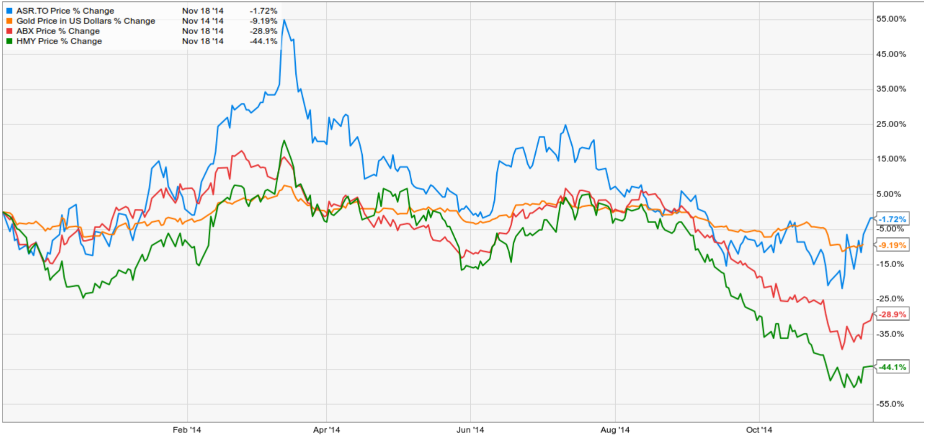
Applications of Linear Programming

lecturer: András London

University of Szeged
Institute of Informatics
Department of Computational Optimization

Lecture 8

The portfolio selection problem



The portfolio selection problem

Given a set of assets (bond, stock, activity, etc.)

Question: How to compose a portfolio of them?

Suppose that the expected return of an investment in stock r is $\mathbb{E}(r)$ (That is an expected value calculated from historical time series data *in some way*)

Goal: Compose a portfolio with maximum expected return. If we invest in n different assets, then

an **LP model** for the problem:

$$\sum_{i=1}^{n} x_i = 1 \qquad \text{[total capital]}$$

$$x_i \geq 0 \qquad \text{[portion invested in } r_i\text{]}$$

$$\max \sum_{i=1}^{n} \mathbb{E}(r_i) x_i \qquad \text{[total expected return]}$$

The portfolio selection problem

- We can assume that $\mathbb{E}(r_1) \geq \mathbb{E}(r_2) \geq \ldots \geq \mathbb{E}(r_n)$, then the optimal solution is $x_1^* = 1$, $x_2^* = \cdots = x_n^* = 0$, and the return is $\mathbb{E}(r_1)$.
- Generally true that if one follows this strategy will go bankrupt with probability 1. why?
- Unfortunately, the return on stocks that yield a **large expected return is usually highly variable**
- Thus, one often approaches the problem of selecting a portfolio by choosing an acceptable minimum expected return and **finding the portfolio with the minimum variance**

$\Rightarrow$ several approach developed (moreover, research on the topic is still very intensive) to solve the problem. Here we discuss two of them:

1. **Markowitz model**, 1952; Nobel-prize in Economy, 1990
2. **MAD model** (Konno and Yamazaki, 1990)

Portfolio selection – example

Let $\sigma(r)$ be the risk of investment in asset r (will be measured by the variance calculated from historical time series data).

Consider the returns of the following investments in the past 3 years:

	Year 1	Year 2	Year 3
Property (e.g. a house)	0.05	-0.03	0.04
Security (e.g. a bond)	-0.05	0.21	-0.10

The expected **returns** are calculated as:

$$\mathbb{E}(r_h) = \frac{0.05-0.03+0.04}{3} = 0.02 \text{ és } \mathbb{E}(r_b) = \frac{-0.05+0.21-0.10}{3} = 0.02$$

The **risks** are:

$$\sigma(r_h) = \sqrt{\frac{(0.02-0.05)^2+(0.02+0.03)^2+(0.02-0.04)^2}{3}} \approx 0.036 \text{ and}$$

$$\sigma(r_b) = \sqrt{\frac{(0.02+0.05)^2+(0.02-0.21)^2+(0.02+0.10)^2}{3}} \approx 0.164$$

Portfolio selection – example

If we invest 75% of the capital to the house and 25% to bond then the **return of the portfolio** is

$$\mathbb{E}(r_p) = \frac{(0.75 \cdot 0.05 + 0.25 \cdot -0.05)}{3} + \frac{(0.75 \cdot -0.03 + 0.25 \cdot 0.21)}{3} + \frac{(0.75 \cdot 0.04 + 0.25 \cdot -0.10)}{3} = \frac{(0.025 + 0.03 + 0.005)}{3} = 0.02$$

The portfolio risk is

$$\sigma(r_p) = \sqrt{\frac{(0.02-0.025)^2+(0.02-0.03)^2+(0.02-0.005)^2}{3}} \approx 0.019$$

However the **average risk** is $0.75 \cdot 0.036 + 0.25 \cdot 0.164 = 0.068$

$\Rightarrow$ Diversification reduces the risk

Portfolio selection – example

	Year 1	Year 2	Year 3
Property	0.05	-0.03	0.04
Security	-0.05	0.21	-0.10

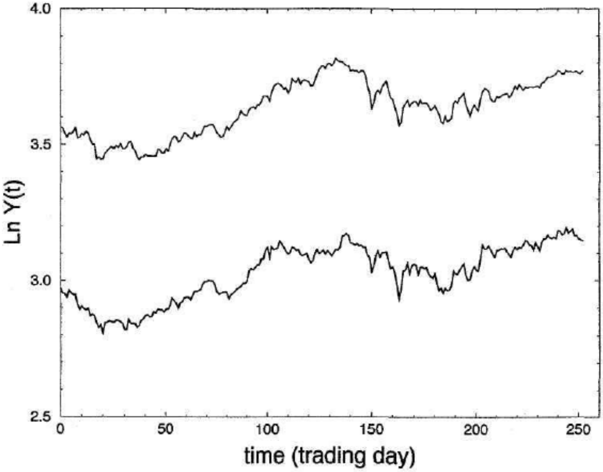
Covariance: is a measure of the joint variability of two random variables:

$$\mathrm{cov}_{p,s} = \frac{(0.02-0.05)\cdot(0.02+0.05)}{3} + \frac{(0.02+0.03)\cdot(0.02-0.21)}{3} + \\ + \frac{(0.02-0.04)\cdot(0.02+0.10)}{3} = -0.005$$

Correlation: Normalized covariance $\mathrm{corr}_{i,e} = \frac{-0.005}{0.036\cdot 0.164} = -0.84$

- $-1 \leq \mathrm{corr} \leq 1$
- $\mathrm{corr} > 0$ positive correlation
- $\mathrm{corr} = 0$ no correlation ($\sim$ independence, but $\neq$ independence)
- $\mathrm{corr} < 0$ anticorrelation

Portfolio selection – example



ábra. Exchange rate of Coca-Cola and Procter&Gamble in 1990

Expected value, variance, covariance

Basic properties:

$$
\begin{aligned}
E(\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n) &= E(\mathbf{X}_1) + E(\mathbf{X}_2) + \cdots + E(\mathbf{X}_n) \\
\text{var}\,(\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n) &= \text{var}\,\mathbf{X}_1 + \text{var}\,\mathbf{X}_2 + \cdots + \text{var}\,\mathbf{X}_n + \sum_{i \neq j} \text{cov}(\mathbf{X}_i, \mathbf{X}_j) \\
E(k\mathbf{X}_i) &= kE(\mathbf{X}_i) \\
\text{var}\,(k\mathbf{X}_i) &= k^2 \text{var}\,\mathbf{X}_i \\
\text{cov}(a\mathbf{X}_i, b\mathbf{X}_j) &= ab\,\text{cov}\,(\mathbf{X}_i, \mathbf{X}_j)
\end{aligned}
$$

Portfolio selection – Markowitz model

The **general model**:

- $(r_1, r_2, \ldots, r_n)$ – the assets in the portfolio
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ – portion of the capital invested in each individual asset
- $\sum_{i=1}^{n} x_i = 1$ and $x_i \geq 0$ $(\forall i)$

Risk: measured via the variance (squared deviation, i.e. $\text{var} = \sigma^2$)

Covariance matrix: contains the pairwise covariances of the (historical daily) stock returns

$$\mathbf{C} = \begin{pmatrix} \text{cov}_{11} & \text{cov}_{12} & \cdots & \text{cov}_{1n} \\ \text{cov}_{21} & \text{cov}_{22} & \cdots & \text{cov}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}_{n1} & \text{cov}_{n2} & \cdots & \text{cov}_{nn} \end{pmatrix}$$

$\text{cov}_{ii} = \sigma^2(r_i) = \text{var}(r_i)$

Portfolio selection – Markowitz model

A Portfolio risk:

$$\mathrm{var}\left(\sum_{i=1}^{n}\mathbb{E}(r_i)x_i\right)=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\mathrm{cov}_{ij}x_ix_j\right)=\mathbf{x}^T\mathbf{C}\mathbf{x}$$

Efficient portfolio: A portfolio that

- provides the greatest expected return for a given level of risk,
- or equivalently, the lowest risk for a given expected return.

Portfolio selection – Markowitz model

Let R be the **minimum expected return** of an investment. We can formulate the following quadratic programming problem:

$$\sum_{i=1}^{n} \mathbb{E}(r_i)x_i \geq R$$

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0 \qquad i = 1,2,\ldots,n$$

$$\min \mathbf{x}^T \mathbf{C} \mathbf{x}$$

That is **minimizing the risk** given a **minimum expected return**.

A solution of the problem called optimal portfolio.

Portfolio selection – Markowitz model

Remarks:

- Non-linear (e.g. quadratic) optimization
- Efficient algorithms exist for solving such problems
- A difficulty: computing (estimating) the elements of the covariance matrix $\mathbf{C}$
- instead, we can maximize e.g. the mean absolute error $\mathbb{E}(|\sum_i (r_i - \mathbf{E}(r_i))x_i|)$[1]

[1] if $r = (r_1, \ldots, r_n)$ follows a multivariate normal distribution then the two method is equivalent

Portfolio selection – MAD model

- **M**ean **A**bsolute **D**eviation
- Developed by Konno and Yamazaki **uses the observed data directly** and avoids the calculation of $\mathbb{E}(r_i)$ and $\mathbf{C}$
- Let T be number of observations (closure prices of T days) of n investments and let r_{it} be the observation of the return of investment i
- Let

$$r_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} \text{ és } a_{it} = r_{it} - r_i$$

be the observed average return, and the difference of the individual returns from the average, respectively

Portfolio selection – MAD model

The following **optimization problem** can be given:

$$\sum_{i=1}^{n} r_i x_i \geq R$$

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0 \qquad i = 1,2,\dots,n$$

$$\min \ \frac{1}{T} \sum_{t=1}^{T} \left| \sum_{i=1}^{n} a_{it} x_i \right|$$

This is not an LP, but can be formulated as LP.

Portfolio selection – MAD model

MAD model as LP:

$$\sum_{i=1}^{n} a_{it} x_i \geq -y_t \qquad t = 1,2,\ldots,T$$

$$\sum_{i=1}^{n} a_{it} x_i \leq y_t \qquad t = 1,2,\ldots,T$$

$$\sum_{i=1}^{n} r_i x_i \geq R$$

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0 \qquad i = 1,2,\ldots,n$$

$$y_t \geq 0 \qquad t = 1,2,\ldots,T$$

$$\min \ \frac{1}{T} \sum_{t=1}^{T} y_t$$

Inventory model[2]

Suppose that a company has to decide an order quantity x of a certain product (newspaper) to satisfy demand d. The cost of ordering is $c > 0$ per unit.

If the demand d is bigger than x, then a back order penalty of $b \geq 0$ per unit is incurred. The cost of this is equal to $b(d - x)$ if $d > x$, and is zero otherwise. On the other hand if $d < x$, then a holding cost of $h(x - d) \geq 0$ is incurred. The total cost is then

$$G(x, d) = cx + b[d - x]_+ + h[x - d]_+,$$

where $[a]_+ = \max\{a, 0\}$. We assume that $b > c$

The objective is to minimize $G(x, d)$. Here x is the decision variable and the demand d is a parameter

[2] based on the lecture notes *A Tutorial on Stochastic Programming* by Alexander Shapiro and Andy Philpott

Inventory model

The non-negativity constraint $x \geq 0$ can be removed if a back order policy is allowed. The objective function $G(x, d)$ can be rewritten as

$$G(x, d) = \max\{(c - b)x + bd, (c + h)x - hd\}$$

which is **piecewise linear** with a minimum attained at $x = d$.

That is, if the demand d is known, then (no surprises) the **best decision** is to order exactly the demand quantity d.

We can write the problem as an LP

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & t \geq (c - b)x + bd, \\ & t \geq (c + h)x - hd, \\ & x \geq 0. \end{aligned}$$

Inventory model

For a numerical instance suppose $c = 1$, $b = 1.5$, and $h = 0.1$. Then

$$G(x,d) = \begin{cases} -0.5x + 1.5d, & \text{if } x < d \\ 1.1x - 0.1d, & \text{if } x \geq d. \end{cases}$$

Let $d = 50$. Then $G(x, 50)$ is the pointwise maximum of the linear functions plotted in Figure 1.

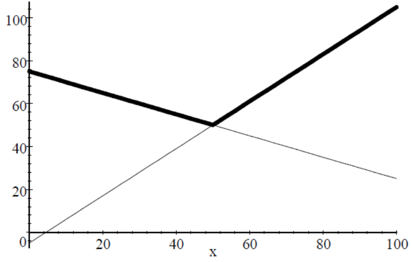


Figure 1: Plot of $G(x, 50)$. Its minimum is at $\bar{x} = 50$

Inventory model

- Consider now the case when the ordering decision should be made before a realization of the demand becomes known.
- One possible way to proceed in such situation is to view the demand D as a **random variable**
- We assume, further, that the *probability distribution* of D *is known* (e.g. estimated from historical data)
- We can consider the expected value $\mathbb{E}[G(x, D)]$ and corresponding optimization problem

$$\min_{x \geq 0} \mathbb{E}[G(x, D)].$$

- it means **optimizing** (minimizing) the total cost **on average**

Inventory model

- Suppose that D has a finitely supported distribution, i.e., it takes values $d_1, \ldots, d_K$ (called scenarios) with respective probabilities $p_1, \ldots, p_K$.
- Then the expected value can be written as

$$\mathbb{E}[G(x, D)] = \sum_{k=1}^{K} p_k G(x, d_k)$$

- The expected value problem can be written as the linear programming problem:

$$\begin{array}{ll} \min\limits_{x, t_1, \ldots, t_K} & \sum_{k=1}^{K} p_k t_k \\ \text{s.t.} & t_k \geq (c-b)x + b d_k, \;\; k = 1, \ldots, K, \\ & t_k \geq (c+h)x - h d_k, \;\; k = 1, \ldots, K, \\ & x \geq 0. \end{array}$$

- We stop here the discussion!