Outline: Risk Measures, Risk Management, and Optimization (10am-1pm)

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Outline: Risk Measures, Risk Management, and Optimization

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Risk and Uncertainty

- Risk and uncertainty are related but are not synonymous.

- Risk is often argued to be a *subjective* phenomenon involving *exposure* and *uncertainty*.\(^1\) Investment risk is related to the uncertain monetary loss to which an investment is exposed.

- Risk is qualified as an *asymmetric phenomenon*—it is related to loss only. Instead, uncertainty is a symmetric notion; it relates to possible positive and negative deviations from the expected price or return.\(^2\)

- Generally, a risk model consists of two parts:
  - Probabilistic models are constructed for the underlying sources of risk, such as market or credit risk factors, and the portfolio loss distribution is described by means of the probabilistic models.
  - Risk is quantified by means of a *risk measure* that associates a real number to the portfolio loss distribution.

---

\(^1\)See Holton (2004) for an analysis of the notion of risk, as well as Knight(1921).

\(^2\)Rachev, Stoyanov, Fabozzi (2008) provide thorough discussions of risk/performance measurement and portfolio analysis.
Measures of Dispersion

Measures of dispersion represent how observations in a dataset are distributed and whether the variability around the mean of the distribution is high or low.

Intuitively, if a quantity is nonrandom, then it is equal to its mean with probability one and there is no fluctuation whatsoever around the mean.

We look at three measures widely used in practice: the standard deviation, the mean absolute deviation, and the semistandard deviation.
Measures of Dispersion: Standard Deviation

- Calculated as the square root of the variance, which itself can be regarded as a measure of uncertainty.
- Denoting by $X$ a generic random variable, the standard deviation is expressed as
  \[ \sigma_X = \sqrt{E(X - EX)^2}, \]
  in which $E$ stands for expectation.
- The standard deviation is always nonnegative and is equal to zero if $X$ is nonrandom.
- The Chebyshev’s inequality helps give an idea of the spread of values of an arbitrary probability distribution:
  \[ P\left( |X - EX| \leq k\sigma_X \right) \geq 1 - 1/k^2, \]
  provided that $X$ has a finite second moment.
- For e.g., we can calculate that if $k = 2$, $P\left( X \in EX \pm 2\sigma_X \right)$ is at least 75%.
In certain cases, the standard deviation may be infinite. A measure, which may be finite even when the standard deviation is not, is the *mean absolute deviation* (MAD).

MAD is defined as the average deviation from the mean in absolute terms,

$$\text{MAD}_X = E|X - EX|.$$ 

It is clear, that as in the case of the standard deviation, both positive and negative deviations are taken into account in the MAD formula. MAD is just an alternative measure of uncertainty.

For some distributions, the two measures can be expressed from one another. For example, if $X$ is Gaussian, then

$$\text{MAD}_X = \sigma_X \sqrt{\frac{2}{\pi}}.$$
The *semistandard deviation* differs from the previous two measures in that it takes into account only the positive or negative deviations from the mean. Therefore, it is not symmetric.

The positive and negative semistandard deviations are defined as

\[
\sigma^+_X = \left( E \left( X - EX \right)_+^2 \right)^{1/2} \quad \text{and} \quad \sigma^-_X = \left( E \left( X - EX \right)_-^2 \right)^{1/2},
\]

where \((x - EX)_+^2\) is the squared difference between an outcome \(x\) and the mean \(EX\) if the difference is positive. Analogously, in the other case.

The two measures can be called, respectively, *upside* (downside) *dispersion measure*.

If \(X\) is a constant, both measures are zero. If the variable is symmetric, they are equal.

If \(X\)’s distribution is positively- (negatively-) skewed, \(\sigma^-_X < \sigma^+_X\) (\(\sigma^-_X > \sigma^+_X\)).
The three measures of dispersion we considered above possess a number of common properties which can be synthesized into axioms.¹

Denote a dispersion measure of the random variable $X$ by $D(X)$.

- **Positive homogeneity.** $D(\lambda X) = \lambda D(X)$ for all $X$ and all $\lambda > 0$.
- **Positivity.** $D(X) \geq 0$ for all $X$ and $D(X) > 0$ for nonconstant $X$.
- **Subadditivity.** $D(X + Y) \leq D(X) + D(Y)$ for all $X$ and $Y$.
- **Translation invariance.** $D(X+C) = D(X)$ for all $X$ and constants $C \in \mathbb{R}$.

In fact, these axioms define a class of convex dispersion measures, called *deviation measures*.

Although a downside deviation measure possesses several of the desirable characteristics of a risk measure, it is *not* a risk measure. We next give an example to demonstrate this.

¹See Rockafellar, Uryasev, Zbarankin (2006) for axiomatic description of dispersion/deviation measures.
Suppose we have in our portfolio a common stock \( X \), with a current market value $95 and an expected monthly return of 0.5\%. Choose a particular deviation measure \( D \) and compute \( D(\bar{r}_X) = 20\% \), where \( \bar{r}_X \) is the stock’s return.

Add a risk-free bond, \( B \), with current value $95, face value $100, and maturity in a month. Bond’s return is nonrandom and equal to 5.26\%.

Portfolio’s return is \( r_p = \bar{r}_X/2 + \bar{r}_B/2 \). Using the axioms above, we obtain \( D(r_p) = D(\bar{r}_X)/2 = 10\% \).

Uncertainty of portfolio return decreases twice. How about risk?

Intuitively, the risk of \( r_p \) must decrease more than twice (compared with \( \bar{r}_X \)), given that half of the new portfolio earns a sure profit of 5.26\%.

This discrepant effect is due to the translation invariance, which makes a deviation measure insensitive to nonrandom profit.
Markowitz (1952) was the first to recognize the relationship between risk and reward and introduced the standard deviation as a measure of risk.

He was also the first to suggest the semistandard deviation as an alternative to deal with the standard deviation’s symmetric nature. We saw, however, that that one cannot be a true risk measure either.

Next, we consider a widely-used risk measure—VaR—and comment on its properties and calculation methods. Afterwards, we explain the concept of coherent risk measure and introduce the average VaR (AVaR) as a representative.
Value-at-Risk

- **Value-at-risk** (VaR) was popularized in the late 1980s by JP Morgan and since the mid-1990s has been used as a regulator-approved valid approach for calculating capital reserves to cover market risk.

- VaR is defined as the minimum level of loss at a given, sufficiently high confidence level for a predefined time horizon. The recommended confidence levels are 95% and 99%.

- Suppose a portfolio’s one-day 99% VaR is equal to $1 million. This means that over the horizon of one day, the portfolio may lose more than $1 million with probability equal to 1%.

- More generally, denote by $(1 - \epsilon)100\%$ the confidence level of the VaR. Losses larger than VaR occur with probability $\epsilon$, called *tail probability*.  


To define VaR formally, suppose that $X$ describes random returns.

VaR at confidence level $(1 - \epsilon)100\%$ (tail probability $\epsilon$) is defined as the negative of the lower $\epsilon$-quantile of the return distribution:

$$\text{VaR}_\epsilon(X) = -\inf_x \{x \mid P(X \leq x) \geq \epsilon\} = F^{-1}_X(\epsilon),$$

where $\epsilon \in (0, 1)$ and $F^{-1}_X(\epsilon)$ is the inverse of the distribution function.
According to the definition, VaR could become a negative number. That would mean that at tail probability $\epsilon$ we do not observe losses but profits. Losses happen with even smaller probability than $\epsilon$.

Consider an aspect which differentiates VaR from all deviation and uncertainty measures. As a consequence of the definition, if we add a nonrandom profit $C$ to the random variable $X$, the resulting VaR is

$$\text{VaR}_\epsilon(X + C) = \text{VaR}_\epsilon(X) - C.$$ 

Furthermore, it holds that $\text{VaR}_\epsilon(\lambda X) = \lambda \text{VaR}_\epsilon(X)$.

In our earlier example, the portfolio return can be expressed as

$$r_p = r_X/2 + 0.0526/2.$$ 

Suppose that the VaR of the stock is 12%. Using the two properties above, we can calculate the VaR of the equally-weighted portfolio to be $\text{VaR}_\epsilon(r_p) \approx 3.365\%$, which is far less than 6% (half of the initial risk).
Suppose a portfolio contains $n$ common stocks. We are interested in calculating the daily 99% VaR.

Denote the random daily stock returns by $X_1, \ldots, X_n$ and the portfolio weights by $w_1, \ldots, w_n$.

The portfolio return is thus $r_p = w_1 X_1 + \cdots + w_n X_n$.

The portfolio VaR is derived from the distribution of $r_p$.

We consider three approaches to VaR calculation, which vary in the assumptions they are based on.
The approach of RiskMetrics is centered on the assumption that the stock returns have a multivariate normal distribution. Then, the distribution of portfolio return is also normal.

The 99% VaR is the negative of the 1% quantile of the $N\left(E_{r_p}, \sigma_{r_p}^2\right)$ distribution.

The portfolio’s expected return and variance are

$$E_{r_p} = \sum_{i=1}^{n} w_i E_{r_i} \quad \text{and} \quad \sigma_{r_p}^2 = w' \Sigma w,$$

where $E_{r_i}$ are the stocks’ expected returns and $\Sigma$ is the returns’ covariance matrix.

Denote the 99% quantile of the standard normal distribution by $q_{0.99}$. Then, using the two VaR properties mentioned earlier, we can write

$$VaR_{0.01}(r_p) = q_{0.99} \sigma_{r_p} - E_{r_p}.$$
VaR Calculation: Parametric Approach

Note that $q_{0.99}$ is a quantity independent of the portfolio composition. The parameters that depend on the portfolio weights are the portfolio return’s standard deviation and expected return.

As a consequence, under the assumption of normality, VaR is symmetric, even though, by construction, it is asymmetric.

The approach of RiskMetrics can be extended for other types of distributions.\(^1\)

A conditional parametric approach to VaR calculation could take the following form:

- Estimate a conditional mean model and predict $Er_p$ one step ahead.
- Estimate a conditional variance model and predict $\sigma_p$ one step ahead.
- Determine $q_{0.99}$ as the 99% quantile of the distribution of standardized residuals.
- Plug these three quantities into the VaR expression.

\(^1\)Lamantia, Ortobelli, Rachev (2006a,b) provide such extensions for Student’s $t$ and stable distributions.
The historical method does not impose any distributional assumptions; the distribution of portfolio returns is constructed from historical data.

The 99% daily VaR, for e.g., is computed as the negative of the 1% quantile of the observed daily portfolio returns. The observations are collected from a predetermined time window, such as the most recent business year.

A number of drawbacks:

- Assumes past trends will continue in the future. This discounts the possibility for extreme future events which have not been observed in the past.

- Observations are treated as $i.i.d.$, which is not realistic. Autocorrelations, clustering, and other features of daily returns need to be accounted for within a conditional approach.

- Not reliable for estimation of VaR at very high confidence levels. An yearly sample of daily data (250 observations) is insufficient for the purpose of 99% VaR estimation.
VaR Calculation: Hybrid Method

- The hybrid method modifies the historical VaR, so that observations are not regarded as i.i.d. Instead, a weighting scheme is applied depending on how close observations are to the present.

- The scheme is exponential smoothing, which emphasizes the more recent observations, thus taking into account the time-varying volatility. The hybrid method consists of the following steps:
  
  - Let \( r_{t-k+1}, \ldots, r_{t-1}, r_t \) be a sequence of \( k \) observations. Then, the \( i \)th observed return is assigned a weight

    \[
    \theta_i = c \lambda^{t-i},
    \]

    where \( 0 < \lambda < 1 \) and \( c = \frac{1-\lambda}{1-\lambda^k} \) is a constant chosen such that the weights sum up to one.

  - The VaR measure is computed, as in the historical approach, from the empirical CDF in which each observation has probability equal to the weight \( \theta_i \).
The Monte Carlo method requires specification of a multivariate statistical model for the stock returns. That multivariate behavior could be specified either through a multivariate distribution or through one-dimensional distributions together with a copula.

The Monte Carlo method consists of the following steps:

- Select a statistical model. The model should be able to explain heavy tails, volatility clustering, etc., which may influence portfolio risk.

- Estimate model parameters. Size of data sample needs to be appropriate for the model complexity.

- Generate scenarios from the fitted model. Each scenario is a vector of returns that depend on each other according to the presumed dependence structure of the model. Scenarios are independent of one another.

- Calculate portfolio risk. Compute portfolio risk on the basis of the portfolio return scenarios from the previous step.
The Monte Carlo method is a very general numerical approach to risk estimation. No closed-form expressions are required and by choosing a flexible statistical model, accurate risk numbers can be obtained.

A disadvantage is that the computed VaR is dependent on the generated sample of scenarios, so that it will fluctuate a little if we regenerate the sample. This effect can be reduced by generating a larger sample.

To illustrate, suppose the daily portfolio return distribution is standard normal. In order to investigate the fluctuations of the 99% VaR about the theoretical value (which we can compute using the RiskMetrics approach), we generate samples of different sizes: from 500 to 100,000 scenarios.

The 99% VaR is computed from these samples and the numbers stored. The experiment is repeated 100 times, so that in the end, we have 100 VaR numbers for each sample size.

The boxplot diagrams on the next slide illustrate best the effect that as the sample size increases, the 100 VaR numbers are more tightly clustered around the theoretical value $\text{VaR}_{0.01}(X) = 2.326 (X \sim N(0, 1))$. 

$$X \sim N(0, 1).$$
The advantage of the Monte Carlo method become obvious in case of a portfolio with complicated instruments, such as derivatives. Closed-form expressions for the portfolio risk are no longer possible, since the portfolio return distribution becomes quite arbitrary.

The Monte Carlo method provides a framework to generate scenarios for the risk-driving factors, revaluate the financial instruments in the portfolio under each scenario, and estimate risk in each state of the world.
A 99% daily portfolio VaR implies that the portfolio loses more than the 99% VaR with a 1% probability. Is that the case in reality? This can be answered by backtesting of VaR.

The procedure generally consists of the following steps:

- Choose a time-window for the backtesting—usually the most recent one or two years.
- For each day in the time window, calculate the daily VaR.
- Check if the loss on a given day is larger or smaller than the predicted VaR. If the loss is larger, we say there is a case of an exceedance.
- Count the number of exceedances. Does the number of exceedances belong to the corresponding 95% confidence interval for the VaR?
- Too many (few) exceedances imply that the VaR numbers produced by the model are too optimistic (pessimistic). Both are undesirable.
Outline: Risk Measures, Risk Management, and Optimization

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Coherent Risk Measures

- Even though VaR has been widely adopted as a risk measure, it does not always satisfy the important property that the VaR of a portfolio should not exceed the sum of the VaRs of the portfolio positions. That is, it does not always represent the diversification effect.

- What are the desirable properties that a good risk measure should satisfy? The so-called coherent risk measure, denoted by $\rho(\cdot)$ has the following defining properties:\(^1\)

  - **Monotonicity.** $\rho(Y) \leq \rho(X)$ if $Y \geq X$ almost surely.
  - **Positive Homogeneity.** $\rho(0) = 0$, $\rho(\lambda X) = \lambda \rho(X)$, for all $X$ and $\lambda > 0$.
  - **Subadditivity.** $\rho(X + Y) \leq \rho(X) + \rho(Y)$, for all $X$ and $Y$.
  - **Invariance.** $\rho(X + C) = \rho(X) - C$, for all $X$ and $C \in \mathbb{R}$.

- An example of a coherent risk measure is the average value-at-risk (AVaR). We discuss it next.

---

\(^1\)Artzner, Delbaen, Eber, Heath (1998) describe axiomatically coherent risk measures.
Average Value-at-Risk: Motivation

Consider the numerical example below:

- $X, Y =$ financial returns.
- $EX = 3\%, \; EY = 1\%,$
  $\sigma_X = \sigma_Y = 10\%.$
- $95\% \text{VaR}_X = 95\% \text{VaR}_Y = 15\%.$
- Do $X$ and $Y$ have the same risk?
- Left tail of $X$ is heavier than left tail of $Y$. Then, losses of $X$ larger than $15\%$ will be larger than corresponding losses of $Y$.
- Moreover, analysis based on the standard deviation and expected return would conclude that $X$ is preferable to $Y$!
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 (AVaR), also known as conditional VaR and expected shortfall.

The AVaR at tail probability $\epsilon$ is defined as the average of the VaRs which are larger than the VaR at tail probability $\epsilon$. The average is computed through the integral:

$$AVaR_\epsilon(X) = \frac{1}{\epsilon} \int_0^\epsilon \text{VaR}_p(X) dp.$$ 

The AVaR satisfies all the properties of coherent risk measures and thus, unlike VaR, always accounts for the diversification effect.
Average Value-at-Risk

Geometric interpretation of $\text{AVaR}_\epsilon(X)$

$\text{AVaR}$ is the value for which the area of the rectangle ($\epsilon \times \text{AVaR}_\epsilon(X)$) is equal to the shaded area, computed by the integral in the earlier definition.
We apply the geometric AVaR interpretation to the earlier example of two financial return distributions with the same 95% VaR but different tail fatness.

The plot shows the magnified left tails of the CDFs. Clearly, the AVaRs at 5% tail probability differ: $\text{AVaR}_{0.05}(X) > \text{AVaR}_{0.05}(Y)$. 

\[ \text{AVaR}_{0.05}(X) \] 
\[ \text{AVaR}_{0.05}(Y) \]
Apart from the definition of AVaR we considered, AVaR can be equivalently represented through a minimization formula,\(^1\)

\[
\text{AVaR}_\epsilon(X) = \frac{1}{\epsilon} \min_{\theta \in \mathbb{R}} \left( \theta \epsilon + E(-X - \theta)_+ \right),
\]

where \((x)_+\) denotes the maximum between \(x\) and zero, \((x)_+ = \max(x, 0)\) and \(X\) is the random portfolio return.

This formula has an important application in optimal portfolio problems based on AVaR as a risk measure, as we will see further below.

\(^1\)A proof that this expression is indeed AVaR can be found in Rockafellar and Uryasev (2002).
AVaR can be represented (assuming there are no discontinuities in the CDF) as the average loss that is larger than the VaR level:

\[
\text{AVaR}_\epsilon(X) = \int_0^\epsilon F_X^{-1}(t)dt = -E(X | X < -\text{VaR}_\epsilon(X)),
\]

which is called expected tail loss (ETL) and denoted by ETL_\epsilon(X).

For some continuous distributions, such as Gaussian and Student’s t, AVaR can be calculated explicitly (from closed-form analytical expressions). In the case of stable distributions, a semi-closed expression is available.

In all three distributional cases, AVaR can be represented in a compact way as

\[
\text{AVaR}_\epsilon(X) = \sigma_X C_\epsilon - \mu_X,
\]

where \( \sigma_X \) and \( \mu_X \) are the scale and location parameters. In the normal and t cases, \( C_\epsilon \) is a constant which depends only on the tail probability \( \epsilon \). Therefore, AVaR is symmetric, even though by construction it is not.

In the stable case, \( C_\epsilon \) depends on the tail index and skewness parameter.
Denote the observed portfolio returns by $r_1, \ldots, r_n$ (in order of occurrence). Denote the returns sorted in ascending order by $r^{(1)}, \ldots, r^{(n)}$.

Suppose we do not impose any distributional assumptions on $r$. Then, AVaR at tail probability $\epsilon$ is estimated according to the formula\(^1\)

$$\hat{\text{AVaR}}_{\epsilon}(r) = -\frac{1}{\epsilon} \left( \frac{1}{n} \sum_{k=1}^{\lceil n\epsilon \rceil - 1} r^{(k)} + \left( \epsilon - \frac{\lceil n\epsilon \rceil - 1}{n} \right) r^{(\lceil n\epsilon \rceil)} \right),$$

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

The formula above can also be applied when the statistical model we work with does not allow a closed-form AVaR expression. Then, we can simply generate scenarios from the distribution and estimate AVaR using those simulations.

---

\(^1\)For a detailed description and derivation, see Rockafellar and Uryasev (2002).
The ideas behind portfolio VaR estimation can be applied to AVaR.

As before, assume that the returns on the $n$ stocks are denoted by $X_1, \ldots, X_n$, so that the portfolio return is $r_p = w_1 X_1 + \ldots w_n X_n$.

**Multivariate Normal Assumption**

AVaR of portfolio return at tail probability $\epsilon$ can be expressed analytically as

$$\text{AVaR}_\epsilon(r_p) = \frac{\sqrt{w' \Sigma w}}{\epsilon \sqrt{2 \pi}} \exp \left( - \frac{(\text{VaR}_\epsilon(Y))^2}{2} - \mu_p \right)$$

$$= C_\epsilon \sqrt{w' \Sigma w} - \mu_p.$$
Historical Method

- Not related to any distributional assumptions. Use historically observed returns and apply the non-parametric approach to AVaR calculation above.

- Very inaccurate for low-tail probabilities. The AVaR estimate is not robust: when the sample changes, the smallest observations tend to fluctuate the most.
Portfolio Average Value-at-Risk

Hybrid Method

- Weights are assigned to observed returns, according to an exponential smoothing scheme, as explained in the VaR estimation discussion.

- Since the weights are interpreted as probabilities, the portfolio AVaR can be estimated from the resulting discrete distribution according to

  \[
  \widehat{\text{AVaR}}_\epsilon(r_p) = -\frac{1}{\epsilon} \left( \sum_{j=1}^{k_\epsilon} p_j r(j) + \left( \epsilon - \sum_{j=1}^{k_\epsilon} p_j \right) r(k_\epsilon+1) \right),
  \]

  where \( r(1) \leq \cdots \leq r(k_m) \) denote the sorted sample of returns and \( p_1, \ldots, p_{k_m} \) denote the probabilities of the sorted observations. The number \( k_\epsilon \) is an integer satisfying the inequalities

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

  \[1\]Rockafellar and Uryasev (2002).
Outline: Risk Measures, Risk Management, and Optimization

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Mean-variance analysis (MV analysis) (referred to as modern portfolio theory (MPT)) is centered on two criteria: expected portfolio return and variance of portfolio return (degree of diversification).\textsuperscript{1}

The portfolio choice problem is typically a one-period (static) problem: at time $t_0$ the investor makes an investment decision and keeps the allocation unchanged until time $t_1$. At that time, a new decision may be made. In contrast, in a dynamic setting, investment decisions are made for several time periods ahead.

The main principle of MV analysis can be expressed in two equivalent ways:

From all feasible portfolios,

\begin{itemize}
  \item find the ones with minimum variance, for a given lower bound on the expected return.
  \item find the ones with maximum expected performance, for a given upper bound on the portfolio variance.
\end{itemize}

\textsuperscript{1}Markowitz (1952, 1959).
Mean-Variance Optimization

The two MV principle formulations give rise to two optimization problems.

First, we introduce some necessary notation:

- Returns on the $n$ financial assets: $X' = (X_1, \ldots, X_n)$.
- Mean of returns: $\mu' = (\mu_1, \ldots, \mu_n)$.
- Covariance between returns: $\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$.
- Portfolio weights: $w' = (w_1, \ldots, w_n)$.

Portfolio attributes:

- Expected portfolio return: $Er_p = \sum_{i=1}^{n} w_i \mu_i = w' \mu$.
- Portfolio return variance: $\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} = w' \Sigma w$. 
The optimization problem behind the first formulation is given by

$$\min_w \ w' \Sigma w$$

subject to

$$w' e = 1$$

$$w' \mu \geq R_*$$

$$w \geq 0,$$

where $w \geq 0$ means all weights are nonnegative (we consider long-only strategies) and $e' = (1, \ldots, 1)$.

The objective function is the variance of portfolio returns and $R_*$ is the lower bound on the expected performance (so-called "required return").
The second formulation of the MV principle is formalized as

\[
\max_{\mathbf{w}} \quad \mathbf{w}' \boldsymbol{\mu}
\]

subject to

\[
\mathbf{w}' \mathbf{e} = 1 \\
\mathbf{w}' \mathbf{\Sigma} \mathbf{w} \leq R^* \\
\mathbf{w} \geq 0,
\]

where \( R^* \) is the upper bound on the variance of the portfolio return.
The set of all optimal portfolios is known as *mean-variance efficient frontier*. It is obtained by solving the optimization problems for varying constraints (on the required expected performance or maximum admissible portfolio risk). For example,

- Suppose we solve the first optimization problem without any constraint on the expected performance. We obtain the *global minimum variance portfolio* ⇒ most diversified portfolio but lowest expected performance.

- Next, we include a constraint on required return and gradually raise the lower bound to obtain optimal portfolios that are less and less diversified but with increasingly high expected performance.

- The portfolio with the highest expected performance also has the highest concentration. It is composed only of the asset with the highest expected return.

The efficient frontier can be obtained from the second problem as well, by varying the upper bound on the variance.
Mean-Variance Efficient Frontier

- Initial, suboptimal portfolio. Two ways to reach the efficient frontier:

1. Solve first problem:
   \[ R^* = \text{portfolio’s expected return}. \]
   \[ \Rightarrow \] Optimal solution: pointed by the horizontal arrow.

2. Solve second problem:
   \[ R^* = \text{portfolio’s variance}. \]
   \[ \Rightarrow \] Optimal solution: pointed by the vertical arrow.

Arc enclosed by the two arrows: portfolios that are more efficient than the initial portfolio (lower variance and higher expected performance).
If we add a risk-free asset to the investment universe, the efficient frontier changes. Efficient portfolios derived in this case are superior to those available to investors without the risk-free asset.

The efficient portfolios consist of a combination of a particular portfolio of risky assets, called market portfolio, and the risk-free asset.

The first portfolio optimization problem transforms into

\[
\min_{w, w_f} w' \Sigma w
\]

subject to

\[
w' e + w_f = 1 \\
w' \mu + w_f r_f \geq R_* \\
w \geq 0, w_f \leq 1,
\]

where \( w_f \) is the risk-free asset's weight, which can be positive or negative. The second optimization problem can be modified as well accordingly.

Changing the lower bound \( R_* \) results in an optimal portfolio with different relative proportions of the same portfolio of risky assets and the risk-free asset.
The portfolio of risky assets is known as the market portfolio, 
\[ w_M = (w_{M1}, \ldots, w_{Mn}) \], and its weights sum up to one.

All efficient portfolios can be represented as 
\[ r_p = aw'_M X + (1 - a)r_f, \]
where \( a \) denotes the proportion of the market portfolio in the efficient portfolio.

The market portfolio is located on the efficient frontier, where a straight line passing through the risk-free rate is tangent to the efficient frontier.

The straight line is called \textit{capital market line} and the market portfolio is also known as the \textit{tangency portfolio}.
The capital market line equation can be derived as

$$E(r_p) = r_f + \left( \frac{E(r_M) - r_f}{\sigma_{RM}} \right) \sigma_{r_p}. $$

The efficient frontier in a universe with a risk-free asset is transformed to the capital market line.

The ratio $\frac{E(r_M) - r_f}{\sigma_{RM}}$ is called the *Sharpe ratio*.

The tangency portfolio is the portfolio with the maximum Sharpe ratio.
MV Analysis: Implications

In general, it can be shown that the MV analysis approach is the natural approach when an elliptical distributional assumption for universe of asset returns is made. Here is what this means:

- Suppose there is a risky portfolio with composition $w = (w_1, \ldots, w_n)$ which is preferred by all risk-averse investors over another portfolio $\nu = (\nu_1, \ldots, \nu_n)$. That is, $w$ dominates $\nu$.

- Does MV analysis identify $\nu$ as not more efficient than $w$? I.e., is it the case that $\nu' \mu \leq w' \mu$ and $\nu' \Sigma \nu \geq w' \Sigma w$?

Answer: No, in general. Yes, if, for e.g., a multivariate normal distribution $N(\mu, \Sigma)$ is assumed.

Alternatively, MV analysis is consistent with investors having risk-return preferences that can be described with a quadratic utility function.
The standard deviation of portfolio returns can only be used as a proxy for risk, as it is not a true risk measure.

If we employ a true risk measure and study the optimal tradeoff between risk and return, we obtain an extension of the MV analysis’ framework called mean-risk analysis (MR analysis).¹

The main principle of MR analysis can be formulated in a similar way to MV analysis. From all feasible portfolios,

- Find the ones with minimum risk, for a given required return.
- Find the ones with maximum expected performance, for a given upper bound on risk.

The key input is the particular risk measure we would like to employ. We denote the generic risk measure by $\rho(X)$.

¹Rachev, Stoyanov, Fabozzi (2008).
Two corresponding optimization problems can be formulated:

\[
\begin{align*}
\min_{w} & \quad \rho(r_p) \\
\text{subject to} & \quad w' e = 1 \\
& \quad w' \mu \geq R^* \\
& \quad w \geq 0
\end{align*}
\]

\[
\begin{align*}
\max_{w} & \quad w' \mu \\
\text{subject to} & \quad w' e = 1 \\
& \quad \rho(r_p) \leq R^* \\
& \quad w \geq 0.
\end{align*}
\]

- In order to calculate the risk \( \rho(r_p) \), we need to know (make an explicit assumption about) the multivariate distribution of asset returns.

- In MV framework, this requirement is not so obvious, since there we only need as an input the covariance matrix.

- However, as we saw, MV analysis leads to reasonable decision-making only under certain distributional hypotheses, such as multivariate normal.

- In MR analysis, the risk measure \( \rho \) may capture completely different characteristics of the portfolio return distribution.
Outline: Risk Measures, Risk Management, and Optimization

- Classical Risk Measures and Estimation Techniques
- Coherent Risk Measures
- Fat-Tailed Extensions to Mean-Variance Optimization
- ETL (AVaR) Optimization
- Benchmark Tracking Problems
- Alternative Performance Measures
The choice of AVaR as a risk measure allows certain simplifications of the optimization problems, especially if there are available scenarios for asset returns.

Denote the scenarios for asset returns by \( r^1, \ldots, r^k \), where \( r^j \) is a vector of observations \( r^j = (r^j_1, \ldots, r^j_n) \).

Denote the returns of all assets observed in a given time period denoted by the index \( j \). Thus, all observations can be arranged in a \( k \times n \) matrix

\[
H = \begin{pmatrix}
  r^1_1 & r^1_2 & \cdots & r^1_n \\
  r^2_1 & r^2_2 & \cdots & r^2_n \\
  \vdots & \vdots & \ddots & \vdots \\
  r^k_1 & r^k_2 & \cdots & r^k_n 
\end{pmatrix}.
\]

\( H \) may not only be a matrix of observed returns. It could contain the \textit{i.i.d.} scenarios produced by any multivariate model (including a copula-based model). Then, \( k \) denotes the number of multivariate scenarios and \( n \)—the number of random asset returns.
The observed returns of a portfolio with composition $w$ are $r^1 w, \ldots, r^n w$, or simply $Hw$.

Recall the second definition of $\text{AVaR}_\epsilon$ we considered earlier,

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

This minimization formula can be restated as a linear optimization problem by introducing auxiliary variables $d_1, \ldots, d_n$, one for each observation in the sample,

$$\min_{\theta, d} \quad \theta + \frac{1}{n \epsilon} \sum_{k=1}^{n} d_k$$

subject to

$$- r_k - \theta \leq d_k, \quad k = 1, \ldots, n$$

$$d_k \geq 0, \quad k = 1, \ldots, n$$

$$\theta \in \mathbb{R}.$$
Then, the AVaR optimization problem is expressed in the following way (using matrix notation):

\[
\begin{align*}
\min_{\theta, d, w} & \quad \theta + \frac{1}{k \epsilon} d' e \\
\text{subject to} & \quad - Hw - \theta e \leq d \\
& \quad w' e = 1 \\
& \quad w' \mu \geq R_* \\
& \quad w \geq 0, d \geq 0, \theta \in \mathbb{R},
\end{align*}
\]

where \( d' = (d_1, \ldots, d_k) \) and \( e = (1, \ldots, 1) \).

Even though the optimization problem may seem involved because of the matrix notation, it has a simple structure. The objective function is linear and the constraints are linear (in)equalities. There are efficient algorithms for solving such linear programming problems.
Three stocks.
Data: 12/31/02—12/31/03.

\[ \mu_1 = 0.17\%, \mu_2 = 0.09\%, \mu_3 = 0.03\% . \]

Only 250 data, so set \( \epsilon = 40\% \) ⇒ greater stability of \( \hat{AVaR}_\epsilon \).

Top plot: Efficient frontier.

AVaR for global minimum risk portfolio \( \approx 1.5\% \). AVaR for maximum expected return portfolio \( \approx 2.8\% \).

Bottom plot: Compositions of efficient portfolios along frontier.

\( w_1 \) increases along frontier.
Outline: Risk Measures, Risk Management, and Optimization

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Benchmark-tracking problems include a benchmark portfolio against which the performance of the managed portfolio is compared.

The arising optimization problems include the distribution of the active portfolio return, defined as the difference $r_p - r_b$, in which $r_b$ denotes the return of the benchmark.

A measure of performance of the portfolio relative to the benchmark is the average active return, also known as the portfolio alpha and denoted by $\alpha_p$. It is calculated as the difference in the sample means: $\hat{\alpha}_p = \bar{r}_p - \bar{r}_b$.

A widely used measure of how close the portfolio returns are to the benchmark returns is the standard deviation of the active return, also known as tracking error (TE). The closer the TE is to zero, the closer the risk profile of the portfolio matches that of the benchmark.

In ex ante analysis, portfolio alpha is the expectation of the active return and the TE is the standard deviation of the active return. In this case, the later is referred to as forward-looking TE.
The Tracking Error Problem

- Active portfolio strategies are characterized by high alphas and TEs, while passive strategies—by very small alphas and TEs. In between are enhanced indexing strategies, with small to medium-sized alphas and TEs.

- The optimal portfolio problem originating from this framework is the minimal TE problem, which has the same structure as MV optimization problems:¹

\[
\begin{align*}
\min_w & \quad \sigma(r_p - r_b) \\
\text{subject to} & \quad w' e = 1 \\
& \quad w' \mu - Er_b \geq R_* \\
& \quad w \geq 0,
\end{align*}
\]

where \( R_* \) denotes the lower bound of the expected alpha.

- The goal is to find a portfolio that is closest to the benchmark in a certain sense, while setting a threshold on the expected alpha. In this case, the "closeness" is determined by the standard deviation.

¹Roll (1992) provides a mean-variance analysis of the tracking error.
The Tracking Error Problem

By varying the limit $R_*$, we obtain the whole spectrum from passive ($R_*$ close to zero), through enhanced indexing ($R_*$ taking medium-sized values), to active strategies ($R_*$ with medium-sized to large values).

- **Plot**: Efficient frontier.
- If investment universe same or larger than benchmark portfolio, global min TE portfolio $= 0$.
- Optimal portfolio $=$ benchmark portfolio.
- Global min TE portfolio represents a passive strategy.
The Tracking Error Problem

- The tracking error suffers from the same disadvantages as the standard deviation. It treats in the same way underperformance and overperformance, while our attitude towards them is asymmetric.

- From an asset management perspective, then, the measure of "closeness" of the managed portfolio to the benchmark should be asymmetric.

- Therefore, the aim is to restate the minimum TE problem in the more general form\(^1\)

\[
\begin{align*}
\min_w & \quad \mu(r_p, r_b) \\
\text{subject to} & \quad w'e = 1 \\
& \quad w'\mu - Er_b \geq R_* \\
& \quad w \geq 0,
\end{align*}
\]

where \(\mu(X, Y)\) is a measure of the deviation of \(X\) relative to \(Y\). That is, we regard \(\mu\) as a function that measures relative deviation and we call it relative deviation metric (\textit{r.d. metric}).

\(^1\)For details, see Rachev, Stoyanov, and Fabozzi (2008).
Benchmark Tracking Problem: r.d. Metrics

- We briefly discuss two r.d metric that have been proposed in the context of financial applications:\(^1\)

\[
\theta^*(X, Y) = \int_{-\infty}^{\infty} \left( \max(F_X(t) - F_Y(t), 0) \right) dt
\]

and

\[
l^*(X, Y) = \int_{0}^{1} \left( \max(F_X^{-1}(t) - F_Y^{-1}(t), 0) \right) dt,
\]

where \(X\) and \(Y\) are zero-mean random variables, \(F_X\) is the CDF of \(X\) and \(F_X^{-1}\) is the inverse of the CDF.

- Both functionals measure the relative deviation of \(X\) and \(Y\) using only the part of the CDFs, or the inverse CDFs, that describes losses.

- If the distributions of \(X\) and \(Y\) are skewed, this is reflected in the functionals, by construction.

\(^1\)See Stoyanov, Rachev, Ortobelli, Fabozzi (2007).
Benchmark Tracking Problem: r.d. Metrics

Intuition behind the r.d. Metrics

- The difference $F_X(t) - F_Y(t)$ is non-negative only for negative values of $t$ \(\Rightarrow \theta^*\) uses the information about losses contained in the CDF. Same for $l^*$. 
- $\theta^*(X, Y) = l^*(X, Y)$.
- The two CDFs coincide if and only if $\theta^*(X, Y) = l^*(X, Y) = 0$.
- In some special cases, for e.g., Gaussian distributional for both $X$ and $Y$, the r.d. metrics can be calculated explicitly.
We illustrate the difference between the optimal solutions of (1) the classical tracking error and (2) the r.d. metric \( l^*(X, Y) \).

The optimal solution is a portfolio whose empirical CDF is closest to the empirical CDF of the benchmark.

- **Data:** 10 S&P 500 stocks; 12/31/02—12/31/03.
- **Benchmark:** S&P 500 index.
- **Solve benchmark-tracking problems for** \( R_* = 0 \), using centered returns.
- **Initial portfolio:** Equal weighting.
- **Plot:** Inverse CDFs of initial portfolio and benchmark.
Both optimizations lead to better solutions than trivial equal-weighting strategy.

r.d. metric better at approximating CDF of benchmark’s returns.

r.d. metric allows for asymmetries in loss versus profit part of CDF.
Outline: Risk Measures, Risk Management, and Optimization

Classical Risk Measures and Estimation Techniques

Coherent Risk Measures

Fat-Tailed Extensions to Mean-Variance Optimization

ETL (AVaR) Optimization

Benchmark Tracking Problems

Alternative Performance Measures
Performance Measures

- Measurement and evaluation of portfolio performance key in investment management process. The formula that quantifies the portfolio performance is called a *performance measure*.¹

- Measuring a strategy’s performance is an ex post analysis. The performance measure is calculated using the realized portfolio returns during a specifies period back in time.

- Alternatively, performance measures can be used in an ex ante analysis, in which certain assumptions about the future behavior of assets must be made. The goal is to find a portfolio with the best characteristics, as calculated by the performance measure.

---

¹A widely used performance measure is the Sharpe ratio (Sharpe (1966))—ratio between average active portfolio return and standard deviation of portfolio return.
Reward-to-Risk Ratios

- A general type of performance measure is the *reward-to-risk* (RR) *ratio*.
- Defined as ratio between a reward measure of the active portfolio return and the risk of the active portfolio return,

\[
RR(r_p) = \frac{\nu(r_p - r_b)}{\rho(r_p - r_b)},
\]

where \( r_p = w'X \) is the return of portfolio with weights \( w \) and returns described by the random vector \( X \). The reward and risk measures are denoted by \( \nu \) and \( \rho \), respectively.

- The benchmark return can be either a fixed target or a random variable (return of another portfolio or reference interest rate).

- In ex ante analysis, a portfolio optimization problem can be solved in which the objective function is the RR ratio. Before that, the joint distribution of the portfolio return and the benchmark return must be hypothesized and calibrated.
**STARR** stands for stable tail-adjusted return ratio.\(^1\) In it, AVaR is selected as a risk measure,

\[
\text{STARR}_\epsilon (w) = \frac{\mathbb{E}(r_p - r_b)}{\text{AVaR}_\epsilon (r_p - r_b)}.
\]

If \(r_b\) is a constant benchmark return, STARR equals

\[
\text{STARR}_\epsilon (w) = \frac{w' \mu - r_b}{\text{AVaR}_\epsilon (r_p) + r_b}.
\]

The value of \(\epsilon\) chosen in practical applications depends on the extent to which we want to emphasize the tail risk in the comparison. For instance,

- \(\epsilon = 0.01 \Rightarrow\) compare the average portfolio return per unit of the extreme average realized losses.
- \(\epsilon = 0.5 \Rightarrow\) compare average portfolio return per unit of total average realized loss.

---

\(^1\) More details in Rachev, Martin, Racheva-Yotova, Stoyanov (2006).
If all empirical AVaRs are positive, then the portfolio with the highest STARR had the best performance over the comparison period.

Notice that it is always possible, if expected return is positive, to find $\epsilon$ for which $\text{AVaR}_\epsilon$ is negative.

In practice, if daily returns are used, the empirical AVaR at $\epsilon \leq 0.5$ is very rarely negative. With monthly returns, this may not be the case.

If there are portfolios with negative AVaRs, then all portfolios should be divided into two groups—one with nonpositive and one with nonnegative AVaRs. Separate ranking should be performed. Additionally, one could argue that the portfolios in the former group have a better performance, since negative risk implies no reserve capital should be allocated.

Thus, if risk is negative, smaller STARR indicates better performance.
The Sortino ratio is defined as the ratio between the expected active portfolio return and the semistandard deviation of the underperformance of a fixed target level $s$. If $r_b$ is constant, the ratio is expressed as

$$\text{SoR}_s(w) = \frac{w'\mu - r_b}{(E(s - r_p)^2)^{1/2}},$$

where $(x)^2_+ = (\max(x, 0))^2$.

- The fixed target $s$ is also called the *minimum acceptable return level*. For e.g., it could be set equal to $r_b$.
- The function in the denominator is a proxy for portfolio risk, however it is not a coherent risk measure.
- In ex post analysis, the Sortino ratio is calculated by substituting expectations with the sample means.
The Rachev ratio uses a reward measure which is not the expectation of active portfolio returns. Instead, reward is defined as the average of the quantiles of the portfolio return distribution that are above a certain target return level.

Formally, this performance measure is defined as

$$RaR_{\epsilon_1, \epsilon_2}(w) = \frac{AVaR_{\epsilon_1}(r_b - r_p)}{AVaR_{\epsilon_2}(r_p - r_b)}.$$ 

Even though AVaR is used in the numerator, the numerator represents a measure of reward because

$$AVaR_{\epsilon_1}(X) = -\frac{1}{\epsilon} \int_0^{\epsilon_1} F_X^{-1}(p) dp = \frac{1}{\epsilon} \int_{1-\epsilon_1}^{1} F_X^{-1}(p) dp,$$

where $X = r_b - r_p$ can be interpreted as benchmark underperformance and $-X$ is the active portfolio return.
Reward-to-Risk Ratios: The Rachev Ratio

The numerator in the Rachev ratio can be interpreted as the average outperformance of the benchmark provided the outperformance is larger than the quantile at $1 - \epsilon_1$ probability of the active return distribution.

Thus, there are two performance levels in the Rachev ratio:\footnote{Biglova, Ortobelli, Rachev, Stoyanov (2004) provide an empirical example.}

- The quantile at probability $\epsilon_2$ in the denominator.
  ⇒ If active return below this quantile, it is counted as a loss.

- The quantile at probability $1 - \epsilon_1$ in the denominator.
  ⇒ If active return is above this quantile, it is counted as reward.
References


Prof. Dr. Zari Rachev  
Risk Measures/Management & Optimization  
Next Generation Risk Management