formula

$$\widehat{AVaR}_{\epsilon}(r) = -\frac{1}{\epsilon}\left(\frac{1}{n}\sum_{k=1}^{\lceil n\epsilon \rceil - 1} r_{(k)} + \left(\epsilon - \frac{\lceil n\epsilon \rceil - 1}{n}\right) r_{(\lceil n\epsilon \rceil)}\right) \tag{6}$$

where $\lceil x \rceil$ stands for the smallest integer larger than x.

AVaR estimation from a sample

We demonstrate how equation (6) 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 AVaR at 30% tail probability.
- Here the sample contains 7 observations and $\lceil n\epsilon \rceil = \lceil 7 \times 0.3 \rceil = 3$. According to equation (6), we calculate

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

- We may want to work with a statistical model for which no closed-form expressions for AVaR are known. Then we can simply sample from the distribution and apply formula (6) to the generated simulations.

AVaR estimation from a sample

- Besides formula (6), there is another method for calculation of AVaR. It is based on the minimization formula (2) in which we replace the mathematical expectation by the sample average,

$$\widehat{AVaR}_{\epsilon}(r) = \min_{\theta \in \mathbb{R}} \left(\theta + \frac{1}{n\epsilon} \sum_{i=1}^{n} \max(-r_i - \theta, 0) \right). \qquad (7)$$

- Even though it is not obvious, equations (6) and (7) are completely equivalent.

- The minimization formula in equation (7) 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 $d_1, \dots, d_n$, one for each observation in the sample,

$$\begin{array}{rl} \min\limits_{\theta, d} & \theta + \dfrac{1}{n\epsilon}\displaystyle\sum_{k=1}^{n} d_k \\ \text{subject to} & -r_k - \theta \leq d_k, \; k = 1, n \\ & d_k \geq 0, \; k = 1, n \\ & \theta \in \mathbb{R}. \end{array} \tag{8}$$

AVaR estimation from a sample

The linear problem (8) is obtained from (7) through standard methods in mathematical programming:

- Let us fix the value of θ to θ^*. Then the following choice of the auxiliary variables yields the minimum in (8).
- If $-r_k - \theta^* < 0$, then $d_k = 0$.
- Conversely, if it turns out that $-r_k - \theta^* \geq 0$, then $-r_k - \theta^* = d_k$.
- The sum in the objective function becomes equal to the sum of maxima in equation (7).

AVaR estimation from a sample

- Applying (8) to the sample in the example above, we obtain the optimization problem,

$$\begin{array}{rl}
\min\limits_{\theta,d} & \theta + \dfrac{1}{7 \times 0.3} \sum\limits_{k=1}^{7} d_k \\
\text{subject to} & 0.98\% - \theta \leq d_1 \\
& -0.31\% - \theta \leq d_2 \\
& -1.91\% - \theta \leq d_3 \\
& 1.37\% - \theta \leq d_4 \\
& 0.38\% - \theta \leq d_5 \\
& 0.26\% - \theta \leq d_6 \\
& -0.19\% - \theta \leq d_7 \\
& d_k \geq 0,\, k = \overline{1,7} \\
& \theta \in \mathbb{R}.
\end{array}$$

AVaR estimation from a sample

- The solution to this optimization problem is the number 1.137% which is attained for $\theta = 0.38\%$.
- 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 lecture.
- Let's verify that the solution of the problem is indeed the number 1.137% by calculating the objective in equation (7) for $\theta = 0.38\%$,

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

  Thus, we obtain the number calculated through equation (6).

Computing portfolio AVaR in practice

- The ideas behind the approaches of VaR estimation can be applied to AVaR. We revisit the four methods focusing on the implications for AVaR.
- We assume that there are n common stocks with random returns described by the random variables $X_1, \ldots, X_n$.
- 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.

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.

The multivariate normal assumption

- Under the assumption of normality the AVaR of portfolio return at tail probability ϵ can be expressed in closed-form through (4),

$$AVaR_{\epsilon}(r_p) = \frac{\sqrt{w'\Sigma w}}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{(VaR_{\epsilon}(Y))^2}{2}\right) - Er_p \\ = C_{\epsilon}\sqrt{w'\Sigma w} - Er_p \tag{9}$$

  where C_{ϵ} is a constant independent of the portfolio composition.
- Due to the limitations of the multivariate normal assumption, the portfolio AVaR appears symmetric and is representable as the difference between the properly scaled standard deviation of the random portfolio return and portfolio expected return.

The Historical Method

- 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 (6) or (7).
- We emphasize that the historical method 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 AVaR 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 AVaR may change a lot because the smallest observations tend to fluctuate a lot.

The Hybrid Method

- In the hybrid method, different weights are assigned to the observations by which the more recent observations get a higher weight. The observations far back in the past have less impact on the portfolio risk at the present time.
- In AVaR estimation, the weights assigned to the observations are interpreted as probabilities and the portfolio AVaR can be estimated from the resulting discrete distribution:

$$\widehat{AVaR}_{\epsilon}(r) = -\frac{1}{\epsilon}\left(\sum_{j=1}^{k_\epsilon} p_j r_{(j)} + \left(\epsilon - \sum_{j=1}^{k_\epsilon} p_j\right) r_{(k_{\epsilon+1})}\right) \qquad (10)$$

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 Hybrid Method

- The number k_ϵ in (10) is an integer satisfying the inequalities,

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

- Equation (10) follows directly from the definition of AVaR under the assumption that the underlying distribution is discrete without the additional simplification that the outcomes are equally probable.

- The basic steps of the Monte Carlo method, given in the previous lecture, are applied without modification.
- We assume and estimate a multivariate statistical model for the stocks return distribution. Then we sample from it, and we calculate scenarios for portfolio return.
- On the basis of these scenarios, we estimate portfolio AVaR using equation (6) in which $r_1, \ldots, r_n$ stands for the vector of generated scenarios.

- Similar to the case of VaR, an artifact of the Monte Carlo method is the variability of the risk estimate.
- Since the estimate of portfolio AVaR is obtained from a generated sample of scenarios, by regenerating the sample, we will obtain a slightly different value.
- We illustrate the variability issue by a simulation example.
- Suppose that the portfolio daily return distribution is the standard normal law, $r_p \in N(0,1)$.
- By the closed-form expression in equation (4), we calculate that the AVaR of the portfolio at 1% tail probability equals,

$$AVaR_{0.01}(r_p) = \frac{1}{0.01\sqrt{2\pi}} \exp\left(-\frac{2.326^2}{2}\right) = 2.665.$$

- In order to investigate how the fluctuations of the 99% AVaR change about the theoretical value, we generate samples of different sizes: 500, 1,000, 5,000, 10,000, 20,000, and 100,000 scenarios.
- The 99% AVaR is computed from these samples using equation (6) and the numbers are stored.
- We repeat the experiment 100 times. In the end, we have 100 AVaR numbers for each sample size. We expect that as the sample size increases, the AVaR values will fluctuate less about the theoretical value which is $AVaR_{0.01}(X) = 2.665$, $X \in N(0,1)$.

The Monte Carlo Method

Table below contains the result of the experiment.

Number of Scenarios	AVaR at 99%	95% confidence interval
500	2.646	[2.2060, 2.9663]
1,000	2.771	[2.3810, 2.9644]
5,000	2.737	[2.5266, 2.7868]
10,000	2.740	[2.5698, 2.7651]
20,000	2.659	[2.5955, 2.7365]
50,000	2.678	[2.6208, 2.7116]
100,000	2.669	[2.6365, 2.6872]

Table: The 99% AVaR of the standard normal distribution computed from a sample of scenarios. The 95% confidence interval is calculated from 100 repetitions of the experiment. The confidence intervals cover the theoretical value $AVaR_{0.01}(X) = 2.665$ and also we notice that the length of the confidence interval decreases as the sample size increases. This effect is illustrated in the next figure with boxplot diagrams.

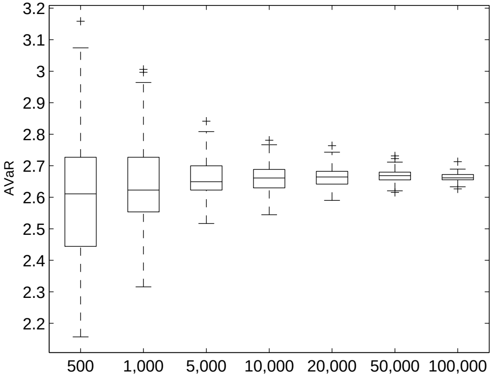

Figure 5. Boxplot diagrams of the fluctuation of the AVaR at 1% tail probability of the standard normal distribution based on scenarios. The horizontal axis shows the number of scenarios and the boxplots are computed from 100 independent samples.

The Monte Carlo Method

- A sample of 100,000 scenarios results in AVaR numbers which are tightly packed around the true value. A sample of only 500 scenarios may give a very inaccurate estimate.
- By comparing the given table to table (VaR computation) on the slide 62 in lecture 6, we notice that the length of the 95% confidence intervals for AVaR are larger than the corresponding confidence intervals for VaR.
- Given that both quantities are at the same tail probability of 1%, the AVaR has larger variability than the VaR for a fixed number of scenarios because the AVaR is the average of terms fluctuating more than the 1% VaR.
- This effect is more pronounced the more heavy-tailed the distribution is.

Back-testing of AVaR

How we can verify if the estimates of daily AVaR are realistic?

- In the context of VaR, a back-testing was used. It consists of computing the portfolio VaR for each day back in time using the information available up to that day only.
- On the basis of the VaR numbers back in time and the realized portfolio returns, we can use statistical methods to assess whether the forecasted loss at the VaR tail probability is consistent with the observed losses.
- If there are too many observed losses larger than the forecasted VaR, then the model is too optimistic.
- If there are too few losses larger than the forecasted VaR, then the model is too pessimistic.
- In this case we are simply counting the cases in which there is an exceedance.

Back-testing of AVaR

Back-testing of AVaR is not so straightforward.

- By definition, the AVaR at tail probability ϵ is the average of VaRs larger than the VaR at tail probability ϵ.
- The most direct approach to test AVaR would be to perform VaR back-tests at all tail probabilities smaller than ϵ. If all these VaRs are correctly modeled, then so is the corresponding AVaR.

$\Rightarrow$ But it is impossible to perform in practice.

- Suppose that we consider the AVaR at tail probability of 1%, for example. Back-testing VaRs deeper in the tail of the distribution can be infeasible because the back-testing time window is too short.
- The lower the tail probability, the larger time window we need in order for the VaR test to be conclusive.
- Even if the VaR back-testing fails at some tail probability ϵ_1 below ϵ, this does not necessarily mean that the AVaR is incorrectly modeled because the test failure may be due to purely statistical reasons and not to incorrect modeling.

Back-testing of AVaR

Why AVaR back-testing is a difficult problem?

- We need the information about the entire tail of the return distribution describing the losses larger than the VaR at tail probability ϵ and there may be too few observations from the tail upon which to base the analysis.
- For example, in one business year, there are typically 250 trading days. Therefore, a one-year back-testing results in 250 daily portfolio returns which means that if $\epsilon = 1\%$, then there are only 2 observations available from the losses larger than the VaR at 1% tail probability.

Back-testing of AVaR

As a result, in order to be able to back-test AVaR, we can assume a certain “structure” of the tail of the return distribution which would compensate for the lack of observations.

There are two general approaches:

1. Use the tails of the Lévy stable distributions as a proxy for the tail of the loss distribution and take advantage of the practical semi-analytic formula for the AVaR[2].
2. Make the weaker assumption that the loss distribution belongs to the domain of attraction of a max-stable distribution. Thus, the behavior of the large losses can be approximately described by the limit max-stable distribution and a statistical test can be based on it.

[2] See the appendix.

Back-testing of AVaR

The rationale of the *first approach*:

- Generally, the Lévy stable distribution provides a good fit to the stock returns data and, thus, the stable tail may turn out to be a reasonable approximation.
- From the Generalized central limit theorem we know that stable distributions have domains of attraction which makes them an appealing candidate for an approximate model.

The *second approach* is based on a weaker assumption:

- The family of max-stable distributions arises as the limit distribution of properly scaled and centered maxima of i.i.d. random variables.
- If the random variable describes portfolio losses, then the limit max-stable distribution can be used as a model for the large losses.
- But then the estimators of poor quality have to be used to estimate the parameters of the limit max-stable distribution, such as the Hill estimator for example. This represents the basic trade-off in this approach.

Spectral risk measures

- By definition, the AVaR at tail probability ϵ is the average of the VaRs larger than the VaR at tail probability ϵ.
- It appears possible to obtain a larger family of coherent risk measures by considering the weighted average of the VaRs instead of simple average.
- Thus, the AVaR becomes just one representative of this larger family which is known as spectral risk measures.

Spectral risk measures

- Spectral risk measures are defined as,[3]

$$\rho_\phi(X) = \int_0^1 VaR_p(X)\phi(p)dp \tag{11}$$

where $\phi(p)$, $p \in [0, 1]$ is the weighting function also known as risk spectrum or risk-aversion function.

- It has the following interpretation. Consider a small interval $[p_1, p_2]$ of tail probabilities with length $p_2 - p_1 = \Delta p$. The weight corresponding to this interval is approximately equal to $\phi(p_1) \times \Delta p$.
- Thus, the VaRs at tail probabilities belonging to this interval have approximately the weight $\phi(p_1) \times \Delta p$.

[3] See Acerbi(2004) for further details.

Spectral risk measures

The risk-aversion function should possess some properties in order for $\rho_\phi(X)$ to be a coherent risk measure, it should be:

Positive	$\phi(p) \geq 0$, $p \in [0, 1]$.
Non-increasing	Larger losses are multiplied by larger weights, $\phi(p_1) \geq \phi(p_2)$, $p_1 \leq p_2$.
Normed	All weights should sum up to 1, $\int_0^1 \phi(p)dp = 1$.

If we compare equations (11) and (1) we notice that the AVaR at tail probability ϵ arises from a spectral risk measure with a constant risk aversion function for all tail probabilities below ϵ.

The Figure below illustrates a typical risk-aversion functions.

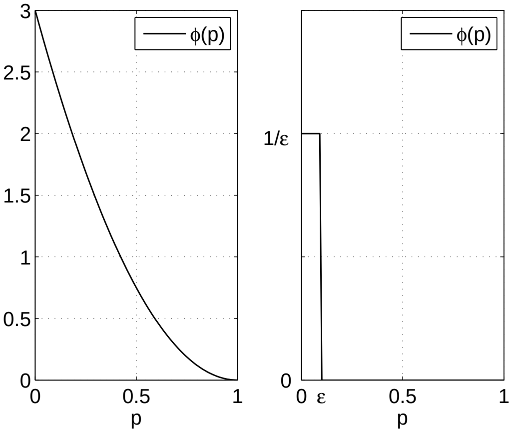

Figure 6. Examples of risk-aversion functions. The right plot shows the risk-aversion function yielding the AVaR at tail probability ϵ.

Spectral risk measures

- In the part of AVaR computation we emphasized that if a sample is used to estimate VaR and AVaR, then there is certain variability of the estimates. We illustrated it through a Monte Carlo example for the standard normal distribution.
- Comparing the results we concluded that the variability of AVaR is larger than the VaR at the same tail probability because in the AVaR, we average terms with larger variability. The heavier the tail, the more pronounced this effect becomes.

Spectral risk measures

- When spectral risk measures are estimated from a sample, the variability of the estimate may become a big issue.
- Note that due to the non-increasing property of the risk-aversion function, the larger losses, which are deeper in the tail of the return distribution, are multiplied by a larger weight.
- The larger losses (VaRs at lower tail probability) have higher variability and the multiplication by a larger weight further increases the variability of the weighted average.
- Therefore, larger number of scenarios may turn out to be necessary to achieve given stability of the estimate for spectral risk measures than for AVaR.

Spectral risk measures

- The distributional assumption for the r.v. X is very important because it may lead to unbounded spectral risk measures for some choices of the risk-aversion function.
- An infinite risk measure is not informative for decision makers and an unfortunate combination of a distributional model and a risk-aversion function cannot be identified by looking at the sample estimate of $\rho_\phi(X)$.
- In practice, when $\rho_\phi(X)$ is divergent in theory, we will observe high variability of the risk estimates when regenerating the simulations and also non-decreasing variability of the risk estimates as we increase the number of simulations.

Spectral risk measures

- We would like to stress that this problem does not exist for AVaR because a finite mean of X guarantees that the AVaR is well defined on all tail probability levels.
- The problem for the spectral measures of risk arises from the non-increasing property of the risk-aversion function. Larger losses are multiplied by larger weights which may result in an unbounded weighted average.

Risk measures and probability metrics

- The probability metrics provide the only way of measuring distances between random quantities.
- A small distance between random quantities does not necessarily imply that selected characteristics of those quantities will be close to each other.
- For example, a probability metric may indicate that two distributions are close to each other and, still, the standard deviations of the two distributions may be arbitrarily different.
- As a very extreme case, one of the distributions may even have an infinite standard deviation.
- If we want small distances measured by a probability metric to imply similar characteristics, the probability metric should be carefully chosen.
- A small distance between 2 random quantities estimated by an ideal metric means that the 2 random variables have similar absolute moments.

Risk measures and probability metrics

- A risk measure can be viewed as calculating a particular characteristic of a random variable.
- There are problems in finance in which the goal is to find a random variable closest to another random variable. For instance, such is the benchmark tracking problem which is at the heart of passive portfolio construction strategies.
- Essentially, we are trying to construct a portfolio tracking the performance a given benchmark; that is finding a portfolio return distribution which is closest to the return distribution of the benchmark.
- Usually, the distance is measured through the tracking error which is the standard deviation of the active return.

Suppose that we have found the portfolio tracking the benchmark most closely with respect to the tracking error.
Can we be sure that the risk of the portfolio is close to the risk of the benchmark?

- Generally, the answer is affirmative only if we use the standard deviation as a risk measure.
- Active return is refined as the difference between the portfolio return r_p and the benchmark return r_b, $r_p - r_b$. The conclusion that smaller tracking error implies that the standard deviation of r_p is close to the standard deviation of r_b is based on the inequality,

$$|\sigma(r_p) - \sigma(r_b)| \leq \sigma(r_p - r_b).$$

- The right part corresponds to the tracking error and, therefore, smaller tracking error results in $\sigma(r_p)$ being closer to $\sigma(r_b)$.

- In order to guarantee that small distance between portfolio return distributions corresponds to similar risks, we have to find a suitable probability metric.
- Technically, for a given risk measure we need to find a probability metric with respect to which the risk measure is a continuous functional,

$$|\rho(X) - \rho(Y)| \leq \mu(X, Y),$$

  where ρ is the risk measure and μ stands for the probability metric.
- We continue with examples of how this can be done for VaR, AVaR, and the spectral risk measures.

1. *VaR*
   Suppose that X and Y describe the return distributions of two portfolios. The absolute difference between the VaRs of the two portfolios at any tail probability can be bounded by,

$$\begin{aligned}|VaR_\epsilon(X) - VaR_\epsilon(Y)| &\le \max_{p\in(0,1)} |VaR_p(X) - VaR_p(Y)| \\ &= \max_{p\in(0,1)} |F_Y^{-1}(p) - F_X^{-1}(p)| \\ &= \mathbf{W}(X,Y)\end{aligned}$$

   where $\mathbf{W}(X,Y)$ is the uniform metric between inverse distribution functions.

   If the distance between X and Y is small, as measured by the metric $\mathbf{W}(X,Y)$, then the VaR of X is close to the VaR of Y at any tail probability level ϵ.

2. *AVaR*
   Suppose that X and Y describe the return distributions of two portfolios. The absolute difference between the AVaRs of the two portfolios at any tail probability can be bounded by,

$$\begin{aligned}|AVaR_\epsilon(X) - AVaR_\epsilon(Y)| &\leq \frac{1}{\epsilon}\int_0^\epsilon |F_X^{-1}(p) - F_Y^{-1}(p)|dp \\ &\leq \int_0^1 |F_X^{-1}(p) - F_Y^{-1}(p)|dp \\ &= \boldsymbol{\kappa}(X, Y)\end{aligned}$$

   where $\boldsymbol{\kappa}(X, Y)$ is the Kantorovich metric.

   If the distance between X and Y is small, as measured by the metric $\boldsymbol{\kappa}(X, Y)$, then the AVaR of X is close to the AVaR of Y at any tail probability level ϵ.

Note that the quantity

$$\kappa_\epsilon(X, Y) = \frac{1}{\epsilon}\int_0^\epsilon |F_X^{-1}(p) - F_Y^{-1}(p)|dp$$

can also be used to bound the absolute difference between the AVaRs.

It is a probability semi-metric giving the best possible upper bound on the absolute difference between the AVaRs.

3. *Spectral risk measures*

Suppose that X and Y describe the return distributions of two portfolios. The absolute difference between the spectral risk measures of the two portfolios for a given risk-aversion function can be bounded by,

$$\begin{aligned}|\rho_\phi(X) - \rho_\phi(Y)| &\leq \int_0^1 |F_X^{-1}(p) - F_Y^{-1}(p)|\phi(p)dp \\ &= \boldsymbol{\kappa}_\phi(X, Y)\end{aligned}$$

where $\boldsymbol{\kappa}_\phi(X, Y)$ is a weighted Kantorovich metric.

If the distance between X and Y is small, as measured by the metric $\boldsymbol{\kappa}_\phi(X, Y)$, then the risk of X is close to the risk of Y as measured by the spectral risk measure ρ_ϕ.

Summary

- We considered in detail the AVaR risk measure. We noted the advantages of AVaR, described a number of methods for its calculation and estimation, and remarked some potential pitfalls including estimates variability and problems on AVaR back-testing. We illustrated geometrically many of the formulae for AVaR calculation, which makes them more intuitive and easy to understand.
- We also considered a more general family of coherent risk measures — the spectral risk measures. The AVaR is a spectral risk measure with a specific risk-aversion function. We emphasized the importance of proper selection of the risk-aversion function to avoid explosion of the risk measure.
- Finally, we demonstrated a connection between the theory of probability metrics and risk measures. Basically, by choosing an appropriate probability metric we can guarantee that if two portfolio return distributions are close to each other, their risk profiles are also similar.

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi
Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures
John Wiley, Finance, 2007.

Chapter 7.