

Quantitative Finance: Mathematical Foundations and Models

Liam O'Shaughnessy

December 2025

Preface

These notes were written to consolidate the mathematical foundations of modern quantitative finance, with an emphasis on arbitrage pricing, stochastic calculus, and practical modeling considerations. They are not intended as a textbook, but as a structured reference bridging theory and implementation. This is more theoretically and academically focused, which I wanted to make sure I had a thorough understanding of - I would like to continue developing practical approaches to effectuate these methods.

Contents

1	Introduction	2
2	Mathematical Foundations	3
2.1	Probability Theory	3
2.1.1	Axioms of Probability	3
2.1.2	Lebesgue Integration	5
2.1.3	Conditioning and Change of Measure	7
2.1.4	Inference	8
2.2	Stochastic Processes	9
2.2.1	Foundations of Stochastic Processes	9
2.2.2	Properties of Stochastic Processes	10
2.2.3	Markov Processes	11
2.2.4	Brownian Motion	12
2.3	Stochastic Calculus	14
2.3.1	Itô Calculus	14
2.3.2	SDE-PDE Connection	16
2.4	Jump Diffusion	20
3	Risk-Neutral Formulation	24
3.1	Risk-Neutral Pricing	24
3.1.1	One Asset	24
3.1.2	Multiple Assets	26
3.2	Change of Numéraire	27
4	Asset Pricing and Derivatives	30
4.1	Forwards and Futures	30
4.2	European Options	31
4.3	Exotic Options	33
4.3.1	Asian Options	33
4.3.2	Barrier and Lookback Options	34
4.4	American Options	35
4.5	Local and Stochastic Volatility	37
4.5.1	Dupire Model	37

4.5.2	Heston Model	39
4.5.3	SABR Model	41
5	Fixed-Income and Credit	43
5.1	Term-Structure Models	43
5.1.1	Short Rate Models	43
5.1.2	Forward Rate Models	45
5.1.3	Market Models	46
5.2	Credit Risk Modeling	48
5.2.1	Structural Credit Risk Models	48
5.2.2	Reduced-Form Credit Models	49
5.2.3	CDSs and CDOs	50
5.3	Counterparty Risk and XVA	54
6	Portfolio Management	58
6.1	Portfolio Theory	58
6.1.1	Mean-Variance Analysis	58
6.1.2	Factor Models and Evaluation	60
6.2	Portfolio Risk	61
7	Markets and Trading	64
7.1	Exchanges	64
7.2	Optimal Execution	65
8	Time Series Analysis	68
8.1	Time Series Foundations	68
8.1.1	Models and Tests	68
8.1.2	Estimation and Model Selection	71
8.2	Volatility Modeling	72
8.2.1	Observations and Models	73
8.2.2	Forecasting versus Pricing	76
8.3	Spectral Analysis and Cointegration	76
9	Numerical Methods	79
9.1	Common Methods in Finance	79
9.1.1	Monte Carlo	79
9.1.2	Finite Difference Methods	81
9.2	Optimization and Root Finding	84
9.2.1	Optimization Techniques	84
9.2.2	Root Finding and Programming	86
10	Machine Learning Methods	87
10.1	Common Methods in Finance	87
10.2	Validation and Backtesting	92
11	Model Risk, Calibration, and Practical Pitfalls	94
11.1	The Nature of Model Risk	94

11.2	Meanings of Measures	95
11.3	Calibration as an Inverse Problem	95
11.4	Regime Dependence	96
11.5	Incomplete Markets	96
11.6	Numerical Error	96
11.7	Backtesting and Validation	97
11.8	Machine Learning and Overfitting	97
11.9	Practical Philosophy of Modeling	97

Chapter 1

Introduction

This has been my attempt to touch on most topics relevant to quantitative finance theory. I start with the mathematics of modeling assets as stochastic processes, which culminates with the Kolmogorov equations/generator framework for Markov processes that gets used for pricing. I then formalize risk-neutral pricing and how the Girsanov's theorem establishes the criteria to change to a martingale measure. In the next chapter, I price flavors of options, and discuss models for fitting implied volatility. Following that, I model rates stochastically, and introduce credit and counterparty risk for complete pricing. I then address portfolio management and risk assessment, as well as market microstructure dynamics. I spend three chapters on statistical and computational techniques to solve equations detailed earlier and better model data, such as time series analysis, numerical methods, and machine learning. Finally, I include a discussion on "when things go wrong".

These notes are drawn from the following books and resources: *Stochastic Calculus for Finance II* by Shreve, *Arbitrage in Continuous Time* by Björk, *Probability Theory* by Sinai and Koralov, *Applied Stochastic Analysis* by Holmes-Cerfon, *Brownian Motion* by Schilling and Partzsch, *Kolmogorov Equations* by Ludvigsson, *All of Statistics* by Wasserman, *Convex Optimization* by Boyd and Vandenberghe, *Monte Carlo Methods in Financial Engineering* by Glasserman, *Advances in Financial Machine Learning* by de Prado, 8.S096 on MIT OCW, ECO362 from Princeton, Time Series Analysis from Baruch's MFE program, *Credit Risk Modeling* by Hurd from McMaster, and, of course, ChatGPT.

Chapter 2

Mathematical Foundations

Motivation. The first part of "quantitative finance" is "quantitative"! The purpose of this section is to develop the mathematical engine that will be applied to assets.

2.1 Probability Theory

Motivation. In order to model random events in the real world, the probabilistic framework must be rigorously constructed. As a result, the Kolmogorov axioms are used as a starting point for a measure-theoretic conception of probability, which will be essential for modeling stochastic processes.

2.1.1 Axioms of Probability

Definition. Let Ω be a non-empty set, the **sample space** of events, and \mathcal{F} a collection of subsets of Ω . \mathcal{F} is called a **σ -algebra** if the empty set is in \mathcal{F} , complements of sets in \mathcal{F} are in \mathcal{F} , and countably infinite unions of sets in \mathcal{F} are in \mathcal{F} . The entries of \mathcal{F} are called **measurable sets**, and (Ω, \mathcal{F}) is called a **measurable space**.

Definition. For A a collection of subsets of Ω , the **σ -algebra generated by A** , $\sigma(A)$, is the intersection of all σ -algebras containing A - that is, the minimal algebra containing A . The **Borel σ -algebra** on \mathbb{R} , $\mathcal{B}(\mathbb{R})$, is the σ -algebra generated by open subsets of \mathbb{R} .

Definition. Let Ω be a non-empty set (the sample space) and \mathcal{F} a σ -algebra on Ω . Define $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ such that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$ for countably infinite, disjoint sets $A_n \in \mathcal{F}$. Then \mathbb{P} is a **probability measure**, and $(\Omega, \mathcal{F}, \mathbb{P})$ is a **probability space**. "**Measure zero**" means that the event has probability zero, and "**almost surely**" means that the event has probability one.

Definition. Let (Ω, \mathcal{F}) be a measurable space. Define a function $g : \Omega \rightarrow \mathbb{R}$. f is **\mathcal{F} -measurable** if $\forall A \in \mathcal{B}(\mathbb{R}), g^{-1}(A) \in \mathcal{F}$. An \mathcal{F} -measurable on a probability space is called a **random variable**. A σ -algebra generated by a random variable X is the preimage of X of Borel sets

$$\sigma(X) = \{X^{-1}(B) | B \in \mathcal{B}(\mathbb{R})\}, \quad (2.1)$$

that is, the smallest σ -algebra that makes X measurable.

Remark. This is similar to continuity in topology, and is particularly nice here because the probability that the random variable takes some set of values is thus the measure of a set that is guaranteed to be in the initial σ -algebra.

Example (Discrete probability space). Consider the experiment of tossing a fair coin twice. The sample space is

$$\Omega = \{HH, HT, TH, TT\}, \quad (2.2)$$

where H and T denote heads and tails, respectively.

Let $\mathcal{F} = 2^\Omega$ be the power set of Ω , and define a probability measure \mathbb{P} by assigning equal mass to each outcome:

$$\mathbb{P}(\{\omega\}) = \frac{1}{4}, \quad \omega \in \Omega. \quad (2.3)$$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space satisfying the axioms of probability. For example,

$$\mathbb{P}(\{HH, HT\}) = \mathbb{P}(\{HH\}) + \mathbb{P}(\{HT\}) = \frac{1}{2}. \quad (2.4)$$

Define the random variable $X : \Omega \rightarrow \mathbb{R}$ by letting $X(\omega)$ be the number of heads observed. The induced σ -algebra $\sigma(X)$ partitions Ω into events corresponding to $X = 0, 1, 2$.

Definition. Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The **distribution measure of X** , μ_X , is a probability measure that assigns to each Borel subset $B \subset \mathbb{R}$ the mass

$$\mu_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\omega : X(\omega) \in B). \quad (2.5)$$

This encodes the equivalent information as the **cumulative distribution function**, $F(x) = \mathbb{P}(X \leq x)$.

Definition. The distribution measure μ_X on X , a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, can admit a **density function** if there exists an $f(x)$ such that

$$\mu_X[a, b] = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx. \quad (2.6)$$

Alternatively, if X takes finite set of numbers as its image, then the **probability mass function** defines $p_i = \mathbb{P}(X = x_i)$ for all output values.

Example. A **Gaussian random variable** has density $\mathcal{N}(\mu, \sigma)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (2.7)$$

This is applicable in almost everything.

An **exponential random variable** has density $\exp(\lambda)$:

$$f(t) = \lambda e^{-\lambda t}, t \geq 0. \quad (2.8)$$

This is applicable in modeling the probability of first arrival times. The exponential random variable is said to be *memoryless*. Consider $\tau \sim \exp(\lambda)$. Then

$$\mathbb{P}(\tau > t + s | \tau > s) = \frac{\mathbb{P}(\tau > t + s, \tau > s)}{\mathbb{P}(\tau > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}. \quad (2.9)$$

A **Poisson random variable** has probability mass $\text{Pois}(\lambda)$:

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}, k \in \mathbb{N}. \quad (2.10)$$

This is useful for counting the number of arrivals in a certain time window by setting $\lambda = rt$.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ sub-algebras. \mathcal{G}, \mathcal{H} are said to be **independent** if, for all $A \in \mathcal{G}, B \in \mathcal{H}$,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (2.11)$$

Two random variables are independent if the σ -algebras generated from them are independent, and a random variable and a σ -algebra are independent if the σ -algebra generated from the random variable is independent of the σ -algebra.

2.1.2 Lebesgue Integration

Motivation. We would like to integrate to find things like expectation, yet there is no obvious way to partition the domain (the sample space Ω) as the Riemann integral would require, in contrast to the simple real x-axis. Since the y-axis is indeed the real numbers, we can partition that instead, and measure how much of the domain is assigned to each y point.

Definition. Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the positive real interval $[a, b]$. Let Π be a set of sub-points y_1, y_2, \dots partitioning that interval. For each sub-interval $[y_{k-1}, y_k]$, define $A_k = \{\omega \in \Omega :$

$y_{k-1} \leq X(\omega) \leq y_k$. Then as the supremum of sub-point distances goes to zero, the following sum converges to the **Lebesgue integral**:

$$\sum_k y_k \mathbb{P}(A_k) \rightarrow \int_{\Omega} X(\omega) d\mathbb{P}(\omega). \quad (2.12)$$

If $\mathbb{P}(X = \infty) > 0$, $\int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \infty$.

Definition. Previously, the range was limited to positive intervals so as to not introduce problems with reordering and convergence. If X takes both positive and negative values, call $X^+(x) = \max(X(\omega), 0)$, $X^-(\omega) = \max(-X(\omega), 0)$. If both of those have finite Lebesgue integrals, X is called **integrable**, and

$$\int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} X^+(\omega) d\mathbb{P}(\omega) - \int_{\Omega} X^-(\omega) d\mathbb{P}(\omega). \quad (2.13)$$

Theorem. Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty \iff X$ is integrable.

Proof: Beginning with the forwards direction, $|X| = X^+ + X^-$; $X^+, X^- \geq 0$ so $X^+, X^- \leq |X|$. Hence, via comparison property (in reference texts), $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) \geq \int_{\Omega} X^{\pm}(\omega) d\mathbb{P}(\omega)$. So, $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty$ implies $\int_{\Omega} X^{\pm}(\omega) d\mathbb{P}(\omega) < \infty$, so X is integrable. Now for backwards direction: using linearity, since $|X| = X^+ + X^-$, and both parts have finite integrals, $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty$. \square

Definition. Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the **expectation of X** is

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega). \quad (2.14)$$

Definition. Let X, Y be a random variables with finite expectation. The **variance of X** is

$$Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]. \quad (2.15)$$

The **covariance of X, Y** is

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]. \quad (2.16)$$

Theorem. Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for some Borel-measurable function g , if $\int_{\Omega} |g(x)| d\mu_X(x) < \infty$, then

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathcal{B}} g(x) d\mu_X(x). \quad (2.17)$$

If X has a density, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx. \quad (2.18)$$

Proof idea (Standard machine). Show this is true for indicator functions, then for nonnegative simple functions, then for non-negative Borel-measurable functions, and finally for general Borel-measurable functions, since it is known that non-negative measurable functions are the monotone limit from below of simple functions.

2.1.3 Conditioning and Change of Measure

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-algebra, and X a random variable. Then the **conditional expectation of X given \mathcal{G}** , $\mathbb{E}[X|\mathcal{G}]$, is any random variable that is both \mathcal{G} -measurable, and for which, for all $A \in \mathcal{G}$,

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega). \quad (2.19)$$

This is known as partial averaging.

Remark. The intuition here is that this is the best estimate of X , restricted by the information of the sub-algebra.

Proposition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ sub-algebras, and X, Y integrable random variables. If X is \mathcal{G} -measurable, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$. Additionally, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ (the tower property). If X is independent of \mathcal{G} , then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

Example. Consider two consecutive fair coin tosses with sample space

$$\Omega = \{HH, HT, TH, TT\}, \quad (2.20)$$

equipped with the uniform probability measure.

Define the random variable X as the total number of heads in the two tosses. Let

$$\mathcal{G} = \sigma(\text{first toss}), \quad (2.21)$$

so that \mathcal{G} represents the information revealed by observing the outcome of the first toss.

The σ -algebra \mathcal{G} has atoms

$$\{HH, HT\} \quad \text{and} \quad \{TH, TT\}. \quad (2.22)$$

We compute the conditional expectation $\mathbb{E}[X | \mathcal{G}]$ by averaging X over each atom. On $\{HH, HT\}$, the first toss is heads, and

$$\mathbb{E}[X | \mathcal{G}] = \frac{2+1}{2} = \frac{3}{2}. \quad (2.23)$$

On $\{TH, TT\}$, the first toss is tails, and

$$\mathbb{E}[X | \mathcal{G}] = \frac{1+0}{2} = \frac{1}{2}. \quad (2.24)$$

Thus,

$$\mathbb{E}[X \mid \mathcal{G}] = \begin{cases} \frac{3}{2}, & \text{if the first toss is } H, \\ \frac{1}{2}, & \text{if the first toss is } T. \end{cases} \quad (2.25)$$

Theorem. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let Z be a non-negative random variable almost surely and $\mathbb{E}(Z) = 1$. Then for $A \subset \mathcal{F}$,

$$\tilde{\mathbb{P}} = \int_A Z(\omega) d\mathbb{P}(\omega) \quad (2.26)$$

is also a probability measure. Z is called the **Radon-Nikodym derivative**, and facilitates a change of measure. Therefore, the expectation can be written as

$$\tilde{\mathbb{E}}[X] = \mathbb{E}[XZ]. \quad (2.27)$$

2.1.4 Inference

Definition. Let $\{X_1, X_2, \dots\}$ be a sequence of random variables. $\{X_n\}$ are called **independent and identically distributed (i.i.d.)** if all variables are independent of each other, and all variables have the same distribution function.

Definition. Let $\{X_n\}$ be a sequence of random variables, and let X be another random variable. Then the sequence **converges in probability** if

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0, \forall \epsilon > 0. \quad (2.28)$$

The sequence **converges in distribution** if, for distribution functions

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad (2.29)$$

for all x where $F(x)$ is continuous.

Theorem (Weak Law of Large Numbers). Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_n] = \mu$. Then the sample mean \bar{X} converges in probability to the true mean:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \epsilon\right) = 0. \quad (2.30)$$

Theorem (Central Limit). Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_n] = \mu, \text{Var}[X_n] = \sigma^2$. Then the following converges in distribution to the standard normal distribution $\mathcal{N}(0, 1)$:

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\text{Var}[\bar{X}_n]}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}. \quad (2.31)$$

Alternatively, this can be stated as

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (2.32)$$

2.2 Stochastic Processes

Motivation. In the real world, random processes occur over ranges of time, rather than simply being one variable. The conception of a family of random variables with a time parameter leads to the idea of a stochastic process, which is used to model the randomness in stock prices.

2.2.1 Foundations of Stochastic Processes

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X(t) : \Omega \rightarrow \mathbb{R}^d$ be a family of random variables, indexed by parameter $t \in T$, a parameter set (take T a field such as the real). Then $X(t)$ is called a **stochastic process**.

Definition. Let $\tilde{\Omega}$ be the set of all functions $\tilde{\omega} : T \rightarrow \mathbb{R}^d$ from the parameter set to the real numbers. For a finite set of points $\{t_1, \dots, t_k\}$ in the parameter set and a Borel set $A \in \mathcal{B}(\mathbb{R}^k)$, call a **finite-dimensional cylinder** the set of functions

$$\pi_{t_1, \dots, t_k}^{-1}(A) = \{\tilde{\omega} : (\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k)) \in A\}. \quad (2.33)$$

Additionally, call $\mathcal{B}_{t_1, \dots, t_k}$ the set of all cylinders of fixed parameter set values, and allow A to vary (this is a σ -algebra). Now call \mathcal{B} the cylindrical σ -algebra, the smallest σ -algebra containing all $\mathcal{B}_{t_1, \dots, t_k}$ for all choices of k, t_1, \dots, t_k :

$$\mathcal{B} = \sigma(\pi_{t_1, \dots, t_k}^{-1}(A) : k \in \mathbb{N}, t_i \in T, A \in \mathbb{R}^{kd}). \quad (2.34)$$

Then denote $(\tilde{\Omega}, \mathcal{B})$ the measurable space.

Proposition. Let $(\Omega, \mathcal{B}(\mathbb{R}^d)), (\tilde{\Omega}, \mathcal{B})$ be two measurable spaces. Define $\Phi : \Omega \rightarrow \tilde{\Omega}$ as $\Phi(\omega)(t) = X(t)(\omega)$. Then this mapping is measurable, and thus there is a probability measure $\tilde{\mathbb{P}} = \mathbb{P} \circ \Phi$ induced on $(\tilde{\Omega}, \mathcal{B})$.

Definition. Let $X(t)$ be a stochastic process. The **finite-dimensional distributions** are push forward measures into \mathbb{R}^k defined by $\mathcal{B}_{t_1, \dots, t_k}$:

$$\mathbb{P}_{t_1, \dots, t_k}^X(A) = \mathbb{P}(\omega \in \Omega : (X(t_1)(\omega), \dots, X(t_k)(\omega)) \in A). \quad (2.35)$$

Theorem (Kolmogorov Extension). Let $\{\mathbb{P}_{t_1, \dots, t_k}\}$ be a family of finite-dimensional probability measures on $\mathcal{B}(\mathbb{R}^{dk})$. If the family of probability measures satisfies the following consistency conditions (permutation invariance and marginal consistency), taking $A \in \mathcal{B}(\mathbb{R}^k)$,

$$\mathbb{P}_{t_1, \dots, t_k}((\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k)) \in A) = \mathbb{P}_{\pi(t_1, \dots, t_k)}(\pi((\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k))) \in A) \quad (2.36)$$

$$\mathbb{P}_{t_1, \dots, t_k}((\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k)) \in A) = \mathbb{P}_{t_1, \dots, t_{k+1}}((\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k)) \in A \times \mathbb{R}), \quad (2.37)$$

then there exists a probability measure $\tilde{\mathbb{P}}$ on $(\tilde{\Omega}, \mathcal{B})$ whose finite-dimensional distributions coincide with the given family.

Remark. This is helpful in arguing that processes that satisfy certain properties exist, and is used to construct the Wiener process.

2.2.2 Properties of Stochastic Processes

Definition. Let Ω be a nonempty set. Assume that $\forall t \in T, \exists \mathcal{F}_t$ a σ -algebra. Additionally, assume that $s \leq t \implies \mathcal{F}_s \subset \mathcal{F}_t$. Then the collection of algebras $\mathcal{F}(t)$ is a **filtration**. If $X(t)$ is a collection of random variables indexed $[0, T]$, then $X(t)$ is called an **adapted stochastic process** if $X(t)$ is $\mathcal{F}(t)$ -measurable for $t \in [0, T]$.

Definition. Let τ be a random variable taking values in T on $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\mathcal{F}(t)$ defined for all $t \in T$. τ is called a **stopping time** if

$$\{\omega : \tau(\omega) \leq t\} = \{\tau \leq t\} \in \mathcal{F}(t), \forall t \in T. \quad (2.38)$$

A **stopped process** for $X(t)$ a stochastic process is

$$X(t \wedge \tau), \quad (2.39)$$

where \wedge denotes the "min" operation.

Remark. $X(t)$ being $\mathcal{F}(t)$ -measurable means that you can resolve all of the random variable's behavior at that time given the information known at that time. This is a way to formulate causality, since it means that information from after time t cannot be used in determining $X(t)$. A stopping time being defined as such means that the decision to stop must be made only with information available at a given time. Any decision made by, say, a trader must occur at a stopping time.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{F}(t)$ a filtration of sub- σ -algebras of \mathcal{F} , and $M(t)$ an adapted stochastic process. $M(t)$ is called a **martingale** if, for all $s \leq t$,

$$\mathbb{E}[M(t) | \mathcal{F}(s)] = M(s). \quad (2.40)$$

A **submartingale** has $\mathbb{E}[M(t) | \mathcal{F}(s)] \geq M(s)$, and a **supermartingale** has $\mathbb{E}[M(t) | \mathcal{F}(s)] \leq M(s)$.

Proposition. A stopped sub/super/martingale is still a sub/super/martingale.

Theorem (Doob's Optional Stopping). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{F}(t)$ a filtration, $M(t)$ a stochastic process, and τ a stopping time. Then $M(t)$ being a martingale is equivalent to, for any almost surely bounded τ with $\mathbb{E}[|M(\tau)|] < \infty$,

$$\mathbb{E}[M(\tau)] = \mathbb{E}[M(0)]. \quad (2.41)$$

Remark. This theorem is often invoked to argue that it is impossible to truly win money on average at a fair game.

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{F}(t)$ a filtration. Let Z be an almost surely positive random variable with $\mathbb{E}(Z) = 1$, and $\tilde{\mathbb{P}}$ defined from

the Radon-Nikodym derivative. Then, call the **Radon-Nikodym derivative process**

$$Z(t) = \mathbb{E}[Z|\mathcal{F}(t)]. \quad (2.42)$$

This is a martingale, as, by conditional identities,

$$\mathbb{E}[Z|\mathcal{F}(t)] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}(s)]|\mathcal{F}(t)] = \mathbb{E}[Z|\mathcal{F}(s)] = Z(s). \quad (2.43)$$

Lemma. Let $Z(t)$ be a Radon-Nikodym derivative process on a filtration, and Y an $\mathcal{F}(t)$ -measurable random variable. Then

$$\tilde{\mathbb{E}}[Y] = \mathbb{E}[YZ(t)], \quad (2.44)$$

and

$$\tilde{\mathbb{E}}[Y|\mathcal{F}(s)] = \frac{1}{Z(s)} \mathbb{E}[YZ(t)|\mathcal{F}(s)]. \quad (2.45)$$

Proof. For the first part, $\tilde{\mathbb{E}}[Y] = \mathbb{E}[YZ] = \mathbb{E}[\mathbb{E}[YZ|\mathcal{F}(t)]]$, via the partial averaging property of conditional expectations. Since Y is $\mathcal{F}(t)$ -measurable, $\mathbb{E}[\mathbb{E}[YZ|\mathcal{F}(t)]] = \mathbb{E}[Y\mathbb{E}[Z|\mathcal{F}(t)]] = \mathbb{E}[YZ(t)]$.

For the second part, let A be an arbitrary subset of $\mathcal{F}(s)$. Then $\tilde{\mathbb{E}}[\mathbb{I}_A \frac{1}{Z(s)} \mathbb{E}[YZ(t)|\mathcal{F}(s)]] = \mathbb{E}[\mathbb{E}[\mathbb{I}_A YZ(t)|\mathcal{F}(s)]]$, by the definition of the Radon-Nikodym derivative. Therefore, by partial averaging, this equals $\mathbb{E}[\mathbb{I}_A YZ(t)] = \tilde{\mathbb{E}}[\mathbb{I}_A Y]$. This is both $\mathcal{F}(s)$ -measurable and partial averages, so this is indeed the conditional expectation. \square

2.2.3 Markov Processes

Definition. Let $\{X(n)\}_{n \in \mathbb{N}}$ be a sequence of random variables taking values in a countable state space $S = \{x_1, x_2, \dots\}$. The process $X(n)$ is called a **discrete-time Markov chain** if

$$\mathbb{P}(X(n+1) = x_j \mid X(n), X(n-1), \dots) = \mathbb{P}(X(n+1) = x_j \mid X(n) = x_i). \quad (2.46)$$

The chain is called **time-homogeneous** if the transition probabilities do not depend on n . In this case, there exists a **transition matrix**

$$P = (p_{ij}), \quad p_{ij} = \mathbb{P}(X(n+1) = x_j \mid X(n) = x_i). \quad (2.47)$$

Proposition. The Markov property implies that the entire probabilistic structure of the chain is determined by the transition matrix and the initial distribution. In particular,

$$\mathbb{P}(X(n) = x_j \mid X(0) = x_i) = (P^n)_{ij}. \quad (2.48)$$

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}(t)\}_{t \geq 0}$, and let $X(t)$ be an adapted stochastic process taking values in \mathbb{R}^d . The process $X(t)$ is called a **Markov process (continuous time, continuous state space)** if, for all bounded Borel-measurable functions f , and all $s \leq t$,

$$\mathbb{E}[f(X(t)) \mid \mathcal{F}(s)] = \mathbb{E}[f(X(t)) \mid X(s)]. \quad (2.49)$$

Equivalently, there exists a Borel-measurable function g such that

$$\mathbb{E}[f(X(t)) \mid \mathcal{F}(s)] = g(X(s)). \quad (2.50)$$

The corresponding **transition probabilities** are defined by

$$p(s, t, x, A) = \mathbb{P}(X(t) \in A \mid X(s) = x). \quad (2.51)$$

If the process is time-homogeneous, then $p(s, t, x, A) = p(t - s, x, A)$.

2.2.4 Brownian Motion

Motivation. When modeling stocks, something must account for the random motion. That is modeled by Brownian motion, the characteristic zig-zag of a random walk.

Theorem. There exists a stochastic process $W(t)$ such that $W(0) = 0$ almost surely, $W(t)$ is almost surely continuous, $W(t)$ has independent increments, and $W(t) - W(s) \sim \mathcal{N}(0, t - s)$. That process is called the **Wiener process** or **Brownian motion**. Multidimensional Brownian motion can likewise be defined as $W(t) = [W_1(t), W_2(t), \dots]$.

Proof idea. Use Kolmogorov extension theorem.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined $W(t)$. A **filtration for Brownian motion**, $\mathcal{F}(t)$ is a standard filtration such that, for $t \leq u$, $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.

Theorem. Brownian motion is a martingale.

Proof. $\mathbb{E}[W(t) \mid \mathcal{F}(s)] = \mathbb{E}[(W(t) - W(s)) + W(s) \mid \mathcal{F}(s)] = 0 + W(s)$, using independence and the $\mathcal{F}(s)$ -measurability of $W(s)$. \square

Theorem. Brownian motion is a Markov process.

Theorem. Brownian motion accumulates **quadratic variation** - that is,

$$[W, W](T) = \sum_j [W(t_{j+1}) - W(t_j)]^2 \rightarrow T. \quad (2.52)$$

For multidimensional Brownian motion, this is instead $[W_i, W_j](t) = \delta^{ij}t$.

Theorem (Lévy). Let $M(t)$ be a martingale with respect to filtration $\mathcal{F}(t)$, such that $M(0) = 0$, $M(t)$ has continuous paths, and $[M, M](t) = t$. Then

$M(t)$ is a Brownian motion. For the multidimensional case, the statements are identical, except the quadratic variation becomes $[M_i, M_j](t) = \delta^{ij}t$

Example (Scaling). Let $(W(t))_{t \geq 0}$ be a standard Brownian motion. For any constant $a > 0$, define the process

$$\widetilde{W}(t) = \frac{1}{a}W(a^2t). \quad (2.53)$$

Then $\widetilde{W}(t)$ is also a standard Brownian motion. Indeed, $\widetilde{W}(0) = 0$, the paths are continuous as the underlying Brownian motion has not been muddled. $\widetilde{W}(t)$ is a martingale, as

$$\mathbb{E}[\widetilde{W}(t)|\mathcal{F}(s)] = \mathbb{E}\left[\frac{1}{a}W(a^2t)|\mathcal{F}(s)\right], \quad (2.54)$$

yet the natural filtration is $\mathcal{F}(s) = \sigma(\widetilde{W}(u), u \leq s) = \sigma(W(u), u \leq a^2s)$, so the expectation is

$$\frac{1}{a}\mathbb{E}[W(a^2t)|\sigma(W(u), u \leq a^2s)]. \quad (2.55)$$

Standard Brownian motion is a martingale with respect to its natural filtration, so this becomes $\frac{1}{a}W(a^2s) = \widetilde{W}(s)$, so this is a martingale.

Finally, we check quadratic variation:

$$\left[\frac{1}{a}W(a^2t), \frac{1}{a}W(a^2t)\right] = \frac{1}{a^2}a^2t = t, \quad (2.56)$$

Since the quadratic variation with arbitrary time input is that time input. Hence, via Lévy's theorem, this is indeed Brownian motion.

If an asset price follows a diffusion driven by Brownian motion, this scaling implies that volatility scales with the square root of time. In particular, if σ_{daily} denotes daily volatility, then annualized volatility satisfies

$$\sigma_{\text{annual}} \approx \sqrt{252} \sigma_{\text{daily}}, \quad (2.57)$$

a convention widely used in financial practice.

Proposition (Reflection). Call $M(t)$ the maximum reached by $W(t)$ in the interval $[0, t]$, and call τ_m the first hitting time of the value m by $W(t)$. Then,

$$\mathbb{P}(M(t) \geq a) = 2\mathbb{P}(W(t) \geq a), \quad (2.58)$$

and

$$\mathbb{P}(\tau_m \leq t, W(t) \leq w) = \mathbb{P}(W(t) \geq (2m - w)). \quad (2.59)$$

2.3 Stochastic Calculus

Motivation. Brownian motion famously does not have a time derivative. The following section will formulate a version of calculus for processes with Brownian motion with consistent notions of differentials and integrals, since that will be the backbone of stochastic modeling of derivatives.

2.3.1 Itô Calculus

Definition. Consider $W(t)$ Brownian motion and $\mathcal{F}(t)$ a filtration for it, defined on $[0, T]$. Let $\Delta(t)$ be an adapted stochastic process, and Π a partition of sub-points on the interval. Then, as the supremum of sub-point distances goes to zero, the sum

$$\sum_j^{k-1} (\Delta(t_j)[W(t_{j+1}) - W(t_j)] + \Delta(t_k)[W(t_{k+1}) - W(t_k)]) \quad (2.60)$$

converges to the **Itô integral**

$$I(T) = \int_0^T \Delta(t) dW(t), \quad (2.61)$$

provided that $\mathbb{E}[\int_0^T \Delta(t)^2 dW(t)] < \infty$.

Theorem. For Itô integral $I(T) = \int_0^T \Delta(t) dW(t)$, the integral is a martingale. The integral is continuous and $\mathcal{F}(t)$ -measurable. The integral accumulates quadratic variation equal to

$$[I, I](t) = \int_0^t \Delta(u)^2 du. \quad (2.62)$$

Additionally, the integral satisfies the following isometry:

$$\mathbb{E}[I^2(t)] = \int_0^t \Delta(u)^2 du. \quad (2.63)$$

Definition. Consider $W(t)$ Brownian motion and $\mathcal{F}(t)$ an associated filtration, defined on $[0, T]$. Let $\Delta(t), \Theta(t)$ be adapted stochastic processes. Then, define an **Itô process** for $X(t)$ non-random as

$$X(t) = X(0) + \int_0^t \Theta(u) du + \int_0^t \Delta(u) dW(u), \quad (2.64)$$

or, in differential notation,

$$dX(t) = \Theta(t)dt + \Delta(t)dW(t). \quad (2.65)$$

Lemma (Itô). Let $f(x, t)$ be a function for which the partial derivatives exist and are continuous, and $W(t)$ be a Brownian motion. Then, in differential form,

$$df(t, W(t)) = \frac{\partial f}{\partial t}(t, W(t))dt + \frac{\partial f}{\partial x}(t, W(t))dW(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, W(t))dt. \quad (2.66)$$

If $f(t, X(t))$ takes a general Itô process, then

$$df(t, X(t)) = \frac{\partial f}{\partial t}(t, X(t))dt + \frac{\partial f}{\partial x}(t, X(t))dX(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X(t))dX(t)^2. \quad (2.67)$$

If the function involves a product of two Itô processes, then the product rule gives

$$d(X(t)Y(t)) = X(t)dY(t) + dX(t)Y(t) + dX(t)dY(t). \quad (2.68)$$

For a multi-dimensional process $X(t)$, where the coefficient functions are now vectors,

$$df(t, X(t)) = \frac{\partial f}{\partial t}(t, X(t))dt + \sum_i \frac{\partial f}{\partial x_i}(t, X(t))dX_i(t) \quad (2.69)$$

$$+ \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X(t))dX_i(t)dX_j(t). \quad (2.70)$$

Proof idea. Use Taylor's formula along with the (informal) rule $dt dt = dW dt = 0, dW dW = dt$.

Example (GBM). One model of stock price motion, for $S(t)$ the asset's price over time, is.

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t). \quad (2.71)$$

The solution to this **stochastic differential equation**, an equation relating differentials of stochastic processes with Brownian motion, is obtained as follows:

$$\frac{dS(t)}{S(t)} = \mu(t)dt + \sigma(t)dW(t). \quad (2.72)$$

If this were ordinary calculus, one could simply use the chain rule to obtain a logarithm, yet this does not work here. Instead, we can propose something similar. By Itô's lemma, the differential of the function $f(t, S(t)) = \log(S(t))$ is

$$d(f(t, S(t))) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dS(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}dS(t)^2 \quad (2.73)$$

$$= 0 + \frac{1}{S(t)}dS(t) + \frac{1}{2} \frac{-1}{S(t)^2}(dS(t))^2 \quad (2.74)$$

$$= \mu(t)dt + \sigma(t)dW(t) - \frac{1}{2} \frac{1}{S(t)^2} \sigma(t)^2 S(t)^2 dt \quad (2.75)$$

$$= (\mu(t) - \frac{1}{2} \sigma(t)^2)dt + \sigma(t)dW(t). \quad (2.76)$$

Fortunately, this choice of substitution yields all $S(t)$ in the differential on one side. Integrating and exponentiating yields

$$S(t) = S(0) \exp\left(\int_0^t (\mu(u) - \frac{1}{2}\sigma(u)^2)du + \int_0^t \sigma(u)dW(u)\right). \quad (2.77)$$

This is known as **geometric Brownian motion**.

Theorem (Girsanov). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $W(t)$ a Brownian motion, and $\mathcal{F}(t)$ a filtration for the Brownian motion from $[0, T]$. Let $\Theta(t)$ be an adapted process, and define

$$Z(t) = \exp\left(-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2 du\right) \quad (2.78)$$

$$\widetilde{W}(t) = W(t) + \int_0^t \Theta(u)du. \quad (2.79)$$

If $\mathbb{E}[\int_0^T \Theta(u)^2 Z(u)^2 du] < \infty$, then, taking $Z = Z(T)$, $\mathbb{E}[Z] = 1$, and $\widetilde{W}(t)$ is a Brownian motion under $\widetilde{\mathbb{P}}$.

The multidimensional case is identical, except the Itô integral uses a dot product between the adapted process and the Brownian motion, and the square turns into the Euclidean norm.

Proof idea. Use Lévy's theorem by showing martingale property.

Theorem (Martingale representation). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $W(t)$ a Brownian motion, and $\mathcal{F}(t)$ a filtration for the Brownian motion from $[0, T]$. Let $M(t)$ be a martingale with respect to the filtration. Then there exists an adapted process $\Gamma(u)$ such that

$$M(t) = M(0) + \int_0^t \Gamma(u)dW(u). \quad (2.80)$$

If $\widetilde{M}(t)$ is likewise a martingale under $\widetilde{\mathbb{P}}$ with Girsanov-generated Brownian motion $\widetilde{W}(u)$, then there exists an adapted process $\widetilde{\Gamma}(u)$ such that

$$\widetilde{M}(t) = \widetilde{M}(0) + \int_0^t \widetilde{\Gamma}(u)d\widetilde{W}(u). \quad (2.81)$$

The multidimensional analogue is identical, except for the Itô integral gaining a dot product.

2.3.2 SDE-PDE Connection

Proposition. Consider the stochastic differential equation

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u) \quad (2.82)$$

Under standard regularity conditions, the solution exists and defines a time-inhomogeneous Markov process. If β, γ are time-independent, the process is time-homogeneous.

Theorem (Feynman-Kac). Let $dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$ be an SDE, and let $h(y)$ be any Borel-measurable function. For $0 \leq t \leq T$, define

$$g(t, x) = \mathbb{E}[h(X(T)) | X(t) = x]. \quad (2.83)$$

This exists as SDE solutions are Markov processes. Then g satisfies the following PDE:

$$\frac{\partial g}{\partial t}(t, x) + \beta(t, x) \frac{\partial g}{\partial x}(t, x) + \frac{1}{2} \gamma(t, x)^2 \frac{\partial^2 g}{\partial x^2}(t, x) = 0, \quad (2.84)$$

with $g(T, x) = h(x)$.

If instead one defines a discounted process

$$f(t, x) = \mathbb{E}[e^{-r(T-t)} h(X(T)) | X(t) = x], \quad (2.85)$$

then f satisfies the following PDE:

$$\frac{\partial f}{\partial t}(t, x) + \beta(t, x) \frac{\partial f}{\partial x}(t, x) + \frac{1}{2} \gamma(t, x)^2 \frac{\partial^2 f}{\partial x^2}(t, x) = r f(t, x), \quad (2.86)$$

with $f(T, x) = h(x)$.

Proof idea. Show that f, g is a martingale, then use Itô's lemma to evaluate df, dg and impose the condition that the coefficient on dt is zero to ensure martingale property.

Remark. This is a link between SDEs and PDEs. Many SDEs, which do not appear to have analytical solutions, can be solved via switching to the dual PDE, which can be solved numerically. The following theorems are more general elaborations on this theme.

Definition. Define operators P_t acting on bounded measurable functions. Then $\{P_t\}_{t \geq 0}$ forms a **semigroup** if

$$P_0 = \text{id}, \quad P_{s+t} = P_s P_t. \quad (2.87)$$

If P_t maps $C_0(\mathbb{R}^d)$ into itself and is strongly continuous, it is called a **Feller semigroup**.

Proposition. For $X(t)$ a homogeneous Markov process, defining operators P on bounded measurable functions f as

$$(P_t f)(x) := \mathbb{E}[f(X(t)) | X(0) = x] \quad (2.88)$$

forms a semigroup. If the process is inhomogeneous, there is instead a two-parameter family with an evolution property:

$$(P_{t,T}f)(x) = \mathbb{E}[f(X(T)) \mid X(t) = x], \quad P_{t,T} = P_{t,s}P_{s,T}. \quad (2.89)$$

Proof. Since $X(0) = x$ almost surely under $\mathbb{P}(\cdot \mid X(0) = x)$, then $(P_0f)(x) = \mathbb{E}[f(X(0)) \mid X(0) = x] = f(x)$. So $P_0 = \text{id}$.

To show the second property, fix $x \in \mathbb{R}^d$ and a bounded measurable function f . By definition, $(P_{s+t}f)(x) = \mathbb{E}[f(X(s+t)) \mid X(0) = x]$. Insert a conditional expectation with respect to \mathcal{F}_s using the tower property: $(P_{s+t}f)(x) = \mathbb{E}[\mathbb{E}[f(X(s+t)) \mid \mathcal{F}_s] \mid X(0) = x]$.

Since $X(t)$ is Markov, we have $\mathbb{E}[f(X(s+t)) \mid \mathcal{F}_s] = \mathbb{E}[f(X(s+t)) \mid X(s)]$. By time-homogeneity, we also have $\mathbb{E}[f(X(s+t)) \mid X(s) = y] = (P_tf)(y)$, therefore $\mathbb{E}[f(X(s+t)) \mid \mathcal{F}_s] = (P_tf)(X(s))$. Substituting back gives: $(P_{s+t}f)(x) = \mathbb{E}[(P_tf)(X(s)) \mid X(0) = x] = (P_s(P_tf))(x)$. Since this holds for all x and all bounded measurable f , the second property holds. The inhomogeneous case is not shown. \square

Definition. The **generator** \mathcal{L} of the semigroup is defined by

$$\mathcal{L}f = \lim_{t \downarrow 0} \frac{P_tf - f}{t}, \quad (2.90)$$

for all f in its domain. For the homogeneous Markov case, substituting the semigroup operators gives

$$\mathcal{L}f(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}[f(X(t)) \mid X(0) = x] - f(x)}{t}, \quad (2.91)$$

or, for the inhomogeneous case,

$$\mathcal{L}_t f = \lim_{h \rightarrow 0} \frac{P_{t,t+h}(f) - f}{h}. \quad (2.92)$$

The **adjoint** of the generator, \mathcal{L}^* , is defined such that $\langle \mathcal{L}f, g \rangle = \langle f, \mathcal{L}^*g \rangle$ for test functions and measures f, g respectively, given that their integrals exist and boundary terms vanish. The adjoint acts on densities or measures.

Proposition. A homogeneous Markov process $X(t)$ is said to be **affine** if

$$\mathbb{E}[e^{uX(t)} \mid X(t) = x] = e^{A(T-t,u) + B(T-t,u)x}, \quad (2.93)$$

for non-random functions A, B . This property is equivalent to the generator mapping exponentials to exponentials times an affine function of the state:

$$\mathcal{L}e^{ux} = (a(u) + b(u)x)e^{ux}. \quad (2.94)$$

Proposition. Let $X(t)$ satisfy the stochastic differential equation $dX(u) = \beta(u, X(u)) du + \gamma(u, X(u)) dW(u)$, and define the Markov semigroup $(P_tf)(x) =$

$\mathbb{E}[f(X(t)) \mid X(0) = x]$, for $f \in C_b^2(\mathbb{R})$. Then the generator of the semigroup for a time-inhomogeneous process, acting on the spatial variable, is

$$\mathcal{L}_t f(x) = \beta(t, x) \frac{\partial f}{\partial x}(x) + \frac{1}{2} \gamma(t, x)^2 \frac{\partial^2 f}{\partial x^2}(x), \quad (2.95)$$

or, in the time-homogeneous case,

$$\mathcal{L} f(x) = \beta(x) \frac{df}{dx}(x) + \frac{1}{2} \gamma(x)^2 \frac{d^2 f}{dx^2}(x). \quad (2.96)$$

Note that this is coefficient in front of dt in Itô's lemma. The adjoint is of the form

$$\mathcal{L}_t^* p(t, x) = -\frac{\partial}{\partial x}(\beta(t, x)p(t, x)) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(\gamma(t, x)^2 p(t, x)). \quad (2.97)$$

Proof idea. Take the differential $df(X(u))$, take the conditional expectation, and take the limit to zero. For the adjoint, use integration by parts.

Theorem (Kolmogorov backward equation). Let $u(t, x) = \mathbb{E}[f(X(t)) \mid X(0) = x]$. Then,

$$\frac{\partial u}{\partial t}(t, x) = \mathcal{L}u(t, x), \quad u(0, x) = f(x). \quad (2.98)$$

By replacing t by $T - t$, as is more common with time flowing backwards from T , this takes the form

$$-\frac{\partial u}{\partial t}(t, x) = \mathcal{L}u(t, x), \quad u(T, x) = f(x). \quad (2.99)$$

If $X(t)$ is defined as $dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$, and the transition density as $p(t, T, x, y) dy = \mathbb{P}(X(T) \in dy \mid X(t) = x)$, then this can be written

$$-\frac{\partial p}{\partial t}(t, T, x, y) = \beta(t, x) \frac{\partial p}{\partial x}(t, T, x, y) + \frac{1}{2} \gamma(t, x)^2 \frac{\partial^2 p}{\partial x^2}(t, T, x, y). \quad (2.100)$$

If instead, in the discounted case, $u(t, x) = \mathbb{E}[e^{-r(T-t)} f(X(T)) \mid X(t) = x]$, then

$$\frac{\partial u}{\partial t}(t, x) + \mathcal{L}u(t, x) - ru(t, x) = 0, \quad u(T, x) = f(x). \quad (2.101)$$

Theorem (Kolmogorov forward equation). For p the probability density of the random process

$$\frac{\partial p}{\partial t}(t, x) = \mathcal{L}_t^*(p(t, x)). \quad (2.102)$$

Considering again the SDE $dX(u) = \beta(u, X(u)) du + \gamma(u, X(u)) dW(u)$, denote the transition density of the process $p(t, T, x, y) dy = \mathbb{P}(X(T) \in dy \mid X(t) = x)$. The transition density then satisfies the **Fokker-Planck equation**:

$$\frac{\partial}{\partial T} p(t, T, x, y) = -\frac{\partial}{\partial y} (\beta(T, y) p(t, T, x, y)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\gamma(T, y)^2 p(t, T, x, y)), \quad (2.103)$$

for $T > t$, with initial condition

$$\lim_{T \downarrow t} p(t, T, x, y) = \delta(y - x). \quad (2.104)$$

Proof. Take the backwards equation explicitly for u

$$\frac{\partial}{\partial t} \int f(x) p(t, x) dx = \int \mathcal{L}_t(f(x)) p(t, x) dx \quad (2.105)$$

$$\int f(x) \frac{\partial p}{\partial t}(t, x) dx = \int \mathcal{L}_t^*(f(x)) p(t, x) dx. \quad (2.106)$$

By definition of the adjoint, this gives

$$\int f(x) \frac{\partial p}{\partial t}(t, x) dx = \int f(x) \mathcal{L}_t^*(p(t, x)) dx. \quad (2.107)$$

Since this holds for all functions in the domain, $\frac{\partial p}{\partial t}(t, x) = \mathcal{L}_t^*(p(t, x))$. Plugging in the form that the adjoint takes yields the Fokker-Planck and transition probability versions. \square

Remark. The backward equation governs the evolution of observables, while the forward equation governs the evolution of probability distributions. Both are determined by the generator \mathcal{L} .

2.4 Jump Diffusion

Motivation. Pure diffusion models imply continuous price paths and fail to capture empirically observed features such as sudden price jumps, volatility clustering, and heavy-tailed return distributions.

Definition: Given a metric space (M, d) , a function f is called **càdlàg** (*continue à droite, limite à gauche*) if both the left and the right limit exist, and the right limit equals the function value at all points.

Definition. On filtered probability space, a stochastic process $X(t)$ is called a **Lévy process** if

- $X(0) = 0$ almost surely,
- the paths of $X(t)$ are càdlàg,
- $X(t)$ is adapted to $\mathcal{F}(t)$,

- the increments $X(t) - X(s)$ are mutually independent,
- the increments are stationary ($X(t) - X(s) = X(t - s)$ in distribution),
- the process is continuous in probability

$$\lim_{h \rightarrow 0} \mathbb{P}(|X(t+h) - X(t)| > 0) = 0. \quad (2.108)$$

Theorem. There exists a stochastic process $N(t)$ such that $N(0) = 0$ almost surely, $N(t)$ is almost surely càdlàg, $N(t)$ has independent increments, and $N(t) - N(s)$ is distributed as a Poisson distribution with intensity $\lambda(t - s)$. That process is called a **Poisson process**.

Proof idea. Use Kolmogorov extension theorem.

Proposition. The sum of independent Poisson processes with intensities λ_1, λ_2 is a Poisson process with intensity $\lambda_1 + \lambda_2$. Likewise, a Poisson process with intensity $\lambda_1 + \lambda_2$ can be decomposed into a sum of independent Poisson processes with intensities λ_1, λ_2 .

Remark. For some intuition, if the idea of a Poisson process is to count the number of events in a window of time, then for some given window of time, the amount of events that should occur within that time for two processes is just the same one process with a larger rate.

Definition. Let $N(t)$ be a Poisson process with intensity $\lambda > 0$, and let $(Y_k)_{k \geq 1}$ be a sequence of i.i.d. real-valued random variables with mean β , independent of $N(t)$. The process

$$Q(t) = \sum_{k=1}^{N(t)} Y_k \quad (2.109)$$

is called a **compound Poisson process**. $N(t)$ counts the number of jump events up to time t , while Y_k represents the size of the k -th jump. The process $Q(t)$ therefore models random jump arrivals with random magnitudes.

Proposition. A compound Poisson process is a Lévy process.

Proof. The increments are stationary and independent due to the corresponding properties of the Poisson process and the independence of the jump sizes. The paths are càdlàg with finitely many jumps on any compact interval. \square

Definition. For a Poisson process of intensity λ , the variable

$$M(t) = N(t) - \lambda t \quad (2.110)$$

is called a **compensated Poisson process**, and is a martingale. For a compound Poisson process, the martingale is

$$Q(t) - \beta t \lambda. \quad (2.111)$$

Theorem (Lévy-Itô decomposition). Let $X(t)$ be a Lévy process. Then $X(t)$ has a canonical decomposition into continuous and jump components,

$$X(t) = bt + \sigma W(t) + \int_{|x|<1} x \tilde{N}(t, dx) + \int_{|x|\geq 1} x N(t, dx), \quad (2.112)$$

where:

- $b \in \mathbb{R}$ is a drift term,
- $\sigma \geq 0$,
- $W(t)$ is a standard Brownian motion,
- $N(t, dx)$ is a Poisson random measure,
- $\tilde{N}(t, dx) = N(t, dx) - t\nu(dx)$ is the compensated Poisson measure,
- $\nu(dx)$ is the **Lévy measure**, satisfying

$$\int_{\mathbb{R}} \min(1, x^2) \nu(dx) < \infty. \quad (2.113)$$

Remark. The Lévy measure ν governs the intensity and size distribution of jumps. Large jumps occur finitely often on finite intervals, while small jumps may occur infinitely often but with finite total quadratic variation.

Definition. A **jump diffusion** is a stochastic process of the form

$$X(t) = X(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s) + \sum_{k=1}^{N(t)} Y_k, \quad (2.114)$$

where:

- $W(t)$ is Brownian motion,
- $N(t)$ is a Poisson process with intensity λ ,
- Y_k are i.i.d. jump sizes, independent of W and N .

Remark. Jump diffusions combine continuous fluctuations with discrete shocks and are capable of modeling sudden price moves, crashes, and news arrivals that cannot be captured by pure diffusion models.

Definition. A **stochastic differential equation with jumps** is written as

$$dX(t) = \mu(t, X(t^-)) dt + \sigma(t, X(t^-)) dW(t) + \int_{\mathbb{R}} \gamma(t, X(t^-), x) N(dt, dx), \quad (2.115)$$

where $X(t^-)$ denotes the left limit of X at time t .

Remark. The presence of jumps destroys pathwise continuity and invalidates classical Itô calculus. Instead, one works with extensions of Itô's formula for semimartingales with jumps.

Theorem (Itô's lemma with jumps). Let $X(t)$ be a jump diffusion and let $f \in C^2(\mathbb{R})$. Then

$$df(X(t)) = \frac{\partial f}{\partial x}(X(t^-)) dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X(t^-)) \sigma^2(t, X(t^-)) dt \quad (2.116)$$

$$+ \sum_{0 < s \leq t} \left(f(X(s)) - f(X(s^-)) - \frac{\partial f}{\partial x}(X(s^-)) \Delta X_s \right), \quad (2.117)$$

where $\Delta X(s) = X(s) - X(s^-)$ denotes the jump size at time s .

Remark. Jump diffusions provide a mathematically tractable framework for incorporating these effects and form the foundation for models such as the Merton jump-diffusion model and more general Lévy-driven asset price processes.

Chapter 3

Risk-Neutral Formulation

3.1 Risk-Neutral Pricing

Motivation. Many economics models use the assumption of no-arbitrage when developing theories of asset pricing. This section elaborates on the mathematical requirements for and implications of no-arbitrage.

3.1.1 One Asset

Definition. A **portfolio** of an agent has value denoted $X(t)$, and is the value of a combination of stock and money-market investments (at risk-free rate r). Let the stock be geometric Brownian motion, and let $\Delta(t)$ be the number of shares of stock held, adapted to the underlying filtration. The differential for the portfolio with one stock is then

$$dX(t) = \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt. \quad (3.1)$$

Definition. For an adapted interest rate process, define the **discount process** as

$$D(t) = e^{-\int_0^t R(u)du}, \quad (3.2)$$

with differential

$$dD(t) = -R(t)D(t)dt. \quad (3.3)$$

The **discounted stock price** (in light of the time value of money) is therefore

$$D(t)S(t) = S(0) \exp\left(\int_0^t (\mu(s) - R(s) - \frac{1}{2}\sigma(s)^2)ds + \int_0^t \sigma(s)dW(s)\right), \quad (3.4)$$

with differential

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)[\Theta(t)dt + dW(t)], \quad (3.5)$$

in which the **market price of risk** is defined as $\Theta(t) = \frac{\mu(t) - R(t)}{\sigma(t)}$.

Definition. In the above setting, using Girsanov's theorem to switch to $\widetilde{W}(t) = W(t) + \int_0^t \Theta(u)du$ yields

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)d\widetilde{W}(t). \quad (3.6)$$

Then $\widetilde{\mathbb{P}}$, the probability measure generated by applying Girsanov's theorem with the market price of risk as the adapted stochastic process, is called the **risk-neutral measure**.

Under a risk-neutral measure, which is equivalent to the initial probability measure, the discounted price of an asset is a martingale, as is visible from the lack of dt term in the differential.

Proposition. Discounted portfolio values are martingales under the risk-neutral measure - they have an average return of $R(t)$:

$$d(D(t)X(t)) = \Delta(t)\sigma(t)D(t)S(t)d\widetilde{W}(t). \quad (3.7)$$

Remark: This is quite helpful for pricing assets. The usual formula, that the present value of a portfolio is equal to the risk-neutral expectation of discounted future cash flows, is reproduced from this. The insistence upon the martingale nature is helpful in that we can get expectations of future states given current information in straightforward expressions.

Theorem (Risk-neutral pricing). For $V(T)$, a $\mathcal{F}(T)$ -measurable random variable, an interest rate process, and a risk-neutral probability measure, the value of $V(t)$ at time $t < T$ is

$$V(t) = \widetilde{\mathbb{E}}[e^{-\int_t^T R(u)du} V(T) | \mathcal{F}(t)]. \quad (3.8)$$

Proof. Under the risk-neutral measure, discounted asset prices are martingales. Hence, for hedge $X(t)$, $D(t)X(t) = \widetilde{\mathbb{E}}[D(T)X(T) | \mathcal{F}(t)]$. If a hedge can be constructed, $X(T) = V(T)$ almost surely, so this equals $\widetilde{\mathbb{E}}[D(T)V(T) | \mathcal{F}(t)]$. One can call this the equivalent price $V(t)$, discounted. Therefore, $V(t) = \widetilde{\mathbb{E}}[e^{-\int_t^T R(u)du} V(T) | \mathcal{F}(t)]$. \square

Proposition (Existence of hedges). Earlier, it was unclear if hedges could be constructed - this proposition qualifies that. Consider $V(T)$, a $\mathcal{F}(T)$ -measurable random variable, representing the payoff of a derivative security. In order to hedge/replicate a short position against the derivative, there must be a portfolio such that $X(T) = V(T)$ almost surely. For a single stock, choosing

$$X(0) = V(0), \quad \Delta(t) = \frac{\widetilde{\Gamma}(t)}{\sigma(t)D(t)S(t)} \quad (3.9)$$

guarantees the hedge, assuming that the volatility is never zero, and that Brownian motion is the only source of uncertainty. If all securities can be hedged, a market is called **complete**.

Proof. Under the risk-neutral measure, $D(t)V(t) = \mathbb{E}[D(T)V(T)|\mathcal{F}(t)]$, as this is a martingale. Therefore, by the martingale representation theorem, this can be written $D(t)V(t) = V(0) + \int_0^t \tilde{\Gamma}(u)d\tilde{W}(u)$. Likewise, for a portfolio, $D(t)X(t) = X(0) + \int_0^t \Delta(u)\sigma(u)D(u)S(u)d\tilde{W}(u)$. For these to be equal at all times (and hence to be a true replication/hedge), match terms to obtain the above expressions. \square

Remark: Note here that the existence of the martingale measure is allowing hedges to be created - this is key! Hedging perfectly in continuous time, however, is dubious in practice.

3.1.2 Multiple Assets

Definition. Under the **multidimensional market model**, consider m stocks and d Brownian motions. Define the stocks as

$$dS_i(t) = \mu_i(t)S_i(t)dt + S_i(t) \sum_{j=1}^d \sigma_{ij}(t)dW_j(t). \quad (3.10)$$

Rewriting as

$$\sigma_i(t) = \sqrt{\sum_{j=1}^d \sigma_{ij}(t)^2}, \quad dB_i(t) = \sum_{j=1}^d \frac{\sigma_{ij}(t)}{\sigma_i(t)} dW_j(t), \quad (3.11)$$

then $B_i(t)$ is Brownian via Lévy's theorem, and

$$dS_i(t) = \mu_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dB_i(t) \quad (3.12)$$

$$dB_i(t)dB_k(t) = \sum_{j=1}^d \frac{\sigma_{ij}(t)\sigma_{kj}(t)}{\sigma_i(t)\sigma_k(t)} dt = \rho_{ik}(t)dt, \quad (3.13)$$

with instantaneously correlated Brownian motions.

Proposition. Discounted stock prices under the multidimensional market model can be made martingales by change of measure if the following m market price of risk equations can be solved for $\Theta_i(t)$:

$$\mu_i(t) - R(t) = \sum_{j=1}^d \sigma_{ij}(t)\Theta_i(t). \quad (3.14)$$

If this is true, analogous to previous results, discounted portfolio values are also martingales under the risk-neutral measure.

Proof. Taking the differential of the discounted stock price via Itô's lemma yields $d(D(t)S_i(t)) = D(t)S_i(t)[(\mu_i(t) - R(t))dt + \sum_{j=1}^d \sigma_{ij}(t)dW_j(t)]$. In order to use Girsanov's theorem to switch into a risk-neutral probability measure where this becomes a martingale, the differential should be of the form $d(D(t)S_i(t)) = D(t)S_i(t)[\sum_{j=1}^d \sigma_{ij}(t)(\Theta_j(t)dt + dW_j(t))]$. This is possible if, by matching these two differentials, the above equations hold. \square

Definition. An **arbitrage** is a portfolio process $X(t)$ such that

$$X(0) = 0, \quad \mathbb{P}(X(T) \geq 0), \quad \mathbb{P}(X(T) > 0) > 0. \quad (3.15)$$

Theorem (First fundamental theorem of asset pricing). If a market model admits a risk-neutral measure, then it does not admit arbitrages.

Proof. Suppose there does exist a risk-neutral measure for a market (a collection of stocks). Then the discounted value of a portfolio is a martingale. Suppose $X(0) = 0$, then $\mathbb{E}[D(T)X(T)] = 0$. Suppose $X(t)$ also satisfies $\mathbb{P}(X(T) \geq 0) = 1$, so that $\mathbb{P}(X(T) < 0) = 0$. Then $\mathbb{P}(X(T) > 0) = 0$ must also be the case for the expectation to hold. Hence, $X(t)$ cannot be an arbitrage. \square

Theorem (Second fundamental theorem of asset pricing). For a market having a risk-neutral probability measure, the completeness of the market is equivalent to the uniqueness of the risk-neutral measure.

3.2 Change of Numéraire

Motivation. Asset prices do not have an intrinsic meaning in absolute terms, but only relative to a chosen unit of account, or numéraire. While risk-neutral pricing is often presented as a special change of probability measure under which discounted asset prices become martingales, this perspective generalizes naturally.

Definition. A **numéraire** is a strictly positive adapted process $N(t)$ such that $N(t) > 0$ almost surely for all $t \geq 0$, and such that N is a traded asset (or replicable portfolio). Given any asset price process $S(t)$, its **relative price** with respect to the numéraire N is defined as

$$S^{(N)}(t) = \frac{S(t)}{N(t)}. \quad (3.16)$$

Definition. Recall that in a frictionless arbitrage-free market, the asset pricing theorems imply the existence of an equivalent probability measure under which discounted asset prices are martingales.

Let $M(t)$ denote the money market account,

$$M(t) = \exp\left(\int_0^t R(u) du\right) \quad (3.17)$$

A probability measure \mathbb{Q} is called a **risk-neutral measure** or **martingale measure** if every traded asset $S(t)$ divided by the numéraire is a \mathbb{Q} -martingale:

$$\frac{S(t)}{N(t)} = \mathbb{E}^{\mathbb{Q}}\left[\frac{S(T)}{N(T)} \mid \mathcal{F}(t)\right]. \quad (3.18)$$

Plugging in the money market account reproduces the usual risk-neutral measure.

Theorem (Stochastic representation of assets). Consider the multidimensional market model of the previous section with m assets, and let $\tilde{\mathbb{P}}$ be the risk-neutral measure with respect to the money market account numéraire, with d -dimensional independent Brownian motion $\tilde{W}(t)$. Let $N(t)$ be a positive price process that does not pay dividends. Then there exists a vector volatility process

$$\nu(t) = (\nu_1(t), \dots, \nu_d(t)), \quad (3.19)$$

such that

$$dN(t) = R(t)N(t)dt + N(t)\nu(t) \cdot \tilde{W}(t), \quad (3.20)$$

or equivalently,

$$D(t)N(t) = N(0) \exp\left(\int_0^t \left(-\frac{1}{2}|\nu(u)|^2\right)du + \int_0^t \nu(u) \cdot d\tilde{W}(u)\right). \quad (3.21)$$

Proof. By construction of risk-neutral pricing, $D(t)N(t)$ is a martingale under the risk-neutral measure. By the martingale representation theorem, there exists a d -dimensional $\tilde{\Gamma}$ such that $d(D(t)N(t)) = \tilde{\Gamma}(t) \cdot d\tilde{W}(t)$. Defining $\nu_i(t) = \frac{\tilde{\Gamma}_i(t)}{D(t)N(t)}$ yields $d(D(t)N(t)) = D(t)N(t)\nu(t) \cdot d\tilde{W}(t)$. Using Itô with a good initial guess produces the above results. \square

Theorem (Change of numéraire). Let $S(t), N(t)$ be two asset prices in common currency with respective vector volatility processes. Denote this as

$$d(D(t)S(t)) = D(t)S(t)\sigma(t) \cdot \tilde{W}(t) \quad (3.22)$$

$$d(D(t)N(t)) = D(t)N(t)\nu(t) \cdot \tilde{W}(t). \quad (3.23)$$

Take $N(t)$ as the numéraire. Then there exists a risk-neutral measure $\tilde{\mathbb{P}}^{(N)}$ where the relative price $S^{(N)}(t)$ is a martingale. That relative price has differential

$$dS^{(N)}(t) = S^{(N)}(t)[\sigma(t) - \nu(t)] \cdot d\tilde{W}^{(N)}(t). \quad (3.24)$$

Proof. First, we establish the risk-neutral measure for the new numéraire. Consider the existing risk-neutral measure with money market account numéraire, and recall multidimensional Girsanov's theorem. Define $\Theta_j(t) = -\nu_j(t)$, such

that the Radon-Nikodym derivative becomes $Z(t) = \exp(\int_0^t \nu(u) \cdot d\widetilde{W}(u) - \int_0^t |\nu(u)|^2 du/2)$, and the new Brownian motion becomes $\widetilde{W}^{(N)}(t) = \widetilde{W}(t) - \int_0^t \nu(u) du$. By comparison with the results from the previous theorem, $Z(t) = D(t)N(t)/N(0)$.

Now, we evaluate the differential to show that the relative prices are martingales. Using the above theorem to obtain formulas for $D(t)S(t)$, $D(t)N(t)$, then dividing yields $S^{(N)}(t) = \frac{S(0)}{N(0)} \exp(-\frac{1}{2} \int_0^t (|\sigma(u)|^2 - |\nu(u)|^2) dt + \int_0^t (\sigma(u) - \nu(u)) \cdot d\widetilde{W}(u))$. To take the differential, apply Itô's formula, taking the exponent as the stochastic process. This reduces to $dS^{(N)}(t) = S^{(N)}(t)(\sigma(t) - \nu(t))(-\nu(t)dt + d\widetilde{W}(t))$. Substituting the previously proposed new Girsanov Brownian motion obtains the result. \square

Theorem. Let $V(T)$ be a contingent claim payable at time T . Its price at time t expressed in units of the numéraire N is given by

$$\frac{V(t)}{N(t)} = \mathbb{E}^{(\mathbb{Q}^N)} \left[\frac{X(T)}{N(T)} \middle| \mathcal{F}(t) \right]. \quad (3.25)$$

Example (Zero-coupon bond numéraire). Let $B(t, T)$ denote the time- t price of a zero-coupon bond maturing at time T . Consider a forward contract written at time t with maturity T on an asset with price $S(T)$.

Using the money market account as numéraire, the time- t price of the forward is

$$V(t) = \mathbb{E}^{(\mathbb{Q})} \left[e^{-\int_t^T r(u) du} (S(T) - K) \middle| \mathcal{F}_t \right]. \quad (3.26)$$

Now change numéraire to the zero-coupon bond $B(t, T)$, and let \mathbb{Q}^T denote the associated forward measure. Under \mathbb{Q}^T , the discounted asset price

$$\frac{S(t)}{B(t, T)} \quad (3.27)$$

is a martingale. The forward price is therefore given by

$$F(t, T) = \mathbb{E}^{(\mathbb{Q}^T)} [S(T) \mid \mathcal{F}_t], \quad (3.28)$$

as will be shown in the following chapters, and the drift term disappears under the T -forward measure.

Chapter 4

Asset Pricing and Derivatives

Motivation. Now that we have developed an asset pricing framework, how are real-world assets valued under the model? This section will develop prices for various derivatives using the risk-neutral model.

4.1 Forwards and Futures

Proposition (cash flows). For an asset generating cash flows $C(u)$, a portfolio holding this asset has value

$$X(t) = \frac{1}{D(t)} \tilde{\mathbb{E}} \left[\int_t^T D(u) dC(u) | \mathcal{F}(t) \right]. \quad (4.1)$$

Proof idea. The discounted portfolio has differential $d(D(t)X(t)) = D(t)dC(t)$. Hence, taking the integral of this from t to T , then applying the risk-neutral conditional expectation to exploit the martingale nature of the left-hand side reproduces the above expression.

Proposition. The price at time t of a **zero-coupon bond** paying 1 at time T is

$$B(t, T) = \frac{1}{D(t)} \tilde{\mathbb{E}}[D(T) | \mathcal{F}(t)]. \quad (4.2)$$

Define its **yield to maturity (YTM)** to be the rate equivalent to continuous compounding a single interest rate to maturity:

$$B(t, T) = e^{Y(t, T)(T-t)}. \quad (4.3)$$

Definition. A **forward contract** is an agreement to pay the amount K at delivery time T to receive an asset whose value at time t is $S(t)$. As the forward does not require upfront payment to enter into the contract, by no arbitrage, the price of the forward is the value of K that makes the forward worth zero at time t .

Theorem. Assuming bonds of all maturities are available, the price of a forward contract is

$$For_S(t, T) = \frac{S(t)}{B(t, T)}. \quad (4.4)$$

Proof. The value of a forward contract at time T is $S(T) - K$. Under the risk-neutral pricing system, the present value is $V(t) = \frac{1}{D(t)} \tilde{\mathbb{E}}[D(T)(S(T) - K) | \mathcal{F}(t)]$. Using linearity of expectation and the fact that discounted prices are martingales, we obtain $V(t) = S(t) - KB(t, T)$. Yet the value of the forward contract at time t must be zero, so setting K as the forward price above guarantees this. \square

Definition. A **futures contract** for asset $S(t)$ with delivery time T is an arrangement such that the contract requires no payment to enter into, the payments are marked-to-market via the establishment of a margin account such that holding a long position on the interval $[s, t]$ gives the holder $Fut_S(t, T) - Fut_S(s, T)$, and the holder at time T pays $Fut_S(T, T)$ to receive the asset.

Theorem. The price of a futures contract is

$$Fut_S(t, T) = \tilde{\mathbb{E}}[S(T) | \mathcal{F}(t)]. \quad (4.5)$$

This satisfies $Fut_S(T, T) = S(T)$ and is a martingale under the risk-neutral measure. The value of a futures contract held over a period of time is zero.

4.2 European Options

Definition. A **European call option** on underlying asset $S(t)$ gives the holder the right, but not obligation, to buy the asset at time T for price K . The value at time T of a European call is thus $\max(S(T) - K, 0)$. Likewise, a **European put option** on underlying asset $S(t)$ gives the holder the right, but not obligation, to sell the asset at time T for price K . The value at time T of a European put is thus $\max(K - S(T), 0)$.

Theorem (Black-Scholes). A European call option for asset $S(t)$, satisfying geometric Brownian motion with constant μ, σ , with expiry date T and strike price K , where the risk-free rate is r , has price $c(t, S(t))$ for $0 \leq t \leq T$. That price, under the no-arbitrage condition, satisfies the **Black-Scholes partial differential equation**:

$$\frac{\partial c}{\partial t}(t, x) + rx \frac{\partial c}{\partial x}(t, x) + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 c}{\partial x^2}(t, x) = rc(t, x). \quad (4.6)$$

Proof. Via the no-arbitrage condition, there exists a hedge to the call. This means that the hedge $\Delta(t)$ is chosen such that the discounted value of the portfolio is equal to the discounted value of the asset at all times, under no-arbitrage. Therefore,

$$d(e^{-rt}X(t)) = d(e^{-rt}c(t, S(t))). \quad (4.7)$$

Using Itô's lemma to evaluate both sides, and recalling that the stock in the portfolio obeys $dS(t) = \mu S(t)dt + \sigma S(t)dW(t)$, then the following is obtained by matching the differentials on both sides:

$$\Delta(t)(\mu - r)S(t)dt = \left[\frac{\partial c}{\partial t} + \mu S(t) \frac{\partial c}{\partial x} + \frac{1}{2} \sigma^2 S(t)^2 \frac{\partial^2 c}{\partial x^2} - rc \right] dt \quad (4.8)$$

$$\Delta(t)\sigma S(t)dW(t) = \sigma S(t) \frac{\partial c}{\partial x} dW(t). \quad (4.9)$$

Hence, the criteria for the delta-hedge is that

$$\Delta(t) = \frac{\partial c}{\partial x}(t, S(t)). \quad (4.10)$$

Substituting this into the first equation yields that $c(t, S(t))$ must satisfy the Black-Scholes PDE. \square

Theorem. The solution to the Black-Scholes PDE is

$$c(t, x) = x\Phi(d_+(\tau, x)) - e^{-r\tau}\Phi(d_-(\tau, x)) \quad (4.11)$$

$$d_-(\tau, x) = \frac{1}{\sigma\sqrt{\tau}} \left[\log\left(\frac{x}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau \right] \quad (4.12)$$

$$d_+(\tau, x) = d_-(\tau, x) + \sigma\sqrt{\tau}, \quad (4.13)$$

where $\tau = T - t$ and Φ is the CDF of the standard normal distribution.

Proof idea. Via the Feynman-Kac formula, the function that satisfies the Black-Scholes PDE is $c(t, x) = \tilde{\mathbb{E}}[e^{r(T-t)} \max(S(T) - K, 0) | \mathcal{F}(t)]$. Plugging in geometric Brownian motion for $S(t)$ and simplifying gives the above solution for the call price. Note that within the Feynman-Kac formula, the volatility does not need to be constant, and a version of Black-Scholes with non-constant volatility can be obtained.

Definition. The **Greeks** are derivatives of the option price function. **Delta**, **theta**, **vega**, and **gamma** are defined as

$$\Delta(t) = \frac{\partial c}{\partial x}(t, S(t)) \quad (4.14)$$

$$\Theta(t) = \frac{\partial c}{\partial t}(t, S(t)) \quad (4.15)$$

$$\mathcal{V}(t) = \frac{\partial c}{\partial \sigma}(t, S(t)) \quad (4.16)$$

$$\Gamma(t) = \frac{\partial^2 c}{\partial x^2}(t, S(t)). \quad (4.17)$$

Theorem (Put-call parity). For a forward, European call, and European put on asset $S(t)$ with expiry time T and strike price K ,

$$f(t, x) = S(t) - e^{rt}S(0) = c(t, S(t)) - p(t, S(t)). \quad (4.18)$$

Proof. At time T , the forward is worth $S(T) - K$, which is equal to $\max(S(T) - K, 0) - \max(K - S(T), 0)$, which is the difference in the call and put payoffs at time T . This must hold at all times for no arbitrage. \square

4.3 Exotic Options

Definition. European options depend only on the terminal value of the underlying asset. In contrast, **exotic options** are derivatives whose payoff depends on the path of the underlying process, on multiple assets, or on additional features such as barriers or averages.

4.3.1 Asian Options

Definition. An **Asian option** depends on previous sampled values of the stock. A fixed-strike Asian call, sampled continuously, pays

$$V(T) = \max\left(\frac{1}{T} \int_0^T S(u) du - K, 0\right). \quad (4.19)$$

Remark. For an exotic option, the underlying process $S(t)$, being an SDE solution, is Markov, yet the payoff depends on more than $S(T)$. In this case, it also depends on $A(T) = \int_0^T S(u) du$. One approach to recovering the Markov property is then to define a *joint process* $Y(t) = (S(t), A(t))$, enlarging the state space. Now the relevant history or "hidden memory" appears explicitly.

Theorem. For an Asian option, define the price function

$$u(t, s, a) = \mathbb{E}^{(\mathbb{Q})} \left[e^{-r(T-t)} h\left(\frac{A(T)}{T}\right) \mid S(t) = s, A(t) = a \right]. \quad (4.20)$$

Then u satisfies

$$\frac{\partial u}{\partial t} + rs \frac{\partial u}{\partial s} + s \frac{\partial u}{\partial a} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 u}{\partial s^2} - ru = 0, \quad (4.21)$$

with terminal condition

$$u(T, s, a) = h\left(\frac{a}{T}\right). \quad (4.22)$$

Proof. Define the auxiliary process $A(t) = \int_0^t S(u) du$. Then the payoff may be written solely in terms of $A(T)$ as $H = h(A(T)/T)$. The joint process

$(S(t), A(t))$ contains all information relevant for pricing the Asian option. Assume that under the risk-neutral measure \mathbb{Q} , the asset price follows $dS(t) = rS(t)dt + \sigma S(t)dW(t)$, where r and σ are constants. The process $A(t)$ then satisfies $dA(t) = S(t)dt$, with $A(0) = 0$. This defines the dynamics of the joint process.

Moreover, the process $(S(t), A(t))$ is Markovian. For $h > 0$, $A(t+h) = A(t) + \int_t^{t+h} S(u)du$. The future increment of A depends only on the future path of S , whose conditional distribution depends only on $S(t)$. Consequently, conditional on $(S(t), A(t))$, the future evolution of the process is independent of the past. Hence, $(S(t), A(t))$ is a time-homogeneous Markov process. This means that we can now use the Kolmogorov/generator formalism.

Now, let $f \in C^2(\mathbb{R}^2)$ be a smooth test function. Applying Itô's formula to $f(S(t), A(t))$ yields $df = \frac{\partial f}{\partial s} dS + \frac{\partial f}{\partial a} dA + \frac{1}{2} \frac{\partial^2 f}{\partial s^2} (dS)^2$. Solving for the coefficient on the dt term, the generator \mathcal{L} of the joint process $(S(t), A(t))$ is therefore:

$$\mathcal{L}f(s, a) = rs \frac{\partial f}{\partial s} + s \frac{\partial f}{\partial a} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 f}{\partial s^2}. \quad (4.23)$$

Define the Asian option price function

$$u(t, s, a) = \mathbb{E}^{(\mathbb{Q})} \left[e^{-r(T-t)} h \left(\frac{A(T)}{T} \right) \middle| S(t) = s, A(t) = a \right]. \quad (4.24)$$

Since $(S(t), A(t))$ is Markovian, u depends only on (t, s, a) . By the Feynman-Kac theorem, u satisfies the backward Kolmogorov equation,

$$\frac{\partial u}{\partial t} + \mathcal{L}u - ru = 0, \quad (4.25)$$

the expansion of which is given in the theorem. \square

4.3.2 Barrier and Lookback Options

Definition. A **barrier option** known as a knock-out option is a standard European option, with the condition that if the asset crosses a barrier, the option is worthless. A knock-in option is a standard European option, with the condition that the option is worthless unless the asset crosses a barrier. A knock-out option where the barrier is above the current asset price is called "up-and-out".

Proposition. Let $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ be a probability space and $\tilde{W}(t)$ a Brownian motion. Define

$$\hat{W}(t) = \alpha t + \tilde{W}(t), \quad \hat{M}(t) = \max_{0 \leq t \leq T} \hat{W}(t). \quad (4.26)$$

Then, the joint density under $\tilde{\mathbb{P}}$ of $(\hat{M}(t), \hat{W}(t))$ is

$$\tilde{f}_{(\hat{M}(t), \hat{W}(t))}(m, w) = \frac{2(2m - w)}{T\sqrt{2\pi T}} \exp\left(\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2m - w)^2\right). \quad (4.27)$$

Proof idea. Use Girsanov to switch into the hat measure, then the joint density is a standard result. Then, use the trick that probability is equal to expectation of an indicator function, and explicitly write the Radon-Nikodym derivative. Evaluating this gives the above expression.

Theorem. An up-and-out call price $v(t, S(t))$ satisfies the Black-Scholes equation in the region $([0, T] \times [0, B])$

$$\frac{\partial v}{\partial t}(t, x) + rx \frac{\partial v}{\partial x}(t, x) + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 v}{\partial x^2}(t, x) = rv(t, x), \quad (4.28)$$

subject to

$$v(t, 0) = v(t, B) = 0, \quad v(T, x) = \max(x - K, 0). \quad (4.29)$$

The risk-neutral expectation can also be taken given the above joint density.

Definition. A **lookback option**, which depends on the maximum of the asset price attained, called a floating strike lookback call pays, for the maximum attained,

$$Y(t) = \max_{0 \leq u \leq t} S(u), \quad (4.30)$$

the following:

$$v(T) = Y(T) - S(T). \quad (4.31)$$

Theorem. A floating strike lookback call price $v(t, S(t), Y(t))$ satisfies Black-Scholes:

$$\frac{\partial v}{\partial t}(t, x, y) + rx \frac{\partial v}{\partial x}(t, x, y) + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 v}{\partial x^2}(t, x, y) = rv(t, x, y), \quad (4.32)$$

subject to

$$v(t, 0, y) = e^{-r(T-t)} y, \quad v(t, y, y) = 0, \quad v(T, x, y) = y - x. \quad (4.33)$$

4.4 American Options

Definition. Let $S(t)$ be an asset whose price under the risk-neutral measure \mathbb{Q} satisfies

$$dS(t) = rS(t) dt + \sigma S(t) d\widetilde{W}(t), \quad (4.34)$$

where $r > 0$ is the constant risk-free rate and $\widetilde{W}(t)$ is a Brownian motion under \mathbb{Q} .

An **American option** written on $S(t)$, with strike price K and expiration date T , differs from a European option in that it may be exercised at any stopping

time prior to maturity. The holder of the option chooses an exercise time so as to maximize the expected discounted payoff.

Formally, the value of an American option with payoff function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is given by the **optimal stopping problem**

$$v(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{(\mathbb{Q})} \left[e^{-r(\tau-t)} h(S(\tau)) \mid S(t) = x \right], \quad (4.35)$$

where $\mathcal{T}_{t,T}$ denotes the set of stopping times taking values in $[t, T]$.

Proposition. The American option pricing problem is an instance of optimal stopping theory. The value function $v(t, x)$ satisfies

- $v(t, x) \geq h(x)$ for all (t, x) , since immediate exercise is always allowed,
- $e^{-rt}v(t, S(t))$ is the smallest supermartingale dominating the discounted payoff process $e^{-rt}h(S(t))$.

This characterization is known as the **Snell envelope** and provides the probabilistic foundation for the PDE formulation of American option pricing.

Theorem (Variational Inequality). Let \mathcal{L} denote the generator of the Markov process $S(t)$,

$$\mathcal{L}f(x) = rx \frac{\partial f}{\partial x}(x) + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 f}{\partial x^2}(x). \quad (4.36)$$

The value function $v(t, x)$ of an American option satisfies the **free boundary problem**

$$\max \left(\frac{\partial v}{\partial t}(t, x) + \mathcal{L}v(t, x) - rv(t, x), h(x) - v(t, x) \right) = 0, \quad (4.37)$$

with terminal condition

$$v(T, x) = h(x). \quad (4.38)$$

The region

$$\mathcal{E} = \{(t, x) : v(t, x) = h(x)\} \quad (4.39)$$

is called the **exercise region**, while

$$\mathcal{C} = \{(t, x) : v(t, x) > h(x)\} \quad (4.40)$$

is the **continuation region**. The boundary separating these regions is an unknown, time-dependent **free boundary**.

Theorem. For an American put option, the payoff is

$$h(x) = \max(K - x, 0). \quad (4.41)$$

Early exercise may be optimal, and the value function satisfies the following **linear complementarity conditions** for $x \geq 0$ and $t \in [0, T]$:

$$rv(t, x) - \frac{\partial v}{\partial t}(t, x) - rx \frac{\partial v}{\partial x}(t, x) - \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 v}{\partial x^2}(t, x) \geq 0, \quad (4.42)$$

$$v(t, x) \geq K - x, \quad (4.43)$$

$$\left(rv(t, x) - \frac{\partial v}{\partial t}(t, x) - rx \frac{\partial v}{\partial x}(t, x) - \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 v}{\partial x^2}(t, x) \right) (v(t, x) - (K - x)) = 0. \quad (4.44)$$

There exists a free boundary $x = b(t)$ such that

- if $x \leq b(t)$, it is optimal to exercise immediately,
- if $x > b(t)$, it is optimal to continue holding the option.

At the free boundary, the value function satisfies the *value-matching* and *smooth-pasting* conditions:

$$v(t, b(t)) = K - b(t), \quad (4.45)$$

$$\frac{\partial v}{\partial x}(t, b(t)) = -1. \quad (4.46)$$

Theorem. If the underlying asset pays no dividends, the price of an American call option equals the price of the corresponding European call option with the same strike and maturity:

$$v(t, x) = c(t, x). \quad (4.47)$$

Remark. In this case, early exercise is never optimal. Consequently, the continuation region coincides with the entire state space, and the American call option satisfies the standard Black–Scholes PDE without a free boundary.

4.5 Local and Stochastic Volatility

Motivation. Using the Black–Scholes model for European options, one can plug in the call prices at different strikes and expiry times to back out an “implied volatility”. Black–Scholes assumes that this is constant, however it varies with strike and expiry in what are known as *volatility smiles/skews*. The following models are attempts to model volatility more dynamically.

4.5.1 Dupire Model

Definition. Market option prices imply a surface of implied volatilities $\sigma_{\text{imp}}(K, T)$ across strikes K and maturities T . A natural question is whether there exists a diffusion process for the underlying asset whose model prices exactly reproduce this surface.

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a risk-neutral probability space. A **local volatility model** assumes that the asset price $S(t)$ follows

$$dS(t) = r(t)S(t) dt + \sigma_{\text{loc}}(t, S(t)) S(t) dW(t), \quad (4.48)$$

where $\sigma_{\text{loc}}(t, s)$ is a deterministic function called the **local volatility**. This model is Markovian, time-inhomogeneous, and arbitrage-free under \mathbb{Q} .

Proposition. Let $c(t, T)$ denote the price at time t of a European call option with strike K and maturity T . Under mild regularity assumptions, the call price satisfies the forward PDE:

$$\frac{\partial c}{\partial T} = -r(T)K \frac{\partial c}{\partial K} + \frac{1}{2} \sigma_{\text{loc}}^2(T, K) K^2 \frac{\partial^2 c}{\partial K^2} \quad (4.49)$$

This equation is the forward Kolmogorov equation expressed in strike space.

Proof idea. Take the forward Kolmogorov equation with the generator evaluated for this particular $S(t)$.

Theorem (Dupire's formula). Assume that European call prices $c(0, T; K)$ are observed for all strikes and maturities and are sufficiently smooth. Then the local volatility function consistent with these prices is given by

$$\sigma_{\text{loc}}^2(T, K) = \frac{\frac{\partial c}{\partial T}(0, T; K) + r(T)K \frac{\partial c}{\partial K}(0, T; K)}{\frac{1}{2} K^2 \frac{\partial^2 c}{\partial K^2}(0, T; K)}. \quad (4.50)$$

Proof. Rearrange the forward equation. □

Definition. The **stochastic volatility inspired (SVI) model** allows for fitting implied volatility as a function of time to expiry and strike price. Define total implied variance as $w(\tau, k) = \sigma_{BS}^2(\tau, k)^\tau$ for k the logarithm of the strike price, and τ the time to expiry. The raw SVI model is then, for a given expiry time,

$$w(k) = a + b(\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}), \quad (4.51)$$

where parameters can be fit. The model is free of calendar arbitrage if its time derivative is nowhere negative.

Remark. Dupire's formula shows that the local volatility surface can be recovered uniquely from the observed implied volatility surface, provided sufficient smoothness. Local volatility models exactly fit the observed implied volatility surface, remain Markovian and one-dimensional, and admit efficient PDE and Monte Carlo pricing methods. However, their volatility dynamics are deterministic, they exhibit poor performance for forward-start and path-dependent options, and they are unable to reproduce realistic volatility dynamics.

4.5.2 Heston Model

Definition. Empirical evidence suggests that volatility exhibits mean reversion, clustering, and randomness independent of asset price movements. Stochastic volatility models address these features by introducing a second source of randomness.

The **Heston model** specifies the joint dynamics of the asset price $S(t)$ and its instantaneous variance $V(t)$ under the risk-neutral measure:

$$dS(t) = rS(t) dt + \sqrt{V(t)} S(t) dW_1(t) \quad (4.52)$$

$$dV(t) = \kappa(\theta - V(t)) dt + \xi \sqrt{V(t)} dW_2(t), \quad (4.53)$$

where $\kappa > 0$ is the rate of mean reversion, $\theta > 0$ is the long-run variance, $\xi > 0$ is the volatility of volatility, and W_1 and W_2 are Brownian motions with correlation ρ .

Proposition. The joint process $(S(t), V(t))$ is a two-dimensional Markov diffusion with generator

$$\mathcal{L}f = rs \frac{\partial f}{\partial s} + \kappa(\theta - v) \frac{\partial f}{\partial v} + \frac{1}{2}vs^2 \frac{\partial^2 f}{\partial s^2} + \rho\xi vs \frac{\partial^2 f}{\partial s \partial v} + \frac{1}{2}\xi^2 v \frac{\partial^2 f}{\partial v^2} \quad (4.54)$$

Proof idea. Use the multidimensional Itô formula on the joint process and read off the coefficient on the dt term.

Theorem. Let $u(t, s, v)$ denote the price of a European derivative with payoff $h(S(T))$. Then u satisfies

$$\frac{\partial u}{\partial t} + \mathcal{L}u - ru = 0, \quad u(T, s, v) = h(s). \quad (4.55)$$

Proof. Given the above generator, this is the Kolmogorov backward equation. \square

Theorem. To ensure strict positivity of the variance process, the parameters must satisfy the **Feller condition**

$$2\kappa\theta \geq \xi^2. \quad (4.56)$$

This guarantees that the boundary $V(t) = 0$ is unattainable.

Remark. Explicitly evaluating the risk-neutral expectation to calculate asset prices will not be successful, as there is no closed-form probability density for the Heston model. There are still many ways to price assets using the Heston model. One popular one relies on the characteristic function of the logarithm of the asset price, which does have a closed form, and the Fourier transform, which can be performed quickly using the Fast Fourier Transform.

Proposition (Fourier pricing). Let $S(t)$ be an asset price over time, and define $X(t) = \log S(t)$. Let the **characteristic function** of $X(t)$ be defined as

$$\varphi(u) = \mathbb{E}^{(\mathbb{Q})}[e^{iuX}]. \quad (4.57)$$

If ϕ denotes the payoff of a derivative as a function of the logarithm of the asset price and admits a Fourier representation,

$$\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ipx} \hat{\phi}(p) dp, \quad (4.58)$$

then the initial price of the derivative is

$$V(0) = e^{-rT} \mathbb{E}^{(\mathbb{Q})}[\phi(X(T))] = \frac{e^{-rT}}{2\pi} \int_{-\infty}^{\infty} \varphi(-p) \hat{\phi}(p) dp. \quad (4.59)$$

Theorem. For $X(t)$, the logarithm of a Heston model asset price, the characteristic function of $X(T)$ is

$$\varphi_{X(T)}(u) = \exp(C(T, u) + D(T, u)V(0) + iuX(0)), \quad (4.60)$$

for C, D solutions to Riccati differential equations (discussed in next section).

Proof idea. Guess a solution similar in form to the above function and plug it into the Kolmogorov backwards equation, then solve the ODEs.

Remark. The payout of a European call is not absolutely integrable, hence there is no Fourier transform for it. Nevertheless, the **Carr-Madan approach** dampens the payoff by an exponential so that the Fourier transform exists. This allows standard options to be computed using the Heston model.

Example (Carr-Madan). Consider a European call option with strike K and maturity T on an asset $S(t)$ following the Heston model.

Define

$$X(T) = \log S(T), \quad k = \log K. \quad (4.61)$$

Introduce a damping factor $\alpha \in (0, 1)$ and define

$$C_\alpha(K) := e^{-\alpha k} \mathbb{E}^{(\mathbb{Q})}[(e^{X(T)} - e^k)^+]. \quad (4.62)$$

Using the Carr-Madan approach, the call price can be written as

$$C(K) = e^{-\alpha k} \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[e^{-ivk} \frac{\varphi_{X(T)}(v - i(\alpha + 1))}{(\alpha + iv)(\alpha + 1 + iv)} \right] dv, \quad (4.63)$$

where $\varphi_{X(T)}(u) = \mathbb{E}[e^{iuX(T)}]$ is the characteristic function of the log-price under the Heston model.

From the previous theorem, we have

$$\varphi_{X(T)}(u) = \exp \left(C(T, u) + D(T, u)V(0) + iuX(0) \right), \quad (4.64)$$

where $C(T, u)$ and $D(T, u)$ solve the Riccati ODEs associated with the Heston model. Compute the integral to solve for the price.

Remark: This method allows efficient computation of European option prices under stochastic volatility models. The damping factor α ensures integrability, and the integral can be evaluated using standard quadrature methods or the Fast Fourier Transform (FFT). The same approach can be extended to other European-style payoffs.

4.5.3 SABR Model

Definition. Let $F(t)$ denote a forward price or forward interest rate. Under a pricing measure, the **SABR (Stochastic Alpha Beta Rho) model** is defined by the system of stochastic differential equations

$$dF(t) = \alpha_t F(t)^\beta dW_1(t), \quad (4.65)$$

$$d\alpha(t) = \nu \alpha(t) dW_2(t), \quad (4.66)$$

$$d\langle W_1, W_2 \rangle(t) = \rho dt, \quad (4.67)$$

where:

- $\alpha_t > 0$ is the instantaneous volatility level,
- $\beta \in [0, 1]$ controls the elasticity of the forward process,
- $\nu \geq 0$ is the volatility of volatility,
- $\rho \in [-1, 1]$ is the instantaneous correlation.

Proposition. The forward price $F(t)$ is a martingale under the chosen measure. The SABR model therefore preserves no-arbitrage conditions for forwards and rates.

Remark. Several classical models arise as special cases:

- $\beta = 1$: lognormal SABR, consistent with Black's model,
- $\beta = 0$: normal SABR, consistent with Bachelier's model,
- $\nu = 0$: constant elasticity of variance (CEV) model,
- $\rho = 0$: uncorrelated volatility and underlying.

The implied volatility smile in SABR arises from stochastic volatility ($\nu > 0$), correlation between price and volatility ($\rho \neq 0$), and nonlinear dependence of volatility on the forward level ($\beta \neq 1$). In particular, negative correlation ($\rho < 0$) produces left-skewed smiles commonly observed in interest rate markets.

Theorem (Hagan et al. approximation). Closed-form option prices are unavailable in the SABR model. However, an asymptotic expansion for implied volatility is available. Let K denote the option strike and T the maturity. For

$F(0), K > 0$ and small T , the Black implied volatility under the SABR model admits the approximation

$$\sigma(F(0), K) = \frac{z\alpha_0}{\chi(z)(F(0)K)^{(1-\beta)/2}} \left[1 + \left(\frac{\alpha_0^2(1-\beta)^2}{24(F(0)K)^{1-\beta}} + \frac{\rho\beta\nu\alpha_0}{4(F(0)K)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2 \right) T \right], \quad (4.68)$$

where

$$z = \frac{\nu}{\alpha_0} (F(0)K)^{(1-\beta)/2} \log\left(\frac{F(0)}{K}\right), \quad (4.69)$$

$$\chi(z) = \log\left(\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right). \quad (4.70)$$

Remark. This formula is used almost universally for calibration in interest rate markets.

Corollary. At-the-money ($F(0) = K$), the implied volatility simplifies to

$$\sigma_{\text{ATM}} = \frac{\alpha_0}{F(0)^{1-\beta}} \left[1 + \left(\frac{(1-\beta)^2}{24} \frac{\alpha_0^2}{F(0)^{2-2\beta}} + \frac{\rho\beta\nu\alpha_0}{4F(0)^{1-\beta}} + \frac{2-3\rho^2}{24}\nu^2 \right) T \right]. \quad (4.71)$$

Chapter 5

Fixed-Income and Credit

5.1 Term-Structure Models

Motivation. The market does not have one single interest rate, nor does that interest rate stay constant over time. This section addresses models attempting to reflect that.

5.1.1 Short Rate Models

Definition. The **short rate** $R(t)$ is the instantaneous yield/interest rate at time t

$$R(t) = \lim_{T \rightarrow t} -\frac{\partial}{\partial T} \log(B(t, T)). \quad (5.1)$$

Definition. The **Vasicek model** assumes that the short rate $R(t)$ follows an Ornstein-Uhlenbeck process:

$$dR(t) = (\alpha - \beta R(t)) dt + \sigma d\widetilde{W}(t). \quad (5.2)$$

Theorem. The rate under a Vasicek process is

$$R(t) = e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}) + \sigma e^{-\beta t} \int_0^t e^{-\beta s} d\widetilde{W}(s), \quad (5.3)$$

with mean $e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t})$ and variance $\frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$.

Proof. This could be immediately integrated, but the rate appears on the right side. To eliminate this, consider the differential $d(e^{\beta t} R(t)) = e^{\beta t} dR(t) + \beta e^{\beta t} R(t) dt$. Substituting the Vasicek differential gives $d(e^{\beta t} R(t)) = \alpha e^{\beta t} dt + e^{\beta t} \sigma d\widetilde{W}(t)$. Integrating this gives the above equation. The only random part of the rate is the Itô integral.

Since this is a martingale, and Brownian motion starts at zero, this has mean zero, so the mean is the constant terms. Using Itô isometry, the variance is $\sigma^2 e^{-2\beta t} \int_0^t e^{2\beta s} ds$, which gives the above result. \square

Theorem. The price of a zero-coupon bond of maturity T , $B(t, r; T)$, under the Vasicek model satisfies

$$\frac{\partial B}{\partial t}(t, r; T) + (\alpha - \beta r) \frac{\partial B}{\partial r}(t, r; T) + \frac{1}{2} \sigma^2 \frac{\partial^2 B}{\partial r^2}(t, r; T) - rB(t, r; T) = 0 \quad (5.4)$$

$$B(T, r; T) = 1. \quad (5.5)$$

The Vasicek model is an affine term structure model, and zero-coupon bonds under them have prices of the form

$$B(t, r; T) = e^{g(t, T) + h(t, T)r}. \quad (5.6)$$

Proof. The value of a zero coupon bond under the standard risk-neutral measure is $D(t)B(t, T) = \mathbb{E}[D(T)1|\mathcal{F}(s)]$, so $B(t, T) = \mathbb{E}[e^{\int_t^T R(u)du}|\mathcal{F}(t)]$. Using the standard Kolmogorov/generator framework with the SDE being the Vasicek process $\mathcal{L}f = (\alpha - \beta r) \frac{\partial f}{\partial r} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial r^2}$ yields the above result. The generator satisfies the affine condition $\mathcal{L}e^{ur} = (\alpha - \beta r)ue^{ur} + \frac{1}{2} \sigma^2 u^2 e^{ur} = [(\alpha u + \frac{1}{2} \sigma^2 u^2) + (\beta u)r]e^{ur}$. This affine form can serve as an ansatz for a closed-solution from Riccati ODEs. \square

Proposition. If, under the risk-neutral measure, a short rate model has the form

$$dR(t) = a(t, R(t))dt + b(t, R(t))d\widetilde{W}(t), \quad (5.7)$$

for

$$a(t, R(t)) = \kappa(t)R(t) + \eta(t), \quad b(t, R(t)) = \sqrt{\gamma(t)R(t) + \delta(t)}, \quad (5.8)$$

then the model is affine, and the functions in the exponent of the bond price satisfy the **Riccati differential equations**

$$\frac{dh}{dt}(t, T) = -\kappa(t)h(t, T) - \frac{1}{2} \gamma(t)h^2(t, T) + 1 \quad (5.9)$$

$$\frac{dg}{dt}(t, T) = -\eta(t)h(t, T) - \frac{1}{2} \delta(t)h^2(t, T), \quad (5.10)$$

with $g(T, T) = h(T, T) = 0$.

Remark. This model allows for negative interest rates.

Definition. The **Cox-Ingersoll-Ross (CIR) model** modifies the diffusion term to preserve positivity:

$$dR(t) = \kappa(\theta - R(t))dt + \sigma\sqrt{R(t)}d\widetilde{W}(t). \quad (5.11)$$

Under the Feller condition $2\kappa\theta \geq \sigma^2$, the short rate remains strictly positive. Like Vasicek, the CIR model is affine and yields closed-form bond prices, but introduces state-dependent volatility.

Definition. The **Hull-White model** extends the Vasicek framework by allowing time-dependent parameters to fit the initial yield curve exactly:

$$dR(t) = (\alpha(t) - \beta(t)R(t))dt + \sigma(t)d\widetilde{W}(t). \quad (5.12)$$

This produces yet another affine, Gaussian model.

Proposition. Given a simplified Hull-White model,

$$dR(t) = (\theta(t) - aR(t))dt + \sigma d\widetilde{W}(t), \quad (5.13)$$

the model can be calibrated given the present forward rate $f^{mkt}(t, T)$:

$$\theta(t) = \frac{\partial f^{mkt}}{\partial t}(0, t) + af^{mkt}(0, t) + \frac{\sigma^2}{2a}(1 - e^{-at})^2. \quad (5.14)$$

a, σ can then be calibrated.

Remark. There likewise exist **multifactor models** that consider the differential of the short rate to depend on multiple processes.

5.1.2 Forward Rate Models

Definition. The **forward rate** at time t over the interval $[T, T + \delta T]$ is

$$e^{f(t, T, T+\delta T)\delta T} B(t, T) = B(t, T + \delta T). \quad (5.15)$$

Limiting δt to zero gives the instantaneous interest rate at time T that can be locked in at time t :

$$f(t, T) = -\frac{\partial}{\partial T} \log(B(t, T)). \quad (5.16)$$

Consequently, the short rate is $f(t, t)$, and

$$B(t, T) = e^{-\int_t^T f(t, v)dv}. \quad (5.17)$$

Definition. The **Heath-Jarrow-Morton (HJM) framework** models the entire forward rate curve directly. Let $f(t, T)$ denote the instantaneous forward rate at time t for maturity T . Under the risk-neutral measure, its dynamics are

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW(t), \quad (5.18)$$

with $f(0, T)$ being set to the initial observed forwards rate curve.

Theorem. Absence of arbitrage requires that there must a process such that

$$\alpha(t, T) = \sigma(t, T)(\sigma^*(t, T) + \Theta(t)), \quad (5.19)$$

for

$$\sigma^*(t, T) = \int_t^T \sigma(t, u) du. \quad (5.20)$$

Proof. $d(D(t)B(t, T))$ should be a martingale for no-arbitrage to hold. This evaluates to $-R(t)D(t)B(t, T)dt + D(t)dB(t, T)$. Evaluating $dB(t, T)$ using Itô's lemma requires considering the exponential as the function, as the exponent as the stochastic process, whose differential is $d(-\int_t^T f(t, v)dv) = f(t, t)dt - \int_t^T df(t, v)dv$.

Evaluating this simplifies the original differential to $D(t)B(t, T)[(-\alpha^*(t, T) + \frac{1}{2}\sigma^*(t, T)^2)dt - \sigma^*(t, T)dW(t)]$, where the stars represent the quantity time integrated in the second argument over $[t, T]$, written explicitly in the statement. To make this a martingale, this must be able to be made into the form $-D(t)B(t, T)\sigma^*(t, T)[\Theta(t)dt + dW(t)]$, so as to use Girsanov's theorem to switch into a risk neutral measure. The characterization of this random process is concretized in the statement above. The fact that this market price of risk is unique means that there is a unique risk-neutral measure, so the model is also complete. \square

Theorem. Assuming that the no-arbitrage condition is satisfied, forwards rates evolve as

$$df(t, T) = \sigma(t, T)\sigma^*(t, T)dt + \sigma(t, T)d\widetilde{W}(t), \quad (5.21)$$

and zero-coupon bond prices follow

$$B(t, T) = \frac{B(0, T)}{D(t)} \exp[-\int_0^t \sigma^*(u, T)d\widetilde{W}(u) - \frac{1}{2}\int_0^t \sigma^*(u, T)^2 du]. \quad (5.22)$$

Remark. This model can reproduce the Hull-White model. Unfortunately, if this formula is used to try to apply Black-Scholes to equity options, forwards rates explode.

5.1.3 Market Models

Motivation. Historically, interest rate derivatives were quoted, hedged, and settled in terms of forward-looking LIBOR (London Inter-Bank Offered Rate) rates rather than instantaneous short rates. Products such as caps, floors, and swaptions reference discretely compounded forward rates over fixed accrual periods, motivating models that directly specify the dynamics of these observable quantities.

Unlike short-rate models, which infer forward rates indirectly through bond prices, the LIBOR Market Model (LMM) specifies the joint dynamics of a finite set of forward rates consistent with market conventions.

Definition. Fix a discrete **tenor**

$$0 = T_0 < T_1 < \dots < T_N, \quad (5.23)$$

with accrual periods $\delta_i = T_{i+1} - T_i$. The **forward LIBOR rate** for the period $[T_i, T_{i+1}]$, observed at time $t \leq T_i$, is defined by

$$L_i(t) = \frac{1}{\delta_i} \left(\frac{B(t, T_i)}{B(t, T_{i+1})} - 1 \right), \quad (5.24)$$

where $B(t, T)$ denotes the time- t price of a zero-coupon bond maturing at T .

Remark. LIBOR rates are discretely compounded and are defined only on the tenor dates, in contrast to continuously compounded instantaneous forward rates.

Proposition. For each maturity T_{i+1} , define the T_{i+1} -**forward measure** $\mathbb{Q}^{(T_{i+1})}$ via the numéraire $B(t, T_{i+1})$. Under the measure $\mathbb{Q}^{(T_{i+1})}$, the forward LIBOR rate $L_i(t)$ is a martingale.

Proof Sketch. Since $L_i(t)$ is a tradable payoff at T_{i+1} expressed in units of the numéraire $B(t, T_{i+1})$, absence of arbitrage implies the martingale property under $\mathbb{Q}^{(T_{i+1})}$.

Assumptions. The **LIBOR Market Model (LMM)** proposes the following assumptions:

Under its associated forward measure $\mathbb{Q}^{(T_{i+1})}$, the LIBOR rate follows

$$dL_i(t) = \sigma_i(t) L_i(t) d\widetilde{W}_i^{(T_{i+1})}(t), \quad t \leq T_i, \quad (5.25)$$

where $\sigma_i(t)$ is a deterministic volatility function.

The Brownian motions satisfy

$$d\widetilde{W}_i^{(T_{i+1})}(t) d\widetilde{W}_j^{(T_{i+1})}(t) = \rho_{ij} dt, \quad (5.26)$$

with a positive semidefinite correlation matrix (ρ_{ij}) .

Proposition. When expressed under a single measure (e.g. the terminal measure $\mathbb{Q}^{(T_N)}$), the dynamics of $L_i(t)$ acquire a drift:

$$dL_i(t) = \mu_i(t) dt + \sigma_i(t) L_i(t) d\widetilde{W}_i^{(T_N)}(t), \quad (5.27)$$

where the drift depends on all subsequent LIBOR rates:

$$\mu_i(t) = -\sigma_i(t) L_i(t) \sum_{j=i+1}^{N-1} \frac{\delta_j L_j(t) \sigma_j(t) \rho_{ij}}{1 + \delta_j L_j(t)}. \quad (5.28)$$

Remark. This coupling is the principal source of analytical complexity in the LMM. While each LIBOR is lognormal under its own measure, the joint system is not jointly lognormal under a single measure.

Definition. A **caplet** on the period $[T_i, T_{i+1}]$ with strike K pays

$$\max(\delta_i(L_i(T_i) - K), 0). \quad (5.29)$$

Proposition. Under the T_{i+1} -forward measure, caplets price as

$$V_{cap}(t) = B(t, T_{i+1}) \tilde{\mathbb{E}}^{(T_{i+1})}[\max(\delta_i(L_i(T_i) - K), 0) \mid \mathcal{F}(t)]. \quad (5.30)$$

Since L_i is lognormal under $\mathbb{Q}^{(T_{i+1})}$, the price is given by the Black formula. Caplets price cleanly because their payoff depends on a *single* LIBOR rate, which is lognormal under the appropriate forward measure.

Definition. The forward **swap rate** for a swap starting at T_α and ending at T_β is

$$S_{\alpha,\beta}(t) = \frac{B(t, T_\alpha) - B(t, T_\beta)}{\sum_{k=\alpha}^{\beta-1} \delta_k B(t, T_{k+1})}. \quad (5.31)$$

The swap rate is a nonlinear function of multiple LIBOR rates. There exists no measure under which it is exactly lognormal.

Procedure. To calibrate the LMM:

- Calibrate individual $\sigma_i(t)$ to caplet implied volatilities.
- Fit the correlation matrix to swaption prices.
- Validate joint dynamics via Monte Carlo pricing of exotic payoffs.

Remark (Post-LIBOR). Although LIBOR has been discontinued, the modeling framework survives in adapted forms (e.g. for SOFR-based forward rates), and the LMM remains foundational for understanding multi-rate interest rate modeling.

5.2 Credit Risk Modeling

Motivation. It is possible that a counterparty fails to meet its contractual obligations. How should parties factor that risk into their prices? The central object of interest is thus the random default time, whose distribution determines survival probabilities, credit spreads, and the valuation of credit-sensitive instruments, which will be investigated in this section.

5.2.1 Structural Credit Risk Models

Motivation. Structural models view default as the outcome of the firm's asset value deteriorating relative to its liabilities. These models provide strong economic intuition and link credit risk to equity dynamics.

Theorem. Let the firm asset process (A_t) follow a geometric Brownian motion

$$dA(t) = \mu A(t) dt + \sigma A(t) dW_t. \quad (5.32)$$

Default occurs at maturity T if

$$A(T) < D, \quad (5.33)$$

where D denotes the face value of the firm's debt. The **distance to default** is defined by

$$\text{DD} = \frac{\log(A(0)/D) + (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \quad (5.34)$$

Under the **Merton model**, the probability of default is

$$\mathbb{P}(A(T) < D) = \Phi(-\text{DD}), \quad (5.35)$$

where Φ denotes the standard normal cumulative distribution function.

Remark. The Merton model implies that default can only occur at maturity and typically underestimates short-term credit spreads. Furthermore, credit spreads are strongly tied to equity volatility, which is inconsistent with observed market behavior for distressed firms.

Definition. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space. A **default time** τ is a non-negative random variable representing the time at which default occurs. In barrier models such as **Black-Cox**, **barrier default** occurs at the stopping time

$$\tau = \inf\{t \geq 0 : A(t) \leq K(t)\}, \quad (5.36)$$

where $K(t)$ is a deterministic default barrier. This becomes a first hitting time problem for Brownian motion.

Remark. Barrier models allow for early default and improve short-term spread behavior relative to the Merton model, but remain sensitive to equity dynamics and the specification of the default barrier.

5.2.2 Reduced-Form Credit Models

Reduced-form models treat default as an exogenous jump event and are the dominant framework for pricing and calibration in practice.

Definition. The survival process is defined by a **survival indicator**

$$\mathbb{I}_{\{\tau > t\}}. \quad (5.37)$$

The **survival probability** from time t to maturity T is

$$Q(t, T) = \mathbb{P}(\tau > T \mid \mathcal{F}(t)). \quad (5.38)$$

Definition. The **recovery rate** $R \in [0, 1]$ is the fraction of notional recovered in the event of default.

Remark. Common modeling choices include constant recovery, stochastic recovery correlated with default, and alternative recovery conventions such as recovery of par or recovery of market value. In practice, constant recovery is widely assumed due to tractability and calibration constraints.

Definition. A non-negative adapted process $(\lambda(t))_{t \geq 0}$ is called the **default intensity** if

$$\mathbb{P}(\tau \in [t, t + dt) \mid \mathcal{F}(t), \tau > t) = \lambda(t) dt. \quad (5.39)$$

Proposition. In the **survival probability representation**, if λ_t is the default intensity, then

$$Q(0, T) = \mathbb{E} \left[\exp \left(- \int_0^T \lambda(s) ds \right) \right]. \quad (5.40)$$

If λ_t is deterministic,

$$Q(0, T) = \exp \left(- \int_0^T \lambda(s) ds \right). \quad (5.41)$$

Remark. Market instruments reveal default intensities under the risk-neutral measure \mathbb{Q} , which embed both expected default losses and risk premia. Physical default probabilities require additional modeling assumptions or historical estimation.

Definition. The **credit spread** $s(T)$ is the yield difference between a defaultable zero-coupon bond and a default-free bond of the same maturity.

Proposition. Under deterministic interest rates, deterministic intensity $\lambda(t)$, and recovery rate R , the credit spread satisfies approximately

$$s(T) \approx (1 - R) \frac{1}{T} \int_0^T \lambda(s) ds. \quad (5.42)$$

5.2.3 CDSs and CDOs

Definition. A **credit default swap (CDS)** on a defaultable security exchanges periodic premium payments for a default-contingent payment of $(1 - R)$ upon default.

Proposition. The CDS spread S must respect no arbitrage. Since the CDS has no value at the initial time,

$$\text{PV}_{\text{premium}} = \text{PV}_{\text{protection}}. \quad (5.43)$$

The present value of the protection leg is the expectation of the discounted future $(1 - R)$ payout

$$\text{PV}_{\text{protection}} = \mathbb{E} \left[(1 - R) \int_0^T D(0, t) d\mathbb{I}_{\{\tau \leq t\}} \right]. \quad (5.44)$$

Since $\mathbb{E}[d\mathbb{I}_{\{\tau \leq t\}}] = d\mathbb{P}(\tau \leq t)$, this is

$$\text{PV}_{\text{protection}} = (1 - R) \int_0^T D(0, t) d(1 - Q(0, t)). \quad (5.45)$$

The buyer pays $S\mathbb{I}_{\{\tau > t\}}dt$ at each moment in time, so the present value of the premium is similarly

$$\text{PV}_{\text{premium}} = \mathbb{E} \left[\int_0^T D(0, t) S\mathbb{I}_{\{\tau > t\}} dt \right] = S \int_0^T D(0, t) Q(0, t) dt. \quad (5.46)$$

In discretized premium payments, this is

$$S \sum_i \Delta_i D(0, T_i) Q(0, T_i) = (1 - R) \int_0^T D(0, t) d(1 - Q(0, t)). \quad (5.47)$$

Remark. If default occurs between payment dates, the **accrued premium** must be included in the valuation.

Procedure (Bootstrapping). Given market CDS spreads:

1. Assume piecewise-constant default intensities.

$$\lambda(t) = \lambda_i, \quad t \in [T_{i-1}, T_i]. \quad (5.48)$$

2. Solve sequentially for λ_i such that model CDS spreads match market quotes.
3. Recover survival probabilities via

$$Q(0, T_i) = \exp \left(- \sum_{j \leq i} \lambda_j \Delta T_j \right). \quad (5.49)$$

Remark. Single-name calibration does not determine joint default behavior. Modeling portfolio credit risk requires explicit assumptions on default dependence.

Definition. A **collateralized debt obligation (CDO)** redistributes the payouts and losses of a credit portfolio into tranches with different seniority levels. Tranche losses depend critically on default correlation.

Call $L(t)$ the cumulative **portfolio loss** at time t

$$L(t) = \sum_{i=1}^d (1 - R_i) \mathbf{1}_{\{\tau_i \leq t\}}. \quad (5.50)$$

For a tranche with attachment point A and detachment point D , the **tranche loss** at time t is

$$L^{A,D}(t) = \min(D - A, \max(L(t) - A, 0)). \quad (5.51)$$

Remark. Expected tranche losses depend on the full joint distribution of default times, and are therefore sensitive to default correlation. In a single-name credit model, $Q_i(t) = \mathbb{P}(\tau_i > t)$, yet in something like a CDO, $\mathbb{P}(\tau_1 > t_1, \dots, \tau_n > t_n)$ would be needed. We would like to add the joint dependence between otherwise independently calibrated default times (marginals).

Definition. Let (X_1, \dots, X_d) be a random vector, with each random variable having continuous marginal cumulative distribution functions $F_i(x) = \mathbb{P}(X_j \leq x)$. Via the probability integral transform,

$$(U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d)) \quad (5.52)$$

has uniform distribution marginals. The **copula** is the joint cumulative distribution function of this distribution

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d). \quad (5.53)$$

Proposition. Let $F_i(t) = \mathbb{P}(\tau_i \leq t)$ denote the marginal default distribution of name i , calibrated from single-name CDS data. Let (U_1, \dots, U_d) be a random vector with copula C . Define default times by

$$\tau_i = F_i^{-1}(U_i). \quad (5.54)$$

Then each τ_i has marginal distribution F_i , and the joint dependence between default times is fully characterized by the copula C .

Definition. A d -dimensional **Gaussian copula** with $d \times d$ correlation matrix Σ is

$$C_R^{Gauss}((u_1, \dots, u_d)) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)), \quad (5.55)$$

for Φ^{-1} the inverse standard normal CDF and Φ_R the multivariate CDF of a Gaussian with mean zero and covariance matrix equal to the correlation matrix.

In the **one-factor Gaussian copula model**, latent variables (Z_1, \dots, Z_d) are defined by

$$Z_i = \sqrt{\rho} X + \sqrt{1 - \rho} \varepsilon_i, \quad (5.56)$$

where $X, \varepsilon_i \sim \mathcal{N}(0, 1)$ are independent, and $\rho \in [0, 1]$ is the asset correlation parameter. The default times are then defined as

$$\tau_i = F_i^{-1}(\Phi(Z_i)). \quad (5.57)$$

The parameter ρ controls the proportion of default risk attributable to a common systemic factor. Equity tranches are primarily driven by idiosyncratic risk, while senior tranches are dominated by systemic risk.

Example (CDS and CDO pricing).

Consider a 5-year CDS on a corporate bond with recovery $R = 40\%$ and risk-free discount factor $D(0, t) = e^{-0.02t}$. Suppose the market quotes the 5-year CDS spread as $S = 100$ bps.

Assume piecewise-constant default intensity $\lambda(t) = \lambda$ over $[0, 5]$ years. The survival probability is

$$Q(0, T) = e^{-\lambda T}. \quad (5.58)$$

The present value of the premium leg is

$$S \int_0^5 D(0, t) Q(0, t) dt = S \int_0^5 e^{-0.02t} e^{-\lambda t} dt \quad (5.59)$$

$$= S \int_0^5 e^{-(\lambda+0.02)t} dt. \quad (5.60)$$

The present value of the protection leg is

$$(1 - R) \int_0^5 D(0, t) d(1 - Q(0, t)) = 0.6 \int_0^5 e^{-0.02t} \lambda e^{-\lambda t} dt \quad (5.61)$$

$$= 0.6\lambda \int_0^5 e^{-(\lambda+0.02)t} dt. \quad (5.62)$$

Equating PVs for no-arbitrage,

$$S \int_0^5 e^{-(\lambda+0.02)t} dt = 0.6\lambda \int_0^5 e^{-(\lambda+0.02)t} dt \implies \lambda \approx 0.0167. \quad (5.63)$$

Thus, the implied default intensity is $\lambda \approx 1.67\%$ per year. The 5-year survival probability is

$$Q(0, 5) = e^{-0.0167 \cdot 5} \approx 0.921. \quad (5.64)$$

Now, consider a 10-name portfolio of identical names with marginal default distributions calibrated as above. Take the 0–3% equity tranche.

Using a one-factor Gaussian copula, the latent variables are

$$Z_i = \sqrt{\rho}X + \sqrt{1 - \rho}\varepsilon_i, \quad X, \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.} \quad (5.65)$$

and default times are

$$\tau_i = F_i^{-1}(\Phi(Z_i)), \quad F_i(t) = 1 - Q(0, t). \quad (5.66)$$

For a given systemic correlation $\rho = 0.2$, we can simulate the tranche loss at $T = 5$:

$$L^{0,0.03}(T) = \min(0.03, \max(L(T) - 0, 0)), \quad (5.67)$$

$$L(T) = \frac{1}{10} \sum_{i=1}^{10} (1 - R) \mathbf{1}_{\{\tau_i \leq 5\}}. \quad (5.68)$$

A Monte Carlo with $N = 10,000$ scenarios yields the expected equity tranche loss

$$\mathbb{E}[L^{0,0.03}(T)] \approx 2.1\%. \quad (5.69)$$

Remark. There are several limitations to the model. The Gaussian copula exhibits zero upper **tail dependence**:

$$\lambda_U = \lim_{u \uparrow 1} \mathbb{P}(U_2 > u \mid U_1 > u) = 0. \quad (5.70)$$

Thus extreme default events do not cluster sufficiently under this model. In practice, different CDO tranches imply different implied correlation parameters, indicating that a single Gaussian copula parameter cannot simultaneously fit all tranches.

5.3 Counterparty Risk and XVA

Definition. **Counterparty risk** is the risk that a trading counterparty defaults prior to the maturity of a contract, causing losses due to the replacement cost of the position at default.

Definition. Let $V(t)$ denote the mark-to-market value of a derivative portfolio at time t from the perspective of the investor. The **exposure** is defined as

$$E(t) = V(t). \quad (5.71)$$

The positive and negative parts are

$$E^+(t) = \max(V(t), 0), \quad E^-(t) = \max(-V(t), 0). \quad (5.72)$$

Definition. The **expected positive exposure (EPE)** and **expected negative exposure (ENE)** are

$$\text{EPE}(t) = \mathbb{E}[E^+(t)], \quad \text{ENE}(t) = \mathbb{E}[E^-(t)], \quad (5.73)$$

where the expectation is taken under the risk-neutral measure.

Remark. In the presence of collateral and funding costs, the curve used for discounting cash flows need not coincide with the risk-free curve. Discounting reflects the remuneration of collateral, while funding curves reflect the institution's actual borrowing and lending costs. This separation breaks the classical risk-neutral pricing framework and leads naturally to valuation adjustments.

Definition. **Collateral agreements** specify the posting of variation margin to cover current exposure and initial margin to cover potential future exposure. Let $C(t)$ denote the collateral account, then the collateralized exposure is

$$\tilde{E}(t) = V(t) - C(t). \quad (5.74)$$

Definition. The following are adjustments to prices:

- **Credit Valuation Adjustment (CVA)** represents the expected loss due to counterparty default:

$$\text{CVA} = (1 - R_c) \int_0^T D(0, t) \mathbb{E}[\tilde{E}^+(t)] d\mathbb{P}(\tau_c \leq t), \quad (5.75)$$

where R_c is the counterparty recovery rate and τ_c is the counterparty default time.

- **Debit Valuation Adjustment (DVA)** accounts for the institution's own default risk:

$$\text{DVA} = (1 - R_o) \int_0^T D(0, t) \mathbb{E}[\tilde{E}^-(t)] d\mathbb{P}(\tau_o \leq t), \quad (5.76)$$

where τ_o denotes the investor's own default time.

- **Funding Valuation Adjustment (FVA)** captures the cost or benefit of funding uncollateralized exposures:

$$\text{FVA} = \int_0^T D(0, t) \mathbb{E}[\tilde{E}(t)] s_f(t) dt, \quad (5.77)$$

where $s_f(t)$ is the funding spread over the discounting rate.

- **Capital Valuation Adjustment (KVA)** reflects the cost of holding regulatory capital against counterparty risk:

$$\text{KVA} = \int_0^T D(0, t) \mathbb{E}[K(t)] h dt, \quad (5.78)$$

where $K(t)$ is the required capital and h is the hurdle rate.

- **Margin Valuation Adjustment (MVA)** accounts for the funding cost of posting initial margin:

$$\text{MVA} = \int_0^T D(0, t) \mathbb{E}[\text{IM}(t)] s_f(t) dt. \quad (5.79)$$

Remark. While CVA represents a true economic cost, DVA reflects an accounting benefit and is controversial from a risk management perspective.

Note. Exposure profiles are typically computed by simulating the joint evolution of market risk factors and valuation paths, followed by computing $V(t)$ and collateral dynamically along each path.

Definition. Wrong-way risk arises when exposure and counterparty credit quality are positively correlated:

$$\text{Corr}(E^+(t), \lambda_c(t)) > 0, \quad (5.80)$$

where $\lambda_c(t)$ is the counterparty default intensity. Wrong-way risk invalidates factorized CVA formulas and requires joint modeling of market and credit risk.

Definition. A **netting set** is a collection of transactions subject to a legally enforceable netting agreement, under which exposures are aggregated before collateral and close-out are applied.

Definition. Upon default, the **close-out** amount specifies the valuation used to settle remaining obligations. Common conventions include risk-free close-out and replacement close-out, leading to different XVA implications.

Proposition. The full valuation of a derivative portfolio is therefore given by

$$V_{\text{total}} = V_{\text{clean}} - \text{CVA} + \text{DVA} - \text{FVA} - \text{KVA} - \text{MVA}. \quad (5.81)$$

Example (XVA for a swap).

Consider a 5-year payer interest rate swap (IRS) with notional \$100M, fixed rate $K = 2\%$, floating payments quarterly, and discounting at $r = 1\%$ flat. Assume the counterparty has recovery $R_c = 40\%$ and the investor has recovery $R_o = 40\%$.

Let $V(t)$ denote the mark-to-market value of the IRS at time t from the investor's perspective. For simplicity, assume a deterministic exposure profile given by the following positive expected exposure (EPE):

$$\text{EPE}(t) = \mathbb{E}[E^+(t)] = 5 e^{-0.05t} \text{ (in \$M)}. \quad (5.82)$$

Similarly, the expected negative exposure (ENE) is:

$$\text{ENE}(t) = \mathbb{E}[E^-(t)] = 2 e^{-0.05t} \text{ (in \$M)}. \quad (5.83)$$

Let the counterparty default intensity be constant $\lambda_c = 1\%$ per year, and the investor's own default intensity be $\lambda_o = 0.5\%$. The survival probability for counterparty up to time t is

$$Q_c(t) = e^{-\lambda_c t}. \quad (5.84)$$

Using the continuous-time approximation for CVA:

$$\text{CVA} = (1 - R_c) \int_0^5 D(0, t) \text{EPE}(t) d(1 - Q_c(t)) \quad (5.85)$$

$$= (1 - 0.4) \int_0^5 e^{-0.01t} \cdot 5e^{-0.05t} \cdot \lambda_c e^{-\lambda_c t} dt. \quad (5.86)$$

$$= 0.03 \int_0^5 e^{-0.07t} dt. \quad (5.87)$$

$$\approx 0.03 \cdot 4.21 \approx 0.126 \text{ (\$M)}. \quad (5.88)$$

Similarly, for the investor's own default (DVA):

$$\text{DVA} = (1 - R_o) \int_0^5 D(0, t) \text{ENE}(t) d(1 - Q_o(t)) \quad (5.89)$$

$$= 0.6 \int_0^5 e^{-0.01t} \cdot 2e^{-0.05t} \cdot 0.005e^{-0.005t} dt. \quad (5.90)$$

$$\approx 0.006 \cdot 4.27 \approx 0.0256 \text{ (\$M)}. \quad (5.91)$$

Now, assume no FVA, KVA, or MVA for simplicity. If the clean (risk-free) value of the IRS is \$0.8M:

$$V_{\text{total}} = V_{\text{clean}} - \text{CVA} + \text{DVA} = 0.8 - 0.126 + 0.0256 \approx 0.6996 \text{ (\$M)}. \quad (5.92)$$

The CVA reduces the portfolio value due to counterparty credit risk, while the DVA partially offsets this, reflecting the investor's own default risk. In practice, exposure profiles are stochastic and would be computed via Monte Carlo simulations of the underlying market risk factors. Netting and wrong-way risk would adjust the EPE and ENE further.

Chapter 6

Portfolio Management

Motivation. How can positions be hedged against risk? Can positions "beat the market" without incurring large risk? This section investigates investment allocation.

6.1 Portfolio Theory

6.1.1 Mean-Variance Analysis

Definition. Consider a single-period economy with m risky assets and one risk-free asset. Let the random returns on the risky assets be

$$\vec{R} = (R_1, \dots, R_m), \quad (6.1)$$

and let the risk-free return be constant,

$$R_0 = r_0. \quad (6.2)$$

Let $\vec{w} = (w_1, \dots, w_m)$ denote the portfolio weights invested in the risky assets. The weight invested in the risk-free asset is

$$w_0 = 1 - \vec{w} \cdot \vec{1}_m, \quad (6.3)$$

which may be negative in the presence of borrowing.

The portfolio return is

$$R_p = \vec{w} \cdot \vec{R} + w_0 r_0. \quad (6.4)$$

Define the vector of expected returns and the covariance matrix by

$$\vec{\mu} = \mathbb{E}[\vec{R}], \quad \Sigma = \text{Var}(\vec{R}) = (\text{Cov}(R_i, R_j))_{i,j=1}^m, \quad (6.5)$$

where Σ is symmetric and positive definite.

Then the mean and variance of the portfolio return under the **Markowitz model** are

$$\mathbb{E}[R_p] = \vec{w} \cdot \vec{\mu} + (1 - \vec{w} \cdot \vec{1}_m)r_0, \quad \text{Var}(R_p) = \vec{w}^\top \Sigma \vec{w}. \quad (6.6)$$

Definition. A rational investor is assumed to have an increasing, concave utility function U . The **portfolio choice problem** is

$$\max_{\vec{w}} \mathbb{E}[U(1 + R_p)]. \quad (6.7)$$

For power utility with constant relative risk aversion $\gamma > 0$,

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma}, \quad (6.8)$$

a second-order Taylor approximation implies that maximizing expected utility is equivalent to the mean-variance problem

$$\max_{\vec{w}} \left(\mathbb{E}[R_p] - \frac{\gamma}{2} \text{Var}(R_p) \right). \quad (6.9)$$

Theorem (Optimal portfolio). The solution to the mean-variance optimization problem is

$$\vec{w}^* = \frac{1}{\gamma} \Sigma^{-1} (\vec{\mu} - r_0 \vec{1}_m), \quad (6.10)$$

with the remaining wealth invested in the risk-free asset.

The set of all optimal portfolios for varying γ forms the **mean-variance efficient frontier**.

Remark. This is the equivalent of the **Kelly criterion** for optimal bet size in continuous time, with the addition of the risk-free asset and the risk aversion.

Definition. The **tangency portfolio** is the unique efficient portfolio that is fully invested in risky assets, i.e. whose weights sum to one. The corresponding risk aversion is

$$\hat{\gamma} = \vec{1}_m^\top \Sigma^{-1} (\vec{\mu} - r_0 \vec{1}_m). \quad (6.11)$$

The tangency portfolio weights are

$$\hat{\vec{w}} = \frac{1}{\hat{\gamma}} \Sigma^{-1} (\vec{\mu} - r_0 \vec{1}_m), \quad (6.12)$$

and its expected return is

$$\mathbb{E}[\hat{R}_p] = \hat{\vec{w}} \cdot \vec{\mu}. \quad (6.13)$$

Theorem (Two-fund separation). Every optimal portfolio can be written as a combination of the risk-free asset and the tangency portfolio:

$$\vec{w} = \frac{\hat{\gamma}}{\gamma} \hat{w}. \quad (6.14)$$

Consequently,

$$\mathbb{E}[R_p] = r_0 + \frac{\hat{\gamma}}{\gamma} (\mathbb{E}[\hat{R}_p] - r_0), \quad \text{Var}(R_p) = \frac{\hat{\gamma}^2}{\gamma^2} \text{Var}(\hat{R}_p). \quad (6.15)$$

Definition. Eliminating γ yields

$$\mathbb{E}[R_p] = r_0 + \frac{\mathbb{E}[\hat{R}_p] - r_0}{\sqrt{\text{Var}(\hat{R}_p)}} \sqrt{\text{Var}(R_p)}, \quad (6.16)$$

which defines the **capital allocation line (CAL)**. Its slope is the **Sharpe ratio** of the tangency portfolio.

6.1.2 Factor Models and Evaluation

Theorem. In equilibrium, asset returns must be consistent with the portfolio choices of all investors. Assume that all investors are mean-variance optimizers, there exists a risk-free asset, investors have homogeneous beliefs, and markets clear. Let the **market portfolio** be the value-weighted portfolio of all risky assets, with weights \vec{w}_m . Its expected return is

$$\mathbb{E}[R_m] = \vec{w}_m \cdot \vec{\mu} + (1 - \vec{w}_m \cdot \vec{1}_m) r_0. \quad (6.17)$$

Then under the **capital asset pricing model (CAPM)**, each asset i has,

$$\mathbb{E}[R_i - r_0] = \beta_i \mathbb{E}[R_m - r_0], \quad (6.18)$$

where

$$\beta_i = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}. \quad (6.19)$$

Thus, only systematic risk, as measured by β_i , is priced in equilibrium.

Definition. More generally, expected returns may depend on exposure to multiple systematic risk factors. Call a **multi-factor model** on factors F_1, \dots, F_J :

$$\mathbb{E}[R_i - r_0] = \sum_{j=1}^J \beta_{i,j} \mathbb{E}[F_j], \quad (6.20)$$

with factor loadings

$$\beta_{i,j} = \frac{\text{Cov}(R_i, F_j)}{\text{Var}(F_j)}. \quad (6.21)$$

Such models include the Fama-French factor models and form the basis for empirical asset pricing.

Definition. In practice, portfolio performance is evaluated relative to a benchmark, often represented as a factor or portfolio. Given a factor model, fund returns satisfy the time-series regression

$$R_{i,t} - r_{0,t} = \sum_{j=1}^J \beta_{i,j} F_{j,t} + \alpha_i + \varepsilon_{i,t}, \quad (6.22)$$

where **alpha**, α_i , measures abnormal performance not explained by factor exposures, and each **beta** measures performance via correlation with the factors.

Definition. The **information ratio** of a fund is defined as

$$\text{IR}_i = \frac{\alpha_i}{\sigma(\varepsilon_{i,t})}, \quad (6.23)$$

and measures risk-adjusted excess performance relative to the benchmark.

6.2 Portfolio Risk

Motivation. Practical investment requires explicit control, measurement, and mitigation of risk. Risk management provides quantitative tools to measure downside exposure, enforce limits, and ensure portfolio robustness under adverse market conditions.

Definition. Let L be a real-valued random variable representing the **loss** of a portfolio over a fixed horizon, where positive values correspond to losses. The distribution of L summarizes the portfolio's risk profile.

Definition. For a confidence level $\alpha \in (0, 1)$, the **Value at Risk (VaR)** at level α is defined as

$$\text{VaR}_\alpha(L) = \inf \{ \ell \in \mathbb{R} : \mathbb{P}(L \leq \ell) \geq \alpha \}. \quad (6.24)$$

Equivalently, VaR_α is the α -quantile of the loss distribution. With probability α , the realized loss will not exceed VaR_α over the given horizon.

Remark. Despite its widespread use, VaR has important theoretical shortcomings. The failure of subadditivity implies that diversification may appear to *increase* risk under VaR, contradicting basic financial intuition.

Definition. The **Expected Shortfall** (ES), also called **Conditional Value at Risk** (CVaR), at level α is

$$\text{ES}_\alpha(L) = \mathbb{E}[L \mid L \geq \text{VaR}_\alpha(L)]. \quad (6.25)$$

It measures the expected loss in the worst $(1 - \alpha)$ fraction of cases.

Definition. A risk measure $\rho(L)$ is called **coherent** if it satisfies:

1. *Monotonicity:* $L_1 \leq L_2 \Rightarrow \rho(L_1) \leq \rho(L_2)$
2. *Translation invariance:* $\rho(L + c) = \rho(L) + c$
3. *Positive homogeneity:* $\rho(\lambda L) = \lambda \rho(L)$ for $\lambda > 0$
4. *Subadditivity:* $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$

Theorem. Expected Shortfall is a coherent risk measure, whereas Value at Risk is not in general.

Definition. Let W_t denote portfolio value. The **maximum drawdown** over $[0, T]$ is

$$\text{MDD} = \max_{0 \leq t \leq T} \left(\sup_{0 \leq s \leq t} W_s - W_t \right). \quad (6.26)$$

Remark. Drawdowns capture path-dependent risk ignored by variance-based measures and are central to capital allocation, leverage control, and investor behavior.

Proposition (Diversification). For a portfolio with weight vector \vec{w} and covariance matrix Σ ,

$$\text{Var}(R_p) = \vec{w}^\top \Sigma \vec{w} \quad (6.27)$$

is reduced when assets exhibit imperfect correlation.

Proof: Expanding the quadratic form gives $\vec{w}^\top \Sigma \vec{w} = \sum_i^m w_i \sigma_i^2 + \sum_{i \neq j} w_i w_j \text{Cov}(\sigma_i, \sigma_j)$, and we know $\text{Cov}(\sigma_i, \sigma_j) = \rho_{ij} \sigma_i \sigma_j$. Hence, this can be written as $\sum_i^m w_i \sigma_i^2 + \sum_{i \neq j} w_i w_j \rho_{ij} \sigma_i \sigma_j$. Since $\rho_{ij} \in [-1, 1]$, we thus have $\vec{w}^\top \Sigma \vec{w} \leq (\sum_i^m w_i \sigma_i)^2$, since the cross terms in the portfolio variance are less than or equal to the square of the sum of weighted standard deviations.

Remark. Diversification is helpful in eliminating **idiosyncratic risk** of individual securities, however it cannot eliminate **systemic risk** that affects the entire market.

Definition. The **marginal risk contribution** contribution of asset i to portfolio variance is

$$\text{MRC}_i = \frac{\partial}{\partial w_i} \text{Var}(R_p) = 2(\Sigma \vec{w})_i. \quad (6.28)$$

The **total risk contribution** of asset i is

$$\text{RC}_i = w_i(\Sigma \vec{w})_i. \quad (6.29)$$

Proposition. Consider a **hedge**, a position constructed to reduce sensitivity to a given risk factor, typically by taking an offsetting exposure. Let a portfolio have factor exposure β to a risk factor F . A hedge position h satisfying

$$\beta_{\text{portfolio}} + \beta_h = 0 \quad (6.30)$$

eliminates first-order exposure to F .

Remark. In practice, hedging reduces variance but may introduce basis risk, transaction costs, and model risk.

Definition. A **risk limit** is a constraint imposed on a portfolio risk measure, such as

$$\text{VaR}_\alpha(L) \leq \ell_{\max}, \quad \text{ES}_\alpha(L) \leq e_{\max}, \quad (6.31)$$

or on factor exposures, leverage, or drawdowns.

Proposition. Imposing convex risk constraints preserves tractability of portfolio optimization problems and prevents excessive concentration or leverage.

Definition. A **stress test** evaluates portfolio performance under extreme but plausible scenarios, such as large volatility shocks, correlation breakdowns, liquidity freezes, or macroeconomic regime shifts. More generally, **scenario analysis** computes portfolio losses under explicitly specified joint moves in risk factors.

Chapter 7

Markets and Trading

Motivation. Unlike frictionless models in asset pricing, liquidity, transaction costs, and strategic behavior impact how traders behave. Market microstructure studies how trades are executed in practice and how prices form through the interaction of market participants.

7.1 Exchanges

Definition. A **securities exchange** is a market venue for the exchange of stocks, bonds, etc. between stockbrokers and traders. Tradable securities are *listed* on the exchange, and transactions are done by *open outcry auction*. Examples are the NYSE and NASDAQ, which trade stocks, bonds, options, exchange-traded funds (ETFs), and more. Stocks do not need to be traded on exchanges, and products such as forwards or swaps are traded off-exchange directly between two parties, known as **over-the-counter**.

Definition. A **broker** is an agent who trades stocks and other investments on an exchange on behalf of someone else for a markup or fee. Since the average investor does not have a seat on the NYSE, brokerage firms such as Fidelity are the vehicle for investments. Stockbrokers in the US must pass Series exams, and generally act as advisors and investment managers for clients.

Definition. **Market liquidity** is a measure of the trade-off between how quickly an asset can be sold and at what price it can be sold. In a liquid markets, assets can be converted into cash easily and quickly, without greatly impacting market price.

Definition. In a transaction of a good or asset, **adverse selection** occurs when one party, either the buyer or the seller, has information about the product quality that the other party does not. This asymmetric information might concern hidden defects, such as the famous "market for Lemons".

Definition. A **central limit order book (CLOB)** is a market mechanism that records outstanding buy and sell limit orders for a given asset. At any time t , the order book consists of:

- A set of **bid** (buyers) prices $\{p_i^b\}$ with associated volumes $\{q_i^b\}$,
- A set of **ask** (sellers) prices $\{p_j^a\}$ with associated volumes $\{q_j^a\}$,

where bids are sorted in decreasing price order and asks in increasing price order.

Definition. The **best bid** p_t^b is the highest outstanding bid price, and the **best ask** p_t^a is the lowest outstanding ask price. The **midprice** and **bid-ask spread** are defined respectively by

$$m_t = \frac{p_t^a + p_t^b}{2}, \quad s_t = p_t^a - p_t^b. \quad (7.1)$$

The midprice is commonly used as a proxy for the “efficient price”.

Definition. Common order types include:

- **Market orders**, which execute immediately against the best available prices,
- **Limit orders**, which specify a maximum (buy) or minimum (sell) execution price,
- **Cancel orders**, which remove existing limit orders.

Remark. Market orders consume liquidity, while limit orders provide liquidity. The interaction between the two drives short-term price dynamics.

Definition. A **market maker** is a firm that stands ready to buy or sell stocks at a publicly quoted price. The market maker provides liquidity to the market and profits via the bid-ask spread, which compensates for adverse selection and inventory risk. Examples of market makers are Jane Street or Optiver.

Definition. In contrast to a CLOB, a **Request for Quote (RFQ)** trade execution model occurs when buyers and sellers request bid and ask prices quoted from a number of market makers. Traders do not trade with each other at chosen prices, but rather with the market makers.

7.2 Optimal Execution

Motivation. Large trades cannot generally be executed instantaneously without incurring significant market impact. Optimal execution models aim to minimize trading costs while managing risk.

Definition (Execution Problem). An agent seeks to execute a total quantity Q of shares over a finite horizon $[0, T]$, by choosing a trading strategy $(q_t)_{t \in [0, T]}$, where q_t denotes the trading rate.

Assumptions. The **Almgren-Chriss model** assumes the following:

The unaffected asset midprice $(S_t)_{t \in [0, T]}$ evolves as a martingale under the agent's pricing measure:

$$dS_t = \sigma dW_t, \quad (7.2)$$

where $\sigma > 0$ is constant volatility and W_t is a standard Brownian motion. The agent's trades do not influence this reference price.

Trading at instantaneous rate q_t affects execution prices through two channels: *temporary impact*, reflecting liquidity consumption, which affects only the current execution price, and *permanent impact*, reflecting information leakage or inventory effects, which shifts all future prices.

In the linear impact model, the execution price is

$$\tilde{S}_t = S_t - \eta q_t - \gamma \int_0^t q_s ds, \quad (7.3)$$

where $\eta \geq 0$ and $\gamma \geq 0$ are coefficients representing the impacts respectively.

Definition. Let X_t denote the remaining inventory at time t . The trading rate q_t satisfies the **inventory process**

$$dX_t = -q_t dt, \quad X_0 = Q, \quad X_T = 0. \quad (7.4)$$

The total **execution cost** relative to the initial mark-to-market value QS_0 is

$$C = \int_0^T q_t \tilde{S}_t dt - QS_0. \quad (7.5)$$

Proposition. Under the above assumptions, the execution cost decomposes as

$$C = \eta \int_0^T q_t^2 dt + \frac{\gamma}{2} Q^2 - \sigma \int_0^T X_t dW_t. \quad (7.6)$$

Remark. The first term represents temporary impact costs, the second term is the deterministic permanent impact cost, and the final term captures price risk due to holding inventory over time.

Proposition. For an agent choosing a trading strategy $(q_t)_{t \in [0, T]}$ to minimize the mean-variance criterion

$$\mathbb{E}[C] + \lambda \text{Var}(C), \quad (7.7)$$

where $\lambda \geq 0$ is a risk-aversion parameter, the objective reduces to

$$\min_{(X_t)} \int_0^T \left(\eta \dot{X}_t^2 + \lambda \sigma^2 X_t^2 \right) dt, \quad (7.8)$$

subject to $X_0 = Q$ and $X_T = 0$.

Theorem (Optimal execution path). The optimal inventory trajectory satisfies the Euler-Lagrange equation

$$\ddot{X}_t = \frac{\lambda\sigma^2}{\eta} X_t, \quad (7.9)$$

with boundary conditions $X_0 = Q$, $X_T = 0$. The unique solution is

$$X_t = Q \frac{\sinh(\kappa(T-t))}{\sinh(\kappa T)}, \quad \kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}}. \quad (7.10)$$

Corollary. The optimal trading speed is thus

$$q_t = -\dot{X}_t = Q \frac{\kappa \cosh(\kappa(T-t))}{\sinh(\kappa T)}. \quad (7.11)$$

As $\lambda \rightarrow 0$, the strategy minimizes expected cost only and trades linearly in time. As $\lambda \rightarrow \infty$, the strategy becomes increasingly front-loaded, approaching instantaneous liquidation.

Remark (Limitations). The Almgren-Chriss model assumes linear impact, continuous trading, and exogenous price dynamics. Extensions incorporate non-linear impact, stochastic liquidity, discrete trading, and limit order book dynamics.

Chapter 8

Time Series Analysis

Motivation. How can asset data over time be understood? Time series analysis concerns the modeling and statistical properties of data indexed in time order. In finance, understanding serial dependence, stationarity, and volatility dynamics is essential for forecasting, risk modeling, and econometric inference.

8.1 Time Series Foundations

8.1.1 Models and Tests

Definition. A discrete-time stochastic process $\{X_t\}$ is a **time series**. The time series $\{X_t\}$ is **strictly stationary** if for all $m \in \mathbb{N}$ and all time shifts τ ,

$$\mathbb{P}(X_{t_1+\tau}, \dots, X_{t_m+\tau}) = \mathbb{P}(X_{t_1}, \dots, X_{t_m}), \quad (8.1)$$

i.e., all finite-dimensional distributions are invariant under time shifts. $\{X_t\}$ is **weakly stationary** if

$$\mathbb{E}[X_t] = \mu, \quad \text{Var}(X_t) = \sigma^2 < \infty, \quad \text{Cov}(X_t, X_{t+\tau}) = \Gamma(\tau), \quad (8.2)$$

where $\mu, \sigma^2, \Gamma(\tau)$ (autocovariance) do not depend on t . The **autocorrelation function** (ACF) is defined by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}. \quad (8.3)$$

Stationary $\{X_t\}$ is called **ergodic** if

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{1 \leq k \leq T} X_{t+k} = \mu \quad (8.4)$$

Definition. Define the **lag operator** L by

$$L^k(X_t) = X_{t-k}, \quad (8.5)$$

and the **difference operator** Δ by

$$\Delta^k = (1 - L)^k. \quad (8.6)$$

Theorem (Wold representation). Any zero-mean weakly stationary process is equivalent to

$$X_t = V_t + \sum_{i=0}^{\infty} \psi_i \eta_{t-i} = V_t + \sum_{i=0}^{\infty} \psi_i L^i \eta_t = V_t + \psi(L) \epsilon_t, \quad (8.7)$$

where $\{\epsilon_t\} \sim \mathcal{N}(0, \sigma^2)$ is linearly unpredictable white noise, $\psi_0 = 1$, $\sum_i \psi_i^2 < \infty$, and V_t is deterministic. This expresses X_t as an infinite moving average.

Definition. For operator $\psi(L)$, if there exists an inverse

$$\psi^{-1}(L) = \sum_{i=0}^{\infty} \psi_i^* L^i, \quad (8.8)$$

such that $\psi(L)\psi^{-1}(L) = L^0$, then a moving average time series is **invertible** with equivalent autoregressive representation

$$X_t = \psi(L) \epsilon_t \Leftrightarrow \psi^{-1}(L) X_t = \epsilon_t. \quad (8.9)$$

Definition. An **MA(q) (moving average) model** takes the form

$$X_t - \mu = \theta(L) \epsilon_t, \quad (8.10)$$

for characteristic polynomial

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q. \quad (8.11)$$

Proposition. MA(q) processes are always weakly stationary.

Proof idea. Taking the expected value of both sides yields $\mathbb{E}[X_t] = \mu$, since white noise has mean zero. Taking the covariance $Cov(X_t, X_{t+\tau})$ will yield zero for the expectation of different white noises, and will give the product of coefficients times σ^2 when white noises overlap. The number of identical white noises is determined by the lag τ , so this is indeed weakly stationary.

Definition. An **AR(p) (autoregressive) model** is

$$\phi(L)(X_t - \mu) = \epsilon_t, \quad (8.12)$$

for characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p. \quad (8.13)$$

Definition. An $\text{AR}(p)$ process is said to possess a **unit root** if the characteristic equation

$$\phi(z) = 0 \quad (8.14)$$

has at least one root z_0 such that $|z_0| = 1$.

Proposition. For a general $\text{AR}(p)$ process, stationarity requires that all roots of $\phi(z)$ lie strictly outside the unit circle.

Proof. A time series having a Wold representation as an $\text{MA}(\infty)$ process is necessary and sufficient for being weakly stationary. This moving average time series is obtainable via inverting the characteristic polynomial evaluated with the lag operator, as per the definition of invertible. Hence, there must exist a representation of $1/\phi(z)$ as a convergent power series.

However, it is known from complex analysis that this series, centered at zero, converges if and only if there are no poles within the unit disk. Hence, there cannot be any roots that lie within the unit circle for the process to admit a moving average representation with summable coefficients. \square

Example. Consider the $\text{AR}(1)$ process

$$X_t = \phi X_{t-1} + \varepsilon_t. \quad (8.15)$$

The characteristic polynomial is

$$\phi(z) = 1 - \phi z. \quad (8.16)$$

This polynomial has a root at $z = 1/\phi$. A unit root occurs when $\phi = 1$, yielding

$$X_t = X_{t-1} + \varepsilon_t, \quad (8.17)$$

which is a random walk. Iterating forward gives

$$X_t = X_0 + \sum_{j=1}^t \varepsilon_j. \quad (8.18)$$

Since the shocks accumulate permanently,

$$\mathbb{E}[X_t] = X_0, \quad \text{Var}(X_t) = t\sigma^2, \quad (8.19)$$

and the variance grows without bound as $t \rightarrow \infty$. Hence, the process is not weakly stationary.

Theorem. Many financial time series exhibit persistent non-stationarity consistent with unit roots. The classic test for a unit root in an $\text{AR}(1)$ context is the **Dickey–Fuller test**. Rewriting an $\text{AR}(1)$ process as

$$X_t = \phi X_{t-1} + \epsilon_t \rightarrow \Delta X_t = (\phi - 1)X_{t-1} + \epsilon_t = \delta X_{t-1} + \epsilon_t, \quad (8.20)$$

an OLS estimator for $\hat{\delta}$ can be calculated. However, under the null, the test statistic has a nonstandard distribution requiring special critical values. This test has null hypothesis $\phi = 1$ (unit root, non-stationary) and alternative $|\phi| < 1$ (stationary).

Remark. Augmented Dickey-Fuller (ADF) and related tests (Phillips–Perron) control for higher-order dynamics, while KPSS tests take stationarity as the null.

Definition. A process $\{X_t\}$ follows an **ARMA**(p, q) **model** if

$$\phi(L)(X_t - \mu) = \theta(L)\epsilon_t, \quad (8.21)$$

where $\{\epsilon_t\}$ is white noise with variance σ^2 , and

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \quad \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q. \quad (8.22)$$

The associated Wold decomposition yields

$$X_t = \mu + \psi(L)\eta_t, \quad \psi(L) = \phi(L)^{-1}\theta(L), \quad (8.23)$$

provided the process is stationary and invertible.

Definition. A unit root at $z = 1$ implies that the process can be factored as

$$\phi(L) = (1 - L)\psi(L), \quad (8.24)$$

leading to the representation

$$(1 - L)X_t = \psi(L)^{-1}\epsilon_t. \quad (8.25)$$

Thus, the first difference $\Delta X_t = X_t - X_{t-1}$ is stationary, while X_t itself is not. Such a process is said to be **integrated** of order one, denoted $I(1)$. More generally, if $(1 - L)^d$ divides $\phi(L)$, the process is integrated of order d , written $I(d)$.

Definition. To model non-stationary series with stochastic trends, the ARMA framework is extended by differencing. If $\Delta^d X_t$ is stationary and follows an ARMA(p, q) model, then X_t is said to follow an **ARIMA**(p, d, q) model:

$$\phi(L)\Delta^d X_t = \theta(L)\epsilon_t, \quad (8.26)$$

Pure random walks and integrated processes $I(1)$ are special cases of ARIMA with $d = 1$.

8.1.2 Estimation and Model Selection

ARMA and ARIMA models are typically estimated using MLE. Under the assumption of Gaussian innovations, the likelihood function is fully characterized

by the conditional mean and variance implied by the model, and MLE is asymptotically efficient. For an ARMA(p, q) process,

$$\phi(L)X_t = \theta(L)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad (8.27)$$

the likelihood depends on the parameter vector $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$ and is maximized numerically, often using exact likelihood methods or a state-space representation with Kalman filtering.

In special cases, simpler estimation techniques are available. For pure autoregressive models, the **Yule-Walker equations** relate the autoregressive coefficients to the theoretical autocovariance function:

$$\gamma(k) = \sum_{j=1}^p \phi_j \gamma(k-j), \quad k \geq 1. \quad (8.28)$$

Replacing population autocovariances with their sample counterparts yields closed-form estimators. While computationally efficient, Yule-Walker estimators are generally less statistically efficient than MLE, particularly in small samples.

Model order selection requires balancing goodness of fit against model complexity. Information criteria provide a formal framework for this trade-off. Let $\hat{\sigma}^2$ denote the estimated innovation variance and n the sample size. Commonly used criteria include:

$$\text{AIC}(p, q) = \log(\hat{\sigma}^2) + \frac{2(p+q)}{n}, \quad (8.29)$$

$$\text{BIC}(p, q) = \log(\hat{\sigma}^2) + \frac{\log n (p+q)}{n}. \quad (8.30)$$

The **Akaike Information Criterion (AIC)** is motivated by minimizing the expected Kullback-Leibler divergence between the true data-generating process and the fitted model, and therefore tends to favor more flexible specifications. The **Bayesian Information Criterion (BIC)** imposes a stronger penalty on complexity and is consistent for model selection under standard regularity conditions. In practice, AIC is often preferred for forecasting applications, while BIC is favored when the goal is parsimonious structural modeling.

8.2 Volatility Modeling

Motivation. Empirical studies of financial return series reveal features that are not captured by linear time-series models such as ARMA or ARIMA. In particular, while raw returns often exhibit little serial correlation, functions of returns - most notably squared or absolute returns - display strong and persistent dependence. This motivates explicit modeling of the conditional variance process.

8.2.1 Observations and Models

Remark. There are several well-documented empirical properties of asset return volatility:

Volatility clustering: Large price movements tend to be followed by large movements (of either sign), and small movements tend to be followed by small movements. Formally, while $\{X_t\}$ may be weakly autocorrelated, the processes $\{|X_t|\}$ or $\{X_t^2\}$ exhibit strong positive autocorrelation over time. This phenomenon is referred to as **volatility clustering**.

Heavy tails: Empirical return distributions exhibit excess kurtosis relative to the Gaussian distribution. That is,

$$\mathbb{P}(|X_t| > x) \sim x^{-\alpha}, \quad \alpha < \infty, \quad (8.31)$$

for large x , indicating power-law or sub-exponential tail behavior. This has important implications for risk measurement and extreme-event modeling.

Leverage effect: Negative returns tend to be followed by higher future volatility than positive returns of the same magnitude. This asymmetric response of volatility to shocks is known as the **leverage effect** and reflects both financial leverage and behavioral mechanisms.

Long memory: Volatility proxies such as $|X_t|$ or X_t^2 often display slowly decaying autocorrelation functions,

$$\text{Corr}(X_t^2, X_{t+k}^2) \sim k^{-\delta}, \quad 0 < \delta < 1, \quad (8.32)$$

suggesting long-range dependence rather than the exponential decay implied by short-memory models.

Hurst exponent: Long memory is often quantified using the **Hurst exponent** $H \in (0, 1)$, defined via scaling behavior

$$\mathbb{E}[|X_{t+h} - X_t|^2] \sim h^{2H}. \quad (8.33)$$

For volatility processes, empirical estimates frequently yield $H > \frac{1}{2}$, consistent with persistent dependence. In contrast, Brownian motion has $H = \frac{1}{2}$.

Definition. While ARMA/ARIMA models address autocorrelation in levels, financial time series often exhibit persistent time-varying conditional variance. An **ARCH(q) (autoregressive conditional heteroskedasticity) model** specifies

$$X_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2, \quad (8.34)$$

where $\{z_t\}$ is an i.i.d. sequence with $\mathbb{E}[z_t] = 0$ and $\mathbb{E}[z_t^2] = 1$, and $\alpha_0 > 0$, $\alpha_i \geq 0$.

ARCH models capture volatility clustering by allowing large past shocks to increase future conditional variance. However, high-order ARCH models are often required in practice, motivating a more parsimonious extension.

Definition. A **GARCH**(p, q) (**generalized ARCH**) model incorporates lagged conditional variances:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (8.35)$$

with $\omega > 0$, $\alpha_i, \beta_j \geq 0$.

Proposition. For a GARCH(1,1) model, a sufficient condition for covariance stationarity is

$$\alpha_1 + \beta_1 < 1, \quad (8.36)$$

in which case the unconditional variance exists and equals

$$\mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha_1 - \beta_1}. \quad (8.37)$$

When $\alpha_1 + \beta_1$ is close to one, volatility shocks decay slowly, producing highly persistent volatility dynamics.

Definition. Standard ARCH and GARCH models are symmetric in the sense that positive and negative shocks of equal magnitude have identical effects on future volatility. This contradicts empirical evidence. An **EGARCH** (**exponential GARCH**) model specifies

$$\log \sigma_t^2 = \omega + \sum_{j=1}^p \beta_j \log \sigma_{t-j}^2 + \sum_{i=1}^q [\alpha_i z_{t-i} + \gamma_i (|z_{t-i}| - \mathbb{E}|z_{t-i}|)], \quad (8.38)$$

where asymmetry is introduced through the α_i terms. Negative shocks can increase volatility more than positive ones without imposing non-negativity constraints on parameters.

Definition. A **GJR-GARCH** model introduces asymmetry directly via an indicator:

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \gamma X_{t-1}^2 \mathbf{1}_{\{X_{t-1} < 0\}} + \beta \sigma_{t-1}^2. \quad (8.39)$$

Here, negative returns have an amplified impact on future volatility when $\gamma > 0$.

Definition. While GARCH-type models generate persistent volatility, they still imply exponential decay of autocorrelations. To capture true long memory, fractional models are introduced. A **FIGARCH** (**fractionally integrated GARCH**) model replaces the lag polynomial with a fractional difference operator:

$$\sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 - \Delta^d] \alpha(L) X_t^2, \quad (8.40)$$

where $d \in (0, 1)$ governs the degree of long memory.

FIGARCH models generate hyperbolically decaying autocorrelations in squared returns, consistent with empirical evidence, at the cost of increased complexity and weaker moment conditions.

Example (GARCH-type Models). Consider a daily return series $\{X_t\}$ for a hypothetical equity with sample unconditional variance $\hat{\sigma}^2 \approx 0.0004$ (i.e. standard deviation $\approx 2\%$). We illustrate different volatility models:

Suppose we fit an ARCH(1) model:

$$X_t = \sigma_t z_t, \quad \sigma_t^2 = 0.0001 + 0.3X_{t-1}^2, \quad z_t \sim \mathcal{N}(0, 1). \quad (8.41)$$

If yesterday's return was $X_{t-1} = 0.04$ (4%), the conditional variance today is

$$\sigma_t^2 = 0.00058, \quad (8.42)$$

so $\sigma_t \approx 0.0241$ (2.41%). This illustrates volatility clustering (a large past shock increases today's variance).

Consider a GARCH(1,1) model:

$$\sigma_t^2 = 0.0001 + 0.3X_{t-1}^2 + 0.6\sigma_{t-1}^2. \quad (8.43)$$

Assume $\sigma_{t-1}^2 = 0.0004$ and $X_{t-1} = 0.02$:

$$\sigma_t^2 = 0.00046. \quad (8.44)$$

The unconditional variance is

$$\mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha_1 - \beta_1} = \frac{0.0001}{1 - 0.3 - 0.6} = 0.001, \quad (8.45)$$

demonstrating persistent volatility when $\alpha_1 + \beta_1$ is close to 1.

Consider now a GJR-GARCH to capture asymmetry:

$$\sigma_t^2 = 0.0001 + 0.3X_{t-1}^2 + 0.2X_{t-1}^2 \mathbf{1}_{\{X_{t-1} < 0\}} + 0.5\sigma_{t-1}^2. \quad (8.46)$$

If yesterday's return was $X_{t-1} = -0.03$, then

$$\sigma_t^2 = 0.00075, \quad (8.47)$$

so negative returns increase volatility more than positive returns, illustrating the leverage effect.

In a FIGARCH(1,d,1) model with $d = 0.4$, the conditional variance is influenced by the fractional difference of past squared returns:

$$\sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 - \Delta^{0.4}] \alpha(L) X_t^2. \quad (8.48)$$

Unlike standard GARCH, shocks decay hyperbolically:

$$\text{Corr}(X_t^2, X_{t+k}^2) \sim k^{-0.4}, \quad k \rightarrow \infty, \quad (8.49)$$

so volatility remains correlated over long horizons, consistent with empirical long memory.

8.2.2 Forecasting versus Pricing

ARCH-family models are primarily designed for *forecasting* conditional variance under the physical probability measure \mathbb{P} . They are widely used for risk management (VaR, ES) and volatility forecasting.

However, these models do not define a full arbitrage-free price process for the underlying asset and do not specify dynamics under a risk-neutral measure. Consequently, they are not directly suitable for derivative pricing without additional assumptions or measure changes.

In contrast, volatility models used for *pricing*, such as stochastic volatility or local volatility models, are embedded within continuous-time, no-arbitrage frameworks and explicitly specify dynamics under equivalent martingale measures. Bridging the gap between statistically calibrated volatility models and arbitrage-free pricing remains a central challenge in quantitative finance.

8.3 Spectral Analysis and Cointegration

Motivation. How can cyclical behavior be captured? Rather than analyzing dependence through autocovariances in the time domain, spectral analysis decomposes variability across frequencies, revealing behavior that may not be evident from finite-lag correlations. Likewise, many economic and financial time series are nonstationary but move together over long horizons. Cointegration formalizes this notion.

Definition. Let $\{X_t\}_{t \in \mathbb{Z}}$ be a weakly stationary process with autocovariance function

$$\Gamma(k) = \text{Cov}(X_t, X_{t-k}). \quad (8.50)$$

The **spectral density** of X_t is defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Gamma(k) e^{-i\lambda k}, \quad \lambda \in [-\pi, \pi]. \quad (8.51)$$

When $\sum_k |\Gamma(k)| < \infty$, the series converges absolutely and $f(\lambda)$ is continuous.

The spectral density provides a decomposition of the variance of X_t across frequencies, with low frequencies corresponding to long-run components and high frequencies corresponding to short-run fluctuations.

Proposition. The autocovariance function can be recovered from the spectral density via the inverse Fourier transform:

$$\Gamma(k) = \int_{-\pi}^{\pi} e^{i\lambda k} f(\lambda) d\lambda, \quad (8.52)$$

establishing a one-to-one correspondence between time- and frequency-domain representations.

Proposition. For linear time-series models, the spectral density admits a closed-form expression. Let $\{X_t\}$ follow an ARMA(p, q) process

$$\phi(L)X_t = \theta(L)\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2), \quad (8.53)$$

where $\phi(z)$ and $\theta(z)$ are polynomials in the lag operator. Then the spectral density is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}. \quad (8.54)$$

Zeros of $\phi(e^{-i\lambda})$ close to the unit circle generate sharp peaks in the spectral density at low frequencies, corresponding to highly persistent dynamics. Poles at $\lambda = 0$ correspond to unit-root or near-unit-root behavior. Flat spectra correspond to white noise, and slowly decaying autocorrelations manifest as high spectral mass near zero frequency.

Definition. Let $X_t \in \mathbb{R}^k$ be a vector of integrated processes of order one, $I(1)$. The components of X_t are said to be **cointegrated** if there exists a nonzero vector $\beta \in \mathbb{R}^k$ such that

$$\beta^\top X_t \sim I(0). \quad (8.55)$$

The vector β defines a long-run equilibrium relationship among the components of X_t .

Remark. Cointegration eliminates low-frequency divergence: while individual components of X_t exhibit infinite variance at frequency zero, the cointegrating combination $\beta^\top X_t$ has finite spectral mass near $\lambda = 0$. In this sense, cointegration is fundamentally a *low-frequency restriction*.

Theorem (Granger representation). If X_t is a cointegrated $I(1)$ vector process, then it admits a **vector error-correction model (VECM)** representation:

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \quad (8.56)$$

where $\Pi = \alpha\beta^\top$ has reduced rank. Here, β contains the cointegrating vectors, α governs the speed at which deviations from long-run equilibrium are corrected, and short-run dynamics are captured by lagged differences.

Remark. Two standard approaches are widely used in practice:

- The **Engle–Granger two-step procedure**, which estimates the long-run relationship via regression and tests the residuals for stationarity;
- **Johansen tests**, which use a likelihood-based framework to infer the number of cointegrating relationships in a multivariate system.

Remark. Cointegration plays a central role in financial modeling such as **pairs trading and statistical arbitrage**, where mean-reverting spreads are constructed from nonstationary price series. However, care must be taken when embedding such models into no-arbitrage pricing frameworks.

Chapter 9

Numerical Methods

Motivation. Many models of interest in quantitative finance do not admit closed-form solutions for derivative prices or risk measures. As a result, numerical methods play a central role in both pricing and calibration. Many also reduce to minimizing or maximizing an objective function, or equivalently solving a system of nonlinear equations. Optimization thus arises in model calibration, portfolio construction, risk management, and optimal control.

9.1 Common Methods in Finance

9.1.1 Monte Carlo

Definition. **Monte Carlo methods** estimate expectations by simulating a large number of independent sample paths of the underlying stochastic process. By the Law of Large Numbers, these estimates converge almost surely to the true expectation.

Under the risk-neutral measure \mathbb{Q} , the time-0 value of a derivative with payoff $\Phi(X_T)$ is

$$V_0 = e^{-rT} \mathbb{E}^{\mathbb{Q}}[\Phi(X_T)]. \quad (9.1)$$

Approximating the expectation with M simulated paths yields the Monte Carlo estimator

$$V_0^{(M)} = e^{-rT} \frac{1}{M} \sum_{m=1}^M \Phi(X_T^{(m)}). \quad (9.2)$$

Theorem (Monte Carlo convergence). If $\mathbb{E}^{\mathbb{Q}}[\Phi(X_T)^2] < \infty$, then

$$V_0^{(M)} \xrightarrow{a.s.} V_0, \quad \sqrt{M}(V_0^{(M)} - V_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad (9.3)$$

where $\sigma^2 = \text{Var}^{\mathbb{Q}}(\Phi(X_T))$.

Consequently, the root mean square error decays at rate $\mathcal{O}(M^{-1/2})$, independent of the dimension of the problem.

Definition. Consider a stochastic differential equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t, \quad (9.4)$$

where W_t is a standard Brownian motion.

In general, exact simulation of X_t is not available, and numerical discretization schemes must be used. Let $\Delta t = T/N$ and $t_n = n\Delta t$. The **Euler-Maruyama approximation** is

$$X_{n+1} = X_n + \mu(t_n, X_n)\Delta t + \sigma(t_n, X_n)\Delta W_n, \quad (9.5)$$

where $\Delta W_n \sim \mathcal{N}(0, \Delta t)$ are independent.

Definition. Let X_T be the true solution and $X_T^{(N)}$ its numerical approximation. The approximation scheme has **strong order** α if

$$\mathbb{E}[|X_T - X_T^{(N)}|] = \mathcal{O}(\Delta t^\alpha). \quad (9.6)$$

The approximation scheme has **weak order** β if for all smooth test functions φ ,

$$|\mathbb{E}[\varphi(X_T)] - \mathbb{E}[\varphi(X_T^{(N)})]| = \mathcal{O}(\Delta t^\beta). \quad (9.7)$$

The Euler–Maruyama scheme has strong order 1/2 and weak order 1 under standard Lipschitz assumptions.

Remark. In option pricing, weak convergence is typically sufficient, since only expectations of payoffs are required.

Definition. Variance reduction techniques aim to decrease $\text{Var}(V_0^{(M)})$ without increasing the number of simulations, thereby improving estimator efficiency.

- **Antithetic variates.** If $X_T(W)$ is simulated using Brownian motion W , then simulating both W and $-W$ and averaging the payoffs can reduce variance when the payoff is monotone in W .
- **Control variates.** Let Y be a random variable with known expectation $\mathbb{E}[Y]$. For a scalar θ ,

$$\Phi^* = \Phi - \theta(Y - \mathbb{E}[Y]). \quad (9.8)$$

The optimal θ minimizing variance is

$$\theta^* = \frac{\text{Cov}(\Phi, Y)}{\text{Var}(Y)}. \quad (9.9)$$

Control variates are among the most effective techniques in financial Monte Carlo, particularly when pricing options with known Black–Scholes prices.

- **Importance sampling.** Importance sampling changes the probability measure to emphasize rare but important events, such as deep out-of-the-money option payoffs. If \mathbb{Q} is replaced by $\tilde{\mathbb{Q}}$, then

$$\mathbb{E}^{\mathbb{Q}}[\Phi] = \mathbb{E}^{\tilde{\mathbb{Q}}} \left[\Phi \frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}} \right]. \quad (9.10)$$

Definition. Quasi-Monte Carlo (QMC) methods replace pseudo-random numbers with low-discrepancy sequences such as Sobol or Halton sequences.

Theorem (Koksma–Hlawka inequality). For a function f of bounded variation,

$$\left| \frac{1}{M} \sum_{m=1}^M f(u_m) - \int_{[0,1]^d} f(u) du \right| \leq V(f) D_M, \quad (9.11)$$

where $V(f)$ is the Hardy-Krause variation, and D_M is the discrepancy of the point set.

In practice, QMC often achieves convergence rates close to $\mathcal{O}(M^{-1})$ for moderate effective dimension, though without the probabilistic error guarantees of standard Monte Carlo.

Definition (Least squares Monte Carlo). Monte Carlo methods naturally handle path-dependent payoffs such as Asian, barrier, and lookback options. However, pricing American options introduces optimal stopping problems. The **Longstaff-Schwartz algorithm** approximates the continuation value by regressing future discounted payoffs onto basis functions of the state variables, yielding an approximate optimal stopping rule.

Remark. Monte Carlo pricing of American options is inherently biased due to approximation of the stopping boundary, but the bias can be controlled with sufficiently rich basis functions and large sample sizes.

9.1.2 Finite Difference Methods

Motivation: For Markovian models, derivative prices admit a representation as solutions to partial differential equations via the Feynman-Kac theorem. Let X_t be a Markov process with infinitesimal generator \mathcal{L} . Then the value function

$$V(t, x) = \mathbb{E}_{t,x}^{\mathbb{Q}} \left[e^{-r(T-t)} \Phi(X_T) \right] \quad (9.12)$$

satisfies the backward pricing PDE

$$\frac{\partial V}{\partial t}(t, x) + \mathcal{L}V(t, x) - rV(t, x) = 0, \quad V(T, x) = \Phi(x). \quad (9.13)$$

It is not immediately obvious how to analytically solve this.

Definition: To solve numerically, **finite difference methods** approximate the PDE by discretizing time and state space. Let $t_n = n\Delta t$ and $x_i = x_{\min} + i\Delta x$. The continuous derivatives are replaced by finite difference quotients, yielding a system of algebraic equations for the grid values $V_i^n \approx V(t_n, x_i)$.

Common discretization schemes include:

- **Explicit** (forward-time, centered-space): simple to implement but conditionally stable,
- **Implicit** (backward-time, centered-space): unconditionally stable but requires solving linear systems,
- **Crank–Nicolson**: second-order accurate in time, combining explicit and implicit methods.

Definition. A finite difference scheme is **consistent** if the local truncation error vanishes as $\Delta t, \Delta x \rightarrow 0$, and is **stable** if numerical errors do not grow uncontrollably during iteration.

Theorem (Lax equivalence). For linear initial-value problems, consistency and stability together imply convergence of the numerical solution to the true solution.

Remark. Explicit schemes are subject to CFL-type stability conditions linking Δt and Δx , whereas implicit and Crank–Nicolson schemes are unconditionally stable and therefore preferred in practice.

Example (American put). Let S_t follow geometric Brownian motion under the risk-neutral measure:

$$dS_t = rS_t dt + \sigma S_t dW_t, \quad (9.14)$$

with constant interest rate r and volatility σ . Consider an American put option with strike K and maturity T , whose payoff is

$$\Phi(S) = \max(K - S, 0). \quad (9.15)$$

The option value $V(t, S)$ satisfies the variational inequality

$$\max \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV, (K - S) - V \right) = 0, \quad (9.16)$$

with terminal condition

$$V(T, S) = \max(K - S, 0). \quad (9.17)$$

At each time and asset level, the option value is the maximum of the continuation value and the intrinsic value.

To discretize, let the spatial grid be $S_i = i\Delta S$, $i = 0, \dots, I$, and the time grid $t_n = n\Delta t$, $n = 0, \dots, N$, with $t_N = T$. Denote the numerical approximation

by $V_i^n \approx V(t_n, S_i)$. Using a backward-time, centered-space discretization, the Black-Scholes operator is approximated by

$$\mathcal{L}V_i^{n+1} = \frac{1}{2}\sigma^2 S_i^2 \frac{V_{i+1}^{n+1} - 2V_i^{n+1} + V_{i-1}^{n+1}}{(\Delta S)^2} + rS_i \frac{V_{i+1}^{n+1} - V_{i-1}^{n+1}}{2\Delta S} - rV_i^{n+1}. \quad (9.18)$$

The fully implicit time discretization yields

$$\frac{V_i^{n+1} - V_i^n}{\Delta t} + \mathcal{L}V_i^{n+1} = 0. \quad (9.19)$$

At each time step, the American constraint is enforced by

$$V_i^n = \max(V_i^n, K - S_i). \quad (9.20)$$

Equivalently, the problem can be written as a discrete variational inequality:

$$V_i^n \geq K - S_i, \quad \frac{V_i^{n+1} - V_i^n}{\Delta t} + \mathcal{L}V_i^{n+1} \leq 0, \quad (9.21)$$

with complementarity at each grid point.

Boundary conditions are chosen as:

$$V(t, 0) = K, \quad (9.22)$$

$$V(t, S_{\max}) = 0, \quad (9.23)$$

reflecting certain exercise at zero asset price and negligible value at large prices.

The numerical procedure proceeds as follows:

1. Initialize $V_i^N = \max(K - S_i, 0)$ at maturity.
2. Step backward in time by solving the linear system for V^{n+1} .
3. Enforce the early exercise condition pointwise:

$$V_i^n \leftarrow \max(V_i^n, K - S_i). \quad (9.24)$$

4. Repeat until $t = 0$.

The early exercise boundary $S^*(t)$ is implicitly defined by

$$V(t, S^*(t)) = K - S^*(t). \quad (9.25)$$

Finite difference methods naturally approximate this boundary as the interface between grid points where the intrinsic value dominates and where continuation is optimal.

Remark. Finite difference methods have high accuracy for low-dimensional problems and handling of early exercise features, yet computational cost grows exponentially with dimension. As a result, finite difference methods are primarily used for one- and two-dimensional problems, while Monte Carlo methods dominate in higher-dimensional settings.

9.2 Optimization and Root Finding

9.2.1 Optimization Techniques

Definition. An **optimization problem** consists of

$$\min_{x \in \mathcal{D}} f(x), \quad (9.26)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function and $\mathcal{D} \subseteq \mathbb{R}^n$ is the feasible set.

A point x^* is a *local minimum* if $f(x^*) \leq f(x)$ for all x in a neighborhood of x^* , and a *global minimum* if the inequality holds for all $x \in \mathcal{D}$.

Definition: A set C is called **convex** if the line segment between any two points in the set also is in C :

$$x, y \in C \Rightarrow (1 - t)x + ty \in C, \forall t \in [0, 1]. \quad (9.27)$$

A **convex function** $f : X \rightarrow \mathbb{R}^n$ defined on a convex subset, likewise, has the line segment connecting any two points on the graph greater than or equal to the function value (the epigraph is convex):

$$x_1, x_2 \in X \Rightarrow f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2), \forall t \in [0, 1]. \quad (9.28)$$

Convex optimization minimizes convex functions over convex subsets.

Theorem. A function f is convex if and only if its domain is convex and it satisfies the **first-order convexity condition**

$$f(y) \leq f(x) + \nabla f(x)^T(y - x), \forall x, y \in \text{dom}(f). \quad (9.29)$$

Alternatively, a function f is convex if and only if its domain is convex and it satisfies the **second-order convexity condition**

$$\nabla^2 f(x) \geq 0, \forall x \in \text{dom}(f). \quad (9.30)$$

Definition. Given a differentiable objective f , **gradient descent** is a first-order method that iterates

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad (9.31)$$

where $\alpha_k > 0$ is a step size. Gradient descent is robust and easy to implement, but can converge slowly, especially for ill-conditioned problems.

Definition. If f is twice differentiable, **Newton's method** is a second-order method that updates

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k), \quad (9.32)$$

where $\nabla^2 f$ is the Hessian. Newton's method converges quadratically near the optimum but requires inversion of the Hessian, which may be unstable or computationally expensive.

Definition. The **BFGS algorithm** is a quasi-Newton method that updates an approximation $H_k \approx \nabla^2 f(x_k)^{-1}$ via rank-two updates that preserve symmetry and positive definiteness. BFGS provides a good balance between robustness and speed, and is widely used in model calibration.

Definition. A **constrained optimization problem** takes the form

$$\min_x f(x) \quad \text{subject to} \quad h_i(x) = 0, \quad g_j(x) \geq 0 \quad (9.33)$$

for arbitrary numbers of constraints.

Theorem (Karush-Kuhn-Tucker). For constrained optimization of a continuously differentiable, convex f , with g_j convex and h_i affine, then for a local minimum x^* , there exist λ_i, μ_j satisfying the following conditions. Any point x^* satisfying the following conditions is also a global minimum.

Define the **Lagrangian** as

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_i^p \lambda_i h_i(x) + \sum_j^q \mu_j g_j(x). \quad (9.34)$$

- $\nabla_x \mathcal{L}(x^*, \lambda, \mu) = 0$ (stationary Lagrangian).
- The point x^* satisfies the constraints (primal feasibility).
- $\mu_j \geq 0$ (dual feasibility).
- $\mu_j g_j(x^*) = 0$ (complementary slackness).

Definition. An optimization problem is **ill-conditioned** if small perturbations in data or parameters lead to large changes in the solution.

Parameters are **identifiable** if distinct parameter values produce distinct model outputs. Non-identifiability is common in over-parameterized models and leads to flat directions in the objective function.

Definition. Regularization augments the objective function to stabilize optimization:

$$f_{\text{reg}}(x) = f(x) + \lambda R(x), \quad (9.35)$$

where $R(x)$ is a penalty term. ℓ_2 **regularization** is defined as

$$R(x) = \|x\|_2^2 \quad (9.36)$$

and improves conditioning, and ℓ_1 **regularization** is defined as

$$R(x) = \|x\|_1 \quad (9.37)$$

and promotes sparsity.

Remark. Calibration objectives are often non-convex due to nonlinear model dynamics, latent variables, discrete market instruments, correlation parameters. As a result, optimization may exhibit multiple local minima, making initialization and regularization critical.

9.2.2 Root Finding and Programming

Definition. A **root-finding problem** asks to find x such that

$$F(x) = 0, \quad (9.38)$$

for a vector-valued function F .

Remark. Calibration of implied volatility or CDS hazard rates is often posed as root-finding rather than optimization.

Definition. A **quadratic program** minimizes

$$\frac{1}{2}x^\top Qx + c^\top x \quad (9.39)$$

subject to linear constraints.

Proposition. Let μ be expected returns and Σ the covariance matrix. The mean-variance problem is:

$$\min_w \frac{1}{2}w^\top \Sigma w - \lambda w^\top \mu \quad \text{subject to} \quad \mathbf{1}^\top w = 1. \quad (9.40)$$

The mean-variance problem is a convex quadratic program.

Definition. A **shrinkage estimator** of the covariance matrix is:

$$\hat{\Sigma} = \alpha \Sigma_{\text{sample}} + (1 - \alpha) \Sigma_0, \quad (9.41)$$

where Σ_0 is a structured target. Shrinkage improves stability in high-dimensional settings.

Chapter 10

Machine Learning Methods

Motivation. How are strategies work-shopped and tested? This section explores how statistical learning is applied to finance, and how models are validated.

10.1 Common Methods in Finance

Definition. Regression analysis seeks to model the relationship between a dependent variable and one or more explanatory variables. In its simplest linear form, the **regression model** is written as

$$Y = X\beta + \varepsilon, \quad (10.1)$$

where $Y \in \mathbb{R}^n$ is the vector of observed responses, $X \in \mathbb{R}^{n \times p}$ is the design matrix, $\beta \in \mathbb{R}^p$ is the vector of unknown coefficients, and $\varepsilon \in \mathbb{R}^n$ is a vector of random errors.

A common assumption is

$$\mathbb{E}[\varepsilon] = 0, \quad \text{Var}(\varepsilon) = \sigma^2 I_n. \quad (10.2)$$

Definition. The **ordinary least squares (OLS)** estimator is defined by

$$\hat{\beta}_{\text{OLS}} = \arg \min_{\beta} \|Y - X\beta\|_2^2 = (X^\top X)^{-1} X^\top Y, \quad (10.3)$$

provided $X^\top X$ is invertible.

Theorem (Gauss–Markov). If the regression model is linear in parameters, $\mathbb{E}[\varepsilon] = 0$, $\text{Var}(\varepsilon) = \sigma^2 I_n$, and X has full column rank, then $\hat{\beta}_{\text{OLS}}$ is the *best linear unbiased estimator* (BLUE) of β .

Definition. Given a parametric model $f(x; \theta)$ and observations $\{x_i\}_{i=1}^n$, the **maximum likelihood estimator (MLE)** is

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta) = \arg \max_{\theta} \prod_{i=1}^n f(x_i; \theta). \quad (10.4)$$

Remark. Under Gaussian errors, the OLS estimator coincides with the MLE.

Definition. In high-dimensional or noisy financial datasets, ordinary least squares often suffers from overfitting or multicollinearity. **Ridge regression** solves

$$\hat{\beta}_{\text{ridge}} = \arg \min_{\beta} (\|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2), \quad (10.5)$$

where $\lambda > 0$ controls the strength of regularization. The **lasso estimator** replaces the ℓ^2 penalty with an ℓ^1 penalty:

$$\hat{\beta}_{\text{lasso}} = \arg \min_{\beta} (\|Y - X\beta\|_2^2 + \lambda \|\beta\|_1). \quad (10.6)$$

Remark. Lasso performs variable selection by setting some coefficients exactly to zero, making it particularly useful for factor selection in asset pricing models.

Definition. Hidden Markov Models (HMMs) are probabilistic models for time series driven by an unobserved discrete state process. They are commonly used to model regime switching in returns or volatility.

An HMM consists of:

- a latent Markov chain $\{Z_t\}$ taking values in $\{1, \dots, K\}$,
- an observation/emission process $\{Y_t\}$ such that

$$Y_t \mid Z_t = k \sim f_k(\cdot), \quad (10.7)$$

- a transition matrix P satisfying

$$\mathbb{P}(Z_t = j \mid Z_{t-1} = i) = P_{ij}. \quad (10.8)$$

Gaussian emission distributions are commonly used for log-returns or log-volatility.

Theorem (Baum-Welch algorithm). Given an HMM with K hidden states, observations $\{Y_1, \dots, Y_T\}$, and unknown parameters

$$\theta = (\pi, P, \{f_k\}_{k=1}^K), \quad (10.9)$$

where $\pi_i = \mathbb{P}(Z_1 = i)$ are initial state probabilities, P is the $K \times K$ transition matrix, and f_k are emission densities, we wish to estimate θ from data.

Define the *forward probability*:

$$\alpha_t(i) = \mathbb{P}(Y_1, \dots, Y_t, Z_t = i \mid \theta), \quad (10.10)$$

computed recursively:

$$\alpha_1(i) = \pi_i f_i(Y_1), \quad \alpha_{t+1}(j) = f_j(Y_{t+1}) \sum_{i=1}^K \alpha_t(i) P_{ij}. \quad (10.11)$$

Also define the *backward probability*:

$$\beta_t(i) = \mathbb{P}(Y_{t+1}, \dots, Y_T \mid Z_t = i, \theta), \quad (10.12)$$

computed recursively:

$$\beta_T(i) = 1, \quad \beta_t(i) = \sum_{j=1}^K P_{ij} f_j(Y_{t+1}) \beta_{t+1}(j). \quad (10.13)$$

E-step: for expectations, compute the posterior probabilities of hidden states and transitions:

$$\gamma_t(i) = \mathbb{P}(Z_t = i \mid Y_{1:T}, \theta) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^K \alpha_t(j) \beta_t(j)}, \quad (10.14)$$

$$\xi_t(i, j) = \mathbb{P}(Z_t = i, Z_{t+1} = j \mid Y_{1:T}, \theta) = \frac{\alpha_t(i) P_{ij} f_j(Y_{t+1}) \beta_{t+1}(j)}{\sum_{i,j} \alpha_t(i) P_{ij} f_j(Y_{t+1}) \beta_{t+1}(j)}. \quad (10.15)$$

M-step: now, update parameters to maximize expected complete-data log-likelihood:

$$\pi_i^{\text{new}} = \gamma_1(i), \quad (10.16)$$

$$P_{ij}^{\text{new}} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad (10.17)$$

$$f_j^{\text{new}} = \text{MLE of emission parameters weighted by } \gamma_t(j). \quad (10.18)$$

For Gaussian emissions $f_j = \mathcal{N}(\mu_j, \sigma_j^2)$:

$$\mu_j^{\text{new}} = \frac{\sum_{t=1}^T \gamma_t(j) Y_t}{\sum_{t=1}^T \gamma_t(j)}, \quad (\sigma_j^2)^{\text{new}} = \frac{\sum_{t=1}^T \gamma_t(j) (Y_t - \mu_j^{\text{new}})^2}{\sum_{t=1}^T \gamma_t(j)}. \quad (10.19)$$

Repeat E-step and M-step until convergence of the log-likelihood

$$\log \mathcal{L}(\theta) = \log \mathbb{P}(Y_{1:T} \mid \theta). \quad (10.20)$$

Remark. The Baum–Welch algorithm is an instance of the Expectation–Maximization (EM) algorithm tailored for HMMs. It guarantees monotone increase of the likelihood, but only convergence to a local maximum. Once trained, the most likely hidden state sequence can be inferred via the Viterbi algorithm.

Theorem. A **linear Gaussian state-space model** is defined by:

$$\text{State (latent) dynamics: } X_t = AX_{t-1} + Bu_t + w_t, \quad w_t \sim \mathcal{N}(0, Q), \quad (10.21)$$

$$\text{Observation: } Y_t = CX_t + v_t, \quad v_t \sim \mathcal{N}(0, R), \quad (10.22)$$

where:

- X_t is the hidden state vector at time t ,
- Y_t is the observed data vector,
- u_t is a known input (control) vector,
- w_t and v_t are mutually independent Gaussian noise processes,
- A, B, C are known matrices of appropriate dimensions.

The goal of the **Kalman filter** is to estimate the hidden state X_t given observations Y_1, \dots, Y_t recursively. Let:

$$\hat{X}_{t|t-1} = \mathbb{E}[X_t | Y_1, \dots, Y_{t-1}] \quad (\text{prediction}), \quad (10.23)$$

$$\hat{X}_{t|t} = \mathbb{E}[X_t | Y_1, \dots, Y_t] \quad (\text{updated estimate}). \quad (10.24)$$

Predict the state and covariance at time t given the estimate at $t-1$:

$$\hat{X}_{t|t-1} = A\hat{X}_{t-1|t-1} + Bu_t, \quad (10.25)$$

$$P_{t|t-1} = AP_{t-1|t-1}A^\top + Q, \quad (10.26)$$

where $P_{t|t-1} = \text{Var}[X_t - \hat{X}_{t|t-1}]$ is the predicted state covariance.

Update by incorporating the new observation Y_t :

$$K_t = P_{t|t-1}C^\top (CP_{t|t-1}C^\top + R)^{-1} \quad (\text{Kalman gain}), \quad (10.27)$$

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (Y_t - C\hat{X}_{t|t-1}), \quad (10.28)$$

$$P_{t|t} = (I - K_tC)P_{t|t-1}. \quad (10.29)$$

K_t determines how much weight to give the new observation versus the prediction, and the term $Y_t - C\hat{X}_{t|t-1}$ is the **innovation** (prediction error).

Repeat prediction and update for each $t = 1, \dots, T$.

Remark. The filter provides the best linear unbiased estimate of the state at each time. The innovation sequence is Gaussian and can be used for likelihood-based parameter estimation. The Kalman filter is particularly powerful in finance because many models (stochastic volatility, term structure models) are

naturally linear in latent states or can be linearized, allowing real-time estimation of unobserved quantities.

Definition. A **Random Forest** is an ensemble learning method that aggregates a collection of decision trees trained on bootstrapped samples of the data.

Each tree is constructed by sampling observations with replacement, selecting a random subset of features at each split, growing the tree to reduce impurity (e.g. variance or Gini index). The final prediction is given by averaging (regression) or majority voting (classification).

Remark. Random Forests are nonparametric and capture nonlinear interactions between variables without requiring explicit model specification. These can help with return and volatility forecasting, credit risk and default prediction, and feature importance analysis.

Definition. Let $X \in \mathbb{R}^{n \times p}$ be a data matrix with n observations and p variables, assumed to be centered (zero mean for each column). **Principal component analysis (PCA)** finds orthogonal directions v_1, \dots, v_p that maximize the variance of projected data:

$$v_1 = \arg \max_{\|v\|=1} \text{Var}(Xv), \quad (10.30)$$

$$v_k = \arg \max_{\|v\|=1, v \perp v_1, \dots, v_{k-1}} \text{Var}(Xv), \quad (10.31)$$

for $k = 2, \dots, p$. The projections $Z = XV$ are called the **principal components**, where $V = [v_1, \dots, v_p]$ is the matrix of eigenvectors.

Compute the sample covariance of X :

$$\Sigma = \frac{1}{n-1} X^\top X \in \mathbb{R}^{p \times p}. \quad (10.32)$$

Solve for eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ and corresponding eigenvectors v_1, \dots, v_p :

$$\Sigma v_k = \lambda_k v_k. \quad (10.33)$$

The first principal component v_1 points in the direction of maximal variance. Each subsequent component explains the maximal remaining variance while being orthogonal to previous components.

Retain only the first $r < p$ components to approximate the data:

$$X \approx Z_r V_r^\top, \quad Z_r = X V_r, \quad V_r = [v_1, \dots, v_r]. \quad (10.34)$$

The proportion of variance explained by the k -th component is

$$\text{PVE}_k = \frac{\lambda_k}{\sum_{j=1}^p \lambda_j}. \quad (10.35)$$

Cumulative explained variance helps choose r such that most variability is preserved.

Remark. PCA assumes linear relationships and maximizes variance, and does not necessarily capture nonlinear structure. Standardization (scaling each variable to unit variance) is often used when variables are on different scales. In finance, PCA is particularly useful for interpretable factor decomposition and stress testing.

10.2 Validation and Backtesting

Motivation. Unlike classical supervised learning settings, financial data are time-ordered, non-stationary, and subject to structural breaks. As a result, standard validation techniques must be adapted to control for temporal dependence, regime changes, and multiple sources of bias.

Definition. The most basic distinction in model evaluation is between **in-sample** (IS) and **out-of-sample** (OOS) performance.

In-sample testing measures performance on the same dataset used to estimate model parameters. While useful for debugging and diagnostics, strong in-sample fit alone is not evidence of predictive power and often reflects over-optimization.

Out-of-sample testing evaluates the model on data not used during training. The full dataset can be partitioned chronologically as

$$\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{test}}, \quad \max(t \in \mathcal{D}_{\text{train}}) < \min(t \in \mathcal{D}_{\text{test}}). \quad (10.36)$$

Definition. The following biases occur often:

- **Look-ahead bias** occurs when future information is inadvertently used in model training or feature construction. Common sources include using revised macroeconomic data rather than real-time releases, normalizing features using full-sample statistics, and using future returns in label construction
- **Data snooping** occurs when repeated experimentation on the same dataset leads to models that exploit noise rather than signal. Data snooping inflates apparent performance and invalidates classical statistical inference.
- **Survivorship bias** arises when datasets include only assets that survive until the end of the sample period. This leads to systematically overstated performance, particularly in equity and credit strategies. Examples include equity indices excluding delisted firms or funds databases omitting liquidated funds. Proper backtests must include dead assets and historically accurate index constituents.

Definition. Standard k -fold cross-validation assumes i.i.d. observations and is generally inappropriate for financial time series. Nevertheless, structured

variants can be used with care. In **blocked** or **time-series k -fold cross-validation**, folds respect temporal ordering, with training data always preceding validation data. This provides multiple OOS estimates while limiting information leakage.

Definition. To reflect real-world deployment, models are often evaluated using **rolling windows** or **walk-forward analysis**. In a rolling window framework, parameters are re-estimated over a fixed-length window:

$$[t_i - \Delta, t_i] \rightarrow \hat{\theta}_i, \quad (10.37)$$

and tested on a subsequent holdout period. Walk-forward analysis chains these windows forward through time, producing a realistic sequence of OOS predictions. This approach allows simultaneous evaluation of predictive performance, parameter stability, and sensitivity to market conditions.

A deployable model should exhibit stable parameters across calibration windows. Let $\hat{\theta}_t$ denote parameters estimated at time t . Instability can be assessed via

$$\|\hat{\theta}_{t+1} - \hat{\theta}_t\| \quad \text{or} \quad \text{Var}(\hat{\theta}_t). \quad (10.38)$$

Remark. Financial markets exhibit structural breaks and regime changes driven by macroeconomic shifts, monetary policy, or crises. Models calibrated in one regime may fail catastrophically in another.

Definition. Stress testing evaluates model behavior under extreme but plausible scenarios, such as volatility spikes, liquidity shocks, and correlation breakdowns. Stress testing complements backtesting by probing regions of the state space not well represented in historical data.

Definition. Over-optimization occurs when model complexity or hyperparameter tuning is excessive relative to the available data. It manifests as strong IS performance, rapid OOS decay, and high parameter instability. Regularization, simplicity constraints, and strict OOS evaluation are primary fixes.

Definition. For risk models such as Value-at-Risk (VaR), validation focuses on exceedance behavior. Let α denote the VaR confidence level and I_t an indicator of violation:

$$I_t = \mathbf{1}\{L_t > \text{VaR}_\alpha(t)\}. \quad (10.39)$$

The **Kupiec test** (unconditional coverage) assesses whether the empirical violation rate matches α . The **Christoffersen test** extends this by testing both correct coverage and independence of violations. Together, these form the basis of regulatory backtesting standards for market risk models.

Chapter 11

Model Risk, Calibration, and Practical Pitfalls

Motivation. Up to this point, the focus of these notes has been on constructing mathematically consistent models for asset prices and using no-arbitrage arguments to derive pricing formulas. In practice, however, all models are approximations. They are calibrated to incomplete and noisy data, embedded in market microstructure frictions, and used for decision-making under uncertainty. This chapter addresses the gap between theoretically sound models and their real-world deployment.

11.1 The Nature of Model Risk

Definition. **Model risk** refers to the risk of financial loss arising from the use of an incorrect, misspecified, or misapplied model. Unlike market risk, model risk is endogenous, created by the modeling choices themselves.

Model risk arises from several sources:

- **Structural assumptions:** continuous paths, lognormal returns, constant parameters, or complete markets.
- **Parametric choices:** selecting a specific volatility or interest-rate model from many plausible alternatives.
- **Calibration error:** fitting parameters to finite, noisy, and regime-dependent data.
- **Implementation error:** numerical discretization, Monte Carlo error, or coding mistakes.

A key distinction is that no-arbitrage conditions constrain prices *relative* to one another, but do not uniquely determine the correct dynamics of the underlying risk factors.

Remark. The existence of a risk-neutral measure guarantees internal consistency of prices, not correctness of the model relative to reality.

11.2 Meanings of Measures

A recurring conceptual distinction is between the real-world probability measure \mathbb{P} and a risk-neutral (or pricing) measure \mathbb{Q} .

Under \mathbb{P} :

- Asset dynamics reflect statistical properties of realized returns.
- Parameters are estimated from historical time series.
- Expectations correspond to forecasts.

Under \mathbb{Q} :

- Discounted asset prices are martingales.
- Parameters are chosen to reproduce observed market prices.
- Expectations correspond to arbitrage-free prices, not forecasts.

These measures coincide only in very special cases. In general, there is no reason to expect risk premia, jump intensities, or volatility dynamics to agree under \mathbb{P} and \mathbb{Q} .

Example. Implied volatility extracted from option prices reflects the market price of volatility risk. Historical volatility estimated from returns measures realized variance under \mathbb{P} . The discrepancy between the two is itself a tradable risk premium.

11.3 Calibration as an Inverse Problem

Model calibration is the process of choosing parameters θ such that model prices $V^{\text{model}}(\theta)$ match observed market prices V^{mkt} .

Typically, this is formulated as an optimization problem:

$$\min_{\theta} \sum_i w_i \left(V_i^{\text{model}}(\theta) - V_i^{\text{mkt}} \right)^2, \quad (11.1)$$

where the weights w_i reflect liquidity, bid-ask spreads, or confidence in the data.

This inverse problem is often:

- **Ill-posed:** multiple parameter sets yield similar prices.

- **Unstable:** small data changes lead to large parameter shifts.
- **Over-parameterized:** more degrees of freedom than informative constraints.

Remark. A good calibration fit does not imply a good model. It only implies interpolation of current prices.

11.4 Regime Dependence

Market parameters are not constant. Volatility surfaces, correlation structures, and credit spreads evolve over time, often discontinuously.

Common manifestations include:

- Volatility smiles shifting shape intraday.
- Heston parameters recalibrating to different values day-to-day.
- Correlations increasing sharply during crises.

This regime dependence implies that calibrated parameters should be viewed as *state variables*, not immutable constants.

Example. A local volatility surface calibrated today may produce arbitrage-free prices for vanilla options, yet generate unrealistic forward smiles or unstable hedging behavior when used dynamically.

11.5 Incomplete Markets

Classical replication arguments rely on market completeness. In practice:

- Jumps destroy perfect hedging.
- Stochastic volatility introduces unhedgeable risk.
- Transaction costs prevent continuous rebalancing.

As a result, hedging strategies minimize risk rather than eliminate it. This reframes pricing as a trade-off between replication error and risk tolerance.

Remark. In incomplete markets, prices are no longer unique. They depend on preferences, constraints, and capital considerations.

11.6 Numerical Error

Even when a model is theoretically correct, numerical implementations introduce additional error:

- Time discretization bias in Monte Carlo simulation.

- Grid instability in finite-difference schemes.
- Poor convergence of optimization routines.

Example. A Monte Carlo estimator with insufficient time steps may systematically misprice barrier options due to discretization of path-dependent events.

11.7 Backtesting and Validation

A model that fits prices today must still be validated against future data. Common validation tools include:

- Out-of-sample pricing error.
- Hedging PnL attribution.
- Stability of calibrated parameters.
- Sensitivity to market perturbations.

Backtesting failures often arise from:

- Look-ahead bias.
- Overfitting to noise.
- Ignoring transaction costs and liquidity.

11.8 Machine Learning and Overfitting

Machine learning methods exacerbate model risk if used without structural constraints. Key pitfalls include:

- High in-sample performance with poor generalization.
- Lack of economic interpretability.
- Sensitivity to regime shifts.

ML methods are most effective when combined with domain structure, such as no-arbitrage constraints or known symmetries.

11.9 Practical Philosophy of Modeling

A useful mental model is:

Models are tools, not truths.

Good models are internally consistent, calibrated transparently, and fail in understood and monitored ways.

Final Remark. The goal of quantitative finance is not to eliminate uncertainty, but to quantify, price, and manage it. Mathematical rigor provides the foundation, but robust skepticism determines success in practice.