Statistics of Financial Markets

Studium Generale

Prof. Dr. Natalie Packham

Berlin School of Economics and Law

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Winter term 2024/25

Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Contents

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
   1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Review of Statistical Foundations

- To build a solid foundation for the course, we review some concepts of probability theory.
- A particular focus lies on the normal distribution and the central limit theorem (including a quite straightforward proof).
- We review the very basics of statistics with a focus on hypothesis testing.
- Finally, we apply our (new) knowledge to testing trading strategies in finance.
- The material in this session cannot substitute a rigorous study of probability theory, but the *selection* in this session should turn out to be useful in many different contexts.

Contents

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
   1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Review of probability theory

- Probability theory deals with modelling and analysing random variables.
- A deeper understanding of statistical principles is impossible without the tools of probability theory.
- Therefore, for easy reference, we collect basic facts from probability theory.
- For a more detailed fresh up we recommend (Larsen *et al.*, 2012; Jacod and Protter, 2003; Shiryaev, 1996; Kallenberg, 2002) or any other textbook on probability theory.

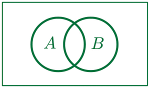
Probability space

- The mathematical model for a random experiment is a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$.
- The set Ω is called the **state space** and its elements $\omega \in \Omega$ are the possible outcomes of the experiment, also named **elementary events**. In the context of finance we consider ω to be a state of the economy or the “world”.
- The set $\mathcal{F}$ contains the possible events; these are combinations of outcomes and thus subsets of Ω.
- The **probability measure** $\mathbb{P}$ assigns a number $\mathbb{P}(A)$ measuring the “certainty” of each event $A \in \mathcal{F}$.
  A probability measure satisfies the following properties:

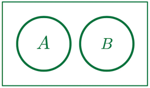
$$\mathbb{P}(A) \in [0,1],$$

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$$

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n), \quad \text{if } A_n \cap A_m = \emptyset,\ n \neq m.$$



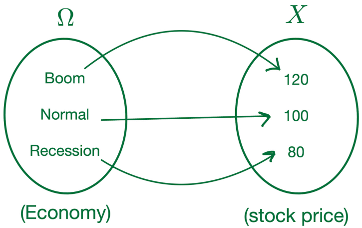
$A \cap B \neq \emptyset$



$A \cap B = \emptyset$

Random variables

- A **random variable** X is a mapping that assigns to every outcome ω a number $X(\omega)$.
- This means the true values x of the random variable X are uncertain – they depend on the outcome.
- For example, the stock price of a company at the end of today can be viewed as a random variable: it depends on the state of the world, and is unknown as of now.

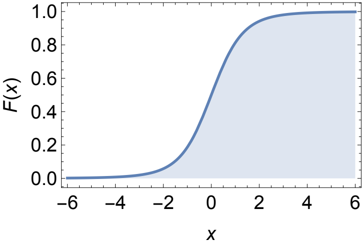


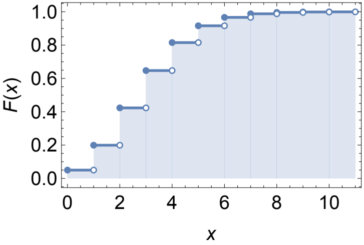
Distribution function

- The probabilities for the possible values x of the random variable X define the so-called **distribution** of X.
- This distribution is often given by the **(cumulative) distribution function** F_X of X:

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}. \tag{1}$$

- The distribution function is an *increasing*, *right-continuous* function taking values between $0 = F_X(-\infty)$ and $1 = F_X(\infty)$.





Discrete and continuous random variables

- We distinguish **discrete** and **continuous** random variables.
- A discrete random variables takes values in a countable set $x_1, x_2, x_3, \ldots$, and F_X is a discontinuous function, whereas a continuous random variable can take any real value (possibly restricted to some intervals), and F_X is a continuous function.
- For a discrete random variable X taking values $x_1 \leq x_2 \leq \ldots$ with probabilities $p_1, p_2, \ldots$, we have

$$\mathbb{P}(X \leq x_k) = F_X(x_k) = \sum_{i=1}^{k} p_i.$$

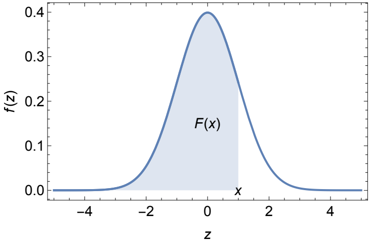
Continuous random variables: density function

- Often, when the function F_X is continuous, it possesses a first derivative $f_X(x) = F'_X(x)$, which is called the **probability density function** or just **density** of X.

- We have the relationship

$$\mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(z)\mathrm{d}z,$$

and intuitively $f_X(z)\mathrm{d}z$ stands for the probability of X taking values in an interval of "length" $\mathrm{d}z$ around the point z.



- Note that a probability distribution gives a complete description of the possible outcomes of a random variable: It specifies which outcomes are possible and how likely they are.

Normal distribution

- An important example for a distribution of a random variable X is given by the **normal distribution** with parameters μ and σ.[1]
- We indicate that X possesses a normal distribution by writing

$$X \sim \mathrm{N}(\mu,\sigma^2).$$

- The corresponding bell-shaped density is given by the formula

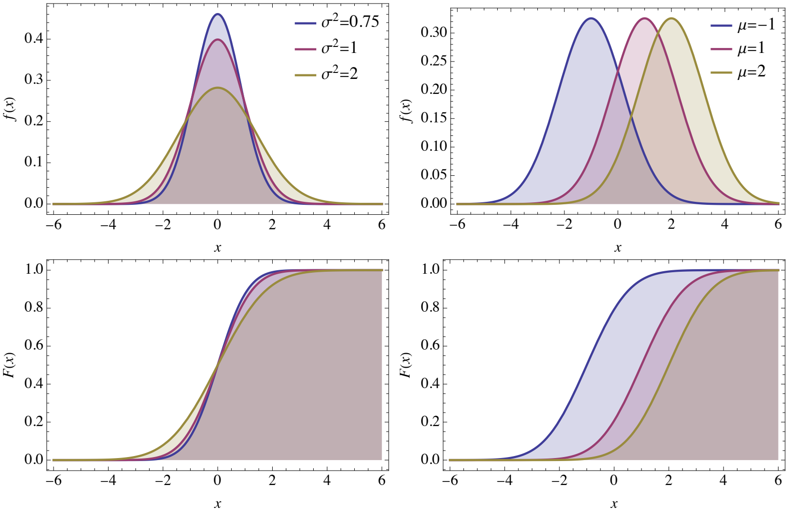
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \tag{2}$$

- In case of $\mu = 0$ and $\sigma = 1$ one obtains the so-called **standard normal distribution** $\mathrm{N}(0,1)$.
- The distribution function of the standard normal distribution is usually denoted by N and its density by n.

[1] The normal distribution is also often called Gaussian distribution. The former 10-DM note shows C. F. Gauss and the density of the normal distribution: `https://www.bundesbank.de/resource/blob/644664/26eb6287c09eb598cabd7ed0d924b845/mL/0010-1999-vs-data.jpg`

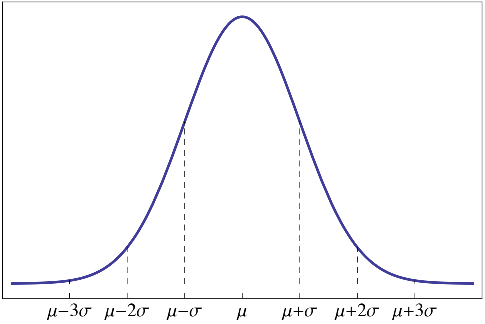
Normal distribution: Example

- Different normal distributions:



Probabilities within standard deviations

- As there is no closed formula for calculating normal probabilities, this is a good rule-of-thumb:



$$\mathbb{P}(|X-\mu| \leq \sigma) = 0.6827$$
$$\mathbb{P}(|X-\mu| \leq 2\sigma) = 0.9545$$
$$\mathbb{P}(|X-\mu| \leq 3\sigma) = 0.9973$$

Normal distribution

- Any normally distributed random variable X can be transformed into a standard normal variable and vice versa, via

$$X \sim \mathrm{N}(\mu,\sigma^2) \iff \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$$

$$X \sim \mathrm{N}(0,1) \iff \sigma X + \mu \sim \mathrm{N}(\mu,\sigma^2).$$

- As a consequence, for $X \sim \mathrm{N}(\mu,\sigma^2)$,

$$F_X(x) = F_{\mathrm{N}(\mu,\sigma^2)}(x) = \mathrm{N}\left(\frac{x-\mu}{\sigma}\right)$$

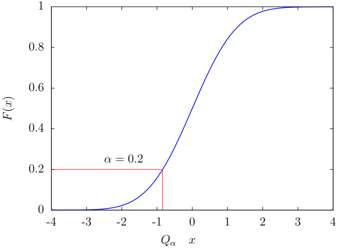
$$f_X(x) = f_{\mathrm{N}(\mu,\sigma^2)}(x) = \frac{1}{\sigma}\,\mathrm{n}\left(\frac{x-\mu}{\sigma}\right).$$

Quantile

- For $\alpha \in (0,1)$, the **α-quantile** Q_α of the **continuous** distribution function F_X or of the **continuous** random variable X is given by

$$Q_\alpha := F^{(-1)}(\alpha), \tag{3}$$

where $F^{(-1)}$ is the **inverse function** of F.[2]



- Intuitively, Q_α is the value such that X is less than or equal this number with probability α.
- In risk management, quantiles are important risk quantifiers – if the random variable X represents the loss of a position, then Q_p is the **value-at-risk**.

[2] The inverse $F_X^{(-1)}$ also plays a central role when simulating random outcomes for X according to a given distribution function F_X.

Expectation and variance

- Important quantities characterizing the distribution F_X of a random variable X are its **expectation** (probability weighted mean, average),

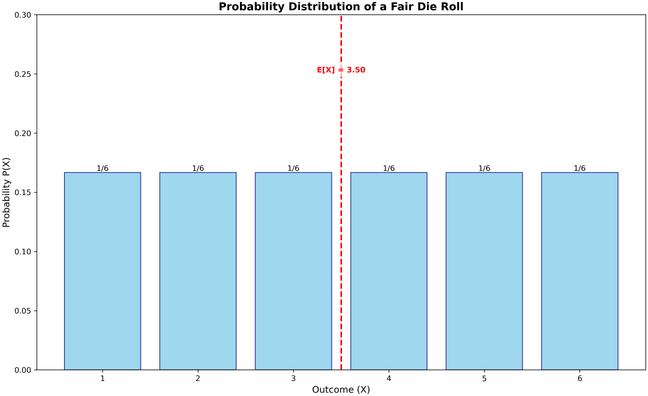
$$\mathbb{E}(X) = \mathbb{E}X = \begin{cases} \sum_{i=1}^{\infty} x_i\, p_i, & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x, & \text{if } X \text{ is continuous (and has a density),} \end{cases}$$

- and its **variance** (mean squared deviation from the expectation),

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f_X(x)\mathrm{d}x \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \end{aligned}$$

- The quantity $\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.
- The parameters μ and σ^2 of the normal distribution denote the expectation and the variance, respectively.

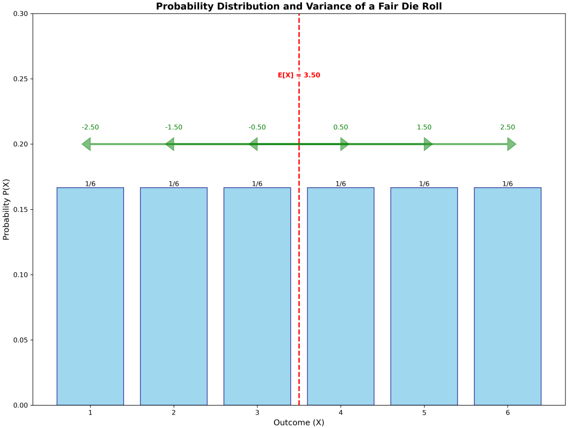
Expectation



x	P(X=x)	x * P(X=x)
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6
	Sum:	3.50

Source: Julius.AI / Claude 3.5 Sonnet

Variance



x	P(X=x)	x - E[X]	(x - E[X])²	(x - E[X])² * P(X=x)
1	1/6	-2.50	6.25	1.0417
2	1/6	-1.50	2.25	0.3750
3	1/6	-0.50	0.25	0.0417
4	1/6	0.50	0.25	0.0417
5	1/6	1.50	2.25	0.3750
6	1/6	2.50	6.25	1.0417
			Sum:	2.9167

Source: Julius.AI / Claude 3.5 Sonnet

Covariance and correlation

- For two random variables X_1,X_2 the **covariance** is defined by

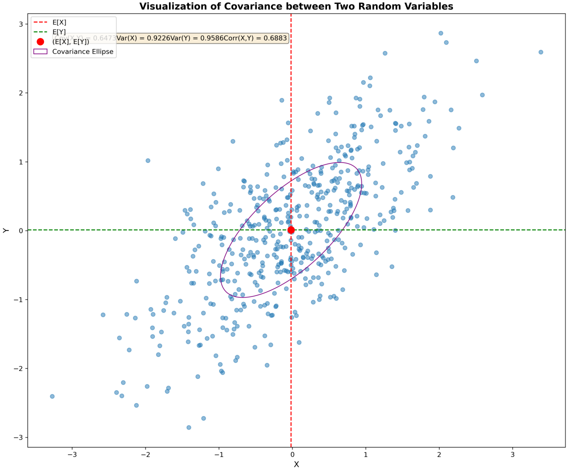
$$\mathrm{Cov}(X_1,X_2) = \mathbb{E}\left((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)\right) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2),$$

- and the **correlation** is

$$\rho(X_1,X_2) = \frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}.$$

- One can show that the correlation can only take values in the range $-1 \leq \rho(X_1,X_2) \leq 1$.
- Correlation measures the degree of **linear dependence** between X_1 and X_2.
- If $|\rho(X_1,X_2)| = 1$, then the random variables are linearly dependent, i.e., $X_2 = aX_1 + b$.
- If $\rho(X_1,X_2) = 0$, then the variables X_1,X_2 are called **uncorrelated**.

Covariance



Source: Julius.AI / Claude 3.5 Sonnet

Independence

- Two random variables X_1,X_2 are called **independent** if

$$F(x_1, x_2) := \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = F(x_1)\, F(x_2), \quad x_1{,}x_2 \in \mathbb{R}.$$

- Independence implies that the variables are uncorrelated, but the reverse is not true in general, i.e., independence is a stronger requirement compared to being uncorrelated.

- Given a random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ the pairwise covariances $\mathrm{Cov}(X_i{,}X_j)$ and correlations are aggregated in a matrix $\mathrm{Cov}(\mathbf{X})$ resp. $\rho(\mathbf{X})$:

$$\mathrm{Cov}(\mathbf{X}) = \left(\mathrm{Cov}(X_i{,}X_j)\right)_{i,j=1,\ldots,n}$$
$$\rho(\mathbf{X}) = \left(\rho(X_i{,}X_j)\right)_{i,j=1,\ldots,n}\,.$$

- These matrices are symmetric and positive semi-definite.

- The diagonal of the covariance matrix contains the variances $\mathrm{Cov}(X_i{,}X_i) = \mathrm{Var}(X_i)$.

Computation rules

- The following **computation rules for expectations and variances** apply:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \tag{4}$$

$$\mathbb{E}(a_1X_1 + \cdots + a_nX_n) = a_1\mathbb{E}(X_1) + \cdots + a_n\mathbb{E}(X_n) = \sum_{i=1}^{n} a_i\mathbb{E}(X_i) \tag{5}$$

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X) \tag{6}$$

$$\mathrm{Var}(a_1X_1 + \cdots + a_nX_n) = a_1^2\mathrm{Var}(X_1) + \cdots + a_n^2\mathrm{Var}(X_n) + \sum_{i,j=1,i\neq j}^{n} a_ia_j\mathrm{Cov}(X_i,X_j). \tag{7}$$

- Rule (7) can be written more compact in matrix notation as

$$\mathrm{Var}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\,\mathrm{Cov}(\mathbf{X})\,\mathbf{a} \tag{8}$$

with $\mathbf{a}^T = (a_1, \ldots, a_n)$ and with $\mathbf{a}^T\mathbf{X} = \sum_{i=1}^{n} a_iX_i$ the inner (scalar) product of vectors.

Computation rules

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- **Computation rules for the covariance** are

$$\mathrm{Cov}(a_1X_1 + b_1, a_2X_2, +b_2) = a_1a_2\mathrm{Cov}(X_1,X_2) \tag{9}$$

$$\mathrm{Cov}(a_1X_1 + a_2X_2, X_3) = a_1\mathrm{Cov}(X_1,X_3) + a_2\mathrm{Cov}(X_2,X_3) \tag{10}$$

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_iX_i, \sum_{j=1}^{m} b_jY_j\right) = \sum_{i=1,\ldots,n,j=1,\ldots,m} a_ib_j\mathrm{Cov}(X_i,Y_j). \tag{11}$$

Multivariate normal distribution

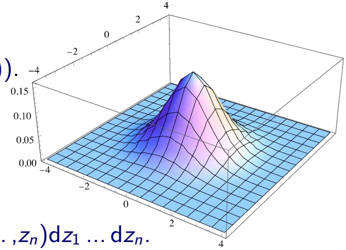
- A random vector $\mathbf{X} = (X_1, \dots, X_n)^T$ follows an **(n-dimensional) normal distribution** with expectation vector μ and covariance matrix $\mathrm{Cov}(\mathbf{X})$, if the joint probability density $f_{\mathbf{X}}(x_1, \dots, x_n)$ is given by[3]

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det \mathrm{Cov}(\mathbf{X})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathrm{Cov}(\mathbf{X})^{-1}(\mathbf{x} - \mu)\right), \tag{12}$$

$\mu = (\mu_1, \dots, \mu_n)^T, \mathbf{x} = (x_1, \dots, x_n)^T.$

- We abbreviate this by writing $\mathbf{X} \sim \mathrm{N}(\mu, \mathrm{Cov}(\mathbf{X}))$.

- The joint density determines the distribution of the random vector by



$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(z_1, \dots, z_n) \mathrm{d}z_1 \dots \mathrm{d}z_n.$$

[3] To simplify we assume a non-singular n-dimensional normal distribution, i.e., $\det \mathrm{Cov}(\mathbf{X}) \neq 0$.

Multivariate normal distribution

Theorem

Let $\mathbf{X}$ be an n-dimensional normally distributed random vector. Then the components of $\mathbf{X}$ are **uncorrelated** *if and only if they are* **independent**.

Theorem

The random vector $\mathbf{X}$ is n-dimensional normally distributed if and only if, for every vector $(a_1, \dots, a_n) \in \mathbb{R}^n$, the random variable

$$Y = a_1 X_1 + \cdots + a_n X_n$$

is normally distributed.

- Let $\mathbf{X} \sim \mathrm{N}(\mu, \mathrm{Cov}(\mathbf{X}))$, then $Y \sim \mathrm{N}(\nu, \sigma^2)$, with

$$\nu = \sum_{i=1}^{n} a_i \mu_i \qquad (13)$$

$$\sigma^2 = \mathbf{a}^T \,\mathrm{Cov}(\mathbf{X})\, \mathbf{a}. \qquad (14)$$

Multivariate normal distribution

- The proofs of both theorems are quite straightfoward when using the so-called **characteristic function** of a probability distribution,

$$\varphi_{\mathbf{X}}(t) = \mathbb{E}\left[e^{it^T\mathbf{X}}\right],$$

  where i is the imaginary number satisfying $i^2 = -1$ and t is an n-dimensional vector.

- A distribution is uniquely characterised by its characteristic function.
- The characteristic function of an n-dimensional normally distributed random vector X is

$$\varphi_{\mathbf{X}}(t) = e^{it^T\mu - \frac{1}{2}t^T\mathrm{Cov}(\mathbf{X})t}.$$

- In particular, in the univariate case, $X \sim \mathrm{N}(\mu,\sigma^2)$, we have

$$\varphi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}.$$

Contents

1. Review of Statistical Foundations
    1.1 Review of Probability Theory
    1.2 Central Limit Theorem
    1.3 Statistics
    1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Central Limit Theorem (CLT)

- The **Central Limit Theorem** is one – if not *the* – most important theorem in probability theory.
- Very loosely speaking, it states that **sums** of independent, identically distributed random variables are **normally distributed**!
- On the one hand, this explains why the normal distribution serves as a good approximation for many variables encountered in practice: whenever a variable can be decomposed into a sum of individual variables, the normal distribution may be an appropriate model.
- On the other hand, the normal distribution is the gateway to **inferential statistics**, as it simplifies many distributional assumptions in e.g. statistical testing.

Central Limit Theorem (CLT)

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Theorem (Central Limit Theorem)

Let $X_1, X_2, \ldots$ *be independent, identically distributed random variables with*

$$\mathbb{E}(X_i) = \mu \quad \textit{and} \quad Var(X_i) = \sigma^2 < \infty.$$

Then, the distribution function $F_n(x) = \mathbb{P}(S_n \leq x)$ *of the standardised sum*

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

converges at each point $x \in \mathbb{R}$ *to the distribution function* $\mathrm{N}(x)$ *of the standard normal distribution:*

$$F_n(x) \to \mathrm{N}(x) \quad \textit{for } n \to \infty.$$

- Notation: $S_n \overset{a}{\sim} \mathrm{N}(0,1)$.

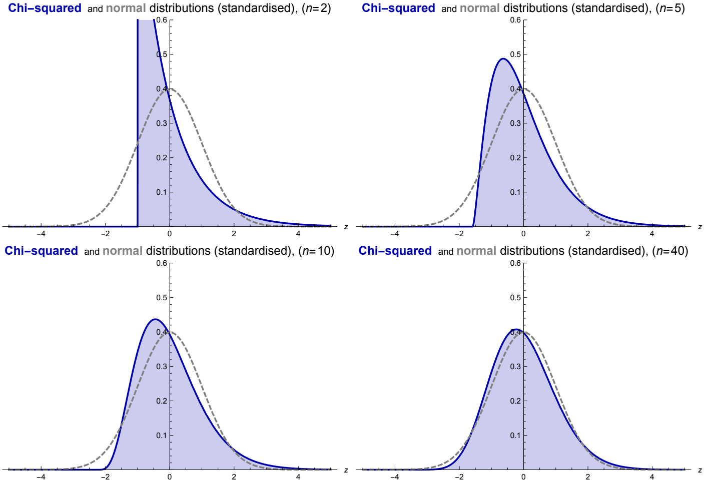
Central Limit Theorem (CLT)

- As a consequence: for sufficiently large n, the expectation of a random sample is approximately normally distributed:

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right).}$$

CLT: Example

- Let $X_1, \ldots, X_n$ be iid N(0,1)-distributed random variables. The sum $X_1^2 + \cdots + X_n^2$ follows a **Chi-squared distribution** with parameter n.



CLT: Sketch of proof

- Using characteristic functions, the proof is surprisingly easy!
- Let $\mathbb{E}X_1 = \mu$ and $\text{Var}X_1 = \sigma^2$ and $\varphi(t) = \mathbb{E}\left[e^{it(X_1-\mu)}\right]$.
- Then, using $S_n = X_1 + \cdots + X_n$, it can be shown that

$$\varphi_n(t) = \mathbb{E}\left[e^{it\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbf{Var}(S_n)}}}\right] = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

- Using a Taylor expansion[4] of the characteristic function, using $\mathbb{E}|X|^2 < \infty$,

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad t \to 0,$$

gives

$$\varphi_n(t) = \left[1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\left(\frac{1}{n}\right)\right]^n \to e^{-t^2/2},$$

as $n \to \infty$ for fixed t.

[4] If $\mathbb{E}|X|^m < \infty$, then $\varphi(t) = \sum_{r=0}^m \frac{(it)^r}{r!}\mathbb{E}[X^r] + o(t^m), \quad t \to 0$.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
   - 1.1 Review of Probability Theory
   - 1.2 Central Limit Theorem
   - 1.3 Statistics
     - Basics of hypothesis testing
     - Gauss test and t-test
   - 1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Descriptive statistics

- In finance, probability distributions and parameters of distributions are typically estimated from historical data using statistical techniques.
- Consider a random variable X with observed (realised) values $x_1, \dots, x_n$.
- In the context of finance, the variable X could be, for example, a stock price return, a change in an interest rate, a return of an exchange rate.
- The values $x_1, \dots, x_n$ are then historical observations.

- The standard estimation rule for the **expectation** $\mu = \mathbb{E}(X)$ of X is

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{15}$$

- The **variance** of X, $\sigma^2 = \mathrm{Var}(X)$, is estimated from the data by

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{16}$$

- For two random variables X, Y with observed pairs of values $(x_1,y_1), \ldots, (x_n,y_n)$ their **covariance** $\mathrm{Cov}(X,Y)$ is estimated according to

$$\hat{c} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \tag{17}$$

Limitations

- Under the assumption that the observed values are realizations of independent trials of one and the same random variable (distribution), the estimates above are **unbiased** and **consistent**, i.e., for an increasing number of observations the estimate converges to the true value.
- In practical applications it is questionable whether these assumptions are really satisfied.
- For example, subsequent daily changes in interest rates or returns of securities may fail to be independent.
- Furthermore, it is not obvious whether the historical data originate from one and the same distribution since there can be structural changes.
- A related problem is that it is not clear whether information gained from past data is valid for the distribution of future variables.

Empirical quantile

- To estimate the quantile Q_p of a random variable X from data $x_1, \ldots, x_n$, the values are sorted according to their size,

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)},$$

and the "$n \cdot p$" greatest value is a reasonable estimate for the quantile:

$$\hat{Q}_p = x_{(\lfloor n \cdot p \rfloor + 1)}, \tag{18}$$

with $\lfloor n \cdot p \rfloor$ denoting the integer part of $n \cdot p$.

- For example, for $n = 250$ observations the $q = 1\%$ quantile is estimated by the $\lfloor 250 \cdot 0.01 \rfloor + 1 = 3$rd smallest value $x_{(3)}$.

Principal idea of inferential statistics

- The principal goal of **inferential statistics** is to draw conclusions about the distribution or distribution measures of a (large) population based on a (smaller) observed sample.
- Examples: Election poll, excess returns of a trading strategy, regression coefficients.
- **Hypothesis testing**:
  - **Informed** decision-making; draw conclusions about different hypotheses.
  - Is some observation of a sample statistic an "effect" or purely random?
  - Example: Rolling a die 100 times yields a sample mean of 3.61. Is the deviation from 3.5 just chance or is the die unfair?

Random sample

Definition

Let $X_1, X_2, \dots, X_n$ be a set of n independent identically distributed random variables from a sample space (population) Ω.

- Then random vector $(X_1, \dots, X_n)$ is a **random sample** of size n.
- The observed outcome $(x_1, \dots, x_n)$ of $(X_1, \dots, X_n)$ is called the **realisation** of the random sample.
- The set $\mathcal{X}$ of all possible realisations is called the **sampling space**.

- Often $\Omega \subseteq \mathbb{R}$, in which case $\mathcal{X} \subseteq \mathbb{R}^n$.

Estimators and estimates

Definition

Let $(X_1, \dots, X_n)$ be a random sample.

- An **estimator** or **statistic** is a function

$$T = g(X_1, \dots, X_n)$$

of the random sample.

- The resulting numerical value $g(x_1, \dots, x_n)$ is called the **estimate**.

- Note that T – as a function of random variables – is itself a random variable.
- Examples of estimators are estimations of distribution parameters or measures as well as test statistics in decision problems.

Contents

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
      Basics of hypothesis testing
      Gauss test and t-test
   1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Hypothesis testing

- Typical problems in **decision-making**:
  - Is the defendant mentally fit or not?
  - Should the new medication be approved or not?
  - Is the normal distribution an acceptable model for daily stock returns?
- **Statistical testing**:
  - Differentiate possible conclusions of an experiment (hypotheses);
  - Make decision using the tools of probability theory.

Hypothesis testing

- In case of two mutually exclusive possibilities:
  - **null hypothesis** H_0
    - will be accepted as long as there is not sufficient evidence against it ("presumption of innocence")
  - **alternative hypothesis** H_1
    - will be accepted if there is "sufficient" evidence against the null hypothesis
- Possible conclusions and their consequences:

	fail to reject	reject
H_0 is true	Correct decision	Incorrect decision (Type I error)
H_0 is false	Incorrect decision (Type II error)	Correct decision

- Goal: Decision *strategy* such that type I and type II errors small

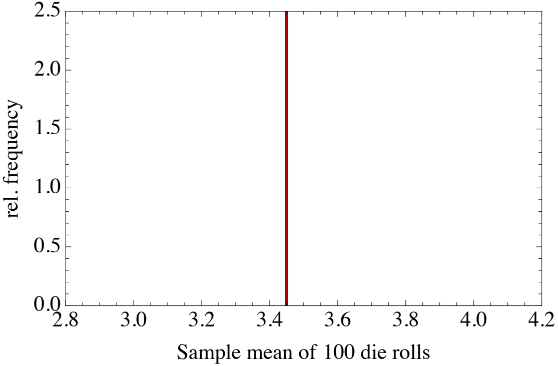
Basics of hypothesis testing

- Example: Testing if a die is fair; use expectation as proxy
- Hypotheses (mutually exclusive)[5]:
  - **Null hypothesis**: Die is fair; formally: $H_0 : \mu = 3.5$
  - **Alternative hypothesis**: Die is not fair; formally: $H_1 : \mu \neq 3.5$.
- Experiment: Roll die 100 times and calculate sample mean.
- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Sampling Space: Ω^{100}
- Random sample: $X_1, \ldots, X_{100}$
- Statistic: $T = g(X_1, \ldots, X_{100}) = \dfrac{1}{100} \displaystyle\sum_{i=1}^{100} X_i$.
- Realisation: $(x_1, \ldots, x_{100})$, $g(x_1, \ldots, x_{100}) = \dfrac{1}{100} \displaystyle\sum_{i=1}^{100} x_i$.

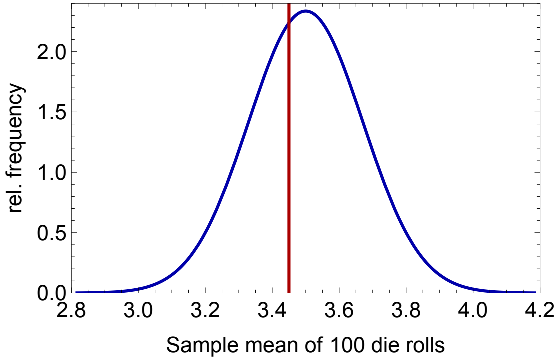
[5] Of course, there are ways in which an unfair die can produce an expectation of 3.5, so further testing would be required.

Basics of hypothesis testing

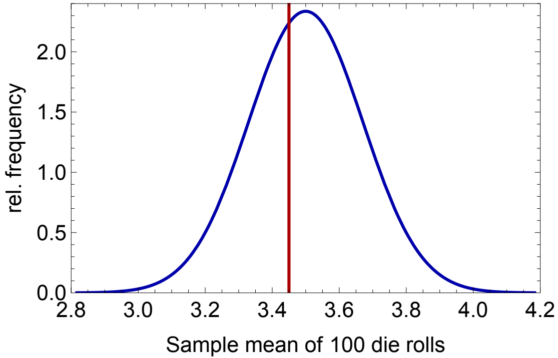
- Observed sample mean: 3.45



- **Central Limit Theorem**: $\frac{1}{100}\sum_{i=1}^{100} X_i \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{100}\right)$, where $\mu = 7/2$ and $\sigma^2 = 35/12$ are the expectation and variance of one die roll.



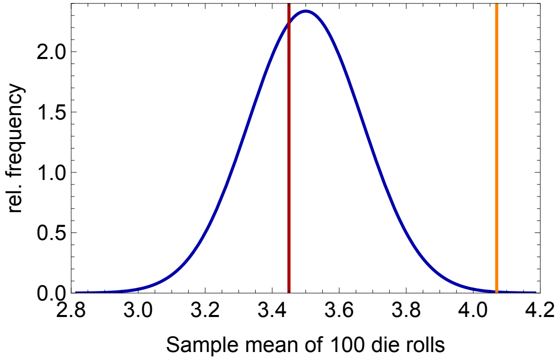
- **Central Limit Theorem**: $\frac{1}{100}\sum_{i=1}^{100} X_i \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{100}\right)$, where $\mu = 7/2$ and $\sigma^2 = 35/12$ are the expectation and variance of one die roll.



- What do you think: Is the observed sample mean of 3.45 typical for a fair die or not?

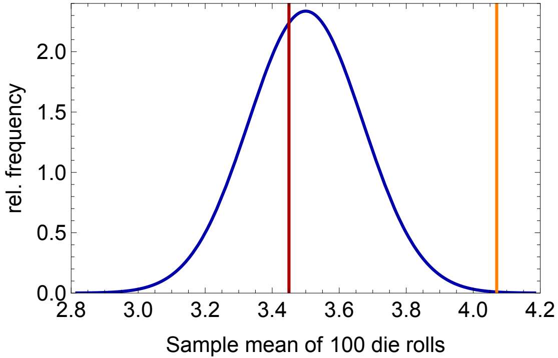
Basics of hypothesis testing

- What do you conclude if the observed sample mean is 3.95?



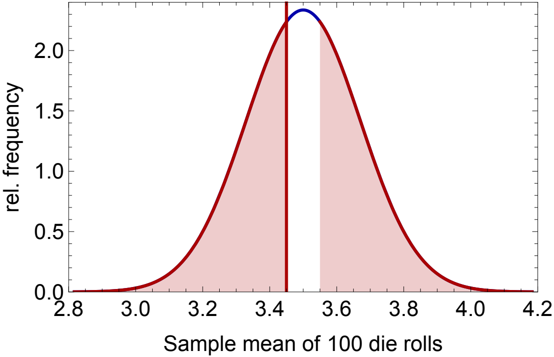
Basics of hypothesis testing

- What do you conclude if the observed sample mean is 3.95?



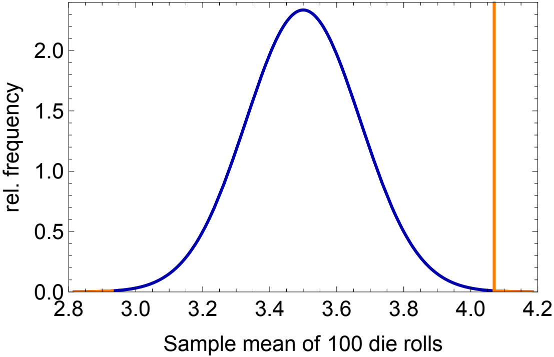
- Note it is impossible to draw a conclusion with certainty!
- How do you express the uncertainty related to your conclusion?

- If H_0 is true: Probability of observing 3.45 or something more extreme away from 3.5 is 77%.



- This probability is called p-**value**. Rejecting H_0 does not seem justified.

- If H_0 is true: Probability of observing 3.95 or something more extreme away from 3.5 is 0.84%.



- p-value small enough to reject H_0 at level of statistical significance of 1%.

Contents

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
      Basics of hypothesis testing
      Gauss test and t-test
   1.4 Testing trading strategies
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Steps in hypothesis testing

1. Quantify the problem and formulate model (e.g. Gauss-test, t-test)
2. Specify the **null hypothesis** H_0 and the **alternative hypothesis** H_1
3. Specify the **significance level** α
   - This is the (risk) tolerance of falsely rejecting H_0.
4. Calculate p-value.
5. Reject H_0 if p-value is smaller than α.
6. Do not reject H_0 if p-value is greater than α.

Gauss test for μ

Gauss-test

Let $X_1, \dots, X_n$ be iid (independent, identically distributed) random variables with

- $X_i \sim \mathrm{N}(\mu,\sigma^2)$,
- OR: X_i with arbitrary distribution, $\mathbb{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, and n sufficiently large (rule of thumb: $n \geq 30$).

Assume that σ^2 is known. Consider the following test problems:

$$\begin{array}{llll} (a) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu \neq \mu_0 \\ (b) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu < \mu_0 \\ (c) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu > \mu_0. \end{array}$$

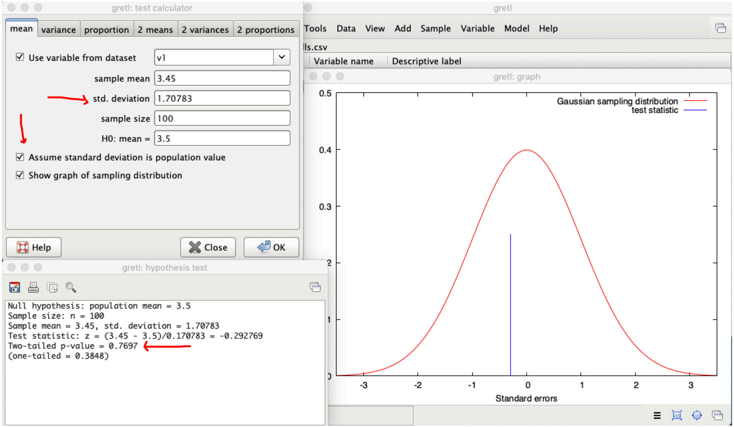
For $\mu = \mu_0$ the test statistic is $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{\text{approx.}}{\sim} \mathrm{N}(0,1)$.

Based on *observed sample statistic* z, H_0 is rejected if p-value is smaller than α:

$$\begin{array}{ll} (a) & \text{if } \mathbb{P}(|Z| > |z| \,|\, H_0) < \alpha \\ (b) & \text{if } \mathbb{P}(Z < z | H_0) < \alpha \\ (c) & \text{if } \mathbb{P}(Z > z | H_0) < \alpha. \end{array}$$

Example in `gretl`

- `Tools` $\rightarrow$ `Test statistic calculator`



- Standard deviation $\sigma = 1.70783$ calculated from fair die.
- Standard deviation as population value implies Gauss test.

t-test for μ

t-test

Assumption; $X_1, \dots, X_n$ are iid with

- $X \sim \mathrm{N}(\mu,\sigma^2)$,
- **OR** follow an arbitrary distribution and $n > 30$.

μ, σ^2 are both unknown. Hypotheses:

$$\begin{array}{llll} (a) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu \neq \mu_0 \\ (b) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu < \mu_0 \\ (c) & H_0 : \mu = \mu_0 & \text{against} & H_1 : \mu > \mu_0. \end{array}$$

Test statistic: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$.

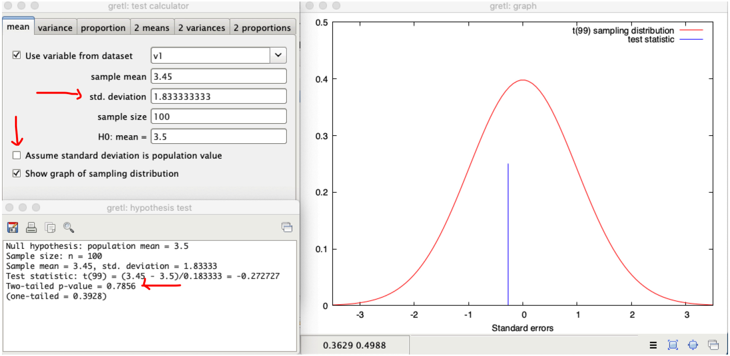
Distribution under $\mu = \mu_0$: Student t-distribution with parameter $n - 1$.

Based on sample statistic t, reject H_0 if p-value smaller than α:

$$\begin{array}{ll} (a) & \mathbb{P}(|T| > |t||H_0) \leq \alpha \\ (b) & \mathbb{P}(T < t|H_0) \leq \alpha \\ (c) & \mathbb{P}(T > t|H_0) \leq \alpha \end{array}$$

Example in gretl

- Tools $\rightarrow$ Test statistic calculator



- Standard deviation S estimated from sample.
- "Standard deviation as population value" unticked implies t-test.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
   1.4 Testing trading strategies
      Testing against a benchmark
      Example: Momentum buying
      Example: Moving average crossover
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Motivation for testing trading strategies

- If a trading strategy was successful in the past: Can its success be attributed to the strategy or was it pure luck?
- Different methods of statistical testing can provide valuable insights.
- Testing methods:
  - Hypothesis testing against benchmark
  - Monte Carlo simulation
  - Bootstrapping (advanced)
- Advanced methods can test for bias from data mining.
- Literature: (Aronson, 2006; West, 1996; White, 2000)

Contents

1. Review of Statistical Foundations
   1.1 Review of Probability Theory
   1.2 Central Limit Theorem
   1.3 Statistics
   1.4 Testing trading strategies
      Testing against a benchmark
      Example: Momentum buying
      Example: Moving average crossover
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Principal ideas

- Trading strategy:
  - Any **rule** or **algorithm** that generates a financial position.
  - Example: trading signal, e.g. from $\{0,1\}$ or $\{-1,1\}$ (so-called reversal rules), for every point in time.
- If trading strategy is built from historical data, then split data set into **in-sample** period for building strategy, and conduct test on **out-of-sample** period.
- Test hypothesis that a trading strategy / rule consistently outperforms a *benchmark*, i.e., has **predictive power**.
- Examples of benchmarks:
  - buy-and-hold in stock index,
  - riskless interest rate,
  - strategy with no predictive power (randomly generated trading signal).

Testing a trading strategy

- Observed returns of underlying security or portfolio: $(X_t)_{t \in \mathbb{Z}}$
- Observed returns of benchmark: $(Y_t)_{t \in \mathbb{Z}}$ (often $X = Y$)
- Strong assumption: independent, identically distributed
- Weaker assumptions possible, e.g. $(X_t)_{t \in \mathbb{Z}}$ is a stationary strong α-mixing sequence, where dependence “fades out” over time (Rosenblatt, 1956).

Testing a trading strategy

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Trading signals are generated for times $R, \dots, T$ where $T = R + n - 1$.
- Signals are $\hat{p}_{R+1}, \dots, \hat{p}_{T+1}$, where $\hat{p}_{R+k+1}$ may be based on $(X_1, \dots, X_{R+k})$.
- Null hypothesis: average excess return f is zero, i.e.,

$$\mathbb{E}[f] = 0, \text{ with } f = p\,X - Y$$

- Alternative hypothesis: strategy outperforms benchmark, i.e. $\mathbb{E}[f] > 0$.
- Build test statistic from sample mean:

$$\bar{f} = \frac{1}{n} \sum_{t=R}^{T} \hat{f}_{t+1},$$

  with $\hat{f}_{t+1} = \hat{p}_{t+1}\,X_{t+1} - Y_{t+1}$.

- Central Limit Theorem implies that t-test can be used (West, 1996; White, 2000).

Building the benchmark: Random permuations of signals

- Alternative method is to build benchmark as

$$Y_{t+1} = p_{R+\pi_t} X_{t+1},$$

  where π is a random permuation of $\{1, \dots, n\}$, with all permuations having equal probability.
- $Y_R, \dots, Y_T$ has same position bias, but no predictive power.
- Monte Carlo simulation of benchmark generates distribution under H_0.
- Calculate p-value as relative frequency of outcomes greater than observed test statistic $\overline{f}$.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

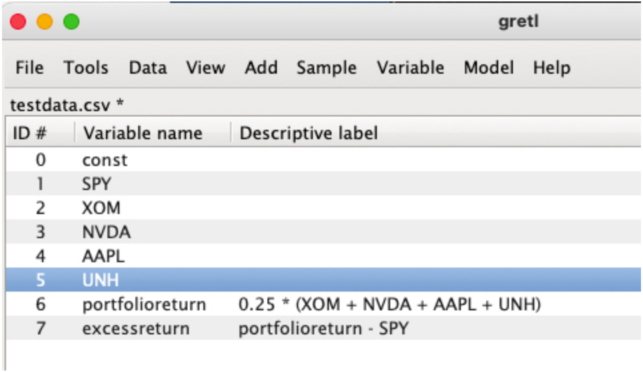
1. Review of Statistical Foundations
    1.1 Review of Probability Theory
    1.2 Central Limit Theorem
    1.3 Statistics
    1.4 Testing trading strategies
        Testing against a benchmark
        Example: Momentum buying
        Example: Moving average crossover
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Example: Momentum buying

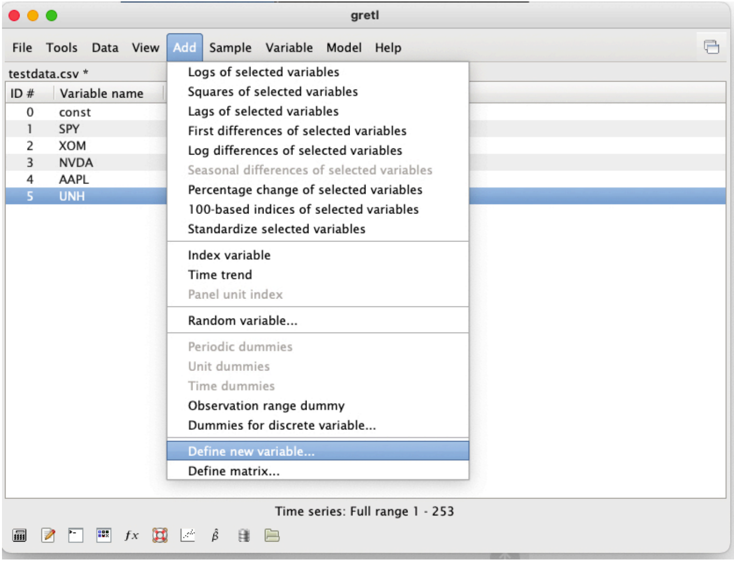
- File `sp500data.csv` contains daily data of 20 S&P 500 stocks with highest market capitalisation.
- Time period: 1 May 2021 – 30 Apr 2024
- In-sample period: 1 May 2021 – 30 Apr 2023
- Out-of-sample period: 1 May 2023 – 30 Apr 2024
- Careful: Forward-looking bias if market cap is determined in 2024 (here it is determined in 2021).
- Invest in four best performing stocks and test if outperform out-of-sample.
- Equal weights in out-of-sample period.
- Stocks: XOM, NVDA, AAPL, UNH
- File `testdata.csv` contains log returns of out-of-sample period.

Example: t-test

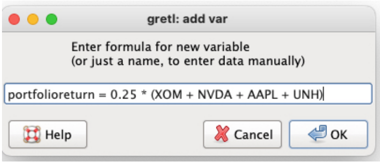
- Create portfolio and calculate excess return.

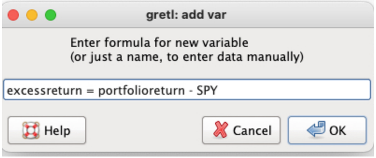


Example: t-test



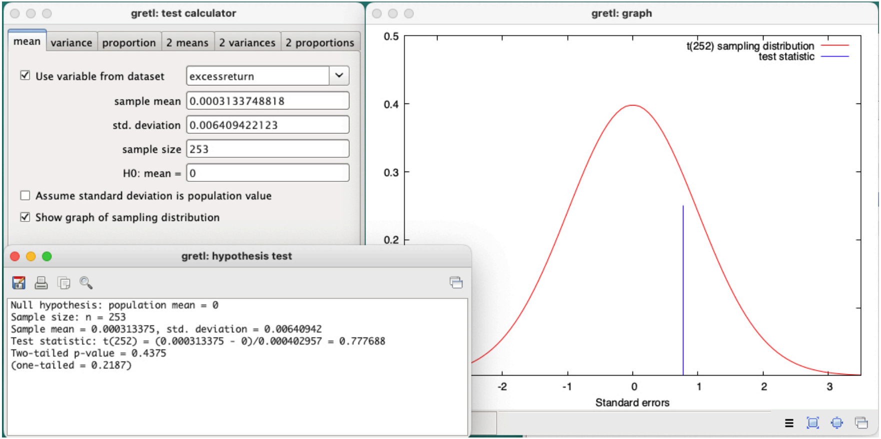
Example: t-test





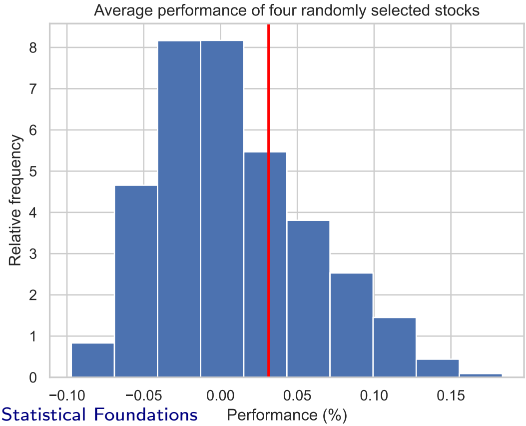
Example: t-test

- t-test on excess return yields p-value of 0.2187.
- Note that one-tailed t-test is appropriate as alternative hypothesis points only in one direction.



Example: Random stock selection

- Histogram of average out-of-sample return of portfolio with four randomly selected stocks.
- Red line: Portfolio of best-performing stocks
- *p*-value: 29.36%



Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
    1.1 Review of Probability Theory
    1.2 Central Limit Theorem
    1.3 Statistics
    1.4 Testing trading strategies
        Testing against a benchmark
        Example: Momentum buying
        Example: Moving average crossover
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Example: Moving average crossover

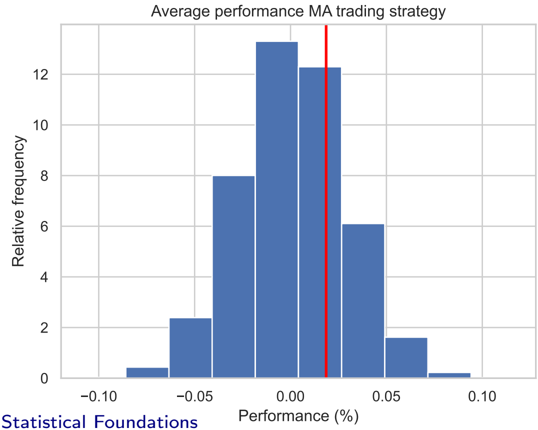
- We test if a trading strategy based on the 30-day moving average (MA) can generate statistically significant excess returns.
- The file `GDAXI.csv` contains two years of daily prices and returns of the German DAX stock index (Data source: Yahoo Finance).
- The strategy is long the DAX if the stock price is below the 30-day MA, and neutral else.
- We test if a statistically significant excess return against a buy-and-hold position in DAX is achieved.
- We also run a permutation test on the number of short/neutral positions to check the strategy against a random strategy.

Example: t-test

- See `GDAX_solution.xlsx` or `dax.ipynb`.
- 215 long positions, 268 neutral positions
- Statistics of excess returns:
  - Average sample mean: 0.0186%
  - Standard deviation: 0.7007%
  - Test statistic: 0.5833
  - p-value (t-test): 28%

Example: permutation test

- Histogram of average excess return of strategies with random permutations of trading signal.
- Red line: MA crossover strategy
- *p*-value: 26.48%



Long/short-bias

- If strategy is subject to long/short bias, then benchmark may need to be corrected for this.
- A strategy with no predictive power that has a **long bias** in an **upward trending market** will easily have a positive expected return.
- Likewise for short bias and downward trending market.
- Bias / trend problem can be eliminated by **de-trending historical data before testing**.
- De-trending:
  - Determine average log-return over historical test period.
  - Subtract this average from daily return.
- Note: De-trended data is used only for testing, not signal generation.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
    2.1 Simple regression and OLS estimation
    2.2 The econometrics of the simple regression model
    2.3 Multiple regression
    2.4 Regression with dummy variables
    2.5 Freeing up the classical assumptions
    2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

   2.1 Simple regression and OLS estimation

      The model

      Measuring the fit of a regression model: R^2

      Nonlinear relationships

   2.2 The econometrics of the simple regression model

   2.3 Multiple regression

   2.4 Regression with dummy variables

   2.5 Freeing up the classical assumptions

   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Introduction

- **(Linear) regression** is the starting point and workhorse for many empirical research questions, especially when observational data is involved.
- The material covered in this section will be good for **prediction** or **forecasting**, but only to a limited extent for establishing **causality**.[6]
- However, the material in this section should arm you to delve deeper into more advanced regression techniques tackling causality, e.g. instrumental variables, diff-in-diff, panel data, and other research designs, see e.g. (Angrist and Pischke, 2008; Imbens and Rubin, 2015; Cunningham, 2021), https://mixtape.scunning.com.
- A solid understanding of regression builds the foundation for data-driven empirical analysis and is therefore a good starting point for more sophisticated methods and algorithms, such as machine learning.

[6] When the rooster crows, the sun rises. But we know that the rooster doesn't cause the sun to rise – had the rooster not crowed, the sun would still have risen. Establishing causal relationships and distinguishing causal relationships from purely correlational relationship is tricky.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
      The model
      Measuring the fit of a regression model: R^2
      Nonlinear relationships
   2.2 The econometrics of the simple regression model
   2.3 Multiple regression
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
   2.6 Time series models

3. Derivatives pricing and hedging

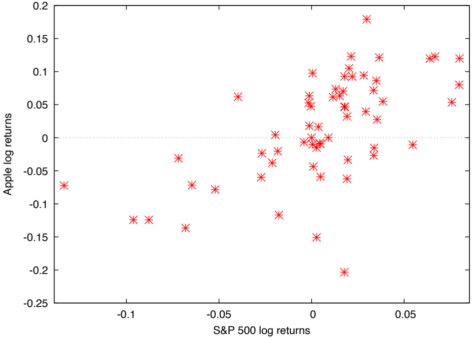
4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Linear regression

- **Correlation** measures the **strength** of the linear relationship between two variables X and Y.
- **Regression** measures the **form** of the linear relationship between X and Y.
- The simple regression model can be extended to non-linear relationships.

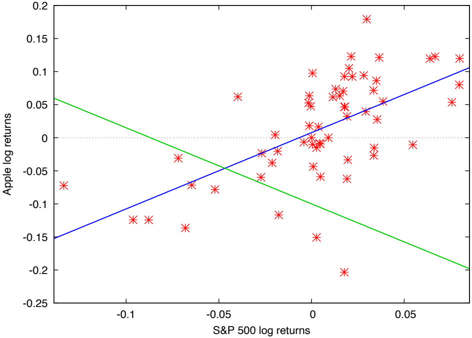
- Example: Apple stock returns against S&P 500 stock returns (monthly, June 2015 – May 2020)[7]
- Correlation: 0.6
- Regression fits line through the scatter plot that best captures the relationship between flat size and monthly rent.



[7] We will see later why it is important to consider returns, not stock prices.

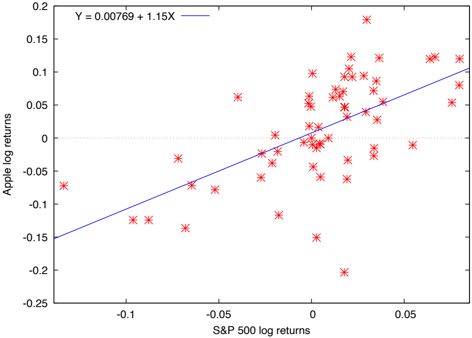
Simple Regression Model: Example

- Example: Apple stock returns against S&P 500 stock returns (monthly, June 2015 – May 2020)[7]
- Correlation: 0.6
- Regression fits line through the scatter plot that best captures the relationship between flat size and monthly rent.



[7] We will see later why it is important to consider returns, not stock prices.

- Example: Apple stock returns against S&P 500 stock returns (monthly, June 2015 – May 2020)[7]
- Correlation: 0.6
- Regression fits line through the scatter plot that best captures the relationship between flat size and monthly rent.



[7] We will see later why it is important to consider returns, not stock prices.

Simple regression: some theory

- Question: What **exactly** do we mean by **"best fitting line"**?
- Assume a linear relationship between $X =$ ld_AppleClose and $Y =$ ld_SPClose:
$$Y = \alpha + \beta X,$$
where α is the intercept of the line and β its slope.
- In practice, even if the straight line relationship were true, deviations from the line due to **measurement error** would typically occur.
- Often, the true relationship is **more complicated**, so the straight line is just an approximation (e.g. important variables affecting Y may be omitted).
- Both factors are captured by adding an error term, ε, giving:

Simple regression model

$$Y = \alpha + \beta X + \varepsilon.$$

Simple regression: some theory

- Given: data sample $(X_1, Y_1), \dots, (X_n, Y_n)$.
- Unknown: α, β and ε.
- Regression analysis uses data to make an estimate of what α and β are.
- Notation: $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of α and β.

True regression line

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, \dots, n$$

with **error** for i-th observation: $\varepsilon_i = Y_i - \alpha - \beta X_i$.

Fitted (or estimated) regression line

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i, \quad i = 1, \dots, n$$

Residuals: $\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$.

Derivation of OLS estimator

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- The method of **ordinary least squares (OLS)** determines α and β by minimising the sum of quadratic error terms:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \varepsilon_i^2 = \min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2.$$

- Setting the derivatives with respect to α and β to zero gives

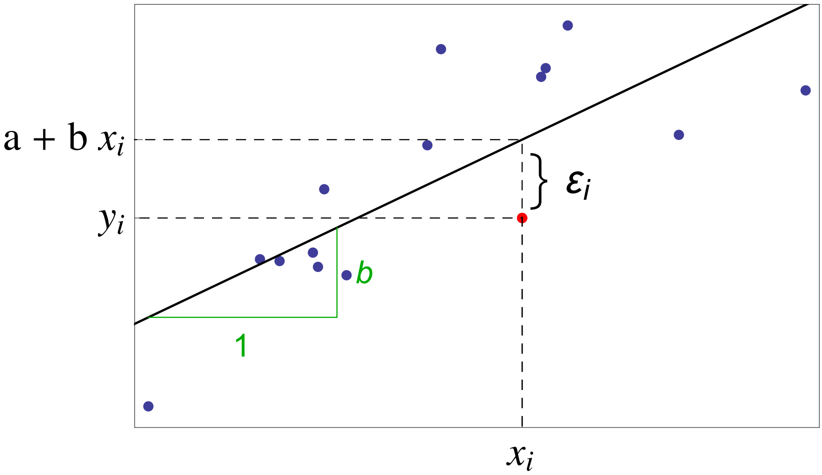
$$\frac{\partial \sum \varepsilon_i^2}{\partial \alpha} = 2 \sum (-\alpha - \beta x_i + y_i) = 0$$
$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta} = 2 \sum (\alpha x_i + \beta x_i^2 - y_i x_i) = 0$$

- Re-arranging gives a system of two linear equations with two unknowns:

$$\alpha\, n + \beta \sum x_i = \sum y_i$$
$$\alpha \sum x_i + \beta \sum x_i^2 = \sum y_i x_i$$

- The OLS estimator for α and β is $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$, $\hat{\beta} = \hat{\rho}_{XY} \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$, where $\overline{x}$, $\overline{y}$ are the sample means of $(x_1, \dots x_n)$ and $(y_1, \dots, y_n)$, respectively.

Illustration



Jargon of regression

- Y: dependent variable
- X: explanatory (or independent) variable
- α and β are coefficients
- $\hat{\alpha}$ and $\hat{\beta}$ are OLS estimates of coefficients
- "Run a regression of Y on X"

Interpreting OLS estimates

- Fitted regression: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$.
- Interpretation of $\hat{\alpha}$: estimated value of Y if $X = 0$.
- $\hat{\beta}$: slope of the best fitting straight line through the scatter plot.
- The following are a few different ways of interpreting $\hat{\beta}$.
  - $\hat{\beta}$ can be interpreted as a **derivative**:

$$\hat{\beta} = \frac{\mathrm{d}\hat{Y}_i}{\mathrm{d}X_i}.$$

  - $\hat{\beta}$ is the **marginal effect** of X on Y.
  - $\hat{\beta}$ is measure of how much Y tends to **change** when X is changed by one unit, where "unit" refers to what the variables are measured in (e.g. \$, €, %, metres, etc.).

Example: Apple vs. S&P 500

- Output from gretl:

Model 1: OLS, using observations 2015:06-2020:05 (T = 60)
Dependent variable: ld_AppleClose

	Coefficient	Std. Error	t-ratio	p-value
const	0.00768812	0.00826986	0.9297	0.3564
ld_SPClose	1.15288	0.200536	5.749	0.0000

Mean dependent var	0.013145	S.D. dependent var	0.079052
Sum squared resid	0.234864	S.E. of regression	0.063635
R^2	0.362995	Adjusted R^2	0.352012
$F(1, 58)$	33.05108	P-value(F)	3.52e−07
Log-likelihood	81.15651	Akaike criterion	−158.3130
Schwarz criterion	−154.1243	Hannan-Quinn	−156.6746
$\hat{\rho}$	−0.168157	Durbin-Watson	2.334075

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

2.1 Simple regression and OLS estimation

The model

Measuring the fit of a regression model: R^2

Nonlinear relationships

2.2 The econometrics of the simple regression model

2.3 Multiple regression

2.4 Regression with dummy variables

2.5 Freeing up the classical assumptions

2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Measuring the fit of a regression model

- The most common measure of fit is the **coefficient of determination** R^2.
- Total variance of dependent variable Y is
  - variance explained by the explanatory variable (X) in the regression
  - plus variance that cannot be explained and is left as error.

Coefficient of determination, R^2

R^2 measures *proportion* of the variance in Y that can be explained by X.

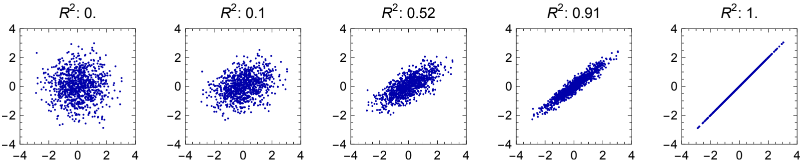
Definition of R^2

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- R^2 is defined as

$$R^2 = 1 - \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_Y^2}.$$

- Note that $0 \leq R^2 \leq 1$ and $R^2 = \hat{\rho}_{XY}$.
- $R^2 = 0$ corresponds to the case of "no correlation".
- $R^2 = 1$ corresponds to the case of "perfect correlation", which describes a perfect linear relationship between X and Y.



Example: Apple vs. S&P 500

- Output from gretl:

Model 1: OLS, using observations 2015:06-2020:05 (T = 60)
Dependent variable: ld_AppleClose

	Coefficient	Std. Error	t-ratio	p-value
const	0.00768812	0.00826986	0.9297	0.3564
ld_SPClose	1.15288	0.200536	5.749	0.0000

Mean dependent var	0.013145	S.D. dependent var	0.079052
Sum squared resid	0.234864	S.E. of regression	0.063635
R^2	0.362995	Adjusted R^2	0.352012
$F(1, 58)$	33.05108	P-value(F)	3.52e−07
Log-likelihood	81.15651	Akaike criterion	−158.3130
Schwarz criterion	−154.1243	Hannan-Quinn	−156.6746
$\hat{\rho}$	−0.168157	Durbin-Watson	2.334075

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

   2.1 Simple regression and OLS estimation

      The model

      Measuring the fit of a regression model: R^2

      Nonlinear relationships

   2.2 The econometrics of the simple regression model

   2.3 Multiple regression

   2.4 Regression with dummy variables

   2.5 Freeing up the classical assumptions

   2.6 Time series models

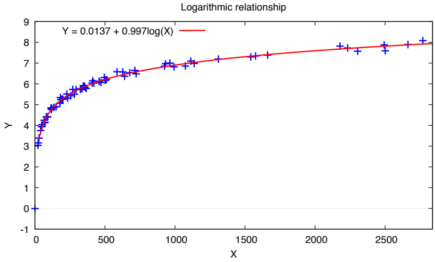
3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Nonlinear relationships

- So far, regression of Y on X: $Y = \alpha + \beta X + \varepsilon$.
- If the relationship between X and Y is non-linear, one can do regression of Y (or $\ln Y$ or Y^2...) on X^2 (or $1/X$ or $\ln X$ or X^3...).
- Example: $Y = \alpha + \beta \ln(X) + \varepsilon$.
- How do you know if a relationship is nonlinear?
- Answer: Careful examination of scatter plots or residual plots or hypothesis testing procedures.



Data transformations

- If X = company earnings and W = number of shares, new variable

$$Y = \frac{X}{W} = \text{earnings per share}$$

eliminates firm **size effects**.

- Asset prices Y_t, $t \geq 0$, are **non-stationary**.[8] Transformations to **returns**:

$$r_t = \frac{Y_{t+1} - Y_t}{Y_t} \text{ (discrete)}, \qquad \ln\left(\frac{Y_{t+1}}{Y_t}\right) \text{ (continuous)}.$$

- To measure **percentage changes**, economic variables are sometimes transformed into **logs**. Example:
  - ln(wage) as apposed to wage when regressing wage on education (measured in number of years).
  - If the estimated regression coefficient is x, then interpretation is that one additional year of education increases wage by $x \cdot 100\%$ on average.

[8] Their mean and variance change over time.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
     - The classical assumptions for the regression model
     - Properties of the OLS estimator
     - Hypothesis tests
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Simple regression model

$$Y = \alpha + \beta X + \varepsilon.$$

- In deriving the regression line, we have not used any economic or statistical theory.
- It is just an optimisation problem that determines the best linear approximation of one variable using another variable.
- In practice, we often want more than just a best linear approximation of a sample of data.
- In order to draw conclusions, we specify the simple regression model to be a **statistical model by adding probability-theoretic assumptions** on the variables and the error term.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
     - The classical assumptions for the regression model
     - Properties of the OLS estimator
     - Hypothesis tests
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

- To keep formulas simple, we will work with the simple regression model (i.e., one explanatory variable) with no intercept:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \dots, n.$$

- Note that (X_i, Y_i), $i = 1, \dots, n$, are random variables, assumed to be independent and identically distributed (iid).
- Derivations for multiple regression model are conceptually similar, but formulas get more complicated (involving matrix algebra).

- **Classical assumptions**:[9]
  1. $\mathbb{E}(Y_i) = \beta X_i$
  2. $\text{var}(Y_i) = \sigma^2$
  3. $\text{cov}(Y_i, Y_j) = 0$, for $i \neq j$
  4. Y_i is normally distributed
  5. X_i is fixed; it is not a random variable
- Compact notation: Y_i are independent $\text{N}(\beta X_i, \sigma^2)$.

[9] When loosening conditions 4 and 5 to requiring that $(X_1, \dots, X_n)$ and $(\varepsilon_1, \dots, \varepsilon_n)$ are independent, the conditions are called **Gauss-Markov assumptions**.

- An equivalent way of writing the classical assumptions is:
  1. $\mathbb{E}(\varepsilon_i) = 0$ (mean zero errors)
  2. $\text{var}(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2) = \sigma^2$ (constant variance errors; **homoskedasticity**)
  3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$
  4. ε_i is normally distributed
  5. X_i is fixed; it is not a random variable

Remarks on the classical assumptions

- Classical assumptions are foundation for inferential statistics, e.g. hypothesis testing.
- $\mathbb{E}(Yi) = \beta X_i$ captures the linearity assumption of the regression model.
- Assumption 2: all observations have the same variance (**homoskedasticity**).
- Example where this might not be a good assumption:
  - Real estate data: Small flats could all be the same, whereas big flats could be more diverse. If so, property prices might be more diverse for big flats (**heteroskedasticity**).
- Assumption 3: observations uncorrelated with one another.
- Usually reasonable with cross-sectional data (e.g. in a survey, response of person 1 and person 2 are unrelated), but not for time series data (e.g. interest rate now and last month are correlated with one another).

Remarks on the classical assumptions

- Assumption 4 – Y is normally distributed – may appear unreasonably restrictive.
- However, asymptotic theory (i.e., central limit theorem) can often be used to relax this assumption, provided sample is large enough.
- Assumption 5 – explanatory variable is not a random variable – is good in experimental sciences, but may be not be reasonable in social sciences.
- However, this simplifies the notation greatly – otherwise we would need to work with conditional expectations.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
   2.2 The econometrics of the simple regression model
      The classical assumptions for the regression model
      Properties of the OLS estimator
      Hypothesis tests
   2.3 Multiple regression
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

The ordinary least squares (OLS) estimator

- Simple regression:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \dots, n.$$

- OLS estimator is chosen to minimize:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta X_i)^2$$

- Taking first derivative with respect to β

$$\begin{aligned} \frac{\mathrm{d} \sum \varepsilon_i^2}{\mathrm{d}\beta} &= \sum 2(Y_i - \beta X_i)(-X_i) = 2 \sum (-X_i Y_i + \beta X_i^2) \\ &= 2\beta X_i^2 - 2 \sum X_i Y_i. \end{aligned}$$

- Setting first derivative to zero and solving for β gives the OLS estimator:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

Properties of OLS estimator

- First, we note an alternative way of writing the OLS estimator, which we shall use several times:

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i(X_i\beta + \varepsilon_i)}{\sum X_i^2} = \beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2} \qquad (*)$$

Property 1: OLS is unbiased under the classical assumptions

$$\mathbb{E}(\hat{\beta}) = \beta$$

- Proof on next slide.

Proof of Property 1

- Use Equation (19) and properties of expectations.
- Remember: X_i is not random and hence can be treated as a constant.

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}\left(\beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) = \beta + \mathbb{E}\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right)$$

$$= \beta + \frac{1}{\sum X_i^2}\mathbb{E}\left(\sum X_i \varepsilon_i\right) = \beta + \frac{1}{\sum X_i^2}\sum X_i \mathbb{E}(\varepsilon_i)$$

$$= \beta$$

Properties of OLS estimator

Property 2: Variance of OLS estimator under the classical assumptions

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum X_i^2}$$

- Remember that variance relates to dispersion.
- This property expresses how dispersed/uncertain/imprecise the OLS estimator is.
- Proof (on next slide) uses Equation (19) and properties of variance operator.
- Remember X_i is not random (hence can be treated as a constant).

Proof of Property 2

$$\begin{aligned}
\text{var}(\hat{\beta}) &= \text{var}\left(\beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) = \text{var}\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= \left(\frac{1}{\sum X_i^2}\right)^2 \text{var}\left(\sum X_i \varepsilon_i\right) = \left(\frac{1}{\sum X_i^2}\right)^2 \sum X_i^2 \text{var}(\varepsilon_i) \\
&= \left(\frac{1}{\sum X_i^2}\right)^2 \sigma^2 \sum X_i^2 \\
&= \frac{\sigma^2}{\sum X_i^2}.
\end{aligned}$$

Properties of the OLS estimator

Property 3: Distribution of OLS estimator under classical assumptions

$\hat{\beta}$ is $\mathrm{N}\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right)$-distributed.

- Proof: Properties 1 and 2 plus "linear combinations of normals are normal" theorem.
- Property 3 is important since we can use it to derive confidence intervals and perform hypothesis tests.
- The assumption that Y_i, resp. ε_i are normally distributed is easily relaxed, by applying the Central Limit Theorem to Equation (19).

Properties of the OLS estimator

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Property 4: The Gauss-Markov Theorem

If the classical assumptions hold, then the OLS estimator is the best, linear unbiased estimator.

- best = minimum variance
- linear = estimator is linear function of the observations $Y_1, \ldots Y_N$.
- Short form: "OLS is BLUE".
- For a proof see e.g. pages 84–85 of (Koop, 2008).fin
- Note: the assumption of normal errors is **NOT** required to prove Gauss-Markov theorem.
- Hence, OLS is BLUE even if errors are not normally distributed.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
     - The classical assumptions for the regression model
     - Properties of the OLS estimator
     - Hypothesis tests
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Hypothesis tests about β

- A hypothesis of special interest is: $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$.
- If $\beta = 0$, then the model $Y = \beta X + \varepsilon$ simplifies to $Y = \varepsilon$.
- If σ^2 is known, then do **Gauss test** with test statistic

$$Z = \frac{\hat{\beta}}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \sim \mathrm{N}(0,1).$$

- If σ^2 is unknown (the common case in practice), then replace σ^2 by an estimate S^2 to give t-statistics

$$T = \frac{\hat{\beta}}{\sqrt{\frac{S^2}{\sum X_i^2}}} \sim t(n-1),$$

  and do t-**test**.
- Check p-value to conclude if H_0 can be rejected.

Note on hypothesis testing

- If we reject the hypothesis that $\beta = 0$, ...
- ... then conclude:

"X has significant explanatory power for Y"

"β is significantly different from zero"

"β is statistically significant".

p-value

p-value

The **p-value** is the probability of observing an estimate (e.g. $\hat{\beta}$) as extreme as the one observed *under the assumption that the null hypothesis is true*.

- "as extreme as": pointing away from H_0 in the direction of H_1.
- Equivalently, p-value is level of significance at which you can reject H_0.
- For example, if 5% level of significance is chosen and software package gives p-value of 0.05 *or smaller*, then reject H_0.
- **Pitfalls:**
  1. False interpretation: p-value measures the probability that $\beta = 0$. *This does not make sense: either* $\beta = 0$ *or* $\beta \neq 0$*.*
  2. p-values only valid under classical assumptions, or more generally, **Gauss-Markov properties**. If not satisfied (e.g. heteroskedasticity, non-stationarity of time series), p-values may be invalid.

Example: Apple vs. S&P 500

- Output from gretl:

Model 1: OLS, using observations 2015:06-2020:05 (T = 60)
Dependent variable: ld_AppleClose

	Coefficient	Std. Error	t-ratio	p-value
const	0.00768812	0.00826986	0.9297	0.3564
ld_SPClose	1.15288	0.200536	5.749	0.0000

Mean dependent var	0.013145	S.D. dependent var	0.079052
Sum squared resid	0.234864	S.E. of regression	0.063635
R^2	0.362995	Adjusted R^2	0.352012
$F(1, 58)$	33.05108	P-value(F)	3.52e−07
Log-likelihood	81.15651	Akaike criterion	−158.3130
Schwarz criterion	−154.1243	Hannan-Quinn	−156.6746
$\hat{\rho}$	−0.168157	Durbin-Watson	2.334075

Hypothesis testing involving R^2: The F-statistic

- Another popular test is **F-test**, which tests the hypothesis $H_0 : R^2 = 0$.
- If $R^2 = 0$, then X does not have any explanatory power for Y.
- Note: for simple regression, this is equivalent to a test of $\beta = 0$.
- However, for multiple regression, the test of $R^2 = 0$ will be different than tests of whether regression coefficients equal zero.
- Most software packages calculate a p-value, which directly gives a measure of the plausibility of $H_0 : R^2 = 0$.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
     - Multiple regression and OLS
     - Omitted variable bias
     - Multicollinearity
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
     - Multiple regression and OLS
     - Omitted variable bias
     - Multicollinearity
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

OLS estimation of the multiple regression model

- **Multiple regression** similar to simple regression except that **several explanatory variables**.

Multiple regression model

With k explanatory variables the model is:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i,$$

where i denote observations, $i = 1, \ldots, n$.

- With multiple regression we need to estimate α and $\beta_1, \ldots, \beta_k$.
- OLS estimates are found by choosing the values of $\hat{\alpha}$ and $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ that minimize the sum of squared error terms:

$$\min_{\alpha,\beta_1,\beta_2,\ldots\beta_k} \sum_{i=1}^{N} \varepsilon^2 = \min_{\alpha,\beta_1,\beta_2,\ldots\beta_k} \sum_{i=1}^{N} \left(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki} \right)^2 .$$

Statistical aspects of multiple regression

- Statistical aspects of multiple regression are largely the same as for simple regression.
- R^2 is still a measure of fit.
- Can test $R^2 = 0$ (F-test) in same manner as for simple regression.
- If you find $R^2 \neq 0$, then conclude that explanatory variables *together* provide significant explanatory power (this does not necessarily mean each individual explanatory variable is significant).
- Can test $\beta_j = 0$ for each individual coefficient ($j = 1, 2, \ldots, k$) as before.
- Have test statistic and p-value for each coefficient.

Interpreting OLS estimates

- **Mathematically**:
  - Simple regression:
  $$\beta = \frac{\mathrm{d}Y}{\mathrm{d}X}$$
  - Multiple regression:
  $$\beta_j = \frac{\partial Y}{\partial X_j},$$
  for j-th coefficient, $j = 1, \dots, k$.
- **Verbally**:
  - With simple regression β is the marginal effect of X on Y.
  - Multiple regression: β_j is the marginal effect of X_j on Y, **ceteris paribus**.[10]
  - β_j is the effect of a small change in the j-th explanatory variable on the dependent variable, *holding all the other explanatory variables constant*.

[10] Ceteris paribus: all else being equal

Interpreting OLS estimates

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- In simple regression, $\hat{\beta} = \hat{\rho}_{XY} \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} = \frac{\text{Cov}(y_i, x_i)}{\text{Var}(x_i)}$.
- One can show that
$$\hat{\beta}_j = \frac{\text{Cov}(y_i, \tilde{x}_{ij})}{\text{Var}(\tilde{x}_{ij})},$$
where $\tilde{x}_{ij}$ are the residuals from regressing x_j on the other regressors $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$.
- This expresses that each coefficient in a regression is the slope coefficient in a simple regression after partialing out the other covariates.
- For more information, see e.g. Section 3.1.2 (Angrist and Pischke, 2008) or look for Frisch-Waugh Theorem.

Example: Fama-French factors

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Returns R_S of a stock or portfolio are explained by several portfolios representing different risk factors.
- **Fama-French factors** (Fama and French, 1993):
  - **Market portfolio** R_{mkt}: A well-diversified portfolio of all stocks in the market (e.g. a large stock index)
  - **Small-minus-big** factor R_{SMB}: Portfolio consisting of small stocks financed by shorting large stocks
  - **High-minus-low** factor R_{HML}: Portfolio consisting of stocks with a high book-to-market ratio financed by shorting stocks with a low book-to-market ratio
- Regression model with excess returns over risk-free interest rate r_f:

$$R_{S,t} - r_{f,t} = \alpha + \beta_{\mathrm{mkt}} \left(R_{\mathrm{mkt},t} - r_{f,t}\right) + \beta_{\mathrm{SMB}}\, R_{\mathrm{SMB},t} + \beta_{\mathrm{HML}}\, R_{\mathrm{HML},t} + \varepsilon_t,$$

where t refers to the time points of the observations.

- Berkshire Hathaway is the investment company founded and managed by Warren Buffet.[11]
- Data:[12]
  - Monthly (excess) returns of BRK-A,
  - excess returns of market,[13]
  - returns of portfolios expressing Fama-French factors SMB and HML.

[11] Example is taken from https://www.kaggle.com/code/yousefsaeedian/introduction-to-factor-investing.

[12] Data from https://www.kaggle.com/datasets/yousefsaeedian/edhec-investment-management-datasets

[13] Presumably from S&P 500.

Example: Factor analysis of Berkshire Hathaway

- Output from gretl:

Model 3: OLS, using observations 1990:01-2018:12 (T = 348)
Dependent variable: brka_exc

	Coefficient	Std. Error	t-ratio	p-value
const	0.00516540	0.00259475	1.991	0.0473
MktRF	0.709605	0.0625200	11.35	0.0000
SMB	−0.482937	0.0847882	−5.696	0.0000
HML	0.405283	0.0901842	4.494	0.0000

Mean dependent var	0.009646	S.D. dependent var	0.057498
Sum squared resid	0.783189	S.E. of regression	0.047715
R^2	0.317293	Adjusted R^2	0.311339
$F(3, 344)$	53.29220	P-value(F)	2.59e−28
Log-likelihood	567.0149	Akaike criterion	−1126.030
Schwarz criterion	−1110.621	Hannan-Quinn	−1119.895
$\hat{\rho}$	−0.088217	Durbin-Watson	2.162132

Example: Factor analysis of Berkshire Hathaway

- Interpretation of coefficient `MktRF`:
  - Recall that excess market returns are expressed in percentages.
  - Coefficient $\hat{\beta}_{\text{mkt}} = 0.71$ means that if excess market return increases by one percentage point, then – on average – Berkshire Hathaway increases by 0.71 percentage points, *keeping everything else constant*.
- Likewise: $\hat{\beta}_{\text{SMB}} = -0.48$ indicates that Berkshire Hathaway decreases by 0.48 percentage points if excess return of small over big businesses is one percentage point – on average, keeping everything else fixed (Berkshire Hathaway is known to invest in large firms.)
- Positive coefficient of high-minus-low confirms the well-known fact that Berkshire Hathaway is a value investor.
- Positive constant indicates that investors earn additional return when investing in Berkshire Hathaway as opposed to investing in factor portfolios.

Which explanatory variables to choose in a multiple regression model?

- Which explanatory variables should we choose in a multiple regression model?
- We will relate this question to the topics of **omitted variables bias** and **multicollinearity**.
- First note that these are two important considerations that pull in opposite directions.
- It is good to include all variables that help explain the dependent variable (include as many explanatory variables as possible).
- Including irrelevant variables (i.e. ones with no explanatory power) will lead to less precise estimates (include as few explanatory variables as possible).
- Playing off these two competing considerations is an important aspect of any empirical exercise. Hypothesis testing procedures can help with this.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
   2.2 The econometrics of the simple regression model
   2.3 Multiple regression
      Multiple regression and OLS
      Omitted variable bias
      Multicollinearity
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Example: Explaining house prices

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- To illustrate this problem we use the house price data set.
- Data on $n = 546$ houses sold in Windsor, Canada.
- Dependent variable, Y, is the sales price of the house in Canadian dollars.
- Four explanatory variables:
  - X_1 = lot size of the property (in square feet)
  - X_2 = number of bedrooms
  - X_3 = number of bathrooms
  - X_4 = number of storeys (excluding the basement).

Omitted variable bias

- Results of the multiple regression of saleprice on several variables:

	Coefficient	Std. Error	*t*-ratio	p-value
const	−4009.55	3603.11	−1.113	0.2663
lotsize	5.42917	0.369250	14.70	0.0000
bedrooms	2824.61	1214.81	2.325	0.0204
bathrooms	17105.2	1734.43	9.862	0.0000
stories	7634.90	1007.97	7.574	0.0000

- Simple regression of saleprice on number of bedrooms:

	Coefficient	Std. Error	*t*-ratio	p-value
const	28773.4	4413.75	6.519	0.0000
bedrooms	13270.0	1444.60	9.186	0.0000

- Why are these two coefficients on the same explanatory variable so different?

Omitted variable bias

- Answer 1: They just come from two different regressions that control for different explanatory variables (different ceteris paribus conditions).
- Answer 2:
  - Imagine a friend asked: "I have 2 bedrooms and I am thinking of building a third. How much will it raise the price of my house?"
  - Simple regression: "Houses with 3 bedrooms tend to cost \$13,269.98 more than houses with 2 bedrooms."
  - Does this mean adding a third bedroom will tend to raise price of house by \$13,269.98? Not necessarily, other factors influence house prices.

Omitted variable bias

- Houses with three bedrooms also tend to be desirable in other ways (e.g. bigger, with larger lots, more bathrooms, more storeys, etc.). Call these "good houses".
- Simple regression notes "good houses" tend to be worth more than others.
- Number of bedrooms acts as a proxy for all these "good house" characteristics and hence its coefficient becomes very big (13,269.98) in simple regression.
- Multiple regression can estimate separate effects due to lot size, number of bedroom, bathrooms and storeys.
- Tell your friend: "Adding a third bedroom will tend to raise your house price by $2,824.61".
- Multiple regressions, which contains all (or most) of house characteristics, will tend to be more reliable than simple regression, which only uses one characteristic.

Omitted variable bias

- This is confirmed by taking a look at the correlation matrix of the data set:

	Price	Lot size	# bed	# bath	# storey
Price	1				
Lot size	0.54	1			
# bed	0.37	0.15	1		
# bath	0.52	0.19	0.37	1	
# storey	0.42	0.08	0.41	0.32	1

- Positive correlations between explanatory variables indicate that houses with more bedrooms also tend to have larger lot size, more bathrooms and more storeys.

Omitted variable bias

Omitted variable bias

Omitted variable bias is a statistical term capturing that coefficients may be misleading due to the omission of other correlated variables.

- **IF** 1. we exclude explanatory variables that should be present in the regression,
- **AND** 2. these omitted variables are correlated with the included explanatory variables,
- **THEN** 3. the OLS estimates of the coefficients on the included explanatory variables will be biased.

Practical advice for selecting explanatory variables

- Include (insofar as possible) all explanatory variables that you think might possibly explain your dependent variable. This will reduce the risk of omitted variable bias.
- However, including irrelevant explanatory variables reduces accuracy of estimation and increases confidence intervals ($\rightsquigarrow$ **multicollinearity**).
- So do t-tests (or other hypothesis tests) to decide whether variables are significant. Run new regressions omitting one-by-one the explanatory variables that are not significant.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
     - Multiple regression and OLS
     - Omitted variable bias
     - Multicollinearity
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Multicollinearity

Multicollinearity

If two or more explanatory variables are **highly correlated** with one another, then regression model has trouble telling which individual variable explains Y.

- Symptom: Individual coefficients may look insignificant, but regression as a whole may look significant (e.g. R^2 big, F-test p-value small, but p-value on individual coefficients high).
- Looking at a correlation matrix for explanatory variables is helpful in revealing extent and source of multicollinearity problem.

Example of multicollinearity

- Example using artifical data; true model:

$$Y = 0.5X_1 + 0.5X_2 + \varepsilon.$$

- Correlation between X_1 and X_2: 0.98.
- `gretl` output:

Model 4: OLS, using observations 1–70
Dependent variable: y

	Coefficient	Std. Error	t-ratio	p-value
const	0.0100544	0.114885	0.08752	0.9305
x1	0.523628	0.548575	0.9545	0.3432
x2	0.578672	0.542956	1.066	0.2903

Mean dependent var	0.005800	S.D. dependent var	1.370131
Sum squared resid	60.28705	S.E. of regression	0.948581
R^2	0.534574	Adjusted R^2	0.520681
$F(2, 67)$	38.47708	P-value(F)	7.46e–12
Log-likelihood	−94.09747	Akaike criterion	194.1949
Schwarz criterion	200.9404	Hannan–Quinn	196.8743

Example of multicollinearity

- Drop X_2 and re-run:

Model 7: OLS, using observations 1–70
Dependent variable: y

	Coefficient	Std. Error	t-ratio	p-value
const	0.0296106	0.113523	0.2608	0.7950
x1	1.09278	0.125626	8.699	0.0000

Mean dependent var	0.005800	S.D. dependent var	1.370131
Sum squared resid	61.30913	S.E. of regression	0.949529
R^2	0.526683	Adjusted R^2	0.519723
$F(1, 68)$	75.66706	P-value(F)	1.18e–12
Log-likelihood	−94.68587	Akaike criterion	193.3717
Schwarz criterion	197.8687	Hannan–Quinn	195.1580

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
     - Simple regression with dummy variable
     - Multiple regression with dummy variables
     - Multiple regression with dummy and non-dummy explanatory variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Regression with dummy variables

- **Dummy variable** is either 0 or 1.
- Turns qualitative "Yes"/"No" data into $1/0$.
- Example: Explaining House Prices
- Data set has 5 potential dummy explanatory variables:
  - $D_1 = 1$ if the house has a driveway ($= 0$ if it does not)
  - $D_2 = 1$ if the house has a recreation room ($= 0$ if not)
  - $D_3 = 1$ if the house has a basement ($= 0$ if not)
  - $D_4 = 1$ if the house has gas central heating ($= 0$ if not)
  - $D_5 = 1$ if the house has air conditioning ($= 0$ if not)

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
 2.1 Simple regression and OLS estimation
 2.2 The econometrics of the simple regression model
 2.3 Multiple regression
 2.4 Regression with dummy variables
  Simple regression with dummy variable
  Multiple regression with dummy variables
  Multiple regression with dummy and non-dummy explanatory variables
 2.5 Freeing up the classical assumptions
 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Simple regression with a dummy variable

- One dummy explanatory variable, D:

$$Y_i = \alpha + \beta\, D_i + \varepsilon_i,$$

for $i = 1, \ldots, N$ observations.

- OLS estimation produces $\hat{\alpha}$ and $\hat{\beta}$, and fitted regression line is:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}\, D_i.$$

- Since D_i is either 0 or 1, we either have

$$\hat{Y}_i = \hat{\alpha}$$

or $$\hat{Y}_i = \hat{\alpha} + \hat{\beta}.$$

Example: Explaining house prices

- Regress $Y =$ house price on $D =$ dummy for air conditioning ($= 1$ if house has air conditioning, $= 0$ otherwise).
- Fitted regression line is:

$$\hat{Y}_i = 59884.85 + 25995.74\, D_i$$

- Average price of house without air conditioning is \$59,885
- Average price of house with air conditioning is \$85,881
- (Remember, however, omitted variables bias;
  this simple regression no doubt suffers from it).

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
     - Simple regression with dummy variable
     - Multiple regression with dummy variables
     - Multiple regression with dummy and non-dummy explanatory variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

- Model:
$$Y_i = \alpha + \beta_1 \, D_{1i} + \cdots + \beta_k \, D_{ki} + \varepsilon_i .$$
- Example **House Prices**: regress Y = house price on D_1 = driveway dummy and D_2 = rec room dummy.

Multiple regression with dummy variables

- Fitted regression line:

$$\hat{Y}_i = 47099.08 + 21159.91\, D_{1i} + 16023.69\, D_{2i}$$

- Putting in either 0 or 1 values for the dummy variables, we obtain the fitted values for Y for the four categories of houses:
  - Houses with a driveway and recreation room ($D_1 = 1$ and $D_2 = 1$) have $\hat{Y}_i = 47099 + 21160 + 16024 = \$84{,}283$.
  - Houses with a driveway but no recreation room ($D_1 = 1$ and $D_2 = 0$) have $\hat{Y}_i = 47099 + 21160 = \$68{,}259$.
  - Houses with a recreation room but no driveway ($D_1 = 0$ and $D_2 = 1$) have $\hat{Y}_i = 47099 + 16024 = \$63{,}123$.
  - Houses with no driveway and no recreation room ($D_1 = 0$ and $D_2 = 0$) have $\hat{Y}_i = \$47{,}099$.
- Multiple regression with dummy variables may be used to classify the houses into different groups and to find average house prices for each group.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
     - Simple regression with dummy variable
     - Multiple regression with dummy variables
     - Multiple regression with dummy and non-dummy explanatory variables
   - 2.5 Freeing up the classical assumptions
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Multiple regression with dummy and non-dummy explanatory variables

- E.g. one dummy variable (D) and one regular non-dummy explanatory variable (X):

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \varepsilon_i.$$

- Example **Explaining House Prices**:
- Regressing Y on $D =$ air conditioning dummy and $X =$ lot size gives $\hat{\alpha} = 32{,}693$, $\hat{\beta}_1 = 20{,}175$ and $\hat{\beta}_2 = 5.638$.
- Two different fitted regression lines:
  - If $D_i = 1$ (i.e. the ith house has an air conditioner):

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 X_i = 52868 + 5.638 X_i$$

  - If $D_i = 0$ (i.e. the house does not have an air conditioner):

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_2 X_i = 32{,}693 + 5.638 X_i.$$

- The two regression lines have the same slope and differ in their intercepts.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
    2.1 Simple regression and OLS estimation
    2.2 The econometrics of the simple regression model
    2.3 Multiple regression
    2.4 Regression with dummy variables
    2.5 Freeing up the classical assumptions
        Heteroskedasticity*
        Autocorrelation*
        Instrumental variable methods*
    2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Freeing up the classical assumptions

- What if some assumptions of Gauss-Markov Theorem are not fulfilled?
- Derivation of p-values required all classical assumptions.
- Material here is based on Chapter 5 (Koop, 2008), but more practical; if you want more background, then look there.
- Topics:
  - **Heteroskedasticity**
  - **Autocorrelated errors**
  - **Instrumental variables**

Basic theoretical results

- Previously derived theoretical results using multiple regression model with classical assumptions:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

1. $\mathbb{E}(\varepsilon_i) = 0$ (mean zero errors)
2. $\text{var}(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2) = \sigma^2$ (constant variance errors, homoskedasticity)
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$
4. ε_i is normally distributed
5. Explanatory variables fixed, not random variables

Basic theoretical results

- Remember: Assumption 1 is innocuous.
  - If error had non-zero mean could include it as part of the intercept – it would have no effect on estimation of slope coefficients in the model.
- Assumption 4 can be relaxed by using asymptotic theory (Central Limit Theorem).
- Assumption 5 we will still maintain (will discuss this later on in the context of "instrumental variables" estimation).
- For now we focus on Assumptions 2 and 3.
- **Heteroskedasticity** relates to Assumption 2.
- **Autocorrelation** (also called **serial correlation**) relates to Assumption 3.

Basic idea

- Under classical assumptions, Gauss Markov theorem says "OLS is BLUE".
- But if Assumptions 2 and 3 are violated, this no longer holds (OLS is still unbiased, but no longer "best", i.e. no longer minimum variance).
- Concepts/proofs/derivations often use following strategy:
  - The model can be transformed to create a new model that satisfies classical assumptions.
  - We know OLS (on the transformed model) will be BLUE.
  - The OLS estimator using such a transformed model is called the **generalized least squares (GLS) estimator**.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   - 2.1 Simple regression and OLS estimation
   - 2.2 The econometrics of the simple regression model
   - 2.3 Multiple regression
   - 2.4 Regression with dummy variables
   - 2.5 Freeing up the classical assumptions
     - Heteroskedasticity*
     - Autocorrelation*
     - Instrumental variable methods*
   - 2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Heteroskedasticity

- **Heteroskedasticity** occurs when the error variance differs across observations.
- Example: Errors reflect mis-pricing of houses:
  - If small houses are all similar, then unlikely to be large pricing errors.
  - Suppose big houses are very different to one another, then possible to have large pricing errors.
  - In this case, expect errors for big houses to have larger variance than for small houses.
  - Statistically: heteroskedasticity is present and is associated with the size of the house.

Testing for heteroskedasticity

- If heteroskedasticity is present, OLS estimator is still **unbiased**, but **variance** of OLS estimator is **different** than under the classical assumptions.
- But $\text{var}(\hat{\beta})$ enters the formula for confidence intervals and test statistics.
- If heteroskedasticity is present and you ignore it, simply using the OLS estimator in a software package, the software package will use the incorrect formula for $\text{var}(\hat{\beta})$.
- **Test statistics (and confidence intervals) will be incorrect.**
- Thus, first, it is important to know if heteroskedasticity is present.
- There are many tests; here, we describe two of the most popular ones: the **White test** and the **Breusch Pagan test**.
- Second, it is important to know how to fix it.

Breusch-Pagan test for heteroskedasticity

- **Breusch-Pagan test**: H_0: error variance is constant (homoskedasticity).
- Run OLS on the original regression (ignoring heteroskedasticity), and using the residuals $\hat{\varepsilon}_i$, $i = 1, \dots, n$, calculate:
$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n}.$$
- Run second regression of the equation:
$$\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} = \gamma_0 + \gamma_1 Z_{1i} + \cdots + \gamma_p Z_{pi} + v_i,$$
with $Z_1, \dots, Z_p$ explanatory variables and functions thereof.
- Calculate Breusch-Pagan test statistic using regression sum of squares (RSS) of second regression:
$$\mathrm{BP} = \frac{\mathrm{RSS}}{2}.$$
- Test statistic has a $\chi^2(p)$ distribution.
- Intution: If value if high, then independent variables have high explanatory power for the errors terms.

White test for heteroskedasticity

- Similar to the Breusch-Pagan test.
- **White test** involves the following steps:
- Run OLS on original regression (ignoring heteroskedasticity) and obtain the residuals, $\hat{\varepsilon}_i$, $i = 1, \ldots, n$.
- Run second regression of equation:

  $$\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 Z_{1i} + \cdots + \gamma_p Z_{pi} + v_i$$

  and obtain the R^2 from this regression.
- Calculate the White test statistic:

  $$W = N\,R^2.$$

- This test statistic has a $\chi^2(p)$ distribution from which a critical value can be obtained.

Testing for heteroskedasticity

- An advantage of the White and Breusch Pagan tests is that you just need to choose $Z_1, \ldots, Z_p$.
- Usually these are just the explanatory variables in the original regression, although with the White test researchers sometimes also include squares and cross-products of the explanatory variables.
- `gretl` will do this automatically for you.
- A disadvantage is that, if the tests indicate that heteroskedasticity is present, they do not offer much guidance on how you should try and transform the model to do GLS.
- All you know is that heteroskedasticity is present and is related to one (or several) of the variables $Z_1, \ldots, Z_p$.

Recommendations for empirical practice

- If you think you might have a heteroskedasticity problem, begin by doing a heteroskedasticity test.
- If your tests indicate heteroskedasticity is present, then do some experimentation to see if you can solve the heteroskedasticity problem.
- Sometimes logging your dependent variable (and some or all of the explanatory variables) will be enough to fix the problem.
- Sometimes multiplying/dividing all your explanatory variables by some variable (Z_j) is enough to fix the problem.
- Try different choices for Z_j.
- Note: Every time you try such a transformation you must do a heteroskedasticity test to check if it has fixed the problem.

Recommendations for empirical practice

- If you cannot find a transformation that fixes the heteroskedasticity problem, then use a **heteroskedasticity consistent estimator (HCE or HC1)**.[14]
- Remember: if heteroskedasticity is present, then hypothesis tests involving β's will be incorrect. So wait until after you have corrected the problem (or are using an HCE) before doing hypothesis testing.
- Remember to be careful with your interpretation of the marginal effects of coefficients:
  - If logging solves heteroskedasticity problem, see page 111 of (Koop, 2008) for interpretation of logged coeffcients.
  - If multiplying/dividing by Z_j solves the problem, see page 129 of (Koop, 2008) for interpretation of coefficients.

[14] Both `gretl` and `statsmodels` in Python support heteroskedasticity consistent estimators.

Example: Explaining house prices

- Data set used before: $n = 546$ houses sold in Windsor, Canada.
- The dependent variable, Y, is the sales price of the house in Canadian dollars.
- The explanatory variables are:
  - X_1 = lot size of the property (in square feet)
  - X_2 = number of bedrooms
  - X_3 = number of bathrooms
  - X_4 = number of storeys (excluding the basement).
  - $D_1 = 1$ if the house has a driveway ($= 0$ if it does not)
  - $D_2 = 1$ if the house has a recreation room ($= 0$ if not)
  - $D_3 = 1$ if the house has a basement ($= 0$ if not)
  - $D_4 = 1$ if the house has gas central heating ($= 0$ if not)
  - $D_5 = 1$ if the house has air conditioning ($= 0$ if not)

Example: Explaining house prices

- Breusch-Pagan and White tests require the selection of variables that might be related to the error variance.
- We set these variables to be the explanatory variables in the original regression.
- We find $\mathrm{BP} = 117.18$ and $W = 55.64$.
- Critical values for both of these tests are taken from the $\chi^2(9)$ distribution.
- p-values are both close to 0 (smaller than 10^{-4}).
- Both tests indicate that heteroskedasticity is present.
- Thus, although the OLS estimates presented in previously for this data set were unbiased, the p-values from hypothesis testing were incorrect.

Example: Explaining house prices

- To try and eliminate the heteroskedasticity problem we can experiment with different transformations until we find one that eliminates heteroskedasticity.
- The one that worked in this case was to take logs of all variables except for the dummies (since the log of zero is undefined, you cannot take the log of a dummy variable).
- In this log-linear regression, the values for the two heteroskedasticity tests are $\mathrm{BP} = 14.03$ and $W = 19.75$, with p-values 12.13% and 10.16%, respectively.
- Hence, both tests fail to reject the hypothesis of homoskedasticity.
- Thus, the log transformation has solved the heteroskedasticity problem in this example.
- Note: β_j in log linear regression is interpreted as an elasticity.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
   2.2 The econometrics of the simple regression model
   2.3 Multiple regression
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
      Heteroskedasticity*
      Autocorrelation*
      Instrumental variable methods*
   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Autocorrelation

- We will continue our discussion of problems by considering **autocorrelation**.
- This is used with time series data, so we will use $t = 1, \ldots, T$ to denote observations (rather than $i = 1, \ldots, N$).
- Reminder of classical assumptions: Slide 157.
- Remember: Slide 158
- **Autocorrelation** (also called **serial correlation**) relates to Assumption 3.
- Basic ideas: Slide 159

Autocorrelated errors

- Use multiple regression model under the classical assumptions, with the exception that the errors follow an **autoregressive process of order 1** (AR(1)):

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

where u_t satisfies classical assumptions.

- So $\mathbb{E}(u_t) = 0$, $\text{var}(u_t) = \sigma^2$ and $\text{cov}(u_t, u_s) = 0$ (for $t \neq s$).
- We also assume $-1 < \rho < 1$.
- To preview later material, this restriction ensures **stationarity** and means you do not have to worry about problems relating to **unit roots** and **cointegration**.
- Focus on the AR(1) cases, but note that the AR(p) errors case is a simple extension:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots + \rho_p\varepsilon_{t-p} + u_t.$$

Variances and covariances of regression error

- Assumptions above specified properties of u_t, but need properties of ε_t.
- Notation:
$$\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2),$$
where the last equality follows since errors have mean zero.
- Derivation of variance of regression errors:
$$\sigma_\varepsilon^2 = \mathrm{var}(\rho\varepsilon_{t-1} + u_t) = \rho^2\,\mathrm{var}(\varepsilon_{t-1}) + \mathrm{var}(u_t) = \rho^2\sigma_\varepsilon^2 + \sigma^2 = \frac{\sigma^2}{1-\rho^2}$$
- We used properties of variance operator, in particular the fact that ε_{t-1} and u_t are independent of one another and that ε_t is homoskedastic.

- The covariance between different regression errors is

$$\operatorname{cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho\sigma_\varepsilon^2.$$

(Exercise!)

- For errors more than one period apart, can show:

$$\operatorname{cov}(\varepsilon_t, \varepsilon_{t-s}) = \rho^s\sigma_\varepsilon^2.$$

- Thus, the regression model with autocorrelated errors violates Assumption 3.
- Regression errors are **not** uncorrelated with one another.
- Hence, need to work with GLS estimator or robust estimator (**HAC** errors).

The GLS estimator for the autocorrelated errors

- Remember: GLS can be interpreted as OLS on a suitably transformed model.
- In this case, the appropriate transformation is referred to as "quasi-differencing".
- To explain what this is, consider the regression model:

$$Y_t = \alpha + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \varepsilon_t.$$

- Holds for every time period so we can take it at period $t-1$ and multiply both sides of the equation by ρ:

$$\rho Y_{t-1} = \rho\alpha + \rho\beta_1 X_{1(t-1)} + \cdots + \rho\beta_k X_{k(t-1)} + \rho\varepsilon_{t-1}.$$

- Subtract this equation from the original regression equation:

$$Y_t - \rho Y_{t-1} = \alpha - \rho\alpha + \beta_1 (X_{1t} - \rho X_{1(t-1)}) + \cdots + \beta_k (X_{kt} - \rho X_{k(t-1)}) + \varepsilon_t - \rho\varepsilon_{t-1}.$$

or

$$Y_t^* = \alpha^* + \beta_1 X_{1t}^* + \cdots + \beta_k X_{kt}^* + u_t.$$

The GLS estimator for the autocorrelated errors

- But u_t satisfies the classical assumptions, so OLS on this transformed model will be GLS (which will be BLUE).
- Note: transformed variables are "quasi-differenced":

$$Y_t^* = Y_t - \rho Y_{t-1}$$
$$X_{1t}^* = (X_{1t} - \rho X_{1(t-1)})$$
$$\text{etc.}$$

- The case with $\rho = 1$ is called "differenced" – this is not quite the same, so we say "quasi-differenced".

The GLS estimator for the autocorrelated errors

- One (relatively minor) issue:
- If our original data is from $t = 1, \dots, T$ then $Y_1^* = Y_1 - \rho Y_0$ will involve Y_0 (and same issue for explanatory variables).
- But we do not observe such "initial conditions".
- There are many ways of treating initial conditions.
- Simplest is to work with data from $t = 2, .., T$ (and use $t = 1$ values for variables as initial conditions).
- Summary: If we knew ρ, then we could quasi-difference the data and do OLS using the transformed data (which is equivalent to GLS).
- In practice, we rarely (if ever) know ρ. Hence, replace ρ by an estimate: $\hat{\rho}$.
- This is what the **Cochrane-Orcutt procedure** does.

The Cochrane-Orcutt procedure

- Remember: with autocorrelated errors, GLS is BLUE. However, OLS (on original data) is still unbiased.
- **Cochrane-Orcutt procedure** begins with OLS and then uses OLS residuals to estimate ρ.
- Cochrane-Orcutt procedure goes through following steps:
  1. Do OLS regression of Y_t on intercept, $X_{1t}, \dots, X_{kt}$ and produce the OLS residuals, $\hat{\varepsilon}_t$.
  2. Do OLS regression of $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}$ which provides $\hat{\rho}$.
  3. Quasi-difference all variables to produce

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1}$$

$$X_{1t}^* = (X_{1t} - \hat{\rho} X_{1(t-1)})$$

etc.

  4. Do OLS regression of Y_t^* on intercept,$X_{1t}^*, \dots, X_{tk}^*$, thus producing GLS estimates of the coefficients.

Autocorrelation consistent estimators

- Remember: with heteroskedasticity we discussed heteroskedasticity consistent estimator (HCE).
- This is less efficient than GLS, but is a correct second-best solution when GLS is difficult to implement.
- Similar issues hold for autocorrelated errors.
- There exist **autocorrelation consistent estimators** that allow for the correct use of OLS methods when you have autocorrelated errors.
- We will not explain these, but many popular econometrics software packages include them. The most popular is the **Newey-West estimator** (which corrects for heteroskedasticity as well).

Testing for autocorrelated errors

- If $\rho = 0$ then doing OLS on the original data is fine (OLS is BLUE).
- However, if $\rho \neq 0$, then a GLS estimator (e.g. so-called Cochrane-Orcutt estimator) is better.
- This motivates testing $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$.
- There are several such tests, here we describe some of the most popular ones.

Breusch-Godfrey test

- This is a test of $H_0 : \rho_1 = 0, \rho_2 = 0, \ldots, \rho_p = 0$ in the regression model with $\text{AR}(p)$ errors:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + u_t.$$

- Breusch-Godfrey test involves the following steps:
  1. Run a regression of Y_t on an intercept, $X_1, \ldots, X_k$ using OLS and produce the residuals, $\hat{\varepsilon}_t$.
  2. Run second regression of $\hat{\varepsilon}_t$ on intercept, $X_1, \ldots, X_k, \hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-p}$ using OLS and produce the R^2.
  3. Calculate the test statistic:

$$\text{LM} = T \cdot R^2.$$

  4. If H_0 is true, then LM has an (approximate) $\chi^2(p)$ distribution.
  5. Thus, critical value are taken from statistical tables for the χ^2-distribution; or check p-value.

The Box-Pierce and Ljung tests

- These test $H_0 : \rho = 0, \rho_2 = 0, \ldots, \rho_p = 0$
- Both based on idea that, if the errors are not autocorrelated, then the correlations between different errors should be zero.
- Replace errors by residuals.
- $\hat{\varepsilon}_t$ are residuals from OLS regression of Y on intercept and $X_1, \ldots, X_k$.
- Correlations between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t-s}$ are:

$$r_s = \frac{\sum_{t=s+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-s}}{\sum_{t=s+1}^{T} \hat{\varepsilon}_t^2}.$$

- **Box-Pierce test statistic** (sometimes called the Q **test statistic**) is:

$$Q = T \sum_{j=1}^{p} r_j^2,$$

with p meaning that $\mathrm{AR}(p)$ errors are being tested for.

The Box-Pierce and Ljung tests

- The **Ljung test statistic** is:

$$Q^* = T(T+2)\sum_{j=1}^{p} \frac{r_j^2}{T-j}.$$

- Critical values for both are taken from $\chi^2(p)$ tables.
- Econometrics software packages such as `gretl` will calculate test statistics.
- Warning: in some cases, one of the explanatory variables will be the dependent variable from a previous period ("lagged dependent variable").
- For instance:

$$Y_t = \alpha + \delta Y_{t-1} + \beta X_t + \varepsilon_t.$$

- The Box-Pierce and Ljung tests are not appropriate in this case. The Breusch-Godfrey test, however, is still appropriate.
- (Koop, 2008) discusses two additional approaches: **Durbin-Watson test statistic** and **Durbin's *h*-test**.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
   2.2 The econometrics of the simple regression model
   2.3 Multiple regression
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
      Heteroskedasticity*
      Autocorrelation*
      Instrumental variable methods*
   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Instrumental variable methods

- Under the classical assumptions (▸ Slide 157), OLS is BLUE.
- If we relax Assumptions 2 or 3 (e.g. allow for heteroskedasticity or autocorrelated errors, ▸ Slide 157), then OLS is no longer BLUE but it is still unbiased (although GLS is better).
- Now we will focus on Assumption 5 and relax assumption that explanatory variables are not random variables.
- In the case we are about to consider, OLS will be biased and an entirely different estimator will be called for – the **instrumental variables (IV) estimator**.
- For simplicity, use simple regression model, but results generalize to the case of multiple regression.

Case 1: Explanatory variable is random, but uncorrelated with error

- If X_i is random, then need to make some assumptions about its distribution.
- Assume X_i are independent and identically distributed (iid) random variables with:
$$\mathbb{E}(X_i) = \mu_X, \qquad \text{var}(X_i) = \sigma_X^2.$$
- Assume explanatory variable and errors are **uncorrelated** with one another:
$$\text{cov}(X_i, \varepsilon_i) = \mathbb{E}(X_i \varepsilon_i) = 0.$$
- Then, result under classical assumptions can be shown to hold approximately:
$$\hat{\beta} \text{ is N} \left(\beta, \frac{\sigma^2}{\sum X_i^2} \right).$$
- Bottom line: If we relax the assumptions of normality and fixed explanatory variables we get the same results as for OLS under the classical assumptions (but here they hold approximately), **provided explanatory variables are uncorrelated with the error term**.

Case 2: Explanatory variable is random and correlated with the error term

- Assume X_i are iid random variables with:

$$\mathbb{E}(X_i) = \mu_X, \qquad \text{var}(X_i) = \sigma_X^2.$$

- Now assume explanatory variable and errors are correlated with one another:

$$\text{cov}(X_i, \varepsilon_i) = \mathbb{E}(X_i \varepsilon_i) \neq 0.$$

- It turns out that, in this case, OLS is biased and a new estimator is called for: the **instrumental variables (IV) estimator**.
- Basic idea: proof that OLS is biased begins as before, giving

$$\mathbb{E}(\hat{\beta}) = \beta + \mathbb{E}\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right).$$

- Important point: if the error and explanatory variable are correlated, then OLS is biased and should be avoided.

The instrumental variables estimator

- An **instrumental variable (IV)**, Z_i, is a random variable that is uncorrelated with the error, but is correlated with the explanatory variable.
- Formally, an instrumental variable is assumed to satisfy:

$$\begin{aligned}
\mathbb{E}(Z_i) &= \mu_Z \\
\text{var}(Z_i) &= \sigma_Z^2 \\
\text{cov}(Z_i, \varepsilon_i) &= \mathbb{E}(Z_i\, \varepsilon_i) = 0 \\
\text{cov}(X_i, Z_i) &= \mathbb{E}(X_i Z_i) - \mu_Z \mu_X = \sigma_{XZ} \neq 0.
\end{aligned}$$

- Assuming an instrumental variable exists,the IV estimator is given as:

$$\hat{\beta}_{\text{IV}} = \frac{\sum_{i=1}^{n} Y_i Z_i}{\sum_{i=1}^{n} X_i Z_i}.$$

The instrumental variables estimator

- The asymptotic derivations in (Koop, 2008) imply – approximately / asymptotically:
$$\hat{\beta}_{\mathsf{IV}} \text{ is } \mathsf{N}\left(\beta, \frac{(\sigma_Z^2 + \mu_Z^2)\sigma^2}{n \cdot (\sigma_{XZ} + \mu_X \mu_Z)^2}\right).$$
- Note: This implies $\mathbb{E}(\hat{\beta}_{\mathsf{IV}}) = \beta$ is unbiased (approximately / asymptotically).
- This formula can be used to calculate confidence intervals, hypothesis tests.
- In practice, the unknown means and variances can be replaced by their sample counterparts.
- No additional details of how this is done, but note that econometrics software packages do IV estimation.
- Note: this is sometimes called the two stage least squares estimator (and `gretl` uses this term).

Testing for IV: The Hausman test

- The **Hausman test** is used to see if IV estimator is necessary.
- Basic idea of Hausman test:
  - H_0: explanatory variables are uncorrelated with error.
  - If H_0 holds, then both OLS and IV are acceptable estimators and should give roughly the same result.
  - However, if H_0 is false, then OLS in not acceptable, but IV is and results can be quite different.
  - Hausman test based on measuring the difference between $\hat{\beta}$ and $\hat{\beta}_{\mathrm{IV}}$.
- Example: simple regression with one instrument.
- Original regression model of interest is: $Y_i = \alpha + \beta X_i + \varepsilon_i$.
- Hausman test uses this regression but adds the instrumental variable:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon_i.$$

- If coefficient on Z is significant, reject H_0 and use IV estimator.
- If coefficient on Z is insignificant, do OLS on original regression model.

An example where the explanatory variable could be correlated with the error

- Suppose we are interested in estimating returns to schooling and have data from a survey of many individuals on:
  - dependent variable: $Y = \text{income}$
  - explanatory variable: $X = \text{years of schooling}$
- To simplify exposition, ignore other explanatory variables like experience, age, occupation, etc.
- The issue is that in such a regression it probably is the case that X is correlated with the error and, thus, OLS will be inappropriate.
- To understand why, first think of how errors are interpreted in this regression.
- An individual with a positive error earns an unusually high level of income. That is, his/her income is more than his/her education would suggest.
- Similar for negative error.
- What might be correlated with this error?
- Perhaps each individual has some underlying quality (e.g. intelligence, ambition, drive, talent, luck – or even family encouragement).

An example where the explanatory variable could be correlated with the error

- This quality would likely be associated with the error (e.g. individuals with more drive tend to achieve unusually high incomes).
- But this quality would also effect the schooling choice of the individual. For instance, ambitious students would be more likely to go to university.
- Summary: Ambitious, intelligent, driven individuals would both tend to have more schooling and more income (i.e., positive errors).
- So both the error and the explanatory variable would be influenced by this quality.
- Error and explanatory variable probably would be correlated with one another.

How do you choose instrumental variables?

- Far from straightforward how to do choose instrumental variables; here are some practical thoughts.
- An instrumental variable should be correlated with explanatory variable (e.g. schooling), but not with error (i.e., explain why individuals have unusually high/low income).
- An alternative way of saying this: we want to find a variable that affects schooling choice, but has no direct effect on income.
- Characteristics of parents or older siblings have been used as instruments.
- Justification: if either of your parents had a university degree, then you probably come from a family where education is valued (increase the chances you go to university). However, your employer will not care that your parents went to university (so no direct effect on your income).
- Other researchers have used geographical location variables as instruments.
- Justification: if you live in a community where a university/college is you are more likely to go to university. However, your employer will not care where you lived so location variable will have no direct effect on your income.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis
   2.1 Simple regression and OLS estimation
   2.2 The econometrics of the simple regression model
   2.3 Multiple regression
   2.4 Regression with dummy variables
   2.5 Freeing up the classical assumptions
   2.6 Time series models

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Time series models

- Time series Y_t, $t \in \mathbb{Z}$ or $t = 0, \dots, T$, take into account flow of time and related causality.
- Often, time series violate classical assumptions, especially Assumption 3, which are necessary for valid interpretation of p-values, …
- Reminder of classical assumptions: Slide 157.
- To fix this, sometimes include **lags** of dependent variable (or other variables).
- Avoid **non-stationary** time series or make them **stationary**.
- The time series Y_t, $t \in \mathbb{Z}$, is **weakly stationary** if
  - $\mathbb{E}Y_t$ is equal for all t;
  - $\mathrm{Var}(Y_t)$ is equal for all t;
  - $\mathrm{Cov}(Y_t, Y_{t-s})$ depends only on s, not on t.
- In words: "Basic statistical properties do not change over time."

AR(1) process

- Simplest time series model is AR(1) (auto-regressive of order 1) process:

$$Y_t = \alpha + \theta\, Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

  where $\mathbb{E}\varepsilon_t = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma^2$.

- If $|\theta| < 1$, then $(Y_t)_{t\in\mathbb{Z}}$ is a stationary process with

$$\begin{aligned} \mathbb{E}Y_t &= \frac{\alpha}{1-\theta} \\ \mathrm{Var}(Y_t) &= \frac{\sigma^2}{1-\theta^2} \\ \mathrm{Cov}(Y_t, Y_{t-s}) &= \frac{\rho^s \sigma^2}{1-\theta^2}, \end{aligned}$$

  where $\rho = \mathrm{Corr}(Y_t, Y_{t-1})$.

Weak stationarity

- For the first property, observe

$$\begin{aligned}\mathbb{E}Y_t &= \alpha + \theta(\alpha + \theta(\alpha + \theta(\alpha + \theta(\ldots)))) \\ &= \alpha + \theta\alpha + \theta^2\alpha + \theta^3\alpha + \cdots \\ &= \alpha(1 + \theta + \theta^2 + \theta^3 + \cdots) \\ &= \frac{\alpha}{1-\theta}\end{aligned}$$

- If $t = 0, \ldots$, then

$$\mathbb{E}Y_t = \alpha \sum_{k=0}^{t-1} \theta^k + \theta^t Y_0 = \frac{\alpha(1-\theta^t)}{1-\theta} + \theta^t Y_0 \quad \to \frac{\alpha}{1-\theta} \text{ as } t \to \infty$$

- Other properties follow similarly.

Unit root

- If $\theta = 1$, then the above properties do not hold.
- If $\alpha \neq 0$, it is easily seen that $\mathbb{E}Y_t = \infty$, $t \in \mathbb{Z}$, or $\mathbb{E}Y_t \to \infty$ as $t \to \infty$.
- However, if $\theta = 1$, then the process is **difference stationary**:

$$\Delta Y_t = Y_t - Y_{t-1} = \alpha + \varepsilon_t,$$

  which is easily seen to be stationary.
- In practice, with financial time series, one would often use returns:

$$\frac{\Delta Y_t}{Y_{t-1}} \quad \text{or} \quad \ln\left(\frac{Y_t}{Y_{t-1}}\right).$$

- Another form of stationarity is **trend stationary**

$$Y_t = \alpha + \delta \cdot t + \varepsilon_t,$$

  with some $\delta \in \mathbb{R}$.

Testing for stationarity

- To test for stationarity, write AR(1) process as

$$\Delta Y_t = Y_t - Y_{t-1} = \alpha + \rho Y_{t-1} + \varepsilon,$$

  where $\rho = \theta - 1$.

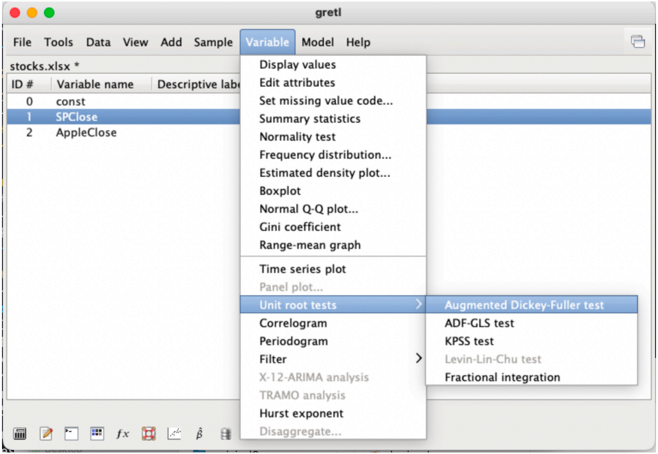
- The process has a unit root if $\rho = 0$.
- Careful: simply testing for $\rho = 0$ in OLS is misleading, because the test assumes stationarity to begin with!
- Use **(Augmented) Dickey-Fuller** test instead:

$$H_0 : \text{time series is non-stationary}$$
$$H_1 : \text{time series is stationary}$$

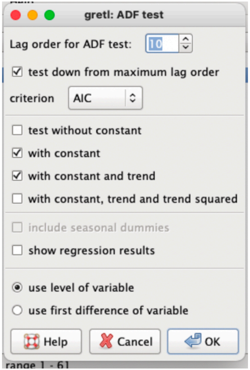
Stock price example

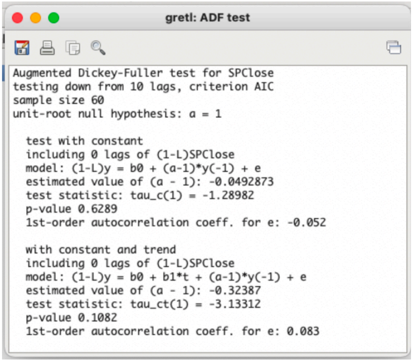
Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Monthly S&P 500 closing prices, May 2015 until April 2020



Stock price example

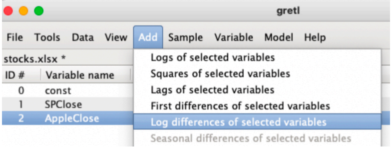


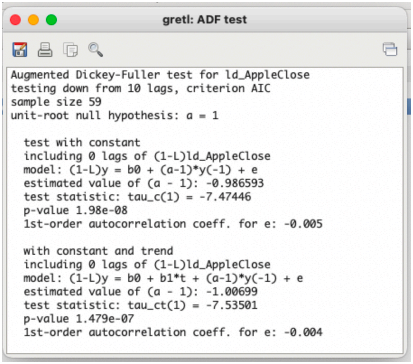


- Test results indicate that S&P 500 closing prices are non-stationary.

Stock price example

- Add log-returns (log differences) of S&P 500 closing prices.





- Test results indicate that log-returns are stationary.

Regression with time series

- Regression:

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

- If both time series are stationary, then regression is OK.
- But if one or both are non-stationary, then possibly p-values, etc., not valid, because the were computed under the assumption that classical assumptions hold. $\rightsquigarrow$ **spurious regression**
- Sole exception: **Co-integration**: If Y_t and X_t are **co-integrated**, then non-stationarity "cancels" in regression model, so error term fullfil classical assumptions.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   3.1 Options and option strategies
   3.2 Binomial tree model
   3.3 The Black-Scholes model
   3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Overview

- Option pricing and risk management is a topic that can easily take up a whole course.
- Here, we will focus on the basic theoretic insights, which hold apply to all, even sophisticated and complex, option pricing models.
- Option pricing theory is a fascinating field that combines
  - Economic theory
  - Stochastic calculus
  - Differential calculus
  - Numerical and computational techniques

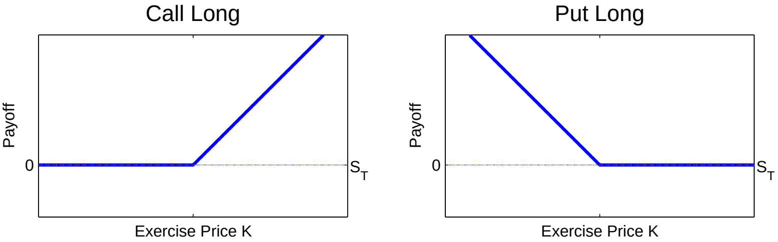
Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
    3.1 Options and option strategies
    3.2 Binomial tree model
    3.3 The Black-Scholes model
    3.4 Hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance

Options

- An **option** is the right, but not the obligation, to purchase or sell a financial security ("the underlying", e.g. a stock) at some time $T \geq 0$ at a pre-defined price K.
- The price K is called the **strike price**.
- The time point T is called the **expiry**.
- A $\left\{\begin{matrix}\textbf{call option}\\ \textbf{put option}\end{matrix}\right\}$ is the right to $\left\{\begin{matrix}\textbf{buy}\\ \textbf{sell}\end{matrix}\right\}$ an asset.
- The $\left\{\begin{matrix}\textbf{buyer}\\ \textbf{seller}\end{matrix}\right\}$ of an option is said to have a $\left\{\begin{matrix}\textbf{long position}\\ \textbf{short position}\end{matrix}\right\}$ in the option.
- The price paid for an option is also called **(option) premium**.

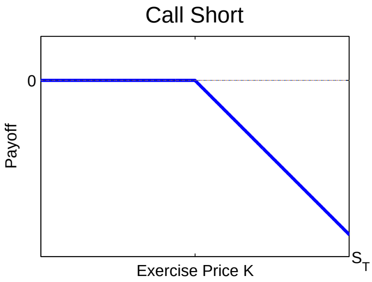
Payoff diagrams: call and put long

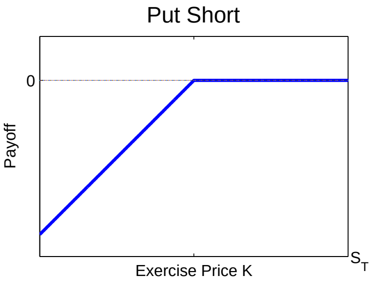


	call long	**put long**
payoff	$C_T = \max(S_T - K, 0)$ $=: (S_T - K)^+$	$P_T = \max(K - S_T, 0)$ $=: (K - S_T)^+$
possible gain	unlimited	limited (to strike)
possible loss	limited (to option premium)	

Payoff diagrams: call and put short

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law





	call short	put short
payoff	$-C_T = -(S_T - K)^+$	$-P_T = -(K - S_T)^+$
possible gain	limited (to option premium)	
possible loss	unlimited	limited (to strike)

Variations

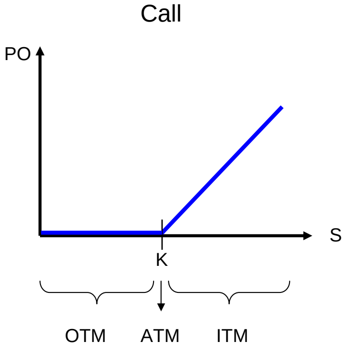
- European versus American options:
  - **European options**: can only be exercised at maturity
  - **American options**: can be exercised at any time
  - **Bermudan options**: can be exercised at a finite number of pre-defined time points before maturity
  - **Asian options**: can only be exercised at maturity, but payoff is related to average price of underlying during lifetime of the option
- Plain-vanilla versus exotic options
  - **Plain-Vanilla options**: standard options, e.g. call, put
  - **Exotic options**: binary, barrier, lookback, ...
  - **Contingent claims**: arbitrary derivative payoffs that depend on realisation of some unknown future event
- Physical Delivery versus Cash Settlement
  - **physical delivery**: underlying is delivered
  - **cash settlement**: gains and losses are settled in cash

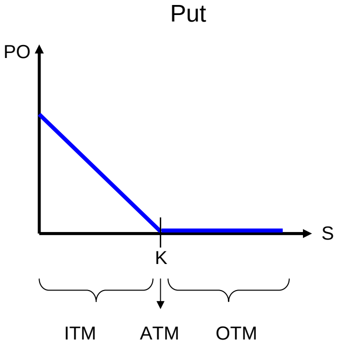
    (only alternative for temperature derivatives, for example)

In-, out- and at-the-money

- If the option matured right now,

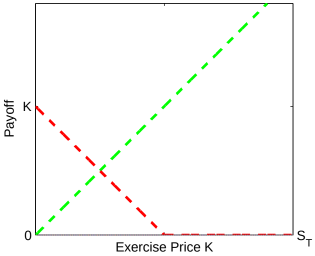
$$\text{one would} \left\{ \begin{array}{c} \textbf{exercise} \\ \textbf{not exercise} \\ \textbf{be indifferent for} \end{array} \right\} \text{an} \left\{ \begin{array}{c} \textbf{in-the-money} \\ \textbf{out-of-the-money} \\ \textbf{at-the-money (ATM)} \end{array} \right\} \text{option.}$$

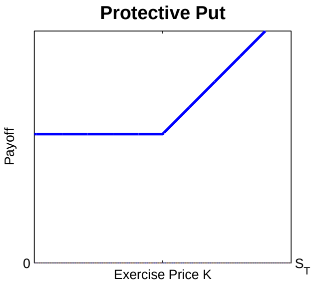




Protective Put

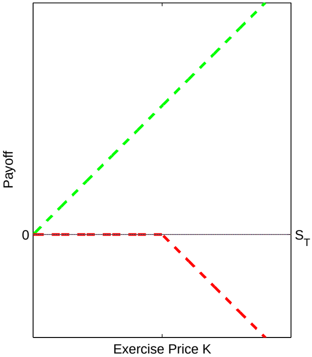
- **Protective Put**: underlying long, put long
- Portfolio value at T: at least equal to strike price
- Portfolio value today: at least equal to discounted strike price
- Objective: protect stock position against decreasing stock prices

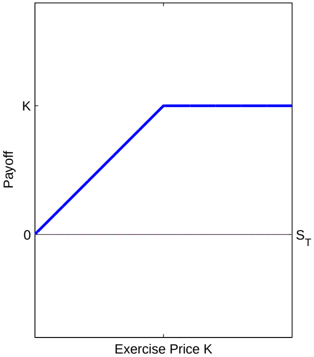




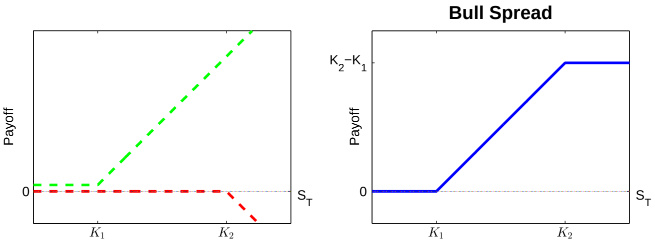
Covered Call Writing

- **Covered Call Writing**: call short, underlying long
- Objective: protect call short against increasing stock prices by taking a long position in the stock
- However: high losses if the stock price decreases (similar to put short!)

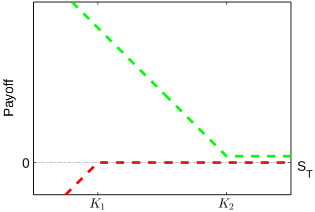


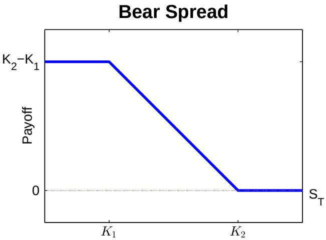


- **Bull Spread**:
  - Alternative 1: call long with strike K_1, call short with strike K_2
  - Alternative 2: amount $(K_2 - K_1)e^{-rT}$ in money market account (risk-free investment), put short with strike K_2, put long with strike K_1
- Objective: speculate on moderate increase of stock price

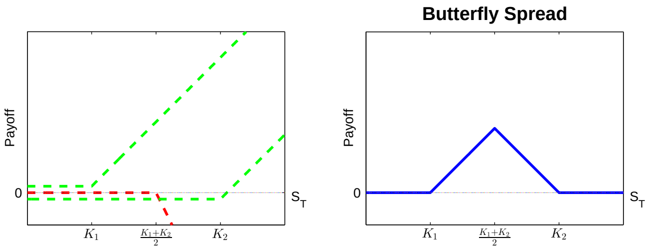


- **Bear Spread**:
  - Alternative 1: put long with strike K_2, put short with strike K_1
  - Alternative 2: amount $(K_2 - K_1)e^{-rT}$ in money market account, call short with strike K_1, call long with strike K_2
- Objective: speculate on moderate decrease of stock price

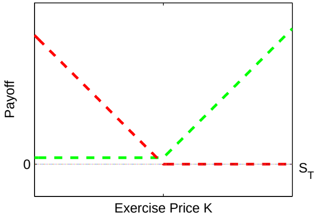


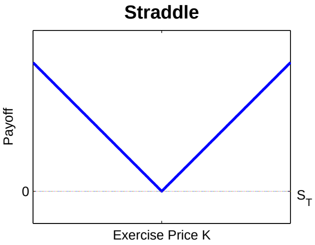


- **Butterfly Spread**:
  - Alternative 1: call long with strike K_1, 2 calls short with strike $\frac{K_1+K_2}{2}$, call long with strike K_2
  - Alternative 2: put long with strike K_2, 2 puts short with strike $\frac{K_1+K_2}{2}$, put long with strike K_1
- Objective: speculate on no (large) change of stock price

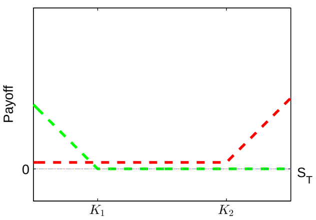


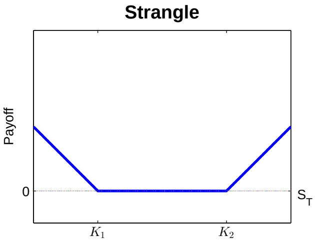
- **Straddle**
  - call long with strike K, put long with strike K
- Objective: speculate on large stock price changes (i.e. high volatility)





- **Strangle**
  - call long with strike K_2, put long with strike K_1 (where $K_1 < K_2$)
- Objective: speculate on very large stock price changes (i.e. very high volatility)





Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
     - The one-period model
     - Portfolios
     - Arbitrage
     - Replication
     - Risk-neutral valuation
     - The multiperiod model
     - The Cox-Ross-Rubinstein model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

The binomial model

- In practice, we mostly focus on continuous-time models, as they are more tractable in practice.
- To fix the basic concepts and ideas without technicalities, we start with a simple discrete-time setup.
- This is a convenient setting to give rigorous proofs of the underlying theory.

The binomial tree model

Example

Suppose you want to buy a **call option** on a stock. The option matures in one year and has a strike price of 105. The current stock price is 100, and it is known that the stock price either increases or decreases by 20% within one year. The risk-free interest rate is 5% (discrete compounding).

- What is the **value** of this option? (Pricing)
- How can you **hedge** this option? (Hedging)

- The **binomial tree model** is the simplest non-trivial model of a financial market giving answers to the questions above.
- It is a discrete-time model; time is indexed by time points $\{0,1,2,3,\dots\}$.
- In the simplest case we restrict attention to two time points $\{0,1\}$ $\rightsquigarrow$ **one-period binomial model**.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

  3.1 Options and option strategies

  3.2 Binomial tree model

    The one-period model
    Portfolios
    Arbitrage
    Replication
    Risk-neutral valuation
    The multiperiod model
    The Cox-Ross-Rubinstein model

  3.3 The Black-Scholes model

  3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

The one-period model

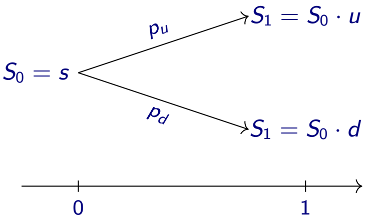
- Two points in time: $t \in \{0,1\}$
- The market consists of two assets, a **bond** and a **stock**
- The **bond price**, denoted by B_t, is deterministic and given by

$$B_0 = 1$$
$$B_1 = 1 + R,$$

with R the spot rate for the time period.

- The **stock price**, denoted by S_t, is a stochastic process and described by

$$S_0 = s$$
$$S_1 = \begin{cases} S_0 \cdot u, & \text{with probability } p_u \\ S_0 \cdot d, & \text{with probability } p_d. \end{cases}$$



The one-period model

- We also write
$$S_1 = S_0 \cdot Z,$$
with
$$Z = \begin{cases} u, & \text{with probability } p_u, \\ d, & \text{with probability } p_d. \end{cases}$$

- Assumptions:
  - S_0, u, d, p_u, p_d are known
  - $p_u, p_d > 0$
  - $d < u$
  - $p_u + p_d = 1$

The one-period model

Example

- The market consists of a bond and a stock, with the following properties:
- The risk-free interest rate is $R = 0.05$ p.a. (discretely compounded), so that, with $B_0 = 1$:

$$B_1 = B_0 \cdot (1 + R) = 1.05.$$

- For the stock price, let $S_0 = 100$, $u = 1.2$, $d = 0.8$, $p_u = 0.6$, $p_d = 0.4$.
- Thus, we have the following one-period tree describing the one-period evolution of the stock price:

$$\begin{tikzcd} & S_1 = 120 \\ S_0 = 100 \arrow[ur, "0.6"] \arrow[dr, "0.4"'] & \\ & S_1 = 80 \end{tikzcd}$$

The one-period model – Roadmap

We will be concerned with finding answers to the following questions:

1. Can we find conditions under which a market (model) is free of arbitrage?
2. Can we use the principle of no-arbitrage to determine prices of contingent claims?
3. In particular, when adding a contingent claim to a market, which price preserves the absence of arbitrage of the market?
4. Is there a *straightforward method* for calculating prices of contingent claims?

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

   3.1 Options and option strategies

   3.2 Binomial tree model

      The one-period model

      Portfolios

      Arbitrage

      Replication

      Risk-neutral valuation

      The multiperiod model

      The Cox-Ross-Rubinstein model

   3.3 The Black-Scholes model

   3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Portfolios

- **Portfolio (position)** is a vector $h = (x, y)$ with
  - x: number of bonds
  - y: number of stocks

Assumption (Frictionless market)

Throughout, we make the following assumptions:

- Short positions and arbitrary holdings are allowed ($h \in \mathbb{R}^2$).
- There are no bid-ask spreads.
- There are no transaction costs.
- The market is completely liquid, i.e., it is always possible to buy and sell unlimited quantities in the market.

Portfolios

Definition

The **value process** of the portfolio h is defined as

$$V_t^h = x\,B_t + y\,S_t\,, \quad t = 0{,}1$$

or, equivalently,

$$V_0^h = x + y\,S_0$$

$$V_1^h = x(1+R) + y\,S_0\,Z$$

Example (▸ cont'd)

- Continuing the earlier example, take a portfolio consisting of 100 units worth of the bond and 2 shares of stock.
- The holdings can be described as $h = (100,2)$.
- At $t = 0$, the value of the portfolio is

$$V_0^h = 100 + 2 \cdot 100 = 300.$$

- At $t = 1$, the value of the portfolio is

$$V_1^h = 100 \cdot 1.05 + 2 \cdot 100 \cdot Z = \begin{cases} 345, & \text{if } Z = u = 1.2, \\ 265, & \text{if } Z = d = 0.8. \end{cases}$$

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

  3.1 Options and option strategies

  3.2 Binomial tree model

    The one-period model

    Portfolios

    Arbitrage

    Replication

    Risk-neutral valuation

    The multiperiod model

    The Cox-Ross-Rubinstein model

  3.3 The Black-Scholes model

  3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Arbitrage

- The first question that arises, is under which conditions the market given by a one-period model is free of arbitrage (Roadmap).

Definition

An **arbitrage** is a portfolio h with properties

$$\begin{aligned} V_0^h &= 0 \\ \mathbb{P}(V_1^h \geq 0) &= 1 \\ \mathbb{P}(V_1^h > 0) &> 0 \end{aligned}$$

- In words: An arbitrage opportunity is the possibility of generating at zero cost today a payoff at some time point in the future that is nonnegative with certainty and positive with positive probability.
- This is the weakest form of arbitrage: it is impossible to make a loss, but there is a chance for a profit, at the cost of zero today.

Arbitrage

Proposition (No-arbitrage)

The one-period model is free of arbitrage if and only if the following condition holds:

$$d < (1 + R) < u. \tag{19}$$

- Interpretation: return on stock is not allowed to dominate return on bond, and vice versa.

Proof.

"$\Rightarrow$":

- Assume that (19) does not hold and show that this implies an arbitrage.
- Suppose $(1+R) \geq u$, so that $S_0(1+R) \geq su$ and $S_0(1+R) > sd$ (recall that $u > d$), so it is always more profitable to invest in the bond than in the stock.
- Arbitrage strategy: $h = (S_0, -1)$, i.e., sell stock and buy bond.
- $V_0^h = 0$
- $V_t^h = S_0(1+R) - S_0\,Z = \begin{cases} S_0(1+R) - S_0\,u \geq 0, & \text{if } Z = u, \\ S_0(1+R) - S_0\,d > 0, & \text{if } Z = d. \end{cases}$
- Now suppose $(1+R) \leq d$, so that $S_0(1+R) \leq S_0\,d < S_0\,u$, so it is always more profitable to invest in the stock than in the bond.
- Arbitrage strategy: $h = (-S_0{,}1)$, i.e., sell bond and buy stock.
- $V_0^h = 0$
- $V_t^h = -S_0(1+R) + S_0\,Z = \begin{cases} -S_0(1+R) + S_0\,u > 0, & \text{if } Z = u, \\ -S_0(1+R) + S_0\,d \geq 0, & \text{if } Z = d. \end{cases}$

Arbitrage*

Proof (cont'd).

"$\Leftarrow$":

- Now suppose that (19) is satisfied.
- Consider an arbitrary portfolio with $V_0^h = 0$.
- This implies $x + y\,S_0 = 0$, i.e., $x = -y\,S_0$.
- Then, by the Inequalities (19),

$$V_t^h = \begin{cases} y\,S_0[u-(1+R)] > 0, & \text{if } Z = u \\ y\,S_0[d-(1+R)] < 0, & \text{if } Z = d, \end{cases}$$

which shows that there is no arbitrage opportunity.

□

Example (cont'd)

- In the market described earlier, we have $d = 0.8$, $u = 1.2$ and $R = 0.05$, hence
$$0.8 < 1.05 < 1.2$$
is fulfilled, and the market is free of arbitrage.

- Suppose that instead the interest rate was $R = 0.25$. In this case, the trading strategy $h = (100, -1)$ is an arbitrage:
$$V_0^h = 100 - 100 = 0$$
$$V_1^h = 100 \cdot 1.25 - 100 \cdot Z \geq 125 - 120 = 5.$$

- Now, suppose the interest rate was $R = 0.05$ and that $d = 1.1$, $u = 1.2$. Then, $h = (-100,1)$ is an arbitrage:
$$V_0^h = -100 + 100 = 0$$
$$V_1^h = -100 \cdot 1.05 + 100 \cdot Z \geq -105 + 110 = 5.$$

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
     - The one-period model
     - Portfolios
     - Arbitrage
     - Replication
     - Risk-neutral valuation
     - The multiperiod model
     - The Cox-Ross-Rubinstein model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Contingent claims

- We now turn to the second question: can we determine prices of contingent claims by the principle of no-arbitrage? (▸ Roadmap)

Definition

A **contingent claim (financial derivative)** is any random variable X of the form $X = \Phi(Z)$, where Z is the random variable driving the stock price process.

- Φ is called the **contract function**.

$$\begin{tikzcd}[column sep=large]
 & S_1 = S_0 \cdot u \quad \boxed{\Phi(u)} \\
S_0 = s \arrow[ur, "p_u"] \arrow[dr, "p_d"'] & \\
 & S_1 = S_0 \cdot d \quad \boxed{\Phi(d)}
\end{tikzcd}$$

- Problem: find **price process** $\Pi(t; X)$ for claim X.

Replication

Definition

- A contingent claim X is said to be **reachable** if there exists a portfolio h such that

$$V_1^h = X,$$

with probability 1.

- Such a portfolio is called a **hedging portfolio** or **replicating portfolio**.
- If all claims can be replicated, then the market is **complete**.

Proposition

Suppose that a claim X is reachable with replicating portfolio h. Then any price at $t = 0$ of claim X other than V_0^h, leads to an arbitrage possibility.

Proof.

"Buy low, sell high". □

Replication

- Let's go about finding replicating portfolios: Fix an arbitrary claim X with contract function Φ.
- We want to find a portfolio $h = (x,y)$ such that

$$V_1^h = \begin{cases} \Phi(u), & \text{if } Z = u, \\ \Phi(d), & \text{if } Z = d. \end{cases}$$

- This requires a solution (x,y) to the system of equations,

$$\begin{aligned} (1+R)x + S_0\, uy &= \Phi(u), \\ (1+R)x + S_0\, dy &= \Phi(d). \end{aligned}$$

- Since $u > d$, the linear system has a unique solution, given by

$$x = \frac{u\Phi(d) - d\Phi(u)}{(1+R)(u-d)} \qquad y = \frac{1}{S_0} \cdot \frac{\Phi(u) - \Phi(d)}{u-d} \tag{20}$$

Replication

- Note we just proved the following statement:

Proposition

Assume that the binomial model is free of arbitrage. Then it is also complete.

Remark

- The portfolio replicating the option payoff is called the **replicating portfolio**.
- The option price is called the **no-arbitrage price** of the option.
- Adding an option to the market that trades at its no-arbitrage price implies that the market remains free of arbitrage (see Question 3 of Roadmap).
- A buyer of the option can **hedge** the exposure by entering into **minus** the replicating portfolio.
- The probabilities p_u, p_d did not enter into the calculation of the price.

Example: Replicating a call option

- We apply the findings to a call option.

Example (cont'd)

Consider a European call option with strike price $K = 105$, so that:

$$\begin{tikzcd}[column sep=huge] & \begin{array}{l} S_1 = 120 \\ \text{Option payoff: } 15 \end{array} \\ \begin{array}{l} S_0 = 100 \\ \text{Option value: } ? \end{array} \arrow[ur, "0.6"] \arrow[dr, "0.4"'] & \\ & \begin{array}{l} S_1 = 80 \\ \text{Option payoff: } 0 \end{array} \end{tikzcd}$$

- Given the two tradeable assets, we would like to build a portfolio satisfying:

$$x\, B_1 + y\, S_0 \cdot u = 15$$
$$x\, B_1 + y\, S_0 \cdot d = 0.$$

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example (cont'd)

- Plugging in the numbers for B_t, $S_0 \cdot u$ and $S_0 \cdot d$ gives

$$x \cdot 1.05 + 120y = 15$$
$$x \cdot 1.05 + \ \ 80y = 0.$$

- The solution to this system of equations is

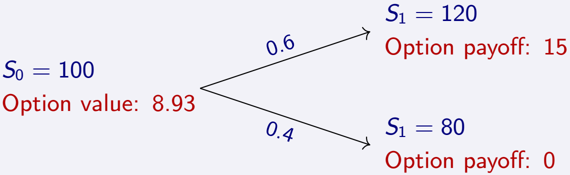
$$x = -\frac{30}{1.05} = -\frac{30}{21/20} = -\frac{200}{7} = -28.57 \quad \text{and} \quad y = \frac{15}{40} = \frac{3}{8} = 0.375.$$

- We have found that the portfolio $h = (-28.57, 0.375)$ **replicates** the option payoff, that is, V_1^h is idential to the option payoff.
- The value of the portfolio today is $V_0^h = -28.57 + 0.375 \cdot 100 = 8.93$.

Replicating a call option

Example (cont'd)

- 8.93 is the **only** price for the option that is consistent with the no-arbitrage principle:
- If the option could be traded at a $\left\{\begin{array}{c}\textbf{higher}\\ \textbf{lower}\end{array}\right\}$ price, then one could create a riskless profit by $\left\{\begin{array}{c}\textbf{selling}\\ \textbf{buying}\end{array}\right\}$ the option and $\left\{\begin{array}{c}\textbf{buying}\\ \textbf{selling}\end{array}\right\}$ the replicating portfolio.
- Hence:



Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

   3.1 Options and option strategies

   3.2 Binomial tree model

      The one-period model

      Portfolios

      Arbitrage

      Replication

      Risk-neutral valuation

      The multiperiod model

      The Cox-Ross-Rubinstein model

   3.3 The Black-Scholes model

   3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Risk-neutral valuation

- We have provided full answers to Questions 1.-3. of the roadmap.
- However, in multiperiod models price calculation by replication portfolios is quite tedious.
- Q. 4: Is there an easier method for price calculation?
- Recall the "no-arbitrage" result:

Proposition (No-arbitrage)

The one-period model is free of arbitrage if and only if the following condition holds:

$$d < (1 + R) < u. \qquad (19)$$

- The Inequalities (19) are equivalent to saying that: there exist $q_u, q_d > 0$, with $q_u + q_d = 1$, such that

$$1 + R = q_u \cdot u + q_d \cdot d.$$

- q_u,q_d can be interpreted as a **probability measure** $\mathbb{Q}$ with

$$\mathbb{Q}(Z = u) = q_u, \qquad \mathbb{Q}(Z = d) = q_d.$$

Risk-neutral valuation

- Denoting expectation wrt. $\mathbb{Q}$ by $\mathbb{E}^{\mathbb{Q}}$ gives the risk-neutral valuation formula:

$$\frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[S_1] = \frac{1}{1+R}\left[q_u S_0\, u + q_d S_0\, d\right] = \frac{1}{1+R}\cdot S_0(1+R) = S_0.$$

Definition

A probability measure $\mathbb{Q}$ is called a **risk-neutral measure** if the following condition holds:

$$S_0 = \frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[S_1].$$

- Risk-neutral measures play a dominant role in pricing derivatives.

Risk-neutral valuation

- We re-state the "no-arbitrage" proposition in a way that is not specific to the one-period model.[15]

Proposition (No-arbitrage, First Fundamental Theorem)

The market model is arbitrage free if and only if there exists a risk-neutral measure $\mathbb{Q}$.

Proposition

For the one-period model, the risk-neutral probabilities are given by

$$q_u = \frac{(1+R)-d}{u-d} \tag{21}$$

$$q_d = 1 - q_u = \frac{u-(1+R)}{u-d}. \tag{22}$$

[15] In continuous-time, existence of a risk-neutral measure $\mathbb{Q}$ is equivalent to the slightly stronger condition "no free lunch with vanishing risk".

Risk-neutral valuation

Proof.

- We have the two conditions

$$q_u \cdot u + q_d \cdot d = 1 + R$$
$$q_u + q_d = 1.$$

- Therefore,

$$q_u \cdot u + (1 - q_u) \cdot d = (1 + R),$$

and the left-hand side is equal to $q_u(u - d) + d$, so that

$$q_u = \frac{(1 + R) - d}{u - d},$$

and

$$q_d = 1 - q_u = \frac{u - d - (1 + R) + d}{u - d} = \frac{u - (1 + R)}{u - d}.$$

□

Risk-neutral valuation

Example (cont'd)

- The risk-neutral probabilities in the example are given by

$$q_u = \frac{1.05 - 0.8}{1.2 - 0.8} = \frac{0.25}{0.4} = \frac{5}{8} = 0.625$$

$$q_d = 1 - q_u = 0.375$$

Risk-neutral valuation

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Proposition

If the binomial model is free of arbitrage, then the arbitrage-free price of a contingent claim X is given by

$$\Pi(0;X) = \frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[X].$$

Proof.

- Since the one-period model is complete, we can price any contingent claim.
- The no-arbitrage price is $\Pi(0;X) = V_0^h$.
- Using the portfolio holdings (x,y) of the replicating portfolio, see (20), gives, after some re-arranging,

$$\Pi(0;X) = x + S_0\, y = \frac{1}{1+R}\left[\frac{(1+R)-d}{u-d}\cdot\Phi(u) + \frac{u-(1+R)}{u-d}\cdot\Phi(d)\right]$$

$$= \frac{1}{1+R}\left[\Phi(u)\cdot q_u + \Phi(d)\cdot q_d\right].$$

□

Risk-neutral valuation

Example (cont'd)

- Consider again the call option with strike price 105, that is, $X = \max(S_1 - 105, 0)$.
- The no-arbitrage price is calculated as

$$\Pi(0, \max(S_1 - 105, 0)) := \frac{1}{1.05}(0.625 \cdot 15 + 0.375 \cdot 0) = 8.93.$$

Risk-neutral probabilities

- Objective probabilities determine which events are possible and which are impossible.
- We compute the arbitrage free price of a financial derivative **as if** we were living in a risk-neutral world.
- This does **not** mean that we believe that we live in a risk-neutral world.
- Rather, investors **do not receive a risk premium** for holding contingent claims that can be entirely risk-managed by replication.
- The valuation formula holds for all investors, regardless of their attitude towards risk (as long as they prefer more money to less).

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
     - The one-period model
     - Portfolios
     - Arbitrage
     - Replication
     - Risk-neutral valuation
     - The multiperiod model
     - The Cox-Ross-Rubinstein model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

The multiperiod model

- Discrete time, with time running from $t = 0$ to $t = T$
- Market consists of two assets, a **bond** and a **stock**
- The **bond price process**, denoted by B_t, is given by

$$B_{t+1} = (1 + R)\, B_t$$
$$B_0 = 1,$$

  with R the interest rate (a simple period rate).
- The dynamics of the **stock process**, denoted by S_t, are

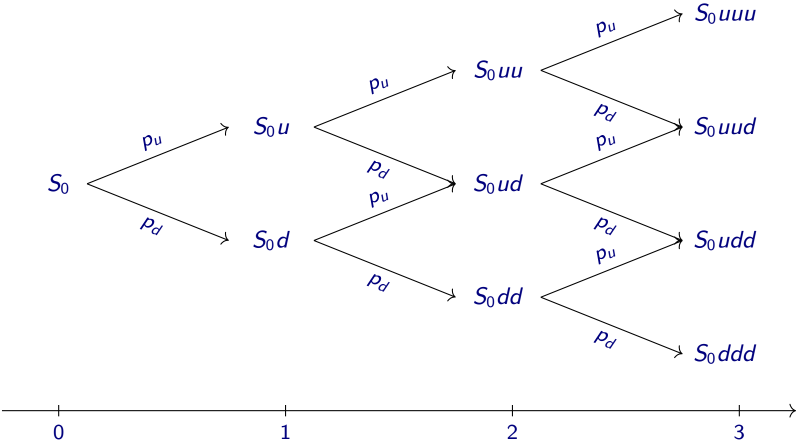
$$S_{t+1} = S_t \cdot Z_t,$$
$$S_0 = s,$$

  where $Z_0, \dots, Z_{T-1}$ are i.i.d. random variables with

$$\mathbb{P}(Z_t = u) = p_u, \qquad \mathbb{P}(Z_t = d) = p_d.$$

The multiperiod model

A three-period binomial tree:



Portfolios

Definition

- A **portfolio strategy** is a stochastic process

$$\{h_t = (x_t, y_t), \quad t = 1, \dots, T\},$$

  such that h_t is a function of $S_0, S_1, \dots, S_{t-1}$.

- The **value process** corresponding to portfolio h is defined by

$$V_t^h = x_t(1+R) + y_t S_t.$$

- x_t **is the amount of money invested at time** $t-1$ **and kept until time** t.
- y_t is the number of shares bought at time $t-1$ and kept until time t.
- At any point in time, the portfolio strategy can depend on all information available at that time...
- ... but of course, one cannot look into the future.

Portfolios

Definition

A portfolio strategy h is said to be **self-financing** if the following condition holds for all $t = 1, \dots, T-1$,

$$x_t(1+R) + y_t S_t = x_{t+1} + y_{t+1} S_t.$$

- This expresses that, at each time t, the market value of the portfolio created at $t-1$ equals the market value of the portfolio created at t.
- In other words: no funds are injected or withdrawn from the portfolio strategy at times $t = 1, \dots, T-1$.
- The portfolio is merely **rebalanced** at every time point, that is, the holdings in the risky asset and risk-free asset are changed subject to keeping the portfolio value constant.

Arbitrage

Definition

An **arbitrage** is a self-financing portfolio h with the following properties:

$$V_0^h = 0$$
$$\mathbb{P}(V_T^h \geq 0) = 1$$
$$\mathbb{P}(V_T^h > 0) > 0.$$

Proposition

The binomial multiperiod model is free of arbitrage if and only if the following condition holds:

$$d < (1+R) < u.$$

Assumption

Henceforth, we assume that $d < (1 + R) < u$.

Definition

The risk-neutral (martingale) probabilities q_u and q_d are defined as the probabilities for which the following relation holds:

$$s = \frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[S_{t+1}|S_t = s].$$

Proposition

The risk-neutral (martingale) probabilities are given by

$$q_u = \frac{(1+R)-d}{u-d}, \qquad q_d = \frac{u-(1+R)}{u-d}.$$

Arbitrage

Proof.

$$
\begin{aligned}
\frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[S_{t+1}|S_t = s] &= \frac{1}{1+R}[q_u \cdot su + q_d \cdot sd] \\
&= \frac{1}{1+R} s \cdot \left[\frac{(1+R)-d}{u-d} \cdot u + \frac{u-(1+R)}{u-d} \cdot d \right] \\
&= \frac{s}{1+R} \cdot \frac{(1+R)u - du + du - (1+R)d}{u-d} \\
&= s.
\end{aligned}
$$

$\square$

Contingent claims

Definition

A **contingent claim** is a random variable X of the form

$$X = \Phi(S_T),$$

where the **contract function** Φ is some given real valued function.

- Note here that we are considering only "simple" claims.
- Path-dependent claims that depend on the price process during $[0, T]$ can be valued as well, but this is a bit more involved.
- Problem: find price process

$$\{\Pi(t; X), t = 0, \dots, T\},$$

for given claim X.

Contingent claims

Definition

- A contingent claim X is **reachable** if there exists a self-financing portfolio h such that
$$V_T^h = X,$$
with probability 1.
- Such a portfolio is called a **hedging portfolio** or **replicating portfolio**.
- If all claims can be replicated, then the market is **complete**.

Contingent claims

- The only reasonable price process for X is given by

$$\Pi(t;X) = V_t^h, \quad t = 0,1,\dots,T.$$

Proposition

Suppose that a claim X is reachable with portfolio h. Suppose furthermore that, at some t, it is possible to buy X at a price cheaper than (or to sell it a price higher than) V_t^h. Then it is possible to make an arbitrage profit.

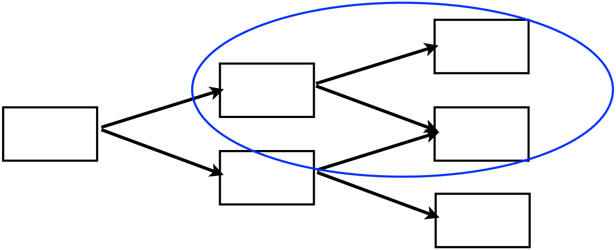
Proposition

The multiperiod binomial model is complete, that is, every claim can be replicated by a self-financing portfolio.

Replication

Idea of proof:

- Break down multiperiod model into one period models



- Price and replicate the claim backwards, step-by-step

Example

Example (Call option)

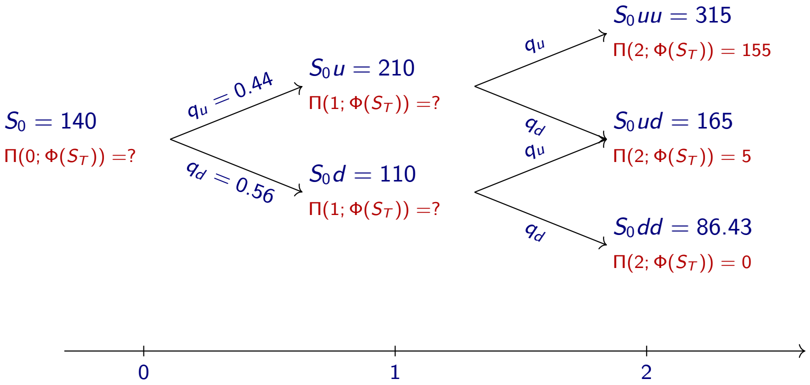
Given:

- spot price $S_0 = 140$
- up and down move factors: $u = 1.5$, $d = 0.78571$
- interest rate $R = 0.1$
- strike $K = 160$
- maturity $T = 2$

First determine:

- Stock price process
- Risk-neutral probabilities: $q_u = 0.44$, $q_d = 0.56$
- Contract function: $\Phi(S_T) = \max(S_T - K, 0)$

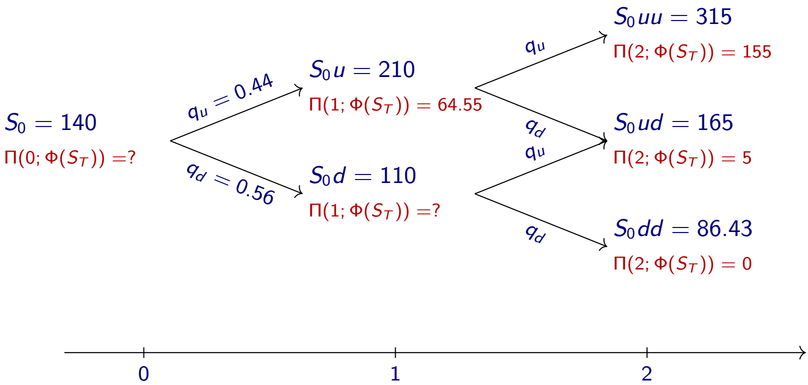
Example


$S_0 = 140$
$\Pi(0; \Phi(S_T)) = ?$
$q_u = 0.44$
$q_d = 0.56$
$S_0 u = 210$
$\Pi(1; \Phi(S_T)) = ?$
$S_0 d = 110$
$\Pi(1; \Phi(S_T)) = ?$
q_u
q_d
q_u
q_d
$S_0 uu = 315$
$\Pi(2; \Phi(S_T)) = 155$
$S_0 ud = 165$
$\Pi(2; \Phi(S_T)) = 5$
$S_0 dd = 86.43$
$\Pi(2; \Phi(S_T)) = 0$
0
1
2

Example

Example (Call option, cont'd)

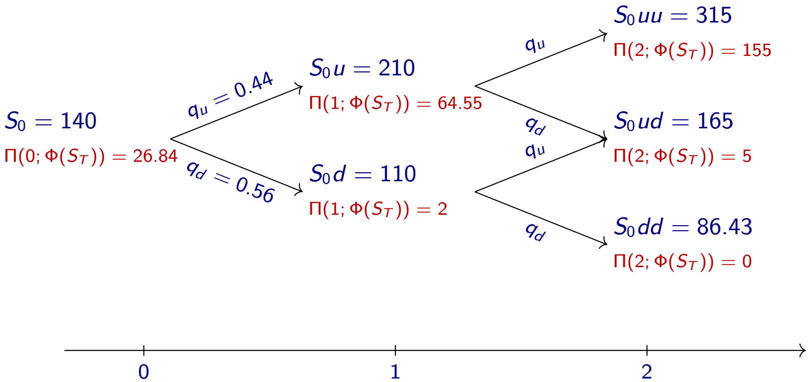
- Next, determine option price at time 1 in the case where $S_1 = S_0 \cdot u$.
- This is just a one period model embedded in the multiperiod model.



Example

Example (Call option, cont'd)

- Repeat for the case when $S_1 = S_0 \cdot d$.
- Given the option values at time 1, calculate option value at time 0, again using the one period binomial tree.



Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example (Call option, cont'd)

- It is easily verified (again by a backward induction argument) that **risk-neutral pricing** holds, yielding

$$
\begin{aligned}
\Pi(0;X) &= \frac{1}{(1+R)^2}\mathbb{E}^{\mathbb{Q}}[\Phi(S_T)] \\
&= \frac{1}{(1+R)^2}\mathbb{E}^{\mathbb{Q}}[q_u^2\Phi(S_0\cdot u^2) + 2q_uq_d\Phi(S_0\cdot ud) + q_d^2\Phi(S_0\cdot d^2)] \\
&= 26.84.
\end{aligned}
$$

- Replication can be achieved – again via backward induction – using the formulas given by (20).
- The replicating portfolio is **self-financing**.

Pricing and replication

- We formalize the replication strategy and pricing formula.
- Let k denote the number of **up-moves** that have occurred.
- Each node in the binomial tree can be represented by a pair (t,k), $t = 0, \dots, T$ and $k = 0, \dots, t$.

Pricing and replication

Proposition

- *An arbitrary claim* $X = \Phi(S_T)$ *can be replicated using a self-financing portfolio. Denote by* $V_t(k)$ *the value of the replicating portfolio at node* (t,k)*, given by*

$$V_t(k) = \frac{1}{1+R}\left[q_u V_{t+1}(k+1) + q_d V_{t+1}(k)\right], \quad k = 0, \dots, T-1$$

$$V_T(k) = \Phi(s\, u^k\, d^{T-k}).$$

- *The risk-neutral probabilities are given by*

$$q_u = \frac{(1+R)-d}{u-d}, \quad q_d = \frac{u-(1+R)}{u-d}.$$

Pricing and replication

Proposition (cont'd)

- *The replicating portfolio is given by*

$$x_t(k) = \frac{1}{1+R} \cdot \frac{uV_t(k) - d\,V_t(k+1)}{u-d}$$

$$y_t(k) = \frac{1}{S_{t-1}} \cdot \frac{V_t(k+1) - V_t(k)}{u-d}.$$

- *In particular, the arbitrage free price of the claim at $t=0$ is given by $V_0(0)$.*

Pricing and replication

Proposition

The arbitrage free price at $t = 0$ of the claim $X = \Phi(S_T)$ is given by

$$\Pi(0; X) = \frac{1}{(1+R)^T} \cdot \mathbb{E}^{\mathbb{Q}}[X],$$

where $\mathbb{Q}$ denotes the risk-neutral measure. More explicitly,

$$\Pi(0; X) = \frac{1}{(1+R)^T} \cdot \sum_{k=0}^{T} \binom{T}{k} q_u^k \, q_d^{T-k} \Phi(su^k \, d^{T-k}).$$

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
     - The one-period model
     - Portfolios
     - Arbitrage
     - Replication
     - Risk-neutral valuation
     - The multiperiod model
     - The Cox-Ross-Rubinstein model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

The Cox-Ross-Rubinstein model

- The **Cox-Ross-Rubinstein model (CRR model)**[16], is a special case of the multiperiod binomial tree model.
- Fix a time interval $[0, T]$ and set $\Delta t = T/N$.
- Trading takes place at times

$$0, \Delta t, 2\Delta t, \dots, (N-1)\Delta t, N\Delta t = T.$$

- The bond price accrues continuously at rate r, so that $B_t = B_0\, e^{rt}$ and $B_{n\Delta t} = B_{(n-1)\Delta t}\, e^{r\Delta t}$, where $n = 1, \dots, N$.
- The risk of the stock price is expressed by the **volatility** σ, the standard deviation of the stock price's log-return.
- The volatility for the time period Δt is $\sigma\sqrt{\Delta t}$, yielding

$$u = e^{\sigma\sqrt{\Delta t}}, \qquad d = 1/u = e^{-\sigma\sqrt{\Delta t}}.$$

[16] Cox, John C., Ross, Stephen A. and Rubinstein, Mark. (1979). "Option Pricing: A Simplified Approach." Journal of Financial Economics 7: 229-263.

The CRR model

- The stock price at time $N\Delta t$ is

$$S_{N\Delta t} = S_0 \, e^{\sigma\sqrt{\Delta t}(N_u - N_d)},$$

where

  - N_u is the number of up moves,
  - $N_d = N - N_u$ the number of down moves.

- The risk-neutral probabilities are, cf. Equations (21)-(22),

$$q_u = \frac{e^{r\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}$$

$$q_d = 1 - q_u = \frac{e^{\sigma\sqrt{\Delta t}} - e^{r\Delta t}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}$$

The CRR model

- According to the risk-neutral pricing formula the value of a claim maturing at $T = N\,\Delta t$ is given by

$$
\begin{aligned}
V_0 &= e^{-rT}\,\mathbb{E}(V_T) \\
&= e^{-rT} \sum_{k=0}^{N} V_T\left(S_0\, e^{\sigma\sqrt{\Delta t}(N_u - N_d)}\right)\,\mathbb{Q}(N_u = k) \\
&= e^{-rT} \sum_{k=0}^{N} V_T\left(S_0\, e^{\sigma\sqrt{\Delta t}(2k - N)}\right)\,\mathbb{Q}(N_u = k) \\
&= e^{-rT} \sum_{k=0}^{N} V_T\left(S_0\, e^{\sigma\sqrt{\Delta t}(2k - N)}\right)\,\binom{N}{k}\, q_u^k\, q_d^{N-k}.
\end{aligned}
$$

The CRR model

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example (Call option)

Let $V_T = \max(S_T - K,0) = (S_T - K)^+$. Denote by a the smallest integer k such that $S_T = S_0\, e^{\sigma\sqrt{\Delta t}(2k-N)} \geq K$. Then,

$$V_0 = e^{-rT}\sum_{k=a}^{N}\left[S_0\, e^{\sigma\sqrt{\Delta t}(2k-N)} - K\right]\binom{N}{k} q_u^k q_d^{N-k}$$

$$= S_0 \sum_{k=a}^{N} e^{-rT+\sigma\sqrt{\Delta t}(2k-N)}\binom{N}{k} q_u^k q_d^{N-k} - e^{-rT} K \sum_{k=a}^{N}\binom{N}{k} q_u^k q_d^{N-k}$$

$$= S_0 \Phi(a; N,\overline{q}_u) - e^{-rT} K\, \Phi(a; N,q_u),$$

where $\Phi(a; N,q_u) := \sum_{k=a}^{N}\binom{N}{k} q_u^k (1-q_u)^{N-k}$ and $\overline{q}_u := q_u\, e^{\sigma\sqrt{\Delta t}}\, e^{-r\Delta T}$.

- Recall that $q_d = 1 - q_u$.

Exercises

Exercise

Assume a one period model with $S_0 = 50$ and $S_T \in \{48,53\}$, where $T = 2$ months. The interest rate is 6% p.a. Determine the price of a call option with strike $K = 50$. What is the price of the put option with the same parameters? How are the call and the put option prices related?

Exercise

In a multiperiod binomial model let $S_0 = 100$ and $u = 1.2$, $d = 0.8$ with probabilities $p_u = 3/4$ and $p_d = 1/4$, respectively. Set $R = 0$. Calculate the price and replication strategy for a digital contract that pays off 100 is the stock ends higher than it started at $T = 3$.

Exercises

Exercise (Path-dependent options)

Assume a CRR model with monthly price jumps. Today's stock price is $S_0 = 100$ and assume an annualised volatility of 33%. Let $r = 5\%$ p.a.

- Determine price of an up-and out barrier option with strike $K = 105$, maturity $T = 3$ and barrier $B = 115$. (An up-and-out barrier pays $(S_T - K)^+$ at maturity if the barrier has not been touched during the lifetime of the option.)
- Determine the price of a Bermudan put option with maturity $T = 5$ and strike $K = 100$.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

3.1 Options and option strategies

3.2 Binomial tree model

3.3 The Black-Scholes model

Brownian motion

Martingales

The market model

Black-Scholes formula from a PDE

Risk-neutral valuation

No-arbitrage and market completeness

Black-Scholes formula as limit in the CRR ($\star$)

3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Motivation

- Let $S_0, S_1, \dots, S_n$ be the prices of a stock observed at the end of days $0, 1, \dots, n$.
- The log-return over one day is $r_k = \ln\left(\frac{S_k}{S_{k-1}}\right)$.
- Log-returns are convenient to work with, because the return over a longer time period is the sum of consecutive intra-period returns:

$$\ln\left(\frac{S_n}{S_0}\right) = \ln\left(\frac{S_n}{S_{n-1}}\right) + \ln\left(\frac{S_{n-1}}{S_{n-2}}\right) + \cdots + \ln\left(\frac{S_1}{S_0}\right)$$

$$= r_n + r_{n-1} + \cdots + r_1.$$

- In turn, each daily log return can be interpreted as the sum of hourly log returns, etc.

Motivation

- Under the assumptions that
  - log returns over disjoint time intervals are **stochastically independent**,
  - log returns over equidistant time intervals are **identically distributed**,

  the **Central Limit Theorem** implies that log returns are (close to) **normally distributed**.
- This is due to the propery of log returns of being the sum of many small independent and identially distributed (iid) random variables with finite variance.
- This property is captured by the **Black-Scholes(-Merton) model**: it is a continuous-time market model where log-returns over arbitrary time intervals are normally distributed.

Motivation

	Discrete time Binomial Model	Continuous time Model
uncertainty	binomial model with up- and down-moves	**Brownian motion**
description of stock price	give stock price in every node of the tree	**stochastic differential equation**
model	Cox/Ross/Rubinstein $S_{t+1} = \begin{cases} S_t\, u & \text{up} \\ S_t\, d & \text{down} \end{cases}$	**Black-Scholes** **Geometric Brownian motion**
pricing derivatives	replication risk-neutral pricing	replication risk-neutral pricing

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
     - Brownian motion
     - Martingales
     - The market model
     - Black-Scholes formula from a PDE
     - Risk-neutral valuation
     - No-arbitrage and market completeness
     - Black-Scholes formula as limit in the CRR ($\star$)
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Stochastic process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition (Stochastic process)

A **stochastic process** $X = (X_t)_{t \in \mathbb{T}}$ is a set of random variables indexed by $t \in \mathbb{T}$ ("time")

- $\mathbb{T} = \{0, t_1, t_2, \ldots, t_N\}$ for a stochastic process in discrete time
- $\mathbb{T} = [0, T]$ for a stochastic process in continuous time

- Interpretation:
  - for each point in time t: X_t is a random variable
  - each realization $X(\omega) = (X_t(\omega))_{t \in \mathbb{T}}$, is a path, where $\omega \in \Omega$

Brownian motion

Definition

The process $W = (W_t)_{t \geq 0}$ is a **Brownian motion** if

(i) $W_0 = 0$ and W has continuous paths,

(ii) W has independent increments, that is, for $r < s \leq t < u$, the random variables $W_u - W_t$ and $W_s - W_r$ are independent,

(iii) For $s < t$, $W_t - W_s \sim \mathrm{N}(0, t - s)$, that is, the random variable $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$.

- W is a Brownian motion with respect to the measure $\mathbb{P}$.
- Brownian motion is also often called **Wiener process**.

Brownian motion

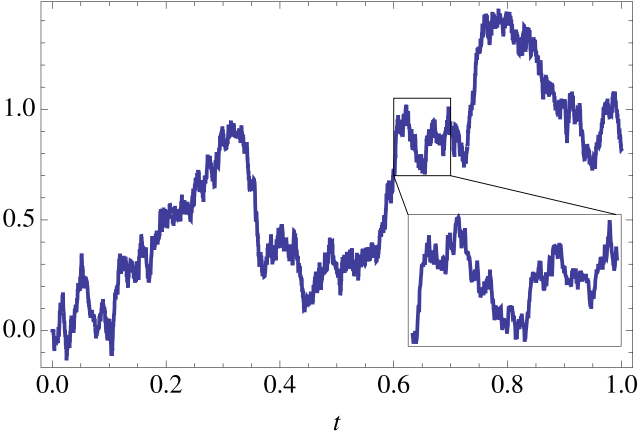


Figure: Path of a Brownian motion

Geometric Brownian motion

- Continuous compounding at interest rate $r > 0$ of a **bond** or **money market account** is governed by the **differential equation**:

$$\mathrm{d}B_t = B_t r \,\mathrm{d}t, \quad t \geq 0.$$

- The solution to the differential equation is given by

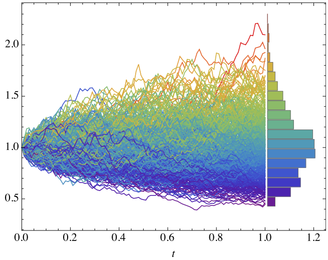
$$B_t = B_0 \exp(rt), \quad t \geq 0.$$

- Similarly, the **Black-Scholes model** specifies the dynamics of a stock price as

$$\mathrm{d}S_t = S_t \,\mu \,\mathrm{d}t + S_t \,\sigma \,\mathrm{d}W_t.$$

- The solution of the differential equation turns out to be

$$S_t = S_0 \,\exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right).$$



Example paths with histogram at $t = 1$.

- Brownian motion is "odd":
  - Brownian motion is continuous, but **nowhere differentiable**, so "$\frac{dW_t}{dt}$" is not defined.
  - Brownian motion is a **fractal** (that is, W is of **unbounded variation**), implying in particular that the integral "$\int_0^T dW_t$" cannot be defined in the usual Riemann style.

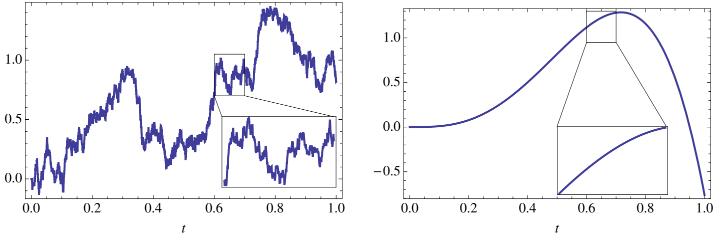


Figure: Brownian motion path vs. smooth function

Stochastic calculus, Itô's formula

- One of the most important results of stochastic calculus is a "Taylor expansion" for Brownian motion.
- **Itô's Lemma / Itô's formula** specifies the dynamics of stochastic processes.

Theorem (Itô's Lemma for Brownian motion)

Let $W = (W_t)_{t \geq 0}$ be a Brownian motion and let f be a deterministic, twice continuously differentiable function. Then,

$$\mathrm{d}f(W_t) = f'(W_t)\,\mathrm{d}W_t + \frac{1}{2}f''(W_t)\,\mathrm{d}t.$$

Example

- Let's determine the dynamics of $f(W_t) = W_t^2$.
- By the Itô formula:

$$dW_t^2 = 2W_t\,dW_t + \frac{1}{2}\cdot 2\,dt,$$

or, in integral form,

$$\begin{aligned} W_t^2 &= W_0^2 + \int_0^t 2W_s\,dW_s + \frac{1}{2}\int_0^t 2\,ds \\ &= \int_0^t 2W_s\,dW_s + t. \end{aligned}$$

Theorem (Itô's Lemma for SDE's)

Let X be a stochastic process satisfying

$$\mathrm{d}X_t = \sigma(t,X_t)\,\mathrm{d}W_t + \mu(t,X_t)\,\mathrm{d}t,$$

and let $f(t,x)$ be a deterministic function, continuously differentiable in t and twice continuously differentiable function in x. Denote the first and second partial derivatives by f_t, f_x and f_{xx}, respectively. Then,

$$\begin{aligned}\mathrm{d}f(t,X_t) &= f_t(t,X_t)\,\mathrm{d}t + f_x(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}f_{xx}(t,X_t)\,\mathrm{d}[X,X]_t \\ &= \left[f_t(t,X_t) + f_x(t,X_t)\mu(t,X_t) + \frac{1}{2}f_{xx}(t,X_t)\sigma^2(t,X_t)\right]\mathrm{d}t \\ &\quad + f_x(t,X_t)\sigma(t,X_t)\,\mathrm{d}W_t,\end{aligned}$$

where $\int \mathrm{d}[X,X]_t = \int \sigma^2(t,X_t)\,\mathrm{d}t$ is the ***quadratic variation*** *of X.*

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

   3.1 Options and option strategies

   3.2 Binomial tree model

   3.3 The Black-Scholes model

      Brownian motion

      Martingales

      The market model

      Black-Scholes formula from a PDE

      Risk-neutral valuation

      No-arbitrage and market completeness

      Black-Scholes formula as limit in the CRR ($\star$)

   3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Martingales

- The modern theory of financial derivatives is closely connected to special stochastic processes, called **martingales**.
- In fact, **risk-neutral pricing** is based on martingale theory.
- First, we need the concept of **information flow** or **filtration** $(\mathcal{F}_t)_{t \geq 0}$.
- $\mathcal{F}_t$ describes the information generated by all events that are observed or revealed until time t.
- For example, $\mathcal{F}_t$ can be generated by the stock price process, containing the information about the stock price path until time t.

Martingales

Definition

A stochastic process X is called an $(\mathcal{F}_t)_{t\geq 0}$-**martingale** if the following conditions hold:

- X is **adapted** to the filtration $(\mathcal{F}_t)_{t\geq 0}$, that is, $(X_s)_{s\leq t}$ is completely determined from $\mathcal{F}_t$, $t \geq 0$;
- For all t
$$\mathbb{E}[|X_t|] < \infty;$$
- For all s and t, with $s \leq t$, the following relation holds:
$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s.$$

Proposition

Under some integrability conditions on the function g, the following holds:

(i)

$$\mathbb{E}\left[\int_s^t g(u)\,\mathrm{d}W_u \middle| \mathcal{F}_s^W\right] = 0,$$

where $(\mathcal{F}_t^W)$ is the information generated by the Brownian motion W.

(ii) *The process X, defined by*

$$X_t = \int_0^t g(s)\,\mathrm{d}W_s$$

is an $(\mathcal{F}_t^W)$-martingale.

Martingales

Proposition

Under some integrability conditions on the function g, the following holds:

(i)

$$\mathbb{E}\left[\int_s^t g(u)\,\mathrm{d}W_u \middle| \mathcal{F}_s^W\right] = 0,$$

where $(\mathcal{F}_t^W)$ is the information generated by the Brownian motion W.

(ii) *The process X, defined by*

$$X_t = \int_0^t g(s)\,\mathrm{d}W_s$$

is an $(\mathcal{F}_t^W)$-martingale.

- "Every stochastic integral is a martingale."

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
     - Brownian motion
     - Martingales
     - The market model
     - Black-Scholes formula from a PDE
     - Risk-neutral valuation
     - No-arbitrage and market completeness
     - Black-Scholes formula as limit in the CRR ($\star$)
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Black-Scholes(-Merton) model

- We assume a frictionless[17] financial market with a bond and a stock.
- In the **Black-Scholes(-Merton) model** the dynamics of the bond price and the stock price are given by

$$dB_t = rB_t\, dt \tag{23}$$

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t, \tag{24}$$

where
- r, μ, σ are deterministic constants,
- $W = (W_t)_{t \geq 0}$ is a Brownian motion (under the real-world probability measure $\mathbb{P}$),
- r is the **risk free rate**,
- μ is the **drift**,
- σ is the **volatility**.

[17] Assets are liquidly traded, there are no short-selling constraints, there are no transaction costs and there are no bid-ask-spreads.

Black-Scholes(-Merton) model

- Applying the the Itô formula to $f(t,S_t) = \ln S_t$ gives

$$\mathrm{d} \ln S_t = \left(\mu - \frac{1}{2}\sigma^2\right) \mathrm{d}t + \sigma \, \mathrm{d}W_t .$$

- Together with $W_0 = 0$,

$$\ln\left(\frac{S_t}{S_0}\right) = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t ,$$

and

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right), \quad t > 0. \tag{25}$$

Portfolios

- A **portfolio strategy** is a 2-dimensional process $h(t) = (h_1(t), h_2(t))$ that depends on (t, S_t), $t \geq 0$.
- The **value process** $V^h(t)$ corresponding to portfolio h is given by

$$V^h(t) = h_1(t)B_t + h_2(t)S_t.$$

- The portfolio strategy h is **self-financing** on $[0, T]$, if V^h satisfies the condition

$$V^h(t) = V^h(0) + \int_0^t h_1(s)\,\mathrm{d}B_s + \int_0^t h_2(s)\,\mathrm{d}S_s, \quad t \leq T,$$

or, in differential form,

$$\mathrm{d}V^h(t) = h_1(t)\,\mathrm{d}B_t + h_2(t)\,\mathrm{d}S_t, \quad t \leq T.$$

- This expresses that money / cash is neither injected not taken out of the strategy.

Portfolios

- More generally, we could make the strategy depend on the whole price path, that is, h is an $(\mathcal{F}_t^S)_{t\geq 0}$-adapted process.
- However, most of the time we shall consider portfolios that depend only on (t, S_t).

Arbitrage

Definition

An **arbitrage** is a self-financing portfolio h with properties

$$V^h(0) = 0,$$
$$\mathbb{P}(V^h(T) \geq 0) = 1,$$
$$\mathbb{P}(V^h(T) > 0) > 0.$$

A market is **arbitrage-free** if there are no arbitrage opportunities.

Contingent claims

- Problem: For a contingent claim $X = \Phi(S_T)$, find price process $\Pi(t; X)$
- The following result states that under absence of arbitrage a riskless portfolio earns the risk-free rate.

Lemma

Let h be a self-financing portfolio with dynamics

$$\mathrm{d}V^h(t) = k_t V^h(t)\,\mathrm{d}t,$$

where k is an adapted process. If the market is free of arbitrage, then $k_t = r$, for all $t \geq 0$.

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
     - Brownian motion
     - Martingales
     - The market model
     - Black-Scholes formula from a PDE
     - Risk-neutral valuation
     - No-arbitrage and market completeness
     - Black-Scholes formula as limit in the CRR ($\star$)
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Black-Scholes formula from a PDE

Here are the steps to derive the **Black-Scholes formula**, giving the price process of European call options:

- We describe the dynamics of a portfolio consisting of a contingent claim and its replicating strategy.
- Because the portfolio is riskless, ...
  - ... the dynamics are described by a **partial differential equation (PDE)** instead of an SDE,
  - ... the portfolio must earn the risk-free rate.
- For a call option payoff, the PDE can be solved explicitly, giving the Black-Scholes formula.

Black-Scholes formula from a PDE

Assumption

(i) The market is free of arbitrage.

(ii) The price process for the derivative X is of the form

$$\Pi(t;X) = F(t,S_t),$$

where F is a smooth function.

Black-Scholes formula from a PDE

Assumption

(i) The market is free of arbitrage.

(ii) The price process for the derivative X is of the form

$$\Pi(t; X) = F(t, S_t),$$

where F is a smooth function.

Let's go...

- At time t, consider a portfolio π_t consisting of
  - short the claim X, maturing at time T, with price $F(t,S_t)$,
  - Δ_t units of the stock,
  - Q_t units of the bond.

  (At this stage, $F(t,S_t)$, Δ_t and Q_t are unknown.)

Black-Scholes formula from a PDE

- At time t, consider a portfolio π_t consisting of
  - short the claim X, maturing at time T, with price $F(t,S_t)$,
  - Δ_t units of the stock,
  - Q_t units of the bond.

  (At this stage, $F(t,S_t)$, Δ_t and Q_t are unknown.)
- Choosing $Q_0 = F(0,S_0) - \Delta_0 S_0$ implies $\pi_0 = 0$.

Black-Scholes formula from a PDE

- At time t, consider a portfolio π_t consisting of
  - short the claim X, maturing at time T, with price $F(t,S_t)$,
  - Δ_t units of the stock,
  - Q_t units of the bond.

  (At this stage, $F(t,S_t)$, Δ_t and Q_t are unknown.)
- Choosing $Q_0 = F(0,S_0) - \Delta_0 S_0$ implies $\pi_0 = 0$.
- The objective is to choose Δ_t and Q_t, $t \geq 0$, such that $\pi = (\pi_t)_{t\in[0,T]}$ consists of the contingent claim and the self-financing replicating portfolio.

Black-Scholes formula from a PDE

- The dynamics of the contingent claim's price process are (Itô formula)

$$\mathrm{d}F(t,S_t) = \left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2 \right) \mathrm{d}t + \frac{\partial F}{\partial S}\,\mathrm{d}S_t.$$

Black-Scholes formula from a PDE

- The dynamics of the contingent claim's price process are (Itô formula)

$$\mathrm{d}F(t,S_t) = \left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2\right)\mathrm{d}t + \frac{\partial F}{\partial S}\,\mathrm{d}S_t.$$

- The dynamics of the portfolio are given by

$$\begin{aligned}
\mathrm{d}\pi_t &= \Delta_t\,\mathrm{d}S_t + Q_t r\,\mathrm{d}t - \mathrm{d}F(t,S_t) \\
&= \left(\Delta_t - \frac{\partial F}{\partial S}\right)\mathrm{d}S_t \\
&\quad + \left(Q_t r - \frac{\partial F}{\partial t} - \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2\right)\mathrm{d}t.
\end{aligned}$$

Black-Scholes formula from a PDE

- Setting $\Delta_t = \dfrac{\partial F}{\partial S}$ eliminates the $\mathrm{d}S_t$-term, that is, the risk.
- This gives a risk-free portfolio with $\pi_0 = 0$.
- To make the portfolio self-financing and assuming no-arbitrage requires

$$\pi_t = 0, \quad t \in [0, T].$$

- This implies

$$Q_t = F(t, S_t) - \Delta_t S_t, \quad t \in [0, T].$$

  (Any surplus or shortage from holding the option and making the portfolio risk-free is put / taken from the bank account).

Black-Scholes formula from a PDE

- Because $\mathrm{d}\pi_t = 0$, this gives

$$\left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2\right)\mathrm{d}t = r\left(F(S_t,t) - \frac{\partial F}{\partial S}\cdot S_t\right)\mathrm{d}t.$$

- Re-arranging gives

$$\boxed{\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2 + rS_t\frac{\partial F}{\partial S} - rF = 0.} \tag{26}$$

- This is a linear parabolic **partial differential equation (PDE)**, and is often referred to as the Black-Scholes differential equation or Black-Scholes PDE.

Black-Scholes formula from a PDE

- The payoff structure has not yet entered the calculation, so the Black-Scholes PDE is satisfied for arbitrary European-style derivatives.
- The concrete payoff is reflected in the so-called **boundary conditions** at $t = T$, given by
$$F(T, S_T) = \Phi(S_T), \quad S_T > 0.$$

Example (Call option)

For a European call option with maturity T and strike K, the boundary conditions are

- at T:

$$\Pi(T;X) = F(T,S_T) = (S_T - K)^+, \tag{27}$$

- for any $t \in [0,T]$:

$$F(t,0) = 0, \qquad \lim_{S_t \to \infty} F(t,S_t)/S_t = 1.$$

Black-Scholes formula from a PDE

- The following two Theorems summarise our findings.

Theorem (Black-Scholes PDE)

Suppose the market follows a Black-Scholes model, Equations (23)-(24), and let $X = \Phi(S_T)$ be a European contingent claim.

The only pricing function $\Pi(t; X) = F(t,S_t)$ that is consistent with absence of arbitrage is when F is the solution of the following boundary value problem in the domain $[0,T] \times \mathbb{R}_+$:

$$\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial s^2}\sigma^2 s^2 + rs\frac{\partial F}{\partial s} - rF = 0$$

$$F(T,s) = \Phi(s).$$

Theorem (Replication)

Any European contingent claim $X = \Phi(S_T)$ *with (smooth) pricing function* $\Pi(t; X) = F(t, S_t)$, $0 \leq t < T$, *can be replicated by a self-financing portfolio strategy consisting at time* t *of*

- $\Delta_t = \dfrac{\partial F}{\partial S}$ *units of stock,*
- $F(t, S_t) - \Delta_t S_t$ *units of the bond.*

Theorem (Black-Scholes formula)

In the Black-Scholes model, the time-t price of a European call option with parameters K, T, r, σ and S_t is given by

$$\begin{aligned}\Pi(t;(S_T-K)^+) &= F(t,S_t)\\ &= S_t\,\mathrm{N}(d_{t,+}) - e^{-r(T-t)}K\,\mathrm{N}(d_{t,-}),\end{aligned}$$

with

$$d_{t,\pm} = \frac{\ln(S_t/K) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}},$$

where $\mathrm{N}(x)$ is the cumulative distribution function of the standard normal distribution.

Black-Scholes formula from PDE

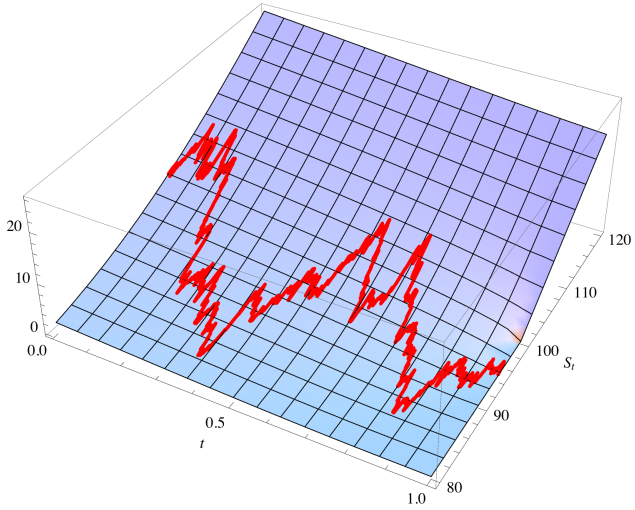


Figure: Black-Scholes formula as a function of time t and the stock price S_t. The red line shows one realisation of the price path.

Black-Scholes formula from PDE

Proof.

- We show that the Black-Scholes formula is indeed a solution to the Black-Scholes PDE (26) with boundary condition (27).
- The sensitivities theta, gamma and delta are given by

$$\frac{\partial F}{\partial t} = -\frac{S_t\,\mathrm{n}(d_{t,+})\,\sigma}{2\sqrt{T-t}} - r\,Ke^{-r(T-t)}\mathrm{N}(d_{t,-})$$

$$\frac{\partial F}{\partial S} = \mathrm{N}(d_{t,+})$$

$$\frac{\partial^2 F}{\partial S^2} = \frac{\mathrm{n}(d_{t,+})}{S_t\sigma\sqrt{T-t}},$$

with $\mathrm{n}(x)$ the standard normal density.

- It is easily verified that the Black-Scholes PDE is satisfied.

[...]

Proof (cont'd).

- For the boundary condition observe that

$$\ln(S_T/K) = \ln(S_T) - \ln(K) \begin{cases} > 0, & \text{if } S_T > K, \\ = 0, & \text{if } S_T = K, \\ < 0, & \text{if } S_T < K. \end{cases}$$

Hence,

$$\lim_{t \to T} d_\pm = \lim_{t \to T} \frac{\ln(S_t/K)}{\sigma\sqrt{T-t}} + \underbrace{\lim_{t \to T} \left(r \pm \frac{1}{2}\sigma^2\right)\sqrt{T-t}}_{=0}$$

$$= \begin{cases} +\infty, & \text{if } S_T > K, \\ 0 & \text{if } S_T = K, \\ -\infty & \text{if } S_T < K. \end{cases}$$

Proof (cont'd).

- Hence,

$$\lim_{t\to T} \mathrm{N}(d_{t,+}) = \lim_{t\to T} \mathrm{N}(d_{t,-}) = \mathbf{1}_{\{S_T>K\}} + \frac{1}{2}\mathbf{1}_{\{S_T=K\}}$$

and

$$F(T,S_T) = S_T\mathrm{N}(d_{T,+}) - K\mathrm{N}(d_{T,-}) = (S_T - K)^+.$$

□

Exercise

Show that the Black-Scholes price of a binary option (digital option) with payoff $\mathbf{1}_{\{S_T > K\}}$ is given by

$$e^{-r(T-t)}\mathrm{N}(d_{t,-}), \quad 0 \le t \le T.$$

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

    3.1 Options and option strategies

    3.2 Binomial tree model

    3.3 The Black-Scholes model

        Brownian motion
        Martingales
        The market model
        Black-Scholes formula from a PDE
        Risk-neutral valuation
        No-arbitrage and market completeness
        Black-Scholes formula as limit in the CRR ($\star$)

    3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Risk-neutral valuation

- The Black-Scholes PDE opens the door to a fundamental understanding of pricing under no-arbitrage.
- Concrete price determination can be challenging, though:
  - It is straightforward to **verify** that a given price process satisfies the Black-Scholes PDE, ...
  - ... but it may be tricky to come up with a solution given a boundary condition.
- Risk-neutral valuation is a different approach to the pricing problem, giving rise to simulation methods for practical valuation.

Risk-neutral valuation: the discrete case

- In the discrete setup, we derived a **risk-neutral probability measure** $\mathbb{Q}$ given by **risk-neutral probabilities** q_u, q_d.
- This was characterised by the fact that discounted asset prices are martingales, that is,
$$\frac{1}{1+R}\mathbb{E}^{\mathbb{Q}}[S_{t+1}|S_t] = S_t.$$
- The no-arbitrage price of a contingent claim was given by the **discounted risk-neutral expected payoff**.
- Can we generalise this to the continuous-time setting?
- The answer is positive.
- Our main tool will be the **Girsanov-Theorem**, which links the drift of a process with the underlying probability measure.

Risk-neutral probability measure

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- Assume a Black-Scholes model, as in Equations (23)-(24):

$$\begin{aligned} \mathrm{d}B_t &= rB_t \,\mathrm{d}t \\ \mathrm{d}S_t &= \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t. \end{aligned}$$

- Here $W = (W_t)_{t \geq 0}$ is a Brownian motion under the probability measure $\mathbb{P}$, that is,

$$\mathbb{P}(W_t \leq x) = \mathrm{N}(x; t),$$

where $\mathrm{N}(\cdot; t)$ is the cumulative normal distribution function with variance t.

Risk-neutral probability measure

- One can easily show that under $\mathbb{P}$, the market does not have the martingale property.
- The goal is to find a probability measure such that the process (24) can be written as

$$dS_t = rS_t\,dt + \sigma S_t\,d\overline{W}_t, \tag{28}$$

  where $\overline{W}$ is a Brownian motion.
- It is important to note that (24) and (28) refer to one and the same process!

Risk-neutral probability measure

- Recall that the process S with dynamics given by (24) is a Geometric Brownian motion.

- Re-write this as follows:

$$\begin{aligned} S_t &= S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right) \\ &= S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)t + (\mu - r)t + \sigma W_t\right) \\ &= S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)t + \sigma \underbrace{\left(\frac{\mu - r}{\sigma}t + W_t\right)}_{=:\,\overline{W}_t}\right). \end{aligned}$$

- Under $\mathbb{P}$, the process $\overline{W}$ is a Brownian motion with drift.

- Using the Girsanov Theorem one can derive a probability measure $\mathbb{Q}$ that makes $\overline{W}$ a Brownian motion without drift.

Risk-neutral probability measure

- Loosely speaking, the **Girsanov Theorem** states that a change of measure adjusts the drift of a Brownian motion.
- Moreover, the measures $\mathbb{P}$ and $\mathbb{Q}$ are **equivalent**, i.e., $\mathbb{P}$ and $\mathbb{Q}$ agree on which events have probability 0 and 1. This is important:
  - It ensures that no information is lost in the measure change, as all "possible" events under $\mathbb{P}$ (that is, event with probability greater than zero) are possible under $\mathbb{Q}$.
  - The measure $\mathbb{P}$ and $\mathbb{Q}$ agree on the sets of arbitrage strategies and replication strategies.

Risk-neutral probability measure

- To summarise:
  - The discounted stock price process is not a martingale under the real-world measure $\mathbb{P}$.
  - With the Girsanov-Theorem, we can determine a measure $\mathbb{Q}$ under which the discounted stock price process is a martingale.
- In the following, we shall use the measure $\mathbb{Q}$ to value contingent claims.

Risk-neutral valuation

- Recall that for risk-neutral valuation it is necessary that the measures $\mathbb{P}$ and $\mathbb{Q}$ **agree** on which events have probability 1 and which events have probability 0, that is, for an arbitrary event $A \in \mathcal{F}$,

$$\mathbb{P}(A) = 1 \iff \mathbb{Q}(A) = 1$$
$$\mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0.$$

- We say that the measures $\mathbb{P}$ and $\mathbb{Q}$ are **equivalent**.
- If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, then
  - there are no arbitrage opportunities under $\mathbb{P}$ if and only if there are no arbitrage opportunities under $\mathbb{Q}$;
  - a replicating strategy under $\mathbb{Q}$ is a replicating strategy under $\mathbb{P}$.
- $\mathbb{Q}$ is called an equivalent martingale measure.

Risk-neutral valuation

Proposition

The arbitrage-free price process of the claim $X = \Phi(S_T)$ *is given by* $\Pi(t;X) = F(t,S_t)$, *where* F *is given by*

$$F(t,s) = e^{-r(T-t)}\,\mathbb{E}^{\mathbb{Q}}[\Phi(S_T)|S_t = s], \quad t \geq 0.$$

- To motivate the result, observe that by construction $\mathbb{E}^{\mathbb{Q}}[\Phi(S_T)|S_t]$ is a martingale, since for $s < t$,

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\Phi(S_T)|S_t]|S_s\right] \overset{\text{tower law}}{=} \mathbb{E}^{\mathbb{Q}}[\Phi(S_T)|S_s].$$

- Hence, if we were to add the contingent claim as another asset, then its price process is again a martingale.

Risk-neutral valuation

- The proof uses the **Martingale representation theorem**, which we state here in simplified form.
- In the following, let $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by the Brownian motion W (i.e., the **information** generated by W).
- A process X is said to be $(\mathcal{F}_t)_{t\geq 0}$-**adapted** if $(X_s)_{s\leq t}$ is determined by $\mathcal{F}_t$.
- **Important:** In-line with the standard literature, from now on $W = (W_t)_{t\geq 0}$ refers to the Brownian motion under the risk-neutral measure (i.e., W_t corresponds to $\overline{W}_t$ from the previous slides).

Theorem (Martingale representation theorem)

Let $M = (M_t)_{t \geq 0}$ *be an adapted continuous martingale. Then there exists a process* $H = (H_t)_{t \geq 0}$ *with* $\int_0^T H_s^2 \, \mathrm{d}s < \infty$ *such that*

$$M_t = M_0 + \int_0^t H_s \, \mathrm{d}W_s, \quad t \leq T.$$

Furthermore, if Y *is a random variable whose value is determined by* $\mathcal{F}_T$*, that is,* $Y = F(W_T)$ *for some function* F*, and if* $\mathbb{E}|Y| < \infty$*, then there exists an adapted process* H *such that*

$$Y = \mathbb{E}Y + \int_0^T H_s \, \mathrm{d}W_s.$$

- We derive the Black-Scholes formula by risk-neutral pricing.

Theorem (Black-Scholes formula)

In the Black-Scholes model, the time-t price of a European call option with parameters K, T, r, σ and S_0 is given by

$$\Pi(t;(S_T-K)^+) = F(t,S_t) = S_t\,\mathrm{N}(d_{t,+}) - e^{-r(T-t)}K\,\mathrm{N}(d_{t,-}), \tag{29}$$

with

$$d_{t,\pm} = \frac{\ln(S_t/K) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}},$$

where $\mathrm{N}(x)$ is the cumulative distribution function of the standard normal distribution.

Proof.

The call option payoff can be separated into an **Asset-or-Nothing** option with payoff $S_T \, \mathbf{1}_{\{S_T > K\}}$ and a **Money-or-Nothing** option with payoff $K \, \mathbf{1}_{\{S_T > K\}}$.

[...]

Proof (cont'd).

A **Money-or-Nothing option** with strike K pays K at maturity T if $S_T - K \geq 0$. By the principle of risk-neutral pricing, the no-arbitrage price of the option at time 0 is

$$\mathbb{E}^{\mathbb{Q}}[e^{-rT}K \cdot \mathbf{1}_{\{S_T>K\}}] = e^{-rT}K\,\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{S_T>K\}}] = e^{-rT}K\,\mathbb{Q}(S_T > K).$$

We show that $\mathbb{Q}(S_T > K) = \mathrm{N}(d_{0,-})$. [...]

Proof (cont'd).

$$
\begin{aligned}
\mathbb{Q}(S_T > K) &= \mathbb{Q}\left(S_0 \cdot \exp\left(\left(r - \frac{1}{2}\sigma^2\right) T + \sigma W_T\right) > K\right) \\
&= \mathbb{Q}\left(\underbrace{\frac{W_T}{\sqrt{T}}}_{\sim \mathrm{N}(0,1)} > \frac{\ln(K/S_0) - \left(r - \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}\right) \\
&= 1 - \mathrm{N}\left(\frac{\ln(K/S_0) - \left(r - \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}\right) \\
&= \mathrm{N}\left(\frac{\ln(S_0/K) + \left(r - \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}\right) \\
&= \mathrm{N}(d_{0,-}).
\end{aligned}
$$

[...]

Proof (cont'd).

An **Asset-or-Nothing option** with strike K pays S_T at maturity T if $S_T - K \geq 0$. By the principle of risk-neutral pricing, the no-arbitrage price of the option at time 0 is

$$\mathbb{E}^{\mathbb{Q}}[e^{-rT} S_T \cdot \mathbf{1}_{\{S_T > K\}}] = e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T \mathbf{1}_{\{S_T > K\}}].$$

We show that $e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T \mathbf{1}_{\{S_T > K\}}] = S_0 \, \mathrm{N}(d_{+,0})$. [...]

Proof (cont'd).

$$
\begin{aligned}
& e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T \mathbf{1}_{\{S_T>K\}}] \\
&\quad = \mathbb{E}^{\mathbb{Q}}\left[S_0 \exp\left(-\frac{1}{2}\sigma^2 T + \sigma W_T\right) \mathbf{1}_{\{S_0 \exp\left(\left(r-\frac{1}{2}\sigma^2\right)T+\sigma W_T\right)>K\}}\right] \\
&\quad = S_0\, \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\frac{1}{2}\sigma^2 T + \sigma W_T\right) \mathbf{1}_{\left\{\frac{W_T}{\sqrt{T}} > \underbrace{\frac{\ln(K/S_0)-\left(r-\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}}_{=:z}\right\}}\right] \\
&\quad = S_0 \int_z^{\infty} e^{-\frac{1}{2}\sigma^2 T + \sigma\sqrt{T}x}\, \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}_{=\mathrm{n}(x)}\, \mathrm{d}x \\
&\quad = S_0 \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-\frac{(x-\sigma\sqrt{T})^2}{2}}\, \mathrm{d}x \\
&\quad = S_0 \frac{1}{\sqrt{2\pi}} \int_{z-\sigma\sqrt{T}}^{\infty} e^{-\frac{y^2}{2}}\, \mathrm{d}y = S_0\, \mathrm{N}(d_{0,+}).
\end{aligned}
$$

□

Risk-neutral valuation: General case

- In general, the pricing formula for a call option is

$$\mathbb{E}^{\mathbb{Q}}[e^{-rT} S_T \mathbf{1}_{\{S_T > K\}}] - e^{-rT} K \, \mathbb{Q}(S_T > K).$$

- $\mathbb{Q}(S_T > K)$ is the risk-neutral probability that the option expires in-the-money.
- The term $\mathbb{E}^{\mathbb{Q}}[e^{-rT} S_T \mathbf{1}_{\{S_T > K\}}]$ can be re-written as

$$S_0 \, \mathbb{E}^{\mathbb{Q}^S}[\mathbf{1}_{\{S_T > K\}}] = S_0 \, \mathbb{Q}^S(S_T > K),$$

  where $\mathbb{Q}^S$ is the martingale measure when prices are expressed in units of the asset price $S = (S_t)_{t \geq 0}$ instead of discounted currency units.
- $\mathbb{Q}^S(S_T > K)$ is the probability that the option expires in-the-money, when prices are expressed in units of the stock (S takes the role of the so-called **numeraire**).

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
     - Brownian motion
     - Martingales
     - The market model
     - Black-Scholes formula from a PDE
     - Risk-neutral valuation
     - No-arbitrage and market completeness
     - Black-Scholes formula as limit in the CRR $(\star)$
   - 3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

No-arbitrage and market completeness

- Two problems that we would like answered:
  - Under what conditions is the market **arbitrage-free**?
  - Under what conditions is the market **complete**?
- First, we need to rule out "unreasonable" strategies, such as unlimited doubling strategies.

Definition

A self-financing portfolio strategy h is called **admissible** if there exists a constant A such that $V^h(t) \geq -A$, for every $t \leq T$.

The following is part of the **First Fundamental Theorem** of Asset Pricing:

Theorem

Existence of an equivalent martingale measure $\mathbb{Q}$ *implies absence of arbitrage.*

- Note that the converse statement does not hold in general, but requires a slightly stronger condition than absence of arbitrage, called **"no free lunch with vanishing risk"**.
- In particular, the Black Scholes models admits no arbitrage.

Proof.

- For the proof here we make some simplifying assumptions:
  - h is bounded,
  - $r = 0$,
  - asset prices follow a Black-Scholes model.
- Suppose there exists a self-financing portfolio strategy h satisfying

$$\mathbb{P}(V^h(T) \geq 0) = 1,$$
$$\mathbb{P}(V^h(T) > 0) > 0.$$

- This is a potential arbitrage portfolio, so, in order to prove absence of arbitrage, we must show that $V^h(0) > 0$.

[...]

Proof (cont'd).

- Because $\mathbb{P}$ and $\mathbb{Q}$ agree on sets of probability 1 and on sets of probability 0,

$$\mathbb{Q}(V^h(T) \geq 0) = 1, \tag{30}$$

$$\mathbb{Q}(V^h(T) > 0) > 0. \tag{31}$$

- Since h is self-financing, and since $r = 0$,

$$V^h(t) = V^h(0) + \int_0^t h_2(s)\,\mathrm{d}S_s = V^h(0) + \int_0^t h_2(s) S_s \sigma\,\mathrm{d}W_s.$$

- Because h is bounded, V^h is a martingale under $\mathbb{Q}$.

  [For unbounded strategies the argument is a bit more delicate and uses that $V_t^h \geq -a$, $t \leq T$.]

- Finally, it follows from Equations (30)-(31) that $V^h(0) = \mathbb{E}^{\mathbb{Q}}[V^h(T)] > 0$.

$\square$

Theorem

The Black-Scholes model is complete, that is, every contingent claim can be replicated.

- For simple claims of the form $X = \Phi(S_T)$ this follows from the Black-Scholes PDE.
- The Theorem holds also for other claims, such as path-dependent options.

No-arbitrage and market completeness

Theorem (Second Fundamental Theorem)

Assume that the market is arbitrage-free. Then the market is complete if and only if the martingale measure is unique.

- A proof is beyond the scope of the course. Here is a sketch of the main arguments:
- If the market is complete, then the unique no-arbitrage price of a contingent claim is given by the price of the replicating strategy.
- In an incomplete market, some claims cannot be replicated,
  - several prices are consistent with the no-arbitrage principle,
  - as certain price differences cannot be exploited by a replicating strategy.

Example (Incomplete market model)

- Here is the simplest example of an **incomplete market**:
- In a simple one-period model, let

$$B_0 = 1 \qquad B_T = 1$$
$$S_0 = 1 \qquad S_T \in \{d, 1, u\}, \quad d < 1 < u.$$

- Then, the contingent claim with payoff $\mathbf{1}_{\{S_T > S_0\}}$ cannot be replicated, but equivalent martingale measures exist.
- First, any replicating portfolio must fulfill

$$V_1^h = \begin{cases} 0, & \text{if } S_T \in \{d, 1\} \\ 1, & \text{if } S_T = u. \end{cases}$$

- In other words,

$$x + dy = 0$$
$$x + y = 0$$
$$x + uy = 1$$

- This is a system of three linear equations with two free variables, which can only be solved if two equations are linearly dependent.
- Hence, the claim cannot be replicated.

[...]

Example (cont'd)

- Second, both assets are martingales if

$$\mathbb{E}^{\mathbb{Q}}[B_T] = q_u + q_n + q_d = 1$$
$$\mathbb{E}^{\mathbb{Q}}[S_T] = uq_u + q_n + dq_d = 1.$$

- Choose

$$q_u = \frac{1-d}{u-1}\, q_d$$

$$q_n = \frac{u - 1 + (d-u)q_d}{u-1},$$

with $0 < q_u, q_n, q_d < 1$.

- It is easily verified that any such measure is a martingale measure.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

   3.1 Options and option strategies

   3.2 Binomial tree model

   3.3 The Black-Scholes model

      Brownian motion
      Martingales
      The market model
      Black-Scholes formula from a PDE
      Risk-neutral valuation
      No-arbitrage and market completeness
      Black-Scholes formula as limit in the CRR ($\star$)

   3.4 Hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

CRR model

- The CRR model is a discrete approximation of the Black-Scholes model.
- First, we take the CRR model to the (continuous-time) limit.
- Second, we show that the CRR pricing formula for call options converges to the Black-Scholes formula.

CRR model

- Time horizon $[0, T]$
- $\Delta t = T/n$
- Discrete trading times $0, \Delta t, 2\Delta t, \dots, (n-1)/\Delta t, n\Delta t = T$
- risk-free interest rate: r
- Set

$$u = e^{r\Delta t} e^{\sigma\sqrt{\Delta t}}$$
$$d = e^{r\Delta t} e^{-\sigma\sqrt{\Delta t}},$$

  with $\sigma > 0$ the stock price volatility.
- The risk-neutral probabilities are

$$q_u = \frac{1 - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}, \qquad q_d = 1 - q_u.$$

CRR model

- Return R_i of underlying over time $[i\Delta t, (i+1)\Delta t]$ under risk-neutral measure $\mathbb{Q}$:

$$R_i(n) = \begin{cases} e^{r\Delta t} e^{\sigma\sqrt{\Delta t}} - 1, & \text{with prob. } q_u \\ e^{r\Delta t} e^{-\sigma\sqrt{\Delta t}} - 1, & \text{with prob. } 1 - q_u \end{cases}$$

CRR model

- Stock price at time T:

$$S_T = S_0 \cdot e^{-rT} \prod_{i=1}^{n} (1 + R_i(n)) = S_0 \cdot \exp\left(\sum_{i=1}^{n} \ln\left(\frac{1 + R_i(n)}{e^{r\Delta t}} \right) \right)$$
$$=: S_0 \cdot e^{Z(n)},$$

with

$$Z(n) = \sum_{i=1}^{n} \ln\left(\frac{1 + R_i(n)}{e^{r\Delta t}} \right).$$

- $Z(n)$ is the stock price's log-return over $[0, T]$.

- To investigate the limit of $Z(n)$ as $n \to \infty$, we require a special version of the **Central Limit Theorem**.

Theorem (Central Limit Theorem)

Let $(Y_k(n))_{k \leq n}$ be a sequence of independent, identically distributed (iid) random variables with

- *mean $\mu(n)$, where $(n\mu(n)) \to \mu < \infty$ as $n \to \infty$, and*
- *variance $\sigma^2/n + R$, where $\lim\limits_{n\to\infty} \dfrac{R}{1/n} = \lim\limits_{n\to\infty} n \cdot R = 0$.*

Then

$$Z(n) = \sum_{k=1}^{n} Y_k(n) \stackrel{d}{\to} Z,$$

with $Z \sim \mathrm{N}(\mu,\sigma^2)$.

- Recall that $\Delta t = T/n$ and set

$$Y_i(n) = \ln\left(\frac{1+R_i(n)}{e^{r\Delta t}}\right) = \begin{cases} \sigma\sqrt{T/n}, & \text{with prob. } q_u, \\ -\sigma\sqrt{T/n}, & \text{with prob. } 1-q_u. \end{cases}$$

- We calculate the mean and variance of $Y_i(n)$.
- For the mean,

$$\mathbb{E}[Y_i(n)] = (2q(n)-1)\sigma\sqrt{T/n} = -\frac{\sigma^2 \cdot T}{2n},$$

  - where the last step is determined by applying Taylor's formula with respect to $\sqrt{T}n$ at 0 (asymptotically) is given by

$$2q(n) - 1 = -\frac{\sigma}{2} \cdot \frac{\sqrt{T}}{\sqrt{n}} + R,$$

    where the remainder R vanishes quickly enough to be ignored as $n \to \infty$.

- For the variance:

$$\mathrm{Var}(Y_i(n)) = \mathbb{E}[Y_i(n)^2] - (\mathbb{E}[Y_i(n)])^2 \qquad = \frac{\sigma^2 T}{n} - \underbrace{\left(-\frac{\sigma^2 T}{2n} - R\right)^2}_{=\overline{R}},$$

  where $\overline{R}$ vanishes quickly enough to be ignored as $n \to \infty$.

- The conditions of the CLT are satisfied, and it follows that $Z(n) \xrightarrow{d} Z$ with $Z \sim \mathrm{N}(-1/2\sigma^2 T, \sigma^2 T)$.

- The CRR model therefore converges to the Black-Scholes model.

- The price of the call option converges to

$$\mathbb{E}\left[(S_0 e^Z - K)^+\right],$$

  and it can be shown that this is the limit of the call price derived for the CRR model in Section 2.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

3.1 Options and option strategies

3.2 Binomial tree model

3.3 The Black-Scholes model

3.4 Hedging

Hedging in the Black-Scholes model

Static hedging

Hedging in incomplete market models*

Robustness of the Black-Scholes formula*

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging
     - Hedging in the Black-Scholes model
     - Static hedging
     - Hedging in incomplete market models*
     - Robustness of the Black-Scholes formula*

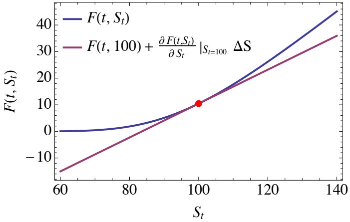
4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Hedging in the Black-Scholes model

- At any point in time $t \leq T$, the **replicating portfolio** of a European call option maturing at T with time-t price $F(t,S_t)$ consists of
  - $\dfrac{\partial F(t,S_t)}{\partial S} = \mathrm{N}(d_{t,+})$ units of stock and
  - $F(t,S_t) - \dfrac{\partial F(t,S_t)}{\partial S} S_t = -e^{-r(T-t)} K \mathrm{N}(d_{t,-})$ units of the bond.
- The partial derivative is the **sensitivity** of the call option with respect to the underlying asset.
- Small movements ΔS_t in either direction in the asset price change the option price by approximately $\dfrac{\partial F(t,S_t)}{\partial S} \Delta S_t$; formally:

$$\Delta F(t,S_t) \approx \frac{\partial F(t,S_t)}{\partial S} \Delta S_t.$$



Greeks

- The **sensitivities** of an option position are called **"Greeks"**.

Definition (Greeks)

The sensitivities of a contingent claim's price process $F(t,S_t)$ with respect to the input parameters are:

$$
\begin{array}{llll}
\text{Delta:} & \Delta_t = \dfrac{\partial F(t,S_t)}{\partial S} & \\[2ex]
\text{Gamma:} & \Gamma_t = \dfrac{\partial^2 F(t,S_t)}{\partial S^2} & \text{(sensitivity of Delta)} \\[2ex]
\text{Theta:} & \Theta_t = \dfrac{\partial F(t,S_t)}{\partial t} & \\[2ex]
\text{Vega:} & \nu_t = \dfrac{\partial F(t,S_t)}{\partial \sigma} & \text{("model risk")} \\[2ex]
\text{Rho:} & \rho_t = \dfrac{\partial F(t,S_t)}{\partial r} &
\end{array}
$$

Greeks

- In the case of a call option, the Greeks are given by:

$$\text{Delta:} \qquad \Delta_t = \mathrm{N}(d_{t,+})$$

$$\text{Gamma:} \qquad \Gamma_t = \frac{\mathrm{N}'(d_{t,+})}{S_0\sigma\sqrt{T}}$$

$$\text{Theta:} \qquad \Theta_t = -\frac{S_0\mathrm{N}'(d_{t,+})\sigma}{2\sqrt{T}} - rKe^{-rT}N(d_{t,-})$$

$$\text{Vega:} \qquad \nu_t = S_0\sqrt{T}\mathrm{N}'(d_{t,+})$$

$$\text{Rho:} \qquad \rho_t = Ke^{-rT}\mathrm{N}(d_{t,-})T$$

Greeks

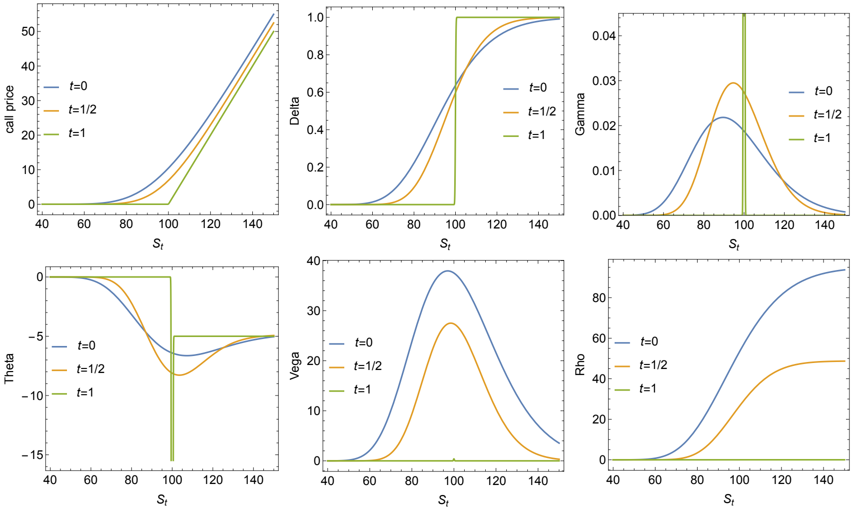


Figure: Call option prices and Greeks.

Greeks

- The option price decreases with decreasing time-to-maturity.
- In other words, aside from the **intrinsic value**, which is the payoff, the option contains a **time value**.
- This reflects the chance (risk in the positive sense) of an attractive payoff, which diminishes over time.
- This is reflected in a **negative Theta (Θ)**.
- With the same argument, the chance of a higher payoff increases with greater volatility (observe in particular that in the Black-Scholes formula, volatility and time-to-maturity always appear together).
- This is reflected in a **positive Vega**.

Greeks

- The call option price increases with the stock price, leading to a **positive Delta (Δ)**.
- Delta itself lies within 0 and 1.
- Viewed as a hedge ratio, this expresses that replicating / hedging a call option never requires holdings of more than one share of stock.
- Replicating an at-the-money call option at the end of its lifetime is tricky, because Delta has a high slope.
- This is reflected in a **increasing Gamma (Γ)** as time approaches maturity.

Greeks

- Finally, **Rho (ρ)** of a call option is always **positive**:
  - The discounted stock price process is a martingale and as such insensitive with respect to the riskfree interest rate.
  - For a call option on the other hand, only large realisations of the stock price process are significant for the payoff.
  - The higher the interest rate, the more such sufficiently large realisations.

- Plugging the sensitivities into a Taylor approximation, yields

$$\Delta F(t,S_t) \approx \Delta_t\, \Delta S + \Theta_t\, \Delta t + \nu_t\, \Delta\sigma + \rho_t\, \Delta r + \frac{1}{2}\Gamma_t\, (\Delta S)^2, \tag{32}$$

  where ΔS, Δt, $\Delta\sigma$, Δr denote (small) changes in the underlying risk factors, respectively.
- This approximation is the key to hedging the risks associated with the risk factors.

Delta-Hedging

Example

- A bank sells 50 call options with exercise price 100 EUR maturing in six months.
- The current stock price is 110 EUR, the volatility of the stock is 25%.
- The risk-free interest rate is 10%.
- How does the bank achieve a delta-neutral position?
- Verify the value change in the portfolio when the stock price increases, resp. decreases, by €1.

Example (Delta-neutral position)

- $S_0 = 110,\ K = 100,\ \sigma = 0.25,\ r = 0.1,\ T = 0.5$
- Option price: $C_0 = 16.9629$
- Furthermore,

$$
\begin{aligned}
d_{t,+} &= 0.9104 \qquad & N(d_{t,+}) &= 0.8187 \\
d_{t,-} &= 0.7336 \qquad & N(d_{t,-}) &= 0.7684
\end{aligned}
$$

- Delta of the option: $\Delta_t = N(d_{t,+}) = 0.8187$
- Delta of 50 call options: $50 \cdot 0.8187 = 40.935$
- Hedging strategy for the short position:
  - buy 40.935 stocks
  - buy $50 \cdot 16.9629 - 40.935 \cdot 110 = -3654.71$ bonds (i.e., sell short) to finance stock price position
- The resulting position is delta-neutral.

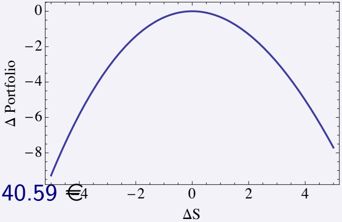
Example (Scenarios)

Case 1: stock price increases by 1 €

- call price increases by 0.8253 €
- position of the bank
  - 50 calls short: loss of $50 \cdot 0.8253 = 41.265$ €
  - 40.935 stocks long: gain of 40.935 €
- net change: -0.33 €

Case 2: stock price decreases by 1 €

- call price decreases by 0.8118 €
- position of the bank
  - 50 calls short: gain of $50 \cdot 0.8118 = 40.59$ €
  - 40.935 stocks long: loss of 40.935 €
- net change: -0.345



Hedge error from discretisation

- Hedging or replicating requires **continuous** rebalancing of the portfolio.
- In practice, this is – of course – not possible.
- Moreover, frequent rebalancing produces high transaction costs.
- Rebalancing the portfolio at discrete time points only generates P&L in the hedged portfolio – the so-called **hedge error**.
- The following example demonstrates this.

Hedge error from discretisation

Example

- Hedging a call option with daily rebalancing
- $n = 252$ time steps, $\Delta t = T/n$
- $S_0 = 100$, $\mu = 0.075$, $K = 100$, $T = 1$, $r = 0.05$, $\sigma = 0.25$
- At time 0, we enter the portfolio

$$V_0 = -F(0, S_0) + \Delta_0 S_0.$$

[This is a hedge, but not a replicating portfolio as $V_0 \neq 0$.]

[..]

Example (cont'd)

- After one time step, at $t = \Delta t$:
  - The portfolio value is

$$V_t = -F(t,S_t) + \Delta_0 S_t.$$

  - Because the hedging strategy is self-financing, the rebalanced portfolio is

$$V_t = -F(t,S_t) + \Delta_t S_t + \underbrace{(\Delta_0 - \Delta_t)S_t}_{\text{bond position}\tilde{B}_t}.$$

[..]

Hedge error from discretisation

Example (cont'd)

- More generally,at time point t:
  - The portfolio value is

$$V_t = -F(t, S_t) + \Delta_{(t-\Delta t)} S_t + \tilde{B}_{(t-\Delta t)} e^{r\Delta t},$$

  - After rebalancing:

$$V_t = -F(t, S_t) + \Delta_t S_t + \underbrace{\tilde{B}_{(t-\Delta t)} e^{r\Delta t} + (\Delta_{(t-\Delta t)} - \Delta_t) S_t}_{\text{bond position } \tilde{B}_t}$$

[..]

Hedge error from discretisation

Example (cont'd)

- Because the hedging portfolio is risk-less it must earn the risk-free rate.

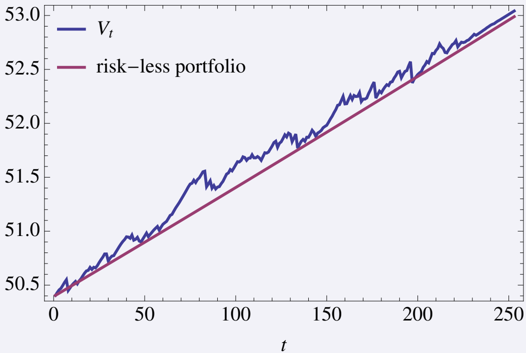


Figure: Hedge portfolio of one realisation vs. bond investment at initial portfolio value.

[..]

Example (cont'd)

- To compare, here is the hedge portfolio when $n = 10{,}000$:

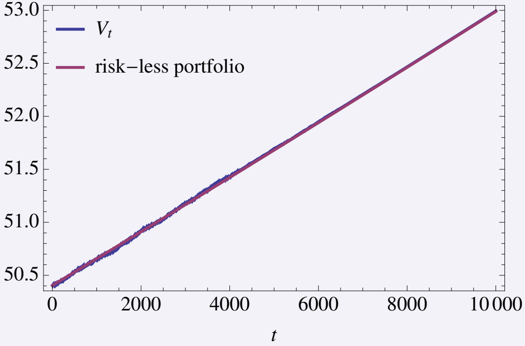


Figure: Hedge portfolio vs. bond investment at initial portfolio value.

Delta-Gamma-Hedging

- The hedge error can be reduced by hedging the **gamma risk**, i.e., by eliminating the sensitivity of Delta.
- This requires that a liquidly traded option is available in the market.
- Then, using the approximation

$$\Delta F(t,S_t) \approx \Delta_t \, \Delta S + \frac{1}{2}\Gamma_t \, (\Delta S)^2,$$

one can make the portfolio
  - **gamma-neutral** using the liquidly traded option, and
  - **delta-neutral** using the underlying asset.

Example (cont'd)

- Continuing the previous example, recall that the bank has a short position in 50 call options with parameters:
  $S_0 = 110,\ K = 100,\ \sigma = 0.25,\ r = 0.1,\ T = 0.5$
- The Delta and Gamma are given by

$$\Delta = 0.8187 \qquad\qquad \Gamma = 0.01356.$$

  This expresses that changes in the stock price translate into changes in Delta changes at a rate of 0.01356.
- Changes in the option price are approximated by

$$\Delta F(0{,}110) \approx 0.8187\,\Delta S + 0.00678\,(\Delta S)^2.$$

Example (cont'd)

- Assume that an at-the-money call option is available in the market that can be used for hedging.
- The option price, Delta and Gamma are

$$F_{\text{ATM}}(0,100) = 10.5405 \qquad \Delta_{\text{ATM}} = 0.6448 \qquad \Gamma_{\text{ATM}} = 0.0192$$

- Gamma-matching means buying $x = 35.4047$ ATM-options, since this solves

$$50 \cdot -0.01356 + x \cdot 0.1915 = 0.$$

- The Delta of this portfolio is

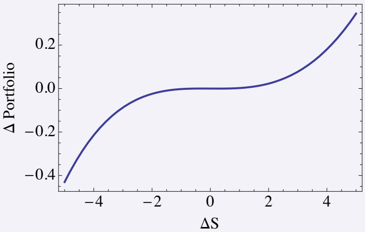
$$50 \cdot -0.8187 + 35.4047 \cdot 0.6448 = -18.106.$$

- Delta-matching means buying 18.106 shares of stock.

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example (cont'd)

- The resulting portfolio consisting of 18.106 shares of stock, 35.4047 ATM-options and the short position in 50 call options with $K = 100$ is delta- and gamma-neutral.
- The value change of the portfolio if the stock price increases by 1 € is 0.0024 €.
- The value change of the portfolio if the stock price decreases by 1 € is -0.0024 €.



Hedge error from misspecified volatility

- A further source of P&L when hedging enters via a possibly **misspecifed volatility** (see also "Robustness of the Black-Scholes formula" below).
- This risk is sometimes called **model risk**.

Example

The setup is the same as in the previous example.

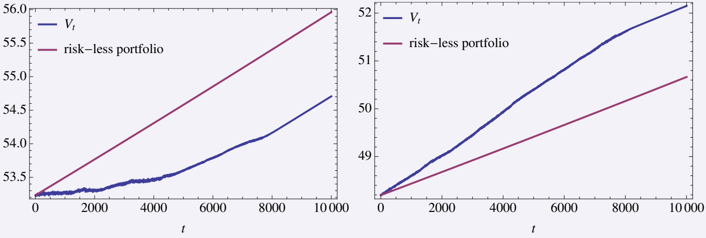


Figure: Hedging with volatility 0.2 (left) and 0.3 (right).

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

   3.1 Options and option strategies

   3.2 Binomial tree model

   3.3 The Black-Scholes model

   3.4 Hedging

      Hedging in the Black-Scholes model

      Static hedging

      Hedging in incomplete market models*

      Robustness of the Black-Scholes formula*

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Static hedging

- In some markets, liquidly traded options can be used as additional instruments for hedging.
- Instead of a dynamic hedge involving only the underlying security, a **static hedge** (i.e., without rebalancing) from a portfolio spanning the liquidly traded options can be implemented.
- Static hedges are more robust than dynamic hedges, because they do typically not require model assumption.
- We derive some key results in the theory of static hedging.
- A good overview article is

Albrecher, H. & Mayer, P.: Semi-static hedging strategies for exotic options. In: Kiesel, R.; Scherer, M. & Zagst, R. (ed.): Alternative Investments and Strategies. World Scientific, Singapore, 2010, pages 345–373.

Static hedging

- Denote the price of some asset by $S = (S_t)_{t \geq 0}$.
- Consider a European-type contingent claim on S with payoff $p(S_T)$.
- Suppose that standard European call and put options with maturity T are liquidly traded for all strike levels $K \geq 0$.
- If the payoff p is twice differentiable, then a Taylor expansion implies

$$p(S) = p(K^*) + p'(K^*)(S - K^*) + \int_{K^*}^{S} p''(x)(S - x)\,\mathrm{d}x.$$

Proof of remainder term

- To prove that the remainder is $\int_{K*}^{S} p''(x)(S-x)\,\mathrm{d}x$, recall first the **integration by parts rule**:

$$\Big[fg\Big]_a^b = \int_a^b f'g + \int_a^b f\,g'$$

- Then, with $f(x) = p'(x)$ and $g(x) = x$,

$$\begin{aligned}
\int_{K*}^{S} p''(x)(S-x)\,\mathrm{d}x &= S\int_{K*}^{S} p''(x)\,\mathrm{d}x - \int_{K*}^{S} p''(x)x\,\mathrm{d}x \\
&= S\,p'(S) - S\,p'(K^*) - \left[\Big[p'(x)x\Big]_{K*}^{S} - \int_{K*}^{S} p'(x)\,\mathrm{d}x\right] \\
&= S\,p'(S) - S\,p'(K^*) - p'(S)S + p'(K^*)K^* + p(S) - p(K^*) \\
&= p(S) - p(K^*) + p'(K^*)(K^* - S).
\end{aligned}$$

Static hedging

- Since $(S-x)=(S-x)^+-(x-S)^+$, rewrite this to obtain

$$
\begin{aligned}
p(S) = {} & p(K^*) + p'(K^*)(S-K^*) \\
& + \int_{K^*}^{\infty} p''(x)(S-x)^+ \,\mathrm{d}x + \int_0^{K^*} p''(x)(x-S)^+ \,\mathrm{d}x.
\end{aligned}
$$

Static hedging

- For hedging purposes, rewrite the Taylor expansion as

$$p(S) = p(K^*) + p'(K^*)(F - K^*) + p'(K^*)(S - F) + \int_{K^*}^{\infty} p''(x)(S - x)^+ \, \mathrm{d}x + \int_0^{K^*} p''(x)(x - S)^+ \, \mathrm{d}x,$$

where F denotes the forward price of the asset.

- The payoff is thus decomposed into the following positions:

Payoff	Hedged by position in
$p(K^*) + p'(K^*)(F - K^*)$	$(p(K^*) + p'(K^*)(F - K^*))$ bonds
$p'(K^*)(S_T - F)$	$p'(K^*)$ forwards
$\int_{K^*}^{\infty} p''(x)(S_T - x)^+ \, \mathrm{d}x$	$p''(x)\,\mathrm{d}x$ calls struck at x, $\forall x \in [K^*,\infty)$
$\int_0^{K^*} p''(x)(x - S_T)^+ \, \mathrm{d}x$	$p''(x)\,\mathrm{d}x$ puts struck at x, $\forall x \in [0,K^*)$

Static hedging

- This expresses that, if European calls and puts with maturity T are liquid for all strikes, then any (sufficiently regular) European contingent claim maturity at T can be perfectly replicated.
- The price of the claim is of course given by the static replication portfolio.
- This is regardless of model assumptions, and holds also in incomplete market models.

Static hedging

- This was already known e.g. by

  Breeden, D. T. & Litzenberger, R. H.: Prices of State-Contingent Claims Implicit in Option Prices. The Journal of Business, 1978, 51, pages 621–651
- See also

  Carr, P. & Madan, D.: Towards a theory of volatility trading. In: R. Jarrow, editor, *Risk Book on Volatility*. Risk Publications, 1998

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

- 3.1 Options and option strategies
- 3.2 Binomial tree model
- 3.3 The Black-Scholes model
- 3.4 Hedging
  - Hedging in the Black-Scholes model
  - Static hedging
  - Hedging in incomplete market models*
  - Robustness of the Black-Scholes formula*

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Hedging in incomplete market models

- In an incomplete market model, we consider hedging a claim that cannot be replicated.
- Market incompleteness could for example be caused by the presence of jumps in the market model.
- One approach is **quadratic hedging** where the variance (globally or one time step ahead) of the portfolio consisting of the claim and the hedging strategy is minimised.
- Minimising the hedge error at maturity, so-called **mean-variance hedging** is achieved by choosing the strategy that satisfies

$$\inf_{\phi} \mathbb{E}^{\mathbb{P}}[(X - V_T^{\phi})^2],$$

where $V_T^{\phi} = V_0^{\phi} + \int_0^T r\phi_t^1 \, dt + \int_0^T \phi_t^2 \, dS_t$ denotes the outcome of the dynamic self-financing trading strategy ϕ.

- It is now important to take expectation with respect to the real-world measure $\mathbb{P}$.

- When the real market is incomplete (e.g. features jumps), then there is a tradeoff in specifying the model choice:
  - hedging an incomplete market model leads to P&L from the inability to replicate a claim;
  - hedging in a misspecified, but complete market model, leads to P&L from model misspecification.

- As an example, we consider the loss from hedging options on futures in energy markets.[18]
- The Schwartz-Smith model[19] is a popular two-factor model for the energy spot price, where the logarithm of the spot price is governed by

$$\log S_t = \Lambda_t + X_t + Y_t,$$

with
  - Λ_t a deterministic seasonality function,
  - X_t a mean-reverting process driven by jumps,
  - Y_t a pure jump process independent of X_t.
- As a complete market model, we consider the same model with the jump processes replaced by Brownian motions.

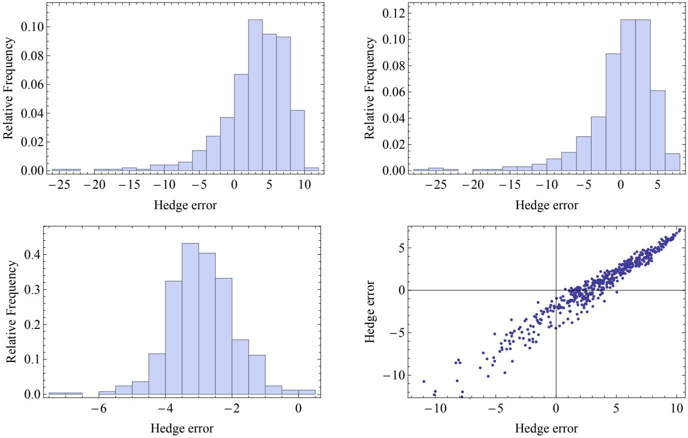
[18] The example is taken from Detering N. and Packham N., Model risk in incomplete markets with jumps. In Innovations in Risk Management, Springer Proceedings in Mathematics & Statistics, edited by K. Glau, M. Scherer and R. Zagst, Vol. 99, pp. 39–56, Springer, 2015.

[19] Schwartz,E. S., Smith,J. E.:Short-term variations and long-term dynamics in commodity prices. Management Science 46(7), 893–911, 2000.

Case study: Mean-variance hedging vs. model misspecification

- Both models are calibrated to future and spot price data from Nord Pool energy exchange, using data from January 2011 until May 2013.
- The claim to be hedged is an option on a future with a one-week delivery period trading one month prior to expiry.
- The following graphs show hedge errors from a Monte Carlo simulation.

Case study: Mean-variance hedging vs. model misspecification



- Top left: hedge error from quadratic risk-minimisation in Schwartz-Smith model; top right: hedge error from hedging in misspecified complete market model; bottom left: Differences in hedge error; bottom right: scatter plot of the hedge error samples.

- The results show that the hedge in the misspecified model generally outperforms the hedge in the incomplete market model.
- This indicates that hedging in a complete market model, even if misspecified, can be very robust.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging
   - 3.1 Options and option strategies
   - 3.2 Binomial tree model
   - 3.3 The Black-Scholes model
   - 3.4 Hedging
     - Hedging in the Black-Scholes model
     - Static hedging
     - Hedging in incomplete market models*
     - Robustness of the Black-Scholes formula*

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

Robustness of the Black-Scholes formula*

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- The **robustness of the Black-Scholes formula** is a famous result by *El Karoui, Jeanblanc, Shreve: Robustness of the Black and Scholes formula. Mathematical Finance, 8(2), 93–126, 1998.*
- Essentially, if a convex payoff is hedged in a Black-Scholes model with greater volatility than the true underlying diffusion process, then the hedge is a **superhedge**.

Robustness of the Black-Scholes formula*

- The **money market price** M and the **stock price** S are processes adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ with

$$M(t) = e^{\int_0^t r(u)\,\mathrm{d}u}$$
$$\mathrm{d}S_t = S_t(r(t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t,$$

  where $r(t)$ is a deterministic interest rate and the volatility is a nonnegative adapted process.[20]

- A **portfolio process** $(\Delta_t)_{0\leq t\leq T}$ is a bounded adapted process.
- The **self-financing value** of a portfolio process Δ with initial value Π_0^Δ is the solution to the SDE

$$\mathrm{d}\Pi_t^\Delta = r(t)[\Pi_t^\Delta - \Delta_t S_t]\,\mathrm{d}t + \Delta_t\,\mathrm{d}S_t.$$

- The **price process** for a European contingent claim paying $h(S_T)$ at T is an adapted process $(P_t)_{0\leq t\leq T}$ satisfying $P_T = h(S_T)$.

[20] The processes considered here satisfy some technical conditions that we omit for the sake of readability and that we thus take for granted.

- Now, let's assume a **misspecified stock price process** governed by

$$\mathrm{d}S_t^{\gamma,x} = S_t^{\gamma,x}[r(t)\,\mathrm{d}t + \gamma(t,S_t^{\gamma,x})\,\mathrm{d}W_t], \tag{33}$$

  where γ is the misspecified volatility and $S_0^{\gamma,x} = x$, $x > 0$.

- The misspecified value of the contingent claim at time 0 is

$$v^{\gamma}(0,x) = \frac{1}{M(T)}\mathbb{E}h(S_T^{\gamma,x}), \quad x > 0.$$

- If the market were governed by (33), then the market would be complete and $v^{\gamma}(0,x)$ would be the no-arbitrage price of the contingent claim.
- One can show that if the payoff is convex, then $v^{\gamma}(0,x)$ is convex too.

Robustness of the Black-Scholes formula*

- The hedger's self-financing portfolio value evolves according to

$$d\Pi_{\Delta_t^\gamma} = r(t)\Pi_{\Delta_t^\gamma}\,dt + \Delta_t^\gamma[dS_t - r(t)S_t\,dt].$$

- The (incorrectly!) computed value of the contingent claim is governed by (Itô formula plus BS-PDE)

$$dP_t^\gamma = r(t)P_t^\gamma\,dt + \Delta_t^\gamma[dS_t - r(t)S_t\,dt] + \frac{1}{2}[\sigma_t^2 - \gamma^2(t,S_t)]S_t^2\,\frac{\partial^2}{\partial x^2}v^\gamma(t,S_t)\,dt.$$

- Thus, the tracking error $e_t^\gamma := \Pi_t^{\Delta^\gamma} - P_t^\gamma$ is given by

$$e_t^\gamma = \frac{1}{2}M(t)\int_0^t \frac{1}{M(u)}[\gamma^2(u,S_u) - \sigma_u^2]S_u^2\,\frac{\partial^2}{\partial x^2}v^\gamma(u,S_u)\,du.$$

- If the misspecified volatility is greater than the true (possibly stochastic) volatility, $\gamma(t,S_t) \geq \sigma_t$, then (P^γ,Δ^γ) is a **superstrategy** and

$$\Pi_T^{\Delta^\gamma} \geq h(S_T)$$
$$v^\gamma(0,S_0) \geq \mathbb{E}[h(S_T)/M(T)].$$

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation
    4.1 Monte Carlo simulation
    4.2 Option pricing
    4.3 Path-dependent options
    4.4 Model discretisation error
    4.5 Variance reduction

5. Introduction to Machine Learning in Finance

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation
   - 4.1 Monte Carlo simulation
   - 4.2 Option pricing
   - 4.3 Path-dependent options
   - 4.4 Model discretisation error
   - 4.5 Variance reduction

5. Introduction to Machine Learning in Finance

Monte Carlo simulation

- **Monte Carlo simulation** is about simulating independent samples of a random variable Z using computer-generated random numbers.
- Once sufficiently many samples have been drawn, these can be used to produce an **estimate** of some quantity that depends on the distribution of Z.
- The quality of an estimate can be quantified by a **confidence interval** around the estimate.
- A comprehensive resource is
  Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

Monte Carlo simulation

- To fix ideas consider the problem of estimating the integral of a function f over the unit interval:
$$\alpha = \int_0^1 f(x)\,\mathrm{d}x.$$
- We may write
$$\alpha = \mathbb{E}[f(U)],$$
with U uniformly distributed between 0 and 1.

Monte Carlo simulation

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Drawing points $u_1, u_2, \ldots, u_n$ independently and uniformly from $[0,1]$, the Monte Carlo estimate is given by

$$\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} f(u_i).$$

- If f is integrable over $[0,1]$ then, by the **strong law of large numbers**,

$$\hat{\alpha}_n \to \alpha \text{ with probability } 1, \text{ as } n \to \infty.$$

Monte Carlo simulation

- If f is square integrable, and setting,

$$\sigma_f^2 = \int_0^1 (f(x) - \alpha)^2 \, dx,$$

then, by the **central limit theorem**, the error $\hat{\alpha}_n - \alpha$ is approximately normally distributed with mean 0 and standard deviation $\sigma_f / \sqrt{n}$.

- σ_f is typically unknown, but can be estimated by the sample standard deviation

$$s_f = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (f(u_i) - \hat{\alpha}_n)^2}.$$

Monte Carlo simulation

- Thus, an (asymptotically) valid $1 - \delta$ confidence interval for α is given by

$$\left[\hat{\alpha}_n - \mathsf{N}_{\delta/2}\frac{s_f}{\sqrt{n}}\ ,\ \hat{\alpha}_n + \mathsf{N}_{\delta/2}\frac{s_f}{\sqrt{n}}\right]$$

  where $\mathsf{N}_{\delta/2}$ denotes the $1 - \delta/2$ quantile of the standard normal distribution.
- For example, for $1 - \delta = 0.95$: $\mathsf{N}_{1-\delta/2} = \mathsf{N}_{0.975} \approx 1.96$.

Monte Carlo simulation

- Thus, from the function value $f(u_1), \ldots, f(u_n)$ we obtain
  - an estimate of the integral α,
  - and a measure of the error of the estimate.
- The form of the standard error $\sigma_f/\sqrt{n}$ implies:
  - to cut the error in half requires increasing the sample size by four;
  - adding one decimal point of precision requires 100 times as many points.
- Monte Carlo simulation it particularly well suited for high-dimensional applications, that is, when integrating over $[0,1]^d$, $d \geq 1$.

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example

- We use Monte Carlo simulation to determine the integral $\int_0^1 (2\sin(15x) + 15x)\,\mathrm{d}x$.
- This can be calculated to be $\frac{1}{30}(229 - 4\cos(15)) = 7.73463$.

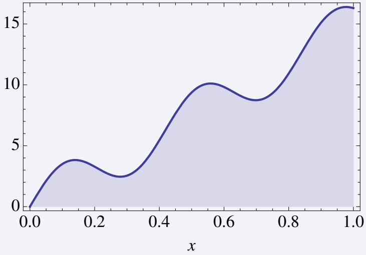


Figure: $f(x) = \int_0^1 (2\sin(15x) + 15x)\,\mathrm{d}x$.

Example (cont'd)

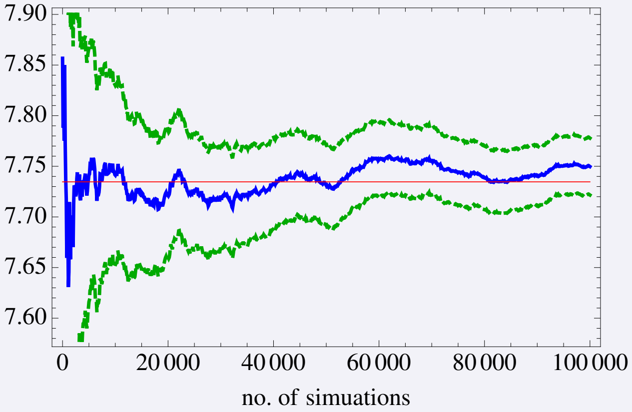


Figure: Monte Carlo estimate as a function of sample size. Dashed lines show 95%-confidence interval, constant line shows correct value.

Example

- We calculate the area of a circle, given by

$$\iint_{[-1,1]^2} \mathbf{1}_{\{u^2+v^2\leq 1\}}\, \mathrm{d}u\, \mathrm{d}v = \pi = 3.14159.$$

- In a Monte Carlo simulation, sample random numbers $u_1, \dots, u_n$ and $v_1, \dots, v_n$, each uniformly from $[-1,1]$.
- Then, count the number of samples that fulfill $u_k^2 + v_k^2 \leq 1$, $k = 1, \dots, n$.
- The fraction of such samples will be near $\pi/4$.
- Given that $[-1,1]^2$ has area 4, we obtain a result near π.

Example (cont'd)

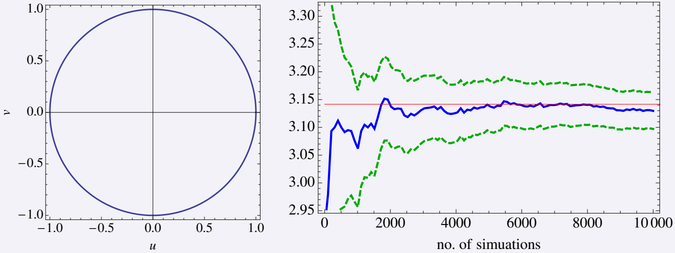


Figure: Left: Circle with radius 1. Right: Monte Carlo estimates of area (solid line), 95%-confidence interval (dashed lines), and target (constant line).

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
   - 4.1 Monte Carlo simulation
   - 4.2 Option pricing
   - 4.3 Path-dependent options
   - 4.4 Model discretisation error
   - 4.5 Variance reduction
5. Introduction to Machine Learning in Finance

- Recall: by risk-neutral pricing, the value of a contingent claim with payoff $X = \Phi(S_T)$ is given by
$$e^{-rT}\mathbb{E}^{\mathbb{Q}}[X],$$
where $\mathbb{Q}$ is the risk-neutral measure.

- Under $\mathbb{Q}$, the bond and stock prices at time T are given by
$$B_T = B_0 \, e^{rT}$$
$$S_T = S_0 \, \exp\left((r - 1/2\sigma^2)T + \sigma W_T\right),$$
with $W_T \sim \mathrm{N}(0, T)$.

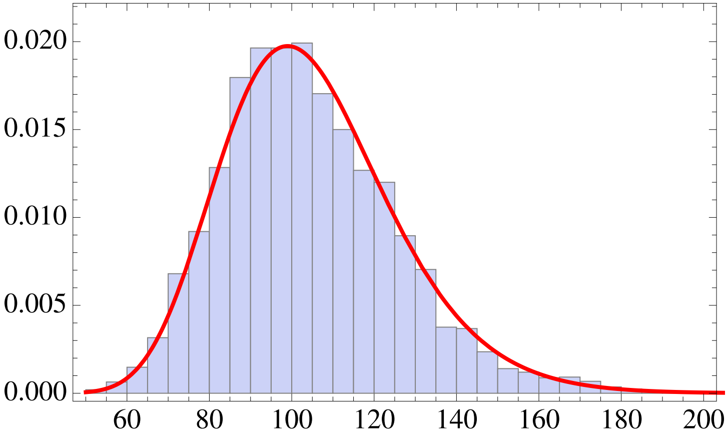


Figure: Histogram of 5000 simulated stock prices S_T. Parameters: $S_0 = 100$, $\sigma = 0.2$, $T = 1$, $r = 0.05$. Red line is the density of S_T.

- $e^{-rT}\mathbb{E}^{\mathbb{Q}}[X]$ is estimated using the following algorithm:
  - Generate uniformly distributed random numbers $u_1, \dots, u_n$.
  - Transform them to normally distributed random numbers by applying the inverse standard normal distribution function: $\mathrm{N}^{(-1)}(u_1), \dots, \mathrm{N}^{(-1)}(u_n)$
  - Set $S_{T,i} = S_0 \exp\left((r - 1/2\sigma^2)T + \sigma\sqrt{T}\mathrm{N}^{(-1)}(u_i)\right)$.
  - Set $X_i = \Phi(S_{T,i})$.
  - Set $\hat{\Pi}_n(0; X) = e^{-rT}(X_1 + \dots + X_n)/n$.

- For any $n \geq 1$, the estimated price $\hat{\Pi}(0; X)$ is **unbiased**, that is,

$$\mathbb{E}^{\mathbb{Q}}[\hat{\Pi}_n(0; X)] = \Pi(0; X) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[X].$$

- The estimator is **strongly consistent**, that is,

$$\hat{\Pi}_n(0; X) \to \Pi(0; X), \text{ as } n \to \infty.$$

Example (Call option)

- Set $X = \Phi(S_T) = (S_T - K)^+$.
- The Black-Scholes model and option parameters are $S_0 = 100$, $K = 100$, $T = 1$, $\sigma = 0.2$, $r = 0.05$.
- The Black-Scholes call option price is given by $\Pi(0; X) = 10.4506$.

Example (cont'd)

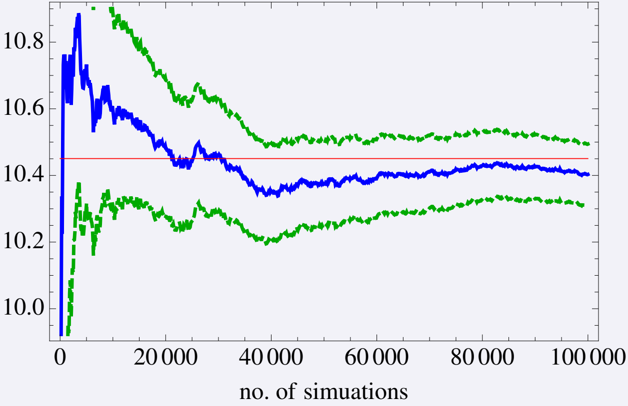


Figure: Monte Carlo estimates of Black-Scholes call option price (solid line), 95%-confidence interval (dashed lines), and target (constant line).

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
   - 4.1 Monte Carlo simulation
   - 4.2 Option pricing
   - 4.3 Path-dependent options
   - 4.4 Model discretisation error
   - 4.5 Variance reduction
5. Introduction to Machine Learning in Finance

Path-dependent options

- Valuing path-dependent options requires simulating whole sample paths.
- Depending on the payoff a **discretisation error** is introduced leading to **bias** in the value of the option.

Example (Asian option with discrete monitoring)

- The payoff of an Asian option depends on the average level of the underlying asset, e.g.

$$\overline{S} = \frac{1}{m} \sum_{j=1}^{m} S_{t_j},$$

for some fixed dates $0 = t_0 < t_1 < \cdots < t_m = T$.

- Calculating $e^{-rT}\mathbb{E}^{\mathbb{Q}}[(\overline{S} - K)^+]$ requires samples of the average $\overline{S}$.
- This is achieved by simulating the path $S_{t_1}, \ldots, S_{t_m}$ via

$$S_{t_{j+1}} = S_{t_j} \exp\left((r - 1/2\sigma^2)(t_{j+1} - t_j) + \sigma\sqrt{t_{j+1} - t_j} Z_{j+1}\right),$$

where $Z_1, \ldots, Z_m$ are independent N(0,1) random variables.

Example (Asian option (cont'd))

- As a concrete example, consider the following setup:
  - $S_0 = 100$
  - $\sigma = 0.2$
  - $r = 0.05$
  - $K = 100$
  - $T = 1$
  - discrete monitoring; every month
- The payoff is thus

$$\left(\sum_{k=1}^{12} S_{k/12} - K \right)^+$$

Example (Asian option (cont'd))

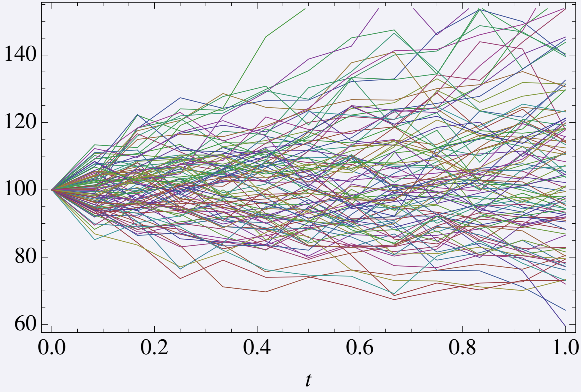


Figure: Sample stock price paths monitored every month.

Example (Asian option (cont'd))

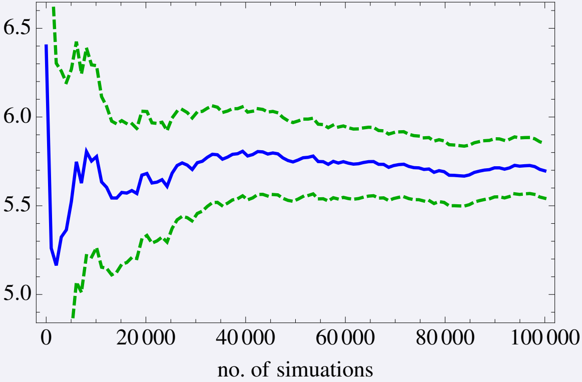


Figure: Price of discretely monitored Asian option (solid line) and 95%-confidence interval (dashed lines).

Path-dependent options

- Now consider the following path-dependent payoffs:
  - **continuously monitored Asian option** with payoff

$$\left(\frac{1}{T}\int_0^T S_u \,\mathrm{d}u - K\right)^+$$

  - **lookback option** with payoff

$$\max_{0\leq t\leq T} S_t - S_T$$

  - **continuously monitored barrier option** with payoff[21]

$$(S_T - K)^+ \mathbf{1}_{\{S_u \leq B, 0\leq u\leq t\}}.$$

- These options can be simulated only with a discretisation error in the payoff or with specialised methods, such as Brownian bridge techniques.

[21] This is an "up-and-out-option". Other variants are "down-and-out", "up-and-in" and "up-and-out".

Path-dependent options

- Sometimes the simulation can be improved by sampling the **running maximum** $M_t = \max_{0 \leq u \leq t} W_t$ of a Brownian motion.
- Conditional on $W_0 = 0$ and W_t,

$$M_t = \frac{W_t + \sqrt{W_t^2 - 2t \ln U}}{2},$$

with $U \sim U(0,1)$ independent of W_t.

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Example (Barrier option)

In each simulation step,

- In each simulation step, draw W_T and M_T and set

$$S_T = S_0\, e^{(r-1/2\sigma^2)T+\sigma W_T}$$

$$S_T^{\max} = S_0\, e^{(r-1/2\sigma^2)T/2+\sigma M_T},$$

  where we have assumed that on average the maximum will be attained when half of the time has elapsed.

- Use $S_T^{\max}$ to determine if the option has knocked out.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation
   - 4.1 Monte Carlo simulation
   - 4.2 Option pricing
   - 4.3 Path-dependent options
   - 4.4 Model discretisation error
   - 4.5 Variance reduction

5. Introduction to Machine Learning in Finance

Model discretisation error

- Consider stock price dynamics given by

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma(S_t)S_t\,\mathrm{d}W_t.$$

- Here, the volatility depends on the current level of S, and often, this equation does not admit an explicit solution.
- Hence, there is no exact mechanism for sampling from S_T.
- In this situation, one may partition $[0,T]$ into subintervals of length Δt and simulate paths via **Euler discretisation**

$$S_{t+\Delta t} = S_t + rS_t\Delta t + \sigma(S_t)S_t\sqrt{\Delta t}Z,$$

  with $Z \sim N(0,1)$.
- The resulting **model discretisation error** introduces a **bias** in the estimated option value.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
   - 4.1 Monte Carlo simulation
   - 4.2 Option pricing
   - 4.3 Path-dependent options
   - 4.4 Model discretisation error
   - 4.5 Variance reduction
5. Introduction to Machine Learning in Finance

Variance reduction techniques

- The result of a Monte Carlo simulation is an **estimator**, which itself can be regarded as a random variable.
- The principal idea of **variance reduction techniques** is to modify the simulation technique in order to reduce the estimator's variance.
- Often, this requires additional information on the problem at hand.

Key ideas

- Suppose the goal is to estimate $\mathbb{E}f(X^1, \dots, X^d)$, where
  - $(X^1, \dots, X^d)$ is a random vector,
  - $f : \mathbb{R}^d \to \mathbb{R}$ an square integrable function
  - with $\mathrm{Var}(f(X^1, \dots, X^d)) < \infty$.
- The Monte Carlo procedure is to generate iid samples $(X_i^1, \dots, X_i^d)$, $i = 1, \dots, n$.
- The MC estimator is

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i^1, \dots, X_i^d).$$

Key ideas

- The MC estimator is **strongly consistent**, that is,

$$\frac{1}{n}\sum_i f(X_i^1,\dots,X_i^d) \longrightarrow \mathbb{E}f(X^1,\dots,X^d) \quad \text{(with prob. 1)}.$$

- The **Central Limit Theorem (CLT)** states:

$$\frac{1}{\sqrt{n}}\sum_i f(X_i^1,\dots,X_i^d) \xrightarrow{\mathcal{L}} \mathrm{N}(0,\sigma^2),$$

  where
  - $\sigma^2 = \mathrm{Var}(f(X^1,\dots,X^d))$ and
  - $\xrightarrow{\mathcal{L}}$ denotes that the convergence is in distribution.

Key ideas

- Take the CLT as an indicator of **speed of convergence**:

$$\frac{1}{n}\sum_i f(X_i^1,\dots,X_i^d) \approx X, \quad X \sim \mathrm{N}(0,\sigma^2/n)$$

- The goal is to find estimators with similar properties (consistency) and smaller limit variance.

Control Variates

- This method exploits information about errors in estimates of known quantities to reduce the error in an estimate of an unknown quantity.
- Let $Y_1, \dots, Y_n$ be n iid outputs from a MC simulation, representing e.g. discounted payoffs of a contingent claim.
- The usual estimator

$$\overline{Y} = \frac{Y_1 + \cdots + Y_n}{n}$$

converges to $\mathbb{E}Y_i$ with probability 1.

Control Variates

- Let X be another random variable whose expectation $\mathbb{E}X$ is known, and calculate X_i along with Y_i.
- For example, X could be the stock price itself.
- Setting
$$Y_i(b) = Y_i - b(X_i - \mathbb{E}X), \quad b \in \mathbb{R},$$
the **control variate estimator** is given by
$$\overline{Y}(b) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(X_i - \mathbb{E}X)) = \overline{Y} - b(\overline{X} - \mathbb{E}X)$$
- Interpretation:
  - Suppose that X and Y are positively correlated.
  - If $X_i - \mathbb{E}X$ is positive, i.e., X_i overestimates $\mathbb{E}X$, then there is a tendency that Y_i is also overestimated, and vice versa.

Control Variates

- The estimator is **unbiased**:

$$\mathbb{E}[\overline{Y}(b)] = \mathbb{E}\left[\overline{Y} - b(\overline{X} - \mathbb{E}X)\right] = \mathbb{E}[\overline{Y}] - b\,\underbrace{\mathbb{E}\left[\overline{X} - \mathbb{E}X\right]}_{=0} = \mathbb{E}[Y]$$

- The estimator is **consistent**:

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i(b) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(X_i - \mathbb{E}X)) \\ &= \mathbb{E}[Y - b(X - \mathbb{E}X)] = \mathbb{E}Y.\end{aligned}$$

Control Variates

- The estimator is **unbiased**:

$$\mathbb{E}[\overline{Y}(b)] = \mathbb{E}\left[\overline{Y} - b(\overline{X} - \mathbb{E}X)\right] = \mathbb{E}[\overline{Y}] - b\,\underbrace{\mathbb{E}\left[\overline{X} - \mathbb{E}X\right]}_{=0} = \mathbb{E}[Y]$$

- The estimator is **consistent**:

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i(b) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(X_i - \mathbb{E}X)) \\ &= \mathbb{E}[Y - b(X - \mathbb{E}X)] = \mathbb{E}Y.\end{aligned}$$

- The estimator therefore fulfills some minimal properties to be classified as a "good" estimator.
- But is it better than the usual MC estimator?

Control Variates

- The variance of $Y_i(b)$ is

$$\begin{aligned}\mathrm{Var}(Y_i(b)) &= \mathrm{Var}(Y_i - b(X_i - \mathbb{E}X)) \\ &= \sigma_Y^2 - 2b\sigma_X\sigma_Y\rho_{XY} + b^2\sigma_X^2 \\ &:= \sigma^2(b),\end{aligned}$$

  where
  - $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$ and
  - ρ_{XY} is the correlation between X and Y.
- The control variate estimator $\overline{Y}(b)$ has variance $\sigma^2(b)/n$.
- The usual MC estimator $\overline{Y}$ has variance σ_Y^2/n.
- Hence the control variate estimator is superior if $b^2\sigma_X < 2b\sigma_Y\rho_{XY}$.

Control Variates

- The optimal coefficient b^* minimises the variance and is given by

$$b^* = \frac{\sigma_Y}{\sigma_X} \rho_{XY}.$$

- The ratio of the variance of the optimally controlled estimator to that of the uncontrolled estimator is given by

$$\frac{\mathrm{Var}(\overline{Y} - b^*(\overline{X} - \mathbb{E}X))}{\mathrm{Var}(\overline{Y})} = 1 - \rho_{XY}^2.$$

Control Variates

- The optimal coefficient b^* minimises the variance and is given by

$$b^* = \frac{\sigma_Y}{\sigma_X} \rho_{XY}.$$

- The ratio of the variance of the optimally controlled estimator to that of the uncontrolled estimator is given by

$$\frac{\mathrm{Var}(\overline{Y} - b^*(\overline{X} - \mathbb{E}X))}{\mathrm{Var}(\overline{Y})} = 1 - \rho_{XY}^2.$$

- In practice, of course, since $\mathbb{E}Y$ is unknown it is unlikely that σ_Y^2 or ρ_{XY} are known...
- One may replace b^* by its sample counterpart:

$$\hat{b}_n = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}.$$

Control Variates

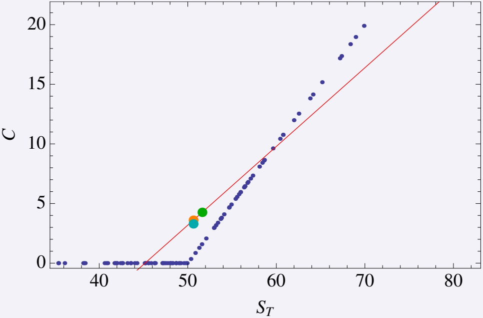
Example (Call option)

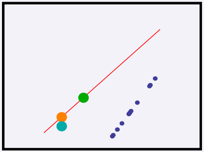
- Let $S_0 = 50$, $\sigma = 0.3$, $r = 0.05$, $T = 0.25$ and choose strikes $K \in \{40, 45, 50, 55, 60, 65, 70\}$.
- Choose S as the control variate.
- These variance ratios are based on 10,000 simulations:

K	40	45	50	55	60	65	70
var. ratio	0.012	0.065	0.200	0.414	0.642	0.814	0.916

Example (Call option (cont'd))

- Outcome of 100 simulated values for $K = 50$
- Red line: regression line with slope $\hat{b}_{100}$
- Green point: MC simulated call price
- Orange point: Controlled simulated call price
- Blue point: Black-Scholes price





Other variance reduction techniques

- **Antithetic variates**:
  - Introduce negative variance between pairs of simulations
  - Example: When sampling $U \sim U(0,1)$, then add $1 - U$ as a sample, too.
- **Stratified sampling**:
  - Sample from subsets (strata) of the sample space
  - Example: Instead of sampling n random numbers uniformly from the $[0,1]$ interval, sample one sample from each interval $[0,1/n], (1/n,2/n], \ldots, ((n-1)/n,1]$.

Other variance reduction techniques

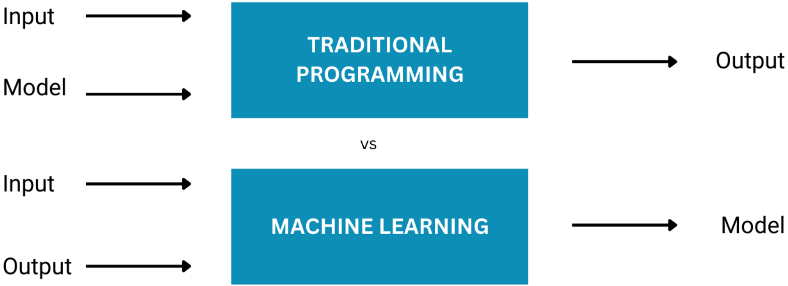
- **Importance sampling**:
  - Change the probability measure from which paths are generated
  - Example: When simulating default events, change the probabilities to attain more samples in the "region of interest" (default) and re-adjust the probability measure accordingly when calculating sample expectations.

Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
    5.1 Machine Learning Fundamentals
    5.2 Application areas of ML in Finance
    5.3 ML techniques
    5.4 Examples in Finance
    5.5 Explainable AI (XAI)

Introduction

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law



Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
     - Model validation and measures of fit
     - Statistical methods in Machine Learning
     - Ridge regression, Lasso and Elastic Net
     - Logistic regression
     - Principal component analysis
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

5.1 Machine Learning Fundamentals

Model validation and measures of fit

Statistical methods in Machine Learning

Ridge regression, Lasso and Elastic Net

Logistic regression

Principal component analysis

5.2 Application areas of ML in Finance

5.3 ML techniques

5.4 Examples in Finance

5.5 Explainable AI (XAI)

Model validation and measures of fit

- In **Machine Learning** (Data Science, Statistical Learning, "Big Data", Artificial Intelligence, ...) there is no one single statistical method that performs *best* across all data sets.
- It is an important – and at times difficult – task to select the appropriate method or model for a given data set.
- We begin by studying a number of measures to assess the quality of fit, which in turn allows to compare methods and models.
- For more, see Chapters 2.2, 5.1 and 6.1.3 of (James *et al.*, 2013):
  - James, Witten, Hastie, Tibshirani: An Introduction to Statistical Learning. Springer, 2013.

Mean-square error and overfitting

- A commonly used measure for assessing how well predictions match observed data is the **mean squared error (MSE)**, which you know e.g. from ordinary least squares (OLS) in linear regression:

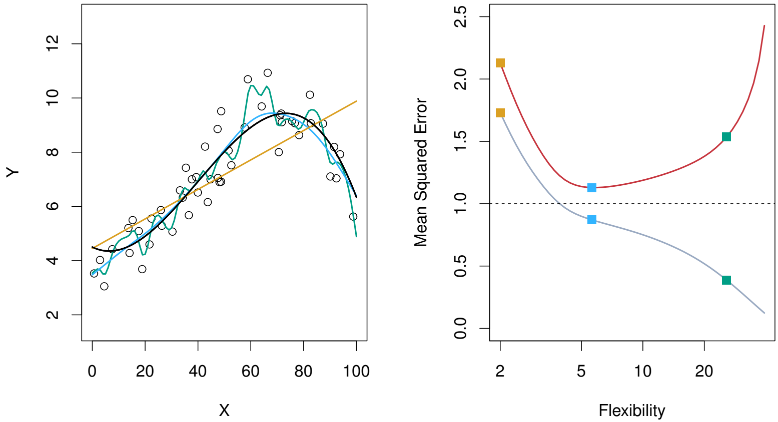
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2,$$

where $\hat{f}(x_i)$ is the prediction that the fitted method / model $\hat{f}$ gives for the i-th observation y_i.

Training and tests MSE

- In Linear Regression (a statistics method): whole data set is used for finding a linear function $\hat{f}$ that minimises the MSE.
- In ML: split the data set into a **training data set** and a **test data set**.
  - Reflects that we do not really care how well a method works on the training data.
  - Rather, we are interested in the accuracy of the prediction when applying the method to previously unseen data (the test data).
- In other words, first fit the training data to obtain the estimate $\hat{f}$, e.g. by minimising the MSE on the training data.
- Second, calculate the MSE on test data, which are data points that were not used in training.
- Choose method / model that gives lowest **test MSE** (or related measure).
- When **hyperparameters** (tuning parameters, e.g. number of hidden units in a neural network) are involved, a **validation data set** is added.

Test and training MSE



Source: James et al.: An Introduction to Statistical Learning. Springer, 2013.

Left: Test data shown as black dots, simulated from f (black smooth line).

Estimates of f: Linear regression (orange), smoothing splines (blue, green)

Right: Training MSE (grey), test MSE (red).

Flexibility denotes complexity of model (e.g. number of parameters)

The classification setting

- MSE applies in a regression setting.
- In a classification setting, we seek to estimate f on the basis of training observations $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where $y_1,\ldots,y_n$ are qualitative.
- Here, the training **error rate**, which denotes the proportion of mistakes when applying the $\hat{f}$ to the training observations, is a measure of **accuracy**:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{y_i\neq\hat{y}_i},$$

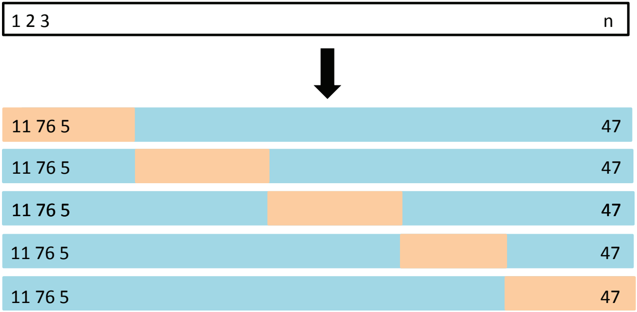
  where $\mathbf{1}_A$ is an *indicator function* taking value 1 if A is true and 0 otherwise.
- The **test error rate** is giving as the error rate from applying $\hat{f}$ to the test data set.

Cross validation

- **Cross validation (CV)** refers to several methods of building the test and training data sets.
- In k-fold CV, the data set is randomly divided in k groups or **folds** of approximately equal size.
- In k iterations, each first fold is treated as the test data set, while the $k-1$ other folds are taken as the training data.
- In this way, k MSE's of the test error are estimated and the k-fold CV estimate is given by

$$\mathrm{CV}_{(k)} = \frac{1}{k}\sum_{i=1}^{k} \mathrm{MSE}_i.$$

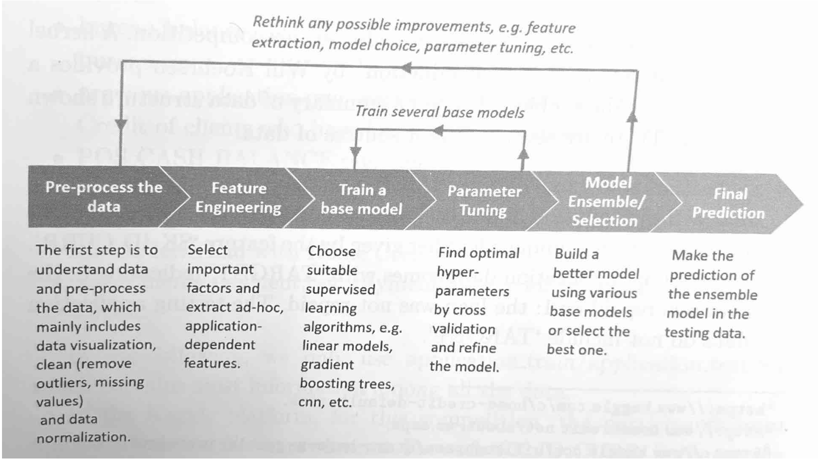
Cross validation



- A schematic display of 5-fold CV.
- A set of n observations is randomly split into five non-overlapping groups.
- Each of these fifths acts as a validation set (shown in beige), and the remainder as a training set (shown in blue).
- The test error is estimated by averaging the five resulting MSE estimates.

ML pipeline

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law



Source: (Ni *et al.*, 2021)

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
    5.1 Machine Learning Fundamentals
        Model validation and measures of fit
        Statistical methods in Machine Learning
        Ridge regression, Lasso and Elastic Net
        Logistic regression
        Principal component analysis
    5.2 Application areas of ML in Finance
    5.3 ML techniques
    5.4 Examples in Finance
    5.5 Explainable AI (XAI)

Statistical methods in ML

- We study a number of statistical methods that are popular in ML:
- ML methods are often classified into:
  - Regression vs. classification
  - Supervised vs. unsupervised learning (vs. reinforcement learning)
- The following methods illustrate these different aspects:
  - Ridge regression and Lasso (regression, supervised)
  - Logistic regression (classification, supervised)
  - Principal components analysis (PCA) (regression, unsupervised)

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

    5.1 Machine Learning Fundamentals

        Model validation and measures of fit

        Statistical methods in Machine Learning

        Ridge regression, Lasso and Elastic Net

        Logistic regression

        Principal component analysis

    5.2 Application areas of ML in Finance

    5.3 ML techniques

    5.4 Examples in Finance

    5.5 Explainable AI (XAI)

Ordinary least squares (OLS)

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- In linear regression, we assume a linear relationship between the **target** Y and the **feature vector** X:

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_m X_m + \varepsilon,$$

where $a, b_1, \ldots, b_m$ are constants and ε is the error term.

- The ordinary least squares (OLS) estimates of a,b minimise the errors

$$\sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (Y_i - a - b_1 X_{i1} - b_2 X_{i2} - \cdots - b_m X_{im})^2.$$

Regularisation

- In machine learning, especially when the number of features is high and when features are highly correlated, overfitting can easily occur.
- One way of dealing with this is known as **regularisation**.
- The most popular regularisation methods are:
  - Ridge regression
  - Lasso
  - Elastic net

Ridge regression

- **Ridge regression** is also known as **Tikhonov regularisation** and L_2 **regularisation**.
- Building on OLS, a term is added to the objective function that places a **penalty** on the size of the coefficients $b_1, \dots, b_m$, by minimising:

$$\sum_{i=1}^{n}(Y_i - a - b_1 X_{i1} - b_2 X_{i2} - \cdots - b_m X_{im})^2 + \lambda \sum_{j=1}^{m} b_j^2 .$$

- The constant λ is called **tuning parameter** or **hyperparameter** and controls the strength of the penalty factor.
- The term $\lambda \sum_{j=1}^{m} b_j^2$ is called the **shrinkage penalty**, as it will shrink the estimates of $b_1, \dots, b_m$ towards zero.
- Selection of an appropriate value of λ can be achieved, for example, by **cross-validation**.

Ridge regression

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- OLS estimates do not depend on the magnitude of the independent variables: multiplying X_j by a constant c leads to a scaling of the OLS-coefficient by $1/c$.
- This is different in regularised versions of regression: the estimated coefficients can change substantially when re-scaling independent variables.
- Therefore, it is custom, to **standardise** the features:

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}},$$

so that all variables are on the same scale, i.e., they all have a standard deviation of one.

Lasso

- **Lasso (Least absolute shrinkage and selection operator)**, also known as L_1 **regularisation** adds a different penalty:

$$\sum_{i=1}^{n}(Y_i - a - b_1X_{i1} - b_2X_{i2} - \cdots - b_mX_{im})^2 + \lambda\sum_{j=1}^{m}|b_j|.$$

- This has the interesting effect that the less relevant features are completely eliminated.
- For this reason, Lasso is also often used as a feature selection or variable selection method.

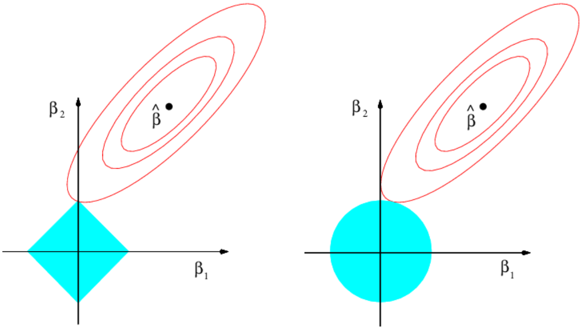
Elastic net regression

- **Elastic net regression** is a mixture of ridge regression and Lasso:

$$\sum_{i=1}^{n}(Y_i - a - b_1X_{i1} - b_2X_{i2} - \cdots - b_mX_{im})^2 + \lambda_1 \sum_{j=1}^{m} b_j^2 + \lambda_2 \sum_{j=1}^{m} |b_j|.$$

- Combining the effects of ridge regression and Lasso means that simultaneously
- some coefficients are reduced to zero (Lasso),
- some coefficients are reduced in size (ridge regression).

Illustration of regularisation constraints



Source: James et al.: An Introduction to Statistical Learning. Springer, 2013.

- $\hat{\beta}$: OLS solution
- Red level curves: mean square error
- Turquoise areas: regularisation constraints (left: Lasso; right: ridge regression)

Example

- The following application predicts house prices based on different features of the property.
- The data set is from (Hull, 2021):
  Hull: Machine Learning in Business. 3rd edition, independently published, 2021.

House price example


```
>>> import pandas as pd # python's data handling package
>>> import numpy as np # python's scientific computing package
>>> import matplotlib.pyplot as plt # python's plotting package
>>> import seaborn as sns
>>> sns.set()

>>> from sklearn.metrics import mean_squared_error as mse
>>> from sklearn.model_selection import train_test_split
>>> # The sklearn library has cross-validation built in
>>> # https://scikit-learn.org/stable/modules/cross_validation.html
>>> from sklearn.model_selection import cross_val_score

>>> #data = pd.read_csv('data/Houseprice_data_scaled.csv')
>>> data = pd.read_csv('https://raw.githubusercontent.com/packham/' \
...                    + 'SFM/main/data/Houseprice_data_scaled.csv')

>>> type(data)
<class 'pandas.core.frame.DataFrame'>
>>> # standardise the data
>>> data=(data-data.mean())/(data.std())
>>> [min(abs(data.mean())), min(abs(data.std()-1))]
[0.0, 4.440892098500626e-16]
```

House price example


```
>>> data.columns
Index(['LotArea', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'BsmtFinSF1', 'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF', '2ndFlrSF',
       'GrLivArea', 'FullBath', 'HalfBath', 'BedroomAbvGr', 'TotRmsAbvGrd',
       'Fireplaces', 'GarageCars', 'GarageArea', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', 'Blmngtn', 'Blueste', 'BrDale', 'BrkSide', 'ClearCr',
       'CollgCr', 'Crawfor', 'Edwards', 'Gilbert', 'IDOTRR', 'MeadowV',
       'Mitchel', 'Names', 'NoRidge', 'NPkVill', 'NriddgHt', 'NWAmes',
       'OLDTown', 'SWISU', 'Sawyer', 'SawyerW', 'Somerst', 'StoneBr', 'Timber',
       'Veenker', 'Bsmt Qual', 'Sale Price'],
      dtype='object')
```

House price example


```
>>> X = data.drop('Sale Price', axis=1)
>>> y = data['Sale Price']
```

- sklearn can split training and testing data randomly.

```
>>> X_train, X_test, y_train, y_test = train_test_split(X, y, \
...                                     test_size=0.25, random_state=123)
```

- If selecting a model and/or hyperparameters, use the training data set and possibly also a validation data set for training.
- Test data set is used to evaluate out-of-sample prediction error.
- Once model and hyperparameters are set, train on whole data set.

Linear Regression


```
>>> from sklearn.linear_model import LinearRegression

>>> lr=LinearRegression()
>>> lr.fit(X,y)
LinearRegression()

>>> pred = lr.predict(X)
>>> mse(y, pred)
0.11054079741914288
```

```
>>> # Create dataFrame with corresponding feature and its respective coefficients

>>> coeffs = pd.DataFrame([['intercept'] + list(X.columns),[lr.intercept_] \
...            + lr.coef_.tolist()]).transpose().set_index(0)
```

Linear Regression


```
>>> coeffs[0:24]
                                    1
0
intercept                    0.000001
LotArea                      0.073815
OverallQual                  0.223312
OverallCond                  0.085409
YearBuilt                    0.156325
YearRemodAdd                 0.036859
BsmtFinSF1                   0.078249
BsmtUnfSF                   -0.053001
TotalBsmtSF                  0.142386
1stFlrSF                     0.164668
2ndFlrSF                     0.143506
GrLivArea                    0.166844
FullBath                    -0.014708
HalfBath                     0.012265
BedroomAbvGr                 -0.07193
TotRmsAbvGrd                 0.047018
Fireplaces                   0.028241
GarageCars                   0.003794
GarageArea                   0.083191
WoodDeckSF                   0.021904
OpenPorchSF                  0.016002
EnclosedPorch                0.005409
Blmngtn         -58830654859.717506
Blueste         -35267738639.175789
```

```
>>> coeffs[24:]
                              1
0
BrDale      -60874365180.827942
BrkSide    -113416663326.511093
ClearCr     -73542987526.013672
CollgCr    -173967552841.168884
Crawfor    -110833830321.921158
Edwards     -149241610149.47345
Gilbert    -139374492502.688263
IDOTRR     -104332634101.465836
MeadowV     -67522090765.366051
Mitchel    -116921411515.306015
Names      -216490452424.313202
NoRidge     -91692890758.494904
NPkVill     -53366073480.986565
NriddgHt   -139374492502.567383
NWAmes     -124954531584.559631
OLDTown    -165463291935.284332
SWISU       -76759486076.153412
Sawyer     -133670409898.126358
SawyerW    -122191228948.987762
Somerst    -145924106522.775238
StoneBr      -79080358844.21637
Timber      -93615505038.571793
Veenker     -54504413657.359413
Bsmt Qual              0.024403
```

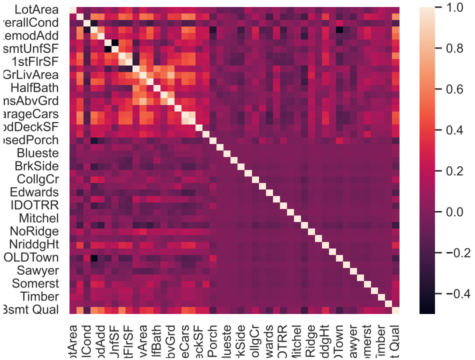
Linear Regression

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Observe how the OLS coefficients are all non-zero.
- Some coefficients are negative where a positive coefficient would be expected.
- Some coefficients are unreasonably large.
- This is an indication that the model is struggling to fit the large number of features.

- Indeed, some correlations are high, as the heatmap indicates, which may cause ill-fitting.

```
>>> sns.heatmap(X_train.corr())
<Axes: >
>>> plt.savefig('_pics/heatmap.pdf')
```



- Likewise some correlations of the sale price with the features are close to zero:

```
>>> plt.clf()
>>> sns.heatmap(pd.DataFrame(data.corr().iloc[:,-1]))
<Axes: >
>>> plt.savefig('_pics/heatmap2.pdf')
```

Ridge regression

- Use cross-validation for training and testing.
- Specify the number of folds ('cv') and MSE as the 'scoring' function.
- On each fold, train model and determine test MSE:

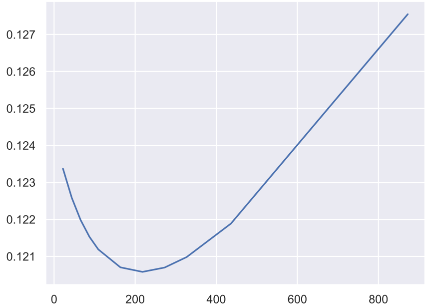
```
>>> # Importing Ridge
>>> from sklearn.linear_model import Ridge

>>> n=int(len(X)*.75) # choose 75% of data for training (=4 folds)

>>> # The alpha used by Python's ridge should be the lambda
>>> # times the number of observations
>>> alphas=[0.01*n, 0.02*n, 0.03*n, 0.04*n, 0.05*n, 0.075*n,0.1*n,0.125*n, \
...         0.15*n,0.2*n, 0.4*n]
>>> mses=[]
>>> for alpha in alphas:
...     scores = cross_val_score(Ridge(alpha=alpha), X, y, cv=4, \
...                              scoring="neg_root_mean_squared_error")**2
...     mses.append(np.mean(scores))
...
```

Ridge regression

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Average test MSE varies with the hyperparameter α.

```
>>> alpha0 = alphas[np.argmin(mses)]
>>> alpha0
218.10000000000002

>>> plt.clf()
>>> plt.plot(alphas, mses)
[<matplotlib.lines.Line2D object at 0x298c37790>]
>>> plt.savefig("_pics/mses_ridge.pdf")
```



Ridge regression

- Once optimal parameter α has been found, train on whole data set.

```
>>> ridge=Ridge(alpha=alpha0)
>>> ridge.fit(X,y)
Ridge(alpha=218.10000000000002)
>>> pred=ridge.predict(X)
>>> mse(y, pred)
0.11193850931696854

>>> coeffs = pd.DataFrame([['intercept'] + list(X.columns),[ridge.intercept_] \
...                           + ridge.coef_.tolist()]).transpose().set_index(0)
```

Ridge regression

```
>>> coeffs[0:24]
                      1
0
intercept           0.0
LotArea        0.068256
OverallQual    0.205581
OverallCond    0.071675
YearBuilt      0.111748
YearRemodAdd   0.050405
BsmtFinSF1     0.097716
BsmtUnfSF     -0.029288
TotalBsmtSF    0.113554
1stFlrSF       0.129195
2ndFlrSF       0.087459
GrLivArea      0.171742
FullBath       0.006301
HalfBath       0.025146
BedroomAbvGr  -0.057341
TotRmsAbvGrd   0.054855
Fireplaces     0.039078
GarageCars     0.025353
GarageArea     0.076272
WoodDeckSF     0.025644
OpenPorchSF    0.020948
EnclosedPorch  0.003147
Blmngtn       -0.013848
Blueste       -0.013227
```

```
>>> coeffs[24:]
                  1
0
BrDale    -0.020517
BrkSide    0.010642
ClearCr   -0.007046
CollgCr   -0.009199
Crawfor    0.034767
Edwards   -0.007784
Gilbert    -0.01416
IDOTRR    -0.003463
MeadowV   -0.016518
Mitchel   -0.030596
Names     -0.029973
NoRidge    0.056606
NPkVill   -0.020497
NriddgHt   0.111686
NWAmes     -0.04545
OLDTown    -0.02373
SWISU     -0.009188
Sawyer    -0.016932
SawyerW   -0.026284
Somerst    0.024082
StoneBr    0.080789
Timber     0.007768
Veenker    -0.00441
Bsmt Qual  0.042287
```

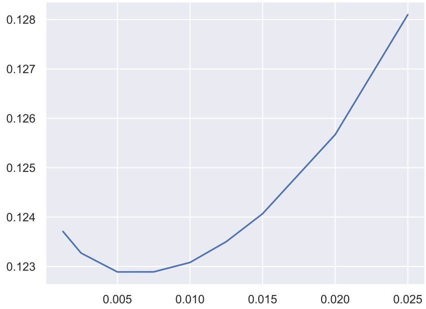
- First, find the parameter with minimal average test MSE.

```
>>> from sklearn.linear_model import Lasso

>>> # The alphas are half the lambdas
>>> alphas=[0.0025/2, 0.005/2, 0.01/2, 0.015/2, 0.02/2, 0.025/2, \
...         0.03/2,0.04/2, 0.05/2]
>>> mses=[]
>>> for alpha in alphas:
...     scores = cross_val_score(Lasso(alpha=alpha), X, y, \
...                              cv=4, scoring="neg_root_mean_squared_error")**2
...     mses.append(np.mean(scores))
...
>>> alpha0=alphas[np.argmin(mses)]
>>> alpha0
0.005
```

Lasso

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

```
>>> plt.clf()
>>> plt.plot(alphas, mses)
[<matplotlib.lines.Line2D object at 0x287c801d0>]
>>> plt.savefig("_pics/mses_lasso.pdf")
```



Lasso

- Now train on optimal α:

```
>>> lasso=Lasso(alpha=alpha0)
>>> lasso.fit(X,y)
Lasso(alpha=0.005)
>>> pred=lasso.predict(X)
>>> mse(y, pred)
0.11182731049018248

>>> coeffs = pd.DataFrame([['intercept'] + list(X.columns),[lasso.intercept_] \
...                       + lasso.coef_.tolist()]).transpose().set_index(0)
```

- Some of the coefficients have now been set to zero.
- Lasso acts as a variable selection method.

Lasso

```
>>> coeffs[0:24]
                       1
0
intercept            0.0
LotArea         0.071922
OverallQual     0.235109
OverallCond     0.073156
YearBuilt       0.134115
YearRemodAdd    0.041335
BsmtFinSF1        0.1115
BsmtUnfSF      -0.016096
TotalBsmtSF     0.106359
1stFlrSF        0.040813
2ndFlrSF             0.0
GrLivArea       0.320877
FullBath            -0.0
HalfBath         0.01179
BedroomAbvGr   -0.058548
TotRmsAbvGrd    0.028973
Fireplaces      0.028746
GarageCars           0.0
GarageArea      0.088191
WoodDeckSF      0.019653
OpenPorchSF     0.013921
EnclosedPorch        0.0
Blmngtn        -0.004362
Blueste        -0.006062
```

```
>>> coeffs[24:]
                 1
0
BrDale   -0.011639
BrkSide   0.019993
ClearCr       -0.0
CollgCr        0.0
Crawfor   0.043509
Edwards   0.002799
Gilbert       -0.0
IDOTRR    0.003519
MeadowV  -0.004346
Mitchel   -0.01507
Names    -0.005214
NoRidge   0.058122
NPkVill  -0.011026
NriddgHt   0.12146
NWAmes   -0.031721
OLDTown  -0.000562
SWISU         -0.0
Sawyer        -0.0
SawyerW  -0.012822
Somerst   0.028818
StoneBr   0.083186
Timber    0.010555
Veenker       -0.0
Bsmt Qual 0.026023
```

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance

5.1 Machine Learning Fundamentals

Model validation and measures of fit

Statistical methods in Machine Learning

Ridge regression, Lasso and Elastic Net

Logistic regression

Principal component analysis

5.2 Application areas of ML in Finance

5.3 ML techniques

5.4 Examples in Finance

5.5 Explainable AI (XAI)

Logistic regression

- In a regression setting, numerical variables are predicted.
- Another application is **classification**, which is about predicting the category a new observation belongs to.
- In supervised learning, and with two categories, a variation of regression, called **logistic regression** can be used.
- Given features $X_1, \dots, X_m$, suppose there are two classes to which observations can belong.
- An example is the prediction of a loan's default risk, given characteristics of the creditor such as age, education, marital status, etc.
- Another example is the classification of e-mails into junk or non-junk e-mails.

Logistic regression

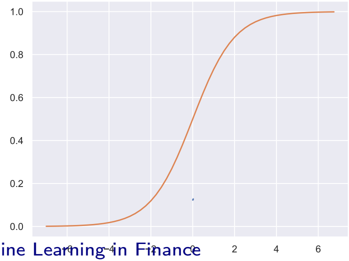
Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Logistic regression can be used to calculate the probability of a positive outcome via the **sigmoid function**

$$P(Y=1|X) = \frac{1}{1+e^{-X}} = \frac{e^X}{1+e^X},$$

where e is the Euler constant.

```
>>> x=np.arange(-7,7,0.25)
>>> plt.plot(x, 1/(1+np.exp(-x)))
[<matplotlib.lines.Line2D object at 0x287cde590>]
>>> plt.savefig("_pics/sigmoid.pdf")
```



Logistic regression

- Setting $X = a + b_1X_1 + b_2X_2 + \cdots + b_mX_m$, the probability of a positive outcome is

$$P(Y = 1|X_1, \dots, X_m) = \frac{1}{1 + \exp(-a - \sum_{j=1}^{m} b_jX_j)}.$$

- The objective is to find the coefficients $a, b_1, \dots, b_m$ that best classify the given data.
- **Maximum likelihood** is a versatile method for this type of problem, when OLS does not apply.
- Without going into detail, the **log likelihood function** is given as

$$\ell(a, b_1, \dots, b_m|x_1, \dots, x_n) = \sum_{k:y_k=1} \ln(p(x_k)) + \sum_{k:y_k=0} \ln(1 - p(x_k)),$$

  and the parameters are chosen that maximise this function.
- (Note: The likelihood function is derived by considering the observations to be independent outcomes of a Bernoulli random variable.)

Example: Credit risk

- The dataset in this example is taken from (James *et al.*, 2013):
  James et al.: An Introduction to Statistical Learning. Springer, 2013.
- It contains simulated data of defaults on credit card payments, on the basis of credit card balance (amongst other things).
- An excellent tutorial with examples on logistic regression in Python is available here: `https://realpython.com/logistic-regression-python/`.
- We will use the `sklearn` package below. Logistic regression can also be performed with the `statsmodels.api`, in which case p-values and other statistics are calculated.

Example: credit risk


```
>>> import matplotlib.pyplot as plt
>>> import numpy as np
>>> import pandas as pd
>>> from sklearn.linear_model import LogisticRegression
>>> from sklearn.metrics import classification_report, confusion_matrix
>>> from sklearn.model_selection import train_test_split

>>> data = pd.read_csv("https://raw.githubusercontent.com/packham/" \
...                    + "SFM/main/data/Default_JamesEtAl.csv")

>>> data.head()
  default student      balance        income
0      No      No   729.526495  44361.625074
1      No     Yes   817.180407  12106.134700
2      No      No  1073.549164  31767.138947
3      No      No   529.250605  35704.493935
4      No      No   785.655883  38463.495879
```

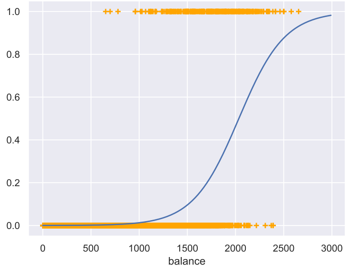
Example: credit risk


```
>>> x=np.array(data["balance"]).reshape(-1,1) # array must be two-dimensional
>>> y=np.array([True if x=="Yes" else False for x in data["default"]])
>>> x_train, x_test, y_train, y_test \
...     = train_test_split(x, y, test_size=0.2, random_state=0)

>>> model = LogisticRegression(solver='liblinear', random_state=0)
>>> model.fit(x_train,y_train)
LogisticRegression(random_state=0, solver='liblinear')


>>> # fitted parameters
>>> a=model.intercept_[0]
>>> b=model.coef_[0,0]
>>> [a,b]
[-8.537515117344613, 0.0041960838665424886]
```

Example: credit risk

- Scatter plot of data and fitted logistic function:

```
>>> plt.clf()
>>> plt.scatter(x,y,c='orange', marker="+")
<matplotlib.collections.PathCollection object at 0x28127fad0>
>>> plt.xlabel('balance')
Text(0.5, 0, 'balance')
>>> xrange=range(0,3000,10)
>>> plt.plot(xrange,1/(1+np.exp(-a-b *xrange)))
[<matplotlib.lines.Line2D object at 0x2811ed6d0>]
>>> plt.savefig("_pics/logistic.pdf")
```



Example: credit risk

- Predictions:

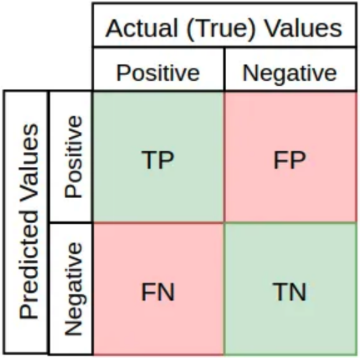
```
>>> model.predict_proba(x_train)[:5]
array([[0.98633215, 0.01366785],
       [0.98600606, 0.01399394],
       [0.98133591, 0.01866409],
       [0.998236  , 0.001764  ],
       [0.99643573, 0.00356427]])
```

- Mean accuracy of the model:

```
>>> [model.score(x_train,y_train), model.score(x_test,y_test)]
[0.972875, 0.968]
```

Example: credit risk

- Confusion matrix:



https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec

```
>>> confusion_matrix(y_test,model.predict(x_test))
array([[1923,    3],
       [  61,   13]])
```

Example: credit risk

- Classification report:

```
>>> print(classification_report(y_test, model.predict(x_test)))
              precision    recall  f1-score   support

       False       0.97      1.00      0.98      1926
        True       0.81      0.18      0.29        74

    accuracy                           0.97      2000
   macro avg       0.89      0.59      0.64      2000
weighted avg       0.96      0.97      0.96      2000
```

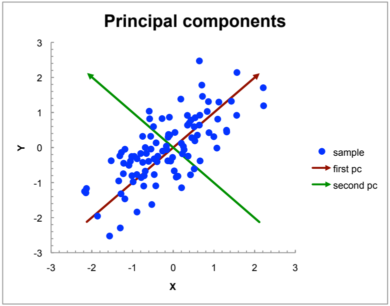
Contents

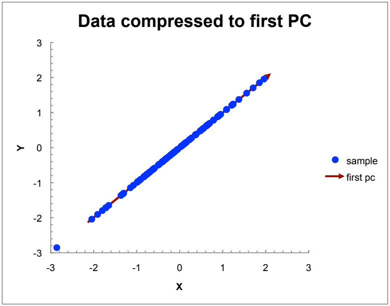
1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
   5.1 Machine Learning Fundamentals
      Model validation and measures of fit
      Statistical methods in Machine Learning
      Ridge regression, Lasso and Elastic Net
      Logistic regression
      Principal component analysis
   5.2 Application areas of ML in Finance
   5.3 ML techniques
   5.4 Examples in Finance
   5.5 Explainable AI (XAI)

Principal Component Analysis

- **Principal Component Analysis (PCA)** summarises a large set of correlated variables by smaller number of representative variables that explain most of the variability of the original data set.
- **PCA** is a standard method for reducing the dimension of high dimensional, highly correlated systems.
- **PCA** is a particular rotation of the axes, driven by random variables or data.
- Key idea is to align random variables / data such that
  - first dimension captures maximal variance,
  - second dimension is orthogonal and captures second-most variance,
  - etc.

- Example:


Principal components
3
2
1
0
-1
-2
-3
Y
-3
-2
-1
0
1
2
3
X
sample
first pc
second pc


Data compressed to first PC
3
2
1
0
-1
-2
-3
Y
-3
-2
-1
0
1
2
3
X
sample
first pc

Principal Component Analysis

- See James *et al.* (2013), Section 10.2, for the following.
- Given $n \times d$ data set $\mathbf{X}$ that is **standardised**.
- **First principal component**: find **scores**

$$z_{i1} = \phi_{11} x_{i1} + \phi_{21} x_{i2} + \cdots + \phi_{d1} x_{id}, \quad i = 1, \dots, n,$$

that have largest sample variance, subject to constraint $\sum_{j=1}^{p} \phi_{j1}^2 = 1$.

- In other words, **first PC vector** solves optimisation problem

$$\max_{\phi_{11},\dots,\phi_{p1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(\sum_{j=1}^{p} \phi_{j1} x_{ij} \right)^2}_{=z_{i1}^2} \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} \phi_{j1}^2 = 1.$$

- Second (and higher) PCs: linear combination of data uncorrelated with first PC(s) and with largest variance (subject to constraint).

Principal Component Analysis

- **Principal components (PCs)** are the **eigenvectors** of covariance / correlation matrix.
- **Eigenvalues** express **amount of variance** captured by each PC.
- Compact notation (recall that $\mathbf{X}$ is standardised):

$$\mathbf{Z} = \mathbf{\Phi}^T \mathbf{X}$$

- PCs can be viewed as factors, giving factor model

$$\mathbf{X} = \mathbf{\Phi}\,\mathbf{Z}.$$

- $\mathbf{\Phi}$ are the eigenvectors of correlation matrix of $\mathbf{X}$.

Principal Component Analysis

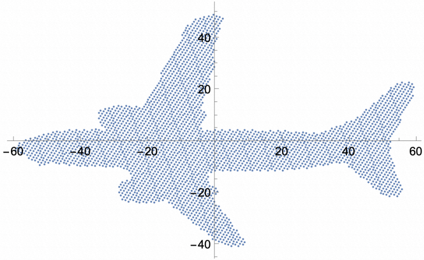
- Example from Mathematica:

In[264]:=

```
shape = Position[ImageData@
```

```
, 1, {2}];
ListPlot[PrincipalComponents[N@shape]]
```

Out[265]=



Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
   5.1 Machine Learning Fundamentals
   5.2 Application areas of ML in Finance
   5.3 ML techniques
   5.4 Examples in Finance
   5.5 Explainable AI (XAI)

Application areas of ML in Finance

- **Risk Management**: Assess and manage financial risks by analyzing market trends, economic indicators, and historical data to predict potential risks and advise on mitigation strategies.

  E.g. XAIFi (Explainable AI for the Risk Management of FinTech) project, https://www.ifaf-berlin.de/projekte/xaifi/; Stress-testing (Packham, 2024; Packham and Woebbeking, 2023)

- **Credit Scoring and Loan Underwriting**: Assess credit quality of applicants by analyzing a wide range of data, including credit history, transaction behaviors, and even social media activity.

  See e.g. (Dumitrescu *et al.*, 2022) and references herein

- **Fraud Detection and Prevention**: Identify unusual patterns and anomalies that may indicate fraudulent activities. These systems analyze transaction data in real-time to detect and prevent fraud ($\mapsto$ Buy-now-pay-later (BNPL)).

  https://en.wikipedia.org/wiki/Data_analysis_for_fraud_detection

Application areas of ML in Finance

- **Anti-Money Laundering (AML)**: Monitor transactions and identify patterns indicative of money laundering, ensuring compliance with regulatory standards.

  (Han *et al.*, 2020; Jensen and Iosifidis, 2023)

Application areas of ML in Finance

- **Calibration, Pricing and Hedging**: Model calibration, rapid derivatives pricing and hedging on the trading floor.
  (Funahashi, 2024; Ruf and Wang, 2020; Büchel *et al.*, 2021; Buehler *et al.*, 2019; De Spiegeleer *et al.*, 2018)
- **Stock Market Prediction**: Analyse historical stock data, market sentiment, and other relevant factors to predict future stock prices, aiding investors in making informed decisions.
  https://en.wikipedia.org/wiki/Stock_market_prediction#Machine_learning; also fractional trading in XAIFi project
- **Algorithmic Trading**: Trading algorithms that analyze market data, predict price movements, and execute trades at optimal times, often within milliseconds.
  https://en.wikipedia.org/wiki/Stock_market_prediction#Machine_learning

Application areas of ML in Finance

- **Portfolio Management (Robo-Advisors)**: Automated investment platforms, known as robo-advisors, use machine learning to create and manage investment portfolios tailored to individual preferences and risk tolerances.
  Literature review: (Torno *et al.*, 2021)
- **Customer Service Chatbots**: AI-driven chatbots handle customer inquiries, provide account information, and assist with transactions.
  https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
     - Decision trees
     - Random forests
     - Gradient Boosting Decison Trees
     - Neural networks
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

ML techniques

- This section primarily uses material from:
- (Ni *et al.*, 2021)

  Ni, H., X. Dong, J. Zheng, and G. Yu. An Introduction to Machine Learning in Quantitative Finance. World Scientific, 2021.
- (Bernard, 2021)

  Bernard, E. Introduction to Machine Learning. Wolfram Media, 2021.

Contents

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

1. Review of Statistical Foundations

2. Regression and time series analysis

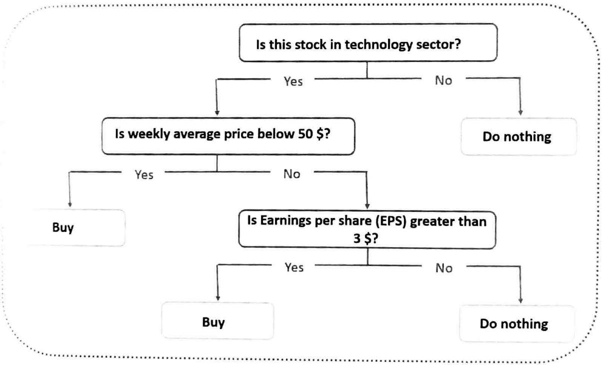
3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
     - Decision trees
     - Random forests
     - Gradient Boosting Decison Trees
     - Neural networks
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

Decision trees

- A **decision tree** is a undirected graph, such that for any two nodes there exists only one path connecting them.
- Fixing root of tree makes it a directed graph (like a flowchart)
- Terminology: root node, decision node, leaf/terminal node, parent and child node



Source: (Ni *et al.*, 2021)

Decision trees

- Pros:
  - Easy to understand (interpretability)
  - Useful in data exploration and for feature selection
  - Data type is not a constraint
  - Non-parametric method
- Cons:
  - Over-fitting
  - Not suitable for continuous variables
- Most decision tress are **binary trees**:
  - Each decision node has a splitting rule that assigns observations to either left or right child nodes.
  - Terminal node thus identifies a partition of the observation space according to splitting rules.

Regression trees

Regression tree:

- Tree with leaves $R_1, \dots, R_m$ gives rise to

$$f(x) = \sum_{m=1}^{M} c_m \, \mathbf{1}_{\{x \in R_m\}},$$

  where $c_1, \dots, c_M$ are constants.

- Given $R_1, \dots, R_m$, finding optimal regression tree boils down to minimising:

$$\sum_{i=1}^{n} (y_i - f(x_i))^2.$$

- Main problem is to find global optimal partition.

Classification trees

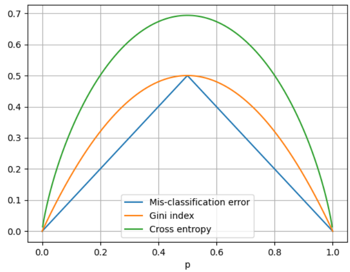
Classification tree:

- Percentage of samples estimated as class k in region R_m:

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} \mathbf{1}_{\{y_i \in k\}},$$

with N_m number of samples belonging to region R_m.

- **Impurity measures**: performance measures:
  - **Missclassification error**: $\sum_{m=1}^{M}(1 - \hat{p}_{mm}$
  - **Gini index**: $\sum_{k=1}^{M} \hat{p}_{mk}(1 - \hat{p}_{mk})$
  - **Cross entropy**: $-\sum_{k=1}^{M} \hat{p}_{mk} \log(\hat{p}_{mk})$
- **Feature importance**:
  - e.g. impurity measure gain



Source: (Ni *et al.*, 2021)

Contents

1. Review of Statistical Foundations

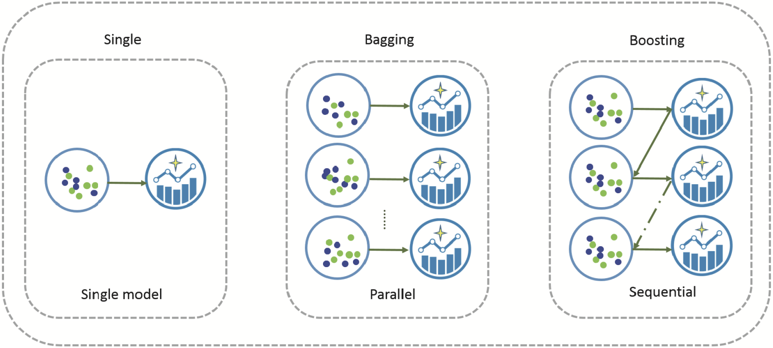
2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
     - Decision trees
     - Random forests
     - Gradient Boosting Decison Trees
     - Neural networks
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

Ensemble methods

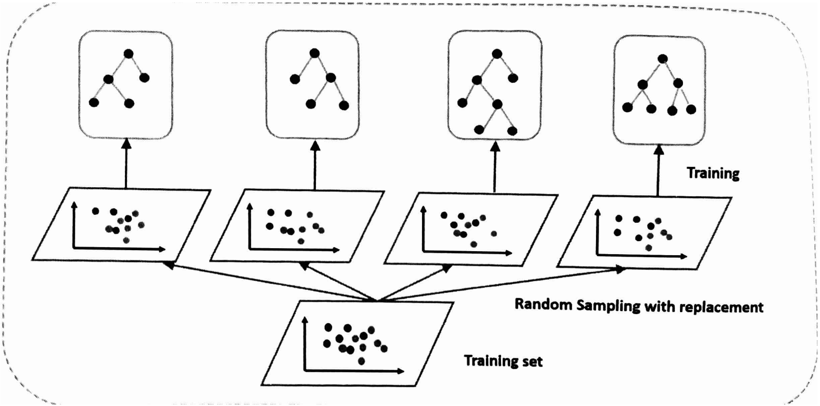


Source: (Ni *et al.*, 2021)

Random Forest

- A **forest** is an undirected graph, all of whose connected components are trees. In other words, the graph consists fo a disjoint union of trees.
- **Random forests** (Breiman, 2001) generate multiple decision trees by
  - random sampling of data (“bootstrap aggregation”);
  - random selection of input features for generating individual base decision trees.
- Prediction is average model prediction (regression) or majority vote (classification).
- Variance of error of random forest models reduced by
  - predictive power of individual decision trees, and
  - correlation among trees.

Random forest



Source: (Ni *et al.*, 2021)

Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
     - Decision trees
     - Random forests
     - Gradient Boosting Decison Trees
     - Neural networks
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

Gradient Boosting Decison Trees

- Main idea of **boosting**: Update subsequent predictors based on the error of previous predictors.
- **Gradient boosting**:
  - Goal is to minimise loss function $L(f,y)$.
  - Sequence of weak learners (here: decision tree) h_m, $m = 1, \dots, M$, is used to build sequence of predictors f_m, $m = 1, \dots, M$.
  - Updates uses gradient, similar to **gradient descent** (a popular optimisation algorithm) for optimal "direction".

Gradient boosting

Algorithm 2: Gradient Boosting Algorithm

Input: $(x_i, y_i)_{i=1}^N$.
Initialize f_0 by a constant γ_0 via the following equation:

$$\gamma_0 = \arg\min_{\gamma} L(y, \gamma);$$

for $m = 1 : M$ **do**
  for $i = 1 : N$ **do**
    Compute the residuals

$$r_{im} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}.$$

  end for
  Fit a base model learner h_m to the target r_{im}, using the data $(x_i, r_{im})_{i=1}^n$.
  Solve the one dimensional optimization problem

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, f_{m-1}(x_i) + \gamma h_m(x_i)).$$

  Update f_m using the following formula:

$$f_m(x) = f_{m-1}(x) + \gamma_m h_m(x).$$

end for
Output: f_M.

Source: (Ni *et al.*, 2021)

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

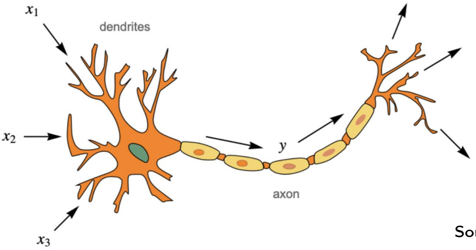
4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
     - Decision trees
     - Random forests
     - Gradient Boosting Decison Trees
     - Neural networks
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

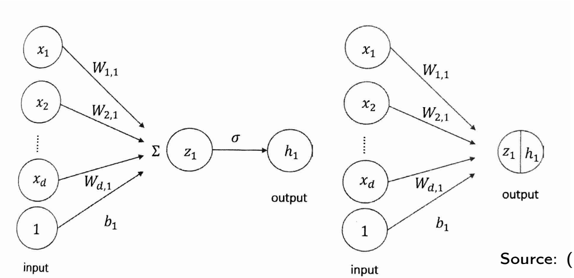
Neural networks – basic concepts

- A **neuron** is the basic unit of a neural network.
- Neurons are connected and transmit **signals** between each other.
- The input (z_1 on the following slide) represents the input received by an (artificial) neuron.
- This is transformed into the output h_1 by an **activation function** σ.
- Several parallel neurons constitute a **layer**.
- Several (connected) layers constitute a multi-layer neural network.
- **Learning** refers to finding the optimal parameters of the network.
- Types of layers:
  - input layers
  - hidden layers
  - output layers

Basic concepts: Neural networks

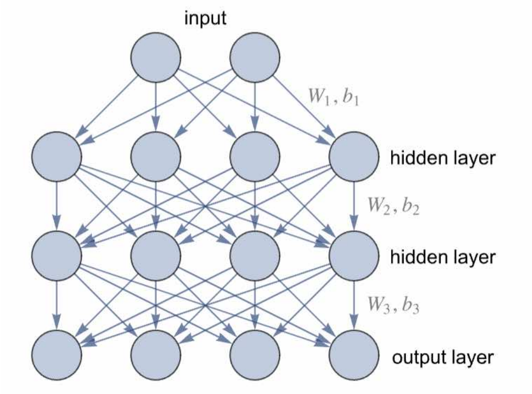


Source: (Bernard, 2021)



Source: (Ni *et al.*, 2021)

- A **deep neural network** has several hidden layers

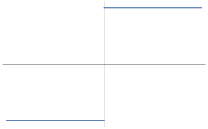


Source: (Bernard, 2021)

Basic concepts: Activation functions

- The **perceptron** is a neuron that acts as a binary linear classifier:

$$f(x) = \mathrm{sign}(Wx + b).$$

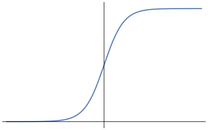


- **Heaviside function (step function)**:

$$f(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x = -1. \end{cases}$$

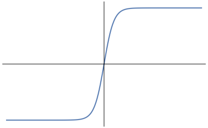
- **Sigmoid function**:

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$



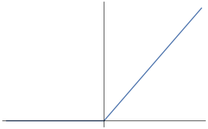
- **Tanh** as a generalisation of σ:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}.$$



- **ReLU** (rectified linear unit):

$$f(x) = \max(x,0).$$



- **Softmax** (normalised exponential function):

$$f(x_1, \dots, x_n) = \frac{1}{\sum_{i=1}^{n} e^{x_i}} \left(e^{x_1}, \dots, e^{x_n}\right)$$

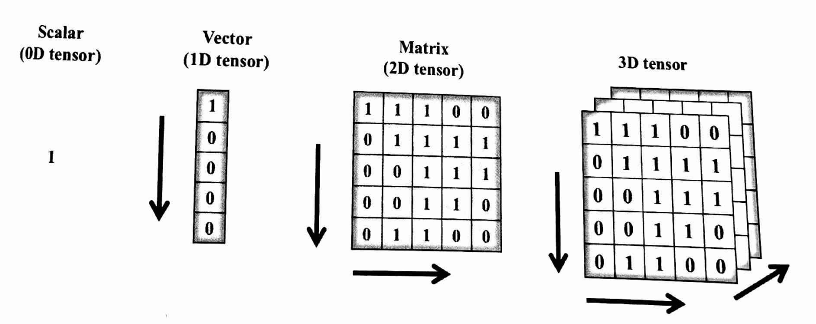
This is often used to model probability distributions and popular for classification problems.

- **Identity function**:

$$f(x) = x.$$

Basic concepts: Tensors

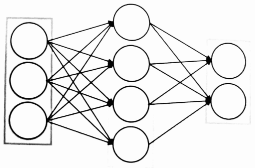
- All computations in neural networks boil down to **tensors**:



Source: (Ni *et al.*, 2021)

- **Shallow neural network** (2-layer artificial NN) has one hidden layer:
  - Input layer ($h^{(0)} : \mathbb{R}^d \to \mathbb{R}^d$) (identity):

$$x = \left(x^{(1)}, x^{(2)}, \dots, x^{(d)}\right) \mapsto x$$



  - Hidden layer ($h^{(1)} : \mathbb{R}^d \to \mathbb{R}^{n_1}$):

$$z^{(1)} = W^{(1)}x + b^{(1)}$$
$$h^{(1)} = \sigma_1(z^{(1)}),$$

  where $W^{(1)}$ is $n_1 \times d$, $b^{(1)}$ is n_1-dimensional;
  - Output layer ($h^{(2)} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$):

$$z^{(2)} = W^{(2)}h^{(1)} + b^{(2)}$$
$$h^{(2)} = \sigma_2(z^{(2)}),$$

E.g. identity for regression, Softmax for classification

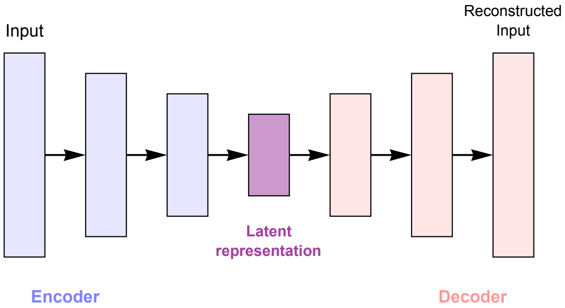
  where $W^{(2)}$ is $n_2 \times n_1$, $b^{(2)}$ is n_2-dimensional.

Shallow neural networks

- For non-linear regression, sigmoid function is natural choice for σ_1.
- One can show that a shallow NN is a **universal function approximator**, meaning it can model any suitably smooth function, given enough hidden units, to any desired level of accuracy, see e.g. (Hornik *et al.*, 1989; Cybenko, 1989; Hornik, 1991).

A simple NN: Autoencoder

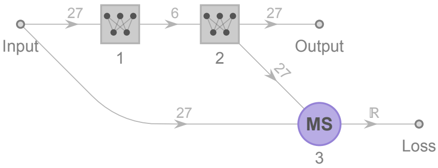
- An **autoencoder (AE)** is a simple feed forward neural network used for dimension reduction.
- Generalising PCA, it can capture nonlinear relationships.
- See e.g. (Bourlard and Kamp, 1988; Kramer, 1991; Hinton and Salakhutdinov, 2006).
- Schematic autoencoder representation:[22]



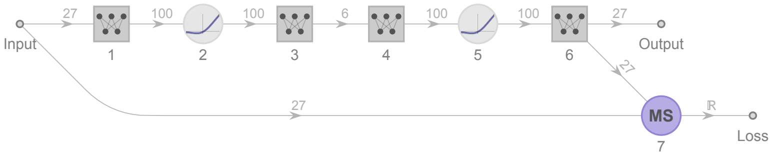
[22] https://reference.wolfram.com/language/ref/method/Autoencoder.html

A simple NN: Autoencoder

- PCA as AE:



- Optimal AE from training various architectures on test/validation data sets.



Source: (Packham, 2024)

Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
    5.1 Machine Learning Fundamentals
    5.2 Application areas of ML in Finance
    5.3 ML techniques
    5.4 Examples in Finance
        Credit risk modelling
        Algorithmic trading
    5.5 Explainable AI (XAI)

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
   - 5.4 Examples in Finance
     - Credit risk modelling
     - Algorithmic trading
   - 5.5 Explainable AI (XAI)

Kaggle competition: Home Credit Default Risk

- A good starting point for examples in credit risk is the **Kaggle competition** on **Home Credit Default Risk**.
- **Home Credit** launched two competitions (2018 and 2024).
- 2018 competition:
  - More than 8000 participants and a total of $70,000 in prizes.
  - https://www.kaggle.com/competitions/home-credit-default-risk/
  - https://www.kaggle.com/code/willkoehrsen/start-here-a-gentle-introduction
- 2024 competition:
  - More than 5000 participants and a total of $105,000 in prizes.
  - https://www.kaggle.com/competitions/home-credit-credit-risk-model-stability/
  - https://www.kaggle.com/code/jetakow/home-credit-2024-starter-notebook

Contents

1. Review of Statistical Foundations
2. Regression and time series analysis
3. Derivatives pricing and hedging
4. Option pricing with Monte Carlo simulation
5. Introduction to Machine Learning in Finance
   5.1 Machine Learning Fundamentals
   5.2 Application areas of ML in Finance
   5.3 ML techniques
   5.4 Examples in Finance
      Credit risk modelling
      Algorithmic trading
   5.5 Explainable AI (XAI)

Kaggle competition: Algorithmic trading

- Open Quant League Trading Competition:
- https://www.quantconnect.com/league/
- Open source, but not necessarily ML algorithms.

Contents

1. Review of Statistical Foundations

2. Regression and time series analysis

3. Derivatives pricing and hedging

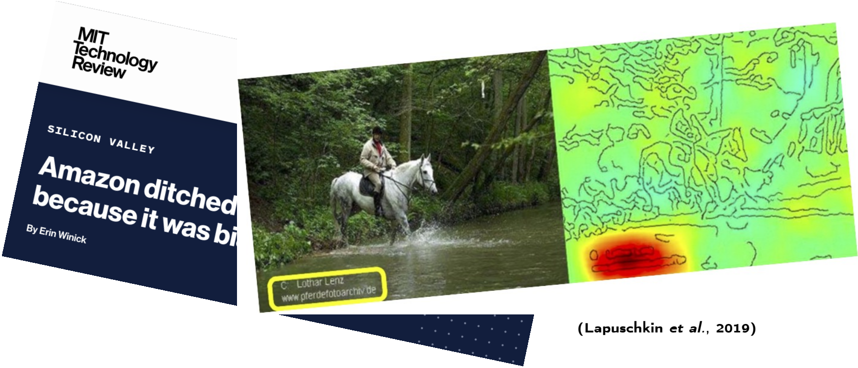
4. Option pricing with Monte Carlo simulation

5. Introduction to Machine Learning in Finance
   - 5.1 Machine Learning Fundamentals
   - 5.2 Application areas of ML in Finance
   - 5.3 ML techniques
   - 5.4 Examples in Finance
   - 5.5 Explainable AI (XAI)

Why explainability?

https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/

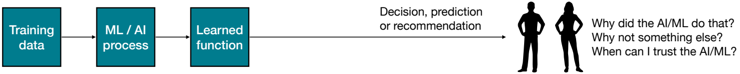
Why explainability?



(Lapuschkin *et al.*, 2019)

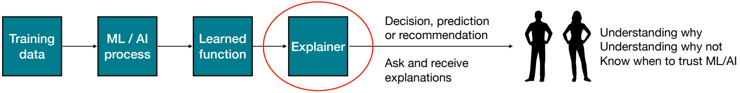
https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/

Explainability in the workflow



following (Gianfagna and Di Cecco, 2021)

Explainability in the workflow



following (Gianfagna and Di Cecco, 2021)

Interpretability, Explainability, causality...

- **Explainability**:
  - Human-agent interaction; explanatory agent revealing underlying causes to its or another agent's decision making (Miller, 2019).
  - Post-hoc interpretation; predictions without elucidating the mechanisms (Lipton, 2018).
- Distinguish explainability and interpretability, e.g. (Gianfagna and Di Cecco, 2021):
  - **Interpretability**: Understand (as a human) what the *model* does on known data.
  - **Explainability**: Broader understanding of the *model*, i.e., ability to understand counterfactuals.
- **Causality**: Understand causal effects in the *real world*.
- Not always possible to disentangle explainability and causality.

Examples from Finance

- **Buy-now-pay-later (BNPL)** systems make loan decisions in real time (Zalando, Klarna)
  $\rightsquigarrow$ unsecured loan to customer with no known credit history; revenue v fraud.
- **Risk management**, esp. stress-testing requires to know relationships between risk factors and portfolio value, e.g. (Packham and Woebbeking, 2019, 2023)
- **Model calibration** using deep neural networks in a real-time trading environment, e.g. (Brigo *et al.*, 2021; Yuan *et al.*, 2024).
- Examples, where explainability is less relevant:
  - A/B testing (???)
  - ???

Financial modelling: Why explainability?

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Observational data
- Noisy
- Heavy-tailed
- Jumps
- Non-stationary
- Autocorrelated

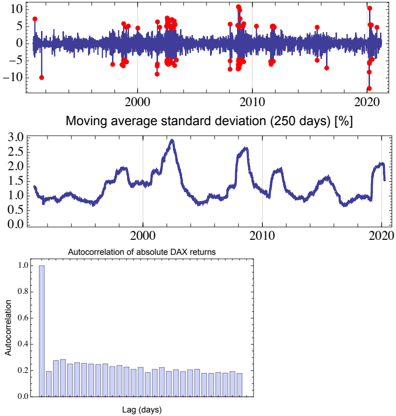
$$\frac{\mathrm{d}S_t}{S_{t-}} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t - \lambda_t^+ \mathsf{E}(\mathbf{e}^{J^+}-1)\,\mathrm{d}t - \lambda_t^- \mathsf{E}(\mathbf{e}^{J^-}-1)\,\mathrm{d}t + (\mathbf{e}^{J^+}-1)\,\mathrm{d}N_t^{(1)} + (\mathbf{e}^{J^-}-1)\,\mathrm{d}N_t^{(2)}$$

$$\begin{pmatrix}\mathrm{d}\lambda_t^+\\ \mathrm{d}\lambda_t^-\end{pmatrix} = \begin{pmatrix}\kappa^+(\theta^+-\lambda_t^+)\\ \kappa^-(\theta^--\lambda_t^-)\end{pmatrix}\mathrm{d}t + \begin{pmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{pmatrix}\begin{pmatrix}J^+\,\mathrm{d}N_t^{(1)}\\ J^-\,\mathrm{d}N_t^{(2)}\end{pmatrix},$$

$$J^+ \sim \varpi^+,\quad J^- \sim \varpi^-$$

(Liu, Packham, Sepp, 2024)

DAX log returns



Regulatory aspects

- **EU-AI Act**: Categorizes risk from AI systems as minimal, limited, high, and unacceptable risk.
- **High-Risk AI** includes access to services (e.g. insurance, banking, credit, ...).
- Key requirements:
  - Transparency
  - Human oversight
  - Accuracy
  - Robustness

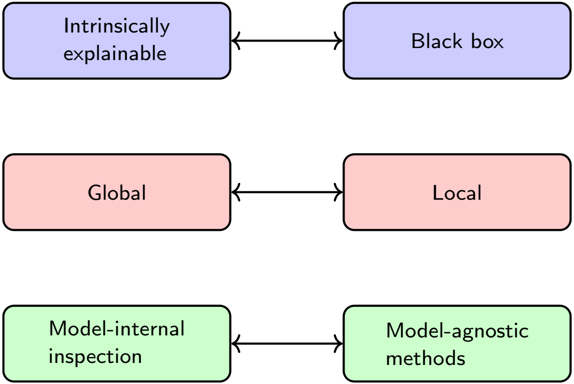
Regulatory aspects

- **EU-AI Act**: Categorizes risk from AI systems as minimal, limited, high, and unacceptable risk.
- **High-Risk AI** includes access to services (e.g. insurance, banking, credit, ...).
- Key requirements:
  - Transparency
  - Human oversight
  - Accuracy
  - Robustness

Do we really need explainability?

- **Model validation**: Does the model work as expected? Does the model manage extreme inputs? Is the model unbiased?
- **Model robustness**: Does the model work reasonably for unknown data? How does the model react to small changes in the inputs? Can we trust the model?
- **Knowledge discovery**: We are typically interested in more than prediction. What conclusions can we draw from model output?
- **"Human-in-the-loop"**: Collaboration of AI/ML and human may be much stronger than AI/ML alone.

Classification of XAI methods



Model-agnostic methods

- **Feature Importance**:
  - Measures how much each input feature contributes to the model's predictions.
  - Examples: **Permutation importance** (Breiman, 2001), **SHAP** (Shapley Additive Explanations) Lundberg and Lee (2017), **LIME** (Local Interpretable Model-Agnostic Explanations) (Ribeiro *et al.*, 2016).
- **Counterfactual Explanations**:
  - Provides "what-if" scenarios (e.g. what minimal change to the input would lead to a different output).
  - Useful for understanding decision boundaries and actionable insights.
- **Partial Dependence Plots (PDP)** (Friedman, 2001) and **Accumulated Local Effects (ALE)** (Apley and Zhu, 2020):
  - Show average effects of features.
- **Surrogate Models**:
  - Simplifies a complex model by approximating it with an interpretable model (e.g. decision trees, linear regression).

SHAP

- **Shapley values** (Shapley, 1953):
  - Concept in cooperative game theory
  - Method that allocates payouts (costs) to players depending on their contribution to the total payout (cost).
  - Players cooperate in a coalition and receive a certain profit from this cooperation.
- Example:
  - Alice wants to go to Adlershof; Bob wants to go to Britz; Charlie wants to go to Charlottenburg.
  - If they share a taxi that drops them off one at a time, what is a fair split-up of the total cost?
  - Shapley values: Marginal contributions of each person, averaged over all possible coalitions.

SHAP values

- **SHAP Values (Shapley Additive Explanations):** replace players by features, (Lundberg and Lee, 2017):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left(f(S \cup \{i\}) - f(S)\right),$$

  where

  - ϕ_i: SHAP value for feature i
  - N: set of all features
  - S: Subset of all features N that does not include feature i
  - $|S|$: Number of features in subset S
  - $f(S)$: Model prediction based on the features in subset S

Example: Real-estate model

- Cooperation with **Scope Ratings GmbH** (and HTW Berlin) on valuing and risk management of real estate portfolios using publicly available data
- Approx. 1.24m transactions in Île-de-France, 2014-2022
- Open Street Map (OSM) data
- Macroeconomic data
- XGBoost (eXtreme Gradient Boosting)

SHAP values XGB

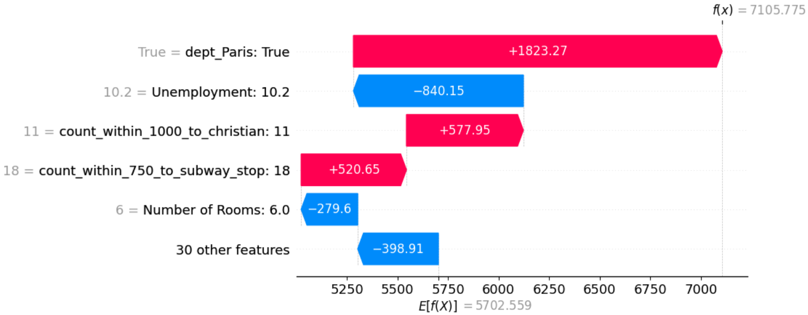


Figure: SHAP waterfall plot for a random apartment in Paris

SHAP values XGB

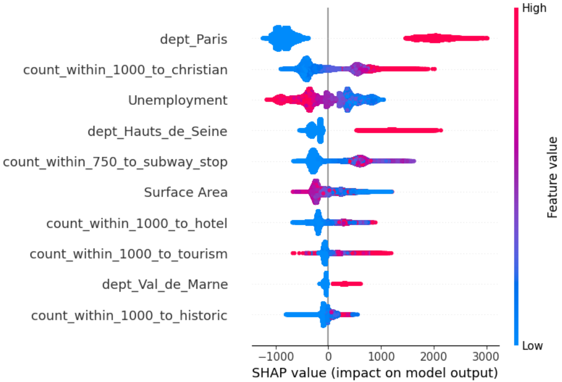
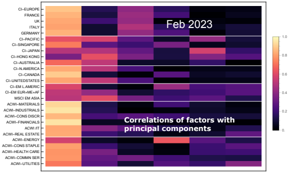
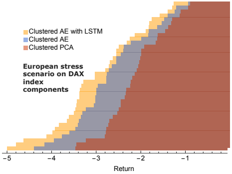


Figure: SHAP summary plot

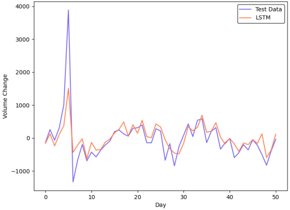
Further applications

- **Stress-testing, reverse stress testing**, (Packham, 2024):
  - Dimension reduction and aggregation of risk factors.
  - Explainable PCA and explainable auto-encoder.





- **Fractional trading**:
  - Cooperation with **Upvest GmbH** and HTW Berlin
  - Risk Management, esp. hedging of fractions
  - Drivers: News sentiment, trading volume, ...



Outlook

- Significance (in statistical sense) when data are dependent, see e.g. (Aas *et al.*, 2021).
- Interpretation of coefficients as in linear regression: Double machine learning, (Chernozhukov *et al.*, 2018)

References

Aas, K., M. Jullum, and A. Løland. Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *Artificial Intelligence*, 298:103502, September 2021.

Angrist, J. D. and J.-S. Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

Apley, D. W. and J. Zhu. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):1059–1086, June 2020.

Aronson, D. R. *Evidence-Based Technical Analysis*. Wiley & Sons, 2006.

Bernard, E. *Introduction to Machine Learning*. Wolfram Media, 2021.

Bourlard, H. and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4-5):291–294, sep 1988.

Breiman, L. Random forests. *Machine learning*, 45:5–32, 2001.

Brigo, D., X. Huang, A. Pallavicini, and H. S. d. O. Borde. Interpretability in deep learning for finance: a case study for the heston model, 2021.

Büchel, P., M. Kratochwil, M. Nagl, and D. Rösch. Deep calibration of financial models: turning theory into practice. *Review of Derivatives Research*, 25(2):109–136, August 2021.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, February 2019.

References

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Cunningham, S. *Causal Inference – The Mixtape*. Yale University Press, 2021.

Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, December 1989.

De Spiegeleer, J., D. B. Madan, S. Reyners, and W. Schoutens. Machine learning for quantitative finance: fast derivative pricing, hedging and fitting. *Quantitative Finance*, 18(10):1635–1643, July 2018.

Dumitrescu, E., S. Hué, C. Hurlin, and S. Tokpavi. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3):1178–1192, March 2022.

Fama, E. F. and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Funahashi, H. Deep learning for derivatives pricing: a comparative study of asymptotic and quasi-process corrections. *Annals of Operations Research*, July 2024.

Gianfagna, L. and A. Di Cecco. *Explainable AI with Python*. Springer, 2021.

References

Han, J., Y. Huang, S. Liu, and K. Towey. Artificial intelligence for anti-money laundering: a review and extension. *Digital Finance*, 2(3-4):211–239, 2020.

Hinton, G. E. and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, jul 2006.

Hornik, K., M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989.

Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

Hull, J. *Machine Learning in Business*. Independently published, 3rd edition, 2021.

Imbens, G. W. and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, April 2015.

Jacod, J. and P. Protter. *Probability Essentials*. Springer, 2nd edition, 2003.

James, G., D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

Jensen, R. I. T. and A. Iosifidis. Fighting money laundering with statistics and machine learning. *IEEE Access*, 11:8889–8903, 2023.

Kallenberg, O. *Foundations of modern probability*. Springer Science & Business Media, 2nd edition, 2002.

Koop, G. *Introduction to Econometrics*. Wiley, 2008.

References

Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, feb 1991.

Lapuschkin, S., S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 10(1), 2019.

Larsen, R. J., M. L. Marx, et al. *An introduction to mathematical statistics and its applications*. Prentice-Hall Englewood Cliffs, NJ, 5th edition, 2012.

Lipton, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, June 2018.

Liu, F., N. Packham, and A. Sepp. Jump risk premia in the presence of clustered jumps. Working paper, 2024.

Lundberg, S. M. and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4768–4777, 2017.

Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, February 2019.

Ni, H., X. Dong, J. Zheng, and G. Yu. *An Introduction to Machine Learning in Quantitative Finance*. World Scientific, 2021.

Packham, N. and C. F. Woebbeking. A factor-model approach for correlation scenarios and correlation stress testing. *Journal of Banking & Finance*, 101:92–103, 2019.

References

Packham, N. and F. Woebbeking. Correlation scenarios and correlation stress testing. *Journal of Economic Behavior & Organization*, 205:55–67, 2023.

Packham, N. Risk factor aggregation and stress testing. *Quantitative Finance*, 24(9):1327–1340, 2024.

Ribeiro, M. T., S. Singh, and C. Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144. ACM, 2016.

Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

Ruf, J. and W. Wang. Neural networks for option pricing and hedging: a literature review. *The Journal of Computational Finance*, 2020.

Shapley, L. S. A value for n-person games. *Contributions to the Theory of Games*, 2, 1953.

Shiryaev, A. N. *Probability*. Springer, Berlin, 2nd edition, 1996.

Torno, A., D. R. Metzler, and V. Torno. Robo-what?, robo-why?, robo-how? –a systematic literature review of robo-advice. In *Proc. Pacific Asia Conference on Information Systems*, 2021.

West, K. D. Asymptotic inference about predictive ability. *Econometrica*, pages 1067–1084, 1996.

White, H. A reality check for data snooping. *Econometrica*, pages 1097–1126, 2000.

Yuan, B., D. Brigo, A. Jacquier, and N. Pede. Deep learning interpretability for rough volatility, 2024.

Prof. Dr. Natalie Packham
Professor of Mathematics and Statistics
Berlin School of Economics and Law
Department of Business and Economics
Badensche Str. 52
10825 Berlin

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law