Lecture Notes MTH6113: Mathematical Tools for Asset Management*

Dr Linus Wunderlich
thanks to Dr Kathrin Glau

January 26, 2023

Contents

*We try to ensure that these lecture notes include all the material covered. Unfortunately, this might not always be the case and there may be mistakes or differences to the lectures. Please email L.Wunderlich@qmul.ac.uk if you note any mistakes. For the exam and in-term assessments only the material presented in the lectures is relevant.

0 Preliminaries (Week 1)

Preliminary remark: This module differs from most other mathematical modules in that we are exploring mathematics as a tool for financial purposes in this lecture rather than an aim. This means that an understanding of the material is not possible by understanding the mathematical bits alone. We have to understand the financial context. Ultimately, an equation or inequality that we encounter in this lecture is not interesting as such, but we have to understand its financial meaning. Chapter 1 gives us a first flavour, as we here introduce some basic lines of thoughts in financial terms.

For organizational issues and preliminaries such as repetition from probability theory and basic information on the financial market see the slides. Here, we give a basic list of notions from probability.

Revision of probability theory Given random variables X, Y, we consider

- the expected value $\mathbb{E}(X)$:
  - $\sum_i x_i P(X = x_i)$ for discrete variables with possible values x_i;
  - $\int_{\mathbb{R}} x f_X(x) \, \mathrm{d}x$ for the probability density function (pdf) f_X.
- the variance $\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$;
- the standard deviation $\sigma_X = \sqrt{\mathrm{Var}(X)}$;
- the covariance $\mathrm{Cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)$;
- the correlation $\mathrm{corr}(X, Y) = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$;
- the distribution function $F_X(x) = P(X \leq x)$;
  - for continuous variables, this is the integral of the density function: $F_X(x) = \int_{-\infty}^{x} f_X(\xi) \, \mathrm{d}\xi$.

With random variables X, Y, Z and a deterministic scalar a, we frequently use:

- linearity of the expected value: $\mathbb{E}(aX + Y) = a\,\mathbb{E}(X) + \mathbb{E}(Y)$;
- variance as the covariance with itself: $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$;
- symmetry and scaling of the covariance $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ and $\mathrm{Cov}(aX, Y) = a\,\mathrm{Cov}(X, Y)$;
- as a result also $\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$ and $\sigma_{aX} = a\sigma_X$;
- bilinearity of the covariance: $\mathrm{Cov}(aX + Y, Z) = a\,\mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$,
- which yields $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}(Y)$,
- and if X and Y are independent also $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

1 EMH: Efficient Market Hypothesis (Week 1)

Efficient markets restrict possibility to strategically make profit that is larger than the market's average.

The general line of thoughts is this one: If whenever you spot a possibility "to beat the market", e.g. by being able to predict the price, you believe that...

1. ...there would have been many others to know the price in advance,
2. ...these others would have bought the stock,
3. ...the resulting bids would yield to a rise of today's stock price,
4. ...this would happen very fast and until the advantage vanishes,

then you believe in the **efficiency of the market**.

In more detail, we discuss three main formulations of market efficiency:

- *Weak form* of the EMH:
  The current stock price reflects all the historical stock prices;
- *Semi-strong form* of the EMH:
  The current stock price reflects all public information;
- *Strong form* of the EMH:
  The current stock price reflects all public and private information.

The weaker formulations are contained in the stronger formulations, i.e.:

$$\text{strong form holds} \;\Rightarrow\; \text{semi-strong form holds} \;\Rightarrow\; \text{weak form holds.}$$

1.1 The Weak Form of EMH

Under the weak form of EMH investments based on past stock prices do not yield superior returns.

"Technical Analysis", i.e. predicting price movements based on past prices is not possible under this hypothesis. A simple example of strategy that would classify as technical analysis follows the idea "the trend is your friend": Here you invest in stocks when there is an upwards trend.

Advice: be careful with such ideas, scientific studies show that this strategy is less profitable than a "buy and hold" strategy, when trading costs are taken into account. See Part 2 of **?** for more examples.

Roughly, the weak form of EMH means that an investor cannot "beat the market" based on the knowledge of past stock prices.

"Beating the market" means to consistently outperform the market. A clear way to "beat the market" is by arbitrage, i.e. by making a profit without risking a loss.

There are investment strategies trying to "beat the market", which are consistent with the weak EMH. For instance:

A) Fundamental analysis: Model the intrinsic value of a company and invest in underrated stocks; then wait for the price to approach the intrinsic value.

B) Quickly react to news with your investment strategy, e.g.:
   - announcements about the company / the market, e.g. higher profit then expected, new CEO, ...
   - rumours, e.g. expected merger, expected contracts
   - political events, e.g. tax and tariffs; strike action, changed regulations.

The general line of thoughts behind the weak EMH is this one: If it is possible to "beat the market" based on knowledge of past prices, then algorithms are produced to do so. Large companies will use these algorithms and trade accordingly. This will rapidly lead to a rise of demands of specific products. This rise in demand in turn will be visible to those ones who sell these products and therefore lead to a rise of the price of these products. This process will continue until the price is finally so high that the strategy is no longer superior. Now, this process is assumed to be very fast and one may assume that it has already been in place once we see the prices.

1.2 The Semi-strong Form of EMH

Under the semi-strong form of the EMH investments based on any publicly available information do not yield superior returns.

The hypothesis assumes that the price adjusts immediately to new information, e.g. the announcement of quarterly earnings, dividends, new stocks.

Public information means is anything that is publicly available and relatively easy to acquire (e.g. press releases, newspapers, financial magazines). Non-public Information: is information that is not publicly available, for instance insider information. Notice that insider trading is usually illegal. However, many cases of insider trading are indeed documented.

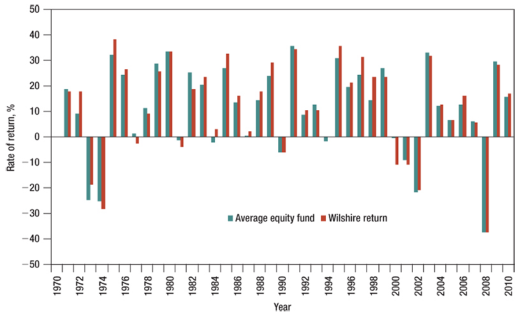
1.3 The Strong Form of EMH

Nobody can consequently outperform the market with their investment.

The line of thoughts is very similar to the other two cases.

1.4 Criticism and Use of the EMH

In general, is difficult to test the hypotheses, as the primary information is not available. In regards to the semi-strong form of efficient market hypothesis, one can study the influence of information releases on prices of financial instruments. There is, stronger criticism against the validity of the strong hypothesis: if we believe that insider trading is not profitable, that has strong consequences. Many cases of insider trading are documented, so one cannot argue that they do not exist. In order to have sufficient insider trading so that prices reflect all insider information, a significant number of insiders need to trade, such that the price can reflect their private information.



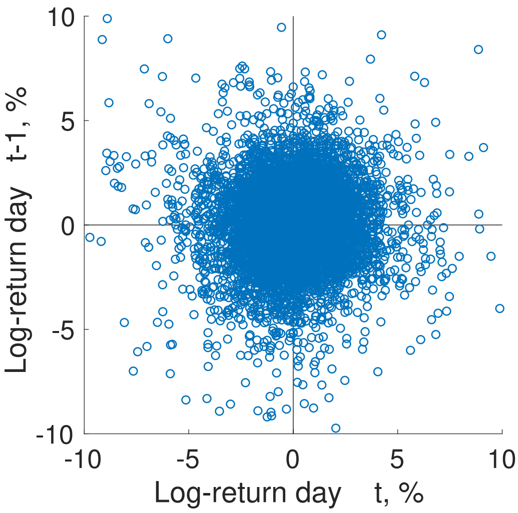


Figure 1: Scatter plot of subsequent plots for GE's returns within over 55 years. The empirical correlation $1,46\%$ is not statistically significant.

There are also arguments based on data that support the EMH: the scatter plot 1 shows that returns of subsequent days are uncorrelated for a specific time series of prices. This means that the autocorrelation of the stock returns $\text{corr}(R_{t+1}, R_t) \approx 0$. This empirical observation has been repeatedly made for other asset price time series as well thus underpinning that the future price cannot be predicted based on the past and today's price. At least this is evidence against a very basic form of predicting the price based on the price history thus supporting the weak EMH. Some studies have investigated the possibility of outperforming the market by comparing the long-term performance of mutual funds with the one of the market, the latter here is represented by the

Wilshire 5000 Total Market Index, see Figure 1.4. In some years mutual funds outperformed the market, but no fund does so consistently, underpinning the strong EMH.

The different forms of EMH follow intuitive rationales. This is highly beneficial for getting a first understanding of trading strategies and modelling purposes. We discussed rationales of trading strategies: Those based on the believe that one can consistently outperform the market using historical prices, those based on the believe that one can consistently outperform the market using publicly/ private information. The position of facing a strongly efficient market is that one of a market participant who does not believe to be able to use information to ”beat the market”.

The benefit of the EMH for modelling purposes is the following: the financial market is utterly complex and simplifications need to be made before one is able to formulate a mathematical model. The different forms of the EMH give a reasonable rational to formulate such simplifications. In this sense, we are with this chapter at a stage where we set the ground for mathematical tools, for being able to formulate and justify mathematical models for the behaviour of financial quantities.

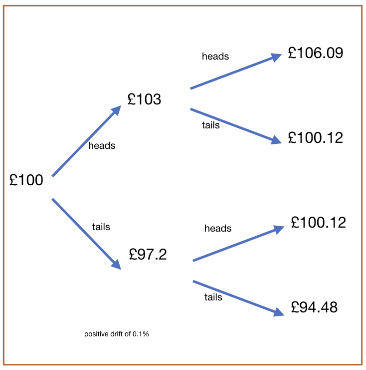


Figure 2: Example of a binomial model over three time periods.

The observation of uncorrelated subsequent returns andthe reasoning underpinning the EMH supports the random walk theory of stock prices. Here, we model stock prices randomly, und in a way that the daily increments are independent of the history of prices. An simple example of a model respecting these features is the binomial model, compare Figure **??**.

1.5 Summary

The implication on potential investments is of large interest for us:

1. Assuming *none of the hypothesis holds*, you can find investments, which are based on
   - patterns found in historical stock prices, or
   - any information concerning the company/the market

   and can consistently expect profits that are larger than the market average.

2. Assuming only the *weak form* is valid, you cannot find investments, which consistently yield superior profit and are based on
   - patterns found in historical stock prices,

   however, it can be based on
   - any further information concerning the company/the market.

3. Assuming the *semi-strong form* (hence also the weak form) is valid, you cannot find investments, which consistently yield superior profit and are based on
   - any public information,

investment based on	investor may believe in	investor *does not* believe in
historical stock prices	no EMH	weak form
public information	weak form	semi-strong form
private information	semi-strong form	strong form
investor needs to increase risk, to increase expected payoff	strong form	-

Table 1: Overview of the EMH

however, it can be based on

- any private information concerning the company/the market.

4. Assuming the *strong form* (hence also the semi-strong and weak forms) is valid, you cannot find

   - any investment that consistently yield superior profit.

   The only way to increase the expected return is to

   - increasing the risk.

An overview on what can/cannot be used to design superior investment strategies is Table 1.

Empirical evidence of the weak formulation of efficient markets can be found, but testing the hypotheses is difficult. The validity of the hypotheses is therefore also criticised. The EMH is useful to get an orientation towards the investment strategies and to simplify the complexity of real markets to set the ground for mathematical modelling.

2 Stochastic Models of Long-Term Behaviour of Security Prices (Week 2)

A consequence of the efficient market hypothesis is the *random walk theorem*, stating that the returns on subsequent days are independent of each other. An important model is the *lognormal model.*

The Log-normal Model With $(S_t)_{t\in\mathbb{N}}$ the daily stock price, we consider $(X_t)_{t\in\mathbb{N}}$ the daily log-returns $X_t = \log(S_{t+1}/S_t)$. Log-returns for several days are obtained by summing up the daily log-returns:

$$\log(S_t) - \log(S_s) = \sum_{i=s}^{t-1} X_i,$$

i.e. $S_t = S_s \exp(\sum_{i=s}^{t-1} X_i)$ for $s < t$.

The key-assumption for the lognormal model is that

- the daily log-returns X_t are iid (i.e. independent and identically distributed), and that
- this distribution is a normal distribution $\mathcal{N}(\mu, \sigma^2)$.

Parameter Estimation in the Log-normal Model Given N iid random variables X_i with an assumed distribution, e.g. $\mathcal{N}(\mu, \sigma^2)$, we need to estimate the model parameters, here μ and σ. Here, X_i for each i represents the daily log-return of a stock, and the model parameters are the mean μ and the volatility σ of the log-returns, and σ^2 is its variance. Parameter estimation of a time series of data is a large and deep area of statistics. The estimation will only approximately represent

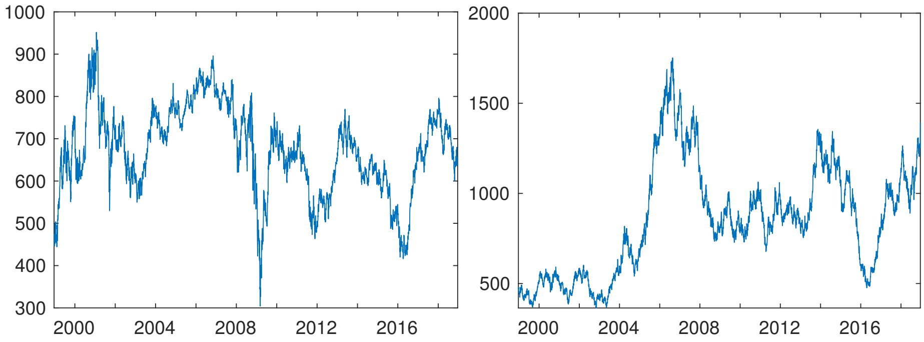


Figure 3: Stock prices. Left: Empirical data; Right: Lognormal model

the real time series, and in view of the limited number of observations, the error needs to be well understood. On the one hand one may choose or build an estimator that fulfils many desirable statistical features, which allow to better judge the quality of the estimation, on the other hand one would like to keep the estimation process as simple as possible. Since mathematical simplification in terms of model assumptions meet reality here, there is a large number of sources of additional errors. Here, we make a very simple and convenient choice, we consider the empirical mean and variance to estimate μ and σ^2 via

$$\mu \approx \overline{X} = (X_1 + \ldots + X_N/N), \quad \text{(empirical mean)}$$

$$\sigma^2 \approx \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2 \quad \text{(empirical variance)}.$$

In Excel, for instance, these formulas can be conveniently implemented through the commands `AVERAGE` (computing the mean of a cell range) and `STDEV.S` (computing the standard deviation of a cell range).

This simple choice of estimators comes with some crucial statistical properties. In the Homework, Coursework 1 you will show the following: If $\mathbb{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/N$, then $\overline{X})$ converges to the mean μ with probability 1. This is called consistency. Essentially, this means that if we pick more and more observations, the empirical mean converges to the true mean. Another basic property of estimators is unbiasedness.

A parameter estimate $\hat{\theta}_N$ of the true parameter θ is called *unbiased*, iff $\mathbb{E}(\hat{\theta}_N) = \theta$. In other words, if the estimated parameter is in mean equal to the true parameter. The Mean Square Error (MSE) can be represented in terms of the bias and the variance of the estimator,

$$\mathrm{MSE} = \mathbb{E}\left((\hat{\theta}_{\mathrm{N}} - \theta)^2\right) = \mathrm{bias}(\hat{\theta}_{\mathrm{N}})^2 + \mathrm{var}(\hat{\theta}_{\mathrm{N}}),$$

if the estimator is unbiased, the bias vanishes. One can show that both the empirical mean and the empirical variance are unbiased estimators. In fact, the first guess for a good estimator of the variance might be $\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$. This estimator is consistent, however is biased. In contrast, the estimator $\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$ is unbiased and consistent, and therefore is the standard estimator of the variance, which is called the empirical variance.

Comparison of the Log-normal Model to Market Data An example of the stock price is given in Figure 3.

When comparing the model with empirical data, we see the limitations, in particular:

- Volatility clustering is observed (large squared daily returns are likely to follow each other), but not present in the lognormal model, see Figure 4
- Large losses are underestimated with the lognormal model, see Figures 4 and 5. While the higher lieklyhood of large gains and losses is visible in the tails ofthe histogram, it is present in form of spikes in the time series of log-returns.

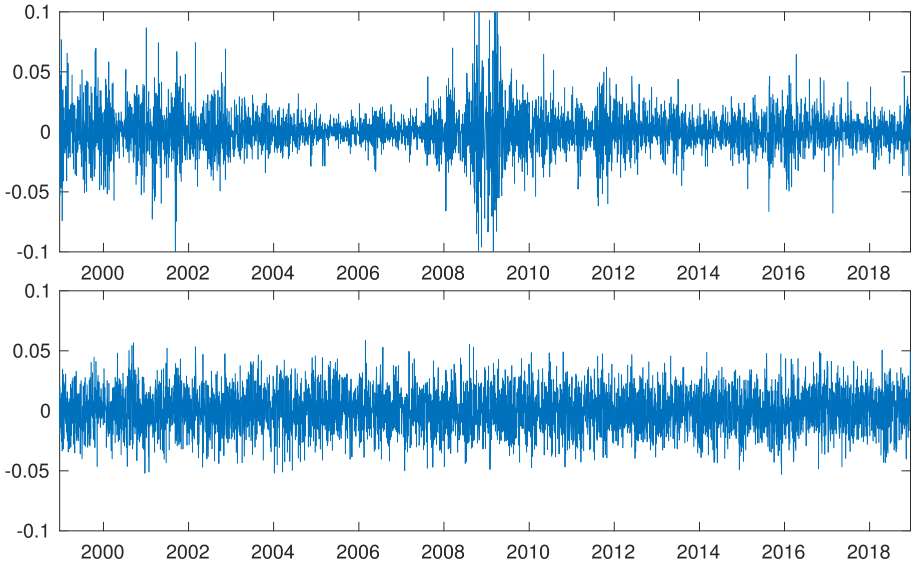


Figure 4: Log-returns Top: Empirical data; Bottom: Lognormal model

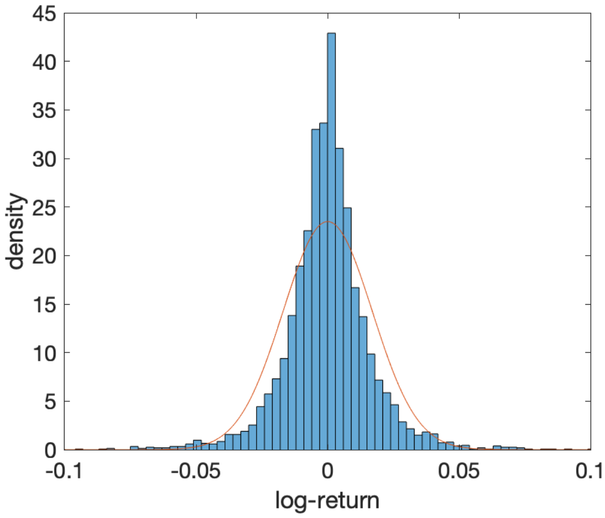


Figure 5: Histogram for log-returns and normal pdf

The consequences of underestimating large losses can be severe! „[...] Large fluctuations in the stock market are far more common than Brownian motion predicts. The reason is unrealistic assumptions – ignoring potential black swans. [...]“ see `https://www.theguardian.com/science/2012/feb/12/black-scholes-equation-credit-crunch`

If the Black-Scholes model is systemically used to estimate the risk and underestimates large losses, this leads to a systemic underestimation of large losses. This can have severe consequences as financial institutions thus may face a lack of risk capital in times of crises. This in turn can lead to a further destabilisation of the system and can advance a crisis. This does not mean that the Black-Scholes model is not a good model. It is a very good model in the sense that it displays some features of the stock market in a very simple way. However, it cannot serve all purposes. It has its clear shortcomings and when it is used systemically in the wrong way this can lead to damages on a global scale. It is therefore highly important that you understand the benefits and shortcomings of the Black-Scholes model. Moreover, it is important to realise that whatever model you use, it has its specific scope and you need to understand its benefits and shortcomings very well. This is of a urgent economic meaning, globally.

(In)dependence and (no) autocorrelation and volatility clusters In Figure 4 we see clusters of high changes in the subsequent returns. This is known as **volatility clusters**. There presence indicate a dependence of subsequence returns, contradicting one of the basic assumptions of the log-normal model.

Next, let us graphically study the autocorrelation of the returns, i.e. the correlation between subsequent returns. To do so, we build pairs (R_t, R_{t-1}) of all subsequent returns observed. We plot the value of R_t on the x-axis and the value of R_{t-1} on the y-axis, thus obtaining the scatter plot Figure 6. The points are centred around zero, radially symmetric. This indicates that there is no linear dependence between R_t and R_{t-1}. Computing the empirical autocorrelation yields -0.016 confirming that this is very low, thus no indication of a linear dependence. Notice that this is a rudimentary approach, only to get a rough idea. To make this mathematically conclusive, one would need to employ statistical techniques, which goes beyond the scope of this lecture. This observation has been made consistently,

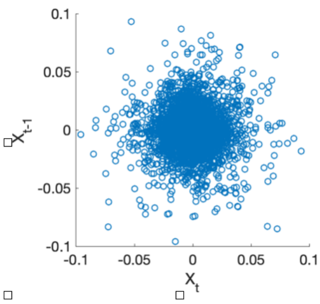


Figure 6: Scatter plot of subsequent returns of the HSBC stock prices.

To summarize our findings, returns of stock prices (and also log-returns as they are very similar) exhibit

- No autocorrelation: $\text{corr}(R_t, R_{t+1}) \approx 0$ (this is in line with the weak form of EMH)
- Volatility clustering: $\text{corr}(R_t^2, R_{t+1}^2) > 0$. We can observe periods of large volatility and of small volatility;
- Heavy tails / spikes: High losses and gains much more likely than for normally distributed random variables.

The presence of volatility clusters indicates a dependence of subsequence returns. However, we also observed no autocorrelation. Correlation and dependence of random variables is closely linked. If two random variables are independent they are uncorrelated. The contrary, however is not always true. For instance consider X standard normally distributed. Clearly X and X^2 are dependent. What is their correlation? For returns, we look for such random variables, which have no autocorrelation, but which are dependent.

Stylized Facts To model the stock price evolution in an appropriate way means to balance model complexity against realistic features. Researchers have established a list of stylized features, that is features that stock prices typically exhibit. This step is helpful in modelling as it establishes the features that a model should reproduce. In practice, the actual goal of the model determines which features are most important and which ones may be ignored. Building a good model is a highly nontrivial task, each model will clearly not be perfect: Each model is flawed. But which model is good enough for the actual tasks at hand? This type of work, modelling, is done in financial institutions when internal models are build and validated, it is also a vivid research area.

Here, we have listed three of the most important stylized features of the daily returns that we have observed. Deeper discussion and more stylised facts: R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, Volume 1, 2001 `https://www.lpsm.paris/pageperso/ramacont/papers/empirical.pdf`

Autoregressive Model A better fit of the data is available with more complex models, e.g. the autoregressive AR(1) process. There the volatility (i.e. standard deviation of the log-returns) is a stationary autoregressive stochastic process:

$$
\begin{aligned}
X_t &= \mu t + \sigma_t Z_t, \quad Z_t \sim \mathcal{N}(0,1) \text{ iid },\\
\sigma_t &= \alpha + \beta \sigma_{t-1} + v\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1) \text{ iid }, |\beta| < 1,
\end{aligned}
$$

with Z_t and ϵ_t being independent of each other and of σ_{t-1}, X_{t-1}. The autoregressive model introduces a positive correlation of the volatility and hence the magnitude of returns. This way volatility clusters are introduced. Challenges are the fitting the parameters and a more complex evaluation compared to the lognormal model.

Comparison AR(1) Model to Market Data In order to obtain a first impression on the behaviour of the $AR(1)$ model compared to market data, we simulate log-returns in the model, for an arbitrary choice of the parameters. Note that we did not fit the parameters, so the comparison is in a preliminary stage and we can only have a glimpse on the behaviour in respect to stylized facts. We display the time series of the related stock prices, the time series of the log-returns, in comparison to one empirically observed time series of market data in Figure 7. From the time

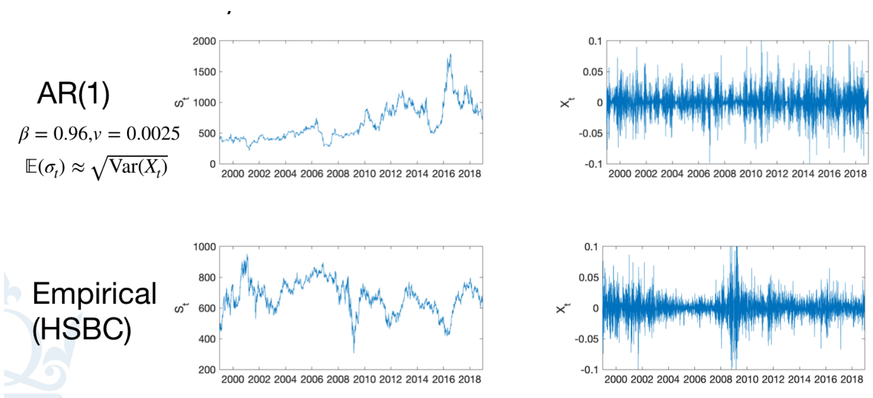


Figure 7: Log-returns Top: AR(1) model; Bottom: Empirical data

series of stock prices itself it is hard to extract stylized facts, similarities or differences. Turning to the time series of log-returns, however, we observe that the AR(1) model reproduces clusters, i.e. periods of a large number of high returns in absolute values and periods of lower numbers of high returns. We also observe some positive and negative spikes. Both features are more extreme in the empirical time series, but note that we did not fit the parameters and we have only one example here, so we should not draw any conclusion from this single observation.

Next, we display the histogram of log-returns from the empirically observed prices, in the AR(1) model, in a log-normal model in Figure 8. We observe that the shape of the empirical distribution of the log-returns is better reproduced, it is steeper around the mean and higher returns are more likely than in the log-normal model. Both the steeper form in the middle and the slower decay of the tails are visually more similar to the market data then the histogram of the log-normal returns.

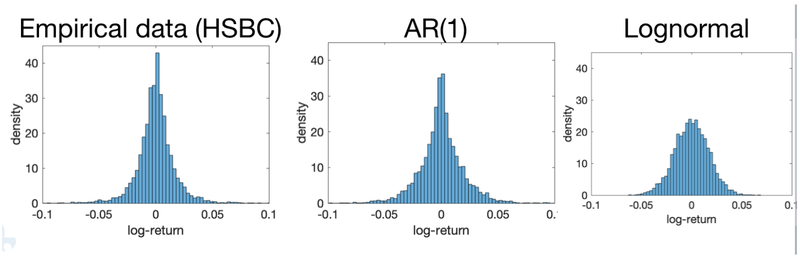


Figure 8: Histogram of Log-returns

Estimation of Parameters in the Autoregressive Model The general approach to derive the model parameters α, β, v in $\sigma_t = \alpha + \beta\sigma_{t-1} + v\epsilon_t$ is the following two-stage procedure. First, estimate the expectation value $\mathbb{E}(\sigma_t)$, the variance $\mathrm{Var}(\sigma_t)$ and autocorrelation $\mathrm{corr}(\sigma_t, \sigma_{t-1})$. Second, derive the parameters

$$\begin{aligned}\beta &= \mathrm{corr}(\sigma_t, \sigma_{t-1}),\\ \alpha &= (1-\beta)\,\mathbb{E}(\sigma_t),\\ v^2 &= (1-\beta^2)\ \mathrm{Var}(\sigma_t).\end{aligned}$$

The challenging step is to estimate the empirical volatility. Remember how we estimate the empirical variance in the log-normal model, $\sigma^2 \approx \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$. The subtle point is that this is a good estimator if the sequence X_i is iid. However, the AR(1) model is build in such a way to create dependence of the log-returns. The main difficulty thus is that

- X_i are not independent in the AR(1) model, and
- σ_t is different for each X_t, but it is impossible to estimate the variance with a single data point only.

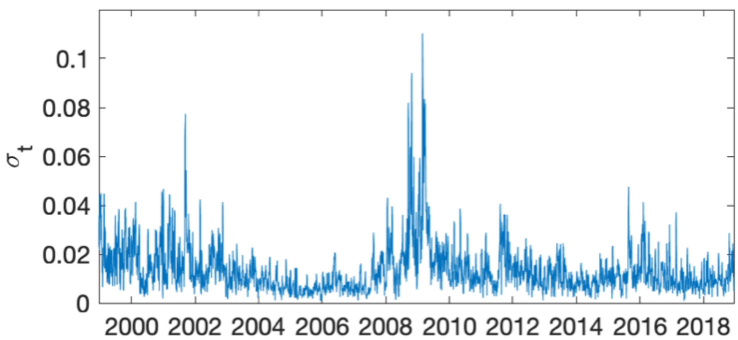


Figure 9: Time series of locally estimated volatilities.

As a compromise we use here a naive parameter fitting approach, a local estimate. We estimate the local variance using 5 neighbouring values of the log-return:

$$\sigma_t^2 \approx 1/4 \sum_{i=t-2}^{t+2} \left(X_i - \bar{X}\right)^2 .$$

The resulting time series of the volatility is shown in Figure 10. Then we use this time-series to estimate $\mathbb{E}(\sigma_t)$, $\mathrm{Var}(\sigma_t)$ and $\mathrm{corr}(\sigma_t, \sigma_{t-1})$. We apply this approach to the time series of HSBC stock prices, and obtain $\mathbb{E}(\sigma_t) \approx 0.0138$, $\mathrm{Var}(\sigma_t) \approx 1.05 \cdot 10^{-4}$, $\mathrm{corr}(\sigma_t, \sigma_{t-1}) \approx 0.9013$. A graphical comparison of the empirical log-returns and log-returns simulated from the AR(1) model is shown in Figure

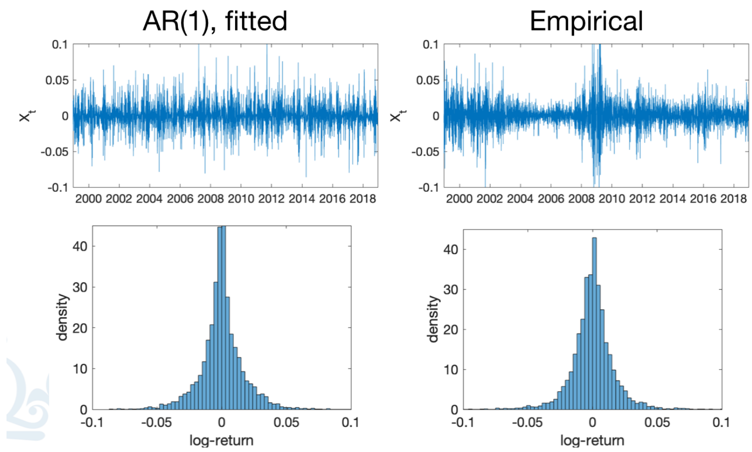


Figure 10: Time series and histogram of log-returns. Left: simulated from the AR(1) model; right: empirical.

Further Alternative Models There is a large and ever growing family of stock price models, each model comes with advantages and disadvantages. Here we list a few approaches and concrete models.

The AR(1) model is discrete in time. Many models are continuous in time, which makes the analysis often much more elegant and therefore easier for complex tasks. Time-continuous extensions of the AR(1) model for the volatility have been developed, the most famous ones are

- Ornstein-Uhlenbeck (OU) process,
- Cox-Ingersoll-Ross (CIR) model.

The latter is used as volatility process in the famous Heston model for option pricing.

3 Risk and Return (Weeks 3 and 4)

Assessment of risk is one of the most important parts of mathematical finance. Quantified risk can be used to evaluate investments, as well as optimize a portfolio of assets. We first discuss the dominance of assets based on mean and variance and then discuss various measures of risk.

First, we consider a basic risk quantity. Comparing a savings account with fixed interest rates with a stock, one basic difference is that we know in advance how the investment of the savings account will change over time, while we do not know how the stock price will change over time, the broader the expected deviation, the more risky the asset feels. This brings us to the first basic notion of risk in finance, the volatility, or its square, the variance. More precisely, we consider the *variance of returns* $\mathrm{Var}(R) = \mathbb{E}\left((R - \mathbb{E}(R))^2\right)$ as a measure of risk. It

- measures uncertainty in terms of scatter around the expectation,

- measures distance between realised and expected return $R - \mathbb{E}(R)$,
- by the square the sign vanishes and larger deviations are weighted higher than smaller ones,
- by taking the outer expectation, the deviations are weighted according to their likelihoods.
- The variance is 0 if there is no risk!

Consider now two opportunities to invest. The first investment is one with a return of high variance, say 20% and low expectation, for instance 0.01. The second possible investment is one without risk, i.e. the variance is zero, and the double return, 0.02. Which investment is more risky? To make it even more extreme, let the expected return of the risky asset be negative. These examples show that for investment decisions, the variance alone cannot measure the risk. A sensible approach is to consider both mean and variance of the returns.

For the investment evaluation based on mean and variance, we consider each investment as a pair (μ, σ) of the returns mean $\mu = \mathbb{E}(R)$ and standard deviation $\sigma = \sqrt{\mathrm{Var}(R)}$.

Definition 1. *An investments* (μ_1, σ_1) dominates *another investment* (μ_2, σ_2)*, iff*

$$\mu_1 \geq \mu_2, \quad \sigma_1 \leq \sigma_2,$$

and one of the inequalities is strict (i.e. not equal). We write $(\mu_1, \sigma_1) \succ (\mu_2, \sigma_2)$*.*

An investment is dominated, when another investment has a higher expected payoff with less risk. Note that not all pairs can be ordered. Investments that are not dominated form the efficient subset

Definition 2. *Given a set of investments* $A = \{(\mu_i, \sigma_i), i \in \mathcal{I}\}$*. An investment* $(\widehat{\mu}, \widehat{\sigma}) \in A$ *is an element of the* efficient subset A_{eff}*, iff it is not dominated, i.e. there is no* $i \in \mathcal{I}$*, such that* $(\mu_i, \sigma_i) \succ (\widehat{\mu}, \widehat{\sigma})$*.*

We can use the efficient subset to determine reasonable investments. If we have a given set of investments and we want to invest according to the mean-variance analysis, only elements of the efficient subset are reasonable.

We can evaluate the efficient subset by testing pairwise dominance and neglecting all elements that are dominated. Graphically dominance means that the dominating asset lies towards the top left in the σ-μ plane, see Figure 11.

To summarize, the pair of expectation and variance of returns,

- represents both the level of return that we can expect and the risk we take.
- together can be investigated by comparing different investments on the (σ, μ)-plane.
- cancelling out the pairs for which we find a better alternative leaves us with the efficient subset.

This is the basis of investment theory!

We also observe some drawbacks of the variance as risk measure, namely,

- unexpected large profit contributes same as a loss Remember.
- we cannot distinguish between frequent small losses and rare huge losses.
- Variance follows historical prices, and does not allow us a tool to include the impact of events (such as the outbreak of a global pandemic or the Brexit or the storming of the US Capitol) which are not reflected in historical price series.

These are severe shortcomings, and therefore further risk measures have been developed. Building good risk measures is, as building models, a highly complex task. Each attempt to pin down the risk in a single number will ultimately fail to assess the true risk completely. The nature of financial risks is too complex. However, quantifying essential aspects of the risk in a single number is utterly important in order to deal with the risk in a responsible manner. When dealing with

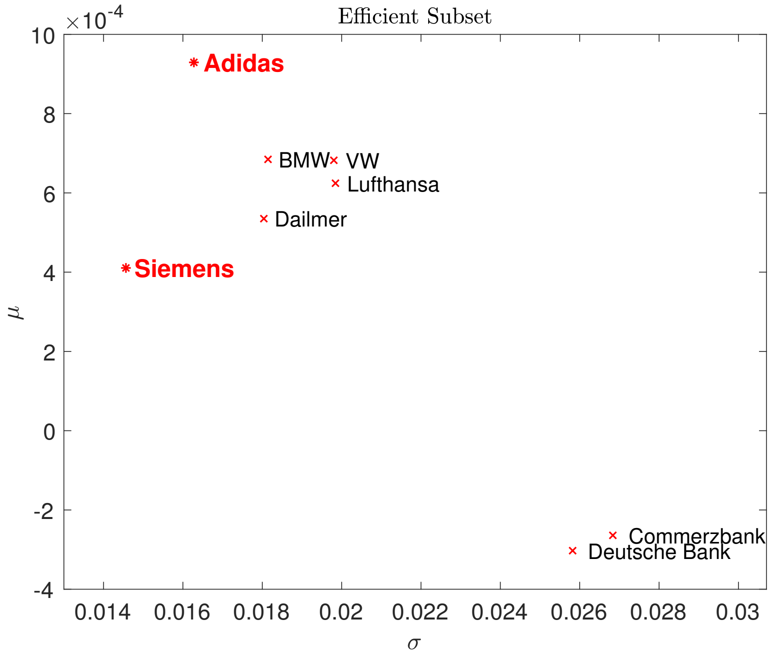


Figure 11: Several stocks in the $\sigma - \mu$-plane and their efficient subset.

risk measures it is crucial to understand the scope of the measure, what does it reflect and what does it not reflect? Remember, there is always something that the risk measure does not capture, so always be sure to understand very well what it does and what it does not reflect.

Ultimately, the development and the understanding of measuring financial risks is highly relevant for financial institutions and sound risk measures are required to control the risk of investments. On a systemic level, controlling the risk of the investments of all institutions is required to guarantee the stability of the financial system and of the economy as a whole.

Next, we consider the semi-variance as a slight adaptation of the variance as a risk measure and then we turn to the most commonly used risk measures in practice, the shortfall probability and the value-at-risk. Finally, we briefly discuss the concept of stress testing, which is one of the pillars of financial risk assessment.

Semi-variance The first drawback of the variance as risk measure listed above is that losses and gains equally contribute to the variance, while investors will welcome gains and suffer from losses. In order to adapt the *semi-variance* is defined by $\mathbb{E}(\min\{0, X - \mu\}^2)$. It measures the downside-risk. As a major drawback, we observe that it is still highly dependent on the mean μ, also the other two criticism listed above are still valid.

3.1 Shortfall probability

The variance is a very simple measure of investment risk. While it enables us to easily compare stocks, for a more detailed investigation more advanced risk measures need to be considered. Shortcomings include:

- due to it's dependency on the expected value, assets with a larger expected value may seem riskier although they are not;
- unexpected large gains are valued the same as unexpected large losses;
- the variance does not give any information about the size of the risk or their probability. A likely small loss can have the same variance as a less likely huge loss.

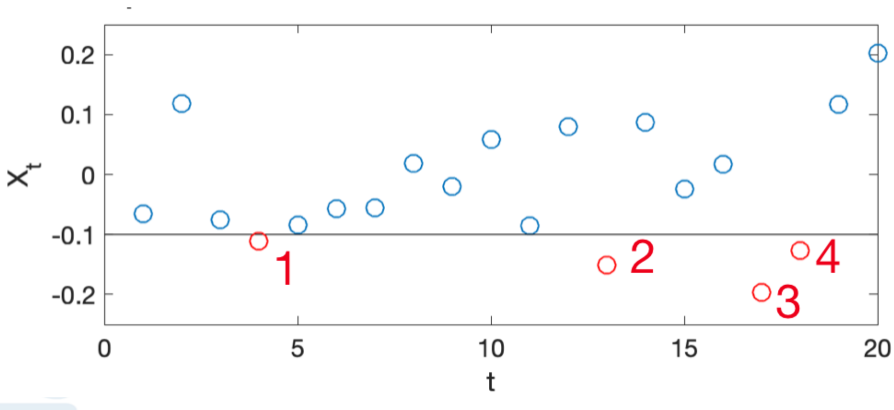


Figure 12: Illustration of the empirical shortfall probability.

To solve the problems, the *shortfall probability* and the *Value at Risk* can be considered. They answer the questions

- how likely are large losses (shortfall probability);
- how large are likely losses (Value at Risk).

Both are based on the realised loss $L = -R$ (note, that we can use either the return R or the log-return X in the definition of the loss, depending on the situation; the results will differ only slightly). The shortfall probability can best be evaluated using the distribution function of the return R: $F_R(x) = P(R \leq x)$:

$$\mathrm{SF}(b, R) = P(L \geq b) = F_R(-b),$$

see Figure 13 for an illustration. Roughly, the shortfall probability measures how likely large losses are. More precisely, it measures how likely losses larger than a pre-specified threshold are.

How to compute the shortfall probability? If we have a model at hand, we can do that with the help of the density, or the distribution function directly. If we have an observation of a time series of daily returns X_t for days $t = 1, \ldots, N$ instead, we need to evaluate the *empirical shortfall probability* instead. This is given by

$$\mathrm{SF}_{\mathrm{e}}(b) = \frac{|\{t : 1 \leq t \leq N, s.t. -X_t > b\}|}{N}.$$

Figure 12 illustrates how to obtain the empirical shortfall probability for 20 samples and the threshold $b = 0.1$. We count 4 samples below the threshold, which is 20%, therefore $\mathrm{SF}_{\mathrm{e}}(0.1) =$ 20%.

3.2 Value at Risk and α-quantiles

The shortfall probability quantifies how likely losses beyond a given threshold are. Asking differently, we may want to know with which level of loss do we have to probably deal? For instance, we would like to be prepared to compensate all likely losses with cash, while we leave it open how we move on when a larger loss happens, because the scenario is unlikely. The notion of value-at-risk makes this mathematically precise. First, we have to specify what we mean with likely losses. This is done by specifying a confidence level, for instance 95%.

The value-at-risk is the maximum amount to be lost with a specified likelyhood, i.e. at a predefined confidence level. For example, if the 95% VaR is 1 million, there is 95% confidence that the portfolio will not lose more than 1 million.

The Value at Risk is defined as

$$\mathrm{VaR}_\alpha = \inf\{b : P(L > b) < 1 - \alpha\}.$$

If the distribution function of the return F_R is continuous and strictly increasing, we can use the inverse function to evaluate the value at risk:

$$\mathrm{VaR}_\alpha = -F_R^{-1}(1-\alpha).$$

Note: we usually evaluate VaR_α for $\alpha > 0.5$, e.g. 95% or 99%, which yields $1-\alpha < 0.5$. See Figure 14 for the illustration of the evaluation using the density function.

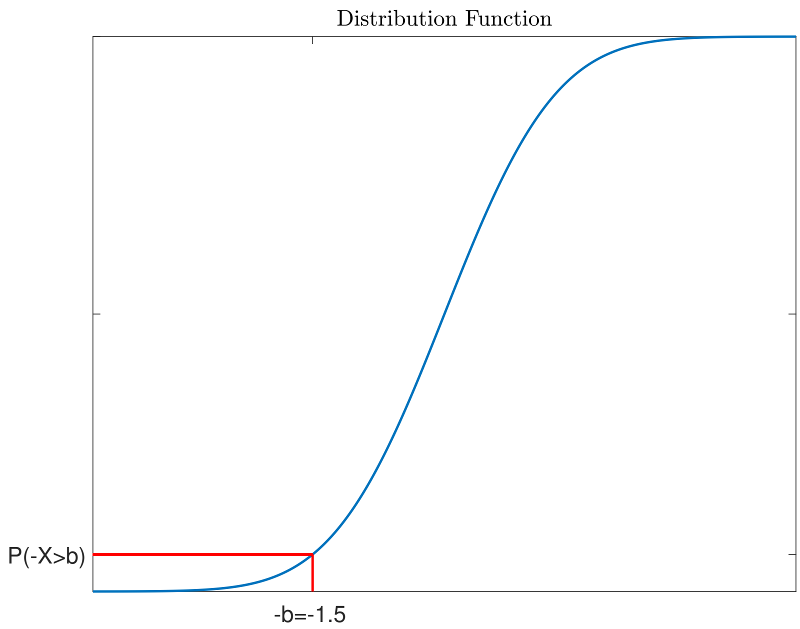


Figure 13: Evaluating the shortfall probability $\mathrm{SF}(X, 1.5)$ using the distribution function (here X log-return)

Reminder: The distribution function $F_X(x) := P(X \leq x)$ is left-continuous

Definition 3. *For $\alpha \in (0,1)$ the number*

$$q^\alpha(X) = \inf\{x : \alpha < F_X(x)\}$$

is called upper α-quantile of X.

$$q_\alpha(X) = \inf\{x : \alpha \leq F_X(x)\}$$

is called the lower α-quantile of X.
Any $q \in [q_\alpha(X), q^\alpha(X)]$ is called α-quantile of X.

- If F_X is continuous and strictly increasing,

$$q_\alpha(X) = q^\alpha(X) = F_X^{-1}(\alpha).$$

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

Note: different notations are used in practice, e.g. $\mathrm{VaR}_{95\%}$ is sometimes denoted $\mathrm{VaR}_{5\%}$.
Examples:

1. uniform distribution → Tutorials
2. normal distribution → Homework

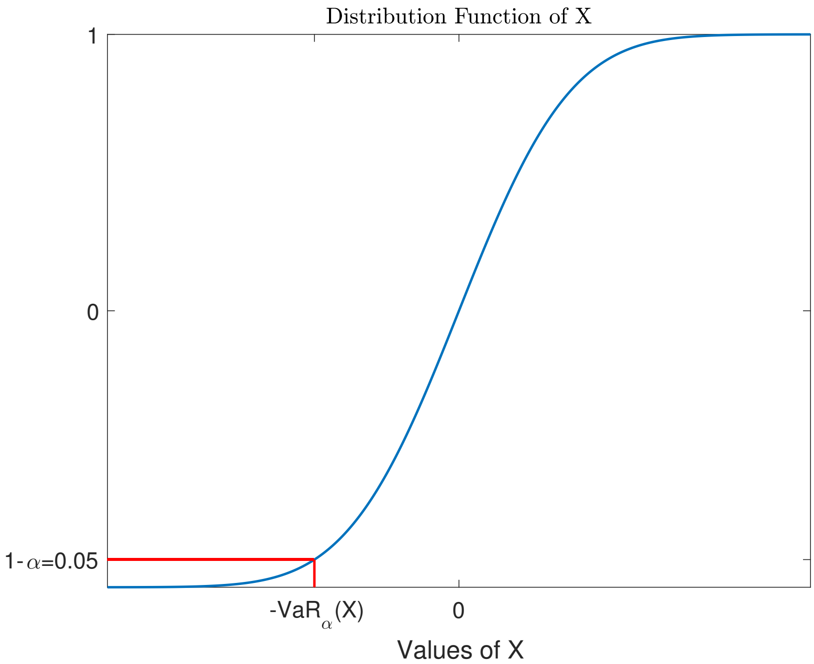


Figure 14: Evaluating the Value at Risk $\text{VaR}_{95\%}$ using the inverse of the distribution function (here X log-return)

3. discrete random variable: **Note: This example is not part of the lecture. It is only included for your own interest.**

Why do we consider discrete random variables? Two examples are

a) Binary options, e.g.

$$\text{Payoff} = \begin{cases} £100, & S_T < £1,200 \\ £0, & S_T \leq 1,200. \end{cases}$$

b) Corporate bond with given probability p for a default. E.g.

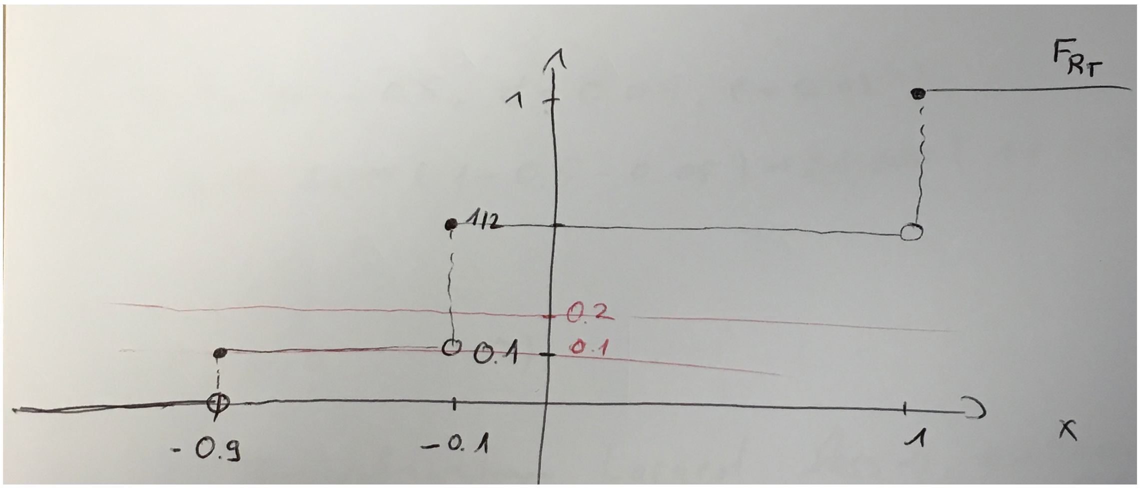
$$\text{Return} = \begin{cases} 1, & \text{with probability } 1-p \\ -1, & \text{with probability } p \quad \text{(default)}. \end{cases}$$

Let's work on an example return of

$$R_T = \begin{cases} -0.9 & p = 0.1, \\ -0.1 & p = 0.4, \\ 1 & p = 0.5. \end{cases}$$

The distribution function is given as

$$F_{R_T}(x) = P(R_T \leq x) = \begin{cases} 0, & x < -0.9, \\ 0.1, & -0.9 \leq x < -0.1, \\ 0.5, & -0.1 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$



With the help of the draft we see

$$\begin{aligned}\mathrm{VaR}_{80\%} &= -q^{1-0.8}(R_T) = -q^{0.2}(R_T) = -\inf\{x : 0.2 < F_{R_T}(x)\}\\ &= -\inf\{x \geq -0.1\} = 0.1,\end{aligned}$$

and also $q_{0.2}(R_T) = -0.1$ for the lower quantile.

For a threshold at 90% we have

$$\mathrm{VaR}_{90\%}(R_T) = -\inf\{x : 0.1 < F_{R_T}(x)\} = 0.1,$$

while the lower quantile yields

$$q_{0.1}(R_T) = \inf\{x : 0.1 \leq F_{R_T}(x)\} = \inf\{x : x \geq -0.9\} = -0.9.$$

This means that any $q \in [-0.9, -0.1]$ is a 10%-quantile of R_T.

Empirical Value-at-risk To compute the empirical value-at-risk, proceed along the following two steps: Let α be the confidence level and R_t for $t = 1, \ldots, N_{\text{samples}}$ the observed daily returns.

1) Sort the values of R_t by magnitude.

2) Consider the smallest $(1-\alpha)N_{\text{samples}}$ elements and choose the value of the largest one.

This process is illustrated in Figure 15. To deepen the understanding of the empirical value-at-risk,

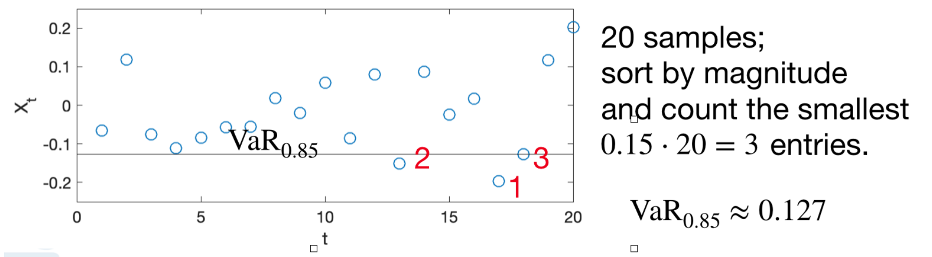


Figure 15: Illustration of the empirical value-at-risk.

remember that sorting the observations by magnitude yields the empirical distribution function. The value-at risk is q quantile of the empirical distribution, compare Figure 16.

Shortcomings of the Value-at-risk Some shortcoming of the Value at Risk is that it does not give us any information about the distribution of the loss in the unlikely case of $(1-\alpha)$. Furthermore it does not enable us to study the influence of possible events without precedent.

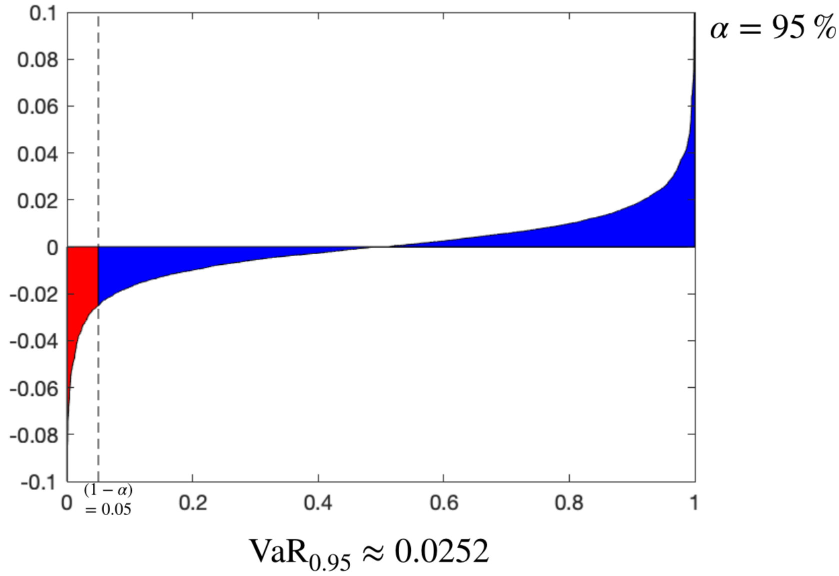


Figure 16: Illustration of the empirical value-at-risk as quantile of the empirical distribution.

3.3 Stress test

The idea behind stress testing is to model important possible scenarios and the compute the related risk. A possible implementation is done along the following steps:

1. Build a factor model for the ingredients of the portfolio.
2. Specify a set of stress scenarios $S \subset \Omega$.

   (for instance high/moderate/low interest rates and high/moderate/low inflation rates)

3. For all $\omega \in S$ compute the future portfolio gain $G(\omega)$.
4. For losses $L = -X$ compute *worst case loss*

$$\varrho(L) = \sup\{L(\omega) | \omega \in S\}$$

   when we restrict our attention to those element of the space of possible events that belong to S, our selected scenarios.

Example: Consider one stock S and a risk-free asset with rate r. We assume the stock-price is given by the random variable

$$S_1 = \begin{cases} S_0(1 + \mu + \sigma), & p = 1/2, \\ S_0(1 + \mu + \sigma), & p = 1/2. \end{cases}$$

where the current mean and variance are $\mu = 0.05$ and $\sigma = 0.1$. The risk-free rate is $r = 4\%$ and we have invested £$1,000$ each in the stock and the risk-free security.

The stress test defines certain scenarios and returns the worst-case lost. Then one needs to check, whether the result is acceptable (passing the stress test) or not (failing it). In our case these scenarios could be

$$\Omega = \{\text{“}\mu = -0.5, \sigma = 0.05, r = 0.03\text{”}, \\ \text{“}\mu = 0, \sigma = 0.2, r = 0.01\text{”}, \ldots\}$$

In each case the maximal loss is computed, e.g.

$$\begin{aligned} R(\text{“}\mu = -0.5, \sigma = 0.05, r = 0.03\text{”}) &= (1000(1 - 0.5 - 0.05) + 1000(1 + 0.03))/2000 - 1 \\ &= 0.26 \\ R(\text{“}\mu = 0, \sigma = 0.2, r = 0.01\text{”}) &= (1000(1 + 0.0 - 0.2) + 1000(1 + 0.01))/2000 - 1 \\ &= 0.095. \end{aligned}$$

In this case the largest loss would be L=-9.5%.

4 Mean-Variance Portfolio Theory (Weeks 5 and 6)

- What is a portfolio?
  - A collection of investments hold (here stocks/risk-free securities)
- Why is portfolio theory so interesting?
  - A lot more can happen than with single assets.

Illustrative example: Ice-cream sellers & umbrella sellers. We model that the following summer will be either rainy or sunny, both equally likely. If it is rainy the umbrella sellers make a larger profit, while ice-cream sellers make a loss. In a sunny summer the situation is the reverse.

	rainy summer ($p = 50\%$)	sunny summer ($p = 50\%$)
Return ice-cream sellers (R^1)	-5%	+10%
Return umbrella corporation (R^2)	+10%	-5%

Both investments have an expectation of 2.5% and a standard deviation of 7.5%. If we buy equal parts of the ice-cream seller and the umbrella corp. we have a return of

$$\frac{1}{2}R^1 + \frac{1}{2}R^2 = \begin{cases} 2.5\%, & p = 50\%, \\ 2.5\%, & p = 50\%. \end{cases},$$

i.e. a safe return of 2.5% (standard deviation zero).

Why does this happen $\rightarrow$ both investments are negatively correlated:

$$\mathrm{corr}(R^1, R^2) = \frac{\mathrm{Cov}(R^1, R^2)}{\sigma_1 \sigma_2} = \frac{\mathbb{E}\left((R^1 - \mu_1)(R^2 - \mu_2)\right)}{\sigma_1 \sigma_2} = -1.$$

Note that a correlation of -1 is a very extreme case unlikely to happen in practice. Let us have a look at an example with no correlation:

$$R^1 = \begin{cases} 10\%, & p = 1/2 \\ -5\%, & p = 1/2 \end{cases}, \qquad R^2 = \begin{cases} 10\%, & p = 1/2 \\ -5\%, & p = 1/2 \end{cases},$$

which are independent of each other. Due to their independence, the joint distribution now has four cases:

probability:	25%	25%	25%	25%
R^1	10%	10%	–5%	–5%
R^2	10%	–5%	10%	–5%
$\frac{1}{2}(R^1 + R^2)$	10%	2.5%	2.5%	–5%

For the portfolio this yields $\mathbb{E}(\frac{1}{2}(R^1 + R^2)) = 2.5\%$ and $\sqrt{\mathrm{Var}(\frac{1}{2}(R^1 + R^2))} \approx 5.3\% < 7.5\%$.

In general we see that a portfolio can have a smaller expected value than the individual assets, while having an expectation value as large as both assets. This is called *diversification*.

4.1 Introduction to portfolios

In the following, we try to optimise our portfolio. We will figure out how to distribute the money and what other choices remain.

Let us consider

- 2 Stocks S^1, S^2 with expected return μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ;
- 2 dates $t \in \{0, 1\}$:
  - $t = 0$ is *today*, i.e. values $S^1(0), S^2(0)$ are deterministic;
  - $t = 1$ is some point in the *future*, i.e. values $S^1(1), S^2(1)$ are random variables.

A portfolio consists of buying/owning x_1 stocks of asset 1 and x_2 of asset 2. The current value is known:

$$P_{(x_1,x_2)}(0) = x_1 S^1(0) + x_2 S^2(0),$$

and the future value is a random variable:

$$P_{(x_1,x_2)}(1) = x_1 S^1(1) + x_2 S^2(1).$$

Note that the amount of shares x_1, x_2 can be quite disproportional to their value, e.g. with $x_1 = 1$ and $S^1(0) = £150$ the value is higher than for $x_1 = 10$ and $S^1(0) = £10$. We therefore introduce weights w_1, w_2, which represent proportion of our wealth $P_{(x_1,x_2)}(0)$ invested in the two assets:

$$w_1 = \frac{x_1 S^1(0)}{P_{(x_1,x_2)}(0)}, \quad w_1 = \frac{x_2 S^2(0)}{P_{(x_1,x_2)}(0)},$$

with $w_1 + w_2 = 1$.

The weights allow us to conveniently express the return $R^P_{(w_1,w_2)}$ of our portfolio in terms of the individual returns $R^1 = S^1(1)/S^1(0) - 1$ and $R^2 = S^2(1)/S^2(0) - 1$:

$$\begin{aligned}
R^{\mathrm{P}}_{(w_1,w_2)} &= \frac{P_{(x_1,x_2)}(1)}{P_{(x_1,x_2)}(0)} - 1 = \frac{x_1 S^1(1) + x_2 S^2(1)}{P_{(x_1,x_2)}(0)} - 1 \\
&= \frac{x_1 S^1(1)}{P_{(x_1,x_2)}(0)} + \frac{x_2 S^2(1)}{P_{(x_1,x_2)}(0)} - 1 \\
&= x_1 \frac{S^1(0)}{P_{(x_1,x_2)}(0)} \frac{S^1(1)}{S^1(0)} + x_2 \frac{S^2(0)}{P_{(x_1,x_2)}(0)} \frac{S^2(1)}{S^2(0)} - 1 \\
&= w_1 \frac{S^1(1)}{S^1(0)} + w_2 \frac{S^2(1)}{S^2(0)} - w_1 - w_2 = w_1 R^1 + w_2 R^2,
\end{aligned}$$

i.e. $R^{\mathrm{P}} = w_1 R^1 + w_2 R^2$.

If we know the proportions we wish to invest in each asset, i.e., we have the weights w_1, w_2 given, we can compute the amount of shares as

$$x_1 = \frac{w_1 P(0)}{S^1(0)}, \quad x_2 = \frac{w_2 P(0)}{S^2(0)}.$$

We typically ignore the fact that we cannot buy fractions of a share.

Example 1. *We wish to invest £1 000 in equal parts in two assets with $S^1(0) = 10$, $S^2(0) = 100$.*

With the initial wealth $P(0) = 1\,000$ and the weights $w_1 = w_2 = 1/2$, we can compute the amount of stocks:

- *Asset 1: $x_1 = \frac{w_1 P(0)}{S^1(0)} = \frac{500}{10} = 50$, i.e. we buy 50 shares of company 1;*
- *Asset 2: $x_2 = \frac{w_2 P(0)}{S^2(0)} = \frac{500}{100} = 5$, i.e. we buy 5 shares of company 2.*

We can compute the return either based on new prices or based on the given return:

1. *Given new prices* $S^1(1) = 12$, $S^2(1) = 100$*:*

$$P(1) = x_1 S^1(1) + x_2 S^2(1) = 600 + 550 = 1\,150,$$

i.e. $R^{\mathrm{P}} = 15\%$*.*

2. *Given returns:* $R^1 = 20\%$ *and* $R^2 = 10\%$*:*

$$R^{\mathrm{P}} = w_1 R^1 + w_2 R^2 = 15\%.$$

4.2 Mean & variance of the portfolio

Given the formula for the return of the portfolio $R^{\mathrm{P}} = w_1 R^1 + w_2 R^2$, we can compute its expected value $\mu_{\mathrm{P}} = \mathbb{E}(R^{\mathrm{P}})$ and the variance $\sigma_{\mathrm{P}}^2 = \mathrm{Var}(R^{\mathrm{P}})$:

$$\mu_{\mathrm{P}} = \mathbb{E}(w_1 R^1 + w_2 R^2) = w_1 \mathbb{E}(R^1) + w_2 \mathbb{E}(R^2) = w_1 \mu_1 + w_2 \mu_2;$$

as well as

$$\begin{aligned}
\sigma_{\mathrm{P}}^2 &= \mathrm{Var}(R^{\mathrm{P}}) = \mathbb{E}\Big(\big(R^{\mathrm{P}} - \mathbb{E}(R^{\mathrm{P}})\big)^2\Big) = \mathbb{E}\Big(\big(w_1 R^1 + w_2 R^2 - \mathbb{E}(w_1 R^1 + w_2 R^2)\big)^2\Big) \\
&= \mathbb{E}\Big(\big(w_1 R^1 - \mathbb{E}(w_1 R^1) + w_2 R^2 - \mathbb{E}(w_2 R^2)\big)^2\Big) \\
&= \mathbb{E}\Big(\big(w_1 R^1 - \mathbb{E}(w_1 R^1)\big)^2\Big) + \mathbb{E}\Big(\big(w_2 R^2 - \mathbb{E}(w_2 R^2)\big)^2\Big) + 2\mathbb{E}\Big(\big(w_1 R^1 - \mathbb{E}(w_1 R^1)\big)\big(w_2 R^2 - \mathbb{E}(w_2 R^2)\big)\Big) \\
&= w_1^2 \,\mathrm{Var}(R^1) + w_2^2 \,\mathrm{Var}(R^2) + 2 w_1 w_2 \,\mathrm{Cov}(R^1, R^2) \\
&= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 \rho.
\end{aligned}$$

Theorem 1. *For a portfolio of two assets* (μ_1, σ_1) *and* (μ_2, σ_2) *with correlation* ρ *and the portions* w_1 *in asset 1, the portfolio's return's expectation* μ_{P} *and variance* σ_{P}^2 *satisfy*

$$\begin{aligned}
\mu_{\mathrm{P}} &= w_1 \mu_1 + (1 - w_1) \mu_2 \\
\sigma_{\mathrm{P}}^2 &= w_1^2 \sigma_1^2 + 2 w_1 (1 - w_1) \sigma_1 \sigma_2 \rho + (1 - w_1)^2 \sigma_2^2.
\end{aligned}$$

4.3 Attainable sets of portfolios

Which points of the (σ, μ)-plane can we attain:

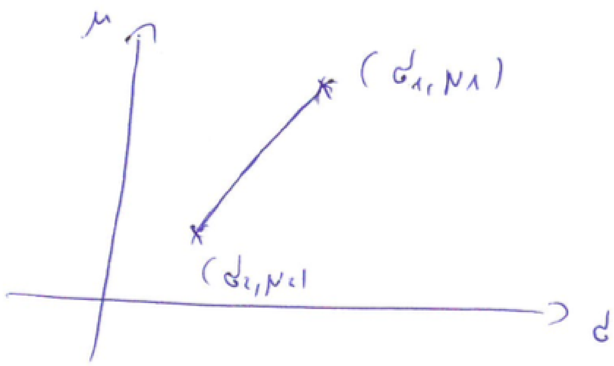
$$\{(\sigma_{\mathrm{P}}(w_1), \mu_{\mathrm{P}}(w_1)) : w_1 \in [0, 1]\}$$

Let's start with some special cases:

- $\rho = 1$: extreme positive correlation (artificial setting)

$$\begin{aligned}
\mu_{\mathrm{P}} &= w_1 \mu_1 + (1 - w_1) \mu_2 \\
\sigma_{\mathrm{P}}^2 &= w_1^2 \sigma_1^2 + 2 w_1 (1 - w_1) \sigma_1 \sigma 2 + (1 - w_1)^2 \sigma_2^2 \\
&= (w_1 \sigma_1 + (1 - w_1) \sigma_2)^2 \\
&\Rightarrow \sigma_{\mathrm{P}} = w_1 \sigma_1 + (1 - w_1) \sigma_2.
\end{aligned}$$

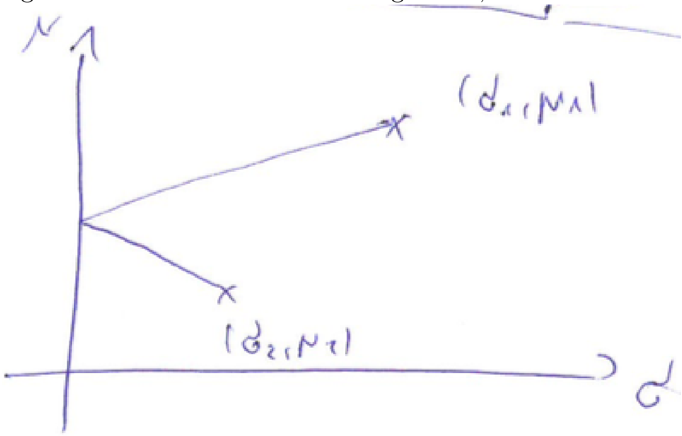
Here the attainable set is a straight line connecting both assets:



- $\rho = -1$: extreme negative correlation (also artificial)

$$\begin{aligned} \mu_{\mathrm{P}} &= w_1\mu_1 + (1-w_1)\mu_2 \\ \sigma_{\mathrm{P}}^2 &= w_1^2\sigma_1^2 - 2w_1(1-w_1)\sigma_1\sigma 2 + (1-w_1)^2\sigma_2^2 \\ &= (w_1\sigma_1 - (1-w_1)\sigma_2)^2 \\ &\Rightarrow \sigma_{\mathrm{P}} = |w_1\sigma_1 - (1-w_1)\sigma_2| \,. \end{aligned}$$

Again the attainable set is a straight line, but with one kink:



See the lecture slides for the non-trivial examples. Now let's summarise the general case:

Theorem 2. *The attainable set for $\mu_1 \neq \mu_2$, $\rho \in (-1, 1)$ is a hyperbola with its centre on the vertical axis.*

Proof e.g. Theorem 2.7 of Capinski, Kopp.

4.4 Minimal Variance Portfolio (MVP)

The attainable portfolio with the minimal risk (i.e. smallest σ_{P}) is called minimal variance portfolio (MPV).

Example 2. *For $\rho = 0$ we have*

$$\begin{aligned} \sigma_{\mathrm{P}}^2 &= w_1^2\sigma_1^2 + (1-w_1)\sigma_2^2, \\ \mu_{\mathrm{P}} &= w_1\mu_1 + (1-w_1)\mu_2. \end{aligned}$$

To find the MVP, we need to find $w_1 \in [0,1]$, such that σ_{P} is minimal (and equivalently σ_{P}^2). I.e. we solve the optimisation problem

$$\min_{w_1 \in [0,1]} w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2$$

Since the function is convex in w_1, we can find the maximum as the root of the first derivative:

$$\frac{\mathrm{d}}{\mathrm{d}w_1}(w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2) = 2(\sigma_1^2 + \sigma_2^2)w_1 - 2\sigma_2^2 = 0,$$

i.e.

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad w_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

This yields

$$\sigma_{\mathrm{MVP}} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad \mu_{\mathrm{MVP}} = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

Theorem 3. *Let $\sigma_1 > \sigma_2$ be the standard deviations for both assets and $\rho \in (-1, 1)$ their correlation. Then the MVP is given for the weights*

$$w_1 = \max\left(\frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, 0\right).$$

and $w_2 = 1 - w_1$.

Proof. See Tutorial 5.

4.5 Short selling

Assuming no restrictions on short-selling, negative weights are positive

$$w_1 \in \mathbb{R}, \quad w_2 = 1 - w_1 \in \mathbb{R}.$$

Negative weights include a leverage (borrow the less profitable asset & sell it to buy the more profitable one.)

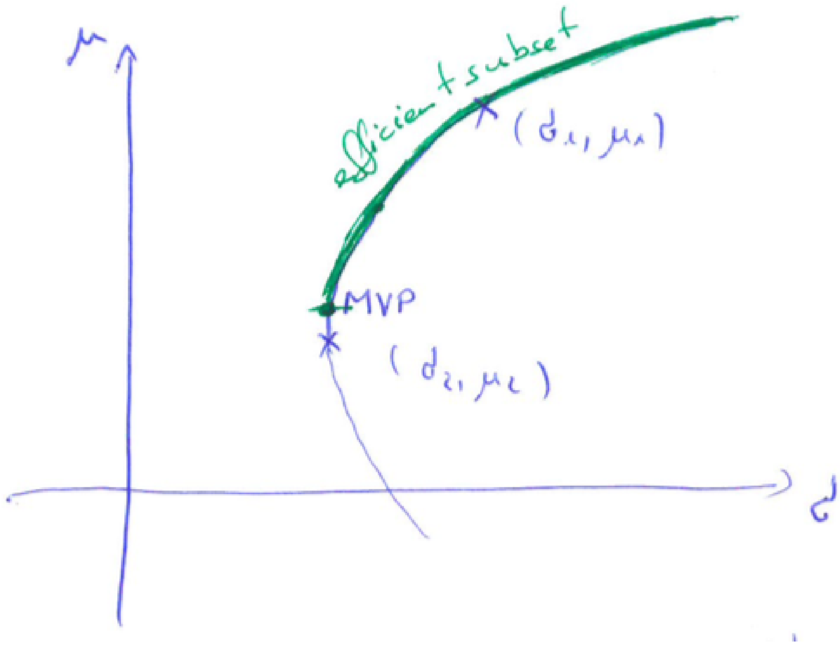
Theorem 4 (MVP, general case)**.** *With no restrictions on short-selling and $\rho \in (-1, 1)$, the MVP is given for the weights*

$$w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

and $w_2 = 1 - w_1$.

4.6 Efficient frontier

Definition 4. *The efficient subset of the attainable set is called* efficient frontier*.*



Observe: The efficient frontier is the part of the attainable set, which connects the MVP with the asset of the highest expectations (and continuing beyond if short-selling is possible).

When no short-selling is possible, it is a closed set (including the end points), otherwise it is half-bounded with the MVP as its end-point.

4.7 Adding a risk-free security

As risk-free asset is like a bank account, where we earn a fixed interest on our money. In the case of short-selling, we assume that we can borrow money for the same rate and without any restrictions.

Combinations of an asset with the risk-free asset form a straight line. We can view any portfolio of assets 1&2 as an individual asset and then apply the known theory.

Example 3. *We compare two cases with a wealth of £1 000:*

a) investing £700 in BP and £300 in Shell

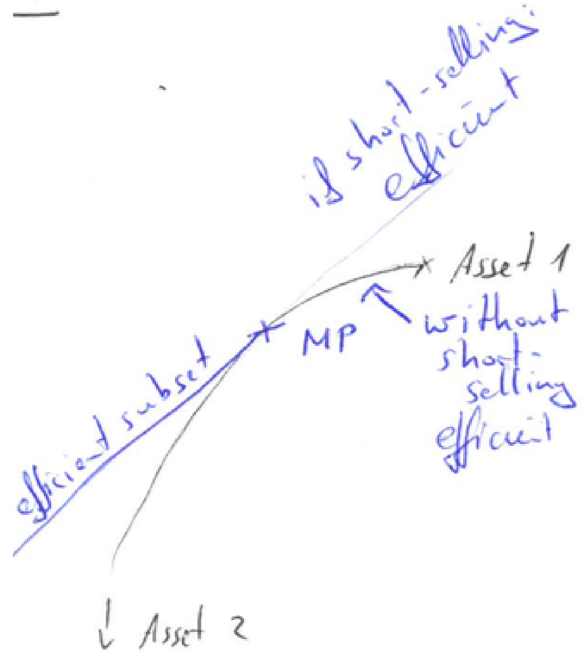
b) investing $w_0 = 1/2$ into the bank account and investing the rest to BP and Shell with the same ratio's as in part a).

Investment (b) then means to put £500 into the bank account, and investing £350 in BP and £150 in Shell.

In the mean-variance diagram, portfolio (b) is in the middle of portfolio (a) and the riks-free portfolio.

See the slides for further examples.

Efficient frontier with a risk-free asset



The efficient frontier is a tangent to the attainable set of all portfolios that include assets 1&2. (Special cases do exist in the case of no short-selling and are outlined in the lecture slides).

The tangent point, which connects the risk-free security with the attainable set of assets 1&2 is called *Market Portfolio* (MP).

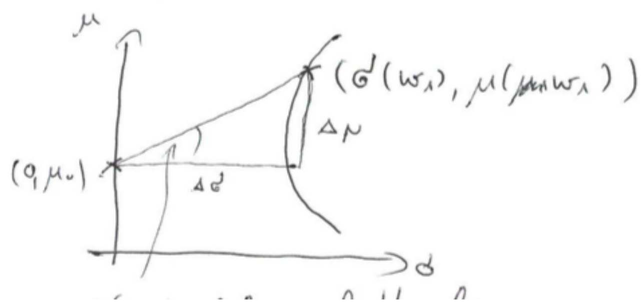
Theorem 5. *If $\mu_0 < \mu_{\text{MVP}}$ the weights of the market portfolio are*

$$w_1 = \frac{c}{c+d}, \quad w_2 = \frac{d}{c+d},$$

where

$$c = \sigma_2^2(\mu_1 - \mu_0) - \rho\sigma_1\sigma_2(\mu_2 - \mu_0),$$
$$d = \sigma_1^2(\mu_2 - \mu_0) - \rho\sigma_1\sigma_2(\mu_1 - \mu_0)$$

Proof sketch (This has not been part of the lecture; it is only included for your own interest) Maximise the slope of the line connecting any portfolio with the risk-free security to find the tangent.



The slope is given in dependency of w_1 as

$$s(w_1) = \frac{\Delta\mu(w_1)}{\Delta\sigma(w_1)},$$

and we need to solve the optimisation problem

$$\max_{w_1 \in \mathbb{R}} s(w_1).$$

We note graphically that there is a unique solution, so we can find it by looking for roots of the derivative:

$$\text{find}\, w_1 \text{ s.t. } \quad s'(w_1) = 0,$$

i.e.

$$0 = s'(w_1) = \frac{\mu'(w_1)\sigma(w_1) - (\mu(w_1) - \mu_0)\sigma'(w_1)}{\sigma^2(w_1)},$$

where

$$\begin{aligned} \mu(w_1) &= w_1\mu_1 + (1 - w_1)\mu_2, \\ \sigma(w_1)^2 &= w_1^2\sigma_1^2 + 2\rho w_1(1 - w_1)\sigma_1\sigma_2 + (1 - w_1)^2\sigma_2^2. \end{aligned}$$

For the derivatives of μ and σ this yields

$$\begin{aligned} \mu'(w_1) &= \mu_1 - \mu_2, \\ \sigma'(w_1 &= \frac{\mathrm{d}}{\mathrm{d}w_1}\left(\sqrt{\sigma^2(w_1)}\right) = \frac{1}{2\sqrt{\sigma^2(w_1)}}\frac{\mathrm{d}}{\mathrm{d}w_1}\left(\sigma^2(w_1)\right) = \frac{1}{2\sqrt{\sigma^2(w_1)}}\left(\sigma^2\right)'(w_1). \end{aligned}$$

This yields

$$\begin{aligned} s'(w_1) &= \frac{\mu'(w_1)\sigma(w_1) - (\mu(w_1) - \mu_0)\frac{1}{2\sigma(w_1)}\left(\sigma^2\right)'(w_1)}{\sigma^2(w_1)} \\ &= \frac{2\mu'(w_1)\sigma^2(w_1) - (\mu(w_1) - \mu_0)(\sigma^2)'(w_1)}{2\sigma^3(w_1)}. \end{aligned}$$

The fraction can only be zero if the nominator is zero, so we solve

$$2\mu'(w_1)\sigma^2(w_1) - (\mu(w_1) - \mu_0)(\sigma^2)'(w_1) = 0,$$

where

$$(\sigma^2)'(w_1) = 2w_1\sigma_1^2 + 2\rho\sigma_1\sigma_2 - 2\rho 2w_1\sigma_1\sigma_2 - 2(1 - w_1)\sigma_2^2.$$

Inserting and solving this finishes the proof.

Why is it called market portfolio? The line connecting the MP with the risk-free asset (called capital market line, CML) forms the efficient frontier of an idealised market. Hence any efficient investment consists of a portion of the MP and the risk-free security. No investor has any reason to divert from this strategy, hence the MP must reflect the whole market.

5 Factor Models of Asset Returns (Weeks 6 and 7)

For further reading, the book of McNeil, Frey and Embrechts, *Quantitative Risk Management*, has a chapter on factor models.

Let's recall some key finding from the past chapters:

- Diversification: Combining two stocks in a portfolio can reduce the risk.
  $\Rightarrow$ We will try to consider as many stocks at the same time as possible
- Risk&Return: Variance has some drawbacks as a risk-model. For more insights, we considered the value-at-risk and the shortfall probability, which can be based on stock price models.

Hence we go back to Chapter 2, and recall e.g. the log-normal model. With the log-normal model, the stock price is modelled as a stochastic process with independent normal log-returns:

$$S(T) = S(0) \exp\left(\sum_{t=0}^{T-1} X(t)\right), \quad X(t) \sim \mathcal{N}(\mu, \sigma) \text{ iid.}$$

For several stocks $S^1, \ldots, S^d$, we can generate several stochastic processes

$$S^i(t) = S^i(0) \exp\left(\sum_{t=0}^{T-1} X^i(t)\right), \quad X^i(t) \sim \mathcal{N}(\mu, \sigma) \text{ iid.}$$

With $X^i(t)$ being independent, also the modelled stock prices for different stocks are independent. This is not what we observe in practice, where significant correlations of different assets can be observed. Therefore, we investigate models that account for the dependence of different stocks.

It is easier to create independent random variables, than dependent ones. Therefore, we try to combine independent random variables to create dependent random variables with certain properties.

Example 4. *Create Y_1, Y_2 normally distributed with*

$$\mathbb{E}(Y_1) = 0, \quad \mathbb{E}(Y_2) = 1, \quad \text{Var}(Y_1) = 1, \quad \text{Var}(Y_2) = 1, \quad \text{Cov}(Y_1, Y_2) = 1/2.$$

Solution: Create $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ iid and set

$$\begin{aligned} Y_1 &= a_1 + b_{11} Z_1 + b_{12} Z_2, \\ Y_2 &= a_2 + b_{21} Z_1 + b_{22} Z_2. \end{aligned}$$

and try to find the correct parameters, i.e.

$$\begin{aligned} 0 &= \mathbb{E}(Y_1) = \mathbb{E}(a_1 + b_{11} Z_1 + b_{12} Z_2) = a_1 \Rightarrow a_1 = 0 \\ 1 &= \text{Var}(Y_1) = \text{Cov}(Y_1, Y_1) = \text{Cov}(a_1 + b_{11} Z_1 + b_{12} Z_2, a_1 + b_{11} Z_1 + b_{12} Z_2) \\ &= b_{11}^2 + b_{12}^2, \\ &\vdots \end{aligned}$$

Compare to the coursework. As we have 6 parameters with 5 (non-linear) conditions, there might exist several possible solutions.

Now let's come back to stock prices and for simplicity consider only a single time-step:

$$X_i = X_i(0), \quad i = 1, \ldots, d.$$

Full model. As seen in the example, we can model this as

$$X_i = a_i + \sum_{j=1}^{d} b_{ij} Z_j, \quad \text{with } Z_j \sim \mathcal{N}(0,1)$$

for $i = 1, \dots, d$. We can rewrite this as a matrix-vector product for

$$\mathbf{X} = (X_i)_{i=1,\dots,d}, \mathbf{B} = (b_{ij})_{i,j=1,\dots,d}, \mathbf{a} = (a_i)_{i=1,\dots,d}, \mathbf{Z} = (Z_j)_{j=1,\dots,d},$$

as

$$\mathbf{X} = \mathbf{a} + \mathbf{BZ}.$$

The *problem* with this full model is the large dimension to consider. This yields a lot of parameters to fit and makes the evaluation more difficult. E.g. when considering all FTSE 100 stocks, $d^2 = 10\,000$ parameters need to be fitted for the covariances alone.

To reduce these numbers, we do not consider the full dependency, but a dependency on a number of common factors

$$\mathbf{F} = (F_j)_{j=1,\dots p}$$

and consider $\mathbf{BF}$ with $\mathbf{B} \in \mathbb{R}^{d\times p}$ for $p \leq d$. This can drastically reduce the amount of parameters, e.g. for $p = 5$ only 500 parameters need to be fitted for the FTSE 100.

However, $\mathbf{BF}$ does not allow for individual risk of each stocks (with 5 random processes for 100 stocks, they will all behave in very similar patterns). Therefore, we add idiosyncratic terms $\boldsymbol{\varepsilon}$ to the model, which are independent, individual risks $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1,\dots d}$: $\mathbf{a} + \mathbf{BF} + \boldsymbol{\varepsilon}$ with ε_i iid and $\mathbb{E}(\varepsilon_i) = 0$.

Linear Factor Model / p-Factor Model This yields

$$\mathbf{X} = \mathbf{a} + \mathbf{BF} + \boldsymbol{\varepsilon},$$

with $\mathbf{a} \in \mathbb{R}^d$ deterministic, $\mathbf{B} \in \mathbb{R}^{d\times p}$ deterministic, $\mathbf{F} = (F_1, \dots, F_p)^\top \in \mathbb{R}^p$ stochastic and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)^\top \in \mathbb{R}^d$ stochastic. We assume $\operatorname{corr}(F_j, \varepsilon_i) = 0$. The result $\mathbf{X} = (X^i)_{i=1,\dots,n}$ is the vector of the modelled log-returns on the assets.

- $\mathbf{F}$ is a random vector of common factors with $p < d$ and a covariance matrix that is positive definite (holds e.g. when F_j are independent of each other);
- $\boldsymbol{\varepsilon}$ is a random vector of idiosyncratic terms, which are uncorrelated and have mean zero;
- $\mathbf{B} \in \mathbb{R}^{d\times p}$ is a matrix of constant factor loadings, $\mathbf{a} \in \mathbb{R}^d$ a vector of constants;

Note: we have formulated a quite general case. An easy example would be a log-normal factor models with $F_j \sim \mathcal{N}(0,1), \varepsilon_i \sim \mathcal{N}(0,1)$ iid and $a_i = \mu_i$. However in practice, the factors may not be independent, e.g. when several market indices are considered. We will look into the choice of the factors next. The idiosyncratic terms may depend on each other (but not correlate), e.g. when volatility clustering is considered.

Types of factors

1. Macroeconomical factors;
2. Fundamental factors;
3. Statistical factors.

In more detail:

1. *Macroeconomical factors* are observable economic time-series, e.g. market indices, industrial indices, interest rates, inflation, etc.

   Example: F_1 is the value of the FTSE 100 index, F_2 of the FTSE 4GOOD index, F_3 the value of the FTSE All-Share index. Then the factor loadings $\mathbf{B} \in \mathbb{R}^{d\times 3}$ are unknown and must be estimated.

2. *Fundamental factors* are *not* observable. Examples are the industry sector, the country or the continent. Here the factor loading is known, but the factors need to be modelled.

   Example: consider the assets

   (a) HSBC,
   (b) Barclays,
   (c) Shell,
   (d) BP,
   (e) Bank of America,

   and the factors

   (a) Company in the financial industry;
   (b) Company in oil&gas;
   (c) Company located in Europe.

   Then we have three factors $\mathbf{F} = (F_1, F_2, F_3)$, where the factor loading is given by

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

   The factors $\mathbf{F}$ need to be modelled.

3. *Statistical factors* are factors purely based on statistical evidence. Neither $\mathbf{B}$ nor $\mathbf{F}$ are observable. Examples include PCA (Principal Component Analysis) or statistical factor analysis.

   As an example, let us consider PCA for a single factor. There the stock prices are combined linearly and we seek the stochastic processes, which can best explain the stocks (for processes with mean zero). This is an optimisation problem, where we minimise the mean-squared-error of the combined process to the individual stocks as follows.

   Based on the historical data $X^i(t), i = 1, \ldots, d, t = 0, \ldots N$ we find $c_i, i = 1, \ldots, d$ and $\mathbf{B} = (b_{i1})_{i=1,\ldots d} \in \mathbb{R}^{d\times 1}$, such that

$$\sum_{t=0}^{N} \sum_{i=1}^{d} \left\| b_i \underbrace{\sum_{j=1}^{d} c_j (X^j(t) - \mu_j)}_{F_1} - (X^i(t) - \mu_i) \right\|^2$$

   is minimal.

   Statistic factors can be very powerful, as they need only few factors for a good approximation. However, the factors lack interpretability.

5.1 Single factor models

A single factor model is the most simple form of factor models. It restricts the possible values of the correlation, but allows for a clearer interpretation. In this subsection, we consider single factor models of different complexities.

With $p = 1$ we have

$$X^i = a_i + b_i F + \varepsilon_i, \quad i = 1, \ldots, d,$$

where F is a stochastic joint factor. An example would be a broad market index, e.g. the FTSE All-Shares index.

Mean, variance and covariance are:

- $\mathbb{E}(X^i) = a_i + b_i \mathbb{E}(F)$;
- $\mathrm{Var}(X^i) = b_i^2 \mathrm{Var}(F) + \mathrm{Var}(\varepsilon_i)$ as ε_i and F are uncorrelated;
- $\mathrm{Cov}(X^i, X^j) = b_i b_j \mathrm{Var}(F)$ for $i \neq j$ (also due to ε_i and F being uncorrelated).

A special one-factor model is *Sharpe's Single-Index Model (SIM)*, which brings an economic interpretation:

$$R^i - \mu_0 = \alpha_i + \beta_i (R^{\mathrm{M}} - \mu_0) + \varepsilon_i.$$

- μ_0 is the risk-free rate;
- R^{M} is the return of the market portfolio;
- $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ independent of each other and of R^{M}.
- α_i: the stock's alpha: abnormal results
- β_i: the stock's beta: responsiveness to the market return.

The model was developed by William Sharpe in 1963 and is widely used in practice.

- All returns are corrected by the risk-free rate
- With the market portfolio as the single factor, the return of the stock is decomposed into three parts:
  - The abnormal return: any returns that outperforms the market consistently is influenced by the alpha α_i;
  - The return may have a different responsiveness to market movements. With a small value of β the stock price reacts only slightly to changes of the market. With β larger than one, the stock price reacts stronger than the market;
  - Each stock has an idiosyncratic risk, independent of the market and other stocks.

We will come across β again in our next chapter on CAPM.

Equicorrelation model (extra reading) A second, even simpler, single factor model is the equicorrelation model:

$$X^i = \sqrt{\rho} F + \sqrt{1 - \rho} \varepsilon_i, \quad i = 1, \dots, d,$$

where F is the single factor with $F, \varepsilon \sim \mathcal{N}(0, 1)$ iid, $\rho \in (0, 1)$.

As the name suggests, the random variables X^i have the same mutual correlation coefficient:

$$\begin{aligned}
\mathbb{E}(X^i) &= \sqrt{\rho} \mathbb{E}(F) = 0; \\
\mathrm{Var}(X^i) &= \rho \mathrm{Var}(F) + (1 - \rho) \mathrm{Var}(\varepsilon_i) = 1; \\
\mathrm{Cov}(X^i, X^j) &= \mathrm{Cov}(\sqrt{\rho} F + \sqrt{1 - \rho} \varepsilon_i, \sqrt{\rho} F + \sqrt{1 - \rho} \varepsilon_j) \\
&= \rho \mathrm{Var}(F) = \rho, \quad i \neq j,
\end{aligned}$$

hence $\mathrm{corr}(X^i, X^j) = \varrho$.

Where is this useful

- if you have several equal risks, e.g. with homogeneous credit portfolios;
- as a (drastically) simplified version of Sharpe's SIM:

  $$R^i = \sqrt{\rho} R^{\mathrm{M}} + \sqrt{1 - \rho} \varepsilon_i,$$

  where $\varepsilon_i, R^{\mathrm{M}} \sim \mathcal{N}(0, 1)$ and $\mu_0 = 0$ are assumed for simplicity.

Due to it's simplicity, the model allows us to consider huge portfolios by hand:
Let us consider d stocks in an equicorrelation model and a portfolio of equal parts:

$$R^{\mathrm{P}} = 1/d \sum_{i=1}^{d} R^i$$

and

$$R^i = \sqrt{\rho} R^{\mathrm{M}} + \sqrt{1-\rho}\varepsilon_i.$$

Then expectation and variance are given as

$$\begin{aligned}
\mathbb{E}(R^{\mathrm{P}}) &= \frac{1}{d}\sum_{i=1}^{d} \mathbb{E}(R^i) = 0,\\
\operatorname{Var}(R^{\mathrm{P}}) &= \frac{1}{d^2}\operatorname{Var}\left(\sum_{i=1}^{d} R^i\right) = \frac{1}{d^2}\sum_{i,j=1}^{d}\operatorname{Cov}(R^i, R^j)\\
&= \frac{1}{d^2}\sum_{i,j=1}^{d}(\rho + (1-\rho)\delta_{ij}) = \frac{1}{d^2}(d^2\rho + d(1-\rho))\\
&= \rho + \frac{1-\rho}{d} \to^{d\to\infty} \rho.
\end{aligned}$$

We immediately note two interesting facts:

1. With large portfolios the individual variance $\operatorname{Var}(R^i)$ becomes negligible in comparison to the covariances $\operatorname{Cov}(R^i, R^j)$, due to the curse of dimension.

2. Investing into many assets eliminates the risk only up to a certain limit (dictated by the market return).

This shows two different kinds of risk:

- The *systemic risk* which is due to the common factor, the market return R^{M}.

- The *specific risk* which is individual to each asset and independent of other assets.

By diversification only the specific risk can be reduced. The systemic risk remains and will be estimated further in the next chapter.

6 Pricing (Weeks 7 and 8)

Pricing models aim to explain the expected return of assets.

6.0 Mean-variance portfolio theory for several assets

Note: This subsection contains interesting insights and prepares for CAPM, but is not relevant for the exam.

Before introducing the Capital Asset Pricing Model (CAPM), let us extend the portfolio theory of Chapter 4 to several risky assets. Let d risky assets be given by:

- vector of their expected returns $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top = \left(\mathbb{E}(R^1), \ldots, \mathbb{E}(R^d)\right)^\top \in \mathbb{R}^d$

- matrix of the pairwise covariances (symmetric and assumed to be invertible)

$$\mathbf{C} = \begin{pmatrix} \operatorname{Cov}(R^1, R^1) & \operatorname{Cov}(R^1, R^2), & \cdots & \operatorname{Cov}(R^1, R^d)\\ \operatorname{Cov}(R^2, R^1) & \operatorname{Cov}(R^2, R^2), & \cdots & \operatorname{Cov}(R^2, R^d)\\ \cdots & \cdots & \ddots & \vdots\\ \operatorname{Cov}(R^d, R^1) & \operatorname{Cov}(R^d, R^2), & \cdots & \operatorname{Cov}(R^d, R^d) \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

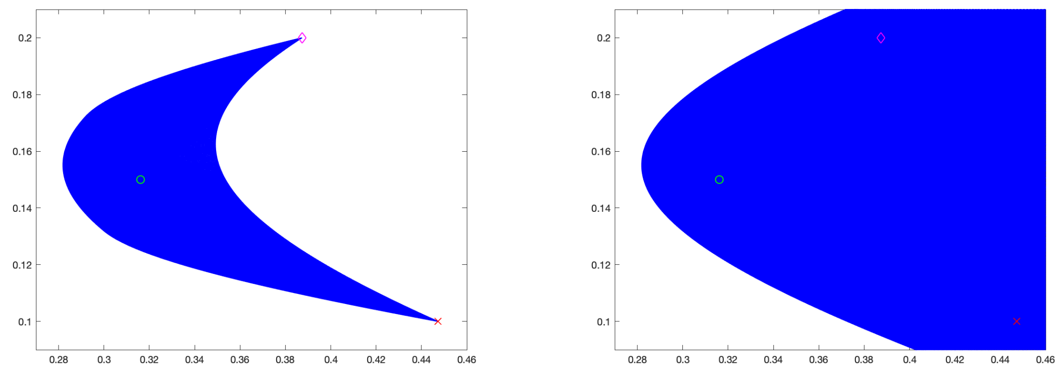


Figure 17: Attainable set for a portfolio composed of three assets (marked in the figure). Left: without short-selling; Right: with short-selling.

A portfolio is given by any weights $\mathbf{w} = (w_1, \dots, w_d)^\top \in \mathbb{R}^d$ with $\sum_{i=1}^d w_i = 1$. Using the vector $\mathbb{1} = (1, \dots, 1)^\top \in \mathbb{R}^d$, we can rewrite it as $\sum_{i=1}^d w_i = \mathbf{w}^\top \mathbb{1}$ (which is the scalar product of the vector of weights $(w_1, \dots, w_d)^\top$ with the vector of ones: $(1, \dots, 1)^\top$).

Analogue to Chapter 4, the portfolio's return is given as $R^{\mathrm{P}} = \sum_{i=1}^d R^i$. We can use this to compute the expected return μ_{P} and variance μ_{P} of the portfolio:

$$\mu_{\mathrm{P}} = \sum_{i=1}^d w_i \mu_i = \mathbf{w}^\top \boldsymbol{\mu},$$
$$\sigma_{\mathrm{P}}^2 = \sum_{i,j=1}^d w_i w_j \operatorname{Cov}(R^i, R^j) = \mathbf{w}^\top \mathbf{C} \mathbf{w}.$$

The attainable set is then given as

$$\left\{(\sigma_{\mathrm{P}}, \mu_{\mathrm{P}}) : \ \sigma_{\mathrm{P}}^2 = \mathbf{w}^\top \mathbf{C} \mathbf{w}, \quad \mu_{\mathrm{P}} = \mathbf{w}^\top \boldsymbol{\mu}, \quad \mathbf{w}^\top \mathbb{1} = 1\right\}.$$

An example of the attainable set with and without short-selling can be found in Figure 17.

The attainable set is no longer a line, but in general an area. For simplicity, we neglect any restrictions on short-selling.

As the interior of the set does not contain any efficient portfolios, it is sufficient to consider its boundary: *The minimal variance line.*

Given a value m for the expectation, find a portfolio with $\mu_{\mathrm{P}} = m$ and minimal variance:

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^d} \ & \mathbf{w}^\top \mathbf{C} \mathbf{w} \\ \text{s.t. } & \mathbf{w}^\top \boldsymbol{\mu} = m, \\ & \mathbf{w}^\top \mathbb{1} = 1. \end{aligned}$$

This is a *constrained optimisation problem*, which requires advanced techniques to solve it. If you are interested, read Chapter 3 of *Capinski, Kopp*. The optimisation problem is then solved in Theorems 4.6–4.9 and it is shown, that the minimal variance line is a hyperbola.

With the minimal variance line as a hyperbola (as the attainable set used to be in Chapter 4), we can apply most of our theory of Chapter 4:

Minimal variance portfolio. The MVP is found solving the constrained optimisation problem

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^d} \ & \mathbf{w}^\top \mathbf{C} \mathbf{w} \\ \text{s.t. } & \mathbf{w}^\top \mathbb{1} = 1. \end{aligned}$$

Theorem 4.4. of *Capinski, Kopp* shows that this is solved by

$$\mathbf{w}_{\min} = \frac{\mathbf{C}^{-1}\mathbb{1}}{\mathbb{1}^\top \mathbf{C}^{-1}\mathbb{1}}.$$

Market Portfolio. If the risk-free return μ_0 is less than the expected return of the minimal variance portfolio μ_{MVP}, the market portfolio exists:

$$\mu_{\mathrm{MP}} = \frac{\mathbf{C}^{-1}(\boldsymbol{\mu} - \mu_0\mathbb{1})}{\mathbb{1}^\top \mathbf{C}^{-1}(\boldsymbol{\mu} - \mu_0\mathbb{1})}$$

(see Theorem 4.10 of *Capinski, Kopp*;. The MP can be found by optimising the slope of the portfolio, constrained by $\mathbf{w}^\top \boldsymbol{\mu} = 1$.)

The capital market line is given as $\mu = \mu_0 + \frac{\mu_{\mathrm{MP}} - \mu_0}{\sigma_{\mathrm{MP}}}\sigma$ and is the efficient frontier.

6.1 The Captial Asset Pricing Method (CAPM)

Asset pricing tries to explains why different assets have different expectation values. The main idea is that a higher expected value is a reward for taking a higher risk.

6.1.1 CAPM formula:

Theorem 6. *Assume $\mu_0 < \mu_{\mathrm{MVP}}$ (then the market portfolio $(\sigma_{\mathrm{MP}}, \mu_{\mathrm{MP}})$ exist). Then the return of the i-th asset is given as*

$$\mu_i = \mu_0 + \beta_i(\mu_{\mathrm{MP}} - \mu_0), \quad i = 1, \dots, d,$$

where $\beta_i = \frac{\mathrm{Cov}(R^i, R^{\mathrm{MP}})}{\sigma_{\mathrm{MP}}^2}$ is the beta factor *of the asset.*

Proof. We consider portfolios consisting of asset i and the market portfolio. For some $w \in \mathbb{R}$, we invest a portion w in asset i and a portion $1 - w$ in the market portfolio. This yields expected return and variance

$$\begin{aligned}
\mu_w &= w\mu_i + (1-w)\mu_{\mathrm{MP}}, \\
\sigma_w^2 &= w^2\sigma_i^2 + (1-w)^2\sigma_{\mathrm{MP}}^2 + 2w(1-w)\,\mathrm{Cov}(R^i, R^{\mathrm{MP}}).
\end{aligned}$$

From Chapter 4, we know that $\{(\sigma_w, \mu_w) : w \in \mathbb{R}\}$ is a hyperbola, which passes through $(\sigma_{\mathrm{MP}}, \mu_{\mathrm{MP}})$ with the CML (capital market line) as the tangent.

Thus the slope of the tangent for $w = 0$ is the same as the slope of the CML:

$$\frac{\Delta\mu}{\Delta\sigma} = \frac{\mu_{\mathrm{MP}} - \mu_0}{\sigma_{\mathrm{MP}}}.$$

We can evaluate the tangent of the hyperbola by its derivative with respect to w:

$$\left.\frac{\partial \mu_w}{\partial w}\right|_{w=0} = \mu_i - \mu_{\mathrm{MP}};$$

as well as

$$\begin{aligned}
\left.\frac{\partial \sigma_w}{\partial w}\right|_{w=0} &= \left.\frac{1}{2\sigma_w}\right|_{w=0} \cdot \left.\frac{\partial(\sigma_w^2)}{\partial w}\right|_{w=0} \\
&= \frac{1}{2\sigma_{\mathrm{MP}}} \cdot \left(2w\sigma_i^2 + 2(w-1)\sigma_{\mathrm{MP}}^2 + 2(1-2w)\,\mathrm{Cov}(R^i, R^{\mathrm{MP}})\right)\Big|_{w=0} \\
&= \frac{\mathrm{Cov}(R^i, R^{\mathrm{MP}}) - \sigma_{\mathrm{MP}}^2}{\sigma_{\mathrm{MP}}}.
\end{aligned}$$

Thus, the slope of the tangent is

$$\left.\frac{\frac{\partial \mu_w}{\partial w}}{\frac{\partial \sigma_w}{\partial w}}\right|_{w=0} = \frac{\mu_i - \mu_{\mathrm{MP}}}{\frac{\mathrm{Cov}(R^i, R^{\mathrm{MP}}) - \sigma_{\mathrm{MP}}^2}{\sigma_{\mathrm{MP}}}}.$$

As it is the same as the slope of the CML, we can set it equal to $\frac{\mu_{\mathrm{MP}} - \mu_0}{\sigma_{\mathrm{MP}}}$ and solve for μ_i:

$$\begin{aligned}\mu_i - \mu_{\mathrm{MP}} &= \frac{\mu_{\mathrm{MP}} - \mu_0}{\sigma_{\mathrm{MP}}} \cdot \frac{\mathrm{Cov}(R^i, R^{\mathrm{MP}}) - \sigma_{\mathrm{MP}}^2}{\sigma_{\mathrm{MP}}} = (\beta_i - 1)(\mu_{\mathrm{MP}} - \mu_0) \\ &= \beta_i(\mu_{\mathrm{MP}} - \mu_0) - (\mu_{\mathrm{MP}} - \mu_0),\end{aligned}$$

Thus

$$\mu_i - \mu_{\mathrm{MP}} = \beta_i(\mu_{\mathrm{MP}} - \mu_0).$$

□

Remark 1. *The same formula holds for portfolios:*

$$\mu_{\mathrm{P}} = \mu_0 + \beta_{\mathrm{P}}(\mu_{\mathrm{MP}} - \mu_0),$$

where $\beta_{\mathrm{P}} = \frac{\mathrm{Cov}(R^{\mathrm{P}}, R^{\mathrm{MP}})}{\sigma_{\mathrm{MP}}^2}$.

In the CAPM formula, we see that the expected return of any investment is only determined by the covariance with the market portfolio.

Interpretation

- $\beta(\mu_{\mathrm{MP}} - \mu_0)$ is called *risk-premium*. It rewards investors, who expose themselves to a higher market risk.
- the idiosynchratic term ε_{P}, such that

  $$R^{\mathrm{P}} - \mu_0 = \beta_{\mathrm{P}}(R_{\mathrm{MP}} - \mu_0) + \varepsilon_{\mathrm{P}},$$

  represents the specific/diversifiable risk. We can compute the variance:

  $$\sigma_{\mathrm{P}}^2 = \beta_{\mathrm{P}}^2 \sigma_{\mathrm{MP}}^2 + \mathrm{Var}(\varepsilon_{\mathrm{P}}).$$

  As the specific risk ε_{P} can be diversified (by buying the market portfolio), it is not rewarded by a higher expected return.

  Computation of the variance:

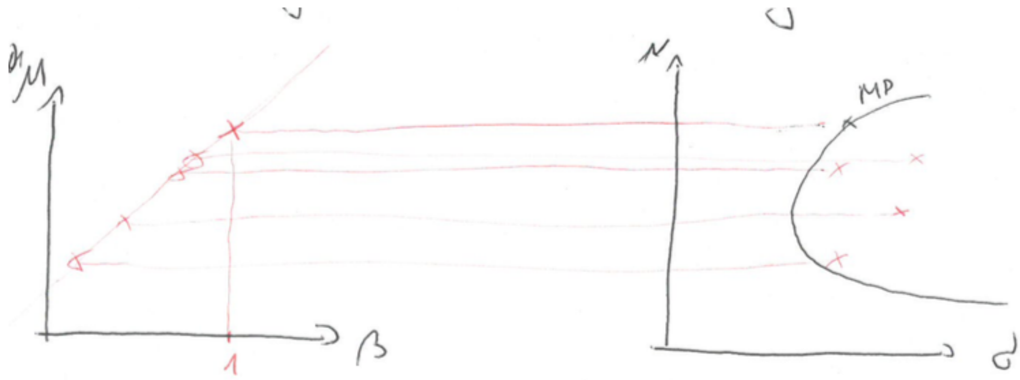
  $$\mathrm{Var}(R^{\mathrm{P}}) = \mathrm{Var}(\mu_0 + \beta_{\mathrm{P}}(R_{\mathrm{MP}} - \mu_0) + \varepsilon_{\mathrm{P}}) = \beta_{\mathrm{P}}^2 \,\mathrm{Var}(R^{\mathrm{MP}}) + \mathrm{Var}(\varepsilon_{\mathrm{P}}) + \beta_{\mathrm{P}} \,\mathrm{Cov}(R^{\mathrm{MP}}, \varepsilon_{\mathrm{P}}),$$

  where

  $$\begin{aligned}\mathrm{Cov}(R^{\mathrm{MP}}, \varepsilon_{\mathrm{P}}) &= \mathrm{Cov}\left(R^{\mathrm{MP}}, R^{\mathrm{P}} - \mu_0 - \beta_{\mathrm{P}}(R^{\mathrm{MP}} - \mu_0)\right) \\ &= \mathrm{Cov}(R^{\mathrm{MP}}, R^{\mathrm{P}}) - \beta_{\mathrm{P}} \,\mathrm{Cov}(R^{\mathrm{MP}}, R^{\mathrm{MP}}) = 0.\end{aligned}$$

6.1.2 The security market line (SML)

CAPM yields a linear relation between the expected return and the beta of a portfolio. Plotting several portfolios in a β-μ-diagram, they should form a line: the security market line.



6.1.3 Efficient portfolios

A portfolio is efficient if and only if $\operatorname{corr}(R^{\mathrm{P}}, R^{\mathrm{MP}}) = 1$, see coursework. This yields

$$\beta_{\mathrm{P}} = \frac{\operatorname{Cov}(R^{\mathrm{P}}, R^{\mathrm{MP}})}{\sigma_{\mathrm{MP}}^2} = \frac{\sigma_{\mathrm{P}}\sigma_{\mathrm{MP}} \operatorname{corr}(R^{\mathrm{P}}, R^{\mathrm{MP}})}{\sigma_{\mathrm{MP}}^2} = \frac{\sigma_{\mathrm{P}}}{\sigma_{\mathrm{MP}}}.$$

Thus for *efficient portfolios* there is no diversifiable risk and

$$\begin{aligned} \mu_{\mathrm{P}} - \mu_0 &= \beta_{\mathrm{P}}(\mu_{\mathrm{MP}} - \mu_0) \\ \sigma_{\mathrm{P}} &= \beta_{\mathrm{P}}\sigma_{\mathrm{MP}}. \end{aligned}$$

6.1.4 How to use CAPM?

In the following, we discuss some ways how CAPM can be used for portfolio analysis.

- Value portfolios using the *Sharpe ratio*:

$$\frac{\mu_{\mathrm{P}} - \mu_0}{\sigma_{\mathrm{P}}}.$$

  The larger the Sharpe ratio is, the more efficient is the portfolio. For efficient portfolios, we have

$$\frac{\mu_{\mathrm{P}} - \mu_0}{\sigma_{\mathrm{P}}} = \frac{\beta_{\mathrm{P}}(\mu_{\mathrm{MP}} - \mu_0)}{\beta_{\mathrm{P}}\sigma_{\mathrm{MP}}} = \frac{\mu_{\mathrm{MP}} - \mu_0}{\sigma_{\mathrm{MP}}},$$

  where the Sharpe ratio of the market portfolio is maximal:

$$\mathbf{w}_{\mathrm{MP}} = \operatorname*{arg\,max}_{\mathbf{w}, \mathbf{w}^\top \mathbb{1} = 1} \frac{\mathbf{w}^\top \boldsymbol{\mu} - \mu_0}{\mathbf{w}^\top \mathbf{C} \mathbf{w}}$$

  (This holds by construction of the market portfolio, as $(\mathbf{w}^\top \boldsymbol{\mu} - \mu_0)/(\mathbf{w}^\top \mathbf{C}\mathbf{w})$ is the slope of the line connecting a portfolio with the risk-free asset.)

- Valuing the stock price:
  CAPM models the required return for the taken risk. Strong deviation from the model can be used as an investment strategy.
  - If $\mu_i - \mu_0 > \beta_i(\mu_{\mathrm{MP}} - \mu_0)$, the stock is underpriced and you could buy it (as the expected return is larger than the required return)
  - If $\mu_i - \mu_0 < \beta_i(\mu_{\mathrm{MP}} - \mu_0)$, the stock is overpriced and you could sell it or short-sell it (as the expected return is less than the required return considering the taken risk).

  Remark: These transactions will push the market price towards predicted value and the market gains efficiency.

- Performance measure of a stock using Jensen's alpha:
  Jensen's alpha is the difference of the realised return and the required return:

  $$\alpha = R^{\mathrm{P}} - (\mu_0 + \beta_{\mathrm{P}}(R^{\mathrm{MP}} - \mathbb{R}^{\mathrm{P}})),$$

  where for this application, R^{P} and R^{MP} are the *realised returns*, i.e. empirical data. Jensen's alpha measures the past performance in comparison to the required return.

6.1.5 Discussion of the validity

As a mathematical model, CAPM has only few assumptions, e.g.,

- No trading costs (including brokerage fees, bid-ask-spread, taxes, etc);
- No restrictions on short-selling and borrowing money for the same rate as lending money;
- Available values for all covariances and expectations.

Knowing the parameters is crucial to compute the market portfolio, but not required beyond that. A practical approach is thus to replace the "mathematically computed" market portfolio w_{MP} by a market-index.

(Reminder: If all investors buy efficient portfolios, everyone buys portions of w_{MP}, hence it reflects the whole market and can be replaced by an index)

This requires severely more assumptions, e.g.,

- all investors have the same time horizon;
- all investors can borrow or lend money with no risk at the same rate;
- all investors are non-satiated, risk-averse and trade purely based on σ and μ;
- "perfect market" – information is freely and instantly available, no investor believes they can affect the price by their actions
- all investors have the same estimates for the parameters
- all investors measure in the same currency (e.g. pounds/dollars/ "value").

We see that the assumptions to have $\mathbf{w}_{\mathrm{MP}}$ reflect the whole market are quite strong and often criticised as being unrealistic.

How can we test CAPM? Idea: Plot the security market line.

Remember: The CAPM formula

$$\mu_i - \mu_0 = \beta_i\,(\mu_{\mathrm{MP}} - \mu_0)$$

predicts a linear relationship between the excess return $\mu_i - \mu_0$ and $\beta_i = \mathrm{Cov}(R^i, R^{\mathrm{MP}})/\sigma^2_{\mathrm{MP}}$. The line in the $\beta - \mu$-plane is called security market line (SML). We can plot the line by estimating these parameters. Let $(S^i_t)_{t=1,\ldots N+1}$ be the historic time-series of asset prices of asset i and their log-returns $(X^i_t)_{t=1,\ldots N}$,

$$X^i_t = \log\left(\frac{S^i_{t+1}}{S^i_t}\right).$$

With the log-returns of a market index $(X^{\mathrm{MP}}_t)_{t=1,\ldots N}$, we can estimate

$$\mu_i = \mathbb{E}(X^i_t) \approx \frac{1}{N}\sum_{t=1}^{N} X^i_t,$$

$$\mu_{\mathrm{MP}} = \mathbb{E}(X^{\mathrm{MP}}_t) \approx \frac{1}{N}\sum_{t=1}^{N} X^{\mathrm{MP}}_t.$$

$\beta_i = \text{Cov}(R^i, R^{\text{MP}})/\sigma^2_{\text{MP}}$ can be estimated with

$$\text{Cov}(X^i_t, X^{\text{MP}}_t) \approx \frac{1}{N-1} \sum_{i=1}^{N} (X^i_t - \mu_i)(X^{\text{MP}}_t - \mu_{\text{MP}}),$$
$$\text{Var}(X^{\text{MP}}_t) \approx \frac{1}{N-1} \sum_{i=1}^{N} (X^{\text{MP}}_t - \mu_{\text{MP}})^2.$$

Finally, the risk-free rate μ_0 needs to be defined, e.g. by a government bond.

Then all data points can be plotted in the $\beta - \mu$-plane, as shown during the lecture. The resulting figure did not resemble a line.

Why is the result of our test so bad? Several effects have an influence:

- The estimates, especially for μ are very sensible
- the risk-free return is unclear
- estimating a complete market portfolio is very difficult. FTSE 100 includes only 100 stocks out of over 2,000 companies traded on the London Stock Exchange.
  - we could go for a global index, e.g. MSCI World Index. However this still only includes 1,643 stocks from 23 countries, excluding e.g. Brazil and China.
  - even if we found a perfect stock index, it lacks alternative investments which should be taken into account, e.g. real-estates, private companies, human capital, ...

Some literature on CAPM and it's validity

- Famous study of Black, Jensen, Scholes: "The Capital Asset Pricing Model: Some Empirical Tests", 1972
- Some further remarks by Black: "Beta and Return".

To get more robust parameter estimates, they consider portfolios instead of single stocks. The result indeed form a line, though the slope is smaller than expected.

6.2 The arbitrage pricing theory (APT)

(Ross et al, 1976)

- APT is a more general approach to asset pricing than CAPM
- it does not require an equilibrium (meaning that every investor holds the same portfolio)
- the price depends on certain factors with the only condition being that it must not allow for arbitrage.

Reminder on factor models:

$$R^i = a_i + b_{i1}I_1 + b_{i2}I_2 + \ldots + b_{iL}I_L + \varepsilon_i,$$

with

- R^i the (stochastic) return of asset i,
- a_i deterministic, ε_i the idiosyncratic risk (stochastic),
- $I_1, \ldots, I_L$ the (stochastic) return of L factors/indices,
- $b_{i1}, \ldots, b_{iL}$ the deterministic factor loadings / sensitivity of security i to the factors,

where $\mathbb{E}(\varepsilon_i) = 0$, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $\text{Cov}(I_k, \varepsilon_j) = 0$.

APT The arbitrage pricing theory considers the factor model and requires that no arbitrage exists. In this case, arbitrage refers to a risk-free profit that is larger than the risk-free rate. This results in the condition $a_i = \mu_0$ for all i. The resulting model then reads

$$\mu_i = \mu_0 + \lambda_1 b_{i1} + \ldots + \lambda_L b_{iL},$$

where $\lambda_k = \mathbb{E}(I_k)$ and μ_0 is the risk-free interest rate.

Why $\mu_0 = a_i$? In an example we can see how arbitrage possibilities can be constructed once the condition does not hold:

Let three diversified portfolios and one factor be given:

- $R^1 = 1 + I_1$,
- $R^2 = 1.5\, I_1$,
- $R^3 = 1/2 I_1$.

As the portfolios are diversified, we can assume $\varepsilon_i = 0$. Now buying R^1 and short-selling $0.5\,(R^2 + R^3)$ yields a safe return of 1, without any necessary investment. This is an arbitrage possibility.

It can be shown in general, that no arbitrage exists if

$$(\mu_i, b_{i1}, \ldots, b_{iL}) \in \mathbb{R}^{L+1}$$

for all i lie on a L-dimensional hyperplane in $\mathbb{R}^{L+1}$. This holds iff $a_i = \mu_i$.

A main advantage compared to CAPM is the higher flexibility. This comes at the cost of the complexity to identify the correct factors and loadings.

Example 5. *One example is the Fama-French three-factor model (not relevant for the exam)*

The three factors are

1. *the excess return of the market:* $\mu_{\mathrm{MP}} - \mu_0$*;*
2. *the out-performance of small companies compared to big companies (measured in terms of market capitalisation, i.e. the product of stock prices and number of stocks);*
3. *the out-performance of companies with a high book-to-market value vs small book-to-market value*

Then the factor model reads

$$\mu_i = \mu_0 + \beta_i\,(\mu_{\mathrm{MP}} - \mu_0) + \lambda_{i1}\mathrm{SMB} + \lambda_{i2}\mathrm{HML},$$

with

- *SMB: small market capitalisation minus big market capitalisation;*
- *HML: high book-to-market ratio minus low book/market.*

7 Utility Theory (Weeks 9 and 10)

Instead of measuring our wealth in absolute value (i.e. money), we consider its utility, i.e. the satisfaction an individual obtains by a particular action.

We can measure the

- utility of the wealth, or
- utility of the payoff of a certain action, which we'll use to value a single investment decision.

Both measures are equivalent, if we assume all other investments to remain at the same value:

$$u(\text{payoff}) = \frac{\widetilde{u}(\text{wealth} + \text{payoff}) - \widetilde{u}(\text{wealth})}{\widetilde{u}(\text{wealth})}.$$

This means that the payoff of a particular payoff is a scaled version of the utility function of the wealth. We therefore only consider the utility of the payoff for a certain action.

Definition 5. *A utility function is a function $u\colon \mathbb{R} \to \mathbb{R}$ which is monotonic increasing.*

Example 6. *With this first example, we demonstrate which utility functions suit different risk attitudes.*

Let two lotteries be given with the payoff

$$L_1 = \begin{cases} £2, & \textit{probability } 50\%, \\ -£1, & \textit{probability } 50\%, \end{cases}$$

$$L_2 = \begin{cases} £2, & \textit{probability } 25\%, \\ £0.5, & \textit{probability } 50\%, \\ -£1, & \textit{probability } 25\%, \end{cases}$$

Both lotteries have the same expected value $£L_1$, while L_1 bears more risk.

- *If you are risk-seeking, you prefer L_1,*
- *If you are risk-averse (risk-avoiding), you prefer L_2,*
- *If you are risk-neutral, you are indifferent.*

To decide for one of the lotteries, we value the expected utility:

$$\mathbb{E}\big(u(L_1)\big) \textit{ vs. } \mathbb{E}\big(u(L_2)\big).$$

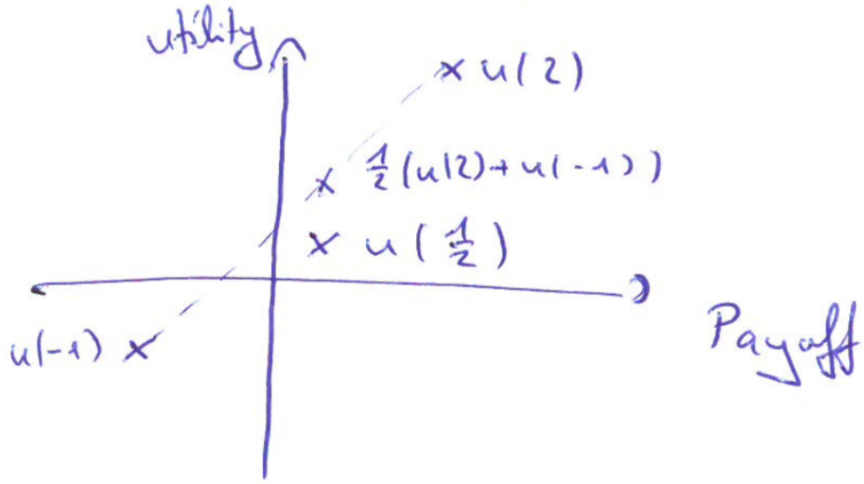
(Note that in general $\mathbb{E}\big(u(L_i)\big) \neq u\big(\mathbb{E}(L_1)\big)$ for nonlinear functions u.)

- $\mathbb{E}\big(u(L_1)\big) = 1/2\, u(2) + 1/2\, u(-1)$
- $\mathbb{E}\big(u(L_2)\big) = 1/4\, u(2) + 1/2\, u(1/2) + 1/4\, u(-1)$

1. *If you are risk-seeking:* $\mathbb{E}\big(u(L_1)\big) > \mathbb{E}(\big(u(L_2)\big)$*:*

$$\Leftrightarrow \quad 1/2\, u(2) + 1/2\, u(-1) > 1/4\, u(2) + 1/2\, u(1/2) + 1/4\, u(-1)$$
$$\Leftrightarrow \quad 1/2\, (u(2) + u(-1)) > u\left(\frac{2-1}{2}\right).$$

This is a convexity condition for u:



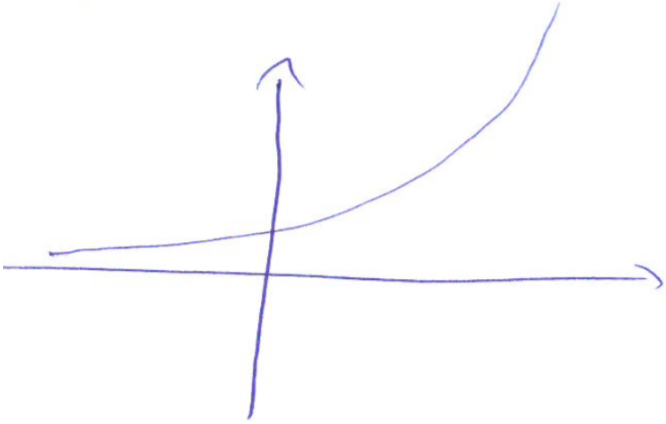
2. *If you are risk-neutral:* $1/2\,\big(u(2)+u(-1)\big) = u\left(\frac{2-1}{2}\right)$*, which means linearity of* u*.*

3. *If you are risk-averse:* $1/2\,\big(u(2)+u(-1)\big) < u\left(\frac{2-1}{2}\right)$*, which is a concavity condition for* u*.*

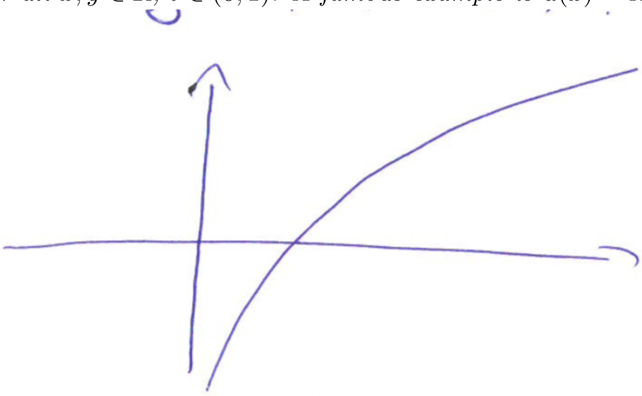
7.1 Reminder: convex and concave functions

Definition 6. *A function* $u\colon \mathbb{R} \to \mathbb{R}$ *is called*

- *strictly convex, iff*
$$u\big(t\,x + (1-t)\,y\big) < t\,u(x) + (1-t)\,u(y),$$
for all $x, y \in \mathbb{R}$*,* $t \in (0,1)$*. A famous example is* $u(x) = \exp(x)$*.*



- *strictly concave, iff*
$$u\big(t\,x + (1-t)\,y\big) > t\,u(x) + (1-t)\,u(y),$$
for all $x, y \in \mathbb{R}$*,* $t \in (0,1)$*. A famous example is* $u(x) = \log(x)$ *(for* $x > 0$*).*



If u is twice differentiable, we can use the sign of the second derivative as a test:

- $u''(x) > 0$ for all $x \in \mathbb{R} \Rightarrow u$ is convex,
- $u''(x) < 0$ for all $x \in \mathbb{R} \Rightarrow u$ is concave.

7.2 Expected utility

Remember, that a utility function is required to be monotonic increasing.

Definition 7.

- *A utility function, that is strictly concave is called risk-averse,*
(Check if $u''(x) < 0$*).*
- *A utility function, that is strictly convex is called risk-seeking,*
(Check if $u''(x) > 0$*).*

- *A utility function, that is linear is called risk-neutral.*

We compare the expected utility for two cases:

1. A lottery L
2. A fixed payments of $\mathbb{E}(L)$.

- If you are risk-averse $\mathbb{E}(u(L)) \leq u(\mathbb{E}(L),$
  i.e. you need to be rewarded for taking risks.
- If you are risk-seeking $\mathbb{E}(u(L)) \geq u(\mathbb{E}(L),$
  i.e. you would pay to get the risky option.
- If you are risk-neutral $\mathbb{E}(u(L)) = u(\mathbb{E}(L),$
  i.e. you do not care about the risk, only about the expected value.

Proof. We proof the inequality for the case of L being a discrete random variable and u being risk-averse (without loss of generality).

Let x_i be the possible outcomes of L, each with a positive probability of p_i, $i = 1, \ldots, N$, and $\sum_{i=1}^{N} p_i = 1$. Then

$$\mathbb{E}\big(u(L)\big) = \sum_{i=1}^{N} p_i \, u(x_i),$$
$$u\big(\mathbb{E}(L)\big) = u\left(\sum_{i=1}^{N} p_i \, x_i\right).$$

Thus, we need to show

$$\sum_{i=1}^{N} p_i \, u(x_i) \leq u\left(\sum_{i=1}^{N} p_i \, x_i\right)$$

for u concave, $p_i > 0$ and $\sum_{i=1}^{N} p_i = 1$.

We proof this by induction over N.[1]

- Base case $N = 1$:
  $p_1 u(x_1) \leq u(p_1 x_1)$ with $p_1 = 1$ is trivially true.
- Induction step $N - 1 \mapsto N$:
  We assume that inequality holds for $N - 1$ terms and then conclude that it is still valid for N terms. This means, we assume

  $$\sum_{i=1}^{N-1} q_i \, u(x_i) \leq u\big(\sum_{i=1}^{N-1} q_i \, x_i\big)$$

  for any $q_i > 0$ with $\sum_{i=1}^{N-1} q_i = 1$. and show that

  $$\sum_{i=1}^{N} p_i \, u(x_i) \leq u\big(\sum_{i=1}^{N} p_i \, x_i\big)$$

  for any $p_i > 0$ with $\sum_{i=1}^{N} p_i = 1$.

  To show this, we first regroup the sum of N items as a sum of $N-1$ item with the remaining item:

  $$\sum_{i=1}^{N} p_i \, u(x_i) = \sum_{i=1}^{N-1} p_i \, u(x_i) + p_N \, u(x_i) = (1 - p_N) \sum_{i=1}^{N-1} \underbrace{\frac{p_i}{1 - p_N}}_{=:q_i} u(x_i) + p_N \, u(x_i)$$

[1] https://en.wikipedia.org/wiki/Mathematical_induction

Defining $q_i = p_1/(1-p_N)$, we have $\sum_{i=1}^{N-1} q_i = 1$ and

$$\sum_{i=1}^{N} p_i\, u(x_i) = (1-p_N) \sum_{i=1}^{N-1} q_i x_i + p_N x_N.$$

Setting $t = 1 - p_N$, $x = \sum_{i=1}^{N-1} q_i x_i$ and $y = x_N$ we can apply the concavity to show

$$\begin{aligned} u\left(\sum_{i=1}^{N} p_i x_i\right) &= u\big(tx + (1-t)y\big) \geq tu(x) + (1-t)u(y) \\ &= (1-p_N)\, u\left(\sum_{i=1}^{N-1} q_i x_i\right) + p_N u(x_N). \end{aligned}$$

On the remaining term, we can apply our induction assumption:

$$\begin{aligned} u\left(\sum_{i=1}^{N} p_i x_i\right) &\geq (1-p_N) u\left(\sum_{i=1}^{N-1} q_i x_i\right) + p_N u(x_N) \\ &\geq (1-p_N) \sum_{i=1}^{N-1} q_i u\left(x_i\right) + p_N u(x_N) \\ &= \sum_{i=1}^{N-1} (1-p_N) q_i u\left(x_i\right) + p_N u(x_N) = \sum_{i=1}^{N} p_i u(x_i). \end{aligned}$$

□

Let's have some examples to see how we can apply the utility theory:

Example 7. *Lotteries L_1 and L_2:*

$$L_1 = \begin{cases} £110, & \textit{probability } 50\%, \\ -£100, & \textit{probability } 50\%, \end{cases}$$

$$L_2 = \begin{cases} £110, & \textit{probability } 25\%, \\ £5, & \textit{probability } 50\%, \\ -£100, & \textit{probability } 25\%. \end{cases}$$

1. *We are risk-seeking with utility function $u(x) = \exp(x/100) - 1$. How are the expected utilities? Which lottery do we prefer?*

$$\begin{aligned} \mathbb{E}\big(u(L_1)\big) &= \frac{1}{2}\left(\exp(1.1) - 1 + \exp(-1) - 1\right) \approx 0.686, \\ \mathbb{E}\big(u(L_2)\big) &= \frac{1}{4}\left(\exp(1.1) - 1 + \exp(-1) - 1 + 2\exp(0.05) - 2\right) \approx 0.11. \end{aligned}$$

   As expected $\mathbb{E}\big(u(L_1)\big) > \mathbb{E}\big(u(L_2)\big)$, we prefer Lottery 1.

2. *We are risk-avoiding. Which lottery do we prefer? We prefer the lottery with the largest expected utility, but must now decide on this question without a concrete function u at hands.*

   With the utility function being concave, we have

$$u(5) > \frac{1}{2}\big(u(110) + u(-100)\big),$$

   which yields

$$\begin{aligned} \mathbb{E}\big(u(L_2)\big) &= \frac{1}{4}u(110) + \frac{1}{2}u(5) + \frac{1}{4}u(-100) \\ &> \frac{1}{4}u(110) + \frac{1}{4}\big(u(110) + u(-100)\big) + \frac{1}{4}u(-100) \\ &= \frac{1}{2}u(110) + \frac{1}{2}u(-100) = \mathbb{E}\big(u(L_1)\big). \end{aligned}$$

7.3 Pricing lotteries based on utility theory

The *certainty equivalent*

$$u^{-1}\Big(\mathbb{E}\big(u(L)\big)\Big)$$

is the single payoff that would result in the same utility as expected for the lottery. Therefore it can be regarded as the amount of money, we would be willing to pay for the lottery.

Note that u^{-1} exists if u is strictly increasing.

Risk-premium If u is risk-averse, we have

$$u^{-1}\Big(\mathbb{E}\big(u(L)\big)\Big) \leq u^{-1}\Big(u\big(\mathbb{E}(L)\big)\Big) = \mathbb{E}(L).$$

The difference

$$\gamma(L) = \mathbb{E}(L) - u^{-1}\Big(\mathbb{E}\big(u(L)\big)\Big) > 0,$$

is called *risk-premium.*

Example 8 (Portfolio optimisation)**.** *From the mean-variance portfolio theory we know that all efficient portfolios lie on the capital market line (CML). However we cannot decide which of these efficient portfolios is optimal. Utility theory can answer this question. Let's assume the following simplified setting:*

With an initial wealth of 1*, we have*

- *Portion* w_0 *invested risk-free with a rate of* $\mu_0 = 10\%$*, i.e. a payoff of* $P_0 = 1.1$.
- *Portion* $1 - w_0$ *invested in the risky market portfolio with payoff*

$$P_1 = \begin{cases} 2 & 50\%, \\ 1/2 & 50\%. \end{cases}$$

We choose a risk-averse utility function for the payoff $u(x) = \log(x)$ *and look for the optimal value of* w_0*:*

$$\max_{w_0 \leq 1} \mathbb{E}\Big(u\big(w_0 P_0 + (1 - w_0)P_1\big)\Big).$$

To compute the expected utility, we first state it as a random variable:

$$u\big(w_0 P_0 + (1 - w_0)P_1\big) = \begin{cases} \log\big(1.1w_0 + 2(1 - w_0)\big), & \textit{probability } 50\%, \\ \log\big(1.1w_0 + 0.5(1 - w_0)\big), & \textit{probability } 50\%. \end{cases}$$

Thus it's expectation is given as

$$\begin{aligned} \mathbb{E}\Big(u\big(w_0 P_0 + (1 - w_0)P_1\big)\Big) &= \frac{1}{2}\log\big(2 - 0.9w_0\big) + \frac{1}{2}\log\big(0.5 + 0.6w_0\big) \\ &= \frac{1}{2}\log\Big(\big(2 - 0.9w_0\big)\big(0.5 + 0.6w_0\big)\Big), \end{aligned}$$

where we used the basic logarithmic law $\log(a \cdot b) = \log(a) + \log(b)$*. Since the logarithm is monotonic increasing, this is equivalent to maximising*

$$\big(2 - 0.9w_0\big)\big(0.5 + 0.6w_0\big),$$

which is again equivalently to maximising

$$\big(20 - 9w_0\big)\big(5 + 6w_0\big) = -54w_0^2 + 75w_0 + 100.$$

The graph of this function is a parabola with opening to the bottom. Therefore we find the maximum as the root of the first derivative:

$$0 = \frac{\mathrm{d}}{\mathrm{d}w_0}\big(-54w_0^2 + 75w_0 + 100\big) = -108w_0 + 75,$$

which yields

$$w_0 = 75/108 \approx 69\%.$$

Thus our expected utility *is maximised with 69% of our money being invested risk-free and 31% in the market portfolio.*

As a comparison, we can see if we can maximise the expected payoff*:*

$$\begin{aligned}\mathbb{E}\big(w_0 P_0 + (1-w_0)P_1\big) &= \frac{1}{2}(2 - 0.9\, w_0) + \frac{1}{2}(0.5 + 0.6\, w_0)\\ &= 1.25 - 0.15\, w_0 \to \infty, \quad \textit{for } w_0 \to -\infty.\end{aligned}$$

The expected payoff can be arbitrarily large, but the (risk-averse) expected utility is bounded.

8 Behavioural Finance (Week 11)

Behavioural finance is a modern extension of classical market models, taking into account the irrationality of the market participants.

Two examples are the following:

Example 9 (Allais paradox)**.** *Named after Maurice Allais, who published it in 1953.*

Consider two different situations, each of them being the choice between two lotteries.

Situation 1:

$$\textit{Lottery 1A:} \qquad \Big\{ £1\,000\,000, \quad \textit{probability } 100\%.$$

$$\textit{vs.}$$

$$\textit{Lottery 1B:} \qquad \begin{cases} £1\,000\,000, & \textit{probability } 89\%,\\ £5\,000\,000, & \textit{probability } 10\%,\\ £0, & \textit{probability } 1\%.\end{cases}$$

Most people prefer Lottery 1A over Lottery 1B

Situation 2:

$$\textit{Lottery 2A:} \qquad \begin{cases} £1\,000\,000, & \textit{probability } 11\%,\\ £0, & \textit{probability } 89\%,\end{cases}$$

$$\textit{vs.}$$

$$\textit{Lottery 2B:} \qquad \begin{cases} £5\,000\,000, & \textit{probability } 10\%,\\ £0, & \textit{probability } 90\%.\end{cases}$$

Most people prefer Lottery 2B over Lottery 2A.

Can we explain this using utility theory?

Let's assume we have a utility function u that explains this choice. The choice L1A vs. L1B yields

$$u(£1\text{m}) > 0.89\, u(£1\text{m}) + 0.1\, u(£5\text{m}) + 0.01\, u(£0) \tag{1}$$

The choice L2B vs L2A yields

$$0.1\, u(£5\text{m}) + 0.9\, u(£0) > 0.11\, u(£1\text{m}) + 0.89\, u(£0),$$

which yields

$$0.01\, u(0) > 0.11\, u(£1\text{m}) - 0.1\, u(£5\text{m})$$

Inserting in (1) *yields*

$$u(£1\text{m}) > 0.89u(£1\text{m}) + 0.1u(£5\text{m}) + 0.11\, u(£1\text{m}) - 0.1\, u(£5\text{m}) = u(£1\text{m}),$$

which is a contradiction.

We see that the choice made by most people cannot be explained by classical utility theory. The reason is a different reception of small probabilities compared to larger ones (i.e. the difference between 0% *and* 1% *is a lot more significant different than the difference between* 89% *or* 90%*)*

Example 10 (Different perception of gains and losses)**.** *Again, we consider two different scenarios.*

Situation 1: *You receive £1 000 and have two options:*

$$\text{Lottery 1:} \quad \begin{cases} £500, & \text{probability } 100\%. \end{cases}$$

$$vs.$$

$$\text{Lottery 2:} \quad \begin{cases} £1,000, & \text{probability } 50\%, \\ £0, & \text{probability } 50\%. \end{cases}$$

Most people prefer L_1 over L_2 (risk-averse).

Situation 2: *You receive £2 000 and have two options:*

$$\text{Lottery 1:} \quad \begin{cases} -£500, & \text{probability } 100\%. \end{cases}$$

$$vs.$$

$$\text{Lottery 2:} \quad \begin{cases} -£1,000, & \text{probability } 50\%, \\ £0, & \text{probability } 50\%. \end{cases}$$

Most people prefer L_2 over L_1 (risk-seeking).

Although the final outcome is exactly the same in both cases:

$$\text{Lottery 1:} \quad \begin{cases} £1\,500, & \text{probability } 100\%. \end{cases}$$

$$vs.$$

$$\text{Lottery 2:} \quad \begin{cases} £1\,000, & \text{probability } 50\%, \\ £2\,000, & \text{probability } 50\%, \end{cases}$$

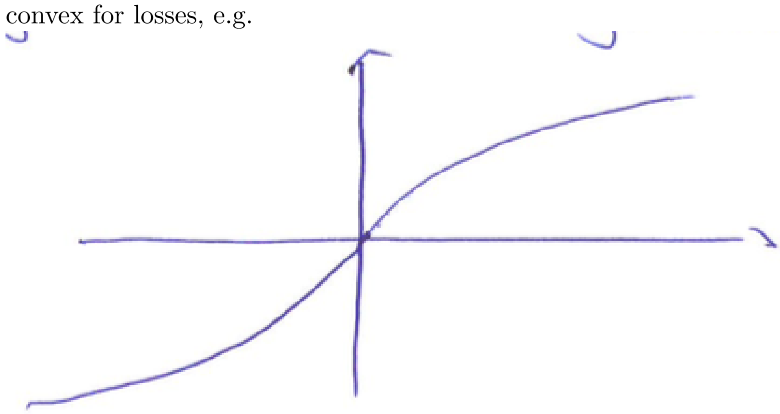
the choice of many people depend on a reference point (here £1 000 or £2 000).

Also we see that gains and losses are valued with a different risk attitude:

- *risk-seeking for losses, and*
- *risk-averse for profits.*

One way to incorporate these effects is the *cumulative prospect theory (CPT)* It has three main features:

- A reference point in wealth, defining profits and losses (framing)
- S-shaped utility functions, i.e. functions that are (locally) concave for profits and (locally) convex for losses, e.g.



- A non-linear transformation of the probability measure which increases the weight of small probabilities.

For more informations, see, e.g. Xun Yu Zhou - Mathematicalising Behavioural Finance, 2010.