MFE Bootcamp

Tobias Ingebrigtsen

September 7, 2022

Abstract

These notes are written for part 2 of the UCLA Anderson Master of Financial Engineering Bootcamp. I have pulled together many concepts that will be useful for the MFE program and tried to make example code where I believe it might be useful. The material is gathered from lectures that I have taken myself, and some are my very own notes. If you find any typos (there are likely a lot of typos) or have any feedback, then please let me know.

Contents

1 Linear Algebra

Motivation: As quantitative economists, we often rely on linear algebra as a tool for solving simple and complex problems. For example, linear algebra allows us to find solutions to linear systems:

$$\begin{aligned} y_1 &= \phi_1 x_{11} + \phi_2 x_{12} + ... + \phi_k x_{1k}, \\ &\vdots \\ y_n &= \phi_1 x_{n1} + \phi_2 x_{n2} + ... + \phi_k x_{nk}, \end{aligned}$$

where we are interested in the unknowns ϕ_i given series of observables x_i and y_i. An example of a system of interest is the famous Capital Asset Pricing Model:

$$r_{i,t} - r^f = \alpha_i + \beta_i (r_{m,t} - r^f),$$

where $r_{i,t}$ is the return of asset i on time t, $r_{m,t}$ is the market return on time t, and r^f is the risk-free rate of interest. In this model we try to explain the excess return of all assets by each assets sensitivity to the market (more on this later). Important things to consider when we work with these problems:

- How many solutions (if any) exists?
- If there are no closed-form solution to this system, can we still approximate it?
- How do we compute the solution?

Linear algebra can also simplify notation. Consider the simple case of computing the variance of a portfolio of two assets: We know that the total variance of two variables, in this case stock 1 and stock 2, s_1 and s_2 can be computed as

$$\text{Var}(s_1 + s_2) = Var(s_1) + Var(s_2) + 2cov(s_1, s_2).$$

Now, let ω_1 and ω_2 be the portfolio weights of stock 1 and stock 2. The portfolio variance is now:

$$\text{Var}(\omega_1 s_1 + \omega_2 s_2) = \omega_1^2 Var(s_1) + \omega_2^2 Var(s_2) + 2\omega_1\omega_2 cov(s_1, s_2).$$

For an n-asset portfolio this becomes

$$\text{Var}(\omega_1 s_1 + ... + \omega_n s_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \omega_j Cov(s_i, s_j).$$

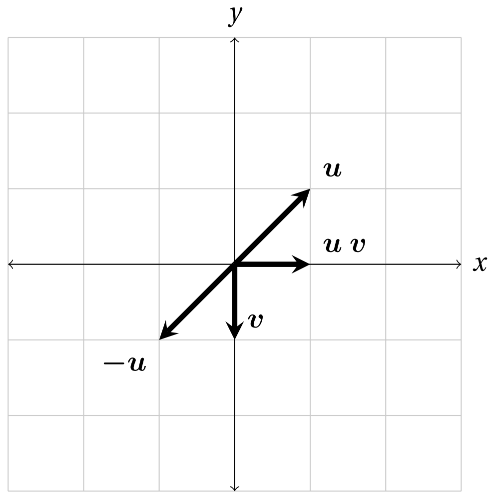
By using matrix notation we get

$$\text{Var}(\omega_1 s_1 + ... + \omega_n s_n) = \boldsymbol{\omega}' \boldsymbol{\Sigma} \boldsymbol{\omega},$$

where $\boldsymbol{\Sigma}$ is and $(n \times n)$ covariance matrix and $\boldsymbol{\omega}$ is a vector of weights.

Vectors

A vector in R^n is an n-tuple of numbers and represent a direction and magnitude.

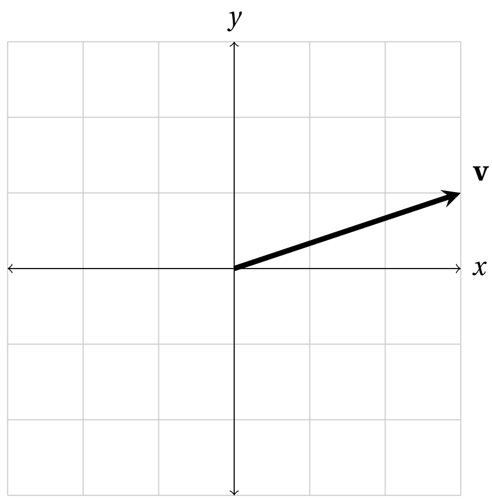


We can express the same vector in two ways:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad \mathbf{x}^t = [x_1, x_2, ..., x_3].$$

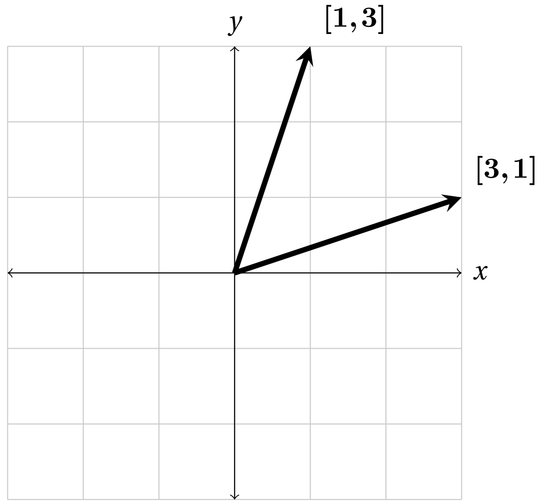
Example: In R^2 :

$$\mathbf{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$



The ordering matters:

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$



What can we do with vectors?

An n-vector, **u**, can be multiplied by a number c. This is called scalar multiplication and yields a new vector n-vector, $c\mathbf{u}$.

Example: (n=3) Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ and $c = 3$. Then,

$$3\mathbf{u} = 3\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3\times 2 \\ 3\times(-1) \\ 3\times 0 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \\ 0 \end{bmatrix}.$$

We can also add and subtract to vectors **u**, **v**:

$$\mathbf{u}' + \mathbf{v}' = (u_1 + v_1, u_2 + v_2, +\ldots+, u_n + v_n),$$
$$\mathbf{u}' - \mathbf{v}' = (u_1 - v_1, u_2 - v_2, -\ldots-, u_n - v_n).$$

Definition 1.1: Vector Space

Formally, a vector space over $\mathbb{R}$ is a tuple $(V, +, \cdot)$ where V is a set, +: $V \times V \to V$ and $\cdot$: $\mathbb{R} \times V \to V$ such that

1. $x + y = y + x$ for all $x, y \in V$,
2. $(x + y) + z = x + (y + z)$ for all $x, y \in V$,
3. there exist $0 \in V$ such that $x + 0$=x for all $x \in V$,
4. for every $x \in V$, there exist $-x \in V$ such that $x + (-x) = 0$,
5. $1 \cdot x = x$ for all $x \in V$,
6. $(ab) \cdot x = a \cdot (b \cdot x)$ for all $a, b \in \mathbb{R}$ and $x \in V$,
7. $a \cdot (x + y) = a \cdot x + a \cdot y$ for all $a \in \mathbb{R}$ and $x, y \in V$.
8. $(a + b) \cdot x = a \cdot x + b \cdot x$ for all $a, b \in \mathbb{R}$ and $x \in V$

Example: Let $n \in \mathbb{N}$. Let

$$V = \mathbb{R}^n = \{(x_1, ..., x_n) | x_1, ..., x_n \in \mathbb{R}\}.$$

Given $(x_1, ..., x_n), (y_1, ..., y_n) \in V$ and $a \in \mathbb{R}$, define $(x_1, ..., x_n) + (y_1, ..., y_n) \in V$ and $a \cdot (x_1, ..., x_n) \in V$ by

$$(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$$

and

$$a \cdot (x_1, ..., x_n) = (ax_1, ..., ax_n).$$

Then $(V, +, \cdot)$ is a vector space.

1.1 Exercise

1. Let S be a nonempty set. Let $V = \{f : S \to \mathbb{R}\}$. Given $f, g \in V$ and $a \in \mathbb{R}$, define $f + g \in V$ and $a \times f \in V$ by

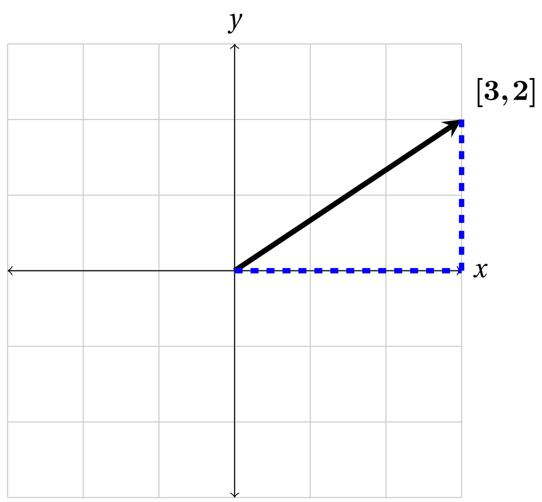
   $$(f+g)(s) = f(s) \quad \text{for } s \in S,$$

   and

   $$(a \cdot f)(s) = af(s) \quad \text{for } s \in S.$$

   Show that $(V, +, \cdot)$ is a vector space.

1.2 Length of a vector, inner product, and the Cauchy-Schwarz inequality



Lenght of a vector: This one is easy, we have a vector of length 2 and we remember the Pythagorean theorem:

$$c^2 = b^2 + a^2.$$

For a vector $\mathbf{x}$ of length n:

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

This is called the **Euclidean norm**. We can verify it for the example above where $\mathbf{x}' = (3,2) : \|\mathbf{x}\| = \sqrt{3^2 + 2^2}$, which is equal to what we learned using the Pythagorean theorem. Also, for any two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} - \mathbf{y}\|$ measures the distance between $\mathbf{x}$ and $\mathbf{y}$.

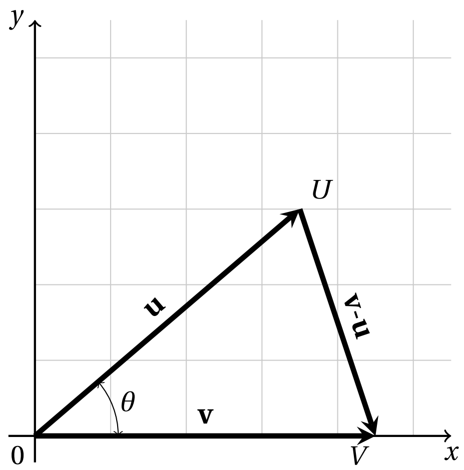
Inner product of two vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$ is defined by $\mathbf{v} \cdot \mathbf{u} = \sum_{i=1}^{n} x_i y_i$. For example, let ω be a vector of portfolio weights and $\mathbf{p}$ be a price vector for every asset in the portfolio. The value of the portfolio is then given by $\omega \cdot \mathbf{p} = \sum_{i=1}^{n} \omega_i p_i$.

Cauchy-Schwarz inequality:

$$\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|$$

In fact,

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cdot cos(\theta)$$



According to Pythagora's theorem, the angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is a right angle $(= 90°)$ if and only if $(OU)^2 + (OB)^2 = (VU)^2$, or $\|\mathbf{u}\|^2 + \|\mathbf{u}\|^2 = \|\mathbf{v} - \mathbf{u}\|^2$. This implies that $\theta = 90°$ if and only if

$$\mathbf{v} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{u} = (\mathbf{v} - \mathbf{u})(\mathbf{v} - \mathbf{u}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v}.$$

Because of symmetry, $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$, so we need $2\mathbf{u} \cdot \mathbf{v} = 0$ for the equality to hold, and so $\mathbf{u} \cdot \mathbf{v} = 0$. When the angle between two vectors $\mathbf{u}$ and $\mathbf{v}$ is $90°$, we say that they are **orthogonal.** If two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal we write $\mathbf{u} \perp \mathbf{v}$. Here we proved that two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are orthogonal if and only if their inner product is 0. For vectors in $\mathbb{R}^n$, we define orthogonality between $\mathbf{u}$ and $\mathbf{v}$ as

$$\mathbf{u} \perp \mathbf{v} \iff \mathbf{u} \cdot \mathbf{v} = 0.$$

Example: Suppose we have n observations of a commodity's price and quantity demanded $(p_1, d_1), (p_2, d_2)$, ..., (p_n, d_n). Define the means as

$$\overline{\mathbf{p}} = \frac{1}{n} \sum_{i=1}^{n} p_i, \quad \overline{\mathbf{d}} = \frac{1}{n} \sum_{i=1}^{n} d_i,$$

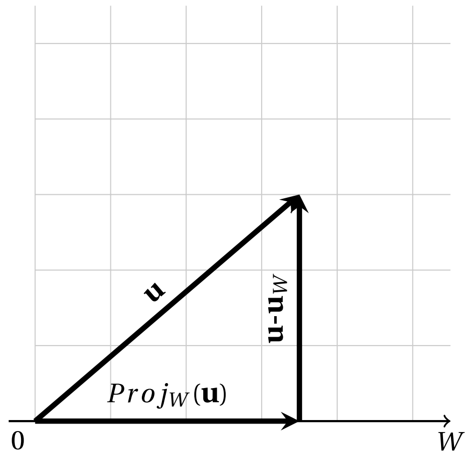
and

$$\mathbf{v} = (p_1 - \overline{\mathbf{p}}, p_2 - \overline{\mathbf{p}}, ..., p_n - \overline{\mathbf{p}}), \quad \mathbf{u} = (d_1 - \overline{\mathbf{d}}, d_2 - \overline{\mathbf{d}}, ..., d_n - \overline{\mathbf{d}}).$$

Then the **correlation coefficient** ρ can be computed as

$$\rho = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\| \cdot \|\mathbf{u}\|} = cos(\theta).$$

Projection



$\mathbf{u}_W$ is such that

1. $\mathbf{u}_W$ is parallell with W: There exist a scalar c such that $\mathbf{u}_W = cW$,
2. $\mathbf{u}_w \perp (\mathbf{u} - \mathbf{u}_W)$

Now we can find c, i.e. find $\mathbf{u}_w$. 2. means that

$$\begin{aligned} \mathbf{u}_W \cdot (\mathbf{u} - \mathbf{u}_W) &= 0, \\ \mathbf{u}_W \cdot \mathbf{u} - \mathbf{u}_W \cdot \mathbf{u}_W &= 0, \quad \text{(Distributive law (VS 7))} \\ \mathbf{u}_W \cdot \mathbf{u} &= \|\mathbf{u}_W\|^2. \quad (*) \end{aligned}$$

From 1. we have $\mathbf{u}_W = cW$. Insert $(*)$ into 1. to get $cW \cdot \mathbf{u} = \|cW\|^2 = c^2\|W\|^2$. Solve for c:

$$c = \frac{W \cdot \mathbf{u}}{\|W\|^2}.$$

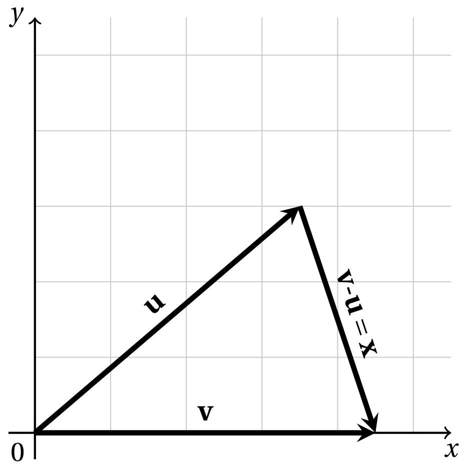
Conclusion:

$$Proj_W(\mathbf{u}) = \mathbf{u}_W = \frac{W \cdot \mathbf{u}}{\|W\|^2} \cdot W.$$

Example: Linear regression as a projection (later).

Linear independence of vectors, basis and linear subspace.

We have already seen that we can create a new vector $\mathbf{x}$ by two other vectors:



A natural question then becomes: how many other vectors can we create by linear combinations of $\mathbf{u}$ and $\mathbf{v}$?

Definition 1.2: Span

Assume $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ are n-vectors. Then, span$(\mathbf{v}_1, ..., \mathbf{v}_m)$ are all the vectors that are linear combinations of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ $(c_1\mathbf{v}_1, c_2\mathbf{v}_2, ..., c_m\mathbf{v}_m)$.

Example: (m=1) span$(\mathbf{v}_1)$ are all vectors $c_1\mathbf{v}_1$ where $c_1 \in \mathbb{R}$.

Exercise: (m=3) $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\3\\0\end{bmatrix}$. What is span$(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$?

Definition 1.3: Linear dependence

The vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ are **linearly dependent** if one or more of the vectors can be written as a linear combination of the others.

Example: $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\3\\0\end{bmatrix}$ are linear dependent since $\mathbf{v}_2 = \frac{1}{3}\mathbf{v}_3 - \frac{2}{3}\mathbf{v}_1$.

Definition 1.4: Basis

If the vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ are not linearly dependent we say that they are **linearly independent**. Then the vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ form a **basis** for span$(\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\})$.

Example: Consider the vectors $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\3\\0\end{bmatrix}$. $\{\mathbf{v}_1, \mathbf{v}_2\}$ are linearly independent (why?) therefore forms a basis for the xy-plane. Note: $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ spans $\mathbb{R}^2$ but is not a basis since they are linearly dependent.

Definition 1.5: Dimension

The number of vectors is in the basis is called the dimension of the linear subspace.

Example: Consider the vectors $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\3\\0\end{bmatrix}$. The dimension of span$(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is 2.

Linear systems

$$\begin{aligned} 2x_1 - 4x_2 + 2x_3 &= 12, \\ 3x_1 - 5x_2 - 2x_3 &= 5, \\ -4x_1 + 7x_2 + 4x_3 &= -2. \end{aligned}$$

Write $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -4 \\ -5 \\ 7 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 12 \\ 5 \\ -2 \end{bmatrix}$. The linear system now becomes

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3 = \mathbf{v}.$$

Question: Does the above system have a solution? (i.e. is $\mathbf{v}$ in span$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$?) $\rightarrow$ Gaussian Elimination.

Definition 1.6: Gaussian elimination

- a method for solving all linear systems.

1. Write the linear system as a matrix
2. From this matrix, create a new matrix in echelon form using elementary row operations
3. Translate the new matrix to a simpler linear system.
4. Solve this system using backwards substitution.

Definition 1.7: Echelon form

1. 0-rows are in the bottom rows of the matrix
2. Every pivot (leading coefficient) is longer to the right than the pivots in the rows above it (a pivot is the first number in a row that is non-zero.)

Definition 1.8: Free variables

The variables that corresponds to columns without pivots (but not the last column).

Example:

$$\begin{bmatrix} 6 & 4 & 0 & 1 \\ 0 & 4 & 3 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The above matrix is in echelon form, where the blue numbers are **pivots** and the variable corresponding to the third column is free.

Exercise: Write the linear system

$$\begin{aligned} 2x_1 - 4x_2 + 2x_3 &= 12, \\ 3x_1 - 5x_2 - 2x_3 &= 5, \\ -4x_1 + 7x_2 + 4x_3 &= -2, \end{aligned}$$

in matrix form and solve it using Gaussian elimination.

1.3 The solutions

Every linear system has either

1. No solution $\iff$ pivot in the last column.
2. A unique solution $\iff$ pivot in all columns except the last.
3. Infinitely many solutions $\iff$ no pivot in last column and at least one free variable.

1.4 Rank of a matrix

The rank of a matrix is the number of pivots in the echelon form.

Example:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & -1 & 3 \\ -4 & -3 & 9 & 5 \\ 6 & 12 & -3 & 9 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 3 \\ 0 & 5 & 7 & 11 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here we have to pivots, so rk(A)=2.

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Here we have three pivots, so rk(B)=3.

```
from numpy.linalg import matrix_rank
A = np.matrix('2 4 -1 3;-4 -3 9 5;6 12 -3 9')

A
Out[10]:
matrix([[ 2,  4, -1,  3],
        [-4, -3,  9,  5],
        [ 6, 12, -3,  9]])

matrix_rank(A)
Out[11]: 2

B = np.eye(3)

B
Out[16]:
array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])

matrix_rank(B)
Out[17]: 3
```

1.5 Applications

Let $\{\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_m\}$ be n-vectors.

1. The vectors are linearly independent $\iff x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + ... + x_m\mathbf{u}_m = 0$ only has the trivial solution $(x_1 = x_2 = ... = x_m = 0)$. [a system on the form $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + ... + x_m\mathbf{u}_m = 0$ is called a homogeneous system]
2. The vectors corresponding to pivot positions in the echelon form gives a basis for span($\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_m$)

Matrices

Definition 1.9: Transpose

The transpose of an (m×n)-matrix A gives an (n×m) matrix A' where the first column in A becomes the first row of A', etc.

Example: $A = \begin{bmatrix} 1 & 0 & -3 \\ 2 & 7 & 5, \end{bmatrix}, \quad A' = \begin{bmatrix} 1 & 2 \\ 0 & 7 \\ -3 & 5 \end{bmatrix}.$

```
import numpy as np

A = np.matrix('1 0 -3; 2 7 5')


Out[4]:
matrix([[ 1,  0, -3],
        [ 2,  7,  5]])

A.T
Out[5]:
matrix([[ 1,  2],
        [ 0,  7],
        [-3,  5]])
```

What can we do with matrices? Almost the same operations as with scalars, however **the order is important. Example:**

$$\begin{aligned} A(B+C) &= AB + AC, \\ (A+B)^2 &= A^2 + AB + BA + B^2, \\ (A')' &= A, \\ c(A+B) &= cA + cB, \quad etc. \end{aligned}$$

Important: AB is not necessarily equal to BA:

```
import numpy as np
A = np.matrix(np.random.randint(0,10, size=(3, 3)))
B = np.matrix(np.random.randint(0,10, size=(3, 3)))

A
Out[28]:
matrix([[0, 0, 1],
        [3, 0, 4],
        [5, 7, 9]])

B
Out[29]:
matrix([[8, 0, 8],
        [4, 5, 9],
        [3, 0, 4]])

C = np.matmul(A,B)
D = np.matmul(B,A)
C
Out[32]:
matrix([[  3,   0,   4],
        [ 36,   0,  40],
        [ 95,  35, 139]])

D
Out[33]:
matrix([[ 40,  56,  80],
        [ 60,  63, 105],
        [ 20,  28,  39]])
```

Linear systems using matrix notation (m equations, n unknowns)

$$
\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m.
\end{gathered}
$$

Write

$$
A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & & \ddots & \vdots \\ a_{11} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.
$$

We can now write the linear system above as

$$
A\mathbf{x} = \boldsymbol{b}.
$$

Definition 1.10: The identity matrix

Let $A = (a_{ij})$ be a **quadratic** $(n \times n)$ matrix where $a_{ij} = 1$ for $i = j$ and 0 otherwise:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is called the **identity matrix**. We then have that

$$AI = A = IA$$

Example:

```
import numpy as np
A = np.matrix(np.random.randint(0,10, size=(3, 3)))
I = np.eye(3)
A
Out[36]:
matrix([[5, 8, 7],
        [7, 1, 2],
        [5, 2, 1]])

I
Out[37]:
array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])

np.matmul(A,I)
Out[38]:
matrix([[5., 8., 7.],
        [7., 1., 2.],
        [5., 2., 1.]])

np.matmul(I,A)
Out[39]:
matrix([[5., 8., 7.],
        [7., 1., 2.],
        [5., 2., 1.]])
```

Definition 1.11: Invertability of a matrix

An $(n \times n)$ matrix A is invertible if there exist an $(n \times n)$ matrix B such that

$$AB = I = BA$$

If A is invertible, B is unique such that $AB = I = BA$.

```
import numpy as np
A = np.matrix(np.random.randint(0,10, size=(3, 3)))
B = np.linalg.inv(A)

A
Out[41]:
matrix([[1, 6, 1],
        [8, 7, 6],
        [4, 1, 7]])

B
Out[42]:
matrix([[-0.25443787,  0.24260355, -0.17159763],
        [ 0.18934911, -0.01775148, -0.01183432],
        [ 0.1183432 , -0.13609467,  0.24260355]])

np.matmul(A,B)
Out[43]:
matrix([[ 1.00000000e+00, -2.77555756e-17,  0.00000000e+00],
        [ 2.22044605e-16,  1.00000000e+00,  5.55111512e-17],
        [ 5.55111512e-17, -2.77555756e-17,  1.00000000e+00]])

np.matmul(B,A)
Out[44]:
matrix([[ 1.00000000e+00,  1.38777878e-16,  0.00000000e+00],
        [-6.93889390e-18,  1.00000000e+00, -5.20417043e-18],
        [-1.11022302e-16, -5.55111512e-17,  1.00000000e+00]])
```

The result above is very useful in economics since it allows us to solve linear systems very easily. Remember that we can write a linear system on the form

$$A\mathbf{x} = \boldsymbol{b}.$$

We might ask ourselves, what **x** will solve the system above? That is, we want to isolate **x** on the left side of the equation. Let B be the inverse of A (we often write A^{-1}), then:

$$\begin{aligned} A\mathbf{x} &= \boldsymbol{b}, \\ BA\mathbf{x} &= B\boldsymbol{b}, \\ I\mathbf{x} &= B\boldsymbol{b}. \end{aligned}$$

⇒ a very easy way to solve a system!

Definition 1.12: Determinants

For every (n×n) matrix A there exist a number det(A) (often $|A|$) so that the following is true:

1. det(AB) = det(A)· det(B)
2. det(A’) = det(A)

Theorem: A invertible $\iff$ det(A)$\neq 0$.

If the determinant of A is not zero, then we say that is **nonsingular**.

Useful results: for an (n×n) matrix A where det(A)= 0:

1. The columns of A are dependent vectors in $\mathbb{R}^n$
2. The rows of A are dependent vectors in $\mathbb{R}^n$
3. A is not invertible

Definition 1.13: Eigenvalues and eigenvectors

The (n×n) matrix A has the eigenvalue λ (a number) if the equation

$$A\mathbf{x} = \lambda\mathbf{x}$$

has a non-trivial solution (i.e. $\mathbf{x} = \boldsymbol{b} \neq \mathbf{0}$). Then the vector $\boldsymbol{b}$ is an eigenvector to A with eigenvalue λ.

Example:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \quad \text{then} \quad A\mathbf{x} = 3\mathbf{x},$$

so $\lambda = 3$ is an eigenvalue to A and all 2-vectors are eigenvectors to A with eigenvalue 3.

Example:

$$A = \begin{bmatrix} 4 & 2 \\ -3 & -1 \end{bmatrix}.$$

For what number λ does the system

$$A\mathbf{x} = \lambda\mathbf{x}$$

have non-trivial solutions?

$$\begin{aligned}
&\text{i.e.} \quad \begin{bmatrix} 4 & 2 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} \\
&\text{i.e.} \quad \begin{cases} 4x_1 + 2x_2 = \lambda x_1 \\ -3x_1 - x_2 = \lambda x_2 \end{cases} \\
&\text{i.e.} \quad \begin{cases} (4-\lambda)x_1 + 2x_2 = 0 \\ -3x_1 - (1+\lambda)x_2 = 0 \end{cases} \\
&\text{i.e.} \quad \underbrace{\begin{bmatrix} 4-\lambda & 2 \\ -3 & -(1+\lambda) \end{bmatrix}}_{B} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \text{a homogeneous system!}
\end{aligned}$$

Generally: A homogeneous system $B\mathbf{x} = \mathbf{0}$ has non-trivial solutions $\iff$ det(B)= 0. **Question:** Why? Therefore,

$$|B| = 0 \iff (4-\lambda)(-(1+\lambda)) - 2 \cdot (-3) = \lambda^2 - 3\lambda + 2 = 0.$$

We solve this second-order equation to get

$$\lambda = \frac{-(-3) \pm \sqrt{(-3)^2 - 4 \cdot 2}}{2} = \begin{cases} 1 \\ 2 \end{cases}$$

So $\lambda = 1$ and $\lambda = 2$ are the eigenvalues of A.

Exercise: Find the eigenvectors of A.

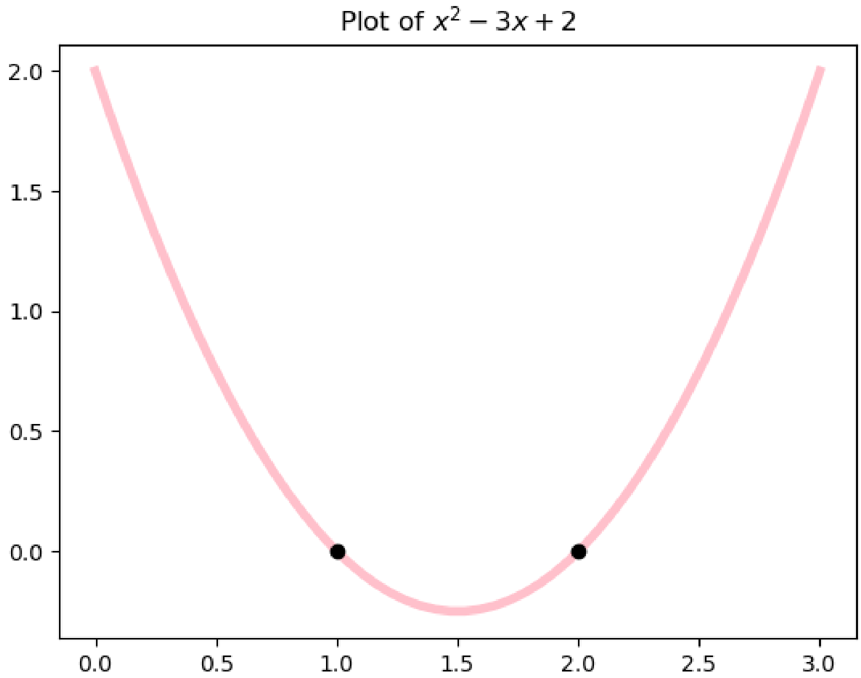


Figure 1: The roots are given by the black dots

1.5.1 Sidenote: Root-finding algorithms

In the previous example we solved a second-order equation of the form

$$x^2 - 3x + 2 = 0.$$

Although these can be straight-forward computed by hand, it's easy to make mistakes (and who solves problems by hand these days anyways?). Moreover, this becomes **very** complicated as the degree of the polynomial increase. A useful tool is therefore a root-finder. A root-finding algorithm is an algorithm for finding zeroes ("roots") of continuous functions.

```
import numpy as np

# Find the roots of f(x) = x^2 -3x +2
coeff = [1,-3,2]
np.roots(coeff)

Out[7]: array([2., 1.])
# The roots are 2 and 1.

# The figure
import matplotlib.pyplot as plt
def f(x):
    return (x**2 -3*x +2 )

x = np.linspace(0,3,50)
y = f(x)

plt.figure()
```

```
plt.plot(x,y,c='pink',linewidth=4, markevery=[1,2])
plt.plot(1,0,"-o",c='black')
plt.plot(2,0,"-o",c='black')
plt.title('Plot␣of␣$x^2-3x␣+2$')
```

1.6 Quadratic forms

Quadratic forms are polynomials where all terms are of degree two. Examples:

1. $Q_1(\mathbf{x}) = 3x_1^2 + 8x_1x_2 + 5x^2$
2. $Q_2(\mathbf{x}) = x_1^2 - 3x_2^2 + 5x_3^2 + 7x_2x_3 + 9x_1x_3 - 11x_1x_2$

For every quadratic form $Q(\mathbf{x})$ in n variables there exist a symmetrix $(n \times n)$ matrix A such that

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{so } \mathbf{x}' = [x_1, x_2, \ldots, x_n].$$

Example:

$$A = \begin{bmatrix} 3 & 4 \\ 4 & 5 \end{bmatrix}.$$

$$\mathbf{x}'A\mathbf{x} = [x_1, x_2] \begin{bmatrix} 3 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 8x_1x_2 + 5x_2^2 = Q(\mathbf{x}).$$

1.7 Definiteness of quadratic forms

Let $Q(\mathbf{x})$ be a quadratic form with symmetric matrix A.

Definition 1.14: Definiteness

$Q(\mathbf{x})$ is positive semidefinite definite if $Q(\mathbf{x}) \geq 0$ for all values of $\mathbf{x}$ (and positive definite if $Q(\mathbf{x}) > 0$ for all $x \neq 0$.)

$Q(\mathbf{x})$ is negative semidefinite definite if $Q(\mathbf{x}) \leq 0$ for all values of $\mathbf{x}$ (and negative definite if $Q(\mathbf{x}) < 0$ for all $x \neq 0$.)

If $Q(\mathbf{x})$ is neither positive or negative semidefinite, then it is indefinite.

Example:

$$Q(\mathbf{x}) = 3x_1^2 + 5x_2^2 > \forall \mathbf{x} \neq \mathbf{0},$$

i.e. $Q(\mathbf{x})$ is positive definite (and positive semidefinite).

What about $Q(\mathbf{x}) = 3x_1^2 + 8x_1x_2 + 5x_2^2$? $8x_1x_2$ is negative if x_1 and x_2 have opposite signs - it's not clear when we have crossterms in $Q(\mathbf{x})$.

Theorem Let $\lambda_1, \lambda_2, ..., \lambda_n$ be the eigenvalues of A. Then, $Q(\mathbf{x})$

$$\begin{aligned} \text{positive semidefinite} &\iff \text{all } \lambda_i \geq 0 \\ \text{positive definite} &\iff \text{all } \lambda_i > 0 \\ \text{negative semdefinite} &\iff \text{all } \lambda_i \leq 0 \\ \text{negative definite} &\iff \text{all } \lambda_i < 0 \\ \text{indefinite} &\iff \lambda_i < 0 \text{ and } \lambda_j > 0 \text{ for some } i, j. \end{aligned}$$

2 Optimization

We are often faced with optimization problems in economics. A classic example is consumption and investment decisions, where an agent can choose to either consume some (or even all) of her wealth, and invest the remaining in assets which she can use to consume next period. This gives rise to yet another problem; what portfolio should she invest in? Should it be equal-weighted in risky stocks? Should she perhaps allocate some of her wealth to risk-free T-bills? These questions can be answered using optimization.

Definition 2.1: Extreme points

If $f(x)$ has domain D then

$$c \in D \text{ is a } \textbf{maximum point} \text{ for } f \iff f(x) \leq f(c) \text{ for all } x \in D,$$
$$d \in D \text{ is a } \textbf{minimum point} \text{ for } f \iff f(x) \geq f(d) \text{ for all } x \in D.$$

If the value of f at c is strictly larger than at any other point in D, then c is a **strict maximum point**. Similarly,d is a **strict minimum point** if $f(x) > f(d)$ for all $x \in D$, $x \neq 0$.

Note: If f is any function with domain D: $f(x) \leq f(c)$ for all x in D iff $-f(x) \geq -f(c)$ for all x in D. Implication: c maximizes f in D iff c minimizes $-f$ in D. This is sometimes useful in numerical optimization if the optimizer only minimize.

Example: find the extreme points for $g(x) = 3-(x-2)^2$.

Solution:

Since $(x-2)^2 \geq 0$ for all x, it follows that $f(x) \leq 3$ for all x. But $f(x) = 3$ when $(x-2)^2 = 0$ at $x = 2$. Therefore, $x = 2$ is a **maximum point** for f. Because $f(x) \to -\infty$ as $x \to \pm\infty$, f has no minimum.

Example: find the extreme points for $f(x) = \sqrt{x-5} - 100, \quad x \geq 5$.

Solution:

Because $\sqrt{x-5}$ is ≥ 0 for all $x \geq 5$, it follows that $f(x) \geq -100$ for all $x \geq 5$. Since $f(5) = -100$, we conclude that $x = 5$ is a **minimum point**. Since $f(x) \to \infty$ as $x \to \infty$, f has no maximum.

Definition 2.2: Derivative of a function

The derivative of a function f at point a, denoted by $f'(a)$, is given by the formula

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

$f'(a)$ = slope of the tangent to the curve $y = f(x)$ at the point $(a, f(a))$.

Definition 2.3: Necessary first-order condition

Suppose that a function f is differentiable in an interval I and that c is an interior point for I. For $x = c$ to be a maximum or minimum point for f in I, a necessary condition is that it is a stationary point for f, i.e. that $x = c$ satisfies the equation

$$f'(x) = 0.$$

We call this the **first-order condition.**

Definition 2.4: First-derivative test for max/min

If $f'(x) \geq 0$ for $x \leq c$ and $f'(x) \leq 0$ for $x \geq c$, then c is a maximum point for f. If $f'(x) \leq 0$ for $x \leq c$ and $f'(x) \geq 0$ for $x \geq c$, then c is a minimum point for f.

Exercise

Consider the function f defined for all x by

$$f(x) = e^{2x} - 5e^{x} + 4$$

(a) Find the zeros of $f(x)$ and compute its derivative $f'(x)$.
(b) Find the intervals where f increases and decreases, and determine possible extreme points and values.
(c) Examine the limit of $f(x)$ as $x \to -\infty$. Sketch the graph of f.
(d) Repeat (a)-(c) on the computer using Python. Try to avoid pre-programmed routines but rather write your own functions.

Definition 2.5: Second-order condition

Suppose that f is a $\mathscr{C}^2$ function (zeroth, first, and second derivative of f are continuous) and x be an interior point of I. Then:

(a) $f'(x) = 0$ and $f''(x) < 0 \Rightarrow$ c = x is a **strict** local maximum point
(b) $f'(x) = 0$ and $f''(x) > 0 \Rightarrow$ c = x is a **strict** local minimum point
(c) $f'(x) = 0$ and $f''(x) = 0 \Rightarrow$?

Exercise

(a) Classify the stationary points of

$$f(x) = {}^{1}\!/_{9}x^3 - {}^{1}\!/_{6}x^2 - {}^{2}\!/_{3}x + 1$$

by using the second-derivative test.
(b) Find a two examples that satisfies (c) while being different local extreme points.

Definition 2.6: Extreme Value Theorem

Suppose that f is a continuous function over a closed and bounded interval $[a, b]$. Then there exist a point d in $[a, b]$ where f has a minimum, and a point c in $[a, b]$ where f has a maximum, so that

$$f(d) \leq f(x) \leq f(c) \qquad \text{for all } x \text{ in } [a, b].$$

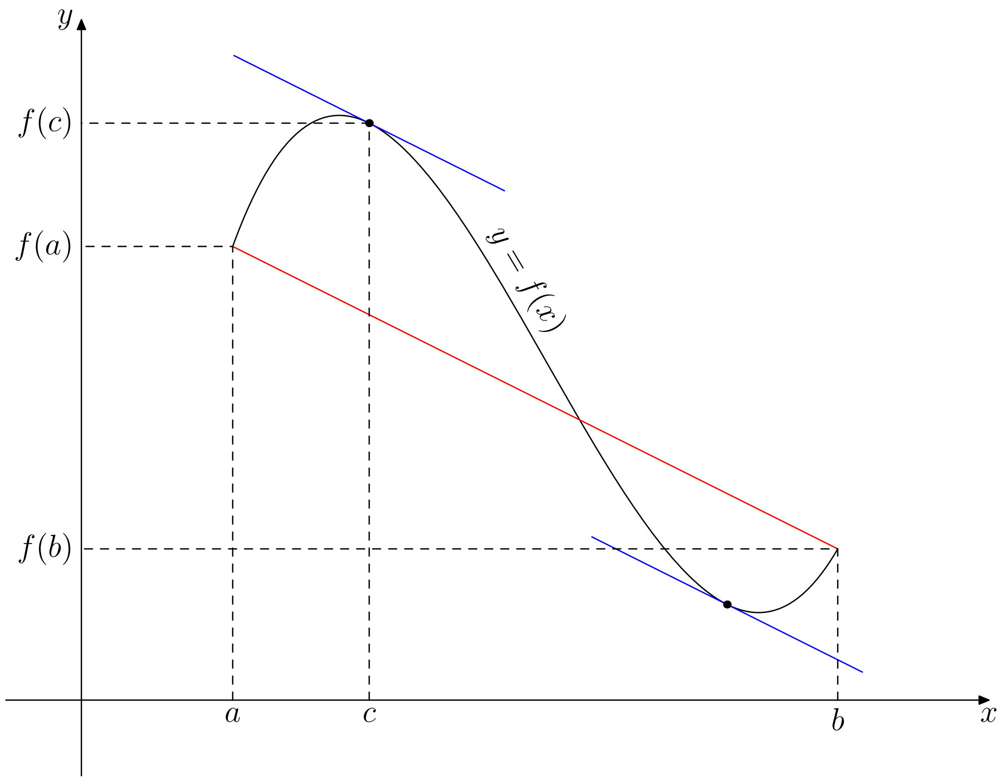
This result is useful in **constrained optimization problems** where we consider optimization problems over closed and bounded intervals.

Definition 2.7: The Mean Value Theorem

If f is continuous in the closed bounded interval $[a, b]$, and differentiable in the open interval (a, b), then there exists at least one interior point y in (a, b) such that

$$f'(y) = \frac{f(b) - f(a)}{b - a}.$$

"Proof":



2.1 Stationary points of quadratic functions

A second-order function:

$$f(\mathbf{x}) = \underbrace{Q(\mathbf{x})}_{\text{quadr. form}} + \underbrace{L(\mathbf{x})}_{\text{linear form}} + \underbrace{c}_{\text{constant}}$$
$$= \mathbf{x}' A\mathbf{x} + B\mathbf{x} + c.$$

Then,

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix} = 2A\mathbf{x} + B'.$$

Example:

$$f(x_1, x_2) = x_1^2 - 4x_1x_2 + 3x_2^2 + 7x_1 - 8x_2 + 5$$

which can be written as

$$[x_1, x_2] \begin{bmatrix} 1 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [7, -8] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 5.$$

$$\frac{\partial f}{\partial x_1} = 2x_1 - 4x_2 + 7,$$
$$\frac{\partial f}{\partial x_2} = 6x_2 - 4x_1 - 8,$$

so

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} 2x_1 - 4x_2 + 7 \\ 6x_2 - 4x_1 - 8 \end{bmatrix}.$$

$$2A\mathbf{x} = \begin{bmatrix} 2x_1 - 4x_2 \\ -4x_1 + 6x_2 \end{bmatrix}, \quad B' = \begin{bmatrix} 7 \\ -8 \end{bmatrix},$$

so

$$\frac{\partial f}{\partial \mathbf{x}} = 2A\mathbf{x} + B'.$$

2.2 Classification of stationary points

$$\frac{\partial f}{\mathbf{x}} = 0 \quad \text{i.e.} \quad 2A\mathbf{x} + B' = 0$$

which is the linear system

$$A\mathbf{x} = -\frac{1}{2}B'.$$

$$f \text{ convex} \iff A \text{ positive semidefinite}$$
$$f \text{ concave} \iff A \text{ negative semidefinite}$$

If f convex:	All stationary points are global minimum.
If f concave:	All stationary points are global maximum.
If f indefinite:	All stationary points are saddle points.

3 Linear Regression

We are often interested in modeling relationships. For example, what is the relationship between returns and risk (what is risk?), what is the relationship between prices and quantities demanded, or does past returns inform us about future return dynamics? One of the workhorses in modeling relationships is linear regression.

3.1 Regression vs correlations

Why do we need regressions when we can easily compute correlations? Correlation between two variables x and y is the degree of linear association between them. This means that we treat the two variables in a completely symmetric way and we do not imply that the are any causal effect (i.e. that changes in x *causes* changes in y, or vice-versa.), but we rather say that there is evidence of a linear relationship, i.e. that they are related to an extent given by the correlation coefficient. When thinking in terms of regressions, the dependent variable (usally denoted by the vector $\boldsymbol{y}$) is treated as random and the independent variable(s) (usally denoted by the matrix X) is treated as fixed. This allows us to model relationships in a more meaningful way.

3.2 The Linear Regression Model

$$y_i = x_i'\beta + \varepsilon_i \quad i = 1, ..., N$$

x_i is a vector of observations on k independent variables. Usually, we include an intercept, so

$$x_i^t = [1, x_{1,i}, ..., x_{k,i}].$$

The conditional mean function, $r(x)$, then maps $\mathbb{R}^k \to \mathbb{R}$. Take the conditional expectation of both sides of the model:

$$\mathbb{E}[y|x] = \mathbb{E}[x'\beta + \varepsilon|x] = x'\beta + \mathbb{E}[\varepsilon|x].$$

If the conditional expectation of the error term, $\mathbb{E}[\varepsilon|x]$, is zero then we can interpret the linear term as the regression function, i.e. $r(x) = x'\beta$. If we include a constant term in the regressor matrix then this exogeneity condition will always be met.

We usually include another assumption, the homoskedasticity assumption:

$$\mathbb{V}[\varepsilon|x] = \sigma^2 < \infty.$$

Homoskedasticity refers to a constant variance (is this a reasonable assumption in financial time series?).

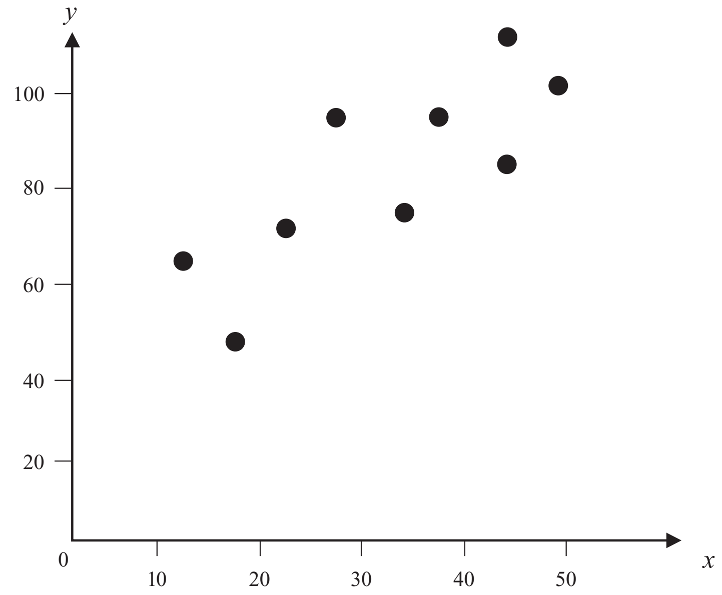
3.3 The Gauss-Markov Theorem

If the following assumptions about the residuals are true:

1. $\mathbb{E}[\varepsilon_i] = 0,$
2. $\mathbb{V}[\varepsilon_i] = \sigma^2 < \infty,$
3. $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j,$

then the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

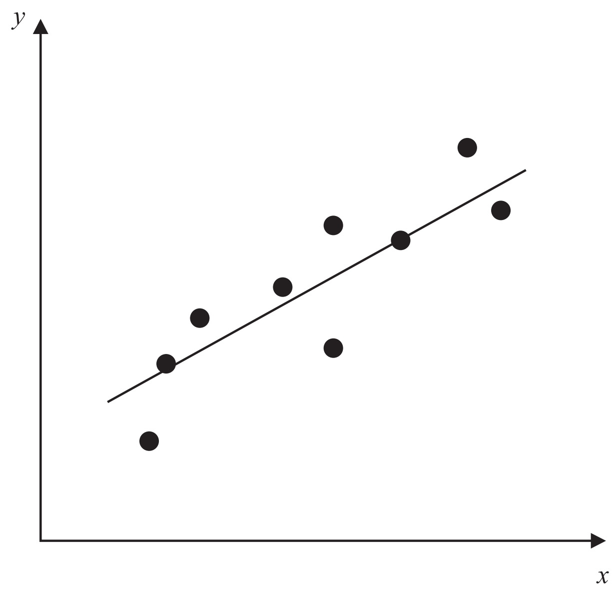
3.4 Estimation of β: Ordinary Least Squares - Univariate case



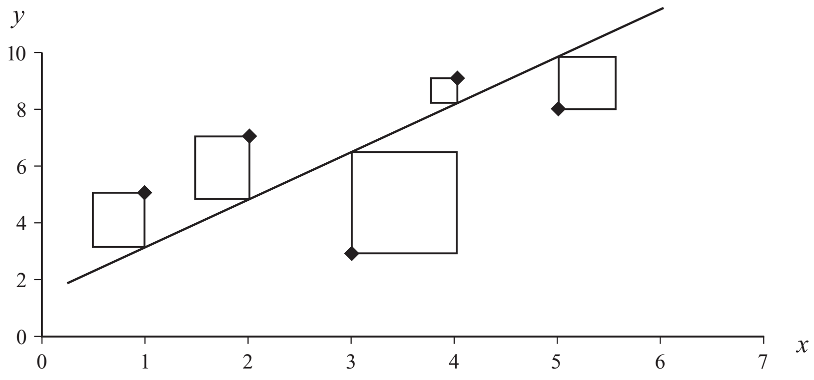
Let's say we are interested in estimating the following model:

$$y_i = \alpha + x_i \beta + \varepsilon_i.$$

What this means is that we are looking for the estimates of α and β that best captures the linear relationship between the two variables:



The method of Ordinary Least Squares (OLS) entails taking each vertical distance from the point to the line, squaring it and then minimizing the sum of the squares:



Write the fitted values of y_i as:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i.$$

The residual, ε_i, is then defined as

$$\varepsilon_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i.$$

The problem of finding the $\hat{\alpha}$ and $\hat{\beta}$ that minimizes the sum of the squares can be written as

$$\underset{\alpha,\beta}{\arg\min} \sum_{i=i}^{N} \varepsilon_i^2 = (y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

Let L denote the objective function, $\sum_{i=i}^{N} \varepsilon_i^2$. FOC:

$$\frac{\partial L}{d\alpha} = \sum_{i=1}^{N} -2(y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \tag{1}$$

$$\frac{\partial L}{d\beta} = \sum_{i=1}^{N} -2x_i(y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \tag{2}$$

From (1):

$$\begin{aligned}
\sum_{i=1}^{N} -2(y_i - \hat{\alpha} - \hat{\beta} x_i) &= 0 \\
\sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) &= 0 \\
N\overline{y} - N\hat{\alpha} - N\hat{\beta}\overline{x} &= 0 \\
\hat{\alpha} &= \overline{y} - \beta\overline{x}.
\end{aligned}$$

Now, insert this into (2):

$$\sum_{i=1}^{N} -2x_i(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$
$$\sum_{i=1}^{N} -2x_i(y_i - (\overline{y} - \hat{\beta}\overline{x}) - \hat{\beta}x_i) = 0$$
$$\sum_{i=1}^{N} -2x_i(y_i - \overline{y} + \hat{\beta}\overline{x} - \hat{\beta}x_i) = 0$$
$$\sum_{i=1}^{N} x_i(y_i - \overline{y} + \hat{\beta}\overline{x} - \hat{\beta}x_i) = 0$$
$$\sum_{i=1}^{N} x_i y_i - \overline{y}x_i + \hat{\beta}\overline{x}x_i - \hat{\beta}x_i^2 = 0$$
$$\sum_{i=1}^{N} x_i y_i - \overline{y}\sum_{i=1}^{N} x_i + \hat{\beta}\overline{x}\sum_{i=1}^{N} x_i - \hat{\beta}\sum_{i=1}^{N} x_i^2 = 0$$
$$\hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}^2}$$
$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2}.$$

We see that that the slope coefficient is the equivalent to the sample covariance between x and y divided by the sample variance of x. We also see from the equation for $\hat{\alpha}$ that the regression line will go through the mean of the observations, i.e. that the point $(\overline{y}, \overline{x})$ lies on the regression line.

3.5 Estimation of β: Ordinary Least Squares - General case

We can write the model as

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon,$$

where $\mathbf{y}$ is a vector of the dependent variable, X is an (N×k) matrix of the independent variables and a constant column, and ε is a vector of the residuals. $\boldsymbol{\beta}$ is the vector of estimates that we are interested in estimating. The objective is still to minimize the sum of the squared residuals:

$$\underset{\boldsymbol{\beta}}{\arg\min}\, \varepsilon'\varepsilon = (\mathbf{y} - \boldsymbol{X}\boldsymbol{\beta})'(\mathbf{y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

FOC:

$$\frac{\partial \varepsilon'\varepsilon}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\mathbf{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0$$
$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\mathbf{y}.$$

Now we need to isolate $\boldsymbol{\beta}$. We do this by using the inverse trick:

$$A^{-1}A = I.$$

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\mathbf{y}$$
$$I\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y}$$
$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y}.$$

3.6 Sampling Properties of OLS

Let $\boldsymbol{b}$ denote the least squares estimator. Then,

$$\begin{aligned}
\boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y},\\
\boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\left(\boldsymbol{X}\beta+\varepsilon\right),\\
\boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon,\\
\boldsymbol{b} &= \beta+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\varepsilon
\end{aligned}$$

Taking the conditional expectation gives us

$$\begin{aligned}
\mathbb{E}[\boldsymbol{b}|\boldsymbol{X}] &= \boldsymbol{\beta}+\mathbb{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\varepsilon|\boldsymbol{X}\right],\\
&= \boldsymbol{\beta}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\mathbb{E}\left[\boldsymbol{X}'\varepsilon|\boldsymbol{X}\right],\\
&= \boldsymbol{\beta}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left[\varepsilon|\boldsymbol{X}\right],\\
&= \boldsymbol{\beta},
\end{aligned}$$

due to the exogeneity condition. Thus, the Least Squares estimator is unbiased. This also holds unconditionally since $\mathbb{E}[b]=\mathbb{E}_X\left[\mathbb{E}[b|X]\right]=\mathbb{E}_X[\beta]=\beta$ by the Law of Iterated Expectations. What is the uncertainty associated with the estimates?

$$\begin{aligned}
\mathbb{V}(\boldsymbol{b}) &= \mathbb{E}\left[(\boldsymbol{b}-\beta)^2|\boldsymbol{X}\right],\\
&= \mathbb{E}\left[(\boldsymbol{\beta}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\varepsilon-\beta)^2|\boldsymbol{X}\right],\\
&= \mathbb{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\varepsilon\varepsilon'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}|\boldsymbol{X}\right],\\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left[\varepsilon\varepsilon'|\boldsymbol{X}\right]\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},\\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{V}(\varepsilon)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},\\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Lambda\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},
\end{aligned}$$

where Λ is the covariance matrix of the residuals. If we assume that the errors are homoskedastic and not autocorrelated, then Λ is simply $I\sigma^2$, thus:

$$\mathbb{V}(\boldsymbol{b}) = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

We are seeking a sample estimator $\boldsymbol{b}$ of the true population coefficient $\boldsymbol{\beta}$. Thus, if we are approaching the population, $\boldsymbol{b}$ should approach β:

$$\lim_{n\to\infty}\mathbb{P}(|\boldsymbol{b}-\beta|\geq\delta)=0 \qquad \text{for every } \delta>0.$$

Equivalently:

$$\lim_{n\to\infty}\mathbb{V}(\boldsymbol{b})=\sigma^2\lim_{n\to\infty}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}=0.$$

This property is called consistency. We can formally show this:

$$\begin{aligned}
b &= \beta+\left(X'X\right)^{-1}X'\varepsilon,\\
&= \beta+\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}.
\end{aligned}$$

Taking the probability limit on both sides gives us

$$\begin{aligned}\text{plim}(b) &= \beta + \text{plim}\left(\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right),\\ &= \beta + \text{plim}\left(\left(\frac{X'X}{n}\right)^{-1}\right)\text{plim}\left(\frac{X'\varepsilon}{n}\right).\end{aligned}$$

First, we know that

$$\text{plim}\left(\left(\frac{X'X}{n}\right)^{-1}\right) = \Omega,$$

which is a positive definite matrix. Next,

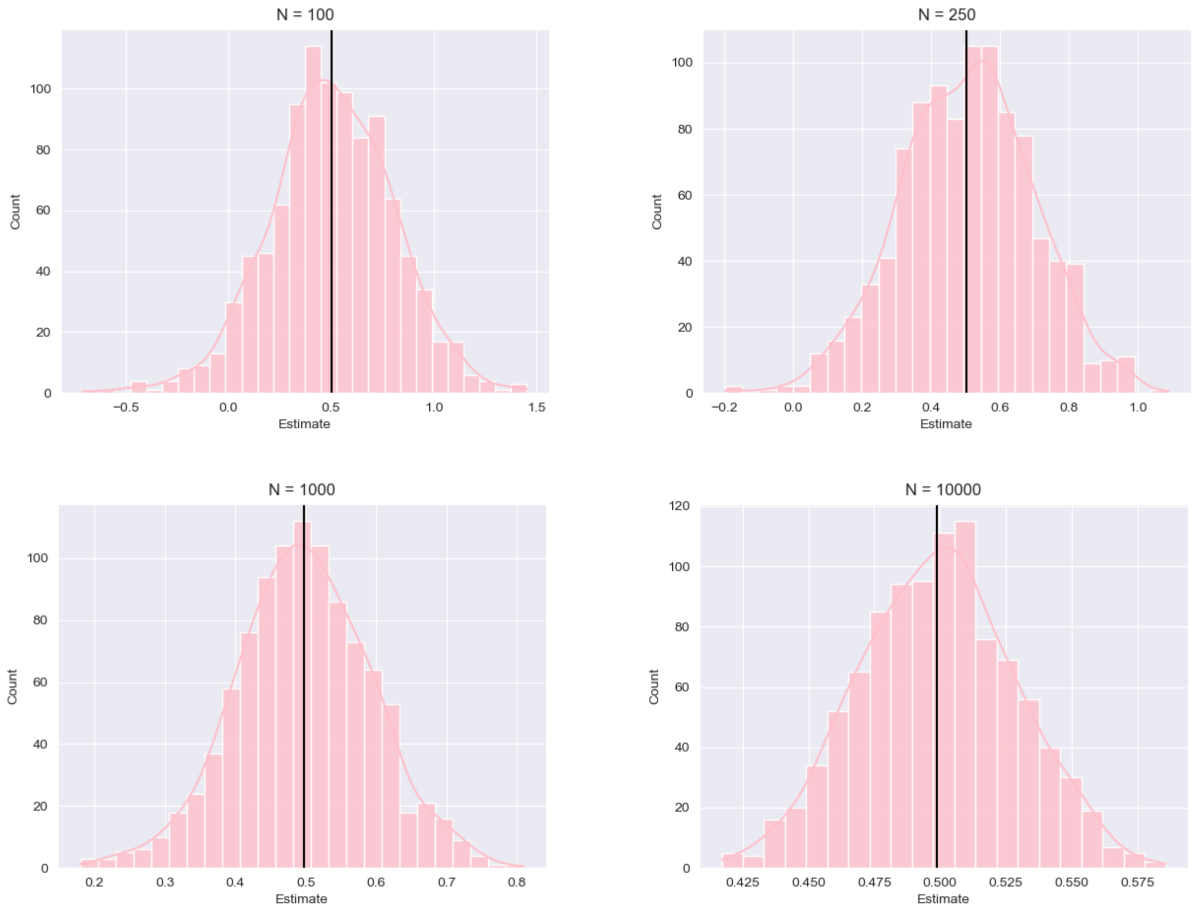
$$\text{plim}\left(\frac{X'\varepsilon}{n}\right) = 0,$$

which follows from the fact that $\mathbb{E}[\varepsilon|X] = 0$ which implies $\mathbb{E}[\varepsilon x] = 0$. This is because $\frac{X'\varepsilon}{n} = n^{-1}\sum_{i=1}^{n} x_i\varepsilon_i$. This leads us to the desired result:

$$\begin{aligned}\text{plim}(b) &= \beta + \text{plim}\left(\left(\frac{X'X}{n}\right)^{-1}\right)\text{plim}\left(\frac{X'\varepsilon}{n}\right),\\ &= \beta + \Omega \times 0 = \beta.\end{aligned}$$

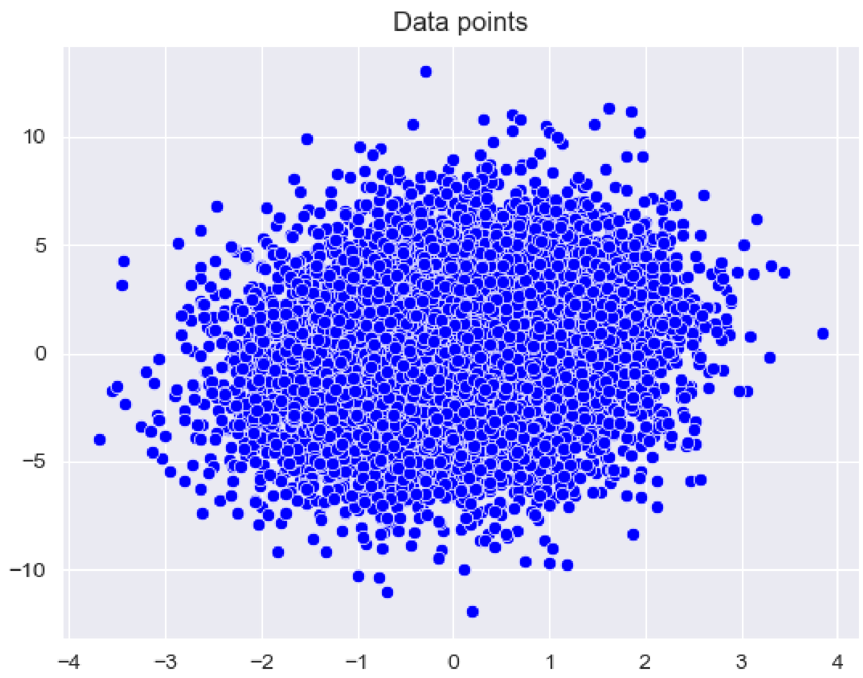
$\Rightarrow$ ***b*** is consistent!

3.7 Estimator uncertainty



The plots above might appear to be very similar at first sight - they are all sort of symmetric and are centered around the same mean value (0.5). However, if we take a look at the spread in each plot, we see that the upper left plot (N=100) includes a range of values between -0.7 and 1.5, but the lower right plot has is spread over over the range 0.425-0.575. These are all estimates from a simulated series of data. The data can be seen in figure 18.

Figure 2: Plot of the data

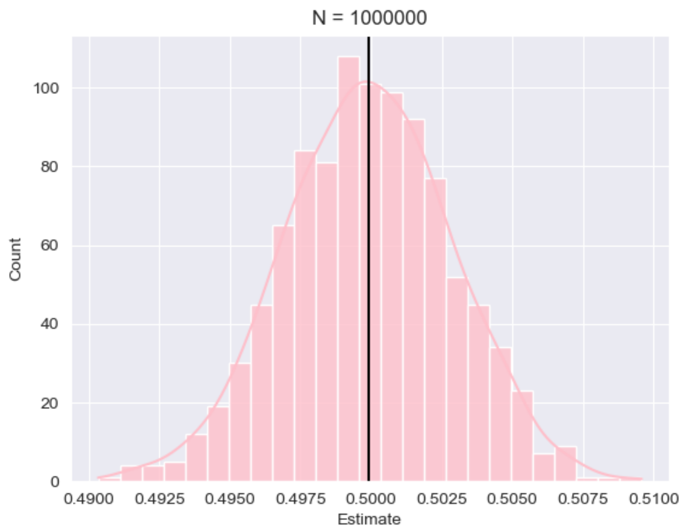


In this example, N refers to the sample length. As N increase we see that also the spread decrease. Put differently, as N increase, the uncertainty about the estimate we have computed decrease. The data generating process in this case was

$$y_i = 0.2 + 0.5 \times x_i + 3 \times u_i,$$

where $x_i \sim \mathcal{N}(0,1), u_i \sim \mathcal{N}(0,1)$. So therefore, the estimates seems to be unbiased and consistent since (i) the estimates are centered around the true value (0.5) (ii) the variance of the estimate appears to go towards zero as the sample size increase. This begs the question: how can we evaluate the precision of our estimate?

Figure 3: Estimates when the sample length is reeaallly long.




```
import numpy             as np
import pandas            as pd
import seaborn           as sns
import matplotlib.pyplot as plt
import statsmodels.api   as sm

N1 = 100
N2 = 250
N3 = 1000
N4 = 10000
N5 = 1000000

N = N4

alpha = 0.2
beta  = 0.5

epsilon = np.random.randn(N)

x = np.random.randn(N)

y = alpha + x*beta + epsilon

X = sm.add_constant(x)

# OLS
XX = np.matmul(X.T,X)

beta_hat = np.matmul(np.linalg.inv(XX),np.matmul(X.T,y))
var_beta = np.var(epsilon)*np.linalg.inv(XX)

# Sampling properties of beta_hat
b = [0]*1000
```

```
a = [0]*1000

for ii in range(1000):
    epsilon = np.random.randn(N)

    x = np.random.randn(N)

    y = alpha + x*beta + epsilon*3


    X = sm.add_constant(x)

    # OLS
    XX = np.matmul(X.T,X)

    a[ii] = np.matmul(np.linalg.inv(XX),np.matmul(X.T,y))[0]
    b[ii] = np.matmul(np.linalg.inv(XX),np.matmul(X.T,y))[1]
    ab = a+b

t1           = ['a']*1000
t2           = ['b']*1000
t            = t1+t2
data         = pd.DataFrame({'Estimate':ab})
data['type'] = t

# Plot histogram
plt.figure()
ax=sns.histplot(data[data.type=='b'].Estimate,kde=True,color='pink',
alpha=0.8)
ax.set(title   = 'N = '+str(N))
plt.axvline(data[data.type=='b'].Estimate.mean(),  color='black')
ax.get_figure().savefig('plot'+str(N) + '.png')


# Scatter plot of data points
plt.figure()
ax=sns.scatterplot(x     = x,
                   y     = y,
                   color = "blue",
                  )
ax.set(title   = "Data points")
ax.get_figure().savefig('plot_data_'+str(N) + '.png')
```

3.8 Hypothesis testing - The test of significance approach

When we have an estimate, $\boldsymbol{b}$, when can easily compute the standard errors of this estimate, SE($\boldsymbol{b}$):

$$SE(\boldsymbol{b}) = \sqrt{\mathbb{V}(\boldsymbol{b})} = \sqrt{\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}}.$$

The diagonal elements of this matrix corresponds to each estimate. Element (1,1) is the standard error of the intercept, element (2,2) is the standard error of the first slope estimate, etc. When doing a test of significance, we usually form some type of hypothesis. A typical hypothesis formulation could be that, under the null hypothesis, the true population parameter is zero. An alternative hypothesis could be that the true population parameter is different from zero:

$$\begin{array}{lll} H_0 & \beta = 0 & \text{(Null hypothesis)} \\ H_1 & \beta \neq 0 & \text{(Alternative hypothesis)} \end{array}$$

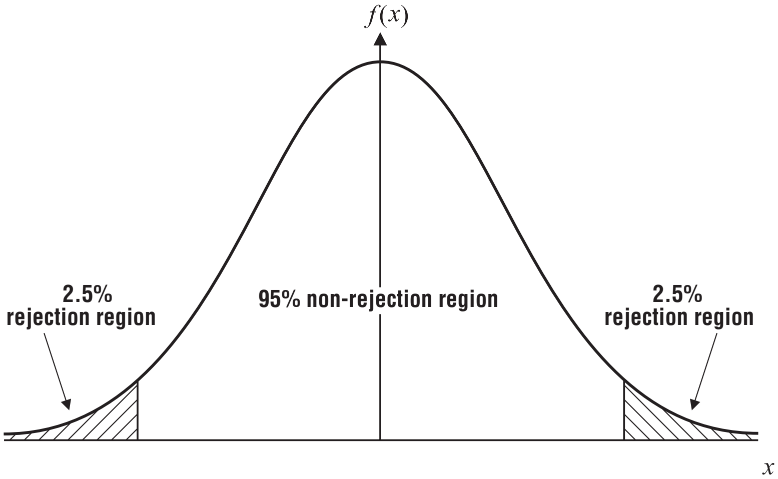
Then, we need to provide the distribution of the test statistic under the null. It is important to keep in mind that the test statistic is indeed a random variable, so we need it to have a well defined distribution so that we can define the rejection region. In our case, our test statistic will follow a t-ditribution with (n-k) degrees for freedom. The test statistic can be computed the following way:

$$\text{test statistic} = \frac{\hat{\beta} - \beta^*}{\text{SE}(\hat{\beta})} \sim t_{\alpha,n-k},$$

where β^* is the parameter value under the null.

We need to define the significance level, α, of our test. What is a "correct" significance level to use? The common value is 5%, however it is debatable if that is always appropriate. For a two-sided test, the rejection regions can be illustrated in the following figure 17.

Figure 4: The rejection regions for a two-sided test with a significance level of 5%.



Look at how the test statistic is computed, it is a ratio of the distance between the null and the computed value divided by the standard error. Since we know the distribution of the test statistic, we can make inferences about, given our null hypothesis, how "likely" it is to observe the estimate that we have computed. If the probability is less that the significance level we have chosen, we reject our null hypothesis. Be careful: this does not mean that we "accept" our alternative hypothesis. Similarly, if we fail to reject the null it means just that - we never accept the null hypothesis. Sometimes it is useful to think about the assumptions that we make for the validity of the test (e.g. Gauss-Markov Assumption + normality of errors).

4 Constrained Optimization

In the previous section we were focusing on optimizing functions over an unconstrained domain. However, usually we have constraints. In economics, an obvious constraint is a budget constraint; the agent might only have ten dollars to spend on either coconuts or bananas which has prices \$1 and \$1.50 respectively. Now the problem becomes much more interesting since the questions we can answer by imposing these restrictions becomes much closer to the ones we are trying to solve in real-life. Another (and perhaps more relevant) examples for people in quantitative finance is that we sometimes want to find optimal portfolios. A common restriction is that all the portfolio weights must sum to one. Perhaps we do not want to allow shorting? Well, then we add another constraint where all weights must be weakly positive.

4.1 The Lagrange Multiplier Method

Assume that we have a consumer that chooses how much of her available income m she will spend on a good x which costs p, and how much she will spend on other goods y. Her budget constraint is $xp + y = m$. Her preference for consuming x and y is given by her utility function $u(x, y)$. Mathematically we can now represent her problem as

$$\max u(x, y) \quad \text{subject to} \quad px + y = m.$$

This is a **constrained maximization problem**. In general, we write

$$\max(\min) \;\; f(x, y) \;\; \text{subject to} \;\; g(x, y) = c.$$

Definition 4.1: Lagrange Multiplier Method

To find the only possible solutions of the problem

$$\text{maximize (minimize)} \quad f(x, y) \;\; \text{subject to} \quad g(x, y) = c$$

proceed as follows:

1. Write down the Lagrangian

   $$\mathscr{L}(x, y) = f(x, y) - \lambda(g(x, y) - c)$$

   where is a constant.

2. Differentiate $\mathscr{L}$ w.r.t. x and y and equate the partial derivatives to 0.

3. The two equations in (2.), together with the constraint, yield the following three equations:

   $$\begin{aligned} \mathscr{L}_1'(x, y) &= f_1'(x, y) - \lambda g_1'(x, y) = 0 \\ \mathscr{L}_2'(x, y) &= f_2'(x, y) - \lambda g_2'(x, y) = 0 \\ g(x, y) &= c. \end{aligned}$$

4. Solve these three equations simultaneously for the three unknowns x, y, and . These triplets (x, y, λ) are the solution candidates, at least one of which solves the problem (if it has a solution).

Example: A consumer has the utlity function $U(x,y) = xy$ and faces the budget constraint $2x+y = 100$. Find the only solution candidate to the consumer demand problem

$$\max xy \text{ subject to } 2x + y = 100.$$

The Lagrangian becomes

$$\mathscr{L}(x,y) = xy - \lambda(2x + y - 100).$$

FOC:

$$\begin{aligned} \mathscr{L}_1'(x,y) &= y - 2\lambda = 0, \\ \mathscr{L}_2'(x,y) &= x - \lambda = 0, \\ 2x + y &= 100. \end{aligned}$$

From the first two solutions we have that $y = 2\lambda$ and $x = \lambda$. So $y = 2x$. Inserting this into the constraint yields $2x + 2x = 100$. So $x = 25$ and $y = 50$, implying that $\lambda = x = 25$.

4.2 Intepreting the Lagrange Multiplier

Consider again the problem

$$\text{maximize (minimize)} \quad f(x,y) \quad \text{subject to} \quad g(x,y) = c.$$

Suppose x^* and y^* are the values of x and y that solve this problem. In general, x^* and x^* depend on c. We assume that $x^* = x^*(c)$ and $y^* = y^*(c)$ are differentiable functions of c. The associated value of $f(x,y)$ is then also a function of c, with

$$f^*(c) = f(x^*(c), y^*(c)).$$

Here $f^*(c)$ is called the (optimal) value function for the problem. The associated value of the Lagrange multiplier also depends on c, in general. Provided that a certain regularity conditions are satisfied, we have the result that

$$\frac{df^*(c)}{dc} = \lambda(c).$$

We therefore have the following result:

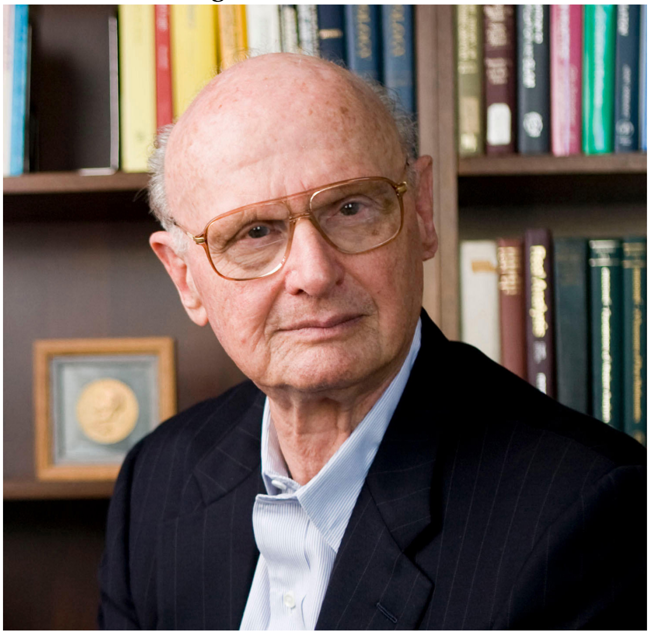
Definition 4.2: Lagrange Multiplier

The Lagrange multiplier $\lambda = \lambda(c)$ is the rate at which the optimal value of the objective function changes with respect to changes in the contraint constant c. In particular, if dc is a small change in c, then

$$f^*(c + dc) - f^*(c) \approx \lambda(c)dc.$$

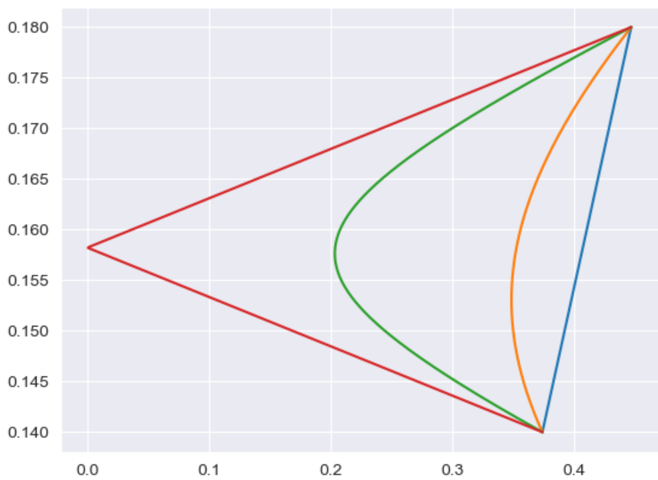
4.3 Application: Portfolio Theory

Figure 5: God himself



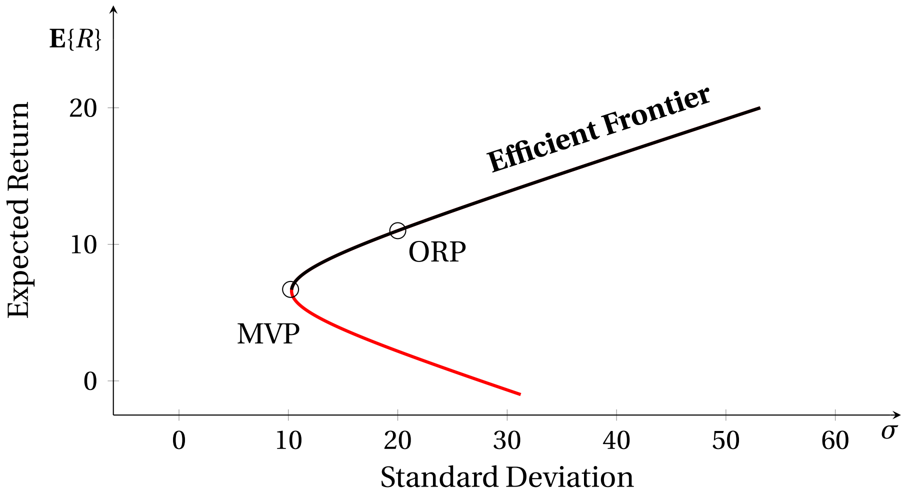
When we are investing we are faced with the problem of deciding how to allocate our limited wealth into assets. In his seminal 1952 paper, Nobel Laureate Harry Markowitz introduced modern portfolio theory. He argued that we should not pay much attention to each assets return-risk characteristics (expected return and standard deviation), but rather how inclusion of assets affects the overall riskiness and return of the portfolio.

Figure 6: The effect of different combination of two assets (correlations: blue =1, orange = 0.5, green = -0.5, and red = -1)



```
w_1    = np.linspace(0,1,500)
w_2    = 1-w_1
var_1  = 0.2
var_2  = 0.14
r_1    = .18
r_2    = .14
rhos   = [1,0.5,-0.5,-1]
def var(weight1,weight2,var1,var2,corr):
    std1 = var1**(.5)
    std2 = var2**(.5)
    return(weight1**2*var1 + weight2**2*var2 + \
                    2*corr*std1*std2*weight2*weight1)
for rho in rhos:
    p_ret = [0]*len(w_1)
    p_var = [0]*len(w_1)
    for ind in range(len(w_1)):
        p_ret[ind] = w_1[ind]*r_1+w_2[ind]*r_2
        p_var[ind] = var(w_1[ind],w_2[ind],var_1,var_2,rho)**.5
    plt.plot(p_var,p_ret, '-')
```

First we will tackle the problem of finding the of the optimal portfolios. We will be interested in two portfolios: (i) The minimum variance portfolio (ii) the optimal risky portfolio. Graphically:



The question is then, how can we find these two portfolios?

4.4 Minimum Variance Portfolio

The minimum variance portfolio is the portfolio that minimize the portfolio with the least variance among all feasible portfolios. Mathematically, we are searching for the weights, ω_i, that minimize the variance of the portfolio p. Since we are constructing a portfolio, we want the weights of this portfolio to sum to one. This gives rise to the following constrained optimization problem:

$$\underset{\omega_i}{\arg\min}\ \sigma_p^2 \quad \text{subject to} \sum \omega_i = 1.$$

Again, remember how we defined portfolio variance:

$$\sigma^2 = \boldsymbol{\omega}'\Sigma\boldsymbol{\omega},$$

where $\boldsymbol{\omega}$ is a vector of weights and Σ is a covariance matrix. We see that this is a simple optimization problem with a quadratic form:

$$\underset{\omega}{\arg\min}\ \ \boldsymbol{\omega}'\Sigma\boldsymbol{\omega} \quad \text{subject to } \boldsymbol{\omega}'\mathbf{1} = 1.$$

Example: Consider the case with two assets, 1 and 2. We want to find the weights, ω_1 and ω_2, that minimizes the portfolio variance. The problem becomes:

$$\underset{\omega_1,\omega_2}{\arg\min}\ \ \omega_1^2 Var(s_1) + \omega_2^2 Var(s_2) + 2\omega_1\omega_2 cov(s_1, s_2) \quad \text{subject to } \omega_1 + \omega_2 = 1.$$

From the constraint we get that $\omega_2 = 1 - \omega_1$. We rewrite the problem as

$$\underset{\omega_1}{\arg\min}\ \ \omega_1^2 Var(s_1) + (1-\omega_1)^2 Var(s_2) + 2\omega_1(1-\omega_1)cov(s_1, s_2).$$

FOC:

$$2\omega_1 Var(s_1) + 2Var(s_2) - 2\omega_1 Var(s_2) + 2cov(s_1, s_2) - 4\omega_1 cov(s_1, s_2) = 0.$$

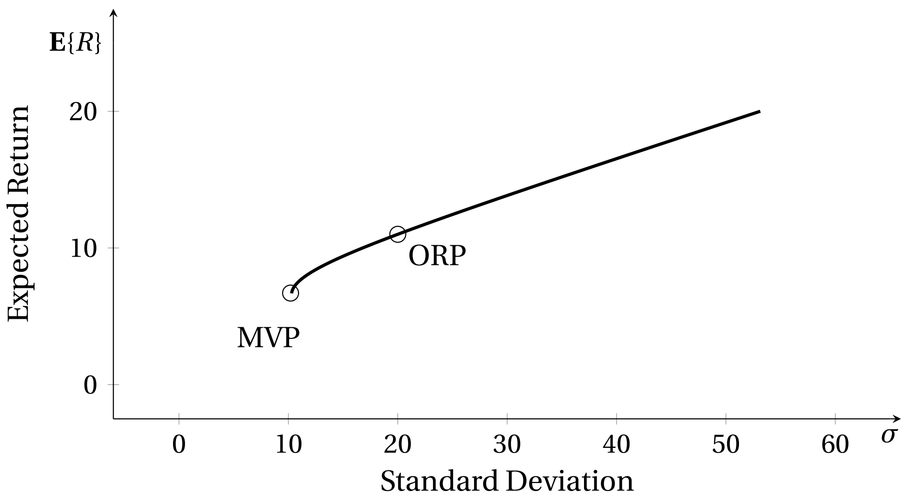
Rearranging yields

$$\omega_1^* = \frac{Var(s_2) - cov(s_1, s_2)}{Var(s_1) + Var(s_2) - 2cov(s_1, s_2)}.$$

and consequently, $\omega_2^* = 1 - \omega_1^*$. How do we know that this actually solves the minimization problem? Since this is a quadratic function, it's sufficient to check the definiteness of the covariance matrix, Σ. Since any covariance matrix is positive semi-definite, we know that the function if convex and that any optimum is a global minimum.

4.5 Optimal Risky Portfolio

We are now interested in the Optimal Risky Portfolio, which is defined as the portfolio that maximizes the expected risk-return tradeoff.



If we look at the efficient frontier, we see that it is constructed by the portfolios that, for a given exptected return, minimize the risk of the portfolio. Clearly, the minimum variance portfolio will be one of these portfolios. However, among these portfolios there exists a portfolio that gives us a better deal in terms of return per unit of risk than any of the other portfolios. We call this the Optimal Risky Portfolio.

For any portfolio we can define the Sharpe Ratio, named after Nobel Laureate William F. Sharpe as the ratio of the excess return of the portfolio divided by its standard deviation:

$$S_p = \frac{\mathbb{E}[r_p] - r^f}{\sigma_p}.$$

The Optimal Risky Portfolio is then the portfolio that maximize this ratio. The problem of finding this portfolio becomes

$$\underset{\omega}{\arg\max}\ S_p \quad \text{subject to} \quad \boldsymbol{\omega}'\mathbf{1} = 1,$$

or

$$\underset{\omega}{\arg\max}\ \omega'\boldsymbol{R}\left(\sqrt{\omega'\Sigma\omega}\right)^{-1} \quad \text{subject to} \qquad \boldsymbol{\omega}'\mathbf{1} = 1,$$

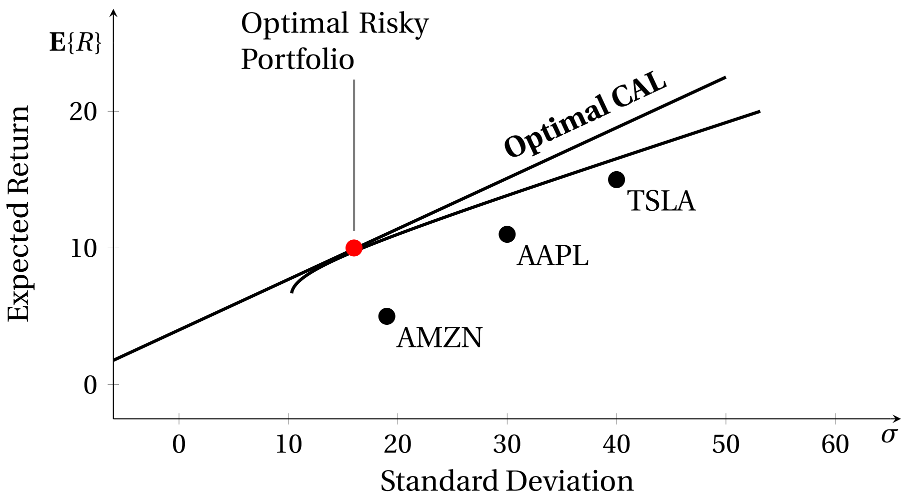
where $\boldsymbol{R}$ is a vector of excess returns for each asset.

Example: For the 2-asset problem, the optimal weights can be computed as

$$\begin{aligned}
\omega_1^* =& \frac{\left(\mathbb{E}[s_1] - r^f\right) Var(s_2) - \left(\mathbb{E}[s_2] - r^f\right) cov(s_1, s_2)}{\left(\mathbb{E}[s_1] - r^f\right) Var(s_2) + \left(\mathbb{E}[s_2] - r^f\right) Var(s_1) - \left(\mathbb{E}[s_1] - r^f + \mathbb{E}[s_2] - r^f\right) cov(s_1, s_2)},\\
\omega_2^* =& 1 - \omega_1^*.
\end{aligned}$$

4.6 Asset allocation with risky assets and T-bill

Now that we have established the optimal portfolio, we can talk about optimal capital allocation. In mean-variance analysis, every investor actually agrees on what this portfolio is. Why? Because the mean-variance investor only cares about the first two moments, the expected return and the variance. If we are all in the same investment universe, then mechanically we should end up with the same optimal portfolio. If we introduce a risk-free asset (usually a T-bill), we can construct a line that is tangent to the efficient frontier.

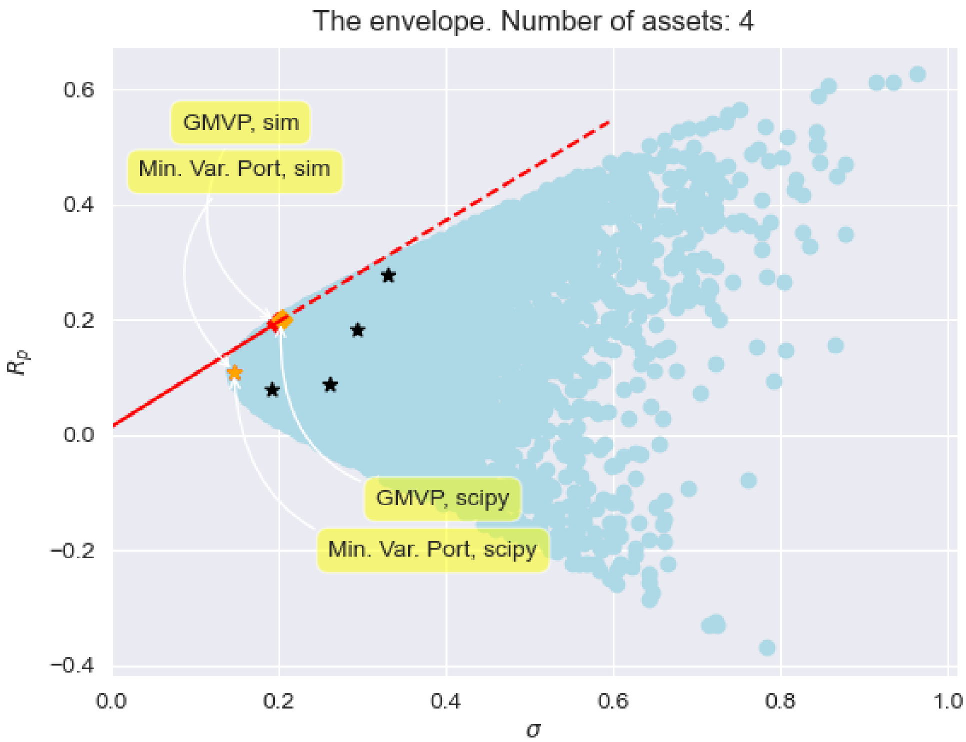


In fact, the exercise of finding the optimal risky portfolio can be viewed as the problem of finding the capital allocation line with the steepest gradient. Moreover, now there exists only one efficient risky portfolio. As an example, it does not any longer make sense to say that the minimum variance portfolio is efficient. Why? Because we can do better. By investing in T-bills and the optimal portfolio, we can generate a return that is higher than the minimum variance portfolio, but with the same amount of risk. Therefore, the problem of an investor can be decomposed into two steps: (i) finding the optimal risky portfolio (ii) decide on the allocation between the risk-free asset and this optimal portfolio. This is called the two-fund separation theorem. Notice that the capital allocation line extends further than the optimal portfolio. This is because can **lend** at the risk-free rate (**short T-bills**) and invest the proceeds in the optimal portfolio. The weight in the risk-free asset will then become negative and the weight in the optimal portfolio will be greater than one, but the sum of the two will still be one.

4.7 Numerical approach to find the portfolios

How can we solve the portfolio construction problem in Python? Two ways are by simulation and by a numerical optimizer.

Figure 7: Using numerical approaches to find the portfolios



We see from the figure above that the two approaches yields approximately the same solution. The simulation approach works by drawing random weights and then constructing portfolios based on these. The numerical optimizer takes an objective function and then optimize.

```
import requests
import sys
import pandas              as pd
import pandas_datareader   as web
import numpy               as np
import matplotlib.pyplot   as plt
import datetime            as dt
```

```python
import yfinance           as yf

from bs4              import BeautifulSoup
from scipy.optimize import minimize


### Define functions ###
def objective_gmvp(w,ret,vCov,rf):
    '''
    Objective function for global mean variance portfolio
    Inputs:
     - w      : vector of weights
     - ret    : vector of returns
     - vCov : covariance matrix
     - rf     : risk-free rate

    Output:
         - -1s_p : The Sharpe ratio times negative one.
    '''
    var = np.matmul(np.matmul(w,vCov),np.transpose(w))
    std  = var**0.5
    s_p  = (np.matmul(w,ret)-rf)/std
    return -1*s_p


def objective_var(w,vCov):
    '''
    Objective function for minimum variance portfolio.
    Inputs:
     - w    : vector of weights
     - ret : vector of returns

    Output:
    - var_p : The portfolio variance.
    '''
    var_p = np.matmul(np.matmul(w,vCov),np.transpose(w))
    return var_p

def constraint(w):
    '''
    Constraint. We want our weights to sum to one.
    Input:
     - w    : vector of weights

    Output:
     - The sum of the weight vector minus one.
    '''
    return sum(w) - 1


def get_ticks():
    # Get all SP500 tickers and industries
    URL = "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies"

    res          = requests.get(URL).text
    soup         = BeautifulSoup(res,'lxml')
    comp_list = []
    ind_list   = []
    for items in soup.find('table', class_='wikitable').find_all('tr')[1::1]:
        row = items.find_all(['th','td'])
        try:
            comp_list.append(row[0].a.text)
            ind_list.append(row[3].text)
        except: continue
    ind_list = [ind.split('\n')[0] for ind in ind_list ]
    return comp_list, ind_list

################################################################################

tickers        = ['AAPL','MSFT','GOOGL','GLD']
data           = pd.DataFrame()
start          = dt.datetime(2005, 1, 1)
end            = dt.datetime(2018,1,1)
data           = yf.download(tickers, start=start, end=end).Close
data.columns = tickers
ret_daily      = np.log(data / data.shift(1))
ret_mean_an   = ret_daily.mean()*252
vcov           = ret_daily.cov()*252
print('.'*100)
print('\n Summary statistics \n')
print('.'*100)
print(' \n Annualized mean returns: \n')
print(ret_mean_an)
print('\n Covariance matrix of returns: \n')
print(vcov)
print('.'*100)


port_std = []
port_ret = []
simLen    = 10000
print('\nProgress:')
print('.'*100 + '\n')
for ii in range(1,simLen+1):
    weights    = np.random.normal(0,1,len(tickers))
    weights    /= weights.sum()
    if np.any(weights>=3) or np.any(weights<=-3): continue
    port_var   = np.matmul(np.matmul(weights,vcov),np.transpose(weights))
    port_std   = np.append(port_std,port_var**0.5)
    port_ret   = np.append(port_ret,np.matmul(weights,ret_mean_an))
    if (ii % 100==0 and ii != simLen+1):
        b=('Finished with iteration ' + str(ii) + ' of ' + str(len(range(1,simLen+1))))
```

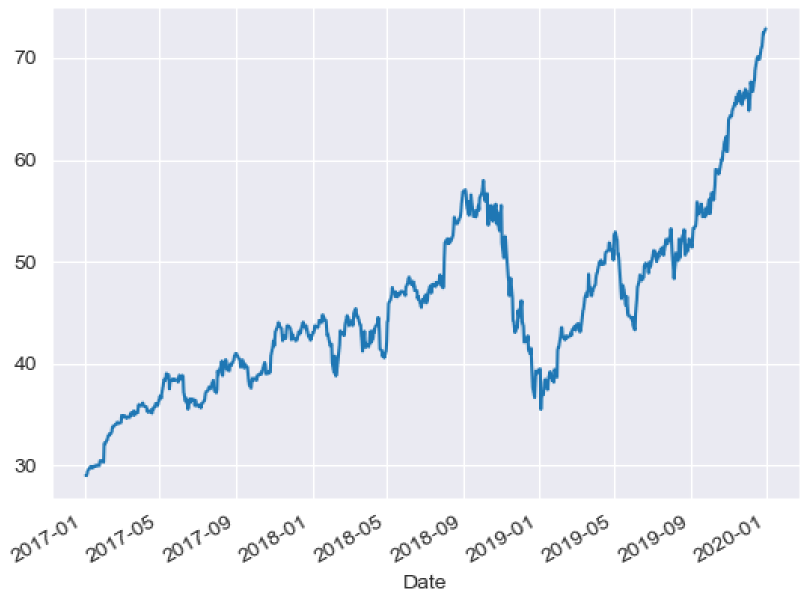
```
        sys.stdout.write('\r'+b)
    if (ii == simLen): sys.stdout.write('\r'+'-'*30+' Done! '+'-'*30+'\n')

## Identify mean-variance portfolio through simulation ##
risk_free    = web.get_data_fred('TB3MS',start=end,end=end)/100
s_p          = (port_ret-risk_free['TB3MS'][0])/port_std
s_p_m_ind    = np.argmax(s_p)
min_var_ind  = np.argmin(port_std)

### Identify GMVP by constrained optimization ###
rf = risk_free.values[0].tolist()
rf = rf[0]

# Initial guesses
x0   = np.ones(len(weights))
x0  /= x0.sum()

# Run optimization routine, find GMVP
b     = (-3,3)
bnds  = (b,)*len(weights)
con   = {'type': 'eq', 'fun': constraint}
gmvp = minimize(objective_gmvp,x0,args=(ret_mean_an,vcov,rf),constraints=con,
                bounds=bnds,options={'disp': True})
var_gmvp = np.matmul(np.matmul(gmvp.x,vcov),np.transpose(gmvp.x))
ret_gmvp = np.matmul(gmvp.x,ret_mean_an)
std_gmvp = var_gmvp**0.5
s_p_gmvp = (ret_gmvp-rf)/std_gmvp
print('.'*100)
print('\nGlobal Mean Variance portfolio:\n')
print('Return = ' + str(ret_gmvp) + '\n')
print('Standard deviation = ' + str(std_gmvp) + '\n')
print('Sharpe ratio = ' + str(s_p_gmvp) + '\n')
print('.'*100)


# Run optimization routine, find MVP
minvar = minimize(objective_var,x0,args=(vcov),constraints=con,
                  options={'disp': True})
ret_minvar = np.matmul(minvar.x,ret_mean_an)
std_minvar = minvar.fun**0.5
print('.'*100)
print('\nMinimum variance portfolio:\n')
print('Return = ' + str(ret_minvar) + '\n')
print('Standard deviation = ' + str(std_minvar) + '\n')
print('.'*100)

fig, ax = plt.subplots()
ax.scatter(port_std, port_ret, c='lightblue')
ax.scatter(port_std[s_p_m_ind], port_ret[s_p_m_ind], c='red',marker='D')
ax.scatter(port_std[min_var_ind], port_ret[min_var_ind], c='red',marker='*')
ax.scatter(std_gmvp, ret_gmvp, c='orange',marker='D')
ax.scatter(std_minvar, ret_minvar, c='orange',marker='*')
ax.scatter(np.diag(vcov**0.5), ret_mean_an, c='black',marker='*')
ax.plot([0,port_std[s_p_m_ind]], [risk_free['TB3MS'][0],port_ret[s_p_m_ind]],c='r')
ax.plot([port_std[s_p_m_ind],port_std[s_p_m_ind]*3], [port_ret[s_p_m_ind],\
        (port_ret[s_p_m_ind]-risk_free['TB3MS'][0])*3],c='r',linestyle='--')
plt.xlim(left=0)
ii=1
for port, x, y in zip(['GMVP, scipy','GMVP, sim','Min. Var. Port, scipy','Min. Var. Port, sim'],
                      [std_gmvp,port_std[s_p_m_ind],std_minvar,port_std[min_var_ind]],
                      [ret_gmvp,port_ret[s_p_m_ind],ret_minvar, port_ret[min_var_ind]]):
    plt.annotate(
        port, xy=(x, y), xytext=(-40*((-1)**(ii)), 80*((-1)**(ii))),
        textcoords='offset points', ha='left', va='bottom',
        bbox=dict(boxstyle='round,pad=0.5', fc='yellow', alpha=0.5),
        arrowprops=dict(arrowstyle = '->', connectionstyle='arc3,rad=' + str((-1)**(ii)*0.5)))
    ii = ii+1
plt.title("The envelope. Number of assets: " + str(len(tickers)))
plt.xlabel(r'$\sigma$')
plt.ylabel(r'$R_p$')
plt.show()
#plt.savefig('plot1.pdf')


### What if we add more assets? ###

sp_tickers,_ = get_ticks()

## Fetch 20 random tickers ##
magic_indicies = np.random.randint(0,500,20)
ext_tickers    = [sp_tickers[ind] for ind in magic_indicies]
data_new       = yf.download(ext_tickers, start=start, end=end).Close

ret_daily_new     = np.log(data_new / data_new.shift(1))
ret_mean_an_new   = ret_daily_new.mean()*252
vcov_new          = ret_daily_new.cov()*252
print('.'*100)
print('\n Summary statistics \n')
print('.'*100)
print(' \n Annualized mean returns: \n')
print(ret_mean_an_new)
print('\n Covariance matrix of returns: \n')
print(vcov_new)
print('.'*100)



port_std_new = []
port_ret_new = []
simLen   = 50000
print('\nProgress:')
print('.'*100 + '\n')
```

```
for ii in range(1,simLen+1):
    weights    = np.random.normal(0,1,len(data_new.columns))
    weights   /= weights.sum()
    if np.any(weights>=3) or np.any(weights<=-3): continue
    port_var_new = np.matmul(np.matmul(weights,vcov_new),np.transpose(weights))
    port_std_new = np.append(port_std_new,port_var_new**0.5)
    port_ret_new = np.append(port_ret_new,np.matmul(weights,ret_mean_an_new))
    if (ii % 100==0 and ii != simLen):
        b=('Finished with iteration ' + str(ii) + ' of ' + str(len(range(1,simLen+1))))
        sys.stdout.write('\r'+b)
    if (ii == simLen): sys.stdout.write('\r'+'-'*30+' Done! '+'-'*30+'\n')

## Identify mean-variance portfolio through simulation ##
risk_free        = web.get_data_fred('TB3MS',start=end,end=end)/100
s_p_new          = (port_ret_new-risk_free['TB3MS'][0])/port_std_new
s_p_m_ind_new    = np.argmax(s_p_new)
min_var_ind_new  = np.argmin(port_std_new)

# Initial guesses
x0   = np.ones(len(data_new.columns))
x0  /= x0.sum()

# Run optimization routine, find GMVP
b        = (-3,3)
bnds     = (b,)*len(data_new.columns)
con      = {'type': 'eq', 'fun': constraint}
gmvp_new = minimize(objective_gmvp,x0,args=(ret_mean_an_new,vcov_new,rf),constraints=con,
                    bounds=bnds,options={'disp': True})
ret_gmvp_new = np.matmul(gmvp_new.x,ret_mean_an_new)
std_gmvp_new = objective_var(gmvp_new.x,vcov_new)**0.5

# Run optimization routine, find MVP
minvar_new = minimize(objective_var,x0,args=(vcov_new),constraints=con,
                      options={'disp': True})
ret_minvar_new = np.matmul(minvar_new.x,ret_mean_an_new)
std_minvar_new = minvar_new.fun**0.5


fig, ax = plt.subplots()
ax.scatter(port_std_new, port_ret_new, c='lightblue')
ax.scatter(port_std_new[s_p_m_ind_new], port_ret_new[s_p_m_ind_new], c='red',marker='D')
ax.scatter(port_std_new[min_var_ind_new], port_ret_new[min_var_ind_new], c='red',marker='*')
ax.scatter(std_gmvp_new, ret_gmvp_new, c='orange',marker='D')
ax.scatter(std_minvar_new, ret_minvar_new, c='orange',marker='*')
ax.plot([0,port_std_new[s_p_m_ind_new]], [risk_free['TB3MS'][0],port_ret_new[s_p_m_ind_new]],c='r')
ax.scatter(np.diag(vcov_new**0.5), ret_mean_an_new, c='black',marker='*')
ii=1
for port, x, y in zip(['GMVP, scipy','GMVP, sim','Min. Var. Port, scipy','Min. Var. Port, sim'],
                      [std_gmvp_new,port_std_new[s_p_m_ind_new],std_minvar_new,port_std_new[min_var_ind_new]],
                      [ret_gmvp_new,port_ret_new[s_p_m_ind_new],ret_minvar_new, port_ret_new[min_var_ind_new]]):
    plt.annotate(
        port, xy=(x, y), xytext=(-40*((-1)**(ii)), 80*((-1)**(ii))),
        textcoords='offset points', ha='left', va='bottom',
        bbox=dict(boxstyle='round,pad=0.5', fc='yellow', alpha=0.5),
        arrowprops=dict(arrowstyle = '->', connectionstyle='arc3,rad=' + str((-1)**(ii)*0.5)))
    ii = ii+1
ax.plot([port_std_new[s_p_m_ind_new],port_std_new[s_p_m_ind_new]*3],\
        [port_ret_new[s_p_m_ind_new],(port_ret_new[s_p_m_ind_new]-risk_free['TB3MS'][0])*3],c='r',linestyle='--')
ax.plot([port_std[s_p_m_ind],port_std[s_p_m_ind]*3], [port_ret[s_p_m_ind],\
         (port_ret[s_p_m_ind]-risk_free['TB3MS'][0])*3],c='g',linestyle='--')
plt.xlim(left=0,right=2)
plt.title("The envelope. Number of assets: " + str(len(data_new.columns)))
plt.xlabel(r'$\sigma$')
plt.ylabel(r'$R_p$')
plt.show()
#plt.savefig('plot2.pdf')


### Figure with both envelopes, for comparison
fig, ax = plt.subplots()
ax.scatter(port_std, port_ret, c='red')
ax.scatter(port_std_new, port_ret_new, c='lightblue' ,alpha=0.6)
ax.legend(['Number of assets: '+ str(len(tickers)),'Number of assets: '+ str(len(data_new.columns))])
plt.xlim(left=0,right=1)
plt.title('The envelope.')
plt.xlabel(r'$\sigma$')
plt.ylabel(r'$R_p$')
plt.show()
#plt.savefig('comparison.pdf')
```

5 Time Series

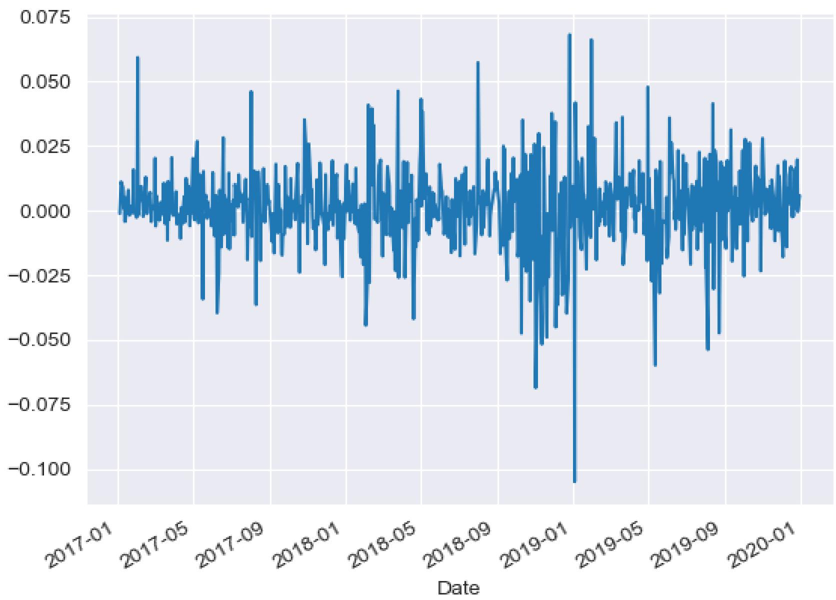
A time series simply a series which realizations occur over time and can therefore be indexed by a time-counter (typically, t). An example is the stock price of Apple:

Figure 8: Stock price evolution of Apple



Another example is the stock price return of Apple:

Figure 9: Stock price returns of Apple



In general, we can write a time series as $\{y_1, y_2, ..., y_T\}$ or $\{y_t\}$ for $t \in [1, T]$.

5.1 Important notation and concepts

5.1.1 Strictly stationary process

A strictly stationary process is one where, for any $t_1, t_2, ..., t_T \in \mathbb{Z}$, any $k \in \mathbb{Z}$ and $T = 1, 2, ...$

$$F_{y_{t1}, y_{t2}, ..., y_{tT}}\left(y_1, ..., y_T\right) = F_{y_{t_1+k}, y_{t_2+k}, ..., y_{t_T+k}}\left(y_1, ..., y_T\right),$$

where F denote the joint distribution function of the set of random variables. We can also say that the probability measure for the sequence $\{y_t\}$ is the same as that for $\{y_{t+k}\}\forall k$.

5.1.2 Weakly stationary process

$$\begin{aligned}
\mathbb{E}[y_t] =& \mu & (3)\\
\mathbb{E}\left(y_t - \mu\right)\left(y_t - \mu\right) =& \sigma^2 < \infty & (4)\\
\mathbb{E}\left(y_{t_1} - \mu\right)\left(y_{t_2} - \mu\right) =& \gamma_{t_2 - t_1} \quad \forall t_1, t_2 & (5)
\end{aligned}$$

If a series satisfies (3)-(5) for $t \in [1, \infty]$, is is said to be weakly or covariance stationary. These equations states that, for a series to be stationary, it should have a constant mean, a constant variance, and a constant autocovariance structure.

The autocovariances captures the dependency that realizations of y has on its previous values. In the case of a stationary process, the autocovariances depends only on the time difference between t_1 and t_2. This means that the covariance between y_t and y_{t-2} is the same as y_{t-5} and y_{t-7}. The moment

$$\mathbb{E}\left(y_t - \mathbb{E}[y_t]\right)\left(y_{t-s} - \mathbb{E}[y_{t-s}]\right) = \gamma_s, \quad s = 0, 1, 2...$$

Note that for $s = 0$ we have the variance of the process. A perhaps more useful measure is the autocorrelation function which is bounded in the interval $[-1, 1]$.

$$\tau_s = \frac{\gamma_s}{\gamma_0}.$$

5.1.3 A white noise process

One of the building blocks for time series modeling is a white noise process, sometimes denotes as ε. The definition of a white noise process is

$$\begin{aligned}
\mathbb{E}[\varepsilon_t] =& \mu & (6)\\
\mathbb{V}[\varepsilon_t] =& \sigma^2 & (7)\\
\gamma_{t-r} =& \begin{cases} \sigma^2 & \text{if } t = r \\ 0 & \text{otherwise.} \end{cases} & (8)
\end{aligned}$$

If $\mu = 0$ and (6)-(7) are satisfied then the process $\{\varepsilon_t\}$ is known as a zero mean white noise. We usually write

$$\varepsilon_t \sim \text{ i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2).$$

5.2 ARMA models

AR(1)	$y_t = \phi y_{t-1} + \varepsilon_t$
MA(1)	$y_t = \theta \varepsilon_{t-1} + \varepsilon_t$
AR(p)	$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$
MA(q)	$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$
ARMA(1,1)	$y_t = \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$
ARMA(p,q)	$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$

5.2.1 Lag operators

In time series, using lag operator notation can be helpful when manipulating ARMA models. The lag operator works why moving the index back one unit:

$$\begin{aligned} Lx_t &= x_{t-1}, \\ L^2 x_t &= x_{t-2}, \quad \text{etc.} \end{aligned}$$

5.2.2 Moving average processes

This is the simplest class of time series models, denoted MA(q) where q refers to the number of lags in the model. Generally, a moving average model of order q (MA(q)) can be written as

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

This is a linear combination of white noise processes. We can also express this using sigma notation:

$$y_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

or by using lag notation:

$$y_t = \theta(L)\varepsilon,$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$.

The properties of the process given above is

1. $\mathbb{E}[y_t] = 0$

2. $\mathbb{V}[y_t] = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2$

3. $\text{Cov}(y_t, y_{t-s}) = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + ... + \theta_q \theta_{t-s})\sigma^2 & \text{for } s = 1, 2, ..., q \\ 0 & \text{for } s > q \end{cases}$

Example: Consider the following moving average process:

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t,$$

where $\varepsilon \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

1. Compute the mean and variance of y_t
2. Derive the autocorrelation function of this process
3. Sketch the autocorrelation function of this process if $\theta_1 = -1/2$ and $\theta_2 = 1/4$.

1.

Expected value:

$$\mathbb{E}[y_t] = \mathbb{E}[\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t] = \theta_1\mathbb{E}[\varepsilon_{t-1}] + \theta_2\mathbb{E}[\varepsilon_{t-2}] + \mathbb{E}[\varepsilon_t] = 0.$$

Variance:

$$\begin{aligned}
\mathbb{V}[y_t] =& \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_t - \mathbb{E}[y_t])] = \mathbb{E}[(y_t)(y_t)] \\
=& \mathbb{E}[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)] \\
=& \mathbb{E}[\theta_1^2\varepsilon_{t-1}^2 + \theta_2^2\varepsilon_{t-2}^2 + \varepsilon_t^2 + \text{ cross-products }] \\
=& \theta_1^2\mathbb{E}[\varepsilon_{t-1}^2] + \theta_2^2\mathbb{E}[\varepsilon_{t-2}^2] + \mathbb{E}[\varepsilon_t^2] + 0 \\
=& \sigma^2(1 + \theta_1^2 + \theta_2^2).
\end{aligned}$$

Here we used two properties: (i) $\mathbb{E}[\varepsilon_t^2] = \sigma^2$ (ii) $\text{Cov}(\varepsilon_t, \varepsilon_{t-s}) = 0$ for all $t \neq s$.

2.

$$\begin{aligned}
\gamma_1 =& \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-1} - \mathbb{E}[y_{t-1}])] \\
=& \mathbb{E}[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3} + \varepsilon_{t-1})] \\
=& \mathbb{E}[(\theta_1\varepsilon_{t-1}^2 + \theta_1\theta_2\varepsilon_{t-2}^2 + \text{cross-products})] \\
=& \sigma^2(\theta_1 + \theta_1\theta_2).
\end{aligned}$$

$$\begin{aligned}
\gamma_2 =& \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-2} - \mathbb{E}[y_{t-2}])] \\
=& \mathbb{E}[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4} + \varepsilon_{t-2})] \\
=& \mathbb{E}[\theta_2\varepsilon_{t-2}^2 + \text{cross-products}] \\
=& \theta_2\sigma^2. \\
\gamma_3 =& \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-3} - \mathbb{E}[y_{t-3}])] \\
=& \mathbb{E}[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-4} + \theta_2\varepsilon_{t-5} + \varepsilon_{t-3})] \\
=& \mathbb{E}[\text{cross-products}] \\
=& 0.
\end{aligned}$$

$$\gamma_s = 0 \text{ for } s \geq 3.$$

The autocorrelation functions are given by

$$\tau_s = \frac{\gamma_s}{\gamma_0}.$$

So,

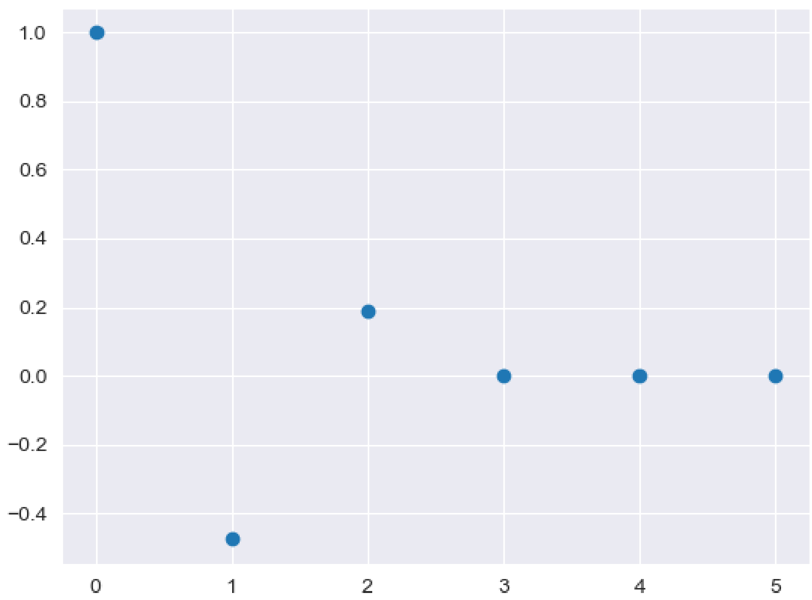
$$\begin{aligned}
\tau_0 =& \frac{\gamma_0}{\gamma_0} = 1, \\
\tau_1 =& \frac{\gamma_1}{\gamma_0} = \frac{\sigma^2(\theta_1 + \theta_1\theta_2)}{\sigma^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\
\tau_2 =& \frac{\gamma_2}{\gamma_0} = \frac{\theta_2\sigma^2}{\sigma^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\
\tau_s =& 0 \quad \forall s \geq 3.
\end{aligned}$$

3.

Using $\theta_1 = -1/2$ and $\theta_2 = 1/4$, we get:

$$\tau_0 = 1,$$
$$\tau_1 = -0.476,$$
$$\tau_2 = 0.190.$$

Figure 10: Autocorrelation function of the MA(2) process with $\theta_1 = -1/2$ and $\theta_2 = 1/4$



5.2.3 Autoregressive processes

An autoregressive process depends on past realizations of itself plus an error term. An autorgressive model of order p can be expressed as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t,$$

or by using the lag operator:

$$\phi(L) y_t = \varepsilon_t \quad \text{where} \quad \phi(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p).$$

5.2.4 Stationarity

Consider the following simple model of log stock returns:

$$r_t = \mu + \sigma \varepsilon_t, \quad \text{for all t}$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$ (We can relax the normality assumption). Assume that all wealth is reinvested. The value of a portfolio, p_t will be:

$$\begin{aligned} p_t &= p_{t-1} + r_t \\ &= p_{t-1} + \mu + \sigma\varepsilon_t. \end{aligned}$$

This is a random walk with drift. A process is called a random walk if it has no discernable structure, it "wanders around" aimlessly, or random. We can write the most simple verison of a random walk as

$$y_t = y_{t-1} + \varepsilon_t.$$

A typical feature of such a process is that the conditional expectation of future realizations of y_t is today's value:

$$\mathbb{E}_t[y_{t+1}] = \mathbb{E}_t[y_{t+1} + \varepsilon_{t+1}] = \mathbb{E}_t[y_t + \varepsilon_t + \varepsilon_{t+1}] = y_t.$$

In other words, the increments are unforecasteable. If we include a drift term to the random walk process we get

$$y_t = \mu + y_{t-1} + \sigma\varepsilon_t,$$

so that

$$\mathbb{E}_t[y_{t+1}] = \mathbb{E}_t[\mu + y_{t+1} + \varepsilon_{t+1}] = y_t + \mu.$$

Consequently,

$$\mathbb{E}_t[\Delta y_{t+1}] = \mu.$$

This is the Efficient Market Hypothesis model of Eugene Fama (1970): If markets are efficient, you cannot forecast returns (other than the risk premium component).

Prices and stationarity

Let's take the unconditional expectation and investigate the limits of the initial observation as $n \to \infty$:

$$\lim_{t\to\infty} \mathbb{E}[p_t|p_0] = \lim_{t\to\infty} p_0 + \mu t = \begin{cases} -\infty & \text{if } \mu < 0 \\ p_0 & \text{if } \mu = 0 \\ \infty & \text{if } \mu > 0 \end{cases}$$

This means that, if μ is different from 0, the unconditional mean does not exist and we see that for any finite t the expectation is a function of t. For $\mu = 0$, it looks like we are fine. What about the variance?

$$\lim_{t\to\infty} \mathbb{V}[p_t|p_0] = \lim_{t\to\infty} t\sigma^2 = \infty.$$

The unconditional variance of a random walk does not exists and we have shown that the price process is nonstationary!

Returns and stationarity

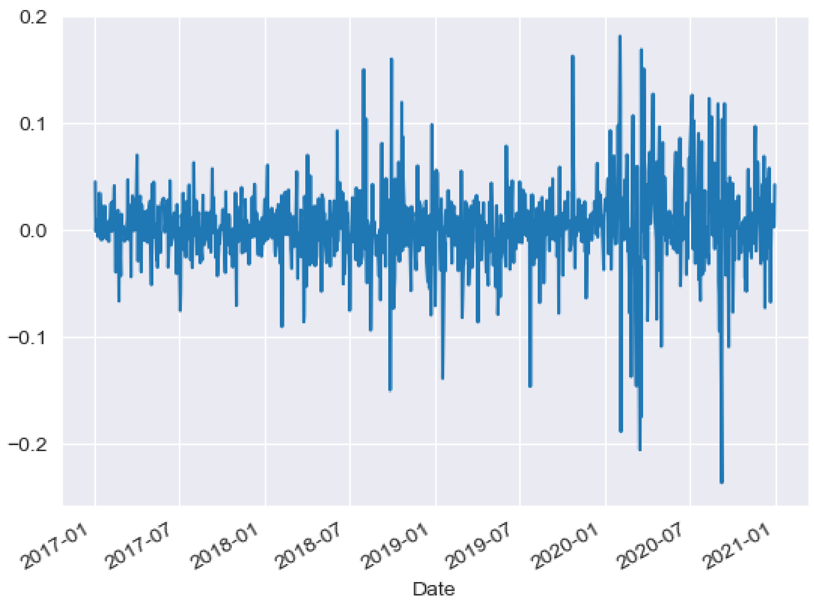
The return process was given by

$$r_t = \mu + \sigma \varepsilon.$$

$$\begin{aligned} \mathbb{E}[r_t] = \mathbb{E}[\mu + \sigma\varepsilon_t] = \mu \quad & \text{for all t,} \\ \mathbb{V}[r_t] = \mathbb{V}[\mu + \sigma\varepsilon_t] = \sigma^2 \quad & \text{for all t.} \end{aligned}$$

Hence, the return process is stationary! This holds in the data as well:

Figure 11: Stock price returns of Tesla



Note that it is the unconditional mean and variance that needs to be constant. The conditional mean and variance can move around.

Stationarity condition for an AR(p) model

Assume we have an AR(p) process, y_t, given by

$$\phi(L) y_t = \varepsilon_t.$$

Then, y_t is stationary if it is possible to write

$$y_t = \phi(L)^{-1} \varepsilon_t,$$

with $\phi(L)^{-1}$ converging to zero. This means that the autocorrelations will decline eventually as the lag length is increased. The condition for testing for the stationarity AR(p) model is that the roots of the characteristic equation

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$$

all lie outside the unit circle.

Example: Is the following model stationary?

$$y_t = y_{t-1} + \varepsilon_t.$$

In order to test whether this model is stationary, we start by rewriting it using the lag operator and rearrange:

$$\begin{aligned} y_t =& y_{t-1} + \varepsilon_t \\ y_t =& L y_t + \varepsilon_t \\ y_t(1-L) =& \varepsilon_t. \end{aligned}$$

The characteristic equation is

$$1 - z = 0,$$

where the root is $z = 1$ which lies on the unit circle. The model, which we know to be a random walk model, is therefore not stationary.

If the autoregressive model has more than one lag, it is not always obvious by eyeballing if the model is stationary or not. We can adopt the same procedure, let's assume that we have the following model:

$$y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + \varepsilon_t.$$

We rewrite:

$$\begin{aligned} y_t = 3Ly_t - 2.75L^2 y_t + 0.75L^3 y_t + \varepsilon_t, \\ y_t(1 - 3L + 2.75L^2 - 0.75L^3) = \varepsilon_t. \end{aligned}$$

The characteristic equation is

$$1 - 3z + 2.75z^2 - 0.75z^3 = 0$$

which factorises to

$$(1-z)(1-1.5z)(1-0.5z) = 0,$$

with roots $z = 2, z = 1$, and $z = 2/3$. Only one of these lie outside the unit circle, hence the process is not stationary. Remember that we can also use a root-finder to solve the polynomials of the characteristic equations:

```
coeff = [1,-3,2.75,-0.75]
np.roots(coeff)
Out[600]: array([1.5,  1. ,  0.5])
```

AR(1) as MA(∞)

Consider the simple AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t.$$

However, we know that $y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1}$. Substitute this into the equation above to get

$$y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \phi^2 y_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t.$$

If we continue this backwards substitution we get

$$y_t = \sum_{i=1}^{\infty} \phi^i \varepsilon_{t-i} + \varepsilon_t,$$

i.e. an infinite series of white noise processes. We therefore now see an example where the stationarity becomes important. If $|\phi| \geq 1$ we see that the sum is exploding as $i \to \infty$.

Example: Consider the following AR(1) model

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t.$$

1. Compute the unconditional mean of y_t
2. For simplicity, assume that $\mu = 0$. Compute the unconditional variance of y_t
3. For simplicity, assume that $\mu = 0$. Derive the autocorrelation function of this process.

1.

$$\begin{aligned}\mathbb{E}[y_t] =& \mathbb{E}[\mu + \phi y_{t-1} + \varepsilon_t] \\ =& \mu + \phi \mathbb{E}[y_{t-1}].\end{aligned}$$

Using backward substitution we get

$$\mathbb{E}[y_t] = \mu(1 + \phi + \phi^2 + ... + \phi^{n-1}) + \phi^n \mathbb{E}[y_{t-n}].$$

Assuming stationarity, i.e. that $|\phi| < 1$, $\lim_{t\to\infty} \phi^n \mathbb{E}[y_{t-n}] = 0$, leaving us with

$$\mathbb{E}[y_t] = \mu(1 + \phi + \phi^2 + ... + \phi^{n-1}).$$

By using the rule of algebra for finite sum of an infinite geometric series we can rewrite this as

$$\mathbb{E}[y_t] = \frac{\mu}{1-\phi}.$$

2.

With $\mu = 0$ we have

$$y_t = \phi y_{t-1} + \varepsilon_t.$$

We know that we can represent this process as an MA(∞):

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ...,$$

and now the unconditional mean is 0.

The variance can be computed by

$$\begin{aligned}\mathbb{V}[y_t] =& \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_t - \mathbb{E}[y_t])] \\ =& \mathbb{E}[(y_t - 0)(y_t - 0)] \\ =& \mathbb{E}[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ...)(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ...)] \\ =& \mathbb{E}[\varepsilon_t^2 + \phi^2\varepsilon_{t-1}^2 + ... + \text{cross-products}]\end{aligned}$$

Since $\mathbb{E}[\varepsilon_t^2] = \sigma^2$ for all t, then this can be rewritten as

$$\begin{aligned}\mathbb{V}[y_t] =& \sigma^2 + \phi^2\sigma^2 + ... \\ =& \sigma^2(1 + \phi^2 + ...) \\ =& \frac{\sigma^2}{1-\phi^2}.\end{aligned}$$

3.

Before we find the autocorrelation function we need to find the autocovariance functions.

$$
\begin{aligned}
\gamma_1 = cov(y_t, y_{t-1}) &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-1} - \mathbb{E}[y_{t-1}])] \\
&= \mathbb{E}[(y_t)(y_{t-1})] \\
&= \mathbb{E}[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ...)(\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + ...)] \\
&= \mathbb{E}[(\phi\varepsilon_{t-1}^2 + \phi^3\varepsilon_{t-2}^2 + ... + \text{cross-products})] \\
&= \phi\sigma^2(1 + \phi^2 + ...) \\
&= \frac{\phi\sigma^2}{1-\phi^2}.
\end{aligned}
$$

$$
\begin{aligned}
\gamma_2 = cov(y_t, y_{t-2}) &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-2} - \mathbb{E}[y_{t-2}])] \\
&= \mathbb{E}[(y_t)(y_{t-2})] \\
&= \mathbb{E}[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ...)(\varepsilon_{t-2} + \phi\varepsilon_{t-3} + \phi^2\varepsilon_{t-4} + ...)] \\
&= \mathbb{E}[(\phi^2\varepsilon_{t-2}^2 + \phi^4\varepsilon_{t-3}^2 + ... + \text{cross-products})] \\
&= \phi^2\sigma^2(1 + \phi^2 + ...) \\
&= \frac{\phi^2\sigma^2}{1-\phi^2}.
\end{aligned}
$$

Notice that $\gamma_2 = \phi\gamma_1 = \phi^2\gamma_0$. If we would have computed γ_3 we would get $\gamma_3 = \phi\gamma_2$. For any lag l, we would have that

$$\gamma_l = \frac{\phi^l\sigma^2}{1-\phi^2}.$$

We can now compute the autocorrelation function:

$$
\begin{aligned}
\tau_0 &= \frac{\gamma_0}{\gamma_0} = 1, \\
\tau_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\phi\gamma_0}{\gamma_0} = \phi, \\
\tau_2 &= \frac{\gamma_2}{\gamma_0} = \frac{\phi^2\gamma_0}{\gamma_0} = \phi^2, \quad \text{etc.}
\end{aligned}
$$

The autocorrelation of lag l is then

$$\tau_l = \phi^l.$$

5.2.5 MA(1) as AR(∞)

Assume we have the following MA(1) process

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad |\theta| < 1.$$

$|\theta| < 1$ is called the invertibility condition. Using the lag operator:

$$
\begin{aligned}
y_t &= \varepsilon_t - \theta_{\varepsilon t-1} \\
y_t &= (1-\theta L)\varepsilon_t \\
\frac{y_t}{(1-\theta L)} &= \varepsilon_t.
\end{aligned}
$$

Now, we recognize that $\frac{y_t}{(1-\theta L)}$ can be written as an infinite geometric series:

$$\frac{y_t}{(1-\theta L)} = y_t + \theta L y_t + \theta^2 L^2 y_t + ...,$$

so

$$\begin{aligned}
\frac{y_t}{(1-\theta L)} &= \varepsilon_t \\
y_t + \theta L y_t + \theta^2 L^2 y_t + ... &= \varepsilon_t \\
y_t &= -\theta L y_t - \theta^2 L^2 y_t - ... + \varepsilon_t,
\end{aligned}$$

which we recognise as an AR(∞) process!

5.2.6 Forecasting with an MA(1) and AR(1)

Assume we have the two following models,

$$\begin{aligned}
y_t &= \phi y_{t-1} + \varepsilon_t \\
x_t &= \theta \varepsilon_{t-1} + \varepsilon_t,
\end{aligned}$$

where y_t is an AR(1) model, x_t is an MA(1) model, and ε is a white noise process. What are the p periods ahead forecasts of these models? Let's first try with the AR(1) model:

1 period:

$$\begin{aligned}
\mathbb{E}_t[y_{t+1}] =& \mathbb{E}_t[\phi y_{t+1} + \varepsilon_{t+1}] \\
=& \mathbb{E}_t[\phi(\phi y_t + \varepsilon_t)] = \phi y_t.
\end{aligned}$$

2 periods:

$$\begin{aligned}
\mathbb{E}_t[y_{t+2}] =& \mathbb{E}_t[\phi y_{t+2} + \varepsilon_{t+2}] \\
=& \mathbb{E}_t[\phi(\phi y_{t+1} + \varepsilon_{t+1})] = \phi^2 y_t.
\end{aligned}$$

p periods:

$$\mathbb{E}_t[y_{t+p}] = \mathbb{E}_t[\phi y_{t+p} + \varepsilon_{t+p}] = \phi^p y_t.$$

What about the MA(1)?

1 period:

$$\begin{aligned}
\mathbb{E}_t[x_{t+1}] =& \mathbb{E}_t[\theta \varepsilon_t + \varepsilon_{t+1}] \\
=& \theta \varepsilon_t.
\end{aligned}$$

2 periods:

$$\begin{aligned}
\mathbb{E}_t[x_{t+2}] =& \mathbb{E}_t[\theta \varepsilon_{t+1} + \varepsilon_{t+2}] \\
=& 0.
\end{aligned}$$

p periods:

$$\begin{aligned}
\mathbb{E}_t[x_{t+p}] =& \mathbb{E}_t[\theta \varepsilon_{t+p-1} + \varepsilon_{t+p}] \\
=& 0.
\end{aligned}$$

5.2.7 ARMA processes

We can combine the moving average and autoregressive models. If we combine an MA(q) and AR(p) we obtain an ARMA(p,q) model, which can be written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

which can be rearranged and written as

$$\phi(L) y_t = \theta(L) \varepsilon_t,$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p,$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q.$$

Example:

1. Explain what stylized shapes would be expected for the autocorrelation function for the following stochastic processes:
   (a) White noise
   (b) AR(2)
   (c) MA(4)
   (d) ARMA(0,1)
   (e) ARMA(1,1)
2. Consider the following ARMA process

$$y_t = 0.21 + 1.32 y_{t-1} - 0.58 u_{t-1} + u_t$$

   (a) Determine whether the MA part of this process is invertible
   (b) Produce one-, two-, three-, and four step-ahead forecasts for the process

5.3 Modeling volatility

Some stylized facts about financial data

- Leptokurtosis - financial returns data tend to exhibit fat tails and excess peakedness at the mean
- Volatility clustering - volatility of financial returns data often appear in bunches. This means that high returns tends to be followed by high returns, and vice-versa. An explanation for this could be that, instead of being evenly spread out over time, news about assets tend to also come in bunches.
- Leverage effects - volatility tends to rise more following large price declines than a price increase of the same magnitude.

Figure 12: Stock price returns of Starbucks

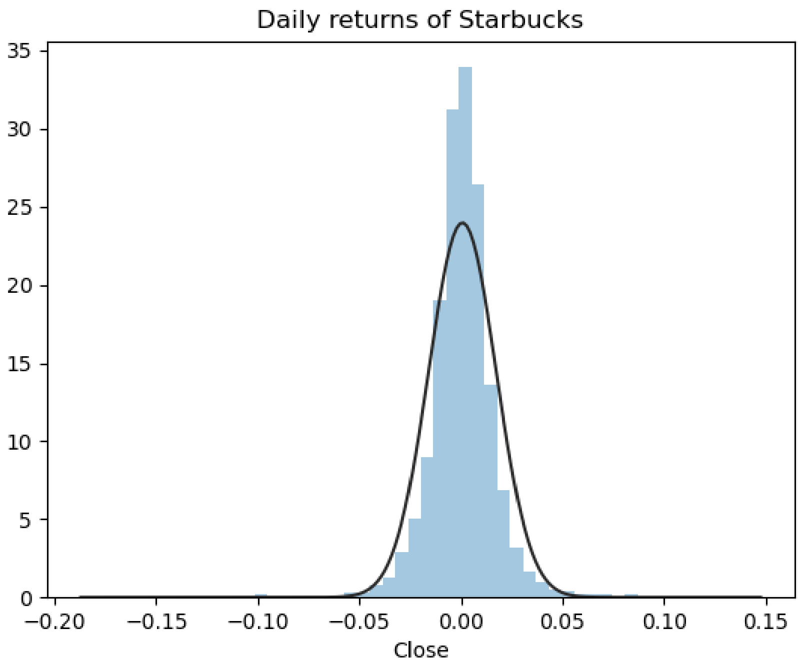


Figure 13: Stock price returns of Starbucks

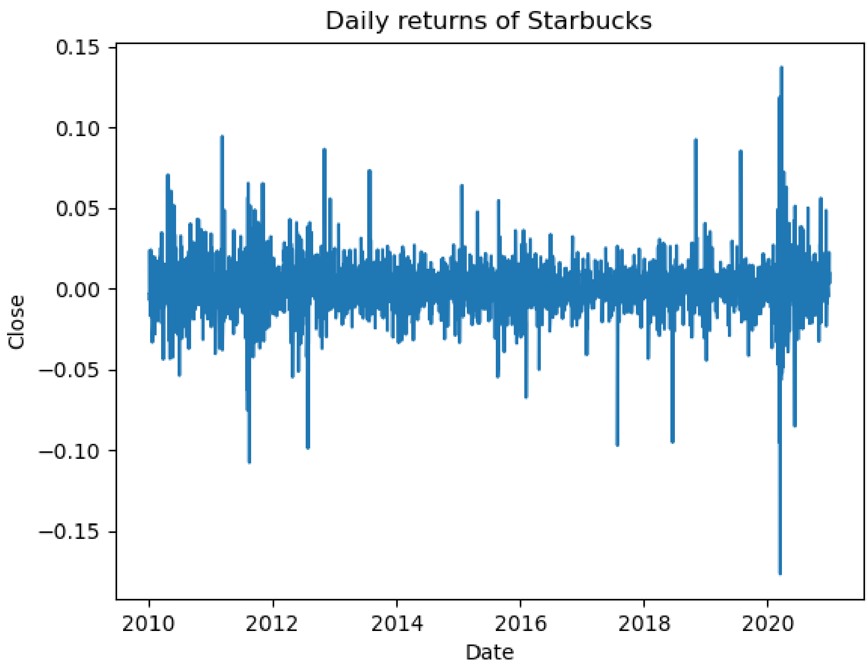


Figure 14: Stock price returns of Tesla

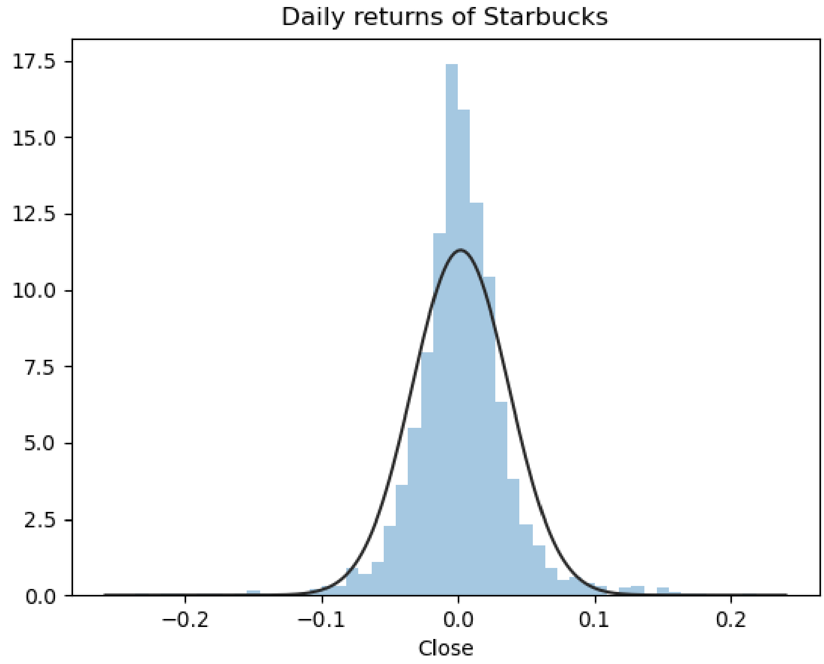
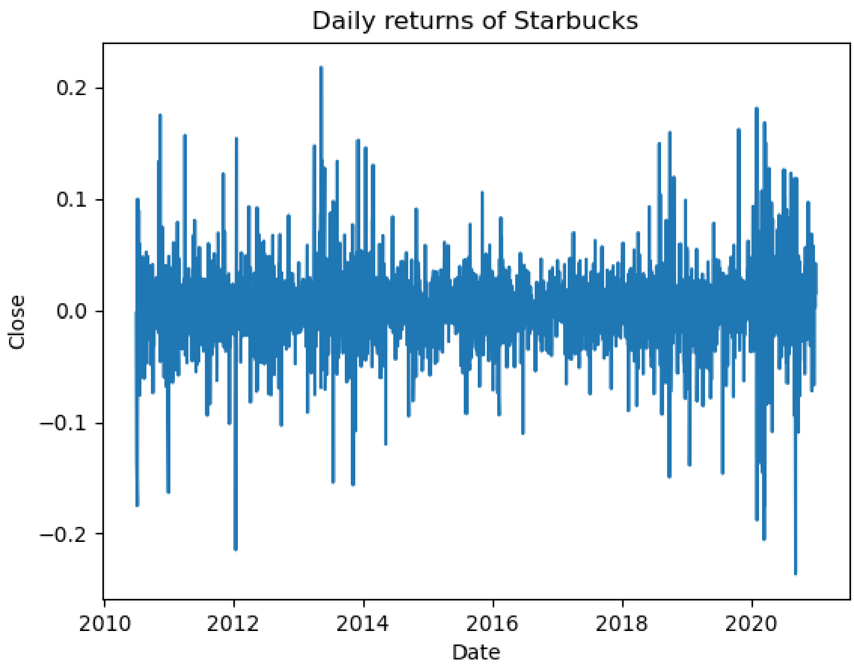


Figure 15: Stock price returns of Tesla



A few number of models have been found useful in modeling the non-linear relationships in volatility. Two of these are the ARCH or GARCH models. These models allows us to model behavior of time varying processes.

Autoregressive Conditional Heteroskedasticity (ARCH) models The distinction of conditional and unconditional variance is the same as that of the conditional and unconditional mean. Let's denote the conditional variance of ε_t as σ_t^2:

$$\sigma_t^2 = \text{var}(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, ...) = \mathbb{E}[(\varepsilon_t - \mathbb{E}[\varepsilon_t])^2|\varepsilon_{t-1}, \varepsilon_{t-2}...].$$

Since $\mathbb{E}[\varepsilon_t] = 0$:

$$\sigma_t^2 = \text{var}(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, ...) = \mathbb{E}[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}...]$$

In this model, the across-time dependence (autocorrelation) of volatility is modelled by allowing the conditional variance of the residual, σ_t^2 depend on the immediate previous value of the squared residual:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad \text{ARCH}(1)$$

This could easily be extended to the general case with q lags:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + ... + \alpha_q \varepsilon_{t-q}^2 \quad \text{ARCH}(q)$$

We can then incorporate this into almost any modeling framework that we want:

$$\begin{aligned} y_t &= \boldsymbol{Xb} + \varepsilon \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \end{aligned}$$

A more common way of expressing the relationship above is

$$\begin{aligned} y_t &= \boldsymbol{Xb} + \varepsilon, \\ \varepsilon_t &= v_t \sigma_t \quad v_t \sim \mathcal{N}(0, 1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \end{aligned}$$

Figure 16: Fitted conditional volatility using an ARCH(1) model

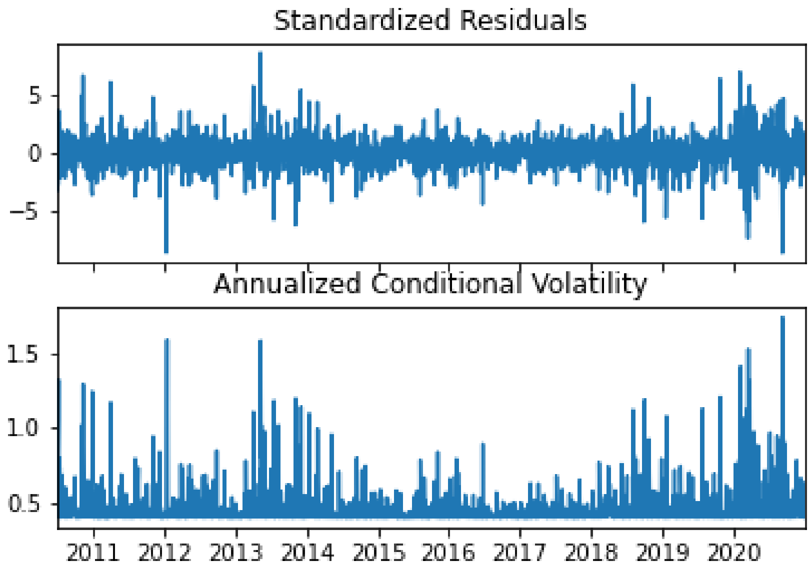
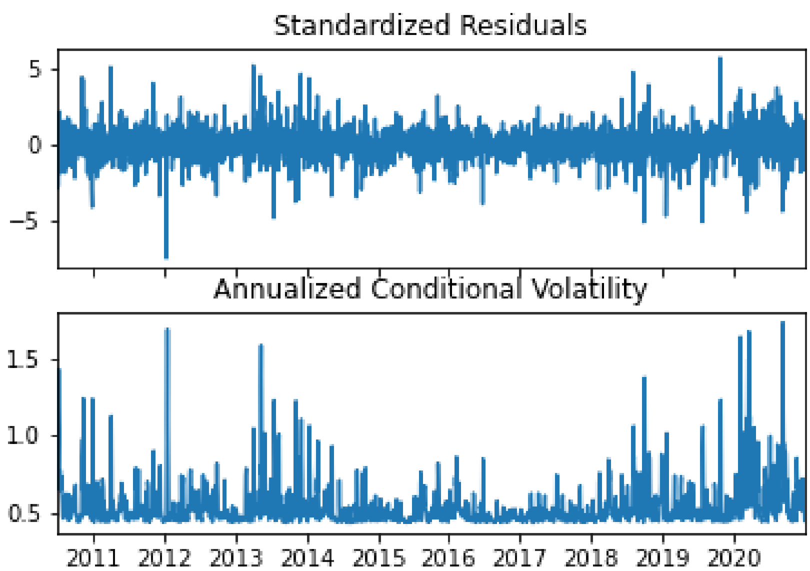


Figure 17: Fitted conditional volatility using an ARCH(5) model



Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models The GARCH model was developed independently by Bollerslev (1986) and Taylor (1986). The GARCH model allows us to model the conditional variance as being dependent upon previous own lags. We can now write

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad \text{GARCH}(1,1)$$

We see that the conditional variance, σ_t^2 is a one-period ahead estimate of the variance computed using any past information thought to be relevant. Using this framework, we can interpret the current fitted variance, σ_t^2, as a weighted function of a long-term average value (α_0), the volatility from the previous period $\alpha_1 \varepsilon_{t-1}^2$, and the fitted variance from the previous period ($\beta \sigma_{t-1}^2$). The GARCH model is usually preferred over the ARCH model, mostly due to the fact that it is parsimonious, and therefore avoids overfitting. To see this, write

$$\sigma^2_{t-1} = \alpha_0 + \alpha_1\varepsilon^2_{t-2} + \beta\sigma^2_{t-2}$$
$$\sigma^2_{t-2} = \alpha_0 + \alpha_1\varepsilon^2_{t-3} + \beta\sigma^2_{t-3}.$$

Now, substitute these expressions into the original expression for σ^2_t:

$$\sigma^2_t = \alpha_0(1+\beta+\beta^2) + \alpha_1\varepsilon^2_{t-1}(1+\beta L+\beta^2 L^2) + \beta^3\sigma^2_{t-3}.$$

Obviously, we can continue with this substitution. An infinite number of successive substitutions yields

$$\sigma^2_t = \alpha_0(1+\beta+\beta^2+...) + \alpha_1\varepsilon^2_{t-1}(1+\beta L+\beta^2 L^2+...) + \beta^\infty\sigma^2_{t-\infty}.$$

Hence, we can write the GARCH(1,1) as

$$\sigma^2_t = \eta_0 + \alpha_1\varepsilon^2_{t-1}(1+\beta L+\beta^2 L^2+...).$$

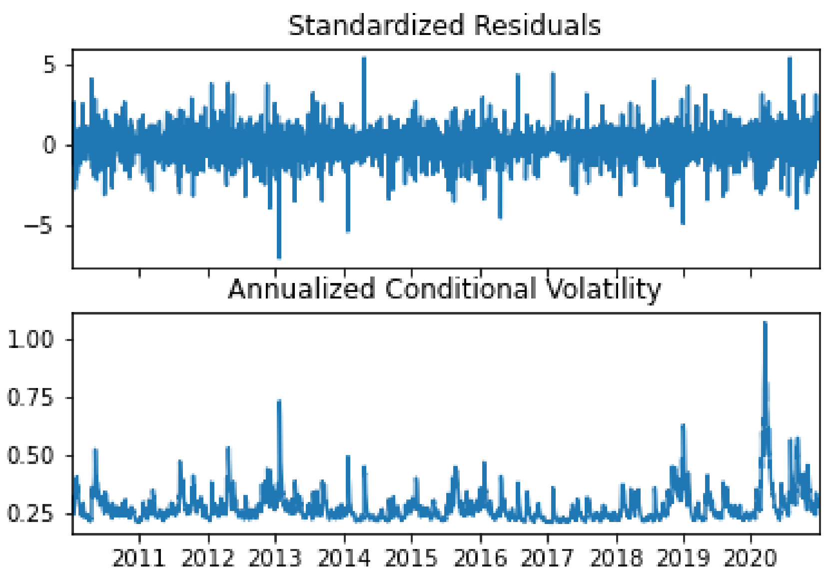
This is an ARCH(∞) representation. Thus, with only three parameters we can capture the past influence of an infinite number of past squared errors on the variance.

The GARCH model can be extended to a GARCH(p,q) formulation:

$$\begin{aligned}\sigma^2_t =& \alpha_0 + \alpha_1\varepsilon^2_{t-1} + \alpha_2\varepsilon^2_{t-2} + ... + \alpha_q\varepsilon^2_{t-q} + \beta_1\sigma^2_{t-1} + \beta_2\sigma^2_{t-2} + ... + \beta_2\sigma^2_{t-p}\\ \sigma^2_t =& \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon^2_{t-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t-j}.\end{aligned}$$

In general, a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data.

Figure 18: Estimated conditional volatility of returns using GARCH(1,1) on Apple



Example: GARCH(1,1)

1. Based on daily return data from Tesla from December 31, 2010 to December 31, 2020, obtain estimates from a GARCH(1,1) model. Also, compute the expected return.
2. Inerpret the parameter estimates.
3. Suppose Tesla's return today is -0.52% and your estimate yesterday for today's conditional volatility was 1.1%. What is your estimate for tomorrow's volatility?
4. Over the same period, redo 1. but for Starbucks.
5. Interpret the estimates. Compare them to the ones from 1. Whate are the main differences in the estimated coefficients?
6. Based on your interpretations from 2. and 5., which of the figures (figure 19-20) belong to which of the companies?

Figure 19: Company 1

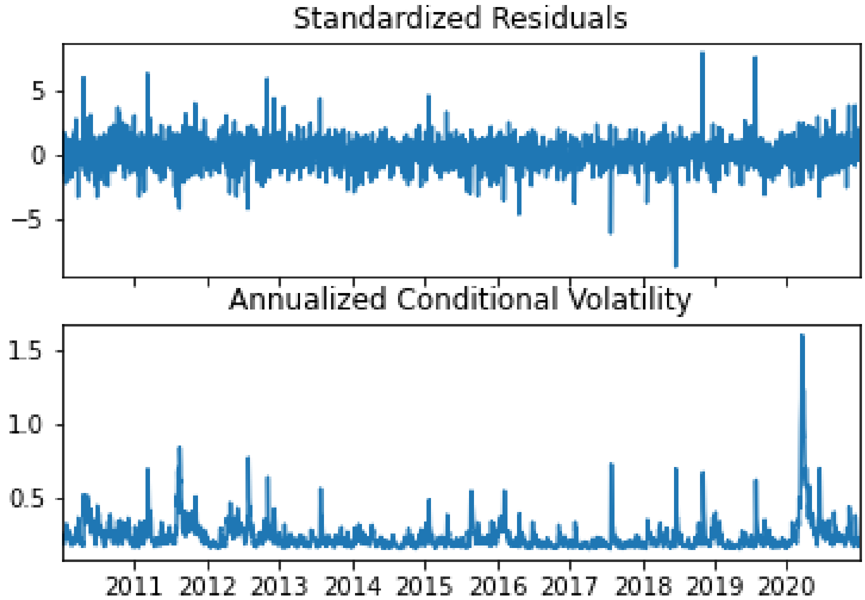
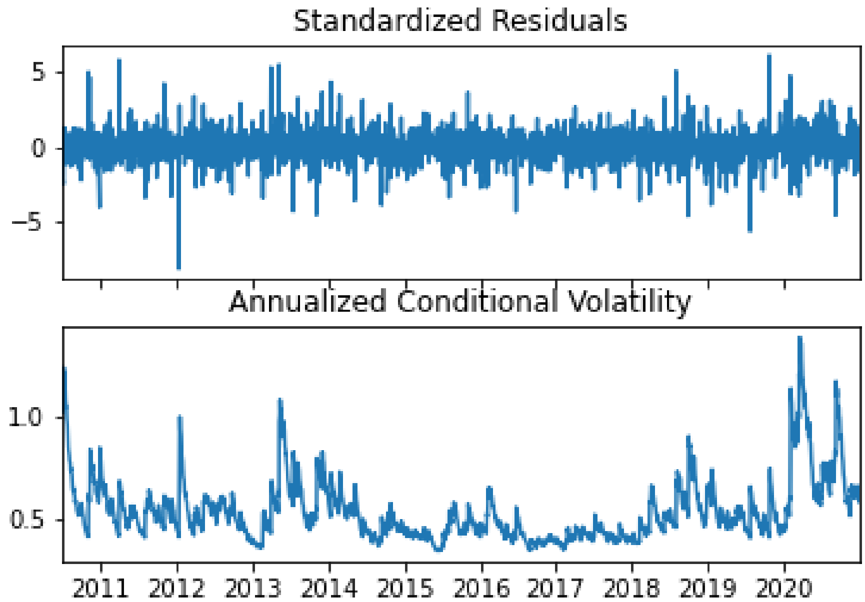


Figure 20: Company 2



1.

$$\hat{r}_t = 0.0018,$$
$$\hat{\sigma}_t^2 = 0.00002496 + 0.05\varepsilon_{t-1}^2 + 0.95\sigma_{t-1}^2.$$

All estimates are statistically significantly different from zero at the 1% significance level.

2.

- Mean: Expected return of Tesla is 0.18% per day.
- ARCH term: weight is 5%. This means that yesterday's squared unexpected shocks to Tesla's return (ε_{t-1}^2) have a small impact on today's conditional variance.
- GARCH term: weight is 93%. This means that yesterday's conditional variance (σ_{t-1}^2) is the main driver of the forecast for today's conditional variance (σ_t^2).

3.

- Unexpected return shock today:

$$\varepsilon_t = r_t - \mu = -0.0052 - 0.0018 = -0.007.$$

- Forecasted conditional variance for tomorrow:

$$\hat{\sigma}_{t+1}^2 = 0.00002496 + 0.05(-0.007)^2 + 0.950.0110^2 = 0.00013746.$$

- Forecasted conditional volatility for tomorrow:

$$\sqrt{0.00013746} \approx 1.17\%.$$

4.

$$\hat{r}_t = 0.0008,$$
$$\hat{\sigma}_t^2 = 0.00002496 + 0.05\varepsilon_{t-1}^2 + 0.95\sigma_{t-1}^2.$$

All estimates are statistically significantly different from zero at the 1% significance level.

5.

- Mean: Expected return of Tesla is 0.08% per day. Tesla seems to have a much higher expected return that Starbucks.
- ARCH term: weight is 20%. This means that yesterday's squared unexpected shocks to Starbucks' return (ε_{t-1}^2) have an impact on today's conditional variance, much more that in the case with Tesla. Therefore, the conditional variance of Starbucks can be much more "jumpy".
- GARCH term: weight is 70%. This means that yesterday's conditional variance (σ_{t-1}^2) is a strong driver of the forecast for today's conditional variance (σ_t^2). Tesla's conditional variance depends more on the previous period's estimate and is therefore more smooth.

6.

Figure 19 - Starbucks / figure 20 - Tesla

```
"""
Created on Tue Aug 31 23:16:30 2021

@author: ruslebiffen
"""

import numpy             as np
import matplotlib.pyplot as plt
import yfinance          as yf
import seaborn           as sns
from scipy.stats         import norm
from datetime            import datetime
from arch                import arch_model

start     = datetime(2010,1,1)
end       = datetime(2020,12,31)
data      = yf.download('SBUX', start=start, end=end)
ret_daily = np.log(data.Close / data.Close.shift(1))

# a standard GARCH(1,1) model
data             = yf.download('AAPL', start=start, end=end)
ret_daily_apple = np.log(data.Close / data.Close.shift(1))
garch = arch_model(ret_daily_apple.dropna(), vol='garch', p=1, o=0, q=1)
garch_fitted = garch.fit()
fig          = garch_fitted.plot(annualize="D")
#fig.savefig('garchsAPPLE.png')

# Scatter plot of data points
plt.figure()
ax=sns.distplot(ret_daily,fit=norm,kde=False)
ax.set(title   = "Daily returns of Starbucks")
#ax.get_figure().savefig('Tesla.png')


plt.figure()
ax=sns.lineplot(data=ret_daily)
ax.set(title   = "Daily returns of Starbucks")
#ax.get_figure().savefig('Tesla_2.png')

data  = yf.download('TSLA', start=start, end=end)
ret_daily_tsla    = np.log(data.Close / data.Close.shift(1))
data  = yf.download('SBUX', start=start, end=end)
ret_daily_sbux    = np.log(data.Close / data.Close.shift(1))

garch = arch_model(ret_daily_tsla.dropna(), vol='garch', p=1, o=0, q=1)
garch_fitted = garch.fit()
mean_tsla    = ret_daily_tsla.mean()
fig          = garch_fitted.plot(annualize="D")
#fig.savefig('garchstsla.png')

garch = arch_model(ret_daily_sbux.dropna(), vol='garch', p=1, o=0, q=1)
garch_fitted = garch.fit()
```

```
mean_sbux      = ret_daily_sbux.mean()
fig            = garch_fitted.plot(annualize="D")
#fig.savefig('garchsbux.png')
```

6 Martingales

Definition 6.1: Martingales

A process $\{S_t, t \in [0,\infty)\}$ is a martingale with respect to the filtration $\{\mathscr{F}_t, t \in [0,\infty)\}$ and probability measure $\mathbb{P}$ if $\forall\ t > 0$:

i. S_t is adapted to the filtration $\mathscr{F}_t$.

ii. S_t is integrable: $\mathbb{E}[|S_t|] < \infty$.

iii. $\mathbb{E}[S_T|\mathscr{F}_t] = S_t\ \forall\ t \leq T$.

Informally, a martingale has the property that *the best forecast of unobserved future values is the last observation.*

a. Submartingale: $\mathbb{E}[S_T|\mathscr{F}_t] \geq S_t\ \forall\ t \leq T$.

b. Supermartingale: $\mathbb{E}[S_T|\mathscr{F}_t] \leq S_t\ \forall\ t \leq T$.

Future increments of martingales are expected to be zero:

$$\mathbb{E}[S_{t+u} - S_t|\mathscr{F}_t] = \mathbb{E}[S_{t+u}|\mathscr{F}_t] - \mathbb{E}[S_t|\mathscr{F}_t] = S_t - S_t = 0$$

6.1 Martingale Problems

1. Show that the process $Y(t) = W^2(t) - t$ is a martingale.
2. Consider the process $Z_t = \int_0^t W_u du,\ t > 0$
   a. Show that $\mathbb{E}[Z_T|\mathscr{F}_t] = Z_t + W_t(T-t)$ for any $t < T$.
   b. Prove that the process $M_t = Z_t - tW_t$ is a martingale.
3. Let $\{W_t\}_{t\geq 0}$, be a Brownian motion and let $\{\mathscr{F}_t\}_{t\geq 0}$, be a filtration associated with this Brownian motion. Show that the process $Y(t) = e^{-t/2}e^{W_t}$ is a martingale/

6.2 Martingale Problems - Solutions

1. Show that the process $Y(t) = W^2(t) - t$ is a martingale.

   (**Adaptability**) $W^2(t)$ is observable at time t. Also, t is known at all times. Thus $Y(t) \in \mathscr{F}_t$.
   (**Integrability**)

$$\mathbb{E}[|Y(t)|] = \mathbb{E}[|W^2(t) - t|] \le \mathbb{E}[|W^2(t) + t|] = \mathbb{E}[W^2(t)] + t = t + t = 2t < \infty$$

   (**Martingale property**) $t \ge s$

$$\begin{aligned}\mathbb{E}[Y(t)|\mathscr{F}_s] &= \mathbb{E}[W^2(t) - t|\mathscr{F}_s] = \mathbb{E}[(W(t) - W(s) + W(s))^2|\mathscr{F}_s] - t \\ &= \mathbb{E}[(W(t) - W(s))^2|\mathscr{F}_s] + \mathbb{E}[W(s)^2|\mathscr{F}_s] - t = t - s + W(s)^2 - t \\ &= W(s)^2 - s = Y(s)\end{aligned}$$

2. Consider the process $Z_t = \int_0^t W_u du$, $t > 0$

   a. Show that $\mathbb{E}[Z_T|\mathscr{F}_t] = Z_t + W_t(T - t)$ for any $t < T$.

$$\begin{aligned}\mathbb{E}[Z_T|\mathscr{F}_t] &= \mathbb{E}\Big[\int_0^T W_u du|\mathscr{F}_t\Big] = \mathbb{E}\Big[\int_0^t W_u du + \int_t^T W_u du|\mathscr{F}_t\Big] \\ &= \mathbb{E}\Big[\int_0^t W_u du|\mathscr{F}_t\Big] + \mathbb{E}\Big[\int_t^T W_u du|\mathscr{F}_t\Big] = \int_0^t W_u du + \int_t^T \mathbb{E}_t[W_u] du \\ &= Z_t + \int_t^T W_t du = Z_t + W_t \int_t^T du = Z_t + W_t(T - t)\end{aligned}$$

   b. Prove that the process $M_t = Z_t - tW_t$ is a martingale.

   (**Adaptability**) Z_t and W_t is observable at time t. Also, t is known at all times. Thus $M_t \in \mathscr{F}_t$.
   (**Integrability**)

$$\begin{aligned}\mathbb{E}[|M_t|] &= \mathbb{E}[|Z_t - tW_t|] \le \mathbb{E}[|Z_t + tW_t|] = \mathbb{E}[|Z_t|] + t\mathbb{E}[|W_t|] \\ &= \int_0^t \mathbb{E}[|W_u|] du + t(\sqrt{2t/\pi}) \text{ (see first Brownian motion problem)} \\ &= \int_0^t \sqrt{\frac{2u}{\pi}} du + \frac{t\sqrt{2t}}{\sqrt{\pi}} = \sqrt{\frac{2}{\pi}} \cdot \frac{2}{3}\Big[u^{3/2}\Big]_0^t + \frac{t\sqrt{2t}}{\sqrt{\pi}} \\ &= \frac{2}{3} \cdot \frac{t^{3/2}\sqrt{2}}{\sqrt{\pi}} + \frac{t^{3/2}\sqrt{2}}{\sqrt{\pi}} = \frac{5}{3}\sqrt{\frac{2}{\pi}} t^{3/2} < \infty\end{aligned}$$

   (**Martingale property**) $t \ge s$

$$\begin{aligned}\mathbb{E}[M_t|\mathscr{F}_s] &= \mathbb{E}\Big[\int_0^t W_u du - tW_t|\mathscr{F}_s\Big] = \mathbb{E}_s\Big[\int_0^t W_u du\Big] - t\mathbb{E}_s[W_t] \\ &= \mathbb{E}_s\Big[\int_0^s W_u du\Big] + \mathbb{E}_s\Big[\int_s^t W_u du\Big] - tW_s \\ &= \int_0^s W_u du + \int_s^t \mathbb{E}_s[W_u] du - tW_s = Z_s + \int_s^t W_s du - tW_s \\ &= Z_s + W_s \int_s^t du - tW_s = Z_s + W_s(t - s) - tW_s = Z_s - sW_s = M_s\end{aligned}$$

   Alternatively,

$$\mathbb{E}[M_t|\mathscr{F}_s] = \mathbb{E}_s[Z_t] - t\mathbb{E}_s[W_t] = Z_s + W_s(t - s) - tW_s = Z_s - sW_s = M_s$$

3. Let $\{W_t\}_{t\geq 0}$, be a Brownian motion and let $\{\mathscr{F}_t\}_{t\geq 0}$, be a filtration associated with this Brownian motion. Show that the process $Y(t) = e^{-t/2}e^{W_t}$ is a martingale.

   (**Adaptability**) W_t is observable at time t. Also, t is known at all times. Thus $Y(t) \in \mathscr{F}_t$.

   (**Integrability**)

$$\mathbb{E}[|Y(t)|] = \mathbb{E}[|e^{-t/2}e^{W_t}|] = e^{-t/2}\mathbb{E}[e^{W_t}] = e^{-t/2}e^{t/2} = 1 < \infty$$

   (**Martingale property**) $t \geq s$

$$\begin{aligned}\mathbb{E}[Y(t)|\mathscr{F}_s] &= \mathbb{E}[e^{-t/2}e^{W_t}|\mathscr{F}_s] = e^{-t/2}\mathbb{E}[e^{W_t}|\mathscr{F}_s] = e^{-t/2}\mathbb{E}[e^{(W_t-W_s)+W_s}|\mathscr{F}_s] \\ &= e^{-t/2}(\mathbb{E}[e^{W_t-W_s}|\mathscr{F}_s]\cdot\mathbb{E}[e^{W_s}|\mathscr{F}_s]) = e^{-t/2}(e^{0+\frac{t-s}{2}}\cdot e^{W_s}) \\ &= e^{-t/2}(e^{\frac{t}{2}-\frac{s}{2}+W_s}) = e^{s/2}e^{W_s} = Y(s)\end{aligned}$$

7 Brownian Motion

Definition 7.1: Brownian Motion

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, there is a continuous function $\{W(t)\}_{t\geq 0}$ that satisfies $W(0) = 0$ and depends on ω. Then $\{W(t)\}_{t\geq 0}$ is a Brownian motion if $\forall\ 0 = t_0 < t_1 < \cdots < t_m$ the increments $W(t_1) - W(t_0), W(t_2) - W(t_1), \cdots, W(t_m) - W(t_{m-1})$ are independent and normally distributed with:

$$\begin{aligned} \mathbb{E}[W(t_{i+1}) - W(t_i)] &= 0 \\ Var(W(t_{i+1}) - W(t_i)) &= t_{i+1} - t_i \end{aligned}$$

Properties of Brownian motion:

i. $W(t_1), \cdots, W(t_m)$ are jointly normally distributed.

ii. Covariance of two Brownian motions with $t > s$.

$$\begin{aligned} Cov(W_t, W_s) &= \mathbb{E}[W_t W_s] - \mathbb{E}[W_t] \cdot \mathbb{E}[W_s] \\ &= \mathbb{E}[(W_t - W_s + W_s)W_s] - 0 \cdot 0 \\ &= \mathbb{E}[(W_t - W_s)W_s] + \mathbb{E}[W_s^2] \\ &= 0 \cdot 0 + s = s \end{aligned}$$

iii. Quadratic variance of a Brownian motion is equal to the time that has passed, $[W, W](t) = t$.

iv. Correlation of a Brownian motion.

$$\begin{aligned} Corr(W_t, W_s) &= \frac{Cov(W_t, W_s)}{\sqrt{Var(W_t)} \cdot \sqrt{Var(W_s)}} \\ Corr(W_t, W_s) &= \begin{cases} \dfrac{s}{\sqrt{t \cdot s}} & \text{if } s \leq t \\ \dfrac{t}{\sqrt{t \cdot s}} & \text{if } t > s \end{cases} \\ &= \begin{cases} \dfrac{\sqrt{s}}{\sqrt{t}} & \text{if } s \leq t \\ \dfrac{\sqrt{t}}{\sqrt{s}} & \text{if } t > s \end{cases} \\ Corr(W_t, W_s) &= \sqrt{\frac{min\{t, s\}}{max\{t, s\}}} \end{aligned}$$

v. Martingale property. For $t \geq s$, $\mathbb{E}[W(t)|\mathscr{F}_s] = W(s)$

vi. Higher order moments. $\mathbb{E}[W_t^3] = 0$ and $\mathbb{E}[W_t^4] = 3t^2$.

Definition 7.2: Brownian Bridge

The process

$$X_t = W_t - \frac{t}{T} W_T$$

is called the Brownian bridge fixed at 0 and T.

Properties of Brownian bridge

i. Brownian bridge is zero at 0 and T.
ii. Variance of the process decreases as you approach T.
iii. Between 0 and T, any path is possible.
iv. Since the Brownian bridge contains W_T is is $\notin \mathscr{F}_t$ and consequently not a martingale.

A Brownian bridge can be used to model bonds.

7.1 Brownian Motion Problems

1. The process $X(t) = |W(t)|$ is called *Brownian motion reflected at the origin.* Show that $\mathbb{E}[X(t)] = \sqrt{2t/\pi}$ and $Var(X(t)) = t(1 - \frac{2}{\pi})$.

2. Consider a symmetric random walk $M_3 = \sum_{j=1}^{3} X_j = \begin{cases} 1, \text{probability } p \\ -1, \text{probability } (1-p) \end{cases}$

   a. Compute the the variance $Var(M_3)$ without assuming $p = \frac{1}{2}$ and the quadratic variation $[M, M]_3$.
   b. Show the variance and the quadratic variation are equal for $p = \frac{1}{2}$.

3. Brownian bridge. The process $X_t = W_t - \frac{t}{T} W_T$ is called the Brownian bridge fixed at 0 and T.

   a. Compute $\mathbb{E}[X_t]$ and $Var(X_t)$.
   b. Let $Y_t = X_t^2$. Show that $Y_0 = Y_T = 0$ and find $\mathbb{E}[Y_t]$ and $Var(Y_t)$.

4. Let W_t be a Brownian motion wrt. a filtration $\mathscr{F}_t$.

   a. Evaluate $Cov(W_t^2, W_s^2)$, for $0 < s \leq t$.
   b. Evaluate $\mathbb{E}[W_t^2 | \mathscr{F}_s]$ for $0 < s \leq t$.

5. Show that $\mathbb{E}[e^{W_s + W_u}] = e^{\frac{u+s}{2}} e^{min\{u,s\}}$.

6. Let $s \leq t$. Find the covariance of the two Brownian bridge processes $X_t = W_t - \frac{t}{T} W_T$ and $Y_s = W_s - \frac{s}{T} W_T$.

7.2 Brownian Motion Problems - Solutions

1. The process $X(t) = |W(t)|$ is called *Brownian motion reflected at the origin.* Show that $\mathbb{E}[X(t)] = \sqrt{2t/\pi}$ and $Var(X(t)) = t(1-\frac{2}{\pi})$.

$$\begin{aligned}
\mathbb{E}[X(t)] = \mathbb{E}[|W(t)|] &= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty}|x|e^{-\frac{x^2}{2t}}dx = \frac{1}{\sqrt{2\pi t}}\int_{0}^{\infty}2xe^{-\frac{x^2}{2t}}dx \\
&\text{Let } y = x^2 \Rightarrow dy = 2xdx \\
&\Rightarrow \frac{1}{\sqrt{2\pi t}}\int_{0}^{\infty}e^{-\frac{y}{2t}}dy = \frac{1}{\sqrt{2\pi t}}\Big[-2te^{-\frac{y}{2t}}\Big]_0^{\infty} = \frac{2t}{\sqrt{2\pi t}} = \sqrt{2t/\pi} \\
Var(X(t)) = Var(|W(t)|) &= \mathbb{E}[|W^2(t)|] - (\mathbb{E}[|W(t)|])^2 = \mathbb{E}[W^2(t)] - (\sqrt{2t/\pi})^2 \\
&= t - \frac{2t}{\pi} = t(1-\frac{2}{\pi})
\end{aligned}$$

2. Consider a symmetric random walk $M_3 = \sum_{j=1}^{3} X_j = \begin{cases} 1, \text{probability } p \\ -1, \text{probability } (1-p) \end{cases}$

   a. Compute the the variance $Var(M_3)$ without assuming $p = \frac{1}{2}$ and the quadratic variation $[M, M]_3$.

$$\begin{aligned}
Var(X_j) &= \mathbb{E}[X_j^2] - (\mathbb{E}[X_j])^2 \\
\mathbb{E}[X_j^2] &= 1^2 \cdot p + (-1)^2 \cdot (1-p) = p + 1 - p = 1 \\
\mathbb{E}[X_j] &= 1 \cdot p + (-1) \cdot (1-p) = 2p - 1 \\
Var(X_j) &= 1 - (2p-1)^2 = 1 - (4p^2 + 4p + 1) = 4p - 4p^2 \\
Var(M_3) &= Var\Big(\sum_{i=1}^{3} X_j\Big) \overset{\text{ind}}{=} \sum_{i=1}^{3}\Big(Var(X_j)\Big) = 3(4p - 4p^2) = 12p - 12p^2 \\
[M, M]_3 &= \sum_{i=1}^{3} X_j^2 = 3
\end{aligned}$$

   b. Show the variance and the quadratic variation are equal for $p = \frac{1}{2}$.

$$\Big[Var(M_3|p = \frac{1}{2})\Big] = 12\frac{1}{2} - 12\Big(\frac{1}{2}\Big)^2 = 6 - 3 = 3 \text{ and } [M, M]_3 = 3 \;\forall\; p$$

3. Brownian bridge. The process $X_t = W_t - \frac{t}{T}W_T$ is called the Brownian bridge fixed at 0 and T.

   a. Compute $\mathbb{E}[X_t]$ and $Var(X_t)$.

$$\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[W_t - \frac{t}{T}W_T] = \mathbb{E}[W_t] - \frac{t}{T}\mathbb{E}[W_T] = 0 - \frac{t}{T}0 = 0 \\
Var(X_t) &= \mathbb{E}[X_t^2] = \mathbb{E}[(W_t - \frac{t}{T}W_T)^2] = \mathbb{E}[W_t^2 - 2\frac{t}{T}W_tW_T + \frac{t^2}{T^2}W_T^2] \\
&= \mathbb{E}[W_t^2] - 2\frac{t}{T}\mathbb{E}[W_tW_T] + \frac{t^2}{T^2}\mathbb{E}[W_T^2] = t - \frac{2t}{T}\cdot t + \frac{t^2}{T^2}\cdot T \\
&= t - \frac{2t^2}{T} + \frac{t^2}{T} = t(1 - \frac{t}{T})
\end{aligned}$$

b. Let $Y_t = X_t^2$. Show that $Y_0 = Y_T = 0$ and find $\mathbb{E}[Y_t]$ and $Var(Y_t)$.

$$\begin{aligned}
Y_0 &= W_0^2 - 2\frac{0}{T}W_0W_T + \frac{0^2}{T^2}W_T^2 = 0 - 2\frac{0}{T}0W_T + \frac{0^2}{T^2}W_T^2 = 0\\
Y_T &= W_T^2 - 2\frac{T}{T}W_TW_T + \frac{T^2}{T^2}W_T^2 = W_T^2 - 2W_T^2 + W_T^2 = 0\\
\mathbb{E}[Y_t] &= \mathbb{E}[X_t^2] = t(1-\frac{t}{T})\\
\mathbb{E}[Y_t^2] &= \mathbb{E}[X_t^4] = \mathbb{E}[(W_t - \frac{t}{T}W_T)^4]\\
&= \mathbb{E}[W_t^4 - 4\frac{t}{T}W_t^3W_T + 6\frac{t^2}{T^2}W_t^2W_T^2 - 4\frac{t^3}{T^3}W_tW_T^3 + \frac{t^4}{T^4}W_T^4]\\
&= \mathbb{E}[W_t^4] - 4\frac{t}{T}\mathbb{E}[W_t^3W_T] + 6\frac{t^2}{T^2}\mathbb{E}[W_t^2W_T^2] - 4\frac{t^3}{T^3}[W_tW_T^3] + \frac{t^4}{T^4}\mathbb{E}[W_T^4]\\
\mathbb{E}[W_t^3W_T] &= \mathbb{E}[W_t^4 - W_t^3(W_T - W_t)] = \mathbb{E}[W_t^4] - \mathbb{E}[W_t^3]\cdot\mathbb{E}[W_T - W_t] = 3t^2 - 0\cdot 0 = 3t^2\\
\mathbb{E}[W_t^2W_T^2] &= \mathbb{E}[(W_t^2 + W_t(W_T - W_t))^2] = \mathbb{E}[W_t^4 + 2W_t^2W_t(W_T - W_t) + W_t^2(W_T - W_t)^2]\\
&= \mathbb{E}[W_t^4] + 2(\mathbb{E}[W_t^3]\cdot\mathbb{E}[(W_T - W_t)]) + \mathbb{E}[W_t^2]\cdot\mathbb{E}[(W_T - W_t)^2]\\
&= 3t^2 + 2(0\cdot 0) + t\cdot(T-t) = 3t^2 + t(T-t)\\
\mathbb{E}[W_tW_T^3] &= \mathbb{E}[((W_T - W_t) + W_t)^3W_t]\\
&= \mathbb{E}[(W_T - W_t)^3W_t + 3(W_T - W_t)^2W_t^2 + 3(W_T - W_t)W_t^3 + W_t^4]\\
&= \mathbb{E}[(W_T - W_t)^3]\cdot\mathbb{E}[W_t] + 3(\mathbb{E}[(W_T - W_t)^2]\cdot\mathbb{E}[W_t^2]) + 3(\mathbb{E}[W_T - W_t]\cdot\mathbb{E}[W_t^3]) + \mathbb{E}[W_t^4]\\
&= 0\cdot 0 + 3((T-t)\cdot t) + 3(0\cdot 0) + 3t^2 = 3t(T-t) + 3t^2\\
\mathbb{E}[Y_t^2] &= 3t^2 - 4\frac{t}{T}3t^2 + 6\frac{t^2}{T^2}(3t^2 + t(T-t)) - 4\frac{t^3}{T^3}(3t(T-t) + 3t^2) + \frac{t^4}{T^4}3T^2\\
&= 3t^2 - \frac{6t^3}{T} + \frac{3t^4}{T^2}\\
Var(Y_t) &= \mathbb{E}[Y_t^2] - (\mathbb{E}[Y_t])^2 = 3t^2 - \frac{6t^3}{T} + \frac{3t^4}{T^2} - (t - \frac{t^2}{T})^2\\
&= 3t^2 - \frac{6t^3}{T} + \frac{3t^4}{T^2} - (t^2 - \frac{2t^3}{T} + \frac{t^4}{T^2}) = 2t^2 - \frac{4t^3}{T} + \frac{2t^4}{T^2}
\end{aligned}$$

4. Let W_t be a Brownian motion wrt. a filtration $\mathscr{F}_t$.

a. Evaluate $Cov(W_t^2, W_s^2)$, for $0 < s \le t$.

$$\begin{aligned}
Cov(W_t^2, W_s^2) &= \mathbb{E}[W_t^2\cdot W_s^2] - \mathbb{E}[W_t^2]\cdot\mathbb{E}[W_s^2]\\
\mathbb{E}[W_t^2\cdot W_s^2] &= \mathbb{E}[((W_t - W_s) + W_s)^2W_s^2] = \mathbb{E}[(((W_t - W_s) + W_s)W_s))^2]\\
&= \mathbb{E}[((W_t - W_s)W_s + W_s^2))^2] = \mathbb{E}[(W_t - W_s)^2W_s^2 + 2(W_t - W_s)W_s^3 + W_s^4]\\
&= \mathbb{E}[(W_t - W_s)^2]\cdot\mathbb{E}[W_s^2] + 2(\mathbb{E}[W_t - W_s]\cdot\mathbb{E}[W_s^3]) + \mathbb{E}[W_s^4]\\
&= (t-s)\cdot s + 2(0\cdot 0) + 3s^2 = ts + 2s^2\\
Cov(W_t^2, W_s^2) &= ts + 2s^2 - ts = 2s^2
\end{aligned}$$

b. Evaluate $\mathbb{E}[W_t^2|\mathscr{F}_s]$ for $0 < s \le t$.

$$\begin{aligned}
\mathbb{E}[W_t^2|\mathscr{F}_s] &= \mathbb{E}[((W_t - W_s) + W_s)^2|\mathscr{F}_s] = \mathbb{E}_s[(W_t - W_s)^2] + 2\mathbb{E}_s[(W_t - W_s)W_s] + \mathbb{E}_s[W_s^2]\\
&= (t-s) + 2(0\cdot W_s) + W_s^2 = W_s^2 + t - s
\end{aligned}$$

5. Show that $\mathbb{E}[e^{W_s + W_u}] = e^{\frac{u+s}{2}}e^{min\{u,s\}}$.

Let $t = max\{s,u\}$ and $v = min\{s,u\}$

$$\begin{aligned}
\mathbb{E}[e^{W_t + W_v}] &= \mathbb{E}[e^{W_t - W_v + 2W_v}] = \mathbb{E}[e^{W_t - W_v}e^{2W_v}] = \mathbb{E}[e^{W_t - W_v}]\cdot\mathbb{E}[e^{2W_v}]\\
&= e^{0 + \frac{t-v}{2}}\cdot e^{0 + \frac{4v}{2}} = e^{\frac{t+v}{2}}e^{v} = e^{\frac{u+s}{2}}e^{min\{u,s\}}
\end{aligned}$$

6. Let $s \leq t$. Find the covariance of the two Brownian bridge processes $X_t = W_t - \frac{t}{T} W_T$ and $Y_s = W_s - \frac{s}{T} W_T$.

$$\begin{aligned}
Cov(X_t, Y_s) &= \mathbb{E}[X_t \cdot Y_s] - \mathbb{E}[X_t] \cdot \mathbb{E}[Y_t] \\
\mathbb{E}[X_t] &= \mathbb{E}[W_t - \frac{t}{T} W_T] = \mathbb{E}[W_t] - \frac{t}{T} \mathbb{E}[W_T] = 0 - \frac{t}{T} 0 = 0 = \mathbb{E}[Y_s] \\
Cov(X_t, Y_s) &= \mathbb{E}[X_t \cdot Y_s] = \mathbb{E}[(W_t - \frac{t}{T} W_T) \cdot (W_s - \frac{s}{T} W_T)] \\
&= \mathbb{E}[W_t W_s - \frac{s}{T} W_t W_T - \frac{t}{T} W_s W_t + \frac{st}{T^2} W_T^2] \\
&= \mathbb{E}[W_t W_s] - \frac{s}{T} \mathbb{E}[W_t W_T] - \frac{t}{T} \mathbb{E}[W_s W_t] + \frac{st}{T^2} \mathbb{E}[W_T^2] \\
&= s - \frac{s}{T} t - \frac{t}{T} s + \frac{st}{T^2} T = s(1 - \frac{t}{T})
\end{aligned}$$

8 Integrals

Theorem 8.1: Riemann-Stieltjes

For a partition $\Pi = t_0 < t_1 < \cdots < t_m$ with $t_0 = 0$ and $t_m = T$. Define $t_i^* \in [t_{i-1}, t_i]$. Then we have

$$RS_n = \sum_{i=1}^{n} f(t_i^*)(W(t_i) - W(t_{i-1}))$$
$$\lim_{n\to\infty} \int_0^T f(s)dW(s)$$

where f is continuously differentiable.

The integration by parts rule is given by

$$\begin{aligned}\int_0^t f(s)dW(s) &= \Big[f(s)\cdot W(s)\Big]_0^t - \int_0^t W(s)\cdot df(s)\\ &= f(t)W(t) - f(0)W(0) - \int_0^t W(s)\cdot f'(s)ds\\ &= f(t)W(t) - \int_0^t W(s)\cdot f'(s)ds\end{aligned}$$

Autocovariance property of Riemann-Stieltjes integral can be stated as

$$\mathbb{E}\Big[\int_0^t f(s)dW(s)\cdot\int_0^{t+h} f(s)dW(s)\Big] = \int_0^t f^2(s)ds$$

Informally, we can think of this as an extension of the covariance of two Brownian motions

$$\mathbb{E}[W(t)W(t+h)] = t$$

Riemann-Stieltjes integral cannot be used if the integrand is a Brownian motion.

Lemma 8.1: Distribution of Riemann-Stieltjes Integral

$$\int_0^t f(s)dW(s) \sim \mathcal{N}(0, \int_0^t f^2(s)ds)$$

Proof:

$$\begin{aligned}
W(t_i) - W(t_{i-1}) &\sim \mathcal{N}(0, t_i - t_{i-1}) \\
\mathbb{E}[RS_n] &= \mathbb{E}\Big[\sum_{i=1}^{n} f(t_i^*) \cdot (W(t_i) - W(t_{i-1}))\Big] \\
&= \sum_{i=1}^{n} f(t_i^*) \cdot \mathbb{E}[W(t_i) - W(t_{i-1})] = 0 \\
Var(RS_n) &= Var\Big(\sum_{i=1}^{n} f(t_i^*) \cdot (W(t_i) - W(t_{i-1}))\Big) \\
&\text{since we have independent increments we can write} \\
&= \sum_{i=1}^{n} f^2(t_i^*) \cdot Var(W(t_i) - W(t_{i-1})) \\
&= \sum_{i=1}^{n} f^2(t_i^*) \cdot (t_i - t_{i-1}) \\
&= \lim_{n\to\infty} \int_0^t f^2(s)ds \qquad \square
\end{aligned}$$

8.1 Riemann-Stieltjes Integral Problems

1. Solve the integral $\int_0^t s dW(s)$.
2. Solve the integral $\int_0^t dW(s)$.
3. Specify distribution of $\int_0^1 s dW(s)$, and calculate its expectation and variance.
4. Specify distribution of $\int_0^1 (1-s)dW(s)$, and calculate its expectation and variance.
5. Specify distribution of $\int_0^1 dW(s)$, and calculate its expectation and variance.
6. A standard Ornstein-Uhlenbeck process is given by

$$X_c(t) = e^{ct}\int_0^t e^{-cs}dW(s) \qquad t \geq 0 \text{ and } X_c(0) = 0$$

   a. Use integration by parts to find an equivalent expression.
   b. Find the distribution of $X_c(t)$.

7. Prove the integration by parts formula for Riemann-Stieltjes integrals.
8. Show that $X_t = \int_0^t (2t-s)dW_s$ and $Y_t = \int_0^t (3t-4s)dW_s$ are Gaussian processes with mean 0 and variance $\frac{7}{3}t^3$.
9. Let W_t be a Brownian motion wrt. the filtration $\mathcal{F}_t$.

   a. Using the 'integration-by-parts' rule show that $\int_0^t g'(s)W_s ds = \int_0^t (g(t)-g(s))dW_s$.
   b. Determine the distribution, expected value and variance of the integral.
   c. Show that if $n = 0, 1, 2, \ldots$ is a non-negative integer, then $\int_0^1 s^n W_s ds \sim \mathcal{N}(0, \frac{2}{(2n+3)(n+2)})$.

10. Specify the integral type and the distribution of the integral $I_t = \int_0^t e^{2s} dW_s$. Compute the expected value and the variance of the integral.

11. Specify the distribution of the integral $I_t = \int_0^t \sqrt{s} e^s dW_s$. Compute the expected value and the variance.

8.2 Riemann-Stieltjes Integral Problems - Solutions

1. Solve the integral $\int_0^t s dW(s)$.

$$\int_0^t s dW(s) = \Big[s \cdot W(s)\Big]_0^t - \int_0^t W(s) \cdot ds = t \cdot W(t) - \int_0^t W(s) 1 \cdot ds = t \cdot W(t) - \int_0^t W(s) ds$$

2. Solve the integral $\int_0^t dW(s)$.

$$\int_0^t dW(s) = \int_0^t 1 \cdot dW(s) = W(t) - W(0) = W(t)$$

3. Specify distribution of $\int_0^1 s dW(s)$, and calculate its expectation and variance.

$$\mathbb{E}\Big[\int_0^1 s dW(s)\Big] = 0$$
$$Var\Big(\int_0^1 s dW(s)\Big) = \int_0^1 s^2 ds = \frac{s^3}{3}\Big|_0^1 = \frac{1}{3}$$
$$\int_0^1 s dW(s) \sim \mathcal{N}(0, 1/3)$$

4. Specify distribution of $\int_0^1 (1-s) dW(s)$, and calculate its expectation and variance.

$$\mathbb{E}\Big[\int_0^1 (1-s) dW(s)\Big] = 0$$
$$Var\Big(\int_0^1 (1-s) dW(s)\Big) = \int_0^1 (1-s)^2 ds = \int_0^1 1 - 2s + s^2 ds$$
$$= s - s^2 + \frac{s^3}{3}\Big|_0^1 = 1 - 1 + \frac{1}{3} = \frac{1}{3}$$
$$\int_0^1 s dW(s) \sim \mathcal{N}(0, 1/3)$$

5. Specify distribution of $\int_0^1 dW(s)$, and calculate its expectation and variance.

$$\mathbb{E}\Big[\int_0^t dW(s)\Big] = 0$$
$$Var\Big(\int_0^t dW(s)\Big) = \int_0^t 1 ds = s|_0^t = t$$
$$\int_0^t dW(s) \sim \mathcal{N}(0, t)$$

6. A standard Ornstein-Uhlenbeck process is given by

$$X_c(t) = e^{ct} \int_0^t e^{-cs} dW(s) \qquad t \geq 0 \text{ and } X_c(0) = 0$$

   a. Use integration by parts to find an equivalent expression.

$$X_c(t) = e^{ct} \int_0^t e^{-cs} dW(s) = e^{ct}\Big(\Big[e^{-cs} W(s)\Big]_0^t - \int_0^t W(s)(-ce^{-cs}) ds\Big)$$
$$= W(t) + c \int_0^t e^{c(t-s)} W(s) ds$$

b. Find the distribution of $X_c(t)$.

$$\mathbb{E}[X_c(t)] = \mathbb{E}\Big[e^{ct}\int_0^t e^{-cs}dW(s)\Big] = 0$$

$$\begin{aligned}
Var(X_c(t)) &= Var\Big(e^{ct}\int_0^t e^{-cs}dW(s)\Big) = e^{2ct}Var\Big(\int_0^t e^{-cs}dW(s)\Big)\\
&= e^{2ct}\int_0^t e^{-2cs}ds = e^{2ct}\Big[-\frac{e^{-2cs}}{2c}\Big]_0^t = -\frac{e^{2ct}}{2c}(e^{-2ct}-1)\\
&= \frac{e^{2ct}-1}{2c}
\end{aligned}$$

$$X_c(t) \sim \mathcal{N}\Big(0, \frac{e^{2ct}-1}{2c}\Big)$$

7. Prove the integration by parts formula for Riemann-Stieltjes integrals.

$$f(t,X) := X\cdot g(t) \Rightarrow f_t(t,X) = X\cdot g'(t);\ f_x(t,X) = g(t);\ f_{xx} = 0$$

Setting $X = W_t$

$$\begin{aligned}
d(g(t)W_t) &= W_t\cdot g'(t)dt + g(t)dW_t + \frac{1}{2}\cdot 0dt\\
&= W_t\cdot g'(t)dt + g(t)dW_t
\end{aligned}$$

Taking integral from 0 to t

$$g(t)W(t) - g(0)W_0 = \int_0^t g'(s)W_s ds + \int_0^t g(s)dW_s$$

$$\int_0^t g(s)dW_s = g(t)W(t) - \int_0^t g'(s)W_s ds$$

8. Show that $X_t = \int_0^t(2t-s)dW_s$ and $Y_t = \int_0^t(3t-4s)dW_s$ are Gaussian processes with mean 0 and variance $\frac{7}{3}t^3$.

Since both X_t and Y_t are Riemann-Stieltjes integrals, they are normally distributed with mean zero by definition. This is easy to see as the randomness enters through the integrator, a Brownian motion which is itself normally distributed with expectation of increments = 0.

$$\begin{aligned}
Var(X_t) &= Var\Big(\int_0^t (2t-s)dW_s\Big) = \mathbb{E}\Big[\int_0^t (2t-s)^2ds\Big] = \int_0^t 4t^2 - 4ts + s^2ds\\
&= 4t^2\int_0^t ds - 4t\int_0^t sds + \int_0^t s^2ds = 4t^3 - 2t^3 + \frac{t^3}{3} = \frac{7}{3}t^3\\
Var(Y_t) &= Var\Big(\int_0^t (3t-4s)dW_s\Big) = \mathbb{E}\Big[\int_0^t (3t-4s)^2ds\Big] = \int_0^t 9t^2 - 24ts + 16s^2ds\\
&= 9t^2\int_0^t ds - 24t\int_0^t sds + 16\int_0^t s^2ds = 9t^3 - 12t^3 + \frac{16t^3}{3} = \frac{7}{3}t^3
\end{aligned}$$

9. Let W_t be a Brownian motion wrt. the filtration $\mathscr{F}_t$.

a. Using the 'integration-by-parts' rule show that $\int_0^t g'(s)W_sds = \int_0^t(g(t)-g(s))dW_s$.

$$\int_0^t g(s)dW_s = g(t)W(t) - \int_0^t g'(s)W_sds$$

$$\int_0^t g'(s)W_sds = g(t)W(t) - \int_0^t g(s)dW_s$$

Since $W(t) = \int_0^t dW_s$ we can write.

$$\int_0^t g'(s)W_sds = \int_0^t g(t)dW_s - \int_0^t g(s)dW_s = \int_0^t (g(t)-g(s))dW_s$$

b. Determine the distribution, expected value and variance of the integral.

The integral is a Riemann-Stieltjes integral. Consequently we know that it is normally distributed with mean zero and variance according to Itô isometry;

$$\mathcal{N}\Big(0, Var\Big(\int_0^t (g(t)-g(s))dW_s\Big)\Big) = \mathcal{N}\Big(0, \int_0^t (g(t)-g(s))^2 ds\Big)$$

c. Show that if $n = 0,1,2,...$ is a non-negative integer, then $\int_0^1 s^n W_s ds \sim \mathcal{N}(0, \frac{2}{(2n+3)(n+2)})$.

Here we can use the formula from part a. Let $g'(s) = s^n$ Then $g(s) = \frac{s^{n+1}}{n+1}$ and $g(1) = \frac{1}{n+1}$.

$$\int_0^1 s^n W_s ds = \int_0^1 \Big(\frac{1}{n+1} - \frac{s^{n+1}}{n+1}\Big) dW_s = Y$$

Since Y is a RS-integral, we know that it is normally distributed with mean 0.

$$\begin{aligned}
Var(Y) &= \mathbb{E}\Big[\int_0^1 \Big(\frac{1}{n+1} - \frac{s^{n+1}}{n+1}\Big)^2 ds\Big] = \int_0^1 \frac{1}{(n+1)^2} - \frac{2s^{n+1}}{(n+1)^2} + \frac{s^{2n+2}}{(n+1)^2} ds \\
&= \frac{1}{(n+1)^2}\int_0^1 1 - 2s^{n+1} + s^{2n+2} ds = \frac{1}{(n+1)^2}\Big[s - \frac{2s^{n+2}}{n+2} + \frac{s^{2n+3}}{2n+3}\Big]_0^1 \\
&= \frac{1}{(n+1)^2}\Big[1 - \frac{2}{n+2} + \frac{1}{2n+3}\Big] = \frac{1}{(n+1)^2}\Big[\frac{(2n+3)(n+2) - 2(2n+3) + n+2}{(2n+3)(n+2)}\Big] \\
&= \frac{1}{(n+1)^2}\Big[\frac{2n^2 + 7n + 6 - 4n - 6 + n + 2}{(2n+3)(n+2)}\Big] = \frac{1}{(n+1)^2}\Big[\frac{(2n^2+4n+2)}{(2n+3)(n+2)}\Big] \\
&= \frac{1}{(n+1)^2}\frac{2(n+1)^2}{(2n+3)(n+2)} = \frac{2}{(2n+3)(n+2)}
\end{aligned}$$

10. Specify the integral type and the distribution of the integral $I_t = \int_0^t e^{2s} dW_s$. Compute the expected value and the variance of the integral.

Since we have a deterministic integrand with a Brownian motion integrator, this is a Riemann-Stieltjes integral. The Riemann-Stieltjes integral is normally distributed.

$$\begin{aligned}
\mathbb{E}\Big[\int_0^t e^{2s} dW_s\Big] &= 0 \\
Var\Big(\int_0^t e^{2s} dW_s\Big) &= \mathbb{E}\Big[\int_0^t e^{4s} ds\Big] = \int_0^t e^{4s} ds = \Big[\frac{e^{4s}}{4}\Big]_0^t = \frac{e^{4t}-1}{4}
\end{aligned}$$

11. Specify the distribution of the integral $I_t = \int_0^t \sqrt{s} e^s dW_s$. Compute the expected value and the variance.

Since we have a deterministic integrand with a Brownian motion integrator, this is a Riemann-Stieltjes integral. The Riemann-Stieltjes integral is normally distributed.

$$\begin{aligned}
\mathbb{E}\Big[\int_0^t \sqrt{s} e^s dW_s\Big] &= 0 \\
Var\Big(\int_0^t \sqrt{s} e^s dW_s\Big) &= \mathbb{E}\Big[\int_0^t s e^{2s} ds\Big] = \int_0^t s e^{2s} ds
\end{aligned}$$

Let $u = s \Rightarrow u' = 1$ and $v' = e^{2s} \Rightarrow v = \frac{e^{2s}}{2}$. Then, using integration-by-parts we get

$$\Big[se^{2s}\Big]_0^t - \frac{1}{2}\int_0^t e^{2s} ds = te^{2t} - \frac{1}{2}\Big[\frac{e^{2s}}{2}\Big]_0^t = te^{2t} - \frac{1}{4}\Big(e^{2t}-1\Big) = \frac{4te^{2t}}{4} + \frac{1-e^{2t}}{4} = \frac{e^{2t}(2t-1)+1}{4}$$

Theorem 8.2: Itô Integral

An Itô integral is of the form

$$I_t(w) = \int_0^t X(s,W)dW(s)$$

Itô integral is the limit of the Itô sum

$$\lim_{n\to\infty}(IS_n) = \int_0^t X(s)dW(s)$$

Properties of Itô integral

i. Martingale property. Consider a stochastic process $\{X_t\}_{t\geq 0}$ such that

   1. X_t is adapted to the filtration $\mathscr{F}_t$ of the Brownian motion $\{W_t\}_{t\geq 0}$.
   2. $\int_0^T \mathbb{E}[X^2(t)]dt < \infty \; \forall \; T$

   Then $IS_n(w) = \sum_{i=1}^n X_{t_{i-1}}(w)(W_{t_i}(w) - W_{t_{i-1}}(w))$ is a martingale.

ii. Expectation of an Itô integral is $\mathbb{E}\big[\int_0^t X(s)dW(s)\big] = 0$.

iii. Variance of an Itô integral is $Var\big(\int_0^t X(s)dW(s)\big) = \mathbb{E}\big[\int_0^t X^2(s)ds\big] = \int_0^t \mathbb{E}\big[X^2(s)\big]ds$ by Itô isometry.

iv. The distribution of an Itô integral is generally not known.

v. Covariance of two Itô integrals for $s < t$,

$$\mathbb{E}\Big[\int_0^t f(W_u,u)dW_u|\mathscr{F}_s\Big] = \int_0^s f(W_u,u)dW_u + \int_s^t f(W_u,u)dW_u = \int_0^s f(W_u,u)dW_u$$

vi. The quadratic accumulated up to time t by an Itô integral

$$[I,I](t) = \int_0^t X^2(s)ds$$

   This can also be seen if we start from the differential form of the integral $dI(t) = X(t)dW(t)$. Then the quadratic variation becomes $dI(t)dI(t) = X^2(t)dt$.

Itô integral is the limit of the portfolio gain achieved by continuous trading.

8.3 Itô Integral Problems

1. Let W_t be a Brownian motion wrt. to the filtration $\mathscr{F}_t$.

   a. For positive integer n, prove that $Cov(W_t, \int_0^t u^n dW_u) = \frac{t^{n+1}}{n+1}$
   b. Compute the variance of $I_t = \int_0^t e^{W_s}dW_s$.

8.4 Itô Integral Problems - Solutions

1. Let W_t be a Brownian motion wrt. to the filtration $\mathscr{F}_t$.

   a. For positive integer n, prove that $Cov(W_t, \int_0^t u^n dW_u) = \frac{t^{n+1}}{n+1}$

$$Cov(W_t, \int_0^t u^n dW_u) = \mathbb{E}[W_t \cdot \int_0^t u^n dW_u] - \mathbb{E}[W_t] \cdot \mathbb{E}\Big[\int_0^t u^n dW_u\Big]$$
$$= \mathbb{E}\Big[\int_0^t dW_t \cdot \int_0^t u^n dW_u\Big] - 0 \cdot 0 = \int_0^t u^n du = \frac{t^{n+1}}{n+1}$$

   b. Compute the variance of $I_t = \int_0^t e^{W_s} dW_s$.

$$Var(I_t) = \mathbb{E}\Big[\int_0^t e^{2W_s} ds\Big] \text{ ,by Itô isometry}$$
$$= \int_0^t \mathbb{E}[e^{2W_s}] ds = \int_0^t e^{0+\frac{1}{2}4t} ds \text{ ,by expectation of a log-normal RV}$$
$$= \int_0^t e^{2t} ds = \Big[\frac{e^{2s}}{2}\Big]_0^t = \frac{e^{2t}}{2} - \frac{1}{2} = \frac{e^{2t}-1}{2}$$

9 Itô Calculus

Lemma 9.1: Itô-Doeblin's Lemma

Chain rule for stochastic functions.

$$f(W_t) - f(W_0) = \int_0^t f'(W_s)dW_s + \frac{1}{2}\int_0^t f''(W_s)ds$$

On differentiable form

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt$$

Lemma 9.2: Itô-Doeblin's Lemma with Time

$$f(T, W_T) - f(0, W_0) = \int_0^T f_t(t, W_t)dt + \int_0^T f_x(t, W_t)dW_t + \frac{1}{2}\int_0^t f_{xx}(t, W_t)dt$$

On differentiable form

$$df(t, W_t) = f_t(t, W_t)dt + f_x(t, W_t)dW_t + \frac{1}{2}f_{xx}(t, W_t)dt$$

Lemma 9.3: Itô-Doeblin's Lemma w.r.t. an Itô Process

If we have the Itô process

$$\int_0^t \Gamma(u)dX(u) = \int_0^t \Gamma(u)\alpha(u)du + \int_0^t \Gamma(u)\sigma(u)dW(u)$$

Then the dynamics of the process is given by

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))dX(t)dX(t)$$

Lemma 9.4: Itô-Doeblin's Lemma for Multidimensional Processes

A d-dimensional Brownian motion is a process $W(t) = \{W_1(t), \cdots, W_d(t)\}$ with the following properties

i. Each $W_i(t)$ is a one-dimensional Brownian motion.

ii. If $i \neq j$, then processes $W_i(t)$ and $W_j(t)$ are independent.

iii. Quadratic variation is given by

$$\begin{cases} [W_i, W_i](t) = t \\ [W_i, W_j](t) = 0 \text{ if } i \neq j \end{cases} \qquad \begin{cases} dW_i dW_i = dt \\ dW_i dW_j = 0 \text{ if } i \neq j \end{cases}$$

Let $X(t)$ and $Y(t)$ be two Itô processes with dynamics

$$dX(t) = \alpha_1(t)dt + \sigma_{1,1}(t)dW_1(t) + \sigma_{1,2}(t)dW_2(t)$$
$$dY(t) = \alpha_2(t)dt + \sigma_{2,1}(t)dW_1(t) + \sigma_{2,2}(t)dW_2(t)$$

Then we get

$$\begin{aligned}
dX(t)dX(t) &= (\sigma_{1,1}^2(t) + \sigma_{1,2}^2(t))dt \\
dX(t)dY(t) &= (\sigma_{1,1}^2(t) \cdot \sigma_{2,1}^2(t) + \sigma_{1,2}^2(t) + \sigma_{2,2}^2(t))dt \\
dY(t)dY(t) &= (\sigma_{2,1}^2(t) + \sigma_{2,2}^2(t))dt
\end{aligned}$$

Itô's lemma for the two-dimensional process becomes (with arguments suppressed)

$$df(t, X, Y) = f_t dt + f_x dX + f_y dY + \frac{1}{2} f_{xx} dX dX + f_{xy} dX dY + \frac{1}{2} f_{yy} dY dY$$

Lemma 9.5: Itô Product Rule

Let $X(t)$, $Y(t)$ be Itô processes. Then

$$d(X(t) \cdot Y(t)) = Y(t)dX(t) + X(t)dY(t) + dX(t)dY(t)$$

Proof:

$$\begin{aligned}
f(t, x, y) = xy &\Rightarrow f_t = 0;\ f_x = y;\ f_y = x;\ f_{xx} = 0;\ f_{xy} = 1 \\
df(t, X(t), Y(t)) &= 0 \cdot dt + Y(t)dX_t + X(t)dY_t + \frac{1}{2} \cdot 0 \cdot (dX_t)^2 + 1 \cdot dX_t dY_t + \frac{1}{2} \cdot 0 \cdot (dY_t)^2 \\
&= Y(t)dX_t + X(t)dY_t + dX_t dY_t
\end{aligned}$$

□

9.1 Itô Calculus Problems

1. Find the following differentials
   a. $d(e^{t+W_t^2})$.
   b. $d(tW_t^2)$.
   c. $d(\sqrt{W_t})$.
2. Given the dynamics $dX_t = \alpha X_t dt + \sigma X_t dW_t$ and the constant r, derive the following functions of X and t.
   a. $Y_t = e^{r(T-t)} X_t$.
   b. $Y_t = X_t^{-1}$.
   c. $Y_t = e^{X_t}$.
3. Find the increment $W_t^2 - W_0^2$.
4. Suppose $\{W_t\}_{t \geq 0}$ is a Brownian motion. Determine the stochastic differential equation satisfied by
   $X_t = e^{\mu t + \sigma W_t}$.
5. Given $dSt = \alpha S_t dt + \sigma S_t dW_t$ find an expression of $f(t, S) = \ln(S)$. Assume that μ and σ are constant.
6. Find the differential of a discounted stock price when we assume that the stock price is modelled by a Geometric Brownian motion, i.e. $dS_t = \alpha S_t dt + \sigma S_t dW_t$.

7. Find the dynamics of the process $Y(t, S_1, S_2) = \frac{S_1}{S_2}$ when

$$\begin{aligned}
dS_1 &= \alpha_1 S_1(t)dt + \sigma_1 S_1(t)dW_1(t) \\
dS_2 &= \alpha_2 S_2(t)dt + \sigma_2 S_2(t)[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)] \\
dW_1 dW_2 &= 0
\end{aligned}$$

8. Find the dynamics of the process $Y_t = \frac{S_t P_t}{B_t}$ given the dynamics

$$\begin{aligned}
dS_t &= \mu_S S_t dt + \sigma_S S_t dW_t^S \\
dP_t &= \mu_P P_t dt + \sigma_P P_t dW_t^P \\
dB_t &= rBtdt \\
dW_t^S dW_t^P &= \alpha dt
\end{aligned}$$

9. Let $S_t = S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}$ be a geometric Brownian motion.

   a. Derive the stochastic differential equation satisfied by S_t.
   b. Let p be a positive constant. Compute $d(S_t^p)$ (i.e. the differential to the power p).

10. Consider the dynamics of two Itô processes :

$$\begin{aligned}
dX_t &= \alpha_X X_t dt + \sigma_X X_t dW_t^X \\
dY_t &= \alpha_Y Y_t dt + \sigma_Y Y_t dW_t^Y
\end{aligned}$$

Where Brownian motions are W_t^X, W_t^Y are correlated with $[W_t^X, W_t^X] = \rho_{XY} t$. Derive the stochastic differential equations for the following processes:

   a. $U_t = X_t Y_t$.
   b. $V_t = X_t / Y_t$.

9.2 Itô Calculus Problems - Solution

1. Find the following differentials

   a. $d(e^{t+W_t^2})$.

$$\begin{aligned} f(t,x) &= e^{t+x^2} \Rightarrow f_t = e^{t+x^2}; f_x = 2xe^{t+x^2}; f_{xx} = 2e^{t+x^2} + 2xe^{t+x^2}2x = 2e^{t+x^2} + 4x^2e^{t+x^2} \\ d(e^{t+W_t^2}) &= e^{t+W_t^2}dt + 2W_te^{t+W_t^2}dW_t + \frac{1}{2}(2e^{t+W_t^2} + 4W_t^2e^{t+W_t^2})dt \\ &= 2e^{t+W_t^2}(1+2W_t^2)dt + 2W_te^{t+W_t^2}dW_t \end{aligned}$$

   b. $d(tW_t^2)$.

$$\begin{aligned} f(t,x) &= t\cdot x^2 \Rightarrow f_t = x^2; f_x = 2tx; f_{xx} = 2t \\ d(tW_t^2) &= W_t^2dt + 2tW_tdW_t + \frac{1}{2}2tdt \\ &= (t+W_t^2)dt + 2tW_tdW_t \end{aligned}$$

   c. $d(\sqrt{W_t})$.

$$\begin{aligned} f(t,x) &= \sqrt{x} \Rightarrow f_t = 0; f_x = \frac{1}{2\sqrt{x}}; f_{xx} = -\frac{1}{4x\sqrt{x}} \\ d(\sqrt{W_t}) &= 0\cdot dt + \frac{1}{2\sqrt{W_t}}dW_t + \frac{1}{2}\Big(-\frac{1}{4}\frac{1}{W_t\sqrt{W_t}}\Big)dt \\ &= \frac{1}{2\sqrt{W_t}}dW_t - \frac{1}{8W_t\sqrt{W_t}}dt \end{aligned}$$

2. Given the dynamics $dX_t = \alpha X_t dt + \sigma X_t dW_t$ and the constant r, derive the following functions of X and t.

   a. $Y_t = e^{r(T-t)}X_t$.

$$\begin{aligned} f(t,x) &= e^{r(T-t)}x \Rightarrow f_t = -re^{r(T-t)}x; f_x = e^{r(T-t)}; f_{xx} = 0 \\ d(Y_t) &= -re^{r(T-t)}X_tdt + e^{r(T-t)}dX_t + \frac{1}{2}\cdot 0(dX_t)^2 \\ &= -re^{r(T-t)}X_tdt + e^{r(T-t)}(\alpha X_tdt + \sigma X_tdW_t) \\ &= e^{r(T-t)}X_t(-r+\alpha)dt + e^{r(T-t)}\sigma X_tdW_t = Y_t(\alpha - r)dt + \sigma Y_tdW_t \end{aligned}$$

   b. $Y_t = X_t^{-1}$.

$$\begin{aligned} f(t,x) &= x^{-1} \Rightarrow f_t = 0; f_x = -x^{-2}; f_{xx} = 2x^{-3} \\ d(Y_t) &= 0dt - \frac{1}{X_t^2}dX_t + \frac{1}{2}\frac{2}{X_t^3}(dX_t^2) \\ &= -\frac{1}{X_t^2}(\alpha X_tdt + \sigma X_tdW_t) + \frac{1}{X_t^3}(\sigma^2X_t^2dt) \\ &= \frac{1}{X_t}(\sigma^2 - \alpha)dt - \sigma\frac{1}{X_t}dW_t = Y_t(\sigma^2-\alpha)dt - \sigma Y_tdW_t \end{aligned}$$

   c. $Y_t = e^{X_t}$.

$$\begin{aligned} f(t,x) &= e^x \Rightarrow f_t = 0; f_x = e^x; f_{xx} = e^x \\ d(Y_t) &= 0dt + e^{X_t}dX_t + \frac{1}{2}e^{X_t}(dX_t)^2 \\ &= e^{X_t}(\alpha X_tdt + \sigma X_tdW_t) + \frac{1}{2}e^{X_t}\sigma^2X_t^2dt \\ &= e^{X_t}X_t(\alpha + \frac{1}{2}\sigma^2X_t)dt + \sigma X_te^{X_t}dW_t \end{aligned}$$

3. Find the increment $W_t^2 - W_0^2$.

$$f(x) = x^2 \Rightarrow f'(x) = 2x; f''(x) = 2$$
$$W_t^2 - W_0^2 = \int_0^t 2W_s dW_s + \frac{1}{2}\int_0^t 2ds = 2\int_0^t W_s dW_s + \int_0^t ds$$
$$\int_0^t W_s dW_s \Rightarrow f'(W_s) = W_s \Rightarrow f(W_s) = \frac{W_s^2}{2} \text{ and } f''(W_s) = 1$$
$$f(W_t) - f(W_0) = \int_0^t W_s dW_s + \frac{1}{2}\int_0^t 1ds$$
$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{W_0^2}{2} - \frac{t}{2} = \frac{W_t^2 - t}{2}$$
$$W_t^2 - W_0^2 = 2\cdot\frac{W_t^2 - t}{2} + t = W_t^2$$

4. Suppose $\{W_t\}_{t\geq 0}$ is a Brownian motion. Determine the stochastic differential equation satisfied by
$X_t = e^{\mu t + \sigma W_t}$.

$$d(t, W_t) = \mu e^{\mu t+\sigma W_t} dt + \sigma e^{\mu t+\sigma W_t} dW_t + \frac{1}{2}\sigma^2 e^{\mu t+\sigma W_t} dt$$
$$= \mu X_t dt + \sigma X_t dWt + \frac{1}{2}\sigma^2 X_t dt = (\mu + \frac{1}{2}\sigma^2) X_t dt + \sigma X_t dWt$$

5. Given $dSt = \alpha S_t dt + \sigma S_t dW_t$ find an expression of $f(t,S) = \ln(S)$. Assume that μ and σ are constant.

$$f_t(t,S) = 0,\ f_S(t,S) = \frac{1}{S},\ f_{SS}(t,S) = -\frac{1}{S^2}\ (dS_t)^2 = \sigma^2 S_t^2 dt$$
$$d(\ln(S_t)) = 0dt + \frac{1}{S_t} dS_t + \frac{1}{2}\cdot\Big(-\frac{1}{S_t^2}(dSt)^2\Big)$$
$$= \alpha dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt = (\alpha - \frac{1}{2}\sigma^2) dt + \sigma dW_t$$

If we integrate this expression from 0 to t, we get

$$\ln S_t = \ln S_0 + (\alpha - \frac{1}{2}\sigma^2)t + \sigma(W_t - W_0)$$
$$S_t = S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}$$
$$\ln S_t \sim \mathcal{N}(\ln S_0 + (\alpha - \frac{1}{2}\sigma^2)t, \sigma^2 t)$$
$$S_t \sim \mathcal{LN}(\ln S_0 + (\alpha - \frac{1}{2}\sigma^2)t, \sigma^2 t)$$

6. Find the differential of a discounted stock price when we assume that the stock price is modelled by a Geometric Brownian motion, i.e. $dS_t = \alpha S_t dt + \sigma S_t dW_t$.

$$f(t,x) = e^{-rt}x \Rightarrow f_t = -re^{-rt}x;\ f_x = e^{-rt};\ f_{xx} = 0$$

Applying Itô's lemma we get

$$df(t,S_t) = -re^{-rt} S_t dt + e^{-rt} dS_t + 0\cdot(dS_t)^2$$
$$= -re^{-rt} S_t dt + e^{-rt}(\alpha S_t dt + \sigma S_t dW_t)$$
$$= e^{-rt} S_t(-r + \alpha) dt + e^{-rt}\sigma S_t dW_t$$

7. Find the dynamics of the process $Y(t,S_1,S_2)=\frac{S_1}{S_2}$ when

$$
\begin{aligned}
dS_1 &= \alpha_1 S_1(t)dt+\sigma_1 S_1(t)dW_1(t)\\
dS_2 &= \alpha_2 S_2(t)dt+\sigma_2 S_2(t)[\rho dW_1(t)+\sqrt{1-\rho^2}dW_2(t)]\\
dW_1 dW_2 &= 0
\end{aligned}
$$

We get

$$
\begin{aligned}
\frac{\partial Y}{\partial t} &= 0;\ \frac{\partial Y}{\partial S_1}=\frac{1}{S_2};\ \frac{\partial Y}{\partial S_2}=-\frac{S_1}{S_2^2}\\
\frac{\partial^2 Y}{\partial S_1^2} &= 0;\ \frac{\partial^2 Y}{\partial S_2^2}=\frac{2S_1}{S_2^3};\ \frac{\partial^2 Y}{\partial S_1 S_2}=-\frac{1}{S_2^2}\\
dS_1 dS_1 &= \sigma_1^2 S_1^2(t)dt\\
dS_1 dS_2 &= \rho\sigma_1\sigma_2 S_1(t)S_2(t)dt\\
dS_2 dS_2 &= \sigma_2^2 S_2^2(t)dt
\end{aligned}
$$

Applying Itô's Lemma for multidimensional processes with $d=2$

$$
\begin{aligned}
d(Y(t,S_1,S_2)) &= 0dt+\frac{1}{S_2}dS_1+\left(-\frac{S_1}{S_2^2}\right)dS_2+\frac{1}{2}\cdot 0(dS_1)^2+\left(-\frac{1}{S_2^2}\right)dS_1dS_2+\frac{1}{2}\frac{2S_1}{S_2^3}(dS_2)^2\\
&= \frac{1}{S_2}(\alpha_1 S_1(t)dt+\sigma_1 S_1(t)dW_1(t))-\frac{S_1}{S_2^2}(\alpha_2 S_2(t)dt+\sigma_2 S_2(t)[\rho dW_1(t)+\sqrt{1-\rho^2}dW_2(t)])\\
&-\frac{1}{S_2^2}(\rho\sigma_1\sigma_2 S_1(t)S_2(t)dt)+\frac{S_1}{S_2^3}(\sigma_2^2 S_2^2(t)dt)\\
&= Y\alpha_1 dt+Y\sigma_1 dW_1(t)-Y\alpha_2 dt-Y\sigma_2[\rho dW_1(t)+\sqrt{1-\rho^2}dW_2(t)]-Y\rho\sigma_1\sigma_2 dt+Y\sigma_2^2 dt\\
&= Y(\alpha_1-\alpha_2+\sigma_2^2-\rho\sigma_1\sigma_2)dt+Y(\sigma_1-\rho\sigma_2)dW_1(t)-Y\sigma_2\sqrt{1-\rho^2}dW_2(t)\\
&= Y[(\alpha_1-\alpha_2+\sigma_2^2-\rho\sigma_1\sigma_2)dt+(\sigma_1-\rho\sigma_2)dW_1(t)-\sigma_2\sqrt{1-\rho^2}dW_2(t)]
\end{aligned}
$$

8. Find the dynamics of the process $Y_t=\frac{S_tP_t}{B_t}$ given the dynamics

$$
\begin{aligned}
dS_t &= \mu_S S_t dt+\sigma_S S_t dW_t^S\\
dP_t &= \mu_P P_t dt+\sigma_P P_t dW_t^P\\
dB_t &= rBtdt\\
dW_t^S dW_t^P &= \alpha dt
\end{aligned}
$$

We get

$$
\begin{aligned}
\frac{\partial Y}{\partial t} &= 0;\ \frac{\partial Y}{\partial S_t}=\frac{P_t}{B_t};\ \frac{\partial Y}{\partial P_t}=\frac{S_t}{B_t};\ \frac{\partial Y}{\partial B_t}=-\frac{S_tP_t}{B_t^2}\\
\frac{\partial^2 Y}{\partial S_t^2} &= 0;\ \frac{\partial^2 Y}{\partial P_t^2}=0;\ \frac{\partial^2 Y}{\partial B_t^2}=\frac{2S_tP_t}{B_t^3};\ \frac{\partial^2 Y}{\partial S_tP_t}=\frac{1}{B_t}\\
\frac{\partial^2 Y}{\partial S_tB_t} &= -\frac{P_t}{B_t^2};\ \frac{\partial^2 Y}{\partial B_tP_t}=-\frac{S_t}{B_t^2}\\
dS_tdS_t &= \sigma_S^2 S_t^2 dt\\
dP_tdP_t &= \sigma_P^2 P_t^2 dt\\
dB_tdB_t &= 0\\
dS_tdP_t &= \alpha\sigma_S\sigma_P S_tP_t dt
\end{aligned}
$$

Applying Itô's Lemma for multidimensional processes with $d = 3$. We directly skip terms that involve 0.

$$\begin{aligned}
d(Y_t) &= \frac{P_t}{B_t}(\mu_S S_t dt + \sigma_S S_t dW_t^S) + \frac{S_t}{B_t}(\mu_P P_t dt + \sigma_P P_t dW_t^P) - \frac{S_t P_t}{B_t^2}(r B_t dt) + \frac{1}{B_t}(\alpha\sigma_S\sigma_P S_t P_t dt) \\
&= Y_t(\mu_S + \mu_p - r + \alpha\sigma_S\sigma_P)dt + Y_t(\sigma_S dW_t^S + \sigma_P dW_t^P)
\end{aligned}$$

9. Let $S_t = S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}$ be a geometric Brownian motion.

   a. Derive the stochastic differential equation satisfied by S_t.

$$\begin{aligned}
dS_t &= (\alpha - \frac{1}{2}\sigma^2)S_t dt + \sigma S_t dW_t + \frac{1}{2}\sigma^2 S_t dt \\
&= \alpha S_t dt + \sigma S_t dW_t
\end{aligned}$$

   b. Let p be a positive constant. Compute $d(S_t^p)$ (i.e. the differential to the power p).

$$\begin{aligned}
& f(t,x) = x^p \Rightarrow f_t = 0;\ f_x = p x^{p-1};\ f_{xx} = p(p-1)x^{p-2} \\
d(S_t^p) &= 0dt + pS_t^{p-1} dS_t + \frac{1}{2}p(p-1)S_t^{p-2}(dS_t)^2 \\
&= pS_t^{p-1}(\alpha S_t dt + \sigma S_t dW_t) + \frac{1}{2}p(p-1)S_t^{p-2}(\sigma^2 S_t^2 dt) \\
&= pS_t^p(\alpha + \frac{1}{2}(p-1)\sigma^2)dt + pS_t^p \sigma dW_t
\end{aligned}$$

10. Consider the dynamics of two Itô processes :

$$\begin{aligned}
dX_t &= \alpha_X X_t dt + \sigma_X X_t dW_t^X \\
dY_t &= \alpha_Y Y_t dt + \sigma_Y Y_t dW_t^Y
\end{aligned}$$

Where Brownian motions are W_t^X, W_t^Y are correlated with $[W_t^X, W_t^X] = \rho_{XY} t$. Derive the stochastic differential equations for the following processes:

   a. $U_t = X_t Y_t$.

$$\begin{aligned}
dX_t dX_t &= \sigma_X^2 X_t^2 dt \\
dX_t dY_t &= \rho_{XY}\sigma_X\sigma_Y X_t Y_t dt \\
dY_t dY_t &= \sigma_Y^2 Y_t^2 dt
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial U_t}{\partial t} = 0;\ \frac{\partial U_t}{\partial X_t} = Y_t;\ \frac{\partial U_t}{\partial Y_t} = X_t \\
&\frac{\partial^2 U_t}{\partial X_t^2} = 0;\ \frac{\partial^2 U_t}{\partial Y_t^2} = 0;\ \frac{\partial^2 U_t}{\partial X_t Y_t} = 1
\end{aligned}$$

$$\begin{aligned}
d(U_t) &= 0dt + Y_t dX_t + X_t dY_t + \frac{1}{2}0(dX_t)^2 + 1dX_t dY_t + \frac{1}{2}0(dY_t)^2 \\
&= Y_t(\alpha_X X_t dt + \sigma_X X_t dW_t^X) + X_t(\alpha_Y Y_t dt + \sigma_Y Y_t dW_t^Y) + \rho_{XY}\sigma_X\sigma_Y X_t Y_t dt \\
&= U_t(\alpha_X + \alpha_Y + \rho_{XY}\sigma_X\sigma_Y)dt + U_t(\sigma_X dW_t^X + \sigma_Y dW_t^Y)
\end{aligned}$$

b. $V_t = X_t / Y_t$.

$$\frac{\partial V_t}{\partial t} = 0;\ \frac{\partial V_t}{\partial X_t} = \frac{1}{Y_t};\ \frac{\partial V_t}{\partial Y_t} = -\frac{X_t}{Y_t^2}$$
$$\frac{\partial^2 V_t}{\partial X_t^2} = 0;\ \frac{\partial^2 V_t}{\partial Y_t^2} = \frac{2X_t}{Y_t^3};\ \frac{\partial^2 V_t}{\partial X_t Y_t} = -\frac{1}{Y_t^2}$$

$$\begin{aligned}
d(V_t) &= \frac{1}{Y_t} dX_t + -\frac{X_t}{Y_t^2} dY_t - \frac{1}{Y_t^2} dX_t dY_t + \frac{1}{2}\frac{2X_t}{Y_t^3}(dY_t)^2 \\
&= \frac{1}{Y_t}(\alpha_X X_t dt + \sigma_X X_t dW_t^X) - \frac{X_t}{Y_t^2}(\alpha_Y Y_t dt + \sigma_Y Y_t dW_t^Y) \\
&- \frac{1}{Y_t^2}\rho_{XY}\sigma_X\sigma_Y X_t Y_t dt + \frac{1}{2}\frac{2X_t}{Y_t^3}\sigma_Y^2 Y_t^2 dt \\
&= V_t(\alpha_X dt + \sigma_X dW_t^X) - V_t(\alpha_Y dt + \sigma_Y dW_t^Y) \\
&- V_t \rho_{XY}\sigma_X\sigma_Y dt + V_t \sigma_Y^2 dt \\
&= V_t(\alpha_X - \alpha_Y - \rho_{XY}\sigma_X\sigma_Y + \sigma_Y^2)dt + V_t(\sigma_X dW_t^X - \sigma_Y dW_t^Y)
\end{aligned}$$

10 Major Models of SDEs

Definition 10.1: Aritmetic Brownian Motion

An arithmetic Brownian motion is a model on the form

$$dX_t = \mu dt + \sigma dW_t$$

where μ, σ are constant.

On integral from

$$\begin{aligned} X_t &= X_0 + \int_0^t dX_u = X_0 + \int_0^t \mu du + \sigma \int_0^t dW_u \\ &= X_0 + \mu_t + \sigma W_t \end{aligned}$$

Distribution of X_t is

$$X_t \sim \mathcal{N}\Big(X_0 + \mu_t, \sigma^2 t\Big)$$

Arithmetic Brownian motion can be used to model returns or log-prices. It is ill-suited to model stock prices as the process may become negative.

Definition 10.2: Geometric Brownian Motion

A Geometric Brownian motion is a model on the form

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

where μ is the drift coefficient and σ is the volatility coefficient.

On integral from

$$X_t = X_0 + \int_0^t \mu X_s ds + \int_0^t \sigma X_s dW_s$$

This SDE can be solved by applying Itô's Lemma on the function $f(t, X_t) = ln(X_t)$.

$$\begin{aligned} d(ln(X_t)) &= \frac{1}{X_t}(\mu X_t dt + \sigma X_t dW_t) - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 X_t^2 dt \\ &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t \end{aligned}$$

Integrating from 0 to t.

$$\begin{aligned} ln(X_t) - ln(X_0) &= (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t \\ ln(X_t) &= lnln(X_0) + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t \end{aligned}$$

Taking exponential of both sides

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

Distribution of X_t is

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[X_0 e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W_t}] = X_0 e^{(\mu-\frac{1}{2}\sigma^2)t}\mathbb{E}[e^{\sigma W_t}] \\
&= X_0 e^{(\mu-\frac{1}{2}\sigma^2)t} e^{0+\frac{\sigma^2 t}{2}} = X_0 e^{\mu t} \\
Var(X_t) &= \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 \\
\mathbb{E}[X_t^2] &= \mathbb{E}[X_0^2 e^{2(\mu-\frac{1}{2}\sigma^2)t+2\sigma W_t}] = X_0^2 e^{2(\mu-\frac{1}{2}}\mathbb{E}[e^{2\sigma W_t}] \\
&= X_0^2 e^{2(\mu-\frac{1}{2}\sigma^2)t} e^{0+\frac{4\sigma^2 t}{2}} = X_0^2 e^{(2\mu+\sigma^2)t} \\
Var(X_t) &= X_0^2 e^{(2\mu+\sigma^2)t} - (X_0 e^{\mu t})^2 = X_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1) \\
X_t &\sim \mathscr{LN}\Big(X_0 e^{\mu t}, X_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)\Big)
\end{aligned}
$$

Geometric Brownian motion can be used to model stock prices. It has the desirable properties that the process 1) cannot become negative and 2) the terms depend on the size of X_t. For interpretability, it is nice that the process can be decomposed into a predictable part (i.e. the expected return), and an unexpected, noisy part (the volatility).

We can generalize the model to allow for time-dependent drift and volatility, $\mu = \mu(t)$, and $\sigma = \sigma(t)$. We write the process now as

$$
\begin{aligned}
dS_t &= \mu(t)S_t dt + \sigma(t)S_t dW_t \\
S_t &= S_0 e^{\int_0^t(\mu(s)-\frac{1}{2}\sigma^(s))ds+\int_0^t \sigma(s)dW_s}
\end{aligned}
$$

Its expectation and variance becomes

$$
\begin{aligned}
\mathbb{E}[S_t] &= S_0 exp\Big(\int_0^t \mu(s)ds\Big) \\
Var(S_t) &= S_0^2 exp\Big(2\int_0^t \mu(s)ds\Big)\Big[exp\Big(\int 0^t \sigma^2(s)ds\Big) - 1\Big]
\end{aligned}
$$

Definition 10.3: Ornstein-Uhlenbeck Process

An Ornstein-Uhlenbeck process is an example of a linear stochastic differential equation and is on the form

$$
dX_t = \kappa(\theta - X_t)dt + \sigma dW_t
$$

where θ is the long-term mean and $\kappa > 0$ is the speed of mean reversion.

We can solve for X_t be applying Itô's lemma to the function $f(X_t, t) = e^{\kappa t} X_t$.

$$
\begin{aligned}
d(e^{\kappa t} X_t) &= \kappa e^{\kappa t} X_t dt + e^{\kappa t} dX_t \\
&= \kappa e^{\kappa t} X_t dt + e^{\kappa t}(\kappa(\theta - X_t)dt + \sigma dW_t) \\
&= \kappa e^{\kappa t} X_t dt + \kappa e^{\kappa t}\theta dt - \kappa e^{\kappa t} X_t dt + e^{\kappa t} dW_t \\
&= \kappa e^{\kappa t}\theta dt + \sigma e^{\kappa t} dW_t
\end{aligned}
$$

Integrating from 0 to t

$$
\begin{aligned}
e^{\kappa t} X_t &= X_0 e^{0\cdot t} + \kappa\theta \int_0^t e^{\kappa s} ds + \sigma \int_0^t e^{\kappa s} dW_s \\
&= X_0 + \kappa\theta \cdot \frac{e^{\kappa t} - 1}{\kappa} + \sigma \int_0^t e^{\kappa s} dW_s \\
X_t &= X_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{\kappa(s-t)} dW_s
\end{aligned}
$$

Expectation of X_t is

$$\mathbb{E}[X_t] = X_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t})$$

The expectation is the weighted average of the initial value and long-term mean. As $t \to \infty$ we move from X_0 towards θ. A higher κ means that we approach θ faster. Variance of X_t is

$$
\begin{aligned}
Var(X_t) &= Var(\sigma e^{-\kappa t} \int_0^t e^{\kappa s} dW_s) = \sigma^2 e^{-2\kappa t} \int_0^t e^{2\kappa s} ds \\
&= \sigma^2 e^{-2\kappa t} \cdot \frac{1}{2\kappa}(e^{2\kappa t} - 1) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})
\end{aligned}
$$

The Ornstein-Uhlenbeck process can be used to model interest rates (the mean-reverting behavior is desirable). However, since $X_t \sim \mathcal{N}$, the process can become negative and we can get unbounded large values of X_t.

Definition 10.4: Cox-Ingersoll-Ross Model

Assume that the spot rate follows the SDE

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t} dW_t, \ 2\kappa\theta \geq \sigma^2$$

where μ, σ are constant.

Features of this model:

i. Mean reversion
ii. Not possible for interest rates to go negative.
iii. Volatility depend on the level of the interest rate.

Integrating the dynamics from 0 to t we get

$$X_t = X_0 + \kappa\theta t - \kappa \int_0^t X_s ds + \sigma \int_0^t \sqrt{X_s} dW_s$$

11 Change of Measure

Definition 11.1: Change of Measure

Discrete case: Let Ω be a finite sample space and $\mathbb{P},\mathbb{Q}$ be two probability measures such that $\mathbb{P}(\omega) > 0, \mathbb{Q}(\omega) > 0 \ \forall\ \omega \in \Omega$. Then $Z(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}$, a random variable that changes the probability measure on Ω.

Continuous case: $\mathbb{Q}(\omega) = \mathbb{P}(\omega) \cdot Z(\omega)$. To change from $\mathbb{P}$ to $\mathbb{Q}$ we beed to reassign probabilities using the new construct $Z(\omega)$.

$Z(\omega) > 1$ revise probabilities upward.

$Z(\omega) < 1$ revise probabilities downward.

Theorem 11.1: Change of Measure

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let Z be a nonnegative random variable with $\mathbb{E}[Z] = 1$. For $A \in \mathscr{F}$ define

$$\mathbb{Q}(A) = \int_A Z(\omega)\, d\mathbb{P}(\omega)$$

then $\mathbb{Q}$ is a probability measure.

If X is a nonnegative random variable, then

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[XZ]$$

If $\mathbb{P}(Z > 0) = 1$ then we can write $\mathbb{E}^{\mathbb{P}}[X] = \mathbb{E}^{\mathbb{Q}}[X/Z]$. Z is called a **Radon-Nikodym** derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ and we use the notation

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$$

Properties of Radon-Nikodym derivative:

i. $\mathbb{P}(Z > 0) = 1$

ii. $\mathbb{E}[Z] = 1$

iii. Radon-Nikodym derivative process is a martingale,

$$Z_t = \mathbb{E}[Z|\mathscr{F}_t] \ 0 \le t \le T$$

This can be shown using the law of iterated expectations

$$\mathbb{E}[Z_t|\mathscr{F}_s] = \mathbb{E}[\mathbb{E}[Z|\mathscr{F}_t]|\mathscr{F}_s] = \mathbb{E}[Z|\mathscr{F}_s] = Z_s$$

Theorem 11.2: Girsanov Theorem

Let $W_t^{\mathbb{P}}, 0 \le t \le T$ be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{F}_t$ be a associated filtration. Let $\lambda_t, 0 \le t \le T$ be an adapted process to $\mathscr{F}_t$.

Define:

$$Z_t = exp(-\int_0^t \lambda_s dW_s - \frac{1}{2}\int_0^t \lambda^2 ds)$$

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t \lambda_s ds$$

And assume that

$$\mathbb{E}\Big[\int_0^t \lambda_s^2 Z_s^2 ds\Big] < \infty$$

Set $Z = Z_T$, then $\mathbb{E}[Z] = 1$ and under the probability measure $\mathbb{Q}$ the process $W_t^{\mathbb{Q}}, 0 \leq t \leq T$ is a Brownian motion.