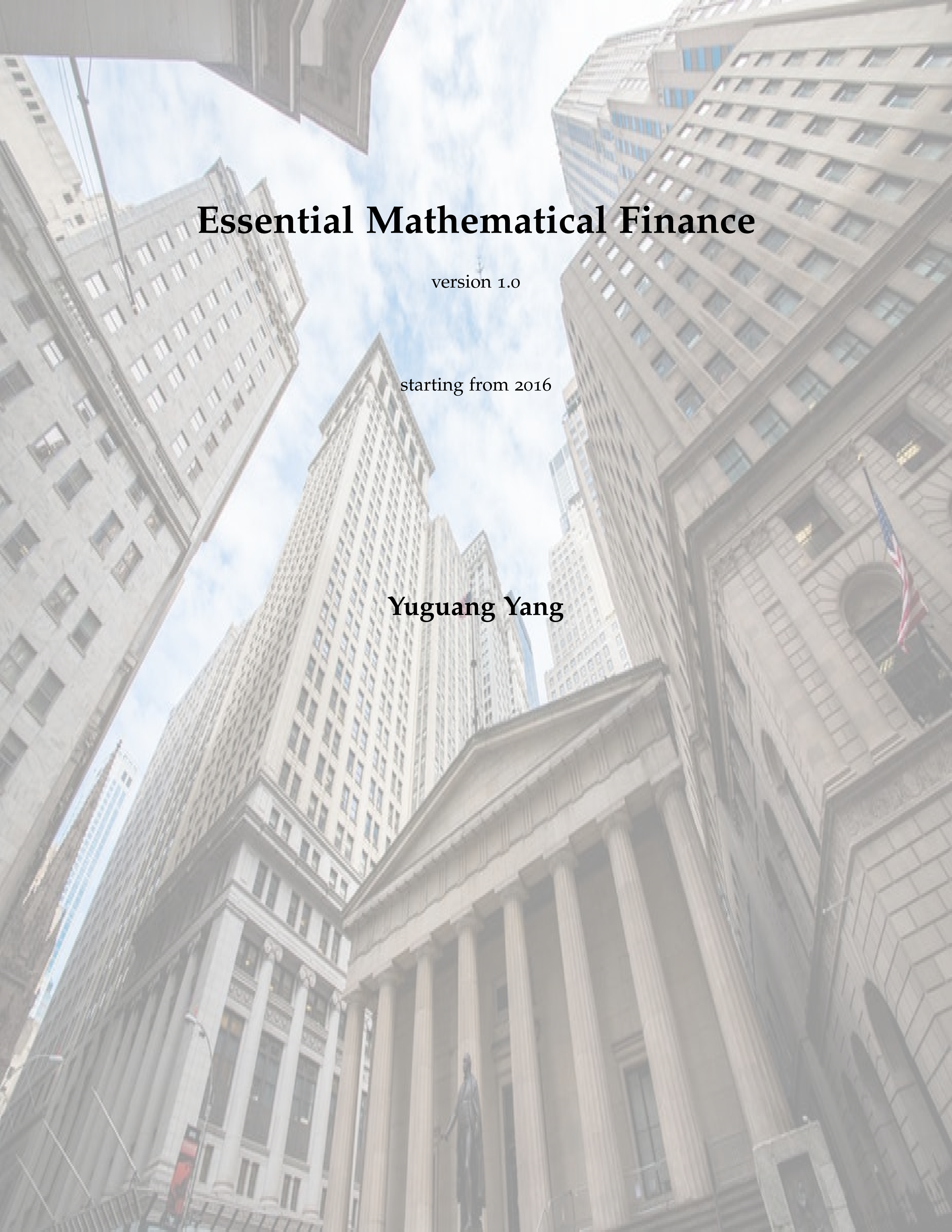


Essential Mathematical Finance

version 1.0

starting from 2016

Yuguang Yang

Notations

Mathematical notation

- $\mathbb{R}$: real numbers.
- $\mathbb{R}_+$: nonnegative real numbers.
- $\mathbb{R}_{++}$: positive real numbers.
- $\bar{\mathbb{R}}$: extended real numbers.
- $\mathbb{C}$: complex numbers.
- $\mathbb{F}$:real or complex numbers.
- $\mathbb{Q}$: rational numbers.
- $\mathbb{Z}$:integer numbers.
- $\mathbb{P}$:positive numbers.
- $\mathcal{P}(\mathbb{P})$:positive numbers.
- $\mathcal{P}_n$: polynomial of degree of n.
- $\mathbb{N}$:natural numbers.
- $\mathcal{R}(A)$: the range of matrix A.
- $\mathcal{N}(A)$: the null space of matrix A.
- V: vector space.
- $det(A)$:the determinant of matrix A.
- $rank(A)$:the rank of matrix A.
- $\rho(A)$:the spectral radius of matrix A.
- $Tr(A)$:the trace of matrix A.
- $L^2[a,b]$:Lebesgue integrable function on $[a,b]$.
- $L^1[a,b]$:Lebesgue integrable function on $[a,b]$.
- $N(0,1)$: standard Gaussian distribution.
- $N(\mu,\sigma^2)$: Gaussian distribution with mean μ and variance σ^2.
- $MN(\mu,\Sigma)$: multivariate Gaussian distribution with mean vector μ and covariance matrix Σ.

Mathematical finance notation

- $E_Q[\cdot]$: expectation taken with respect to risk-neutral measure Q.
- $\mathcal{F}_t$: σ algebra geneated by a driving stochastic process, say Bronwian motion, up to time t.

- r: risk-free short rate.
- $B(t), M(t)$:stochastic bank account or money market account; $B(t) = M(t) = \exp(\int_0^t r(s)ds)$.
- $DF(t_1, t_2)$: discount factor (a deterministic quantity) discounting cash flow from t_2 to t_1; $DF(t_1, t_2) = E_Q[\exp(-\int_{t_1}^{t_2} r(s)ds)|\mathcal{F}_{t_1}]$.
- $P(t, T)$: time-t price of a zero coupon bond maturing at T; $P(t, T) = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t]$.
- $df(t_1, t_2)$: stochastic discount factor discounting cash flow from t_2 to t_1; $df(t_1, t_2) = \exp(-\int_{t_1}^{t_2} r(s)ds)$.
- $DF(t_1, t_2)$: discount factor (a deterministic quantity) discounting cash flow from t_2 to t_1; $DF(t_1, t_2) = E_Q[\exp(-\int_{t_1}^{t_2} r(s)ds)|\mathcal{F}_{t_1}]$.
- $L(S, T)$: LIBOR interest rate over period from S to T.
- $F(t, S, T)$: forward LIBOR interest rate over period from S to T observed at time t.
- $S(t, T_0, T_n)$ par swap rate for a fixed-floating swap starting from T_0 to T_n.
- $A(t, T_0, T_n)$ annuity with payments starting from T_0 to T_n.
- $EE(t)$ undiscounted expected exposure; $EE(t) = E_N[V(t)|\mathcal{F}_0]$, where E_N is the expectation with respect to an appropriate martingale measure.
- $EPE(t), ENE(t)$ undiscounted expected positive exposure and negative exposure; $EPE(t) = E_N[V(t)^+|\mathcal{F}_0]$, $ENE(t) = E_N[V(t)^-|\mathcal{F}_0]$, where $E_N[\cdot]$ is the expectation taken with respect to the equivalent martingale measure associated with numeriare $N(t)$.
- CCY1/CCY2 spot rate, where CCY2 is the quoting currency and CCY1 is the base currency, shows the amount of CCY1 needed to exchange for 1 unit of CCY2.

CONTENTS

LIST OF ALGORITHMS

LIST OF FIGURES

LIST OF TABLES

CONTENTS (DETAILED)

Part I

MATHEMATICAL FOUNDATIONS

1

THEORY OF PROBABILITY

1.1 Sigma algebra

1.1.1 sigma algebra concepts

Definition 1.1.1 (σ algebra). *Given a set Ω, a σ-field, or σ-algebra is a collection $\mathcal{F}$ of subsets of Ω, with the following properties:*

1. $\emptyset \in \mathcal{F}$
2. *if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$*
3. *(countable union)if $A \in \mathcal{F}$, then $\cup_{i=0}^{\infty} A_i \in \mathcal{F}$*

Example 1.1.1.

1. The trivial σ-field $\mathcal{F} = \{\emptyset, \Omega\}$
2. The collection $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, where A is a fixed subset of Ω
3. The set of all the subsets of finite set Ω.
4. For a finite sample space Ω, the power set of Ω is the largest σ field, $\{\emptyset, \Omega\}$ is the smallest σ field.

Remark 1.1.1. The pair $(X, \mathcal{F})$ is called **measurable space**, the members $e \in \mathcal{F}$ are called **measurable sets** or $\Sigma-$measurable sets.

Lemma 1.1.1 (intersection theorem). *[1] If $\{\mathcal{F}_\alpha\}_{\alpha \in T}$ is a collection of σ fields on Ω, then $\cap_{\alpha \in T} \mathcal{F}_\alpha$ is σ field on Ω*

Proof. We consider the special case $T = \{1, 2\}$. Let $A = \mathcal{F}_1 \cap \mathcal{F}_2$. It is easy to see $\emptyset \in A$; since $A \in \mathcal{F}_1 \cap \mathcal{F}_2$, then $A \in \mathcal{F}_1, A \in \mathcal{F}_2$, then $A^c \in \mathcal{F}_1, A^c \in \mathcal{F}_2$, then $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$; Similarly, we can prove the union property. □

1.1.2 Generation of sigma algebra

Lemma 1.1.2 (Existence of smallest σ field, σ algebra generation). *If $\mathcal{A}$ is a collection of subsets of Ω, then there exist a unique smallest σ field on Ω, containing $\mathcal{A}$, which is contained by all the σ fields that contains $\mathcal{A}$. We denote this by $\mathcal{F}(\mathcal{A})$, and called the σ field generated by $\mathcal{A}$.[1]*

Proof. Consider $\mathcal{B}$ as the set of all σ fields that contains $\mathcal{A}$. The intersections of all these sets will lead to $\mathcal{F}(\mathcal{A})$ due to theorem 1.1.1. □

Definition 1.1.2 (sigma algebra generated by an event). *Let A be a subset of a sample space Ω. The sigma algebra generated by A, denoted by $\sigma(A)$, is a set given by*

$$\sigma(A) = \{\emptyset, \Omega, A, A^c\}.$$

Remark 1.1.2 (sigma algebra generated by random variable and stochastic process)**.** The generation of sigma algebra by random variables and stochastic processes are discussed in Definition 1.3.5Definition 5.1.5.

Corollary 1.1.0.1. *(Properties of generated σ algebra) If $\mathcal{A}, \mathcal{A}_1$ and $\mathcal{A}_2$ are subsets of 2^Ω, then we have [1]*

- *If $\mathcal{A}_1 \subset \mathcal{A}_2$, then $\mathcal{F}(\mathcal{A}_1) \subset \mathcal{F}(\mathcal{A}_2)$*
- *If $\mathcal{A}$ is a σ field, then $\mathcal{F}(\mathcal{A}) = \mathcal{A}$*
- *If $\mathcal{F}(\mathcal{F}(\mathcal{A})) = \mathcal{F}(\mathcal{A})$*

1.1.3 Partition of sample space

Definition 1.1.3. *A collection of subsets of Ω, $\{\mathcal{A}_i\}_{i\in I}$ (I can have size of uncountable infinite)is called a partition of Ω if*

$$\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \text{ if } i \neq j$$

and

$$\cup A_i = \Omega.$$

Lemma 1.1.3. *[1] If $\mathcal{P} = \{A_i\}_{i\in 1}^{\infty}$ is a countable partition of Ω, then the σ field generated from $\mathcal{P}$, $\mathcal{F}(\mathcal{P})$, consists of all sets of the form $\cup_{n\in M} A_n$ where M ranges over all subsets of $\mathbb{N}$.*

Lemma 1.1.4. *Let $\mathcal{P}_1, \mathcal{P}_2$ be the partitions of the same sample space Ω. If $\mathcal{P}_2$ is obtained by subdividing sets in $\mathcal{P}_1$ (i.e. $\mathcal{P}_2$ is finer), then we have*

$$\mathcal{F}(\mathcal{P}_1) \subseteq \mathcal{F}(\mathcal{P}_2) \Leftrightarrow \mathcal{P}_1 \subseteq \mathcal{P}_2$$

1.1.4 Filtration & information

Definition 1.1.4 (filtration). *Let $(\Omega, \mathcal{F})$ denote a measurable space.*

- *A **continuous filtration** is defined as: A family of σ algebras $\{\mathcal{F}_t | t >= 0\}$ where*

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, 0 \leq s \leq t$$

- *A **discrete filtration** on $(\Omega, \mathcal{F})$ is an increasing sequence of σ fields $\{\mathcal{F}_n\}$ such that:*

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 ... \subseteq \mathcal{F}$$

Note that as the time progresses, the finer the σ algebra will be. We call $\mathcal{F}_n$ the history up to time n.

Remark 1.1.3. Note that usually not all the subsets of X can be defined a measure with above properties. For example, all irrational numbers in the real line, the root to polynomial equation, are not measurable sets.[2]

Remark 1.1.4 (filtration and information)**.**

- Let $\mathcal{F}_1, \mathcal{F}_2$ be two σ field on Ω, then $\mathcal{F}_1 \subseteq \mathcal{F}_2$ mean $\mathcal{F}_2$ contains more information than $\mathcal{F}_1$; For any A is measurable with respect to $\mathcal{F}_1$ then A is measurable with respect to $\mathcal{F}_2$. That is, if $A \subset \mathcal{F}_1$ then $A \subset \mathcal{F}_2$.

Example 1.1.2. For example, in a die toss example, $\mathcal{F}_1$ is generated by the events of odd number or even number, while $\mathcal{F}_2$ is generated by the event of all possible outcomes. Then, we have $\mathcal{F}_1 \subset \mathcal{F}_2$,i.e., knowing the probability measure on $\mathcal{F}_2$ will enable us to calculate the probability measure on $\mathcal{F}_1$.[1]

Now consider a series of experiment: Let Ω denote the set of all outcomes resulting from tossing a coin three times, the $\Omega = \{(H, H, H), (T, H, H) ..., (T, T, T)\}$. Let $\mathcal{F}_i$ denote the events that have been determined by the end of the i toss. Then $\mathcal{F}_1 = \mathcal{F}(\{(H, \cdot, \cdot), (T, \cdot, \cdot)\})$, where $\cdot$ represent it will range over H, T,i.e., $\mathcal{F}_1$ is generated from a partition of 2. Since we have more information later, we have

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$$

Note that if $\mathcal{F}_2$ represents events determined by the i toss instead of tosses upto i, the above will not hold.

1.1.5 Borel σ algebra

Definition 1.1.5. *[3] A Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union,* *countable intersection, and relative complement. For a topological space X, the collection of all Borel sets on X forms a σ-algebra, known as the Borel σ-algebra. The Borel σ algebra on X is the smallest σ-algebra generated by open sets.*

Remark 1.1.5. Note that the elements like low-dimensional manifold $S \subset \mathbb{R}^m, m < n$ in $\mathbb{R}^n$ will not be in the $\mathcal{B}(\mathbb{R}^n)$, i.e., they cannot be obtained from open set operation defined above.

Note 1.1.1 (open interval close interval conversion). Using countable union and intersection properties, we can convert between open interval and close intervals, for example

- $(a, b) = \cup_{n=1}^{\infty}[a + 1/n, b - 1/n]$
- $[a, b] = \cap_{n=1}^{\infty}(a - 1/n, b]$
- $(a, \infty) = \cup_{n=1}^{\infty}[a, a + n]$
- singleton: $\{a\} = [a, a]$

1.1.6 Measurable set and measurable space

Definition 1.1.6. *Given a set X with its σ field Σ, a function $\mu : \Sigma \to \mathbb{R}$ is called a **measure** if it satisfies: [4]:*

- *Non-negativity: For all $E \in \Sigma$, $\mu(E) >= 0$*
- $\mu(\emptyset) = 0$
- *Countable additivity: For all countable collections $\{E_i\}$ of pairwise disjoint sets in Σ:*

$$\mu(\cup_{i=1} E_i) = \sum_{i=1} \mu(E_i)$$

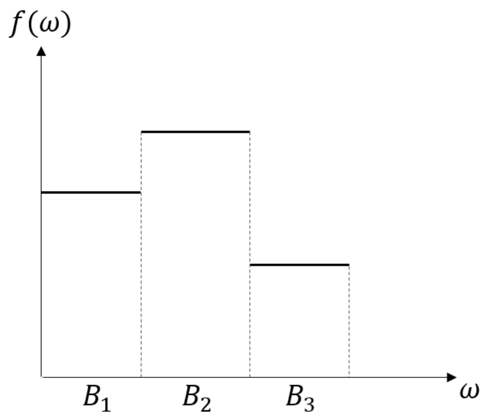
*The pair (X, Σ) is called **measurable space**, the members $e \in \Sigma$ are called **measurable sets** or $\Sigma-$measurable sets. A triple (X, Σ, μ) is called measure space. A **probability measure** is a measure with total measure one $\mu(X) = 1$. A probability space (i.e. a triple) is a measure space with a probability measure.[1]*

Remark 1.1.6. A **measure** on a set is a systematic way to assign a number of each suitable subset of set, as a generalization of the concepts of length, area, and volume.

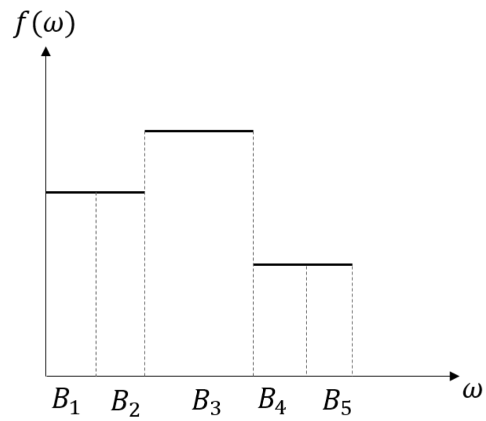
Definition 1.1.7 (measurable function). *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if $f^{-1}(B) \in \mathcal{F}, B \in \mathcal{B}(R)$.*

Theorem 1.1.1 (measurable function with coarse sigma field). *Let $\mathcal{F}$ generated by a finite partitions $B_1, B_2, ..., B_m$ of Ω; let function $f : \Omega \to \mathbb{R}$ be $\mathcal{F}$-**measurable**. Then f takes constant value on each element of $B_i, 1 \leq i \leq m$.*

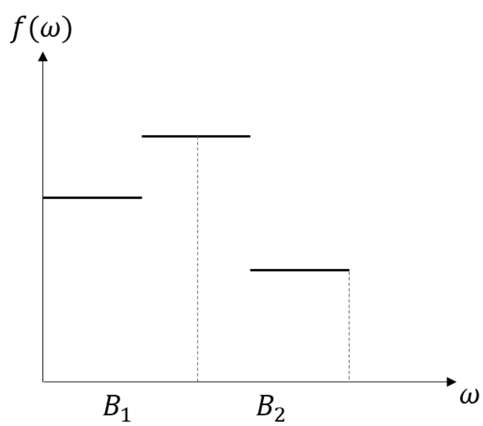
Proof. Suppose f can take different values, say a_1, a_2, then the inverse image of the interval $[a_1, 0.5(a_1 + a_2)]$ is not a subset of $\mathcal{F}$ (note that $\mathcal{F}$ can only contain $\emptyset$ plus subsets due to unions of partition subset. See previous sections on partition of sample space), which contradicts the fact of f is measurable. □



(a) A mesurable function measurable with respect to the σ algebra generated by partition B_1, B_2, B_3.



(b) A mesurable function measurable with respect to the σ algebra generated by partition B_1, B_2, B_3, B_4, B_5.



(c) A function not measurable with respect to the σ algebra generated by partition B_1, B_2.

Figure 1.1.1: An illustration of measurable functions.

Note 1.1.2 (measurable functions vs. ordinary functions)**.**

- Ordinary functions from set A to set B simply establish a relationship between elements in A and elements in B. A measurable function from set A to set B also establish a relationship between elements in A and elements in B, however, under the constraint of measurability of σ fields.
- The level of coarseness constrain the number of values a measurable function can take. For a trivial $\mathcal{F}$, its measurable function can only take one value.
- For random variables, they are required to be measurable functions.

1.2 Probability space

1.2.1 Event, sample point and sample space

Definition 1.2.1 (event,sample point and sample space). *Consider a random experiment. The collection of all outcomes is the sample space Ω. Given a sample space Ω with its σ field $\mathcal{F}$, an event is simply an element in $\mathcal{F}$.*

Remark 1.2.1 (interpretation).

- The results of experiments or observations are called events. For example, the result of a measurement will be called an event. We shall distinguish between *compound*(or decomposable) and *simple*(or indecomposable) *events*. For example, saying that a throw with two dice resulted in "sum six" amounts to saying that it resulted in (1,5) or (2,4) ..., which can be decomposes to five simple events.
- The simple events will be called sample points. Every dis-decomposable result of the experiment is represented by one and only one, sample point. The aggregate of *all* sample points will be called sample space.

1.2.2 Probability space

Definition 1.2.2 (probability space). *[2] A probabilistic model is defined formally by a triple $(\Omega, \mathcal{F}, P)$, called a probability space, where*

1. *Ω is the sample space, the set of possible outcomes of the experiment*
2. *$\mathcal{F}$ is a σ-field, a collection of subsets of Ω, containing Ω itself and the empty set $\emptyset$, and closed under the formation of complements, countable unions, and countable intersections*
3. *P is a **probability measure** defined on σ-field $\mathcal{F}$, and has the property of:*
   - $P(A) \geq 0, \forall A \in \mathcal{F}$
   - *if $A_1, A_2, ... \in \mathcal{F}$ are **disjoints** subsets of Ω, we have **countable additivity** as:* $P(A_1 \cup A_2 \cup A_3 ...) = P(A_1) + P(A_2) + P(A_3) + ...$
   - $P(\Omega) = 1.$

Remark 1.2.2 (interpretation).

- Note that the σ-algebra is the collections of *measurable sets*. These are the subsets $A \subseteq \Omega$ where $P(A)$ is defined. In general, σ-field might not contain *all* subsets of Ω(For example, let Ω be a interval on the real line, then the set of all rational number in the interval is not in σ-field)

- **Note that we cannot extend to *uncountable unions*;** in this case, $\mathbb{F}$ would contain every subset A, since every subset can be written as $A = \cup_{x \in A}\{x\}$ and since the singleton sets $\{x\}$ are all in $\mathbb{F}$.

Definition 1.2.3 (discrete probability space). *A discrete probability space is a triplet* $(\Omega, \mathbb{F}, \mathbb{P})$ *such that*

1. *the sample space is finite or countable*
2. *the σ-field is the set of all subsets of Ω*
3. *the probability measure (a function) assigns a number in the set [0,1] to every pairwise disjoint subset of $A \subseteq \Omega$, given as*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$$

and

$$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$$

Example 1.2.1 (Infinite coin toss process(infinite Bernoulli experiments)). [5, p. 4]

- Consider the probability space for tossing a coin infinitely many time. We can define the sample space as $\Omega_\infty =$ the set of infinite sequences of Hs and Ts. A generic element of Ω_∞ will be denoted as $\omega = \omega_1\omega_2....$, where ω_n indicates the result of the nth coin toss $\omega_n = H \text{ or } T$.
- Example subsets in Ω are
  - A_H: the set of all sequences beginning with H. $A_H = \{\omega : \omega_1 = H\}$.
  - A_T: the set of all sequences beginning with H. $A_T = \{\omega : \omega_1 = H\}$.
  - A_{HT}: the set of all sequences beginning with HT. $A_{HT} = \{\omega : \omega_1 = H, \omega_1 = T\}$.
  - A_{TH}: the set of all sequences beginning with TH. $A_{TH} = \{\omega : \omega_1 = T, \omega_2 = H\}$.
- Possible σ algebra includes:
  - $\mathcal{F}_0 = \{0, \Omega_\infty\}$.
  - $\mathcal{F}_1 = \{0, \Omega_\infty, A_H, A_T\}$.
  -

$$\begin{aligned}\mathcal{F}_2 = \\ & 0, \Omega_\infty A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A_{HH}^C, A_{HT}^C, A_{TH}^C, A_{TT}^C, \\ & A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}\end{aligned}$$

1.2.3 Properties of probability measure

Lemma 1.2.1 (basic properties of probability measure). *[6, p. 11]*

- $P(\emptyset) = 0$.
- *(finite additivity)if* $A_1, A_2, ..., A_n \in \mathcal{F}$ *are* ***disjoints*** *subsets of* Ω*, we have:* $P(A_1 \cup A_2 ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$.
- *For each* $A \in \mathcal{F}$*,* $P(A^C) = 1 - P(A)$*, where* A^C *is the complement of* A *with respect to* Ω.
- *If* $A_1, A_2 \in \mathcal{F}$ *and* $A_1 \subset A_2$*, then* $P(A_1) \leq P(A_2)$.
- *For* $B \subset A$*,* $P(A - B) = P(A) - P(B)$.

Proof. (1) Directly from

$$P(\cup\emptyset) = P(\emptyset) = \sum P(\emptyset)$$

and $P(\emptyset) \geq 0$, we have $P(\emptyset) = 0$. (2) Set $A_{n+1}, A_{n+2}, ... = \emptyset$ and use (1). (3) from (2). (4) note that $A_2 = A_1 + (A_2 - A_1)$ and $P(A_2 - A_1) \geq 0$. (5) Note that $(A - B) \cup B = A$ such that

$$P(B) + P(A - B) = P(A).$$

□

Lemma 1.2.2 (union bound). *For any sequence* $A_1, A_2, ... \in \mathcal{F}$ $P(A_1 \cup A_2 \cup A_3 \cup ...) \leq P(A_1) + P(A_2) + ...$

Proof. Based on countable additivity of probability function, we have:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup ...) &= P(A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2 \backslash A_1) ...) \\ &= P(A_1) + P(A_2 \backslash A_1) ... <= P(A_1) + P(A_2) + ... \end{aligned}$$

□

1.2.4 Conditional probability

1.2.4.1 *Basics*

Remark 1.2.3 (general remarks).

- In some random experiments, we are interested only in those outcomes that are elements of a subset C_1 of the sample space Ω. Then given the probability space $(\Omega, \mathcal{F}, P)$, and $C_1, C_2 \in \mathcal{F}$, the conditional probability of the event C_2, given C_1 is defined as

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}.$$

- Note that usually, for two events C_1, C_2 both occur, we can define an new event $C_3 = C_1 \cap C_2$, then we write $P(C_1, C_2) = P(C_3) = P(C_1 \cap C_2)$. $P(C_1 \cap C_2)$ is quite formal since it is based on set theory.

Definition 1.2.4 (conditional probability measure). *Given a probability space $(\Omega, \mathcal{F}, P)$ and an event $A \in \mathcal{F}, P(A) \neq 0$, we can define a conditional probability measure*

$$P_A(B) \triangleq P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Lemma 1.2.3 (basic properties of conditional probability measure). *[6] Consider the conditional probability measure conditioned on event A. We have*

- $P(B|A) \geq 0, \forall B \in \mathcal{F}$.
- $P(B|A) = 0, \forall B \in \mathcal{F}, A \cap B$.
- $P(A|A) = 1$.
- $P(\cup_{j=1}^{\infty} B_i|A) = \sum_{j=1}^{\infty} P(B_i|A)$, *provided that $B_1, B_2, ... \in \mathcal{F}$ are mutually exclusive event.*
- $\sum_{i=1}^{\infty} P(C_i|A) = 1$, *where $C_1, C_2, ... \in \mathcal{F}$ are the partition of Ω.*

Proof. (4)Use countable additivity property of the definition of probability space Definition 1.2.2, we have

$$P(\cup_{j=1}^{\infty} B_i|A) = \frac{P(\cup_{j=1}^{\infty} B_i \cap A)}{P(A)} = \sum_{j=1}^{\infty} \frac{P(B_i \cap A)}{P(A)} = \sum_{j=1}^{\infty} P(B_i|A).$$

(5) Note that $\cup_{i=1}^{\infty}(C_i \cap A) = A$. $\square$

Lemma 1.2.4 (Law of total probability). *Given a set of subsets $C_1, C_2, ..., C_k$, which are mutual disjoint and partition the sample space Ω, then we have*

$$P(C) = P(C \cap C_1) + P(C \cap C_2) ... + P(C \cap C_k) = \sum_{i=1}^{k} P(C_i) P(C|C_i)$$

Proof. Note that we have $P(C \cap C_i) = P(C_i)P(C|C_i)$, then we get the law of total probability as:

$$P(C) = P(C_1)P(C|C_1) + P(C_2)P(C|C_2) + .. + P(C_k)P(C|C_k) = \sum_{i=1}^{k} P(C_i)P(C|C_i)$$

$\square$

Theorem 1.2.1 (Bayes' theorem). *From the definition of the conditional probability, we have Bayes' theorem as:*

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C_i)P(C|C_i)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)}$$

Proof. The law of total probability has been in the denominator. □

Theorem 1.2.2 (Conditional Bayes' theorem). *From the definition of the conditional probability, we have Bayes' theorem as:*

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C_i)P(C|C_i)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)}$$

Proof. The law of total probability has been in the denominator. □

1.2.4.2 *Independence of events and sigma algebra*

Definition 1.2.5 (independence of event). *Given the probability space* $(\Omega, \mathcal{F}, P)$, *and* $C_1, C_2 \in \mathcal{F}$, *then we say* C_1 *and* C_2 *are independent if*

$$P(C_1 \cap C_2) = P(C_1)P(C_2).$$

Definition 1.2.6 (independence of σ algebras). *Given the probability space* $(\Omega, \mathcal{F}, P)$, *and* $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$, *then we say* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are independent if*

$$P(A_1 \cap A_2) = P(A_1)P(A_2), \forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2.$$

Remark 1.2.4. Note that this is mathematical equivalent definition, which does not reveal the nature of independence in terms of set relationship. **The nature is that if two events are independent, then the occurrence of one event will not change our brief on the occurrence of the other event.**

Example 1.2.2. Consider the sample space of a random experiment is given as $\{(0,0),(0,1),(1,0),(1,1)\}$, with its σ field consists of all its subsets, and we define event $C_1 = \{(0,0),(0,1)\}$, and $C_2\{(0,1)\}$. So the occurrence of C_1 will change the our brief of C_2 from 1/4 to 1/2. Therefore, C_1 and C_2 are not independent to each other. Also, consider $C_3 = C_1^c$, then the occurrence of C_1 change our brief of C_1^c to 0. If $C_4 = \Omega$, then the occurrence of C_4 will not change, and thus C_4 is always independent of other events. In summary, **independence between events is far more complicated than the simple set relations between events**

Remark 1.2.5. Here is an non-trivial example of independence. Consider the sample space as the product of two coin toss sample space, the event that the first toss get 1 is $\{(1,0),(1,1)\}$, which is independent of the other event that the second toss get 1(i.e. $\{(0,1),(1,1)\}$). The two events have finite intersections, but they are independent. Therefore, it seems that simply considering the set relationships between events can not yield complete information of independence. The nature of the random experiment, i.e., the probability measure, dictates the Independence. The intuition way to judge independence will be whether the occurrence of one events provides useful information, i.e., changes our brief, for the occurrence of the other event.

Lemma 1.2.5. *If C_1, C_2 are independent, then C_1 and C_2^c,C_1^c and C_2, C_1^c and C_2^c are independent.*

Lemma 1.2.6. *[1]*

- *If $P(A) > 0$, then A and B are independent if and only if $P(B|A) = P(B)$*
- *If A and B are independent, then A and B^C are independent.*
- *If $P(A) = 0$ or 1, then for any $B \in \mathcal{F}, B \neq A$, A and B are independent.*

Proof. (1) Suppose A and B are independent, then

$$P(A \cap B) = P(A)P(B) = P(A)P(B|A) \implies P(B|A) = P(B).$$

(2)

$$P(A \cap B^C) = P(A \cap (\Omega - B)) = P(A \cap \Omega) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(B^C).$$

(3) If $A = \Omega$ such that $P(A) = 1$,then

$$P(A \cap B) = P(B) = P(A)P(B).$$

If $A = \emptyset$ such that $P(A) = 0$,then

$$P(A \cap B) = P(A) = 0 = P(A)P(B).$$

$\square$

1.3 Measurable map and random variable

1.3.1 Random variable

Definition 1.3.1 (measurable map). *[7] Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable space. A map $T : \Omega \to S$ is called $(\mathcal{F}, \mathcal{S})$-**measurable map** if*

$$T^{-1}(A) \in \mathcal{F}, \forall A \in \mathcal{S}$$

We can also write it as

$$T : (\Omega, \mathcal{F}) \to (S, \mathcal{S}).$$

Definition 1.3.2 (random variables). *Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable space. A measurable map $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is also called a **random variable** if $S = \mathbb{R}^n, \mathcal{S} = \mathcal{B}(\mathbb{R}^n)$; X is called a **n-dimensional real-valued random vector**if $S = \mathbb{R}, \mathcal{S} = \mathcal{B}(\mathbb{R})$.*

Remark 1.3.1 (interpretation).

- Note that the measurability property of the map enables us to measure some suitable subsets of the image S of random variable X *consistently* with the measure in the origin measurable space $(\Omega, \mathcal{F})$.

Lemma 1.3.1 (basic properties of measurable map). *Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable space. Let a map $T : \Omega \to S$ be a measurable map. Then we have:*

- *For any two disjoint sets $S_1, S_2 \in \mathcal{S}$, $T^{-1}(S_1)$ and $T^{-1}(S_2)$ are disjoint.*
- *$T^{-1}(S) = \Omega$.*
- *(measurable composition preserves measurability)Let G be a measurable map from $(S, \mathcal{S})$ to $(S, \mathcal{S})$. Then $G \circ T : \Omega \to S$ is a measurable map.*

Proof. (1) Suppose their inverse image intersection M is nonempty, then $T(m), m \in M$ will map a single element to two different elements in S, which violates the definitions of mapping. (2) Suppose $T^{-1}(S) = \Omega_1 \subset \Omega$ and $\Omega_1 \neq \Omega$, then $T(\Omega - \Omega_1) = \emptyset$(otherwise T will map a single element to two different elements). Therefore $T^{-1}(S \cup \emptyset) = T^{-1}(S) = \Omega$. (3) Note that $(G \circ T)^{-1}(B) = T^{-1} \circ G^{-1}(B) = T^{-1}(G^{-1}(B))$ Because G is measurable map, $G^{-1}(B) \in \mathcal{S}$. Because T is measurable map, $T^{-1}(G^{-1}(B)) \in \mathcal{F}$. Therefore, $G \circ T$ is a measurable from Ω to S. □

Lemma 1.3.2 (basic measurability properties of random variables). *Let $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B})$ be two measurable space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then we have:*

- *(measurable composition preserves measurability)Let f be a measurable map from $(S, \mathcal{S})$ to $(S, \mathcal{S})$. Then $f(X) : \Omega \to \mathbb{R}$ is a measurable map.*
- *Let Y be another random variable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$. Then $\alpha X + \beta Y : \Omega \to \mathbb{R}, \alpha, \beta \in \mathbb{R}$ is also a measurable map.*
- *Let Y be another random variable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$. Then $XY : \Omega \to \mathbb{R}$ is also a measurable map.*
- *Let $Y, Y \neq 0$ be another random variable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$. Then $1/Y : \Omega \to \mathbb{R}$ is also a measurable map.*

Proof. (1) Use the composition property of measurable map(Lemma 1.3.1). (2)(3)(4) use **??**. □

Remark 1.3.2 (implications). This theorem provides the foundation of when X and Y are random variables, usually, $f(X), X+Y, XY, X/Y, ...$ are also random variables.

1.3.2 Image measure

Definition 1.3.3 (image measure). *[7] Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ denote a random variable and P a measure on the measurable space $(\Omega, \mathcal{F})$. Then*

$$P_X(A) := P(X^{-1}(A)), \forall A \in \mathcal{S}$$

defines a probability measure on $(S, \mathcal{S})$, which we call the image measure of P_X with respect to X.

Lemma 1.3.3 (generation of probability space via random variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. For each Borel set $B \in \mathcal{B}, B \subset \mathbb{R}$, we have $X^{-1}(B) \in \mathcal{F}$, then we can define $P_X(B) = P(X^{-1}(B))$. Then prove $(\mathbb{R}, \mathcal{B}, P_X)$ is a probability space.*

Proof. First $(\mathbb{R}, \mathcal{B})$ form a measurable space. So we only need to check the axiom property of P_X: (1) $P_X(A) \geq 0, \forall A \in \mathcal{B}$; (2) For any two disjoint sets A_1, A_2, then

$$P_X(A_1 \cup A_2) = P(X^{-1}(A_1 \cup A_2)) = P(X^{-1}(A_1) \cup X^{-1}(A_2)) = P(X^{-1}(A_1) + P(X^{-1}(A_2)).$$

We can directly generalize to countable additivity. (3) $P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$(from lemma on basic properties of measurable maps) □

Remark 1.3.3. This lemma has significant consequences that **it enables us to directly work on this generated probability space $(\mathbb{R}, \mathcal{B}, P_X)$ and investigate distribution, density functions etc without referring back to the original probability space.**

1.3.3 Product measure

Definition 1.3.4. *[8] Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure space. Then there exists an unique measure $\omega = \mu \times \nu$, called product measure, defined on $(X \times Y, \mathcal{X} \vee \mathcal{Y})$ such that*

$$\omega(A \times B) = \mu(A)\nu(B), A \in \mathcal{X}, B \in \mathcal{Y}$$

The σ field $\mathcal{X} \vee \mathcal{Y}$ is defined formally as the smallest σ field containing all sets $A \times B$ with $A \in \mathcal{X}$ and $B \in \mathcal{Y}$.

1.3.4 σ algebra of random variables

Definition 1.3.5 (σ algebra generated by random variables). *[5, p. 52] Let X be a random variable map from nonempty Ω to $\mathbb{R}$. The σ**algebra generated** by X, denoted by $\sigma(X)$, is the collection of all subsets of Ω of the form $\{\omega \in \Omega : X(\omega) \in B\}$, or equivalently $X^{-1}(B)$, where B ranges over all Borel subsets of $\mathbb{R}$.*

Remark 1.3.4 (interpretation)**.**

- When we define the measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$, usually $\sigma(X) \subseteq \mathcal{F}$. For example, if $X = const$, then $\sigma(X) = \mathcal{F}_0 = \{\emptyset, \Omega\}$.
- We cannot have $\mathcal{F} \subset \sigma(X), \mathcal{F} \neq \sigma(X)$ because the definition of random variable require measurability.

Definition 1.3.6 (measurable random variables with respect to a σ algebra). *[5, p. 53] Let X be a random variable map from nonempty Ω to $\mathbb{R}$. Let $\mathcal{G}$ be the σ algebra defined on Ω. We say X is $\mathcal{G}$ measurable if $\sigma(X) \subseteq \mathcal{G}$.*

Remark 1.3.5 (interpretation)**.**

- Note that for any $B \in \mathcal{B}$, $X^{-1}(B) \in \sigma(X) \subseteq \mathcal{G}$, therefore X is also $\mathcal{G}$-measurable.
- Given a set Ω, we can define different σ algebra, including $\mathcal{F}_0 = \{\emptyset, \Omega\}$. But only σ algebra finer than $\sigma(X)$ can measure the mapping X.

1.3.5 Adapted process

Definition 1.3.7. *Let X denote a stochastic process on $(\Omega, \mathcal{F})$ and $\{\mathcal{F}_t\}$ is a filtration on $(\Omega, \mathcal{F})$. The process X is called $\{\mathcal{F}_t\}$-adapted, if X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$*

Remark 1.3.6.

- Adapted process is a special process that calculation of the probability measure for X_t requires the information upto i, therefore we need $\mathcal{F}_t$ to contain all the information before such that X_t is $\mathcal{F}_t$-measurable. For example, consider a discrete stochastic process $X_n \sum_{i=1}^{n} W_i$, where W_i is independent random variable taking value 0 or 1. If we want to calculate $P(X_4 = 3)$, then the $\mathcal{F}_4$ should contains every event/path that leads to $X_4 = 3$ (for example, $e = \{W_1 = 1, W_2 = 1, W_3 = 1, W_4 = 0\}$). And obviously $\mathcal{F}_3 \subset \mathcal{F}_4$.
- For a Bernoulli process, the calculation of probability measure for X_t does not requires the all information before. Therefore, even if $\mathcal{F}_t$ does not contain all previous information, X_t is still $\mathcal{F}_t$-measurable.

1.3.6 Independence of random variables

Definition 1.3.8 (independence of random variables). *Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $Y : (\Omega, \mathcal{F}) \to (S, \mathcal{S}))$ denote two random variables. We say X, Y are independent, if for all $A, B \in \mathcal{S}$ the events $X^{-1}(A)$ and $Y^{-1}(B)$ are independent in the sense that $P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A))P(Y^{-1}(B))$.*

Definition 1.3.9 (independence of random variables, alternative). *[5, p. 54] Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $Y : (\Omega, \mathcal{F}) \to (S, \mathcal{S}))$ denote two random variables. We say X, Y are independent, if*

$$P(A \cap B) = P(A)P(B), \forall A \in \sigma(X), B \in \sigma(Y).$$

Remark 1.3.7. Note that

- independence of random variables are much more than independence of events, because it requires *all* the pre-images (i.e. events) are independent to each other.
- if X, Y are map from different sample space, then they are independent.

Lemma 1.3.4 (function composition preserves random variable independence). *Let X, Y be independent random variables defined from Ω to , and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.*

Proof. Note that for any $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{B}$ since f is Borel measurable. Then $X^{-1}(f^{-1}(B)) \in \sigma(X)$ based on the definition of σ generation. Therefore, $\sigma(f(X)) \subset \sigma(X)$. Similarly, $\sigma(g(Y)) \subset \sigma(Y)$. Since every events in $\sigma(X)$ and $\sigma(Y)$ are independent, then every events in $\sigma(f(X))$ and $\sigma(g(Y))$ are independent; that is, $f(X)$ and $g(Y)$ are independent random variables. $\square$

1.3.7 Conditional independence

Definition 1.3.10 (conditional independence). *Given discrete random variables X, Y, and Z, we say X and Y are conditionally independent on Z if we can be write:*

$$P(X, Y|Z = z) = P(X|Z = z)P(Y|Z = z).$$

If not conditionally independent, we will have

$$P(X, Y|Z = z) = P(X|Y, Z = z)P(Y|Z = z).$$

Remark 1.3.8. Intuitively, two random variable X, Y are conditional independence given Z is that: if the value of Z is known, X, Y are independent to each other,i.e., the occurrence of events about Y will not give extra information to the occurrence of events abouts X. We need to distinguish two different cases:

- If X, Y are independent, then they are conditionally independent to each other.
- If events about Z already gives information contained in events about Y, then X, Y are conditionaly independent given Z.

Remark 1.3.9. Conditionally independence will help us simplify calculation, for example:

$$P(X = x|Y = y, Z = z) = P(X = x|Z = z)$$

if X, Y are conditionally independent given Z.

1.4 Distributions of random variables

1.4.1 Basic concepts

1.4.1.1 *Probability mass function*

Definition 1.4.1 (random variable, random vector). *[6, p. 75]*

- *Let X be a random variables maps from the probability space $(\Omega, \mathcal{F}, P)$ to $\mathbb{R}$. The **space** of the random variable X is the set*

$$\{(X(\omega) : \omega \in \Omega\}.$$

- *Let $X_1, X_2, ..., X_n$ be random variables maps from the probability space $(\Omega, \mathcal{F}, P)$ to $\mathbb{R}$. We say $(X_1, X_2, ..., X_n)$ is a random vector. The **space** of the random vector $(X_1, X_2, ..., X_n)$ is the set*

$$\{(X_1(\omega), X_2(\omega), ..., X_n(\omega)) : \omega \in \Omega\}.$$

Definition 1.4.2 (probability mass function). *[9]*

- *For a discrete random variable X with space $\mathcal{D}$, the **probability mass function** to characterize its distribution is given by*

$$f_X(x) = P(X = x), \forall x \in \mathcal{D}.$$

- *For a discrete random vector $(X_1, X_2, ..., X_n)$ with space $\mathcal{D}$, the **joint probability mass function** to characterize its distribution is given by*

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n), \forall (x_1, x_2, ..., x_n) \in \mathcal{D}.$$

1.4.1.2 *Distributions on $\mathbb{R}^n$*

Definition 1.4.3 (cumulative distribution functions).

- *Let X be a random variable with space $\mathcal{D} \subset \mathbb{R}$. The cumulative distribution function for X is given by*

$$F_X(x) = Pr(X \leq x).$$

- *Let $(X_1, X_2, ..., X_n)$ be a random vector with space $\mathcal{D} \subset \mathbb{R}^n$. The joint cumulative distribution function for X is given by*

$$F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = Pr(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n).$$

Remark 1.4.1 (A rigorous interpretation). [7]

- Let P denotes a probability measure of the original probability space $(\Omega, \mathcal{F}, P)$. Let X be the random variable, then

$$F_X(x) := Pr(X \leq x) = P(X^{-1}((-\infty, x))).$$

  where X^{-1} maps a measurable subset in $\mathcal{B}(\mathbb{R})$ to a measurable set in $\mathcal{F}$.
- Note that every subset of such form $(-\infty, x), x \in \mathbb{R}$ is a member of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and therefore $Pr(X \leq x) = P(X^{-1}((-\infty, x)))$ has a well-defined value.

Definition 1.4.4 (marginal cdf). *Let $(X_1, X_2, ..., X_n)$ be a random vector with joint cdf $F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. The marginal cdf of X_i is defined by*

$$F_{X_i}(x_i) = Pr(X_1 < \infty, ..., X_i \leq x_i, ..., X_n < \infty).$$

Lemma 1.4.1 (area probability formula). *[6, p. 76] Let X_1, X_2 be random variables with joint cdf $F_{X_1,X_2}(x_1, x_2)$. Then*

$$\begin{aligned} &Pr(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) \\ =&F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2) \end{aligned}$$

Proof. Straight forward. □

1.4.1.3 *Probability density function*

Definition 1.4.5 (probability density function, pdf).

- *Let X be a random variable with cdf $F_X(x)$. The probability density function for X is defined by*

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

- *Let* $(X_1, X_2, ..., X_n)$ *be a random vector with joint cdf* $F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. *The joint probability density function for* $(X_1, X_2, ..., X_n)$ *is given by*

$$F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \frac{\partial^n F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

Definition 1.4.6 (support of a random variable). *Let* X *be a random variable with pdf* f_X *and space* $\mathcal{D}$. *The support of* X *is defined as the set*

$$S_X = \{x \in \mathcal{D} : f_X(x) > 0\}.$$

Definition 1.4.7 (marginal pdf). *Let* $(X_1, X_2, ..., X_n)$ *be a random vector with joint pdf* $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. *The marginal pdf of* X_i *is defined by*

$$f_{X_i}(x) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f_{X_1,X_2,...,X_n} dx_1...dx_{i-1}dx_{i+1}...dx_n.$$

If the marginal cdf of X_i *is* F_{X_i}, *then*

$$f_{X_i}(x) = \frac{dF_{X_i}(x)}{dx}.$$

1.4.1.4 *Conditional distributions*

Definition 1.4.8 (conditional pdf). *[6, p. 97]*

$$f_{X_2|X_1}(x_2; x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)}.$$

Lemma 1.4.2 (basic properties). *[6, p. 97]*

- $f_{X_2|X_1}(x_2; x_1) > 0$
- $\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2; x_1) dx_2 = 1$.
- $\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2; x_1) f_{X_1}(x_1) dx_1 = f_{X_2}(x_2)$.
- $E[u(X_2)|x_1] = \int_{-\infty}^{\infty} f_{X_2|X_1}(x_2; x_1) u(x_1) dx_2$

Theorem 1.4.1 (Bayesian law for random variables). *Let* X, Y, Z *be random variables. It follows that*

- *(unconditional Bayesian law)*

$$f(X|Y) = \frac{f(Y|X)f(X)}{\int f(Y|X)f(X)dx}$$

- *(conditional Bayesian law)*

$$f(X|Y,Z) = \frac{f(X|Z)f(Y|X)}{\int f(Y|X)f(X|Z)dx}$$

Proof. (1) Note that the denominator $\int f(Y|X)f(X)dx = f(Y)$. Therefore

$$f(X|Y)\int f(Y|X)f(X)dx = f(X,Y) = f(Y|X)f(X).$$

(2) Note that

$$\int f(Y|X)f(X|Z)dx = \int f(Y,X|Z)dx = f(Y|Z).$$

Therefore,

$$f(X|Y,Z)\int f(Y|X)f(X|Z)dx = f(X|Y,Z)f(Y|Z) = f(X,Y|Z).$$

□

1.4.2 Independence

Definition 1.4.9 (independence of random variables). *[6, pp. 112, 115] Let the random variables X_1 and X_2 have joint pdf $f(x_1,x_2)$ and marginal pdfs $f_1(x_1), f_2(x_2)$.*

- *The random variables X_1 and X_2 are said to be independent if and only*

$$Pr(a < X_1 \leq b, c < X_2 \leq d) = Pr(a < X_1 \leq b, c < X_2 \leq d),$$

for every $a < b, c < d$, where a,b,c,d are constants.
- *The random variables X_1 and X_2 are said to be independent if and only*

$$f(x_1,x_2) = f_1(x_1)f_2(x_2).$$

Lemma 1.4.3 (conditions for independence). *[6, p. 113] Let the random variables X_1 and X_2 have supports S_1 and S_2, and have joint pdf $f(x_1,x_2)$. Then X_1 and X_2 are independent if and only if $f(x_1,x_2)$ can be written as*

$$f(x_1,x_2) = g(x_1)h(x_2),$$

where $g(x_1) > 0, x_1 \in S_1$, zero elsewhere, and $h(x_2) > 0, x_2 \in S_2$, zero elsewhere.

Proof. (1) If $f(x_1, x_2) = g(x_1)h(x_2)$, then

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = c_2 g(x_1).$$

$$f_{X_2}(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = c_1 h(x_2).$$

Further we have

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1dx_2 = c_1c_2.$$

Therefore, $f(x_1, x_2) = c_1c_2g(x_1)h(x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$; that is, X_1, X_2 are independent.

(2) The other direction directly from definition. □

Lemma 1.4.4. *[6, p. 114] Let the random variables X_1 and X_2 have joint cdf $F(x_1, x_2)$ and marginal cdfs $F_1(x_1), F_2(x_2)$. Then X_1 and X_2 are independent if and only if*

$$F(x_1, x_2) = F_1(x_1)F_2(x_2).$$

Proof. (1) From Lemma 1.4.1, we have

$$\begin{aligned}
&Pr(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) \\
=&F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2) \\
=&F_{X_1}(b_1)F_{X_2}(b_2) - F_{X_1}(a_1)F_{X_2}(b_2) - F_{X_1}(a_2)F_{X_2}(b_1) + F_{X_1}(a_1)F_{X_2}(a_2) \\
=&(F_{X_1}(b_1) - F_{X_1}(a_1))(F_{X_2}(b_2) - F_{X_2}(a_2)) \\
=&Pr(a_1 \leq X_1 \leq b_1)Pr(a_2 \leq X_2 \leq b_2)
\end{aligned}$$

Since a_1, a_2, b_1, b_2 are arbitrary, X_1 and X_2 are independent. (2)The other direction directly from definition. □

Lemma 1.4.5 (independence from moment generating functions). *link Let X and Y be two random variables with space $\mathbb{R}$. Assume the moment generating functions for X, Y and $X + Y$ exist at the neighborhood of 0. If for all t_X, t_Y in the neighborhood of 0 we have*

$$E[\exp(t_X X + t_Y Y)] = E[\exp(t_X X)]E[\exp(t_Y Y)],$$

then X and Y are independent.

Proof. Let (U, V) be such that U and V are independent; moreover, U and X have the same distribution and V and Y have the same distribution.

$$E[\exp(t_X X + t_Y Y)] = E[\exp(t_X X)]E[\exp(t_Y Y)] = E[\exp(t_X U)]E[\exp(t_Y V)] = E[\exp(t_X U + t_Y V)],$$

Therefore (X, Y) and (U, V) have the same joint distribution; that is, X and Y are independent. □

1.4.3 Transformations

1.4.3.1 *Transformation for univariate distribution*

Lemma 1.4.6 (change of variable). *[9, p. 77] Let X have cdf $F_X(x)$ and Let $Y = g(X)$, where g is a* ***monotonely increasing*** *function. Then,*

$$F_Y(y) = F_X(g^{-1}(y)).$$

If g is a ***monotonely decreasing function****, then*

$$F_Y(y) = 1 - F_X(g^{-1}(y)).$$

Proof. If g is increasing function

$$P(Y < y) = P(g(X) < y) = P(X < g^{-1}(y)) = F_X(g^{-1}(y)).$$

If g is decreasing function

$$P(Y < y) = P(g(X) < y) = P(X > g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

□

Lemma 1.4.7 (change of variable). *[9, p. 77] Let X have pdf $f_X(x)$ and Let $Y = g(X)$, where g is a* ***monotone*** *function. Let X and Y be defined as*

$$\mathcal{X} = \{x : f_X(x) > 0\}, \mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$$

then we have

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, y \in \mathcal{Y} \\ 0, otherwise \end{cases}$$

Remark 1.4.2 (why monotonity). We require the $g(X)$ to be monotone because if $g'(x)$ has different sign on different regions, then $g'(x_0) = 0$ for some x_0 and $g(x)$ is not invertible near the neighborhood of x_0.

Corollary 1.4.1.1. *Let $Y = g(X)$, where g is a monotone function, let $m(x)$ be a function, then*

$$\int_{\mathcal{Y}} m(y) f_Y(y) dy = \int_{\mathcal{X}} m(g(x)) f_X(x) dx$$

Proof.

$$\int_{\mathcal{Y}} m(y) f_Y(y) dy = \int_{\mathcal{Y}} m(y) f_X(g^{-1}(y)) \left| dx/dy \right| dy$$

Let $y = g(x), dy = (dy/dx)dx$, then

$$\int_{\mathcal{Y}} m(y) f_X(g^{-1}(y)) \left| dx/dy \right| dy = \int_{\mathcal{X}} m(g(x)) f_X(x) dx$$

□

1.4.3.2 *Location-scale transformation*

Definition 1.4.10. *[9] (Location-scale family)Let $f(x)$ be any pdf. Then for any $\mu, -\infty < \mu < \infty$, and any $\sigma > 0$, the family of pdfs*

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is the location-scale family indexed by μ, σ.

Remark 1.4.3. When $\sigma > 1$, we strech the original pdf; when $\sigma < 1$, we contract it.

Lemma 1.4.8 (location-scale transformation). *[9, p. 116] Let $f(\cdot)$ be any pdf. Then for any $\mu, -\infty < \mu < \infty$, and any $\sigma > 0$:*

- *X is a random variable with pdf*

  $$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

  if and only if $X = \sigma Z + \mu$ and Z has pdf $f(z)$.
- $F_X(x) = F_Z(\frac{x-\mu}{\sigma})$
- $F_X^{-1}(\alpha) = \mu + \sigma F_Z^{-1}(\alpha), \forall \alpha \in [0,1]$, *where* $F_X^{-1}(\alpha) = \inf\{Pr(X < x) \geq \alpha\}$.

- $EX = \sigma EZ + \mu, Var(X) = \sigma^2 Var(Z)$

Proof. Directly from Lemma 1.4.7Corollary 1.4.1.1. For (1)(2)

$$\begin{aligned} F_X(x) &= Pr(X < x) \\ &= Pr(\mu + \sigma Z < x) \\ &= Pr(Z < (x-\mu)/\sigma) \\ &= F_Z((x-\mu)/\sigma) \end{aligned}$$

Then

$$f_X(x) = dF_X(x)/dx = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}).$$

(3)

$$\begin{aligned} \alpha &= Pr(X < F_X^{-1}(\alpha)) \\ &= Pr(\sigma Z + \mu < F_X^{-1}(\alpha)) \\ &= Pr(Z < (F_X^{-1}(\alpha) - \mu)/\sigma) \\ \alpha &= F_Z((F_X^{-1}(\alpha) - \mu)/\sigma) \end{aligned}$$

$$\begin{aligned} F_Z^{-1}(\alpha) &= (F_X^{-1}(\alpha) - \mu)/\sigma \\ \mu + \sigma F_Z^{-1}(\alpha) &= F_X^{-1}(\alpha). \end{aligned}$$

□

Example 1.4.1. Consider the random variable $X \sim N(0,1)$ with

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}).$$

Let $Y = \sigma X + \mu$, then

$$f_Y(y) = \frac{1}{\sigma} f_X(\frac{y-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

1.4.3.3 *Transformation for multivariate distribution*

Lemma 1.4.9 (multivariate transformation). *[6, p. 128] Let $(X_1, X_2, ..., X_n)$ be a random vector with support $\mathcal{S}$. Let*

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

define a set of transformations with inverse

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

Let $\mathcal{T}$ be the image of $\mathcal{S}$ under the transformation.

Let $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ be the joint pdf of $(X_1, X_2, ..., X_n)$. Then the joint pdf for the random vector $(Y_1, Y_2, ..., Y_n)$ is given by

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = f_{Y_1,Y_2,...,Y_n}(y_1(x_1, x_2, ..., x_n), ..., y_n(x_1, x_2, ..., x_n))|J|$$

or

$$f_{X_1,X_2,...,X_n}(x_1(y_1, y_2, ..., y_n), ..., x_n(y_1, y_2, ..., y_n)) = f_{Y_1,Y_2,...,Y_n}(y_1, y_2, ..., y_n)|J|$$

where

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

Moreover,

$$\int_{\mathcal{T}} f_{Y_1,...,Y_n}(y_1, ..., y_n)|J|\, dy_1...dy_n = 1$$

Proof. (1)For S be a measurable subset in $\mathcal{S}$, let $T \in \mathcal{T}$ denote the its image under the transformation. We have

$$Pr((Y_1, Y_2, ..., Y_n) \in T) = Pr((X_1, X_2, ..., X_n) \in S)$$

Note that

$$Pr((X_1, X_2, ..., X_n) \in S) = \int_S f_{X_1,...,X_n}(x_1, ..., x_n)dx_1...dx_n$$

and

$$dx_1 dx_2 = |J|\, dy_1 dy_2.$$

Then

$$\int_S f_{X_1,...,X_n}(x_1, ..., x_n)dx_1...dx_n = \int_T f_{Y_1,...,Y_n}(y_1(x_1, ..., x_n), ..., y_n(x_1, ..., x_n))|J|\, dy_1...dy_n.$$

Because S is arbitrary, we have

$$f_{X_1,...,X_n}(x_1, ..., x_n) = f_{Y_1,...,Y_n}(y_1(x_1, ..., x_n), ..., y_n(x_1, ..., x_n))|J|\,.$$

(2)

$$\int_{\mathcal{T}} f_{Y_1,...,Y_n}(y_1,...,y_n)|J|\,dy_1...dy_n = \int_{\mathcal{S}} f_{X_1,...,X_n}(x_1,...,x_n)dx_1...dx_n = 1$$

□

Remark 1.4.4.

- We interpret dx_1dx_2 as infinitesimal area in the original $\mathcal{S}$, and this area is mapped to an area in $\mathcal{T}$. Note that we divide $\mathcal{S}$ and $\mathcal{T}$ into the same number small areas and the sum them up to calculate the integral. The areas in both $\mathcal{T}$ and $\mathcal{S}$ have the following relation:

$$dx_1dx_2 = |J|\,dy_1(x_1,...,x_2)dy_2(x_1,...,x_n) = |J|\,dy_1dy_2.$$

- If we maps from larger support to a smaller support, for example, from $\mathbb{R}^2$ to $[0,\infty)\times[0,2\pi]$, the density will increase.
-

Lemma 1.4.10 (polar transformation). *Let (X_1,X_2) be a random vector with support $S=\mathbb{R}^2$. Let $R=\sqrt{X_1^2+X_2^2}, \Theta = \arctan(X_1/X_2)$. Then*

- $$f_{R,\Theta}(r,\theta)r = f_{X_1,X_2}(r\cos(\theta), r\sin(\theta)).$$
- $$f_{X_1,X_2}(x_1,x_2)dx_1dx_2 = f_{R,\Theta}(r(x_1,x_2),\theta(x_1,x_2))rdrd\theta.$$
- $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)dx_1dx_2 = \int_0^{\infty}\int_0^{2\pi} f_{R,\Theta}(r,\theta)rdrd\theta = 1.$$
- *The support for (R,Θ) is*
$$\{(0,+\infty)\times[0,2\pi]\}$$

Proof. Note that

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} \implies |J| = r.$$

Therefore

$$f_{X_1,X_2}(x_1,x_2) = f_{R,\Theta}(r(x_1,x_2),\theta(x_1,x_2))r.$$

□

Example 1.4.2. Let $X \sim N(0,1)$ and $Y \sim N(0,1)$. Let $R = \sqrt{X^2+Y^2}, \Theta = \arctan(X/Y)$.

Then

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2}).$$

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos(\theta), r\sin(\theta)) = \frac{1}{2\pi}\exp(-\frac{r^2}{2}).$$

$$\int_0^\infty \int_0^{2\pi} f_{R,\Theta}(r,\theta) r dr d\theta = \int_0^\infty \int_0^{2\pi} \frac{1}{2\pi}\exp(-\frac{r^2}{2}) r dr d\theta = 1.$$

Lemma 1.4.11 (convolution formula). *[6, p. 95] Let X_1 and X_2 be continuous random variables with joint pdf $f_{X_1,X_2}(x_1,x_2)$ with $\mathcal{D} = \mathbb{R}^2$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2$. Then*

- $f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(y_1 - y_2, y_2)$.
- *The pdf of Y_1 is given by*

$$f_{Y_1}(y) = \int_{-\infty}^{\infty} f_{X_1,X_2}(y - y_2, y_2) dy dy_2.$$

Proof. It is easy to see $|J| = 1$. □

1.5 Expectation

1.5.1 Failure of elementary approach

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, if Ω is finite, we can simply define the expectation as

$$E[X] = \sum_{\omega \in \Omega} P(\omega) X(\omega)$$

However, if Ω is countably infinite, we can still list a sequence of $\omega_1, \omega_2, ...$ such that

$$E[X] = \sum_{i=1}^{\infty} P(\omega) X(\omega_i)$$

However, if Ω is uncountably infinite, then **uncountable** summation is not defined, and we need Lebesgue integral.

1.5.2 Formal definitions

Definition 1.5.1 (Lebesgue integral). *[5, p. 15] Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, assume $0 \leq X(\omega) \leq \infty$ for every $\omega \in \Omega$, and let $\Pi : 0 = y_0 < y_1 < ...$ be a partition on the range of $X(\omega)$. For each subinterval $[y_k, y_{k+1}]$, we set*

$$A_k = \{\omega \in \Omega : y_k \leq X(\omega) \leq y_{k+1}\} = X^{-1}([y_k, y_{k+1}])$$

We define the lower Lebesgue sum to be

$$LS^-_\Pi = \sum_{k=1}^{\infty} y_k P(A_k)$$

We further define the limit

$$\lim_{\|\Pi\| \to 0} LS^-_\Pi = \int_\Omega X(\omega) dP(\omega)$$

Remark 1.5.1.

- Because X is measurable maps, its inverse image of any Borel set in $\mathbb{R}$ is measurable, i.e., $P(A)$ has value.
- For $X(\omega)$ that takes positive and negative part, we can simply decompose into two parts and use the linearity.

Definition 1.5.2 (expectation). *Let X be a random variable on a probability space $(\Omega, \mathbb{F}, P)$. The expectation of X is defined to be*

$$EX = \int_{\Omega} X(\omega) dP(\omega)$$

This definition makes sense if X is integrable, i.e., if

$$E|X| = \int_{\Omega} |X(\omega)| \, dP(\omega) < \infty$$

Remark 1.5.2. Note that the integral is defined using Lebesgue integral, and based on this definition we can recover the elementary definitions.

- If X takes only finitely many $x_0, x_1, ..., x_n$, but Ω is uncountable, then

$$EX = \sum_{x_k} x_k P(X = x_k)$$

  and $P(X = x_k)$ is the probability measure of all the subsets $X^{-1}(\{x_k\})$
- In particular, if Ω is finite, then

$$EX = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

Example 1.5.1. Let $\Omega = [0,1]$, and let P be the Lebesgue measure on $[0,1]$. Consider $X(\omega) = 1$, if ω is irrational; 0 otherwise. Then $E[X] = 1P(\omega \in [0,1] : \omega \text{ is irrational}) + 0P(\omega \in [0,1] : \omega \text{ is rational}) = 1$ since $P(\omega \in [0,1] : \omega \text{ is irrational}) = 1$ = 1, $P(\omega \in [0,1] : \omega \text{ is rational}) = 0$

1.5.3 Properties of expectation

Lemma 1.5.1 (linearity of expectation). *Let X, Y be two random variables over the same probability space. Then*

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$$
$$E[cY] = cE[X]$$

Lemma 1.5.2 (law of total expectation). *Let X be a random variable, Let $A_1, ..., A_n \in \mathcal{F}$ be the partition of the sample space, then*

$$E[X] = \sum_{i=1}^{n} E[X|A_i] P(A_i)$$

In concise form, we have

$$E[E[X|Y]] = E[X]$$

where Y is the random variable defined on measure space $(\Omega, \sigma(A_1, ..., A_n))$.

Definition 1.5.3 (expectation of function of random variable). *[10] Let $h : \mathbb{R} \to \mathbb{R}, f \in \mathcal{C}$, and let $X : \Omega \to \mathbb{R}$ be a random variable with probability density $f(x)$. Then the expectation of $h(X) : \Omega \to \mathbb{R}$ is given as:*

$$E[h(X)] = \int_{-\infty}^{+\infty} h(x) f(x) dx$$

1.5.4 Dominated convergence and monotone convergence

[dominated convergence]

Theorem 1.5.1 (Dominated convergence). *[5, p. 27] Let $X_1, X_2, ...$ be a sequence of random variables converging almost surely to a random variable X. If there is another random variable Y such that $E[Y] < \infty$ and $|X_n| \leq Y$ almost surely for every n, then*

$$\lim_{n \to \infty} E[X_n] = E[X]$$

Theorem 1.5.2 (Monotone convergence). *[5, p. 26] Let $X_1, X_2, ...$ be a sequence of random variable converging almost surely to another random variable X. If*

$$0 \leq X_1 \leq X_2 ... \textit{almost surely}$$

then

$$\lim_{n \to \infty} E[X_n] = EX$$

Corollary 1.5.2.1. *Suppose the non-negative random variable X take countable many values $x_1, x_2, ...,$ then*

$$E[X] = \sum_{k=1}^{\infty} x_k P(X = x_k)$$

1.6 Variance and covariance

1.6.1 Basic properties

Definition 1.6.1 (variance, covariance). *The variance of random variable X is defined as*

$$Var[X] = E[(X - EX)^2]$$

The covariance of random variable X and Y is defined as

$$cov(X, Y) = E[(X - EX)(Y - EY)]$$

The covariance matrix $Cov(Z)$ of a random vector $Z = [Z_1, ..., Z_m]^T$ is defined as

$$Cov(Z)_{ij} = cov(Z_i, Z_j)$$

Lemma 1.6.1 (basic properties). *Let X and Y be random variables, let $a, b \in \mathbb{R}$*

- $Var[X] = E[X^2] - E[X]^2$
- $cov(X, Y) = E[XY] - E[X]E[Y]$
- $cov(\sum_i^m a_i X_i, \sum_j^n b_j Y_j) = \sum_i^m \sum_j^n a_i b_j cov(X_i, Y_j)$
- $Var[X + a] = Var[X]$
- $Var[aX] = a^2 Var[X]$
- $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abcov(X, Y)$
- $Var[aX - bY] = a^2 Var[X] + b^2 Var[Y] - 2abcov(X, Y)$
- *More generally,*

$$Var[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i^2 Var[X_i] + 2 \sum_{i=1}^n \sum_{j>1}^n a_i a_j cov(X_i, X_j) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j cov(X_i, X_j)$$

Proof. Straight forward from definitions. For (3), use linearity of expectation. □

Lemma 1.6.2 (basic properties). *Let X be a random vector, let A, B be non-random matrices, we have*

- $Cov(AX) = ACov(X)A^T$
- $Cov(X + B) = Cov(X)$

Proof. Straight forward from definitions. □

Lemma 1.6.3 (variance of a function of a random varible). *Let X be a random variable taking value in $\mathcal{X}$ with pdf $f(x)$, let g be a continuous function, then*

$$Var[g(X)] = E[(g(X) - E[g(X)])^2]] = \int_{\mathcal{X}} (g(x) - E[g(x)])^2 f(x) dx$$

Proof. Note that $E[g(X)]$ is a constant. We can calculate $Var[g(X)]$ using the expectation of a function of a random variable definition(Definition 1.5.3). $\square$

1.6.2 Conditional variance

Theorem 1.6.1 (conditional variance identity). *[9, p. 193] For any two random variables X and Y,*

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]],$$

provided that the expectation exists.

Example 1.6.1. Suppose the random variable $Y \sim Binomial(n, X)$, where $X \sim Uniform(0, 1)$ and n is a given constant. Then we can calculate

$$E[Y] = E[E[Y|X]] = E[nX]$$

and

$$Var[Y] = Var[E[Y|X]] + E[Var[Y|X]] = Var[nX] + E[nX(1 - X)].$$

1.6.3 Partial correlation

Definition 1.6.2 (partial correlation between random variables). *[11, p. 96] Let $X = (X_1, X_2, ..., X_d)$ be a zero mean random vector. The **partial covariance** between X_i and X_j is the covariance conditioned on the other elements in the vector,which is defined as*

$$PCov(X_i, X_j, X_{-ij}) = Cov(X_i - P_{X_{-ij}}[X_i], X_j - P_{X_{-ij}}[X_j]).$$

*The **partial correlation** between X_i and X_j is defined as*

$$PCorr(X_i, X_j, X_{-ij}) = \frac{Corr(X_i - P_{X_{-ij}}[X_i], X_j - P_{X_{-ij}}[X_j]}{\sqrt{Var[X_i - P_{X_{-ij}}[X_i]}\sqrt{Var[X_j - P_{X_{-ij}}[X_i]}}.$$

Theorem 1.6.2 (calculating partial correlation via best linear predictor).

- *Let $X = (X_1, X_2, ..., X_d)$ be a zero mean random vector. The **partial covariance** between X_i and X_j given by*

$$PCov(X_i, X_j, X_{-ij}) = Cov(X_i - Var[X_{-ij}]^{-1}E[X_{-ij}X_i]X_{-ij}, X_j - Var[X_{-ij}]^{-1}E[X_{-ij}X_j]X_{-ij}).$$

- *The **partial correlation** between X_i and X_j is given by*

$$\begin{aligned} &PCorr(X_i, X_j, X_{-ij}) \\ &= \frac{Corr(X_i - Var[X_{-ij}]^{-1}E[X_{-ij}X_i]X_{-ij}, X_j - Var[X_{-ij}]^{-1}E[X_{-ij}X_j]X_{-ij})}{\sqrt{Var[X_i - Var[X_{-ij}]^{-1}E[X_{-ij}X_i]}\sqrt{Var[X_j - Var[X_{-ij}]^{-1}E[X_{-ij}X_j]}}. \end{aligned}$$

Proof. Note that linear predictor on Y based on random vector X is given by (Theorem 1.9.2),

$$L(Y|X) = E[Y] + (X - E[X])^T\beta,$$

where $X = (X_1, X_2, ..., X_n)^T, \beta = \Sigma_{XX}^{-1}\Sigma_{XY}$. In particular, □

Lemma 1.6.4 (detecting partial correlation for multivariate Gaussian random vector). *Let $X = (X_1, X_2, ..., X_d)$ be a zero mean random vector. If $PCorr(X_i, X_j, X_{-ij}) = 1$, then X_i and X_j are conditionally independent given X_{-ij}.*

Proof. □

1.7 Characteristic function and Moment generating functions

1.7.1 Moment generating function

Definition 1.7.1 (moment generating function). *[9, p. 62] The moment generating function of a random variable X is given as*

$$M_X(t) = E[e^{tX}],$$

provided that the expectation exists for t in some neighborhood of 0.

Remark 1.7.1 (existence of moment generating function). If the expectation does not exist for some t in the neighborhood of 0, then moment generating function does not exist.

Lemma 1.7.1 (generating moments). *[9, p. 62] Let X be a random variable with moment generating function $M_X(t)$. Under the assumption of exchange expectation and differential is legitimate, for $n > 1$, then*

- $$E[X^n] = M_X^{(n)}(0) = \frac{d^n M_X(t)}{dt^n}|_{t=0}.$$
- $$M_X(t) = 1 + \sum_{n=1}^{\infty} \frac{M^{(n)}(0)}{n!} t^n = 1 + \sum_{n=1}^{\infty} \frac{E[X^n]}{n!} t^n.$$

Proof. (1)

$$\begin{aligned} M_X(t) &= \int e^t x f(x) dx \\ M_X^{(n)}(t) &= \int x^n e^t x f(x) dx \\ M_X^{(0)}(t) &= \int x^n f(x) dx \end{aligned}$$

(2) Use Taylor expansion. □

Theorem 1.7.1 (fundamental relationship between distribution and moment generating functions). *[9, p. 65] Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist. We have*

- *If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if*

$$E[X^r] = E[Y^r]$$

for all integers $r = 0, 1, 2, ...$

- ***(uniqueness)****If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.*

Remark 1.7.2 (non-uniqueness of moments)**.**

- Two distinct random variables might have the same moments.[9, p. 64].
- The problem of uniqueness of moments does not occur if the random variables have bounded support.

Remark 1.7.3. The assumption that expectation and differentiation operands can be exchanged holds whenever the moment generating function exists in a neighborhood of zero, which will be the case for common distributions.[12]

Lemma 1.7.2 (basic properties). *[9, p. 67] Let X and Y be two independent random variables, then*

- $M_{X+Y}(t) = M_X(t)M_Y(t)$
- *If $Z = aX + b$, then $M_Z(t) = e^{bt}M_X(at)$*

Proof. (1) Let $Z = X + Y$, $f_Z(z) = \int f_X(z-y)f_Y(y)dy$, then

$$M_Z = \int e^{zt}f_Z(z)dz = \int e^{zt}\int f_X(z-y)f_Y(y)dy = \int e^{(z-y)t}f_X(z-y)dz\int e^{yt}f_Y(y)dy$$

let $w = z - y$, then $dw = dz - dy, dzdy = dydw((dy)^2 = 0)$, we have

$$\int e^{(z-y)t}f_X(z-y)dz\int e^{y}f_Y(y)dy = \int e^{wt}f_X(w)dw\int e^{yt}f_Y(y)dy = M_X(t)M_Y(t)$$

(2) From Corollary 1.4.1.1, $M_Z(t) = \int e^{zt}f_Z(z)dz = \int e^{axt+bt}f_X(x)dx = e^{bt}M_X(at)$ $\square$

1.7.2 Characteristic function

Definition 1.7.2 (characteristic function). *Given a random variable X with probability measure P, its characteristic function is given as*

$$\psi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx}dP(x) = \int_{-\infty}^{\infty} e^{itx}f(x)dx.$$

Remark 1.7.4 (interpretation and existence)**.**

- We can interpret the characteristic function as the Fourier transform of the density function $f(x)$.
- Because $\left|e^{itx}f(x)\right| \leq |f(x)|$ is L^1 integrable, then characteristic function always exists.

Lemma 1.7.3 (characteristic function as bijections). *Every distribution has a unique characteristic function; and to each characteristic function there corresponds a unique distribution of probability.*

Remark 1.7.5 (Moment generating functions vs characteristic functions)**.**

- Characteristic function always exists, whereas moment generating function not necessarily exists.
- Characteristic function is useful when we want to develop theory for more general pdf.

Lemma 1.7.4 (recovering probability distribution from characteristic function). *Let $\psi_X(t)$ be the characteristic function of random variable X. Then we can obtain its probability density function via*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_X(t)\exp(-itx)dt$$

Proof. Use the property of Fourier transform(**??**):

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_X(t)\exp(-itx)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{itx}dP(x) = \int_{-\infty}^{\infty} e^{itx'}f(x')\exp(-itx)dtdx' = \int_{-\infty}^{\infty} f(x')\delta(x-$$

□

1.7.3 Joint moment generating functions for random vectors

Definition 1.7.3 (joint moment generating function). *The joint moment generating function for a random vector $X = (X_1, ..., X_n)^T$ is defined as*

$$m_X(t) = E[\exp(t^T X)]$$

where $t \in \mathbb{R}^n, m_X(t) \in \mathbb{R}$, if the expectation exists in the neighborhood of the origin.

Lemma 1.7.5 (constructing joint moment generating function). *Let X be a K-dimensional random vector with a joint mgf $M_X(t)$, then we have*

- *If $X_1, X_2, ..., X_K$ are mutually independent of each other, then $M_X(t) = M_{X_1}(t_1)...M_{X_K}(t_K)$*
- *Let A be a matrix and b a vector, then $Z = AX + b$ has joint mgf given as*

$$M_Z(t) = e^{t^T b}M_X(A^T t)$$

Proof. Directly from definitions. □

Lemma 1.7.6 (cross moment generation). *Let X be a K dimensional random vector possessing a joint mgf $M_X(t)$, then*

$$\mu_X(n_1, n_2, ..., n_K) = E[X_1^{n_1} X_2^{n_2} ... X_K^{n_K}]$$

is given by

$$\mu_X(n_1, n_2, ..., n_K) = \frac{\partial^{n_1+...+n_K} M_X(t_1, ..., t_K)}{\partial t_1^{n_1} ... \partial t_K^{n_K}}|_{t=0}$$

Remark 1.7.6 (some applications). With joint mgf, we can evaluate the mean and covariance easily. For example, $E[X_1]$ can be obtained by setting $n_1 = 1, n_2 = 0, n_K = 0$. $E[X_i X_j]$ can be obtained by setting $n_i = n_j = 1$.

1.7.4 Probability generating function

Definition 1.7.4 (sequence generating function). *[13, p. 148] Given a real-valued sequence $a = \{a_1, a_2, ...\}$. The generating function G of the sequence is*

$$G(s) = \sum_{i=0}^{\infty} a_i s^i$$

for $s \in \mathbb{R}$ such that the sum converges.

Definition 1.7.5 (convolution of real sequence). *The convolution of the real sequences $a = \{a_i, i \geq 0\}$ and $b = \{b_i, i \geq 0\}$ is the sequence $c = \{c_i, i \geq 0\}$ defined by*

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

Lemma 1.7.7 (convolution theorem of real sequence). *[13, p. 150] If sequences $\{a_n\}$ and $\{b_n\}$ have generating function G_a and G_b respectively, then*

$$c = a * b, G_c(s) = G_a(s) G_b(s).$$

Proof.

$$G_c(s) = \sum_{n=0}^{\infty} c_n s^n = \sum_{n=0}^{\infty} (\sum_{i=0}^{n} a_i b_{n-i}) s^n = \sum_{i=0}^{\infty} a_i s^i \sum_{n=i}^{\infty} b_{n-i} s^{n-i} = G_a(s) G_b(s).c$$

□

Lemma 1.7.8 (term-by-term operation property). *Consider a sequence generating function $G(s)$. If s^* is the convergence radius, then*

- $\sum_{i=0}^{\infty} a_i s^i$ *is uniformly convergent within* $|s| < s^*$.
- $\sum_{i=0}^{\infty} a_i s^i$ *can be differentiated and integrated term-by-term within* $|s| < s^*$.

Proof. Note that the generating function is the power sequence. See **??**. □

Definition 1.7.6 (probability generating function). *[13, p. 150] The probability generating function of the random variable X is defined to be the generating function $G(s) = E[s^X]$ of its probability mass function.*

Example 1.7.1.

- Bernoulli variable X.

$$G(s) = E[s^X] = (1-p) + ps.$$

- Binomial distribution X with parameter n and p.

$$G(s) = E[s^X] = ((1-p) + ps)^n.$$

- Poisson distribution $Poisson(\lambda)$ random variable X:

$$G(s) = E[s^X] = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{s\lambda} e^{-\lambda} = e^{\lambda(s-1)}.$$

Lemma 1.7.9. *[13, p. 151] If X has generating function $G(s)$ then*

- $E[X] = G'(1)$.
- $E[X(X-1)\cdots(X-k+1)] = G^{(k)}(1)$.

Proof. (1) Note that $G(s) = E[s^X]$, $G'(s) = E[Xs^{X-1}], G'(1) = E[X]$.(use term-by-term differentiation property Lemma 1.7.8.) (2) Same as (1). □

Lemma 1.7.10. *[13, p. 153] If X and Y are independent, then*

$$G_{X+Y}(s) = G_X(s) G_Y(s)$$

Proof. Use Lemma 1.7.7. □

1.7.5 Cumulants

Definition 1.7.7 (cumulant-generating function, cumulant).

- *The* ***cumulant-generating function*** $K(t)$ *of a random variable* X *is defined by*

$$K(t) = \ln E[\exp(tX)] = \ln M_X(t),$$

where $M_X(t)$ *is the moment generating function of* X*.*
- *The* ***cumulants*** κ_n *are obtained via*

$$\kappa_n = K^{(n)}(0),$$

such that

$$K(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}.$$

Lemma 1.7.11 (connections between cumulants and moments). *Let* $\mu_i, i = 1, 2, ...$ *denote the central moments,i.e.* $\mu_i = E[(X - E[X])^i]$ *of a distribution of a random variable* X*.Let* $m_i, i = 1, 2, ...$ *denote the cumulants of the same distribution. Let* $\kappa_i, i = 1, 2, ...$ *denote the cumulants of the same distribution. Assume the existence of moment generating function. Then*

- $$\ln(1 + \sum_{n=1}^{\infty} \frac{m_n}{n!} t^n) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} t^n$$
- *explicitly, we have*

$$\begin{aligned}
\kappa_1 &= m_1 \\
\kappa_2 &= m_2 - m_1^2 = \mu_2 \\
\kappa_3 &= \mu_3 \\
\kappa_4 &= \mu_4 - 3\mu_2^2 \\
\kappa_5 &= \mu_5 - 10\mu_3\mu_2.
\end{aligned}$$

Proof. (1) Based on the definition,

$$\sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!} = K(t) = \ln E[\exp(tX)] = \ln M_X(t) = \ln(1 + \sum_{n=1}^{\infty} \frac{m_n}{n!} t^n),$$

where we use the properties of moment generating functions(Lemma 1.7.1). (2) Use Taylor expansion for $\ln(1+x)$(**??**) given by

$$\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ...$$

and then match the coefficients for t^n. □

Example 1.7.2. Consider a Gaussian distribution given by

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

Then

- the cumulant generating function is given by

$$K(t) = \ln(e^{\mu}e^{\sigma^2\frac{t^2}{2}}) = \mu t + \frac{\sigma^2 t^2}{2}.$$

- the cumulants are given by

$$\kappa_1 = \mu, \kappa_2 = \sigma^2, \kappa_n = 0, \forall n > 2.$$

1.8 Conditional expectation

1.8.1 General intuitions & comments

Consider a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$ ($\mathcal{G}$ is a σ-algebra and $\mathcal{G} \subset \mathcal{F}$). We have the following situations:[5]

1. If X is independent of $\mathcal{G}$, then the information in $\mathcal{G}$ provides no help in determining the value X. In this case, $E[X|\mathcal{G}] = E[X]$.
2. If X is $\mathcal{G}$ measurable, then the information in $\mathcal{G}$ can fully determine X. In this case, $E[X|\mathcal{G}] = X$.
3. In the intermediate case, we can use information in $\mathcal{G}$ to estimate but not precisely evaluate X. The *conditional expectation* of X given $\mathcal{G}$ is such an estimate.
4. If $\mathcal{G}$ is the trivial σ algebra $\{\emptyset, \Omega\}$, then $\mathcal{G}$ barely contains any information: $E[X|\mathcal{G}] = E[X]$.

Another understanding in terms of random variables are: $E[X|Y]$ is the function of Y that bests approximates X. We consider a extreme case. Suppose that X is itself a function of Y, then the function of Y that best approximates X is X itself, i.e., $E[g(Y)|Y] = X = g(Y)$; If X is independent of Y, then the best estimate we can give is $E[X|Y] = E[X]$.

As a summary, we have

Lemma 1.8.1. *[14] [conditional expectation as least-squared-best predictor] If $E[X^2] < \infty$, then the conditional expectation $Y = E[X|\mathcal{G}]$ is a version of the orthogonal projection of X onto the space $\mathcal{L}^2(\Omega, \mathcal{G}, P)$. Hence, Y is the lease-squared-best $\mathcal{G}$-measurable predictor of X: among all $\mathcal{G}$-measurable functions, Y minimizes*

$$E[(Y - X)^2]$$

Remark 1.8.1. Note that the discussion on the existence and uniqueness of such Y can be found at [14][15, p. 28].

1.8.2 Formal definitions

Definition 1.8.1 (sub σ algebra). *Let X be a set and let $\mathcal{F}, \mathcal{G}$ be two σ algebras on X. then $\mathcal{G}$ is said to be sub-σ algebra of $\mathcal{F}$ if $\mathcal{G} \subseteq \mathcal{F}$.*

Definition 1.8.2 (conditional expectation as a random variable). *[5, p. 68] Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a **sub-σ algebra** of $\mathcal{F}$, and let X be a random* *variable that is either non-negative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $E[X|\mathcal{G}]$ is a **random variable** that satisfies:*

1. *(**measurability**) $E[X|\mathcal{G}]$ is $\mathcal{G}$ measurable*
2. *(partial averaging): For any element A in $\mathcal{G}$,*

$$\int_A E[X|\mathcal{G}](\omega)dP(\omega) = \int_A X(\omega)dP(\omega).$$

In particular,

- *if $\mathcal{G} = \mathcal{F}$, then $E[X|\mathcal{G}] = X$.*
- *If $G = \{\emptyset, \Omega\}$, then $E[X|\mathcal{G}] = E[X]$.*

[a]

[a] The meaning of X is $\mathcal{G}$ measurable can be understood as $\sigma(X) \subseteq \mathcal{G}$.

Remark 1.8.2.

- the filtration $\mathcal{G}$ in $E[X|\mathcal{G}]$ has to be $\mathcal{G} \subseteq \mathcal{F}$, otherwise P is defined for some elements in cG.
- the Partial averaging property reflects the **consistence** requirement between the new random variable $E[X|\mathcal{G}]$ and the old random variable X.
- If $\mathcal{G}$ is the σ algebra generated by some other random variable W, then we generally write $E[X|W]$ instead of $E[X|\sigma(W)]$.
- if $G = \{\emptyset, \Omega\}$, then the only $\mathcal{G}$- measurable function is a constant function. Among all the constant functions, the function that satisfies the partial averaging property is the expectation.

Note 1.8.1 (interpreting partial averaging property in partition set). Consider the case where $\mathcal{G}$ is countable. Let $\mathcal{P}$ be the smallest partition set of $\mathcal{G}$. Then the random variable $E[X|\mathcal{G}]$ can only take countable many values. In particular, the partial averaging property implies

$$E[X|\mathcal{G}](A_i) = \int_A X(\omega)dP(\omega)\forall A_i \in \mathcal{P}.$$

That is, $E[X|\mathcal{G}]$ can be viewed as a mapping from Ω to $\mathbb{R}$ that has been **coarsened via local averaging**.

Note 1.8.2 (Generalization on expectation). When we talk about expectation, there are two items we should consider: which measure the expectation is taken with respect to and which filtration the expectation is taken with respect to.

- We can view expectation as a special case of conditional expectation: for example
$$E[X] = E[X|\mathcal{G}], \mathcal{G} = \{\emptyset, \Omega\}.$$
- Conditional expectations with respect to different measure can equal if the two measures agree on the filtration. For example,
$$E_P[X|\mathcal{G}] = E_Q[X|\mathcal{G}],$$
if $P(A) = Q(A), \forall A \in \mathcal{G}$.

1.8.3 Different versions of conditional expectation

Remark 1.8.3. For different versions of conditional expectation, see [15, p. 17] for details.

1.8.3.1 *Conditioning on an event*

Definition 1.8.3. *For any integrable random variable η and any event $B \in \mathcal{F}$ such that $P(B) \neq 0$, the conditional expectation given B is defined as*
$$E[\eta|B] = \frac{\int_B \eta dP}{\int_B dP} = \frac{1}{P(B)} \int_B \eta dP$$

1.8.3.2 *Conditioning on a discrete random variable as a new random variable*

Definition 1.8.4. *Let X be an integrable random variable, let Y be a discrete random variable. Then the conditioning expectation of X given Y is defined to be a random variable $E[X|Y]$ such that*
$$E[X|Y](\omega) = E[X|\{Y(\omega) = y_i\}]$$

Lemma 1.8.2. *If X is an integrable random variable, and Y is a discrete random variable, then*

- *$E[X|Y]$ is $\sigma(Y)$-measurable*
- *For any $A \in \sigma(Y)$:*
$$\int_A E[X|Y]dP = \int_A XdP$$

Proof. When Y is a discrete random variable, $E[X|Y]$ can only take discrete values. For any Borel set on $\mathbb{R}$, we find the inverse image $B \in \sigma(Y)$. Therefore it is measurable. (2) directly form partial averging property of conditional expectation. □

1.8.3.3 *Condition on random variable vs. event vs σ algebra*

- Conditional expectations for discrete random variables, such as $E[X|Y = 2], E[X|Y = 5]$ are numbers. These are examples of condition on events. $E[X|Y]$ can be interpreted as $E[X|Y = y]$, a function depends on y.
- When we write $E[X|Y]$, we should interpret as conditioning on the σ algebra generated by Y.

1.8.4 Properties

For a comprehensive treament, see [15, p. 70]. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$,let X, Y be integrable random variables. We have:

1.8.4.1 *Linearity*

Lemma 1.8.3 (linearity of conditional expectation). *[5, p. 69] Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$,let X, Y be integrable random variables. We have:*

$$E[c_1X + c_2Y|\mathcal{G}] = c_1E[X|\mathcal{G}] + c_2E[Y|\mathcal{G}].$$

Proof. (1) First, $c_1E[X|\mathcal{G}]$ is $\mathcal{G}$ measurable, $c_2E[Y|\mathcal{G}]$ is $\mathcal{G}$ measurable, therefore, $E[c_1X + c_2Y|\mathcal{G}] = c_1E[X|\mathcal{G}] + c_2E[Y|\mathcal{G}]$ is $\mathcal{G}$ measurable(Lemma 1.3.2). (2) For every $A \in \mathcal{G}$,

$$\begin{aligned}
&\int_A (c_1E[X|\mathcal{G}](\omega) + c_2E[Y|\mathcal{G}](\omega))dP(\omega) \\
=&\int_A (c_1E[X|\mathcal{G}](\omega) + c_2E[Y|\mathcal{G}](\omega))dP(\omega) \\
=&c_1\int_A (E[X|\mathcal{G}](\omega)dP(\omega) + c_2\int_A E[Y|\mathcal{G}](\omega))dP(\omega) \\
=&c_1\int_A X(\omega)dP(\omega) + c_2\int_A Y(\omega)dP(\omega) \\
=&\int_A c_1X(\omega) + c_2Y(\omega)dP(\omega)
\end{aligned}$$

that is $E[c_1X + c_2Y|\mathcal{G}]$ satisfies the partial averaging property. □

1.8.4.2 *Taking out what is known*

Lemma 1.8.4. *[5, p. 70]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$,let X, Y be integrable random variables. If XY is integrable, X is $\mathcal{G}$-measurable, then*

- $$E[XY|\mathcal{G}] = XE[Y|\mathcal{G}], E[g(X)Y|\mathcal{G}] = g(X)E[Y|\mathcal{G}].$$
- $$E[X|\mathcal{G}] = X, E[X|X] = X, E[g(X)|X] = g(X)$$

Proof. Note that from Lemma 1.3.2, $g(X), XY, g(X)Y$ are all $\mathcal{F}$ measurable random variables. □

1.8.4.3 *Law of iterated expectations*

Lemma 1.8.5 (iterated conditioning). *If $\mathcal{H}, \mathcal{G}$ are both σ algebra on Ω, and $\mathcal{G} \subset \mathcal{H}$ (in some sense $\mathcal{G}$ has less information), then for random variable X, we have*

$$E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$$

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{G}].$$

In particular,

$$E[E[X|\mathcal{G}]] = E[X],$$

or equivalently, in terms of conditioning on random variables, we have

$$E[E[X|Y]] = E[X].$$

Proof. (1)(a) First $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable. (b) For any $A \in \mathcal{G} \subseteq H$, we have

$$\begin{aligned} & \int_A E[E[X|\mathcal{H}]|\mathcal{G}](\omega)dP(\omega) \\ = & \int_A E[X|\mathcal{H}](\omega)dP(\omega) \\ = & \int_A X(\omega)dP(\omega) \\ = & \int_A E[X|\mathcal{G}](\omega)dP(\omega) \end{aligned}$$

(2) Note that the random variable $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable therefore $\mathcal{H}$-measurable.

□

Example 1.8.1. Let B_t be a Brownian motion, let $\mathcal{F}_t = \sigma(B_s, s \leq t)$ be the filtration, then

$$E[B_s|\mathcal{F}_t, t > s] = B_s, E[B_s|\mathcal{F}_t, t < s] = E[B_t + (B_s - B_t)|\mathcal{F}_t] = B_t$$

1.8.4.4 *Conditioning on independent random variable/σ algebra*

Lemma 1.8.6. *[5, p. 70]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$,let X, Y be integrable random variables. Let f be Borel measurable function and $f(X)$ be integrable.*

- *If $\sigma(X)$ and $\mathcal{G}$ are independent, then*

$$E[X|\mathcal{G}] = E[X], E[g(X)|\mathcal{G}] = E[g(X)].$$

- *If X and Y are independent, then*

$$E[X|Y] = E[X|\sigma(Y)] = E[X].$$

Proof. (1)(a) $E[X]$ is a constant, therefore is $\mathcal{G}$ measurable. (b)(informal) Consider the special case where $X = \mathbf{1}_B$, where $B \in \mathcal{F}$ but B is independent of $\mathcal{G}$. Then

$$\int_A X(\omega)dP(\omega) = P(A \cap B) = P(A)P(B) = E[X]P(A) = E[X]\int_A dP(\omega) = \int_A E[X]dP(\omega).$$

Since X can be represented by the sum of indicator function, such relation can hold when X is an arbitrary random variable.(See reference for more details). □

1.8.4.5 *Least Square minimizing property*

Lemma 1.8.7 (least square minimizing property of conditional expectation). *Let $Y \in \mathcal{L}_2(\Omega, \mathcal{G}, P)$ and $\mathcal{F}$ be a sub-σ of $\mathcal{G}$, then*

$$E[(Y - E[Y|\mathcal{F}])^2] = \min\{E[(Y - Z)^2], \forall Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)\}$$

Proof. For any $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$, we have

$$\begin{aligned}
E[(Y - E[Y|\mathcal{F}])^2] &= E[(Y - Z + Z - E[Y|\mathcal{F}])^2] \\
&= E[(Y - Z)^2] + E[(Z - E[Y|\mathcal{F}])^2] + 2E[(Y - Z)(Z - E[Y|\mathcal{F}])] \\
&= E[(Y - Z)^2] + E[(Z - E[Y|\mathcal{F}])^2] + 2E[E[(Y - Z)(Z - E[Y|\mathcal{F}])|\mathcal{F}]] \\
&= E[(Y - Z)^2] + E[(Z - E[Y|\mathcal{F}])^2] + 2E[E[(Y - Z)|\mathcal{F}](Z - E[Y|\mathcal{F}])] \\
&= E[(Y - Z)^2] + E[(Z - E[Y|\mathcal{F}])^2] - 2E[(Z - E[Y|\mathcal{F}])(Z - E[Y|\mathcal{F}])] \\
&= E[(Y - Z)^2] - E[(Z - E[Y|\mathcal{F}])^2] \leq E[(Y - Z)^2]
\end{aligned}$$

Note that we use $E[(Y - Z)(Z - E[Y|\mathcal{F}])|\mathcal{F}] = (Z - E[Y|\mathcal{F}])E[(Y - Z)|\mathcal{F}]$ since $(Z - E[Y|\mathcal{F}])$ is $\mathcal{F}$ measurable. $\square$

1.9 The Hilbert space of random variables

1.9.1 Definitions

Definition 1.9.1. *The vector space $L^2(\Omega, \mathcal{F}, P)$ of real-valued random variables on $(\Omega, \mathcal{F}, P)$ can be defined as the Hilbert space of random variables with finite second moment(see* **??***). The inner product is then defined as*

$$\langle x, y \rangle = E[xy].$$

The norm of a random variable is

$$\|X\| = \sqrt{E[X^2]}$$

Lemma 1.9.1 (correlation and orthogonality for zero mean random variables). *Let X and Y be two zero mean random variables in the Hilbert space $L^2(\Omega, \mathcal{F}, P)$. Then X and Y are uncorrelated if and only if they are orthogonal, i.e., $\langle X, Y \rangle = 0$.*

Proof. (1) If $\langle X, Y \rangle = 0$, then

$$E[XY] = E[X]E[Y] + Cov(X, Y) = 0 \implies Cov(X, Y) = 0$$

. (2) If $Cov(X, Y) = 0$, then

$$\langle X, Y \rangle = E[XY] = E[X]E[Y] + Cov(X, Y) = 0.$$

□

1.9.2 Subspaces, projections, and approximations

Theorem 1.9.1 (projection onto closed subspace , recap). *Let U be a closed subspace of L^2 and $X \in L^2$. Then the projection of X onto U is the vector/random variable $V \in U$ such that*

- $$\langle X - V, u \rangle = E[(X - V)u] = 0, \forall u \in U$$
- *V is unique;*
- *V is minimizer,i.e., $\|X - V\|^2 \leq \|X - u\|^2, \forall u \in U$.*

[a].

[a] Note that in a Hilbert space(also a normed linear space), any finite-dimensional subspace is closed(**??**)

Proof. See the projection theorem(**??**) guarantees the existence of solution. □

Lemma 1.9.2 (projection onto the subspace of constant random variables).

- *Let real-valued random variable $X \in L^2$, we define the root mean square error function by*

$$d_2(X,t) = \|X - t\|_2 = \sqrt{E[(X-t)^2]}, t \in \mathbb{R}$$

then $d_2(X,t)$ is minimized when $t = E[X]$ and that the minimum value is $\sqrt{Var[X]}$.
- *Let real-valued random variable $X \in L^2$, we define the 1d subspace $W = \{a : a \in \mathbb{R}\}$(the subspace spanned by constant random variable 1). Then the projection of X onto W is $E[X]$.*

Proof. (1)directly minimize with respect t. (2) We can see that the orthogonality condition implies that

$$0 = \langle X - a, b \rangle = E[(X-a)b] = 0, \forall b \in \mathbb{R},$$

which gives $a = E[X]$. □

Theorem 1.9.2 (best linear predictor for random variables).

- *Given $X, Y \in L^2$, the best linear predictor for Y given X is to find a projection onto the subspace $W = \{a + bX : a \in \mathbb{R}, b \in \mathbb{R}\}$(the subspace spanned by random variable 1 and X), given as*

$$L(Y|X) = E[Y] + \frac{Cov(X,Y)}{Var(X)}(X - E[X])$$

and the variance/mean square error for the prediction is

$$Var(Y - L(Y|X)) = Var(Y) - \frac{Cov(X,Y)^2}{Var(X)}.$$

- *Given $X_1, X_2, ..., X_n, Y \in L^2$, the best linear predictor for Y given $X_1, X_2, ..., X_n$ is*

$$L(Y|X) = E[Y] + \sum_{i=1}^n (X_i - E[X_i])[\sum_{j=1}^n (\Sigma_{XX}^{-1})_{ij} Cov(X_j, Y)], ;$$

or in vector form

$$L(Y|X) = E[Y] + (X - E[X])^T\beta,$$

where $X = (X_1, X_2, ..., X_n)^T, \beta = \Sigma_{XX}^{-1}\Sigma_{XY}$. *In particular, if* $Cov(X_iX_j) = Var[X_i]\delta_{ij}$, *then*

$$L(Y|X) = E[Y] + \sum_{i=1}^{n} \frac{Cov(X_i, Y)}{Var(X_i)}(X_i - E[X_i]).$$

- *The estimation error is given by*

$$E[(Y - L(Y|X))^2] = Var[Y] - \Sigma_{XY}^T(\Sigma_{XX}^{-1})\Sigma_{XY}.$$

- *The single coefficient associated with* X_i *is given by*

$$\beta_i = \frac{Cov(Y - L(Y|X_{-i}), X_i - L(X_i|X_{-i})}{Var[X_i - L(X_i|X_{-i})]},$$

where X_{-i} *denotes the subspace associated with span* $\{1, X_1, ..., X_{i-1}, X_{i+1}, ..., X_n\}$.
- *un-correlation of the residual and X:*

$$Cov(Y - L(Y|X), X) = 0.$$

Proof. (1) To verify that $L(Y|X)$ is the projection, we only need to verify the orthogonality conditions(**??**):

$$\langle Y - L(Y|X), X \rangle = 0, \langle Y - L(Y|X), 1 \rangle = 0.$$

We have

$$\begin{aligned}
\langle Y - L(Y|X), X \rangle &= E[(Y - L(Y|X))X] \\
&= E[(Y - E[Y] - \frac{Cov(X,Y)}{Var(X)}(X - E[X]))X] \\
&= E[(Y - E[Y])X] - E[\frac{Cov(X,Y)}{Var(X)}(X - E[X]))X] \\
&= Cov(X,Y) - Cov(X,Y) \\
&= 0
\end{aligned}$$

where we used the fact that $E[X(X - E[X])] = Var[X]$. For another,

$$\begin{aligned}
\langle Y - L(Y|X), 1 \rangle &= E[(Y - L(Y|X))] \\
&= E[(Y - E[Y] - \frac{Cov(X,Y)}{Var(X)}(X - E[X]))] \\
&= 0 - 0 \\
&= 0.
\end{aligned}$$

The variance is given by

$$\begin{aligned}
Var[Y - L(Y|X)] &= E[(Y - L(Y|X))^2] \\
&= E[(Y - E[Y] - \frac{Cov(X,Y)}{Var(X)}(X - E[X]))(Y - E[Y] - \frac{Cov(X,Y)}{Var(X)}(X - E[X]))] \\
&= E[(Y - E[Y])^2] + \frac{Cov(X,Y)^2}{Var(X)} - 2E[(Y - EY)(\frac{Cov(X,Y)}{Var(X)}(X - E[X]))] \\
&= E[(Y - E[Y])^2] + \frac{Cov(X,Y)^2}{Var(X)} - 2\frac{Cov(X,Y)^2}{Var(X)} \\
&= E[(Y - E[Y])^2] - \frac{Cov(X,Y)^2}{Var(X)}
\end{aligned}$$

(2)

We can obtain the vector form via the optimization

$$\min f = E[(Y - \beta_0 - \beta^T X)^2]$$

over β_0, β_1, we have

$$f(\beta_0, \beta_1) = E[(Y^2 + \beta_0^2 + (\beta^T X)^2 + 2\beta_0 \beta^T X - 2\beta_0 Y - 2Y\beta^T X)]$$

The first order condition on β_0 gives that

$$\beta_0 = E[Y] - \beta^T E[X];$$

Plug in β_0 and the first order condition on β_1 gives that

$$\begin{aligned}
f(\beta_0, \beta_1) &= E[((Y - EY)^2 - 2\beta^T (X - EX)(Y - EY) + \beta^T E[(X - EX)(X - EX)^T]\beta \\
\implies \partial f / \partial \beta &= -2E[(X - EX)(Y - EY)] + 2E[(X - EX)(X - EX)^T]\beta = 0 \\
\implies \beta &= (E[(X - EX)(X - EX)^T])^{-1} E[(X - EX)(Y - EY)] = (\Sigma_{XX}^{-1})\Sigma_{XY}
\end{aligned}$$

Note that the problem has semi-positive definite Hessian we are sure that the minimizer exists.

From the Hilbert space projection point of view, we can also verify the orthogonality conditions(**??**):

$$\left\langle Y - L(Y|X), X_k \right\rangle = 0, k = 1, 2, ..., n.$$

We have

$$\begin{aligned}\langle Y - L(Y|X), X_k \rangle &= E[(Y - E[Y] + \sum_{i=1}^{n}(X_i - E[X_i])[\sum_{j=1}^{n}(\Sigma_{XX}^{-1})_{ij}Cov(X_j, Y)]))X_k] \\ &= E[(Y - E[Y])X_k] - E[(\sum_{i=1}^{n}(X_i - E[X_i])[\sum_{j=1}^{n}(\Sigma_{XX}^{-1})_{ij}Cov(X_j, Y)])X_k] \\ &= Cov(X_k, Y) - Cov(X_j, Y)\delta_{jk} \\ &= 0\end{aligned}$$

where we used the fact that

$$\sum_{i=1}^{n}(X_i - E[X_i])\sum_{j=1}^{n}(\Sigma_{XX}^{-1})_{ij}X_k = \delta_{jk}.$$

Note that for an invertible matrix A, $\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}A_{jk}^{-1} = \delta_{ik}$.

(3) Note that $L(Y|X)$ is unbiased because of the orthogonality condition

$$\langle Y - L(Y|X), 1 \rangle = 0 \implies E[(Y - L(Y|X))1] = E[Y] - E[L(Y|X)] = 0.$$

In the following we use the notation

$$\hat{Y} = L(Y|X), E[Y] = \mu_Y = E[L(Y|X)] = \mu_{\hat{Y}}.$$

We have

$$\begin{aligned}E[(Y - \hat{Y})^2] &= E[((Y - \mu_Y) - (\hat{Y} - \mu_{\hat{Y}}))^2] \\ &= E[((Y - \mu_Y))^2] - 2E[(Y - \mu_Y)(\hat{Y} - \mu_{\hat{Y}})] + E[((\hat{Y} - \mu_{\hat{Y}}))^2] \\ &= Var[Y] - 2E[(Y - \mu_Y)(\hat{Y} - \mu_{\hat{Y}})] + Var[\hat{Y}] \\ &= Var[Y] - 2E[(Y - \mu_Y)(\beta^T X - \mu_{\hat{Y}})] + Var[\hat{Y}] \\ &= Var[Y] - 2Cov(Y, \beta^T X) + Var[\beta^T X] \\ &= Var[Y] - 2\beta^T Cov(Y, X) + \beta^T Cov(X, X)\beta \\ &= Var[Y] - 2\Sigma_{XY}^T(\Sigma_{XX}^{-1})Cov(Y, X) + \Sigma_{XY}^T(\Sigma_{XX}^{-1})Cov(X, X)(\Sigma_{XX}^{-1})\Sigma_{XY} \\ &= Var[Y] - \Sigma_{XY}^T(\Sigma_{XX}^{-1})\Sigma_{XY}\end{aligned}$$

(4) Direct generalization from Hilbert space approximation theory(**??**).

(5)

$$\begin{aligned}E[Y - L(Y|X), X - E[X]] &= E[Y - E[Y] - (X - E[X])^T\beta, X - E[X]] \\ &= Cov(X, Y) - Var[X]\beta \\ &= Cov(X, Y) - Cov(X, Y) \\ &= 0.\end{aligned}$$

□

Remark 1.9.1.

- The more correlated X and Y are, the more information X can provide to predict Y
- The more volatile X is, the less information X can provide.
- The magnitude of $\frac{Cov(X,Y)^2}{Var(X)}$ reflects the importance of X in prediction.

1.9.3 Connection to conditional expectation

Theorem 1.9.3 (conditional expectation with respect to a σ algebra as a projection).

- *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, then the set*
$$U = \{X \in L^2 | X \text{ is measurable with respect to } \mathcal{G}\}$$
is a subspace of L^2.
- *If $X \in L^2$, then $E[X|\mathcal{G}]$ is the projection of X onto the subspace U defined as*
$$U = \{X \in L^2 | X \text{ is measurable with respect to } \mathcal{G}\}.$$

Proof. (1) the zero element 0 is both $\mathcal{F}$ and $\mathcal{G}$ measurable. (2) If X, Y are $\mathcal{G}$ measurable, then $cX, X + Y$ are $\mathcal{G}$ measurable (**??**). (2) directly from the definition of conditional expectation(Definition 1.8.2). □

Definition 1.9.2 (conditional expectation and projection). *The conditional expectation of $X \in L^2$ given $X_1, X_2, ..., X_n \in L^2$ is defined to be the projection of X onto the closed subspace $M(X_1, X_2, ..., X_n)$ spanned by* ***all random variables of the form*** *$g(X_1, X_2, ..., X_n)$, where g is some measurable function $g : \mathbb{R}^n \to \mathbb{R}$, i.e.*
$$E[X|X_1, X_2, ..., X_n] = P_S[M(X_1, X_2, ..., X_n)].$$

Definition 1.9.3 (conditional expectation and projection onto a subspace, special case). *The conditional expectation of $X \in L^2$ given a closed subspace $S \subseteq L^2$, which contains the constant random variable 1, is defined to be the projection of X onto S,i.e.,*
$$E[X|S] = P_S[X].$$

Remark 1.9.2. Note that the subspace has to contains the constant random variable to make the definition and conditional expectation and projection match.

Remark 1.9.3.

- $$span(1, X_1, ..., X_n) \subseteq M(X_1, X_2, ..., X_n),$$

  therefore

  $$\|X - E[X|X_1, X_2, ..., X_n]\|^2 \leq \|X - E[X|span(1, X_1, X_2, ..., X_n)]\|^2.$$

- The definition of

  $$E[X|X_1, X_2, ..., X_n] = P_S[M(X_1, X_2, ..., X_n)]$$

  coincides with the usual definition of conditional expectation with respect to a σ algebra(Definition 1.8.2).
- The conditional expectation with respect to a subspace is not the general definition of conditional expectation.

Lemma 1.9.3 (conditional expectation and best predictor for multivariate normal random variables). *Let $(Y, X_1, X_2, ..., X_n), X = (X_1, X_2, ..., X_n)$ be a random vector with multivariate normal distribution with parameter*

$$\mu = [\mu_Y^T, \mu_X^T]^T, \Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$$

Then

- *$Y|X_1, X_2, ..., X_n$ has the same distribution of*

  $$\hat{Y} = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X) + \epsilon,$$

  conditioning on X and $\epsilon \sim N(0, \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY})$.
- $$E[Y|X_1, X_2, ..., X_n] = E[Y|span(1, X_1, X_2, ..., X_n)] = P_{span(1, X_1, X_2, ..., X_n)}[Y].$$

  That is, the best predicator(in terms of minimum variance) given $X_1, X_2, ..., X_n$ is the best linear predictor.

Proof. From Theorem 2.2.2, the martinal distribution is Gaussian given by

$$N(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}).$$

Therefore, Y has the conditional expectation of

$$\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X),$$

and the conditional variance of

$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}.$$

□

1.10 Probability inequalities

1.10.1 Some common inequalities

Remark 1.10.1. See [9, p. 137] for more details.

Theorem 1.10.1 (General Chebychev's inequality). *Let X be a random variable, and let $g(x) \geq 0$. Then, for any $r > 0$, we have:*

$$P(g(X) > r) \leq \frac{E[g(X)]}{r}.$$

Proof.

$$\begin{aligned} E[g(x)] &= \int_{\infty}^{\infty} g(x) f_X(x) dx \\ &\geq \int_{x:g(x)\geq r} g(x) f_X(x) dx \\ &\geq r \int_{x:g(x)\geq r} f_X(x) dx \\ &= rP(g(X) \geq r) \end{aligned}$$

□

Corollary 1.10.1.1 (Chebychev's inequality).

$$P(\frac{(X-EX)^2}{\sigma^2} \geq t^2) \leq \frac{1}{t^2} E[\frac{(X-\mu)^2}{\sigma^2}]$$

Proof. Let $g(X) = (X-\mu)^2/\sigma^2$ and use above theorem. □

Corollary 1.10.1.2. *Let $g : [0, \infty) \to [0, \infty)$ be a strictly increasing non-negative function, and set $h(x) = g(|x|)$ to obtain*

$$P(|X| \geq a) \leq \frac{E[g(|X|)]}{g(a)}$$

where $a > 0$.

Corollary 1.10.1.3 (Markov's inequality). *If $X \geq 0$, then*

$$P(X \geq r) \leq E[X]/r$$

Proof. Let $g(X) = X$ in the general Chebychev's inequality. □

Lemma 1.10.1 (Jensen's inequality). *For any random variable X, if $g(x)$ is a convex function then*

$$E[g(X)] \geq g(E[X])$$

Proof. Note that for convex function

$$g(\sum_{i=1}^{n} w_i x_i) \leq \sum_{i=1}^{n} w_i g(x_i), \forall w_i \geq 0, \sum_{i=1}^{n} w_i = 1, i = 1, ..., n.$$

□

Example 1.10.1 (Jensen's inequality application). Use Jensen's inequality, it can be showed that

$$E[X]^2 \leq E[X^2]$$

with $g(x) = x^2$.

Theorem 1.10.2 (Holder's inequality). *[13, p. 319] If $p, q > 1$ and $1/p + 1/q = 1$, then*

$$E[\![XY|] \leq (E[\![X|^p])^{1/p}(E[\![Y|^q])^{1/q}.$$

The equality holds when there exists real numbers $\alpha, \beta > 0$ such that $\alpha|X|^p = \beta|Y|^q$ almost everywhere.

Proof. Let $A = (\int |x|^p \, dP)^{1/p} = E[\![X^p|])^{1/p}$ and $B = (\int |y|^q \, dP)^{1/q} = (E[\![Y^q|])^{1/q}$. Then let $a = |X| / A$, $b = |Y| / B$, and then apply Young's inequality:

$$ab = |XY| / AB \leq \frac{|X|^p}{pA^p} + \frac{|Y|^q}{qA^q} = \frac{a^p}{p} + \frac{b^q}{q}$$

Integrate(Lebesgue) both sides use probability measure and notice that A, B are constant, $A^p = E[\![X^p|]$, then

$$\frac{E[\![XY|]}{(E[\![X^p|])^{1/p}(E[\![Y^q|])^{1/q}} \leq 1/p + 1/q = 1$$

□

Remark 1.10.2. Let $q = p = 2$, and we get the Cauchy-Schwarz inequality.

Theorem 1.10.3 (Minkowski's inequality). *[13, p. 319] If $p \geq 1$, then*

$$(E[\![X+Y|^p])^{1/p} \leq (E[\![X|^p])^{1/p} + (E[\![Y|^p])^{1/p}$$

Proof. Because L^p space are normed vector space, we can prove this using triangle inequality. □

Theorem 1.10.4 (Cauchy-Schwarz inequality). *[2][9, p. 187][16]*

- *Let X and Y be random variables with $E[X^2] < \infty, E[Y^2] < \infty$. Then*

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}.$$

The equality holds when there exists real numbers $\alpha, \beta > 0$ such that $\alpha|X|^2 = \beta|Y|^2$ almost everywhere.
Further more,

$$(Cov(X,Y))^2 \leq Var[X] \cdot Var[Y].$$

- *Let X and Y be two p dimensional random vectors with bounded variance. Then*

$$Var[Y] \geq Cov(Y,X)Var[X]^{-1}Cov(X,Y).$$

Proof. (1)(a) define inner product between two random variable as $\langle X, Y \rangle = \int xy\rho(x,y)dxdy$, since each random variable can be viewed as a functional. (b) Similarly we can use Holder's inequality.(Theorem 1.10.2)

A simple derivation: Since the covariance matrix of random vector (X,Y) much be positive semi-definite, we have

$$|Cov([XY])| = \begin{vmatrix} Var[X] & Cov(X,Y) \\ Cov(X,Y) & Var[Y] \end{vmatrix} \geq 0.$$

Expand the determinant, we have

$$Cov(X,Y)^2 \leq Var[X] \cdot Var[Y].$$

(2) See reference. □

Corollary 1.10.4.1 (bounds on correlations).

- *Let X and Y be two random variables with mean μ_X and μ_Y. Define correlation by*

$$\rho \triangleq \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E[(X-\mu_X)^2]}\sqrt{E[(Y-\mu_Y)^2]}}.$$

Then

$$|\rho| \leq 1$$

.

- *Let $X_1, X_2, ..., X_n$ be the iid random sample of X. Let $Y_1, Y_2, ..., Y_n$ be the iid random sample of Y. Define sample correlation by*

$$\hat{\rho} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(Y - \overline{Y})}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2}}.$$

For each realizations of $X_1, ..., X_n, Y_1, ..., Y_n$, we have

$$|\hat{\rho}| \leq 1.$$

Proof. (1) From Cauchy-Schwarz inequality, we have

$$|\rho| = \frac{|E[(X-\mu_X)(Y-\mu_Y)]|}{\sqrt{E[(X-\mu_X)^2]}\sqrt{E[(Y-\mu_Y)^2]}} \leq 1.$$

(2) Suppose we have a realization of $x_1, x_2, ..., x_n, y_1, y_2, ..., y_n$, we can define a random variable X with probability $1/n$ in taking discrete values $x_1, x_2, ..., x_n$;similarly define a random variable Y. Then

$$|\hat{\rho}| = |\rho_{XY}| \leq 1.$$

□

Lemma 1.10.2 (Popoviciu's inequality for variance). link *Consider a random variable X with support on a finite interval $[m, M]$. Then*

its variance is bounded via

$$Var[X] = \frac{(M-m)^2}{4}.$$

- *the bound is tight and can be achieve by a discrete distribution of*

$$p(X) = \begin{cases} \frac{1}{2}, X = m \\ \frac{1}{2}, X = M \end{cases}$$

Proof. Define a function $g(t) = E[(X-t)^2]$. The derivative of g with respect to t is given by $g'(t) = -2E[X] + 2t = 0$. And the g achieves its minimum at $t = E[X]$(note that $g''(E[X]) > 0$) with minimum value $g(E[X]) = Var[X]$. Consider the special point $t = \frac{M+m}{2}$, we have

$$Var[X] = g(E[X]) \leq g(\frac{M+m}{2}) = E[(X - \frac{M+m}{2})^2].$$

Now our goal is to find an upper bound on $E[(X - \frac{M+m}{2})^2] = \frac{1}{4}E[((X-m)+(X-M))^2]$.

Since $X - m \geq 0, X - M \leq 0$, we have

$$\begin{aligned}(X-m)^2 + 2(X-m)(X-M) + (X-M)^2 &\leq (X-m)^2 - 2(X-m)(X-M) + (X-M)^2 \\ ((X-m)+(X-M))^2 &\leq ((X-m)-(X-M))^2 = (M-m)^2 \\ \implies \frac{1}{4}E[((X-m)+(X-M))^2] &\leq \frac{1}{4}E[((X-m)-(X-M))^2] = \frac{(M-m)^2}{4}.\end{aligned}$$

We therefore have

$$Var[X] = \frac{(M-m)^2}{4}.$$

□

Example 1.10.2. Consider a discrete random variable X with support on $[-1,1]$, then the upper bound for its variance is given by

$$\frac{1}{4}(2)^2 = 1.$$

The bound can be achieve by a discrete distribution of

$$p(X) = \begin{cases} \frac{1}{2}, X = -1 \\ \frac{1}{2}, X = 1 \end{cases}$$

1.10.2 Chernoff bounds

Theorem 1.10.5 (Chernoff bounds). *[12]The Chernoff bound for a random variable X: for $t > 0$,*

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq E[e^{tX}]/e^{ta}$$

minimize t, we have

$$P(X \geq a) \leq \min_{t>0} E[e^{tX}]/e^{ta}$$

similarly, for $t < 0$, we have

$$P(X \leq a) \leq \min_{t<0} E[e^{tX}]/e^{ta}$$

Proof. We use the Markov inequality(Corollary 1.10.1.3) □

Remark 1.10.3. (Poisson trials)

- Poisson trials are a series of independent 0-1 random variables
- The distributions of the random variables in Poisson trials are not necessarily identical. Bernoulli trials are a special case of Poisson trials where the independent 0-1 random variables have the same distribution.

1.11 Convergence of random variables

1.11.1 Different levels of equivalence among random variables

Remark 1.11.1. Given two random variables A and B defined on the same probability space $(\Omega, \mathcal{F}, P)$, we can have the following different levels of equivalence:

- We say A is identical to B if
$$A(\omega) = B(\omega), \forall \omega \in \Omega.$$
- We say A is almost surely identical to B if
$$P(\mathcal{N}) = 0, \mathcal{N} = \{\omega, A(\omega) \neq B(\omega)\}.$$
- We say A and B have the same distribution if
$$P(A < x) = P(B < x).$$
- We say A and B have the same moments upto K if
$$E[A^k] = E[B^k], k = 1, 2, ..., K.$$

Remark 1.11.2 (element).

1.11.2 Convergence almost surely

Definition 1.11.1. *[13, p. 308]Let $\{X_n\}$ be a sequence of random variables. Then X_n converges to X almost surely if, for arbitrary $\delta > 0$ and for all $\omega \in \Omega$, we have:*
$$P(\lim_{n\to\infty} \left|X_n(\omega) - X(\omega)\right| < \delta) = 1;$$
or
$$X_n(\omega) \to X(\omega), \text{ as } n \to \infty, \forall \omega \in \Omega.$$

Remark 1.11.3 (interpretation).

- X_n converges to X almost surely if the functions $X_n(\omega)$ converges to $X(\omega)$ for all $\omega \in \Omega$ except perhaps for $s \in N, N \subset \Omega, P(N) = 0$. The probability measure 0 for the non-convergent point is the key point here.
- Note that if we view X_n as a type of function mapping, then the almost surely convergence says that X_n and X are the same (in the limit) when maps from sample space to $\mathbb{R}^n$.

Remark 1.11.4 (convergence almost surely vs. converge pointwise). If the partition the sample space Ω into two sets D and N such that $P(D) = 1$ and $P(N) = 0$. Then $X_1, X_2, ...$ converges to X almost surely is equivalently to $X_1, X_2, ...$ converges to X **pointwise** on the set of D.

Example 1.11.1. For example, let $\Omega = [0,1]$, and $X_n(\omega) = \omega + \omega^n$ and $X(\omega) = \omega$. For every $s \in [0,1)$, X_n converges to X; the non-convergent point 1 has measure of 0.

1.11.3 Convergence in probability

1.11.3.1 *Basics*

Definition 1.11.2 (convergence in probability). *[6]Let $\{X_n\}$ be a sequence of random variables and let X be a random variable defined on a sample space. We say that X_n converges in probability to X if, $\forall \epsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| >= \epsilon) = 0,$$

or equivalently

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1,$$

then we write

$$X_n \xrightarrow{P} X$$

Remark 1.11.5. Note that if the random variable X is degenerate, i.e., X has close to 1 but not 1 probability of taking a constant value a. Not 1 probability means that there will infinitely often $X_n \neq a$ as $n \to \infty$. The convergence in probability is NOT like the real sequence convergence in which when n is large enough, X_n will be arbitrarily closer to a, but in probability convergence, X_n might have small chances to **take value far from** a.

Lemma 1.11.1. *[17][9]Convergence almost surely will imply convergence in probability.*

Proof. Convergence almost surely says that given $\epsilon > 0$, there exist an N such that for all $n > N$, we have $\left|X_n(\omega) - X(\omega)\right| < \epsilon, \forall \omega \in A \in \mathcal{F}, P(A) \neq 0$. Therefore, $P(|X_n - X| < \epsilon) = 1, \forall n > N$, therefore, $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$ $\square$

Remark 1.11.6. Convergence in probability cannot imply convergence almost surely. For example, consider $X_n(\omega) = \omega + I_{[0,1/n]}(\omega), \omega \in [0,1]$, $P_n = 1 - 1/n$ therefore it converges in probability but not almost surely since the non-convergent region has measure greater than 0.

However, **If sequence $\{X_n\}$ converges to X in probability, then there is a subsequence converges to X almost surely.**

1.11.3.2 *Algebraic properties*

Theorem 1.11.1 (Algebraic properties of convergence in probability). *[6, p. 297][18, p. 1165] If $X_n \xrightarrow{P} x$ and $Y_n \xrightarrow{P} y$, then*

- $X_n + Y_n \xrightarrow{P} x + y$
- $aX_n \xrightarrow{P} ax$ *for any constant a.*
- $X_n \xrightarrow{P} x \Rightarrow g(X_n) \Rightarrow g(x)$*, for any real valued function g continuous at x*
- $X_nY_n \xrightarrow{P} xy$
- $X_n/Y_n \xrightarrow{P} x/y$*,if $y \neq 0$.*
- *If W_n is a matrix whose elements are random variables and if $\operatorname{plim} W_n = \Omega$, then*
$$\operatorname{plim} W_n^{-1} = \Omega^{-1}.$$
- *If X_n, Y_n are random matrices with $\operatorname{plim} X_n = A, \operatorname{plim} Y_n = B$, then*
$$\operatorname{plim} X_nY_n = AB.$$

Proof. (1)

$$\begin{aligned} P(|X_n + Y_n - x - y| > \epsilon) &\leq P(|X_n - x| + |Y_n - y| > \epsilon) \\ &\leq P(|X_n - x| > \epsilon/2) + P(|Y_n - y| > \epsilon/2) \to 0 \end{aligned}$$

where we have used the fact that **probability measure is monotone relative to set containment**. For the first line, $|X_n - x| + |Y_n - y| \geq |X_n + Y_n - x - y| > \epsilon$, therefore when we randomly sample X_n, Y_n, we have a higher chance to have $|X_n - x| + |Y_n - y| > \epsilon$, therefore $P(|X_n + Y_n - x - y| > \epsilon) \leq P(|X_n - x| + |Y_n - y| > \epsilon)$. For the second line, $|X_n - x| > \epsilon/2, |Y_n - y| > \epsilon/2 \Rightarrow |X_n + Y_n - x - y| > \epsilon$

(2) $P(|aX_n - ax| > \epsilon) = P(|a||X_n - x| > \epsilon) = P(|X_n - x| > \epsilon/|a|) \to 0$

(3) For any $\epsilon > 0$, there exist a δ such that $|x_n - x| < \delta \Rightarrow |g(x_n) - g(x)| < \epsilon$, therefore
$$P(|g(X_n) - g(x)| < \epsilon)) \leq P(|X_n) - x| < \delta) \to 0$$
where we have used the fact that **probability measure is monotone relative to set containment**.

(4) $X_nY_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2$, use (1)(2)(3) to prove.

(5) use (3) to prove $1/Y_n \xrightarrow{P} 1/y$. (6)(7) We can approximately view matrix inversion and matrix multiplication as a series of algebraic operations on the matrix elements. $\square$

1.11.4 Mean square convergence

Definition 1.11.3. *Let $\{X_n\}$ be a sequence of random variables. Then X_n converges to a random variable X in mean square if:*

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0.$$

Theorem 1.11.2 (mean square convergence to a constant). *Let $\{X_n\}$ be a sequence of random variables and c be a constant. We say X_n converges to c if*

- $\lim_{n\to\infty} E[X_n] = c.$
- $\lim_{n\to\infty} Var[X_n] = 0.$

Proof. Use notation $\mu_n = E[X_n]$. Note that

$$\begin{aligned}
\lim_{n\to\infty} E[(X_n - c)^2] &= \lim_{n\to\infty} E[(X_n - \mu_n + \mu_n - c)^2] \\
&= \lim_{n\to\infty} E[(X_n - \mu_n)^2] + 2\lim_{n\to\infty} E[(X_n - \mu_n)(\mu_n - c)] + \lim_{n\to\infty} E[(\mu_n - c)^2] \\
&= \lim_{n\to\infty} E[(X_n - \mu_n)^2] + 0 + 0 \\
&= 0
\end{aligned}$$

$\square$

Theorem 1.11.3 (convergence in mean square implies convergence in probability.). *Let $\{X_n\}$ be a sequence of random variables. If X_n converges to X in mean square, then X_n converges to X in probability.*

Proof. Given $\epsilon > 0$, we have

$$P(|X_n - X| > \epsilon) = P(|X_n - X|^2 > \epsilon^2) < E[(X_n - X)^2]/\epsilon^2 \to 0.$$

$\square$

1.11.5 Convergence in rth mean

Definition 1.11.4. *[13, p. 308] Let $\{X_n\}$ be a sequence of random variables. Then X_n converges to X in rth mean $r \geq 1$, if $E[X_n^r] < \infty$:*

$$\lim_{n\to\infty} E[(X_n - X)^r] = 0$$

1.11.6 Convergence in distribution

Definition 1.11.5 (Convergence in distribution). *[6, p. 300]Let $\{X_n\}$ be a sequence of random variables and let X be a random variable. Let F_{X_n} and F_X be the cumulative distribution function of X_n and X. Let $C(F_X)$ denote the set of all points where F_X is continuous. We say X_n converges in distribution to X if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \forall x \in C(F_X)$$

We denote as

$$X_n \xrightarrow{D} X$$

1.11.6.1 *Convergence in probability vs in distribution*

Theorem 1.11.4. *[6, p. 304][13, p. 311] If X_n converges to X in probability, then X_n converges to X in distribution.*

Proof. Let x be a point of continuity of $F_X(x)$. For every $\epsilon > 0$,

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x \cap |X_n - X| < \epsilon) + P(X_n \leq x \cap |X_n - X| \geq \epsilon) \\ &\leq P(X_n < x + \epsilon) + P(|X_n - X| \geq \epsilon) \end{aligned}$$

where the inequality is established by using a containing set. Then we have

$$\limsup_{n\to\infty} F_{X_n}(x) \leq P(X_n < x + \epsilon) = F_X(x + \epsilon)$$

since the second term can be arbitrarily small. Similarly, we have

$$\liminf_{n\to\infty} F_{X_n}(x) \geq P(X < x - \epsilon) = F_X(x - \epsilon)$$

We therefore have

$$F_X(x - \epsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x + \epsilon)$$

As $\epsilon \to 0$, we have $\liminf_{n\to\infty} F_{X_n}(x) = \limsup_{n\to\infty}$ as required by the continuity of F_X, then $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. □

Remark 1.11.7.

- In the above proof, we cannot directly use $\lim_{n\to\infty} F_{X_n}$ because it might not exist; however $\limsup_{n\to\infty} F_{X_n}$ always exists for bounded sequence.
- Convergence in distribution is weaker than convergence almost surely, because **it says nothing on the mapping from random experiment outcomes to** $\mathbb{R}$. For example, let X be a normal random variable, let $Y = -X$, then Y and X are the same in distribution, but X and Y are totally different mappings.

Theorem 1.11.5 (Convergence to a constant). *[6, p. 305]If X_n converges to a constant b in distribution, then X_n converges to b in probability.*

Proof. for any $\epsilon > 0$, we have $P(|X_n - b| > \epsilon) = F_{X_n}(b+\epsilon) - F_{X_n}(b-\epsilon) \to 1 - 0 = 1$. □

1.11.7 Convergence of random vectors

[6, p. 320]

1.12 Finite sampling models

The major reference for this section is link.

1.12.1 Counting principles

Theorem 1.12.1 (Fundamental counting principle). *Suppose that two events occur in order. If the first can occur in m ways and the second in n ways(after the first has occurred), then the two events can occur in order in $m \times n$ ways.*

Definition 1.12.1 (permutation).

- *A* ***permutation*** *of any r elements taken from a set of n elements is an arrangement of the r elements. We denote the number of such permutations by $P(n, r)$.*
- *A* ***permutation*** *is an arrangement of objects. For example, the permutations of three letters abc are the six arrangements:*

$$abc, acb, bac, bca, cab, cba.$$

Theorem 1.12.2 (number of permutations).

- *The number of permutations for n objects is*

$$P(n, n) = n!$$

.

- *The number of permutations of n objects taken from r at a time is*

$$P(n, r) = \frac{n!}{(n-r)!}$$

Proof. Choosing r elements from a set of size n, we have:

- the first element can be selected n ways.
- the second element can be selected $n-1$ ways(since now there are $n-1$ left).
- the third element can be selected $n-2$ ways.
- Continue the process, and the r^{th} element can be selected $n-r+1$ ways.

Using the fundamental counting principle(Theorem 1.12.1), we have

$$P(n, r) = n(n-1)(n-2)\cdots(n-r+1).$$

□

Lemma 1.12.1 (number of distinguishable permutations). *If a set of n objects consists of k different kinds of objects with n_i objects of the i kind such that $\sum_{i=1}^{k} = n$.* ***Objects from the same kind is not distinguishable****. Then then the number of distinguishable permutations of these objects is*

$$\frac{n!}{n_1!n_2!...n_k!}.$$

Definition 1.12.2 (combination). *A combination is a subset of elements of a set.*

Example 1.12.1. The combinations of size $r = 1, 2, 3$ taken from the set $\{a, b, c\}$ is given in the following table.

$r = 1$	$r = 2$	$r = 3$
$\{a\}$	$\{a, b\}$	$\{a, b, c\}$
$\{b\}$	$\{a, c\}$	
$\{c\}$	$\{b, c\}$	

Theorem 1.12.3 (number of combinations). *The number of combinations(or subsets) of size r which can be selected from a set of size n, denoted by $C(n, r)$ or $\binom{n}{r}$, is*

$$C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Proof. Because combinations are essentially permutations where order does not matter. Then

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}.$$

□

Lemma 1.12.2.

- *Given a set of n objects, the number of ways to divide them into k groups, each with n_i objects such that $\sum_{i=1}^{k} n_i = n$, is given by*

$$\frac{n!}{n_1!n_2!...n_k!}$$

- *Select n_1 objects from n objects to form a group, the number of ways is given by*

$$\frac{n!}{n_1!(n-n_1)!}.$$

Example 1.12.2. Assume 365 days a year. Among N people, the probability of exact 2 people has the same day of birthday is given as

$$365 \times \frac{1}{365} \times \frac{1}{365} \times (364 \cdot 363 \cdot \ldots (364-(n-2)+1)/365^{n-2}.$$

Example 1.12.3. Assume 365 days a year. Among N people, the probability of at least 2 people has the same day of birthday is given as

$$1 - \frac{365 \cdot 364 \cdots 365 - n + 1}{365^n}.$$

Example 1.12.4. 52 cards are randomly distributed to 4 players with each player getting 13 cards. What is the probability that each of them will have an ace.
Solution: The total possibilities are

$$N_0 = \frac{52!}{13!13!13!13!}.$$

The possibilities that each of them has an ace is

$$N_1 = \frac{48!}{12!12!12!12!}4!.$$

Then, we have

$$p = \frac{N_1}{N_0}.$$

Example 1.12.5. Imagine you have the following setup:

```
_A1_A2_A3_A4_
```

Each ace separated out evenly and we are interested in the pile that's before A1. For a standard deck of cards you have 52 cards - 4 aces = 48 cards left, and

$$\frac{48}{5} = 9.6,$$

cards for each pile. So basically you would have to turn all 9.6 cards + the A1 card in order to see the first ace. So the answer is

$$1 + \frac{48}{5}.$$

1.12.2 Matching problem

Example 1.12.6. A secretary randomly stuffs 5 letters into 5 envelopes. We want to find the probability of exactly k matches, with $k \in \{0, 1, ..., 5\}$.

Lemma 1.12.3 (sampling with replacement). *Define* $I_j = 1(X_j = j)$*.*

- $(I_1, I_2, ..., I_n)$ *is a sequence of n Bernoulli trials, with success probability* $\frac{1}{n}$*.*
- *The number of matches* N_n *is binomial distribution with parameter n and* $1/n$*.*

Lemma 1.12.4 (probability of the union of n events). *For any n events* $E_1, E_2, ..., E_n$ *that are defined on the same sample space, we have the following formula:*

$$P(E_1 \cup E_2 \cup ... \cup E_n) = \sum_{m=1}^{n} (-1)^{m+1} S_m,$$

where

$$\begin{aligned}
S_1 &= \sum_{i=1}^{n} P(E_i) \\
S_2 &= \sum_{1 \le j < k \le n} P(E_i \cap E_j) \\
&\cdots\cdots \\
S_m &= \sum_{1 \le i_1 < i_2 < ... < i_m \le n} P(E_{i_1} \cap E_{i_2} \cap ... \cap E_{i_m}).
\end{aligned}$$

In particular,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2),$$

and

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_1 \cap E_3) + P(E_1 \cap E_2 \cap E_3).$$

Lemma 1.12.5 (The matching problem). link *Suppose that the n letters are numbered $1, 2, \cdots, n$. Let E_i be the event that the i^{th} letter is stuffed into the correct envelop.*

- *$P(E_1 \cup E_2 \cup \cdots \cup E_n)$ is the probability that at least one letter is matched with the correct envelop.*
- *$1 - P(E_1 \cup E_2 \cup \cdots \cup E_n)$ is the probability that **all** letters matched incorrectly.*
- *The probability of the intersection of m events is:*

$$P(E_{i(1)} \cap E_{i(2)} \cap \cdots \cap E_{i(m)}) = \frac{(n-m)!}{n!}.$$

- *$P(E_1 \cup E_2 \cup \cdots \cup E_n)$ can be calculated using the probability of event union lemma(Lemma 1.12.4).*

Proof. (3) The calculation of the probability of intersection of m events can use the following model. There are totally $n!$ ways putting letters into envelops; there are totally $(n-m)!$ ways putting letters into envelops such that at least m specified letters are in the correct envelops. Therefore,

$$P(E_{i(1)} \cap E_{i(2)} \cap \cdots \cap E_{i(m)}) = \frac{(n-m)!}{n!}.$$

□

1.12.3 Birthday problem

Definition 1.12.3. *The sampling experiment as a distribution of n balls into m cells; X_i is the cell number of ball i. In this interpretation, our interest is in the number of empty cells and the number of occupied cells.*

Example 1.12.7. In a set of n randomly chosen people, some pair of them will have the same birthday.

Lemma 1.12.6. *Let Y_i to denote the number of balls falling into the i box, then*

$$p(Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m) = \frac{n!}{y_1! y_2! \dots y_m!} \frac{1}{m^n}, \sum_{i=1}^{m} y_i = n$$

That is, the random vector $(Y_1, ..., Y_m)$ has the multinomial distribution with parameter n and $(1/m, ..., 1/m)$.

Example 1.12.8. Assume 365 days a year.

- Among N people, the probability of exact 2 people has the same day of birthday is given as
$$365 \times \frac{1}{365} \times \frac{1}{365} \times \frac{364 \cdot 363 \cdot \ldots (364 - (n-2) + 1)}{/} 365^{n-2}.$$
- The probability that at least 2 have the same birthday is
$$1 - \frac{1}{365^n} \frac{365!}{(365-n)!}.$$

Example 1.12.9. If you randomly put 18 balls into 10 boxes, what is the expected number of empty boxes? For each box, the probability of being empty is $(\frac{9}{10})^1 8$, then the expected number of empty boxes is $10(\frac{9}{10})^1 8$.

Lemma 1.12.7 (generalized birthday problem). *link Given a year with d days, the generalized birthday problem asks for the minimal number $n(d)$ such that, in a set of n randomly chosen people, the probability of a birthday coincidence is at least 50%. It follows that $n(d)$ is the minimal integer n such that*
$$1 - (1 - \frac{1}{d})(1 - \frac{2}{d} \cdots (1 - \frac{n-1}{d})) \geq 1/2.$$

1.12.4 Coupon collector problem

Definition 1.12.4 (coupon collector problem). *Suppose that there is an urn of n different coupons. How many coupons do you expect you need to draw **with replacement** before having drawn each coupon at least once?*

Lemma 1.12.8. *Consider the coupon collector problem with m different coupons. Let Z_i denote the number of additional samples needed to go from $i-1$ distinct coupons to i distinct coupons. Let W_k denote the number of samples needed to get k distinct coupons. Then*

- *Then $Z_1, ..., Z_m$ is a sequence of independent random variables, and Z_i has the geometric distribution with parameter $p_i = \frac{m-i+1}{m}$.*
- $W_k = \sum_{i=1}^k Z_i$.

- $E[W_k] = \sum_{i=1}^{k} \frac{m}{m-i+1}$.

Proof. (1) When $i = 1$, Z_1 has a geometric distribution with parameter $p_1 = 1$. Similarly, Z_2 has a geometric distribution with parameter $p_2 = (m-1)/m$; Z_3 has a geometric distribution with parameter $p_3 = (m-2)/m$. Then, we can generalize to Z_i has a geometric distribution with parameter $p_i = (m-(i-1))/m$. (3) From the property of geometric distribution(Lemma 2.2.27),

$$E[W_k] = E[Z_1] + E[Z_2] + ... + E[Z_k] = \frac{1}{p_1} + \frac{1}{p_2} + ... + \frac{1}{p_k}.$$

□

Lemma 1.12.9. *Consider the coupon collector problem with m different coupons. Among m different coupons, there are $n, n \leq m$ are special coupons. Let Z_i denote the number of additional samples needed to go from $i-1$ distinct special coupons to i distinct special coupons. Let W_k denote the number of samples needed to get k distinct special coupons. Then*

- *Then $Z_1, ..., Z_m$ is a sequence of independent random variables, and Z_i has the geometric distribution with parameter $p_i = \frac{n-i+1}{m}$.*
- *$W_k = \sum_{i=1}^{k} Z_i$.*
- *$E[W_k] = \sum_{i=1}^{k} \frac{m}{n-i+1}$.*

Proof. (1) When $i = 1$, Z_1 has a geometric distribution with parameter $p_1 = n/m$. Similarly, Z_2 has a geometric distribution with parameter $p_2 = (n-1)/m$; Z_3 has a geometric distribution with parameter $p_3 = (n-2)/m$. Then, we can generalize to Z_i has a geometric distribution with parameter $p_i = (n-(i-1))/m$. (3) From the property of geometric distribution(Lemma 2.2.27),

$$E[W_k] = E[Z_1] + E[Z_2] + ... + E[Z_k] = \frac{1}{p_1} + \frac{1}{p_2} + ... + \frac{1}{p_k}.$$

□

1.12.5 Balls into bins model

Definition 1.12.5 (balls into bins problem). *Suppose there are m balls and n bins, balls are thrown into bins where each ball is thrown into a bin uniformly at random.*

- *Pick a bin. What is the probability for this box to be empty? What is the expected number of bins that are empty?*

- *Pick a bin. What is the probability for this box to contain exactly 1 ball? What is the expected number of bins that contain exactly 1 ball.*
- *Pick a bin. What is the probability for this box to contain exactly i balls? What is the expected number of bins that contain exactly i balls?*

Example 1.12.10. Suppose there are N types of coupons in a box. If a child draws with replacement m times from the box, what is the expected number of distinct coupon types?

View each coupon as a box. And this problem is equivalent to throw m balls into N boxes and ask the expected number of non-empty boxes.

- For each box, the probability of being empty is $(\frac{N-1}{N})^m$; therefore, the probability of being non-empty is $1-(\frac{N-1}{N})^m$.
- The expected number of empty boxes is $N(\frac{N-1}{N})^m$, and nonempty boxes is $N-N(\frac{N-1}{N})^m$.

Definition 1.12.6 (balls-into-bins distribution problems).

- *(distribution of distinguishable balls into indistinguishable bins without restriction) Suppose we want to put m distinguishable balls into n labeled bins. What is the number of ways that the balls are in different bins?*
- *(distribution of indistinguishable balls into indistinguishable bins without restriction) Suppose we want to put m indistinguishable balls into n labeled bins. What is the number of ways that the balls are in different bins?*
- *(distribution of distinguishable balls into distinguishable bins without restriction) Suppose we want to put m indistinguishable balls into n labeled bins. What is the number of ways that the balls are in different bins?*
- *(distribution of indistinguishable balls into distinguishable bins without restriction) Suppose we want to put m indistinguishable balls into n labeled bins. What is the number of ways that the balls are in different bins?*
- *(distribution without restriction I) Suppose we want to put m labeled balls into n labeled bins. What is the number of ways that the balls are in different bins such that each bin has at least has one ball?*
- *(distribution without restriction II) Suppose we want to put m labeled balls into n labeled bins. What is the number of ways that the balls are in different bins such that each bin has at least k balls?*

Lemma 1.12.10. *link*

The number of ways of putting m distinguishable balls in n distinguishable bins is m^n.

- *The number of ways of putting m distinguishable balls in n indistinguishable bins is* $m^n/n!$.
- *The number of ways of putting m indistinguishable balls in n distinguishable bins is*

$$\binom{m+n-1}{n-1}.$$

- *The number of ways of putting m indistinguishable balls in n indistinguishable bins is*

$$\binom{m+n-1}{n-1}\frac{1}{n!}.$$

$$\binom{m-1}{n-1}$$

$$\binom{m-(k-1)n-1}{n-1}$$

Proof. (1) The number of ways of putting m balls in n bins is m^n since each ball has n bins to go. □

Example 1.12.11. If there are 200 students in the library, how many ways are there for them to be split among the floors of the library if there are 6 floors?

1.13 Law of Large Number and Central Limit theorem

1.13.1 Law of Large Numbers

Theorem 1.13.1 (Weak Law of Large Numbers). *[9, p. 232]Let $\{X_n\}$ be a sequence of **iid random variables** having common mean $EX_i = \mu$ and the variance $\sigma^2 < \infty$. Let $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. Then for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(\left|\bar{X}_n - \mu\right| < \epsilon) = 1;$$

that is, X_n converges in probability to μ.

Proof. The weak law can be easily proved using probability Markov inequalities. □

Remark 1.13.1 (Cauchy random variable does not hold)**.** An example where the law of large numbers does not apply is the standard Cauchy distributionLemma 2.2.48, which does not have the expectation. And the average of n such variables has the same distribution as one such variable. The probability of the averaging deviation from μ does not tend toward zero as n goes to infinity.

Theorem 1.13.2 (Strong Law of Large Numbers). *[9, p. 235]Let $\{X_n\}$ be a sequence of iid random variable having common mean $EX_i = \mu$,$E\|X_i|] < \infty$ and the variance $\sigma^2 < \infty$. Let $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. for arbitrary $\delta > 0$:*

$$P(\lim_{n\to\infty}\left|X_n - \mu\right| < \delta) = 1$$

that is, $\bar{X}_n$ converges almost surely to μ.

Remark 1.13.2 (discussion)**.**

- Compared to weak law, strong law requires one more moment condition $E\|X_i|] < \infty$
- The weak law states that for a specified large n, the average $\overline{X}_n$ will be concentrated on μ. However, it may still have nonzero possibility that $\left|\overline{X}_n - \mu\right| > \epsilon$; that is, such situation will happen an infinite number of times, although at infrequent intervals.
- The strong law shows with probability 1, we have that for any $\epsilon > 0$, there exists an $N > 0$ such that the inequality $\left|\overline{X}_n - \mu\right| < \epsilon$ holds for all large enough $n > N$, except possible at zero-measure set.

Remark 1.13.3 (Cauchy random variable does not hold)**.** An example where the law of large numbers does not apply is the standard Cauchy distributionLemma 2.2.48, which does not have the expectation. And the average of n such variables has the same distribution as one such variable. The probability of the averaging deviation from μ does not tend toward zero as n goes to infinity.

1.13.2 Central limit theorem

Theorem 1.13.3 (central limit theorem). *[9, p. 236][6, p. 313] Let $X_1, X_2, ..., X_n$ be a sequence* ***iid random variables*** *that have mean μ and variance $\sigma^2 < \infty$. Then the random variable*

$$Y_n = \frac{(\sum_{i=1}^n X_i/n - \mu)}{\sigma/\sqrt{n}} = \frac{(\sum_{i=1}^n X_i - n\mu)}{\sqrt{n}\sigma} = \sqrt{n}(\bar{X}_n - \mu)/\sigma$$

converges in distribution to $N(0,1)$.

Proof. Use moment generating function(if exists) or characteristic function to prove.

Let $\phi(t) = E[\exp(it(X-\mu))] = \exp(\frac{i\sigma^2 t^2}{2})$ be the characteristic function of X. Then the characteristic function for Y_n can be derived via

$$\begin{aligned}
\Phi(t,n) &= E[\exp(it\frac{(\sum_{j=1}^n X_j/n - \mu)}{\sigma/\sqrt{n}})] \\
&= \phi(\frac{t}{\sigma\sqrt{n}})^n \\
&= (1 - \frac{t^2}{2n} + O((t/\sqrt{n})^3))^n \\
&\to \exp(-\frac{t^2}{2}), n \to \infty
\end{aligned}$$

where we use the Taylor expansion of $\phi(t)$ given by

$$\phi(t) = \phi(0) + \phi'(0)t + \phi''(0)\frac{t^2}{2} + O(t^3) = 1 - \sigma^2\frac{t^2}{2} + O(t^3),$$

and the limit theorem to e(**??**).

That is, as $n \to \infty$, Y_n will have its characteristic function converge to the characteristic function of the standard normal. □

Remark 1.13.4 (convergence rate). We can view the sample mean $\overline{X}_n$ has distribution similar to $N(\mu, \sigma/\sqrt{n})$ at large n. Therefore, the convergence rate is $O(1/\sqrt{N})$.

Remark 1.13.5 (Situations where central limit theorem breaks down).

- The sample mean of the iid standard Cauchy distribution random variable will not converge in distribution to standard normal; instead, the sample mean will converge to standard Cauchy distribution(Lemma 2.2.48). Note that standard Cauchy does not have finite mean and variance.

Example 1.13.1 (application of CLT for normal approximation).

- Let $X_1, ..., X_n$ be independent iid random variable of $Exp(\lambda)$, then
$$Y = \sum_{i=1}^{n} X_i$$
can be approximated(when $n \to \infty$) by
$$\frac{Y - n\mu}{\sqrt{n\sigma}} \sim N(0,1),$$
where $\mu = n/\lambda$,and $\sigma = 1/\lambda^2$.
- Let $X_1, ..., X_n$ be independent iid random variable of $Poisson(\theta)$, then
$$Y = \sum_{i=1}^{n} X_i$$
can be approximated by
$$\frac{Y - n\theta}{\sqrt{n\theta}} \sim N(0,1),$$
or equivalently
$$Y \sim N(n\theta, \theta/n).$$

Theorem 1.13.4 (general central limit theorem). *Let $X_1, X_2, ...$ be independent random variables. Suppose they have*
$$E[X_k] = \mu_k, Var[X_k] = \sigma_k^2.$$
Further let
$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2.$$
If there exists a $\delta > 0$ such that as $n \to \infty$, we have
$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E[|X_k - \mu_k|^{2+\delta}] \to 0,$$

then

$$Z = \frac{\sum_{k=1}^{n} X_k - \mu}{\sigma}$$

converges to $N(0,1)$ *in distribution, where* μ *and* σ^2 *are the mean and variance of* $\sum_{k=1}^{n} X_k$.

Lemma 1.13.1 (Slutsky's theorem). *[9, p. 239] If* $X_n \to X$ *is distribution and* $Y_n \to a$*, a constant, in probability, then*

- $Y_n X_n \to aX$ *in distribution*
- $X_n + Y_n \to X + a$ *in distribution.*

Remark 1.13.6 (linearity in convergence in distribution)**.** This importance of Slutsky's theorem is that it provides the sufficient condition for linearity in convergence in distribution.

1.13.3 Delta method & generalized CLT

Lemma 1.13.2 (first-order approximation to mean and variance of a function). *[9, p. 242] Let* $T_1, ..., T_k$ *be random variables with mean* $\mu_1, ..., \mu_k$*, and define* $T = (T_1, ..., T_k)$ *and* $\mu = (\mu_1, ..., \mu_k)$*. Define a differentiable function* $g : \mathbb{R}^k \to \mathbb{R}$*. Then we have the following first-order approximate mean and variance:*

$$\begin{aligned} E[g(T)] &\approx g(\mu) \\ Var[g(T)] &\approx \sum_{i=1}^{k} [g_i'(\mu)]^2 Var[T_i] + 2\sum_{i=1}^{k}\sum_{j>i}^{k} g_i'(\mu) g_j'(\mu) Cov_{ij}[T]. \end{aligned}$$

Proof. (1)

$$\begin{aligned} g(T = t) &= g(\mu) + \sum_{i=1}^{k} g_i'(\mu)(t - \mu) + o((t - \mu)) \\ &\approx g(\mu) + \sum_{i=1}^{k} g_i'(\mu)(t - \mu) \\ E[g(T)] &\approx g(\mu) + \sum_{i=1}^{k} g_i'(\mu)(E[T] - \mu) = g(\mu) \end{aligned}$$

(2)

$$g(T=t) \approx g(\mu) + \sum_{i=1}^{k} g_i'(\mu)(t-\mu)$$
$$Var[g(T)] \approx Var[\sum_{i=1}^{k} g_i'(\mu)(T-\mu)] = \sum_{i,j} g_i'(\mu) g_j'(\mu) Cov_{ij}$$

□

Corollary 1.13.4.1. *Let $T_1, ..., T_k$ be a iid random sample of T. Assume $E[T] = \mu$. Let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Then, we have the first-order approximation:*

$$E[g(T)] \approx g(\mu)$$
$$Var[g(T)] \approx [g'(\mu)]^2 Var[T].$$

Moreover, let $\overline{T}$ be the sample mean. Then,

$$E[g(\overline{T})] \approx g(\mu)$$
$$Var[g(\overline{T})] \approx [g'(\mu)]^2 \frac{Var[T]}{k}.$$

Example 1.13.2. Let X and Y are random variables with means μ_X and μ_Y, respectively. Let $g(x,y) = x/y$. $\frac{\partial g}{\partial x} = \frac{1}{\mu_X}, \frac{\partial g}{\partial y} = -\frac{\mu_X}{\mu_Y^2}$.

We have
$$E[\frac{X}{Y}] \approx \frac{\mu_X}{\mu_Y}$$
and
$$E[\frac{X}{Y}] \approx \frac{1}{\mu_Y^2} Var[X] + \frac{\mu_X^2}{\mu_Y^4} Var[Y] - 2\frac{\mu_X}{\mu_Y^3} Cov(X,Y).$$

Theorem 1.13.5 (Delta method for central limit theorem). *[9, p. 243] Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

$$\sqrt{n}(g(Y_n) - g(\theta)) \to N(0, \sigma^2 [g'(\theta)]^2)$$

in distribution.

1.14 Order statistics

1.14.1 Basic theory

Definition 1.14.1. *The order statistics of a random sample $X_1, ..., X_n$ are the sample values placed in ascending order. And they are denoted by $X_{(1)}, X_{(n)}, ..., X_{(n)}$, where $X_{(1)} \leq X_{(2)} ... \leq X_{(n)}$*

Theorem 1.14.1 (Discrete order statistics). *[9] Let $X_1, ..., X_n$ be a random sample from a discrete distribution with pmf $f_X(x_i) = p_i$, where $x_1 < x_2 < ...$ are possible values of X in ascending order. Define*

$$
\begin{aligned}
P_0 &= 0 \\
P_1 &= p_1 \\
&... \\
P_i &= \sum_{k=0}^{i} p_i
\end{aligned}
$$

Let $X_{(1)}, X_{(n)}, ..., X_{(n)}$ denote the order statistics from the sample. Then

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \tag{1}$$

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{n-k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \tag{2}$$

Proof. We can treat P_i as discrete version of cdf, and it means the probability of one X satisfies the inequality. The order statistics connected to binomial distribution as:

- If the minimum of Xs are less then x, then there are 1,2,...,n out of n are less than x.
- If the second minimum of Xs are less then x, then there are 2,3,...,n out of n are less than x.

□

Theorem 1.14.2 (Continuous order statistics). *[9] Let $X_1, ..., X_n$ be a random sample from a continuous distribution with pmf f_X and cdf $F_X(x)$. Let $X_{(1)}, X_{(n)}, ..., X_{(n)}$ denote the order statistics from the sample. Then*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j}$$

Proof. We can use

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} [F_X(x)]^k (1 - F_X(x))^{n-k}$$

and take derivative.
Another proof: When jth order statistic at x, that means we from n variables, we first select 1 variable to be at x, then from rest of $n-1$ variables, we select $j-1$ to be smaller than x then the rest greater than x. From combinatorics, we know

$$f_j(x) = nf(x)\binom{n-1}{j-1}(F(x))^{j-1}(1 - F(x))^{n-j}$$

□

Lemma 1.14.1 (Two order statistics). *Let $X_1, ..., X_n$ be a random sample from a continuous distribution with pmf f and cdf $F(x)$. Let $X_{(1)}, X_{(n)}, ..., X_{(n)}$ denote the order statistics from the sample. Then the joint density for $X_{(r)}$ and $X_{(s)}$ is:*

$$f_{r,s}(u, v) = \frac{n!}{(r-1)!(n-s)!(s-r-1)!} f(u)f(v)(F(u))^{r-1}(1 - F(v))^{n-s}(F(v) - F(u))^{s-r-1}$$

Proof. Use the argument similar to above: just divide the variables into five groups. □

Corollary 1.14.2.1. *Let $X_1, ..., X_n$ be a random sample from a continuous distribution with pmf f and cdf $F(x)$. Let $X_{(1)}, X_{(n)}, ..., X_{(n)}$ denote the order statistics from the sample. Then we have*

- $$f_{1,n}(u, v) = n(n-1)(F(v) - F(u))^{n-2} f(u)f(v)$$
- *(density of range) Let $W = X_{max} - X_{min}$, then*

  $$f_W(w) = \int_u f_{1,n}(u, u + w)du$$

Lemma 1.14.2 (joint density of all the order statistics). *Let $X_1, ..., X_n$ be a random sample from a continuous distribution with pmf f and cdf $F(x)$. Let $X_{(1)}, X_{(n)}, ..., X_{(n)}$ denote the order statistics from the sample. Then the conditional joint density function of $X_{(1)}, X_{(n)}, ..., X_{(n)}$ is given by*

$$f_{1,2,...,n}(y_1, ..., y_n) = n! f(y_1) f(y_2) ... f(y_n) I_{y_1 < y_2 ... < y_n}$$

Proof. The sample space of $(X_1, ..., X_n)$ can be partition into $n!$ **equal-sized** subspaces such that $X_1 < X_2 < ... < X_n, ...$ In each this subspace, there exist a map from $(X_1, ..., X_n)$ to $(X_{(1)}, X_{(n)}, ..., X_{(n)})$ with the Jacobian being 1(since it is a permutation matrix). The density for $(X_{(1)}, X_{(n)}, ..., X_{(n)})$ is $f(y_1) f(y_2) ... f(y_n) I_{y_1 < y_2 ... < y_n}$. Use the law of total probability(Lemma 1.2.4). □

Lemma 1.14.3 (distribution of max and min). *Let X be a random variable with cdf $F_X(x)$. Let $Y_n = \min(X_1, ..., X_n)$ and $Z_n = \max(X_1, ..., X_n)$, where $X_1, ..., X_n$ are n iid random sample of X. Then*

$$f_{Y_n}(x) = n f_X(x)(1 - F_X(x))^{n-1}$$

and

$$f_{Z_n}(x) = n f_X(x)(F_X(x))^{n-1}.$$

Proof.

$$P(Y_n \geq x) = (P(X \geq x))^n \implies 1 - F_{Y_n}(x) = (1 - F_X(x))^n \implies f_{Y_n}(x) = n f_X(x)(1 - F_X(x))^{n-1}$$

and

$$P(Z_n \leq x) = (P(X \leq x))^n \implies F_{Z_n}(x) = (F_X(x))^n \implies f_{Z_n}(x) = n f_X(x)(F_X(x))^{n-1}.$$

□

Corollary 1.14.2.2 (order statistics of uniform random variables). *Let X be a uniform random variable at [0,1]. Let $Y_n = \min(X_1, ..., X_n)$ and $Z_n = \max(X_1, ..., X_n)$, where $X_1, ..., X_n$ are n iid random sample of X. Then*

- $$f_{Y_n}(x) = n f_X(x)(1 - F_X(x))^{n-1} = n(1 - x)^{n-1}$$
- $$f_{Z_n}(x) = n f_X(x)(F_X(x))^{n-1} = n x^{n-1}$$

- $$f_j = \frac{n!}{(j-1)!(n-j)!}[x]^{j-1}[1-x]^{n-j} = Beta(j, n-j+1)$$

Proof. Note that we use the fact that for $U(0,1)$ distribution, $F_X(x) = x, f_X(x) = 1$. □

1.14.2 Conditional distribution and Markov property

Theorem 1.14.3. *Let $X_1, ..., X_n$ be iid random sample with cdf F and pdf f. Fix $1 < i < j < n$. Then*

- *The conditional distribution of $X_{(i)}$ given $X_{(j)} = x$ is*

$$f_{X_{(i)}|X_{(j)}=x}(u) = \frac{(j-1)!}{(i-1)!(j-1-i)!}(\frac{F(u)}{F(x)})^{i-1}(1-\frac{F(u)}{F(x)})^{j-1-i}\frac{f(u)}{F(x)}, u < x$$

  note that $f_{X_{(i)}|X_{(j)}=x}(u = x) = 0$.
- *The conditional distribution of $X_{(j)}$ given $X_{(i)} = x$ is*

$$f_{X_{(j)}|X_{(i)}=x}(u) = \frac{(n-i)!}{(n-j)!(j-1-i)!}(\frac{1-F(u)}{1-F(x)})^{n-j}(1-\frac{F(u)-F(x)}{1-F(x)})^{j-1-i}\frac{f(u)}{1-F(x)}, u < x$$

  note that $f_{X_{(i)}|X_{(j)}=x}(u = x) = 0$.

Theorem 1.14.4 (Markov property). *Let $X_1, ..., X_n$ be iid random sample with cdf F and pdf f. Fix $1 < i < j < n$. Then,*

- *the conditional distribution of $X_{(j)}$, given $X_{(1)} = x_1, ..., X_{(i)} = x_i$ is the same as the conditional distribution of $X_{(j)}$, given $X_{(i)} = x_i$.*
- *The joint distribution of $X_{(1)}, ...X_{(i)}$ given $X_{(i+1)} = x_{i+1}, ..., X_{(n)} = x_n$ is*

$$f(u_1, ..., u_{n-1}) = (n-1)!\frac{f(u_1)}{F(x_{i+1})}\frac{f(u_2)}{F(x_{i+1})}...\frac{f(u_i)}{F(x_{i+1})}I_{u_1<u_2...<u_i}$$

1.14.3 generating order statistics

Definition 1.14.2 (naive sorting method). *Suppose we want to generate $X_{(1)}, ..., X_{(n)}$, we can first generate $X_1, ..., X_n$ and then sort them.*

Definition 1.14.3 (sequential method for uniform order statistics). *[19] Suppose we want to generate $X_{(1)}, ..., X_{(n)}$ of uniform random variables, we can*

- *Generate a sample U from uniform distribution.*
- *Set $U_{(n+1)} = 1$*
- *Set $U_{(i)} = U^{1/i} U_{(i+1)}, i = n, n-1, ..., 1$*

Remark 1.14.1 (interpretation).

- $U_{(n)}$ has cdf of $F_n = x^n$, then we can use inverse transform method (**??**)to generate $U_{(n)}$.
- Use Markov property Theorem 1.14.4 to generate the rest.

1.15 Notes on bibliography

For excellent treatment on the whole topic, see[20][21]. For clear treatment on conditional expectation, see [22],[15].

For clear treatment on σ field and measure, see[1][23].

For problems in probability, see [24][25].

For treatment on measure and integral, see [26].

An excellent online resource is http://www.math.uah.edu/stat/, including random variable vector space theory(http://www.math.uah.edu/stat/expect/Spaces.html), finite sampling model(http://www.math.uah.edu/stat/urn/index.html), Brownian motion (http://www.math.uah.edu/stat/brown/Standard.html)

BIBLIOGRAPHY

1. Dineen, S. *Probability theory in finance: a mathematical guide to the Black-Scholes formula* (American Mathematical Soc., 2013).
2. Rosenthal, J. S. *A first look at rigorous probability theory* (World Scientific, 2006).
3. Wikipedia. *Borel set — Wikipedia, The Free Encyclopedia* [Online; accessed 18-May-2016]. 2016. %5Curl%7Bhttps://en.wikipedia.org/w/index.php?title=Borel_set&oldid=697808826%7D.
4. Wikipedia. *Measure (mathematics) — Wikipedia, The Free Encyclopedia* [Online; accessed 18-May-2016]. 2016. https://en.wikipedia.org/w/index.php?title=Measure_(mathematics)&oldid=705572763.
5. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
6. Hogg, R. V., McKean, J. & Craig, A. T. Introduction to Mathematical Statistics, 7 ed (2012).
7. Fries, C. *Mathematical finance: theory, modeling, implementation* (John Wiley & Sons, 2007).
8. Keener, R. *Theoretical Statistics: Topics for a Core Course* ISBN: 9780387938394. https://books.google.com/books?id=aVJmcega44cC (Springer New York, 2010).
9. Casella, G. & Berger, R. L. *Statistical inference* (Duxbury Pacific Grove, CA, 2002).
10. Stefanica, D. *A primer for the Mathematics of Financial Engineering* (Fe Press, 2008).
11. Subba Rao, S. *A course in time series analysis lecture notes* (Texas A & M, 2017).
12. Mitzenmacher, M. & Upfal, E. *Probability and computing: Randomized algorithms and probabilistic analysis* (Cambridge University Press, 2005).
13. Grimmett, G. & Stirzaker, D. *Probability and Random Processes* ISBN: 9780198572220. https://books.google.com/books?id=G3ig-0M4wSIC (OUP Oxford, 2001).
14. Williams, D. *Probability with martingales* (Cambridge university press, 1991).
15. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
16. Tripathi, G. A matrix extension of the Cauchy-Schwarz inequality. *Economics Letters* **63,** 1–3 (1999).
17. De Micheaux, P. L. & Liquet, B. Understanding convergence concepts: A visual-minded and graphical simulation-based approach. *The American Statistician* (2012).

18. Greene, W. *Econometric Analysis* ISBN: 9780134461366. https://books.google.com/books?id=Gqh2AQAACAAJ (Pearson, 2017).

19. Pasha, G., Aslam, M. & Akbar, A. Sampling Algorithm of Order Statistics for Conditional Lifetime Distributions. *Pakistan Journal of Statistics and Operation Research* **1** (2005).

20. Shiryaev, A. N. Probability: Volume 1 (Graduate Texts in Mathematics) (1996).

21. Feller, W. *An introduction to probability theory and its applications* (John Wiley & Sons, 2008).

22. Brzezniak, Z. & Zastawniak, T. *Basic stochastic processes: a course through exercises* (Springer Science & Business Media, 1999).

23. Koralov, L. & Sinai, Y. G. *Theory of probability and random processes* (Springer Science & Business Media, 2007).

24. Capinski, M. & Zastawniak, T. J. *Probability through problems* (Springer Science & Business Media, 2013).

25. Grimmett, G. & Stirzaker, D. *One thousand exercises in probability* (Oxford University Press, 2001).

26. Capinski, M. & Kopp, P. E. *Measure, integral and probability* (Springer Science & Business Media, 2013).

2

THEORY OF STATISTICAL DISTRIBUTION

2.1 General remarks

2.1.1 The goal and task of statistical inference

- Given observations of random variable X, the **ultimate goal** of statistical inference is to infer the distribution of X.
- Limited observation data usually make direct inference on the distribution of X impractical.
- We subjectively **assume** the distribution of X is given by some parameterized statistical model(e.g. Gaussian, Binomial,Poisson). More commonly, a statistical model is written as a set of distributions for X, $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$, where Ω is the parameter space containing all possible values of θ. Note that a statistical model is a hypothesis, which might be correct or incorrect.
- With the statistical model proposed, we estimate the model parameter θ from the data. Once estimated, we have a way to describe the distribution of X, which finishes the inference task.

2.1.2 Major components in statistical inference

- There are two major components in statistical inference: **statistical models** and **estimating model parameters**.
- We usually restrict ourselves to **mathematically convenient models**, such as exponential families, such that useful properties of the distribution, such as mean and variance, can be easily derived.
- We design statistic δ, which is a function of random sample X, such that δ is closed to θ or $g(\theta)$. In this way, we use statistic to relate data X to model parameter θ.
  - Not all statistics are equally good. Some are biased: $E\delta_\theta \neq \theta$. Some are more efficient in terms of using information to reduce uncertainty.
  - We can use mean-square-error(MSE), or more general risk functions to evaluate of a statistic.

2.2 Common distributions and properties

2.2.1 Bernoulli distribution

Definition 2.2.1 (Bernoulli distribution). *A random variable Y with sample space $\{0,1\}$ is said to have Bernoulli distribution $Ber(\theta)$ with parameter θ if it has a pmf given as*

$$p(y) = \theta^y(1-\theta)^{1-y}, y \in \{0,1\}.$$

Lemma 2.2.1 (basic properties). *Let X be a random variable with distribution $Ber(p)$. Then*

- $M_X(t) = (1 - p + pe^t)$.
- $E[X] = p, E[X^2] = p, Var[X^2] = p - p^2 = p(1-p)$.

Proof. Straight forward. □

2.2.2 Normal distribution

Definition 2.2.2. *[1, p. 171][normal distribution] A random variable X is said to have normal distribution $N(\mu,\sigma^2)$ with parameter μ and σ if its pdf is given as*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2}(x-\mu)^2/\sigma^2), -\infty < x < \infty$$

Lemma 2.2.2 (moment generating function). *Let X be a random variable with normal distribution $N(0,1)$, then the moment generating function is*

$$m_X(t) = \exp(\frac{1}{2}t^2).$$

If Y is a random variable with normal distribution $N(\mu,\sigma^2)$, then the moment generating function is

$$m_Y(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2).$$

Proof. (1) $m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ complete the square and get the result. (2)Let $Y = \sigma X + \mu$ and use Lemma 1.7.2. Then $m_Y(t) = e^{\mu t}m_X(\sigma t)$ □

Lemma 2.2.3 (Basic properties). *Consider* $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$.

- *If* X, Y ***are independent***, *then we have*
  - $$aX + b \sim N(a\mu + b, a^2\sigma_x^2)$$
  - $$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$
    $$aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$$
- *If* X ***and*** Y ***are not independent but jointly normal***, *then* $X + Z$ *will be normal, and*
  $$aX + bZ \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_z^2 + 2abCov(X, Z)).$$
- *Assume* X, Y *are independent. Further let* $W = \rho X + \sqrt{1-\rho^2}Y, \rho \in [-1, 1]$. *Then* W *is normal, correlated with* X *and* Y, *and the sum* $X + W$ *is also normal; that is,*
  $$W \sim N, aX + bW \sim N.$$
- ***In general, the sum of two dependent normal random variable is not necessarily normal.*** *See Lemma 2.2.10.*

Proof. (1) Directly from the properties of moment generating functions at Lemma 1.7.2. (2)The proof of two general jointly normal random variable will be showed in Lemma 2.2.10. (3) Note that

$$(X, W)^T = (X, \rho X + \sqrt{1-\rho^2}Y)^T = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} (X, Y)^T,$$

therefore (X, W) are jointly normal(Theorem 2.2.1). Then we use (2). $\square$

Lemma 2.2.4 (moments of standard normal distribution). *Let* $X \sim N(0, 1)$, *then*

$$E[X] = 0, E[X^2] = 1, E[X^3] = 0, E[X^4] = 3$$

Moreover, all odd moments are 0.

Proof. The mgf is $m(t) = e^{t^2/2}$, then

$$m'(t) = te^{t^2/2}$$

$$m''(t) = e^{t^2/2} + t^2e^{t^2/2}$$

$$\dots$$

For all odd moments
$$\int x^{2k+1} f(x) dx$$
has integrand as odd function. $\square$

Corollary 2.2.0.1 (moments of normal distribution). *Let $X \sim N(0, \sigma^2)$, then*
$$E[X] = 0, E[X^2] = \sigma^2, E[X^3] = 0, E[X^4] = 3\sigma^4$$
Moreover, all odd moments are 0.

Proof. The mgf is $m(t) = e^{\sigma^2 t^2/2}$, then
$$m'(t) = \sigma^2 t e^{\sigma^2 t^2/2}$$
$$m''(t) = \sigma^2 e^{\sigma^2 t^2/2} + \sigma^4 t^2 e^{t^2/2}$$
$$\dots$$

For all odd moments
$$\int x^{2k+1} f(x) dx$$
has integrand as odd function. $\square$

2.2.3 Half-normal distribution

Definition 2.2.3 (half-normal distribution). *Let X follow an ordinary normal distribution $N(0, \sigma^2)$. Then $Y = |X|$ follows a half-normal distribution with parameter σ. It has probability density function*
$$f_Y(y; \sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp(-\frac{y^2}{2\sigma^2}).$$

Lemma 2.2.5 (basic properties of half normal distribution). *Let Y follow a half-normal distribution with parameter σ. Then*

- $E[Y] = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$.
- $Var[Y] = \sigma^2(1 - \frac{2}{\pi})$

2.2.4 Laplace distribution

Definition 2.2.4 (Laplace distribution). *A random variable X has a Laplace distribution, denoted by $Lap(\mu, b)$, if its probability density function is*

$$f(x|\mu,b) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b}) = \begin{cases} \frac{1}{2b}\exp(-\frac{\mu-x}{b}), if\ x < \mu \\ \frac{1}{2b}\exp(-\frac{x-\mu}{b}), if\ x \geq \mu \end{cases}.$$

Lemma 2.2.6 (properties of Laplace distribution). *Let X be a random variable with Laplace distribution with parameter μ, b. It follows that*

- *The mean and the median are μ.*
- *The variance is $2b^2$.*
- *The cdf is given by*

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} \frac{1}{2}\exp(-\frac{\mu-x}{b}), if\ x < \mu \\ 1-\frac{1}{2}\exp(-\frac{x-\mu}{b}), if\ x \geq \mu \end{cases} = \frac{1}{2} + \frac{1}{2}sgn(x-\mu)(1-\exp(-\frac{|x-\mu|}{b})).$$

- *The inverse cdf is given by*

$$F^{-1}(p) = \mu - b \cdot sgn(p-0.5)\ln(1-2|p-0.5|).$$

2.2.5 Multivariant Gaussian distribution

2.2.5.1 *Basic definitions*

Definition 2.2.5. *[1, p. 180] [2, p. 40]The multivariate Gaussian/normal distribution on $\mathbb{R}^n$ is defined as*

$$\rho(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|det\Sigma|^{1/2}}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

with mean $\mu \in \mathbb{R}^n$ and covariance matrix Σ. A random vector is said be multivariate Gaussian/normal if it's pdf is multivariate Gaussian/normal distribution.

Lemma 2.2.7 (alternative definition via mgf). *[1, p. 181] An n-dimensional random vector X has a multivariate normal distribution with mean vector μ and covariance matrix Σ if its mgf is*

$$M_X(t) =\triangleq E[\exp(t^T X)] = \exp(t^T \mu + \frac{1}{2} t^T \Sigma t)$$

for all $t \in \mathbb{R}^n$.

Proof.

$$\begin{aligned} M_X(t) &= E[\exp(t^T X)] \\ &= \exp(E[t^T X] + \frac{1}{2} Var[t^T X]) \\ &= \exp(t^T \mu + \frac{1}{2} t^T \Sigma t) \end{aligned}$$

where we use the fact that $t^T X \sim N(t^T \mu, t^T \Sigma t)$ from Theorem 2.2.1. □

Remark 2.2.1 (implication). Given a random vector X, if we want to check whether X is a multivariate Gaussian, we can check its mgf. If its mgf is the exponential of a linear form plus a quadratic form, then it is multivariate Gaussian.

Lemma 2.2.8 (alternative definition via linear combination). *A vector $X = (X_1, X_2, ..., X_n)^T$ is a multivariate Gaussian distribution if every linear combination*

$$S = a^T X, a \in \mathbb{R}^n$$

has a normal distribution.

Proof. Because $a^T X$ is normal, then it has characteristic function

$$E[\exp(ita^T X)] = \exp(ita^T \mu_X - \frac{1}{2} t^2 (a)^T \Sigma_X a).$$

Since a is arbitrary, we can say for any $t' \in \mathbb{R}^n$, we have

$$E[\exp(it' X)] = \exp(i[t']^T \mu_X - \frac{1}{2} [t']^T \Sigma_X t).$$

That is, X is multivariate Gaussian. □

Example 2.2.1 (bivariate Gaussian distribution). Let $f(x,y)$ be the density of a bivariate Gaussian distribution $MN(\mu,\Sigma)$, where

$$\mu = \begin{Bmatrix} \mu_X \\ \mu_Y \end{Bmatrix}, \Sigma = \begin{Bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{Bmatrix}.$$

Then,

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}]).$$

2.2.5.2 *Affine transformation and its consequences*

Theorem 2.2.1 (affine transformation for Multivariate normal distribution). *[1, p. 183] Let X be a n dimensional random vector with $MN(\mu,\Sigma)$ distribution. Let $Y = AX+b, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$. Then Y is an m dimensional random vector having a $MN(A\mu+b, A\Sigma A^T)$ distribution.*

Proof. Use moment generating function to prove. Let $Y = AX+b$, then from Lemma 1.7.5

$$M_Y(t) = e^{t^Tb}M_X(A^Tt) = e^{t^T(A\mu+b)+\frac{1}{2}t^TA\Sigma A^Tt}$$

which suggesting $Y \sim MN(A\mu+b, A\Sigma A^T)$ □

Lemma 2.2.9 (orthonormal transformation maintains independence). *Let X be a n dimensional random vector with $MN(0,I)$. If C is an orthonormal matrix, then $Y = CX$ has distribution $MN(0,I)$. That is, orthonormal transformation will preserve independence.*

Proof. $Cov(Y) = C^TIC = I$. □

Lemma 2.2.10 (sum of two multivariate normal random variable). *Let $X_1 \sim MN(\mu_1,\Sigma_1)$ and $X_2 \sim MN(\mu_1,\Sigma_1)$ be two n dimensional multivariate normal random variable. It follows that*

- *If X_1 and X_2 are independent, then $Y = X_1+X_2$ is a normal random variable with $MN(\mu_1+\mu_2,\Sigma_1+\Sigma_2)$.*

- *If X_1 and X_2 are dependent but (X_1, X_2) are joint normal with covariance matrix given as*

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12} & \Sigma_2 \end{bmatrix}$$

then $Y = X_1 + X_2$ is a normal random variable with $MN(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2 + 2\Sigma_{12})$.

Proof. (1) Consider a $2n$ dimensional multivariate normal random variable Z with distribution $\mu = [\mu_1; \mu_2]$, $\Sigma = \Sigma_1 \oplus \Sigma_2$. Then construct transformation matrix

$$A = \begin{bmatrix} I_n & I_n \end{bmatrix}$$

Then (Theorem 2.2.1) $Y = AZ$, and $Y \sim MN(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$ (2) same as (1). □

Note 2.2.1. caution! The joint distribution of two Gaussian margins are not necessarily joint Gaussian:

- Two multivariate normal random variables are not necessarily joint normal.[a]. For example, consider two marginal distribution of Gaussian. For Gaussian copula, the joint distribution is multivariate Gaussian; however, for other copulas including Frank copula and Clayton copula, the joint distribution is not multivariate Gaussian.
- If two multivariate normal random variables are independent, then they are joint normal.

[a] link

2.2.5.3 *Marginal and conditional distribution*

Lemma 2.2.11 (marginal distribution). *[2, p. 41]The multivariant Gaussian distribution $\rho(x; \mu, \Sigma)$ on $\mathbb{R}^n$ has marginal distribution on $\mathbb{R}^k, k \leq n$ given as $\rho(x_1; \mu_1, \Sigma_{11}), x_1 \in \mathbb{R}^k$ where we decompose*

$$\mu = [\mu_1, \mu_2]^T, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Proof. Use above Theorem 2.2.1. Let

$$A = \begin{bmatrix} I & 0 \end{bmatrix}$$

Then $X_1 = AX$. □

Lemma 2.2.12 (full joint distribution can be constructed from pair joint distribution). *Let $X = (X_1, X_2, ..., X_n)^T$ be a random multivariate Gaussian vector with mean $\mu = (\mu_1, \mu_2, ..., \mu_n)^T$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. The the pair $(X_i, X_j), i \neq j$ has joint distribution*

$$\hat{\mu} = (\mu_i, \mu_j), \hat{\Sigma} \in \mathbb{R}^{2\times 2}, \hat{\Sigma}_{11} = \Sigma_{ii}, \hat{\Sigma}_{12} = \Sigma_{ij}.$$

That is, all the pair joint distribution can construct the full joint distribution.

Proof. Directly from Lemma 2.2.11. □

Remark 2.2.2 (caution! not all the distribution has this property)**.** If the full joint distribution is not Gaussian, then such property (reconstruct full distribution from pair distribution) will not generally hold.

Theorem 2.2.2 (conditional distribution). *[2, p. 43] The multivariate Gaussian distribution $\rho(x; \mu, \Sigma)$ on $\mathbb{R}^n$ has marginal distribution on $\mathbb{R}^k, k \leq n$ given as*

$$\frac{\rho(x_1, x_2)}{\rho(x_2)} = \rho(x_1; \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

where we decompose

$$\mu = [\mu_1^T, \mu_2^T]^T, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with $\mu_1 \in \mathbb{R}^k, \mu_2 \in \mathbb{R}^{n-k}$.

Proof. See link □

Remark 2.2.3 (gaining information)**.** From the conditional distribution, we can see that given the information of x_2, the mean of x_1 will be corrected and the variance of x_1 will be reduced.

Example 2.2.2 (bivariate Gaussian distribution). Let $f(x, y)$ be the density of a bivariate Gaussian distribution $MN(\mu, \Sigma)$, where

$$\mu = \begin{Bmatrix} \mu_X \\ \mu_Y \end{Bmatrix}, \Sigma = \begin{Bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{Bmatrix}.$$

Then,
$$X|Y \sim N(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y), (1-\rho^2)\sigma_X^2).$$

2.2.5.4 *Box Muller transformation*

Lemma 2.2.13 (Box Muller transformation). *Let $X, Y \sim N(0,1)$ and X, Y be independent. Let*
$$R = \sqrt{X^2+Y^2}, \Theta = \arctan(Y/X)$$
. Then

- *R and Θ are independent.*
- *$\Theta \sim U(0, 2\pi)$ and $F_R(r) = 1 - \exp(-r^2/2)$.*
- *Suppose we have U_1, U_2 being independent uniform on $[0,1]$. Then $2\pi U_1$ and $\sqrt{-2\ln(1-U_2)}$ are independent and have the same distribution of R and Θ.*
- *Further, $\sqrt{-2\ln(1-U_2)}\cos(2\pi U_1)$ and $\sqrt{-2\ln(1-U_2)}\sin(2\pi U_1)$ are independent and have the same distribution of X and Y.*

Proof. (1)Using polar transformation Lemma 1.4.10, we have
$$Pr(R < r, \Theta < \theta) = \int_0^r \int_0^\theta \frac{1}{2\pi}\exp(-\frac{r^2}{2}) r dr d\theta = \int_0^r \int_0^\theta \exp(-\frac{r^2}{2}) r dr \frac{1}{2\pi} d\theta = F_R(R < r) F_\Theta(\Theta < \theta).$$
Using independence condition Lemma 1.4.3, we know that R and Θ are independent. (2) Integrate directly in (1). (3) Let $U = 1 - \exp(-R^2/2)$. Based on probability integral transform(**??**), we know U is an uniform random variable. Or equivalently, $R = \sqrt{-2\ln(1-U)}$ has the same distribution of R. (4) Note that $X = R\cos(\Theta), Y = R\sin(\Theta)$. □

2.2.6 Lognormal distribution

2.2.6.1 *Univariate lognormal distribution*

Definition 2.2.6 (lognormal distribution). *A random variable Y has a lognormal distribution with parameters μ and σ^2, written as*
$$Y \sim LN(\mu, \sigma^2)$$
if $\log(Y)$ is normally distributed as $N(0, \sigma^2)$. Several equivalent definitions are:

- *$Y \sim LN(\mu, \sigma^2)$ if and only if $\log(Y) \sim N(\mu, \sigma^2)$.*
- *$Y \sim LN(\mu, \sigma^2)$ if and only if $Y = e^X$ with $X \sim N(\mu, \sigma^2)$.*

- *The distribution function is given as*

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp(-\frac{(\ln y - \mu)^2}{2\sigma^2}).$$

Lemma 2.2.14 (basic properties of lognormal distribution). *Let $Y \sim LN(\mu, \sigma^2)$, or equivalently $Y = \exp(X), X \sim N(\mu, \sigma^2)$ then*

- *The distribution function for Y is given as*

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp(-\frac{(\ln y - \mu)^2}{2\sigma^2}).$$

- $E[Y] = \exp(E[X] + \frac{1}{2}Var[X^2]) = \exp(\mu + \sigma^2/2)$.
- $E[Y^2] = \exp(2\mu + 2\sigma^2), E[Y^m] = \exp(m\mu + \frac{1}{2}m^2\sigma^2)$
- $Var[Y] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. *In particular $\mu = 0$, we have*

$$E[Y] = \exp(\frac{1}{2}\sigma^2), E[Y^m] = \exp(\frac{1}{2}m^2\sigma^2), Var[Y] = \exp(2\sigma^2) - \exp(\sigma^2).$$

- *If $X_1 \in N(\mu_1, \sigma_1^2), X_2 \in N(\mu_2, \sigma_2^2)$, then*

$$E[\exp(X_1 + X_2)] = \exp(E[X_1] + E[X_2] + \frac{1}{2}Var[X] + \frac{1}{2}Var[X_2] + Cov(X_1, X_2)).$$

- $$\mu = \log(\frac{E[Y]^2}{\sqrt{E[Y^2]}}), \sigma^2 = \ln(\frac{E[Y^2]}{E[Y]^2}).$$

- *The median of Y is $\exp(\mu)$.*
- *skewness*

$$(\exp(\sigma^2) + 2)\sqrt{\exp(\sigma^2) - 1} > 0.$$

Proof. (1) Note that

$$x = \ln y, f_Y(y) = f_X(\ln y)\left|\frac{d \ln y}{dy}\right|.$$

(2)(3) Note that for $X \sim N(\mu, \sigma^2)$, $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. Then

$$E[Y] = E[\exp(X)] = M_X(1) = \exp(\mu + \frac{1}{2}\sigma^2),$$

and

$$E[Y^2] = E[\exp(2X)] = M_X(2) = \exp(2\mu + 2\sigma^2),$$

and
$$Var[Y] = E[Y^2] - (E[Y])^2 = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$

(4) Note that $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$. Then we use (1). (5) Note that the exponential is a monotone function, the median of Y will be $\exp(median\ X) = \exp(\mu)$, where we used the fact that median of X is μ. □

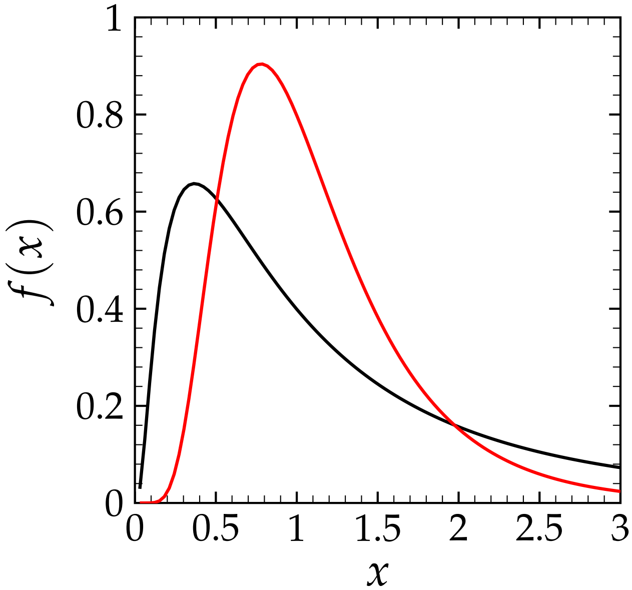


Figure 2.2.1: Density of $LN(0,1)$(black and $LN(0,0.5)$(red). Note the positive skewness.

2.2.6.2 *Extension to univariate lognormal distribution*

Definition 2.2.7. *[3]*

- ***regular log-normal distribution*** *with parameter* (μ, σ^2) *is given by*
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{(\ln x - \mu)^2}{2\sigma^2}), x > 0.$$
- ***negative log-normal distribution*** *with parameter* (μ, σ)*, denoted by* $NLN(\mu, \sigma^2)$ *is given by*
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{(\ln -x - \mu)^2}{2\sigma^2}), x < 0.$$

- ***shifted log-normal distribution*** *with parameter* (μ, σ, τ),*denoted by* $SLN(\mu, \sigma^2, \tau)$ *is given by*

$$f(x) = \frac{1}{(x-\tau)\sigma\sqrt{2\pi}} \exp(-\frac{(\ln x - \tau - \mu)^2}{2\sigma^2}), x > \tau.$$

- ***negative shifted log-normal distribution*** *with parameter* (μ, σ^2, τ) *is given by*

$$f(x) = \frac{1}{(-x-\tau)\sigma\sqrt{2\pi}} \exp(-\frac{(\ln(-x-\tau) - \mu)^2}{2\sigma^2}), x < -\tau.$$

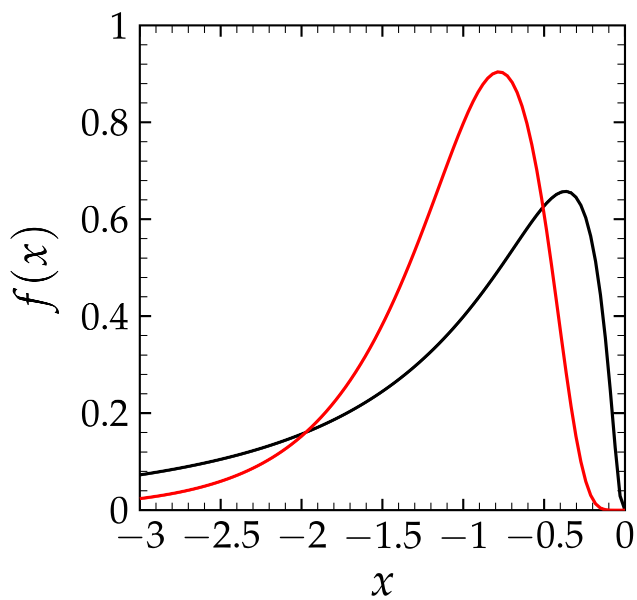


Figure 2.2.2: Density of $NLN(0,1)$(black and $NLN(0,0.5)$(red). Note the negative skewness.

Lemma 2.2.15. *Let* $X \sim LN(\mu, \sigma^2)$. *It follows that*

- *Let* $Y = -X$. *Then* $Y \sim NLN(\mu, \sigma^2)$.
- *Let* $Z = X + \tau$. *Then* $Y \sim NLN(\mu, \sigma^2, \tau)$.
- *Let* $W = -X - \tau$. *Then* $Y \sim NSLN(\mu, \sigma^2, \tau)$.

Proof. Straight forward from definition and transformation. □

Lemma 2.2.16 (basic properties of shifted lognormal distribution). *Let* $X \sim SLN(\mu, \sigma^2, \tau)$. *Then*

- $$E[X] = \tau + \exp(\mu + \frac{1}{2}\sigma^2).$$
- $$E[X^2] = \tau^2 + 2\tau\exp(\mu + \frac{1}{2}\sigma^2) + \exp(2\mu + 2\sigma^2).$$
- $$E[X^3] = \tau^3 + 3\tau^2\exp(\mu + \frac{1}{2}\sigma^2) + 3\tau\exp(2\mu + 2\sigma^2) + \exp(3\mu + \frac{9}{2}\sigma^2).$$

Proof. Note that from Lemma 2.2.14, we have if $Y \sim LN(\mu, \sigma^2)$, then $E[Y^m] = \exp(m\mu + \frac{1}{2}m^2\sigma^2)$. Then, we use

$$E[X] = E[Y + \tau] = E[Y] + \tau,$$
$$E[X^2] = E[(Y + \tau)^2] = E[Y^2] + 2\tau E[Y] + \tau^2,$$
$$E[X^3] = E[(Y + \tau)^3] = E[Y^3] + 3\tau E[Y^2] + 3\tau^2 E[Y] + \tau^3.$$

□

2.2.6.3 *Moment matching approximation*

Lemma 2.2.17 (2 parameter Log-normal approximation via moment matching). *Suppose we have a random variable X having moments given by*

$$E[X] = M_1, E[X^2] = M_2.$$

Let Y be a log-normal random variable defined by

$$Y = M_1 \exp(-\frac{1}{2}v^2 + vZ), Z \in N(0,1),$$

where

$$v^2 = \log(M_2/M_1^2)$$

Then Y has the same first two moments as X; that is

$$E[Y] = M_1, E[Y^2] = M_2.$$

Proof. Using moment generating function of Z, we know that

$$E[Y] = M_Z(v) M_1 \exp(-\frac{1}{2}v^2) = M_1.$$

and

$$E[Y^2] = M_Z(2v) M_1^2 \exp(-v^2) = \exp(v^2) M_1^2 = \frac{M_2}{M_1^2} M_1^2 = M_2.$$

□

Lemma 2.2.18 (3 parameter shifted lognormal approximation via moment matching). *Suppose we have a random variable X having moments given by*

$$E[X] = M_1, E[X^2] = M_2, E[X^3] = M_3.$$

Let Y be a shifted log-normal random variable with parameter $SLN(\mu, \sigma^2, \tau)$ such that

$$E[Y] = \tau + \exp(\mu + \frac{1}{2}\sigma^2),$$

$$E[Y^2] = \tau^2 + 2\tau\exp(\mu + \frac{1}{2}\sigma^2) + \exp(2\mu + 2\sigma^2),$$

$$E[Y^3] = \tau^3 + 3\tau^2\exp(\mu + \frac{1}{2}\sigma^2) + 3\tau\exp(2\mu + 2\sigma^2) + \exp(3\mu + \frac{9}{2}\sigma^2).$$

If we can find (μ, σ, τ) such that

$$E[X] = E[Y], E[X^2] = E[Y^2], E[X^3] = E[Y^3],$$

then X and Y have matched moments.

Proof. For moments of Y, see Lemma 2.2.16. □

Note 2.2.2 (choice of approximating distribution). [3] We can based on the target distribution's location and skewness to choose the type of lognormal distribution we want to use. The table is good summary.

skewness	$\eta > 0$	$\eta > 0$	$\eta < 0$	$\eta < 0$
location	$\tau \geq 0$	$\tau < 0$	$\tau \geq 0$	$\tau < 0$
choice of approximation	regular	shifted	negative	negative shifted

2.2.6.4 *Multivariate lognormal distribution*

Definition 2.2.8 (multivariate lognormal distribution). *If $X = (X_1, X_2, ..., X_n) \sim MN(\mu, \Sigma)$, then $Y = \exp(X) = (\exp(X_1), \exp(X_2), ..., \exp(X_n)) \sim MLN(\mu, \Sigma)$, i.e., Y has multivariate lognormal distribution*

Lemma 2.2.19 (basic properties of multivariate lognormal distribution). *Let* $X = (X_1, X_2, ..., X_n) \sim MN(\mu, \Sigma)$ *and* $Y = \exp(X) = (\exp(X_1), \exp(X_2), ..., \exp(X_n))$. *Then*

- $E[Y_i] = \exp(\mu_i + \frac{1}{2}\Sigma_{ii})$.
- $E[Y_iY_j] = \exp(\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj} + 2\Sigma_{ij})) = E[Y_i][Y_i]\exp(\Sigma_{ij})$.
- $Var[Y_i] = \exp(2\mu_i + \Sigma_{ii})(\exp(\Sigma_{ii}) - 1)$.
- $Cov[Y_iY_j] = \exp(\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj}))(\exp(\Sigma_{ij}) - 1)$.

Proof. (1)Note that $M_X(t) = \exp(t^T\mu + \frac{1}{2}t^T\Sigma t), t \in \mathbb{R}^n$, and $E[Y_i] = M_X(e_i)$. (2) Let $t = e_i + e_j$. Then

$$E[Y_iY_j] = M_X(t) = \exp(\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj} + 2\Sigma_{ij})).$$

(3)

$$Var[Y_i] = E[Y_iY_i] - E[Y_i]E[Y_i].$$

(4)

$$Cov[Y_i, Y_j] = E[Y_iY_j] - E[Y_i]E[Y_j].$$

□

2.2.7 Exponential distribution

Definition 2.2.9 (exponential distribution). *A random variable* X *is said to have an exponential distribution* $Exp(\lambda)$ *with parameter* λ *if it has pdf given as*

$$p(x|\lambda) = \lambda\exp(-\lambda x)$$

with $x \in [0, \infty)$.

Lemma 2.2.20 (basic properites). *Let* X *be a random variable with exponential distribution with parameter* λ, *then we have*

- $E[X] = 1/\lambda$
- $Var[X] = 1/\lambda^2$
- *memoryless:*
$$P(X > s + t|X > s) = P(X > t)$$
(even though $P(X > s + t) < P(X > t)$*)*

Proof. (1)(2) are straightforward. (3) The cmf is given as

$$F(t) = \int_0^t \lambda \exp(-\lambda\tau)d\tau = 1 - \exp(-\lambda t)$$

$$P(X > s+t|X > s) = \frac{P(X > s+t \cap X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)} = \frac{\exp(-\lambda(s+t))}{\exp(-\lambda t)} = \exp(-\lambda t)$$

□

Remark 2.2.4 (interpretation of memorylessness). Supppose we are waiting for an event to occur, and we model the waiting time as a random variable X with $Exp(\lambda)$. If we already wait for s time, the distribution that we need to wait an extra of t time is the same as the distribution of the waiting time at time 0.

Remark 2.2.5. Exponential distribution is the only memoryless continuous distribution.[4].

Lemma 2.2.21 (Normal approximate sum of Exponential). *Let $X_1, ..., X_n$ be independent iid random variable of $Exp(\lambda)$, then*

$$Y = \sum_{i=1}^{n} X_i$$

can be approximated(when $n \to \infty$) by

$$\frac{Y - n\mu}{\sqrt{n\sigma}} \sim N(0,1),$$

where $\mu = n/\lambda$,and $\sigma = n/\lambda^2$.

Proof. Directly from Central Limit Theorem(Theorem 1.13.3). Also see Gamma distribution properties, since exponential distribution is a special case of Gamma distribution. □

2.2.8 Poisson distribution

Definition 2.2.10 (Poisson distribution). *A discrete random variable X is said to have a Poisson distribution $Poisson(\lambda)$ with parameter λ if it has pmf given as*

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

with $x \in \{0, 1, 2, ...\}$.

Lemma 2.2.22 (basic property of Poisson distribution). *[1, p. 154] Let X be a random variable with distribution Poisson*(λ)*. Then*

- $M(t) = \exp(\lambda(e^t - 1))$.
- $E[X] = \lambda, Var[X] = \lambda$.

Proof. (1)

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{\lambda e^t}. \end{aligned}$$

(2) $E[X] = M'_X(0) = \lambda, E[X^2] = M''_X(0) = \lambda^2 - \lambda$. □

Lemma 2.2.23 (sum of Poisson distribution). *Assume* $X_1, ..., X_n$ *to be independent random variables, and* $X_i \sim Poisson(\theta_i), i = 1, ..., n$*. Then*

$$Y = \sum_{i=1}^{n} X_i \sim Poisson(\sum_{i=1}^{n} \theta_i).$$

Proof. Note that

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \exp \sum_{i=1}^{n} \theta_i (e^t - 1).$$

□

Lemma 2.2.24 (Normal approximate sum of Poisson). *Let* $X_1, ..., X_n$ *be independent iid random variable of Poisson*(θ)*, then*

$$Y = \sum_{i=1}^{n} X_i$$

can be approximated by

$$\frac{Y - n\theta}{\sqrt{n\theta}} \sim N(0, 1),$$

or equivalently

$$Y \sim N(n\theta, n\theta).$$

Proof. Directly from Central Limit Theorem(Theorem 1.13.3). □

2.2.9 Gamma distribution

Definition 2.2.11 (Gamma distribution). *[5, p. 42] A random variable X is said to have a Gamma distribution Gamma(a, b) with parameter a, b if it has pdf given as*

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

with support $x \in (0, \infty)$.

Remark 2.2.6 (exponential distribution is a special case). An exponential distribution with parameter b is a Gamma distribution $Gamma(1, b)$with

$$f(x) = be^{-bx}.$$

Remark 2.2.7 (Application in arrival times of Poisson process). If $N(t)$ is a Poisson process with rate λ, then the arrival time $T_1, T_2, ...$ have $T_n \sim Gamma(n, \lambda)$ distribution.(See Lemma 7.1.4)

Caution! Gamma distribution is different from Gamma function $\Gamma(t)$, which is given as

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

Remark 2.2.8 (conjugate prior for Poisson distribution). Gammad distribution conjugate prior for the parameter of Poisson distribution. When integrate out x in $\Gamma(t)$, we have

$$\int_0^\infty x^{a-1} e^{-bx} dx = \Gamma(a)/b^2$$

Lemma 2.2.25 (mean and variance). *The Gamma distribution Gamma(a, b) has mean a/b and variance a/b^2.*

Proof. Using the property of

$$\int_0^\infty x^{a-1} e^{-bx} dx = \Gamma(a)/b^a,$$

we can show the result. □

Theorem 2.2.3 (sum of Gamma random variables). *[1, p. 163] Let $X_1, ..., X_n$ be independent random variables. Suppose $X_i \sim Gamma(a_i, b), \forall i = 1, ..., n$. Then*

$$Y = \sum_{i=1}^{n} X_i \sim Gamma(\sum_{i=1}^{n} a_i, b)$$

Proof. This can be proved using moment generating functions. □

Lemma 2.2.26 (Normal approximate sum of Gamma). *Let $X_1, ..., X_n$ be independent iid random variables of $Gamma(a, b)$, then*

$$Y = \sum_{i=1}^{n} X_i$$

can be approximated(when $n \to \infty$) by

$$\frac{Y - n\mu}{\sqrt{n\sigma}} \sim N(0, 1),$$

where $\mu = na/b$,and $\sigma = na/b^2$.

Proof. Directly from Central Limit Theorem(Theorem 1.13.3). □

2.2.10 Geometric distribution

Definition 2.2.12 (geometric distribution). *A discrete random variable X is said to have a geometric distribution $Geo(\theta)$ with parameter θ if it has pmf given as*

$$p(X = k) = (1 - \theta)^{k-1}\theta$$

with $k \in \{1, 2, ...\}$.

Remark 2.2.9 (relation to Bernoulli trias). The geometric distribution The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set 1, 2, 3, ...

Lemma 2.2.27 (basic statistics of geometric distribution). *The expected value of a geometrically distributed random variable X with parameter p is $1/p$ and the variance is $(1 - p)/p^2$.*

Proof. (1)

$$E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$(1-p)E[X] = \sum_{k=1}^{\infty} k(1-p)^{k}p$$

subtract and get $pE[X] = 1$.
(2)

$$Var[X] = \sum_{k=1}^{\infty} (k-1/p)^2(1-p)^{k-1}p$$

can be proved similarly. □

2.2.11 Binomial distribution

Definition 2.2.13 (binomial distribution). *A discrete random variable X is said to have a Binomial distribution $Binomial(n,p)$ with parameter n,p if it has pmf given as*

$$f(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

with $x \in \{0,1,2,...,n\}$.

Remark 2.2.10 (interpretation). Binomial distribution represents the probability distribution of the number of successes in a sequence of n independent binary experiments, each of which yields 1 with probability p.

Remark 2.2.11 (relation to Bernoulli distribution). Let X_i be iid random variables with Bernoulli distribution of parameter p, then

$$Y = \sum_{i=1}^{n} X_i$$

is a random variable of binomial distribution with parameter (n,p).

Lemma 2.2.28 (sum of independent binomial random variable). *Let $X_1, X_2, ..., X_K$ be the independent binomial random variables with parameter $(n_1,p),(n_2,p),...,(n_K,p)$. Let $Y = \sum_{i=1}^{K} X_i$. Then*

- $M_{X_i}(t) = (1-p+pe^t)^{n_i}, i = 1,...,K.$
- $M_Y(t) = (1-p+pe^t)^{\sum_{i=1}^{K} n_i}$
- $Y \sim Binomial(\sum_{i=1}^{K} n_i, p).$

Proof. (1) Use the mgf of Bernoulli distribution(Lemma 2.2.1). (2)(3) Consider $X_1 \sim Binomial(n_1, p)$ and $X_2 \sim Binomial(n_2, p)$, each has momemt generating function of $(1 - p + pe^t)^{n_1}$ and $(1 - p + pe^t)^{n_2}$. $X_1 + X_2$ will have mgf of $(1 - p + pe^t)^{n_1+n_2}$(Lemma 1.7.2), corresponding to $Binomial(n_1 + n_2, p)$. It is straight forward to extend multiple cases. □

Lemma 2.2.29 (convergence of binomial distribution to Poisson distribution). *Suppose that $p_n \in (0,1)$ for $n \in \mathcal{N}_+$ and $np_n \to \lambda$ as $n \to \infty$. Then the binomial distribution with parameters n and p_n converges to the Poisson distribution with parameter λ **in distribution** as $n \to \infty$. That is, for fixed $k \in \mathcal{N}$,*

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}$$

as $n \to \infty$.

Proof. (direct method)Note that

$$\begin{aligned}
\binom{n}{k} p_n^k (1 - p_n)^{n-k} &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} (p_n)^k (1 - p_n)^{n-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} \\
&\approx \frac{n^k}{k!} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} \\
&= \frac{\lambda^k}{k!} (1 - \frac{\lambda}{n})^{n-k} \\
&\to e^{-\lambda \frac{n-k}{n}} \frac{\lambda^k}{k!} \\
&\approx e^{-\lambda} \frac{\lambda^k}{k!}
\end{aligned}$$

(use generating function) Note that binomial distribution has probability generating function(Definition 1.7.6)

$$((1 - p_n) + p_n s)^n = (1 + (pns - p_n)n/n)^n \to e^{(s-1)a}, n \to \infty$$

where $e^{(s-1)a}$ is the generating function of Poisson distribution. □

Remark 2.2.12 (Poisson distribution as an approximate for large n and small k)**.** Note that the lemma requires that k fixed. In other words, when $n \gg k$, we can use Poisson distribution to approximate binomial distribution.

2.2.12 Hypergeometric distribution

Definition 2.2.14 (hypergeometric distribution distribution). *[1, p. 148] A random variable X is said to have a hypergeometric distribution $HG(N, K, n)$ with parameter N, K, n if it has pmf given as*

$$p(x = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

with support $x \in \{0, 1, ..., \min(n, K)\}$. Note that the parameters should be non-negative integers and satisfying

$$N \geq K, N \geq n.$$

Remark 2.2.13 (interpretation). $p(x = k)$ describes the probability of k successes in n draws, without replacement, from a finite population of size N that contains exactly K successes.

Lemma 2.2.30 (combinatorial identities). *Assuming $K \geq n$, we have*

$$\sum_{0 \leq k \leq n} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = 1$$

Lemma 2.2.31 (mean of a hypergeometric distribution). *[1, p. 148] Let X be a random variable with $HG(N, K, n)$, then its mean is*

$$E[X] = n\frac{K}{N}$$

2.2.13 Beta distribution

Definition 2.2.15 (Beta distribution). *[5, p. 43] A random variable X is said to have a Beta distribution $B(a, b)$ with parameter a, b if it has a pdf given as*

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

with support $x \in [0, 1]$.

Remark 2.2.14.

- $$\int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

- Beta distribution is commonly **used as the conjugate prior for binomial distribution**, where
$$p(y_1,...,y_n|\theta) = \theta^{\sum_i y_i}(1-\theta)^{n-\sum_i y_i}, y_i \in \{0,1\}$$
then the posterior distribution will also be Beta.

Lemma 2.2.32 (basic property). *Let X be a random variable with distribution $B(a,b)$.*

- $$E[X] = \frac{a}{a+b}.$$
- $$E[X^2] = \frac{a(a+1)}{(a+b)(a+b+1)}.$$
- $$E[X^r] = \frac{a(a+1)\cdots(a+r-1)}{(a+b)(a+b+1)\cdots(a+b+r-1)}.$$
- $$Var[X] = \frac{ab}{(a+b)^2(a+b+1)}.$$
- *The mode of X, i.e., the value x that has the maximum probability is*
$$x^* = \frac{a-1}{a+b-2}.$$

Proof. (1)This can be proved using properties of Gamma distribution.

$$\begin{aligned}
E[X] &= \int_0^1 xf(x)dx \\
&= \int_0^1 \frac{x^a(1-x)^{b-1}}{B(a,b)} \\
&= \frac{B(a+1,b)}{B(a,b)} \\
&= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} / \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\
&= \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a+b+1)\Gamma(a)} \\
&= \frac{a}{a+b}
\end{aligned}$$

(2)

$$\begin{aligned}
E[X^2] &= \int_0^1 x^2 f(x)dx \\
&= \int_0^1 \frac{x^{a+1}(1-x)^{b-1}}{B(a,b)} \\
&= \frac{B(a+2,b)}{B(a,b)} \\
&= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} / \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\
&= \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a+b+2)\Gamma(a)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}$$

(3) Use $Var[X] = E[X^2] - E[X]^2$. (4) To find the maximizer for $x^{a-1}(1-x)^{b-1}$, we take the log and maximize it. We have

$$\ln f(x) = (a-1)\ln x + (b-1)\ln(1-x).$$

Take the derivative with respect to x and set to 0, we have

$$\begin{aligned}
\frac{a-1}{x} &= \frac{b-1}{1-x} \\
(a-1)(1-x) &= x(b-1) \\
\implies x^* &= \frac{a-1}{a+b-2}.
\end{aligned}$$

□

2.2.14 Multinomial distribution

Definition 2.2.16. *[5, p. 35] A discrete random vector $X = (X_1, ..., X_n)$ is said to have multinomial distribution with parameters $(p_1, ..., p_n)$ and m if its pmf is given as*

$$f(x_1, x_2, ..., x_n) = \frac{m!}{x_1!...x_n!} p_1^{x_1} ... p_n^{x_n}$$

where we require $x_i \in \{0, ..., m\}, \sum x_i = m, \sum p_i = 1$.

Remark 2.2.15. Consider m independent experiment, each has n outcomes with probability p_i to occur. The outcome distribution is given as[6]

$$f(x_1, x_2, ..., x_n) = \frac{m!}{x_1!...x_n!} p_1^{x_1} ... p_n^{x_n}$$

where $\sum x_i = m, \sum p_i = 1$.

Lemma 2.2.33 (basic properties of Multinomial Distribution). *Let $X = (X_1, ..., X_n)$ discrete random vector with multinomial distribution with parameters $p = (p_1, ..., p_n)$ and m.*

-
- $$E[X_i] = np_i.$$
- $$Var[X_i] = np_i(1 - p_i), Cov(X_i, X_j) = np_i(1 - p_i),$$

  or in vector form
  $$Var[X] = n(diag(p) = pp^T).$$

Proof. (1)This can be proved using properties of Gamma distribution.

$$\begin{aligned} E[X_i] &= \int_0^1 x_i f(x) dx \\ &= \frac{\prod_{k=1}^K \Gamma(a_k + \delta_{ik})}{\Gamma(a_0 + 1)} / \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(a_0)} \\ &= \frac{a_i}{a_0} \end{aligned}$$

(2)

$$\begin{aligned} E[X_i^2] &= \int_0^1 x_i^2 f(x) dx \\ &= \frac{\prod_{k=1}^K \Gamma(a_k + 2\delta_{ik})}{\Gamma(a_0 + 2)} / \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(a_0)} \\ &= \frac{a_i(a_i + 1)}{(a_0 + 1)a_0} \end{aligned}$$

(3) Use $Var[X] = E[X^2] - E[X]^2$. (4) To find the maximizer for $f(x)$, we take the log and maximize it. The optimality condition requires that $x_i^* \propto a_i - 1$ and $\sum_{i=1}^K a_i = 1$. $\square$

2.2.15 Dirichlet distribution

Definition 2.2.17. *[5, p. 49] A random vector $X = (X_1, ..., X_K)$ is said to have a Dirichlet distribution distribution with parameter $a = (a_1, ..., a_K)$ if it has pdf given as*

$$f(x_1, ..., x_K) = \frac{1}{B(a)} \prod_{k=1}^{K} x_k^{a_k - 1}$$

with support $x \in \{x : 0 \leq x_k \leq 1, \sum_k x_k = 1, \forall k = 1, 2..., K\}$, and $B(a)$ is a normalization constant given as

$$B(a) = \frac{\prod_{k=1}^{K} \Gamma(a_k)}{\Gamma(\sum_k a_k)},$$

where $\Gamma(\cdot)$ is the Gamma function.

Remark 2.2.16.

- Dirichlet distribution can be viewed as multivariate generalization of Beta distribution.
- Dirichlet distribution is usually **used as the conjugate prior** for multinomial distribution.

Lemma 2.2.34 (basic properties of Dirichlet Distribution). *Let $X = (X_1, X_2, ..., X_K), x_i \in (0, 1), \sum_{i=1}^{K} x_i = 1$, be a random vector with distribution $B(a), a \in \mathbb{R}^K$. Let $a_0 = \sum_{i=1}^{K} a_i$.*

- $$E[X_i] = \frac{a_i}{\sum_{i=1}^{K} a_i}.$$
- $$E[X_i^2] = \frac{a_i(a_i + 1)}{(a_0)(a_0 + 1)}.$$
- $$Var[X_i] = \frac{a_i(a_0 - a_i)}{a_0^2(a_0 + 1)}.$$
- *The mode of X, i.e., the value x that has the maximum probability is*

$$x_i^* = \frac{a_i - 1}{a_0 - K}.$$

Proof. (1)This can be proved using properties of Gamma distribution.

$$\begin{aligned} E[X_i] &= \int_0^1 x_i f(x) dx \\ &= \frac{\prod_{k=1}^K \Gamma(a_k + \delta_{ik})}{\Gamma(a_0+1)} / \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(a_0)} \\ &= \frac{a_i}{a_0} \end{aligned}$$

(2)

$$\begin{aligned} E[X_i^2] &= \int_0^1 x_i^2 f(x) dx \\ &= \frac{\prod_{k=1}^K \Gamma(a_k + 2\delta_{ik})}{\Gamma(a_0+2)} / \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(a_0)} \\ &= \frac{a_i(a_i+1)}{(a_0+1)a_0} \end{aligned}$$

(3) Use $Var[X] = E[X^2] - E[X]^2$. (4) To find the maximizer for $f(x)$, we take the log and maximize it. The optimality condition requires that $x_i^* \propto a_i - 1$ and $\sum_{i=1}^K a_i = 1$. □

2.2.16 χ^2-distribution

2.2.16.1 *Basic properties*

Definition 2.2.18. *A random variable X is said to have a $\chi^2(n)$ distribution with parameter $n \in \mathbb{Z}_+$ if it has pdf given as*

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-y/2}$$

with $x \in (0, +\infty)$

Remark 2.2.17 (special case of Gamma distribution)**.** $\chi^2(n)$ has the same distribution of Gamma(n/2,2).

Definition 2.2.19 (alternative). *The χ^2-distribution with k degrees of freedom is the distribution of a sum of squares of k independent standard normal random variables. Mathematically, if $X_1, X_2, ..., X_k$ are iid random variable with $X_i \sim N(0,1)$, the random variable*

$$Q = \sum_{i=1}^k X_i^2$$

is distributed according to the χ^2 distribution with k degrees of freedom, writen as $Q \sim \chi^2(k)$.

Lemma 2.2.35 (basic property). *[1, pp. 161–163] Let X_1, X_2 be independent random variables. Suppose $X_1 \sim \chi^2(a_1), X_2 \sim \chi^2(a_2)$. Then*

- $Y = X_1 + X_2 \sim \chi^2(a_1 + a_2)$
- $\lambda X_1 \sim \lambda^2 \chi^2(a_1)$
- *The moment generating function is given by*

$$M(t) = (1-2t)^{-r/2}.$$

Proof. (1)This can be proved using properties of Gamma distribution. (2) λX_1 can be viewed as the sum of squares of normal random variables Y_i with $N(0, \lambda^2)$. Then $\sum_{i=1}^{n} (Y_i/\lambda)^2 \sim \chi^2(n)$.

□

Lemma 2.2.36 (expectation and variance). *Let random variable X has distribution of $\chi^2(n)$, then*

$$E[X] = n, Var[X] = 2n$$

In particular,

$$E[X/n] = 1, Var[X/n] = 0 \text{ as } n \to \infty.$$

that is the random variable X/n becomes deterministic constant as $n \to \infty$.

Proof. (1)Let $Z \sim \chi^2(1), Z = Y^2, Y \sim N(0,1)$, then $E[Z] = Var[Y] + (E[Y])^2 = 1$. $Var[Z] = E[Z^2] - (E[Z])^2 = E[Y^4] - 1 = 3 - 1 = 2$. (2) Use linearity of expectation that $E[X/n] = E[X]/n = 1$. Use $Var[X/n] = Var[X]/n^2 = 2/n$. □

2.2.16.2 *Quadratic forms and chi-square distribution*

Definition 2.2.20 (quadratic forms of random vectors). *[1, p. 485] Let $X = (X_1, X_2, ..., X_n)^T$ be a random vector, we called*

$$Q = X^T \Sigma X, \Sigma \in \mathbb{R}^{n \times n},$$

a quadratic form of random vector X.

Note that Q is also a random variable.

Lemma 2.2.37. *Let X be a m-dimensional random vector with multivariate Gaussian distribution, i.e., $X \sim N(\mu, \Sigma)$. It follows that*

- $$\Sigma^{1/2}(x-\mu) \sim N(0, I).$$
- $$(x-\mu)\Sigma^{-1}(x-\mu) \sim \chi^2(m).$$

Proof. (1) Directly from affine transformation property of multivariate Gaussian random variable (Theorem 2.2.1). (2) Use the definition that sum of iid normal random variable square is chi-square random variable. □

Theorem 2.2.4 (chi-square orthogonal decomposition). *Let $X_1, X_2, ..., X_n$ be independent standard normal variables such that*

$$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n).$$

Denote $X = (X_1, ..., X_n)^T$. If there exists an orthogonal projector $P \in \mathbb{R}^{n\times n}$ such that $Y = PX, Z = (I-P)X$, then

- *$Y \sim MN(0, P), Z \sim MN(0, I-P)$, and Y, Z are independent of each other.*
- *$Y^T Y \sim \chi^2(r), r = rank(P)$; or equivalently, the quadratic form $Q = X^T PX \sim \chi^2(r)$.*
- *$Z^T Z \sim \chi^2(n-r)$;or equivalently, the quadratic form $Q = X^T(I-P)X \sim \chi^2(n-r)$*

In summary, for a quadratic form $Q = X^T \Sigma X$, if Σ is idempotent and symmetric, then $Q \sim \chi^2(rank(\Sigma))$.

Proof. (1) From affine transform of multivariate normal(Theorem 2.2.1),

$$Y \sim MN(0, P\Sigma_X P^T) = MN(0, P^2) = MN(0, P).$$

To show independence, we have $E[YZ^T] = E[PXX^T(I-P)^T] = E[P(I-P)] = 0$.

(2) Let U be the eigen-decomposition of P such that $P = UU^T$. Let $Z = U^T X, Z \in \mathbb{R}^r, Z \sim MN(0, I_r)$. Let V be the eigen-decomposition of $I-P$ such that $I-P = VV^T$.

Let $W = V^T X, W \in \mathbb{R}^{n-r}, W \sim MN(0, I_{n-r})$. We want to show that the characteristic function of the random quantity $Y^T Y$ is the same as the characteristic function of $\chi^2(r)$.

$$\begin{aligned}
&E[\exp(itY^TY)] \\
=&\frac{1}{(2\pi)^{n/2}}\int\int\cdots\int \exp(it(U^TX)^T(U^TX)\exp(-\frac{1}{2}X^T(I-P+P)X))dx_1dx_2\cdots dx_n \\
=&\frac{1}{(2\pi)^{n/2}}\int\int\cdots\int \exp(itZ^TZ)\exp(-\frac{1}{2}(Z^TZ+W^TW))dz_1\cdots dz_r dw_{r+1}\cdots dw_n \\
=&\frac{1}{(2\pi)^{r/2}}\int\int\cdots\int \exp(itZ^TZ)\exp(-\frac{1}{2}(Z^TZ))dz_1\cdots dz_r
\end{aligned}$$

where we change the integral variable such that

$$[dz_1\cdots dz_r dw_{r+1}\cdots dw_n]^T = [U\ V](dx_1dx_2\cdots dx_n)^T$$

. The last line is the characteristic function of $\chi^2(r)$. (3) similar to (2). □

Lemma 2.2.38 (moment generating functions for Gaussian quadratic forms). *[1, p. 523] Let $X = (X_1, X_2, ..., X_n)^T$ where $X_1, X_2, ..., X_n$ are iid $N(0,1)$. Consider the quadratic form $Q = X^T AX$ for a symmetric matrix A of rank $r \leq n$. It follows that*

- *Q has the moment generating function $M(t) = \prod_{i=1}^r (1-2t\lambda_i)^{-1/2} = |I-2tA|^{-1/2}$, where $\lambda_1, \lambda_2, ..., \lambda_r$ are the nonzero eigenvalues of A, $|t| < \frac{1}{\max|\lambda_i|}$.*
- *If A is an orthogonal projector such that $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$, then*

$$M(t) = M_{\chi^2(r)}.$$

Proof. (1)Let the eigen-decomposition of A be

$$A = U\Lambda U^T, U \in \mathbb{R}^{n\times r}, \Lambda \in \mathbb{R}^{r\times r}.$$

Then

$$Q = X^TAX = X^TU\Lambda U^TX = X^T(\sum_{i=1}^r \lambda_i u_i u_i^T)X = \sum_{i=1}^r \lambda_i (u_i^T X)^T.$$

Let $Y_i = u_i^T X, i = 1,2,...,r$. It can be shown that $Y_i \sim N(0,1), E[Y_iY_j] = u_i^T E[XX^T]u_j^T = \delta_{ij}$; that is $Y_1, Y_2, ..., Y_r \sim MN(0, I_r)$. Therefore, $Y_i^2 \sim \chi^2(1)$.

The moment generating function is given by

$$\begin{aligned}
M(t) &= E[\exp(tQ)] \\
&= E[\exp(t\sum_{i=1}^{r}\lambda_i Y_i^2)] \\
&= \prod_{i=1}^{r} E[\exp(t\lambda_i Y_i^2)] \\
&= \prod_{i=1}^{r} M_{\chi^2(1)}(\lambda_i t) \\
&= \prod_{i=1}^{r}(1-2\lambda_i t)^{-1/2}
\end{aligned}$$

where we use the moment generating function of $\chi^2(1)$ from Lemma 2.2.35. (2) straight forward. □

Lemma 2.2.39 (independence of quadratic forms). *[1, p. 528] Let $X = (X_1, X_2, ..., X_n)$ be a random vector where $X_1, X_2, ..., X_n$ are iid $N(0,1)$. For real symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, let $Q_1 = X^T AX$ and $Q_2 = X^T BX$. Then Q_1 and Q_2 are independent if and only if $AB = 0$.*

Proof. Let $rank(A) = r, rank(B) = s$. Let the eigendecomposition of A, B be such that

$$A = \sum_{i=1}^{r}\lambda_i u_i u_i^T, A = \sum_{i=1}^{s}\beta_i v_i v_i^T.$$

If $AB = 0$, then $u_1, ..., u_r, v_1, ..., v_r$ will be orthogonal to each other. Then

$$Q_1 + Q_2 = \sum_{i=1}^{r+s}\lambda_i u_i u_i^T,$$

where $u_{r+i} = v_i, \lambda_{r+i} = \beta_i$.

It is easy to see that(Lemma 2.2.38)

$$M_{Q_1,Q_2}(t_1,t_2) = M_{Q_1}(t_1)M_{Q_2}(t_2).$$

Then from independence-from-mgf(Lemma 1.4.5), we can prove Q_1 and Q_2 are independent.

□

2.2.16.3 *Noncentral chi-squared distribution*

Definition 2.2.21 (noncentral chi-squared distribution). *Let $(X_1, X_2, ..., X_k)$ be k independent, normally distributed random variables with mean μ_i and unit variances. Then the random variable*

$$Y = \sum_{i=1}^{k} X_i^2$$

is distributed according to the ***noncentral chi-squared distribution*** *with parameter k specifying the degree of freedom and λ, known as the* ***noncentrality parameter****, given by*

$$\lambda = \sum_{i=1}^{k} \mu_i^2.$$

2.2.17 Wishart distribution

Definition 2.2.22 (Wishart distribution). *Let $X_1, ..., X_n$ be independent p dimensional multivariate normal random vector with distribution $MN(0, V)$. Let $X = [X_1, ..., X_n]$. Then $M = XX^T$ is said to have Wishart distribution with parameter (n, p, V).*

Definition 2.2.23 (Wishart distribution). *A random matrix $M \in \mathbb{R}^{p\times p}$ is said to have the Wishart distribution with parameters $W_p(n, V)$ if it has pdf*

$$f(M) = \frac{1}{2^{np/2}\Gamma_p(\frac{n}{2}|V|^{n/2})}|M|^{n-p-1/2}\exp(\frac{1}{2}Tr[V^{-1}M]),$$

with the support M be the set of all symmetric positive definite matrices. Here $\Gamma_p(\alpha)$ is the multivariate gamma function.

Lemma 2.2.40 (basic properties).

- *(reduction to χ^2) If $M \in \mathbb{R}^{1\times 1}$, then*

$$M \sim W_1(n, \sigma^2) = \sigma^2\chi^2(n).$$

- *For $M \sim W_p(n, V)$, then $B^T MB \sim W_m(n, B^T VB)$, where $B \in \mathbb{R}^{p\times m}$.*
- *For $M \sim W_p(n, V)$, then $V^{-1/2}MV^{-1/2} \sim W_m(n, I)$.*
- *If M_i are independent $W_p(n_i, V)$, then $\sum_{i=1}^k \sim W_p(\sum_{i=1}^k n_i, V)$.*
- *If $M \sim W_p(n, V)$, then $E[M] = nV$.*
- *If M_1, M_2 are independent and $M_1 + M_2 = M \sim W_p(n, V)$. Further if $M_1 \sim W_p(n_1, V)$, then $M_2 \sim W_p(n - n_1, V)$.*

Lemma 2.2.41 (sample covariance). *The sample covariance*

$$\hat{Cov} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T$$

where X_i are iid $MN(0, V)$, and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, has the property of

$$E[\hat{Cov}] = V.$$

2.2.18 t-distribution

2.2.18.1 *Standard t distribution*

Definition 2.2.24 (t distribution). *[1, p. 192] A random variable X is said to have a $t(n)$ distribution with parameter $n \in \mathbb{Z}_+$ if it has pdf given as*

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi n}}(1 + \frac{t^2}{n})^{-(n+1)/2}$$

with $x \in (-\infty, +\infty)$

Definition 2.2.25 (alternative). *Let random variable $W \sim N(0,1)$, Let random variable $V \sim \chi^2(n)$* ***independent of*** *V. Define a new random variable T as*

$$T = \frac{W}{\sqrt{V/n}}$$

Then T has a t-distribution with degree of freedom n, denoted by T_n or t_n.

Remark 2.2.18 (comparison with normal distribution).

- t distribution generally have shorter peak and fatter tails than normal distribution.
- $t_n \to N(0,1)$ as $n \to \infty$.

Lemma 2.2.42 (mean and variance of t-distribution). *The mean for a t-distribution with degree of n is given by*

$$E[t_n] = \begin{cases} 0, n > 1 \\ \infty(undefined), n = 1 \end{cases}.$$

The variance for a t-distribution with degree of n is given by

$$Var[t_n] = \begin{cases} \frac{n}{n-2}, n > 2 \\ \infty, n = 1,2 \end{cases}.$$

2.2.18.2 *classical t distribution*

Definition 2.2.26. *[7, p. 95] If Y has a standard t_n distribution, then*

$$Z = \mu + \lambda Y$$

is said to have a $t_n(\mu, \lambda^2)$ distribution.

Lemma 2.2.43 (mean and variance of classical t-distribution). *Let Z be a random variable $t_n(\mu, \lambda^2)$. Then*

$$E[Z] = \begin{cases} \mu, n > 1 \\ \infty(undefined), n = 1 \end{cases},$$

and

$$Var[Z] = \begin{cases} \lambda^2 \frac{n}{n-2}, n > 2 \\ \infty, n = 1,2 \end{cases}.$$

2.2.18.3 *Multivariate t distribution*

Definition 2.2.27 (multivariate t distribution). *[7]*

- *Let Z be a d dimensional multivariate Gaussian $MN(0, \Sigma)$, and $\mu \in \mathbb{R}^d$. The d dimensional random vector Y, defined as,*

  $$X = \mu + \sqrt{\frac{n}{W}} Z,$$

  *where $W \sim \chi^2(n)$ and W is **independent** of Z, has a $t_n(\mu, \Sigma)$ multivariate distribution.*

- *Let $X \sim t_n(\mu, \Sigma)$. Then X has the density given by*

$$f(x) = \frac{\Gamma((n+d)/2)}{\Gamma(n/2)n^{d/2}\pi^{d/2}|\Sigma|^{1/2}}(1 + \frac{1}{n}(x-\mu)^T\Sigma^{-1}(x-\mu))^{-(n+d)/2}.$$

Lemma 2.2.44 (mean and variance of multivariate t-distribution). *Let Z be a random variable $t_n(\mu, \Sigma)$. Then*

$$E[Z] = \begin{cases} \mu, n > 1 \\ \infty(undefined), n = 1 \end{cases},$$

and

$$Cov[Z] = \begin{cases} \Sigma\frac{n}{n-2}, n > 2 \\ \infty, n = 1, 2 \end{cases}.$$

2.2.18.4 *Student's Theorem*

Theorem 2.2.5 (Student's Theorem). *[1, p. 194] Let $X_1, X_2, ..., X_n$ be iid random variables each having a normal distribution with mean μ and variance σ^2. Define random variables as:[1]*

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{n} X_i, S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

1. *$\bar{X}$ has a $N(\mu, \sigma^2/n)$ distribution*
2. *$\bar{X}$ and S^2 are independent.*
3. *$(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution*
4. *The random variable*

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has t-distribution with $n-1$ degrees of freedom.

Proof. (1)From Lemma 2.2.3. (2) We can prove $\overline{X}$ and the random vector $Y = (X_1 - \overline{X}, ..., X_n - \overline{X})$ are independent. Note that

$$\overline{X} = \frac{1}{n}\mathbf{1}^T X, Y = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X,$$

and hence $\overline{X}$ and Y are both normal.

$$\begin{aligned}
Cov(\overline{X}, Y) &= X^T(\frac{1}{n}\mathbf{1}^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T))X \\
&= X^T\frac{1}{n}(\mathbf{1}^T - \frac{1}{n}\mathbf{1}^T\mathbf{1}\mathbf{1}^T)X \\
&= X^T\frac{1}{n}(\mathbf{1}^T - \mathbf{1}^T))X = 0
\end{aligned}$$

where we use the fact that $\mathbf{1}^T\mathbf{1} = n$.

Then $S^2 = \frac{1}{n-1}Y^TY$ will be independent of $\overline{X}$ because S^2 is a function of Y(Lemma 1.3.4). (3) See reference and Corollary 3.10.1.1. (4) From the definition of the t distribution, we have

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is the $N(0,1)$. $W = (n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution. Then

$$\frac{Y}{\sqrt{W/n-1}}$$

has $t(n-1)$ distribution. □

2.2.19 F-distribution

Definition 2.2.28 (F distribution). *[1, p. 192] A random variable X is said to have a $F(n_1, n_2)$ distribution with parameter $n_1, n_2 \in \mathbb{Z}_+$ if it has pdf given as*

$$f(x) = \frac{\Gamma((n_1+n)2)/2)(n_1/n_2)^{n_1/2}y^{n_1/2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1x/n_2)]^{(n_1+n_2)/2}}$$

with $x \in (0, +\infty)$

Definition 2.2.29 (alternative). *Given two **independent** chi-squared random variables W and V having r_1 and r_2 degrees of freedom. We define a new random variable*

$$W = \frac{U/r_1}{V/r_2}$$

Then W has a F-distribution with parameter (r_1, r_2).

Lemma 2.2.45 (inverse relationship). *Let X be a random variable with distribution $F(n_1, n_2)$, then $1/X$ is a random variable with distribution $F(n_2, n_1)$.*

Proof. Directly from definition. □

Lemma 2.2.46 (relationship to t distribution). *Let X be a random variable with standard t distribution with n degrees of freedom. Then*

$$X^2 \sim F(1, n).$$

That is, X^2 has the distribution of $F(1, n)$.

Proof. Directly from definition. □

Definition 2.2.30 (noncentral F distribution). *Given two chi-squared random variables W and V such that V is a noncentral chi-squared random variable with non-centrality parameter λ and degree of freedom r_1, and W is a chi-squared random variable having r_1 r_2 degrees of freedom. We define a new random variable*

$$W = \frac{U/r_1}{V/r_2}$$

Then W has a noncentral F-distribution with parameter (λ, r_1, r_2). .

2.2.20 Empirical distributions

Definition 2.2.31 (empirical cumulative distribution function(CDF)). *Given N iid random variables $Y_1, Y_2, ..., Y_N$ with common cdf $F(t)$, the empirical CDF is defined by*

$$\hat{F}_N(t) = \frac{number\ of\ elements\ in\ the\ sample \leq t}{N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{Y_i \leq t}$$

.

Lemma 2.2.47 (basic statistic properties). *Let $\hat{F}_N(t)$ be the empirical cdf of a random sample of size N. For a fixed t, we have*

- *$N\hat{F}_N(t)$ is a binomial random variable with parameter (N, p),where $p = F(t)$.*
- *$N\hat{F}_N(t)$ is an unbiased estimator for $NF(t)$.*

- $N\hat{F}_N(t)$ *has variance* $NF(t)(1-F(t))$.

Proof. (1) Note that based on the definition of $\hat{F}_N(t)\frac{1}{N}\sum_{i=1}^N \mathbf{1}_{Y_i \leq t}$, $\mathbf{1}_{Y_i \leq t}$ is a Bernoulli random variable with parameter $p = F(t)$. Therefore, $N\hat{F}_N(t) = \sum_{i=1}^N \mathbf{1}_{Y_i \leq t}$ will follow a binomial distribution of parameter (N, p). (2)

$$E[N\hat{F}_N(t)] = Np = NF(t).$$

(3)

$$Var[N\hat{F}_N(t)] = Np(1-p) = NF(t)(1-F(t)).$$

□

2.2.21 Heavy-tailed distributions

2.2.21.1 *Basic characterization*

Definition 2.2.32 (Heavy-tailed distribution). *The distribution of a random variable X with distribution function F is said to have a heavy right tail if*

$$\lim_{x\to\infty} e^{\lambda x} Pr(X > x) = \infty, \forall \lambda > 0.$$

Remark 2.2.19 (interpretation). Heavy-tailed distributions have densities decaying slower in the tails than the normal.

2.2.21.2 *Pareto and power distribution*

Definition 2.2.33 (Pareto distribution). *A random variable X is said to have Pareto distribution with scale parameter* $x_m > 0$ *and shape parameter* $\alpha > 0$ *if its has pdf*

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, x \geq x_m, \\ 0, x < x_m. \end{cases};$$

or cdf

$$f_X(x) = \begin{cases} 1 - (\frac{\alpha x_m}{x})^\alpha, x \geq x_m, \\ 0, x < x_m. \end{cases}.$$

X has support $[x_m, \infty)$.

Definition 2.2.34 (power law distribution). *A random variable X is said to have power law distribution with parameters K, α if its has probability characterization on its tail given by*

$$Pr(X > x) = Kx^{-\alpha}.$$

Remark 2.2.20 (Pareto distribution and power law distribution are heavy-tailed distribution)**.** Note that since power grows much slower than the exponentialA.7.1, therefore

$$\lim_{x\to\infty} e^{\lambda x} Pr(X > x) = \infty, \forall \lambda > 0.$$

2.2.21.3 *Student t distribution family*

Definition 2.2.35 (Student's t-Distribution family). *The t distribution has a single parameter, $\nu > 0$, known as* degrees of freedom*. The density function is given as*

$$f_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{x^2}{\nu})^{-\frac{1}{2}(\nu+1)}$$

The first two members of family are

1. $f_1(x) = \frac{1}{\pi(1+x^2)}$
2. $f_2(x) = \frac{1}{2\sqrt{2}}(1 + x^2/2)^{-3/2}$

The $\nu = 1$ density is known as Cauchy's density*. As $\nu \to \infty$, the density distribution tends to the standard normal density.*

Definition 2.2.36 (Cauchy distribution). *The Cauchy distribution with parameter (x_0, γ) has the probability density function*

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma}(\frac{\gamma^2}{(x - x_0)^2 + \gamma^2}),$$

where x_0 is the location parameter, specifying the location of the peak of the distribution, and γ is the scale parameter which specifies the half-width at half-maximum. ***Standard Cauchy distribution*** *is Cauchy distribution with parameter $(0, 1)$.*

Remark 2.2.21 (nonexistence of moments)**.**

- The Cauchy distribution is an example of a distribution which has no mean, variance or higher moments. And therefore the moment generating function does not exist. However, the **mode and median** are well defined and both equal to x_0.
- The nonexistence of expectation is because of the $E[\![X|] < \infty$.

Lemma 2.2.48 (sum of Cauchy distribution). *If $X_1, ..., X_n$ are independent and identically distributed random variables, each with a standard Cauchy distribution, then the sample mean*

$$\overline{X} = (X_1 + ... + X_n)/n$$

has the same standard Cauchy distribution.

Proof. Note that we need to use characteristic function to prove, since the moment generating function does not exist. □

2.2.21.4 *Gaussian mixture distributions*

Definition 2.2.37 (normal scale mixture distribution). *[7, p. 99] The normal scale mixture distribution is the distribution of the random variable*

$$Y = \mu + \sqrt{U}Z,$$

where μ is constant equal to the mean, and $Z \sim N(0,1)$, U is a positive random variable giving the variance of each component, and Z and U are independent.

*If U can assume only a finite number of values, then Y has a **discrete scale mixture distribution**. If U is continuously distributed, then Y has a **continuous scale mixture distribution**.*

Example 2.2.3 (discrete Gaussian mixture distribution). Let $\mu = 0$, and U have the following distribution

$$P(U = 25) = 0.1, P(U = 1) = 0.9.$$

Then

$$Y = \mu + \sqrt{U}Z,$$

is the mixture of 10% of $N(0, 25)$ and 90% of $N(0, 1)$.

Example 2.2.4 (t distribution). The t_n distribution with n degrees of freedom is a continuous Gaussian mixture with

$$\mu = 0, U = \frac{n}{W},$$

where $W \sim \chi^2(n)$.

Definition 2.2.38 (multivariate normal variance mixtures). *The random vector X has a multivariate normal variance mixture distribution if*

$$X \triangleq \mu + \sqrt{W} A Z$$

where

- $Z \sim MN(0, I_k)$
- *W is a **positive** scalar random variable which is independent of Z*
- *$A \in \mathbb{R}^{d \times k}$ and $\mu \in \mathbb{R}^d$ are a matrix and a vector of constants*

Remark 2.2.22 (conditional distribution)**.**

Example 2.2.5 (special case: multivariate t distribution). The t_n distribution with n degrees of freedom is a continuous Gaussian mixture with

$$\mu = 0, U = \frac{n}{W},$$

where $W \sim \chi^2(n)$.

2.3 Characterizing distributions

2.3.1 Skewness and kurtosis

Definition 2.3.1 (skewness). *The skewness of an univariate population for random variable X is defined by*

$$\gamma_1 = E[(\frac{X-\mu}{\sigma})^3] = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{\mu_3}{\mu_2^{3/2}}$$

where μ_2 and μ_3 are the second and the third ***central moments****.*

Remark 2.3.1 (interpretation).

- Intuitively, the skewness is a measure of symmetry.
- **Negative skewness** indicates that the mean of the data values is less than the median, and the data distribution is **left-skewed**.
- **Positive skewness** indicates that the mean of the data values is greater than the median, and the data distribution is **right-skewed**.

Example 2.3.1. Let $X \sim N(\mu, \sigma^2)$. Then the skewness of X distribution is $\gamma_1 = E[(\frac{X-\mu}{\sigma})^3] = E[Z^3] = 0, Z \sim N(0,1)$, where we use the fact the third moment for a standard normal is zero(Lemma 2.2.4).

Definition 2.3.2 (kurtosis, excess kurtosis).

- *The* ***kurtosis*** *of a univariate population is defined by*

$$\gamma_2 = E[(\frac{X-\mu}{\sigma})^4] = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} = \frac{\mu_4}{\mu_2^2},$$

  where μ_2 and μ_4 are the second and the fourth central moments.
- *The* ***excess kurtosis*** *of a univariate population is defined by*

$$\gamma_2^{ex} = \gamma_2 - 3.$$

Remark 2.3.2 (interpretation).

- Intuitively, the kurtosis is a measure of tail shape of a distribution.

Remark 2.3.3 (Three types of kurtosis).

- **mesokurtic**: zero kurtosis as standard normal distribution.

- **leptokurtic**: kurtosis greater than 0. This type of distribution is one with extremely thick tails and a very thin and tall peak. t-distributions are leptokurtic.
- **platykurtic**:kurtosis smaller than 0. This type of distribution has a short and broad-looking peak. Uniform distributions are platykurtic.

Example 2.3.2. Let $X \sim N(\mu, \sigma^2)$. Then the kurtosis of X distribution is $\gamma_2 = E[(\frac{X-\mu}{\sigma})^4] = E[Z^4] = 3, Z \sim N(0,1)$, where we use the fact the fourth moment for a standard normal is 3(Lemma 2.2.4).

2.3.2 Quantiles and percentiles

2.3.2.1 *Basics*

Definition 2.3.3 (percentile of a distribution). *The α **percentile**($\alpha \in [0,1]$) of a probability distribution of random number X is a number p in the support D of the support such that*

$$Pr(x < p) = \alpha, Pr(x > p) = 1 - \alpha.$$

Or equivalently, the α percentile is given by

$$p = F_X^{-1}(\alpha).$$

Definition 2.3.4 (percentile in a set of sample values). *The α **percentile**($\alpha \in [0,1]$) of a set of values is a value in $\mathbb{R}$ that divides them so that $100\alpha\%$ of values lie below and $100(1-\alpha)\%$ of the values lie above.*

Definition 2.3.5 (quantiles of a distribution). *Quantiles are the cutpoints dividing the range of a probability distribution into contiguous intervals with equal probabilities.*

Lemma 2.3.1 (linear relationship between percentiles from two distributions). *Let X and Y be two random variables with cdf F_X and F_Y. Let $p_X = F_X^{-1}(\alpha)$ and $p_Y = F_Y^{-1}(\alpha)$ for $\alpha \in [0,1]$. It follows that*

- *If $Y = aX + b$, then*
$$p_Y = ap_X + b$$
- *If $Y = \alpha X^\beta$, then*
$$p_{\ln Y} = \beta p_{\ln X} + \ln \alpha,$$

where $p_{\ln Y} = F^{-1}_{\ln Y}(\alpha), p_{\ln X} = F^{-1}_{\ln X}(\alpha)$,

Proof. (1) We know that
$$\alpha = F_Y(p_Y) = F_X(p_X).$$
From scale-location transformation(Lemma 1.4.8), we have
$$p_X = (p_Y - b)/a.$$
(2) From $Y = \alpha X^{\beta}$, we have $\ln Y = \beta \ln X + \ln \alpha$. □

Remark 2.3.4 (QQplot and applications)**.**

- When we plotting the percentiles from two samples together, an approximate linear relation suggests $Y = aX + b$.
- When we plotting the percentiles from two log-value samples together, an approximate linear relation suggests a power relationship $Y = \alpha X^{\beta}$.

2.3.2.2 *Cornish-Fisher expansion*

Theorem 2.3.1 (Cornish-Fisher expansion). *Consider a distribution with mean μ and variance σ^2. Then its α quantile can be approximate by*
$$\mu + \sigma z_{\alpha}^{cf}$$
where
$$z_{\alpha}^{cf} = q_{\alpha} + \frac{(q_{\alpha}^2 - 1)S(X)}{6} + \frac{(q_{\alpha}^3 - 3q_{\alpha})K(X)}{24} - \frac{(2q_{\alpha}^3 - 5q_{\alpha})S^2(X)}{36},$$
where $S(X)$ is skewness, $K(X)$ is kurtosis, z_{α}^{cf} is the Cornish-Fisher approximate quantile value for the confidence level α, and q_{α} is the quantile value for the standard normal distribution with confidence level α.

Remark 2.3.5 (motivation)**.** The Cornish-Fisher expansion enables us to approximate quantiles of a random variable based only on its skewness and cumulants.

2.3.3 Exponential families

Definition 2.3.6. *[6, p. 111] A family of pdfs or pmfs is called an exponential family if it can be expressed as*

$$f(x|\theta) = h(x)c(\theta)\exp(\sum_{i=1}^{k} w_i(\theta)t_i(x)), \theta \in \mathbb{R}^n$$

where $h(x) \geq 0$ *and* $t_1(x), ..., t_k(x)$ *are real-valued functions of the observations, the* $c(\theta) \geq 0$ *and* $w_1(\theta), ..., w_k(\theta)$ *are real-valued functions of the vector* θ.

Theorem 2.3.2 (mean and variance of exponential family). *[6, p. 112] The mean and variance of the exponential family is*

$$E[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)] = -\frac{\partial \log(c(\theta))}{\partial \theta_j} \quad (3)$$

$$Var[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)] = -\frac{\partial^2 \log(c(\theta))}{\partial \theta_j^2} - E[\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)] \quad (4)$$

Proof: see [6, p. 132] See [6, p. 622] for a complete review.

Definition 2.3.7. *A class of distributions is in the exponential family if it can be written in the form*

$$p(y; \eta) = b(\eta)\exp(\eta^T T(y) - a(\eta))$$

where η *is called the natural parameter of the distribution;* $T(y)$ *is the sufficient statistic; and* $a(\eta)$ *is the log partition function.*

2.4 Notes on bibliography

For decision theory, [8][9] For overall graduate level treatment, see [6][1]. For likelihood based methods, see [10] For large sample theory(asymptote analysis), see [11].

For introductory level Bayesian statistics, see [12].

For good treatment on statistical estimation theory, see [13].

For linear regression models, see [14][15].

For multivariate statistical analysis, see [16].

For mixed models, see [17].

For an informal but deep treatment on robust statistics, see [18].

For an extensive discussion on statistical distribution, see [19][20].

BIBLIOGRAPHY

1. Hogg, R. V., McKean, J. & Craig, A. T. Introduction to Mathematical Statistics, 7 ed (2012).
2. Chirikjian, G. S. *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods* (Springer Science & Business Media, 2011).
3. Borovkova, S., Permana, F. J. & Weide, H. V. A closed form approach to the valuation and hedging of basket and spread options. *Journal of Derivatives* **14,** 8 (2007).
4. Brzezniak, Z. & Zastawniak, T. *Basic stochastic processes: a course through exercises* (Springer Science & Business Media, 1999).
5. Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT press, 2012).
6. Casella, G. & Berger, R. L. *Statistical inference* (Duxbury Pacific Grove, CA, 2002).
7. Ruppert, D. *Statistics and data analysis for financial engineering, 2ed* (Springer, 2015).
8. Young, G. A. & Smith, R. L. *Essentials of statistical inference* (Cambridge University Press, 2005).
9. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
10. Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood* (Oxford University Press, 2001).
11. Lehmann, E. L. *Elements of large-sample theory* (Springer Science & Business Media, 1999).
12. Hoff, P. D. *A first course in Bayesian statistical methods* (Springer Science & Business Media, 2009).
13. Kay, S. M. Fundamentals of statistical signal processing, volume I: estimation theory (1993).
14. Kutner, M., Nachtsheim, C. & Neter, J. *Applied Linear Regression Models* ISBN: 9780072955675. https://books.google.com/books?id=0nAMAAAACAAJ (McGraw-Hill Higher Education, 2003).
15. Seber, G. A. & Lee, A. J. *Linear regression analysis* (John Wiley & Sons, 2012).
16. Johnson, R. & Wichern, D. *Applied Multivariate Statistical Analysis* ISBN: 9780131877153. https://books.google.com/books?id=gFWcQgAACAAJ (Pearson Prentice Hall, 2007).

17. McCulloch, C. E. & Neuhaus, J. M. *Generalized linear mixed models* (Wiley Online Library, 2001).
18. Wilcox, R. R. *Fundamentals of modern statistical methods: Substantially improving power and accuracy* (Springer Science & Business Media, 2010).
19. Forbes, C., Evans, M., Hastings, N. & Peacock, B. *Statistical distributions* (John Wiley & Sons, 2011).
20. Krishnamoorthy, K. *Handbook of statistical distributions with applications* (CRC Press, 2016).

3

STATISTICAL ESTIMATION THEORY

3.1 Estimator theory

3.1.1 Statistic

Definition 3.1.1 (statistic). *Let $X_1, X_2, ..X_n$ denote a random sample on a random variable X. Let $T = T(X_1, X_2, ..., X_n)$ be a function of the sample. Then T is called a statistic.*

Remark 3.1.1. T is also a random variable.

Definition 3.1.2 (common statistic). *Given a random sample $X_1, ..., X_n$ from X, we have following definitions:*

- *Sample mean:*
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
- *Sample variance:*
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$
- *Sample standard deviation:*
$$S = \sqrt{S^2}.$$

Remark 3.1.2 (another equivalent form of sample variance). Note that $\sum_{i=1}^{n}(X_i - \overline{X})^2$ can also be written by
$$\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$
.

We have

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \sum_{i=1}^{n}(X_i^2-2\overline{X}X_i+\overline{X}^2)\\ &= \sum_{i=1}^{n}X_i^2-2\overline{X}n\overline{X}+n\overline{X}\\ &= \sum_{i=1}^{n}X_i^2-n\overline{X}\overline{X}.\end{aligned}$$

We have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2 &= \sum_{i=1}^{n}2nX_i^2-\sum_{i=1}^{n}\sum_{j=1}^{n}2X_iX_j\\ &= \sum_{i=1}^{n}2nX_i^2-\sum_{i=1}^{n}2X_in\overline{X}\\ &= \sum_{i=1}^{n}2nX_i^2-2n^2\overline{X}^2\\ &= 2n(\sum_{i=1}^{n}X_i^2-n\overline{X}\overline{X})\end{aligned}$$

3.1.2 Different types of estimators

3.1.2.1 *Basic concepts*

Definition 3.1.3 (unbiased Estimator). *Let $X_1, X_2, ..X_n$ denote a random sample on a random variable X with pdf $f(x;\theta), \theta \in \Omega$. Let $T = T(X_1, X_2, ..., X_n)$ be a statistic. We say that T is an unbiased estimator of θ if $E(T) = \theta$.*

Definition 3.1.4 (consistent Estimator). *Let X be a random variable with cdf $F(x,\theta)$. Let $X_1, X_2, ..., X_n$ be a sample from the distribution of X, and let T_n denote a statistic. T_n is a consistent estimator of θ if*

$$T_n \xrightarrow{P} \theta$$

Definition 3.1.5 (bias of an estimator). *The bias of an estimator $\hat{\theta}$ is*

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta,$$

where θ is the true value. If $Bias(\hat{\theta}) = 0$, then estimator $\hat{\theta}$ is said to be unbiased.

Definition 3.1.6 (variance of an estimator). *The variance of an* ***unbiased*** *estimator $\hat{\theta}$ is*

$$Var(\hat{\theta}) = E[\hat{\theta}] - \theta)^2],$$

and the covariance is

$$Var(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T],$$

where θ is the true value.

Definition 3.1.7 (mean squared error of an estimator). *The mean squared error(MSE) of an estimator is*

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

3.1.2.2 *Variance-bias decomposition*

Theorem 3.1.1 (variance bias decomposition). *The MSE of an estimator is related to its variance and bias via*

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + (Bias(\hat{\theta}))^2 \tag{5}$$

where $Var[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$. Particularly, if the estimator is unbiased (i.e. $Bias(\hat{\theta}) = 0$), we have

$$MSE(\hat{\theta}) = Var[\hat{\theta}]$$

Proof. Make $(\hat{\theta} - \theta) = (\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)$ and note that $\hat{\theta}$ is a random variable. Specifically,

$$\begin{aligned}
MSE(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\
&= E[((\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta))^2] \\
&= E[((\hat{\theta} - E[\hat{\theta}])^2] + 2E[((\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta))] + E[(E[\hat{\theta}] - \theta))^2] \\
&= Var[\hat{\theta}] + 2E[((\hat{\theta} - E[\hat{\theta}])](E[\hat{\theta}] - \theta)) + Bias[\hat{\theta}]^2 \\
&= Var[\hat{\theta}] + 0 + Bias[\hat{\theta}]^2
\end{aligned}$$

□

Remark 3.1.3 (biasedness can be useful). At first glance, it may seem that biasedness is always undesired. However, biased estimator might have smaller variance (Figure Figure 3.1.1). As a consequence, biased estimator can have smaller MSE than unbiased estimator. Also consider the following example.

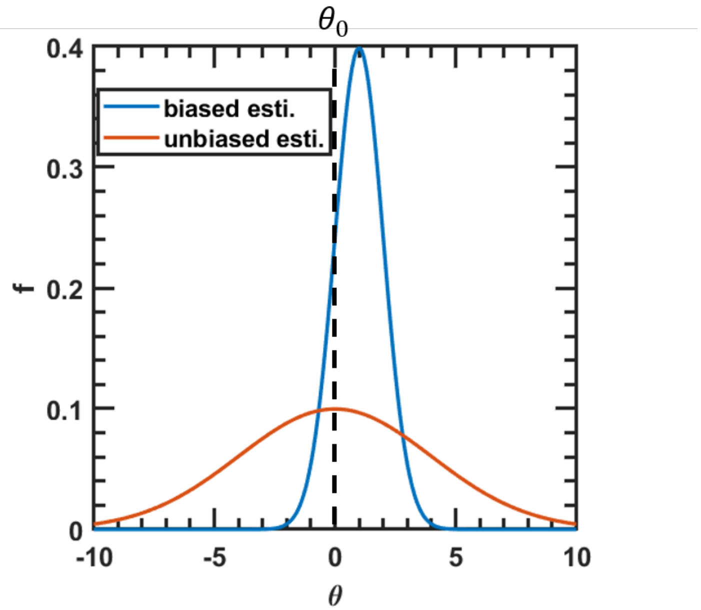


Figure 3.1.1: An example of biased estimator with smaller variance than unbiased estimator

Example 3.1.1. Consider a sample $X_1, X_2, ..., X_n$ of iid normal random variable with unknown mean and variance. Consider two variance estimator

$$S_1^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})}{n-1}, S_2^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})}{n}.$$

Then

- The MSE for S_1^2 is

$$\begin{aligned} MSE[S_1^2] &= Var[S_1^2] + [Bias]^2 \\ &= \frac{2\sigma^4}{n-1} + 0 = \frac{2\sigma^4}{n-1} \end{aligned}$$

  where we use the fact that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

  from Theorem 2.2.5 and

$$Var[\frac{(n-1)S^2}{\sigma^2}] = \frac{(n-1)^2}{\sigma^4} Var[S^2] = 2(n-1),$$

where use $Var[\chi^2(n)] = 2n$ in Lemma 2.2.35.

- The MSE for S_2^2 is

$$\begin{aligned} MSE[S_2^2] &= (Var[S_2^2] + [Bias])^2 \\ &= \frac{2(n-1)\sigma^4}{n} + (E[S_2^2] - \sigma^2)^2 \\ &= \frac{2(n-1)\sigma^4}{n} + (\frac{(n-1)\sigma^2}{n} - \sigma^2)^2 \\ &= \frac{2n-1}{n^2}\sigma^4 \end{aligned}$$

- $MSE[S_2^2] < MSE[S_1^2]$. That is, the maximum-likelihood estimator has smaller MSE than the unbiased estimator.

3.1.2.3 *Consistence*

Definition 3.1.8 (consistent estimator). *We say $\hat{\theta}$ is a consistent estimator of θ if $\hat{\theta}$ converges to θ in probability, i.e.,*

$$\lim_{n\to\infty} P(\left|\hat{\theta}(X_1, X_2, ..., X_n) - \theta\right| < \epsilon) = 1, \forall \epsilon > 0.$$

Theorem 3.1.2 (MSE criterion for consistent estimator). *An* ***unbiased*** *estimator $\hat{\theta}$ is consistent if*

$$\lim_{n\to\infty} Var(\hat{\theta}(X_1, X_2, ..., X_n)) = 0$$

More generally, an estimator $\hat{\theta}$ is consistent if

$$\lim_{n\to\infty} MSE(\hat{\theta}(X_1, X_2, ..., X_n)) = 0$$

Proof. Overall, we can use Theorem 1.11.3 (convergence in mean square implies convergence in probability)

□

Remark 3.1.4 (consistence vs. unbiasedness).

- A consistent estimator is at least **asymptotically unbiased**.However, some unbiased estimators can be inconsistent(i.e. the variance does not converge to 0).
- If the sample size is larger, consistent estimators are consider better than unbiased estimator because consistent estimator ensure that estimator variance is sufficiently smaller.

- Inconsistent estimator usually should be avoided, since increasing the number of samples will not necessarily reduce the variance.

Theorem 3.1.3 (sample mean estimator is consistent). *[1, p. 1160] Let $X_1, ..., X_n$ be a random sample from any population with finite mean μ and finite variance σ^2. Let $\overline{X}_n$ be the sample mean. It follows that*

- *$\overline{X}_n$ is the consistent estimator of μ.*
- *For any function $g(x)$, if $E[g(x)]$ and $Var[g(x)]$ are finite, then the quantity*

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

is the consistent estimator of $E[g(X)]$.

Proof. Because $E[\overline{X}_n] = \mu$ and $Var[\overline{X}_n] \to 0$, then $\overline{X}_n$ converges to μ in mean square and thus in probability (Theorem 1.11.2 and Theorem 1.11.3). $\square$

3.2 Method of moments

Definition 3.2.1 (moments of the random sample). *[2, p. 312] Let $X_1, X_2, ..., X_n$ be a sample from a population with pdf or pmf $f(x|\theta_1, ..., \theta_k)$. The first k moments are given by*

$$
\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
&\cdots \\
m_k &= \frac{1}{n}\sum_{i=1}^{n} X_i^k
\end{aligned}
$$

Definition 3.2.2 (moments of the random sample). *[2, p. 312] Let $X_1, X_2, ..., X_n$ be a random sample of X from a population with pdf or pmf $f(x|\theta_1, ..., \theta_k)$. Define $\mu_i = E[X^i], i = 1, 2, ..., k$. The method of moments is aimed at solving $\theta_1, \theta_2, ..., \theta_k$ from the k equations*

$$
\begin{aligned}
m_1 &= \mu_1(\theta_1, ..., \theta_k) \\
m_2 &= \mu_2(\theta_1, ..., \theta_k) \\
&\cdots \\
m_k &= \mu_k(\theta_1, ..., \theta_k)
\end{aligned}
$$

where $m_i, i = 1, 2, ..., k$ are the k moments of the sample.

Lemma 3.2.1 (estimating normal distribution parameter via method of moments). *[2, p. 312] Suppose $X_1, X_2, ..., X_n$ are iid random variable with $N(\mu, \sigma^2)$. It follows that*

- *$m_1 = \mu, \mu^2 + \sigma^2 = m_2$.*
- *The moment of method estimators for (μ, σ) are*

$$\hat{\mu} = m_1, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Proof. (1) note that $E[X^2] = Var[X] + E[X]^2$. (2) note that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

□

Lemma 3.2.2 (estimating t distribution parameter via method of moments). *Suppose $X_1, X_2, ..., X_n$ are iid random variable with $t_v(\mu, \sigma^2), v > 2$. It follows that*

- $m_1 = \mu, \mu^2 + \sigma^2 \frac{v}{v-2} = m_2$.
- *The moment of method estimators for (μ, σ) are*

$$\hat{\mu} = m_1, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \frac{v-2}{v}.$$

Proof. (1)(2)From subsection 2.2.18 note that $E[X] = \mu$, $E[X^2] = Var[X] + E[X]^2 = \sigma^2 \frac{v}{v-2} + \mu^2$. note that

$$\hat{\sigma}^2 \frac{v}{v-2} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

$\square$

3.3 Maximum likelihood estimation

3.3.1 Basic concepts

Definition 3.3.1 (estimator). *[2, p. 315] A point estimator is any function $W(\boldsymbol{X}) = W(X_1, X_2, ..., X_n)$ of a random sample; that is, any statistic is a point estimator.*

Definition 3.3.2 (likelihood function). *[3, p. 22] Assuming a statistical model parametrized by a fixed and unknown θ, the likelihood $L(\boldsymbol{x}|\theta)$ is the probability of the observations $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ of iid random samples $X_1, X_2, ..., X_n$ as a function of θ. It can be written as*

$$L(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(X = x_i|\theta)$$

And the corresponding log-likelihood function is defined as

$$\log L(\boldsymbol{x}|\theta) = \sum_{i=1}^{n} f(X = x_i|\theta)$$

Definition 3.3.3 (maximum likelihood estimator and score function). *[2, p. 316] A maximum likelihood estimator(MLE) of the parameter θ based on given observations $\boldsymbol{x}$ is*

$$\hat{\theta} = \max_{\theta} \log L(\boldsymbol{x}|\theta),$$

Or alternatively, $\hat{\theta}$ satisfies

$$s(\theta, \boldsymbol{x}) = \frac{\partial \log L}{\partial \theta} = 0,$$

*where $s(\theta, \boldsymbol{x})$ is called **score function**.*

Example 3.3.1 (Bernoulli trial MLE). Consider a series of independent Bernoulli trials with success probability θ such that we have probability mass function given by

$$Pr(Y_i = y) = (1-\theta)^{1-y}\theta^y, y \in \{0, 1\}.$$

- The log-likelihood function based on n observations $Y = \{Y_1, ..., Y_N\}$ can be written by

$$\log L(\theta; Y) = \sum_{i=1}^{n}((1-y_i)\log(1-\theta) + y_i \log\theta) = n((1-\overline{y})\log(1-\theta) + \overline{y}\log(\theta)),$$

where $\overline{y}$ is the sample mean.

- The MLE is given by

$$\hat{\theta} = \overline{y}.$$

Example 3.3.2 (Normal distribution MLE). The log-likelihood function for n iid observations $x_1, ..., x_n$ drawn from normal distribution is given by

$$\begin{aligned}
\log L(\theta_1, \theta_2) &= \prod_{i=1}^{n} f(x_i \theta_1, \theta_2) \\
&= \theta_2^{-n/2} (2\pi)^{-n/2} \exp(-\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2) \\
&= -\frac{n}{2} \log \theta_2 - \frac{n}{2} \log(2\pi) - \frac{\sum_{i=1}^{n} (x_i - \theta_1)^2}{2\theta_2}
\end{aligned}$$

where $\theta_1 = \mu, \theta_2 = \sigma^2$. Setting derivatives to zeros, we have

$$\begin{aligned}
\frac{\partial \log L(\theta_1, \theta_2)}{\partial \theta_1} &= \frac{\sum_{i=1}^{n} (x_i - \theta_1)}{\theta_2} = 0 \\
\frac{\partial \log L(\theta_1, \theta_2)}{\partial \theta_2} &= -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^{n} (x_i - \theta_1)^2}{\theta_2^2} = 0
\end{aligned}$$

which produces

$$\hat{\mu} = \hat{\theta}_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}, \hat{\sigma}^2 = \hat{\theta}_2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x}}{n}.$$

Example 3.3.3 (exponential distribution MLE). Consider an exponential distribution with parameter α such that its pdf is given by

$$f(x; \alpha) = \alpha e^{-\alpha x}, x \geq 0.$$

The MLE for α from an iid random sample $X_1, ..., X_n$ is given by $\hat{\alpha} = 1/\overline{X}$ since

$$\begin{aligned} \log L(\alpha) &= n \log \alpha - \alpha \sum_{i=1}^{n} X_i \\ \partial \log L(\alpha)/\partial\alpha &= \frac{n}{\alpha} - \sum_{i=1}^{n} X_i \\ \partial \log L(\alpha)/\partial\alpha = 0 &\implies \hat{\alpha} = 1/\overline{X}. \end{aligned}$$

3.4 Information and efficiency

3.4.1 Fish information

Assumption 3.1 (Fisher information regularity assumption). *For a pdf $f(x;\theta)$ of random variable X with parameter θ. We make the following regularity assumptions:*

- *The set $A = \{x|p(x;\theta) > 0\}$ does not depend on θ. For all $x \in A, \theta \in \Theta$, $\frac{\partial}{\partial\theta}\log p(x;\theta)$ exists and is finite. Here Θ is the parameter space.*
- *If T is any statistic of X such that $E\|T|] < \infty$ for all $\theta \in \Theta$, then integration and differentiation by θ can be interchanged in the following way:*

$$\frac{\partial}{\partial\theta}[\int T(x)f(x;\theta)dx] = \int T(x)\frac{\partial}{\partial\theta}f(x;\theta)dx,$$

whenever the right-hand side is finite.

Definition 3.4.1 (Fisher information). *For one dimensional parametric family of pdf or pmf $f(x;\theta)$, we define the Fisher information for $\theta \in \mathbb{R}$ as*

$$I(\theta) = E[(\frac{d}{d\theta}\log f(x;\theta))^2],$$

where the expectation is ***taken with respect to*** x. *In particular, if $\theta \in \mathbb{R}^N$, we have Fisher information matrix defined as*

$$I(\theta)_{ij} = -E[\frac{\partial^2 \log(f(x;\theta))}{\partial\theta_i\theta_j}].$$

Remark 3.4.1. This definition holds both for discrete or continuous random variables, as long as f is differentiable respect to θ.

Theorem 3.4.1 (Basic properties of Fisher information). *Let $f(x;\theta)$ be a pdf parameterized by $\theta \in \mathbb{R}$ with Assumption 3.1 holds, then*

- $$E[\frac{d}{d\theta}\log f(x;\theta)] = 0$$
- $$I(\theta) = Var[\frac{d}{d\theta}\log f(x;\theta)].$$

- *Further assume $f(x;\theta)$ is twice differentiable and interchange between integration and differentiation is permitted. Then*

$$I(\theta) = E[(\frac{d}{d\theta}\log f(x;\theta))^2] = -E[\frac{\partial^2 \log(f(x;\theta))}{\partial\theta^2}].$$

- *For $\theta \in \mathbb{R}^N$,*

$$I(\theta) = E[\frac{\partial \log(f(x;\theta))}{\partial\theta}(\frac{\partial \log(f(x;\theta))}{\partial\theta})^T] = -E[\frac{\partial^2 \log(f(x;\theta))}{\partial\theta\theta^T}].$$

Proof. (1) The equivalence of these two expressions can be showed as:

$$\begin{aligned} E[\frac{d}{d\theta}\log f(x;\theta)] &= \int \frac{1}{f(x;\theta)}\frac{d}{d\theta}f(x;\theta)f(x;\theta)dx \\ &= \int \frac{d}{d\theta}f(x;\theta)dx \\ &= \frac{d}{d\theta}\int f(x;\theta)dx \\ &= \frac{d}{d\theta}1 \\ &= 0 \end{aligned}$$

(2) Based on definition, we have

$$\begin{aligned} Var[\frac{d}{d\theta}\log f(x;\theta)] &= E[(\frac{d}{d\theta}\log f(x;\theta))^2] - E[\frac{d}{d\theta}\log f(x;\theta)]^2 \\ &= E[(\frac{d}{d\theta}\log f(x;\theta))^2] - 0 \\ &= I(\theta) \end{aligned}$$

(3)

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}\log f(x;\theta) &= \frac{\partial}{\partial\theta}\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta) \\ &= -\frac{\partial}{\partial\theta}\frac{1}{f(x;\theta)^2}\frac{\partial}{\partial\theta}f(x;\theta) + \frac{1}{f(x;\theta)}\frac{\partial^2}{\partial\theta^2}f(x;\theta) \\ &= -(\frac{\partial}{\partial\theta}\log f(x;\theta))^2 + \frac{1}{f(x;\theta)}\frac{\partial^2}{\partial\theta^2}f(x;\theta) \end{aligned}$$

Take expectation with respect to x on both sides and note that

$$E[\frac{1}{f(x;\theta)}\frac{\partial^2}{\partial\theta^2}f(x;\theta)] = \int \frac{\partial^2}{\partial\theta^2}f(x;\theta)dx = \frac{\partial^2}{\partial\theta^2}\int f(x;\theta)dx = \frac{\partial^2}{\partial^2\theta}1 = 0$$

□

3.4.2 Information matrix for common distributions

3.4.2.1 *Bernoulli distribution*

Lemma 3.4.1 (Fisher information for Bernoulli distribution). *Let the pmf of Bernoulli distribution parameterized by $f(x;\theta) = \theta^x(1-\theta)^{1-x}, x \in \{0,1\}$. Then*

$$I(\theta) = \frac{1}{\theta(1-\theta)}.$$

Proof.

$$\begin{aligned} I(\theta) &= -E[\frac{\partial^2 \log(f(x;\theta))}{\partial\theta^2}] \\ &= E[\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}] \\ &= \theta(1/\theta^2) + (1-\theta)(1/(1-\theta)^2) \\ &= \frac{1}{\theta(1-\theta)} \end{aligned}$$

□

3.4.2.2 *Normal distribution*

Lemma 3.4.2 (Fish information matrix for univariate normal distribution). *[1, p. 548] Let the pdf of normal distribution parameterized by*

$$f(x;\theta) = (2\pi\theta_2)^{-1/2}\exp(-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x-\theta_1)^2).$$

$$\begin{aligned} \frac{\partial^2 \log f}{\partial\theta_1^2} &= -\frac{1}{\theta_2} = -\frac{1}{\sigma^2} \\ \frac{\partial^2 \log f}{\partial\theta_2^2} &= \frac{1}{2\theta_2^2} - \frac{1}{\theta_2^3}(x-\theta_1)^2 \\ \frac{\partial^2 \ln f}{\partial\theta_1\partial\theta_2} &= -\frac{1}{\theta_2^2}(x_i-\theta_1) \end{aligned}$$

Finally, take expectation with respect to x and we have

$$I(\theta_1, \theta_2) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{bmatrix}$$

3.4.2.3 *Multivariate Gaussian distribution*

Lemma 3.4.3.

3.4.3 Cramer-Rao lower bound

3.4.3.1 *Information inequality*

Theorem 3.4.2 (information inequality for statistic). *Let $T(X)$ be any statistic such that $Var[T(X)] < \infty$ for all θ. Denote $E[T(X)]$ by $\phi(\theta)$. Suppose Assumption 3.1 holds and $0 < I(\theta) < \infty$. Then **for all** θ*

$$Var[T(X)] \geq \frac{[\phi'(\theta)]^2}{I(\theta)}.$$

Proof. Based on the Assumption 3.1, ,we have

$$\phi'(\theta) = \int T(x) \frac{\partial}{\partial \theta} f(x;\theta) dx = \int T(x) \frac{\partial \log f(x;\theta)}{\partial \theta} f(x;\theta) dx.$$

Therefore, we can view

$$\phi'(\theta) = E[T(X) \frac{\partial \log f(x;\theta)}{\partial \theta}] = Cov[T(X), \frac{\partial \log f(x;\theta)}{\partial \theta}],$$

since $E[\frac{\partial \log f(x;\theta)}{\partial \theta}] = 0 \implies E[T(X)]E[\frac{\partial \log f(x;\theta)}{\partial \theta}] = 0$.

Using Cauchy-Schwartz inequality (Theorem 1.10.4), we have

$$|\phi'(\theta)|^2 = Cov[T(X), \frac{\partial \log f(x;\theta)}{\partial \theta}]^2 \leq Var[T(X)] Var[\frac{\partial \log f(x;\theta)}{\partial \theta}].$$

At last, use the fact (Theorem 3.4.1) that $I(\theta) = Var[\frac{\partial \log f(x;\theta)}{\partial \theta}]$, we can get the final result. □

Corollary 3.4.2.1 (information lower bound for general estimators). *Let $T(X)$ be a (generally biased) estimator of θ such that*

$$\phi(\theta) \triangleq E[T(X)] = \theta + \underbrace{b(\theta)}_{bias}.$$

Then

- *the variance of $T(X)$ is*
$$Var[T(X)] \geq \frac{\left|1 + b'(\theta)\right|}{I(\theta)}.$$
- *the MSE of $T(X)$ is*
$$MSE[T(X)] \geq \frac{\left|1 + b'(\theta)\right|}{I(\theta)} + b(\theta)^2.$$

3.4.3.2 *Cramer-Rao lower bound: univariate case*

Theorem 3.4.3 (Cramer-Rao lower bound in univariate estimation). *Let $\hat{\theta}$ be an arbitrary univariate estimator as a function of iid random samples $X_1, ..., X_n$, whose distribution is parameterized by single parameter θ. Let θ_0 be the true value. Then the variance of the estimator $\hat{\theta}$ is bounded by*

$$Var(\hat{\theta}) \geq \frac{(\frac{d}{d\theta}E[\hat{\theta}])^2}{nI_1(\theta^0)},$$

where $I_1(\theta)$ is the Fish information associated with distribution $f(x;\theta)$ and the expectation is taken with respect to x. Particularly, if the estimator $\hat{\theta}$ is unbiased (that is $E[\hat{\theta}] = \theta$), we have

$$Var(\hat{\theta}) \geq \frac{1}{nI_1(\theta^0)}.$$

Proof. Note that the Fisher information $I(\theta)$ associated with the joint distribution of $(X_1, ..., X_n)$ can be expressed by $I(\theta) = nI_1(\theta)$, where $I_1(\theta)$ is the Fisher information associated with $f(x;\theta)$. This is because under iid assumption,

$$E[\log f(x_1, ..., x_n; \theta)] = nE[\log f(x;\theta)].$$

Then use the information inequality (Theorem 3.4.2), we have

$$Var(\hat{\theta}) \geq \frac{(\frac{d}{d\theta}E[\hat{\theta}])^2}{nI_1(\theta)}.$$

□

Example 3.4.1 (univariate estimation for normal distributions).

- Consider an unbiased mean estimator $\hat{\mu}$ and an unbiased variance estimator $\hat{\sigma}^2$ for normal distribution with unknown mean μ and variance σ^2. Because the information matrix is given by (Lemma 3.4.2)
  The mean estimator has bounded variance given by (using information matrix from
  $$Var[\hat{\mu}] \geq \frac{1}{nI_1(\theta)} = \sigma^2/n.$$
- Consider anfor normal distribution with known mean μ. The mean estimator has bounded variance given by
  $$Var[\hat{\sigma}^2] \geq \frac{1}{nI_1(\theta)} = 2\sigma^4/n.$$
- It is clear that
  - Increasing sample size n will reduce the estimator variance.
  - Mean/variance estimators of random samples drawn from small-variance distributions have inherent smaller variances.

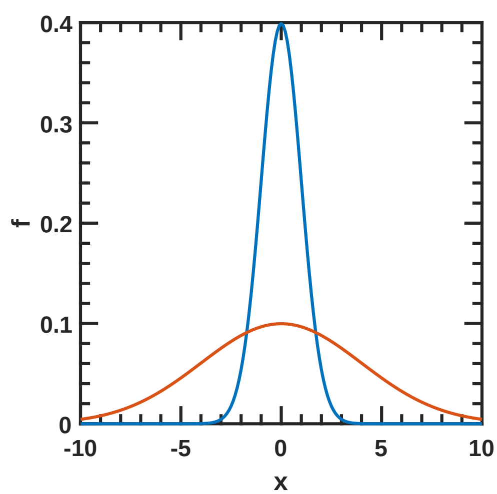


Figure 3.4.1

Theorem 3.4.4 (sufficient and necessary condition to achieve Cramer-Rao lower bound). *Under the regularity Assumption 3.1, there exists an unbiased estimator $\hat{\theta}$ of $\phi(\theta)$ whose variance attains Cramer-Rao lower bound (i.e., there exists a most efficent unbiased estimator $\hat{\theta}$ of $\phi(\theta)$) if and only if the score statistic $S(X)$ can be expressed in the form*

$$S(X) = \alpha(\theta)[\hat{\phi(X)} - \phi(\theta)],$$

where

$$\alpha(\theta) = \frac{I(\theta)}{\phi'(\theta)};$$

or equivalently if and only if the function

$$S(X)\frac{\phi'(\theta)}{I(\theta)} + \phi(\theta)$$

is independent of θ and is only dependent on X.

3.4.3.3 *Cramer-Rao lower boundïijŽ multivariate case*

Theorem 3.4.5 (information inequality for statistic: multivariate case). *Let $T(X)$ be any statistic such that $Var[T(X)] < \infty$ for all θ. Denote $E[T(X)]$ by $\phi(\theta)$. Suppose Assumption 3.1 holds and $0 < I(\theta) < \infty$. Then **for all** θ*

$$Var[T(X)] \geq [\nabla_\theta \phi]^T [I(\theta)]^{-1} [\nabla_\theta \phi].$$

Proof. Based on the Assumption 3.1, ,we have

$$\phi'(\theta) = \int T(x)\frac{\partial}{\partial\theta} f(x;\theta)dx = \int T(x)\frac{\partial \log f(x;\theta)}{\partial\theta}dx.$$

Therefore, we can view

$$\phi'(\theta) = E[T(X)\frac{\partial \log f(x;\theta)}{\partial\theta}] = Cov[T(X), \frac{\partial \log f(x;\theta)}{\partial\theta}],$$

since $E[\frac{\partial \log f(x;\theta)}{\partial\theta}] = 0 \implies E[T(X)]E[\frac{\partial \log f(x;\theta)}{\partial\theta}] = 0.$

Using Cauchy-Schwartz inequality (Theorem 1.10.4), we have

$$\left|\phi'(\theta)\right|^2 = Cov[T(X), \frac{\partial \log f(x;\theta)}{\partial\theta}]^2 \leq Var[T(X)]Var[\frac{\partial \log f(x;\theta)}{\partial\theta}].$$

At last, use the fact (Theorem 3.4.1) that $I(\theta) = Var[\frac{\partial \log f(x;\theta)}{\partial\theta}]$, we can get the final result. □

Proof. Similar to Theorem 3.4.5, we can show that

$$\frac{\partial\phi(\theta)}{\partial\theta_j} = Cov(T(X), \frac{\partial \log f(x;\theta)}{\partial\theta_j}).$$

For constants $c_1, c_2, ..., c_p$, note that

$$Var[T(X) - \sum_{j=1}^{p} c_j \frac{\partial \log f(x;\theta)}{\partial \theta_j}] = Var[T(X)] + c^T I(\theta) c - 2c^T[\nabla_\theta \phi]$$
$$\geq 0$$

Particularly, the minimum is achieved at $c* = [I(\theta)]^{-1}\delta$. Then

$$Var[T(X) - \sum_{j=1}^{p} c_j \frac{\partial \log f(x;\theta)}{\partial \theta_j}] = Var[T(X)] - [\nabla_\theta \phi]^T [I(\theta)]^{-1} [\nabla_\theta \phi] \geq 0.$$

□

Theorem 3.4.6 (Cramer-Rao lower bound in multivariate estimation). *Let $\hat{\theta}$ be a p-dimension* ***unbiased*** *estimator as a function of iid random samples $X_1, ..., X_n$, whose distribution is parameterized by parameter vector $\theta \in \mathbb{R}^p, \theta = (\theta_1, ..., \theta_p)$. Let θ^0 be the true value.*

Then the variance matrix of the estimator $\hat{\theta}_i$ is bounded by

$$Var(\hat{\theta}) \geq [n[I_1(\theta)]^{-1},$$

where $I_1(\theta^0)$ is the Fish information maxtrix associated with distribution $f(x;\theta)$ and the expectation is taken with respect to x.

Proof. Note that the Fisher information $I(\theta)$ associated with the joint distribution of $(X_1, ..., X_n)$ can be expressed by $I(\theta) = nI_1(\theta)$, where $I_1(\theta)$ is the Fisher information associated with $f(x, \theta)$. This is because under iid assumption,

$$E[\log f(x_1, ..., x_n; \theta)] = nE[\log f(x;\theta)].$$

Then use the information inequality (Theorem 3.4.5), we have

$$Var(\alpha^T\hat{\theta}) = \alpha^T Var[\hat{\theta}]\alpha \geq [\nabla_\theta \alpha^T \hat{\theta}]^T [nI_1(\theta)]^{-1} [\nabla_\theta \alpha^T \hat{\theta}] = \alpha^T [nI_1(\theta)]^{-1} \alpha,$$

where $\alpha \in \mathbb{R}^p$ is an arbitrary vector. □

Example 3.4.2 (multivariate estimation for normal distributions).

- Consider an unbiased mean estimator $\hat{\mu}$ for normal distribution with known variance σ^2. The information matrix is given by (Lemma 3.4.2)

$$I(\theta_1, \theta_2) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{bmatrix}.$$

Therefore,

$$Var[\hat{\mu}] \geq \sigma^2/n, Var[\hat{\sigma}^2] \geq 2\sigma^4/n,$$

- It is clear that
  - Increasing sample size n will reduce the estimator variance.
  - Mean/variance estimators of random samples drawn from small-variance distributions have inherent smaller variances.

3.4.3.4 *Uniformly minimum-variance unbiased estimator*

Definition 3.4.2. *Consider a statistic θ as a function of iid random samples $X_1, X_2, ..., X_n$ with pdf $f(x, \theta)$. The estimator is a* ***uniformly minimum-variance unbiased estimator(UMVUE)*** *if*

- *it is unbiased, i.e., $\hat{\theta} = \theta$*
- *$\forall \theta \in \Theta$, we have*
$$var(\hat{\theta}) \leq var(\hat{\theta}')$$
for any other unbiased estimator $\hat{\theta}'$.

Remark 3.4.2 (interpretation). In terms of efficiency in using data to reduce uncertainty, UMVUE has the optimal estimation efficiency.

How to find UMVUE?

There is no simple, general procedure for finding the MVUE estimator. Here are some several approaches:

- Find a sufficient statistic and apply the Rao-Blackwell theorem.
- Determine the so-called Cramer-Rao Lower Bound (CRLB) and verify that the estimator achieves it.
- Further restrict the estimator to a class of estimators (e.g., linear or polynomial functions of the data)
- The existence of UMVUE is in discussed [4, p. 62].

Lemma 3.4.4 (normal distribution estimators). *For a normal distribution with unknown mean and variance, the sample mean and the unbiased sample variance are the MUVEs for the population mean and population variance.*

Proof. TODO

□

Definition 3.4.3 (efficient estimator). *An estimator $\hat{\theta}$ if variance achieves equality in the Cramer Rao lower bound for all $\theta \in \Theta$.*

Remark 3.4.3. An efficient estimator is optimal in the sense of using information to reduce uncertainty.

Example 3.4.3. Let the pmf of Bernoulli distribution parameterized by $f(x;\theta) = \theta^x(1-\theta)^{1-x}, x \in \{0,1\}$. Then

$$\begin{aligned} I(\theta) &= -E[\frac{\partial^2 \log(f(x;\theta))}{\partial \theta^2}] \\ &= E[\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}] \\ &= \theta(1/\theta^2) + (1-\theta)(1/(1-\theta)^2) \\ &= \frac{1}{\theta(1-\theta)} \end{aligned}$$

Consider the estimator $\hat{\theta} = \overline{X}$, then

$$E[\hat{\theta}] = E[\overline{X}] = E[\sum_{i=1}^n X_i/n] = \theta$$

and

$$Var[\hat{\theta}] = E[\overline{X}^2] - E[\hat{\theta}]^2 = \theta/n + \theta^2(n-1)/n - \theta^2 = \theta(1-\theta)/n$$

Therefore, the variance of the estimator is achieving the lower bound.

Lemma 3.4.5. *An efficient estimator is UMVUE; that is, an unbiased estimator with variance achieving the Cramer-Rao lower bound is UMVUE. The converse might not be true.*

Proof. If the variance achieve the lower bound, then it has unformly minimum variance, and therefore it is UMVUE. On the other hand, it is likely that for some estimator the lower bounded is not achieved, but it has smaller variance than any other competing estimators. □

Remark 3.4.4 (implications and practical issues)**.**

- The Cramer-Rao lower Bound enables us to judge whether one estimator is efficient: the closer to lower bound, the more efficient.
- Efficient estimators are usually difficult to find in practice.

Theorem 3.4.7 (Rao-Blackwell theorem). *Let $\hat{\theta}(X_1, ..., X_n)$ be an estimator, and T be any sufficient statistic. Then the new estimator $\hat{\theta}^* = E[\hat{\theta}(X_1, ..., X_n)|T(X_1, ..., X_n)]$ will satisfy*

$$MSE(\hat{\theta}^*) \leq MSE(\hat{\theta}), \forall \theta.$$

Particularly, if both estimators are unbiased, we have

$$Var(\hat{\theta}^*) \leq Var(\hat{\theta}), \forall \theta.$$

The process of creating a new improved estimator via conditioning is called ***Rao-Blackwellization****.*

Proof. Since $Var(X) \geq 0$ implies $E[X^2] \geq E[X]^2$, we have

$$E[(\hat{\theta} - \theta)^2|T] \geq (E[\theta \hat{-} \theta|T])^2 = (E[\hat{\theta}|T] - \theta)^2 = (\hat{\theta}^* - \theta)^2$$

Take expectation on both sides and use the Tower property of conditional expectation, we have

$$E[(\hat{\theta}^* - \theta)^2] \leq E[(\hat{\theta} - \theta)^2]$$

□

Lemma 3.4.6 (unbiasedness inheritance). *The improved estimator is unbiased if and only if the original estimator is unbiased.*

Proof.

$$E[E[\hat{\theta}(X_1, ..., X_n)|T(X_1, ..., X_n)]] = E[\hat{\theta}(X_1, ..., X_n)]$$

□

Lemma 3.4.7 (idempotent operation). *Rao-Blackwellization is an idempotent operation. More precisely, let Rao-Blackwellization operator denoted as R, let $\hat{\theta}$ be the original estimator, then*

$$R^2[\hat{\theta}] = R[\hat{\theta}]$$

Proof.

$$E[E[\hat{\theta}|T]|T] = E[\hat{\theta}|T]$$

via the law of iterated expectation. □

Remark 3.4.5 (Rao-Blackwell theorem vs. Cramer-Rao lower bound).

- An estimator achieving the Cramer-Rao lower bound is performing better than any other estimators in all the possible θ.

- Rao-Blackwellization provides a way to improve current estimator such that the improved estimator is no worst than the current estimator in all the possible θ. The improved estimator cannot guarantee to achieve the Cramer-Rao lower bound, i.e., being better than any other estimators.

3.4.4 Efficiency

Definition 3.4.4 (relative efficiency). *The relative efficiency of two **unbiased** estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ is the ratio of their variance*

$$\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$$

.

Remark 3.4.6. The more efficient estimator is that: **given fixed number of samples**, it has a lower MSE/variance.

3.4.5 Fisher information characterization of MLE

3.4.5.1 *Properties of score function*

Lemma 3.4.8.

$$Es(\theta, \boldsymbol{X}) = 0$$

Proof. This is a restatement of Theorem 3.4.1 □

Lemma 3.4.9 (useful properties of score function). *[5, p. 550] Let t be any vector-valued function $\boldsymbol{X}$ and θ, then*

$$E[s(\theta, \boldsymbol{X})t^T(\theta, \boldsymbol{X})] = \frac{\partial}{\partial\theta}E[t^T] - E[\frac{\partial}{\partial\theta}t^T]$$

And particularly, if $t = s$, we have

$$E[s(\theta, \boldsymbol{X})s^T(\theta, \boldsymbol{X})] = -E[\frac{\partial}{\partial\theta}t^T]$$

Proof. (1) Use the fact that

$$E[t^T] = \int t^T f(x, \theta) dx$$

and differentiate with respect to θ on both sides. (2) use the fact $Es(\theta, \boldsymbol{X}) = 0$ in above lemma. □

3.4.5.2 *Fisher information and MLE*

Definition 3.4.5 (Fisher information). *The covariance matrix of the score function is called **Fisher information matrix**, denoted by $I(\theta)$, and is given by*

$$I(\theta) = E[s(\theta, \boldsymbol{X})s^T(\theta, \boldsymbol{X})],$$

where

$$s(\theta, \boldsymbol{X}) = \nabla_\theta \log L(\boldsymbol{x}|\theta), \log L(\boldsymbol{x}|\theta) = \sum_{i=1}^{n} f(X = x_i|\theta).$$

3.4.5.3 *MLE efficiency*

Definition 3.4.6 (efficient estimator). *An estimator $\hat{\theta}$ is said to be **efficient** if it is **unbiased** and the covariance of $\hat{\theta}$ achieves the Cramer-Rao lower bound; i.e., it satisfies*

$$E[\hat{\theta}] = \theta, E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] = I^{-1}(\theta).$$

Theorem 3.4.8 (sufficient and necessary condition for MLE be efficient). *[5, p. 552] An unbiased estimator $\hat{\theta}$ is efficient (i.e., achieving the Cramer-Rao lower bound) if and only if*

$$I(\theta)(\hat{\theta} - \theta) = s(\theta, \boldsymbol{X})$$

*where J is the Fisher information matrix. Furthermore, **any unbiased maximum-likelihood estimator is an efficient estimator.***

Proof. (forward) Suppose $I(\theta)(\hat{\theta} - \theta) = s(\theta, \boldsymbol{X})$, then

$$I = E[ss^T] = IE[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T]I$$

by the definition of I. Multiply both sides by I^{-1}, and rearrange we get

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] = I^{-1}.$$

(converse) Assume $\hat{\theta}$ is efficient (also unbiased). First we can show

$$E[s(\theta, X)(\hat{\theta} - \theta)^T] = \mathcal{I},$$

where $\mathcal{I}$ is identity matrix and the expectation is taken with respect to X. This is because $E[s(\theta, X)\theta^T] = E[s(\theta, X)]\theta = 0$, where we use Theorem 3.4.1. In addition,

$$\begin{aligned} E[s(\theta, X)\hat{\theta}^T] &= \int \frac{\partial \log f(x|\theta)}{\partial \theta} \hat{\theta}^T f(x|\theta) dx \\ &= \int \frac{\partial f(x|\theta)}{\partial \theta} \hat{\theta}^T dx \\ &= \frac{\partial}{\partial \theta} E[\hat{\theta}^T] \\ &= \frac{\partial}{\partial \theta} \theta^T \\ &= \mathcal{I}. \end{aligned}$$

Now use Cauchy-Schwartz inequality (Theorem 1.10.4)

$$\begin{aligned} \mathcal{I} &= (E[s(\theta, X)(\hat{\theta} - \theta))^T])^2 \\ &= E[ss^T]E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \\ &= IE[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \\ &= II^{-1} \\ &= \mathcal{I} \end{aligned}$$

Because the equality in Cauchy-Schwartz inequality hold if and only if

$$s(\theta, X) = K(\theta)(\hat{\theta} - \theta),$$

for some $K(\theta)$, which can be $I(\theta)$. □

Remark 3.4.7 (a maximum-likelihood estimator is not necessary efficient)**.**

- Consider a biased MLE, then

  $$s(\theta, X) = 0 \neq I(\theta)(\hat{\theta} - \theta).$$

  Therefore, the variance of $\hat{\theta}$ does not achieve the Cramer-Rao lower bound (Theorem 3.4.8).
- However, as the sample size becomes sufficiently large, MLE is consistent and asymptotic efficient (Theorem 3.6.1) no matter it is unbiased or not.

Example 3.4.4. Let the pmf of Bernoulli distribution parameterized by $f(x;\theta) = \theta^x(1-\theta)^{1-x}, x \in \{0,1\}$. Then

$$\begin{aligned}
I(\theta) &= -E[\frac{\partial^2 \log(f(x;\theta))}{\partial\theta^2}] \\
&= E[\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}] \\
&= \theta(1/\theta^2) + (1-\theta)(1/(1-\theta)^2) \\
&= \frac{1}{\theta(1-\theta)}
\end{aligned}$$

3.4.6 One sample estimation

Lemma 3.4.10 (mean and variance estimator of normal samples). *[6] Let $X_1, X_2, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$. Define two statistics*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

- *$\overline{X}$ is the* ***uniformly minimum variance unbiased*** *(UMVU) estimator of μ. And*

$$E[\overline{X}] = \mu, Var[\overline{X}] = \frac{\sigma^2}{n}.$$

- *S^2 is the uniformly minimum variance unbiased (UMVU) estimator of σ^2. And*

$$E[S^2] = \sigma^2, Var[S^2] = \sigma^4\frac{2}{n-1}.$$

Proof. (1) To prove UMVU, see reference. Note that from Theorem 2.2.5,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

and

$$Var[\frac{(n-1)S^2}{\sigma^2}] = \frac{(n-1)^2}{\sigma^4}Var[S^2] = 2(n-1),$$

where use $Var[\chi^2(n)] = 2n$ in Lemma 2.2.35.

The information matrix is given by

$$I_1()$$

□

Example 3.4.5. Consider a sample $X_1, X_2, ..., X_n$ of iid normal random variable with unknown mean and variance. Consider two variance estimator

$$S_1^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})}{n-1}, S_2^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})}{n}.$$

Then

- The MSE for S_1^2 is

$$\begin{aligned} MSE[S_1^2] &= Var[S_1^2] + [Bias]^2 \\ &= \frac{2\sigma^4}{n-1} + 0 = \frac{2\sigma^4}{n-1} \end{aligned}$$

  where we use the fact that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

  from Theorem 2.2.5 and

$$Var[\frac{(n-1)S^2}{\sigma^2}] = \frac{(n-1)^2}{\sigma^4} Var[S^2] = 2(n-1),$$

  where use $Var[\chi^2(n)] = 2n$ in Lemma 2.2.35.
- The MSE for S_2^2 is

$$\begin{aligned} MSE[S_2^2] &= (Var[S_2^2] + [Bias])^2 \\ &= \frac{2(n-1)\sigma^4}{n} + (E[S_2^2] - \sigma^2)^2 \\ &= \frac{2(n-1)\sigma^4}{n} + (\frac{(n-1)\sigma^2}{n} - \sigma^2)^2 \\ &= \frac{2n-1}{n^2}\sigma^4 \end{aligned}$$

- $MSE[S_2^2] < MSE[S_1^2]$. That is, the maximum-likelihood estimator has smaller MSE than the unbiased estimator.

3.5 Sufficiency and data reduction

Motivation for sufficient statistic:

- When we estimate a model parameter θ, **not all the information in the data are relevant** to the estimation procedure. For example, if we want to estimate the mean, then the order of the sample is irrelevant.
- A sufficient statistic for a model parameter θ represents the **summary of all information from the data that are useful** for estimation of θ.
- **A statistic sufficient for one parameter θ_1 might not be sufficient for other parameters θ_2.**

Definition 3.5.1 (sufficient statistics). *Let X be a random sample of size n. A statistic $T(X)$ is a sufficient statistic for θ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ; that is*

$$P(X|T,\theta) = P(X|T)$$

In otherwise, X and θ are conditional independent given T.

Remark 3.5.1 (sufficient statistic as a lossless data compression). A statistic is sufficient means that $T(X)$ itself can capture all the information useful in estimating θ; the sample X might contain more information than $T(X)$(since $T(X)$ is usually not 1-1), but this additional information does not provide additional usefulness in estimating θ.

Remark 3.5.2 (trivial sufficient statistic). The statistic $T(X_1, ..., X_n) = (X_1, ..., X_n)$ is always sufficient for any estimation task.

Theorem 3.5.1 (Neyman-Fisher Factorization theorem). *[2, p. 276] Let $f(\boldsymbol{x}|\theta)$ denote the joint pdf or pmf of a sample $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic for θ if and only if there exist functions $g(T(\boldsymbol{x}|\theta))$ such that for all sample points $\boldsymbol{x} \in \mathcal{X}$ and all parameter points $\theta \in \Theta$, we have*

$$f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x})|\theta)h(\boldsymbol{x})$$

Proof. (1) Assume $T(\boldsymbol{X})$ is sufficient, then we have $f(x|T(x),\theta) = f(x|T(x))$. Then we have

$$\begin{aligned} f(x|\theta) = f(x|\theta)f(T(x)|x,\theta) = f(x,T(x)|\theta) &= f(T(x)|\theta)f(x|T(x),\theta) \\ &= f(T(x)|\theta)f(x|T(x))(\text{use sufficiency}) \\ &= h(x)g(T(x)|\theta) \end{aligned}$$

(2)

□

[5, p. 451] for sufficient statistic examples

Definition 3.5.2 (minimal sufficient). *[4, p. 48] A sufficient statistic is minimal sufficient if it can be represented as a function of any other sufficient statistic. That is, if $S(X)$ is minimal sufficient, then there exists a function f such that $S(X) = f(T(X))$, where $T(X)$ is another sufficient statistic.*

Remark 3.5.3.

- The minimal sufficient statistic represents **the smallest amount of information yet still sufficient for estimation of the parameter.** Therefore, a minimal sufficient statistic can be represented by any other sufficient statistic, which always contains more information.
- A minimal sufficient statistic represents the maximal (and hence optimal) lossless compression of data.

Definition 3.5.3 (complete statistic). *[4, p. 48]*

Theorem 3.5.2. *[4, p. 49] If statistic T is complete and sufficient, then T is minimal sufficient.*

3.6 Asympotic properties of estimators

3.6.1 Asymptotic properties of MLE

Definition 3.6.1 (asymptotic efficiency). *[1, p. 542] An estimator is asymptotically efficient if it is consistent, asymptotically normally distributed, and has an asymptotic covariance matrix that is not larger than the asymptotic covariance matrix of any other consistent, asymptotically normally distributed estimators.*

Theorem 3.6.1 (asymptotic properties of MLE). *[5, p. 553][1, p. 478] Let $\hat{\theta}$ be the MLE of coefficient associated with distribution $f(x;\theta)$. Let θ_0 be the true value of the parameter. Let $I_1(\theta)$ be the Fisher information matrix associated with distribution $f(x;\theta)$. It follows that*

- *Maximum-likelihood estimators are consistent; that is*

$$\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta_0.$$

- *Maximum-likelihood estimators are asymptotic normal; that is,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to MN(0, [I_1(\theta_0)]^{-1})$$

- *Maximum-likelihood estimators are asymptotic efficient.*

Proof. (sketch) (1) Define a scaled log-likelihood function

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i|\theta).$$

Consider the function $L(\theta) = \int (\log f(x;\theta) f(x;\theta_0) dx$, we can show that the true parameter θ_0 is the maximizer of $L(\theta)$; that is, for any θ, we have

$$L(\theta) \leq L(\theta_0).$$

$$\begin{aligned}
L(\theta) - L(\theta_0) &= E_{\theta_0}[\log f(X;\theta) - \log f(X;\theta_0)] \\
&= E_{\theta_0}[\log \frac{f(X;\theta)}{f(X;\theta_0)}] \\
&\leq E_{\theta_0}[\frac{f(X;\theta)}{f(X;\theta_0)} - 1] \\
&= \int (\frac{f(X;\theta)}{f(X;\theta_0)} - 1) f(x;\theta_0) dx \\
&= \int f(x;\theta) dx - \int f(x;\theta_0) dx \\
&= 1 - 1 = 0
\end{aligned}$$

where we use inequality $\log x \leq x - 1$. From the law of large numbers, $L_n(\theta)$ converge to $L(\theta)$ in probability. Since MLE $\hat{\theta}$ is the maximizer for $L_n(\theta)$, $\hat{\theta}$ converges to θ_0 in probability. (2) Since as $n \to \infty$, $\hat{\theta} \to \theta_0$ (becomes unbiased), we use Theorem 15.6.3 to show that $\hat{\theta} - \theta_0$ has the variance reaching the Cramer-Rao lower bound. Further, use central limit theorem can arrive at the conclusion. (3) It is asymptotic efficient because its variance reaches Cramer-Rao lower bound. □

Remark 3.6.1 (interpretation and implications). For finite-sample MLE, only unbiased MLE is efficient(Theorem 3.4.8). At the large sample limit, MLE is always consistent (asymptotic unbiased) therefore efficient.

Example 3.6.1. Consider an exponential distribution with parameter α such that its pdf is given by

$$f(x;\alpha) = \alpha e^{-\alpha x}, x \geq 0.$$

- The MLE for α from an iid random sample $X_1, ..., X_n$ is given by $\hat{\alpha} = 1/\overline{X}$ since

$$\log L(\alpha) = n \log \alpha - \alpha \sum_{i=1}^{n} X_i$$

$$\partial \log L(\alpha)/\partial\alpha = \frac{n}{\alpha} - \sum_{i=1}^{n} X_i$$

$$\partial \log L(\alpha)/\partial\alpha = 0 \implies \hat{\alpha} = 1/\overline{X}.$$

- The Fisher information is given by $I(\alpha) = \frac{1}{\alpha^2}$ since

$$\begin{aligned} \log f(x;\alpha) &= \log \alpha - \alpha x \\ \partial^2 \log L(\alpha)/\partial \alpha^2 &= -\frac{1}{\alpha^2} \\ I(\alpha) = -E[\partial^2 \log L(\alpha)/\partial \alpha^2] &= \frac{1}{\alpha^2}. \end{aligned}$$

- The MLE $\hat{\alpha} = 1/\overline{X}$ is asymptotic normal and

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \to N(0, \alpha_0^2).$$

Example 3.6.2 (Fish information matrix for normal distributions). [1, p. 548]

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu^2} &= -\frac{n}{\sigma^2} \\ \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 \\ \frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu) \end{aligned}$$

For the asymptotic variance of the maximum likelihood estimator, we need the expectations of these derivatives. Using $E[X_i] = \mu$, we have

$$[-E_0[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0 \partial \theta_0^T}]]^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

Let $\hat{\theta}$ be the maximum-likelihood estimator, then the gradient of the log-likelihood function equals zero at $\hat{\theta}$; that is,

$$g(\hat{\theta}) = 0.$$

Expand this equation in a second-order Taylor series around θ_0, we have

$$0 = g(\hat{\theta}) = g(\theta_0) + H(\overline{\theta})(\hat{\theta} - \theta_0),$$

where θ_0 is the true value, and $\overline{\theta} = w\hat{\theta} + (1-w)\theta_0$ for some $0 < w < 1$. Rearranging this function and multiply it by n and we get

$$\sqrt{n}(\hat{\theta} - \theta_0) = [-H(\overline{\theta})]^{-1}(\sqrt{n}g(\theta_0))$$

3.7 Interval estimation

Definition 3.7.1 (interval estimator). *[2, p. 418] An interval estimator of a real-valued parameter θ is a pair of statistic $L(\boldsymbol{X}$ and $U(\boldsymbol{X})$ satisfying $L(\boldsymbol{x}) \leq U(\boldsymbol{x}), \forall x \in \mathcal{X}$. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, then the inference $L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$ is made. The random interval estimator $[L(\boldsymbol{X}U(\boldsymbol{X})]$ is called interval estimator.*

Remark 3.7.1. It is important to note that the interval estimator is a random quantity.

Definition 3.7.2 (Confidence interval). *Let $X_1, X_2, ..X_n$ denote a random sample on a random variable X, where X has pdf $f(x;\theta)$. Let $\alpha(0 < \alpha < 1)$ be given. Let $L = L(X_1, X_2, ..X_n), U = U(X_1, X_2, ..X_n)$ be two statistics. We say that the interval (L, U) is a $(1-\alpha)$ confidence interval for θ if*

$$1 - \alpha = P_\theta(\theta \in (L, U))$$

Definition 3.7.3 (interval estimation, decision). *An interval estimation is a decision rule that assert the parameter to be estimated in the interval $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ when $\boldsymbol{X} = \boldsymbol{x}$ is given.*

Definition 3.7.4 (coverage probability). *[2, p. 418] For an interval estimator $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ of a parameter θ, the coverage probability of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is the probability that the random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ covers the true parameter θ, i.e., $P(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]|\theta)$*

Definition 3.7.5 (confidence coefficient). *[2, p. 418] For an interval estimator $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ of a parameter θ, the confidence coefficient of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is*

$$\inf_{\theta \in \Theta} P(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]|\theta)$$

Remark 3.7.2. Given different θ, we have different coverage probability for $[L(\boldsymbol{X}), U(\boldsymbol{X})]$. And the confidence coefficient is the lower bound.

Theorem 3.7.1 (optimal length for unimodal pdf). *[2, p. 441] Let $f(x)$ be a unimodal pdf. If the interval $[a, b]$ satisfies:*

- $\int_a^b f(x)dx = 1 - \alpha$
- $f(a) = f(b) > 0$
- *$a \leq x^* \leq b$, where x^* is a mode of $f(x)$*

then $[a,b]$ is the shortest length among all intervals that satisfies $\int_a^b f(x)dx = 1-\alpha$.

Proof: TODO

3.7.1 Evaluating interval estimators

Definition 3.7.6 (unimodal pdf). *Let $f(x)$ be a pdf. We say $f(x)$ is unimodal if there exist a x^* such that $f(x)$ is nondecreasing for $x \leq x^*$ and nonincreasing for $x \geq x^*$.*

3.7.2 Interval estimation for normal distribution

Lemma 3.7.1 (confidence interval for mean of normal random sample). *Let X be a normal random variable $N(\mu, \sigma^2)$, Let $X_1, ..., X_n$ be the random sample, let S^2 and $\overline{X}$ be the sample variance and sample mean, then*

- *If σ is known, then the $(1-\alpha)$ confidence interval for μ is*
$$(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$
- *If σ is unknown, then the $(1-\alpha)$ confidence interval for μ is*
$$(\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1), \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1))$$

where $z_{\alpha/2}, t_{\alpha/2)}(n-1)$ are the upper critical point of $\alpha/2$ for standard normal distribution and $t(n-1)$ distribution.

Proof. (1) Use the fact that
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
(2) Use the fact that
$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$
□

Remark 3.7.3 (knowing σ reduce uncertainty). Note that t distribution is wider(has big tails) than normal, which suggest larger confidence interval when σ is unknown.

3.7.3 Large sample confidence interval

Lemma 3.7.2. *[7, p. 220] Let $X_1, ..., X_n$ be the random sample of a random variable with mean μ and variance σ^2. (Note that X is not necessarily normal). Then the $(1-\alpha)$ confidence interval for μ for large sample size is given as*

$$(\overline{X} - \frac{S}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{S}{\sqrt{n}} z_{\alpha/2})$$

Proof. When n is large, $S \approx \sigma$. Based on central limit theorem Theorem 1.13.3.

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

□

3.8 Bayesian estimation theory

3.8.1 Basics

Definition 3.8.1 (Bayesian statistical model). *A Bayesian statistical model is composed of a **data generation model**, $p(x|\theta)$, and a **prior distribution model** on the model parameters, $p(\theta), \theta \in \mathbb{R}^n$.*

Definition 3.8.2 (maximum a posterior estimator). *Maximum A Posteriori estimator is given as*

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x) = \arg\max_{\theta} p(x|\theta)p(\theta)$$

Lemma 3.8.1 (three rules). *[8, p. 15]Let $H_1, ..., H_k$ be the partition of the sample space H and $P(H) = 1$, then we have*

1. *Rule of total probability:*
$$\sum_k P(H_k) = 1$$
2. *Rule of marginal probability:*
$$P(E) = \sum_k P(E \cap H_k) = \sum_k P(E|H_k)P(H_k)$$
3. *Bayes' rule:*
$$Pr(H_k|E) = \frac{P(E|H_k)P(H_k)}{\sum_k P(E|H_k)P(H_k)}$$

3.8.2 Conjugate prior

Definition 3.8.3 (conjugate prior). *$p(\theta)$ is a conjugate prior for $p(x|\theta)$ if*

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|x) \in \mathcal{P}$$

where $\mathcal{P}$ is a family of pdf parameterized by θ. In other words, $p(\theta)$ and $p(x|\theta)$ are in the same family.

Remark 3.8.1 (Benefits of using prior).

- When prior and posterior distribution are in the same family, it is easy to interpret how the observations x changes the prior distribution

- For a comprehensive account of conjugate priors, see [9][10].

Lemma 3.8.2 (Bernoulli). *Let Bernoulli distribution* $p(x|\theta) = \theta^x(1-\theta)^{1-x}, x \in \{0,1\}$ *be the data generation model, let Beta distribution* $B(\alpha,\beta)$ *be the prior model, let the observation data be* $x_1, x_2, ..., x_n \in \{0,1\}$*, then the posterior distribution is given as*

$$B(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$$

Proof. □

3.8.3 Bayesian prediction

Lemma 3.8.3 (prediction based on observations). *Given* n *observations* $y_1, y_2, ..., y_n$ *of an experiment, assuming the underlying statistical model is parameterized by* θ*, then the prediction for the outcome* y^* *of a new experiment is*

$$\int p(y^*|y_1, ..., y_n, \theta)p(\theta|y_1, ..., y_n)d\theta$$

In particular, if y^* *is conditionally independent of* $y_1, ..., y_n$ *given* θ*(that is,* $y_1, ..., y_n$ *only provides information through* θ*), then the prediction is*

$$\int p(y^*|\theta)p(\theta|y_1, ..., y_n)d\theta$$

Proof. (1)

$$\begin{aligned} p(y^*|y_1, ..., y_n) &= \int p(y^*, \theta|y_1, ..., y_n)d\theta \\ &= \int p(y^*|\theta, y_1, ..., y_n)p(\theta|y_1, ..., y_n)d\theta \end{aligned}$$

where we have used the chain rule(**??**,$P(\theta, y) = P(\theta)P(y|\theta)$). (2) direct consequence of conditional independence. □

3.9 Decision theory framework

3.9.1 The framework

[11, p. 4] A statistical decision problems involves the following elements:

- A **parameter space** Θ. Usually, $\Theta \in \mathbb{R}^d$. The unknown parameter value $\theta \in \Theta$ is the quantity we wish to make inference about.
- A **sample space** $\mathcal{X}$, the space in which the observation data x lie. Usually, $x \in \mathbb{R}^n, \forall x \in \mathcal{X}$.
- A **family of probability distributions** on sample space $\mathcal{X}$, indexed by values $\theta \in \Theta$, written as $\{P_\theta(x), x \in \mathcal{X}, \theta \in \Theta\}$. In practice, we will use probability density function or probability mass function $f(x;\theta)$
- An **action space** $\mathcal{A}$. The set of actions available to decision maker. For example, in hypothesis testing, the actions might be "accept or reject"; in point estimation, the action is select $\theta \in \Theta$.
- A **Loss funtion** $L : \mathcal{A} \times \Theta \to \mathbb{R}$. Given a θ and an action d, we use L is assess an action.
- A set $\mathcal{D}$ of **decision rules**. An element $d : \mathcal{X} \to \mathcal{A}, d \in cD$, specifies which action to make when a point $x \in \mathcal{X}$ is given.

Definition 3.9.1 (risk function). *Let X be a random variable, let observation data $x \in \mathcal{X}$ be the realized value of X. For a parametric value $\theta \in \Theta$, the risk function is defined as*

$$R(\theta, d) = E_\theta[L(\theta, d(X))] \tag{6}$$

$$= \begin{cases} \int_{\mathcal{X}} L(\theta, d(x)) f(x;\theta) dx \\ \sum_{\mathcal{X}} L(\theta, d(x)) f(x;\theta) \end{cases} \tag{7}$$

Example 3.9.1 (Examples of actions space). [12, p. 17]

- **Point estimation**: The action space will usually be $\mathbb{R}^n$
- **Interval estimation**: The action space will be a tuple $\mathbb{R}^2$
- **Prediction**: The action space will be a space of function that gives a prediction when new observation is available.

Remark 3.9.1.

- The density $f(x;\theta)$ gives the distribution of the sample data x
- The risk function gives the average loss value associated with a decision rule d under $f(\theta, x)$.

- In practice, we will select some decision rules,say Bayes rule, then we can use $R(\theta, d)$ to infer the value θ via $\theta^* = \arg\min R(\theta, d)$.
- In practice, there will be no way to find a decision rule d that makes the risk function $R(\theta, d)$ uniformly smallest for all values of θ.

Definition 3.9.2 (dominating rules). *[11, p. 7] Given two decision rules d and d', d strictly dominates d' if $R(\theta, d) \leq R(\theta, d') \forall \theta \in \Theta$, and $R(\theta, d) < R(\theta, d')$ for at least one value θ.*

Remark 3.9.2. The dominance condition can be thought as the optimality condition; however, that might not exist and even exist that will usually difficult to find.

Definition 3.9.3 (admissible vs inadmissible). *[11, p. 7] Any decision rule which is strictly dominated by another decision rule is inadmissible; any decision rules not dominated by others are admissible.*

Remark 3.9.3. Admissibility can be thought as **weak** condition for optimality. It represents the lack of negative properties, and does not specify positive properties.

Definition 3.9.4 (Minimax decision rule). *[11, p. 8] $d^* \in \mathcal{D}$ is called the minimax decision rule if*

$$d^* = \arg \inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d)$$

Remark 3.9.4.

- Minimax decision rule is a very conservative rule, which is used to guarantee performance even in the worst cases.
- Minimax decision rule might give undesired result if the worst situation only happens with tiny probability.

Definition 3.9.5 (Bayesian decision rule). *[11, p. 10] The Bayes risk of a decision rule d is defined as $r(\pi, d) = \int_{\theta \in \Theta} R(\theta, d)\pi(\theta) d\theta$ A decision rule is said to be Bayesian decision rule if*

$$d^* = \arg \inf_{d \in \mathcal{D}} r(\pi, d)$$

3.9.2 General Bayesian methods

Definition 3.9.6. *The posterior distribution is given as*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta) d\theta}$$

where $\pi(\theta)$ is the prior distribution

Remark 3.9.5. In Bayesian statistics, X and θ are both viewed as random variables, with joint probability given as $\pi(\theta)f(x|\theta)$.

$$\begin{aligned}
r(\pi, d) &= \int_{\Theta}\int_{\mathcal{X}} L(\theta, d(x))f(x|\theta)\pi(\theta)d\theta dx \\
&= \int_{\Theta}\int_{\mathcal{X}} L(\theta, d(x))f(x)\pi(\theta|x)d\theta dx \\
&= \int_{\mathcal{X}} f(x)[\int_{\Theta} L(\theta, d(x))\pi(\theta|x)d\theta]dx
\end{aligned}$$

Therefore, if we can find d^* that minimizes $\int_{\Theta} L(\theta, d(x))\pi(\theta|x)d\theta$, then d^* is the Bayesian decision rule.

3.10 Cochran's theorem

Lemma 3.10.1. *Let $X \sim MN(\mu, \Sigma)$ be a n dimensional random vector, then*

$$(X-\mu)^T \Sigma^{-1} (X-\mu) \sim \chi^2(n)$$

Proof. Let $Y = \Sigma^{-1/2}(X-\mu)$, then $Y \sim MN(0, I)$. Then

$$Y^T Y \sim \chi^2(n).$$

□

Lemma 3.10.2. *Let $X_1, X_2, ..., X_n$ be real numbers. Suppose that $\sum_{i=1}^n X_i^2$ can be decomposed into a sum of positive semi-definite quadratic forms, that is*

$$\sum_{i=1}^n X_i^2 = Q_1 + ... + Q_k$$

where $Q_i = X^T A_i X$ with $rank(A_i) = r_i$. If $\sum_{i=1}^k r_i = n$, then there exists an orthonormal matrix C such that $X = CY$ and

$$\begin{aligned} Q_1 &= Y_1^2 + ... + Y_{r_1}^2 \\ Q_2 &= Y_{r_1+1}^2 + ... + Y_{r_1+r_2}^2 \\ &\cdots \end{aligned}$$

Proof. (informal)Note that when we decompose a matrix, its sum of rank of the decomposed matrix will increase(**??**), i.e.,

$$rank(A+B) \leq rank(A) + rank(B)$$

and the equality only holds when $\mathcal{R}(A) \cap \mathcal{R}(B) = \emptyset$.
Since in our case $rank(\sum A_i) = \sum rank(A_i)$, then we must have $\mathbb{R}^n = \mathcal{R}(A_1) \oplus \mathcal{R}(A_2) ... \oplus \mathcal{R}(A_k)$. Take the basis of each $\mathcal{R}(A_i)$ and make it to be orthonormal matrix C. Then Y_i are just the orthonormal projection to subspace $\mathcal{R}(A_j)$. An accessible proof is at Theorem 2.2.4 □

Theorem 3.10.1 (Cochran's theorem). *Let $X_1, X_2, ..., X_n$ be iid $N(0, \sigma^2)$ random variables. Suppose that $\sum_{i=1}^n X_i^2$ can be decomposed into a sum of positive semi-definite quadratic forms, that is*

$$\sum_{i=1}^n X_i^2 = Q_1 + ... + Q_k$$

where $Q_i = X^T A_i X$ with $rank(A_i) = r_i$. If $\sum_{i=1}^k r_i = n$, then there exists an orthonormal matrix C such that $X = CY, Y = C^T X$ ($Y_1, Y_2, ..., Y_n$ are independent random variables with $N(0, \sigma^2)$)and

$$Q_1 = Y_1^2 + ... + Y_{r_1}^2$$
$$Q_2 = Y_{r_1+1}^2 + ... + Y_{r_1+r_2}^2$$
$$\cdots$$

Moreover, we have

- *$Q_1, Q_2, ..., Q_k$ are independent*
- *$Q_i \sim \sigma^2\chi^2(r_i)$.*

Proof. (1) Use above lemma. Note that $Y_1, Y_2, ..., Y_n$ are still independent normal because of Lemma 2.2.9. (2) Since Q_i and Q_j have non-overlapping Y_is, they are independent to each other. (3) From properties of χ^2 distribution(Lemma 2.2.35). □

Corollary 3.10.1.1 (distribution of sample variance). *Let $Y_1, ..., Y_n$ be iid random variable with $N(\mu, \sigma^2)$, then*

$$\sum_{i=1}^n (Y_i - \overline{Y})^2 \sim \sigma^2\chi^2(n-1)$$

and

$$\sum_{i=1}^n (Y_i - \mu)^2/n \sim \sigma^2\chi^2(1)$$

Proof.

$$\sum_{i=1}^n (Y_i - \mu)^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2 + \sum_{i=1}^n (Y_i - \mu)^2/n$$

And

$$(Y - \mu)^T(Y - \mu) = (Y - \mu)^T(I - \frac{1}{n}J)(Y - \mu) + (Y^T - \mu)^T(\frac{1}{n}J)(Y - \mu)$$

and $rank(\frac{1}{n}J)$ has rank 1 and $rank(I - \frac{1}{n}J) = n - 1$. □

Remark 3.10.1.

- The matrix $\frac{1}{n}J$ has rank 1 is because it only has one linearly independent column.
- The matrix $I - \frac{1}{n}J$ is because $rank(I - \frac{1}{n}J) \geq rank(I) - rank(\frac{1}{n}J) = n - 1$(**??**)Also $I - \frac{1}{n}J$ has eigenvector 1 associated with eigenvalue 0. Therefore, $rank(I - \frac{1}{n}J) < n$. In summary, we have $rank(I - \frac{1}{n}J) = n - 1$.
- The matrix $I - \frac{1}{n}J$ has rank $n - 1$ because it is orthogonal projector($P^T = P, P^2 = P$) and $rank(I - \frac{1}{n}J) = Tr(I - \frac{1}{n}J) = n - 1$.(**??**).

3.11 Notes on bibliography

For decision theory, [11][5] For overall graduate level treatment, see [2][7]. For likelihood based methods, see [3] For large sample theory(asymptote analysis), see [13].

For introductory level Bayesian statistics, see [8].

For good treatment on statistical estimation theory, see [14].

For linear regression models, see [15][16].

For multivariate statistical analysis, see [17].

For mixed models, see [18].

For an informal but deep treatment on robust statistics, see [19].

For an extensive discussion on statistical distribution, see [20][21].

BIBLIOGRAPHY

1. Greene, W. *Econometric Analysis* ISBN: 9780134461366. https://books.google.com/books?id=Gqh2AQAACAAJ (Pearson, 2017).
2. Casella, G. & Berger, R. L. *Statistical inference* (Duxbury Pacific Grove, CA, 2002).
3. Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood* (Oxford University Press, 2001).
4. Keener, R. *Theoretical Statistics: Topics for a Core Course* ISBN: 9780387938394. https://books.google.com/books?id=aVJmcega44cC (Springer New York, 2010).
5. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
6. Sciacchitano, A. *et al.* Collaborative framework for PIV uncertainty quantification: comparative assessment of methods. *Measurement Science and Technology* **26,** 074004 (2015).
7. Hogg, R. V., McKean, J. & Craig, A. T. Introduction to Mathematical Statistics, 7 ed (2012).
8. Hoff, P. D. *A first course in Bayesian statistical methods* (Springer Science & Business Media, 2009).
9. Fink, D. A compendium of conjugate priors. *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf,* 46 (1997).
10. Wikipedia. *Conjugate prior — Wikipedia, The Free Encyclopedia* [Online; accessed 7-September-2016]. 2016. https://en.wikipedia.org/w/index.php?title=Conjugate_prior&oldid=728478330.
11. Young, G. A. & Smith, R. L. *Essentials of statistical inference* (Cambridge University Press, 2005).
12. Bickel, P. J. & Doksum, K. A. *Mathematical statistics: basic ideas and selected topics* (CRC Press, 2015).
13. Lehmann, E. L. *Elements of large-sample theory* (Springer Science & Business Media, 1999).
14. Kay, S. M. Fundamentals of statistical signal processing, volume I: estimation theory (1993).

15. Kutner, M., Nachtsheim, C. & Neter, J. *Applied Linear Regression Models* ISBN: 9780072955675. https://books.google.com/books?id=0nAMAAAACAAJ (McGraw-Hill Higher Education, 2003).
16. Seber, G. A. & Lee, A. J. *Linear regression analysis* (John Wiley & Sons, 2012).
17. Johnson, R. & Wichern, D. *Applied Multivariate Statistical Analysis* ISBN: 9780131877153. https://books.google.com/books?id=gFWcQgAACAAJ (Pearson Prentice Hall, 2007).
18. McCulloch, C. E. & Neuhaus, J. M. *Generalized linear mixed models* (Wiley Online Library, 2001).
19. Wilcox, R. R. *Fundamentals of modern statistical methods: Substantially improving power and accuracy* (Springer Science & Business Media, 2010).
20. Forbes, C., Evans, M., Hastings, N. & Peacock, B. *Statistical distributions* (John Wiley & Sons, 2011).
21. Krishnamoorthy, K. *Handbook of statistical distributions with applications* (CRC Press, 2016).

4

HYPOTHESIS TESTING

4.1 Hypothesis testing

4.1.1 Basics

Decision making with uncertainty
Given the observation data, we make a finite choice on which partition θ lies in.

Decision is based on $p(\theta|x)$, the posterior distribution can be calculation as $p(x|\theta \in H_i)$

We have two hypothesis about how the data x might have been generated; We have two models $p(x|H_0)$ and $p(x|H_1)$. The goal is to decide which model is more appropriate.

Definition 4.1.1 (decision region). *The range of X is partitioned into regions R_i and we accept H_i when $x \in R_i$.*

Definition 4.1.2 (hypothesis). *[1, p. 371] A hypothesis is a statement about a population parameter θ.*

Definition 4.1.3 (null and alternative hypothesis). *The two **complementary** hypothesis in a hypothesis testing problem are called null hypothesis and the alternative hypothesis, denoted as H_0 and H_1.*

Remark 4.1.1. Let θ denote a population parameter. The general format of the null and alternative hypothesis is $H_0 : \theta \in \Theta_0$, and $H_1 : \theta \in \Theta_0^C$, where Θ_0 and Θ_1 are disjoint subsets of the parameter space Θ.

Definition 4.1.4 (hypothesis testing procedure). *A hypothesis testing procedure or hypothesis test is a* ***decision rule*** *that specifies decisions based on some test statistic $W(\boldsymbol{X})$*

- *For which value of $W(\boldsymbol{X})$ accept H_0 is true and reject H_1*
- *For which value of $W(\boldsymbol{X})$ reject H_0 and accept H_1.*

Caution!

- We only test if H_0 is true or not.
- We **do not test** H_1.

Definition 4.1.5 (Four types of results in binary hopothesis testing).

1. *Detection: H_0 true, decide H_0*
2. *False alarm/* ***type I error:*** *H_0 true, we reject H_0, decide H_1.*
3. *Miss/* ***type II error:*** *H_1 true, decide H_0(or H_0 is false, we do not reject H_0.)*
4. *Correctly rejection: H_1 true,decide H_1*

Remark 4.1.2 (general remarks on two types of errors). [2, p. 245]

- Generally, given a fixed size of samples, it is not possible to minimize both types of error(like situations in multi-objective optimization).
- We usually consider type I error to be worse and try to minimize or bound type I error first and then minimize type II error.
- H_0 is usually conservative statement such that reject H_0 when it is true will have significant consequence.

Remark 4.1.3 (design of hypothesis).

- The null hypothesis is usually a simple hypothesis **the contradiction** to what we would like to prove.
- The alternative hypothesis is usually a hypothesis what we would like to prove.

Example 4.1.1. In drug clinical trial, it is presumed ineffective(H_0) until sufficient significant evidence is shown for **not being ineffective**. Rejecting H_0 means the drug is ineffective with probability at most α.

Definition 4.1.6 (critical region). *[2, p. 245] A critical region C is the set of the values taken by the random sample $(X_1, X_2, ..., X_n)$ such that the* ***null hypothesis*** *H_0* ***will be rejected.*** *The complement C^C is the region such that H_0 will not be rejected.*

Definition 4.1.7 (size of critical region). *A critical region C is of size α if*

$$\max_{\theta \in H_0} P_\theta[(X_1, ..., X_n) \in C]$$

that is, under the null hypothesis H_0, the probability measure of the critical region where H_0 will be rejected.

Remark 4.1.4 (interpretation, boundedness of type I error).

- The size of critical region bounds the probability of making type I error. By definition, under the assumption of H_0, the probability measure of events rejecting H_0 is smaller or equal than α.
- The size of critical region is also called significance level.(See below)

Definition 4.1.8 (significance level). *The* ***significance level*** *α is a probability threshold below which the* ***null hypothesis will be rejected under the assumption that H_0 is true****. Common values are 0.05 and 0.01.*

Remark 4.1.5 (significance level and type I error). **Significance level is the upper bound of type I error**(note that type I error is the error of rejecting H_0 but H_0 is true. And significance level is the probability of rejection.)

4.1.2 Power of a statistical test

Definition 4.1.9 (statistical power of a test). *The power of a hypothesis test is the probability of making the correct decision if the alternative hypothesis is true. That is, the power of a hypothesis test is the probability of rejecting the null hypothesis H_0 is incorrect(or when the alternative hypothesis H_1 is true):*

$$power = P(reject\ H_0 | H_1\ true).$$

Remark 4.1.6 (motivations). Whenever we conduct a hypothesis test, we prefer a test of high quality. One way of quantifying the quality of a hypothesis test is to ensure that it is a "powerful" test, which makes smaller type II error.

Remark 4.1.7 (type I,II error, significance level, and power).

- **Minimize the probability of committing a Type I error.** That, is minimize $\alpha = P(TypeIError)$. Typically, $\alpha \leq 0.1$.
- **Maximize the power, or reduce the type II error** Note that $\beta = P(TypeIIError) = 1 - power$, typically $\beta \leq 0.2$.

Remark 4.1.8 (factors affecting statistical power). Statistical power may depend on a number of factors.

- the statistical significance criterion used in the test.
- the magnitude of the effect of interest in the population.
- the sample size used to detect the effect

Remark 4.1.9 (applications of power analysis to determine sample size). Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size. For example: âĂIJhow many times do I need to toss a coin to conclude it is unfair?

Definition 4.1.10 (most powerful test). *For a given size or significance level, the test with the greatest power (probability of rejection) for a given value of the parameter(s) being tested, contained in the alternative hypothesis.*

4.1.3 Test methods

4.1.3.1 *Likelihood ratio test statistic*

Definition 4.1.11 (likelihood ratio test statistic). *The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_0^C$ is*

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\Theta_0} L(\theta|\boldsymbol{x})}{\sup_{\Theta} L(\theta|\boldsymbol{x})} \tag{8}$$

4.1.3.2 *Likelihood ratio test*

Definition 4.1.12 (likelihood ratio test). *A likelihood ratio test is any test with a decision rule that reject H_0 if the test statistic $\lambda(\boldsymbol{x}) \leq c, 0 \leq c \leq 1$.*

Remark 4.1.10.

- The likelihood ratio test is in nature a decision strategy.
- The larger the likelihood ratio is, the more likely $\theta \in \Theta_0$.

The threshold value selection has consequences.

Definition 4.1.13 (simple hypothesis test). *A test for $\theta \in \Theta_i, i = 0, 1, ..., k-1$ is said to simple if each Θ_i consists of exactly one element. Otherwise, the test is said to be **composite**.*

4.1.4 p value method

Definition 4.1.14 (p value). *The p value is the probability, assuming the null hypothesis is true, of observing a result at least as extreme as (equal to or "more extreme" than) the observed test statistic.*
p value can also be interpreted as the smallest significant value that H_0 will be reject.

Remark 4.1.11 (How to use p value). Given a significance level α:

- If $p \leq \alpha$, then reject H_0.
- If $p > \alpha$, then accept H_0.

4.1.5 Likelihood ratio test

Definition 4.1.15 (likelihood ratio test). *Let*

- *$L(\hat{w})$ denote the maximum of the likelihood function with respect to θ when θ is in the null parameter space Ω_0.*
- *$L(\hat{\Omega})$ denote the maximum of the likelihood function with respect to θ when θ is in the entire parameter space Ω.*

*Then the **likelihood ratio** is defined by*

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})}.$$

The decision rule to reject the null hypothesis with confidence level α is

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})} \leq k,$$

where k is a critical number.

4.1.6 Hypothesis testing on normal distributions

4.1.6.1 *t Test*

Definition 4.1.16. *The general form of a test statistic is*

$$test - stat = \frac{observed(orestimated)value - hypothesizedvalue}{standarderrorofobserved}$$

Definition 4.1.17 (t-Test). *The t-test is a hypothesis test in which the test statistic follows a Student t-distribution on the subset of the sample space where the null hypothesis H_0 is supported.*

4.1.6.2 *Hypothesis test for normal distribution*

Common notations in this sections:

- sample mean $\overline{X}$
- sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - E[X])^2$

4.1.6.3 *Hypothesis testing on sample mean with known variance*

Consider we have n samples $X_1, ..., X_n$ for a random variable with $N(\mu, \sigma^2)$ with σ^2 known. The hypothesis testing involving the mean can be obtained by using the fact that

$$Z = \frac{\overline{X} - \mu}{\sigma/n}$$

is $N(0, 1)$. We can summarize the test as:

Table 4.1.1: Test on mean with known variance σ^2

H_0	test statistic	H_1	critical(rejection) region
$\mu \leq \mu_0$	$Z = \frac{\overline{X}-\mu}{\sigma/n}$	$\mu > \mu_0$	$z \geq z_\alpha$
$\mu \geq \mu_0$	$Z = \frac{\overline{X}-\mu}{\sigma/n}$	$\mu < \mu_0$	$z \leq -z_\alpha$
$\mu = \mu_0$	$Z = \frac{\overline{X}-\mu}{\sigma/n}$	$\mu \neq \mu_0$	$\lvert z\rvert \geq z_\alpha$

4.1.6.4 *Hypothesis testing on sample mean with unknown variance*

Consider we have n samples $X_1, ..., X_n$ for a random variable with $N(\mu, \sigma^2)$ with σ^2 being unknown. The hypothesis testing involving the mean can be obtained by using the fact that

$$T = \frac{\overline{X} - \mu}{\sigma/n}$$

is $t(n-1)$.(Theorem 2.2.5) We can summarize the test as:

Table 4.1.2: Test on mean with unknown variance σ^2

H_0	test statistic	H_1	critical(rejection) region
$\mu \leq \mu_0$	$T = \frac{\overline{X}-\mu}{S/n}$	$\mu > \mu_0$	$t \geq t_\alpha(n-1)$
$\mu \geq \mu_0$	$T = \frac{\overline{X}-\mu}{S/n}$	$\mu < \mu_0$	$t \leq -t_\alpha(n-1)$
$\mu = \mu_0$	$T = \frac{\overline{X}-\mu}{S/n}$	$\mu \neq \mu_0$	$\|t\| \geq t_\alpha(n-1)$

4.1.6.5 *Hypothesis testing on the variance*

Consider we have n samples $X_1, ..., X_n$ for a random variable with $N(\mu, \sigma^2)$ with σ^2 being unknown. The hypothesis testing involving the mean can be obtained by using the fact that

$$T = \frac{\overline{X} - \mu}{\sigma/n}$$

is $t(n-1)$.(Theorem 2.2.5) We can summarize the test as:

Table 4.1.3: Test on variance

H_0	test statistic	H_1	critical(rejection) region
$\sigma^2 \leq \sigma_0^2$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 \leq \sigma_0^2$	$t \geq \chi^2_\alpha$
$\sigma^2 \geq \sigma_0^2$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 \leq \sigma_0^2$	$t \leq \chi^2_{1-\alpha}$
$\sigma^2 = \sigma_0^2$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 \neq \sigma_0^2$	$t \leq \chi^2_{1-\alpha}$ or $t \geq \chi^2_\alpha$

4.1.6.6 *Hypothesis testing on variance comparison*

Consider we have n samples $X_1, ..., X_n$ for a random variable with $N(\mu, \sigma^2)$ with σ^2 being unknown. The hypothesis testing involving the mean can be obtained by using the fact that

$$T = \frac{\overline{X} - \mu}{\sigma/n}$$

is $t(n-1)$.(Theorem 2.2.5) We can summarize the test as:

Table 4.1.4: Test on variance comparison between two samples

H_0	test statistic	H_1	critical(rejection) region
$\sigma_1^2 \leq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$\sigma_1^2 \leq \sigma_2^2$	$t \geq \chi_\alpha^2$
$\sigma^2 \geq \sigma_0^2$	$F = \frac{S_1^2}{S_2^2}$	$\sigma_1^2 \leq \sigma_2^2$	$t \leq \chi_{1-\alpha}^2$
$\sigma_1^2 = \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$\sigma_1^2 \neq \sigma_2^2$	$F \leq \chi_{1-\alpha}^2$ or $F \geq \chi_\alpha^2$

4.1.7 Two sample tests

4.1.7.1 *Two-sample z test*

Note 4.1.1 (basic setup of two-sample z test)**.**

- $X_1, X_2, ..., X_m$ is a random sample from a distribution with mean μ_1 and variance σ_1^2.
- $Y_1, Y_2, ..., Y_n$ is a random sample from a distribution with mean μ_1 and variance σ_1^2.
- X and Y samples are independent of each other.

Lemma 4.1.1 (mean difference estimator). *[3, p. 363]Let $\overline{X}$ and $\overline{Y}$ denote the sample mean.*

- $E[\overline{X} - \overline{Y}] = \mu_1 - \mu_2$*, i.e.,* $\overline{X} - \overline{Y}$ *is the unbiased estimator of* $\mu_1 - \mu_2$*.*
- $$Var[(\overline{X} - \overline{Y})] = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}.$$

Proof. (1) Straight forward. (2) Using independence, we have

$$Var[\overline{X} - \overline{Y}] = Var[\overline{X}] + Var[\overline{Y}] = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}.$$

□

4.1.7.2 *Two-sample t test*

Note 4.1.2 (basic setup two-sample t test)**.**

- $X_1, X_2, ..., X_m$ is a random sample from a distribution with mean μ_1 and variance σ_1^2.

- $Y_1, Y_2, ..., Y_n$ is a random sample from a distribution with mean μ_1 and variance σ_1^2.
- X and Y samples are independent of each other.

4.1.7.3 *Paired Data*

Note 4.1.3 (basic setup for paired data)**.**

- The data consists of n independently selected pairs $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$, with $E[X_i] = \mu_1$ and $E[Y_i] = \mu_2$.
- Let $D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, ..., D_n = X_n - Y_n$ so that $D_i's$ are the differences within pairs.
- The $D_i's$ are assumed to be normally distributed within mean value μ_D and variance σ_D^2.

4.2 Common statistical tests

4.2.1 Chi-square goodness-of-fit test

Theorem 4.2.1 (Pearson's theorem). *Consider r boxes $B_1, ..., B_r$ and throw n balls $X_1, X_2, ..., X_n$ into these boxes independently of each other with probabilities*

$$P(X_1 \in B_1) = p_1, ..., P(X_r \in B_r) = p_r,$$

such that $p_1 + ... + p_r = 1$.

Let v_j be the number of balls in the jth box, i.e. $v_j = \sum_{i=1}^n \mathbf{1}_{X_i = B_j}$.

It follows that

- *The random variable*
$$\frac{v_j - np_j}{\sqrt{np_j}} \to N(0, 1 - p_j) \textit{ in distribution, as } n \to \infty$$
- *The random vector $Y = (Y_1, Y_2, ..., Y_r)$, $Y_j = \frac{v_j - np_j}{\sqrt{np_j}}$ will converge to $MN(0, \Sigma)$ in distribution, where*
$$\Sigma_{ii} = 1 - p_i, \Sigma_{ij} = -\sqrt{p_i p_j}.$$
- *The random variable*
$$\sum_{j=1}^r \frac{(v_j - np_j)^2}{np_j} \to \chi^2(r-1) \textit{ in distribution, as } n \to \infty.$$

Proof. (1) Note that from Bernoulli distribution

$$E[\mathbf{1}(X_1 \in B_j)] = p_j, Var[\mathbf{1}(X_1 \in B_j)] = p_j(1 - p_j).$$

By the central limit theorem

$$\frac{v_j - np_j}{\sqrt{np_j(1-p_j)}} \to N(0,1) \textit{ in dist } \implies \frac{v_j - np_j}{\sqrt{np_j}} \to N(0, 1 - p_j) \textit{ in dist}.$$

(2)

$$\begin{aligned} E[\frac{v_i - np_i}{\sqrt{np_i}} \frac{v_j - np_j}{\sqrt{np_j}}] &= \frac{1}{n\sqrt{p_i p_j}}(E[v_i v_j] - n^2 p_i p_j) \\ E[v_i v_j] &= E[\sum_{l=1}^{n} \mathbf{1}(X_l \in B_i) \sum_{k=1}^{n} \mathbf{1}(X_k \in B_i)] \\ &= E[\sum_{l=1}^{n} \sum_{k=1, k\neq l}^{n} \mathbf{1}(X_l \in B_i)\mathbf{1}(X_k \in B_i)] \\ &= 2E[\sum_{l=1}^{n} \sum_{k>1}^{n} \mathbf{1}(X_l \in B_i)\mathbf{1}(X_k \in B_i)] \\ &= n(n-1)p_i p_j \\ E[\frac{v_i - np_i}{\sqrt{np_i}} \frac{v_j - np_j}{\sqrt{np_j}}] &= -\sqrt{p_i p_j}. \end{aligned}$$

(3) Note that

$$Y^T Y = Z^T(I - UU^T)Z, Z \in MN(0, I_r), U = (\sqrt{p_1}, \sqrt{p_2}, ..., \sqrt{p_r}),$$

where UU^T is an rank 1 orthogonal projector($U^T U = p_1 + p_2 + \cdots + p_r = 1$).

From the chi-square decomposition theorem(Theorem 2.2.4), we know that $Y^T Y \to \chi^2(r-1).in.dist.$ □

Theorem 4.2.2 (chi-square goodness-of-fit test). *Suppose that we observe an iid sample $X_1, X_2, ..., X_n$ of random variable that take a finite number of values $B_1, B_2, ..., B_r$ with unknown probabilities $p_1, p_2, ..., p_r$. Consider hypotheses*

$$\begin{aligned} H_0 :& p_i = p_i^0, for\ i = 1, 2, ..., r \\ H_1 :& for\ some\ i, p_i \neq p_i^0 \end{aligned}$$

and the test statistic

$$T = \sum_{i=1}^{r} \frac{(v_i - np_i^0)^2}{np_i^0},$$

where $v_j = \sum_{i=1}^{n} \mathbf{1}_{X_i = B_j}$.

It follows that

- *If H_0 is true, then $T \to \chi^2(r) - 1$ in dist.*
- *If H_1 is true, then $T \to \infty$, as $n \to \infty$.*
- *The decision rule is reject H_0 if $T > c$ where $c = \inf\{z : F(z) \geq 0.99\}$.*

Proof. (1) From Pearson's theorem(Theorem 4.2.1). (2) If we write

$$\frac{(v_i - np_i^0)}{\sqrt{np_i^0}} = \sqrt{\frac{p_i}{p_i^0}}\frac{(v_i - np_i)}{\sqrt{np_i^0}} + \sqrt{n}\frac{(v_i - n(p_i - p_i^0)}{\sqrt{np_i^0}},$$

then the second quantity will diverge as $n \to \infty$. □

Note 4.2.1 (p value method for chi-square test). The p-value for a chi-square test is defined as the **tail area above the calculated test statistic**.

For example, consider an experiment with test statistic result

$$T = \sum_{i=1}^{r} \frac{(v_i - np_i^0)^2}{np_i^0}.$$

Then

$$p - value = Pr(\chi^2(r-1) \geq T).$$

Given a significance level α:

- If $p \leq \alpha$, then reject H_0.
- If $p > \alpha$, then accept H_0.

4.2.2 Chi-square test for statistical independence

Lemma 4.2.1. *link*

Denote

$$p_i = \sum_{j=1}^{c} \frac{O_{ij}}{N}, q_j = \sum_{i=1}^{r} \frac{O_{ij}}{N}, E_{ij} = Np_iq_j$$

$$\sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{i,j} - E_{ij})^2}{E_{ij}} \sim \chi^2(p),$$

where $p = (r-1)(c-1)$.

The hypothesis is given by

- *H_0: U is independent of V;*
- *H_1: there exists an statistical relationship between U and V.*

4.2.3 Kolmogorov-Smirnov goodness-of-fit test

Definition 4.2.1 (Kolmogorov-Smirnov(KS) goodness-of-fit test). *The Kolmogorov-Smirnov goodness-of-fit test for a random sample of size N has the following elements:*

- *Hypothesis:*
  - H_0: *the data follow a specified* ***continuous*** *distribution with cdf* $F(t)$.
  - H_1: *the data do not follow the specified distribution.*
- *For* ***ascending ordered*** *sample* $Y_1, Y_2, ..., Y_N$. *KS test statistic is defined as*

$$D = \max_{1 \leq i \leq N} (F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)).$$

- *The significance level* α *and critical value* K_α.
- *If* $D > K_\alpha$, *reject* H_0.

Remark 4.2.1 (interpretation and usage)**.**

- The KS test statistic is measuring the distance of proposed distribution F is the empirical cdf given by $(i-1)/N$ and i/N.
- KS test is used for continuous distribution test. For discrete distribution test, see chi-square goodness-of-fit test(Theorem 4.2.2).
- For the KS critical value table, see link.

4.3 Resampling methods

4.3.1 Bootstrap

4.3.1.1 *Motivation and foundations*

general remarks and motivations

- Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random sample of size n. Let $\hat{\theta}(\boldsymbol{X})$ be a statistic of interest. The goal of bootstrap methods is to measure the standard deviation, confidence interval, or even distributions of $\hat{\theta}$.
- In simple cases, we might be able to directly derive the distribution of the estimator. For example, let $\boldsymbol{X}$ be random samples of a normal distribution, the sample variance S^2 will have $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
- In complex cases where obtaining standard deviation, confidence interval or even distributions of $\hat{\theta}$ is difficult, we use bootstrap methods.

Definition 4.3.1 (Bootstrap sample). *Given a set of N samples, a bootstrap sample is a set of N sample drawn from the original samples with replacement.*

Remark 4.3.1 (Bootstrap sample as a sample drawn from empirical sample distribution). We can view the bootstrap sample as a new set of samples drawn from the empirical sample distribution(the joint distribution of $(X_1, ..., X_N)$). Further, consider we have B bootstrap samples, let n_i be the frequency of i sample, and $\sum_{i=1}^N n_i = BN$. Then the sample distribution can be approximated as

$$\frac{1}{B}\sum_{i=1}^{N} n_i.$$

Lemma 4.3.1 (resampling property). *Given a sample of size n, we re-draw n sample with replacement. Then*

- *The probability that ith sample is not being resampled is $(1-\frac{1}{n})$ at the first time.*
- *The probability that ith sample is not being resampled is $(1-\frac{1}{n})^n$ at the new sample of size n.*
- *The probability that ith sample is not being resampled is e^{-1} at the new sample when $n \to \infty$.*
- *On average, about ne^{-1} of original samples will not show in the new sample as $n \to \infty$.*

Proof. (1)(2) straightforward. (3) The definition of e(Lemma A.7.3).(4) The expectation of binomial random variable. □

Definition 4.3.2 (general bootstrap estimation). *Let $\hat{\theta}$ be a statistic as a function of $(X_1, ..., X_N)$. Let θ_i^* be the estimation evaluated at bootstrap sample i. Then*

$$\frac{1}{B}\sum_{i=1}^{B}\theta_i^* \approx E[\hat{\theta}].$$

Let f be a function on $\hat{\theta}$, we have

$$\frac{1}{B}\sum_{i=1}^{B} f(\theta_i^*) \approx E[f(\hat{\theta})].$$

Remark 4.3.2. We can view every bootstrap sample as a sample drawn from the sample distribution. And the theoretical value $Var[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$ is taking expectation with respect to the sample distribution(the joint distribution of $(X_1, ..., X_N)$).

4.3.1.2 *Bootstrap application examples*

Definition 4.3.3 (bootstrap estimation of variance). *Let $\hat{\theta}$ be a statistic as a function of $(X_1, ..., X_N)$. Let θ_i^* be the estimation evaluated at bootstrap sample i. Then the estimated variance of the $\hat{\theta}$ is given as*

$$Var[\hat{\theta}] = \frac{1}{B-1}\sum_{i=1}^{B}(\theta_i^* - \overline{\theta^*})^2,$$

where

$$\overline{\theta^*} = \frac{1}{B}\sum_{i=1}^{B}\theta_i^*$$

Remark 4.3.3 (interpretation).

- Usually, B is chosen to be $B \approx N^2$.
- No matter how large B is chosen, $Var[\hat{\theta}]$ will not vanish. As $B \approx \infty$, $Var[\hat{\theta}]$ will converge.
- The limiting value of $Var[\hat{\theta}]$ is determined by the original sample size. If $N, B \to \infty$, then $Var[\hat{\theta}] \to \infty$.
- We can view every bootstrap sample as a sample drawn from the sample distribution. And the theoretical value $Var[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$ is taking expectation with respect to the sample distribution(the joint distribution of $(X_1, ..., X_N)$)

Definition 4.3.4 (bootstrap confidence level). *Let $\hat{\theta}$ be a statistic as a function of $(X_1, ..., X_N)$. Let θ_i^* be the estimation evaluated at bootstrap sample i. Let $\theta_1^*, \theta_2^*, ..., \theta_B^*$ be sorted. Denote $k_1 = (int)(B \times \frac{\alpha}{2}), k_2 = (int)(B \times (1 - \frac{\alpha}{2}))$. Then $[\theta_{k_1}^*, \theta_{k_2}^*]$ is the α confidence interval such that*

$$Pr(\theta_{k_1}^* \leq \theta \leq \theta_{k_2}^*) = 1 - \alpha.$$

Definition 4.3.5 (bootstrap t test).

4.4 Robust statistics

Definition 4.4.1 (breakdown point). *The finite sample* ***breakdown point*** *of an estimator is the smallest fraction α of data points such that if $[n\alpha]$ points approach ∞, then the estimator approach ∞.*

Example 4.4.1. The sample mean $x_1, x_2, ..., x_n$ is

$$\begin{aligned}\overline{x}_n &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{1}{n}(\sum_{i=1}^{n} x_i + x_n) \\ &= \frac{n-1}{n}(\overline{x}_{n-1}) + \frac{1}{n}x_n\end{aligned}$$

Example 4.4.2.

- (mean) Given sample size n, the breakdown point for the mean using the arithmetic mean is $1/n$; that is one point can ruin the mean.
- (median) The sample median, as an estimate of a population median, can tolerate up to 50% bad values.

Lemma 4.4.1 (minimizing property of mean and median).

Definition 4.4.2 (bootstrap standard deviation). *Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random sample of size n. Let $\hat{\theta}(\boldsymbol{X})$ be a statistic of interest. Let $\boldsymbol{x_1}$ be one realization of the random sample, from which we can have an estimation via $\hat{\theta}(\boldsymbol{x_1})$. We can use replacement sampling to draw another n samples of $\boldsymbol{x_1}$, which we will denote $\boldsymbol{x_2}$ and have an new estimation $\hat{\theta}(\boldsymbol{x_2})$. We continue this process until we have B estimations of $\hat{\theta}$, from which, we can have an estimation on the standard deviation as*

$$\sigma(\hat{\theta}(\boldsymbol{X})) = \frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}(\boldsymbol{x_i}) - \overline{\hat{\theta}})^2$$

Definition 4.4.3 (bootstrap confidence interval). *Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random sample of size n. Let $\hat{\theta}(\boldsymbol{X})$ be a statistic of interest. Let $\hat{\theta_1}, ..., \hat{\theta_B}$ be the B bootstrap samples. Order the B samples in increasing order, and take out $k_1(\approx B\alpha/2)$ and $k_2(\approx B(1-\alpha/2))$ values in the ordered sequence. Then the interval*

$$(\hat{\theta}_{k_1}, \hat{\theta}_{k_2})$$

is the confidence interval of confidence level of α.

Definition 4.4.4 (bootstrap empirical distribution for hypothesis testing). *Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random sample of size n. Let $\hat{\theta}(\boldsymbol{X})$ be a statistic of interest. Let $\hat{\theta_1}, ..., \hat{\theta_B}$ be the B bootstrap samples. The empirical distribution can be constructed from these bootstrap samples.*

Remark 4.4.1. Under the assumption of H_0, we can calculate the p value of the observation, and then decide whether we can reject H_0.

4.4.1 Robust measures of location

Definition 4.4.5 (α trimmed mean). *Let $k = n\alpha$ rounded to an integer(k is the number of observation removed from both ends for calculation). The **α-trimmed mean** is defined as*

$$\overline{X}_\alpha = \sum_{i=k+1}^{n-k} \frac{X_i}{n-2k}.$$

Definition 4.4.6 (median absolute deviation). *[4, p. 122] A robust estimator of standard deviation of iid random sample $X_1, X_2, ..., X_n$ is the **MAD(median absolute deviation)***

$$\hat{\sigma}^{MAD} = 1.4826 \times median\{\left|X_i - median(X_i)\right|\}.$$

Remark 4.4.2 (interpretation).

- For normally distributed data, $median\{\left|X_i - median(Y_i)\right|\}$ is the estimator of $\Phi^{-1}(0.75)\sigma = \sigma/1.4826$.
- For a iid normal random sample, as sample size $n \to \infty$, the MAD is the unbiased estimates of σ.

4.5 Information criterion for model selection

4.5.1 Bayesian information criterion (BIC)

Definition 4.5.1 (Bayesian information criterion(BIC)). *The BIC is formally defined by*

$$BIC = k\ln(n) - 2\ln(\hat{L}),$$

where

- *$\hat{L}$ is the maximized value of the likelihood function of the model M, i.e., $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function and x is the observed data set.*
- *n is the number of the observation data*
- *k is the number of parameters estimated by the model.*

Lemma 4.5.1 (BIC for multiple linear regression). *The BIC for a multiple linear regression model is given by*

$$BIC(M) = k\log(n) + n\log(RSS/n),$$

where

$$RSS = \sum_{i=1}^{n}(Y_i - b\hat{eta}^T X_i)^2$$

Proof. In the linear regression model, we have

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon = f(X) + \epsilon,$$

where ϵ is normal variable with zero mean and a variance of σ. The likelihood function for parameter $\beta_0, ..., \beta_n, \sigma$ is given by

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(Y_i - f(X_i))^2}{2\sigma^2}).$$

The maximum likelihood will be achieved(**??**) at

$$\hat{\sigma}^2 = \frac{1}{n}RSS,$$

where $RSS = \sum_{i=1}^{n}(Y_i - f(X_i))^2$ such that

$$\ln(\hat{L}) = -\frac{n}{2}\ln(RSS/n) - n/2 - \frac{n}{2}\ln 2\pi.$$

Then $-2\ln(\hat{L}) = n\log(RSS/n) + const.$ □

4.6 Expectation Maximization algorithm

Definition 4.6.1 (problem statement). *Given a random vector that can be partitioned into (X, Z), the observation $x_1, ..., x_N$ of X, the observation generation model $P(X|\theta, Z)$, the goal is to infer the unobserved z_i associated with each observed x_i and the parameter θ.*

Algorithm 1: General EM algorithm

Input: Data set consists of $x_1, x_2, ..., x_N, x_i \in \mathbb{R}^D$

Initialize $\theta = \theta^0, n = 1$

E-step: Calculate the probability distribution $p(z_i^n|x_i, \theta^n)$ of z_i^n associated with each x_i.

M-step: Update $\theta = \theta^{n+1}$ by maximizing the marginal likelihood function using the distribution on z_i^n in E-step:

$$L(x, \theta^{n+1}) = \prod P(x_i|z_i^n, \theta^{n+1})p(z_i^n|x_i, \theta^n)$$

Set $n = n + 1$

Iterate between E-step and M-step

Output: The posterior distribution of z_i and θ

Lemma 4.6.1 (non-decreasing in each step). *Consider the marginal likelihood function*

$$L(X|Z, \theta) = \prod P(x_i|\theta) = \prod \sum_z P(x_i|z_i = z, \theta)p(z_i = z)$$

The E-step and the M-step are both maximizing L.

Proof. (1) (2) M-step is just maximizing L. □

Remark 4.6.1. If X is conditional independent of Z given θ, then E-step is unnecessary. And we can simply do M step alone.

Remark 4.6.2.

- The algorithm will converge a stationary point, which can be local maximum, saddle point or global maximum.
- The algorithm can be viewed as block coordinate optimization iteration as

$$(\theta^n, z^{n+1}) = \arg\max_z F(\theta^n, z^n)$$

and

$$(\theta^{n+1}, z^{n+1}) = \arg\max_\theta F(\theta^n, z^{n+1})$$

4.7 Notes on bibliography

For decision theory, [5][6] For overall graduate level treatment, see [1][2]. For likelihood based methods, see [7] For large sample theory(asymptote analysis), see [8].

For introductory level Bayesian statistics, see [9].

For good treatment on statistical estimation theory, see [10].

For linear regression models, see [11][12].

For multivariate statistical analysis, see [13].

For mixed models, see [14].

For an informal but deep treatment on robust statistics, see [15].

For an extensive discussion on statistical distribution, see [16][17].

BIBLIOGRAPHY

1. Casella, G. & Berger, R. L. *Statistical inference* (Duxbury Pacific Grove, CA, 2002).
2. Hogg, R. V., McKean, J. & Craig, A. T. Introduction to Mathematical Statistics, 7 ed (2012).
3. Devore, J. L. *Probability and Statistics for Engineering and the Sciences* (Cengage learning, 2015).
4. Ruppert, D. *Statistics and data analysis for financial engineering, 2ed* (Springer, 2015).
5. Young, G. A. & Smith, R. L. *Essentials of statistical inference* (Cambridge University Press, 2005).
6. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
7. Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood* (Oxford University Press, 2001).
8. Lehmann, E. L. *Elements of large-sample theory* (Springer Science & Business Media, 1999).
9. Hoff, P. D. *A first course in Bayesian statistical methods* (Springer Science & Business Media, 2009).
10. Kay, S. M. Fundamentals of statistical signal processing, volume I: estimation theory (1993).
11. Kutner, M., Nachtsheim, C. & Neter, J. *Applied Linear Regression Models* ISBN: 9780072955675. https://books.google.com/books?id=0nAMAAAACAAJ (McGraw-Hill Higher Education, 2003).
12. Seber, G. A. & Lee, A. J. *Linear regression analysis* (John Wiley & Sons, 2012).
13. Johnson, R. & Wichern, D. *Applied Multivariate Statistical Analysis* ISBN: 9780131877153. https://books.google.com/books?id=gFWcQgAACAAJ (Pearson Prentice Hall, 2007).
14. McCulloch, C. E. & Neuhaus, J. M. *Generalized linear mixed models* (Wiley Online Library, 2001).
15. Wilcox, R. R. *Fundamentals of modern statistical methods: Substantially improving power and accuracy* (Springer Science & Business Media, 2010).

16. Forbes, C., Evans, M., Hastings, N. & Peacock, B. *Statistical distributions* (John Wiley & Sons, 2011).

17. Krishnamoorthy, K. *Handbook of statistical distributions with applications* (CRC Press, 2016).

5

STOCHASTIC PROCESS

5.1 Stochastic process

5.1.1 Basic definition and concepts

Definition 5.1.1 (stochastic process). *[1] A stochastic process X is a collection of random variables $\{X_t\}_{t\in T}$ on some fixed probability triple $(\Omega, \mathcal{F}, P)$, indexed by a subset T of the real numbers.*

*If the index set is the positive integers, we call X a **discrete-time stochastic process**. If the T is an open interval on $\mathbb{R}$, it is called a **continuous-time stochastic process***

Definition 5.1.2 (sample path). *[2, p. 45] For a discrete-time stochastic process, the sequence of numbers $X_1(\omega), X_2(\omega), ...$ for any fixed $\omega \in \Omega$ is called a sample path. For continuous stochastic process, the mapping*

$$t \in T \to X_t(\omega) \in \mathbb{R}$$

is a sample path.

Remark 5.1.1 (interpretation). A stochastic process involves two variables, $t \in T, \omega \in \Omega$. For each fixed t, the mapping

$$\omega \in \Omega \to X_t(\omega) \in \mathbb{R}$$

is a random variable, and for each fixed ω, the mapping

$$t \in T \to X_t(\omega) \in \mathbb{R}$$

is a sample path.

Remark 5.1.2 (sample path examples).

- One trivial case is that $X_1, X_2, ...$ are the same mapping from sample space, then the sample path associated with a ω will be a horizontal line. However, if X_1, X_2 are different mapping from the sample space, then the sample path will not be a horizontal line.
- For a non-trivial case: consider $X_t(\omega) = Z(\omega)\sin(t)$. If $Z(\omega) = 0.5$, then $X_t = 0.5\sin(t)$
- For another non-trivial case: consider $X_t(\omega) = \omega^t$ assuming $\Omega = [0, 1]$

Note 5.1.1 (interpretation on sample space and σ algebra). [3, p. 97] Use random walk as example.

- Let $\omega \in \Omega$. One way to think of ω is as the random sample path. A random experiment is performed, and its outcome is the path of the random walk of horizon T. This random experiment outcome can be thought as a long sequence coin-toss outcome such that we map this long sequence coin-toss outcome to a random walk path, a function parameterized by time. See Figure 5.1.1.

- If time index is from 0 to T, then total number of sample points in Ω is 2^T.
- Some example random events in Ω are: (1) coin-toss sequences starting with H; (2) coin-toss sequence starting with HT.
- Then the σ-algebra is the σ-algebra on the sample-path space such that some 'suitable' subsets of all possible paths can be evaluated. For example, we can evaluate $P(W_t < 0.5)$ for some $t \geq 0$.

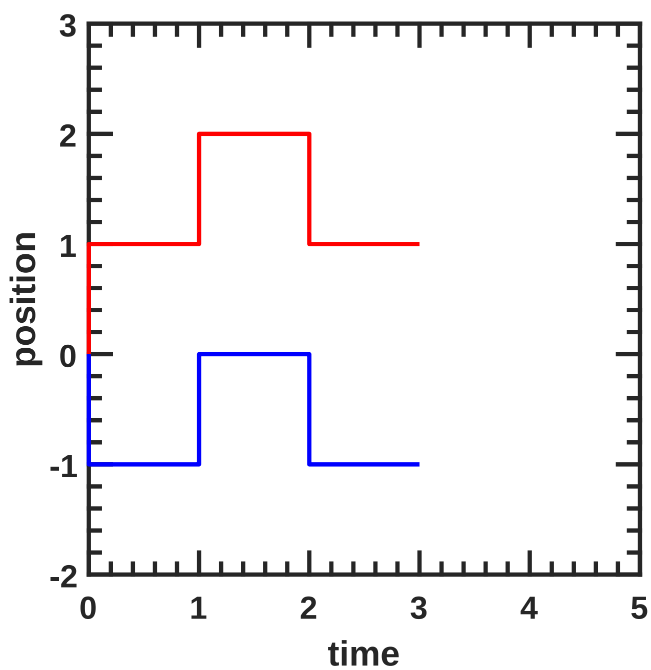


Figure 5.1.1: An illustration of a random walk mapping a sample point, ω, to a trajectory parameterized by time, where red trajectory sample point HHT, and blue trajectory has sample point THT.

5.1.2 Filtration and adapted process

Definition 5.1.3 (filtration). *The collection* $\{\mathcal{F}_t, t \geq 0\}$ *of* σ*-field on sample space* Ω *is called a filtration if*

$$\mathcal{F}_s \subseteq \mathcal{F}_t, \forall 0 \leq s \leq t.$$

Remark 5.1.3. A filtration represents an increasing stream of information.

Definition 5.1.4 (adapted process). *[4, p. 77]Consider a stochastic process $\{X_t\}_{t\in I}$ with a filtration $\{\mathcal{F}_t\}_{t\in I}$ on its σ field. The process is said to be **adapted to the filtration** $\{\mathcal{F}_t\}_{t\in I}$ if the random variable X_t is $\mathcal{F}_t$ measurable for all $t \in I$, or equivalently, $\sigma(X_t) \subseteq \mathcal{F}_t$.*

Remark 5.1.4. [5]

- Examples of 'non-adapted' process. Consider a stochastic process X with $I = \{0, 1\}$. Let $\mathcal{F}_0, \mathcal{F}_1$ be σ field generated by X_0, X_1. And $\mathcal{F}_0$ and $\mathcal{F}_1$ are independent to each other, i.e. $\mathcal{F}_0 \nsubseteq \mathcal{F}_1$.
- For a discrete stochastic process $\{X_n\}$, let $F_n = \sigma(X_0, X_1, ..., X_n)$, then $\{X_n\}$ is an adapated process. Here $\sigma(X_0, X_1, ..., X_n)$ is the smallest σ algebra on Ω such that $X_0, X_1, ..., X_n$ is measurable.

Remark 5.1.5.

5.1.3 Natural filtration of a stochastic process

Definition 5.1.5 (natural filtration generated by a stochastic process). *Let (S, Σ) be a measurable space. Let X_t be a stochastic process such that $X : I \times \Omega \to S$, then natural filtration of $\mathcal{F}$ with respect to X is the filtration $\{\mathcal{F}_t\}_{t\in I}$ given by*

$$\mathcal{F}_t = \sigma(X_t^{-1}(A)|s \in I, s \leq t, A \in \Sigma)$$

here σ is the σ field generation operation. Or equivalently, we write

$$\mathcal{F}_t = \sigma(X_s, s \leq t).$$

Remark 5.1.6.

- In discrete setting, we have $\mathcal{F}_n = \sigma(X_0, X_1, ..., X_n)$.
- **Any stochastic process X_t is an adapted process with respect to its natural filtration $\mathcal{F}_t$ because** $\sigma(X_t) \subseteq \mathcal{F}_t$.

Remark 5.1.7 (interprete natural filtration). [6, p. 43]

- Let the symbol $\mathcal{F}_t^X$ denotes the σ-algebra (i.e., information) generated by X_t on the interval $[0, t]$, or alternatively 'what has happened to X over the interval $[0, t]$'. Note that $\mathcal{F}_t^X$ is one element in the natural filtration.
- (**interpretation of adaptivity**) Informally, if, based upon observations of the trajectory $\{X(s); 0 \leq s \leq t\}$, it is possible to decide whether a given **event** A has occurred or not, then we write this as $\sigma(A) \in F_t^X$, or say that 'A is F_t^X-measurable'.
- If the value of a given **random variable** Z can be completely determined by given observations of the trajectory $\{X(s); 0 \leq s \leq t\}$, then we also write $\sigma(Z) \in F_t^X$.

- If Y_t is a stochastic process such that we have $\sigma(Y(t)) \in \mathcal{F}_t^X, \forall t \geq 0$, then we say that Y is adapted to the filtration $\{\mathcal{F}_t^X, t \geq 0\}$.

We have the following simple examples:

- If we define the event A by $A = \{X(s) \leq 3.14, \forall s \leq 9\}$, then we have $A \in \mathcal{F}_9^X$.
- For the event $A = \{X(10) > 8\}$, we have $A \in \mathcal{F}_{10}^X$ but not $A \notin \mathcal{F}_9^X$ since it is impossible to decide A has occured or not based on the trajectory of X_t over the interval $[0, 9]$.
- For the random variable Z defined by

$$Z = \int_0^5 X(s)ds,$$

we have $\sigma(Z) \in \mathcal{F}_5^X$.

Example 5.1.1 (Trivial adaptive process: single Bernoulli experiment). Consider a stochastic process $\{X_n\}$ represents a single toss experiment. We then have a trivial adapted process by defining $\mathcal{F}_1 = \mathcal{F}_2 = ... = \mathcal{F}_n = \mathcal{F} = \sigma(X_1)$. For this filtration, the stochastic process $Z_n = \sum_{i=1}^n X_i$ is not adapted to it.

Example 5.1.2 (Infinite coin toss process(infinite Bernoulli experiments)). Consider the probability space for tossing a coin infinitely many time. We can define the sample space as $\Omega_\infty =$ the set of infinite sequences of Hs and Ts. A generic element of Ω_∞ will be denoted as $\omega = \omega_1\omega_2....$, where ω_n indicates the result of the nth coin toss.

We can define a stochastic process $\{X_n\}, X_n = f(W_1, W_2, ..., W_n)$, and its filtration $\mathcal{F}_n = \sigma(W_1, W2, ..., W_n)$. Then every X_n is $\mathcal{F}_n$ measurable. A simple event in $\mathcal{F}_n$ is the random experiment value of $W_1, W_2, ..., W_n$. Note that as n increase, $\mathcal{F}_n$ becomes finer and finer, and $\mathcal{F}_n$ can measure any previous $X_m, m < n$.

Remark 5.1.8 (σ algebra for a stochastic process). From 5.1.1, we know that $\mathcal{F}$ is the σ algebra for the set of all possible sample paths. And $\mathcal{F}_t$ can be viewed as the σ algebra for the set of all possible sample paths upto t.

5.1.4 Continuity of sample path

Definition 5.1.6 (continuity of sample path). *A stochastic process with almost all sample paths continuous is called a continuous process. Similarly, a stochastic process is said to be right-continuous if almost all of its sample paths are right-continuous functions.*

Example 5.1.3.

- The Brownian motion is a stochastic process with continuous sample path.
- The white noise process has discontinuous sample path.
- The Poisson process has discontinuous sample path.

Definition 5.1.7 (right-continuous with left limit, cadlag).

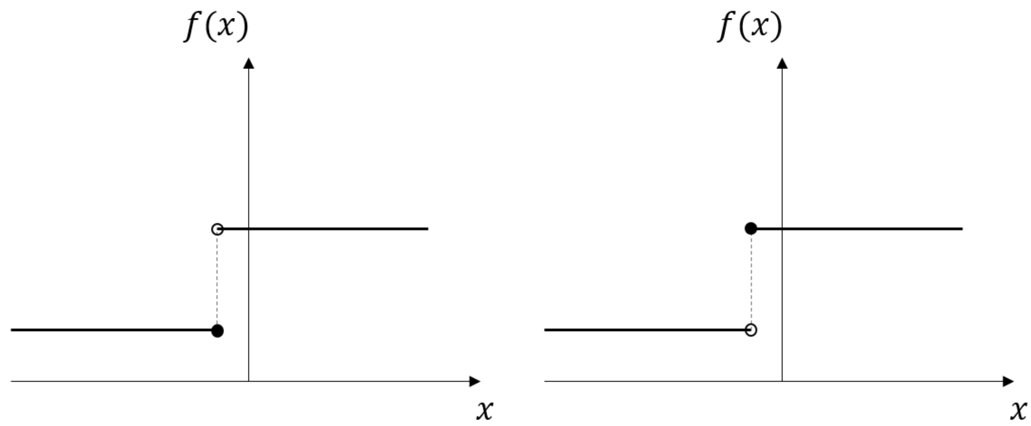
- *A sample path* $X_t : [0,\infty) \to \mathbb{R}$ *is called a right-continuous with left limit if for every* $t \in [0,\infty)$ *if*
  - *the left limit* $\lim_{s\to t^-} X(s)$ *exists;*
  - *the right limit* $\lim_{s\to t^+} X(s)$ *exists and* $\lim_{s\to t^+} X(s) = f(t)$.
- *A stochastic process with almost all sample paths being right-continuous with left limit is called a cadlag stochastic process.*

Example 5.1.4. The sample path of a Poisson process is cadlag.

5.1.5 Predictable process

Definition 5.1.8 (predictable process).

- *Given a probability space* $(\Omega, \mathcal{F}, P)$ *and a filtration* $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, *a discrete-time stochastic process* $\{X_n\}_{n\in\mathbb{N}}$ *is* ***predictable*** *if* X_{n+1} *is measurable with respect to* $\mathcal{F}_n$ *for each* n.
- *Consider a probability space* $(\Omega, \mathcal{F}, P)$ *and a filtration* $\{\mathcal{F}_t\}_{t\geq 0}$. *Let* $\mathcal{P}$ *denote the predictable/previsible* σ*-algebra, i.e. the* σ*-algebra generated by all left-continuous process adapted to* $\{\mathcal{F}_t\}_{t\geq 0}$. *If a process* X_t *is measurable with respect to* $\mathcal{P}$ *for all* t, *then* X_t *is a predictable process.*



(a) A left continuous function. (b) A right continuous function.

Figure 5.1.2: An illustration of left and right continuous functions.

Example 5.1.5.

- Every deterministic process is a predictable process.
- Every continuous-time stochasti process with continuous smaple path (e.g., Brownian motion) is adapted to and predictable with respect to its natrual filtration.

Example 5.1.6 (trading strategies as predictable processes).

5.2 Stationary process

5.2.1 Stationarity concepts

Definition 5.2.1 (strictly stationary process). *[7, p. 231][8, p. 30] A random process $\{X_t\}_{t\in T}$ over probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ is called strictly stationary if for any $t_1, t_2, ..., t_k \in T$ and $A_1, ..., A_k \in \mathcal{B}(\mathbb{R})$ the probabilities*

$$P(X_{t_1+t} \in A_1, ..., X_{t_k+t} \in A_k)$$

do not depend on arbitrary t, where $t \in T$.

Example 5.2.1 (a sequence of iid random variables). A sequence of independent identically distributed random variables is a strictly stationary process.

Lemma 5.2.1 (A Markov chain starting from stationary distribution is strictly stationary process). *[7, p. 231] Consider a finite state, irreducible and aperiodic Markov chain characterized by matrix P. Let the initial state distribution be π_0. If the stationary distribution $\pi = \pi_0$, then the Markov chain P is a strictly stationary process.*

Proof. The stationary distribution will exist(**??**). By iteration, we know that the distribution at every time step is π. □

Definition 5.2.2 (weakly stationary process). *[7, p. 209][8, p. 32] A random process $\{X_t\}_{t\in T}$ is called a weakly stationary process if there exist a constant m and $b(t), t \in T$ function, such that*

$$E[X_t] = m, Var[X_{t_1}] = \sigma^2, cov(X_{t_1}, X_{t_2}) = r(t_1 - t_2), \forall t_1 \neq t_2 \in T,$$

where $b(0) = \sigma^2$.

That is, the mean and the covariance structure a weakly stationary process can be fully characterized by a constant mean parameter and a covariance function $r : \mathbb{R} \to \mathbb{R}$.

Lemma 5.2.2 (properties of a covariance function). *[8, p. 35] For a weakly stationary stochastic process, the covariance function $r(t_1 - t_2) \triangleq Cov(X(t_1), X(t_2))$ has the following properties:*

- $r(0) = Var[X(t)] \geq 0, \forall t$
- $Var[X(t+h) \pm X(t)] = E[(X(t+h) \pm X(t))^2] = 2(r(0) \pm r(h))$
- *(even function)*$r(\tau) = r(-\tau)$.

- $|r(\tau)| \leq r(0)$.
- *If* $|r(\tau)| = r(0)$*, for some* $\tau \neq 0$*, then* r *is periodic. In particular,*
  - *If* $r(\tau) = r(0)$*, then* $X(t+\tau) = X(t), \forall t$.
  - *If* $r(\tau) = -r(0)$*, then* $X(t+\tau) = -X(t) = X(t-\tau), \forall t$*(periodicity of* 2π*).*
- *If* $r(\tau)$ *is continuous for* $\tau = 0$*, then* $r(\tau)$ *is continuous everywhere.*

Proof. (1)(2)(3) Straight forward.

(4) Use Cauchy inequality for random variables(Theorem 1.10.4)

$$E[|X(t)-\mu||X(t+\tau)-\mu|] \leq \sqrt{Var[X(t)]Var[X(t+\tau)]} = \sqrt{r(0)^2} = r(0).$$

(5)

(a) If $r(\tau) = r(0)$, then from (2) we have $E[(X(t+\tau)-X(t))^2] = 0$. It can be showed via contradiction that having $E[(X(t+\tau)-X(t))^2] = 0$ implies $X(t+\tau) = X(t)$(that is the two maps are exactly the same). (b) If $r(\tau) = -r(0)$, then from (2) we have $E[(X(t+\tau)+X(t))^2] = 0$. It can be showed via contradiction that having $E[(X(t+\tau)+X(t))^2] = 0$ implies $X(t+\tau) = -X(t)$(that is the two maps are exactly the same). (6) For any t, consider

$$\begin{aligned}(r(t+h)-r(t))^2 &= (Cov(X(0), X(t+h)) - Cov(X(0), X(t)))^2 \\ &= (Cov(X(0), X(t+h)-X(t)))^2 \\ &\leq Var[X(0)]Var[X(t+h)-X(t)] = 2r(0)(r(0)-r(h))\end{aligned}$$

If $h \to 0$, then $r(0) - r(h) \to 0^+$ due to the continuity of $r(\tau)$ at $\tau(0)$,which implies $r(t+h) \to r(t)$(that is, $r(t)$ is continuous for any t). □

Lemma 5.2.3 (a strictly stationary process is a weakly stationary process). *A strictly stationary process* X_t *will be a weakly stationary process.*

Proof. For mean, use translation invariant property of of marginal distribution. For covariance, use translational invariant property of two variable joint distribution. □

Lemma 5.2.4. *A weakly stationary Gaussian process is a strictly stationary Gaussian process.*

Proof. To a process is strictly stationary, we need to show $(X_{t_1+\tau}, X_{t_2+\tau}, ..., X_{t_n+\tau}), \forall \tau \in \mathbb{R}$ has the same distribution as $(X_{t_1+\tau}, X_{t_2+\tau}, ..., X_{t_n+\tau})$. Because the full distribution of a multivariate Gaussian can be constructed from its pair distribution(Lemma 2.2.12), we only need to show that $(X_{t_1+\tau}, X_{t_2+\tau}), \tau \in \mathbb{R}$ has the same distribution as (X_{t_1}, X_{t_2}). From weak stationarity, we know that the mean vector and covariance matrix are the same; that is, their joint distribution are the same. Note that for a Gaussian distribution, mean and covaraince matrix fully determines the joint distribution. □

5.2.2 Random phase and amplitude

Definition 5.2.3 (random harmonic function). *[8, p. 35] Let $A > 0$ and ϕ be independent random variable, with ϕ uniform over $[0, 2\pi]$. The stochastic process $X(t)$ defined by*

$$X(t) = A\cos(2\pi f_0 t + \phi).$$

Lemma 5.2.5 (random harmonic function). *[8, p. 36] The random harmonic function $X(t) = A\cos(2\pi f_0 t + \phi)$ is a **stationary process** if A and ϕ are independent and ϕ is uniformly distributed in $[0, 2\pi]$. Then*

- $E[X(t)] = 0$
- $Var[X(t)] = \frac{1}{2}E[A^2] = \sigma^2$
- $r(\tau) = \sigma^2\cos(2\pi f_0 \tau)$

Proof. (1)

$$E[X(t)] = E[A\cos(2\pi f_0 t + \phi)] = E[A]E[\cos(2\pi f_0 t + \phi)]$$

$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)] &= \int_0^{2\pi} \cos(2\pi f_0 t + x)\frac{1}{2\pi}dx \\ &= \frac{1}{2\pi}(\sin(2\pi f_0 t + 2\pi) - \sin(2\pi f_0 t + x)) \\ &= 0. \end{aligned}$$

(2)

$$Var[X(t)] = E[X(t)^2] = E[A^2\cos(2\pi f_0 t + \phi)^2] = E[A^2]E[\cos(2\pi f_0 t + \phi)^2].$$

Note that

$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)^2] &= \int_0^{2\pi} \cos(2\pi f_0 t + x)^2\frac{1}{2\pi}dx \\ &= \frac{1}{2\pi}(\pi) \\ &= 1/2. \end{aligned}$$

(3)

$$Cov(X(t), X(s)) = E[X(t)X(s)] = E[A^2]E[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 s + \phi)].$$

Note that

$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 s + \phi)] &= \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}(\cos(2\pi f_0(s+t) + 2x) + \cos(2\pi f_0(s-t)))dx \\ &= \frac{1}{2}\cos(2\pi f_0(s-t)) \end{aligned}$$

□

Lemma 5.2.6 (linear supposition of random harmonic function). *[8, p. 38] The linear supposition of random harmonic functions*

$$X(t) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)$$

with deterministic constants $A_0, A_1, ..., A_k$ *and independent uniform random variables* $\phi_1, \phi_2, ..., \phi_k$ *has the following properties:*

- $E[X(t)] = A_0$.
- $r(\tau) = \sigma_0^2 + \sum_{k=1}^{n} \sigma_k^2 \cos(2\pi f_k \tau)$.

Proof. (1)

$$E[A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)] = E[A_0] + 0.$$

(2) Note that

$$\begin{aligned}
&E[\sum_{k=1}^{n} A_j \cos(2\pi f_k t + \phi_k) \sum_{k=1}^{n} A_k \cos(2\pi f_k (t+\tau) + \phi_k)] \\
&= E[\sum_{j=1}^{n} A_j \cos(2\pi f_j t + \phi_j) \sum_{k=1}^{n} A_k \cos(2\pi f_k (t+\tau) + \phi_k)] \\
&= \sum_{k=1}^{n} E[A_k^2] E[\cos(2\pi f_k t + \phi_k) \cos(2\pi f_k (t+\tau) + \phi_k)] \\
&= \sum_{k=1}^{n} \sigma_k^2 \cos(2\pi f_k \tau)
\end{aligned}$$

□

Lemma 5.2.7 (power and average power). *[8, p. 39] Consider the linear supposition of random harmonic functions*

$$X(t) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)$$

with deterministic constants $A_0, A_1, ..., A_k$ *and independent uniform random variables* $\phi_1, \phi_2, ..., \phi_k$.

Then the average power is given by

$$E_T = \lim_{T\to\infty} \frac{1}{T}\int_0^T X(t)^2 dt = A_0^2 + \frac{1}{2}\sum_{k=1}^n A_k^2.$$

Proof. Note that

$$\begin{aligned}
E_T &= \frac{1}{T}\int_0^T X(t)^2 dt \\
&= \frac{1}{T}\{TA_0^2 + 2\sum_{k=1}^n A_0 A_k \int_0^T \cos(2\pi f_k t + \phi_k) dt \\
&+ \sum_{k=1}^n A_k^2 \int_0^T \cos^2(2\pi f_k t + \phi_k) dt \\
&+ 2\sum_{k=1}^n \sum_{l=k+1}^n A_k A_l \int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l) dt\}.
\end{aligned}$$

Further note that:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos(2\pi f_k t + \phi_k) dt = 0$$

since $\int_0^T \cos(2\pi f_k t + \phi_k) dt$ is bounded.

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l) dt = 0$$

since $\int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l) dt$ is bounded.

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos^2(2\pi f_k t + \phi_k) dt = \frac{1}{2}$$

since $\int_0^T \cos^2(2\pi f_k t + \phi_k) dt = \frac{T}{2}$ is bounded.

□

Remark 5.2.1 (average power is independent of sample path)**.** Note that the integral is independent of $\phi_1, ..., \phi_k$. That is, no matter what values $\phi_1, ..., \phi_k$ are taking in a sample path, we get the same average power.

5.2.3 Ergodicity

5.2.4 Spectral analysis

Definition 5.2.4 (spectral distribution, spectral density). *If a stationary stochastic process X_t has autocovariance function $\gamma(\tau)$ satisfying $\int_{-\infty}^{\infty} |\gamma(\tau)| \, d\tau$, then we define the* ***spectral density function*** *as*

$$f(\omega) = \int_{-\infty}^{\infty} \gamma(\tau) \exp(-2\pi i \omega \tau) d\tau,$$

where $-\infty < \omega < \infty$.

Lemma 5.2.8 (properties of spectral density function).

- *$f(\omega)$ is real-valued.*
- *$f(\omega) > 0$.*
- *$f(\omega)$ is even.*
- *$f(\omega)$ is periodic with period 1.*
- *$\gamma(\tau) = \int_{=1/2}^{1/2} f(\omega) \exp(2\pi i \omega \tau) d\omega$.*

Theorem 5.2.1 (Wiener-Khintchine Theorem).

5.2.5 Monte carlo simulation

5.3 Gaussian process and Finite dimension distributions

5.3.1 One-dimensional Gaussian process

5.3.1.1 *Definitions and properties*

Definition 5.3.1 (One-dimensional Gaussian process). *A stochastic process $\{X_t\}_{t\in T}$ is Gaussian process if for any $t_1, t_2, ..., t_n \in T$, the joint distribution on $(X_{t_1}, ..., X_{t_n})$ is Gaussian, i.e.,*

$$p(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|det\Sigma|^{1/2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)).$$

Definition 5.3.2 (One-dimensional Gaussian process, alternative). *[9, p. 19] A stochastic process $\{X_t\}_{t\in T}$ is Gaussian process if every linear combination*

$$S = \sum_k a_k x(t_k), t_k \in T, a_k \in \mathbb{R}$$

has a Gaussian distribution.

Remark 5.3.1 (equivalence of two definitions).

- The first definition can imply the second by affine transformation Theorem 2.2.1.
- The second can imply the first by using the multivariate Gaussian definition via linear combination(Lemma 2.2.8).

Lemma 5.3.1 (affine transformation of a Gaussian process is a Gaussian process). *Let X_t be a Gaussian process, then $aX_t + b, a, b \in \mathbb{R}$ is also a Gaussian process.*

Proof. Directly from definitions. □

Note 5.3.1 (Gaussian processes can be stationary or non-stationary). A Gaussian process can be stationary(white noise process) or non-stationary(Wiener process).

5.3.1.2 *Stationarity*

Lemma 5.3.2. *If $X(t)$ is a stationary Gaussian process with mean m and covariance function $r(\tau)$. Then*

- *For all t, $X(t) \sim N(m, r(0)^2)$.*
- *For all t_1, t_2, $(X(t_1), X(t_2)) \sim MN(\mu, \Sigma)$, where*

$$\mu = \begin{bmatrix} m \\ m \end{bmatrix}, \Sigma = \begin{bmatrix} r(0) & r(t_1 - t_2) \\ r(t_1 - t_2) & r(0) \end{bmatrix}$$

Proof. Straight forward. □

5.3.1.3 *Examples*

Example 5.3.1 (white noise process). A white noise process W_t is a Gaussian process with zero mean and $cov(W_t, W_s) = \sigma^2 \delta(s - t)$.

Example 5.3.2 (a discrete random walk is not a Gaussian process). A random walk B_n is not a Gaussian process. For example, B_1 is a Bernoulli distribution, not a Gaussian.

Example 5.3.3 (Ornstein-Uhlenbeck process is stationary Gaussian process in the long run). An OU process is a Gaussian process. As $t \to \infty$, the OU process becomes a stationary Gaussian process(Lemma 6.4.1).

Example 5.3.4 (a Wiener process(Brownian motion) is a Gaussian process). From Lemma 6.1.1, a Wiener process is the integral of a white noise Gaussian process. It is not stationary, but it has stationary increments.

Example 5.3.5 (a geometric Brownian motion is not a Gaussian process). Let X_t be a geometric Brownian motion process, then X_t is not Gaussian, thus not a Gaussian process.

Example 5.3.6 (a stable AR(1) process). A stable AR(1) process of X_k can be written as

$$X_k = \sum_{i=0}^{\infty} \beta^i W_{k-i},$$

where $W_k = w(t_k)$ is the discrete sampling of wiener process $w(t)$.

Because any linear combination of samples of a Gaussian process $w(t)$ is a normal random variable, X_k has a normal distribution.

5.3.2 finite dimensional distribution

Definition 5.3.3. *The finite dimensional distribution of a stochastic process X is the joint distribution of $X_{t_1}, X_{t_2}, ..., X_{t_k}$, where $t_1, t_2, ..., t_k \in \mathcal{I}$, we represent the probability measure as*

$$\mu_{t_1,t_2,...,t_k}(F_1 \times F_2 \times ...F_k) = P(X_{t_1} \in F_1 ... X_{t_k} \in F_k)$$

where F_i are measurable events/subsets in $\mathbb{R}$.

Remark 5.3.2 (purpose). We usually characterize a stochastic process by studying the joint distribution of a finite set of random variables in the stochastic process.

consistence of finite-dimensional distributions

Given a set of finite-dimensional distributions on different index set, to make sure that these fdd are derived from the same stochastic process, **we require these fdd to be consistent based on two criteria:**[10, lec 4]

- Invariant under permutations, i.e., two probability measure on two different set of $X_1, X_2, ..., X_k$ and $X_{\pi(1)}, .., X_{\pi(k)}$, then the measure should be the same given the same measurable subset.

- Marginal distribution are consistent,i.e., marginalizing different probability measure to the marginal measure to the same subset of random variables should be the same.

Theorem 5.3.1 (Kolmogorov Extension theorem). *If the family of measures $\{\mu_{t_1,...,t_k}\}$ satisfies the consistent condition, then there exists a stochastic process with the corresponding finite-dimensional distribution.*

5.3.3 Gaussian process generated by Brownian motin

Lemma 5.3.3 (Gaussian process stochastic differential equation). *A stochastic process X_t governed by*

$$dX_t = a(t)dt + b(t)dW_t,$$

where W_t is a Wiener process, is a Gaussian process.

Proof. It can be showed that $X(t) \sim N(\mu(t), \int_0^t b(s)^2 ds)$(Lemma 6.3.9). Also note that the increment is independent and Gaussian; that is $X(t_1) - X(t_2)$ is independent of $X(t_2) - X(t_3), t_1 > t_2 > t_3$. Therefore, the random vector $(X(t_1), X(t_2), ..., X(t_n))$ is multivariate normal since it can be constructed by affine transformation of $(X(t_1) - X(t_2), X(t_2) - X(t_3), ..., X(t_n)$(Theorem 2.2.1). □

Theorem 5.3.2 (linear combination of multiple Brownian-motion-generated Gaussian processes is a Gaussian process). *Consider N stochastic processes generated by M Brownian motions, given by*

$$dX_i(t) = \mu_i(t)dt + \sum_{j=1}^{M} \sigma_{ij}(t)dW_j,$$

where $W_1, W_2, ..., W_M$ are independent Brownian motion, $\mu_i(t), \sigma_{ij}(t)$ are state-independent deterministic function of t.

Then

- *the joint distribution of $X_1, X_2, ..., X_N$ is multivariate Gaussian.*
- *for any linear combination of $X_1(t), X_2(t), ..., X_N(t)$, given by*

$$Y(t) = \sum_{i=1}^{M} a_i X_i(t), a_i \in \mathbb{R},$$

$Y(t)$ is a Gaussian process.

Proof. (1) We only show a zero drift 2 by 2 case. Consider

$$X_1 = \int_0^t \sigma_{11}(s)dW_1 + \int_0^t \sigma_{12}(s)dW_2, X_2 = \int_0^t \sigma_{21}(s)dW_1 + \int_0^t \sigma_{22}(s)dW_2.$$

Denote

$$A = \int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))dW_1(s)$$
$$B = \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))dW_2(s).$$

And we can see immediately that

$$E[A] = E[B] = 0, Var[A] = \int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))^2 ds, Var[B] = \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))^2 ds.$$

More important, $A + B$ is a Gaussian random variable.

To show (X_1, X_2) is joint Gaussian, we can check its mgf, given by

$$\begin{aligned}
\phi(\lambda_1, \lambda_2) &= E[\exp(\lambda_1 X_1 + \lambda_2 X_x)] \\
&= E[\exp(A + B)] \\
&= E[\exp(E[A + B] + \frac{1}{2}Var[A + B])] \\
&= E[\exp(\frac{1}{2}(Var[A] + Var[B]))] \\
&= E[\exp(\frac{1}{2}(\int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))^2 ds + \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))^2 ds))]
\end{aligned}$$

where we eventually will get a quadratic form of λ_1 and λ_2. Then using Lemma 2.2.7, we can show (X_1, X_2) are joint normal.

We can similarly prove cases containing multiple variables and drifting terms.

(2) Directly use affine transformation of multivariate Gaussian vector. □

Corollary 5.3.2.1. *Consider N stochastic processes generated by M Brownian motions, given by*

$$dX_i(t) = \mu_i(t)dt + \sigma_i(t)dW_j,$$

where $W_1, W_2, ..., W_M$ are ***correlated Brownian motions*** *such that $E[dW_i dW_j] = \rho_{ij}dt$, $\mu_i(t), \sigma_{ij}(t)$ are state-independent deterministic function of t.*

Then

- *the joint distribution of $X_1, X_2, ..., X_N$ is multivariate Gaussian.*
- *for any linear combination of $X_1(t), X_2(t), ..., X_N(t)$, given by*

$$Y(t) = \sum_{i=1}^{M} a_i X_i(t), a_i \in \mathbb{R},$$

$Y(t)$ is a Gaussian process.

Proof. use the fact that the two forms are equivalent(6.3.1). □

Remark 5.3.3 (Caution!). Note that if $X_1, X_2, ..., X_N$ are more general Gaussian processes not generated by Brownian motion, then Y is not necessarily Gaussian, since $X_1, X_2, ..., X_N$ are not necessarily joint normal.

5.4 Markov process

Definition 5.4.1 (Markov process). *[3]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t)$ be a filtration of $\mathcal{F}$. Consider an adapated stochastic process $X(t)$, assume that for all $0 \leq s \leq t$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g such that*

$$E[f(X(t))|\mathcal{F}(s)] = g(X(s))$$

then we say $X(t)$ is a Markov process.

Definition 5.4.2 (Alternative definition). *[11]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t)$ be a filtration of $\mathcal{F}$. Consider an adapated stochastic process $X(t) : \Omega \to S$,for all $0 \leq s \leq t$, for each $A \in \mathcal{S}$, then*

$$P(X_t \in A|\mathcal{F}(s)) = P(X_t \in A|X_s)$$

A Markov process is stochastic process satisfies above Markov property with respect to its natural filtration.

Remark 5.4.1.

- The above two definition emphasizes that in a Markov process, only *the most immediate history* is useful when we make predictions on the future based on the past history.
- For discrete-time Markov chains, we have alternative definition as

$$P(X_n = x_n|X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_n = x_n|X_{n-1} = x_{n-1}).$$

- A martingale is not necessarily a Markov process. See link.
- Examples of non-Markovian include long memory auto-regressive processes.

Remark 5.4.2 (Gaussian processes, stationary processes and Markov processes are different characterization of stochastic process.)**.** We have:

- A Gaussian process is not necessarily a Markov process.
- A stationary is not necessarily a Markov process.

5.5 Martingale theory

5.5.1 Basics

Definition 5.5.1 (martingale). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on $\mathcal{F}$. Let X_t be a stochastic process. X_t is called a $\mathcal{F}_t$-martingale, if*

- *X_t is adapted to $\{\mathcal{F}_t\}$;*
- *$E[|X(t)|] < \infty, \forall t$;*
- *$E[X_t|\mathcal{F}_s] = X_s$ almost surely, for all $0 \leq s \leq t$.*

Definition 5.5.2 (supermartingale, submartingale). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on $\mathcal{F}$. Let X_t be a stochastic process.*

- *X_t is called a $\mathcal{F}_t$-supermartingale, if X_t is adapted to $\{\mathcal{F}_t\}$;$E[|X(t)|] < \infty, \forall t$; $E[X_t|\mathcal{F}_s] \leq X_s$ almost surely, for all $0 \leq s \leq t$.*
- *X_t is called a $\mathcal{F}_t$-submartingale, if X_t is adapted to $\{\mathcal{F}_t\}$;$E[|X(t)|] < \infty, \forall t$; $E[X_t|\mathcal{F}_s] \geq X_s$ almost surely, for all $0 \leq s \leq t$.*

Remark 5.5.1.

- Martingale is always an adapted process with respect to some filtration.
- Note that for discrete setting, we have $E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$.

Definition 5.5.3 (discrete-time martingale). *[2, p. 49] A sequence $X_1, X_2, ...$ of random variables is called a martingale with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, ...$ if*

1. *$E[|X_n|] < \infty$;*
2. *$X_1, X_2, ...$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, ...$;*
3. *$E[X_{n+1}|\mathcal{F}_n] = X_n$*

Example 5.5.1 (Sum of independent zero-mean RVs as martingale). Let $X_1, X_2, ...$ be a sequence of independent integrable RVs with $E[|X_k|] < \infty$, and

$$E[X_k] = 0, \forall k.$$

Define

$$S_n = \sum_{i=1}^{n} X_i,$$

such that

$$E[|S_n|] = E[|X_1 + X_2 + ... + X_n|] \leq E[|X_1| + E[|X_2|] + ... + E[|X_n|]] < \infty;$$

and

$$\mathcal{F}_n = \sigma(X_1, X_2, ..., X_n), \mathcal{F}_0 = \{\emptyset, \Omega\}$$

Then the sequence $S_1, S_2, ..., S_n$ is a martingale with respect to $\mathcal{F}_1, \mathcal{F}_2, ...$ Note that a simple event in $\mathcal{F}_n$ should specify the value of $X_1, X_2, ..., X_n$, otherwise we cannot measure S_n.

Definition 5.5.4 (continous-time martingale). *[2, p. 49] A sequence stochastic process X_t is called a martingale with respect to a filtration $\{\mathcal{F}_t\}$ if*

1. $E[|X_t|] < \infty$;
2. $\{X_t\}$ *is adapted to* $\{\mathcal{F}_t\}$;
3. $E[X_t|\mathcal{F}_s] = X_s, s \leq t$.

Lemma 5.5.1 (martingales have constant expectation).

- *A discrete-time martingale X_n has the property that its expectation $E[X_t]$ is constant $E[X_1]$.*
- *A continuous-time martingale X_t has the property that its expectation $E[X_t]$ is constant $E[X_0]$.*

Proof. From property (2), using iterated expectation($E[E[X|\mathcal{F}]] = E[X]$,subsection 1.8.4), we can have $E[X_{n+1}] = E[E[X_{n+1}|\mathcal{F}_n]] = E[X_n] = ... = E[X_1]$. □

Theorem 5.5.1 (conditional expectation process as Martingale). *Let $(\Omega, P, \mathcal{F})$ be a probability space, and let $\{\mathcal{F}_t\}$ be a filtration on $(\Omega, P, \mathcal{F})$. Let Z be a random variable defined on $(\Omega, P, \mathcal{F})$.*

Define $Z(t) = E[Z|\mathcal{F}_t]$, then $Z(t)$ is a martingale with respect to $\mathcal{F}_t$.

Proof.

$$E[Z(t)|\mathcal{F}_s] = E[E[Z|\mathcal{F}_t]|\mathcal{F}_s] = Z(s).$$

□

5.5.2 Martingales with continuous path

Definition 5.5.5 (continuous martingale). *A martingale $M_t$with respect to $\mathcal{F}_t$ is a continuous martingale if almost all sample paths are continuous.*

Example 5.5.2.

- The Brownian motion is a stochastic process with continuous sample path.
- The white noise process has discontinuous sample path.
- The Poisson process has discontinuous sample path.

Definition 5.5.6 (right-continuous with left limit, cadlag).

- *A sample path $X_t : [0,\infty) \to \mathbb{R}$ is called a right-continuous with left limit if for every $t \in [0,\infty)$ if*
  - *the left limit $\lim_{s\to t^-} X(s)$ exists;*
  - *the right limit $\lim_{s\to t^+} X(s)$ exists and $\lim_{s\to t^+} X(s) = f(t)$.*
- *A martingale with almost all sample paths being right-continuous with left limit is called a cadlag martingale.*

Example 5.5.3. The sample path of a compensated Poisson process(Lemma 7.1.9) is cadlag.

5.5.3 Exponential martingale

Lemma 5.5.2 (Exponential martingale).

- *Let $W(t)$ be the Wiener process, define $Z(t) = \exp(\sigma W(t) - \frac{1}{2}\sigma^2 t)$. Then $Z(t)$ is martingale; moreover, $E[Z(t)] = E[Z(0)] = 1$.*
- *Let $X \sim N(0,\sigma^2)$, then*
$$E[\exp(\pm X - \frac{1}{2}\sigma^2)] = 1.$$
- *Let $X \sim MN(0,\Sigma), \Sigma \in \mathbb{R}^{n\times n}, t \in \mathbb{R}^n$, then*
$$E[\exp(\pm t^T X - \frac{1}{2}t^T \Sigma t)] = 1.$$
- *Let $W(t)$ be a n-dimensional correlated Wiener process with $dW(t)dW(t)^T = \rho(t)dt \in \mathbb{R}^{n\times n}$. Let $\{\mathcal{F}_t\}$ be the filtration generated by $W(t)$. Let $\theta(t) \in \mathbb{R}^n$ be a process adapted to $\mathcal{F}_t$. It follows that*
$$Z(T) = \exp(\pm \int_0^T \theta(u)^T dW(u) - \frac{1}{2}\int_0^T \theta(u)^T \rho(u)\theta(u)du)$$
is martingale such that

$$Z(t) = E[Z(T)|\mathcal{F}_t], E[Z(t)] = E[Z(T)] = 1.$$

Proof. (1) (a)

$$\begin{aligned} E[Z(t)|\mathcal{F}_s] &= E[\exp(\sigma(W(t)-W(s))\exp(\sigma W(s)-\frac{1}{2}\sigma^2 t)|\mathcal{F}_s] \\ &= E[\exp(\sigma(W(t)-W(s))|\mathcal{F}_s]\exp(\sigma W(s)-\frac{1}{2}\sigma^2 t) \\ &= \exp(\frac{1}{2}\sigma^2(t-s))\exp(\sigma W(s)-\frac{1}{2}\sigma^2 t) \\ &= Z(s) \end{aligned}$$

where we use by the fact that

$$E[\exp(\sigma(W(t)-W(s))|\mathcal{F}_s] = \int \exp(\sigma x)f(x)dx = \exp(\frac{1}{2}\sigma^2(t-s)), X \sim N(0,(t-s)).$$

To calculate the expectation, we have

$$E[Z(t)] = \exp(-1/2\sigma^2 t)E[\exp(\sigma W(t))] = \exp(-1/2\sigma^2 t)M_X(\sigma\sqrt{t}) = 1$$

where M_X is the moment generating function of standard normal random variable X. (b) We can also use conclusion from (2). Note that $\sigma W(t) \sim N(0,\sigma^2 t)$.

(2)

$$\begin{aligned} E[\exp(-X-\frac{1}{2}\sigma^2)] &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2})\exp(-x-\frac{1}{2}\sigma^2)dx \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\sigma^2)}{2\sigma^2})dx) \\ &= 1 \end{aligned}$$

(3) Note that $t^T X \sim N(0, t^T\Sigma t)$(Theorem 2.2.1) (4) Introduce

$$Y_t \triangleq \pm\int_0^t \theta(u)^T dW(u) - \frac{1}{2}\int_0^t \theta(u)^T\rho(u)\theta(u)du, Z_t = \exp(Y_t).$$

Then

$$dY_t = -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t),$$

and

$$dZ_t = \exp(Y_t)dY_t + \frac{1}{2}\exp(Y_t)(\theta(t)^T\theta(t))dt = \pm Z_t\theta(t)^T dW(t).$$

Note that $Z_0 = 1$, Z_t can be written in integral form as

$$Z_t = 1 \pm \int_0^t Z_s \theta(s)^T dW(s).$$

Because the expectation of Ito integral is zero, we have

$$E[Z_T] = E[Z_t] = 1.$$

To show $Z_t = E[Z_T|\mathcal{F}_t]$, we have

$$\begin{aligned}
&E[Z_T|\mathcal{F}_t] \\
=&E[Z_t \pm \int_t^T Z_s \theta(s)^T dW(s)|\mathcal{F}_t] \\
=&Z_t
\end{aligned}$$

□

Note 5.5.1 (understanding $\theta(t)$ process). Note that we said $\theta(t)$ is adapted to $\mathcal{F}_t$ generally means that it is governed by SDE

$$d\theta(t) = \mu(\theta, t)dt + \Sigma(\theta)dW(t).$$

This representation includes the following cases.

- $\theta(t)$ is a deterministic process. In this case, we can prove in this way:Note that $\int_0^T \theta(u)^T dW(u) \sim (0, \int_0^T \theta^T(u)\rho(u)\theta(u)du)$ from Lemma 6.3.9. Note that

$$E[(\int_0^T \theta^T dW(u))(\int_0^T \theta^T dW(v))^T] = \int_0^T \int_0^T \theta^T(u)\rho(u)\delta(u-v)\theta(u)dudv.$$

- $\theta(t)$ is a stochastic process driven by $W(t)$. Note that

$$\begin{aligned}
dY_t &= -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t) - \int_0^t (\rho(t)\theta(t))^T d(\theta)dt \pm \int_0^t [d\theta(t)]^T dW(t) \\
&= -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t)
\end{aligned}$$

where we use the fact that $d\theta(t) \sim O((dt)^{1/2})$ to ignore the higher order terms.

Example 5.5.4 (application in finance). Under risk-neutral measure, the stock price is given as

$$S_t = S_0 \exp((r - \sigma^2/2)t + \sigma W_t)$$

where r is risk-free rate, σ is the volatility and W_t is the Brownian motion. It can be showed that $\exp(-rt)S_t = \exp(\sigma W_t - \sigma^2 t/2)$ is an martingale(exponential martingale).

5.5.4 Martingale transformation

Definition 5.5.7 (Predictable/previsible process). *Let $\{Y_t\}$ be a sequence random variables adapted to filtration $\{\mathcal{F}_t\}$. The sequence Y_t is said to be predictable if for every $t \geq 1$, the random variable Y_t is $\mathcal{F}_{t-1}$ measurable, or equivalently, $\sigma(Y_t) \subseteq \mathcal{F}_{t-1}$*

Definition 5.5.8 (Martingale transform). *[4, p. 83] Let $\{X_t\}$ be a martingale, let $\{Y_t\}$ be a predictable sequence. The martingale transform $\{(Y \cdot X)_t\}$ is the*

$$(Y \cdot X)_t = X_0 + \sum_{j=1}^{t} Y_j (X_j - X_{j-1})$$

Lemma 5.5.3 (Martingale transformation is a martingale). *Assume that $\{X_t\}$ is an adapted sequence and $\{Y_t\}$ a predictable sequence, both relative to a filtration $\{\mathcal{F}_t\}$. If $\{X_t\}$ is a martingale, then the martingale transform $\{(Y_t \cdot X_t)\}$ is a martingale with respect to $\{\mathcal{F}_t\}$ if $E[X_j^2] < \infty, \forall j$*

Proof. $E[(Y \cdot X)_t - (Y \cdot X)_{t-1} | cF_{t-1}] = E[Y_t(X_t - X_{t-1}) | \mathcal{F}_{t-1}] = 0$ □

Lemma 5.5.4 (connection to Ito integral). *Let $S_n = X_1 + ... + X_n$ be a random walk, then the new random process*

- $Y_n = \sum_{i=1}^{n} X_{i-1}(X_i - X_{i-1})$ *is a martingale. Moreover,* $E[Y_n] = 0$.
- $Z_n = \sum_{i=1}^{n} f(X_{i-1})(X_i - X_{i-1})$ *is a martingale for any function* $f(y)$*.Moreover,* $E[Z_n] = 0$.

Proof. It is easy to see that X_{i-1} is measurable respect to $\mathcal{F}_i$. Therefore they are martingale transformation and they are martingales. □

Remark 5.5.2 (interpretation as discrete version of Ito integral). Later we will see these two examples are discrete version of Ito integral of $\int_0^t W_t dW_t, \int_0^t f(W_t) dW_t$.

5.6 Stopping time

Definition 5.6.1 (stopping time, continuous version). *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, P)$, $I = [0, \infty)$ *be a filtered probability space. Then a random variable* $\tau : \Omega \to I$ *is called a* $\mathcal{F}_t$ *stopping time if*

$$\{\omega \in \Omega : \tau(\omega) \leq t)\} \in \mathcal{F}_t,$$

that is, the subset of Ω, $\{\omega \in \Omega : \tau(\omega) \leq t)\}$ *is measurable respect to* $\mathcal{F}_t$.

Definition 5.6.2 (stopping time, discrete version). *[12] Let* $X = \{X_n, n \geq 0\}$ *be a stochastic process. A stopping time* τ *with respect to* X *is a discrete random variable on the same probability space of* X*, taking values in the set* $\{0, 1, 2, ...\}$*, such that for each* $n \geq 0$*, the event* $\{\tau = n\}$ *is completed determined by the information up to* n*,i.e., the values of* $\{X_0, X_1, ..., X_n\}$*, or equivalently, the subset in* Ω*:* $\{\omega \in \Omega : \tau(\omega) \leq n\}$ *is* $\mathcal{F}_n$ *measurable.*

Remark 5.6.1. If X_n denote the price of the stock at time n, τ denotes the time at which we will sell it. If our selling decision is based on past information, then τ will be a function of past 'states' characterized by $\{X_0, X_1, X_2, ..., X_{\min(\tau,n)}\}$. Moreover, the amount of past information it depends on is restricted by τ.

5.6.1 Stopping time examples

5.6.1.1 *First passage time*

[12]Let stochastic process X has a discrete state space, and let i be a fixed state, then the first passage time defined as

$$\tau = \min\{n \geq 0 : X_n = i\}$$

is stopping time. At first, τ is a random variable; second, the event $\{\tau = n\}$ is completely determined by the value of $\{X_0, X_1, ..., X_n\}$, i.e., the information up to n. Therefore, it is a stopping time.

5.6.1.2 *Trivial stopping time*

Let X be any stochastic process, and let τ be a deterministic function. The real world example is that a gambler decides that he will only play 10 games regardless of the outcome. τ is a stopping time.

5.6.1.3 *Counter example: last exit time*

Consider the rat in a open maze, a stochastic process X, taking discrete values representing states. Let τ denote the last time the rat visits state i:

$$\tau = \max\{n \geq 0 : X_n = i\}$$

Clearly, we need to know the future to determine the value of τ.

5.6.2 Wald's equation

Theorem 5.6.1 (Wald's equation). *If τ is a stopping time with respect to an iid sequence $\{X_n : n \geq 1\}$, and if $E[\tau] < \infty, E\|X_n|] < \infty$, then*

$$E[\sum_{n=1}^{\tau} X_n] = E[\tau]E[X_1]$$

Proof.

$$E[\sum_{n=1}^{\tau} X_n] = E[\sum_{n=1}^{\infty} X_n I(\tau > n-1)] = E[\sum_{n=1}^{\infty} X_n I(\tau > n-1)] = E[X_1]E[\tau]$$

where $I(\tau > n-1)$ is an indicator function. Note that the event $\{\tau > n-1\}$ only depends on the values of $\{X_1, X_2, ..., X_{n-1}\}$ since its complement event $\{\tau \leq n-1\}$ only depends on the values of $\{X_1, X_2, ..., X_{n-1}\}$. And we have

$$\begin{aligned} E[I(\tau > n-1)] &= \sum_{n=1}^{\infty} P(\tau > n-1) \\ &= \sum_{n=0}^{\infty} P(\tau > n) \\ &= \sum_{n=0}^{\infty} \sum_{i=n+1}^{\infty} P(\tau = i) \\ &= \sum_{i=0}^{\infty} \sum_{n=0}^{i} P(\tau = i) \\ &= \sum_{i=0}^{\infty} iP(\tau = i) = E[\tau] \end{aligned}$$

□

5.6.3 Optional stopping

Theorem 5.6.2 (optional stopping theorem). *Let $X = \{X_n, n \geq 0\}$ be a martingale, let τ be a stopping time with respect to X. Define a stochastic process $\bar{X} = \{X_{n\wedge\tau}\}$, then $\bar{X}$ is a martingale.*

Proof. Let $\mathcal{F}_n = \sigma(X_0, X_1, ...X_n)$,we can rewrite $\bar{X}_{n+1} = \bar{X}_n + I(\tau > n)(X_{n+1} - X_n)$ (this can be verified by consider events of $\{\tau > n\}$ and $\{\tau \leq n\}$),then $E[\bar{X}_{n+1}|\mathcal{F}_n] = \bar{X}_n + 0 = \bar{X}_n$. □

Remark 5.6.2 (stopping time strategy in fair game is still fair)**.** Since $\bar{X}_0 = X_0, E[\bar{X}_n] = X_0$, the implication is using any stopping time as a gambling strategy yields on average, no benefit; the game is still fair.

5.7 Random walk

5.7.1 Basic concepts and properties

Definition 5.7.1 (random walk). *The stochastic process $\{B_n, n \in \mathbb{Z}_+\}$ is called a random walk $S_n = X_1 + X_2 + ... + X_n$ and X_is are iid discrete random variables taking σ and $-\sigma, \sigma > 0$ with probability p and $1-p$ respectively. If $p = 1/2$, then B_n is called symmetric random walk.*

Lemma 5.7.1 (basic properties). *Let B_n be a random walk with step size σ, then*

- $E[B_n] = 0$
- $Var[B_n] = n\sigma^2$.
- $cov(B_t, B_s) = \min(s,t)\sigma^2$

Proof. Note that $B_n = \sum_{i=1}^n X_n$ with $E[X_n] = 0$ and $Var[X_n] = \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2$. Then use linearity of expectation, we can get (1)(2). For (3), let $s < t$, then

$$cov(B_s, B_t) = cov(B_s, B_s + \sum_{i=s+1}^{t} X_i) = Var[B_s] = s\sigma^2.$$

□

Lemma 5.7.2 (martingale property). *Let $\mathcal{F}_n$ be the filtrations associated with the random walk(i.e., $\mathcal{F}_n = \sigma(X_1, ..., X_n)$). Then we have*

- *B_n is a martingale.*
- *$B_n^2 - n$ is a martingale*

Proof. (1) $E[B_{n+1}|\mathcal{F}_n] = E[B_n + X_{n+1}|\mathcal{F}_n] = B_n$ (2) similar to (1). □

5.7.2 Persistent random walk

Definition 5.7.2 (persistent random walk). *The stochastic process $\{B_n, n \in \mathbb{Z}_+\}$ is called a random walk $S_n = X_1 + X_2 + ... + X_n$ and X_is are discrete random variables with properties*

$$E[X_i] = 0, Var[X_i] = \sigma^2, Cov(X_i, X_j) = \sigma^2\rho^{|i-j|}, |\rho| < 1.$$

X_i will take σ and $-\sigma, \sigma > 0$ with equal probability. ρ is called step size correlation coefficient and σ is called step size.

Lemma 5.7.3 (basic properties of persistent random walk). *Let B_n be a persistent random walk with correlation coefficient $\rho, |\rho| < 1$ and step size σ, then*

- $E[B_n] = 0$
- $$Var[B_n] = \sigma^2(n + \frac{2\rho(n-1)}{1-\rho} - \frac{2\rho^2(1-\rho^{n-1})}{(1-\rho)^2})$$
  .
- $$cov(B_n, B_{n+h}) = Var[B_n] + \sum_{j=1}^{h} \rho^j (\frac{\rho^{n+1}-\rho}{\rho - 1}), h > 0$$

Proof. Note that $B_n = \sum_{i=1}^n X_n$ with $E[X_n] = 0$. Then use linearity of expectation, we can get (1).(2)

$$\begin{aligned}
Var[B_n] &= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} Cov(X_i, X_j) \\
&= n\sigma^2 + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} \sigma^2\rho^{j-i} \\
&= n\sigma^2 + 2\sigma^2\sum_{i=1}^{n-1}(n-i)\rho^i \\
&= \sigma^2(n + \frac{2\rho(n-1)}{1-\rho} - \frac{2\rho^2(1-\rho^{n-1})}{(1-\rho)^2}).
\end{aligned}$$

(3)

$$\begin{aligned}
Cov(B_n, B_{n+h}) &= Cov(B_n, B_n + \sum_{i=n+1}^{n+h} X_i) \\
&= Cov(B_n, B_n) + Cov(B_n, \sum_{i=n+1}^{n+h} X_i) \\
&= Var[B_n] + \sigma^2 \sum_{j=1}^{h} \rho^j (\rho + \rho^2 + ... + \rho^n) \\
&= Var[B_n] + \sum_{j=1}^{h} \rho^j (\frac{\rho^{n+1}-\rho}{\rho-1})
\end{aligned}$$

□

5.7.3 Asymptotic properties

Lemma 5.7.4 (unboundedness of random walk). *With probability 1 (i.e. almost surely)*

$$\limsup_n |B_n| = \infty$$

Proof. One intuitive way to prove: given any number M, we can find a N, such that when $n > N$, any trajectories that contains $3/4n$ of up steps will be greater than M. And these trajectories have $\binom{3/4n}{n}/2^n > 0$ probability to happen. □

Corollary 5.7.0.1 (finiteness of first passage time). *Define $T_a = \inf\{t : B_n = a, a \in \mathbb{Z}\}$. $T_a < \infty$ almost surely.(Note that the expected first passage time will be infinite.)*

5.7.4 Gambler's ruin problems

Definition 5.7.3. *A player with some initial money plays a game. He get 1 if he wins and loses 1 otherwise. He will continue to play the games until he reach a total fortune of N, or he gets ruined(running out of money).*

Lemma 5.7.5 (winning probability). *Let P_i being probability he will reach fortune N without being ruined with initial money i. The probability he wins a single game is p; the probability he loses a single game is $q = 1 - p$. Then*

$$P_i = \begin{cases} 0, if\ i = 0 \\ 1, if\ i = N \\ pP_{i+1} + (1-p)P_{i-1}, if\ 0 < i < N \end{cases}$$

Moreover, solving this recurrence equation, we have

$$P_i = \begin{cases} \frac{1-r^i}{1-r^N}, if\ p \neq q \\ i/N, if\ p = q = 0.5 \end{cases}$$

where $r = q/p$.

Proof. We can rewrite the recurrence relationship as

$$P_{i+1} - P_i = q/p(P_i - P_{i-1})$$

and we have

$$P_{i+1} - P_i = r^i P_1$$

Add equations of different i, we have

$$P_{i+1} - P_1 = \sum_{k=1}^{i} r^k P_1$$

Use the boundary condition of $P_N = 1$, and we can solve it. □

Lemma 5.7.6 (winning probability for symmetric case, martingale method). *[13, p. 220] Let P_i being probability he will reach fortune N without being ruined with initial money i. The probability he wins a single game is $p = 0.5$; the probability he loses a single game is $q = 1 - p = 0.5$.Further let M_n be the money he has after n steps, then*

- *M_n is a martingale.*
- *M_τ is a martingale, where τ is the stopping time.*
- *The winning probability is i/N.*

Proof. (1) Straight forward. (2) using optional stopping theorem.(Theorem 5.6.2). (3)

$$E[M_\tau] = NP_i + 0(1 - P_i) = M_0 = i \implies P_i = i/N.$$

□

Lemma 5.7.7 (winning probability for asymmetric case, martingale method). *[13, p. 220] Let P_i being probability he will reach fortune N without being ruined with initial money i. The probability he wins a single game is p; the probability he loses a single game is $q = 1 - p$.Further let X_n be the money he has after n steps, and let*

$$M_n = (\frac{q}{p})^{X_n}.$$

- *M_n is a martingale.*
- *M_τ is a martingale, where τ is the stopping time.*
- *The winning probability is $\frac{r^i - 1}{r^N - 1}, r = q/p$.*

Proof. (1)

$$E[M_{n+1}|\mathcal{F}_n] = E[(\frac{q}{p})^{X_n}(\frac{q}{p})^{X_{n+1}-X_n}|\mathcal{F}_n] = (\frac{q}{p})^{X_n}E[(\frac{q}{p})^{X_{n+1}-X_n}|\mathcal{F}_n] = (\frac{q}{p})^{X_n}((q/p)p + (q/p)^{-1}q) = M_n.$$

(2) using optional stopping theorem.(Theorem 5.6.2). (3)

$$E[M_\tau] = r^N P_i + (1 - P_i) = M_0 = r^i \implies P_i = \frac{r^i - 1}{r^N - 1}.$$

□

Lemma 5.7.8 (mean game duration, symmetric case). *[13, p. 220] Consider the case $p = q = 0.5$. Let X_n be the money he has after n steps. Let M_i be the mean game duration with initial wealth i. Let P_i being probability he will reach fortune N without being ruined with initial money i. We have:*

- $X_n^2 - n$ *is a martingale.*
- $X_\tau^2 - -\tau$ *is a martingale, where τ is the stopping time.*
- $M_i = i(N - i)$.

Proof. (1) from Lemma 5.7.2. (2) using optional stopping theorem.(Theorem 5.6.2). (3)

$$E[X_\tau^2 - \tau] = (N^2)P_i + (0)(1 - P_i) - M_i = i^2 \implies M_i = i(N - i),$$

where $P_i = i/N$ is used. □

Lemma 5.7.9 (mean game duration, asymmetric case). *[13, p. 220] Consider the case $p \neq q$. Let X_n be the money he has after n steps. Let M_i be the mean game duration with initial wealth i. Let P_i being probability he will reach fortune N without being ruined with initial money i. We have:*

- $X_n - n(p - q)$ *is a martingale.*
- $X_\tau - \tau(p - q)$ *is a martingale, where τ is the stopping time.*
- $$M_i = \frac{1}{p - q}(N\frac{r^i - 1}{r^N - 1} - i)$$
  .

Proof. (1) from Lemma 5.7.2. (2) using optional stopping theorem.(Theorem 5.6.2). (3)

$$E[X_\tau - \tau(p - q)] = (N)P_i + 0(1 - P_i) - M_i(p - q) = i \implies M_i = \frac{1}{p - q}(N\frac{r^i - 1}{r^N - 1} - i),$$

where $P_i = i/N$ is used. □

Corollary 5.7.0.2 (asymptotic properties).

- *If $p > 0.5, r < 1$, then*
  $$\lim_{N\to\infty} P_i = 1 - r^i > 0$$
  That is, the player has non-zero probability to get infinitely rich.
- *If $p \leq 0.5, r \geq 1$, then*
  $$\lim_{N\to\infty} P_i = 0$$

That is, the player will go broke with probability 1.

Remark 5.7.1 (interpretation).

- When $N \to \infty$, we are considering the case that the player does not set a winning criterion and he will continue to play until he is ruined.
- If $p \leq 0.5$, and if **the player does not set a winning criterion**, then he will definitely go broke, with no chances of being infinitely rich.

Corollary 5.7.0.3 (lose probability). *Let P_i being probability he will go bankrupt without reaching fortune N with initial money i. Then*

$$P_i = \begin{cases} 1, if\ i = 0 \\ 0, if\ i = N \\ pP_{i+1} + (1-p)P_{i-1}, if\ 0 < i < N \end{cases}$$

Moreover, solving this recurrence equation, we have

$$P_i = \begin{cases} \frac{1-r^{N-i}}{1-r^N}, if\ p \neq q \\ 1 - i/N, if\ p = q \end{cases}$$

where $r = q/p$.

Proof. Use symmetry. □

Remark 5.7.2. Gambler's ruin problem is just special cases of absorbing Markov chain problem.

Lemma 5.7.10 (asymptotic behavior). *Let T denote the first time the player's fortune reaches finite N or gets ruined. Then $T < \infty$ almost surely, that is, the player will stop playing (because either he wins N or gets ruined)after finite number of games.*

Proof. (informal idea) in the finite state absorbing chain, the expected number of step to adsorbing state cannot be infinite. □

Remark 5.7.3 (finite money is impossible in a finite state game in the long run). If N is finite, then the player must hit N or 0 in the long run, the player cannot have finite money if he play infinitely.

Lemma 5.7.11. *Let S denotes the number of games the player played when his fortune first reach N or gets ruined. Then the expected time to win or lose given we start with o dollars is*

$$E_i = \begin{cases} 0, if\ i = 0 \\ 0, if\ i = N \\ 1 + pE_{n+1} + (1-p)E_{n-1}, if\ 0 < i < T \end{cases}$$

Proof. Use conditional expectation. □

Remark 5.7.4. Even though the player will definitely stop after finite number of games. The expected number of games played might be infinite.

Lemma 5.7.12. *If you start with n dollars and $p = 1/2$, and you play until you go broke, then for all $n > 0$, $P(gobroke) = 1$(that is, eventually, you will go broke no matter how rich you are initially).However, the* ***expected number of games played*** *is infinity.*

5.8 Random field theory

Definition 5.8.1 (random field). *[9, p. 216] A random field is a stochastic process $X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^p$ with multi-dimensional parameter*

$$\boldsymbol{t} = (t_1, ..., t_p).$$

Example 5.8.1. [9, p. 216]

- If $\boldsymbol{t} = (t_1, t_2)$ is a two-dimensional, then $X(t_1, t_2)$ is a random variable for every pair of (t_1, t_2). We the think of $(t_1, t_2, X(t_1, t_2))$ as a random surface with height $X(t_1, t_2)$ at location with coordinates (t_1, t_2).
- A time-dependent random surface is a field $(t, s_1, s_2, X(t, s_1, s_2))$ with $\boldsymbol{t} = (t, s_1, s_2)$, where t is the time and $(s_1, s_2) \in \mathbb{R}^2$ is the location.

Definition 5.8.2 (homogeneous fields, isotropic fields). *[9, pp. 217, 221]*

- *A random field is a **homogeneous field** is $m(\boldsymbol{t}) = E[X(\boldsymbol{t})]$ is a constant, and the covariance function $r(\boldsymbol{t}, \boldsymbol{u}) = Cov(X(\boldsymbol{t}), X(\boldsymbol{u})$ depends only on the differences $\boldsymbol{t} - \boldsymbol{u}$(a vector having direction and magnitude), that is*

$$r(\boldsymbol{t}, \boldsymbol{u}) = r(\boldsymbol{t} - \boldsymbol{u}).$$

- *A homogeneous random field is **isotropic** if the covariance function is given by*

$$r(\boldsymbol{t}, \boldsymbol{u}) = r(\|\boldsymbol{t} - \boldsymbol{u}\|).$$

5.9 Notes on bibliography

For elementary level treatment on stochastic process, see [2][4][14] and intermediate level [15]. For general SDE, see [16][15]. For treatment on forward and backward SDE, see [17].

For numerical algorithm for SDE, see [18].

For Fokker-Planck equation, see [19].

For comprehensive and advanced treatment of stochastic methods, see [20] and [21].

For martingale, see [22]. For application of martingale theory to probability theory, see **??**.

For stationary stochastic process, see [9][8].

A good source on simulating SDE with code is at [23].

For treatment on Levy processes, see [24][25][26][27]link.

BIBLIOGRAPHY

1. Dineen, S. *Probability theory in finance: a mathematical guide to the Black-Scholes formula* (American Mathematical Soc., 2013).
2. Brzezniak, Z. & Zastawniak, T. *Basic stochastic processes: a course through exercises* (Springer Science & Business Media, 1999).
3. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
4. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
5. Williams, D. *Probability with martingales* (Cambridge university press, 1991).
6. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2009).
7. Koralov, L. & Sinai, Y. G. *Theory of probability and random processes* (Springer Science & Business Media, 2007).
8. Lindgren, G., Rootzén, H. & Sandsten, M. *Stationary stochastic processes for scientists and engineers* (CRC press, 2013).
9. Lindgren, G. *Stationary Stochastic Processes: Theory and Applications* ISBN: 9781466557796. https://books.google.com/books?id=XROYk50wZCMC (Taylor & Francis, 2012).
10. Miranda, H.-C. *Applied Stochastic Analysis lecture notes* (NewYork University, 2015).
11. Wikipedia. *Markov process — Wikipedia, The Free Encyclopedia* [Online; accessed 28-June-2016]. 2016. https://en.wikipedia.org/w/index.php?title=Markov_process&oldid=712058457.
12. Karl, S. *Stochastic modeling I lecture notes* (Columbia University, 2009).
13. Privault, N. *Understanding Markov Chains: Examples and Applications* ISBN: 9789814451505. https://books.google.com/books?id=THyZmAEACAAJ (Springer Singapore, 2013).
14. Wiersema, U. F. *Brownian motion calculus* (John Wiley & Sons, 2008).
15. Klebaner, F. C. *Introduction to Stochastic Calculus with Applications* (Imperial College Press, 2005).
16. Oksendal, B. *Stochastic differential equations: an introduction with applications* (Springer Science & Business Media, 2013).

17. Ma, J. & Yong, J. *Forward-backward stochastic differential equations and their applications* **1702** (Springer Science & Business Media, 1999).

18. Higham, D. J. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review* **43,** 525–546 (2001).

19. Risken, H. & Frank, T. *The Fokker-Planck Equation: Methods of Solution and Applications* ISBN: 9783642615443. https://books.google.com/books?id=6B5uCQAAQBAJ (Springer Berlin Heidelberg, 2012).

20. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127. https://books.google.com/books?id=otg3PQAACAAJ (Springer Berlin Heidelberg, 2009).

21. Gardiner, C. *Handbook of stochastic methods for physics, chemistry, and the natural sciences* ISBN: 9783540156079. https://books.google.com/books?id=ZfwaAQAAIAAJ (Springer, 1994).

22. Ladd, W. *Some Applications of Martingales to Probability Theory* tech. rep. (Technical Report, University of Chicago, 2011).

23. Iacus, S. M. *Simulation and inference for stochastic differential equations: with R examples* (Springer Science & Business Media, 2009).

24. Schoutens, W. *Levy Processes in Finance: Pricing Financial Derivatives* ISBN: 9780470851562. https://books.google.com/books?id=Be4NqIz7h-kC (Wiley, 2003).

25. Applebaum, D. *Lévy processes and stochastic calculus* (Cambridge university press, 2009).

26. Manuge, D. J. L\'evy Processes For Finance: An Introduction In R. *ArXiv e-prints.* arXiv: 1503.03902 [stat.AP] (Mar. 2015).

27. Tankov, P. *Financial modelling with jump processes* (CRC press, 2003).

6

WIENER PROCESS AND STOCHASTIC CALCULUS

6.1 Wiener process (Brownian motion)

6.1.1 Basics

Definition 6.1.1 (Brownian motion). *A stochastic process $W(t)$ is a called a Wiener process or a Brownian motion if:*

- $W(0) = 0$
- *each sample path is continuous almost surely*
- $W(t) \sim N(0, t)$
- *for all* $0 < t_1 < t_2 < ...$ *the random variables:*

$$W(t_1), W(t_2) - W(t_1), ...$$

are independent and have $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$

Lemma 6.1.1 (basic properties of one-D Brownian motion). *Let $W(t)$ be a Brownian motion, then we have:*

- $E[W(t)] = 0$;
- $Var[W(t)] = t$;
- $cov(W(s), W(t)) = \min(s, t)$;
-
$$\rho(t,s) = \sqrt{1 - \frac{\tau}{t}}, t \geq s, \tau = t - s$$
; therefore, $W(t)$ is a nonstationary Gaussian process.

Proof. (1)(2) directly from definition. (3) Let $s < t$, and $cov(W(s), W(t)) = cov(W(s), W(t) - W(s) + W(s)) = cov(W(s), W(s)) = \min(s, t)$. (4) The joint distribution of $W(s), W(t), t > s$ can be constructed from joint distribution of $W(s), W(t) - W(s)$, which are Multivariate Gaussian,via affine transformation. We can similarly extend to arbitrary joint distributions. It is nonstationary because the autocorrelation function depends both on t and τ. □

Definition 6.1.2 (multi-dimensional independent Brownian motion). *A stochastic process $W(t) = (W_1(t), W_2(t), ..., W_n(t))$ is a called a n-dimensional Wiener process or Brownian motion if:*

- *Each $W_i(t)$ is a Wiener process.*
- *If $i \neq j$, then $W_i(t)$ and $W_j(t)$ are independent.*

Lemma 6.1.2 (basic properties of multidimensional independent Brownian motion). *Consider the vector $W(t) = (W_1(t), W_2(t), ..., W_m(t))^T$ denotes an $m-$dimensional independent Brownian motion/Wiener process, each component is uncorrelated with other components for all values of time t. We have*

$$cov(W_i(s)W_j(t)) = \delta_{ij}\min(s,t)$$

and

$$cov(dw_i(t_i)dw_j(t_j)) = \sigma_i^2\delta_{ij}dt_j = \sigma_i^2\delta(t_i - t_j)dt_i dt_j \delta_{ij}$$

where $dw(t) = w(t+dt) - w(t)$,Dirac delta function can be viewed as having a value of $1/dt$.

Definition 6.1.3 (multi-dimensional correlated Brownian motion). *A stochastic process $W(t) = (W_1(t), W_2(t), ..., W_n(t))$ is a called a n-dimensional Wiener process or Brownian motion with constant instantaneous correlation matrix ρ if:*

- *Each $W_i(t)$ is a Wiener process.*
- $$cov(W_i(s)W_j(t)) = \rho_{ij}\min(s,t).$$

  or in matrix form

  $$Cov(W(s), W(t)) = \rho\min(s,t).$$

Remark 6.1.1 (interpreting correlation Brownain motion). Let $X(t) = (X_1(t), X_2(t), ..., X_n(t))$ be a n-dimensional Brownian motion with constant instantaneous correlation matrix ρ. Then we can view $X(t)$ as the solution to the following SDE:

$$dX_i(t) = dW_i(t), X_i(0) = 0, i = 1,2,...,n,$$

where W_is are Brownian motions and

$$E[dW_i(t)dW_j(s)] = \rho_{ij}dt\delta(s-t).$$

6.1.2 Filtration for Brownian motion

Definition 6.1.4 (Brownian motion filtration). *[1][2]Let W_t be a Brownian motion, the filtration for the Brownian motion can be defined as $\mathcal{F}_t = \sigma(\{\mathcal{F}_s\}_{s\leq t})$.*

Remark 6.1.2.

- This filtration is also the natural filtration.

- W_t is $\mathcal{F}_t$ adapted, but W_{2t} is not $\mathcal{F}_t$ adapted.
- Any stochastic process S_t as solution of Ito SDE is also $\mathcal{F}_t$ adapted, which, intuitively,means that given the values of $\{W_s\}_{s \leq t}$, S_t is known.

Remark 6.1.3 (stochastic process adapted to Brownian filtration). Let $B_t, t \geq 0$ be a Brownian motion and $\{\mathcal{F}_t\}$ be the Brownian filtration.

- The stochastic process $X_t = f(t, B_t), t \geq 0$, where f is a function of two variables, are adapted to the Brownian filtration.
  - $X_t = B_t$, $X_t = B_t^2 - t$
  - $X_t = \max_{0 \leq s \leq t} B_s$ and $X_t = \max_{0 \leq s \leq t} B_s^2$
- Examples that are not adapted to the Brownian motion filtration are: $X_t = B_{t+1}$ and $X_t = B_t + B_T, T > 0$.

6.1.3 Quadratic variation

Remark 6.1.4 (purpose). We introduce the concept of **quadratic variation** to measure how jagged the paths of a Brownian motion are.

Definition 6.1.5 (quadratic variation). *[3, p. 101] The quadratic variation of a function $f : [0, T] \to \mathbb{R}$is defined to be*

$$Q^* = \lim_{l(\Delta) \to 0} \sum_{i=0}^{n-1} (f(t_{i+1} - f_i))^2$$

where Δ is a partition of the interval $[0, T]$ with $0 = t_0 < t_1 ... < t_n = T$, and $l(\Delta) = \max_i(t_{i+1} - t_i)$.

Theorem 6.1.1 (continuously differentiable functions have zero quadratic variations). *[3, p. 101]Given a continuously differentiable function on a closed interval, then its quadratic variation is zero.*

Proof. Using the mean value theorem, $f(t_{i+1}) - f(t_i) = f'(x)(t_{i+1} - t_i)$ for some $x \in (t_i, t_{i+1})$. Because $|f'(x)| \leq M$, then

$$(f(t_{i+1}) - f(t_i))^2 \leq M^2(t_{i+1} - t_i)^2$$

As $l(\Delta) \to 0$, we have $Q^* = 0$. $\square$

Theorem 6.1.2 (Brownian motion quadratic variation). *[3, p. 102] The Brownian motion W on the interval $[0, T]$ has quadratic variation of T in the sense of convergence in mean square.[1]*

Proof. We first prove that

$$EQ(\Delta) = E\sum_{i=0}^{n-1}(W_{i+1} - W_i))^2 = \sum_{i=0}^{n-1}(t_{i+1} - t_i) = T$$

Then we can show $VarQ(\Delta) = 0$, as $l(\Delta) \to 0$. □

Lemma 6.1.3. *Let $W(t)$ be a Brownian motion, then*

$$dW(t)dW(t) = dt, dt \to 0$$

by which we mean

$$E[W(t+dt) - W(t))(W(t+dt) - W(t))] = dt$$

and

$$Var[W(t+dt) - W(t))(W(t+dt) - W(t))] = 2(dt)^2 = o(dt)$$

(that is, the variance will vanish as $dt \to 0$).

Proof. Let $X = dW(t) = W(t+dt) - W(t)$. Then $X \sim N(0, dt)$. Therefore,

$$E[X^2] = dt, Var[X^2] = E[X^4] - E[X^2]^2 = 2(dt)^2,$$

where we use the moment property of Gaussian random variable(subsection 2.2.2). Note that $dW(t)dW(t)$ is just a random variable with mean dt, and variance approaches 0 . □

Remark 6.1.5 (implications)**.**

- Here the conclusion holds in statistical sense, not every sample path holds.

6.1.4 Symmetries and scaling laws

Lemma 6.1.4. *Let $W(t)$ be a standard Brownian motion, then the following are also Brownian motions:*

1. $-W(t)$
2. $W(t+s) - W(s)$

3. $aW(t/a^2)$
4. $tW(1/t)$

Proof. (3) $Var(a(W(s/a^2) - W(t/a^2))) = a^2(t-s)/a^2 = (t-s)$; (4)Assume $t > s$, then $Var(tW(1/t) - sW(1/s))) = Var((t-s)W(1/t) + s(W(1/t) - W(1/s)))) = (t-s)^2/t - s^2/t + s = t - s$. □

6.1.5 Non-differentiability and unbounded variation of path

Theorem 6.1.3. *The Brownian path $W(t)$ is almost surely not differentiable at $t \geq 0$.*

Proof. (informal) consider the differential at $t = 0 : \lim_{t\to 0} W(t)/t = \lim_{s\to\infty} sW(1/s) = \lim_{t\to\infty} \tilde{W}_t = \infty$, where $sW(1/s)$ is another Brownian motion. (For formal proof for non-differentiability anywhere, check [2]). □

Theorem 6.1.4. *[4, p. 189] For almost all Brownian sample path,*

$$\sup_{\tau} \sum_{i=1}^{n} \left| B_{t_i}(\omega) - B_{t_{i-1}}(\omega) \right| = \infty$$

where the supremum is taken over all possible partitions

Remark 6.1.6. Here use almost all is because there is some path, e.g. a path that $B_t(\omega) = const$, that variation will be zero; however, such path has zero probability measure.

6.1.6 The reflection principle

6.1.6.1 *Driftless case*

Lemma 6.1.5 (reflection principle). *[3, p. 208] Let W_t be a Brownian motion. Let m_T denote the minimum value of W_t over the interval $[0, T]$(the minimum value might occur at any time between $[0, T]$). Then*

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x),$$

where $x \geq y$ and $y < 0$. Moreover,

$$P(W_T \geq x, m_T \geq y) = P(W_T \geq x) - P(W_T \leq 2y - x)$$

Proof. Consider all trajectories hitting y at some time $\tau \in [0, T]$ and finally reaching $[x, x + dx]$. There are same number of trajectories that hit y at some time $\tau \in [0, T]$ and finally reaching $[2y - x, 2y - x + dx]$, that is

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x, m_T \leq y).$$

Note that $W_T \leq 2y - x \implies W_T \leq y$ since $x \geq y$. Therefore,

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x).$$

□

Remark 6.1.7 (interpretation). Let y be a barrier level, then

- $P(W_T \geq x, m_T \leq y)$ represents the probability that a random walker hitting the barrier and finally reaching above x.
- $P(W_T \geq x, m_T \leq y)$ represents the probability that a random walker **successfully avoid the barrier** and finally reaching above x.

Lemma 6.1.6 (path excursion distribution). *Let W_t be a driftless Brownian motion. Let m_T denote the minimum value of W_t over the interval $[0, T]$(the minimum value might occur at any time between $[0, T]$). Then*

$$P(m_T \leq y) = 2P(W_T \leq y) = 2N(\frac{y}{\sigma\sqrt{T}}), y \leq 0,$$

$$P(m_T \geq y) = 1 - 2N(\frac{y}{\sigma\sqrt{T}})$$

where W_T is zero mean Gaussian with variance $\sigma^2 T$. In particular, if $T \to \infty$, the $P(m_T \leq y) \to 1$; that is, the Brownian motion will hit any level y with probability 1.

Proof. Use reflection principle(Lemma 6.1.5), we have

$$\begin{aligned} P(m_T \leq y) &= P(m_T \leq y, W_T \leq y) + P(m_T \leq y, W_T \geq y) \\ &= P(m_T \leq y, W_T \leq y) + P(m_T \leq y, W_T \leq y) = 2P(m_T \leq y, W_T \leq y) = 2P(W_T \leq y). \end{aligned}$$

□

Remark 6.1.8 (interpretation & maximum excursion).

- Given a time T, this lemma gives the probability distribution of the excursion of a trajectory during time T.
- It is not possible to know exactly maximum excursion for all possible trajectories. We only know that the larger the excursion, the smaller the probability.

6.1.6.2 *Drifting case*

Lemma 6.1.7 (reflection principle with drift). *[3, p. 213] Let stochastic process Z_t be governed by*

$$dZ_t = vdt + \sigma dW_t,$$

where W_t is the Brownian motion. Let m_t denotes the minimum value of Z_t up to time t. Then, for $y < 0$ and $x > y$, we have

$$P(Z_t \geq x, m_t \leq y) = e^{2vy/\sigma^2} P(Z_t \leq 2y - x + 2vt)$$

Proof. □

Lemma 6.1.8 (path excursion with drift). *[3, p. 213] Let stochastic process Z_t be governed by*

$$dZ_t = vdt + \sigma dW_t,$$

where W_t is the Brownian motion. Let m_t denotes the minimum value of Z_t up to time t. Then, for $y < 0$, and we have

$$P(m_t \leq y) = P(Z_t \leq y) + P(m_t \leq y, Z_t \geq y) = P(Z_t \leq y) + e^{2vy/\sigma^2} P(Z_t \leq 2y - x + 2vt).$$

Proof. Use the fact that

$$P(m_t \leq y) = P(m_t \leq y, Z_t \leq y) + P(m_t \leq y, Z_t \geq y)$$

and note $P(m_t \leq y, Z_t \leq y) = P(Z_t \leq y)$ □

6.1.7 First passage time and stopping time

6.1.7.1 *Minimum and maximum of a Wiener process*

Lemma 6.1.9. *[5, pp. 88, 214] Let W_t be a Wiener process and let $M_t = \max_{0\leq s\leq t} W_s$ be the maximum level reached by W_t during hte time interval $[0, t]$ and let $m_t = \min_{0\leq s\leq t} W_s$ be the minimum level reached by W_t during the time interval $[0, t]$.*

Then for all $t \geq 0$, we have

- $$P(M_t \leq a, W_t \leq x) = \begin{cases} \Phi(\frac{x}{\sqrt{t}}) - \Phi(\frac{x-2a}{\sqrt{t}}), a \geq 0, x \leq a \\ \Phi(\frac{x}{\sqrt{t}}) - \Phi(-\frac{a}{\sqrt{t}}), a \geq 0, x \geq a \\ 0, a \leq 0 \end{cases},$$

the associated joint density function is

$$f_{M_t,W_t}(a,x) = \begin{cases} \frac{2(2a-x)}{t\sqrt{2\pi t}} \exp(-\frac{1}{2}(\frac{(2a-x)^2}{t})), a \geq 0, x \leq a \\ 0, \ otherwise \end{cases}.$$

- $$P(m_t \geq b, W_t \geq x) = \begin{cases} \Phi(-\frac{x}{\sqrt{t}}) - \Phi(\frac{2b-x}{\sqrt{t}}), b \leq 0, x \geq b \\ \Phi(-\frac{b}{\sqrt{t}}) - \Phi(-\frac{b}{\sqrt{t}}), b \leq 0, x \leq b \\ 0, b \geq 0 \end{cases},$$

  the associated joint density function is

$$f_{m_t,W_t}(b,x) = \begin{cases} \frac{2(2b-x)}{t\sqrt{2\pi t}} \exp(-\frac{1}{2}(\frac{(2b-x)^2}{t})), b \leq 0, x \geq b \\ 0, \ otherwise \end{cases}.$$

- $$P(M_t \leq a) = \Phi(\frac{a}{\sigma\sqrt{t}}) - \Phi(\frac{-a}{\sigma\sqrt{t}}), a \geq 0.$$
- $$P(m_t \leq b) = \Phi(\frac{b}{\sigma\sqrt{t}}) + \Phi(\frac{-b}{\sigma\sqrt{t}}), b \leq 0.$$

Note 6.1.1 (common pitfall in extending to non-standard Wiener process).

- Consider a new process Y_t is defined by $Y_t = \mu t + W_t$. It is **wrong** to use $M_t^Y = \mu t + M_t$ to derive the distribution associated with the process Y_t. Becuase

$$M_t^Y = \max_{0 \leq s \leq t} (\mu t + W_t) \neq \mu t + \max_{0 \leq s \leq t} W_t = \mu t + M_t.$$

  Instead, we need to use change of measure to derive the results associated with Y_t.
- However, for process Y_t defined by $Y_t = Y_0 + \sigma W_t$, we can use the relation $M_t^Y = Y_0 + M_t$ to derive the distribution associated with the process Y_t.

Lemma 6.1.10 (running min and maximum of a Wiener Process). *[5, p. 214] Let W_t be a Wiener process and let*

$$X_t = X_0 + \mu t + \sigma W_t.$$

let $M_t^X = \max_{0 \leq s \leq t} X_s$ be the maximum level reached by X_t during hte time interval $[0,t]$ and let $m_t^X = \min_{0 \leq s \leq t} W_s$ be the minimum level reached by X_t during the time interval $[0,t]$. Then

- $$P(M_t^X \leq x) = \Phi(\frac{x - X_0 - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(x - X_0)}{\sigma^2})\Phi(\frac{-x + X_0 - \mu t}{\sigma\sqrt{t}}), x \geq X_0.$$

- $$P(m_t^X \leq x) = \Phi(\frac{x - X_0 - \mu t}{\sigma\sqrt{t}}) + \exp(\frac{2\mu(x - X_0)}{\sigma^2})\Phi(\frac{-x + X_0 - \mu t}{\sigma\sqrt{t}}), x \leq X_0.$$

Lemma 6.1.11 (running min and maximum of a geometric Brownian Process). *Let W_t be a Wiener process and let*

$$S_t = S_0 \exp(\mu t + \sigma W_t).$$

Let $M_t^S = \max_{0 \leq u \leq t} S_u$ be the maximum level reached by S_t during hte time interval $[0, t]$ and let $m_t^X = \min_{0 \leq u \leq t} W_u$ be the minimum level reached by S_t during the time interval $[0, t]$. Then

- $$P(M_t^S \leq x) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}).,x \geq S_0.$$

- $$P(m_t^S \leq x) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) + \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}), x \leq S_0.$$

Proof. (1)Note that $S_t = S_0 \exp(X_t), X_t = \mu t + \sigma W_t$. Therefore

$$M_t^S = S_0 \exp(M_t^X),$$

and then use Lemma 6.1.10:

$$P(M_t^S \leq x) = P(M_t^X \leq \ln(\frac{x}{S_0})) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}).$$

(2) similar to (1). □

6.1.7.2 *Martingale method*

Lemma 6.1.12 (first hitting time in bounded region). *Let X_t be a Brownian motion with no drift. Consider two levels $\alpha > 0$ and $-\beta, \beta > 0$. Then*

- *The probability p_α hitting α before hitting $-\beta$ is $\frac{\beta}{\alpha+\beta}$; The probability p_β hitting $-\beta$ before hitting α is $\frac{\alpha}{\alpha+\beta}$*
- *the expected time to reach level α, or level β is $\alpha\beta$.*

Proof. (1) Let B_τ be process with τ being the stopping time hitting α or $-\beta$. B_τ is a martingale by optional stopping theorem(Theorem 5.6.2). Then we have

$$E[B_\tau] = p_\alpha \alpha + p_\beta(-\beta) = 0, p_\alpha + p_\beta = 1.$$

We can solve to get $p_\alpha = \beta/(\alpha+\beta)$,and $p_\beta = \alpha/(\alpha+\beta)$. (2) $E[B_t^2 - t] = 0 \implies E[\tau] = E[B_\tau^2] = p_\alpha \alpha^2 + p_\beta \beta^2 = \alpha\beta$. □

Lemma 6.1.13 (first hitting time of single level in unbounded region). *[1, p. 112] Let X_t be a Brownian motion with no drift. Consider one level $\alpha > 0$. Then*

- *The probability density of stopping time τ is*

$$\frac{\alpha \exp(-\alpha^2/2\tau)}{\tau\sqrt{2\pi\tau}}$$

- *The probability p_α hitting α before hitting $-\infty$ is 1*
- *The expected time to reach level α is ∞.*

Proof. Let τ be the stopping time, $\tau = \min\{t : X(t) = \alpha\}$. (1) $P(\tau < t) = P(\tau < t, X(t) \geq \alpha) + P(\tau < t, X(t) < \alpha) = 2P(\tau < t, X(t) \geq x) = 2P(X(t) \geq \alpha)$. Note that the event $X(t) \geq x$ already contains $\tau < \alpha$ (2) $P(\tau < \infty) = 1$. (3)

$$E[\tau] = \int_0^\infty \frac{\alpha \exp(-\alpha^2/2t)}{t\sqrt{2\pi t}} dt$$

and the integral will diverge. Another proof: We can use results from Lemma 6.1.12 and set $\beta = \infty$. □

6.1.7.3 *General method via Feyman Kac formula*

Lemma 6.1.14 (General method via Feyman Kac formula, fixed boundary and infinite time horizon). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given two levels $a > 0$ and $-b, b > 0$, and let the probability $P(t, x)$ denote the probability that the process starting at $X(t) = x$ hits a before hitting $-b$. Then we have

- *$P(t,x)$ is independent of time t.*
- *The governing equation for $P(x)$ is given by*

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(x=a,t) = 1, P(x=-b) = 0$.

Proof. (1) Note that this is a Markov process, therefore $P(t,x)$ will be depend on time. (2) Consider a value function $P(x,t) = E[P_T|X(t) = x]$ with final condition $P(x,T) = P_T(P_T$ will take value 1 at target sites and take 0 elsewhere). Then from Feyman Kac theorem(Theorem 8.3.1), $P(x,t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $p(t,x) = P_T$. □

Remark 6.1.9 (generalization to high dimensional). This method can be easily generalized to high dimensional state space first passage time problem(Theorem 8.3.2).

Example 6.1.1. Let $a = 3, b = 5, \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(3) = 1, P(-5) = 0.$$

The general solution gives

$$P(x) = c_1e^{0x} + c_2e^{-2mx} = c_1 + c_2e^{-2mx}.$$

With boundary condition, we can solve c_1, c_2 therefore any $P(x)$ can be evaluated.

Example 6.1.2. Let $m = 0, \sigma = 1$. Then

$$P(x) = \frac{x+b}{a+b}, x \in [-b, a].$$

Example 6.1.3. Let $a = \infty, b = 1, m = \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(\infty) = 0, P(-1) = 1.$$

The general solution gives

$$P(x) = c_1e^{0x} + c_2e^{-2mx} = c_1 + c_2e^{-2mx}.$$

With boundary condition, we can solve c_1, c_2 as

$$c_1 + c_2 e^2 = 1, c_1 + c_2 e^{-\infty} = c_1 = 0$$

resulting $c_1 = 0, c_2 = e^{-2}, P(x) = e^{-2(x+1)}$.

Example 6.1.4 (boundary condition issue). **warning!**
Let $a = 1, b = \infty, m = \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(1) = 1, P(-\infty) = ?.$$

Note that because the X is upward drifting, $P(-\infty)$ is not clear which value to set.

Lemma 6.1.15 (General method via Feyman Kac formula, fixed boundary, dynamic boundary and finite time horizon). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given one level $a(t)$ as a function of time. Let the probability $P(t,x)$ denote the probability that the process starting at $X(t) = x$ hits $a(t)$. Then the governing equation for $P(x,t)$ is given by

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(x(t) = a(t), t \leq T) = 1, P(x,T) = 0$.

Proof. Let τ be an arbitrary time satisfying $\tau \leq T$. Then,

$$P(x(\tau) = a(\tau)|x(t) = x) = E[\mathbf{1}_{x(\tau)=a(\tau)}|x(t) = x].$$

Then from Feyman Kac theorem(Theorem 8.3.1), $P(x,t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(\tau, x(\tau) = a(\tau), \tau \leq T) = 1$. Since we limit the time horizon to T, we have $P(x,T) = 0$. □

6.1.8 Asymptotic behaviors

Theorem 6.1.5 (law of iterated logarithms). *As $t \to \infty$, we have with probability 1 (i.e. almost surely):*

- $\lim_{t\to\infty} W_t/t = 0$
- $\limsup_{t\to\infty} W_t/\sqrt{t} = \infty$
- $\limsup_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$
- $\liminf_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$

Proof. See [2] for proofs. □

Corollary 6.1.5.1 (unboundedness of Brownian motion). *With probability 1 (i.e. almost surely)*

$$\limsup_t |W_t| = \infty$$

Proof. Use contradiction. If it does not hold, then the law of iterated logarithm cannot hold. □

Corollary 6.1.5.2 (finiteness of first passage time). *Define $T_a = \inf\{t : W_n > a\}$. $T_a <= \infty$ almost surely(but the mean first passage time will be infinite).*

Remark 6.1.10. Lemma 6.1.13 also shows that the hitting probability is 1 given infinite amount of time.

6.1.9 Levy characterization of Brownian motion

Lemma 6.1.16. *Let $M(t), t \geq 0$ be a stochastic process adapted to a filtration $\{\mathcal{F}_t\}$. Assume that*

- *$M(t), t \geq 0$, be a martingale with respect to a filtration $\mathcal{F}_t$.*
- *$M(0) = 0$, $M(t)$ has continuous paths.*
- *quadratic variation $[M, M](t) = t, \forall t \geq 0$, or $dM_t dM_t = dt$*

Then for all $0 \leq s < t$ and a C^2 function f, we have

$$E[f(X_t)|\mathcal{F}_s] = X_s + \frac{1}{2}\int_s^t E[f''(X_u)|\mathcal{F}_u]$$

Theorem 6.1.6 (Levy characterization, one dimension). *[6, p. 87][1, p. 168] Let $M(t), t \geq 0$ be a stochastic process adapted to a filtration $\{\mathcal{F}_t\}$. Assume that*

- *$M(t), t \geq 0$, be a martingale with respect to a filtration $\mathcal{F}_t$.*

- *$M(0) = 0$, $M(t)$ has continuous paths.*
- *quadratic variation $[M, M](t) = t, \forall t \geq 0$, or $dM_t dM_t = dt$*

Then $M(t)$ is a Brownian motion.

Proof. (idea) Calculate the moment generating function and show $M(t) - M(s) \sim N(0, t-s)$.

Consider the function $\exp(\lambda M_t)$. We have (from Ito lemma)

$$d\exp(\lambda M_t) = \exp(\lambda M_t)\lambda dM_t + \frac{1}{2}\exp(\lambda M_t)\lambda^2 dt,$$

where we use the assumption $dM_t dM_t = dt$.

Therefore,

$$\exp(\lambda M_T) = 1 + \int_0^T \exp(\lambda M_t) dM_t + \frac{1}{2}\lambda^2 \int_0^T \exp(\lambda M_t) dt.$$

Take expectation, we have

$$E[\exp(\lambda M_T)] = 1 + \frac{1}{2}\lambda^2 \int_0^T E[\exp(\lambda M_t)] dt,$$

where we use the fact that $E[\int_0^T \exp(\lambda M_t) dM_t] = 0$ from Lemma 6.2.1.

Set $g(T) = E[\exp(\lambda M_T)]$, we have $dg = \frac{1}{2}\lambda g(t) dt$, therefore, $g(t) = \exp(\frac{1}{2}\lambda^2 t), g(0) = 1$.

That is, $M_t \sim N(0, t)$.

Similarly, we can show $M_t - M_s \sim N(0, t-s)$. □

Remark 6.1.11 (implication).

- **continuous path requirement is essential**, because Brownian motion has continuous paths.
- Given a stochastic process $X(t)$, Levy characterization enables to test whether $X(t)$ is a Brownian motion.
- Normality can be resulted even no Gaussian distribution is explicitly involved.

Theorem 6.1.7 (Levy characterization, multiple dimension). *[1, p. 168] Let $M_1(t), ..., M_n(t), t \geq 0$ be stochastic processes. Assume that*

- *$M_i(t), t \geq 0, \forall i$, be a martingale with respect to a filtration $\mathcal{F}_t$.*
- *$M_i(0) = 0$, $M_i(t)$ has continuous paths for i.*

- *quadratic variation* $[M_i, M_j](t) = t\delta_{ij}, \forall t \geq 0$, *or* $dM_{it}dM_{jt} = dt\delta_{ij}$

Then $[M_1(t), ..., M_n(t)]$ *is a multi-dimensional Brownian motion.*

6.1.10 Discrete-time approximations

Lemma 6.1.17 (discrete-time approximation of white noise). *Consider a white noise* $w(t)$ *satisfying*

$$E[w(t)] = 0, E[w(t)w(\tau)] = \sigma^2\delta(t-\tau)$$

Then its discrete-time approximation white noise process $\{w_1, w_2, ...,\}$ *is given as*

$$E[w_i] = 0, E[w_i w_j] = \frac{1}{\Delta t}\sigma^2\delta_{ij}$$

where w_i *approximate the* $w(t), t \in [t_0 + k\Delta t, t_0 + (k+1)\Delta t]$. *Note that* $\delta(x)$ *is the Dirac delta function, whereas* δ_{ij} *is the Kronecker delta function.*

Moreover, the random walk

$$S_N = \sum_{i=1}^{N} w_i,$$

where $N = \frac{T}{\Delta t}$, *has the distribution of* $N(0, T\sigma^2)$, *which is the same as the Brownian motion distribution at time* T, *given as* $B(T) \sim N(0, \sigma^2 T)$.

Proof. Note that as $\Delta t \to 0$, we recover the covariance for the while noise process. For the distribution S_N, use $N(0, N\frac{1}{\Delta t}\sigma^2) = N(0, T\sigma^2)$ and central limit theorem. □

Remark 6.1.12 (implications). The approximation scheme is important in simulating stochastic differential equations.

6.1.11 Geometric Brownian motion

Definition 6.1.6 (geometric Brownian motion). *Suppose* Z_t *is standard Brownian motion and* $\mu \in \mathbb{R}, \sigma > 0$, *then*

$$X_t = X_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t), t \in [0, \infty)$$

is a stochastic process called geometric Brownian motion with drift μ and volatility parameter σ. Moreover, X_t is the solution to the Ito stochastic differential equation given as

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

Lemma 6.1.18 (distribution). *The geoemtric Brownian motion has the lognormal distribution with parameter $(\mu - \frac{1}{2}\sigma^2)t$ and $\sigma\sqrt{t}$. The pdf is given as*

$$f_t(x) = \frac{1}{\sqrt{2\pi t}\sigma x}\exp\left(-\frac{(\ln(x/x_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}\right)$$

Lemma 6.1.19 (expectation and variance). *Let S_t be a geometric Brownian motion with initial condition S_0, then*

- $E[S_t] = S_0 e^{\mu t}$
- $Var[S_t] = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

Proof. See Definition 2.2.6. □

Remark 6.1.13 (martingale property)**.**

- If $\mu \neq 0$, then Geometric Brownian motion is **not** a martingale since $E[X_t]$ is not a constant.
- If $\mu = 0$, then it is an exponential martingale(Lemma 5.5.2).

Remark 6.1.14. The geometric Brownian motion is **Not a Gaussian process.**

6.2 Ito integral

6.2.1 Construction of Ito integral

Definition 6.2.1 (simple process). *The stochastic process $C_t, t \in [0, T]$ is said to be simple if: there exista a partition*

$$\tau_n : 0 = t_0 < t_1 < ... < t_n = T$$

and a sequence of random variables $Z_i, i = 1, 2, ..., n$ such that

$$C_t = \begin{cases} Z_n, if\ t = T, \\ Z_i, if\ t_{i-1} \leq t < t_i, i = 1, 2, ..., n \end{cases}$$

The sequence Z_i is adapted to $\mathcal{F}_{t_{i-1}}$ and $E[Z_i^2] < \infty$,i.e., the sequence Z_i is a previsible process.

Definition 6.2.2 (Ito integral for simple processes). *Let C_t be a simple process on $[0, T]$, the Ito integral is defined as*

$$\int_0^T C_s dB_s = \sum_{i=1}^{n} C_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$$

Theorem 6.2.1 (Ito integral as martingale transform). *The sequence of Ito integral*

$$\int_0^{t_k} C_s dB_s, k = 0, 1, 2, ..., n$$

of a simple process C_s is a martingale transform with respect to the Brownian filtration $\mathcal{F}_{t_k}$. And the stochastic process $I_t(C) = \int_0^t C_s dB_s$ is a martingale with respect to the Brownian filtration $\mathcal{F}_t$.

Proof. directly from martingale transformation in discrete time.(Lemma 5.5.3) □

Theorem 6.2.2 (Isometry for simple process). *The Ito stochastic integral satisfies the isometry property:*

$$E[(\int_0^t C_s dB_s)^2] = E[\int_0^T C_s^2 ds]$$

Proof. By definition, for simple process:

$$\int_0^t C_s dB_s = \sum_{i=1}^n Z_i \Delta_i B$$

where $\Delta_i B = B(t_i) - B(t_{i-1})$. Then

$$\begin{aligned} E[(\int_0^t C_s dB_s)^2] &= E[\sum_{i=1}^n \sum_{j=1}^n Z_i \Delta_i B Z_j \Delta_j B] = E[\sum_{j=1}^n (Z_i)^2 (\Delta_i B)^2] \\ &= \sum_{j=1}^n E[(Z_i)^2](t_i - t_{i-1}) = \int_0^t E[C_s^2] dt \end{aligned}$$

where we use $E[(\Delta_i B)^2] = t_i - t_{i-1}$ □

Corollary 6.2.2.1. *Consider the stochastic process X_t defined as the Ito integral of a simple process:*

$$X_t = \int_0^t C_s dB_s$$

Then we have

- $E[X_t] = 0$
- $E[X_t^2] = Var[X_t] = E[\int_0^2 C_s^2 dB_s]$

Theorem 6.2.3 (existence of approximating simple process). *Let C be a stochastic process that (1) adapted to Brownian filtration on $[0, T]$ and $\int_0^T E[C^2]dt$ is finite. Then there exist a sequence of simple process C_s^n such that*

$$\lim_{n\to\infty} \int_0^T E[(C_s^n - C)]dt = 0$$

Proof. see[4, p. 109]. □

Definition 6.2.3 (Ito integral of general process). *Let C be a stochastic process that (1) adapted to Brownian filtration on $[0,T]$ and (2) $\int_0^T E[C^2]dt$ is finite. Then*

$$\int_0^T CdB = \lim_{n\to\infty}\int_0^T C_s^n dB$$

where C_s^n is a sequence of simple process approximating C in the mean square sense.

6.2.2 Properties of Ito integral

Definition 6.2.4 (non-anticipating process). *A stochastic process F_t is said to be a non-anticipating process with respect to the Brownian motion W_t if F_t is independent of $B_{t'} - B_t$ with $t' > t$.*

Lemma 6.2.1 (basic properties).

- $\int_0^T cdW_t = cW_T$, *where c is a constant.*
- $I_s = \int_0^s W_t dW_t = 0.5(W_s^2 - s)$ *is a martingale, and* $E[I_s] = E[I_0] = 0$.
- *Let g be an square-integrable adapted process to the Brownian filtration $\{\mathcal{F}_t\}$ generated by Brownian motion $W(s)$. Then $I(t) = \int_0^t g(s)dW(s)$ is a continuous square-integrable martingale.* [a]

[a] an adapted process means either deterministic process or stochastic process represented by $dg_t = \mu(g(t),t)dt + \sigma(g(t),t)dW(t)$.

Proof. (1) recognize that this is a Wiener integral(Lemma 6.2.2) on the left, which will produce a normal distribution of $N(0, \int_0^T c^2 dt)$. The Right side has the exact same distribution. (2) Let $Y_t = W_t^2$, and then

$$dY_t = 2W_t dW_t + dt$$

Integrate both sides, we have

$$W_T^2 = 2\int_0^T W_t dW_t + T.$$

(3)

$$dI_t = g(t)dW(t) + \int_0^t dg(s)dW(s) = g(t)dW(t)$$

where we ignore $\int_0^t dg(s)dW(s)$ since it is of order $(O(t))$.

□

Theorem 6.2.4 (Properties of Ito integral). *[7, p. 100] Let $f(W_t,t), g(W_t,t)$ be nonanticipating processes and $c \in \mathbb{R}$, then we have*

1. *partition property:*
$$\int_S^T f dW_t = \int_S^u f dW_t + \int_u^T f dW_t$$
if $S < u < T$
2. *linearity:*
$$\int_S^T (cf + dg) dW_t = c\int_S^T f dW_t + d\int_S^T g dW_t$$
3. *zero mean:*
$$E[\int_S^T f(W_t,t) dW_t] = 0$$
4. *Isometry:*
$$E[(\int_a^b f(W_t,t) dW_t)^2] = E[\int_a^b f(W_t,t)^2 dt]$$
5. *Convariance:*
$$E[(\int_a^b f(W_t,t) dW_t)(\int_a^b g(W_t,t) dW_t)] = E[\int_a^b f(W_t,t) g(W_t,t) dt]$$

Remark 6.2.1.

- See [2][4]
- (3) does not hold for Stratonvich integral.
- all properties hold for both simple process and general process, since general process is defined via simple process.

6.2.3 Wiener integral and Riemman integral with Wiener process

Lemma 6.2.2 (Wiener integral). *[7, p. 112] Suppose $g : [0,\infty) \rightarrow \mathbb{R}$ is a bounded, piecewise continuous function in L^2. Let B_t be a Brownian motion, then*

$$\int_0^t g(s) dB_s$$

is a random variable which has a

$$N(0, \int_0^t g^2(s) ds)$$

*distribution. This integral is also known as **Wiener integral**. In particular,*

$$\int_0^t dB_s = B_t \sim N(0,t)$$

Proof. (informal) Directly from zero mean and Isometry properties(see Theorem 6.2.4). The resulting process is Gaussian can be derived from the sum of independent Gaussian random variables is Gaussian. □

Example 6.2.1. Consider stochastic process

$$X_t = \int_0^t \frac{1}{1-s} dW_s,$$

where W_t is the Wiener process. Then we have

- X_t is a Gaussian process.
- $E[X_t] = 0$.
- $Var[X_t] = E[X_t^2] - E[X_t]^2 = E[X_t^2]$, and

$$E[X_t^2] = \int_0^t \frac{1}{(1-s)^2} ds$$

  via Ito Isometry.

Lemma 6.2.3 (linearity of Wiener integral). *let $B(t)$ be the Wiener process, let $g, h, m : [0, \infty) \to \mathbb{R}$ be a bounded, piecewise continuous function in L^2. Then*

$$\alpha \int_0^t g(s) dB_s + \beta \int_0^t h(s) dB_s = \int_0^t \alpha g(s) + \beta h(s) dB_s.$$

In particular,

$$m(t)B_t - \int_0^t h(s) dB_s = \int_0^t (m(t) - h(s)) dB_s \sim N(0, \int_0^t (m(t) - h(s))^2 ds)$$

Proof. The linearity can be derived directly from the linearity of Ito integral. □

caution!
We know that

$$Z_t = \int_0^t \alpha g(s) + \beta h(s) dB_s \sim N(0, \int_0^t (\alpha g(s) + \beta h(s))^2 dt)$$

however,

$$X_t = \alpha \int_0^t g(s)dB_s \sim N(0, \alpha^2 \int_0^t g(s)^2 ds), Y_t = \beta \int_0^t h(s)dB_s \sim N(0, \beta^2 \int_0^t h(s)^2 ds)$$

Note that X_t and Y_t are **not independent to each other because they are generated from the same Wiener process.**

Theorem 6.2.5 (Riemann integral with Wiener process). *Let g be a smooth function, and let $W(t)$ be the Wiener process, then*

$$\int_0^t g(W(s))ds = g(W(t))t - \int_0^t sg'(W(s))dW_s - \int_0^t \frac{1}{2}g''(W(s))ds.$$

In particularly, if $g(x) = x$, then

$$\int_0^t W(s)ds = W(t)t - \int_0^t sdW_s \sim N(0, \int_0^t (t-s)^2 ds).$$

Proof. (1)Consider $f(W_t, t) = g(W_t)t$ and apply Ito rule (Lemma 6.3.1), we will get

$$d(g(W_t)t) = g(W_t)dt + tg'(W_t)dW_t + \frac{1}{2}tg''(W_t)dt$$

where $dW_t dW_t = dt$ is used. Then we integrate both sides. (2) To show

$$\int_0^t W(s)ds = W(t)t - \int_0^t sdW_s \sim N(0, \int_0^t (t-s)^2 ds),$$

note that $W(t) = \int_0^t dW_s$ and then use linearity of Wiener integral(Lemma 6.2.3). □

Corollary 6.2.5.1. *Let $W(t)$ be the Wiener process, then*

- $$\int_0^1 W(s)ds = W(1) - \int_0^1 sdW_s \sim N(0, \int_0^1 (1-s)^2 ds) = N(0, \frac{1}{3}).$$
- $$\int_0^T W(s)ds = \frac{T}{\sqrt{3}}W(T) \sim N(0, \frac{T^3}{3})$$
- $$\int_0^t g'(s)W_s ds \sim N(0, \int_0^t [g(t) - g(s)]^2 ds)$$

- $$\int_0^1 s^n W_s ds \sim N(0, \frac{2}{(2n+3)(n+2)}), n = 0, 1, 2, ...$$

Proof. (2) Since
$$d(sW(s)) = W_s ds + s dW_s,$$
we have
$$TW_T) = \int_0^T W_s ds + \int_0^T s dW_s.$$
Rearrange, we have
$$\int_0^T W_s ds = \int_0^T (T-s) dW_s \sim N(0, \int_0^T (T-s)^2 ds) = N(0, \frac{T^3}{3}).$$

(3) Let $f(W_t, t) = g(t)W_t$.(4) from (2). □

6.2.4 Quadratic variations

Lemma 6.2.4 (Quadratic variations for Ito process). *Consider an Ito process given by $dX_t = a(t, X_t)dt + b(t, X_t)dB_t$, then the quadratic variation of a process given by*

$$\int_0^T dX_t dX_t = \int_0^T b^2 dt = b^2 t.$$

Proof. Directly compute $dX_t dX_t$ and ignore $o(dt)$ terms. □

6.3 Stochastic differential equations

6.3.1 Ito Stochastic differential equations

Definition 6.3.1 (Ito SDE). *[4, p. 137]An Ito stochastic differential equation is defined as*

$$dX_t = a(t, X_t)dt + b(t, X_t)dB_t$$

which could be interpret as

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dB_s$$

where the first integral is Riemann integral, the second is Ito integral.

Remark 6.3.1. The integral equation is **Not** a solution, because it contains $X(t)$ itself.

Theorem 6.3.1 (existence). *[4, p. 138] Assume the initial condition X_0 has a finite second moment: $EX_0^2 < \infty$, and is independent of $(B_t, t \geq 0)$. Assume that, for all $t \in [0, T], x, y \in \mathbb{R}$, the coefficient functions $a(t, x)$ and $b(t, x)$ satisfy the following conditions:*

- *They are continuous*
- *They satisfy a Lipschitz condition with respect to the second variable:*

$$\left|a(t,x) - a(t,y)\right| + \left|b(t,x) - b(t,y)\right| \leq K\left|x - y\right|$$

Then the Ito stochastic differential equation has a unique solution X on $[0, T]$.

Theorem 6.3.2 (linear stochastic differential equation).

$$X_t = X_0 = \int_0^t (c_1 X_s + c_2)ds + \int_0^t (\sigma_1 X_s + \sigma_2)dB_s$$

for constants c_i and σ_i is called linear SDE. The Linear SDE has an unique solution.

Proof. It is easy to show that the continuous condition and Lipschitz condition are satisfied. □

6.3.2 Ito's lemma

6.3.2.1 *one-dimensional version*

Lemma 6.3.1. *[6, p. 79] Let $f(B_t, t)$ be a function of Brownian motion B_t, then*

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial B_t}dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial B_t^2}dt$$

Lemma 6.3.2. *Let $f(X_t, t)$ be a function of stochastic process X_t governed by $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$, then*

$$\begin{aligned} df &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}(dX_t)^2 \\ &= (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}\sigma^2 + \frac{\partial f}{\partial X_t}\mu)dt + \frac{\partial f}{\partial X_t}\sigma dB_t \end{aligned}$$

Example 6.3.1. $X_t = W_t^3$, then $dX_t = 3W_t^2 dW_t + 3W_t dW_t dW_t = 3W_t^2 dW_t + 3W_t dt$

Example 6.3.2. $Y_t = \ln(W_t)$, then $dY_t = dW_t/W_t - \frac{1}{2}dt/W_t^2$

6.3.2.2 *Multi-dimensional version*

Lemma 6.3.3. *Let $f(B_{1,t}, B_{2,t}, ..., B_{n,t}, t)$ be a function of Brownian motion $B_{1,t}, B_{2,t}, ..., B_{n,t}$, then*

$$df = \frac{\partial f}{\partial t}dt + \sum_{i=1}^{n}\frac{\partial f}{\partial B_{i,t}}dB_{i,t} + \sum_i\sum_j \frac{1}{2}\frac{\partial^2 f}{\partial B_{i,t}\partial B_{j,t}}D_{ij}dt$$

where we assume $E[dB_{i,t}dB_{j,t}] = D_{ij}dt$.

6.3.2.3 *Product rule and quotient rule*

Lemma 6.3.4 (product rule and quotient rule). *[6, p. 79] Consider*

$$\begin{aligned} dX_t/X_t &= r_1 dt + \sigma_1 dW_1 \\ dY_t/Y_t &= r_2 dt + \sigma_2 dW_2 \\ dW_1 dW_2 &= \rho dt \end{aligned}$$

It follows that

- *Given* $Z_t = X_t Y_t$, *we have*

$$dZ_t = d(X_tY_t) = X_tdY_t + Y_tdX_t + dX_tdY_t$$
$$= X_tY_t((r_1 + r_2 + \rho\sigma_1\sigma_2)dt + (\sigma_1 dW_1 + \sigma_2 dW_2))$$

- *Given* $Z_t = X_t/Y_t$, *we have*

$$dZ_t = d(X_t/Y_t) = dX_t/Y_t - X_tdY_t/(Y_t)^2 - dX_tdY_t/(Y_t)^2 + X_t(dY_t)^2/(Y_t)^3$$
$$= (X_t/Y_t)((r_1 - r_2 - \rho\sigma_1\sigma_2 + \sigma_2^2)dt + (\sigma_1 dW_1 - \sigma_2 dW_2))$$

- *Given* $Z_t = 1/X_t$, *we have*

$$dZ_t = d(1/X_t) = -dX_t/(X_t)^2 + (dX_t)^2/(X_t)^3$$
$$= (1/X_t)((-r_1 + \sigma_1^2)dt - \sigma_1 dW_1)$$

Note that we have to calculate the Hessian for $f(x,y) = x/y$, *and there are two terms for the cross-term.*

Proof. (1)

$$dZ_t = \frac{\partial Z}{\partial X}dX + \frac{\partial Z}{\partial X}dX + \frac{1}{2}\frac{\partial^2 Z}{\partial X^2}dXdX + \frac{1}{2}\frac{\partial^2 Z}{\partial Y^2}dYdY + \frac{\partial^2 Z}{\partial X \partial Y}dXdY.$$

(2)(3) Same as (1). □

6.3.2.4 *Logorithm and exponential*

Lemma 6.3.5 (Ito lemma applied to logorithm and exponential). *Let* $X(t)$ *be an Ito stochasti process.*

- *If* $Y_t = \exp(X(t))$, *then*

$$dY_t = Y_tdX_t + \frac{1}{2}Y_tdX_tdX_t.$$

- *If* $Z_t = \ln(X(t))$, *then*

$$dZ_t = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}dX_tdX_t.$$

Proof. (1)

$$\begin{aligned} dY_t &= \exp(X_t)dX_t + \frac{1}{2}\exp(X_t)dX_t dX_t \\ &= Y_t dX_t + \frac{1}{2}Y_t dX_t dX_t. \end{aligned}$$

(2)

$$dZ_t = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}dX_t dX_t.$$

□

6.3.2.5 *Integrals of Ito process*

Lemma 6.3.6 (integrand is an Ito stochastic process). *Let $r(t)$ be an Ito stochasti process.*

- *If $X_t = \int_0^t r(s)ds$, then*
$$dX_t = r(t)dt.$$
- *If $Y_t = \exp(X_t)$, then*
$$dY_t = Y_t r(t)dt.$$

Proof. (1) Let Ω be the sample space associated with the stochastic process $r(t)$. Then for each sample path $\omega \in \Omega$, we have $X_t(\omega) = \int_0^t r(s,\omega)ds$ and $dX_t(\omega) = r(t,\omega)ds$. Since $dX_t(\omega) \triangleq \lim_{dt\to 0} X(t+dt,\omega) - X(t,\omega)$ and $r(t,\omega)ds$ are both random variables for fixed t, if they are equal for each $\omega \in \Omega$, we can write

$$dX_t = r(t)dt.$$

(2)

$$dY_t = \exp(X_t)dX_t = \exp(X_t)r(t)dt = Y_t r(t)dt.$$

□

Remark 6.3.2 (common pitfalls).

- It is worth noting that when $X_t = \int_0^t r(s)ds$ and $r(t)$ is an Ito stochastic process, X_t is not an Ito integral process.
- Similarly, for $Y_t = \exp(X_t)$, Y_t is not an Ito integral, and the Ito lemma does not apply.

Lemma 6.3.7 (Ito lemma applied to integral of Ito processes). *Let $X(t)$ be an Ito stochasti process. Let $r(t)$ be a deterministic function.*

- *If $Y_t = \int_0^t r(s)dX(s)$, then*
$$dY_t = r(t)dX(t).$$
- *If $Z_t = \exp(Y_t)$, then*
$$dZ_t = Z_t r(t)dX(t) + \frac{1}{2}Z_t r(t)^2 dX(t)dX(t).$$

Proof. (1) by definition. (2) Using Ito rule (Lemma 6.3.5), we have

$$\begin{aligned} dZ_t &= Z_t dY_t + \frac{1}{2}Z_t dY_t dY_t \\ &= Z_t r(t)dX(t) + \frac{1}{2}Z_t r(t)^2 dX(t)dX(t) \end{aligned}$$

□

6.3.2.6 *Ito Integral by parts*

Lemma 6.3.8. *[6, p. 79] Let $X(t), Y(t)$ be two Ito processes. Then*

$$\int_s^u Y(t)dX(t) = X(u)Y(u) - X(s)Y(s) - \int_s^u X(t)dY(t) - \int_s^u dX(t)dY(t)$$

Proof. From the product rule, we have

$$\begin{aligned} \int_s^u d[X(t)Y(t)] &= \int_s^u Y(t)dX(t) + \int_s^u X(t)dY(t) + \int_s^u dX(t)dY(t) \\ X(u)Y(u) - X(s)Y(s) &= \int_s^u Y(t)dX(t) + \int_s^u X(t)dY(t) + \int_s^u dX(t)dY(t) \\ \int_s^u Y(t)dX(t) &= X(u)Y(u) - X(s)Y(s) - \int_s^u X(t)dY(t) - \int_s^u dX(t)dY(t) \end{aligned}$$

Note that This integral-by-part formula is the same as Riemann integral except for the extra term $\int_s^u dX(t)dY(t)$. □

6.3.2.7 *Fundamental theorem of Ito stochastic calculus*

Theorem 6.3.3 (Fundamental theorem of Ito stochastic calculus). *[8, p. 79] Let $h(W_t)$ be a function on $W(t)$, then*

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

Proof. Note that

$$dh = h'dW_t + \frac{1}{2}h''dt.$$

□

Example 6.3.3. If $h(x) = 0.5x^2$, we have

$$\frac{1}{2}(W_t^2 - W_0^2) = \int_{t_0}^t W_s dW_s + \frac{1}{2}\int_{t_0}^t sds$$

6.3.3 Solutions to Ito stochastic differential equations

Definition 6.3.2 (strong solution to SDE). *[4, p. 136]A strong solution to the Ito SDE is a stochastic process X_t which satisfies the following conditions:*

- *X_t is adapted to the Brownian motion filtration, i.e., X_t is a function of $B_s, s \leq t$.*
- *X is a function of the underlying Brownian motion sample path and of the coefficient functions $a(t,x)$ and $b(t,x)$.*

Example 6.3.4 (solution to geometric Brownian motion). The solution to the GBM SDE:

$$X_t = X_0 + c\int_0^t X_s ds + \sigma\int_0^t X_s dB_s$$

or

$$dX_t = cX_t dt + \sigma X_t dB_t$$

is

$$X_t = X_0 \exp((c - 0.5\sigma^2)t + \sigma B_t)$$

Verification: Let $X_t = f(t, B_t)$, then

$$dX_t = f_t dt + f_{B_t} dB_t + \frac{1}{2} f_{B_t B_t} dB_t dB_t = (c - 0.5\sigma^2)X_t dt + \sigma X_t dB_t + \frac{1}{2}\sigma^2 X_t dt$$

Example 6.3.5 (solution to Ornstein-Uhlenbeck process). [4, p. 141] The Ornstein-Uhlenbeck process SDE is:

$$dX_t = cX_t dt + \sigma dB_t$$

and it has solution:

$$X_t = e^{ct} X_0 + \sigma \int_0^t \exp(c(t-s)) dB_s$$

Example 6.3.6 (solution to mean reversion with square-root diffusion). [6, p. 112] The SDE is

$$dr(t) = -\lambda(r(t) - \bar{r}) + \sigma\sqrt{r(t)} dB(t)$$

This SDE has no closed form expression, for its mean and variance property, we can refer to [6, p. 112].

6.3.4 Solution method to linear SDE

6.3.4.1 *State-independent linear arithmetic SDE*

Lemma 6.3.9 (state independent/general arithmetic SDE). *[7, p. 146][6, p. 116]The solution X_t of the stochastic differential equation*

$$dX_t = a(t)dt + b(t)dW(t)$$

is given by

$$X_t = X_0 + \int_0^t a(s)ds + \int_0^t b(s)dW(s),$$

which is a ***Gaussian distribution*** *with mean $X_0 + \int_0^t a(s)ds$ and variance $\int_0^t b^2(s)ds$.*

Moreover, X_t is a Gaussian process(Lemma 5.3.3).

Proof. The integral form is

$$X_t - X_0 = \int_0^t a(s)ds + \int_0^t b(s)dW(s).$$

X_t is a Gaussian because it is a deterministic term plus a Gaussian random process $\int_0^t b(s)dW(s)$. The mean is

$$E[X_t] = X_0 + \int_0^t a(s)ds$$

where the fact of expectation of Ito integral is zero is used. For the calculation of variance, we use

$$E[(\int_0^t b(s)dW(s))^2] = \int_0^t b^2(s)ds$$

via Ito isomery. □

6.3.4.2 *State-independent linear geometric SDE*

Lemma 6.3.10 (general geometric SDE). *[6, p. 116] Consider the SDE*

$$dX(t) = \mu(t)X(t)dt + \sigma(t)X(t)dB(t).$$

It follows that

- *It has the equivalent form*

$$Y_t = \ln X_t$$
$$dY_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t$$

- *The solution for $X(t)$ is given by*

$$X(t) = X(0)\exp(\int_0^t [\mu(s) - \frac{1}{2}\sigma(s)^2]ds + \int_0^t \sigma(s)dB(s)).$$

- *Particularly, if $\mu(t) = 0$ and $\sigma(t)$ is a constant, then*

$$X(t) = X(0)\exp(-\frac{1}{2}\sigma^2 t + \sigma B(t))$$

is a martingale.

Proof. (1)(2) use $Y_t = f(X_t)\ln(X_t)$ and Ito rule, we have

$$\begin{aligned} dY_t &= \frac{\partial f}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}dX_t dX_t \\ &= \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}(\sigma X_t)^2 dt \\ &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t \end{aligned}$$

Then Y_t will have solution

$$Y_t = Y_t + \int_0^t (\mu - \frac{1}{2}\sigma^2)ds + \int_0^t \sigma dB_s.$$

(3) We want to prove $E[X(t)|\mathcal{F}_s] = X(s)$, where $\mathcal{F}_t$ is the filtration associated with Brownian motion. See Lemma 5.5.2. □

Corollary 6.3.3.1 (state independent geometric SDE, conversion to driftless SDE). *Consider SDE for X*

$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$

with constant μ, σ, and let $Y = \exp(-\mu t)X$, then the SDE for Y is

$$dY = \sigma Y(t)dB(t)$$

with solution of $Y(t)$ being an exponential martingale as

$$Y(t) = Y(0)\exp(-\frac{1}{2}\sigma^2 t + \sigma B(t)).$$

Then, $X(t)$ is given by

$$X(t) = X(0)\exp(\mu t - \frac{1}{2}\sigma^2 t + \sigma B(t))$$

Proof. From Ito lemma, we have

$$dY = -\mu \exp(-\mu t)Xdt + \exp(-\mu t)dX = \sigma Y(t)dB(t).$$

The rest can be proved using above lemma. □

Corollary 6.3.3.2 (mean and variance of a state-independent geometric SDE). *Consider SDE for X*

$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$

with constant μ, σ. Then,

- $E[X(t)] = X(0)e^{\mu t}$
- $Var[X(t)] = X(0)^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

Proof. Note that

$$\ln(\frac{X(t)}{X(0)}) = (\mu - \frac{\sigma^2}{2})t + \sigma W_t.$$

That is, $\frac{X(t)}{X(0)} \sim LN((\mu - \frac{\sigma^2}{2})t, \sigma^2 t)$.Then we can use Definition 2.2.6. □

Remark 6.3.3 (some implications).

- For geometric Brownian motion, once discounted with drift term, it becomes a martingale.

- This result is consistent with the finance in which the stock price S under martingale measure has the same drift r as risk free rate. As a consequence, when S is discounted with the risk-free rate, it becomes a martingale. In otherwise, martingale measure is the measure that make S a martingale when discounted.

6.3.4.3 *Integral of state-independent linear arithimetic SDE*

Lemma 6.3.11 (Integral of state independent arithmetic SDE). *Let X_t be governed by stochastic differential equation*

$$dX_t = a(t)dt + b(t)dW(t).$$

Further define a integral

$$I(t,T) = \int_t^T X(s)ds.$$

It follows that

- $$X_s = X_t + \int_t^s a(u)du + \int_t^s b(u)dW(u),$$

  which is a Gaussian distribution with mean $X_0 + \int_0^t a(s)ds$ and variance $\int_0^t b^2(s)ds$.
- *$I(t,T)$ has explicit form*

  $$I(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du + \int_t^T (T-u)b(u)dW(u).$$

- *$I(t,T)$ is a Gaussian distribution with mean and variance given by*

  $$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$

  $$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$

- *If $b(u) = b_0, a(u) = a_0$, then*

  $$M(t,T) = X_t(T-t) + \frac{1}{2}a_0(T-t)^2,$$

  $$V(t,T) = \frac{1}{3}b_0(T-t)^2.$$

Proof. (1) See Lemma 6.3.9. (2)

$$
\begin{aligned}
&\int_t^T X_s ds \\
=&\int_t^T X_t ds + \int_t^T \int_t^s a(u) du ds + \int_t^T \int_t^s b(u) dW(u) ds \\
=&X_t(T-t) + \int_t^T \int_u^T a(u) ds du + \int_t^T \int_u^T a(u) ds dW(u) \\
=&X_t(T-t) + \int_t^T (T-u) a(u) du + \int_t^T (T-u) b(u) dW(u)
\end{aligned}
$$

where we changed the order of integral. (3)(4) Use Lemma 6.3.9 again, we can see that $I(t,T)$ is actually a Gaussian process. □

Lemma 6.3.12 (Integral of sum of two state independent arithmetic SDE). *Let $X_1(t), X_2(t)$ be governed by stochastic differential equations*

$$
\begin{aligned}
dX_1(t) &= a_1(t)dt + b_1(t)dW_1(t) \\
dX_2(t) &= a_2(t)dt + b_2(t)dW_2(t) \\
E[dW_1 dW_2] &= \rho dt
\end{aligned}
$$

Further define a integral

$$I(t,T) = \int_t^T X_1(s) + X_2(s) ds.$$

It follows that

- $$X_1(s) + X_2(s) = X_1(t) + X_2(t) + \int_t^s a_1(u) + a_2(u) du + \int_t^s b_1(u) + b_2(u) dW(u),$$
- *$I(t,T)$ has explicit form*

  $$I(t,T) = (X_1(t) + X_2(t))(T-t) + \int_t^T (T-u) a(u) du + \int_t^T (T-u) b(u) dW(u),$$

  where

  $$a(u) = a_1(u) + a_2(u), b(u) = \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u) b_2(u)}$$

- *$I(t,T)$ is a Gaussian distribution with mean and variance given by*

$$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$

$$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$

- *If $b_1(u) = b_{10}, b_2(u) = b_{20}, a_1(u) = a_{10}, a_2(u) = a_{20}$, then*

$$M(t,T) = X_t(T-t) + \frac{1}{2}a_0(T-t)^2,$$

$$V(t,T) = \frac{1}{3}b_0(T-t)^2,$$

where

$$a_0 = a_{10} + a_{20}, b_0 = \sqrt{b_{10}^2 + b_{20}^2 + 2\rho b_{10}b_{20}}$$

Proof. Note that

$$\begin{aligned} d(X_1(t)+X_2(t)) &= (a_1(t)+a_2(t))dt + b_1(t)dW_1(t) + b_2(t)dW_2(t) \\ dZ(t) &= a(t)dt + \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u)b_2(u)}dW_3 \end{aligned}$$

where $Z(t) \triangleq X_1(t) + X_2(t)$, the W_3 is a new Brownian motion. We arrive at

$$b_1(t)dW_1(t) + b_2(t)dW_2(t) = \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u)b_2(u)}dW_3,$$

via the fact that two independent Gaussian random variable will sum to another Gaussian random variable. Then we use Lemma 6.3.11. □

Lemma 6.3.13 (Integral of sum of multiple state independent arithmetic SDE). *Let $X_1(t), X_2(t), ..., X_n$ be governed by stochastic differential equations*

$$\begin{aligned} dX_1(t) &= a_1(t)dt + b_1(t)dW_1(t) \\ dX_2(t) &= a_2(t)dt + b_2(t)dW_2(t) \\ &\cdots\cdots \\ dX_n(t) &= a_n(t)dt + b_n(t)dW_n(t) \\ E[dW_i dW_j] &= \rho_{ij}dt \end{aligned}$$

Further define a integral

$$I(t,T) = \int_t^T X_1(s) + X_2(s) + \cdots + X_n(s)ds.$$

It follows that

- $$X_1(s) + X_2(s) + \cdots + X_n(s) = \sum_{i=1}^n X_i(t) + \int_t^s \sum_{i=1}^n a_i(u)du + \int_t^s \sum_{i=1}^n b_i(u)dW(u),$$

- $I(t,T)$ *has explicit form*

  $$I(t,T) = (\sum_{i=1}^n X_i(t))(T-t) + \int_t^T (T-u)a(u)du + \int_t^T (T-u)b(u)dW(u),$$

  where

  $$a(u) = \sum_{i=1}^n a_i(u), b(u) = \sqrt{\sum_{i=1}^n b_i(u)^2 + 2\sum_{1\leq i<j\leq n} \rho_{ij} b_i(u) b_j(u)}$$

- $I(t,T)$ *is a Gaussian distribution with mean and variance given by*

  $$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$

  $$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$

6.3.4.4 *Multiple dimension extension*

Lemma 6.3.14 (multi-dimensional state independent/general arithmetic SDE). *[7, p. 146][6, p. 116]Consder a N dimensional stochastic differential equation(SDE) given by*

$$dX_i = a_i(t)dt + b_i(t)dW_i(t),$$

where $E[dW_i dW_j] = \rho_{ij}dt, E[dWdW^T] = \Sigma dt$. *It follows that*

- *The solution for* $X_i(t), i = 1,2,...,N$ *is given by*

  $$X_i(t) = X_i(0) + \int_0^t a_i(s)ds + \int_0^t b_i(s)dW_i(s),$$

  which is a ***Gaussian distribution*** *with mean* $X_i(0) + \int_0^t a_i(s)ds$ *and variance* $\int_0^t b_i^2(s)ds$.

- *The covariance structure for different $X_i(t), X_j(s), s \geq t$ is given by*

$$Cov(X_i(t), X_j(s)) = \int_0^t b_i(u)b_j(u)\rho_{ij}du.$$

Proof. (1) See Lemma 6.3.9. (2)

$$\begin{aligned}
Cov(X_i(t), X_j(s)) &= \int_0^t \int_0^s b_i(u)b_j(v)dW_i(u)dW_j(v) \\
&= \int_0^t \int_0^s b_i(u)b_j(v)\rho_{ij}\delta(u-v)du \\
&= \int_0^t b_i(u)b_j(u)\rho_{ij}du
\end{aligned}$$

□

Lemma 6.3.15 (general multi-dimensional geometric SDE). *[6, p. 116] Consider a N-dimensional SDE*

$$dX_i(t) = \mu_i(t)X_i(t)dt + \sigma_i(t)X_i(t)dW_i(t),$$

where It follows that

- *The solution for $X_i(t), i = 1, 2, ..., N$, is given by*

$$X_i(t) = X_i(0)\exp(\int_0^t [\mu_i(s) - \frac{1}{2}\sigma_i(s)^2]ds + \int_0^t \sigma_i(s)dW_i(s)).$$

- *Particularly, if $\mu_i(t) = 0$ and $\sigma_i(t)$ is a constant, then*

$$X_i(t) = X_i(0)\exp(-\frac{1}{2}\int_0^t \sigma_i^2(s)ds + \int_0^t \sigma_i(s)dW_i(s))$$

  is a martingale.
- *The covariance structure for different $X_i(t), X_j(s), s \geq t$ is given by*

$$Cov(X_i(t), X_j(s)) = X_i(0)X_j(0)\exp(m_i(t) + m_j(s) + \frac{1}{2}(\Sigma_{ii}(t,t) + \Sigma_{jj}(s,s)))(\exp(\Sigma_{ij}(t,s)) - 1),$$

  where

$$m_i(t) = \int_0^t [\mu_i(u) - \frac{1}{2}\sigma_i(u)^2]du,$$

$$\Sigma_{ij}(t,s) = \int_0^t \sigma_i(u)\sigma_j(u)dt.$$

- *Particularly, if $\mu_i(t) = 0$ and $\sigma_i(t)$ is a constant, then*

$$Cov(X_i(t), X_j(s)) = X_i(0)X_j(0)\exp(\Sigma_{ij}(t,s)) - 1.$$

$$E(X_i(t), X_j(s)) = X_i(0)X_j(0)\exp(\Sigma_{ij}(t,s)),$$

Proof. (1)(2) See Lemma 6.3.10. (3) Lemma 2.2.19 (4) Note that when $\mu_i = 0$, we have $m_i(t) + \frac{1}{2}\Sigma_{ii}(t,t) = 0$. □

6.3.5 Exact SDE

Definition 6.3.3 (exact SDE). *[7, p. 151] The SDE*

$$dX_t = a(t, W_t)dt + b(t, W_t)dW_t$$

is called exact if there is a differentiable function $f(t, W_t)$ such that

$$a(t, W_t) = f_t + \frac{1}{2}f_{WW}, b(t, W_t) = f_W$$

Lemma 6.3.16. *The solution to an exact SDE is given as*

$$X_t = f(t, W_t) + C$$

Proof. Use Ito's lemma, we have

$$dX_t = df = f_t dt + f_W dW_t + \frac{1}{2}f_{WW}dt$$

□

Remark 6.3.4. Not every SDE is exact. With a, b given, we can try to first solve for f(not necessarily solvable). If we can get f then obtain an easy way to solve SDE.

Example 6.3.7. [7, p. 151] We have

$$dX_t = e^t(1 + W_t^2)dt + (1 + 2e^t W_t)dW_t$$

We can find $f(t, W_t) = W_t + e^t W_t^2$

Theorem 6.3.4 (exact SDE criterion,necessary condition). *[7, p. 152] If SDE is exact, then*

$$a_x = b_t + \frac{1}{2}b_{xx}$$

6.3.6 Calculation mean and variance from SDE

Theorem 6.3.5 (Fubini's theorem). *[9, p. 53]Let $X(t)$ be a stochastic process with continuous sample paths, then*

$$\int_0^T E[|X(t)|]dt = E[\int_0^T |X(t)|\, dt]$$

furthermore if this quantity is finite, then

$$\int_0^T E[X(t)]dt = E[\int_0^T X(t)dt]$$

Remark 6.3.5. This theorem gives us the condition to exchange expectation and integral.

Given a SDE

$$X_t = X_0 = \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s$$

we are just interested in the mean and variance of X_t. We can use the fact that **expectation of Ito integral is zero** to simplify our calculation:

$$E[X_t] = X_0 + \int_0^t E[a(X_s, s)]ds$$

where the integral of expectation and integral is justified by **Fubini's theorem**.

Using the fundamental theorem of calculus, we know that

$$\frac{dE[X_t]}{dt} = E[a(X_t, t)]$$

Lemma 6.3.17. *[7, p. 142] Let $dX_t = a(t)X_t dt + c(t)dt + b(t)dW(t)$, then*

$$E[X_t] = \Phi(t, 0)X_0 + \int_0^t \Phi(t, \tau)c(\tau)d\tau,$$

where
$$\Phi(t,s) = \exp(\int_s^t a(u)du)$$

Proof. It is easy to find the governing equation for $E[X_t]$ is

$$dE[X_t]/dt = a(t)E[X_t] + c(t),$$

then use solution methods in linear dynamical system to solve. □

Lemma 6.3.18. *[7, p. 142] Let $dX_t = a(t)X_t dt + b(t)dW(t)$, then*

$$E[X_t] = \Phi_1(t,0)X_0$$

$$Var[X_t] = \int_0^t \Phi_2(t,\tau)b(\tau)^2 d\tau$$

where
$$\Phi_1(t,s) = \exp(\int_s^t a(u)du), \Phi_2(t,s) = \exp(\int_s^t 2a(u)du)$$

Proof. Let $Y_t = X_t^2$, then

$$dY_t = 2X_t dX_t + b(t)^2 dt = 2aY_t dt + b^2(t)dt + bdW(t),$$

use the above lemma, we have

$$E[Y_t] = E[X_t^2] = \Phi_2(t,0))X_0^2 + \int_0^t \Phi_2(t,\tau)b(\tau)^2 d\tau$$

Use $Var[X_t] = E[X_t^2] - E[X_t]^2 = \int_0^t \Phi_2(t,\tau)b(\tau)^2 d\tau$. □

Remark 6.3.6. We can also obtain the result using solutions to Ornstein-Uhlenbeck process Lemma 6.4.1.

6.3.7 Multi-dimensional Ito stochastic differential equations

Definition 6.3.4 (Multi-dimensional Ito SDE). *A system of N-dimensional Ito SDEs is defined as*

$$dx_i(t) = \mu_i(\boldsymbol{x}(t),t)dt + \sum_{j=1}^{M} \sigma_{ij}(\boldsymbol{x}(t),t)dw_j(t), \forall i = 1,...,d$$

where the dynamical system is driven by M Wiener process, and $dw_i(t_1)dw_j(t_2) = \delta_{ij}dt\delta(t_1 - t_2)$.

Note 6.3.1 (equivalence of two forms and redundancy of Brownian motion)**.**

- Consider the system of N-dimensional Ito SDEs:

$$dx_i(t) = \mu_i(\boldsymbol{x}(t), t)dt + h_j(\boldsymbol{x}(t), t)dz_j(t), \forall i = 1, ..., N$$

where the dynamical system is driven by N Wiener process, and $dz_i(t_1)dz_j(t_2) = \rho_{ij}dt\delta(t_1 - t_2)$.
To ensure the equivalence of the two forms, we have the following identity:

$$\sigma\sigma^T = H\rho H^T$$

If we know $H\rho H^T$, then we can calculate σ using Cholesky decomposition or eigendecomposition. If we know $\sigma\sigma^T$, then it is exactly the covariance matrix of N dimensional random vector $d\boldsymbol{z}$. It is possible that the covariance matrix is non-singular.
- If $M > N$(more Brownian motions than Ito process) and assume σ has full row rank, then there are redundant Brownian motion, and we can restructure such that the system is driven by N independent Brownian motion.

6.3.8 Ito vs. Stratonovich stochastic differential equations

Consider the system of d SDEs:

$$dx_i(t) = h_i(x_1(t), ..., x_d(t), t)dt + \sum_{j=1}^{m} H_{ij}(x_1(t), ..., x_d(t), t)dw_j(t), \forall i = 1, ..., d$$

where the dynamical system is driven by m Wiener process. When we say this is the Ito SDE, we mean that when we are doing the integral, we use the Ito version of the integral. And we can directly use the Euler-Maruyama method to do the numerical integration.

If we write it as Stratonovich version, we have

$$dx_i(t) = h_i(x_1(t), ..., x_d(t), t) - \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{d}\frac{\partial H_{ij}}{\partial x_k}H_{kj} + \sum_{j=1}^{m} H_{ij}(x_1(t), ..., x_d(t), t)dw_j(t), \forall i = 1, ..., d$$

Usually, Stratonovich SDE is used to describe/formulate a physical process, and then convert to Ito SDE when we want to numerically integrate it.

6.4 Ornstein-Uhlenbeck(OU) process

6.4.1 OU process

6.4.1.1 *Constant coefficient OU process*

Definition 6.4.1 (Ornstein-Uhlenbeck process). *A stochastic process*

$$X_t = e^{-at}x_0 + \sigma \int_0^t e^{-a(t-s)} dB_s,$$

where a, σ, x_0 are constant parameters and B_t is the Brownian motion, is called Ornstein-Uhlenbeck process with parameter (a, σ) and initial value x_0.

The differential form of the OU process is given by

$$dX_t = \sigma dB_t - aX_t dt, X_0 = x_0.$$

Lemma 6.4.1 (OU process solution). *Consider the SDE*

$$dX_t = \sigma dB_t - aX_t dt$$

with $\sigma > 0, a > 0$, and initial condition $X_0 = x_0$.

It follows that

- *It has the solution*

$$X_t = x_0 e^{-at} + \int_0^t \exp(-a(t-s))\sigma dB_s.$$

- *X_t has Gaussian distribution, i.e.,*

$$X_t \sim N(x_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

- *X_t has the stationary distribution given by*

$$X_t \sim N(0, \frac{\sigma^2}{2a}).$$

Proof. (1)(2) Use $Y_t = X_t e^{at}$, then Ito rule gives

$$dY_t = aY_t + e^{at} dX_t = e^{at}\sigma dB_t$$

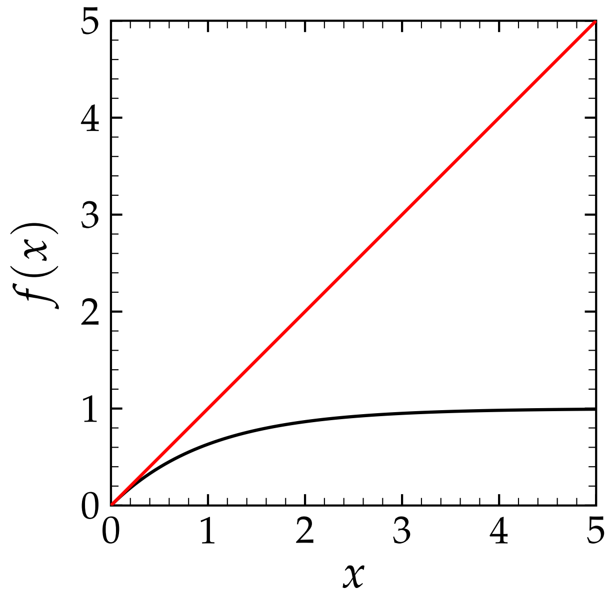


Figure 6.4.1: The variance function $Var[X(t)]$ for Brownian motion (red) and OU process(black) with $a = 0.5, \sigma = 1$.

We have

$$Y_T - Y_0 = \int_0^T e^{at}\sigma dB_t \Leftrightarrow X_T = \exp(-aT)X_0 + \int_0^T e^{-a(T-t)}dB_t.$$

Use Lemma 6.2.2, we have

$$Y_T - Y_0 \sim N(0, \int_0^T (e^{at}\sigma)^2 dt).$$

Then

$$X^T \sim e^{-aT}N(X_0, \int_0^T (e^{at}\sigma)^2 dt)$$

simplifies to

$$X^T \sim N(X_0, e^{-2aT}\int_0^T (e^{at}\sigma)^2 dt).$$

(3) Take $t \to \infty$ will get the result. □

Lemma 6.4.2 (constant shifted OU process). *Consider the constant shifted OU process*

$$dX_t = \sigma dB_t - a(X_t - \mu)dt$$

with $\sigma > 0, a > 0$, *and initial condition* $X_0 = x_0$.

- *It has the solution*

$$X_t \sim N((x_0-\mu)e^{-at}+\mu, \frac{\sigma^2(1-e^{-2at})}{2a})$$

and the stationary distribution is given as

$$X_t \sim N(\mu, \frac{\sigma^2}{2a}).$$

- *the constnat shifted OU process can be re-written as*

$$\begin{aligned} X_t &= Z_t + \mu \\ dZ_t &= \sigma dB_t - aZ_t dt \end{aligned}$$

Proof. (1) Use $Y_t = (X_t-\mu)e^{at}$. The rest is similar to Lemma 6.4.1. (2) Note that $dZ_t = dX_t + d\mu = dX_t$. Therefore

$$\begin{aligned} & dZ_t = \sigma dB_t - aZ_t dt \\ \implies & dX_t = \sigma dB_t - a(X_t-\mu)dt \end{aligned}$$

It can also be verified that:

$$X_t = \mu + Z_t, x_0 = \mu + z_0, Z_t \sim N(z_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a})$$

gives

$$X_t \sim N(\mu + (x_0-\mu)e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

□

Lemma 6.4.3 (scaling property of OU process). *Consider the SDE*

$$dX(t) = \sigma dB(t) - a(X(t)-\mu)dt,$$

and let $X(t)$ be the solution. Then $Y(t) = \lambda X(mt)$ is the solution for

$$dY(t) = \sqrt{m}\lambda\sigma dB(t) - ma(Y(t)-\lambda\mu)dt.$$

Note that we interpret m as the time scaling factor and λ the spatial scaling factor.

Proof. Note that $X(mt)$ will satisfy

$$dX(mt) = \sigma dB(mt) - a(X(mt)-\mu)dmt,$$

or equivalently
$$dX(mt) = \sqrt{m}\sigma dB(t) - ma(X(mt) - \mu)dt.$$
Multiply both sides by λ, we have
$$d\lambda X(mt) = \sqrt{m}\lambda\sigma dB(t) - ma(\lambda X(mt) - \lambda\mu)dt.$$
Plug in $\lambda X(mt) = Y(t)$, we have
$$dY(t) = \sqrt{m}\lambda\sigma dB(t) - ma(Y(t) - \lambda\mu)dt.$$
□

Remark 6.4.1 (applications of scaling property). Suppose we have the dynamics of an asset with time unit day and value unit dollar, we can use the scaling property to find out the coefficients associated with time unit year and value unit JPY.

Lemma 6.4.4 (Stationary Gaussian process). *An Ornstein-Uhlenbeck process (a, σ) with Gaussian initial distribution $\eta \sim N(0, \sigma^2/2a)$(i.e., stationary distribution) is a strictly/weakly stationary Gaussian process.*

Proof. (1)
$$E[X_t] = E[e^{-at}\eta + \sigma\int_0^t e^{-a(t-s)}dW_s] = 0$$
since $E[\eta] = 0$ and $\int_0^t e^{-a(t-s)}dW_s$ is Ito integral(Theorem 6.2.4). (2) Let $s < t$, we have
$$\begin{aligned}
cov(X_t, X_s) = E[X_tX_s] &= e^{-a(s+t)}E[\eta^2] + \sigma^2 E[\int_0^s e^{-a(t-s)}dW_u \int_0^s e^{-a(t-m)}dW_m] \\
&= e^{-a(s+t)}\frac{\sigma^2}{2a} + \sigma^2\int_0^t e^{-2a(t-s)}dt \\
&= e^{-a(s+t)}\frac{\sigma^2}{2a} + \frac{\sigma^2}{2a}(e^{-2as} - 1) \\
&= e^{-a(s+t)}e^{-2as}\frac{\sigma^2}{2a} = \frac{\sigma^2}{2a}e^{-a(t-s)}
\end{aligned}$$
Note that a weakly stationary Gaussian process is strictly Gaussian process(Lemma 5.2.4).
□

6.4.1.2 *Time-dependent coefficient OU process*

Definition 6.4.2 (Time-dependent coefficient Ornstein-Uhlenbeck process). *A stochastic process with differential form*
$$dX_t = (\phi(t) - \lambda X_t)dt + \sigma dW_t,$$

where $\psi(t)$ is time dependent coefficient, a, σ, x_0 are constants, and W_t is Brownian motion., is called time-dependent coefficient Ornstein-Uhlenbeck process with parameter (a, σ) and initial distribution x_0.

Lemma 6.4.5. *Consider a stochastic process with differential form*

$$dX_t = (\psi(t) - \lambda X_t)dt + \sigma dW_t, X_t = x_0$$

where $\psi(t)$ is time dependent coefficient, a, σ are constants, and W_t is Brownian motion. It follows that

- *It has the equivalent form*

$$X_t = Y_t + \int_0^t \exp(-a(t-s))\psi(s)ds$$
$$dY_t = -aY_t dt + \sigma dW_t$$

- *It has solution*

$$X_t = x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds + \int_0^t \sigma \exp(-a(t-s))dW_t.$$

- *X_t has mean and covariance given by*

$$E[X_t] = x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds$$

$$Var[X_t] = \frac{\sigma^2(1-e^{-2at})}{2a}$$

- *X_t has Gaussian distribution at any t, we have*

$$X_t \sim N(x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds, \frac{\sigma^2(1-e^{-2at})}{2a})$$

- *$X_t, t \to \infty$ is generally not a stationary process since its mean depends on t.*

Proof. (1)Note that

$$\frac{d}{dt}\int_0^t \exp(-a(t-s))\psi(s)ds = -a\int_0^t \exp(-a(t-s)) + \psi(t);$$

Therefore,

$$dX_t = dY_t + \frac{d}{dt}\int_0^t \exp(-a(t-s))\psi(s)ds = (\psi(t) - aX_t)dt + \sigma dW_t.$$

(2)(3) Note that Y_t has solution and distribution

$$Y_t = x_0 e^{-at} + \int_0^t \exp(-a(t-s))\sigma dB_s,$$

$$Y_t \sim N(x_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

Then we use relation in (1). □

Lemma 6.4.6. *Let $\psi(t) = \phi_n, t_n \leq t < t_{n+1}, t_0 = 0$, then*

$$E[X(t_n)] \triangleq \mu(t_n) = x_0 \exp(-at_n) + \exp(-at_n) \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1$$

$$\exp(at)\mu_n = x(0) + \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1$$

$$\begin{aligned}
\exp(at)\mu_n &= x(0) + \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1 \\
a\exp(at)\mu_n &= ax(0) + \sum_{k=1}^{n} \exp(at_k) - \exp(at_{k-1})\psi_{k-1} \\
a\exp(at)\mu_{n-1} &= ax(0) + \sum_{k=1}^{n-1} \exp(at_k) - \exp(at_{k-1})\psi_{k-1}, \\
\implies \psi_{n-1} &= \lambda \frac{\exp(at_n)\mu_n - \exp(\lambda t_{n-1})\mu_{n-1}}{\exp(at_n) - \exp(at_{n-1})}
\end{aligned}$$

6.4.1.3 *Integral of OU process*

Lemma 6.4.7 (integral of OU process). *Consider an OU process given by*

$$dx(t) = -ax(t)dt + \sigma dW(t), x(0) = x_0$$

where $a\sigma$ are constants,W is a Brownian motion. For each t, T, the random variable

$$I(t,T) = \int_t^T x(u)du$$

conditioned to the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \frac{1-\exp(-a(T-t))}{a}x(t),$$

and variance

$$V(t,T) = \frac{\sigma^2}{a^2}(T-t+\frac{2}{a}\exp(-a(T-t)) - \frac{1}{2a}\exp(-2a(T-t)) - \frac{3}{2a}).$$

Proof. See the proof of Lemma 6.4.8. □

Lemma 6.4.8 (integral of sum of two OU process). *[10, p. 145][11, p. 64] Consider two OU processes given by*

$$\begin{aligned} dx_1(t) &= -a_1x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

For each t, T, the random variable

$$I(t,T) = \int_t^T (x_1(u) + x_2(u))du$$

conditioned to the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \frac{1-\exp(-a_1(T-t))}{a_1}x_1(t) + \frac{1-\exp(-a_2(T-t))}{a_2}x_2(t),$$

and variance

$$\begin{aligned}V(t,T) = &\frac{\sigma_1^2}{a_1^2}(T-t+\frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}) \\ &+ \frac{\sigma_2^2}{a_2^2}(T-t+\frac{2}{a_2}\exp(-a_1(T-t)) - \frac{1}{2a_2}\exp(-2a_2(T-t)) - \frac{3}{2a_2}) \\ &+ \frac{2\rho\sigma_1\sigma_2}{a_1a_2}(T-t+\frac{\exp(-a_1(T-t))-1}{a_1} + \frac{\exp(-a_1(T-t))-1}{a_1} + \frac{\exp(-(a_1+a_2)(T-t))-1}{a_1+a_2}\end{aligned}$$

Proof. (1)Note that given the observation $x_1(t)$ at t, we have

$$x_1(u) = x_1(t)\exp(-a_1(u-t)) + \int_t^u \sigma\exp(-a_1(u-s))dW(s)$$

Therefore,

$$\begin{aligned}&\int_t^T x_1(u)du \\ &= \int_t^T x_1(t)\exp(-a_1(u-t))du + \int_t^T\int_t^u \sigma\exp(-a_1(u-s))dW(s)du \\ &= x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1} + \int_t^T\int_s^T \sigma\exp(-a_1(u-s))dudW(s) \\ &= x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1} + \int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)\end{aligned}$$

where we changed the order of integration. From this, we note that

$$E[\int_t^T x_1(u)du] = x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1}.$$

Similarly, we can get eh expectation for $\int_t^T x_2(u)du$.

(2) To get the variance, we have

$$Var[\int_t^T x_1(u)+x_2(u)du] = Var[\int_t^T x_1(u)du] + Var[\int_t^T x_1(u)du] + 2Cov(\int_t^T x_1(u)du, \int_t^T x_1(u)du).$$

For $Var[\int_t^T x_1(u)du]$, we have

$$\begin{aligned}
&Var[\int_t^T x_1(u)du \\
&= E[\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)] \\
&= \frac{\sigma_1^2}{a_1^2}(\int_t^T ds + \int_t^T \exp(-2a_1(T-s))ds - 2\int_t^T \exp(-a_1(T-s))ds) \\
&= \frac{\sigma_1^2}{a_1^2}(T-t+\frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}).
\end{aligned}$$

We can similarly evaluate $Var[\int_t^T x_2(u)du]$.

For $Cov[\int_t^T x_1(u)du, \int_t^T x_2(u)du]$, we have

$$\begin{aligned}
&Cov[\int_t^T x_1(u)du, \int_t^T x_2(u)du] \\
&= E[\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_2(s)] \\
&= \frac{\rho\sigma_1\sigma_2}{a_1a_2}(\int_t^T (1-\exp(-a_1(T-s)-\exp(-a_1(T-s))+\exp(-(a_1+a_2)(T-s))))ds \\
&= \frac{2\rho\sigma_1\sigma_2}{a_1a_2}(T-t+\frac{\exp(-a_1(T-t))-1}{a_1}+\frac{\exp(-a_2(T-t))-1}{a_2}+\frac{\exp(-(a_1+a_2)(T-t))-1}{a_1+a_2}).
\end{aligned}$$

□

Lemma 6.4.9 (integral of sum of multiple OU process). *[10, p. 145][11, p. 64] Consider n OU processes given by*

$$\begin{aligned}
dx_1(t) &= -a_1x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\
dx_2(t) &= -a_2x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \\
&\cdots\cdots \\
dx_n(t) &= -a_nx_2(t)dt + \sigma_n dW_1(t), x_2(0) = x_{n0}
\end{aligned}$$

where $a_1,...,a_n,\sigma_1,...,\sigma_n$ are constants, and $W_1,W_2,...,W_n$ are correlated Brownian motions such that

$$dW_i(t)dW_j(t) = \rho_{ij}dt.$$

For each t, T, the random variable

$$I(t,T) = \int_t^T (x_1(u) + x_2(u) + ... + x_n(u))du$$

conditioned on the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \sum_{i=1}^{n} \frac{1-\exp(-a_i(T-t))}{a_i} x_i(t),$$

and variance

$$V(t,T) = \sum_{i=1}^{n} \frac{\sigma_i^2}{a_i^2}(T - t + \frac{2}{a_i}\exp(-a_i(T-t)) - \frac{1}{2a_i}\exp(-2a_i(T-t)) - \frac{3}{2a_i})$$
$$+ \sum_{1\leq i<j\leq n} \frac{2\rho\sigma_i\sigma_j}{a_i a_j}(T - t + \frac{\exp(-a_i(T-t)) - 1}{a_i} + \frac{\exp(-a_j(T-t)) - 1}{a_j} + \frac{\exp(-(a_i+a_j)(T-t}{a_i + a_j}$$

6.4.2 Exponential OU process

Definition 6.4.3 (exponential constant coefficient Ornstein-Uhlenbeck process). *A stochastic process with differential form*

$$d(\ln X_t) = -a\ln X_t dt + \sigma dB_t, X_0 = x_0,$$

where a, σ, x_0 are constant parameters and B_t is the Brownian motion, is called exponential constant coefficient Ornstein-Uhlenbeck process with parameter (a, σ) and initial distribution x_0.

It also has the equivalent form

$$X_t = \exp(Y_t)$$
$$dY_t = -aY_t dt + \sigma dB_t, Y_0 = \ln x_0$$

Lemma 6.4.10 (exponential OU process solution). *Consider the SDE*

$$d(\ln X_t) = -a\ln X_t dt + \sigma dB_t, X_0 = x_0,$$

with $\sigma > 0, a > 0$, and initial condition $X_0 = x_0$.

It follows that

- *It has the solution*
$$X_t = \exp(Y_t), Y_t \sim\sim N(\ln(x_0)e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$
and the stationary distribution is given as
$$X_t = \exp(Y_t), Y_t \sim N(0, \frac{\sigma^2}{2a}).$$
- *(mean and variance property)*
$$\begin{aligned} E[X_t] &= \exp(\mu_Y + \sigma_Y^2/2) \\ Var[X_t] &= (\exp(\sigma_Y^2) - 1)\exp(2\mu_Y + \sigma_Y^2) \end{aligned}$$
where
$$\mu_Y = \ln(x_0)e^{-at}, \sigma_Y^2 = \frac{\sigma^2(1-e^{-2at})}{2a}.$$

Proof. (1) Let $Y_t = \ln X_t$, then we have
$$dY_t = -aY_t dt + \sigma dB_t, Y_0 = \ln x_0.$$
From Lemma 6.4.1, we know that
$$Y_t \sim N(Y_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$
(2) Use the property of log normal distribution(Lemma 2.2.14) □

Remark 6.4.2 (sanity check with Ito rule). Note that the SDE
$$d(\ln X_t) = -a \ln X_t dt + \sigma dB_t, X_0 = x_0,$$
will give the SDE for X_t via the equivalent form
$$\begin{aligned} X_t &= f(Y_t) = \exp(Y_t) \\ dY_t &= -aY_t dt + \sigma dB_t, Y_0 = \ln x_0. \end{aligned}$$

Using Ito rule, we have

$$\begin{aligned} dX_t &= \frac{\partial f}{\partial Y_t} dY_t + \frac{1}{2}\frac{\partial^2 f}{\partial Y_t^2} dY_t dY_t \\ &= \exp(Y_t) dY_t + \frac{1}{2}\exp(Y_t)\sigma^2 dt \\ \implies dX_t/X_t &= dY_t + \frac{1}{2}\sigma^2 dt \\ dX_t/X_t &= (-a\ln X_t + \frac{1}{2}\sigma^2)dt + \sigma dB_t \end{aligned}$$

Lemma 6.4.11 (solving ψ from imposed term structure mean).

$$\psi_{n-1} = \lambda \frac{\exp(\lambda t_n)\ln \mu_n - \exp(\lambda t_{n-1})\ln \mu_{n-1} - \frac{\sigma^2}{4\lambda}(\exp(\lambda t_n) - \exp(-\lambda t_n) - \exp(\lambda t_{n-1}) + \exp(-\lambda t_n}{\exp(\lambda t_n) - \exp(\lambda t_{n-1})}$$

6.4.3 Parameter estimation for OU process

Note 6.4.1. The OU process

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

can be discretized at times $n\Delta t, n = 1, 2, ..., \infty$ which gives

$$X_{k+1} - X_k = k\theta\Delta t - kX_k\Delta t + \sigma(W_{k+1} - W_k),$$

or equivalently,

$$X_{k+1} = k\theta\Delta t - (k\Delta t - 1)X_k + \sigma\sqrt{\Delta t}\epsilon_k,$$

where $\epsilon_k \sim WN(0, 1)$.

The discrete-time form can be viewed as an AR(1) process, and least square method can be used to estimate k, θ, σ.

6.4.4 Multiple factor extension

Definition 6.4.4 (two-factor OU process). *The two-factor OU process is given by the following SDE*

$$\begin{aligned} r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\ dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

Lemma 6.4.12 (basic properties). *Consider a The two-factor OU process is given by the following SDE*

$$\begin{aligned} r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\ dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

It follows that

- *It has solution given by*

$$\begin{aligned} r(t) &= x_{10}\exp(-a_1 t) + x_{20}\exp(-a_2 t) \\ &\quad + \sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s) + \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s) + \psi(t). \end{aligned}$$

- $$E[r(t)] = x_{10}\exp(-a_1 t) + x_{20}\exp(-a_2 t) + \psi(t).$$

$$Var[r(t)] = \frac{\sigma_1^2}{2a_1}(1-\exp(-2a_1 t)) + \frac{\sigma_2^2}{2a_2}(1-\exp(-2a_2 t)) + \frac{2\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t)).$$

- *$r(t)$ has Gaussian distribution; that is,*

$$r(t) \sim N(E[r(t)], Var[r(t)]).$$

Proof. (1) From single factor OU process result(Lemma 6.4.1), we know that

$$x_1(t) = x_{10}\exp(-a_1 t) + \sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s).$$

Similarly, we can evaluate $x_2(t)$ and eventually $r(t)$. (2) The expectation can be easily evaluated based on the fact that Ito integral has zero mean. To evaluate the variance we have

$$\begin{aligned} Var[r(t)] &= Var[\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s)] + Var[\sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)] \\ &+ 2Cov(\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s), \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)) \\ &= \int_0^t \sigma_1^2 \exp(-2a_1(t-s))ds + \int_0^t \sigma_2^2 \exp(-2a_1(t-s))ds + \int_0^t \sigma_1\sigma_2\rho \exp(-a_1(t-s))\exp(-a_2(\\ &= \frac{\sigma_1^2}{2a_1}(1-\exp(-2a_1 t)) + \frac{\sigma_2^2}{2a_2}(1-\exp(-2a_2 t)) + \frac{2\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t)) \end{aligned}$$

where we use Ito isometry in the evaluation, for example,

$$\begin{aligned} &E[\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s) \cdot \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)] \\ &= E[\sigma_1\sigma_2 \int_0^t \int_0^t \exp(-a_1(t-s))\exp(-a_2(t-u))dW_1(s)dW_2(u)] \\ &= E[\sigma_1\sigma_2 \int_0^t \int_0^t \exp(-a_1(t-s))\exp(-a_2(t-u))\rho dt\delta(u-s)] \\ &= E[\sigma_1\sigma_2\rho \int_0^t \exp(-(a_1+a_2)(t-s))\rho dt\delta(u-s)] \\ &= \frac{\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t)) \end{aligned}$$

(3) The random variable $r(t) = x_1(t) + x_2(t)$ is a Gaussian process has been discussed in Theorem 5.3.2. □

6.5 Brownian bridge

6.5.1 Constructions

Definition 6.5.1 (standard Brownian bridge). *A Brownian bridge is a stochastic process $\{X_t, t \in [0,1]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- $X_0 = 0, X_1 = 0$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = 0$.
- $Cov(X_s, X_t) = \min(s,t) - st, \forall s,t \in [0,1]$.
- $Var[X_s] = s - s^2$.
- X_t *is almost surely continuous.*

Remark 6.5.1 (calculate covariance using conditional distribution). The joint distribution of (X_t, X_1) is a multivariate Gaussian with mean

$$\mu = (0,0)^T, \Sigma = \begin{bmatrix} t & t \\ t & 1 \end{bmatrix}$$

based on the property of standard Brownian motion(Lemma 6.1.1). Then

$$(X_t|X_1) \sim MN(0, t - t^2)$$

from Theorem 2.2.2. Similarly, the joint distribution of (X_s, X_t, X_1) is normal, and $(X_s, X_t|X_1) \sim MN(0, \min(s,t) - st)$.

Definition 6.5.2 (Brownian bridge, general state space). *A Brownian bridge is a stochastic process $\{X_t, t \in [0,1]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- $X_0 = a, X_1 = b$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = (1-t)a + tb$.
- $Cov(X_s, X_t) = \min(s,t) - st, \forall s,t \in [0,1]$.
- X_t *is almost surely continuous.*

Definition 6.5.3 (Brownian bridge, general temporal space). *A Brownian bridge is a stochastic process $\{X_t, t \in [p,q]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- $X_p = 0, X_q = 0$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = 0$.

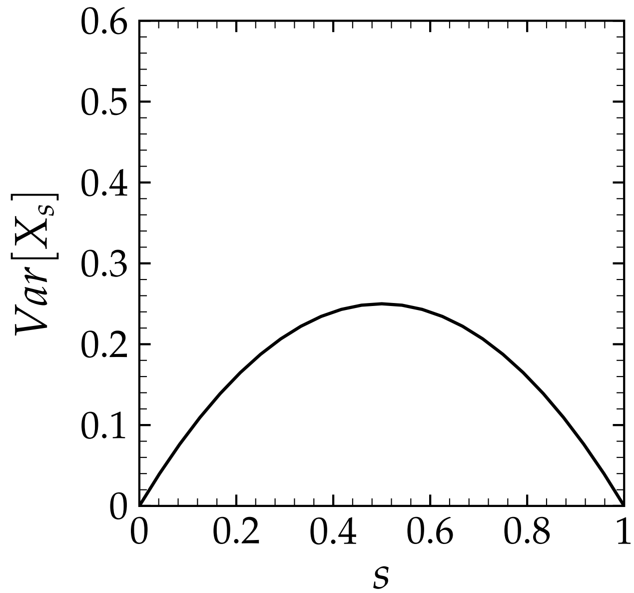


Figure 6.5.1: Variance of X_t in a Brownian bridge

- $$Cov(X_s, X_t) = \min(s-p, t-p) - \frac{(s-p)(t-p)}{q-p}, \forall s,t \in [p,q]$$
  .
- *X_t is almost surely continuous.*

Definition 6.5.4 (Brownian bridge, general state space and temporal space). *A Brownian bridge is a stochastic process $\{X_t, t \in [p,q]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- *$X_p = a, X_q = b$ almost surely.*
- *X_t is a Gaussian process.*
- $E[X_t] = (1 - \frac{t-p}{q-p})a + \frac{t-p}{q-p}b.$
- $$Cov(X_s, X_t) = \min(s-p, t-p) - \frac{(s-p)(t-p)}{q-p}, \forall s,t \in [p,q]$$
  .
- *X_t is almost surely continuous.*

Lemma 6.5.1 (construction of standard Brownian bridge).

- *Suppose Z_t is a standard Brownian motion. Let $X_t = Z_t - tZ_1, t \in [0,1]$.Then X_t is a Brownian bridge process.*

- *Suppose that* $\{Z_t, t \in [0,\infty)\}$ *is standard Brownian motions. Define* $X_1 = 0$, *and*
$$X_t = (1-t)\int_0^t \frac{1}{1-s} dZ_s, t \in [0,1).$$
Then X_t *is a Brownian Bridge. Moreover, the stochastic process has the differential form as*
$$dX_t = dZ_t - \frac{X_t}{1-t} dt.$$

Proof. (1)(a)$X_0 = X_1 = 0$. (b) The random vector $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ can be constructed using affine transformation using random vector $(Z_{t_1}, Z_{t_2}, ..., Z_{t_n}, Z_1)$. Therefore, X_t is also a Gaussian process.(Theorem 2.2.1). (c) $E[X_t] = 0$. (d) $Cov(Z_s - sZ_1, Z_t - tZ_1) = \min(s,t) - st$.(e) X_t is continuous since Z_t, tZ_1 is continuous.

(2) (d)Note that X_t is a zero Gaussian process(Lemma 6.2.2,Lemma 5.3.3). Then
$$Cov(X_t, X_s) = (1-t)(1-s)\int_0^s \frac{1}{(1-u)^2} du = s - st.$$

To prove the differential form, we have
$$\begin{aligned}
X_t &= (1-t)\int_0^t \frac{1}{1-s} dZ_s \\
dX_t &= \int_0^t \frac{1}{1-s} dZ_s d(1-t) + (1-t)d(\int_0^t \frac{1}{1-s} dZ_s) \\
&= -\int_0^t \frac{1}{1-s} dZ_s + (1-t)\frac{1}{1-t} dZ_t \\
&= -\frac{X_t}{1-t} + dZ_t
\end{aligned}$$
□

Lemma 6.5.2 (construction generalized Brownian bridge). *Let* $W(t)$ *be a standard Brownian motion.*

- *Fix* $a \in \mathbb{R}, b \in \mathbb{R}$. *We can construct the Brownian bridge from a to b on* $[0,1]$ *to be the process*
$$Y(t) = a + (b-a)t + X(t),$$
where $X(t)$ *is a standard Brownian bridge from 0 to 0 in time* $[0,1]$.
- *Fix* $p, q \in \mathbb{R}$. *We can construct the Brownian bridge from* 0 *to* 0 *on* $[p,q]$ *to be the process*
$$Y(t) = X(\frac{t-p}{q-p}),$$

where $X(t)$ is a standard Brownian bridge from 0 to 0 in time $[0,1]$.

- *Fix $a,b,p,q \in \mathbb{R}$. We can construct the Brownian bridge from a to b on $[p,q]$ to be the process*
$$Y(t) = a + (b-a)\frac{t-p}{q-p} + X(\frac{t-p}{q-p}),$$
where $X(t)$ is a standard Brownian bridge from 0 to 0 in time $[0,1]$.

Proof. (1) straight forward.(2) □

6.5.2 Simulation

Remark 6.5.2 (simulation of Brownian bridge). We can simulate a Brownian bridge by first simulating a Wiener process W_t and then using

$$X_t = W_t - tW_1$$

to construct.[12, p. 27]

6.5.3 Applications

Remark 6.5.3 (general remarks). A Brownian bridge is used when you know the values of a Wiener process at the beginning and end of some time period, and want to understand the probabilistic behavior in between those two time periods.

Remark 6.5.4 (Brownian bridge as interpolation method). Suppose we have generated a number of points $W(0), W(1), W(2), W(3)$, etc. of a Wiener process path by computer simulation. We can use Brownian bridge simulation will interpolate path between $W(1)$ and $W(2)$.

Remark 6.5.5 (applications of Brownian bridge in bond). In the case of a long-term discount bond with known payoff at final term, we need to simulate values of the asset over a longer period of time such that the stochastic process is conditional on reaching a given final state. For example, take the case of a discount bond such as a 10 year Treasury bond. If we model a discount bond price as a stochastic process, then this process should be tied to the final state of the process.

6.6 Martingale representation theorem

Theorem 6.6.1. *[1, p. 221]Let $W(t)$ be a Brownian motion, and let $\mathcal{F}(t)$ be the Brownian motion filtration. Let $M(t)$ be a martingale with respect to $\mathcal{F}(t)$, then there is **a stochastic process** $h(t)$ **adapted to** $\mathcal{F}(t)$ such that*

$$M(t) = M(0) + \int_0^t h(u)dW(u)$$

Remark 6.6.1 (interpretation)**.**

- If Brownian motion is the only source of randomness, then a continuous martingale, say M, can be expressed as a driftless SDE driven by Brownian motion, i.e.,

  $$dM(t) = h(t, \omega)dB(t)$$

  or

  $$M(t) = M(0) + \int_0^t h(s, \omega)dB(s)$$

  where $h(s, \omega)$ is a random process.
- Similarly, if a stochastic process is represented by

  $$dM(t) = h(t, \omega)dB(t)$$

  then $M(t)$ is a martingale.

Remark 6.6.2. Due to the martingale property of Ito integral, we can show that $E[M(t)] = M(0)$.

Example 6.6.1. If the random process M is known to be a martingale which is a function of t and random process X, whose SDE is $dX(t) = \mu(t, X_t)dt + \sigma(t, X_t)dB(t)$ then by taking derivative using Ito formula $dM(t) = df(t, X_t)$, we can get

$$dM(t) = \sigma(t, X_t)\frac{\partial M}{\partial X}dB(t)$$

and

$$h(t, \omega) = \sigma(t, X_t)\frac{\partial M}{\partial X}$$

where we have setting the drift term to zero.

6.7 Change of measure

6.7.0.1 *Change of measure concepts for probability space*

Definition 6.7.1 (absolute continuity). *[13, p. 34] Let $(\Omega, \mathcal{F})$ be a measurable space. Let P and Q be two different probability measures. We say Q is **absolutely continuous** with respect to P if for any subset $A \in \mathcal{F}$,*

$$P(A) = 0 \implies Q(A) = 0,$$

and it is denoted as $Q \ll P$.

Definition 6.7.2 (equivalent probability measure). *[1, p. 35]Let $(\Omega, \mathcal{F})$ be a measurable space. Let P and $\tilde{P}$ be two different probability measures. We say these two probability measure are equivalent if they agree which sets in $\mathcal{F}$ have probability zero.*

In other words, P and Q are equivalently if $P \ll Q$ and $Q \ll P$.

Remark 6.7.1. Equivalent probability measure means that for $A \in \mathcal{F}$, if $P(A) > 0$ then $\tilde{P}(A) > 0$; if $P(A) = 0$ then $\tilde{P}(A) = 0$.

Theorem 6.7.1 (existence of Radon-Nikodym derivatives). *[1, p. 39] Let P and $\tilde{P}$ be two equivalent probability measure. Then there exist an almost surely positive random variable Z such that $E[Z] = 1$ and*

$$\tilde{P}(A) = \int_A Z(\omega) dP(\omega)$$

6.7.1 Change of measure for random variables

Theorem 6.7.2 (generating new probability measure and equivalence). *[1, p. 33]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let Z be an almost non-negative random variable with $E[Z] = 1$, for $A \in \mathcal{F}$, define*

$$Q(A) = \int_A Z(\omega) dQ(\omega)$$

then

- *Q is a probability measure.*

- *For any random variable X, we have*

$$E_Q[X] = E_P[XZ].$$

- *If $Z > 0$ almost surely, we also have*

$$E_P[Y] = E_Q[\frac{Y}{Z}].$$

- *If $Z > 0$ almost surely, then P and Q are equivalent.*

Proof. (1)(2)(3) See reference. (4) Suppose $A \in \mathcal{F}$ and $P(A) = 0$. Then

$$\tilde{P}(A) = \tilde{E}[\mathbf{1}_A A] = \int \mathbf{1}_A Z(\omega) dP(\omega) = \int_A Z(\omega) dP(\omega) = 0,$$

since $P(A) = \int_A dP(\omega) = 0, Z(\omega) \geq 0$. On the other hand, suppose $B \in \mathcal{F}, (B) = 0$, then

$$P(B) = E[\frac{\mathbf{1}_B}{Z}] = 0.$$

□

Remark 6.7.2.

- For a measurable space, different probability measure can be defined, as long as they satisfy the requirement of probability measure (Definition 1.2.2). Different probability measures only differ at assigning different values to measurable sets, i.e., elements in $\mathcal{F}$.
- On a measurable space, different random variables can be defined. In nature, random variables are just measurable maps, and are not associated with a particular probability measure. Therefore, when taking the expectation of a random variable, we can specify which probability measure is with respect to.
- This theorem provides a way to take expectation with respect to different probability measures.

Theorem 6.7.3 (change of measure for Gaussian variables). *Let $X = \mu + \sigma W, W \sim N(0,1)$, and $Z = \exp(\sigma_Y Y - \frac{1}{2}\sigma_Y^2), Y \sim N(0,1)$. Let W and Y be jointly normal with correlation be ρ. It follows that*

- $E[Z] = 1.$
- *Under the equivalent measure Q generated by Z and for a function g, we have*

$$E_Q[g(X)] = E[Zg(X)] = E[g(X')],$$

where $X' \sim N(\mu + \rho\sigma_Y\sigma, \sigma^2)$.

- *Particularly, if $Y = W$, then under the equivalent measure Q generated by Z, $X \sim N(\mu + \sigma_Y, \sigma_Y^2)$.*

Proof. (1) $E[Z] = E[\exp(\sigma_Y Y - \frac{1}{2}\sigma_Y^2)] = M_Y(\sigma_Y)\exp(-\frac{1}{2}\sigma_Y^2) = 1$, where $M_Y(t) = \exp(\frac{1}{2}t^2)$ is the mgf of Y. (2) The measure is equivalent because $Z > 0$ almost surely(Theorem 6.7.2). We want to the mgf of X under measure generated from Z, we have

$$\begin{aligned}
M_X(t) &= E_Q[\exp(tX)] \\
&= E[Z\exp(tX)] \\
&= E[\exp(\sigma_Y Y - \frac{1}{2}\sigma_Y^2)\exp(t\mu + t\sigma W)] \\
&= E[\exp((\sigma_Y Y + t\sigma W)]\exp(-\frac{1}{2}\sigma_Y^2 + t\mu) \\
&= M_{Y,W}(\sigma_Y, t\sigma))\exp(-\frac{1}{2}\sigma_Y^2 + t\mu) \\
&= \exp(\frac{1}{2}(\sigma_Y^2 + 2\rho t\sigma_Y\sigma + t^2\sigma^2))\exp(-\frac{1}{2}\sigma_Y^2 + t\mu) \\
&= \exp(t(\mu + \rho\sigma\sigma_Y) + \frac{1}{2}t^2\sigma^2)
\end{aligned}$$

where $M_{Y,W}(t_1, t_2) = \exp(\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2))$ is the moment generating function for joint Gaussian. From the above, we can see, under the measure generated by Z, $X \sim N(\mu + \rho\sigma_Y\sigma, \sigma^2)$. (3) Take $\sigma_Y = \sigma, \rho = 1$. $\square$

Lemma 6.7.1 (change of measure for Bivariate Gaussian variables). *Let $X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$ and (X, Y) has correlation coefficient ρ. Then for function g, we have*

$$E[\exp(X)g(Y)] = E[\exp(X)]E[g(Y')],$$

where $Y' \sim N(\mu_Y + \rho\sigma_X\sigma_Y, \sigma_Y^2)$

Proof. Let $X = \mu_X + \sigma_X\rho Z_2 + \sigma_X\sqrt{1-\rho^2}Z_1, Y = \mu_Y + \sigma_Y Z_2$, where Z_1, Z_2 are independent standard normals. Then

$$\begin{aligned}
E[\exp(X)g(Y)] &= E[\exp(\mu_X + \sigma_X\rho Z_2 + \sigma_X\sqrt{1-\rho^2}Z_1)g(\mu_Y + \sigma_Y Z_2)] \\
&= E[\exp(\mu_X + \sigma_X\rho Z_2 + \sigma_X\sqrt{1-\rho^2}Z_1)]E[\exp(\sigma_X\rho Z_2)g(\mu_Y + \sigma_Y Z_2)] \\
&= E[\exp(\mu_X + \sigma_X\sqrt{1-\rho^2}Z_1)]E[\exp(\sigma_X\rho Z_2)g(\mu_Y + \sigma_Y Z_2)] \\
&= \exp(\mu_X + \frac{1}{2}\sigma_X^2(1-\rho^2))]E[\exp(\sigma_X\rho Z_2)g(\mu_Y + \sigma_Y Z_2)] \\
&= \exp(\mu_X + \frac{1}{2}\sigma_X^2)]E[\exp(\sigma_X\rho Z_2 - \frac{1}{2}\sigma_X\rho^2)g(\mu_Y + \sigma_Y Z_2)] \\
&= E[\exp(X)]E[\exp(\sigma_X\rho Z_2 - \frac{1}{2}\sigma_X\rho^2)g(\mu_Y + \sigma_Y Z_2)] \\
&= E[\exp(X)]E[g(Y')], Y' \sim N(\mu_Y + \rho\sigma_X\sigma_Y, \sigma_Y^2)
\end{aligned}$$

where in the last line we use the change of measure for Gaussian random variable(Theorem 6.7.3). □

6.7.2 Change of measure for stochastic process

Definition 6.7.3 (Radon-Nikodym derivative). *[1, p. 37]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\tilde{P}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to P, and let Z be an **almost surely positive** random variable such that $E[Z] = 1$ and relates P and $\tilde{P}$ via*

$$\tilde{P}(A) = \int_A Z(\omega)dP(\omega)$$

then Z is called a Radon-Nikodym derivatives of $\tilde{P}$ with respect to P, and we write

$$Z = \frac{d\tilde{P}}{dP}.$$

Definition 6.7.4 (Radon-Nikodyn derivative process). *[1, p. 211] Let Z be an almost surely positive random variable satisfying $EZ = 1$, then the stochastic process*

$$Z(t) = E[Z|\mathcal{F}_t]$$

is called the Radon-Nikodyn derivative process.

Lemma 6.7.2. *If P and Q are equivalent measures, and X_t is an $\mathcal{F}_t$ adapted process, then*

$$E_Q[X_t] = E_P[\frac{dQ}{dP}X_t].$$

Proof.

$$E_Q[X_t] = \int_\Omega X_t dQ = \int_\Omega X_t \frac{dQ}{dP} dP.$$

□

Remark 6.7.3. This is the foundation of importance sampling in Monte Carlo methods (**??**).

Theorem 6.7.4 (Radon-Nikodyn derivative process applied to change of measure). *[1, p. 212] Let Z be an almost surely positive random variable satisfying $EZ = 1$. It follows that*

- *The stochastic process defined via conditional expectation, which is the Radon-Nikodyn derivative process,*
  $$Z(t) = E[Z|\mathcal{F}_t]$$
  is a martingale.
- *$Z(t)$ is almost surely positive and $E[Z(t)] = 1$.*
- *The probability measure Q generated by Z and the probability measure Q_t generated by $Z(t)$ agree on the $\mathcal{F}_t$; that is*
  $$Q_t(A) = Q(A), \forall A \in \mathcal{F}_t.$$
- *If Y is $\mathcal{F}_t$ measurable random variable, then*
  $$E_Q[Y] = E[YZ] = E[YZ(t)],$$
  that is, we can evaluate the expectation under measure Q generated by Z or under measure $Q(t)$ generated by $Z(t)$.

Proof. (1) Note the conditional expectations are martingales (Theorem 5.5.1). Briefly,

$$E[Z(t)|\mathcal{F}(s)] = E[E[Z|\mathcal{F}_t]|\mathcal{F}_s] = E[Z|\mathcal{F}_s] = Z(s).$$

Also, see Theorem 5.5.1. (2)$E[Z(t)] = E[E[Z|\mathcal{F}_t]] = E[Z] = 1$. Because $Z > 0$ almost surely, its partial averaging

$$0 < E[Z|\mathcal{F}_t] = \int_A Z(\omega)dP(\omega), \forall A \in \mathcal{F}_t, P(A) > 0.$$

(3) The new measure Q_t generated by $Z(t)$ and the new measure generated by $Z(T)$ agrees on the sets in $\mathcal{F}_t$. Let $A \in \mathcal{F}_t \subset \mathcal{F}_T, t \leq T$, and use $Z(t) = E[Z(T)|\mathcal{F}_t]$, we have

$$\begin{aligned}
Q_t(A) &= \int_\Omega I_A Z(t) dP = E[Z(t) I_A] \\
&= E[I_A E[Z(T)|\mathcal{F}_t]] \\
&= E[E[I_A Z(T)|\mathcal{F}_t]] \\
&= E[I_A Z(T)] = \int_\Omega I_A Z(T) dP = Q_T(A)
\end{aligned}$$

Note that we use the fact that the random variable I_A is $\mathcal{F}_t$ measurable(i.e. $I_A^{-1}(1) = A \in \mathcal{F}_t, I_A^{-1}(0) = A^C \in \mathcal{F}_t$). (4)

$$E_Q[Y] = E[YZ] = E[E[YZ|\mathcal{F}_t]] = E[YE[Z|\mathcal{F}_t]] = E[YZ(t)],$$

where we use the fact that Y is $\mathcal{F}_t$ measurable and can be taken out of expectation. $\square$

Theorem 6.7.5 (Bayes theorem for conditional expectation, change of measure for conditional expectation). *[1, p. 212] Let Z be an almost surely positive random variable satisfying $E[Z] = 1$. Let Q be a new probability measure generated by Z. Let Y be a $\mathcal{F}_t$ measurable random variable, then*

$$E_Q[Y|\mathcal{F}_s] = E[YZ|\mathcal{F}_s] = \frac{E[YZ(t)|\mathcal{F}_s]}{E[Z(t)|\mathcal{F}_s]} = \frac{1}{Z(s)} E[YZ(t)|\mathcal{F}_s]$$

where $s \leq t$.

Particularly, if $\mathcal{F}_0 = \{\emptyset, \Omega\}$, we have

$$E_Q[Y|\mathcal{F}_s] = E[YZ|\mathcal{F}_s] = E[YZ(t)|\mathcal{F}_s]$$

Proof. First, $E_Q[Y|\mathcal{F}_s]$ and $\frac{E[YZ(t)|\mathcal{F}_s]}{E[Z(t)|\mathcal{F}_s]}$ are both $\mathcal{F}_s$ measurable random variables. To show that these two random variables are equal, we need to check for every $A \in \mathcal{F}_s$, we have

$$E_Q[Y|\mathcal{F}_s](A) = \frac{E[YZ(t)|\mathcal{F}_s]}{E[Z(t)|\mathcal{F}_s]}(A).$$

We have

$$\begin{aligned}\frac{E[YZ(t)|\mathcal{F}_s]}{E[Z(t)|\mathcal{F}_s]}(A) &= E_Q[I_A\frac{E[YZ(t)|\mathcal{F}_s]}{E[Z(t)|\mathcal{F}_s]}] \\ &= E_Q[I_A\frac{E[YZ(t)|\mathcal{F}_s]}{Z(s)}] \\ &= E[I_A E[YZ(t)|\mathcal{F}_s]] \\ &= E[I_A YZ(t)] \\ &= E[I_A YZ] \ (*) \\ &= E_Q[I_A Y] \\ &= E_Q[Y](A) \\ &= \int_A Y(\omega)dQ(\omega) \\ &= E_Q[Y|\mathcal{F}_s](A)\end{aligned}$$

where in line *, we use Theorem 6.7.4 and the fact that Y is $\mathcal{F}_t$ measurable. □

Example 6.7.1. [1, p. 37] Let P be a probability measure, let X be a random variable and it has normal distribution under P. Let $Y = X + \theta$, where θ is a positive constant. Under probability measure P, $E[Y] = \theta$, $Var[Y] = 1$, and therefore Y has a shifted normal distribution. We can introduce a new probability measure $\tilde{P}$, which assign less weight to subsets $X^{-1}(X(\omega) > 0)$, such that Y is normal under $\tilde{P}$.

The probability density function under the P measure is $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$. Let $Z(\omega) = \exp(-\theta X(\omega) - \frac{1}{2}\theta^2)$, the probability density function for the new probability measure is

$$\tilde{f}(x) = Zf(x) = \exp(-\theta x - \frac{1}{2}\theta^2)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2) = \exp(-\frac{1}{2}(\theta + x)^2)$$

Under this new probability density function for X, we have

$$E[X] = \int_{-\infty}^{\infty}(x+\theta)\tilde{f}(x)dx = \theta$$

note that $f_Y(y) = \tilde{f}(y-\theta)\frac{dY}{dX} = \tilde{f}(y-\theta)$, $E[Y] = \int_{-\infty}^{\infty}(y)f_Y(y)dy = \int_{-\infty}^{\infty} y\exp(-\frac{1}{2}(\theta + y - \theta)^2)dy = 0$

6.7.3 Change of measure for Brownian motions: Girsanov's theorem

Theorem 6.7.6 (Girsanov's theorem for 1d Brownian motion). *[1, p. 212] Let $W(t), 0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{F}(t)$ be the Brownian motion filtration. Let $\theta(t)$ be an adapted process. For a* ***drifting*** *stochastic process defined as*

$$\tilde{W}(t) = W(t) + \int_0^t \theta(s) ds,$$

and a stochastic process given as

$$Z(t) = \exp(-\int_0^t \theta(s) dW(s) - \frac{1}{2}\int_0^t \theta^2(s) ds),$$

we can change $\tilde{W}(t)$ to be a driftless Brownian motion under a new equivalent[a] probability measure Q generated by $Z(T) \triangleq Z$. Specifically, we have

- *$Z > 0$ almost surely, $E[Z] = 1$ such that $Z(t) = E[Z|\mathcal{F}_t]$ is Radon-Nikodyn derivative process and a martingale.*
- *$\tilde{W}(t)Z(t)$ is a martingale under the original probability measure.*
- *$\tilde{W}$ is a martingale under the new probability measure Q. That is*

$$E_Q[\tilde{W}(t)|\mathcal{F}_s] = \tilde{W}(s)$$

a The measure is equivalent because $Z > 0$ almost surely(Theorem 6.7.2).

Proof. (1) $Z(t)$ is an exponential martingale such that $E[Z(t)] = 1$ Lemma 5.5.2. We can also show that $E[Z(T)|\mathcal{F}_t] = E[Z(t)\exp(-\int_t^T \theta(s)dW(s) - \frac{1}{2}\int_t^T \theta^2(s)ds)|\mathcal{F}_t] = Z(t)$. (2) Let $X(t) = -\int_0^t \theta(s)dW(s) - \frac{1}{2}\int_0^t \theta^2(s)ds$, then $dX(t) = -\theta(t)dW(t) - \frac{1}{2}\theta^2(t)dt$, $Z(t) = e^{X(t)}$,

$$dZ(t) = Z(t)dX_t + \frac{1}{2}Z(t)\theta(t)^2 dt = -Z(t)\theta(t)dW(t).$$

Note that

$$d\tilde{W}(t)Z(t) = \tilde{W}(t)dZ(t) + Z(t)d\tilde{W}(t) + d\tilde{W}(t)dZ(t)$$

Then use $d\tilde{W} = dW(t) + \theta(t)dt$ and $dZ(t) = -Z(t)\theta(t)dW(t)$. Note that

$$d(\tilde{W}(t)Z(t)) = \tilde{W}(t)dZ(t) + Z(t)d\tilde{W}(t) + d\tilde{W}(t)dZ(t)$$

Then use $d\tilde{W} = dW(t) + \theta(t)dt$ and $dZ(t) = -Z(t)\theta(t)dW(t)$. We have

$$d(\tilde{W}(t)Z(t)) = (-\tilde{W}(t)\theta(t) + 1)Z(t)dW(t)$$

, which indicates that $\tilde{W}(t)Z(t)$ is a martingale. (3) Use the properties of change of measure with conditional expectation(Theorem 6.7.4)

$$E_Q[\tilde{W}(t)|\mathcal{F}_s] = \frac{1}{Z(s)}E[\tilde{W}(t)Z(t)|\mathcal{F}_s] = \frac{1}{Z(s)}\tilde{W}(s)Z(s) = \tilde{W}(s),$$

where we use the fact that $\tilde{W}(t)Z(t)$ is martingale with respect to $\mathcal{F}_t$.

(moment generating function method) Consider the moment generating function of $\tilde{W}(t)$ under the measure Q:

$$
\begin{aligned}
& E_Q[\exp(\lambda \tilde{W}_t)] \\
& = E[\exp(\lambda \tilde{W}_t) Z] \\
& = E[E[\exp(\lambda \tilde{W}_t) Z | \mathcal{F}_t] \\
& = E[\exp(\lambda \tilde{W}_t) Z(t)] \\
& = E[\exp(\lambda \tilde{W}_t - \int_0^t \theta(s) dW(s) - \frac{1}{2}\int_0^t \theta^2(s) ds)] \\
& = E[\exp(\int_0^t (\lambda - \theta(s)) dW(s) + \int_0^t \lambda\theta(s) ds - \frac{1}{2}\int_0^t \theta^2(s) ds)] \\
& = E[\exp(\int_0^t (\lambda - \theta(s)) dW(s) + \int_0^t \lambda\theta(s) ds - \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds) \\
& + \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds) - \frac{1}{2}\int_0^t \theta^2(s) ds)] \\
& = E[\exp(\int_0^t (\lambda - \theta(s)) dW(s) - \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds)] \\
& E[\exp(\int_0^t \lambda\theta(s) ds + \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds - \frac{1}{2}\int_0^t \theta^2(s) ds)] \\
& = E[\exp(\int_0^t \lambda\theta(s) ds + \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds - \frac{1}{2}\int_0^t \theta^2(s) ds))] \\
& = E[\exp(\frac{1}{2}\int_0^t \lambda^2 ds)] \\
& = E[\exp(\frac{1}{2}\lambda^2 t)]
\end{aligned}
$$

where we use the fact that $E[\exp(\int_0^t (\lambda - \theta(s)) dW(s) - \frac{1}{2}\int_0^t (\lambda - \theta(s))^2 ds)]$ is a martingale(Lemma 5.5.2). That is, the moment generating function shows that $\tilde{W}(t)$ is a Brownian motion following $N(0, t)$. □

Remark 6.7.4.

- $\tilde{W}(t)$ can be thought as the SDE given as

$$d\tilde{W} = \theta(t) dt + dW(t)$$

Remark 6.7.5 (implications).

- Via Girsanov's theorem, we can change **any drifting** Ito SDE to a driftless SDE, i.e., martingale, which has many nice properties. Girsanov's theorem tells us how to construct the new measure Q via Radon-Nikodyn derivative.

- For example, if we make the dynamics of an asset a martingale, then we can relate this current price(the goal) to its future expected payoff with respect to the new probability measure Q, which is known from Girsanov's theorem.

Note 6.7.1. The new measure Q_t generated by $Z(t)$ and the new measure generated by $Z(T)$ agrees on the sets in $\mathcal{F}_t$. See Theorem 6.7.4.

Consider an Ito process $X(t)$ driven by the Brownian $W(t)$, then the distribution $X(t)$ under the measure Q_t will be the same as under the measure Q_T. Therefore, for a drifting Brownian motion, we only need a final time $Z(T)$ to change the measure.

Corollary 6.7.6.1 (one Dimension Girsanov theorem with constant drift). *[6, p. 155][14, p. 221]Given a SDE as:*

$$dX_t = bdt + dW_t$$

define

$$Z_t = \exp(-bW(t) - \frac{1}{2}b^2 t),$$

then the dX_t is a driftless Brownian motion under the new probability measure Q generated by the Radon-Nikodym derivative Z_T. More specifically, under Q, we have

- *$dX_t = d\tilde{W}_t, d\tilde{W}_t = dW_t - bdt$, where $\tilde{W}(t)$ is a Brownian motion under Q at time $[0, T]$.*
- *Given function $f(X_t)$, we can evaluate its expectation under Q via*

$$E_Q[f(X_t)] = E[f(X_t)Z_T] = E[f(X_t)Z_t].$$

Proof. (2) Let $\{\mathcal{F}_t\}$ be the natural filtration associated with $W(t)$. Use the fact that $f(X_t)$ is $\mathcal{F}_t$ measurable, $Z(t) = E[Z_T|\mathcal{F}_t]$, and Theorem 6.7.4. □

6.7.4 Multi-dimensional Girsanov theorem

Theorem 6.7.7. *[1, p. 224][13, p. 37] Let T be fixed positive time, let $\theta(t) = (\theta_1(t), ..., \theta_d(t))$ be a d dimensional adapted process. Define*

$$Z(t) = \exp(-\int_0^t \theta(u) \cdot dW(u) - \frac{1}{2}\int_0^t \|\theta(u)\|^2 du)$$

$$\tilde{W}(t) = W(t) + \int_0^t \theta(u)du.$$

where $W(t)$ is a d-dimensional uncorrelated Brownian motion such that $dW_t dW_t^T = I_d dt \in \mathbb{R}^{d\times d}$. Further assume that

$$E[\int_0^T \|\theta(u)\|^2 Z^2(u)du] < \infty$$

Set $Z = Z(T)$. Then $E[Z] = 1$, and under the new equivalent[a] probability measure Q given by

$$Q(A) = \int_A Z(\omega)dP(\omega), \forall A \in \mathcal{F}$$

the process $\tilde{W}(t)$ is a d-dimensional Brownian motion.

a The measure is equivalent because $Z > 0$ almost surely(Theorem 6.7.2).

Proof. (moment generating function method) Consider the moment generating function of $\tilde{W}(t)$ under the measure Q:

$$
\begin{aligned}
& E_Q[\exp(\lambda^T \tilde{W}_t)] \\
&= E[\exp(\lambda^T \tilde{W}_t) Z] \\
&= E[E[\exp(\lambda^T \tilde{W}_t) Z | \mathcal{F}_t] \\
&= E[\exp(\lambda^T \tilde{W}_t) Z(t)] \\
&= E[\exp(\lambda^T \tilde{W}_t - \int_0^t \theta(s)^T dW(s) - \frac{1}{2}\int_0^t \theta^T \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) + \int_0^t \lambda^T \theta(s) ds - \frac{1}{2}\int_0^t \theta^T \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) + \int_0^t \lambda^T \theta(s) ds - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds) \\
&+ \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds) - \frac{1}{2}\int_0^t \theta^T \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds)] \\
& E[\exp(\int_0^t \lambda \theta(s) ds + \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds - \frac{1}{2}\int_0^t \theta^T \theta(s) ds)] \\
&= E[\exp(\int_0^t \lambda \theta(s) ds + \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds - \frac{1}{2}\int_0^t \theta^T \theta(s) ds)] \\
&= E[\exp(\frac{1}{2}\int_0^t \lambda^T \lambda ds)] \\
&= E[\exp(\frac{1}{2}\lambda^T \lambda t)]
\end{aligned}
$$

where we use the fact that

$$E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds)]$$

is a martingale(Lemma 5.5.2). That is, the moment generating function shows that $\tilde{W}(t)$ is a Brownian motion following $N(0, I_d t)$. □

Remark 6.7.6 (interpretation)**.**

- Under the original probability measure P, $\hat{W}$ will not a martingale because it contains a drift $\int_0^t \theta(u) du$.
- Under the new probability measure Q, $\hat{W}$ will be a martingale. And **the Girsanov theorem provides a method to find such measure**.

Corollary 6.7.7.1 (Girsanov theorem for correlated Brownian motion). *Let T be fixed positive time. Let $W(t)$ is a d-dimensional uncorrelated Brownian motion such that $dW_t dW_t^T = \Sigma dt \in \mathbb{R}^{d\times d}$. Let $\theta(t) = (\theta_1(t), ..., \theta_d(t))$ be a d dimensional adapted process. Define*

$$Z(t) = \exp(-\int_0^t \theta(u) \cdot dW(u) - \frac{1}{2}\int_0^t \theta(u)^T \Sigma \theta(u) du).$$

Then under the new equivalent[a] probability measure Q generated by $Z \triangleq Z(T)$:

- $W(t) \sim MN(-\int_0^t \Sigma(u)\theta(u)du, \int_0^t \Sigma(u)du)$.
- *$W(t)$ becomes a drifting Brownian motion following*

$$dW(t) = d\tilde{W}(t) - \Sigma\theta(t)dt,$$

where $\tilde{W}(t)$ is a Brownian motion under Q.

a The measure is equivalent because $Z > 0$ almost surely(Theorem 6.7.2).

Proof. (moment generating function method) Consider the moment generating function of $\tilde{W}(t)$ under the measure Q:

$$\begin{aligned}
&E_Q[\exp(\lambda^T W_t)] \\
&= E[\exp(\lambda^T W_t) Z] \\
&= E[E[\exp(\lambda^T W_t) Z | \mathcal{F}_t] \\
&= E[\exp(\lambda^T W_t) Z(t)] \\
&= E[\exp(\lambda^T W_t - \int_0^t \theta(s)^T dW(s) - \frac{1}{2}\int_0^t \theta^T \Sigma \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \theta^T \Sigma \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds) \\
&+ \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|^2 ds) - \frac{1}{2}\int_0^t \theta^T \Sigma \theta(s) ds)] \\
&= E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|_\Sigma^2 ds)] \\
&E[\exp(\frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|_\Sigma^2 ds - \frac{1}{2}\int_0^t \theta^T \Sigma \theta(s) ds)] \\
&= E[\exp(\frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|_\Sigma^2 ds - \frac{1}{2}\int_0^t \theta^T \Sigma \theta(s) ds)] \\
&= E[\exp(\frac{1}{2}\int_0^t \lambda^T \Sigma \lambda ds) - \int_0^t \lambda^T (\Sigma \theta(s)) ds)]
\end{aligned}$$

where we use the fact that

$$E[\exp(\int_0^t (\lambda - \theta(s))^T dW(s) - \frac{1}{2}\int_0^t \|(\lambda - \theta(s))\|_\Sigma^2 ds)]$$

is a martingale(Lemma 5.5.2). That is, the moment generating function shows that $W(t)$ is a Brownian motion following $MN(-\int_0^t \Sigma(u)\theta(u)du, \int_0^t \Sigma(u)du)$. $\square$

Remark 6.7.7 (calculate the θ function given target drift)**.** Suppose we want to change a driftless Brownian motion to a Brownian motion with drift μ, then we should choose $\theta = \Sigma^{-1}\mu$.

Corollary 6.7.7.2 (correlation invariance under change of measure). *Let $W = (W_1, W_2, ..., W_n)^T$ be a correlated Brownian motion ($dW_t dW_t^T = \Sigma dt$) under measure Q_1. Let Q_2 be an equivalent measure generated by*

$$Z = \exp(-\int_0^T \theta(u) \cdot dW(u) + \frac{1}{2}\int_0^T \theta(u)^T \Sigma \theta(u) du),$$

where $\theta(t) = (\theta_1, \theta_2, ..., \theta_n)^T$, Then under the new measure Q_2, we have

$$dW_t dW_t^T = \Sigma dt.$$

6.8 Fractional Brownian motion

Definition 6.8.1 (fractional Brownian motion). *[15, p. 16] A normalized fractional Brownian motion $W_t^H, t \geq 0$, with Hurst parameter $H \in (0,1)$ is a zero mean Gaussian process with continuous sample paths whose covariance is given by*

$$E[W_t^H W_s^H] = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}).$$

Remark 6.8.1 (interpretation).

- The Hurst exponent controls the correlations between increments of a fractional Brownian motion as well as the regularity of the paths: they become smoother as H increases.
- If $H = 0.5$, then the process is in fact a Brownian motion.
- If $H > 0.5$, then the increments of the process are positively correlated.
- If $H < 0.5$, then the increments of the porcess are negatively correlated.

Lemma 6.8.1. *[15, p. 17] A fractional Brownian motion has the following properties*

- *When $H = \frac{1}{2}$, then $W_t^{\frac{1}{2}}$ becomes standard Brownian motion.*
- $W_0^H = 0, EW_t^H = 0, E[(W_t^H)^2] = |t|^{2H}, t \geq 0.$
- *It has the zero mean stationary increments, and $E[(W_t^H - W_s^H)^2] = |t-s|^{2H}$.*

Proof. Straight forward. □

6.9 Notes on bibliography

For treatment on Stratonovich integral, see [4].

For treatment on calculating mean and various from SDE, see [7].

For treatment on the techniques for solving SDE, see [7][4][6].

For finance related treatment, see [6].

See [14] for treatment on Girsanov theory and Feynman-Kac connection.

7

JUMP-DIFFUSION PROCESS

7.1 Poisson process

7.1.1 Basics

Definition 7.1.1 (Poisson process). *Let $\lambda > 0$ be fixed. The stochastic process $\{N(t), t \in [0, \infty)\}$ is called a Poisson process with rates λ if all the following conditions hold:*

- $N(0) = 0$.
- $N(t)$ *has independent increments.*
- *The number of arrivals in any interval of length* $N(t_2) - N(t_1) \sim Poisson(\lambda(t_2 - t_1))$.

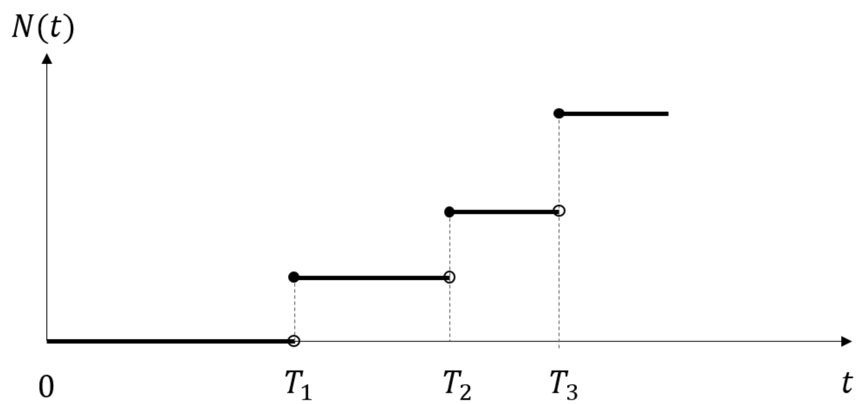


Figure 7.1.1: A typical realized trajectory from the Poisson process with jumps at T_1, T_2, and T_3.

Lemma 7.1.1 (basic properties of Poisson process). *Let $N(t)$ be a Poisson process with rate λ, then:*

- $N(t) \sim Poission(\lambda t)$, *that is*
$$P(N(t) = k) = \frac{e^{\lambda t}(\lambda t)^k}{k!}$$
- $N(t_2) - N(t_1) = N(t_2 - t_1) \sim Poisson(\lambda(t_2 - t_1))$
- $$E[N(t)] = \lambda t, Var[N(t)] = \lambda t, M_{N(t)}(s) = \exp(\lambda t(e^s - 1))$$

- *Jump probability within* $[t, t+\Delta t]$: *let* $\Delta N = N(t+\Delta t) - N(t)$, *we have*

$$Pr(\Delta N = n) = \frac{(\lambda \Delta t)^n}{n!}\exp(-\lambda \Delta t) = \frac{(\lambda \Delta t)^n}{n!}(1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2} + \cdots),$$

or explicitly

$$Pr(\Delta N = n) = \begin{cases} 1 - \lambda \Delta t + O((\Delta t)^2, n = 0 \\ \lambda \Delta t + O((\Delta t)^2, n = 1 \\ O((\Delta t)^2, n \geq 2 \end{cases}.$$

Proof. Directly from definition and the sum property of independent Poisson distribution(Lemma 2.2.23)and basic property of Poisson distribution(Lemma 2.2.22). □

Lemma 7.1.2 (additivity of Poisson process). *Let* $N_1(t)$ *and* $N_2(t)$ *be independent Poisson processes with rate* λ_1 *and* λ_2, *then* $N(t) = N_1(t) + N_2(t)$ *is a Poisson process with rate* $\lambda_1 + \lambda_2$.

Proof. Use the moment generating function for $N_1(t)$ and $N_2(t)$.(Lemma 2.2.23). □

7.1.2 Arrival and Inter-arrival Times

Lemma 7.1.3 (waiting time distribution). *Let* $N(t)$ *be a Poisson process with rate* λ. *Let* X_1 *be the time of the first arrival. Then*

$$P(X_1 > t) = \exp(-\lambda t), f_{X_1}(t) = \lambda \exp(-\lambda t)$$

Similarly, let X_n *be the waiting time between the arrival of* n *after the* $n-1$ *arrival, then*

$$P(X_n > t) = \exp(-\lambda t)$$

Proof. (1)From the definition of Poisson process, the $N(t) - N(0) \sim Poisson(\lambda t)$. Then

$$P(X_1 > t) = P(N(t) - N(0) = 0) = (\lambda t)^0 e^{-\lambda t}/0! = e^{-\lambda t}$$

(2) Using the independent increment property of Poisson process. □

Remark 7.1.1. Note that the waiting time distribution is an exponential distribution with parameter λ, whose mean is $1/\lambda$.

Lemma 7.1.4 (Arrival times for Poisson processes). *If $N(t)$ is a Poisson process with rate λ, then the arrival time $T_1, T_2, ...$ have $T_n \sim Gamma(n, \lambda)$ distribution:*

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

Moreover, we have $E[T_n] = n/\lambda, Var[T_n] = n/\lambda^2$.

Proof. Let random variables $X_1, X_2, ...$ be the interarrival time, then

$$\begin{aligned} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ T_2 &= X_1 + X_2 + X_3 \\ &\cdots\cdots \end{aligned}$$

Since X_i has exponential distribution(which is $Gamma(1, \lambda)$), the T_n will be $Gamma(n, \lambda)$ distribution(which can be showed that the nth power of mgf of exponential function equal to the mgf of Gamma distribution.) Also see property of Gamma distribution(Theorem 2.2.3.) □

Remark 7.1.2 (Simulating a Poisson process). We first generate iid random variables $X_1, X_2, X_3, ...,$ where $X_i \sim Exp(\lambda)$. Then the arrival times are given as

$$\begin{aligned} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ T_2 &= X_1 + X_2 + X_3 \\ &\cdots\cdots \end{aligned}$$

Lemma 7.1.5. *[1, p. 468] Let $N(t)$ be a Poisson process with intensity λ. We define the **compensated Poisson process** as*

$$M(t) = N(t) - \lambda t.$$

Then $M(t)$ is a martingale.

Proof.

$$\begin{aligned} E[M(t)|\mathcal{F}_s] &= E[M(t) - M(s)|\mathcal{F}_s] + E[M(s)|\mathcal{F}_s] \\ &= E[N(t) - N(s) - \lambda(t-s)|\mathcal{F}_s] + M(s) \\ &= E[N(t) - N(s)] - \lambda(t-s) + M(s) \\ &= M(s). \end{aligned}$$

□

7.1.3 Compound Poisson process

Definition 7.1.2 (compound Poisson process). *A compound Poisson process, parameterized by a rate* $\lambda > 0$ *and jump size distribution* G*, is a* ***continuous-time*** *process* $\{Y(t), t \geq 0\}$ *given by*

$$Y(t) = \sum_{i=1}^{N(t)} D_i,$$

where $\{N(t), t \geq 0\}$ *is a Poisson process with rate* λ*, and* $\{D_i : i \geq 1\}$ *are iid random variables, with distribution* G*, which are also independent of* $\{N(t)\}$*.*

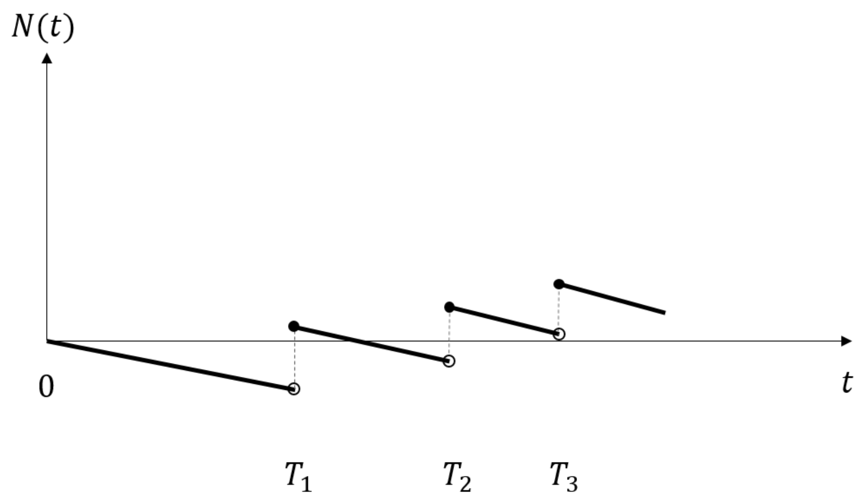


Figure 7.1.2: A typical realized trajectory from the compensated Poisson process with jumps at T_1, T_2, and T_3.

Remark 7.1.3 (interpretation). The jump arrival time distribution and waiting time distribution between two jump events is the same of simple Poisson process; compound Poisson distribution only differs from simple Poisson process in the jump size.

Remark 7.1.4 (reduction to simple Poisson process). If we take D_i to be constant 1, then

$$Y(t) - Y(s) = N(t) - N(s),$$

that is, $Y(t)$ is the Poisson process.

Definition 7.1.3 (simple-scaling compound Poisson process). *Let $N(t)$ be a simple Poisson process, then $Y(t) = \alpha N(t), \alpha \in \mathbb{R}$ is called a simple-scaling compound Poisson process with parameter α. Or equivalently,*

$$Y(t) = \sum_{i=1}^{N(t)} \alpha.$$

Example 7.1.1.

Example 7.1.2 (real-world compound Poission process). Assume in the stock exchange, the coming of orders follows a compound Poission distribution with strength λ. For every order, the amount of buying Z_i is iid random sample from random variable Z such that $E[Z] = \mu, Var[Z] = \sigma^2$. Then, the total amount of buying as a function of time is the compound Poission process.

Lemma 7.1.6 (jump probability within an interval).

Definition 7.1.4 (compound Poisson distribution). *Suppose that $N \sim Poisson(\lambda)$ and $X_1, X_2, ...$ are iid random variables independent of N. Let $Y|N = \sum_{n=1}^{N} X_n$, then the compound Poisson distribution is distribution of Y, which can be obtained by marginalizing the joint distribution (Y, N) over N.*

Lemma 7.1.7. $E_Y[Y] = E_N[E_{Y|N}[Y|N]] = E_N[\sum_{i=1}^{N} E_X[X_i]] = E_N[NE_X[X]] = E_N[N]E_X[X]$.

$$Var[Y] = E_N[Var[Y|N]] + Var_N[E[Y|N]]$$

Lemma 7.1.8 (basic properties of compound Poisson process). *For a compound Poisson process given as*

$$Y(t) = \sum_{i=1}^{N(t)} D_i.$$

We have

- $$E[Y(t)] = E[E[Y(t)|N(t)]] = E[N(t)E[D]] = E[N(t)]E[D] = \lambda t E[D].$$
- $$Var[Y(t)] = \lambda t E[D^2].$$

In particular, if $E[D] = 0$,

$$Var[Y(t)] = \lambda t Var[D].$$

- *The moment generating function*

$$M_Y(s) = \exp(\lambda t(M_D(s) - 1)),$$

where $M_D(s)$ *is the moment generating function of random variable D.*
- *If Y is a simple-scaling compound Poisson process with parameter* α, *then*

$$M_Y(s) = \exp(\lambda t(e^\alpha s - 1)).$$

- *Jump probability within* $[t, t + \Delta t]$: *let* $\Delta N = N(t + \Delta t) - N(t)$, *we have*

$$Pr(\Delta N = n) = \frac{(\lambda \Delta t)^n}{n!} \exp(-\lambda \Delta t) = \frac{(\lambda \Delta t)^n}{n!}(1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2} + \cdots),$$

or explicitly

$$Pr(\Delta N = n) = \begin{cases} 1 - \lambda \Delta t + O((\Delta t)^2, n = 0 \\ \lambda \Delta t + O((\Delta t)^2, n = 1 \\ O((\Delta t)^2, n \geq 2 \end{cases}.$$

Proof. (1)

$$E[Y(t)] = E[E[Y(t)|N(t)]] = E[N(t)E[D]] = E[N(t)]E[D] = \lambda t E[D].$$

(2)

$$\begin{aligned} Var[Y(t)] &= E[Var[Y(t)|N(t)]] + Var[E[Y(t)|N(t)]] \\ &= E[N(t)Var[D]] + Var[N(t)E[D]] \\ &= Var[D]E[N(t)] + E[D]^2 Var[N(t)] \\ &= Var[D]\lambda t + E[D]^2 \lambda t \\ &= \lambda t(Var[D] + E[D]^2) \\ &= \lambda t E[D^2] \end{aligned}$$

(3)

$$\begin{aligned}
E[e^{sY}] &= \sum_i e^{si} Pr(Y(t) = i) \\
&= \sum_i e^{si} \sum_n Pr(Y(t) = i | N(t) = n) Pr(N(t) = n) \\
&= \sum_n Pr(N(t) = n) \sum_i e^{si} \sum_n Pr(Y(t) = i | N(t) = n) \\
&= \sum_n Pr(N(t) = n) \sum_i e^{si} Pr(D_1 + D_2 + ... + D_n = i) \\
&= \sum_n Pr(N(t) = n)(M_D(s))^n \\
&= \sum_n Pr(N(t) = n) e^{n \ln(M_D(s))} \\
&= M_{N(t)}(\ln(M_D(s))) \\
&= \exp(\lambda t (M_D(s) - 1))
\end{aligned}$$

where we use the fact (Lemma 1.7.2)that the moment generating of $D_1 + D_2 + ... + D_n$ is $(M_D(s))^n$. Further, we can use the fact that the Poisson process has moment generating function(Lemma 7.1.1) given as

$$M_{N(t)}(s) = \exp(\lambda t (e^s - 1)).$$

(4) For a simple-scaling Poisson compound process, we know that

$$M_D(u) = E[\exp(u\alpha)] = \exp(u\alpha).$$

□

Lemma 7.1.9. *[1, p. 470] Let $Q(t)$ be a compound Poisson process and let $0 \leq t_0 < t_1 < \cdots < t_n$ be given. The increments*

$$Q(t_1) - Q(t_2), Q(t_2) - Q(t_1), \ldots, Q(t_n) - Q(t_{n-1}),$$

are independent and stationary. In particular, the distribution of $Q(t_j) - Q(t_{j-1})$ is the same as the distribution of $Q(t_j - t_{j-1})$.

Proof. □

Lemma 7.1.10 (decomposition of compound Poisson process when jump size is discrete random variable). *[1, p. 473] Assume the jump size Y is a discrete random variable can take finite set of nonzero numbers given by $y_1, \ldots, y_M$, and let $p(y_1), ..., p(y_M)$*

be the associated probability mass. Let $Y_1, \ldots$ be a iid random sample of Y. Let $N(t)$ be a Poisson process and define the compound Poisson process

$$Q(t) = \sum_{i=1}^{N(t)} Y_i.$$

For $m = 1, \ldots, M$ let N_m denote the number of jumps in Q of size y_m up to and including t. Then,

$$N(t) = \sum_{m=1}^{M} N_m(t), Q(t) = \sum_{m=1}^{M} y_m N_m(t).$$

The processes $N_m(t), m = 1, \ldots, M$ defined this way are independent simple Poisson processes with intensity $\lambda p(y_m)$, and $y_1 N_1(t), \ldots, y_m N_m(t)$ are independent simple-scaling compound Poisson processes.

Proof. Note that the moment generating function for Y is given by

$$M_Y(u) = \sum_{i=1}^{m} p(y_i) \exp(u y_i).$$

And the moment generating function for $Q(t)$ is given by

$$M_{Q(t)}(u) = \exp(\lambda t (\sum_{i=1}^{m} p(y_i) \exp(u y_i) - 1)).$$

On the other hand, the moment generating function for $y_m N_m(t)$ is

$$M_m(u) = \exp(\lambda p(y_m) t (\exp(u y_m) - 1)).$$

It is easy to see

$$M_{Q(t)}(u) = \prod_{i=1}^{M} M_m(u).$$

□

Remark 7.1.5 (interpretation). For a compound Poisson process with finite jump size space, we can decompose it into the summation of 'simple' compound Poisson processes, each with constant jump size.

7.1.4 Compensated Poisson process

Definition 7.1.5 (compensated Poisson process).

- *Let $N(t)$ be a Poisson process with intensity λ. The compensated Poisson process is defined by*
$$M(t) = N(t) - \lambda t.$$
- *Let $Q(t)$ be the compound Poisson process. Then the **compensated compound Poisson process** is defined as*
$$M_Q = Q(t) - \beta\lambda t,$$
where $\beta = E[D]$, is a martingale.

Lemma 7.1.11 (compensated Poisson process is a martingale). *[1, p. 467] Let $N(t)$ be a Poisson process with intensity λ. The compensated Poisson process $M(t) = N(t) - \lambda t$ is a martingale.*

Proof. Let $0 \leq s < t$ be given.

$$\begin{aligned} E[M(t)|\mathcal{F}(s)] &= E[M(t) - M(s) + M(s)|\mathcal{F}(s)] \\ &= E[M(t) - M(s)|\mathcal{F}(s)] + M(s) \\ &= E[N(t) - N(s) + \lambda(t-s)|\mathcal{F}(s)] + M(s) \\ &= 0 + M(s) \\ &= M(s) \end{aligned}$$

□

Lemma 7.1.12 (compensated compound Poisson process is a martingale). *[1, p. 470] Let $Q(t)$ be the compound Poisson process. Then the **compensated compound Poisson process***
$$Q(t) - \beta\lambda t$$
where $\beta = E[D]$, is a martingale.

Proof. □

7.2 Jump process

Definition 7.2.1 (right-continuous pure jump process). *[1, p. 475] We assume J does not jump at $t = 0$ and has only finitely many jumps on each finite interval, and is constant between jumps.*

Remark 7.2.1 (right-continuous vs. left continuous)**.**

- By right-continuous, we mean $J(t) = \lim_{s\to t^+}(s)$(but not $J(t) = \lim_{s\to t^-}(s)$) for all $t \geq 0$. The left-continuous counterpart of $J(t)$, denoted as $J(t-)$ can be constructed/modified from $J(t)$ such that at all jumps $J(t) = \lim_{s\to t^-}(s)$ but not $J(t) = \lim_{s\to t^+}(s)$.
- If J has a jump at time t, then $J(t)$ is the value of J immediately after the jump, and $J(t-)$ is the value immediately before the jump.

Example 7.2.1.

- A Poisson process and a compound Poisson process are pure jump processes because of the Constancy between jumps.
- A compensated Poisson process is not a pure jump process because it is decreasing between jumps.

Definition 7.2.2 (jump difference operator). *Given a stochastic process $X(t)$, we define the jump difference operator Δ as*

$$\Delta X(t) = X(t) - X(t-).$$

If $X(t)$ is a right-continuous process containing jumps at $t_1, t_2, ..., t_N$ with jump size $D_1, D_2, ..., D_N$, then

$$\Delta X(t) = \begin{cases} D_i, t = t_i, i = 1, 2, ..., N \\ 0, otherwise \end{cases}$$

Note that for a continuous process $R(t)$, $\Delta R(t) = 0$.

Lemma 7.2.1. *Let $J(t)$ be a right-continuous pure jump process, let X be a left-continuous process(possibly contains jumps), then*

$$\int_0^t X(s-)dJ(s) = \int_0^t X(s)dJ(s) + \int_0^t \Delta X(s)dJ(s),$$

where $\Delta X(s) = X(s) - X(s-)$. In particular,

$$\int_0^t J(s-)dJ(s) = \int_0^t J(s)dJ(s) + \int_0^t \Delta J(s)dJ(s).$$

$$\int_0^t X(s)dJ(s-) = 0.$$

$$\int_0^t X(s-)dJ(s-) = 0.$$

$$\int_0^t J(s)dX^c = 0, \int_0^t J(s-)dX^c = 0$$

Proof. (1)Based on definition, we have

$$X(s) = \Delta X(s) + X(s-).$$

$$\int_0^t X(s-)dJ(s) = \sum_{0<s\leq t} (X^c(s-) + J(s-))\Delta J(s) = \sum_{0<s\leq t} X^c(s-)\Delta J(s),$$

where we use the fact that at any jump time τ, $J(\tau-) = 0$, $\Delta J(\tau) \neq 0$. (2)From the definition,

$$\int_0^t X(s)dJ(s-) = \sum_{0<s\leq t} X(s)\Delta J(s-) = \sum_{0<s\leq t} X(s)(J(s-) - J(s-)) = 0.$$

(3) From Riemann integral, to integrate a function with finitely many jumps is equivalent to integrate the continuous portion of the integral. Since the continuous portion of J is zero, therefore $\int_0^t J(s)dX^c = 0$. The same arguments apply to $\int_0^t J(s-)dX^c = 0$. □

Definition 7.2.3 (jump process). *[1, p. 475] A stochastic process $X(t)$ defined as*

$$X(t) = X(0) + I(t) + R(t) + J(t),$$

*is called a **jump process**, where $X(0)$ is a nonrandom initial condition, $I(t) = \int_0^t \Gamma(s)dW(s)$ is the Ito integral of an adapted process $\Gamma(s)$ with respect to a Brownian motion $W(t)$. $R(t)$ is a Riemann integral given as $R(t) = \int_0^t \theta(s)ds$ for some adapted process $\theta(t)$ with respect to the Brownian motion. And $J(t)$ is an adapted, **right-continuous*** *pure jump process with $J(0) = 0$. The continuous part of $X(t)$ is denoted by $X^c(t)$, given by*

$$X^c(t) = X(0) + I(t) + R(t).$$

Remark 7.2.2 (understand continuity). A jump process $X(t)$ is a right-continuous and adapted. The left-continuous counterpart is defined as

$$X(t-) = X(0) + I(t) + R(t) + J(t-).$$

The jump size of X at time t is denoted as

$$\Delta X(t) = X(t) - X(t-) = J(t) - J(t-) = \Delta J.$$

If $X(t)$ is continuous, then $\Delta X(t) = 0$. If $X(t)$ has a jump at t, then $\Delta X(t)$ is the size of the jump, which equals $\Delta X(t) = J(t) - J(t-)$. We always assume there is no jump at $t = 0$; therefore $\Delta X(0) = 0$.

Definition 7.2.4 (stochastic integral respect to jump process). *[1, p. 475] Let $X(t)$ be a jump process given by*

$$X(t) = X(0) + I(t) + R(t) + J(t).$$

Let $\Phi(s)$ be an adapted process. The stochastic integral of Φ respect to X is defined as

$$\int_0^t \Phi(s)dX(s) = \int_0^t \Phi(s)\Gamma(s)dW(s) + \int_0^t \Phi(s)\theta(s)ds + \sum_{0<s\leq t} \Phi(s)\Delta J(s).$$

In differential form, we have

$$\Phi(t)dX(t) = \Phi(t)dI(t) + \Phi(t)dR(t) + \Phi(t)dJ(t)$$

where

$$\Phi(t)dI(t) = \Phi(t)\Gamma(t)dW(t), \Phi(t)dR(t) = \Phi(t)\Theta(t)dt.$$

Remark 7.2.3 (interpretation).

- $dJ(t)$ should be interpreted as $dJ(t) = J(t + \Delta t) - J(t)$, which is a random variable. If there is a jump at $\tau \in [t, t + \Delta t]$, then $dJ(t) = J(\tau) - J(\tau-)$; If there is no jumps, then $dJ(t) = 0$; moreover, the probability of one jump increment is given by λdt, the probability of no jump is $1 - \lambda dt$; more than one jump is zero as $\Delta t \to 0$(See the definition of Poisson process in terms of rate functions Lemma 7.1.8).
- The summation $\sum_{0<s\leq t} \Phi(s)\Delta J(s)$ is equivalent to summation of finite terms since there are only finitely many jumps in the interval $[0, t]$.

Lemma 7.2.2 (basic properties stochastic integral with jump process).

- *Let $N(t)$ be a simple Poisson process. let $\theta(t)$ be a continuous function. Then*

$$\int_0^t \theta(s)dN(t) = \sum_{0<s\leq t} \theta(s)\Delta N(s),$$

where $\Delta N(s) = 1$ if there is a jump at time s.

- *Let $N(t)$ be a simple Poisson process, then we have*

$$\int_0^t \Delta N(s)dN(s) = N(t)$$

$$\int_0^t \Delta(N(s)+\theta(s))dN(s) = N(t)$$

where $\theta(t)$ is a continuous nonrandom function.

- *Let $J(t)$ be a compound Poisson process $J(t) = \sum_{i=1}^{N(t)} D_i$, then*

$$\int_0^t \Delta(J(s)+\theta(s))dJ(s) = G(t),$$

where $G(t)$ is the new compound Poisson process $G(t) = \sum_{i=1}^{N(t)} D_i^2$.

Proof. (1)Note that $\Delta N(t)$ is a jump indicator process: it takes value 1 when there is a jump; otherwise 0. Therefore, $\int_0^t \Delta N dN = \sum_{0<s\leq t} \Delta N(s)\Delta N(s) = N(t)$ is counting the total jump up to t. In addition, use $\Delta\theta = 0$. (2)

$$\int_0^t \Delta(J(s)+\theta(s))dJ(s) = \sum_{0<s\leq t} \Delta J(s)\Delta J(s) = \sum_{i=1}^{N(t)} D_i^2.$$

□

Example 7.2.2. Let $X(t) = N(t) - \lambda t$, where $N(t)$ is a Poisson process with intensity λ. In this case, we have $I(t) = 0, R(t) = -\lambda t$ and $J(t) = N(t)$. Let $\Phi(s) = N(s)$(i.e., $\Phi(s)$ is 1 if N has a jump at time s and $\Phi(s) = 0$ otherwise). Then

$$\int_0^t \Phi(s)dN(s) = \sum_{0<s\leq t} \Phi(s)\Delta J = \sum_{0<s\leq t} \Delta N \Delta N = N(t),$$

where use the fact that $\Delta R = 0$. Note that $\Delta N(t)$ is a jump indicator process: it takes value 1 when there is a jump; otherwise 0. Therefore, $\int_0^t \Delta N dN = \sum_{0<s\leq t} \Delta N(s)\Delta N(s) = N(t)$ is counting the total jump up to t.

Theorem 7.2.1 (Martingale property for left-continuous integrand). *[1, p. 477] Assume that the jump process $X(t) = X(0) + I(t) + R(t) + J(t)$ is a martingale, the integral $\Phi(s)$ is **left-continuous** and adapted, and*

$$E[\int_0^t \Gamma^2(s)\Phi(s)^2 ds < \infty, \forall t \geq 0],$$

then the stochastic integral $\int_0^t \Phi(s)dX(s)$ is also a martingale. That is

$$E[\int_0^t \Phi(s)dX(s)|\mathcal{F}_h] = \int_0^h \Phi(s)dX(s)$$

$$E[\int_0^t \Phi(s)dX(s)] = \Phi(0)X(0).$$

Remark 7.2.4 (generalization of Ito integral property)**.** We know that the Ito integral

$$I(t) = E[\int_0^t \Phi(s)dW(s)]$$

is a martingale(Theorem 6.2.1).

7.2.1 Quadratic variations

Lemma 7.2.3. *[1, p. 481] Let $X(t) = X(0) + I(t) + R(t) + J(t)$ be the jump process, then*

$$dXdX = dIdI + dJdJ = \Gamma^2 dt + dJdJ, dJdI = 0$$

The quadratic variation is given as

$$\int_0^t dX(s)dX(s) = \int_0^t \Gamma^2 ds + \sum_{0<s\leq t} (\Delta J_1(s))^2$$

Proof. To show $dJdI = 0$, we can partition the interval $[0, t]$ into n intervals, then

$$\begin{aligned}
&\left|\sum_{j=0}^{n-1}(I(t_{j+1}) - I(t_j))(J(t_{j+1}) - J(t_j)\right| \\
&\leq \max_{0\leq j\leq n-1}\left|I(t_{j+1}) - I(t_j)\right| \cdot \sum_{j=0}^{n-1}\left|J(t_{j+1}) - J(t_j)\right| \\
&\leq \max_{0\leq j\leq n-1}\left|I(t_{j+1}) - I(t_j)\right| \cdot \sum_{0<s\leq T}\left|\Delta J(s)\right|
\end{aligned}$$

Note that as $n \to \infty$, $\max_{0\leq j\leq n-1}\left|I(t_{j+1}) \to 0\right|$ due to continuity, but $\sum_{0<s\leq T}\left|\Delta J(s)\right|$ remains finite. □

Lemma 7.2.4. *[1, p. 481] Let $X_1(t) = X_1(0) + I_1(t) + R_1(t) + J_1(t)$ and $X_2(t) = X_2(0) + I_2(t) + R_2(t) + J_2(t)$ be the jump processes. We further assume I_1 and I_2 are driven by different Brownian motions with correlation coefficient ρ_{12}. Then*

$$dX_1dX_2 = dI_1dI_2 + dJ_1dJ_2 = \Gamma_1\Gamma_2\rho_{12}dt + dJ_1dJ_2.$$

The quadratic variation is given as

$$\int_0^t dX(s)dX(s) = \int_0^t \Gamma_1(s)\Gamma_2(s)\rho_{12}ds + \sum_{0<s\leq t}\Delta J_2(s)\Delta J_1(s)$$

Remark 7.2.5 (interpretation). Note that $\Delta J_1(s)\Delta J_2(s)$ has value only if the two pure jump process has the same jump time.

7.2.2 Ito rule

Lemma 7.2.5 (1D Ito rule). *[1, p. 484] Let $X(t)$ be a jump process and $f(x,t)$ a function for which $\frac{\partial}{\partial t}f$, $\frac{\partial}{\partial x}f$ and $\frac{\partial^2}{\partial x^2}f$ are defined and continuous. Then*

$$df = \frac{\partial}{\partial t}fdt + \frac{\partial}{\partial x}fdX^c + \frac{1}{2}\frac{\partial^2}{\partial x^2}fdX^cdX^c + \Delta f(X(t), t)$$

where $\Delta f(X(t),t) = f(X(t),t) - f(X(t-),t-)$ and $X^c(t) = X^c(0) + I(t) + R(t)$. In the integral form, we have

$$f(X(t)) = f(X(0)) + \int_0^t \frac{\partial}{\partial s} f ds + \int_0^t \frac{\partial}{\partial x} dX^c + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2} f dX^c dX^c + \sum_{0<s\leq t} f(X(s),s) - f(X(s-),s-).$$

Remark 7.2.6 (interpretation). If the process $X(t)$ does not contain any jumps, or discontinuities, then

$$\sum_{0<s\leq t} f(X(s),s) - f(X(s-),s-) = 0.$$

As a consequence, the Ito rule reduce to the canonical Ito rule for continuous processes.

Remark 7.2.7 (differentiation). $\frac{\partial}{\partial x} f(t,x)$ is equivalent to $\frac{\partial}{\partial x^c} f(t,x)$;that is, we only take the differential for the continuous process part. This is because for pure jump process will take constant value except at finitely many values, and such effect has been taken care by the final summation.

Example 7.2.3 (geometric Poisson process). Consider the geometric Poisson process

$$S(t) = S(0)\exp(N(t)\log(\sigma+1) - \lambda\sigma t) = S(0)e^{-\lambda\sigma t}(\sigma+1)^{N(t)},$$

where $\sigma > -1$. We can show the geometric Poisson process is a martingale. Then,

$$\begin{aligned} S(t) &= S(0) + \int_0^t S_t ds + \sum_{0<u\leq t} [S(u) - S(u-)] \\ &= S(0) - \lambda\sigma \int_0^t S(u) du + \sum_{0<u\leq t} \sigma S(u-) \\ &= S(0) - \lambda\sigma \int_0^t S(u-) du + \int_0^t \sigma S(u-) dN(u) \\ &= S(0) - \int_0^t S(u-) dM(u), M(u) = N(u) - \lambda\sigma t. \end{aligned}$$

Note that we use the fact that $\partial_{X^c} S = 0$ and $S(u) - S(u-) = S(0)e^{-\lambda\sigma t} = \sigma S(u-)$, where $S(u-) = S(0)e^{-\lambda\sigma t}$ since $N(t-) = 0$(Remark 7.2.1).

Then, we can show

$$E[S(t)] = S(0) + E[\int_0^t S(u-) dM(u)] = S(0)$$

due to Theorem 7.2.1.

Lemma 7.2.6 (2D Ito rule). *[1, p. 489] Let $X(t)$ be a jump process and $f(x,t)$ a function for which $\frac{\partial}{\partial t}f$, $\frac{\partial}{\partial x}f$ and $\frac{\partial^2}{\partial x^2}f$ are defined and continuous. Then*

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

where $\Delta f(X(t),t) = f(X(t),t) - f(X(t-),t-)$ and $X^c(t) = X^c(0) + I(t) + R(t)$. In the integral form, we have

$$\begin{aligned} f(t, X_1(t), X_2(t)) = f(0, X_1(0), X_2(0)) + \int_0^t f_t ds + \\ \int_0^t f_{x_1} dX_1^c + \int_0^t f_{x_2} dX_2^c + \frac{1}{2}\int_0^t f_{x_1,x_1} dX_1^c dX_1^c + \frac{1}{2}\int_0^t f_{x_2,x_2} \\ dX_2^c dX_2^c + \int_0^t f_{x_1,x_2} dX_1^c dX_2^c + \\ \sum_{0<s\leq t} f(s, X_1(s), X_2(s)) - f(s-, X_1(s-), X_2(s-)). \end{aligned}$$

Lemma 7.2.7 (Ito product rule).

$$\begin{aligned} X_1(t)X_2(t) &= X_1(0)X_2(0) + \int_0^t X_2(s)dX_1^c + \int_0^t X_1(s)dX_2^c(s) + \int_0^t dX_1^c(s)dX_2^c(s) + \\ &\quad \sum_{0<s\leq t}[X_1(s)X_2(s) - X_1(s-)X_2(s-)] \\ &= X_1(0)X_2(0) + \int_0^t X_2(s-)dX_1 + \int_0^t X_1(s-)dX_2(s) + \int_0^t dX_1(s)dX_2(s) \end{aligned}$$

7.3 Levy processes

Definition 7.3.1 (infinitely divisible). *A random variable X is infinitely divisible if its law p_x is infinitely divisible; that is, $X = Y_1^{(n)} + Y_2^{(n)} + \cdots + Y_n^{(n)}$, where $Y_1^{(n)}, ..., Y_n^{(n)}$ are iid random variables for each $n \in \mathcal{N}$. In other words, the characteristic function of X can be written as*

$$\Psi_X(u) = (\Phi_{Y_1^{(n)}}(u))^n.$$

Remark 7.3.1. The superscript (n) does not represent any power exponent; it just means when we choose different n to divide the Levy process, we need to have different Y as building blocks.

Example 7.3.1 (Gaussian random variable). Let X have distribution $N(\mu, \sigma^2)$. Then X has characteristic function

$$\Phi_X(u) = \exp(i\mu t - \frac{1}{2}\sigma^2 u).$$

X is infinitely divisible with $Y_j^{(n)} \sim N(\mu/n, \sigma^2/n)$, with

$$\Phi_Y(u) = \exp(i\mu t/n - \frac{1}{2}\sigma^2 u/n) = \Phi_X(u)^{1/n}.$$

Example 7.3.2 (Poisson random variable). Let X have distribution $Poisson(\lambda)$. Then X has characteristic function

$$\Phi_X(u) = \exp(\lambda(e^{iu} - 1)).$$

X is infinitely divisible with $Y_j^{(n)} \sim Poisson(\lambda/n$, with

$$\Phi_Y(u) = \exp(\lambda(e^{iu} - 1)/n) = \Phi_X(u)^{1/n}.$$

Definition 7.3.2 (Levy symbol, characteristic exponent). *Let X be a random variable, let $\Psi_X(u)$ be its characteristic function, then*

$$\eta = \ln(\Psi_X(u))$$

is called the Levy symbol or Characteristic exponent.

Definition 7.3.3 (cadlag). *Let (M, d) be a metric space, and let $E \subseteq \mathbb{R}$. A function $f : E \to M$ is called a cadlag function if, for every $t \in E$,*

- *the left limits $f(t^-) = \lim_{s \to t^-}$ exists;*
- *the right limits $f(t^+) = \lim_{s \to t^+}$ exists and equals $f(t)$.*

That is, f is right-continuous with left limits.

Definition 7.3.4 (Levy process). *Let X_t be a stochastic process. Then X_t is a Levy process if the following condition are satisfied:*

1. *$X_0 = 0$.*
2. *X_t has independent increments: $L_t - L_s$ is independent of $\mathcal{F}_s, 0 \leq s < t < \infty$.*
3. *L has stationary increments: $P(L_t - L_s \leq x) = P(L_{t-s} \leq x), 0 \leq s < t < \infty$.*
4. *L_t is continuous in probability: $\lim_{t \to s} L_t = L_s$; that is for all $\epsilon > 0$ and for all $s \geq 0$,*

$$\lim_{t \to s} P(\left|X(t) - X(s)\right| > a) = 0.$$

Remark 7.3.2 (continuity of sample path).

- Note that in the Levy processes, the sample path is continuous in probability, while in Wiener process, the sample path is almost surely continuous.
- If X_t is a Levy process, then one may show that the sample path is almost surely right continuous with left limits(cadlag).

Remark 7.3.3 (Levy process is infinitely divisible). From property(2)(3) in the definition, we know that a Levy process X_t is infinitely divisible.

Lemma 7.3.1. *If X_t is a Levy process, then X_t is infinitely divisible for each $t \geq 0$. Furthermore,*

$$\Psi_{X(t)}(u) = e^{t\eta(u)}$$

where η is the Levy symbol of $X(1)$, i.e., $\eta = \ln(\Psi_{X(1)}(u)$.

Proof. From property(2)(3) in the definition, we know that a Levy process X_t is infinitely divisible. □

Definition 7.3.5 (compound Poisson process). *A compound Poisson process, parameterized by a rate $\lambda > 0$ and jump size distribution G, is a process $\{Y(t), t \geq 0\}$ given by*

$$Y(t) = \sum_{i=1}^{N(t)} D_i,$$

where $\{N(t), t \geq 0\}$ is a Poisson process with rate λ, and $\{D_i : i \geq 1\}$ are iid random variables, with distribution G, which are also independent of $\{N(t)\}$.

Definition 7.3.6 (compound Poisson distribution). *Suppose that $N \sim Poisson(\lambda)$ and $X_1, X_2, ...$ are iid random variables independent of N. Let $Y|N = \sum_{n=1}^{N} X_n$, then the compound Poisson distribution is distribution of Y, which can be obtained by marginalizing the joint distribution (Y, N) over N.*

Lemma 7.3.2. $E_Y[Y] = E_N[E_{Y|N}[Y|N]] = E_N[\sum_{i=1}^{N} E_X[X_i]] = E_N[NE_X[X]] = E_N[N]E_X[X]$.

$$Var[Y] = E_N[Var[Y|N]] + Var_N[E[Y|N]]$$

Lemma 7.3.3. *Let $\{Z_n, n \in \mathcal{N}\}$ be a sequence of iid random variables taking values in $\mathbb{R}^d$ with common law μ_Z and let $N \sim Poisson(\lambda)$. Let $X = \sum_{i=1}^{N} Z_i$. Then, for each $u \in \mathbb{R}^d$,*

$$\Psi_X(u) = E[e^{iuX}] = \exp(\int_{\mathbb{R}^d} (e^{iu^T y} - 1)\lambda\mu_Z(dy)).$$

Proof.

$$\begin{aligned}
\Psi_X(u) &= \sum_{n=0}^{\infty} E[e^{iu^T X}|N = n]P(N = n) \\
&= \sum_{n=0}^{\infty} E[e^{iu^T X}|N = n]e^{-\lambda}\frac{\lambda^n}{n!} \\
&= e^{-\lambda}\sum_{n=0}^{\infty} \frac{(\lambda\Psi_Z(u))^n}{n!} \\
&= \exp(\lambda(\Psi_Z(u) - 1))
\end{aligned}$$

where

$$\Psi_Z(u) = \int_{\mathbb{R}^d} (e^{iu^T y} - 1)\mu_Z(dy).$$

□

Theorem 7.3.1. *Suppose that $\mu \in \mathbb{R}, \sigma \geq 0$, and ν is a measure concentrated on $\mathbb{R}/\{0\}$ such that $\int_{\mathbb{R}} \min(1, x^2)\nu(dx) < \infty$. A probability law of a real-valued random variable L has characteristic exponent $\Psi(u) = -\frac{1}{t}\log(E[e^{iuL_t}])$ given by*

$$\Psi(u; t) = \int_{\mathbb{R}} e^{iux}\eta(dx) = e^{-t\Psi(u)}, \forall u \in \mathbb{R}$$

if and only if there exists a triple (μ, σ, ν) such that

$$\Psi(u) = i\mu u + \frac{1}{2}\sigma^2 u^2 + \int_R (1 - e^{iux} + iux\mathbf{1}_{|x|<1})\nu(dx)$$

for every $u \in \mathbb{R}$.

Definition 7.3.7 (Levy measure). *Levy measure is a measure ν on $\mathbb{R}^d/\{0\}$ such that*

$$\int \min(|y|^2, 1)\nu(dy) < \infty$$

Remark 7.3.4 (interpretation). [16] The theorem says that there exists a probability space where $L = L^{(1)} + L^{(2)} + L^{(3)}$; $L^{(1)}$ is the standard Brownian motion with drift, $L^{(2)}$ is a compound Poisson process, and $L^{(3)}$ is a square integrable martingale with countable number of jumps of magnitude less than 1(almost surely).

Example 7.3.3 (Brownian motion). Let $B(t)$ be a standard Brownian motion in $\mathbb{R}^d$. Then, $C(t) = bt + \sigma B(t), b \in \mathbb{R}^d, \sigma \in \mathbb{R}^{d\times d}$ is a Levy process with its characteristic function given by

$$\Psi_{B(t)}(u) = \exp(t\eta_C(u)),$$

where $\eta_C(u)$ is Levy symbol of $C(1)$ of the form

$$\eta_C(u) = ib^T - \frac{1}{2}u^T\sigma^T\sigma u.$$

7.4 Notes on bibliography

For Fokker-Planck equation, see [17][18].

BIBLIOGRAPHY

1. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
2. Miranda, H.-C. *Applied Stochastic Analysis lecture notes* (NewYork University, 2015).
3. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
4. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
5. Chin, E., Nel, D. & Ãlafsson, S. *Problems and Solutions in Mathematical Finance: Equity Derivatives* (John Wiley & Sons, 2017).
6. Wiersema, U. F. *Brownian motion calculus* (John Wiley & Sons, 2008).
7. calin, O. *An Introduction to Stochastic Calculus with Applications to finance* (Eastern Michigan University, 2012).
8. Chang, F.-R. *Stochastic optimization in continuous time* (Cambridge University Press, 2004).
9. Klebaner, F. C. *Introduction to Stochastic Calculus with Applications* (Imperial College Press, 2005).
10. Brigo, D. & Mercurio, F. *Interest rate models-theory and practice: with smile, inflation and credit* (Springer Science & Business Media, 2007).
11. McInerney, D. & Zastawniak, T. *Stochastic Interest Rates* ISBN: 9781107002579. https://books.google.com/books?id=8JTwCAAAQBAJ (Cambridge University Press, 2015).
12. Iacus, S. M. *Simulation and inference for stochastic differential equations: with R examples* (Springer Science & Business Media, 2009).
13. Wu, L. *Interest rate modeling: Theory and practice* (CRC Press, 2009).
14. Steele, J. *Stochastic Calculus and Financial Applications* ISBN: 9780387950167. https://books.google.com/books?id=H06xzeRQgV4C (Springer, 2001).
15. Pavliotis, G. A. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, 2014).
16. Manuge, D. J. L\'evy Processes For Finance: An Introduction In R. *ArXiv e-prints.* arXiv: 1503.03902 [stat.AP] (Mar. 2015).

17. Risken, H. & Frank, T. *The Fokker-Planck Equation: Methods of Solution and Applications* ISBN: 9783642615443. https://books.google.com/books?id=6B5uCQAAQBAJ (Springer Berlin Heidelberg, 2012).

18. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127. https://books.google.com/books?id=otg3PQAACAAJ (Springer Berlin Heidelberg, 2009).

8

FOKKER-PLANCK EQUATION

8.1 Fokker-Planck equations

8.1.1 FormulationsïijŽ one dimension

Lemma 8.1.1 (one dimensional Fokker-Planck equation). *[1, p. 121][2, p. 79] The Fokker-Planck equation is associated with an Ito SDE*

$$dx = \mu(x,t)dt + \sigma(x,t)dW_t,$$

where W_t is a Wiener process, is given as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)^2 p),$$

note that μ, σ can have dependence on both x and t.

Proof. Let $f(x)$ be an arbitrary smooth function of x, then

$$\int f(x)p(x,t)dx = \langle f \rangle \implies \int f(x)\frac{\partial p}{\partial t}dx = \left\langle \frac{df}{dt} \right\rangle.$$

Note that from Ito rule, we have

$$\begin{aligned} df &= \frac{df}{dx}dx + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 dt \\ df &= \frac{df}{dx}(\mu(x,t)dt + \sigma(x,t)dW_t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 dt \\ \implies \langle df \rangle &= (\frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2)dt \end{aligned}$$

Then

$$\begin{aligned} \int f(x)\frac{\partial p}{\partial t}dx &= \left\langle \frac{df}{dt} \right\rangle \\ &= \left\langle \frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 \right\rangle \\ &= \int (\frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2)p(x,t)dx \end{aligned}$$

Integrating by parts(twice) and drop out the surface terms($p(x \to \infty) = 0$), we have

$$\int f(x)\frac{\partial p}{\partial t}dx = \int f(x)(-\frac{\partial \mu p}{\partial x} + \frac{1}{2}\frac{\partial^2 \sigma^2 p}{\partial x^2})dx$$

Because $f(x)$ is an arbitrary function, we have

$$\frac{\partial p}{\partial t} = -\frac{\partial \mu p}{\partial x} + \frac{1}{2}\frac{\partial^2 \sigma^2 p}{\partial x^2}$$

using fundamental lemma of calculus of variations(**??**). □

Remark 8.1.1. Fokker-Planck equation is also known as Kolmogorov forward equation.

Remark 8.1.2 (common boundary conditions). [3, p. 92]

- Adsorbing boundary condition. For example at interval $[0,1]$, we have $p(x = 0, t|x_0, t_0) = p(x = 1, t|x_0, t_0) = 0$.
- Reflecting boundary condition. For example at interval $[0,1]$, we have $\partial_x p(x = 0, t|x_0, t_0) = \partial_x p(x = 1, t|x_0, t_0) = 0$, which are interpreted as zero flux at the boundary.
- Periodic boundary condition. For example at interval $[0,1]$, we have $p(x = 0, t|x_0, t_0) = p(x = 1, t|x_0, t_0)$.

8.1.1.1 *Formulations: multiple dimension*

Lemma 8.1.2 (multi-dimensional Fokker-Planck equation). *[1, p. 121][2, p. 83] The Fokker-Planck equation is associated with a d dimensional Ito SDE*

$$dx_i = \mu_i(x,t)dt + \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t), i = 1,2,...,d$$

where $W_i(t), i = 1,...,d$ is are multiple dimensional independent Wiener process, is given as

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m} \sigma_{ik}\sigma_{jk}p).$$

Proof. Let G_t be a function of $x = (x_1, x_2, ..., x_d)$ given by $G_t = g(x_1, x_2, ..., x_d)$. From Ito rule, we have

$$\begin{aligned}
dG_t &= \sum_{i=1}^{d} \frac{\partial g}{\partial x_i} dx_i + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} dx_i dx_j \\
&= \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu_i(x,t)dt + \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t)) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} \sum_{k=1}^{m} \sigma_{ik}\sigma_{jk} dt \\
&= (\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} \sum_{k=1}^{m} \sigma_{ik}\sigma_{jk})dt + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i} \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t).
\end{aligned}$$

Note that

$$\frac{d}{dt}E[G_t] = \frac{d}{dt}\int_{\mathbb{R}^d} p(x,t)g(x)dx = \int_{\mathbb{R}^d} \frac{d}{dt}p(x,t)g(x)dx.(*)$$

From another aspect, we have

$$\begin{aligned}
\frac{d}{dt}E[G_t] &= E[\frac{d}{dt}G_t] \\
&= E[(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})] \\
&= \int_{\mathbb{R}^d} p(x,t)(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})]dx \\
&= \int_{\mathbb{R}^d} p(x,t)(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})]dx \\
&= \int_{\mathbb{R}^d} [-\sum_{i=1}^{d}\frac{\mu_i(x,t)p(x,t)}{dx_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}\sigma_{ik}\sigma_{jk}p(x,t)]g(x)dx
\end{aligned}$$

Combine with $(*)$, we have

$$\int_{\mathbb{R}^d}[\frac{d}{dt}p(x,t) - \sum_{i=1}^{d}\frac{\mu_i(x,t)p(x,t)}{dx_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}\sigma_{ik}\sigma_{jk}p(x,t)]g(x)dx$$

holds for any function $g(x)$. Therefore

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}p).$$

□

8.1.2 Steady state & detailed balance

Lemma 8.1.3 (1d steady state solution). *Given a 1d Fokker-Planck*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)^2 p),$$

the steady state distribution p^{eq} will satisfy

$$-\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)p) = 0.$$

Let $J = up - \frac{1}{2}\frac{\partial}{\partial x}(\sigma(x,t)^2 p)$, *then the steady state solution is satisfying*

$$\frac{\partial}{\partial x} J = 0.$$

In particular,

- *If domain $D = [a,b]$, and the boundary are reflecting, then we have stronger result of $J = 0$.*
- *If domain $D = [a,b]$, and the boundary are periodic, then we have stronger result of $J = const$, const might not be zero.*

Proof. On 1D, $\partial_x J = 0 \implies J = const.$ □

Lemma 8.1.4 (2d/3d steady state solution). *Given a 2d/3d Fokker-Planck equation*

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m} \sigma_{ik}\sigma_{jk}p),$$

the steady state solution is given as

$$-\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m} \sigma_{ik}\sigma_{jk}p)$$

Let $Q_i = \sum_k \sigma_{ik}^{-1}(2\mu_k - 2\sum_j \frac{\partial}{\partial x_j}\sigma_{kj})$

8.1.3 Averages and adjoint operator

Lemma 8.1.5. *In the function space of square-integrable functions and the scalar product is defined by*

$$\langle f, g \rangle = \int_a^b f(x)\bar{g(x)}dx.$$

Consider a linear differential operator T given as

$$Tu = \sum_{k=0}^{n} a_k(x) \frac{d^k}{dx^k} u.$$

If f or g vanishes for $x \to a, b$, then the adjoint of T is

$$T^\dagger u = \sum_{k=0}^{n} (-1)^k \frac{d^k}{a_k(x) dx^k} u$$

Proof. Use integration-by-parts. □

Example 8.1.1. The Sturm-Liouville operator is defined as

$$Lu = -(pu')' + qu.$$

It can be showed that $Lu = L^\dagger u, L = L^\dagger$.

Corollary 8.1.0.1 (adjoint operators).

- $$Af = b(x)\frac{d}{dx} f(x), A^\dagger f = -\frac{d}{dx} b(x) f(x)$$

  that is,

  $$\int_{-\infty}^{\infty} g(x) b(x) \frac{d}{dx} f(x) dx = -\int_{-\infty}^{\infty} f(x) \frac{d}{dx} b(x) f(x) dx$$

- $$Af = D(x) \frac{d^2}{dx^2} f(x), A^\dagger f = \frac{d^2}{dx^2} D(x) f(x)$$

- $$Af = \frac{d}{dx} D(x) \frac{d}{dx} f(x), A^\dagger f = \frac{d}{dx} D(x) \frac{d}{dx} f(x) c$$

Proof. (1) use integration by parts, note that $\lim_{x\to\infty} f(x) = 0$.(2)(3) use integration by parts, note that $\lim_{x\to\infty} f(x) = 0$, $\lim_{x\to\infty} \frac{d}{dx} f(x) = 0$. □

Corollary 8.1.0.2 (Adjoint of Fokker-Planck operator). *Define*

$$Af = \sum_i b_i(x) \frac{\partial}{\partial x} f(x) + \frac{1}{2} \sum_{i,j} (\sigma(x)\sigma(x)')_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

and its adjoint is

$$A^{\dagger} f = -\sum_i \frac{\partial}{\partial x} b_i(x) f(x) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma(x)\sigma(x)')_{ij} f(x)$$

8.1.4 Backward equation

Lemma 8.1.6. *The backward Fokker-Planck equation is given as*

$$\frac{\partial}{\partial t_1} P(t_2, y_2 | t_1, y_1) = -A(y_1, t_1) \frac{\partial}{\partial y_1} P(t_2, y_2 | t_1, y_1) - \frac{1}{2} B(y_1, t_1) \frac{\partial^2}{\partial y_1^2} P(t_2, y_2 | t_1, y_1),$$

where

$$\int (z - y_1) P(t_1, z | t - \Delta t, y_1) dz = \Delta t A(y_1, t_1 - \Delta t)$$

and

$$\int (z - y_1)^2 P(t_1, z | t - \Delta t, y_1) dz = \Delta t B(y_1, t_1 - \Delta t)$$

.

Proof. From Chapman-Kolmogorov equation, we have

$$P(t_2, y_2 | t - \Delta t, y_1) = \int P(t_2, y_2 | t_1, z) P(t_1, z | t_1 - \Delta t, y_1) dz.$$

Using Taylor expansion, we have

$$P(t_2, y_2 | t_1, z) = P(t_2, y_2 | t_1, y_1) + \frac{\partial}{\partial y_1} P(t_2, y_2 | t_1, y_1)(z - y_1) + \frac{1}{2} \frac{\partial^2}{\partial y_1^2} (z - y_1)^2 + O(\int (z - y_1)^3.$$

Plug in, we have

$$\begin{aligned} P(t_2, y_2 | t_1 - \Delta t, y_1) = {} & P(t_2, y_2 | t_1, y_1) - \Delta t \frac{\partial}{\partial t_1} P(t_2, y_2 | t_1, y_1) \\ & + \frac{\partial}{\partial y_1} P(t_2, y_2 | t_1, y_1) \int (z - y_1) P(t_1, z | t - \Delta t, y_1) dz \\ & + \frac{1}{2} \frac{\partial^2}{\partial y_1^2} \int (z - y_1)^2 P(t_1, z | t_1 - \Delta t, y_1) dz + \\ & + O(\int (z - y_1)^3 P(t_1, z | t - \Delta t, y_1) dz) \end{aligned}$$

Let $\int(z-y_1)P(t_1,z|t-\Delta t,y_1)dz = \Delta t A(y_1,t_1-\Delta t)$ and $\int(z-y_1)^2P(t_1,z|t-\Delta t,y_1)dz = \Delta t B(y_1,t_1-\Delta t)$, we have

$$\frac{\partial}{\partial t_1}P(t_2,y_2|t_1,y_1) = -A(y_1,t_1)\frac{\partial}{\partial y_1}P(t_2,y_2|t_1,y_1) - \frac{1}{2}B(y_1,t_1)\frac{\partial^2}{\partial y_1^2}P(t_2,y_2|t_1,y_1)$$

□

Remark 8.1.3 (interpretation).

- The backward Fokker-Planck equation describes the dependence of $P(y_2,t_2|y_1,t_1)$ on the initial condition (y_1,t_1).
- To be well-posed, the backward Fokker-Planck equation needs a final condition rather than an initial condition.

Lemma 8.1.7 (Backward equation using adjoint operator). *If*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \mathcal{L}_x P,$$

then

$$\frac{\partial}{\partial t_0}P(x,t|x_0,t_0) = \mathcal{L}_{x_0} P,$$

Proof. From Chapman-Kolmogorov equation, we have

$$P(x_1,t_1|x_0,t_0) = \int P(x_1,t_1|x,t)P(x,t|x_0,t_0)dx,$$

where $t_0 < t < t_1$. Take the derivative with respect to t, then

$$\begin{aligned}
\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t)]P(x,t|x_0,t_0) + \int P(x_1,t_2|x,t)\frac{\partial}{\partial t}P(x,t|x_0,t_0) &= 0 \\
\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t)]P(x,t|x_0,t_0) + \int P(x_1,t_2|x,t)\mathcal{L}_x P(x,t|x_0,t_0) &= 0 \\
\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t) - \mathcal{L}_x^\dagger P(x_1,t_2|x,t)]P(x,t|x_0,t_0) &= 0
\end{aligned}$$

where $\mathcal{L}_x^\dagger$ is the adjoint operator satisfying

$$\int P(x_1,t_1|x,t)\mathcal{L}_x P(x,t|x_0,t_0)dx = \int [\mathcal{L}_x^\dagger P(x_1,t_1|x,t)]P(x,t|x_0,t_0)dx.$$

Then

$$\frac{\partial}{\partial t}P(x_1,t_2|x,t) - \mathcal{L}_x^\dagger P(x_1,t_2|x,t) = 0.$$

□

Lemma 8.1.8 (Backward equation on expectation quantity). *Let* $u(t,x) = E[g(X_T)|X(t) = x]$. *Assume*

$$\int P(t,y|t-h,x)(y-x)dy = b(x,t)h + o(h)$$

and

$$\int P(t,y|t-h,x)(y-x)^2 dy = \sigma(x,t)h + o(h),$$

and higher moment vanish as $h \to 0$. *Then* u *satisfy*

$$\frac{\partial u}{\partial t} + b(x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2},$$

with final condition $u(T,x) = g(x)$.

Proof. Note that

$$u(t,x) = E[g(X_T)|X(t) = x] = \int g(X_T)P(T, X_T|t,x)dX_T$$

and

$$u(t-h,x) = \int E[g(X_T)|X(t) = y]P(t,y|t-h,x)dy$$

$$u(t,x) = \int E[g(X_T)|X(t) = x]P(t,y|t-h,x)dy$$

Therefore,

$$\begin{aligned}\frac{u(t) - u(t-h)}{h} &= \frac{\int P(t,y|t-h,x)(u(t,y) - u(t,x)}{h} \\ &= \frac{1}{h}\int P(t,y|t-h,x)[(y-x)\frac{\partial u}{\partial x} + \frac{1}{2}(y-x)^2\frac{\partial^2 u}{\partial x^2}]\end{aligned}$$

□

Remark 8.1.4. Backward equation can also be derived using Feyman Kac theorem by assuming the stochastic process of $X(t)$(see Theorem 8.3.4).

Lemma 8.1.9 (Backward equation on expectation quantity using adjoint operator). *Let* $u(t,x) = E[g(X_T)|X(t) = x]$. *Assume*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \mathcal{L}_x P(x,t|x_0,t_0).$$

Then

$$\frac{\partial u}{\partial t} = \mathcal{L}_x^\dagger u$$

with final condition $u(T,x) = g(x)$.

Proof. Note that

$$u(t,x) = E[g(X_T)|X(t) = x] = \int g(X_T)P(T, X_T|t,x)dX_T$$

and Therefore,

$$\begin{aligned}
\frac{\partial u(t,x)}{\partial t} &= \int g(X_T)\frac{\partial}{\partial t}P(T, X_T|t,x)dX_T \\
&= \int g(X_T)\mathcal{L}_x^\dagger P dX_T = \mathcal{L}_x^\dagger \int g(X_T)PdX_T = \mathcal{L}_x^\dagger u(x,t).
\end{aligned}$$

□

Remark 8.1.5 (analog in Markov chain). For analog in discrete-time Markov chain, see **??**.

8.1.5 Mean first passage time problem

Definition 8.1.1 (survival probability). *The survival probability up to* t *in domain* Ω *is given as*

$$G(t|x_0,t_0) = \int_\Omega dx P(x,t|x_0,t_0).$$

Or equivalently,

$$G(t|x_0,0) = Pr(\tau > t) = \int_t^\infty \rho(\tau|x_0,0)$$

where $\rho(\tau|x_0,0)$ *is the probability density for the first passage time* τ *starting from* x_0 *at* $t = 0$.

Lemma 8.1.10 (survival probability governing equation). *Assume*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = -\mathcal{L}_x P(x,t|x_0,t_0).$$

Then

$$\frac{\partial}{\partial t}G(t|x,0) = -\mathcal{L}_x^\dagger G(t|x,0),$$

where $\mathcal{L}_x^\dagger$ is the adjoint operator of $\mathcal{L}_x$.

Proof. Use the theorem(Lemma 8.1.9) and define $g(X_T)$ as the indicator function. □

Lemma 8.1.11. *Define $T(x)$ be the mean first passage time, we have*

$$T(x) = -\int_0^\infty t\partial_t G(t|x_0 = x, t_0 = 0)dt = \int_0^\infty G(t|x,0)dt.$$

The governing equation for $T(x)$ is

$$\mathcal{L}^\dagger T = -1$$

with boundary condition $T(\partial D) = 0$.

Proof. Use integration by parts, we have

$$T(x) = \int_0^\infty G(t|x,0)dt.$$

Then

$$\begin{aligned}\mathcal{L}^\dagger T = \mathcal{L}^\dagger \int_0^\infty G(t|x,0) = \int_0^\infty \mathcal{L}^\dagger G(t|x,0) = \int_0^\infty \partial_t G(t|x,0)dt \\ = G(\infty|x,0) - G(0|x,0) = 0 - 1 = -1.\end{aligned}$$

□

Remark 8.1.6. For the mean first passage time calculation using adjoint Fokker-Planck equation, see [4].

8.2 Smoluchowski/advection-diffusion equation

Definition 8.2.1 (Smoluchowski/advection-diffusion equation). *The advection-diffusion for density $p(x,t), x \in \mathbb{R}^N$ is given as*

$$\frac{\partial p}{\partial t} = -\nabla \cdot (vp) + \nabla \cdot D\nabla p$$

where $v \in \mathbb{R}^N, D \in \mathbb{R}^{N\times N}$. Note that we can interpret the flux vector as $j = vp - D\nabla p$.

Remark 8.2.1 (Relation to computational fluid mechanics)**.** We can use the PDE to describe the solute transport in the flowing solutions. The v is the velocity field of the solution, which is further determined by the Navier-Stokes equation. When the velocity field describes an incompressible flow, i.e. $\nabla \cdot v = 0$, we can have simplification as $\nabla \cdot (vp) = v \cdot \nabla p$.

Remark 8.2.2 (Relation to advection equation)**.**

The advection equation for a conserved quantity described by a scalar field ϕ is expressed mathematically by a continuity equation

$$\phi_t + \nabla \cdot (\phi u) = 0$$

where u is the flow vector field. If we assume the flow is incompressible, we have

$$\phi_t + u \cdot \nabla\phi = 0.$$

In particular, if the flow is steady, we have

$$u \cdot \nabla\phi = 0.$$

Lemma 8.2.1 (Relation to Fokker-Planck equation). *The equation*

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left(\frac{D(x)}{kT}\frac{dU(x)}{dx}p\right) + \frac{\partial}{\partial x}D(x)\frac{\partial p}{\partial x}$$

has the following equilibrium solution

$$p \propto \exp\left(-\frac{U}{kT}\right).$$

This PDE can also be written as the form of Fokker-Planck equation as

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left(\frac{D(x)}{kT}\frac{dU}{dx}p - p\frac{\partial D(x)}{\partial x}\right) + \frac{\partial^2}{\partial x^2}(D(x)p)$$

with the associated Ito SDE as

$$dx(t) = (-\frac{D(x)}{kT}\frac{dU}{dx} + \frac{\partial D(x)}{\partial x})dt + \sqrt{2D(x)}dw(t)$$

Proof. Set $\frac{\partial p}{\partial t} = 0$, then we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{\partial U}{\partial x}p) + \frac{\partial}{\partial x}D(x)\frac{\partial p}{\partial x} \\
0 &= \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{\partial U}{\partial x}p + D(x)\frac{\partial p}{\partial x}) \\
0 &= (\frac{D(x)}{kT}\frac{dU}{dx}p + D(x)\frac{\partial p}{\partial x}) \\
\implies 0 &= \frac{D(x)}{kT}\frac{\partial U}{\partial x}p + D(x)\frac{\partial p}{\partial x} \\
\frac{d\ln p}{dx} &= -\frac{1}{kT}\frac{dU}{dx} \\
p &\propto \exp(-\frac{U}{kT})
\end{aligned}$$

□

8.3 Feyman Kac theorem and backward equation

8.3.1 Feyman Kac theorem

Theorem 8.3.1 (Feyman Kac theorem). *Consider the 1D parabolic*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)\frac{\partial^2 V}{\partial S^2} + m(S,t)\frac{\partial V}{\partial S} - rV = 0.$$

The solution is given as

$$V(S_t,t) = E_Q[e^{\int_t^T -r(\tau)d\tau} V(S_T,T)|\mathcal{F}_t]$$

where S_t is a stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS = mdt + \sigma dW_t$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau} V(S_t,t)$$

is a martingale under the measure Q.

Proof.

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= d(e^{\int_0^t r(\tau)d\tau})V(t) + e^{\int_0^t r(\tau)d\tau} dV(t) + d(e^{\int_0^t r(\tau)d\tau})dV \\ &= -e^{\int_0^t -r(\tau)d\tau} r(t)Vdt + e^{-\int_0^t r(\tau)d\tau} dV. \end{aligned}$$

Use the fact that

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}(dS)^2\frac{\partial^2 V}{\partial S^2} = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt.$$

Then plug in dV, we have

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= e^{\int_0^t -r(\tau)d\tau}(-r(t)Vdt + \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt) \\ &= e^{\int_0^t -r(\tau)d\tau}\sigma\frac{\partial V}{\partial S}dW_t \end{aligned}$$

Therefore, $e^{\int_0^t r(\tau)d\tau} V(S_t, t))$ is a martingale. Then we can easily show using martingale property:

$$V(s) = E_Q[e^{-\int_s^t r(\tau)d\tau} V(S_t, t)|\mathcal{F}_s].$$

□

Remark 8.3.1 (interpretation and financial applications in path-independent derivatives).

- Feyman Kac theorem shows that certain types of parabolic equation can be solved using stochastic differential equation method(by simulating trajectories and take expectations.) **Note that in parabolic differential equation S is not a random variable, S is simply a variable.**

Remark 8.3.2 (special case of $r = 0$). When $r = 0$, the dynamics of S will not change(i.e. Q will not change), then $V(S_t, t)$ is a martingale. And the parabolic equation becomes Kologorov backward equation.

Theorem 8.3.2 (Feyman Kac theorem, multi-dimensional). *Consider the multidimensional parabolic*

$$\frac{\partial V}{\partial t} + \sum_{i=1}^{N} \mu_i \frac{\partial V}{\partial S} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_{ij} \frac{\partial^2 V}{\partial S_i \partial S_j} + -rV = 0$$

where $\gamma_{ij} = \sum_{k=1}^{N} \sigma_{ik}\sigma_{jk}$. The solution is given as

$$V(s,t) = E_Q[e^{\int_s^t r(\tau)d\tau} V(S_t, t)|\mathcal{F}_s]$$

where S_t is a N dimensional stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS_i = \mu_i dt + \sum_{j=1}^{N} \sigma_{i,j} dW_j(t), i = 1, 2, ..., N$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau} V(S_t, t)$$

is a martingale under the measure Q.

Proof. Similar to 1D case. □

Theorem 8.3.3 (Feyman-Kac formula for PDE with source term). *Consider the PDE*

$$\frac{\partial u(x,t)}{\partial t} + \mu(x,t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 u(x,t)}{\partial x^2} - V(x,t)u(x,t) + f(x,t) = 0$$

with $x \in \mathbb{R}, t \in [0,T]$ and terminal condition of

$$u(x,T) = \psi(x)$$

then the solution to the PDE is

$$u(x,t) = E^Q[\int_t^T e^{-\int_t^r V(X_\tau,\tau}d\tau]f(X_r,r)dr - e^{-\int_t^T V(X_\tau,\tau}d\tau\psi(X_T)|X_t = x]$$

under the probability measure Q such that X is an Ito process driven by the equation

$$dX = \mu(X,t)dt + \sigma(X,t)dW^Q$$

with W^Q being the a Wiener process under Q.

Remark 8.3.3 (solving PDE by simulation)**.**

- It offers a method of solving certain PDEs by simulating random paths of a stochastic process. The procedure for simulating a trajectory will be: simulate X_t from t to T using $dX = \mu(X,t)dt + \sigma(X,t)dW^Q$ with this trajectory we can evaluate

$$e^{-\int_t^r V(X_\tau,\tau}d\tau$$

  and

$$\int_t^T e^{-\int_t^r V(X_\tau,\tau)d\tau}f(X_r,r)dr$$

  to obtain one sample.
- The simulation evaluation approach clearly demonstrates the expectation will depends on $X_t = x$, i.e., x and t.

Remark 8.3.4. an important class of expectations of random processes can be computed by deterministic methods.

Corollary 8.3.3.1 (Black-Scholes equation). *The Black-Scholes equation given as*

$$\frac{\partial V(s,t)}{\partial t} + rs\frac{\partial V(s,t)}{\partial s} + \frac{1}{2}\sigma^2 s^2\frac{\partial^2 V(s,t)}{\partial s^2} - rV(s,t) = 0$$

with $s \in \mathbb{R}, t \in [0,T]$ and terminal condition of $V(s,T) = \psi(S_T)$ has the solution in conditional expectation form as

$$V(s,t) = E^Q[\psi(S_T)\exp(-r(T-t))|S_t = s]$$

under probability measure Q such that S_t is an Ito process given by

$$dS = rSdt + \sigma SdW.$$

Moreover, under probability measure Q, S has solution given as

$$S(T) = S(t)\exp((r - \frac{1}{2}\sigma^2)(T-t) + \sigma(B(T) - B(t)))$$

and therefore

$$V(s,t) = \int_{-\infty}^{\infty} \psi(S_T)\exp(-r(T-t))f(B(T) - B(t) = x)dx$$

Remark 8.3.5 (simulation approach to Black-Scholes equation). We can evaluate the expectation by simulating trajectories of S_t. More precisely, starting with initial condition $S_t = s$ and use Euler algorithm to integrate

$$dS = rSdt + \sigma SdW.$$

8.3.2 Backward equation

Theorem 8.3.4 (Kolmogorov backward equation, expectation pricing). *Assume X_t is governed by the following SDE*

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

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x,t) = \mu\frac{\partial}{\partial x}f(x,t) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}f(x,t)$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t,t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t, t)dt.$$

then we have (use the notation $E[f(X_s, s)|X_t = t] = E_{x,t}[f(X_s, s)]$)

$$\begin{aligned} E_{x,t}[f(X_{t+dt}, t+dt)] &= E_{x,t}[f(x + dX_t, t + dt)] \\ &\approx E_{x,t}[f(x,t) + \partial_x f(x,t)dX_t + \frac{1}{2}\partial_x^2 f(x,t)dX_t^2 + \partial_t f(x,t)dt] \\ &= f(x,t) + \partial_x f(x,t)E_{x,t}[dX_t] + \frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2] + \partial_t f(x,t)dt \\ &= f(x,t) + dt(\partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t)) \end{aligned}$$

$$E_{x,t}[f(X_{t+dt}, t+dt)] = f(x,t) \implies \partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t) = 0$$

□

Remark 8.3.6 (interpretation).

- We can interpret $f(x,t)$ as the price at t and the state is at x. When we cannot hedge the risk, we can price the asset by the expected value of payoff with respect to the real probability. Note that such pricing method does not take into account of the risk-aversion.
- The only difference to Black-Scholes is the extra source decreasing term.

Theorem 8.3.5 (Kolmogorov backward equation with discount). *Assume X_t is governed by the following SDE*

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

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[e^{-\int_t^T r(tau)d\tau}V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x,t) = \mu\frac{\partial}{\partial x}f(x,t) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}f(x,t) - rf$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t, t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t, t)dt.$$

then we have (use the notation $E[f(X_s,s)|X_t=t]=E_{x,t}[f(X_s,s)]$)

$$\begin{aligned}E_{x,t}[(1-rdt)f(X_{t+dt},t+dt)] &= E_{x,t}[(1-rdt)f(x+dX_t,t+dt)]\\ &\approx (1-rdt)E_{x,t}[f(x,t)+\partial_x f(x,t)dX_t+\frac{1}{2}\partial_x^2 f(x,t)dX_t^2+\partial_t f(x,t)dt]\\ &= f(x,t)-rf(x,t)dt+\partial_x f(x,t)E_{x,t}[dX_t]+\frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2]+\partial_t f(x,t)dt\\ &= f(x,t)-rf(x,t)dt+dt(\partial_t f(x,t)+\mu(x,t)\partial_x f(x,t)+\frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t))\end{aligned}$$

$$E_{x,t}[(1-rdt)f(X_{t+dt},t+dt)]=f(x,t)$$

$$\implies \partial_t f(x,t)+\mu(x,t)\partial_x f(x,t)+\frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t)-rf(x,t)=0$$

□

Theorem 8.3.6 (Kolmogorov backward equation, multi-dimensional version). *Assume X_1,X_2 is governed by the following SDE*

$$dX_1(t)=\mu_1 dt+\sigma_1 dW_1(t).$$

and

$$dX_2(t)=\mu_2 dt+\sigma_2 dW_2(t).$$

Suppose we are given a payoff $V(X_1(T),X_2(T))$ at time T. Define

$$f(x_1,x_2,t)=E[e^{-\int_t^T r(\tau)d\tau}V(X_1(T),X_2(T)|X_1(t)=x_1,X_2(t)=x_2]$$

for all $t\leq T$. Then $f(x_1,x_2,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x_1,x_2,t)=\mu_1\frac{\partial}{\partial x_1}f+\mu_2\frac{\partial}{\partial x_2}f++\frac{1}{2}\sigma_1^2\frac{\partial^2}{\partial x_2^2}f+\frac{1}{2}\sigma_2^2\frac{\partial^2}{\partial x_2^2}f+\sigma_1\sigma_2\frac{\partial^2}{\partial x_1\partial x_2}f-rf$$

for $t\leq T$, subject to the final condition $f(x_1,x_2,T)=V(X_1(T),X_2(T))$.

Proof. Similar to 1D case. □

Remark 8.3.7 (Kolmogorov forward and backward equation).

- the Kolmogorov forward equation addresses the following problem: We have information about the state x of the system at time t as $P(x,t)$, we want to know $P(x,s), s>t$.
- The Kolmogorov backward equation addresses the problem: at time t in whether at a future time s the system will be in a given subset of states B, sometimes called the

target set. The target is described by a indicator function for the set B. We want to know for every state x at time t, $(t < s)$ what is the probability of ending up in the target set at time s (sometimes called the hit probability). In this case $u_s(x)$ serves as the final condition of the PDE, which is integrated backward in time, from s to t.

8.3.3 Application to first hitting probability

Lemma 8.3.1 (General method via Feyman Kac formula). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given two levels $a > 0$ and $-b, b > 0$, and let the probability $P(t,x)$ denote the probability that the process starting at $X(t) = x$ hits a before hitting $-b$. Then we have

- *$P(t,x)$ is independent of time t.*
- *The governing equation for $P(x)$ is given by*

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(x = a) = 1, P(x = -b) = 0$.

Proof. (1) Note that this is a Markov process, therefore $P(t,x)$ will be depend on time. (2) Consider a value function $P(x,t) = E[P_T|X(t) = x]$ with final condition $P(x,T) = P_T$(P_T will take value 1 at target sites and take 0 elsewhere). Then from Feyman Kac theorem(Theorem 8.3.1), $P(x,t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $p(t,x) = P_T$. □

8.4 Advanced analysis for Brownian motion

Lemma 8.4.1 (Kramers equation). *The following SDE for a particle subject to Brownian force*

$$\begin{aligned} dv &= -\gamma v dt + \sigma dW_t \\ dx &= v dt \end{aligned}$$

is associated with the Fokker-Planck equation on $p(x, v, t)$, given as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(vp) + \frac{\partial}{\partial v}(\gamma v p) + \frac{1}{2}\sigma \frac{\partial^2 p}{\partial v^2}.$$

Proof. Direct from Lemma 8.1.2. Note that the variance coefficient for x is zero. □

Theorem 8.4.1 (fluctuation-dissipation theorem). *[5, p. 3] Consider the SDE for velocity*

$$dV = -\gamma V dt + \sigma dW_t.$$

Under the assumption $\frac{1}{2}m\left\langle V(t)^2 \right\rangle = \frac{1}{2}kT$, the coefficient σ and γ are connected via

$$\sigma = \sqrt{2\gamma kT/m}.$$

Proof. From the solution of OU process(Lemma 6.4.1), we know that

$$\left\langle V(t)^2 \right\rangle = \frac{\sigma^2}{2\gamma} \implies \frac{\sigma^2}{2\gamma} = \frac{1}{m}kT.$$

□

Lemma 8.4.2 (auto-correlation of velocities). *[5, p. 34] Suppose v is governed by*

$$dv = -\gamma v dt + \sigma dW_t.$$

Then

$$\langle v(t_1)v(t_2) \rangle = v_0^2 e^{-\gamma(t_1+t_2)} + \frac{\sigma}{2\gamma}(e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}).$$

For large t_1, t_2 such that $\gamma t_1 \gg 1, \gamma t_2 \gg 1$, we have

$$\langle v(t_1)v(t_2) \rangle = \frac{\sigma}{2\gamma} e^{-\gamma|t_1-t_2|}.$$

Proof. Note that from Lemma 8.1.2, then

$$v(t_1) = v_0 e^{-\gamma t_1} + \int_0^{t_1} e^{-\gamma(t_1 - t)} \sigma dW_t.$$

We can then proceed. □

Lemma 8.4.3 (mean displacement analysis and fluctuation-dissipation theorem for colloidal particles). *For a colloidal particle with hydrodynamic drag given as*

$$\gamma = \frac{6\pi\mu a}{m}$$

And the equation of motion is given as

$$\begin{aligned} dv &= -\frac{\gamma}{m} v dt + \frac{\sigma}{m} dW_t \\ dx &= v dt. \end{aligned}$$

Then, it can be showed that

$$\left\langle (x(t) - x(0)^2 \right\rangle = \frac{\sigma^2}{\gamma^2} t = 2Dt$$

where

$$D = \frac{kT}{\gamma} = \frac{kT}{6\pi\mu a}.$$

Proof. Using Theorem 8.4.1.

$$\begin{aligned} \left\langle (x(t) - x(0)^2 \right\rangle &= \int_0^t \int_0^t \langle v(t_1) v(t_2) \rangle dt_1 dt_2 \\ &= \left\langle v_0(t)^2 \right\rangle \frac{2}{\gamma} t = \frac{\sigma^2}{\gamma^2} = \frac{2kT}{\gamma} t = 2Dt \end{aligned}$$

□

Remark 8.4.1. There are two versions of fluctuation-dissipation theorem(Theorem 8.4.1,Lemma 8.4.3) Note that The variance of random forces is fundamentally different from the variance of the random displacements.

Lemma 8.4.4 (Maxwell distribution). *Consider the SDE for velocity*

$$dv = -\gamma v dt + \sigma dW_t.$$

Assume the fluctuation-dissipation theorem holds. The equilibrium distribution of the velocity is given as

$$p(v) = \sqrt{\frac{m}{2\pi kT}} \exp(-\frac{mv^2}{2kT}).$$

This distribution is also known as Maxwell distribution.

Proof. From the solution of OU process(Lemma 6.4.1), we know that

$$\left\langle V(t)^2 \right\rangle = \frac{1}{m} kT.$$

□

Theorem 8.4.2 (microscopic to continuum description,Green-Kubo formula). *[6, p. 177] From Lemma 8.4.3, we can claim that at the long time limit, the probability distribution of a Brownian particle is given by*

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}.$$

From which we can show

$$\frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle = 2D$$

$$D = \frac{1}{2}\int_0^\infty \langle v(t)v(0) \rangle dt$$

Proof. (1)From SDE in Lemma 8.4.3, we know that the probability distribution of $x(t)$ is Gaussian, which is also the solution to this equation in this theorem. (2)

$$\begin{aligned}
\frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle &= \frac{d}{dt}\|x(t)\|^2 \rho(x,t)dx = D\int \|x\|^2 \nabla^2 \rho(x,t)dx \\
&= D\int \nabla \cdot (\|x\|^2 \nabla\rho)dx - D\int (\nabla\|x\|^2)\cdot\nabla\rho dx \\
&= -2D\int x\cdot\nabla\rho dx \\
&= 2D\int (\nabla\cdot x)\rho dx = 2D
\end{aligned}$$

Using $x(t) = \int_0^t v(t_1)dt_1$, we have

$$\begin{aligned}\left\langle \|x(t)\|^2 \right\rangle &= \left\langle \int_0^t \int_0^t v(t_1)v(t_2)dt_1dt_2 \right\rangle \\ &= \int_0^t \int_0^t \langle\, v(t_1-t_2)v(0)\,\rangle dt_1dt_2 \\ \implies \frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle &= \int_0^t \langle\, v(t)v(0)\,\rangle dt.\end{aligned}$$

□

Remark 8.4.2 (interpretation). This equation relates a macroscopic transport coefficient D to a quantity in microscopic system. Note that the velocity autocorrelation function $g(t) = \langle\, v(t)v(0)\,\rangle$ can be calculated using microscopic simulations such as molecular dynamics. Similar expressions can be found for viscosity, heat conductivity, etc.

8.4.1 Kramers problem: barrier escape

Lemma 8.4.5.

$$\frac{dx}{dt} = -\frac{1}{\gamma}U' + \frac{1}{\gamma}\xi$$

The corresponding Fokker-Planck equation is given as

$$\frac{\partial}{\partial t}P(x,t|x_0,0) = \frac{\partial}{\partial x}\frac{1}{\gamma}U'P + \frac{\partial^2}{\partial x^2}DP.$$

Let $T(x)$ be the mean first passage time, we have

$$-\frac{1}{\gamma}U'\frac{d}{dx}T(x) + D\frac{d^2}{dx^2}T(x) = -1$$

with boundary condition given as $T(x_{target}) = 0$.

Proof. Multiply both sides by $\exp(-\beta U)$, we have

$$(\frac{d}{dx}e^{-\beta U})\frac{d}{dx}T + e^{-\beta U}\frac{d^2}{dx^2}T(x) = -\frac{1}{D}e^{-\beta U}$$

or

$$\frac{d}{dx}[e^{-\beta U}\frac{d}{dx}T(x)] = -\frac{1}{D}e^{-\beta U}.$$

Integrating from $-\infty$ to x we have

$$\frac{d}{dx}T(x) = -\frac{1}{D}e^{\beta U}\int_{-\infty}^{x} dz e^{-\beta U(z)}.$$

Finally, integrating from x_{target} to x, we have

$$T(x) = \int_{x_{target}}^{x} \frac{1}{D} dy e^{\beta U(y)} \int_{-\infty}^{x} dz e^{-\beta U(z)}.$$

□

Remark 8.4.3 (approximate solution). In the first integral, the integrand is largest around the barrier top x_b, and we expand the external potential aournd this point as

$$U(y) = U_b - \frac{1}{2}mw_b^2(y - x_b)^2.$$

In the second integral, the integrand is largest around x_a and

$$U(z) = U_a + \frac{1}{2}mw_b^2(z - x_{target})^2.$$

Further assume D is a constant, then the mean time from a valley to another valley over the barrier is

$$\tau = \frac{1}{D}\int_{-\infty}^{\infty} dy e^{\beta U_b} e^{-\frac{1}{2}\beta m w_b^2 (y-x_b)^2} \int_{-\infty}^{\infty} e^{-\beta U_{target}} e^{-\frac{1}{2}\beta m w_{target}^2 (z-x_{target})^2} = \frac{1}{D}\frac{2\pi k_B T}{m w_{target} w_b} e^{\beta(U_b - U_{target})}.$$

8.5 Notes on bibliography

For Fokker-Planck equation, see [5][7].

BIBLIOGRAPHY

1. Chirikjian, G. S. *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods* (Springer Science & Business Media, 2011).
2. Clark, I. J. *Foreign exchange option pricing: a practitioner's guide* (John Wiley & Sons, 2011).
3. Pavliotis, G. A. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, 2014).
4. Liu, C., Brennan, D. P., Bhattacharjee, A. & Boozer, A. H. Adjoint Fokker-Planck equation and runaway electron dynamics. *Physics of Plasmas* **23,** 010702 (2016).
5. Risken, H. & Frank, T. *The Fokker-Planck Equation: Methods of Solution and Applications* ISBN: 9783642615443. https://books.google.com/books?id=6B5uCQAAQBAJ (Springer Berlin Heidelberg, 2012).
6. Weinan, E. *Principles of multiscale modeling* (Cambridge University Press, 2011).
7. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127. https://books.google.com/books?id=otg3PQAACAAJ (Springer Berlin Heidelberg, 2009).

Part II

PORTFOLIO THEORY AND ASSET PRICING

9

PORTFOLIO OPTIMIZATION & ASSET PRICING

9.1 Utility function pricing principles

9.1.1 Utility function

Definition 9.1.1 (utility function). *[1, p. 229] A utility function $U : \mathbb{R} \to \mathbb{R}$ is an increasing continuous function.*

Example 9.1.1. [1, p. 229]

- $U(x) = -e^{-ax}$ for some $a > 0$.
- $U(x) = \ln(x)$.
- $U(x) = bx^b$ for some $b \leq 1, b \neq 0$.
- $U(x) = x - bx^2$ for some $b > 0$ when $x < 1/(2b)$.
- risk-neutral utility $U(x) = x$.

Definition 9.1.2 (concavity, risk aversion). *[1, p. 233] A utility function is strictly concave in its domain is said to be* ***risk averse****. The* ***degree of risk aversion*** *is defined by* ***Arrow-Pratt absolute risk aversion coefficient****, given by*

$$\alpha(x) = -\frac{U''(x)}{U'(x)}.$$

Example 9.1.2.

- For $U(x) = -e^{-ax}$, $\alpha(x) = a$.
- For $U(x) = \ln(x)$, $\alpha(x) = 1/x$; that is, the larger x, the less sensitive of risk aversion.
- For $U(x) = x$, $\alpha(x) = 0$; that is, linear utility function has no risk aversion

Lemma 9.1.1 (invariance of linear transformation). *Suppose that the decision-maker's preferences over* ***all monetary gambles, each with different random payoff****, can be represented by a utility function U. Then the new utility function V, defined by*

$$V(x) = AU(x) + B, \forall x \in \mathbb{R},$$

then $V(x)$ is also a valid utility function for representing this decision-maker preferences over all monetary gambles.

Proof. For any two gambles with random payoffs X and Y, if $E[U(X)] > E[U(Y)]$, then

$$\begin{aligned} E[V(X)] &= E[AU(X) + B] \\ &= AE[U(X)] + B \\ &> AE[U(Y)] + B \\ &= E[AU(Y) + B] \\ &= V(Y). \end{aligned}$$

□

9.1.2 Portfolio optimization & pricing

9.1.2.1 *General theory*

Definition 9.1.3 (utility maximization problem for investors). *Let $U : \mathbb{R} \to \mathbb{R}$ be the increasing continuous function. Consider n securities with one period random payoff $d_1, d_2, ..., d_n$ and prices $p_1, p_2, ..., p_n$. The utility maximization problem is given by*

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} E[U(\sum_{i=1}^n \theta_i d_i)] \\ \text{subject to } &\sum_{i=1}^n \theta_i d_i \geq 0 \\ &\sum_{i=1}^n \theta_i p_i \leq W. \end{aligned}$$

Definition 9.1.4 (arbitrage opportunity). *[1, p. 241] The existence of an arbitrage opportunity refers to the existence of a portfolio vector θ such that either one of the following holds. Let $(P, \Omega, \mathcal{F})$ be the probability model of the returns.*

- $\theta^T p \leq 0$, *but* $\theta^T r > 0$ *for all* $\{A \in \mathcal{F} : P(A > 0)\}$.
- $\theta^T p < 0$, *but* $\theta^T r >= 0$ *for all* $\{A \in \mathcal{F} : P(A > 0)\}$.

Theorem 9.1.1 (no arbitrage condition and the existence of solutions). *[1, p. 243] Let $U(x)$ be an increasing and continuous utility function such that $x \to \infty \implies U(x) \to \infty$. Suppose there exists a feasible solution x^0 such that $\sum_{i=1}^n \theta_i d_i > 0$.*

*It follows that the utility maximization problem for investors has a **finite** solution if and only if there is no arbitrage.*

Proof. (informal) If there are arbitrage opportunities, then we can constantly increase $U(x)$ and make the constraint $\sum_{i=1}^n \theta_i p_i \leq W$ satisfied. □

Lemma 9.1.2 (necessary optimality condition and pricing). *[1, p. 243] The necessary KKT condition for the utility maximization problem is that there exists a Lagrangian multiplier $\lambda \geq 0$ such that*

$$E[U'(\sum_{i=1}^n \theta_i d_i] = \lambda p_i, i = 1, 2, ..., n$$
$$\sum_{i=1}^n \theta_i = 1$$

Lemma 9.1.3 (pricing equation). *[1, p. 244] Let $x^* \triangleq \sum_{i=1}^n \theta_i^* d_i$ be the solution to the utility maximization problem, then*

$$E[U'(x^*)d_i] = \lambda P_i, i = 1, 2, ..., n. \lambda > 0.$$

If there exists a risk-free asset with return R, then

$$P_i = \frac{E[U'(x^*)d_i]}{RE[U'(x^*)]}, i = 1, 2, ..., n.$$

9.1.2.2 *Log-optimal pricing*

Theorem 9.1.2 (Log-optimal pricing). *[1, p. 246]*

$$P = E[\frac{d}{R^*}]$$

where R^* is the return on the log-optimal portfolio.

9.1.3 Certainty equivalent

Definition 9.1.5 (certainty equivalent). *Consider a gamble game with random payoff X. Consider a risk taker with utility function $U(x)$. The certain equivalent CE for this risk taker is a real-valued number CE such that*

$$U(CE) = E[U(X)].$$

Lemma 9.1.4 (linearity of certainty equivalent for constant risk tolerance utility). *Consider n independent gamble games with random payoff $X_i, i = 1, 2, ..., n$. Consider a risk taker with utility function $U(x) = -\exp(-x/\rho)$. The certain equivalent CE for the aggregate game with payoff $X = \sum_{i=1}^n X_i$ for this risk taker is sum of certainty equivalent for each individual game*

$$CE = \sum_{i=1}^n CE_i.$$

Proof.

$$\begin{aligned}
U(CE) &= E[U(\sum_{i=1}^n X_i)] \\
CE &= -\rho \ln E[\exp(-\sum_{i=1}^n X_i/\rho)] \\
&= -\rho \ln E[\exp(-\sum_{i=1}^n X_i/\rho)] \\
&= -\rho \ln \prod_{i=1}^n E[\exp(X_i/\rho)] \\
&= -\rho \sum_{i=1}^n E[\exp(X_i/\rho)] \\
&= \sum_{i=1}^n CE_i
\end{aligned}$$

□

Lemma 9.1.5 (certainty equivalent for normal payoff with constant risk tolerance utility). *Consider a gamble game with random payoff $X \sim N(\mu, \sigma^2)$. Consider a risk taker with utility function $U(x) = -\exp(-x/\rho)$. The certain equivalent CE for this risk taker is*

$$CE = \mu - \frac{\sigma^2}{2\rho},$$

such that $U(CE) = E[U(X)]$.

Proof. First we have:

$$\begin{aligned}
E[U(X)] &= \int_{-\infty}^{\infty} -\exp(-x/\rho)\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -\exp(-x/\rho)\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -\exp(-\frac{(x+\sigma^2/\rho-\mu)^2}{2\sigma^2})\exp(\frac{\mu}{\rho}-\frac{\sigma^2}{2\rho^2}) \\
&= -\exp(-\frac{\mu}{\rho}+\frac{\sigma^2}{2\rho^2}).
\end{aligned}$$

Then

$$U(CE) = -\exp(-\frac{CE}{\rho}) = -\exp(-\frac{\mu}{\rho}+\frac{\sigma^2}{2\rho^2}) \implies CE = \mu - \frac{\sigma^2}{2\rho}.$$

□

Note 9.1.1 (market price of risk, risk premium)**.**

9.1.4 Optimal risk sharing

Definition 9.1.6 (risk premium). *The difference between the expected monetary value of a gamble and a risk-averse decision-maker's certainty equivalent of the gamble is called the decision-maker's* ***risk premium.***

Example 9.1.3. Consider a lottery paying 20000 or 0, each with probability 1/2. Suppose A's certainty equivalent is 7000; while B's certainty equivalent is 6000(that is, B is more risk averse than A). Finally, the lottery has risk premium 3000 for A, and 4000 risk premium for B.

Example 9.1.4 (risk sharing example among individuals with constant risk tolerance).

- Consider a risky project with random payoff $X \sim N(\mu, \sigma^2), \mu = 35000, \sigma = 25000$.
- Consider two individuals A and B with constant risk tolerance given by

$$U_1(x) = -\exp(-\frac{x}{r_1}), U_2(x) = -\exp(-\frac{x}{r_2}),$$

where $r_1 = 20000, r_2 = 30000$.

- The certainty equivalent of the project for A and B(Lemma 9.1.5)

$$CE_A = \mu - \frac{\sigma^2}{2r_1} = 19375,$$

and

$$CE_B = \mu - \frac{\sigma^2}{2r_1} = 24583.$$

- Now suppose A and B cooperate in such a way that each participates 50% of the project and the payoff for each is $N(\mu/2, \sigma^2/4)$. The certainty equivalent of the project for A and B

$$CE'_A = \mu/2 - \frac{\sigma^2/4}{2r_1} = 13594,$$

and

$$CE'_B = \mu/2 - \frac{\sigma^2/4}{2r_1} = 14896.$$

Theorem 9.1.3 (optimal risk sharing). *link Consider an investment partnership with N members. Suppose they hold assets which will have random payoff Y. Let $x_i(y)$ denote member i planned payoff when the investment has realized payoff $Y = y$. The functions $x_i(y), i = 1, 2, ..., N$ will subject to constraints $\sum_{i=1}^N x_i(y) = y, \forall y \in \mathbb{R}$.*

It follows that

- *The optimal risk sharing problem is formalized as*

$$\max_{x_i(\cdot), i=1,...,N} \sum_{i=1}^N \lambda_i E[u_i(x_i(Y))]$$
$$s.t. \quad \sum_{i=1}^N x_i(y) = y, \forall y \in \mathbb{R}$$

where $\lambda_1, \lambda_2, ..., \lambda_N$ are given positive utility weights.
- *The optimal solution for $x_i(\cdot), i = 1, 2, ..., N$ is given by*

$$\frac{\partial x_i}{\partial y} = \frac{\tau_i(x_i(y))}{\sum_{i=1}^N \tau_j(x_i(y))},$$

where $\tau_i(w) \triangleq -u'_i(w)/u''_i(w)$.

Proof. For each outcome y, the first order KKT condition for equality constraint (Theorem 15.6.3) gives

$$\lambda_i u_i'(x_i(y)) = v(y), i = 1, 2, ..., N, y \in \mathbb{R}, (*)$$

where $v(y)$ is the Lagrange multiplier parameterized by y.

Now differentiate with respect to y, we have

$$\lambda_i u_i''(x_i(y)) \frac{\partial x_i}{\partial y} = v'(y), i = 1, 2, ..., N, y \in \mathbb{R}.$$

Note that equation (*) implies $\lambda_i = v(y)/u_i'(x_i(y))$. Plug into above, we get

$$u_i''(x_i(y))/u_i'(x_i(y)) \frac{\partial x_i}{\partial y} = v'(y)/v(y), i = 1, 2, ..., N, y \in \mathbb{R}.$$

Rearrange, we get

$$\frac{\partial x_i}{\partial y} = -v'(y)/v(y)\tau_i(x_i), i = 1, 2, ..., N, \forall y \in \mathbb{R}.$$

Differentiate the constraint $\sum_{i=1}^N x_i(y) = y$ gives

$$\sum_{i=1}^N \partial x_i(y)/\partial y = 1.$$

Finally, we get

$$\frac{\partial x_i}{\partial y} = \frac{\tau_i(x_i(y))}{\sum_{i=1}^N \tau_j(x_i(y))}.$$

□

Corollary 9.1.3.1 (optimal risk sharing with constant risk tolerance). *Consider an investment partnership with N members. Suppose they hold assets which will have random payoff Y. Let $x_i(y)$ denote member i planned payoff when the investment has realized payoff $Y = y$. Suppose the N investors each has constant risk tolerance of parameter T_i, then the optimal risk sharing is given by*

$$\frac{\partial x_i}{\partial y} = \frac{\tau_i(x_i(y))}{\sum_{i=1}^N \tau_j(x_i(y))} = \frac{T_i}{\sum_{j=1}^N T_j}, i = 1, 2, ..., N,$$

with solution

$$x_i(y) = x_i(0) + y\frac{T_i}{\sum_{j=1}^N T_j}.$$

Proof. Note that for a constant risk tolerance utility function $u_i(x) = -\exp(-x/T_i)$, we have $T_i = -u_i'(x)/u_i''(x)$. □

Remark 9.1.1. For investors each has constant risk tolerance of parameter T_i the result of

$$x_i(y) = x_i(0) + y\frac{T_i}{\sum_{j=1}^N T_j},$$

indicates that these investors should linearly share all risks in proportion to their risk tolerances.

9.2 Markovitz Portfolio Optimization Model

9.2.0.1 *The optimization framework*

Definition 9.2.1. *A* ***portfolio vector*** *is a vector $w \in \mathbb{R}^d$, with the constraint $\sum_i w_i = 1$.*

Lemma 9.2.1 (return and variance of a portfolio). *[1, p. 150] Suppose in our universe, there are n stocks. We are further given n estimated return $E(r_i)$, and the covariance matrix $Cov(r_i, r_j) = \Sigma$. For a portfolio characterized by $w^T r$, where w is the portfolio vector $w \in \mathbb{R}^n, \sum_{i=1}^n w_i = 1$ and r is the random variable vector $r \in \mathbb{R}^n$ with each component r_i characterized the return rate of asset i, the expected total return and the variance are given as*

$$E(w^T r) = w^T E(r) = \sum_{i=1}^n w_i E(r_i)$$

$$Var(w^T r) = w^T Cov(i,j) w = \sum_{i,j}^n w_i w_j \sigma_{ij}$$

Proof. See Lemma 1.5.1,Lemma 1.6.1. □

Remark 9.2.1 (interpretation).

- The return is a 'normalized' quantity; that is, for every unit money invested, how much extra profit is generated in a single period. The return can be calculated as

$$r_i = \frac{X_{1i} - X_{0i}}{X_{0i}}$$

  where X_{1i} is the value of asset i after a single period and X_{0i} is the initial value of the asset i.
- We require $\sum_{i=1}^n w_i = 1$ can be thought as we split a unit money into investment of different assets.

Definition 9.2.2 (mean-variance optimization formulation, minimum variance at fixed return). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n$ and the covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$. Given an arbitrary* *value μ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\begin{aligned} \min_w \frac{1}{2}\sum_{i,j}^n w_i w_j \sigma_{ij} \\ s.t. \sum_{i=1}^n w_i \mu_i = \mu_0 \\ \sum_i^n w_i = 1. \end{aligned}$$

Definition 9.2.3 (mean-variance optimization formulation, alternative formulation). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n$ and the covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$.*

- *(return bounded formulation) Given an arbitrary value μ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with at least such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\begin{aligned} \min_x \frac{1}{2}\sum_{i,j}^n w_i w_j \sigma_{ij} \\ s.t. \sum_{i=1}^n w_i \mu_i \geq \mu_0 \\ \sum_i^n w_i = 1. \end{aligned}$$

- *(maximum return at fixed variance level) Given an arbitrary value σ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with maximum mean return and variance no greater than σ_0. The mean-variance optimization problem is given as*

$$\begin{aligned} \max_x \sum_i^n w_i \mu_i \\ s.t. \sum_{i,j=1}^n w_i w_j \sigma_{ij} = \sigma_0^2 \\ \sum_i^n w_i = 1. \end{aligned}$$

- *(maximum return with bounded variance level) Given an arbitrary value σ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with maximum mean return and variance no greater than σ_0. The mean-variance optimization problem is given as*

$$\begin{aligned} \max_x & \sum_i^n w_i \mu_i \\ s.t. & \sum_{i,j=1}^n w_i w_j \sigma_{ij} \leq \sigma_0^2 \\ & \sum_i^n w_i = 1. \end{aligned}$$

- *(risk-reversion formulation) Given a positive number λ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with maximum utility $u(x)$ given by $u(x) = E[x] - \lambda Var[X]$. The mean-variance optimization problem is given as*

$$\begin{aligned} \max_x & \sum_i^n w_i \mu_i - \sum_{i,j=1}^n \lambda w_i w_j \sigma_{ij} \\ s.t. & \sum_i^n w_i = 1. \end{aligned}$$

Theorem 9.2.1 (sufficient and necessary conditions for efficient portfolio). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n > 0$ and the* ***positive-definite*** *covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$. Consider the mean-variance optimization problem Given an arbitrary value μ_0, we want to construct with such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\begin{aligned} \min_w & \frac{1}{2} \sum_{i,j}^n w_i w_j \sigma_{ij} \\ s.t. & \sum_{i=1}^n w_i \mu_i = \mu_0 \\ & \sum_i^n w_i = 1. \end{aligned}$$

where the portfolio is characterized by weight vector $w \in \mathbb{R}^n$ and μ_0 is preselected number.

It follows that

- *There exists weights $w_i, i = 1, ..., n$ and the two Lagrange multipliers λ_1 and λ_2 satisfying*

$$\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda_1 \mu_i - \lambda_2 = 0, i = 1, ..., n$$

$$\sum_{i=1}^{n} w_i \mu_i = \mu_0$$

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

- *Because Σ is positive definite, there exists unique $(w, \lambda_1, \lambda_2)$ to the above linear system; Moreover, w is the unique strict global minimizer.*

Proof. See the KKT condition for quadratic optimization **??????**. □

Theorem 9.2.2 (equivalence of different formulations). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n > 0$ and the* ***positive-definite*** *covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$. Consider the mean-variance optimization problem Given an arbitrary value μ_0, we want to construct with such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\max_{w} \sum_{i=1}^{n} w_i \mu_i - \frac{Q}{2} \sum_{i,j}^{n} w_i w_j \sigma_{ij}$$
$$s.t. \sum_{i}^{n} w_i = 1.$$

where the portfolio is characterized by weight vector $w \in \mathbb{R}^n$, Q is the risk aversion parameter and μ_0 is preselected number.

It follows that

- *The KKT necessary condition is that There exists weights $w_i, i = 1, ..., n$ and the two Lagrange multipliers λ satisfying*

$$\mu - Q \sum_{j=1}^{n} \sigma_{ij} w_j - \lambda_2 = 0, i = 1, ..., n$$

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

or equivalently, the linear system

$$\begin{bmatrix} Q\Sigma & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} -w \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mu \\ 1 \end{bmatrix}$$

- *Let* $[f(\Sigma,\mu), g(\Sigma,\mu)]^T, f \in \mathbb{R}^n$ *be the solution of the above linear system; that is*

$$\begin{bmatrix} Q\Sigma & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} f(\Sigma,\mu) \\ g(\Sigma,\mu) \end{bmatrix} = \begin{bmatrix} -\mu \\ 1 \end{bmatrix},$$

  then the unique solution(which is strict global optimizer) is given by

$$w = -\frac{f(\Sigma,\mu)}{Q}, \lambda = g(\Sigma,\mu)$$

- *Consider the optimization problem*

$$\begin{aligned} \max_x \sum_i^n w_i\mu_i \\ s.t. \sum_{i,j=1}^n w_iw_j\sigma_{ij} = \sigma_0^2 \\ \sum_i^n w_i = 1. \end{aligned}$$

  - *The KKT condition associated with optimization is given by*

$$\begin{aligned} \mu - \lambda_1\Sigma w - \lambda_2 = 0 \\ \mu^T w = \mu_0 \\ w^T\mathbf{1} = 1 \end{aligned}$$

  - *The KKT system has unique solutions*

$$\lambda_1 = \sqrt{\frac{f^T\Sigma f}{\sigma_0^2}}, w = -\frac{f(\Sigma,\mu)}{\lambda_1}, \lambda_2 = g(\Sigma,\mu).$$

  - *If we take* $Q = \lambda_1 = \sqrt{\frac{f^T\Sigma f}{\sigma_0^2}}$*, risk-aversion formulation is the same as maximium return formulation.*

- *Consider the optimization problem*

$$\min_w \frac{1}{2}\sum_{i,j}^{n} w_i w_j \sigma_{ij}$$
$$s.t. \sum_{i=1}^{n} w_i\mu_i = \mu_0$$
$$\sum_{i}^{n} w_i = 1.$$

  – *The KKT condition associated with this optimization is given by*

$$\begin{aligned}\Sigma w - \lambda_1\mu - \lambda_2 &= 0\\ \mu^T w &= \mu_0\\ w^T\mathbf{1} &= 1\end{aligned}$$

  – *The KKT system has unique solutions*

$$\lambda_1 = \sqrt{\frac{\sigma_0^2}{f^T\Sigma f}}, w = \lambda_1 f(\Sigma,\mu), \lambda_2 = -\lambda_1 g(\Sigma,\mu).$$

  – *If we take $Q = 1/\lambda_1 = \sqrt{\frac{f^T\Sigma f}{\sigma_0^2}}$, risk-aversion formulation is the same as maximium return formulation.*

Remark 9.2.2 (interpretation)**.**

- The efficient frontier plot in (σ_x, μ_x) is a half bullet shape.
- Because the covariance matrix is usually positive definite, then any portfolio will have an intrinsic risk/volatility that cannot be diversified away. This **intrinsic risk** is the tip of the bullet shape.
- Any efficient investor will choose a point on the efficient frontier based on their risk preference.

Note 9.2.1. general remarks on reduction of risks:

- In the **derivative pricing**, we are able to use correlation between derivative and its underlying asset to make the cash flow (of either derivative or the underlying asset) to be deterministic; that is, totally eliminate the risks.
- In the stock market with no derivatives, we can reduce the risks by: (1) buy the multiple uncorrelated to reduce volatility (Law of large number); (2) take advantages of correlations to form an efficient portfolio.

9.2.1 Two assets optimization

Lemma 9.2.2 (the efficient frontier two assets). *Consider two assets with mean return μ_1, μ_2, variance σ_1^2, σ_2^2, and correlation coefficient ρ.*

It follows that

- *the efficient frontier is* ***a parabola*** *given by*
$$\sigma^2(\mu) = x^2\sigma_1^2 + 2x(1-x)\rho\sigma_1\sigma_2 + (1-x)^2\sigma_2^2,$$
where
$$x = \frac{\mu - \mu_2}{\mu_1 - \mu_2}.$$
- *the efficient frontier is* ***an arc in a parabola*** *with short sale constraint is given by*
$$\sigma^2(\mu) = x^2\sigma_1^2 + 2x(1-x)\rho\sigma_1\sigma_2 + (1-x)^2\sigma_2^2, 0 \leq x \leq 1,$$
where
$$x = \frac{\mu - \mu_2}{\mu_1 - \mu_2}.$$

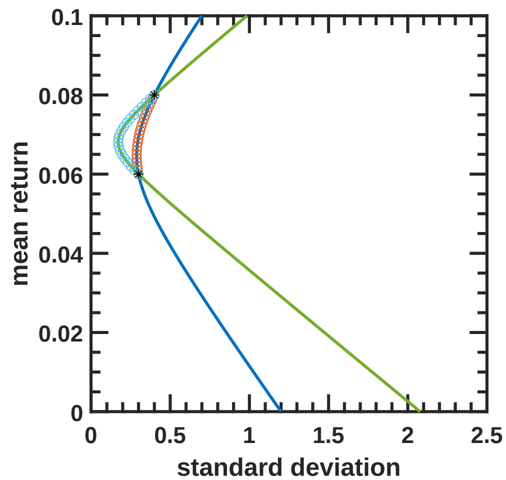
Proof. Let x denote the weight of portfolio one. Then the variance of a portfolio $(x, 1-x)$ is given by
$$\sigma^2(x) = x^2\sigma_1^2 + 2x(1-x)\rho\sigma_1\sigma_2 + (1-x)^2\sigma_2^2,$$
with the target mean constraint
$$x\mu_1 + (1-x)\mu_2 = \mu \implies x = \frac{\mu - \mu_2}{\mu_1 - \mu_2}.$$

Note that when the target mean is specified, the portfolio is determined, and no optimization is needed. □

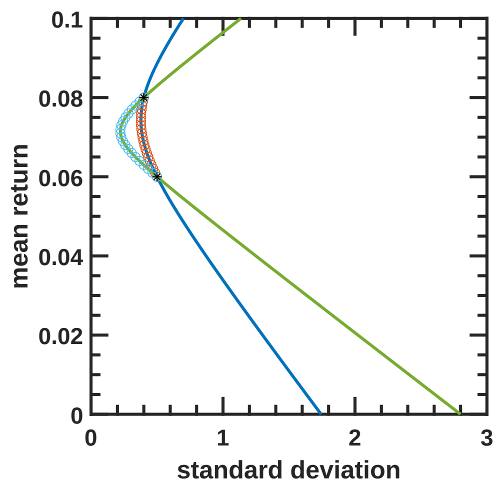
Note 9.2.2 (understanding two asset frontier). Consider two assets' efficient frontier with different parameter settings(Figure 9.2.1).

- (short sale effect) When the short sale is not allowed, the efficient frontier is a short arc; when short sale is allowed, the efficient frontier is the whole parabola.
- (influence of correlation) In the portion where short sale is not allowed, negative correlation can generate smaller standard deviation at given mean; in the portion where short sale exists, positive correlation can generate smaller standard deviation at given mean.

- Even one asset is inferior,i.e., smaller mean return given the same volatility, the inferior one can be shorted to the fund the better asset.
- When there is risk-free asset, the efficient frontier might contains two assets when short sale is allowed; however, the efficient frontier might only contain the better asset when the short sale is not allowed.



(a) Efficient frontier of two assets with parameters $\mu_1 = 0.06, \mu_2 = 0.08, \sigma_1 = 0.3, \sigma_2 = 0.4, \rho = 0.5$(green), -0.5(blue). The symbols are efficient frontier with short sale constraints. The black symbols are the two assets.



(b) Efficient frontier of two assets with parameters $\mu_1 = 0.06, \mu_2 = 0.08, \sigma_1 = 0.5, \sigma_2 = 0.4, \rho = 0.5$(green), -0.5(blue). The symbols are efficient frontier with short sale constraints. The black symbols are the two assets.

Figure 9.2.1: Two asset efficient frontier

Lemma 9.2.3 (two asset efficient frontier with risk-free asset). *Consider two assets consisting of one risky asset(mean return μ_1, variance σ_1^2) and one risk-free asset(mean return r_f).*

It follows that(see Figure 9.2.2)

- *the efficient frontier is **a line** given by*
$$\mu = \sigma(\frac{\mu_1 - r_f}{\sigma_1}) + r_f,$$
where
$$x = \frac{\mu - r_f}{\mu_1 - \mu_2}.$$
- *the efficient frontier is **an arc in a line** with short sale constraint is given by*
$$\mu = \sigma(\frac{\mu_1 - r_f}{\sigma_1}) + r_f, 0 \leq \sigma \leq \sigma_1.$$

Proof. From Lemma 9.2.2, we know that
$$\sigma(\mu) = x\sigma_1, x = \frac{\mu - r_f}{\mu_1 - r_f}.$$
□

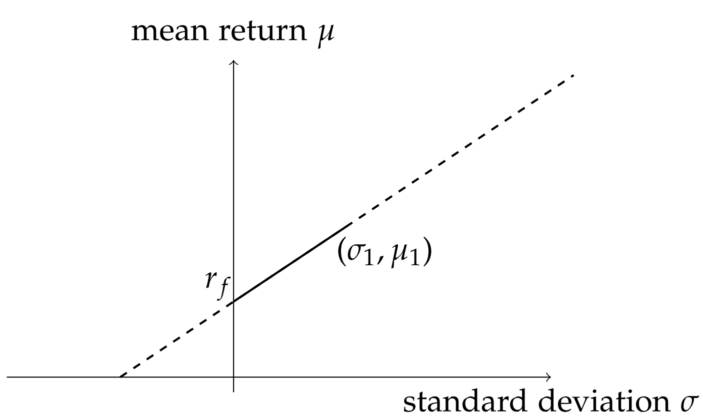


Figure 9.2.2: Efficient frontier for two assets(one risky, one risk-free). The solid line is the efficient frontier under short-sale constraint. The dashed line has no short-sale constraint.

9.2.2 Efficient frontier without risk free asset

Definition 9.2.4 (efficient portfolio, efficient frontier). *An portfolio vector is efficient if it has the minimum risk at the given return.* ***Efficient frontier*** *is the set of portfolio vectors with the maximum return at given risk.*

Theorem 9.2.3 (Two Fund theorem, affine set characterization of efficient portfolio). *[1, p. 163]*

- *The necessary and sufficient condition in Theorem 9.2.1 can be viewed as linear equation* $Ax = b$*, where* $x \in \mathbb{R}^{n+3}, A \in \mathbb{R}^{(n+2)\times(n+3)}$ *is the concatenation of* $w, \lambda, \mu, \bar{r}$*. Then the solution to* $Ax = b$ *form a 1-dimensional affine set. That is, let* x_1 *and* x_2 *be two solutions to* $Ax = b$*, then*

$$\alpha x_1 + (1-\alpha)x_2, \alpha \in \mathbb{R}$$

will also the solutions of $Ax = b$

- *Two efficient funds/portfolio can be established so that any efficient portfolio can be duplicated, in terms of mean and variance, as an affine combination of these two. In other words, all investors seeking efficient portfolios need only invest in combinations of these two funds.*

Proof. (1)See the affine set characterization of linear equations **??**.(2) is just a restatement of (1). □

Algorithm 2: Calculation of efficient fronter in the mean-variance framework

Input: A set of evenly-spaced mean values $\mu_{min} = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n = \mu_{max}$, the mean vector $\mathbb{M}$ and the convariance matrix $\mathbb{V}$.

For $k = 1, 2, ..., n$ solve the following optimization problem

$$\begin{aligned} \sigma_i &= \arg\min_x x^T V x \\ \text{s.t.} &\sum_{i=1}^{n} \mathbb{M}^T x = \mu_i \\ &\sum_i x_i = 1 \end{aligned}$$

Output: $(\mu_i, \sigma_i), i = 1, 2, ..., n$ will form the efficient frontier

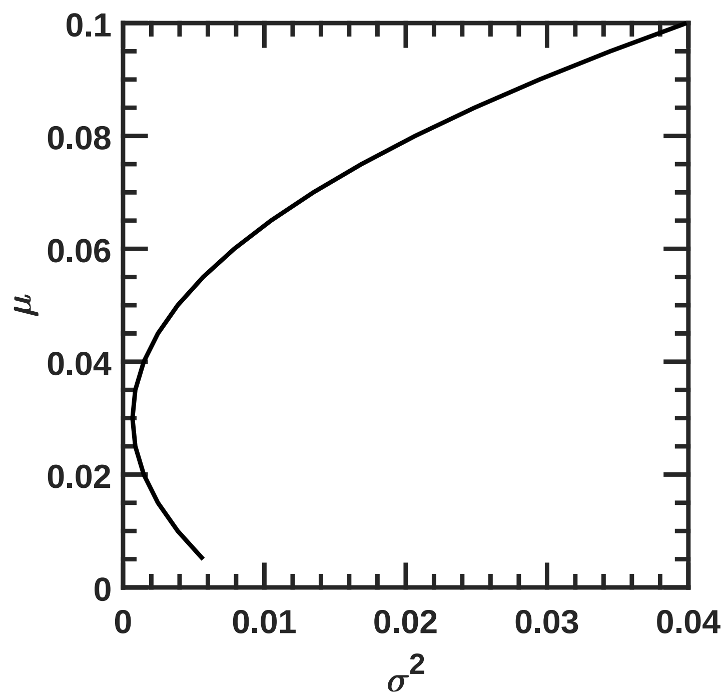


Figure 9.2.3: mean variance efficient frontier for three assets

Remark 9.2.3 (instability in practical applications). [2, p. 629]

- In practice, the means and covariance matrix are estimated from historical data and subject to random variation.
- The optimization algorithm seeks to maximumly exploit the differences in mean and covariance among assets. When these differences are statistically insignificant(i.e. representing as random variations), the resulting minimum-variance frontier are misleading and not practically useful.
- In other words, it overfits the data: it does too much with differences in mean and covariance that are actually not meaningful. In an optimization with no limitation on short sales, some assets can appear with very large negative weights.

9.2.3 Efficient frontier with risk free asset

general remarks:
When there is a risk free asset, the efficient frontier calculated using above framework to lead to a **linear efficient frontier** that intercept at r_f and tangent to the bullet shape of the risky only portfolios.

Definition 9.2.5 (Sharpe ratio). *The **Sharpe ratio** of a portfolio is the ratio of expected excess return over the volatility. The **optimal Sharpe portfolio** is a portfolio that maximizes the Sharpe ratio. The optimal Sharpe ratio can be expressed as*

$$s^* = \arg\min_x (\mu_x - r_f)/\sigma_x$$

Theorem 9.2.4 (sufficient and necessary conditions for efficient portfolio with risk-free asset). *The $n+1$ portfolio weights $w_i, i = 0, 1, ..., n$ ($i = 0$ is the risk-free asset, which has zero variance and zero correlation with other assets) and the two Lagrange multipliers λ and μ for an efficient portfolio have mean return $\bar{r}$ will satisfy*

$$\lambda r_0 + \mu = 0$$

$$\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda r_i - \mu = 0, i = 1, ..., n$$

$$\sum_{i=1}^{n} w_i r_i + w_0 r_0 = \bar{r}$$

$$\sum_{i=1}^{n} w_i + w_0 = 1$$

Proof. See the KKT condition for convex optimization **??**. □

Theorem 9.2.5 (One-Fund theorem).

- *$w_0 = 1, \bar{r} = r_0$ is one solution to equations in Theorem 9.2.4.*
- *Consider a risky portfolio with random return r and variance σ, then the affine combination of risky asset and the risk-free asset has linearity in mean return and standard deviation; that is,*

$$E[(\alpha r_0) + (1-\alpha)r] = (\alpha r_0) + (1-\alpha)E[r]$$

$$\sqrt{Var[(\alpha r_0) + (1-\alpha)r]} = (1-\alpha)\sigma = (\alpha\sigma_0) + (1-\alpha)\sigma$$

  where $\sigma_0 = 0$
- *Given a risky efficient portfolio, then any affine combination between the risk-free asset and the risky portfolio will be an efficient portfolio; moreover, all efficient portfolios will have linearity in their mean return and standard deviation.*

- *Consider a market with risk-free asset. There is a single efficient fund/portfolio such that any efficient portfolio can be constructed as an affine combination of the efficient portfolio and the risk-free asset.*

Proof. (1) It can be showed that $w_0, \bar{r} = r_0, \lambda = \mu = 0$ will satisfy the condition. (2) Direct use linearity of expectation and variance property. (3) From (2) and two-fund theorem(Theorem 9.2.3). Moreover part: note that every risky efficient portfolio can be expressed as affine combination of the risk-free asset and a specified risky portfolio. (4) This is a restatement of (3). □

9.2.4 Market portfolio and efficiency

Definition 9.2.6 (market portfolio). *Let $C_i, i = 1, 2, ..., d$, denote the market capitalization of the d assets. Then the* ***market portfolio*** *$x \in \mathbb{R}^d$ is defined as:*

$$x_i = \frac{C_i}{\sum_i C_i}$$

Theorem 9.2.6 (market portfolio is efficient). *If*

- *all investors in the market are mean-variance optimizers with the same perfect information on mean and covariance of all assets.*
- *all investors are rational such that they will invest in a portfolio as an affine combination of the risk-free asset and the market portfolio.*

Then the ***market portfolio is efficient****(that is, a portfolio has the maximum mean return with given risk).*

Proof. Suppose there are N investors, and they are investing in an arbitrary portfolio on the efficient frontier. Note that from one-fund theorem(Theorem 9.2.5), any efficient portfolio can be viewed as the affine combination of a risk-free asset and a specific same risky portfolio $s \in \mathbb{R}^n$. Suppose each investor invest in the risky portfolio with wealth w^i. Because

$$\sum_i^N w^i s_j = C_j,$$

then $s_j = \frac{C_j}{\sum_{j=1}^d C_j}, j = 1, 2, ..., n$.Therefore s must be the market portfolio. □

Remark 9.2.4 (underlying assumptions for efficient market portfolio)**.** The theorem requires the following:

- All investors have identical information about the return and volatility.
- All investors are mean-variance optimizers.
- There is an unique risk-free rate of borrowing and lending that is available to all investors.
- There is no transaction cost.

Remark 9.2.5 (equilibrium interpretation). To interpret this result from the perspective of market equilibrium, see [1, p. 175].

Corollary 9.2.6.1 (efficient frontier is capital market line). *When a risk free asset exists, the efficient frontier, also known as the* ***capital market line*** *is given as*

$$r - r_f = \frac{r_M - r_f}{\sigma_M}\sigma$$

where r_M, σ_M are the mean return and volatility of the (efficient)market portfolio.

Corollary 9.2.6.2 (market price of risk). *Assume all investors in the market are mean-variance optimizers with the same perfect information on mean and covariance. For any* ***efficient portfolio on the capital market line****, its mean return r is connected to its volatility as:*

$$r - r_f = \beta(r_M - r_f)$$

where r_M, σ_M are the mean return and volatility of the (efficient)market portfolio. The quantity

$$K = \frac{r_M - r_f}{\sigma_M}$$

is known as the ***market price of the risk****. Then we can write*

$$r = r_f + \beta\sigma_M K$$

Proof. This is a restatement of Theorem 9.2.5 and Theorem 9.2.6. □

Remark 9.2.6 (interpretation: rewards on taking risks).

- Only **efficient portfolio** will be rewarded by taking risks.
- For an arbitrary portfolio, its mean return will be discussed in Theorem 10.1.1.

9.3 Extending portfolio optimization

9.3.1 Transaction cost issue

Definition 9.3.1 (turnover). *The **turnover** is the sum of the absolute values of the differences between each positions at time t and time $t+1$; that is*

$$turnover(t) = \sum_{i=1}^{n} \left| x_i(t+1) - x_i(t) \right|,$$

where n is the number of total assets and $x_i(t)$ refers to the position in asset i at time t.

Definition 9.3.2 (booksize). *The **booksize** at time t is the sum of the absolute values of positions at time t; that is*

$$booksize(t) = \sum_{i=1}^{n} \left| x_i(t) \right|,$$

where n is the number of total assets and $x_i(t)$ refers to the position in asset i at time t.

Definition 9.3.3 (optimization problem with limited booksize and turnover).

$$\begin{aligned} \max_x \sum_i^n x_i \mu_i - \sum_{i,j=1}^{n} \lambda x_i x_j \sigma_{ij} \\ s.t. \sum_{i=1}^{n} |x_i| \leq M. \\ \sum_{i=1}^{n} \left| x_i - x_i^{old} \right| \leq L \sum_{i=1}^{n} \left| x_i^{old} \right|. \end{aligned}$$

9.3.2 Conditional value-at-risk constraint

9.3.3 Tracking error

[3].

9.3.4 Risk-based portfolio selection

9.3.4.1 *Risk measure and decomposition*

Definition 9.3.4. *We denote by $w = \{w_1, w_2, ..., w_N\}$ the portfolio weights, by σ_i^2 the variance of asset i, by σ_{ij} the covariance between asset i and asset j, and by Σ the covariance matrix. We have*

- *the standard deviation (volatility) of the portfolio is defined by*

$$\sigma_P = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}}.$$

- *the* ***marginal risk contribution*** *MRC_i of the i asset defines the sensitivity of the portfolio total risk to a change in the weight of asset i, defined by*

$$MRC_i = \frac{\partial \sigma_P}{\partial w_i} = \frac{\sum_{j=1}^{N} w_j \sigma_{ij}}{\sigma_P}.$$

- *the* ***total risk contribution of the*** *i* ***asset*** *is defined by*

$$TRC_i = w_i \cdot MRC_i.$$

- *the* ***total risk decomposition***(*Lemma 9.3.2*) *is given by*

$$\sigma_P = \sum_{j=1}^{N} TRC_j.$$

Lemma 9.3.1 (Euler's principle for homogeneous of degree one functions). *Let $f(x_1, x_2, ..., x_N)$ be a continuous, differentiable, and homogeneous of degree one function. Then it can be decomposed as the following:*

$$f(x_1, x_2, ..., x_N) = x_1 \frac{\partial f(x)}{x_1} + x_2 \frac{\partial f(x)}{x_2} + \cdots + x_N \frac{\partial f(x)}{x_N}.$$

Proof. From the definition of first-order homogeneous function, we have

$$f(tx, ty) = tf(x, y).$$

Take derivative with respect to t, we have

$$f(x, y) = \frac{\partial f}{\partial(tx)} \frac{\partial(tx)}{\partial t} + \frac{\partial f}{\partial(ty)} \frac{\partial(ty)}{\partial t} = \frac{\partial f}{\partial(tx)} x + \frac{\partial f}{\partial(ty)} y.$$

Let $t = 1$, then

$$f(x,y) = \frac{\partial f}{\partial x}x + \frac{\partial f}{\partial y}y.$$

This can be generalized to multiple variables. □

Lemma 9.3.2 (risk decomposition via Euler principle). *[4, p. 31] Let us denote by $\mathcal{R}(w_1, w_2, ..., w_N)$ the* ***first-order homogeneous risk measure*** *for the portfolio with weights $(w_1, w_2, ..., w_N)$. Define the marginal contribution to risk of asset i by*

$$MRC_i(w) = \frac{\partial \mathcal{R}(w_1, w_2, ..., w_N)}{\partial w_i},$$

with RC_i denoting the risk contribution of the i asset

$$RC_i(w_1, w_2, ..., w_N) \triangleq w_i \frac{\partial \mathcal{R}(w_1, w_2, ..., w_N)}{\partial w_i}.$$

Then the Euler decomposition yields

$$\mathcal{R}(w_1, w_2, ..., w_N) = \sum_{i=1}^{N} w_i RC_i(w_1, w_2, ..., w_N).$$

Proof. Use Euler's principle Lemma 9.3.1. □

Note 9.3.1. Note that any coherent risk measure is a first-order homogeneous function(Definition 25.1.5).

Example 9.3.1. [4, p. 185] Let us assume that the asset returns r are normally distributed as $N(\mu, \Sigma)$, and for a portfolio w we have

$$\mu(w) = w^T\mu, \sigma(w) = \sqrt{w^T\Sigma w}.$$

- If the risk measure is given by volatility risk

$$\mathcal{R}(w) = \sigma(w) = \sqrt{w^T\Sigma w},$$

  then

$$\frac{\partial R(w)}{\partial w_i} = \frac{\sum_{j=1}^{N} \Sigma_{ij} w_j}{\sigma(w)}, RC_i(w) = w_i \frac{\sum_{j=1}^{N} \Sigma_{ij} w_j}{\sigma(w)}.$$

- If the risk measure is given by standard deviation-based measure

$$\mathcal{R}(w) = -\sigma(w)^T \mu + c\sigma(w),$$

then

$$\frac{\partial R(w)}{\partial w_i} = -\mu_i + c\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)}, RC_i(w) = -w_i\mu_i + w_i c\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)}.$$

- If the risk measure is given by VaR(Lemma 25.2.3)

$$\mathcal{R}(w) = -\sigma(w)^T \mu + \Phi^{-1}(\alpha)\sigma(w),$$

then

$$\frac{\partial R(w)}{\partial w_i} = -\mu_i + \Phi^{-1}(\alpha)\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)}, RC_i(w) = -w_i\mu_i + w_i\Phi^{-1}(\alpha)\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)}.$$

- If the risk measure is given by expected shortfall(Lemma 25.2.4),

$$\mathcal{R}(w) = -\sigma(w)^T \mu + c\sigma(w),$$

then

$$\frac{\partial R(w)}{\partial w_i} = -\mu_i + c\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)}, RC_i(w) = -w_i\mu_i + w_i c\frac{\sum_{j=1}^N \Sigma_{ij} w_j}{\sigma(w)},$$

where

$$c = \frac{\exp(-y^2/2)}{\sqrt{2\pi}(1-\alpha)}, y = \Phi^{-1}(\alpha).$$

Definition 9.3.5 (risk budget portfolio). *Let us denote by $\mathcal{R}(w_1, w_2, ..., w_N)$ the* ***first-order homogeneous risk measure*** *for the portfolio with weights $(w_1, w_2, ..., w_N)$. Further denote the marginal contribution to risk and total risk contribution of asset i by MRC_i, RC_i.*

*The**risk budgeting portfolio** is then defined by the following constraints*

$$\begin{aligned} RC_1(w_1, ..., w_N) &= b_1 \\ &\cdots \\ RC_i(w_1, ..., w_N) &= b_i \\ &\cdots \\ RC_N(w_1, ..., w_N) &= b_N \end{aligned}$$

9.3.4.2 *Risk-parity portfolio*

Definition 9.3.6 (condition for a risk-parity portfolio). *Let $x = (x_1, x_2, ..., x_n)$ denote the portfolio position vector. Let $\Sigma \in \mathbb{R}^{n\times n}$ denote the covariance matrix of the return. A risk-parity portfolio is such that*

$$x_i \times \sum_{k=1}^{n} \Sigma_{ik} x_k = x_j \times \sum_{ik} \Sigma_{ik} x_k, \forall i, j.$$

Lemma 9.3.3 (scale invariance of risk-parity portfolio). *Let $(x_1, x_2, ..., x_n)$ be risk-parity portfolio vector. Let $\lambda \in \mathbb{R}^+$. Then*

$$(\lambda x_1, \lambda_2, ..., \lambda_n)$$

is also a risk-parity portfolio vector.

Proof. If

$$x_i \times \sum_{k=1}^{n} \Sigma_{ik} x_k = x_j \times \sum_{ik} \Sigma_{ik} x_k, \forall i, j,$$

then

$$\lambda x_i \times \sum_{k=1}^{n} \Sigma_{ik} \lambda x_k = \lambda x_j \times \sum_{ik} \Sigma_{ik} \lambda x_k, \forall i, j.$$

□

Lemma 9.3.4 (special case solution: two assets risk-parity portfolio). *[5] Consider two assets with volatility in returns given by σ_1, σ_2. Let ρ be the correlation between two assets returns and the portfolio position vector be $x = (w, 1 - w)$. It follows that*

- *The risk-parity portfolio requires that*

$$w^2 \sigma_1^2 = (1 - w)^2 \sigma_2^2.$$

- *The unique solution for a risk-parity portfolio is given by*

$$x^* = (\frac{\sigma_1^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}}, \frac{\sigma_2^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}}).$$

Proof. (1) The total risk contribution from asset 1 is

$$TRC_1 = \frac{w}{\sigma(x)} (w\sigma_1^2 + (1 - w)\rho\sigma_1\sigma_2),$$

where $\sigma(x) = x^T \sigma x$.

The total risk contribution from asset 2 is

$$TRC_2 = \frac{1-w}{\sigma(x)}((1-w)\sigma_2^2 + w\rho\sigma_1\sigma_2).$$

Let $TRC_1 = TRC_2$, we have

$$w^2\sigma_1^2 = (1-w)^2\sigma_2^2.$$

(2) Directly solve the equation in (1). □

Lemma 9.3.5 (special case solution: multiple asset with constant correlation). [5] *Consider n assets with volatility in returns given by $\sigma_i, i = 1, 2, ..., n$. Assume that the correlation for every pair of returns is $\rho_{ij} = \rho, \forall i, j$. Let $x = (x_1, x_2, ..., x_n)$ be the portfolio position. It follows that*

- *The risk-parity portfolio requires that*

$$x_i\sigma_i = x_j\sigma_j, \forall i, j.$$

- *The solution for a risk-parity portfolio with constraint $\sum_{i=1}^n x_i = 1$ is given by*

$$x_i^* = \frac{\sigma_i^{-1}}{\sum_{i=1}^n \sigma_i^{-1}}.$$

Lemma 9.3.6 (special case solution: multiple asset with constant volatility). [5] *Consider n assets with constant volatility in returns given by $\sigma_i = \sigma, i = 1, 2, ..., n$. Assume that the correlation for every pair of returns is $\rho_{ij}, \forall i, j$. Let $x = (x_1, x_2, ..., x_n)$ be the portfolio position. It follows that*

- *The risk-parity portfolio requires that*

$$x_i\sigma_i = x_j\sigma_j, \forall i, j.$$

- *The solution for a risk-parity portfolio with constraint $\sum_{i=1}^n x_i = 1$ is given by*

$$x_i^* = \frac{(\sum_{k=1}^n x_k\rho_{ik})^{-1}}{\sum_{j=1}^n(\sum_{k=1}^n x_k\rho_{jk})^{-1}}.$$

Methodology 9.3.1 (numerical optimization formulation for risk-parity portfolio). *[5]link*

- $$\begin{aligned} x^* &= \arg\min_x \sum_{i=1}^n \sum_{j=1}^n (x_i \sum_{k=1} \Sigma_{ik} x_k - x_j \sum_{k=1} \Sigma_{jk} x_k)^2 \\ s.t. &\quad 1^T x = 1, 0 \leq x \leq 1 \end{aligned}$$

- $$\begin{aligned} x^* &= \arg\min_x \sum_{i=1}^n (x_i \sum_{k=1} \Sigma_{ik} x_k - \frac{1}{n} x^T \Sigma x)^2 \\ s.t. &\quad 1^T x = 1, 0 \leq x \leq 1 \end{aligned}$$

9.4 Continuous-time wealth growth

Lemma 9.4.1 (logarithmic performance). *[1, p. 420]If $X_1, X_2, ...$ is the random sequence of capital values generated by the process*

$$X_k = R_k X_{k-1},$$

where $R_k, k = 1, 2, ...$ is a sequence of iid random variables, then

$$\ln(\frac{X_n}{X_0})^{1/n} \to m, in\ dist,$$

or equivalently

$$X_n \to X_0 e^{mn}, in\ dist$$

as $n \to \infty$, where

$$m = E[\ln R_1]$$

Proof. Note that

$$\frac{1}{k}\ln(X_{k+1}/X_k) + ... + \ln(X_2/X_1) = \frac{1}{k}(\ln R_k + ... + \ln R_1) \to N(m, Var[\ln R_1]/k),$$

from central limit theorem(Theorem 1.13.3). □

Example 9.4.1 (volatility pumping). [1, p. 422] Suppose there are two asset available for investment.

- One asset is a stock that in each period either doubles or reduces by half with probability 1/2.
- Another asset just retains its value.

One trading strategy is that we invest one-half of our capital in each asset each period and **rebalance** the portfolio in the beginning of each period. The growth factor R_k is a random variable such that

- $R_k = 2 \times 1/2 + 1 \times 1/2 = 3/2$ with probability 0.5.
- $R_k = \frac{1}{2} \times 1/2 + 1 \times 1/2 = 3/4$ with probability 0.5.

Therefore

$$m = E[\ln R_k] = \frac{1}{2} \times \ln 3/2 + \frac{1}{2} \times \ln 3/4 \approx 0.059.$$

That is, in the long run the total capital is given by

$$X_n = X_0 \exp(0.059n).$$

interpretation:

Lemma 9.4.2. *[1, p. 428]Suppose there are n assets governed by the SDE given by*

$$dS_i/S_i = \mu_i dt + dz_i, i = 1, 2, ..., n$$

where z_i denote Brownian motion with correlation $cov(dz_i, dz_j) = \sigma_{ij}dt$. We denote the whole covariance matrix by the S. Now consider a portfolio consists of the n assets with weight $w_i, i = 1, 2, ..., n, \sum w_i = 1$. It follows that

- *The value V of the portfolio is governed by*

$$dV/V = \sum_{i=1}^{n} w_i \mu_i dt + \sum_{i=1}^{n} w_i dz_i.$$

- *The value V has solution*

$$V(t) = V(0)\exp(\sum_{i=1}^{n} w_i \mu_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} t + \sum_{i=1}^{n} w_i z_i(t)).$$

- *The logarithm of $V(t)$ is given by*

$$\log(\frac{V(t)}{V(0)}) = \sum_{i=1}^{n} w_i \mu_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} t + \sum_{i=1}^{n} w_i z_i(t),$$

  with mean

$$E[\ln(\frac{V(t)}{V(0)})] = \sum_{i=1}^{n} w_i \mu_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} t.$$

Definition 9.4.1 (log-optimal portfolio). *An log-optimal portfolio is a portfolio with optimizer weights $w_1, w_2, ..., w_n$ such that $E[\ln(\frac{V(t)}{V(0)})]$ is maximized.*

Remark 9.4.1. The reason we maximize $E[\ln V(t)]$ instead of $E[V(t)]$ is because maximizing $E[\ln V(t)]$ is maximizing the log utility function $u(x) = \ln(x)$

Lemma 9.4.3 (log-optimal portfolio when there exists risk-free asset). *[1, p. 432] Suppose there are n assets governed by the SDE given by*

$$dS_i/S_i = \mu_i dt + dz_i, i = 1,2,...,n$$

where z_i denote Brownian motion with correlation $cov(dz_i, dz_j) = \sigma_{ij}dt$. We denote the whole covariance matrix by the S. Now consider a portfolio consists of the n assets with weight $w_i, i = 1,2,...,n$. When there is a risk-free asset with weight $w_0 = 1 - \sum w_i$, the log-optimal portfolio has weights for the risky assets that satisfy

$$\sum_{j=1}^{n} \sigma_{ij} w_j^* = \mu_i - r_f,$$

for $i = 1,2,...,n$, such that

$$(1 - \sum_{i=1}^{n} w_i) r_f + \sum_{i=1}^{n} \mu_i w_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_{ij} w_j$$

is maximized.

Proof. Directly take derivative of

$$(1 - \sum_{i=1}^{n} w_i) r_f + \sum_{i=1}^{n} \mu_i w_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_{ij} w_j$$

with respect to $w_i, i = 1,2,...,n$. □

Lemma 9.4.4 (log-optimal pricing formula). *[1, p. 435] Consider a world with n risky asset governed by geometric SDE and a risk-free asset with growth rate r_f. It follows that for each risky asset, we have*

$$\mu_i - r_f = \sigma_{i,opt},$$

where $\sigma_{i,opt}$ is the instantaneous covariance between asset i and the log-optimal portfolio.

Or equivalently,

$$\mu_i - r_f = \beta_{i,opt}(\mu_{opt} - r_f).$$

where $\beta_{i,opt} = \sigma_{i,opt}/\sigma_{opt}^2$, and

$$\sigma_{opt}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i^* \sigma_{ij} w_j^*, \mu_{opt} = \sum_{i=1}^{n} w_i^* \mu_i + (1 - \sum_{i=1}^{n} w_i^*) r_f.$$

Proof. (1)From Lemma 9.4.3, we know that

$$\sum_{j=1}^{n} \sigma_{ij} w_j^* = \mu_i - r_f.$$

Let $V(t)$ be the value of log-optimal portfolio, then

$$dV/V = \sum_{i=1}^{n} w_i^* \mu_i dt + \sum_{i=1}^{n} w_i^* dz_i.$$

and

$$\sigma_{i,opt} = E[dV/V \ dS_i/S] = E[\sum_{i=1}^{n} w_i^* dz_i dz_i] = \sum_{j=1}^{n} \sigma_{ij} w_j^*.$$

Therefore, we have

$$\mu_i - r_f = \sigma_{i,opt}.$$

(2) Note that from Lemma 9.4.3, we know that

$$\sum_{j=1}^{n} \sigma_{ij} w_j^* = \mu_i - r_f, i = 1, 2, ..., n.$$

Therefore

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_i^* \sigma_{ij} w_j^* = \sum_{i=1}^{n} w_i^* (\mu_i - r_f) = \mu_{opt} - r_f.$$

□

9.5 Corporate finance theory

9.5.1 Cost of capital

Definition 9.5.1 (cost of equity). *[6, p. 440]*

$$r_E = r_f + \beta(r_{Mkt} - r_f)$$

Definition 9.5.2 (cost of debt). *The yield to maturity of the bonds/debts is the cost of debt.*

Definition 9.5.3 (unlevered cost of capital). *[6, p. 454]*

$$r_U = \frac{E}{E+D} r_E + \frac{D}{E+D} r_D$$

Definition 9.5.4 (weighted average cost of capital). *[6, p. 461]*

$$r_{AWCC} = \frac{E}{E+D} r_E + \frac{D}{E+D} r_D (1 - t_C)$$

where t_C is the corporate tax rate

Example 9.5.1 (calculating after-tax cost of debt,136). Suppose a company is issuing 1 10Y, 5 percent semiannual coupon bond with face value of 1000\$. Upon issue, the bond sells at 1025\$.

The yield to maturity $r = 4.684$ percent, is obtained by solving the following equation

$$1025 = \sum_{t=1}^{20} \frac{25}{(1+r/2)^i} + \frac{1000}{(1+r/2)^20}.$$

If the marginal tax rate is 35 percent, then the after-tax cost of debt is

$$r(1 - 0.35) = 0.03045.$$

Remark 9.5.1 (issues in estimating cost of debt). The cost of debt estimation can be complicated in practice because the company may issue different forms of bonds like floating-rate debt, callable bond, convertible bond.

Definition 9.5.5 (preferred stock).

- *A preferred stock is a class of ownership in a corporation that has a higher claim on its assets and earnings than common stock. Preferred shares generally have a dividend that must be paid out before dividends to common shareholders, and the shares usually do not carry voting rights.*
- *The* ***cost of preferred stock*** *is the cost that a company has committed to pay preferred stockholder.*

Lemma 9.5.1 (138). *Consider a nonconvertible, noncallable preferred stock that has a fixed dividend rate and no maturity date. The cost of preferred stock is given by*

$$r_p = \frac{D_p}{P_p},$$

where

- *P_p is the current preferred stock price per share.*
- *D_p is the preferred stock dividend per share.*
- *r_p is te cost of preferred stock.*

Example 9.5.2. Consider a company. Suppose the risk-free rate in the market is 5 percent, the company's equity beta is 1.5, and the market risk premium is 7 percent. Then the cost of the common stock is given by

$$C = 5\% + 1.5 \times 7\% = 15.5\%.$$

9.5.2 Project valuation and decision

Definition 9.5.6 (internal rate of return). *[7, p. 285] The discounting rate that makes net present value equal to zero is called* ***internal rate of return*** *or* ***yield to maturity****.*

9.5.3 Optimal capital structure

9.5.3.1 *Modigliani-Miller theory*

Definition 9.5.7 (capital structure). *[6, p. 521] The relative proportions of debt, equity, and other outstanding securities of a firm is called the firm's* ***capital structure****.*

Definition 9.5.8 (perfect market condition). *[6, p. 525] The* ***perfect capital market conditions*** *are*

- *Investors and firms can trade the same set of securities at competitive market price equal to the present value of their future cash flows.*
- *There are no taxes, transaction costs, or issuance costs associated with security trading.*
- *A firm's financing decisions do not change the cash flows generated by its investments, nor do they reveal new information about them.*

Theorem 9.5.1 (Modigliani-Miller theorem). *[6, p. 525]*

- *In a perfect capital market, the total value of a firm's securities is equal to the market value of the total cash flows generated by its assets and is not affected by its choice of capital structure. Mathematically,*

$$E + D = U = A$$

where A is the market value of firm assets, E is the market value of levered equity, U is the market value of unlevered equity, and D is market value of debt.
- *The cost of capital of levered equity increases with the firm's market-valued debt-equity ratio, that is*

$$r_E = r_U + \frac{D}{E}(r_U - r_D).$$

where R_E is the return on levered equity, R_U is the return on unlevered equity, and R_D is the return on debt.

9.6 Notes on bibliography

The major reference are [1].

[8][9][10][11]

For factor models, see [12].

BIBLIOGRAPHY

1. Luenberger, D. G. *et al.* Investment science. *OUP Catalogue* (1997).
2. DeFusco, R. A., McLeavey, D. W., Pinto, J. E., Anson, M. J. & Runkle, D. E. *Quantitative investment analysis* (John Wiley & Sons, 2015).
3. Jorion, P. Portfolio optimization with tracking-error constraints. *Financial Analysts Journal* **59,** 70–82 (2003).
4. Homescu, C. *Many risks, one (optimal) portfolio* (SSRN, 2014).
5. Maillard, S., Roncalli, T. & Teïletche, J. On the properties of equally-weighted risk contributions portfolios (2008).
6. Berk, J. & DeMarzo, P. *Corporate Finance* ISBN: 9780134101415. https://books.google.com/books?id=6fnEDAAAQBAJ (Pearson Education, 2016).
7. Vernimmen, P., Quiry, P., Dallocchio, M., Le Fur, Y. & Salvi, A. *Corporate finance: theory and practice* (John Wiley & Sons, 2014).
8. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
9. Primbs, J. A. *A Factor Model Approach to Derivative Pricing* (CRC Press, 2016).
10. Cornuejols, G. & Tütüncü, R. *Optimization methods in finance* (Cambridge University Press, 2006).
11. Burmeister, E., Roll, R. & Ross, S. A. A practitioner's guide to arbitrage pricing theory (1994).
12. Abbas, N., Khan, J., Aziz, R. & Sumrani, Z. A Study to Check the Applicability of Fama and French, Three-Factor Model on KSE 100-Index from 2004-2014. *International Journal of Financial Research* **6,** 90 (2014).

10

CAPITAL ASSET PRICING MODEL

10.1 Capital Asset Pricing Model

10.1.1 Capital asset pricing model

Note 10.1.1 (recap, efficient market portfolio and implications). From the Portfolio optimization model, we can obtain an efficient portfolio M. Suppose every rational investor is holding this efficient portfolio M, then M is essentially the combination of all stocks weighted by its market capitalization, in other words, M is the market portfolio. The capital asset pricing model is linking the risk premium of the stock i to the market risk premium as

$$E(r_i) = r_f + \beta(E(r_M) - r_f)$$

where β is $cov(R_i, R_m)/\sigma_m^2$. This CAPM models says that if the investor wants to earn profit more than r_f, then he has to bear extra risk.

When we assume market portfolio is efficient, its underlying assumptions are(Theorem 9.2.6):

- All investors have identical information about the return and volatility.
- All investors are mean-variance optimizers.
- There is a unique risk-free rate of borrowing and lending that is available to all investors.
- There is no transaction cost.

Theorem 10.1.1 (Capital asset pricing model(CAPM)). *[1, p. 177] Assume* ***the market portfolio*** *M* ***is efficient****(or equivalently, the CAPM assumption holds(Theorem 9.2.6)) with expected return $E[r_M]$, the expected return $E[r_i]$ of* ***any asset*** *i satisfies*

$$E[r_i] - r_f = \beta(E(r_M) - r_f)$$

where β is $cov(R_i, R_m)/\sigma_m^2$, r is the risk-free rate.

Proof. Let r_a be the random return of a portfolio with a portion of r_i and $(1-a)$ portion of r_M, then

$$E[r_a] = aE[r_i] + (1-a)E[r_M]$$

and

$$\sigma_a = \sqrt{a^2\sigma_i^2 + 2a(1-a)\sigma_{iM} + (1-a)^2\sigma_M^2}$$

We require

$$\frac{dE[r_a]}{\sigma_a}|_{a=0} = \frac{E[r_M] - r_f}{\sigma_M}$$

where the right-hand side is the price of risk(Corollary 9.2.6.2). Eventually, we get

$$E[r_i] - r_f = \frac{cov(R_i, R_m)}{\sigma_M^2}(E(r_M) - r_f)$$

□

Remark 10.1.1 (interpretation via investment behavior)**.**

- In the CAPM world, all investors are rational such that they seek maximum mean return with given risks.
- If an asset is lying above the efficient frontier(the asset is undervalued [1]), investors will tend to buy it and push up its price(or lower its return).
- Similarly, if an asset is lying below the efficient frontier(the asset is overvalued), investors will tend to short it and push down its price(or increase its return).

Remark 10.1.2 (interpretations on β)**.**

- $\beta = 0$: If the asset is uncorrelated, then its mean return is equal to risk-free rate, **no matter how large its volatility σ_i will be**. This is because its volatility can be diversified away using other uncorrelated assets, and **the market will not reward any diversifiable risks.**
- $\beta < 0$: Such asset can be used to reduce the market portfolio, and therefore investors are willing to accept its lower expected return than the risk-free rate. **A put option in-the-money has negative β, the put option will even charge money from the buyer.**
- $\beta = \sigma_i/\sigma_M$: If the asset i is efficient (which consists of $(1-\alpha)$ risk-free asset and α M), then $\sigma_i = \alpha\sigma_M$, $Cov(r_i, r_M) = \alpha\sigma_M^2$, and $\beta = \sigma_i/\sigma_M$. Eventually, we recover the

  $$E[r_i] - r_f = \frac{E[r_M] - r_f}{\sigma_M}\sigma_i$$

  in Corollary 9.2.6.2.

Remark 10.1.3 (The market only reward nondiversifiable risks)**.** [1, p. 181]

- The CAPM **implies** that the random return r_i of asset i is related to the random return of r_M as

  $$r_i = r_f + \beta(r_M - r_f) + \epsilon_i$$

  where $E[\epsilon_i] = 0, cov(\epsilon_i, r_M) = 0$. Then

  $$\sigma_i^2 = \beta^2\sigma_M^2 + var[\epsilon_i]$$

1 for the relationship between return and its price behavior, see the equity valuation section(section 12.3)

or

$$\beta^2 = \frac{\sigma_i^2 - Var[\epsilon_i]}{\sigma_M^2} < \frac{\sigma_i^2}{\sigma_M^2}$$

where β is less the optimal β. The volatility due to ϵ_i is **nonsystematic risk** that can be diversified away.

- If any asset carries nonsystematic risk, they will not fall onto the efficient frontier(the capital market line).

10.1.2 CAPM as a pricing tool

Theorem 10.1.2 (CAPM pricing). *[1, p. 187] Suppose assumptions of CAPM hold. Suppose the* ***random*** *payoff from an investment in a single period is X. Let P be the current price, then the return is $r_X = \frac{X}{P} - 1$. The β of X is given by*

$$\beta_X = \frac{Cov(r_X, r_m)}{\sigma_m^2} = \frac{1}{P}\frac{Cov(X, r_m)}{\sigma_m^2}$$

and its price P is given as

$$P = \frac{E[X]}{1 + r_f + \beta(E[r_m] - r_f)}.$$

In particular,

- *if X is deterministic payoff(such that $\beta = 0$),*
$$P = \frac{X}{1 + r_f}.$$
- *if X is the market portfolio such that $\beta = 1, E[X] = 1 + E[r_m]$, then $P = 1$.*

Proof. From Theorem 10.1.1,we have

$$E[X]/P - 1 - r_f = E[r_X - r_f] = \beta(E[r_m] - r_f)$$

then solve P and will get the result. □

Remark 10.1.4 (interpretation). We can view the discount factor $1 + r_f + \beta(E[r_m] - r_f)$ as the risk adjusted discount factor of $(1 + r_f)$.

Remark 10.1.5 (comparisons between pricing alternatives).

- One alternative method, known as **discount expectation pricing**, for pricing an asset with payoff X is
$$P = \frac{E[X]}{1+r_f}$$
where the price is usually higher than CAPM pricing(Theorem 10.1.2). This is because in CAPM pricing, we are using the correlations with other asset to reduce nonsystematic risk, thus we should get more return $r = X/P - 1$ or lower price. From arbitrage point of view, if the price is higher than the CAPM price, we can use portfolio optimization method to construct a new portfolio with the same expected return yet lower price.
- For different pricing alternatives, see [1, p. 253].

Lemma 10.1.1 (equivalent pricing formula and linearity of CAPM pricing). *[1, p. 189] Suppose assumptions of CAPM hold. Suppose the* ***random*** *payoff from an investment in a single period is X. Let P be the current price, then the return is* $r_X = \frac{X}{P} - 1$. *Its price P is given as*

$$P = \frac{1}{1+r_f}[E[X] - \frac{cov(X, r_m)(E[r_m - r_f])}{\sigma_m^2}].$$

Moreover, P is linear on X.

Proof. (1) Note that $P = \frac{E[X]}{1+r_f+\beta(E[r_m]-r_f)}$. Then use $\beta_X = \frac{Cov(r_X,r_m)}{\sigma_m^2} = \frac{Cov(X/P-1,r_m)}{\sigma_m^2} = \frac{Cov(X/P,r_m)}{\sigma_m^2}$ (2) Note that $E[X]$ and $cov(X, r_m)$ are linear in X. □

10.1.3 Empirical test of CAPM model

Remark 10.1.6 (failure of CAPM).

- Under CAPM assumption(Theorem 9.2.6), the market portfolio is efficient, and the line passing the risk-free rate and the (σ_M, μ_M) should be the efficient frontier(see Figure 10.1.1).
- The CAPM results shows that all assets should lie in the efficient frontier; however in reality it is barely the case. The Figure 10.1.1 shows that portfolio/asset A is more efficient(generate more mean return at the same risk) than the market portfolio; and B is less efficient than the market portfolio.
- The CAPM result does not hold is because the CAPM assumption does not completely hold.
- If the security's expected return versus risk is plotted above the efficient frontier, it is undervalued since the investor can expect a greater return for the inherent risk.

And a security plotted below the efficient frontier, is overvalued since the investor would be accepting less return for the amount of risk assumed.

Remark 10.1.7 (taking advantage of CAPM for active portfolio management)**.** When assets do not follow the CAPM prediction

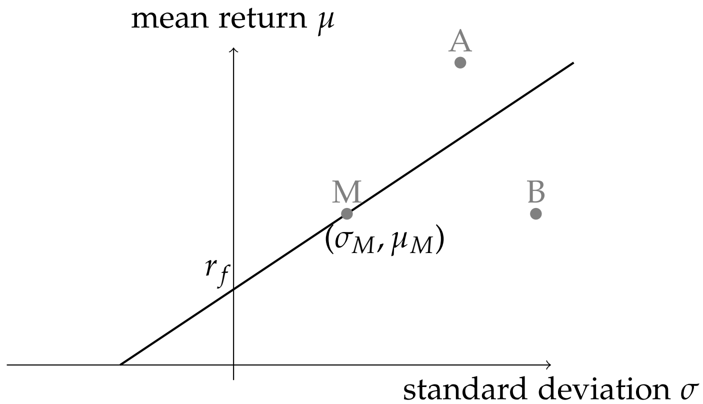


Figure 10.1.1: Failure of CAPM. The figure shows the asset/portfolio A is above the efficient frontier and asset/portfolio B is below the efficient frontier.

10.1.4 The Single-Factor/Index Model

Definition 10.1.1 (Single-Factor model). *In the single-factor model, we model the random return r_i of stock i as as*

$$r_i = E(r_i) + \beta_i m + e_i, i = 1, 2..., n$$

where m, a random variable, is the macroeconomic factor measures the unanticipated macro surprises, and it has $E(m) = 0$ and σ_m; e_i measures the firm-specific surprise, and it has $E(e_i) = 0, \sigma(e_i)$, and $Cov(e_i, e_j) = 0$; β_i characterize the sensitivity of i firm to the macro economic factor.

Remark 10.1.8 (drawback of CAPM)**.** One drawback of the Markkowitz portfolio optimization model is the number of parameters needed to be estimated is a large number.

Remark 10.1.9 (interpretation)**.**

- Then the covariance between stocks are

$$cov(r_i, r_j) = \beta_i \beta_j \sigma_m^2.$$

The variance for stock i is

$$\sigma_i^2 = \beta_i^2 \sigma_m + \sigma^2(e_i).$$

Example 10.1.1 (**Special case: market index to be the factor**). When we use the market index to be the single factor, we can re-formulate the return of the stock i as:

$$R_i = \alpha_i + \beta_i R_M + e_i$$

where $R_i = r_i - r_f, R_M = r_M - r_f$.

- Note that here α is not $E(R_i)$ and **the CAPM implies that $\alpha_i = 0$.**
- The value α_i and β_i can be directly estimated from market data.
- If the estimated $\alpha_i > 0$, this means that the expected return of the stock is more than equilibrium value(the equilibrium value is $\beta_i R_M$ from CAPM pricing theorem(Theorem 10.1.2)), or equivalently, the stock is underpriced(since if the price goes higher, the return will decrease to equilibrium value).
- If the estimated $\alpha_i > 0$, we can long the stock since it is underpriced. Then in the statistical sense, it will get excess return taken the level risks taken (outperform the market) .

  The interpretation of β_i is:

- $0 \leq \beta_i < 1$: a beta less than 1 indicates that the investment is less volatile than the market.
- $\beta_i > 1$: a beta more than 1 indicates that the investment is more volatile than the market.

10.2 Notes on bibliography

The major reference are [1].

[2][3][4][5]

For factor models, see [6].

BIBLIOGRAPHY

1. Luenberger, D. G. *et al.* Investment science. *OUP Catalogue* (1997).
2. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
3. Primbs, J. A. *A Factor Model Approach to Derivative Pricing* (CRC Press, 2016).
4. Cornuejols, G. & Tütüncü, R. *Optimization methods in finance* (Cambridge University Press, 2006).
5. Burmeister, E., Roll, R. & Ross, S. A. A practitioner's guide to arbitrage pricing theory (1994).
6. Abbas, N., Khan, J., Aziz, R. & Sumrani, Z. A Study to Check the Applicability of Fama and French, Three-Factor Model on KSE 100-Index from 2004-2014. *International Journal of Financial Research* **6,** 90 (2014).

11

FACTOR MODELS AND ARBITRAGE PRICING THEORY

11.1 Arbitrage pricing theory

11.1.1 Arbitrage pricing theory for single assets

Lemma 11.1.1 (single factor arbitrage theory). *[1, p. 208] Suppose there are 2 assets and the rate of return(**random**) satisfies*

$$r_1 = a_1 + b_1 f$$
$$r_2 = a_2 + b_2 f$$

*where f is a random variable. Under **no-arbitrage** condition, the coefficient a_1, b_1, a_2, b_2 should satisfy*

$$a_1 = r_f + \lambda b_1$$
$$a_2 = r - f + \lambda b_2$$

where r_f is the risk-free rate, and λ is a constant. Then we have

$$r_1 = r_f + \lambda b_1 + b_1 f$$
$$r_2 = r_f + \lambda b_2 + b_2 f.$$

Proof. See the single factor multiple stocks case Lemma 11.1.2. □

Note 11.1.1 (interpretation). Depending on the parameter b_1 and b_2, we have two situations.

- (hedgeable case) If $b_1, b_2 \neq 0$, then we can construct a risk-free portfolio
$$k = (k_1, k_2)^T = (-b_2, b_1)^T,$$
such that
$$k^T r = (b_1 - b_2) r_f + (\lambda + f)(-b_1 b_2 + b_1 b_2) = k^T f_f \mathbf{1},$$
with risk-free growth rate. This is consistent with the no-arbitrage principle that risk free portfolio should grow at risk-free rate.
- (non-hedgeable case) If $b_1 \neq 0, b_2 \neq 0$, then we cannot construct a risk-free portfolio. However, the asset dynamics is given by
$$r_1 = r_f + \lambda b_1 + b_1 f$$
$$r_2 = r_f.$$

is obeying the arbitrage principle such that (a) risk-free asset grows at risk-free rate;(2) risky asset should grow at a rate faster than risk-free rate.

Lemma 11.1.2 (single factor multiple stocks arbitrage theory). *Suppose are a total n assets and the rate of return on asset i satisfies*

$$r_i = a_i + b_i f, \forall i = 1, 2, ..., n$$

*where f is a random variable. Under no-arbitrage condition, then there must exist a constant $\lambda \in \mathbb{R}$, also known as **market price of risk**, such that the coefficient $a_i, b_i, \forall i = 1, ..., n$ will satisfy*

$$a_i = r_f + \lambda b_i, \forall i = 1, ..., n,$$

where r_f is the risk-free rate.

Proof. Denote $A = (a_1, ..., a_n)^T$, $B = (b_1, ..., b_n)^T$, and **any** portfolio vector $k = (k_1, ..., k_n)^T$. To construct a risk-free portfolio, we have

$$k^T B = 0,$$

such that

$$r = k^T(A + Bf) = k^T A.$$

This risk-free portfolio should grow at risk-free rate r_f, we have

$$k^T A = k^T r_f.$$

Combine $k^T B = 0, k^T(A - r_f) = 0$, we know that $A - r_f = B\lambda$. (from $k^T B = 0$, we know that $k \in \mathcal{N}(B^T)$; from the fact that $k^T(A - r_f) = 0$ and k is arbitrage, we know that $(A - r_f) \in \mathcal{N}(B^T)^\perp = \mathcal{R}(B)$, see fundamental theorem of linear algebra(**??**) for more details.). □

Lemma 11.1.3 (multiple factors multiple stocks arbitrage theory). *[1, p. 209][2, p. 51] Suppose there are a total n assets and the rate of return on asset i satisfies*

$$r_i = a_i + \sum_{j=1}^{m} b_{ij} f_j, \forall i = 1, 2, ..., n$$

where $m < n$ and $f_1, ..., f_m$ are random variables. Under no-arbitrage condition, then there must exist a set of constants $\lambda_j \in \mathbb{R}, \forall j = 1, ..., m$, such that the coefficient $a_i, b_i, \forall i = 1, ..., n$ satisfy

$$a_i = r_f + \sum_{i=1}^{m} b_{ij} \lambda_j, \forall i = 1, ..., n,$$

where r_f is the risk-free rate.

Proof. Denote $A = (a_1, ..., a_n)^T$, $B_{ij} = b_{ij}$, and a portfolio vector $k = (k_1, ..., k_n)^T$. To construct a risk-free portfolio, we have

$$k^T B = 0,$$

such that

$$r = k^T(A + Bf) = k^T A.$$

This risk-free portfolio should grow at risk-free rate r_f, we have

$$k^T A = k^T r_f.$$

Combine $k^T B = 0, k^T(A - r_f) = 0$, we know that $A - r_f = B\lambda$. (from $k^T B = 0$, we know that $k \in \mathcal{N}(B^T)$; from the fact that $k^T(A - r_f) = 0$ and k is arbitrage, we know that $(A - r_f) \in \mathcal{N}(B^T)^\perp = \mathcal{R}(B)$, see fundamental theorem of linear algebra(**??**) for more details.); therefore

$$A - r_f = B\lambda, \lambda \in \mathbb{R}^m.$$

□

Note 11.1.2 (on the properties of the factors)**.**

- Note that we do not require the factors to have zero mean.
- If we have zero mean factors, i.e., $E[f_i] = 0$, then we can interpret λ_i as a risk-premium associated with per unit exposure to factor i. For example, consider an asset has exposure only to factor 1, that is,

  $$r_i = r_f + b_{i1}(\lambda_1 + f_1).$$

  Then we have

  $$E[r_i] = r_f + b_{i1}\lambda_1$$
- Also we do not require the factors to be independent.

11.1.2 Arbitrage pricing theory for well-diversified portfolios

Definition 11.1.1 (APT assumptions). *[3, p. 637]*

- *A factor model can exactly describe asset returns.*
- *There are many assets, so investors can form well-diversified portfolios that eliminate asset-specific risk; in other words, the market price of risk for these asset-specific risk is zero.*

- *No arbitrage opportunities exist among well-diversified portfolios; that is, the market is in equilibrium.*

Theorem 11.1.1 (arbitrage pricing theory(APT)). *[4][5] Assume the APT assumptions hold such that a factor model can exactly describe the $n \gg 1$ assets' return via*

$$r_i = E[r_i] + \sum_{j=1}^{m} \beta_{ij} f_j + \epsilon_i, \forall i = 1, 2, ..., n.$$

where $m < n$ and $f_1, ..., f_m$ are ***zero mean*** *random variables, and ϵ_i are independent random variables independent of both $f_i, i = 1, ..., m$ and $\epsilon_j, j \neq i$.* ***Assume the market is free of arbitrage.*** *It follows that*

- *for a single asset, there exists constants $\lambda_1, \lambda_2, ..., \lambda_m, \lambda_\epsilon^{(i)}$ such that each asset is given by*

  $$r_i = r_f + \sum_{i=1}^{m} \beta_{ij}(\lambda_j + f_j) + \epsilon_i + \lambda_\epsilon^{(i)},$$

  where r_f is the risk-free rate, λ_i is the risk-premium for factor i; or the market price of risk associated with factor i. Note that usually $\lambda_\epsilon^{(i)} = 0$ for diversifiable risks(Lemma 15.5.5). In other words

  $$E[r_i] = r_f + \sum_{i=1}^{m} \beta_{ij} \lambda_j.$$

- *For a* ***well-diversified portfolio****, its expected return is given by*

  $$E[r_P] = r_f + \lambda_1 \beta_{P,1} + \cdots + \lambda_K \beta_{P,K},$$

  such that the portfolio's return is given by

  $$r_P = r_f + \sum_{i=1}^{m} \beta_{Pj}(\lambda_j + f_j).$$

Proof. (1) Directly from Lemma 11.1.3. (2) For a well-diversified portfolio, there is no idiosyncratic risk. In other words, a well-diversified portfolio can be viewed as an asset with return given by

$$r_P = E[r_P] + \sum_{j=1}^{m} \beta_{ij} f_j,$$

then we use Lemma 11.1.3 again. □

Lemma 11.1.4 (theoretical risk premium for internal factors in APT). *Assume the APT assumptions hold such that a factor model can exactly describe the $n \gg 1$ assets' return via*

$$r_i = E[r_i] + \sum_{j=1}^{m} \beta_{ij}(f_j - E[f_j]) + \epsilon_i, \forall i = 1,2,...,n.$$

*where $m < n$ and $f_1,...,f_m$ are random variables **internally** constructed via linear combinations such that*

$$f_i = \sum_{j=1}^{n} w_j^{(i)} r_j, \sum_j w_j^{(i)} = 1, i = 1,2,...,m,$$

and ϵ_i are random variables independent of both $f_i, i = 1,...,m$ and $\epsilon_j, j \neq i$.

It follows that

- *Let $W = [w^{(1)}, w^{(2)}, ..., w^{(m)}] \in \mathbb{R}^{n\times m}$. Then the coefficient matrix $B \in \mathbb{R}^{n\times m}$ is **uniquely** determined by*
$$B = W(W^TW)^{-1}$$
such that $B^TW = I_{m\times m}$ or, $[w^{(i)}]^T B = e_i^T$.
- *the risk premium associated with factor $f_j - E[f_j]$ is $E[f_j] - r_f$, j=1,2,...,m, such that*
$$r_i = E[r_i] + \sum_{j=1}^{m} \beta_{ij}(f_j - r_f) + \epsilon_i, \forall i = 1,2,...,n.$$

Proof. (1) In vector form, we have

$$r - E[r] = B(f - E[f]).$$

Multiply both sides by W^T, we have

$$W^T(r - E[r]) = W^T B(f - E[f]).$$

The left hand side is $f - E[f]$. Therefore, we have

$$f - E[f] = W^T B(f - E[f]) \implies (W^T B - I_m)(f - E[f]) = 0.$$

Because $f - E[f]$ is random variable, to we must have $W^T B = I_m$ to ensure the equality holds for all possible realizations of $f - E[f]$.

Using the pseudoinverse of a full rank matrix(**??**), we have

$$B^T = (W^T W)^{-1} W^T \Leftrightarrow B = W(W^T W)^{-1}.$$

(2) From APT Theorem 11.1.1, we know that

$$\begin{aligned} r - r_f &= B(\lambda + f - E[f]) \\ [w^{(i)}]^T(r - r_f) &= [w^{(i)}]^T B(\lambda + f - E[f]) \\ f_i - r_f &= \lambda_i + f_i - E[f_i] \\ &\implies \lambda_i = E[f_i] - r_f \end{aligned}$$

where we use the factor that $\sum_j w_j^{(i)} = 1$ □

Example 11.1.1 (revisit CAPM). In CAPM model(Theorem 11.1.1), the expected return $E[r_i]$ of any asset i satisfies

$$E[r_i] - r_f = \beta(E(r_M) - r_f)$$

where β is $cov(R_i, R_m)/\sigma_m^2$, r is the risk-free rate.

In the APT framework(Lemma 11.1.4), we interpret $(E(r_M) - r_f)$ as the risk premium, and we can write the asset return as

$$\begin{aligned} r_i &= E[r_i] + \beta(r_M - E[r_M]) + \epsilon_i \\ r_i &= r_f + \beta(r_M - r_f) + \epsilon_i. \end{aligned}$$

Lemma 11.1.5 (solving factor premium parameter from market data). *Assume the APT assumptions hold such that a factor model can exactly describe well-diversified portfolios' return via*

$$r_P = E[r_P] + \sum_{j=1}^{m} \beta_{ij} f_j,$$

where $f_1, ..., f_m$ are ***zero mean*** *random variables,*

$$E[r_P] = r_f + \sum_{i=1}^{m} \beta_{ij}\lambda_j,$$

r_f is the risk-free rate, and λ_i is the risk-premium for factor i.

Suppose we have observations of expected return and sensitivities β of n portfolios; that is,

$$\begin{bmatrix} E[r_1] \\ E[r_2] \\ \cdots \\ E[r_n] \end{bmatrix} = \begin{bmatrix} r_f \\ r_f \\ \cdots \\ r_f \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \cdots & \cdots & \cdots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nm} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_m \end{bmatrix},$$

or simply

$$R = R_f + B\lambda.$$

We have the following situation

- *If the linear system $R = R_f + B\lambda$ is consistent, $n = m$ and B is full row rank*[a]*, then*

$$\lambda = B^{-1}(R - R_f).$$

- *If the linear system $R = R_f + B\lambda$ is consistent but under-determined(i.e., $n < m$), then there are infinitely many solutions of λ forming a subspace.*
- *If the linear system $R = R_f + B\lambda$ is inconsistent and $m \leq n$, then there are arbitrage opportunities. An arbitrage portfolio given by will have no risk but grows at a rate faster than risk-free rate.*

a If the system is over-determined, we can simply eliminate redundant rows

Remark 11.1.1.

Example 11.1.2 (determine factor premium from data using APT). [3, p. 641] Suppose that APT assumptions hold. Suppose that a two factors, surprise in inflation and surprise in GDP growth, can explain the asset returns. Suppose we have the following portfolio parameters

portfolio	expected return	sensitivity to inflation factor	sensitivity to GDP factor
J	0.14	1.0	1.5
K	0.12	0.5	1.0
L	0.11	1.3	1.1

The APT theory gives that

$$E[R_P] = R_F + \lambda_1 \beta_{P,INF} + \lambda_2 \beta_{P,GDP}.$$

Therefore, we can form the following linear system

$$\begin{aligned} E[R_J] &= 0.14 = R_F + 1.0\lambda_1 + 1.5\lambda_2 \\ E[R_K] &= 0.12 = R_F + 0.5\lambda_1 + 1.0\lambda_2 \\ E[R_L] &= 0.11 = R_F + 1.3\lambda_1 + 1.1\lambda_2, \end{aligned}$$

from which we can uniquely solve

$$R_F = 0.07, \lambda_1 = -0.02, \lambda_2 = 0.06.$$

Example 11.1.3 (checking no arbitrage condition). [3, p. 639] Suppose we are have **well-diversified** portfolios A, B, C that can be **exactly** described by a single factor model,

$$r_i = E[r_i] + \beta_i f, i = A, B, C,$$

where $E[f] = 0$ and

$$E[r_i] = r_f + \beta_i \lambda.$$

Suppose we have observed the following expected return and factor sensitivity associated with each portfolio

portfolio	expected return $E[r_i]$	factor sensitivity β_i
A	0.075	0.5
B	0.150	2.0
C	0.070	0.4

It follows that we can solve $\lambda = 0.05, r_f = 0.05$, such that the expected return can be describe consistently via

$$E[r_i] = r_f + \lambda \beta_i = 0.05 + 0.05\beta_i.$$

Therefore, the expected returns obey the no-arbitrage principle.

Example 11.1.4 (exploiting arbitrage opportunities). [3, p. 639] Suppose a one-factor model can describe the returns of four portfolios. Suppose we have observed returns and sensitives given in the following table.

portfolio	expected return	factor sensitivity
A	0.075	0.5
B	0.15	2.0
C	0.07	0.4
D	0.08	0.45
0.5C+0.5D	0.0725	0.45

Note that the linear system

$$\begin{bmatrix} 0.075 \\ 0.15 \\ 0.07 \\ 0.08 \end{bmatrix} = + \begin{bmatrix} 1 & 0.5 \\ 1 & 2 \\ 1 & 0.4 \\ 1 & 0.45 \end{bmatrix} \begin{bmatrix} r_f \\ \lambda_1 \end{bmatrix}$$

is inconsistent.

We can long D and short $0.5A + 0.5C$ to create a arbitrage portfolio, as showed in the following table. Note that our strategy has **expected return rate** of 0.0075, but with **zero risk** in our one factor framework.

portfolio	expected return	initial cash flow	final cash flow	factor sensitivity
D	0.08	-10000	10800	0.45
0.5A+0.5C	0.0725	+10000	-10725	-0.45
sum	0.0075	0	75	0

11.2 Factor models

11.2.1 General remarks and factors

- In the factor model, we assume that the correlation between any two assets is explained by systematic components/factors, one can restrict attention to only K (non-diversifiable) factors.
- Advantages:
  - Drastically reduces number of input variables.
  - simplify the systematic risk parameter estimation.
- (Drawbacks):
  - Purely statistical model (no theory).
  - parameter estimation relies on past data and assumes stationarity.

  Models expected returns (priced risk) Allows to estimate systematic risk (even if it is not priced, i.e. uncorrelated with SDF) Analysts can specialize along factors

Remark 11.2.1 (how to find factors)**.** Consider asset returns X_1, X_2, X_3(they are random variables). We can describe their relationship by introducing a common factorFigure 11.2.1.

- This common factor can be external, such that GDP, inflation, etc.
- This common factor can also be internal, such as linear combination of X_1, X_2, X_3; such internal factor can be obtained via PCA or factor modeling.

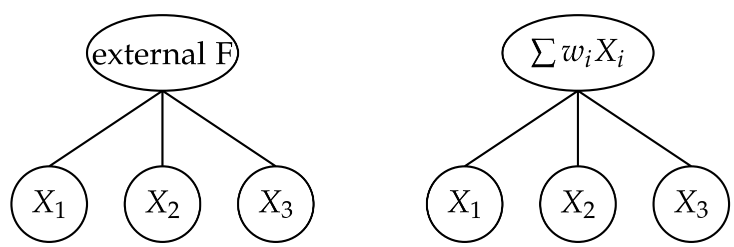


Figure 11.2.1: Factor model concept. (a) introducing relation using external factors; (b) introducing relation using internal factors.

Definition 11.2.1 (momentum factor). *[6, p. 207] The momentum of stock i measured at the end of month t, denoted by $Mon_{i,t}$, is calculated via*

$$Mon_{i,t} = (\prod_{m\in\{t-11:t-1\}})(R_{i,m}+1) - 1,$$

where $R_{i,m}$ represents the return of stock i in month m.

11.2.2 Single-factor Model

Remark 11.2.2 (motivation for single-factor model). One drawback of the Markowitz portfolio optimization model of N is that the number of parameters needed to be estimated is about N^2.

Definition 11.2.2 (single-factor model). *Suppose there are n assets. We assume the rate of return is given by*

$$r_i = a_i + b_i f + e_i, \forall i = 1, 2, ..., n$$

where r_i, f, e_i are random variables, and a_i, b_i are constants. We further assume $E[e_i] = 0, E[e_i e_j] = 0, \forall i \neq j$ and $E[(f - E[f])e_i] = 0, \forall i$.

Lemma 11.2.1 (single factor model with market portfolio excess return as the factor). *[7, p. 14]Let $r_P(t)$ denote the* ***random excess return*** *of a portfolio and r_M the random excess return of the market portfolio at time t. Assume the following linear model,*

$$r_P(t) = \alpha_P + \beta_P r_M(t) + \epsilon_P(t),$$

where we require $E[\epsilon] = 0$, $\alpha_P \in \mathbb{R}$ and $\alpha_P + \beta_P r_M$ be the best(in terms of expected residual square) linear predictor of r_P. It follows that

- $\alpha_P = \mu_P - \beta_P \mu_M$.
- $$\beta_P = \frac{Cov(r_P, r_M)}{Var[r_M]}$$
- *uncorrelation between residual and factors:*

$$Cov(r_M, \epsilon_P) = 0.$$

Proof. See Theorem 1.9.2. □

Remark 11.2.3 (model fitting and implicit assumption). Assuming the excess return process is stationary, we can use collected sample points $r_M(t), r_P(t), t = 1, 2, ..., T$ to do a simple linear regression to determine the coefficient α_P and β_P.

Note 11.2.1 (choice of the factor and connection to CAPM). [7, p. 16]

- One approximate method is to decompose the return of stock i, a random variable, as

  $$r_i = E(r_i) + \beta_i m + e_i$$

  where m, a random variable, is the macroeconomic factor measures the unanticipated macro surprises, and it has $E(m) = 0$ and σ_m; e_i measures the firm-specific surprise, and it has $E(e_i) = 0, \sigma(e_i)$, and e_i, e_j independent to each other; β_i characterize the sensitivity of i firm to the macro economy. Then the covariance between stocks are

$$cov(r_i, r_j) = \beta_i \beta_j \sigma_m^2$$

The variance for stock i is

$$\sigma_i^2 = \beta_i^2 \sigma_m + \sigma^2(e_i)$$

When we use the market index to approximate the market index, we can reformulate the return of the stock i as:

$$R_i = \alpha_i + \beta R_M + e_i$$

where $R_i = r_i - r_f, R_M = r_M - r_f$. Note that here α is not $E(R_i)$, but is the risk premium when the market premium is zero (in the case of $\beta_i = 0$ via hedging) The variance of the stock via the Single-Index model is: $\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma^2(e_i)$.

- **Note that CAPM implies that the residual $\alpha_P = 0$ for any portfolio**; our single factor model only implies $\alpha_P = 0$ for the market portfolio.

Remark 11.2.4 (parameters for mean-variance analysis). In the mean-variance analysis, we require the mean and covariance matrix for the n assets. They can be related to the single-factor model parameters as:

$$\begin{aligned} E[r_i] &= a_i + b_i E[f] \\ \sigma_i^2 &= b_i^2 \sigma_f^2 + \sigma_{e_i}^2 \\ \sigma_{ij} &= b_i b_j \sigma_f^2, i \neq j \\ b_i &= cov(r_i, f)/\sigma_f^2 \end{aligned}$$

Remark 11.2.5 (data preparation for linear regression).

11.2.3 Multiple-factor models

Definition 11.2.3 (multiple-factor model). *In the multiple-factor model, we assume the return for stock i is given by:*

$$r_i(t) = a_i(t) + \sum_{j=1}^{m} b_{i,j}(t) f_j(t) + e_i(t), \forall i = 1, 2, ..., n$$

where r_i, f_i, e_i are random variables, and a_i, b_{ij} are constants. We further assume $E[e_i] = 0, E[e_i e_j] = 0, \forall i \neq j$ and $E[(f_i - E[f])e_i] = 0, \forall i$.

Remark 11.2.6 (parameters for mean-variance analysis). In the mean-variance analysis, we require the mean and covariance matrix for the n assets. They can be related to the single-factor model parameters as:

$$E[r_i] = a_i + \sum_{j=1}^{m} b_{ij} E[f_j]$$

$$\sigma_i^2 = Var[r_i] = \sum_{j=1}^{m} b_{ij}^2 \sigma_{f_j}^2 + \sum_{k<j}^{m} 2 b_{ik} b_{ij} cov(f_k, f_j) + \sigma_{e_i}^2$$

$$\sigma_{ij} = Cov(r_i, r_j) = \sum_{k=1}^{m} \sum_{p=1}^{m} b_{ik} b_{jp} cov(f_k, f_p), i \neq j$$

Remark 11.2.7 (relation to arbitrage pricing theory).

Definition 11.2.4 (statistical factor models). *[8, p. 248]*

Remark 11.2.8 (interpretation of factor loading matrix). [7, p. 55]

- For the factor loading matrix b_{ij} characterizes the **exposure** of asset i to the factor j, estimated at time t.
- For industry categorical factor, the value b_{ij} is usually either 0 or 1, indicating whether the stock belongs to that industry or not.
- For other common factors, the exposures are standardized so that the exposure over all stocks is 0, and the standard deviation across stock is 1.

Remark 11.2.9 (interpretation of factors and idiosyncratic return). [7, p. 56]

- $f_i(t)$ is the **factor return** due to factor i during the period from time t and time $t+1$.
- For example, if we want factor i represents the return of a stock for being in an industry, then f_i is the weighted average return of stocks in the industry and the factor loading matrix element b_{ij} only takes usually either 0 or 1, indicating whether the stock belongs to that industry or not.
- $\epsilon_i(t)$ is the **specific return** during during the period from time t and time $t+1$ that cannot be explained by the model.

Remark 11.2.10 (common factors). [8, p. 240]

- **external macroeconomic factors**: gross domestic product (GDP), consumer price index (CPI), unemployment rate, credit spreads on bonds, and the steepness of the yield curve, etc.
- **fundamental factors for firms**: price-earnings ratio, the dividend-payout ratio, the earnings growth forecast, and financial leverage, etc.

- **extracted statistical model**: return on the market portfolio, the average of the returns of stocks in a particular industry,etc.

Remark 11.2.11 (industry factors vs. risk indices(non-industry factors)). [7, pp. 60, 62]

- Industry factors are created by first partitioning stocks into different nonoverlapping group. Each group contains a reasonable number of companies(with reasonable fraction of capitalization) in that industry. The f_i is the weighted average return of stocks in the industry and the factor loading matrix element b_{ij} only takes usually either 0 or 1, indicating whether the stock belongs to that industry or not;
- risk indices includes:
  - Volatility: distinguishes stocks by their volatilities.
  - Momentum: distinguishes stocks by recent performance.
  - Size: Distinguish large stocks from small stocks.
  - Liquidity: Distinguishes stocks by how much their shares trade.
  - Growth: Distinguishes stocks by past and anticipated earnings growth.
  - Value: distinguishes stocks by their fundamentals, in particular, ratios of earnings, dividends, cash flows, book value, sales, etc.
  - Earnings volatility: Distinguishes stocks by their earnings volatility.
  - Financial leverage: Distinguishes firms by debt-to-equity ratio and exposure to interest-rate risk.

Example 11.2.1 (Barra industry factor model). Suppose the m assets separate into K industry groups. For each asset i, define the factor loadings

$$\beta_{i,k} = \begin{cases} 1, \text{if asset i is in industry group K} \\ 0, otherwise. \end{cases}.$$

Further we denote the K factors via $f = (f_1, f_2, ..., f_K)$. Then the industry factor model is given by

$$X_i = \sum_{j=1}^{K} \beta_{i,j} f_j, i = 1, 2, ..., m.$$

Remark 11.2.12 (interpretation of factor). [7, p. 74]

- Note that we can write the component of the factor vector b via

$$b_k = \sum_{n=1}^{N} c_{k,n} r_n.$$

- We can interpret each factor return b_k as the return to a portfolio, called **factor portfolios**, with portfolio weights $c_{k,n}$.

- The factor portfolio can be interpreted as **factor mimicking portfolio** because it mimics the behavior of some underlying basic factor.

Example 11.2.2 (a macroeconomic factor model). [3, p. 635] Consider a factor model in which the returns of stocks are correlated with surprises in interest rates and surprises in GDP growth. For stock $i = 1, 2, ..., N$, the return is modeled by

$$R_i = a_i + b_{i1}F_{INT} + b_{i2}F_{GDP} + \epsilon_i,$$

where

- R_i is the return of stock i
- a_i is the expected return to stock i
- b_{i1} is the sensitivity of the return to stock i to interest rate surprise
- b_{i2} is the sensitivity of the return to stock i to GDP growth surprise
- F_{INT} is the surprise in interest rate
- F_{GDP} is the surprise in GDP
- ϵ_i an error term with a zero mean that represents the portion of the return to stock i not explained by the factor model

Note that we define **surprise** in general as the actual value minus the predicted(or exptected) value. For example,

$$actual\ inflation = \ predicted\ inflation\ +\ inflation\ surprise.$$

11.2.4 Statistical factor model

11.2.5 Factor model parameter estimation

11.2.5.1 *Parameter estimation with observable factors*

Remark 11.2.13 (cross-panel parameter fitting). [7, p. 74] Consider a model of return

$$r = Xb + u,$$

where r is an N vector of excess returns, X is an N by K matrix of factor exposures, b is a K dimension vector of factor returns, and u is an N dimensional vector of specific returns, and we further assume the factor loading matrix X is given.

Then we can obtain factor vector b from the weighted least square minimization given by

$$\min_b (Xb - r)^T \Delta^{-1} (Xb - r).$$

The final result is from (**??**):

$$b = (X^T \Delta^{-1} X)^{-1} X^T \Delta^{-1} r.$$

11.2.5.2 *Parameter estimation with hidden factors*

Example 11.2.3 (Barra industry factor model).

- Suppose the m assets separate into K industry groups. For each asset i, define the factor loadings

$$\beta_{i,k} = \begin{cases} 1, \text{if asset i is in industry group K} \\ 0, otherwise. \end{cases}.$$

  Further we denote the K factors via $f_t = (f_{1,t}, f_{2,t}, ..., f_{K,t})$. Then the industry factor model is given by

$$X_i = \sum_{j=1}^{K} \beta_{i,j} f_j + \epsilon_{i,t}, i = 1, 2, ..., m,$$

  where

$$Var[\epsilon_{i,t}] = \sigma_i^2, \forall i Cov(\epsilon_{i,t}, f_{k,t}) \qquad = 0, \forall i, k, t Cov(f_t, f_t) = \Omega, \forall t.$$

- For each time period t, consider the cross-sectional regression for the factor model

$$X_t = B f_t + \epsilon_t,$$

  with

$$X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \cdots \\ X_{m,t} \end{bmatrix}, f_t = \begin{bmatrix} f_{1,t} \\ f_{2,t} \\ \cdots \\ f_{m,t} \end{bmatrix}.$$

  And we can estimate

$$\hat{f}_t = (B^T B)^{-1} B^T X_t$$

  For

11.2.6 Case study: the Fama-French 3 factor model

11.2.6.1 *The model*

Remark 11.2.14 (motivations for three factor model).

- It has been observed that Fama and French observed that two classes of stocks tended to outperform the market as a whole: small cap companies and companies with high book-to-market ratio.
- In the single factor(market portfolio return) model, these out-performance cannot be explained.

Definition 11.2.5 (Fama-French 3 factor model). *[9, p. 69][10][11] In the Fama-French 3 factor model, the asset i return is given by*

$$r_i = E[r_i] + \beta_{i1}(R_m - E[R_m]) + \beta_{i2}(SMB - E[SMB]) + \beta_{i3}(HML - E[HML]) + \epsilon_i$$

where

- *$R_m - r_f$ is the return on a market value-weighted index in excess of the one-month T-bill rate.*
- *SMB ='small [market capitalization] minus big' factor. Mathematically, SMB is the average* ***raw***[a] *return on three small-cap portfolios minus the average return on three large-cap portfolio. When small stocks do well relative to large stocks this will be positive, and when they do worse than large stocks, this will be negative. SMB also refers to as* ***size premium***. *Note that $E[SMB] \neq 0$.*
- *HML = 'high [book/price] minus low' factor. Mathematically, HML is the average* ***raw*** *return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios. HML also refers to a* ***value premium***. *$E[HML] \neq 0$.*
- *Arbitrage pricing theorem gives the factor premium(Lemma 11.1.4) such that*

  $$E[r_i] = r_f + \beta_{i1}(E[R_M] - r_f) + \beta_{i2}(E[SMB] - r_f) + \beta_{i3}(E[HML] - r_f).$$

  Equivalently, we can write

  $$r_i = r_f + \beta_{i1}(R_m - r_f) + \beta_{i2}(SMB - r_f) + \beta_{i3}(HML - r_f) + \epsilon_i.$$

[a] when say here raw return to emphasize that it is not excess return

Remark 11.2.15 (interpretation). [6, pp. 147, 146, 175]

- **(market capitalization)** The market capitalization of a stock is defined as the total value of all of the outstanding shares of that stock.
- **(value stock)**, are loosely defined as stocks with low prices relative to earnings, dividends, or book value of equity.

- **(Growth stocks)**, are loosely defined as stocks with high prices relative to earnings, dividends, or book value of equity.
- **(size effect)**, refers to the observation that stocks with large market capitalizations(large stocks) tend to have lower returns than stocks with small market capitalizations(small stocks).

Remark 11.2.16 (factor portfolios as proxy risk drivers). [9, p. 70]

- The fundamental reasons determines a stock's return is the company's management system, technology, the ability to adaptive, efficiency etc. However, these reasons/factors are hard to quantify and observe. Therefore, we use the performance of different companies as the proxy to these factors.
- Te Fama-French model views the size and value factors as representing ('proxying for') a set of underlying risk factors. For example, small market-cap companies may be subject to risk factors such as less ready access to private and public credit markets and competitive disadvantages. High book-to-market may represent shares with depressed prices because of exposure to financial distress. The model views the return premiums to small size and value as compensation for bearing types of systematic risk.
- Fama and French create a portfolio designed to have returns that mimic the returns associated with the size effect and propose using the returns of this portfolio as a risk factor.

Remark 11.2.17 (interpretation of book-to-market ratio).

- The book-to-market ratio attempts to identify undervalued or overvalued securities by taking the book value and dividing it by market value. It helps to determine the market value of a company relative to its actual worth. Investors and analysts use this comparison ratio to differentiate between the true value of a publicly traded company and investor speculation.
- In basic terms, if the ratio is above 1 then the stock is undervalued; if it is less than 1, the stock is overvalued. A ratio above 1 indicates that the stock price of a company is trading for less than the worth of its assets. A high ratio is preferred by value managers who interpret it to mean that the company is a value stock, that is, it is trading cheaply in the market compared to its book value.
- A book-to-market ratio below 1 implies that investors are willing to pay more for a company than its net assets are worth. This could indicate that the company has healthy future profit projections and the investors are willing to pay a premium for that possibility. Technology companies, and other companies in industries which do not have a lot of physical assets, tend to have a low book-to-market ratio.
- It has been found that high book-to-market companies usually have higher returns than the low book-to-market companies[12, p. 2].

11.2.6.2 *The method and the data*

Remark 11.2.18 (typical statistics on factors)**.** Table 11.2.1 shows the monthly excess return statistics of Fama-french 3 factor portfolios. We can see that

- SMB and HML portfolios have positive premium, which agree with the size and value premium we discussed.
- HML has the largest Sharpe ratio than the market portfolio and the SMB. Therefore, simply investing in HML factor portfolio can out-perform the market in the long run.

Table 11.2.1: statistics on Fama-French 3 factors from July 1963 to Dec. 1991.

factor name	**mean**	**std**	**correlation**		
			$r_M - r_f$	$SMB - r_f$	$HML - r_f$
$r_M - r_f$	0.43	4.54	1		
$SMB - r_f$	0.27	2.89	-0.38	1	
$HML - r_f$	0.40	2.54	0.34	-0.08	1

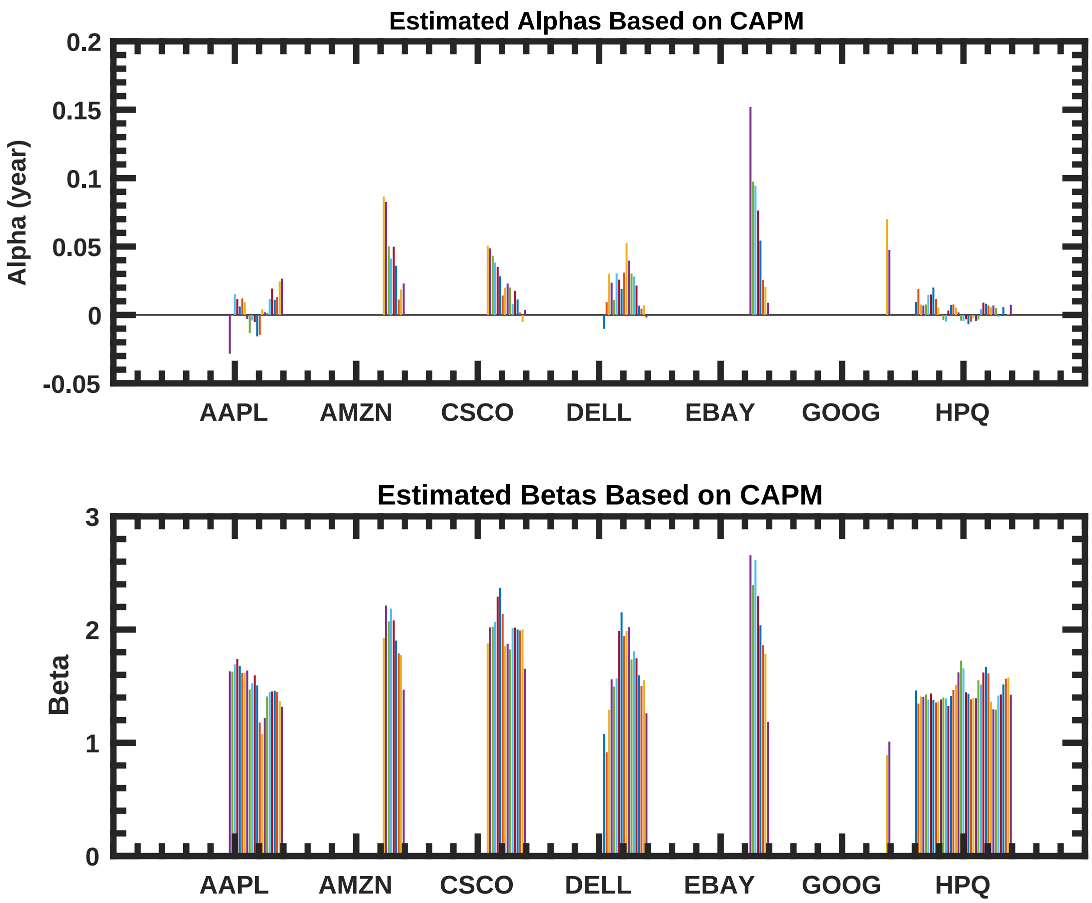

Estimated Alphas Based on CAPM
Alpha (year)
0.2
0.15
0.1
0.05
0
-0.05
AAPL
AMZN
CSCO
DELL
EBAY
GOOG
HPQ
Estimated Betas Based on CAPM
Beta
3
2
1
0
AAPL
AMZN
CSCO
DELL
EBAY
GOOG
HPQ

Figure 11.2.2

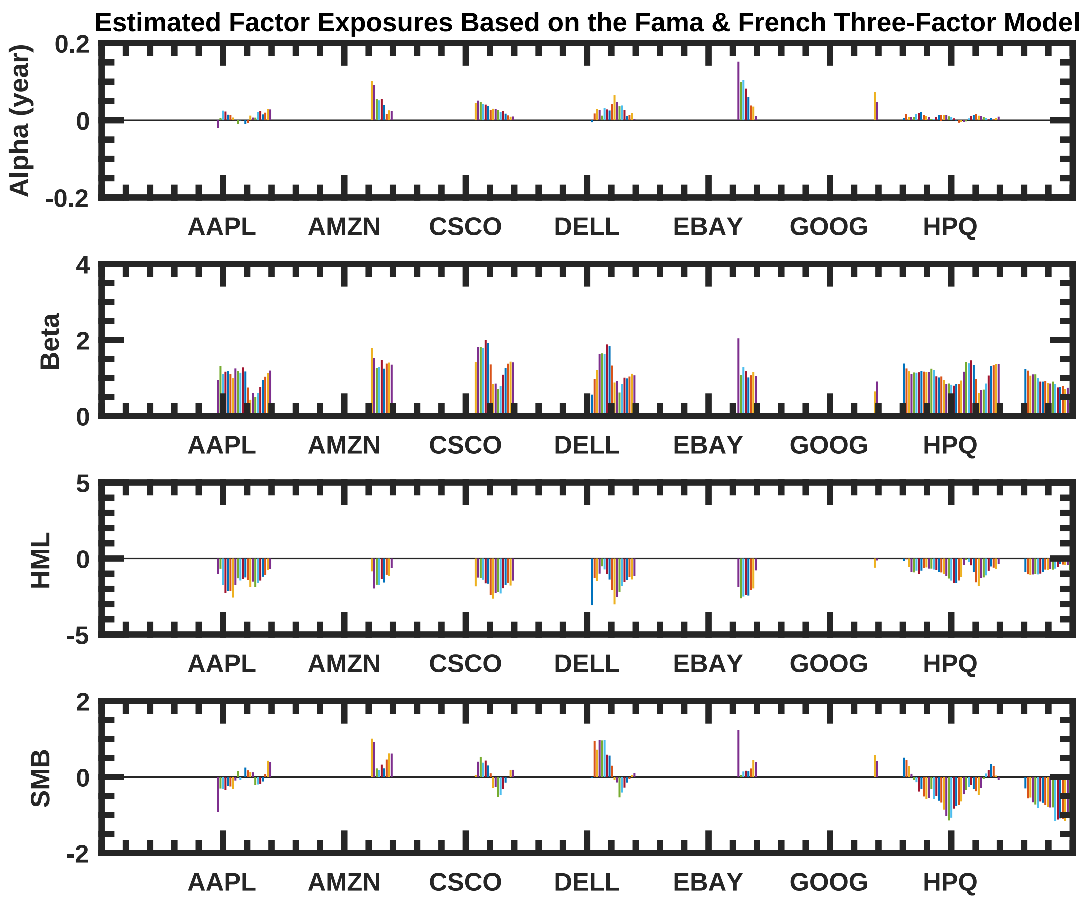
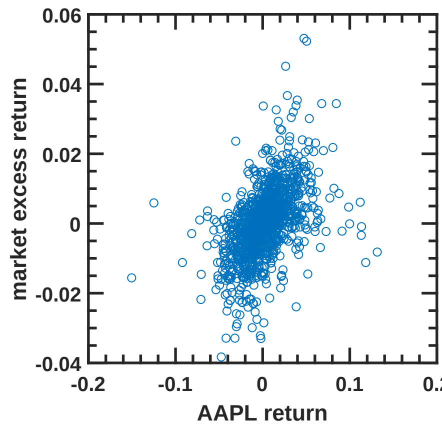

Estimated Factor Exposures Based on the Fama & French Three-Factor Model
Alpha (year)
0.2
0
-0.2
AAPL
AMZN
CSCO
DELL
EBAY
GOOG
HPQ
Beta
4
2
0
HML
5
0
-5
SMB
2
0
-2

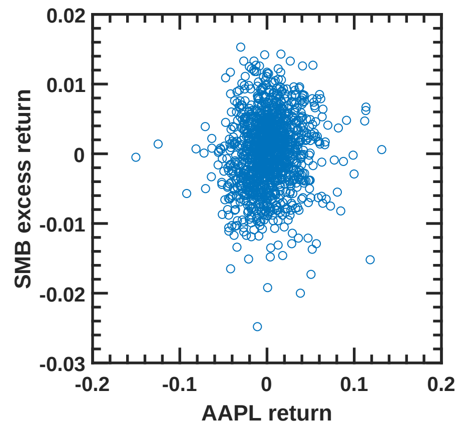
Figure 11.2.3

11.2.6.3 *Results*

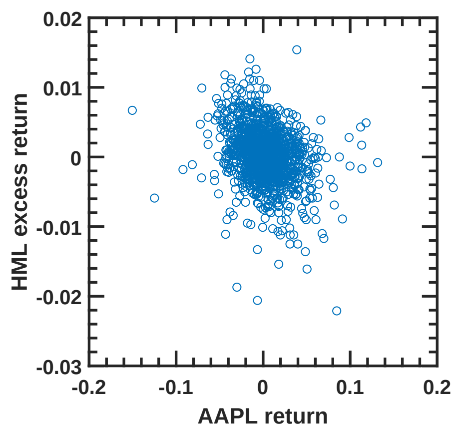
11.2.7 Multiple-factor model zoo



(a) Scatter plot for AAPL daily return vs. market excess daily return from 2001-Oct to 2006-Oct.



(b) Scatter plot for AAPL daily return vs. SMB factor excess daily return from 2001-Oct to 2006-Oct.



(c) Scatter plot for AAPL daily return vs. HML factor excess daily return from 2001-Oct to 2006-Oct.

Figure 11.2.4: Scatter plot of AAPL return vs. market excess return, SMB excess return and HML excess return.

11.2.7.1 *Multiple-factor model: The Fama-French extended 4 factor model*

Table 11.2.2: My caption

	estimate	**SE**	**tStat**	**pValue**
α	0.00186	0.00065	2.877	0.0041
β_{MKT}	1.2058	0.07012	17.196	1.4e-59
β_{SMB}	0.3753	0.12183	3.0806	0.00211
β_{HML}	-0.6583	0.17566	-3.7475	0.000187
R^2	0.264	adjusted R^2	0.263	

Table 11.2.3: My caption

	estimate	**SE**	**tStat**	**pValue**
α	0.00170	0.00065	2.6329	0.0086
β_{MKT}	1.3174	0.065	20.246	6e-79
R^2	0.249	adjusted R^2	0.249	

Table 11.2.4: My caption

	estimate	SE	tStat	pValue
α	0.000546	0.0002280	2.399	0.0168
β_{MKT}	0.943	0.041	23.017	9.6e-82
β_{SMB}	-0.092	0.062	-1.4878	0.1374
β_{HML}	-0.857	0.078	-11.034	1.27e-25
R^2	0.651	adjusted R^2	0.649	

Definition 11.2.6 (the extended Fama-French 4 factor model). *[9, p. 73][13] In the extended Fama-French 4 factor model, the asset i return is given by*

$$r_i = E[r_i] + \beta_{i1}(R_m - E[R_m]) + \beta_{i2}(SMB - E[SMB]) + \beta_{i3}(HML - E[HML]) + \beta_{i,4}(LIQ - E[LIQ]) + \epsilon$$

where

- *$R_m - r_f$ is the return on a market value-weighted index in excess of the one-month T-bill rate.*
- *SMB ='small [market capitalization] minus big' factor. Mathematically, SMB is the average **raw**[a] return on three small-cap portfolios minus the average return on three large-cap portfolio. When small stocks do well relative to large stocks this will be positive, and when they do worse than large stocks, this will be negative. SMB also refers to as **size premium**. Note that $E[SMB] \neq 0$.*
- *HML = 'high [book/price] minus low' factor. Mathematically, HML is the average **raw** return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios. HML also refers to a **value premium**. $E[HML] \neq 0$.*
- *LiQ representing the returns to a portfolio that invests the proceeds from shorting high-liquidity stocks in a portfolio of low-liquidity stocks.*
- *Arbitrage pricing theorem gives the factor premium(Lemma 11.1.4) such that*

  $$E[r_i] = r_f + \beta_{i1}(E[R_M] - r_f) + \beta_{i2}(E[SMB] - r_f) + \beta_{i3}(E[HML] - r_f) + \beta_{i4}(E[LIQ] - r_f).$$

  Equivalently, we can write

$$r_i = r_f + \beta_{i1}(R_m - r_f) + \beta_{i2}(SMB - r_f) + \beta_{i3}(HML - r_f) + \beta_{i4}(LIQ - r_f) + \epsilon_i.$$

[a] when say here raw return to emphasize that it is not excess return

Remark 11.2.19 (motivations for liquidity). [9, p. 72]

- It has been observed that investors demand a return premium for assets that are relatively illiquidâĂŤassets that cannot be quickly sold in quantity without high explicit or implicit transaction costs.
- The risk premium is to compensate for the liquidity risk.

Example 11.2.4. [9, p. 73] Consider a common stock has the following characteristics:

Factors	sensitivities
market beta	1.50
size beta	0.15
value beta	-0.52
liquidity beta	0.20

We can deduce that this appears to be a small-cap(0.15 size beta) and have a growth orientation(negative value beta). It also has a positive liquidity beta, this is typically because small-cap stocks usually trade in less liquid markets than do large-cap stocks.

11.2.7.2 *Carhart four-factor model*

Definition 11.2.7 (the Carhart 4 factor model). *[14] In the extended Fama-French 4 factor model, the asset i return is given by*

$$r_i = E[r_i] + \beta_{i1}(R_m - E[R_m]) + \beta_{i2}(SMB - E[SMB]) + \beta_{i3}(HML - E[HML]) + \beta_{i,4}(UMD - E[UMD])$$

where

- $R_m - r_f$ *is the return on a market value-weighted index in excess of the one-month T-bill rate.*
- *SMB ='small [market capitalization] minus big' factor. Mathematically, SMB is the average* ***raw***[a] *return on three small-cap portfolios minus the average return on three large-cap portfolio. When small stocks do well relative to large stocks this will be positive, and when they do worse than large stocks, this will be negative. SMB also refers to as* ***size premium***. *Note that* $E[SMB] \neq 0$.
- *HML = 'high [book/price] minus low' factor. Mathematically, HML is the average* ***raw*** *return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios. HML also refers to a* ***value premium***. $E[HML] \neq 0$.
- *UMD representing the returns to a portfolio that long previous 12-month return winners and short previous 12-month loser stocks.*
- *Arbitrage pricing theorem gives the factor premium(Lemma 11.1.4) such that*

$$E[r_i] = r_f + \beta_{i1}(E[R_M] - r_f) + \beta_{i2}(E[SMB] - r_f) + \beta_{i3}(E[HML] - r_f) + \beta_{i4}(E[UMD] - r_f).$$

Equivalently, we can write

$$r_i = r_f + \beta_{i1}(R_m - r_f) + \beta_{i2}(SMB - r_f) + \beta_{i3}(HML - r_f) + \beta_{i4}(UMD - r_f) + \epsilon_i$$

a when say here raw return to emphasize that it is not excess return

Remark 11.2.20 (motivations for liquidity).
- It has been observed that stocks return has correlation to the past; winners/losers tend to be the winners/losers in the future.
- The factor portfolio UMD is design to capture such effect.

11.2.7.3 *Multiple-factor model: The Fama-French 5 factor model*

11.3 Corporate finance theory

11.3.1 Cost of capital

Definition 11.3.1 (cost of equity). *[15, p. 440]*

$$r_E = r_f + \beta(r_{Mkt} - r_f)$$

Definition 11.3.2 (cost of debt). *The yield to maturity of the bonds/debts is the cost of debt.*

Definition 11.3.3 (unlevered cost of capital). *[15, p. 454]*

$$r_U = \frac{E}{E+D} r_E + \frac{D}{E+D} r_D$$

Definition 11.3.4 (weighted average cost of capital). *[15, p. 461]*

$$r_{AWCC} = \frac{E}{E+D} r_E + \frac{D}{E+D} r_D (1 - t_C)$$

where t_C is the corporate tax rate

Example 11.3.1 (calculating after-tax cost of debt,136). Suppose a company is issuing 1 10Y, 5 percent semiannual coupon bond with face value of 1000$. Upon issue, the bond sells at 1025$.

The yield to maturity $r = 4.684$ percent, is obtained by solving the following equation

$$1025 = \sum_{t=1}^{20} \frac{25}{(1+r/2)^i} + \frac{1000}{(1+r/2)^2 0}.$$

If the marginal tax rate is 35 percent, then the after-tax cost of debt is

$$r(1-0.35) = 0.03045.$$

Remark 11.3.1 (issues in estimating cost of debt). The cost of debt estimation can be complicated in practice because the company may issue different forms of bonds like floating-rate debt, callable bond, convertible bond.

Definition 11.3.5 (preferred stock).

- *A preferred stock is a class of ownership in a corporation that has a higher claim on its assets and earnings than common stock. Preferred shares generally have a dividend that must be paid out before dividends to common shareholders, and the shares usually do not carry voting rights.*
- *The **cost of preferred stock** is the cost that a company has committed to pay preferred stockholder.*

Lemma 11.3.1 (138). *Consider a nonconvertible, noncallable preferred stock that has a fixed dividend rate and no maturity date. The cost of preferred stock is given by*

$$r_p = \frac{D_p}{P_p},$$

where

- *P_p is the current preferred stock price per share.*
- *D_p is the preferred stock dividend per share.*
- *r_p is te cost of preferred stock.*

Example 11.3.2. Consider a company. Suppose the risk-free rate in the market is 5 percent, the company's equity beta is 1.5, and the market risk premium is 7 percent. Then the cost of the common stock is given by

$$C = 5\% + 1.5 \times 7\% = 15.5\%.$$

11.3.2 Project valuation and decision

Definition 11.3.6 (internal rate of return). *[16, p. 285] The discounting rate that makes net present value equal to zero is called* ***internal rate of return*** *or* ***yield to maturity****.*

11.3.3 Optimal capital structure

11.3.3.1 *Modigliani-Miller theory*

Definition 11.3.7 (capital structure). *[15, p. 521] The relative proportions of debt, equity, and other outstanding securities of a firm is called the firm's* ***capital structure****.*

Definition 11.3.8 (perfect market condition). *[15, p. 525] The* ***perfect capital market conditions*** *are*

- *Investors and firms can trade the same set of securities at competitive market price equal to the present value of their future cash flows.*
- *There are no taxes, transaction costs, or issuance costs associated with security trading.*
- *A firm's financing decisions do not change the cash flows generated by its investments, nor do they reveal new information about them.*

Theorem 11.3.1 (Modigliani-Miller theorem). *[15, p. 525]*

- *In a perfect capital market, the total value of a firm's securities is equal to the market value of the total cash flows generated by its assets and is not affected by its choice of capital structure. Mathematically,*

$$E + D = U = A$$

where A is the market value of firm assets, E is the market value of levered equity, U is the market value of unlevered equity, and D is market value of debt.

- *The cost of capital of levered equity increases with the firm's market-valued debt-equity ratio, that is*

$$r_E = r_U + \frac{D}{E}(r_U - r_D).$$

where R_E is the return on levered equity, R_U is the return on unlevered equity, and R_D is the return on debt.

11.4 Notes on bibliography

The major reference are [1].

[17][2][18][4]

For factor models, see [19].

BIBLIOGRAPHY

1. Luenberger, D. G. *et al.* Investment science. *OUP Catalogue* (1997).
2. Primbs, J. A. *A Factor Model Approach to Derivative Pricing* (CRC Press, 2016).
3. DeFusco, R. A., McLeavey, D. W., Pinto, J. E., Anson, M. J. & Runkle, D. E. *Quantitative investment analysis* (John Wiley & Sons, 2015).
4. Burmeister, E., Roll, R. & Ross, S. A. A practitioner's guide to arbitrage pricing theory (1994).
5. Ross, S. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13,** 341–360 (1976).
6. Bali, T., Engle, R. & Murray, S. *Empirical Asset Pricing: The Cross Section of Stock Returns* ISBN: 9781118095041. https://books.google.com/books?id=svKlCgAAQBAJ (Wiley, 2016).
7. Grinold, R. & Kahn, R. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk* ISBN: 9780071376952. https://books.google.com/books?id=aSDkFQtysy8C (McGraw-Hill Education, 1999).
8. Fabozzi, F. & Pachamanova, D. *Portfolio Construction and Analytics* ISBN: 9781119238164. https://books.google.com/books?id=0OTFCwAAQBAJ (Wiley, 2016).
9. Henry, E., Robinson, T. R., Stowe, J. D., *et al. Equity asset valuation* (John Wiley & Sons, 2010).
10. Fama, E. F. & French, K. R. The cross-section of expected stock returns. *the Journal of Finance* **47,** 427–465 (1992).
11. Fama, E. F. & French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of financial economics* **33,** 3–56 (1993).
12. Fama, E. F. & French, K. R. A five-factor asset pricing model. *Journal of Financial Economics* **116,** 1–22. ISSN: 1556-5068 (2014).
13. Pástor, L. & Stambaugh, R. F. Liquidity risk and expected stock returns. *Journal of Political economy* **111,** 642–685 (2003).
14. Carhart, M. M. On persistence in mutual fund performance. *The Journal of finance* **52,** 57–82 (1997).
15. Berk, J. & DeMarzo, P. *Corporate Finance* ISBN: 9780134101415. https://books.google.com/books?id=6fnEDAAAQBAJ (Pearson Education, 2016).

16. Vernimmen, P., Quiry, P., Dallocchio, M., Le Fur, Y. & Salvi, A. *Corporate finance: theory and practice* (John Wiley & Sons, 2014).
17. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
18. Cornuejols, G. & Tütüncü, R. *Optimization methods in finance* (Cambridge University Press, 2006).
19. Abbas, N., Khan, J., Aziz, R. & Sumrani, Z. A Study to Check the Applicability of Fama and French, Three-Factor Model on KSE 100-Index from 2004-2014. *International Journal of Financial Research* **6,** 90 (2014).

12

MARKET STATISTICAL MODELING AND TRADING STRATEGIES

12.1 Market modeling

12.1.1 Market stylized facts

Definition 12.1.1. *[1, p. 10]*

- ***clustering of volatility***: *large price changes tend to be followed by large price changes and small price changes tend to be followed by small price changes.*
- ***autoregressive behavior***: *price changes depend on price changes in the past, e.g., positive price changes tend to be followed by positive price changes.*
- ***negative skew***: *there is an asymmetry in the upside and downside potential of price changes. Large declines are more common than large rallies.*
- ***fat tails***: *probability of extreme profits or losses is much larger than predicted by the normal distribution. We should note that tail thickness varies from asset to asset.*
- ***temporal behavior of tail thickness***: *probability of extreme profits or losses can change through time; it is smaller in regular markets and much larger in turbulent markets.*
- ***tail thickness varies across frequencies***: *high-frequency data tends to be more fat-tailed than lower-frequency data.*

Definition 12.1.2. *[1, p. 11] for univariate financial time series:*

- *Time series data of returns, in particular daily return series, are in general not independent and identically distributed (i.i.d.).;*
- *The volatility of return processes is not constant with respect to time;*
- *The absolute or squared returns are highly autocorrelated;*
- *The distribution of financial market returns is leptokurtic, and thus occurrence of extreme events is (much) more likely compared to the normal distribution;*
- *Extreme returns are observed closely in time (volatility clustering).*
- *The empirical distribution of returns is skewed to the left; negative returns are more likely to occur than positive one*

Definition 12.1.3. *[1, p. 11] for multivariate financial time series*

- *The absolute value of cross-correlations between return series is less pronounced and the contemporaneous correlations are in general the strongest;*
- *In contrast, the absolute or squared returns do show high cross-correlations. This empirical finding is similar to the univariate case;*
- *The contemporaneous correlations are not constant over time;*

- *Extreme observations in one return series are often accompanied by extremes in the other return series.*

12.1.2 Market statistics

Definition 12.1.4 (periodic return).

- *simple return*
$$r_i = \frac{P_{i+1}}{P_i} - 1.$$
- *log return*
$$r_i = \ln \frac{P_{i+1}}{P_i}.$$

Definition 12.1.5 (arithmetic and geometric mean return).

- $$\bar{r}_A = \frac{1}{n}\sum_{i=1}^{n} r_i.$$
- $$\bar{r}_G = (\prod_{i=1}^{n}(r_i + 1))^{1/n} - 1,$$
such that $(1+\bar{r}_G)^n = \prod_{i=1}^{n}(r_i+1)$.

Definition 12.1.6 (volatility, lognormal volatility, normal volatility).

- $$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r}_A)}.$$
-

Definition 12.1.7 (Beta).

Remark 12.1.1 (Beta vs. volatility).
- Volatility directly measures the uncertainty of return.
- Beta measures the relative risk with respect to the market.
  - Assets with $\beta > 1$ magnify fluctuations of the market.
  - Assets with $0 < \beta < 1$ mitigate fluctuations of the market.

12.1.3 Market invariants

Definition 12.1.8 (market invariant). *[2, p. 108]Consider a starting point t_0 and a time interval τ. Consider the set of equally-spaced dates*

$$D = \{t_0, t_0 + \tau, t_0 + 2\tau, ...\}.$$

Consider a set of random variables

$$X_t, t \in D.$$

*The set of random variables X_t are called **market invariants** if they are independent and identically distributed when the realization x_t of X_t becomes available at time t.*

In other words,

$$E[X_t|\mathcal{F}_t], E[X_{t+\tau}|\mathcal{F}_{t+\tau}], ...$$

are iid random variables.

Note 12.1.1 (equity invariant). [2, p. 108] Let P_t be the random variable representing stock price at time t. The invariant for the stock is

$$C_{t,\tau} \triangleq \ln \frac{P_{t+\tau}}{P_t}.$$

Note 12.1.2 (zero-coupon bond invariant). [2, p. 112] Let $Z(t, T)$ be the random variable representing the zero-coupon bond price at time t with maturity T. The **yield to maturity** for the zero-coupon bond is

$$F_{t,v} \triangleq \frac{1}{\tau} \ln Z(t, t + v).$$

$$R_{t,\tau,v} \triangleq \frac{Z(t + \tau, t + \tau + v)}{Z(t, t + v)}$$

is a market invariant.

The invariant, called **changes in yield to maturity**,is given by

$$X_{t,\tau,v} \triangleq F_{t+\tau,v} - Y_{t,v} = -\frac{1}{v} \ln R(t, \tau, v).$$

Remark 12.1.2 (other candidate quantities).

- The zero-coupon bond price itself cannot be a market invariant because it converges to 1 when approaching maturity.

Note 12.1.3 (derivative invariant). [2, p. 108]

12.2 Volatility and covariance structure modeling

12.2.1 Volatility concepts

Definition 12.2.1 (lognormal volatility).

Definition 12.2.2 (normal volatility).

12.2.2 GARCH model

12.2.2.1 *GARCH(1,1)*

Definition 12.2.3 (GARCH(1,1) model).

$$\sigma_t^2 = w + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2; w \geq 0, \alpha, \beta \geq 0, \alpha + \beta < 1.$$

12.3 Equity asset valuation

12.3.1 Discounted Dividend Valuation

Remark 12.3.1 (when discount dividend valuation model is appropriate). [3, p. 236] Discount dividend valuation model is most suitable when

- the company is dividend-paying (i.e., the analyst has a dividend record to analyze);
- the board of directors has established a dividend policy that bears an understandable and consistent relationship to the companyâĂŹs profitability; and
- the investor takes a noncontrol perspective.

Example 12.3.1. [3, p. 237] Consider the following time series characterizing the profitability and dividend payment of COKE and HRL. Note the EPS stands for earnings per share, DPS stands for dividend per share, and payout is DPS over EPS.

Year	COKE			HRL		
	EPS	DPS	Payout,%	EPS	DPS	Payout,%
2012	3.08	1	32	1.86	0.60	32
2011	3.08	1	32	1.74	0.51	29
2010	3.94	1	25	1.51	0.42	28
2009	3.56	1	28	1.27	0.38	30
2008	1.77	1	56	1.04	0.37	36
2007	2.17	1	46	1.07	0.30	28
2006	2.55	1	39	1.03	0.28	27
2005	2.53	1	40	0.91	0.26	29
2004	2.41	1	41	0.78	0.23	29
2003	3.40	1	29	0.67	0.21	31
2002	2.56	1	39	0.68	0.20	29
2001	1.07	1	93	0.65	0.19	29
2000	0.71	1	141	0.61	0.18	30
1999	0.37	1	270	0.54	0.17	31
1998	1.75	1	57	0.41	0.16	39

For the company COKE, there is no consistent relationship of dividends and earnings; therefore, it is inappropriate to use discount dividend model for COKE. For the company HRL, there is consistent relationship, and it is appropriate.

12.3.1.1 *Required equity return*

Definition 12.3.1 (required return on single stock from CAPM). *[3, p. 61] Let $E[r_M]$ denote the expected value of one-period return of the market portfolio. Let $E[r_i]$ denote the expected value of one-period return of the stock i.*

Then the ***required equity return for stock*** *i, denoted by $E[r_i]$, is given by*

$$E[r_i] = r_f + \beta_{M,i}(E[r_M] - r_f),$$

where r_f is the risk-free rate, $\beta_{M,i} = Cov[r_M, r_i]/Var[r_M]$.

Remark 12.3.2 (parameter estimation). [3, p. 237]

- Risk-free rate and market portfolio return can be directly estimated from historical data of bonds and index.
- β can be estimated from linear regression of historical stock return on market portfolio return. Note that the β from linear regression is the unbiased estimatorTheorem 15.6.3.

Example 12.3.2. Suppose current expected risk-free return is 3 percent, the asset's beta is 1.20, and the equity risk premium is 4.5 percent, then the asset's required return is

$$E[r_i] = 3\% + 1.2 \times (4.5\%) = 8.4\%.$$

Definition 12.3.2 (required return on single stock from multifactor model). *[3, p. 61] Let $E[r_M]$ denote the expected value of one-period return of the market portfolio. Let $E[r_i]$ denote the expected value of one-period return of the stock i.*

*Then the **required equity return for stock** i, denoted by $E[r_i]$, is given by*

$$E[r_i] = r_f + \beta_{M,i}(E[r_M] - r_f),$$

where r_f is the risk-free rate, $\beta_{M,i} = Cov[r_M, r_i]/Var[r_M]$.

Definition 12.3.3 (required return on bond and premium model). *[3, p. 61] Let $E[r_M]$ denote the expected value of one-period return of the market portfolio. Let $E[r_i]$ denote the expected value of one-period return of the stock i.*

*Then the **required equity return for stock** i, denoted by $E[r_i]$, is given by*

$$E[r_i] = r_f + \beta_{M,i}(E[r_M] - r_f),$$

where r_f is the risk-free rate, $\beta_{M,i} = Cov[r_M, r_i]/Var[r_M]$.

12.3.1.2 *The gordon growth model*

Lemma 12.3.1. *[3, p. 245] Assume*

- *one period dividend growth rate is constant g such that*

$$D_t = D_{t-1}(1 + g).$$

- *the market required return is constant r*

Let current time be 0 and current dividend be D_0. Then current equity value is given by

$$V_0 = \frac{D_1}{1+r} + \frac{D_2}{(1+r)^2} + \cdots + \frac{D_n}{(1+r)^n} + \cdots = \frac{D_1}{r-g} = \frac{D_0(1+g)}{r-g}.$$

Proof. Using geometric summation series, we have

$$V_0 = \frac{D_0}{1 - \frac{1+g}{1+r}} = D_0 \frac{1+g}{r-g}.$$

□

Remark 12.3.3 (valid parameter range).

- It must be specified that the required return on equity must be greater than the expected growth rate $r > g$.
- if $r \leq g$, we will get $V_0 = +\infty$. This means that the company is consistently growing at a speed faster than it should be,i.e., r, which is impossible. or r < g, equation 10 as a compact formula for value assuming constant growth is not valid. if r = g, dividends grow at the same rate at which they are discounted, so the value of the stock (as the undiscounted sum of all expected future dividends) is infnite. if r < g, dividends grow faster than they are discounted, so the value of the stock is infnite. of course, infnite values do not make economic sense; so constant growth with r = g or r < g does not make sense.

12.3.1.3 *Required equity return risk*

Remark 12.3.4.

- Suppose the risk-free rate increases due to monetary policy, then the required equity return will increase and devalue the equity.
- Suppose the market risk aversion increases, then the required equity return will increase(since factor premium will increase) and devalue the equity.

12.4 Futures strategies

Lemma 12.4.1 (value path for a portfolio of futures and spots). *Denote an underlying asset's spot price by S_t and its futures by $F(t,T)$. Assume the interest rate is constant r.*

- *Suppose the futures position was entered at t_0. The value of a portfolio consisting of 1 long position of spot and 1 short position of the futures will evolve as*

$$V(t) = F(t_0,T)\exp(-r(T-t)) = S(t_0)\exp(r(t-t_0)), t_0 \leq t \leq T.$$

- *If the underlying asset has proportional dividend rate q. The value of a portfolio consisting of 1 long position of spot and $\exp(q(T-t_0))$ short position of the futures will evolve as*

$$V(t) = F(t_0,T)\exp(-(r-q)(T-t)) = S(t_0)\exp((r-q)(t-t_0)), t_0 \leq t \leq T.$$

Proof. From Theorem 15.7.3, we know that the evolution of a forward/futures contract at time t is given by

$$V_F(t) = S_t\exp(-q(T-t)) - F(t_0,T)\exp(-r(T-t)).$$

Then the whole portfolio has value evolution given by

$$\begin{aligned} V(t) &= S_t - \exp(q(T-t))V_F(t) \\ &= S_t - S_t - F(t_0,T)\exp(-(r-q)(T-t)). \end{aligned}$$

Take $q = 0$ will give (1). □

Remark 12.4.1 (implication). [4, p. 49]

- To hedge a long position of the no-dividend underlying, we can enter a short position of the futures such that the asset value is locked to grow at risk-free rate.

Remark 12.4.2 (use financially settled futures to bet the market).

- Consider an underlying asset S_t has futures price $F(t,T)$ for a futures contract expiring at time T. The value $F(t,T)$ reflects the market's view on the price of S_t at time T.
- If we bet S_T will be higher than $F(t,T)$, then we can long the future. When S_T eventually rise above $F(t,T)$, we will gain.
- If we bet S_T will be lower than $F(t,T)$, then we can short the future. When S_T eventually decline below $F(t,T)$, we will gain.

12.5 Option strategies

12.5.1 Spread strategies

12.5.1.1 *Introduction*

Example 12.5.1 (cash-and-carry strategy). [5, p. 159]

- Given the current cash price, interest rate, and storage and insurance costs, a commodity trader can calculate the value of a forward contract.
- If the actual market price of the forward contract is higher than the calculated value, the trader will create a spread by purchasing the commodity, selling the overpriced forward contract, and carrying the position to maturity.
- For example, consider a commodity trading at a spot price of 700, adding the funding cost, insurance cost, and storage cost, we can calculate the forward price should be 715; If in the market now the forward contract is sold at 730, then we can short the forward contract and buy the commodity(by borrowing money). At the delivery date, we deliver commodity as required by the contract and earn 15.

Example 12.5.2 (cash-and-carry strategy). [5, p. 159]

- Given the current cash price, interest rate, and storage and insurance costs, a commodity trader can calculate the value of a forward contract.
- If the actual market price of the forward contract is higher than the calculated value, the trader will create a spread by purchasing the commodity, selling the overpriced forward contract, and carrying the position to maturity.
- For example, consider a commodity trading at a spot price of 700, adding the funding cost, insurance cost, and storage cost, we can calculate the forward price should be 715; If in the market now the forward contract is sold at 730, then we can short the forward contract and buy the commodity(by borrowing money). At the delivery date, we deliver commodity as required by the contract and earn 15.

Example 12.5.3 (spread with different maturities). [5, p. 159]

- Given the current cash price, interest rate, and storage and insurance costs, a commodity trader can calculate the value of a forward contract of **different maturities**.

- Suppose the trader calculates that the four-month forward prices should be greater than two-months contract by 15;however, the market shows that price difference is 20.
- The trader can short the spread(buy 2M and sell 4M) and wait for the price difference revert back to 15. When the price reverts, the trader can buy the spread(sell 2M and buy 4M) to close his position and earn 5.

Example 12.5.4 (spread with different commodities). [5, p. 159]

- Suppose that a trader observes the prices of two commodities, Commodity A and Commodity B, over an extended period and concludes that

$$Price_B = 3 \times Price_A.$$

- Suppose that in the market, $Price_A = 120, Price_B = 390$; the trader can buy a spread consists of long 3 units of A and short 1 unit of B.
- The trader can wait for the price ratio reverts back to 3; when it reverts, the trader will buy the spread to close its position and earn 30.
- Cash flow analysis: initial position $(3Price_A(0) - Price_B(0))$, final position

$$(3Price_A(0) - Price_B(0)) - (3Price_A(t) - Price_B(t)) = (3Price_A(0) - Price_B(0)) = 30.$$

- Suppose the trader calculates that the four-month forward prices should be greater than two-months contract by 15;however, the market shows that price difference is 20.
- The trader can short the spread(buy 2M and sell 4M) and wait for the price difference revert back to 15. When the price reverts, the trader can buy the spread(sell 2M and buy 4M) to close his position.

12.5.1.2 *Volatility spread*

12.5.1.2.1 SPREAD IN DIFFERENT STRIKE

Definition 12.5.1 (straddle). *[5, p. 170] A straddle consists of a call and a put where both options have the same exercise price and expiration date.*

Definition 12.5.2 (strangle). *[5, p. 171] A* ***strangle*** *consists of a long call and a long put (a long strangle) or a short call and a short put (a short strangle), where both options expire at the same time but the options have different exercise prices*

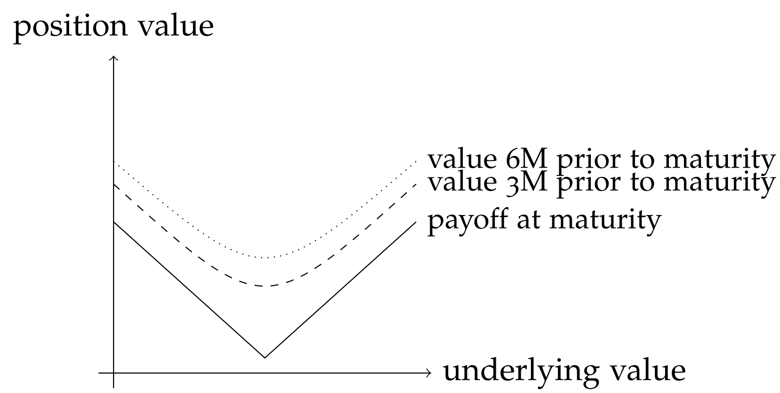


Figure 12.5.1: position value of a straddle(long)

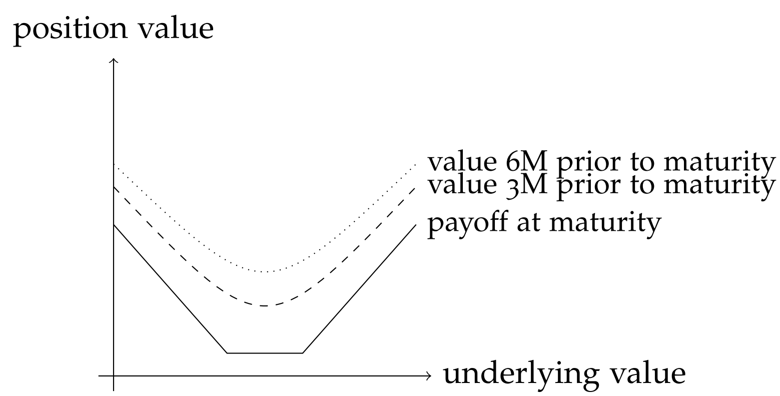


Figure 12.5.2: position value of a strangle(long)

Remark 12.5.1 (value change analysis for straddle and strangle). Consider a long position of a straddle or a strangle.

- The value will increase if there is increase of movement in the underlying asset.
- The value will decay as time passes.
- The value will increase if the implied volatility rises.

Definition 12.5.3 (butterfly). *[5, p. 173] A* ***butterfly*** *is a common three-sided spread consisting of options with equally spaced exercise prices, where all options are of the same type (either all calls or all puts) and expire at the same time.*

For example, long 1 unit call with $K = 100$, long 1 unit call with $K = 110$, and short 2 units of call with K=105.

Remark 12.5.2 (value change analysis for butterfly spread). [5, p. 176]

- Consider a long position of butterfly, a trader wants the underlying market to sit still (negative gamma, positive theta) and implied volatility to fall (negative vega).

The reason we have positive theta is because the short position of the two call with intermediate strike will have relatively large negative theta.

- Consider a short position of butterfly, a trader wants the underlying market to make a large move (+gamma, âĂŞtheta) and implied volatility to rise (+vega).
- There is one important difference between butterfly and straddle. While a straddle is open-ended in terms of either profit potential or risk, a butterfly is strictly limited.

Remark 12.5.3 (volatility vs. implied volatility and Gamma vs. Vega)**.**

- Volatility is used to characterize current movement of price, as a result of buying and selling forces in the market.
- Implied volatility reflects market's view/prediction on the future prices' distribution.
- Gamma characterizes the sensitivity to the second order variation of the price; Vega characterizes the sensitivity to the implied volatility.

12.5.1.2.2Spread in different maturity

Definition 12.5.4 (calendar spread). *[5, p. 182]*

- *Calendar spreads, sometimes referred to as time spreads or horizontal spreads, consist of option position having the same strike but expiring in different months.*
- *The most common type of calendar spread consists of opposing positions in two options of the same type (either both calls or both puts) where both options have the same exercise price.*
- *A typical strategy is long a long-term option and short a short-term option; this is called long the calendar spread. If the trader shorts a long-term option and long short-term option, the trader is short the calendar spread.*

Note 12.5.1 (value change analysis of calender spread)**.** [5, p. 186] A long position of a calender spread has negative theta(getting closer to expiry will increase value), negative gamma, and positive vega. |

- A calendar spread will increase in value if time passes with no movement in the underlying contract. This is because the short position of the short-termed option decreases value faster than the long-termed option.
- A calendar spread will increase in value if implied volatility rises but the underlying actually not making large movement because of negative gamma and positive vega. This 'contradictory' situation can occur in the following scenario: Consider a company that announces that an important statement will be made one week from today. If no one knows the content of the statement, there is unlikely to be any significant change in the companyâĂŹs stock price prior to the statement. But traders will assume that the statement, when it

is made, will have an effect, perhaps a dramatic one, on the stock price. The possibility of future movement in the stock price will cause implied volatility to rise. This combination of conditionsâĂŤthe lack of movement in the underlying stock together with rising implied volatilityâĂŤwill cause calendar spreads to increase value.

Note 12.5.2 (market strategy and principle for calender spread)**.**

- Long calender spreads are likely to be profitable when implied volatility is low but is expected to rise;
- Short calender spreads are likely to be profitable when implied volatility is high but is expected to fall.

12.5.1.3 *Bull and bear spreads*

12.5.1.4 *Risk consideration*

Note 12.5.3. [5, p. 228]

- **Delta(directional) risk** This risk that the underlying market will move in direction and cause the value of the option change. A delta-neutral position is typically immune to the directional risk **within a limited range**.
- **Gamma(curvature) risk** The risk of a **large move** in the underlying contract price, regardless of direction and cause the value of option to change.
- **Theta(Time decay) risk** The risk of option value change as time passes. In general, a positive gamma always goes hand in hand with a negative theta;a negative gamma always goes hand in hand with a positive theta. A trader in a position with positive Gamma and negative theta has to determine how much time can he wait for large movement of underlying to limit the loss.
- **Vega(volatility) risk**.The risk of a change in the underlying contract price volatility and cause the value of option to change.
- **Rho(volatility) risk**.The risk of a change in the interest rate and cause the value of option to change. Usually, interest rate risk is the least important to consider.

12.5.2 options for hedging(protection)

12.5.2.1 *Protective put and protective call*

Definition 12.5.5 (protective put strategy). *[6, p. 506] The protective put involves the purchase of a put option combined with a long stock position. Protective put strategies are valuable to portfolio managers who currently hold a long position in the underlying security or investors who desire upside exposure and downside protection.*

Definition 12.5.6 (protective call strategy). *[6, p. 506]The protective call is a hedging strategy whereby the trader, who has an existing short position in the underlying security, buys call options to guard against a rise in the price of that security.*

Another example is the commodity end user prefers a falling prices the commodity and will buy a out-of-money call to hedge the loss due to rising prices.

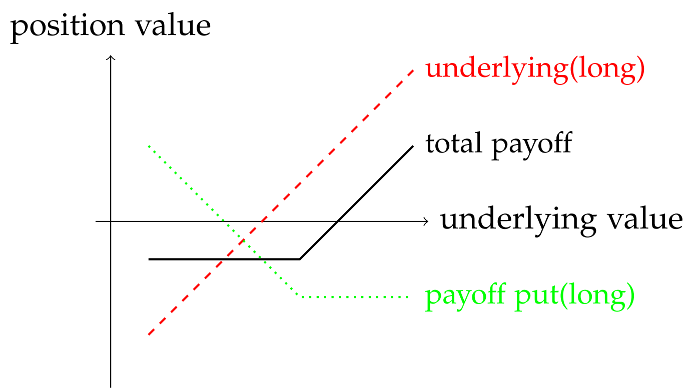


Figure 12.5.3: protective put strategy

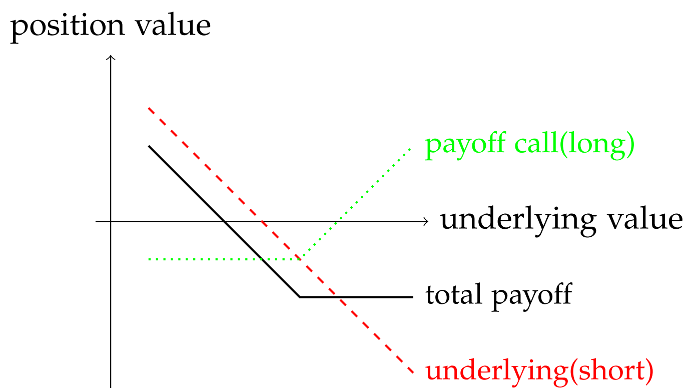


Figure 12.5.4: protective call strategy

12.5.2.2 *Collar strategy*

Definition 12.5.7 (collar strategy). *[6, p. 507]*

- *A collar strategy consists of a long stock position, a long put(out-of-money), and a short call(out-of-money);*
- *The long put is completely and partially financed by the short call position.*
- *Collars are designed for investors who currently hold a long equity position and want to achieve a level of risk reduction.The put strike price establishes a ïňĆoor and the call strike price a ceiling.*

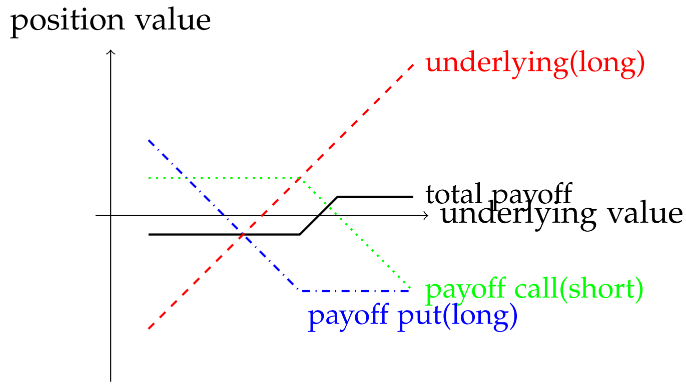


Figure 12.5.5: collar strategy

12.5.2.3 *Covered call and covered put strategy*

Definition 12.5.8 (covered call strategy). *[6, p. 506]*

- *A* ***covered call*** *is an options strategy that the trader buys (or already owns) a stock, then sells call options(usually out of money) for the same amount (or less) of stock, and then waits for the options contract to be exercised or to expire.*
- *The strategy is appropriate for slightly bullish investors who donâĂŹt expect much out of the stock and want to produce additional income from selling the call option.*

Definition 12.5.9 (covered put strategy). *[6, p. 506]*

- *A* ***covered put*** *is an options strategy that the trader sells (or already shorts) a stock, then sells put options(usually out of money) for the same amount (or less) of stock, and then waits for the options contract to be exercised or to expire.*
- *The strategy is appropriate for slightly bearish investors who donâĂŹt expect much out of the short position of the stock and want to produce additional income from selling the put option.*

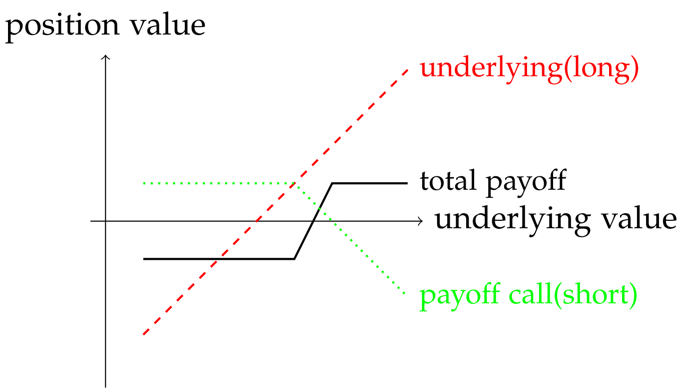


Figure 12.5.6: covered call strategy

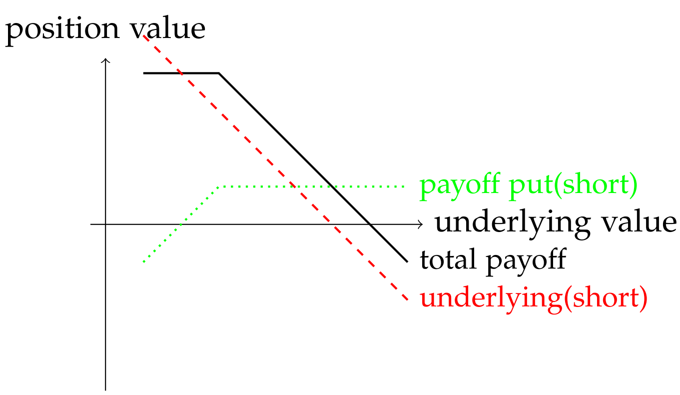


Figure 12.5.7: covered put strategy

12.6 Fixed income strategies

12.7 Commodity strategies

strategies for contango and backwardation market link.

[7][8][9]

12.8 Equity portfolio management

12.8.1 Factor model theory and application

12.8.1.1 *Factor model risk management*

12.8.1.2 *Multiple-factor model: alpha construction*

12.8.1.3 *Return attribution*

Definition 12.8.1 (active return, tracking error, active risk). *[10, p. 655]*

- *The active return associated with a portfolio is the difference between the return of the portfolio and the return of the benchmark portfolio; that is,*

$$R_P^A = R_P - R_B.$$

- *The **tracking error** or **active risk** associated with a portfolio is the standard deviation of the active return.*
- *The **active risk squared** associated with a portfolio is the variance of the active return.*

Definition 12.8.2 (information ratio).

$$IR = \frac{R_P^A}{\sigma(R_P^A)}$$

12.8.2 alpha seeking theory

12.8.2.1 *characteristic portfolio*

Definition 12.8.3 (asset characteristic). *[11, p. 28] An attribute is any quantity associated with an asset. Examples of attributes include volatility, excess return, beta etc. The attributes of a set of N assets can be represented by a vector $a \in \mathbb{R}^N$.*

- *The exposure of a portfolio w_P to attribute a is $a_P = w_P^T a$.*
- *A portfolio vector w_a which has **unit exposure to attribute** a **and minimum risk** is called **characteristic portfolio** of attribute a.*

$$w_a = \arg\min_{w \in \mathbb{R}^N} w^T V w, \ s.t. \ a^T w = 1.$$

Definition 12.8.4 (beta). *The beta of a portfolio A with respect to a portfolio B is defined by*

$$\beta = \frac{Cov(r_A, r_B)}{\sigma_B^2}.$$

Theorem 12.8.1 (basic properties of a characteristic portfolio). *[11, p. 29] Let $a \in \mathbb{R}^N$ be the vector of characteristics. Let A denote its characteristic portfolio*

- *The characteristic portfolio weight is given by*

$$w_a = \frac{V^{-1}a}{a^T V^{-1} a}.$$

- *(scaling properties of characteristic portfolio)If K is a positive scalar, then the characteristic portfolio of Ka is w_a/K.*
- *The variance of the characteristic portfolio w_a*

$$\sigma_a^2 = w_a^T V w_a = \frac{1}{a^T V^{-1} a}.$$

Then we can write

$$w_a = V^{-1} a \sigma_a^2.$$

- *The beta of asset i with respect to portfolio A is equal to its attribute vector a. Mathematically,*

$$\beta_i = \frac{\sum_{j=1}^N V_{ij} w_{a,j}}{\sigma_a^2},$$

or in vector form for all assets

$$\beta = \frac{V w_a}{\sigma_a^2} = a.$$

Proof. (1) The Lagrange associated with the optimization problem

$$\min_{w \in \mathbb{R}^N} w^T V w, \ s.t. \ a^T w = 1,$$

is

$$L(w) = w^T V w - \lambda(a^T w - 1),$$

and the condition for unique minimizer is

$$\begin{aligned}
Vw_a &= \lambda a, a^T w = 1 \\
\implies w_a &= \lambda V^{-1} a \\
\implies w_a^T V w_a &= \lambda a^T V^{-1} a = 1 \implies \lambda = \frac{1}{a^T V^{-1} a}, \\
\implies w_a &= \frac{V^{-1} a}{a^T V^{-1} a}
\end{aligned}$$

(3)

$$\sigma_a^2 = w_a^T V w_a = \frac{a^T V^{-1} V V^{-1} a}{a^T V^{-1} a a^T V^{-1} a} = \frac{1}{a^T V^{-1} a}.$$

(4)

$$\beta = \frac{Cov(r, w_a^T r)}{\sigma_a^2} = \frac{Vw_a}{a^T V^{-1} a}.$$

□

Lemma 12.8.1 (relationship between different characteristic portfolios). *[11, p. 29] Consider two attributes a and d with characteristic portfolio $w_a, w_d \in \mathbb{R}^N$. Let a_d and d_a be the exposure of portfolio w_d to attribute a and the exposure of the portfolio w_a to attribute d. It follows that*

- *The covariance of the two portfolio w_a and w_d are*

$$\sigma_{a,d} \triangleq Cov(r_a, r_d) = a_d \sigma_d^2 = d_a \sigma_a^2.$$

- *if an attribute a is a weighted combination of characteristic d and f such that $a = k_d d + k_f f$, then the characteristic portfolio of a is related to*

$$w_a = \frac{k_d \sigma_a^2}{\sigma_d} w_d + \frac{k_f \sigma_a^2}{\sigma_f^2} w_f,$$

where

$$\frac{1}{\sigma_a^2} = \frac{k_d a_d}{\sigma_d^2} + (\frac{k_f a_f}{\sigma_f^2}).$$

Proof. (1)

$$\begin{aligned}
\sigma_{a,d} &= w_a^T V w_d \\
&= [w_a^T V] w_d \\
&= \sigma_a^2 a^T w_d \\
&= \sigma_a^2 a_d
\end{aligned}$$

where we use the property(Theorem 15.6.3)

$$w_a = V^{-1}a\sigma_a^2 \Leftrightarrow Vw_a = a\sigma_a^2.$$

Similarly we can prove the rest. □

Lemma 12.8.2 (characteristic portfolio of excess return and maximum Sharpe ratio). *[11, pp. 32, 38]*

- *The characteristic portfolio R of excess return is given by*

$$w_R = \frac{V^{-1}r}{r^T V^{-1} r}$$

with variance of

$$\sigma_R^2 = w_R^T V w_R = \frac{1}{r^T V^{-1} r}.$$

- *Portfolio R has the maximum Sharpe ratio given by*

$$SR^2 = \frac{1}{\sigma_R^2},$$

among all the portfolio.

- ***If the CAPM assumptions hold, then the market portfolio is equivalent to the characteristic portfolio of R.***

Proof. (1) See Theorem 12.8.1. (2) Note that for any fixed mean excess return $K \in \mathbb{R}^+$, the portfolio Kw_R has mean excess return of K and minimum variance. Therefore, w_R and its positive scaling has the maximum Sharpe ratio. (3) □

Lemma 12.8.3 (characteristic portfolio of alpha). *[11, p. 135]*

- *The characteristic portfolio A of alpha is given by*

$$w_\alpha = \frac{V^{-1}\alpha}{\alpha^T V^{-1}\alpha}$$

with variance of

$$\sigma_\alpha^2 = w_\alpha^T V w_\alpha = \frac{1}{\alpha^T V^{-1}\alpha}.$$

- *Portfolio A has a zero beta with respect to the benchmark portfolio.*

- *Portfolio A has the maximum information ratio given by*

$$IR = \frac{1}{\sigma_\alpha^2},$$

among all the portfolio.
- *The Sharpe ratio of the portfolio A is given by*

$$SR^2 = SR_B^2 + IR^2,$$

where SR_B is the Sharpe ratio of the benchmark portfolio.

Proof. (1) See Theorem 12.8.1. (2)Lemma 12.8.1 (3) Note that for any fixed alpha $K \in \mathbb{R}^+$, the portfolio Kw_α has alpha of K and minimum variance. Therefore, w_α and its positive scaling has the maximum information ratio. □

Note 12.8.1 (mean excess return characteristic portfolio, alpha characteristic portfolio and benchmark market portfolio)**.**

12.8.2.2 *alpha in single factor model*

Definition 12.8.5 (alpha and residual return). *[11, p. 111] Assume the stochastic time series of* ***excess return*** *of a portfolio $r_P(t)$ is related by the stochastic time series excess return of the benchmark portfolio $r_B(t)$ via*

$$r_P(t) = \alpha_P + \beta_P r_B(t) + \epsilon_P(t).$$

where $\epsilon_P(t)$ is the white noise process with variance function $\sigma^2(t)$. Then

- *α_P, a deterministic quantity, is called the* ***alpha*** *of the portfolio P.*
- *$\theta_P(t) = \alpha_P + \epsilon_P(t)$, a stochastic process, is called the* ***residual return*** *of portfolio P.*
- *the* ***portfolio residual risk*** *is the standard deviation of $\theta_P(t)$.*

Lemma 12.8.4. *[11, p. 111]*

Assume in the universe, all assets' excess returns are given by

$$r_i(t) = \alpha_i + \beta_i r_B(t) + \epsilon_i(t), i = 1, 2, ..., N,$$

where $\epsilon_i(t)$ is the white noise process with variance function $\sigma_i^2(t)$. Let w_B be the portfolio weight vector of the benchmark portfolio. Let w_P be the portfolio weight vector of the our own portfolio. Define $w_{PA} \triangleq w_P - w_B$. Denote the vector of alphas as α. It follows that

- *(alpha neutrality of the benchmark)If the number of assets N is sufficiently large, then the alpha for the benchmark portfolio is*

$$\alpha_B \triangleq \alpha^T w_B \approx 0$$

. Also, we have constraints for the β such that $w_B^T \beta = 1$.
- *(additivity of alpha)The alpha of a portfolio is*

$$\alpha_P = \alpha^T w_P = \alpha^T w_{PA}.$$

Proof. (1) In vector notation, we have

$$r_B(t) = w_B^T r(t) = w_B^T \alpha + w_B^T \beta r_B + w_B^T \epsilon_i(t).$$

If the number of assets $N \to \infty$, then $w_B^T \epsilon_i(t)$ has zero mean and approximately zero variance. To make the two sides equal, we must have

$$w_B^T \alpha = 0, w^T \beta = 1.$$

(2)In vector notation, we have

$$r_P(t) = w_P^T r(t) = w_P^T \alpha + w_P^T \beta r_B + w_P^T \epsilon_i(t).$$

From the definition of alpha, we have

$$\alpha_P = w_P^T \alpha = \alpha^T (w_B + w_{PA}) = \alpha^T w_{PA}$$

. □

Lemma 12.8.5 (calculation active variance). *[11, p. 90]*

Assume in the universe, all assets' excess returns are given by

$$r_i(t) = \alpha_i + \beta_i r_B(t) + \epsilon_i(t), i = 1, 2, ..., N,$$

where $\epsilon_i(t)$ is the white noise process with variance function $\sigma_i^2(t)$. Let w_B be the portfolio weight vector of the benchmark portfolio. Let w_P be the portfolio weight vector of the our own portfolio. Define $w_{PA} \triangleq w_P - w_B$. Denote the covariance matrix of the excess asset return by $V(t) = Cov(r(t), r(t))$.

Then active variance is the variance of the active position, which is given by

$$\psi_P^2(t) = w_{PA}^T V(t) w_{PA},$$

or equivalently,

$$\psi_P^2(t) = \beta_P^2 \sigma_B^2 + \sigma^2(\epsilon_P(t)),$$

Example 12.8.1. [11, p. 115] Consider an universe with only 4 stocks and the risk free asset. We define the benchmark portfolio as the equal-weighted portfolio. We also construct two different portfolios whose weights are different from the benchmark.

stock	α	benchmark weight	portfolio P total weight	portfolio P active weight	portfolio L total weight	portfolio L active weight
1	1.5%	25%	35%	10%	40%	15%
2	-2.00%	25%	10%	-15%	2.50%	-22.5%
3	1.75%	25%	40%	15%	47.5%	22.5%
4	-1.25%	25%	15%	-10%	10%	-15%
benchmark	0.0%					
cash	0.0%					

Definition 12.8.6 (active management objective). *[11, p. 119] The objective of active management is to maximize the value added from the residual return, where the value added is measured by*

$$VA[P] = \alpha_P - \lambda_R Var[\theta_P],$$

where λ is the risk aversion parameter.

Lemma 12.8.6 (residual frontier). *[11, p. 117]*

Lemma 12.8.7 (active risk decomposition).

$$active\ return = \sum_{j=1}^{K}[(portfolio\ sensitivity)_j - (benchmark\ sensitivity)_j] \times (Factor\ return)_j + asset\ selecti$$

- ***active factor risk*** *is the contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.*
- ***active specific risk or asset selection risk*** *is the contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interest with assets' residual risk.*

$$active\ specific\ risk = \sum_{i=1}^{n} w_i^a \sigma^2(\epsilon_i)$$

where

- w_i^a *is the i asset's active weight in the portfolio(that is, the difference between the asset's weight in the portfolio and its weight in the benchmark portfolio).*
- $\sigma^2(\epsilon_i$ *is the residual risk of the i asset; that is, the variance of the i asset return left unexplained by the factor model.*

Example 12.8.2 (calculating active risk and tracking error). [12, p. 506]

month	portfolio return,%	benchmark return,%	active return,%
Jan	-1.05	-0.04	-1.01
Feb	2.13	1.54	0.59
March	0.37	0.00	0.37
April	1.01	0.54	0.47
May	-1.44	-0.76	-0.68
June	-0.57	-0.30	-0.27
July	1.95	0.83	1.12
Aug	1.26	1.23	0.03
Sep	2.17	0.76	1.41
Oct	1.80	0.90	0.90
Nov	2.13	1.04	1.09
Dec	-0.32	0.28	-0.60
Sum			3.42
mean			0.285
tracking error = standard deviation			0.7913

12.8.2.3 *Risk attribution*

12.8.2.4 *Tracking portfolio*

Definition 12.8.7 (tracking portfolio,665). *A **tracking portfolio** is a portfolio having factor sensitivities that are matched to those of a benchmark or other portfolio.*

Lemma 12.8.8 (creating a tracking portfolio).

Lemma 12.8.9 (creating a factor portfolio).

Remark 12.8.1 (motivations for creating a factor portfolio)**.** Suppose we create a well-diversified factor portfolio corresponding to the inflation factor risk such that this portfolio's return can be described by

$$R_P = R_f + \lambda\beta + \beta f_{INF}.$$

Note that this factor portfolio has no asset-specific risk. Then this factor portfolio can be used to bet the inflation.

Example 12.8.3 (creating a tracking portfolio,665).

portfolio	expected return	sensitivity to inflation factor	sensitivity to GDP factor
J	0.14	1.0	1.5
K	0.12	0.5	1.0
L	0.11	1.3	1.1
benchmark		1.3	1.975

The linear system for us to construct a tracking portfolio is given by

$$\begin{aligned} 1 &= w_1 + w_2 + w_3 \\ 1.3 &= 1.0w_1 + 0.5w_2 + 1.3w_3 \\ 1.975 &= 1.5w_1 + w_2 + 1.1w_3 \end{aligned}$$

which produce result

$$w_1 = 2, w_2 = -0.75, w_3 = -0.25.$$

12.8.2.5 *Factor portfolio and investment*

12.8.3 Fundamental law of active portfolio management

12.9 Notes on bibliography

The major reference are [13]. [5]

Factor investment [14].

Active portfolio management, see [15].

Good resources on quantitative trading linklinklink.

Good data source: quandl.

For application of machine learning in financial modeling, see [16].

BIBLIOGRAPHY

1. Homescu, C. *Many risks, one (optimal) portfolio* (SSRN, 2014).
2. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
3. Henry, E., Robinson, T. R., Stowe, J. D., *et al. Equity asset valuation* (John Wiley & Sons, 2010).
4. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
5. Natenberg, S. *Option Volatility and Pricing Workbook, Second Edition* ISBN: 9780071819053. https://books.google.com/books?id=5calngEACAAJ (McGraw-Hill Education, 2014).
6. Fabozzi, F. & Pachamanova, D. *Portfolio Construction and Analytics* ISBN: 9781119238164. https://books.google.com/books?id=0OTFCwAAQBAJ (Wiley, 2016).
7. Erb, C. B. & Harvey, C. R. The Tactical and Strategic Value of Commodity Futures. *National Bureau of Economic Research* **62,** 1–61. ISSN: 1556-5068 (2006).
8. Miffre, J. & Rallis, G. Momentum strategies in commodity futures markets. *Journal of Banking & Finance* **31,** 1863–1886. ISSN: 03784266 (2007).
9. Fuertes, A. M., Miffre, J. & Rallis, G. Tactical allocation in commodity futures markets: Combining momentum and term structure signals. *Journal of Banking and Finance* **34,** 2530–2548. ISSN: 03784266 (2010).
10. DeFusco, R. A., McLeavey, D. W., Pinto, J. E., Anson, M. J. & Runkle, D. E. *Quantitative investment analysis* (John Wiley & Sons, 2015).
11. Grinold, R. & Kahn, R. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk* ISBN: 9780071376952. https://books.google.com/books?id=aSDkFQtysy8C (McGraw-Hill Education, 1999).
12. Fabozzi, F. *Bond Markets, Analysis, and Strategies* ISBN: 9780133797138. https://books.google.com/books?id=a7WgBwAAQBAJ (Pearson Education, 2015).
13. Luenberger, D. G. *et al.* Investment science. *OUP Catalogue* (1997).
14. Foundations of Factor Investing. *Msci,* 33 (2013).
15. De Silva, H., Thorley, S. & Clarke, R. G. The fundamental law of active portfolio management (2006).
16. De Prado, M. L. *Advances in financial machine learning* (John Wiley & Sons, 2018).

13

MARKET MICROSTRUCTURE

13.1 Order book and market microstructure

13.1.1 Order book

Definition 13.1.1 (liquidity). *Ability to buy or sell significant quantities of a security quickly, anonymously, and with minimal or no price impact.*

Definition 13.1.2 (bid price, ask, price, midprice, bid-ask spread).

- ***Bid price*** *is the price buyers offer to buy.*
- ***Ask price*** *is the price sellers offer to sell.*
- ***Midprice*** *is the middle point between bid price and the ask price.*
- *A* ***bid-ask spread*** *is the amount by which the ask price exceeds the bid price for an asset in the market. The bid-ask spread is essentially the difference between the highest price that a buyer is willing to pay for an asset and the lowest price that a seller is willing to accept to sell it.*

Example 13.1.1. link

Let's assume that Morgan Stanley Capital International (MSCI) wants to purchase 1,000 shares of XYZ stock at $10, and Merrill Lynch & Co. wants to sell 1,500 shares at $10.25. The spread is the difference between the asking price of $10.25 and the bid price$10, or 25 cents.

- An individual investor looking at this spread would then know that, if he wants to sell 1,000 shares, he could do so at $10 by selling to MSCI(if no bid-ask spread exists, he can sell more at $10.25). Conversely, the same investor would know that he could purchase 1,500 shares from Merrill Lynch at $10.25(if no bid-ask spread exists, he can buy less at $10).

Definition 13.1.3 (market making).

- *A market maker or liquidity provider is a company or an individual that quotes both a buy and a sell price in a financial instrument or commodity held in inventory, hoping to make a profit on the bid-offer spread, or turn.*
- *The market maker provide liquidity by taking the opposite side of a transaction.*

Note 13.1.1 (the risks of market making). the primary risk a Market Maker can face is a decline in the value of a security after it has been purchased from a seller and before itâĂŹs sold to a buyer.

Definition 13.1.4 (trading priority rule). *Trading Priority Rules consist of the price priority rule and the time priority rule.*

- *The Price Priority Rule: With price priority, buyers posting higher bids have priority over buyers posting lower bids and sellers posting lower asks have priority over sellers posting higher asks.*
- *In addition to price priority, a secondary trading priority rule is often established to specify the sequence to be followed for orders submitted at the same price. With time priority, the first order placed is the first to execute.*

Definition 13.1.5 (types of orders).

- *A **market order** to buy or sell is to be executed at the best price established on the market at a given point in time. For a market order seller, the best price is the highest bid posted by a potential buyer; For a market order buyer, the best price is the lowest ask posted by a potential seller.*
- *In a **limit order**, an individual places a limit order to sell or buy a certain amount of stock at a given price or better. A limit buy order specifies the maximum price at which the trader will buy; a limit sell order specifies the minimum price at which the trader will sell.*
- *A stop-loss order is an order placed with a broker to sell a security when it reaches a certain price. Stop loss orders are designed to limit an investorâĂŹs loss on a position in a security. A stop order is used to capture a specific price (or higher) on a buy order, or to capture a specific price (or lower) on a sell order.*
  A buy stop limit order is used to buy at a specific price or lower (or within a range), while a sell stop limit is used to sell at a specific price or higher (or within a range). This combines elements of the basic stop and limit order types.
  Market if touched orders trigger a market order if a certain price is touched. A limit if touched order sends out a limit order if a specific trigger price is reached.

Definition 13.1.6 (limit order book operations). *[1] Consider a financial asset traded in an order-driven market. Market participants can post two types of buy/sell orders. A limit order is an order to trade a certain amount of a security at a given price.*

- ***Limit orders** are posted to a electronic trading system and the state of outstanding limit orders can be summarized by stating the quantities posted at each price level: this is known as the **limit order book.***
- *The lowest price for which there is an outstanding limit sell order is called the **ask price** and the highest buy price is called the **bid price**.*

Table 13.1.1: Message Book

	Time(sec)	Price($)	Volume	Event Type	Direction
k-1	34203.011926972	585.68	18	execution	ask
k	34203.011926973	585.69	16	execution	ask
...	...	...	...	...	...
k+4	34203.011988208	585.74	18	cancelation	ask
k+5	34203.011990228	585.75	4	cancelation	ask
...	...	...	...	...	...
k+8	34203.012050158	585.70	66	execution	bid
k+9	34203.012287906	585.45	18	submission	bid
k+10	34203.089491920	586.68	18	submission	ask

- *A **market order** is an order to buy/sell (a certain quantity of) the asset at the best available price in the limit order book.When a market order arrives it is matched with the best available price in the limit order book and a trade occurs. The quantities available in the limit order book are updated accordingly.*
- *A limit order sits in the order book until it is either executed against a market order or it is canceled.Alternatively, a limit order can be canceled at any time before it is fulfilled.ïijŇ*

A limit order may be executed very quickly if it corresponds to a price near the bid and the ask, but may take a long time if the market price moves away from the requested price or if the requested price is too far from the bid/ask.

link

limit order submission, limit order cancellation, or market order execution as shown in column âĂIJEvent TypeâĂİ

Table 13.1.2: Order Book

Ask		Bid	
Price	Volume	Price	Volume
585.68	18	585.70	66
585.69	16	585.70	66
$\cdots$	$\cdots$	$\cdots$	*cdots*
585.71	118	585.45	18
585.71	118	585.45	18
585.71	118	585.44	167

13.2 Notes on bibliography

Major references are .

[2][3].

[4][5]

forex factory, quantocracy, quantnews

Some books on algo trading: Trades, Quotes and Prices: financial markets under the microscope

Mean Reversion Trading systems

Foundation of trading: develoing profitable trading systems using scientific techniques

BIBLIOGRAPHY

1. Cont, R., Stoikov, S. & Talreja, R. A stochastic model for order book dynamics. *Operations research* **58,** 549–563 (2010).
2. Kissell, R. *The science of algorithmic trading and portfolio management* (Academic Press, 2013).
3. Guéant, O. *The financial mathematics of market liquidity: From optimal execution to market making* (CRC Press, 2016).
4. Chan, E. *Algorithmic Trading: Winning Strategies and Their Rationale* ISBN: 9781118746912. https://books.google.com/books?id=CIwCTVqEj4oC (Wiley, 2013).
5. Moore, H. M. *Advanced Algorithmic Trading* (Moore, H. Michael, 2017).

14

ALGORITHMIC TRADING STRATEGIES

14.1 Classical trading strategies

14.1.1 Buy and hold strategy

Definition 14.1.1. *Buy and hold is a passive investment strategy in which an investor buys stocks and holds them for a long period of time, regardless of fluctuations in the market. An investor who employs a buy-and-hold strategy actively selects stocks, but once in a position, is not concerned with short-term price movements and technical indicators.*

14.1.2 Trending following

Definition 14.1.2 (golden crossing rule). *[1, p. 466]*

- *At every bar calculate the 10-day and 30-day simple moving averages (SMA).*
- *If the 10-day SMA exceeds the 30-day SMA and the strategy is not invested, then go long.*
- *If the 30-day SMA exceeds the 10-day SMA and the strategy is invested, then close the position.*

Remark 14.1.1 (intuition).

- When the 10-day SMA exceeds the 30-day SMA, it suggests that there is a trend of continuing growing.
- When the 30-day SMA exceeds the 10-day SMA, it suggests that the growing trend will stop.

Definition 14.1.3 (turtle trading rules).

- *Go long (short) when the price exceeds the highest (lowest) of the preceding 20 days.*
- *Close the trade when there was a 10 day low for long positions and a 10 day high for short positions.*

14.1.3 Event-driven trading strategy

Definition 14.1.4. *link Some common micro-level events to watch include:*

Earnings Releases -Corporate earnings tend to move markets when they come in above or below the market's expectations, which means that it's important for active traders to understand the expected figures beforehand.

- *Mergers & Acquisitions âĂŞ M&A tends to produce dramatic increases or decreases in share prices depending on the terms of the deal, while creating an opportunity for arbitrage strategies between the buyer and seller.*
- *Spin-Offs - Spin-offs tend to see an initial decline in share price as institutional investors who received shares sell off their stake to comply with regulatory requirements or other rules, thereby creating opportunities for traders.*

  Macro-level events to watch include:

  Natural Disasters - Natural disasters can spark dramatic movements in the equity markets, especially in certain sectors that are exposed. For instance, a hurricane in the Gulf of Mexico could hurt oil companies with rigs in the region.
- *Politics âĂŞ Political issues can have a dramatic impact on some equities, especially in parts of the world where policies can change dramatically. A new regime in an emerging market, for instance, can have a big impact on the country's ETFs.*
- *Monetary Policy - Central bank monetary policy changes can have a big impact on broad equity indexes, since interest rates directly influence portfolio allocations, which means that these events are important for traders to monitor closely.*

14.1.4 Sentiment trading strategy

Definition 14.1.5 (long-only sentiment trading strategy). *[1, p. 466]*

- *Long a stock if its sentiment value reaches a high value.*
- *Close the long position if its sentiment value reaches a low value.*

14.1.5 Time series prediction trading strategy

Definition 14.1.6 (ARIMA+GARCH trading strategy). *[1, p. 369] The strategy is carried out on a rolling basis:*

- *For each day, n, the previous k days of the differenced logarithmic returns of a stock market index are used as a window for fitting an optimal ARIMA and GARCH model.*
- *The combined model is used to make a prediction for the next day returns.*
- *If the prediction is negative the stock is shorted at the previous close, while if it is positive it is longed.*
- *If the prediction is the same direction as the previous day then nothing is changed.*

14.1.6 Equity long-short strategy

Definition 14.1.7. *[2, p. 84]link Long/short equity is an investing strategy of taking long positions in stocks that are expected to appreciate and short positions in stocks that are expected to decline. A long/short equity strategy seeks to minimize market exposure, while profiting from stock gains in the long positions, along with price declines in the short positions. Although this may not always be the case, the strategy should be profitable on a net basis. The long/short equity strategy is popular with hedge funds, many of which employ a market-neutral strategy, in which dollar amounts of both long and short positions are equal.*

Remark 14.1.2. The reason to go long and short at the same time is that we can use the cash from short-selling to fund the long position.

14.1.7 Convertible arbitrage strategy

Definition 14.1.8. *link[2, p. 86] a hedge fund using convertible arbitrage will buy a companyâĂŹs convertible bonds at the same time as it shorts the companyâĂŹs stock. If the companyâĂŹs stock price falls, the hedge fund will benefit from its short position; it is also likely that the companyâĂŹs convertible bonds will decline less than its stock, because they are protected by their value as fixed-income instruments. On the other hand, if the companyâĂŹs stock price rises, the hedge fund can convert its convertible bonds into stock and sell that stock at market value, thereby benefiting from its long position, and ideally, compensating for any losses on its short position.*

14.2 Dynamic model of trading strategies

14.3 Mean reversion trading

14.3.1 Testing mean reversion

14.3.2 Parameter estimation

Remark 14.3.1 (implication of the mean reversion parameter). [3, p. 47]

- First, if we find that λ is positive, this means the price series is not at all mean reverting, and we shouldn't even attempt to write a mean-reverting strategy to trade it.
- Second, if λ is very close to zero, this means the half-life will be very long, and a mean-reverting trading strategy will not be very prof table because we won't be able to complete many round-trip trades in a given time period.
- Third, this λ also determines a natural time scale for many parameters in our strategy. For example, if the half life is 20 days, we shouldn't use a look-back of 5 days to compute a moving average or standard deviation for a mean-reversion strategy. Often, setting the lookback to equal a small multiple of the half-life is close to optimal, and doing so will allow us to avoid brute-force optimization of a free parameter based on the performance of a trading strategy.

14.3.3 Trading strategy

Lemma 14.3.1 (linear trading strategy for stationary process). *Consider a stationary mean-reversion process characterizing a portfolio price given by*

$$dX_t = -\alpha X_t dt + \sigma dW_t, X_0 = 0,$$

where $\alpha, \sigma > 0$ and W_t is a Brownian motion.

Consider a trading strategy that maintain the $-kX_t, k > 0$ units of this portfolio via continuous rebalancing. Assume zero transaction cost. It follows that

- *The accumulating profit is from time 0 up to time T is a random variable given by*

$$Y_T = \int_0^T -kX_t dX_t = k\alpha \int_0^T X_t^2 dt - \int_0^T k\sigma X_t dW_t.$$

- *The mean and the variance of the profit is given by*

$$E[Y_T] = k\alpha \int_0^T E[X_t^2]dt = \frac{k\sigma^2}{2}(T - \frac{1}{2\alpha}(1 - \exp(-2\alpha T))).$$

$$Var[Y_T] = k^2\alpha^2 \int_0^T E[X_t^2]dt = \frac{k^2\alpha\sigma^2}{2}(T - \frac{1}{2\alpha}(1 - \exp(-2\alpha T))).$$

Proof. (1) The profits from a trading strategy is defined in Definition 15.6.3. (2)

$$\begin{aligned}
E[Y_T] &= E[k\alpha \int_0^T X_t^2 dt - \int_0^T k\sigma X_t dW_t] \\
&= E[k\alpha \int_0^T X_t^2 dt] \\
&= k\alpha \int_0^T E[X_t^2]dt \\
&= k\alpha \int_0^T \frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha t))dt \\
&= \frac{k\sigma^2}{2}(T - \frac{1}{2\alpha}(1 - \exp(-2\alpha T))).
\end{aligned}$$

where we use the solution of OU process Lemma 6.4.1. □

14.4 Pair trading

14.4.1 Distance method

Definition 14.4.1 (pair selection). *[4]*

- *Normalize the prices of all stocks. The normalized price is defined as the cumulative return index, adjusted for dividends and other corporate actions, and scaled to \$1 at the beginning of the formation period.*
- *Calculate the spread between the normalized prices of all possible combinations of stock pairs during the formation period. The formation period is chosen to be 12 months.*
- *We then select 20 of those combinations that have the least sum of squared spreads, or sum of squared differences (SSD), to form the nominated pairs to trade in the following trading period, that is chosen to be 6 months.*

Definition 14.4.2 (pair trading strategy). *[4]*

- *At the beginning of the trading period, prices are once again rescaled to \$1 and the spread is recalculated and monitored.*
- *When the spread diverges by 2 or more historical standard deviation (calculated in the formation period), we simultaneously open a long and a short position in the pair depending on the direction of the divergence.*
- *The two positions are closed (reversed) once the spread converges to zero again.*
- *The pair is then monitored for another potential divergence and therefore can complete multiple round-trip trades during the trading period.*

14.4.2 Cointegration method

Remark 14.4.1 (intuition: pair trading strategy in cointegration framework)**.**

- Wait large deviations from the long-term equilibrium of a cointegrated pair to open long and short positions.
- Positions are unwound once the equilibrium is restored, as a consequence of being stationary.

14.5 Backtesting

14.5.1 Bias in backtesting

Definition 14.5.1 (survivorship bias). *link Survivorship bias is the tendency to view the fund performance of existing funds in the market as a representative comprehensive sample. Survivorship bias can result in the overestimation of historical performance and general attributes of a fund.*

Definition 14.5.2 (look-ahead bias). *link*

Look-ahead bias is present if the test relates a variable to data that where not available at the points in time when the variable's outcomes were observed.

- *To avoid look-ahead bias, if an investor is backtesting the performance of a trading strategy, it is vital that he or she only uses information that would have been available at the time of the trade. For example, if a trade is simulated based on information that was not available at the time of the trade - such as a quarterly earnings number that was released three months later*

Definition 14.5.3 (time period bias).

- *Time period bias occurs when the period chosen is so short that it shows relationships that are unlikely to recur, or so long that it includes fundamental changes in the realtionhip being observed.*

14.6 Notes on bibliography

Major references are .

[5][6].

[3][1]

forex factory, quantocracy, quantnews

Some books on algo trading: Trades, Quotes and Prices: financial markets under the microscope

Mean Reversion Trading systems

Foundation of trading: develoing profitable trading systems using scientific techniques

BIBLIOGRAPHY

1. Moore, H. M. *Advanced Algorithmic Trading* (Moore, H. Michael, 2017).
2. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
3. Chan, E. *Algorithmic Trading: Winning Strategies and Their Rationale* ISBN: 9781118746912. https://books.google.com/books?id=CIwCTVqEj4oC (Wiley, 2013).
4. Rad, H., Low, R. K. Y. & Faff, R. The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance* **16,** 1541–1558 (2016).
5. Kissell, R. *The science of algorithmic trading and portfolio management* (Academic Press, 2013).
6. Guéant, O. *The financial mathematics of market liquidity: From optimal execution to market making* (CRC Press, 2016).

Part III

NO ARBITRAGE PRICING

15

NO-ARBITRAGE PRICING PRINCIPLE

15.1 Introduction

15.1.0.1 *Overview*

The no-arbitrage pricing framework developed in this work is based on a fundamental assumption that market participants cannot gain riskless profits by buying and selling instruments traded in an active and integrated market, as we emphasize in Assumption 15.1.

> **Assumption 15.1 (the fundamental no-arbitrage assumption).** *There is no opportunities for market participants to gain riskless profits by buying and selling instruments traded in an active and integrated market*

If a real-world market or a market mathematical model constructed satisfies such assumption, we say the market (model) is no-arbitrage market.

Assuming markets are free of arbitrage is a reasonable assumption, particularly to active markets where any 'free lunch' will be quickly arbitraged away by arbitragers. Without constructing any sophisticated model, this no-arbitrage assumption can lead to Law of one price, which essentially states that financial instruments with same payoff at a specified future time should have same prices; otherwise arbitrage opportunities exist.

Law of One Price can be readily applied to derive no-arbitrage price for instruments and construct no-arbitrage market models. For example, linear pricing theorem, a direct consequence of Law of One Price, can be used to price complex instruments whose payoff can be linearly decomposed to payoff functions of simple and actively traded instruments. Law of One Price is also useful in deriving price bounds and relations among different financial instruments.

Despite all the merits of Law of One Price, more exotic instruments and portfolio risk management (where how price will change as response to market conditions) requires sophisticated mathematical market model to be constructed.

The breakthrough in option pricing came with the famous Black and Scholes paper in 1973[1], where they show that option could be priced by dynamically creating a riskless portfolio via buy and sell the underlying stock continuously.

The other side of the idea of dynamical hedging is dynamical replication. Consider managing a portfolio based on a trading strategy such that the portfolio at the expiration date of an option exactly replicate the payoff, the Law of One Price gives that the initial capital of the replicating portfolio should be the same of the price of the option.

Such dynamical replication based pricing methodology has been formalized and greatly extended by risk-neutral pricing framework developed by M. Harrison, D. Kreps, and S. Pliska [2]. The construction and application of this types of models usually require additional assumption, as we note in

Assumption 15.2 (additional assumptions for the market).

- *We always consider an economy with continuous and frictionless (no transaction cost) trading take place within a finite horizon $[0, T]$.*
- *We can borrow and lend at the same rate without limit.*
- *The market participants are subject to the same tax rate on all net trading profits.*

They pointed out that once a model is built, one can change the probability measure on the space on which the stock price is defined so that it has mean rate of return equal to the interest rate. This is akin to building a model based on tosses of a fair coin, and then pretending for computational purposes that the coin is biased. Under this socalled risk-neutral measure, both the stock and the money-market account held by the portfolio replicating a call have mean rate of return equal to the interest rate, and so the portfolio itself has this mean rate of return. Therefore, the initial value of the portfolio, which is the Black-Scholes price of the call, can be obtained by discounting the call payoff at the interest rate and taking the expected value under the risk-neutral measure.

In this chapter, instead of introducing concrete market models, we cover different types generic market model frameworks, including single-period finite-state model, multiple-period finite state-state model and continuous-time infinite-stat model driven by Brownian motion.

For these models framework, we demonstrate how the no-arbitrage conditions are ensured using the martingale theory. These no-arbitrage martingale pricing frameworks can then serve as model factories to enable us to construct different types of concrete model based on specific requirements of markets and instruments.

15.1.1 The concept of price

Definition 15.1.1 (different types of prices).

- ***Instrinsic price*** *of an instrument is the amount of money that rational buyers and rational sellers agree on in a liquid market. It is the result of market forces.*
- ***No-arbitrage price*** *of an instrument is the price that arbitrage opportunities cannot occur.*
- *In an efficient market where no arbitrage opportunities exist, the no-arbitrage price will equal the instrinsic price.*
- *In asset pricing, we use no-arbitrage price to approximate the instrinsic price in a no-arbitrage market model.*

15.1.2 The goal of asset price

Remark 15.1.1 (The goal of asset pricing).

- Price of an asset is the amount of money that rational buyers and rational sellers agree on in a liquid market.
- Some asset prices are already existing, such that bonds and stocks.
- Our goal is to
  - Determine prices for untraded instruments based on the prices of instruments currently traded.
  - Understand the relationship between prices and underlying risk factors, laying the foundation for risk managemenet.
- There are generally two types of methods for pricing assets: model-based pricing and model-free pricing.
  - Examples for model-free pricing include pricing via replication, linear pricing theorem.
  - Examples of model based pricing is: Assuming asset dynamics are given by SDE driven by Brownian motions. The model should be arbitrage free.

15.2 The Law of one price

15.2.1 The Law of One Price

Theorem 15.2.1 (The law of One Price). *[3] In an **arbitrage-free** market, consider the values at time t of two portfolio V_X, V_Y (they are random variables parametrized by time $t' > t$): if they are the 'same' random variable at a future time $\tau > t$, in the sense that, they have the equivalent mapping from the random sample space to price, then they must have the same value at time t. If one portfolio at a future time $\tau > t$ is more valuable(or less) regardless of the random outcomes, then one portfolio is more valuable (or less).*

Proof. The simplest proof is that: a portfolio is a good that can generate cash flow, therefore it has value and price. (The portfolio is a good can bring payoff under different state of the world). In a competitive market, same goods must have the same price otherwise there will be arbitrage opportunities. □

Remark 15.2.1 (Implicit assumptions of no-arbitrage condition)**.** Implicit assumptions underlying the no-arbitrage condition:

- Markets are liquid: sufficient number of buyers and sellers
- Price information is available to all buyers and sellers
- Competitions in supply and demand will correct any deviation from no-arbitrage prices.
- Same borrow interest rate and lending rate.

Corollary 15.2.1.1 (valuation of zero-valued portfolio). *If the value of a portfolio is equal to 0 regardless of the random outcomes in future $\tau > t$, then the value of the portfolio is 0 at time t.*

Remark 15.2.2 (implications for pricing)**.**

- The law of one price will give **two ways of pricing a portfolio: portfolio replication and construction of risk-free portfolio.**
- Note that these two method can be converted mutually. Suppose we can replicate a state claim using some portfolio, which has value V_0. Then we can short this portfolio at V_0, and use it to buy the state claim at p. At time T, we return the state claim. In the process, we are guaranteed/deterministically to get zero at T. Therefore, our initial cash flow must be $0 = V_0 - p$.

Corollary 15.2.1.2 (valuation of risk free portfolio). *Consider a no-arbitrage market with risk-free asset. If the value $V(T)$ of a portfolio at time T in the future is independent of the random outcome, then*

- *its current time t value is*
$$V(t) = e^{-r(T-t)}V(T), t < T.$$
where r is the constant risk free rate.
- *the growth rate of $V(t)$ is r.*

Remark 15.2.3. The implication of this corollary is that if we can construct a risk-free portfolio, i.e., a portfolio without randomness and get a deterministic payoff V_T at future time T. Then its initial price V_0 must satisfy $V_0 = \exp(-rT)V_T$. It has to be such, otherwise people can just long(short) infinite amount of this portfolio by borrowing/lending money at interest rate r.

15.2.2 Linear pricing theorem

Theorem 15.2.2 (linear pricing theorem). *Consider a* ***arbitrage-free*** *market with N finite states. Assume there are two assets with payoff vector $D_1, D_2 \in \mathbb{R}^N$ with price p_1, p_2. Then*

- *For an asset with payoff vector $D = D_1 + D_2$, its price is $p = p_1 + p_2$.*
- *For an asset with payoff vector $aD_1, (a \in \mathbb{R})$, its price is $p = ap_1$.*
- *For an asset with payoff vector $aD_1 + bD_2, (a, b \in \mathbb{R})$, its price is $p = ap_1 + bp_2$.*

Proof. (1) If $p < p_1 + p2$, then we can short $D_1 + D_2$ and long D to earn risk-free money. Similar operations can apply to $p > p_1 + p_2$. Therefore, $p = p_1 = p_2$. (2)(3) similar to (1). □

Corollary 15.2.2.1 (Linear pricing theorem with Arrow securities). *Consider a* ***arbitrage-free*** *market with N finite states, assume there are a set of N Arrow securities that produce a unit payoff in each market scenario. If each Arrow security has price P_i, then for any other securities that with payoff vector $D \in \mathbb{R}^N$, its price is $D^T P$.*

Proof. Because the payoff vector is in vector space; that is, a payoff vector can be decomposed as linear combination of Arrow securities payoff vectors. □

15.2.3 Options Put-call parity & bounds

Lemma 15.2.1 (put-call parity).

- *European put-call parity at time t for non-dividend paying stocks:*

$$p(t,K,T) + S_t = c(t,K,T) + Kd(t,T)$$

where $d(t,T) < 1$ is the discount factor from t to T.

- *European put-call parity at time t for dividend paying stocks:*

$$p(t,K,T) + S_t = c(t,K,T) + D(t) + Kd(t,T)$$

where $d(t,T) < 1$ is the discount factor from t to T and $D(t)$ represents the total value of the dividends from one stock share to be paid out over the remaining life of the options, ***discounted to present value.*** *Specifically, if the dividend is continuously paid at a rate a, then*

$$p(t,K,T) + S_t e^{-a(T-t)} = c(t,K,T) + Kd(t,T)$$

Proof.

- Cash flow at time T: $\max\{S_T - K, 0\} - \max\{K - S_T, 0\} - S_T + K = 0$
- Cash flow at time t: $-c(t,K,T) + p(t,K,T) + S_t - Kd(t,T) = 0$

(1)(2)No-arbitrage(Corollary 15.2.1.1) requires once the first one holds, the other must hold. **another proof:** Suppose we long a call and short a put with the same strike K at time t, then at T our payoff is $S_T - K$. In another way, we buy S_t stock and borrow Kd money from bank at time t, at time T, we have payoff $S_T - K$. Based on the law of one price(Theorem 15.2.1), $C - P = S_t - Kd$.
When there are dividends: we long a call and short a put with the same strike K at time t, then at T our payoff is $S_T - K$. In another way, we buy S_t stock and borrow $Kd + D$ money from bank at time t, at time T, we have payoff $S_T + D/d - K - D/d$. Based on the law of one price(Theorem 15.2.1), $C - P = S_t - Kd - D$. (3) under risk-neutral measure, the dynamics of S_t is given as Lemma 15.6.4. Then use

$$(C_t - P_t)D(t) = E_Q[D(T)(S_T - K)|\mathcal{F}_t].$$

□

15.2.3.1 *Bounds on option pricing*

Lemma 15.2.2 (important bounds on option pricing).

- *Price of American option $\geq$ European option:*

$$c_A(t,K,T) \geq c_E(t,K,T), p_A(t,K,T) \geq p_E(t,K,T)$$

- *Lower bounds on European options*

$$c_E(t,K,T) = \max(S_t + p_E(t,K,T) - Kd(t,K), 0) \geq \max(S_t - Kd(t,T), 0)$$

and

$$p_E(t,K,T) = \max(-S_t + c_E(t,K,T) + Kd(t,K), 0) \geq \max(Kd(t,T) - S_t, 0)$$

where we use the put-call parity and the facts that $c_E, p_E \geq 0$.
- *Upper bounds on European options:*

$$c_E(T,K,T) = \max(S_T - K, 0) \leq S_T \Rightarrow c_E(t,K,T) \leq S_t$$

$$p_E(T,K,T) = \max(K - S_T, 0) \leq K \Rightarrow p_E(t,K,T) \leq Kd(t,T)$$

- *Never early exercise an American call on a non-dividend paying stock:*

$$c_A(t,K,T) \geq c_E(t,K,T) \geq \max(S_t - Kd(t,T), 0) > \max(S_t - K, 0)$$

therefore the price of an American call is always strictly greater than its exercise value $\max(S_t - K, 0)$

Corollary 15.2.2.2. *An American call has the same price as European call if they have the same strike price and expiration date.*

Proof. Since it is never optimal to exercise an American call. □

15.2.4 Forward pricing

Definition 15.2.1 (forward price). *When the two parties enter a contract that on a future date T, an asset is traded at the* ***forward price*** *F such that the two parties will not pay extra money to enter the contract.*

Lemma 15.2.3 (forward pricing using no-arbitrage argument). *The fair price should satisfy:*

$$F + \sum_{i=1}^{N} D_i e^{(r-q)(T-t)} = S_0 e^{(r-q)T}$$

where r is the risk free rate, D_i is the dividend guaranteed to pay at t_i, S_0 is the spot price of the underlying asset, q is the cost-of-carry. In particular, for the case of non-dividend and zero cost-of-carry, we have

$$F = S_0 e^{rT}.$$

Proof. The method to replicate a forward for the underlying asset is to buy the asset, and wait to contract expired date. The full cash flow generated within the process discounted to the payment date is the price of forward price. The idea is the asset hold till the expired date is the replication.

For example, If $F < S_0 e^{rT}$, we can short the asset S_0, and put the money into bank to get interest, and then at the expiry date we buy at $F < S_0 e^{rT}$ to return the asset; If $F > S_0 e^{rT}$, we will borrow money S_0 at a rate r to buy the asset, and then sell at $F > S_0 e^{rT}$ and clean the debt at T. $\square$

15.3 Single period finite-state market model

15.3.1 No-arbitrage and asset pricing

15.3.1.1 *The setup and the fundamental theorem*

Setup of general finite-state market

- A probability space (Ω, F, P) describing the world.
- A market in which N assets, labeled $A_1, A_2, ..., A_N$, are freely traded.
- Uncertainty about the behavior of the market is encapsulated in a finite-sized sample space Ω of K possible market scenarios or sample points, labeled $\omega_1, \omega_2, ..., \omega_K$, and $P(\omega) > 0 \forall \omega \in \Omega$. There is an N by K payoff matrix $D \in \mathbb{R}^{N\times K}$ with entries $V_j(\omega_i)$ such that, in scenario i, the value of the asset A_j at after one-period is $V_j(\omega_i)$. The payoff matrix D is given by

$$\begin{pmatrix} V_1(\omega_1) & V_1(\omega_2) & \cdots & V_1(\omega_K) \\ V_2(\omega_1) & V_2(\omega_2) & \cdots & V_2(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ V_N(\omega_1) & V_N(\omega_2) & \cdots & V_N(\omega_K) \end{pmatrix}$$

- We denote a portfolio vector by $\theta \in \mathbb{R}^N$
- The prices of the N assets at current time is represented by a price vector $p \in \mathbb{R}^N$.

Definition 15.3.1 (arbitrage portfolio). *A θ is an arbitrage portfolio is either*

1. *(type I arbitrage) initial price $p^T\theta \leq 0$ but final payoff vector $D^T\theta > 0, D^T\theta \in \mathbb{R}^K$.*
2. *(type II arbitrage) initial price $p^T\theta < 0$ but final payoff vector $D^T\theta \geq 0$.*

Remark 15.3.1. According to this, an arbitrary portfolio θ guarantees some positive return in every possible states, yet it costs nothing to purchase. Or it guarantees a non-negative return whiling have negative cost.

Theorem 15.3.1 (Fundamental theorem of asset pricing). *Consider a single-period finite state market model represented by payoff matrix $D \in \mathbb{R}^{N\times K}$. Further let the prices of the N assets at current time be represented by a price vector $p \in \mathbb{R}^N$. It follows that*

- *The market model has no arbitrage opportunities if and only there exists a nonnegative vector $\pi \in \mathbb{R}^K$ such that the price of each asset i is given as*

$$p_i = \sum_{j=0}^{K} \pi_j D_{ij};$$

or in matrix form

$$p = D\pi.$$

- *And π is also known as **state price**.*

Proof. (1)**Forward (if there is such nonnegative vector, then there is no arbitrage):** If there is a nonnegative vector $\pi \in \mathbb{R}^K$ such that the price of each asset i is given as

$$p_i = \sum_{j=0}^{K} \pi_j D_{ij}$$

Then for any portfolio θ, its initial value is given as

$$V_0 = \theta^T p = \sum_{i=1}^{N} \theta_i p_i = \sum_{i=1}^{N} \sum_{j=0}^{K} \theta_i \pi_j D_{ij} = \theta^T D\pi.$$

- For any final payoff $\theta^T D > 0$(positive payoff), its initial price $V_0 = \theta^T p = \theta^T D\pi > 0$. Therefore, the type I arbitrage opportunity will not exist.
- For any final payoff $\theta^T D \geq 0$(nonnegative payoff), its initial price $V_0 = \theta^T p = \theta^T D\pi \geq 0$. Therefore, the type II arbitrage opportunity will not exist.

(2) **Converse:** (use **??**)

- When there is no arbitrage type I, i.e. the situation that there exists some portfolio vector $\theta \in \mathbb{R}^N$ such that the payoff $V = D^T\theta \geq 0$ and $\theta^T p < 0$ cannot happen, then the situation that there exists some positive $\pi \in \mathbb{R}^K$ such that $D\pi = p$ will definitely happen.
- When there is no arbitrage type II, i.e. the situation that there exists some portfolio vector $\theta \in \mathbb{R}^N$ such that the payoff $V = D^T\theta > 0$ and $\theta^T p \leq 0$ cannot happen, then the situation that there exists some positive $\pi \in \mathbb{R}^K$ such that $D\pi = p$ will definitely happen.

□

Remark 15.3.2 (existence and uniqueness). Note that this theorem only guarantees existence but not uniqueness.

Note 15.3.1 (detecting arbitrage conditions). Given a payoff matrix D and the price vector p observed at the current market. There are following three possibilities:

1. There does not exist a π (no matter positive or negative) satisfying $p = D^T\pi$.
2. There exists a π satisfying $p = D^T\pi$, but π is not fully non-negative for all its components.
3. There exists one or infinite non-negative π satisfying $p = D\pi$.

Example 15.3.1. A simple application of the Arbitrage condition Given $D = \begin{pmatrix} 1+r & 1+r \\ s_{up} & s_{down} \end{pmatrix}$ and $S = (1, s)^T$. Using above theorem, we have $S = D\phi$, or

$$0 = ((1+r) - s_{up}/s)\phi_1 + ((1+r) - s_{down}/s)\phi_2$$

We requires $\phi_1, \phi_2 > 0$, then $s_{down}/s < 1 + r < s_{up}/s$

15.3.1.2 *No-arbitrage pricing*

Lemma 15.3.1 (no-arbitrage pricing using arbitrary state price vector). *Consider a single-period finite state market model ,consisting of N assets, represented by payoff matrix $D \in \mathbb{R}^{N\times K}$. Suppose the current prices of the N assets represented by a price vector $p \in \mathbb{R}^N$* ***is unknown.*** *Then the no-arbitrage prices for these assets are*

$$p = D^T\pi,$$

where π is ***arbitrary non-negative state vector****.*

Proof. From Theorem 15.3.1, we can see that p satisfying $p = D^T\pi$ will not give rise to arbitrage opportunities. □

Remark 15.3.3 (no-arbitrage price vs. instrinsic price)**.** As we discussed in Definition 15.1.1, no-arbitrage price is not necessarily the instrinsic price.

Definition 15.3.2 (replication portfolio). *Consider a market where there are freely trade asset B, and $A_1, A_2, ..., A_N$. The one-period payoff vectors of As are given by $V_1, V_2, ..., V_N \in$*

$\mathbb{R}^K$. *A portfolio* $\theta \in \mathbb{R}^N$ *in the asset As is a replicating portfolio of asset B if there exists a portfolio vector* $\theta = (\theta_1, \theta_2, ..., \theta_N)$ *such that*

$$V^B = \sum_{i=1}^{N} \theta_i V_i.$$

Theorem 15.3.2 (pricing via replication). *indexreplation pricing Suppose that* θ *is a replicating portfolio of B in the asset As. Denote the price vector of asset As by p. If the market is* ***arbitrage-free****, then* $t = 0$ *the no-arbitrage price of B is given as*

$$p^B = \sum_{i=1}^{n} \theta_i p_i$$

Proof. Because the replication portfolio has the same payoff as the asset B, the law of one price (Theorem 15.2.1) requires that they have the same price. □

Remark 15.3.4. The replication pricing procedure only relies on law of one price and does not require additional conditions like existence of risk neutral measure, market completeness.

Lemma 15.3.2 (replication indicates correlation, necessary condition for replication). *Let* X_B *be the random variable representing the payoff of B, let* $Y_1, ..., Y_N$ *be the random variables representing the payoff of assets* $A_1, ..., A_N$*. Suppose an asset B is replicated by the* ***mutually independent*** *assets* $A_1, ..., A_K$ *such that the payoff* X_B *of B is*

$$X_B = \sum_{i=1}^{N} \theta_i Y_i, \theta \in \mathbb{R}^K$$

Then if $\theta_i \neq 0$***, then*** X_B ***is correlated with*** Y_i***.***

Proof. WLOG, we assuming $\theta > 0$ for every component. Then

$$cov(Y_i, X_B) = cov(Y_i, \sum_{i=1}^{N} Y_i) = \theta cov(Y_i, Y_i) \geq 0$$

□

Remark 15.3.5 (implications).

- In general, only economically connected(or correlated) assets can be replicated. Different uncorrelated stocks, for example, can not be replicated. For example, derivatives can be replicated by using its underlying assets.

- **hedging via replicating** For every asset B sold, buy replicating portfolio θ such that the future payoff is neutral to the world state, that is, net gain = net loss = 0.

15.3.2 Risk-neutral measure and pricing

15.3.2.1 *Setup and the risk-neutral measure*

Setup of general finite-state market with risk-free asset

- A probability space (Ω, F, P) describing the world.
- A market in which N assets, labeled $A_1, A_2, ..., A_N$, are freely traded. **Assume that one of these, say A_1, is risk-free.**
- Uncertainty about the behavior of the market is encapsulated in a finite-sized sample space Ω of K possible market scenarios or sample points, labeled $\omega_1, \omega_2, ..., \omega_K$. There is an N by K payoff matrix $D \in \mathbb{R}^{N\times K}$ with entries $V_j(\omega_i)$ such that, in scenario i, the value of the asset A_j at after one-period is $V_j(\omega_i)$. The payoff matrix D is given by

$$\begin{pmatrix} V_1(\omega_1) & V_1(\omega_2) & \cdots & V_1(\omega_K) \\ V_2(\omega_1) & V_2(\omega_2) & \cdots & V_2(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ V_N(\omega_1) & V_N(\omega_2) & \cdots & V_N(\omega_K) \end{pmatrix} = \begin{pmatrix} 1+r & 1+r & \cdots & 1+r \\ V_2(\omega_1) & V_2(\omega_2) & \cdots & V_2(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ V_N(\omega_1) & V_N(\omega_2) & \cdots & V_N(\omega_K) \end{pmatrix},$$

  where r is the risk-free return.
- We denote a portfolio vector by $\theta \in \mathbb{R}^N$.
- The prices of the N assets at current time is represented by a price vector $p \in \mathbb{R}^N$. In particularly, $p_1 = 1$.

Theorem 15.3.3 (Fundamental theorem of asset pricing in a world with risk-free asset). *Consider a single-period finite state market model represented by payoff matrix $D \in \mathbb{R}^{N\times K}$. Further let the prices of N assets at current time be represented by a price vector $p \in \mathbb{R}^N$. It follows that*

- *The market model has no arbitrage opportunities if and only if there exists a nonnegative vector $\pi \in \mathbb{R}^K$ such that the price of each asset i is given as*

$$p_i = \frac{1}{1+r}\sum_{j=0}^{K} \pi_j D_{ij};$$

or in matrix form

$$p = \frac{1}{1+r} D\pi.$$

- π *is known as* ***risk-neutral measure*** Q *since* $\sum_{i=1}^{K} \pi_i = 1, \pi_i \geq 0 \forall i$.
- *Define expectation operator* $E_Q[V] : \mathbb{R}^K \to \mathbb{R}$ *as*

$$E_Q[V] = \sum_{i=1}^{K} \pi_i V(\omega_i),$$

then for the existing assets $V_1, ..., V_2$,

$$p_i = \frac{1}{1+r} E_Q[V_i].$$

- *For any other assets with payoff vector* $V_B \in \mathbb{R}^K$, *its no-arbitrage price is*

$$p_B = \frac{1}{1+r} E_Q[V_i]$$

Proof. Theorem 15.3.1 □

Remark 15.3.6 (risk-neutral measure has to be equivalent to original real-world measure).

- We are implicitly requiring that $\pi(\omega_i) \geq 0, \forall \omega \in \Omega$. Moreover, let the real probability be P, the risk-netural measure π, should satisfy $P(\omega) = 0 \Rightarrow \pi(\omega) = 0; \pi(\omega) = 0 \to P(\omega) = 0, \forall \omega \in \Omega$.

Remark 15.3.7 (existence and uniqueness).

1. Under no-arbitrage condition, the risk-neutral measure must exist. However, it might not be unique.

Remark 15.3.8 (Relation with state price in Theorem 15.3.1).

- We can also apply *Theorem* 15.3.1 to a market with risk-free market, let π be the state price, then $1 = e^{-r} \sum_i \pi_i e^r \implies \sum_i \pi_i = 1$.
- Therefore, we can understand as **risk-neutral measure is the discounted state price**.

15.3.2.2 *Computing risk-netural measure*

Remark 15.3.9 (how to compute risk-neutral measure). Suppose currently thare N actively traded assets and the market has no-arbitrage opportunties. By Theorem 15.3.3, there exists an risk-neutral measure. The risk-netrual measure π is about solving the following linear equation:

$$p = \frac{1}{1+r} D\pi.$$

Since the no-arbitrage condition ensures that the linear equation is consistent(or equivalently, p is in the column space of D), then we have the following situations:

- D has full column rank, then the risk-neutral measure is unique
- D has no full column rank(one situation is $K > N$), then there are infinitely many risk-netural measure.

Example 15.3.2 (Martingale measure for two state spot market). Consider a discrete stock price process consists of S_0 and S_T. We denote the discounted stock price as:

$$S_0^* = S_0, S_T^* = S_T/(1+r)$$

In the martingale measure, we have

$$S_0^* = E_{\mathbb{P}^*}[S_T^*]$$

In the case that the sample space of S^T has only two states, we have

$$S_0 = (1+r)^{-1}(Q(S^u)S^u + Q(S^d)S^d)$$

Solve it and we have

$$Q(S^u) = \frac{(1+r)S_0 - S^d}{S^u - S^d}$$

$$Q(S^d) = \frac{S^u - (1+r)S_0}{S^u - S^d}$$

Now a call option on this stock on the expiry date T will have

$$C_0(1+r) = \mathbb{E}_Q(C^T)$$

Example 15.3.3. [4, p. 10] Consider a market with only one risky asset and one money account. The current discounted price is 5. The one period discounted payoff of the risky asset are 4 and 6. To solve the risk neutral measure, we have

$$5 = 6Q(\omega_1) + 4Q(\omega_2), 1 = Q(\omega_1) + Q(\omega_2).$$

We can get

$$Q(\omega_1) = Q(\omega_2) = 1/2.$$

Example 15.3.4 (infinitely many risk-neutral measure). [4, p. 12] Consider a market with one risky asset and one money account. The current discounted price is 5. The one period discounted payoff of the risky asset are 3, 4 and 6. To solve the risk neutral measure, we have

$$5 = 6Q(\omega_1) + 4Q(\omega_2) + 3Q(\omega_3), 1 = Q(\omega_1) + Q(\omega_2) + Q(\omega_3).$$

We can get the solution

$$Q(\omega_1) = \lambda, Q(\omega_2) = 2 - 3\lambda, Q(\omega_3) = -1 + 2\lambda, 1/2 < \lambda < 2/3.$$

Example 15.3.5 (non-existence of risk-neutral measure). [4, p. 12] Consider a market with two risky asset and one money account. The current discounted price is 5 and 10. The one period discounted payoff of the risky asset are (6,6,4) and (12,8,8). To solve the risk neutral measure, we have

$$\begin{aligned} 5 &= 6Q(\omega_1) + 6Q(\omega_2) + 4Q(\omega_3) \\ 10 &= 12Q(\omega_1) + 8Q(\omega_2) + 8Q(\omega_3) \\ 1 &= Q(\omega_1) + Q(\omega_2) + Q(\omega_3) \end{aligned}$$

We can get the unique solution

$$Q(\omega_1) = 1/2, Q(\omega_2) = 0, Q(\omega_3) = 1/2.$$

We claim that there must exist an arbitrage opportunity.

15.3.3 Numeriares and other measures for pricing

15.3.3.1 *Pricing in arbitrage numeriare*

Setup of general finite-state market with risk-free asset

- A probability space (Ω, F, P) describing the world.
- A market in which N assets, labeled $A_1, A_2, ..., A_N$, are freely traded. **Assume that one of these, say A_1, is risk-free.**
- Uncertainty about the behavior of the market is encapsulated in a finite-sized sample space Ω of K possible market scenarios or sample points, labeled $\omega_1, \omega_2, ..., \omega_K$. There is an N by K payoff matrix $D \in \mathbb{R}^{N\times K}$ with entries $V_j(\omega_i)$

such that, in scenario i, the value of the asset A_j at after one-period is $V_j(\omega_i)$. The payoff matrix D is given by

$$\begin{pmatrix} V_1(\omega_1) & V_1(\omega_2) & \cdots & V_1(\omega_K) \\ V_2(\omega_1) & V_2(\omega_2) & \cdots & V_2(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ V_N(\omega_1) & V_N(\omega_2) & \cdots & V_N(\omega_K) \end{pmatrix} = \begin{pmatrix} 1+r & 1+r & \cdots & 1+r \\ V_2(\omega_1) & V_2(\omega_2) & \cdots & V_2(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ V_N(\omega_1) & V_N(\omega_2) & \cdots & V_N(\omega_K) \end{pmatrix},$$

where r is the risk-free return.

-
- We denote a portfolio vector by $\theta \in \mathbb{R}^N$
- The prices of the N assets at current time is represented by a price vector $p \in \mathbb{R}^N$. In particularly, $p_1 = 1$.

Theorem 15.3.4 (Fundamental theorem of asset pricing in a world with risk-free asset). *Consider a single-period finite state market model represented by payoff matrix $D \in \mathbb{R}^{N\times K}$. Further let the prices of the N assets at current time be represented by a price vector $p \in \mathbb{R}^N$. It follows that*

- *The market model has no arbitrage opportunities if and only if there exists a nonnegative vector $\pi \in \mathbb{R}^K$ such that the price of each asset i is given as*

$$p_i = \frac{1}{1+r}\sum_{j=0}^{K} \pi_j D_{ij};$$

or in matrix form

$$p = \frac{1}{1+r} D^T \pi.$$

- *Define π^B as*

$$\pi_j^B \triangleq \frac{\pi_j}{p_B} V_{B,j};$$

*π is known as **martingale measure** associated with numeriare B since $\sum_{i=1}^K \pi_i = 1$.*
- *Define expectation operator $E_B[V] : \mathbb{R}^K \to \mathbb{R}$ as*

$$E_B[V] = \sum_{i=1}^K \pi_i^B V(\omega_i),$$

then for the existing assets $V_1, ..., V_2$,

$$\frac{p_i}{p_B} = E_Q[\frac{V_i}{V_B}].$$

- *For any other assets with payoff vector $V_S \in \mathbb{R}^K$, its no-arbitrage price is*

$$p_B = \frac{1}{1+r} E_Q[V_i]$$

Proof.

$$p_i = \sum_{i=1}^{K} \pi_j D_{ij}$$
$$p_i = \sum_{i=1}^{K} \pi_j \frac{D_{ij}}{V_{B,j}} \cdot V_{B,j}$$

For the numeriare B, we have

$$p_B = \sum_{j=1}^{K} \pi_j \frac{D_{B,j}}{V_{B,j}} \cdot V_{B,j}$$

therefore

$$p_B = \sum_{j=1}^{K} \pi_j V_{B,j}.$$

Define $\pi_j^B \triangleq \frac{\pi_j}{p_B} V_{B,j}$, then $\sum_{j=1}^K \pi_j^B = 1$.

To pricing, we have

$$\begin{aligned} p_i &= \sum_{i=1}^{K} \pi_j D_{ij} \\ &= \sum_{i=1}^{K} p_B \pi_j^B \frac{D_{ij}}{V_{B,j}} \\ &= p_B E_B[\frac{D_i}{V_B}] \end{aligned}$$

□

Remark 15.3.10. Risk-neutral measure is the martingale measure associated with money account.

15.3.3.2 *Change of numeriare*

Theorem 15.3.5 (change of numeriare in the finite state market model).

15.3.4 Market completeness and pricing

15.3.4.1 *Market completeness*

Definition 15.3.3 (market completeness). *A finite-state market model is complete if an asset with arbitrary payoff vector $V^B \in \mathbb{R}^K$ can be replicated and price.*

Theorem 15.3.6 (Second fundamental theorem of asset pricing). *Consider a no-arbitrage finite-state market model. Suppose there exists a set of actively traded assets whose prices vector p and the associated payoff matrix D are known. The following are equivalent:*

- *the market is complete if and only if D has the row space of dimensionality of K (the number of linearly independent rows is K).*
- *the market is complete if and only if has it has a **unique** risk-neutral measure π such that*
$$p = \frac{1}{1+r} D\pi.$$

Proof. (1) If its row space is of full rank, then for an asset with payoff vector $V \in \mathbb{R}^K$ can be replicated and price because $V \in \mathcal{R}(D^T)$ (Theorem 15.3.2); therefore the linear equation solution must be unique.

(2) (a)When the row space dimensionality equals K, the matrix D must be full column rank(note that the column rank is the number of linerly independent columns, which equals the number of linearly independent rows, see **??**).(b) When the linear equation $p = \frac{1}{1+r} D\pi$ has a **unique** solution, D must have full column rank; because row rank equals column rank(**??**), we get (1). □

Remark 15.3.11 (quick-check necessary condition for completeness)**.** The number of assets N should be greater than the number of states K; otherwise the state space cannot be spanned(or more precisely, the row space dimensionality cannot be greater than K).

Example 15.3.6. Consider there are only two assets in the market, a bond B and a stock S. If there are three market scenarios in $T = 1$, then we cannot **uniquely price** a derivative given the price of B and S at time 0.

Example 15.3.7. Given the following one-period payoff matrix associated with a two-state market

$$D = \begin{bmatrix} 1 & 10.51.2 \end{bmatrix}.$$

D has two linearly independent rows, and the market is complete.

Example 15.3.8. Given the following one-period payoff matrix associated with a three-state market

$$D = \begin{bmatrix} 1 & 1 & 10.5 & 1.2 & 1.7 \end{bmatrix}.$$

D has two linearly independent rows, which is smaller than number of states; therefore the market is incomplete.

15.3.4.2 *Pricing in incomplete market*

Remark 15.3.12 (pricing in incomplete market)**.**

- A complete market gives a unique price for assets with all possible payoff vectors; however, the real-world market is often incomplete.
- In an incomplete market, we have the following pricing properties:
  - there are infinitely many risk-netural measures.
  - the no-arbitrage price for an asset with payoff vector $V \in \mathbb{R}^K$ is given by $p = \frac{1}{1+r}E_\pi[V]$.
  - for an asset that **can be replicated** by existing actively traded assets, its price is unique.
  - for an asset that **cannot be replicated** by existing actively traded assets, its price is not unique.

15.4 Multi-period finite-state market model

15.4.1 Market model setup

Definition 15.4.1 (multi-period finite-state market model).

- *A set of discrete time* $t = 0, 1, 2, ..., T$.
- *A finite sample space* $\Omega = \{\omega_1, ..., \omega_K\}$.
- *A probability measure* P *defined on* Ω *with* $P(\omega) > 0, \forall \omega \in \Omega$.
- *A filtration* $\{cF_t, t = 0, ..., T\}$.
- *A* ***stochastic money account process*** $B(t), t = 0, ..., T$, *where* $B(0) = 1$ *and* $B(t) > 0$ *almost surely.*
- N *risky stochastic security processes* $S_i(t), t = 0, 1, ..., T, i = 1, 2, ..., N$, *where* $S_i(t) > 0, \forall i, t$ *almost surely.*

15.4.2 Self-financing discrete-time trading strategies

Definition 15.4.2 (trading strategies).

- *A trading strategy* $h = (h_0, h_1, ..., h_N)$ *is a* ***vector of processes*** *where* $h_i = \{h_i(t), i = 1, 2, ..., T\}$ *representing the number of units of asset* i *carried from time* $t - 1$ *to time* t.
- $h_0(t)B(t-1)$ *represents the amount of money invested in the money market account from time* $t - 1$ *to time* t.
- $h_i(t)S_i(t-1)$ *represents the long position of* $h_i(t)$ *shares in asset* S_i *from time* $t - 1$ *to time* t.
- *A negative value of* $h_i(t)$ *representing borrowing from the bank or short positions.*

Definition 15.4.3 (value process and gain process).

- *The value process* $V = \{V(t), t = 0, 1, ..., T\}$ *associated with the trading strategy* h *and the market model consists of*
  - *initial value*
    $$V(0) = h_0(1)B(0) + \sum_{i=1}^{N} h_i(1)S_i(0),$$
  - *the value at time* $t = 1, 2, ..., T$
    $$V(t) = h_0(t)B(t) + \sum_{i=1}^{N} h_i(t)S_i(t).$$

- *The gain process* $G(t), t = 1, 2, ..., T$ *is defined as*

$$G(t) = \sum_{u=1}^{t} h_0(u)\Delta B(u) + \sum_{u=1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i(u),$$

where $\Delta B(u) = B(u) - B(u-1), \Delta S_i(u) = S_i(u) - S_i(u-1)$.
- *The single-period gain* $\Delta G(t)$ *is defined as*

$$\Delta G(t) = h_0(t)(B(t) - B(t-1)) + \sum_{i=1}^{N} h_i(t)(S_i(t) - S_i(t-1)),$$

such that $G(t) = \sum_{u=1}^{t} \Delta G(u)$.

Definition 15.4.4 (self-financing strategy). *A trading strategy h is called **self-financing** if for* $t = 1, 2, ..., T-1$,

$$V(t) = h_0(t+1)B(t) + \sum_{i=1}^{N} h_i(t+1)S_i(t),$$

where the LHS represents the time t value of the portfolio just before any transactions take place at that time, while the RHS represents the time t value of the portfolio right after any transactions (i.e. before the portfolio is carried forward to $t+1$*).*

Lemma 15.4.1 (criterion for self-financing strategy). *The following are equivalent:*

- *The trading strategy h is self-financing.*
- *The trading strategy h satisfies*

$$(h_0(t+1) - h_0(t))B(t) + \sum_{i=1}^{N}(h_i(t+1) - h_i(t))S_i(t) = 0, t = 0, 2, ..., T-1.$$

- *The trading strategy h satisfies*

$$V(t) = V(t-1) + \Delta G(t), t = 1, 2, ..., T.$$

- *The trading strategy h satisfies*

$$V(t) = V(0) + G(t), t = 1, 2, ..., T.$$

That is, for a self-financing strategy, any change in the portfolio's value is due to a gain or loss in the investments.

Proof. (1) equivalent to (2): Based on definition, we have

$$\begin{aligned} V(t) &= h_0(t+1)B(t) + \sum_{i=1}^{N} h_i(t+1)S_i(t) \\ &= h_0(t)B(t) + \sum_{i=1}^{N} h_i(t)S_i(t) \end{aligned}$$

Simple algebra can lead to the result. (2) equivalent (3)

$$\begin{aligned} V(t-1) + \Delta G(t) =& h_0(t-1)B(t-1) + \sum_{i=1}^{N} h_i(t-1)S_i(t-1) \\ & h_0(t)(B(t) - B(t-1) + \sum_{i=1}^{N} h_i(t)(S_i(t) - S_i(t-1)) \\ =& h_0(t)B(t-1) + \sum_{i=1}^{N} h_i(t)S_i(t-1) \\ & h_0(t)(B(t) - B(t-1) + \sum_{i=1}^{N} h_i(t)(S_i(t) - S_i(t-1)) \\ =& h_0(t)B(t) + \sum_{i=1}^{N} h_i(t)S_i(t) = V(t) \end{aligned}$$

where we use (2) in the second line. (3) equivalent (4): Use the definition $G(t) = \sum_{i=1}^{t} \Delta G$.

Note that for $t = 1$, we have

$$\begin{aligned} V(0) + \Delta G(1) &= h_0(1)B(0) + \sum_{i=1}^{N} h_i(1)S_i(0) + \\ & h_0(1)(B(1) - B(0)) + \sum_{i=1}^{N} h_i(1)(S_i(1) - S_i(0)) \\ &= h_0(1)B(1) + \sum_{i=1}^{N} h_i(1)S_i(1) \\ &= V(1) \end{aligned}$$

□

Definition 15.4.5 (disocunted asset price process).

- *The* ***discounted price process*** $S_i^*(t), t = 0, ..., T, i = 1, ..., N$ *is defined by*

$$S_i^*(t) = \frac{S_i(t)}{B(t)}, t = 0, 1, ..., T.$$

- *The **discounted value process** $V^*(t), t = 0, ..., T, i = 1, ..., N$ is defined by*

$$V^*(0) = h_0(1) + \sum_{i=1}^{N} h_i(1) S_i^*(0),$$

and

$$V^*(t) = h_0(t) + \sum_{i=1}^{N} h_i(t) S_i^*(t).$$

- *The **discounted gain process** $G^*(t), t = 1, ..., T, i = 1, ..., N$ is defined by*

$$G^*(t) = \sum_{u=1}^{t} \sum_{i=1}^{N} h_i(u) \Delta S_i^*(u),$$

- *A strategy h is self-financing if and only if*

$$V^*(t) = V^*(0) + G^*(t), t = 1, ..., T.$$

- *A strategy h is self-financing if and only if*

$$h_0(t+1) - h_0(t) + \sum_{i=1}^{N} (h_i(t+1) - h_i(t)) S_i^*(t)$$

15.4.3 First fundamental theorem of asset pricing

15.4.3.1 *Arbitrage and self-financing*

Definition 15.4.6 (arbitrage). *In the multi-period finite-state market model, an arbitrage opportunity exists if there is a self-financing strategy h whose value satisfying*

1. $V(0) = 0$;
2. $V(T) \geq 0$;
3. $Pr(V(T) > 0) > 0$.

More generally, the arbitrage condition can be

1. $V(0) = V_0$;
2. $V(T) \geq V_0$;
3. $Pr(V(T) > V_0) > 0$.

where V_0 is a deterministic constant.

Theorem 15.4.1 (arbitrage opportunity, change of numeraire, and change of measure).

- *A self-financing strategy h is an arbitrage opportunity under a probability measure P* ***if and only if*** *its discounted value V^* satisfying*
  1. $V^*(0) = 0$;
  2. $V^*(T) \geq 0$;
  3. $Pr(V^*(T) > 0) > 0$.
- *A self-financing strategy h is an arbitrage opportunity under a probability measure P* ***if and only if*** *h is an arbitrage opportunity under an* ***equivalent*** *probability measure Q.*

Proof. (1) Note that because $B(t) > 0$ almost surely, then

1. $V(0)/B(0) = 0 \Leftrightarrow V(0) = 0$;
2. $V(T)/B(T) \geq 0 \Leftrightarrow V(0) = 0$;
3. $Pr(V(T)/B(T) > 0) > 0 \Leftrightarrow Pr(V(T) > 0) > 0$ since $V(T)/B(T) > 0 \Leftrightarrow V(T) > 0$.

(2) Let h be an arbitrage opportunity under P. Let $\Omega' \subseteq \Omega$ be a set such that $V(T)(\omega) > 0, \omega \in \Omega'$. Then

$$P(\Omega') > 0 \Leftrightarrow Q(\Omega') > 0,$$

due to the equivalence of P and Q. □

Note 15.4.1 (implication).

- The first part of the theorem simply says that an arbitrage is always an arbitrage no matter which numeraire we are using.
- The second part of the theorem simply says that an arbitrage is always an arbitrage under all equivalent probability measure.
- This theorem gives us the latitude to find the most convenient numeriare and measure to pricing assets. As long as under the chosen measure and numeriare no arbitrage exists, then no arbitrage exists in the real world.

Lemma 15.4.2 (martingale value process is not an arbitrage). *Let $V(t)$ be a martingale under a probability measure P with $V(0) = V_0$, where V_0 is a deterministic constant. Then $V(t)$ is not an arbitrage under any equivalent probability measure.*

Proof. If $V'(T) \geq V_0, P(V(T) > V_0) > 0$, we cannot have $E_P[V(t)] = V_0$. Therefore V cannot be an arbitrage. □

15.4.3.2 *First fundamental theorem*

Lemma 15.4.3 (self-financing strategy of martingale price processes cannot be arbitrage). *Let $S_1(t), S_2(t), ..., S_N(t)$ be stochastic price process of assets.*

- *If $S_1, ..., S_N(t)$ are martingale under the probability measure P, then under any equivalent measure any self-financing strategy involving buying and selling these assets cannot be arbitrage.*
- *If the discounted price processes under some numeraire $B(t)$ such that $S_1 * (t), ..., S_N^*(t)$ $(S_i^*(t) = S_i(t)/B(t))$, then under any equivalent measure any self-financing strategy involving buying and selling these assets cannot be arbitrage.*

Proof. (1)We are trying to show that the value processes are martingales:

$$\begin{aligned}
E_P[V(0)+G(t)|\mathcal{F}_s] &= V(0) + E_P[\sum_{u=1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i(u)|\mathcal{F}_s] \\
&= V(0) + E_P[\sum_{u=1}^{s}\sum_{i=1}^{N} h_i(u)\Delta S_i(u)|\mathcal{F}_s] \\
&+ E_P[\sum_{u=s+1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i(u)|\mathcal{F}_s] \\
&= V(0) + E_P[\sum_{u=1}^{s}\sum_{i=1}^{N} h_i(u)\Delta S_i(u)|\mathcal{F}_s] \\
&= V(s)
\end{aligned}$$

where we use the fact that

$$E_P[\Delta S_i(u)|\mathcal{F}_s] = E_P[\Delta S_i(u+1) - S_i(u)|\mathcal{F}_s] = 0, \forall u \geq s.$$

We know that a martingale value process cannot be arbitrage under any equivalent meausre(Lemma 15.4.2). (2) From (1) we know that discounted value process $V^*(t)$ is a martingale that cannot satisfy

1. $V^*(0) = V_0^*$;
2. $V^*(T) \geq V_0^*$;
3. $Pr(V^*(T) > V_0^*) > 0$.

Using Theorem 15.4.1, it is easy to see that $V(t) = B(t)V^*(t)$ cannot satisfy

1. $V(0) = V_0 B(0)$;
2. $V(T) \geq V_0 B(T)$;
3. $P(V(T)B(T) > V_0 B(T)) > 0$.

therefore h is not an arbitrage under P or any equivalent measure. □

Theorem 15.4.2 (first fundamental theorem of asset pricing). *In the multi-period finite-state market model, no arbitrage opportunity exists if and only if there exists a probability measure Q, known as risk-neutral measure, with $Q(\omega) > 0, \forall \omega$, such that every discounted price $S_i^*(t), t = 0, 1, ..., T$ is a martingale under measure Q.*

Proof. (1) (martingale measure implies no arbitrage) (2) (no arbitrage implies existence of martingale)

□

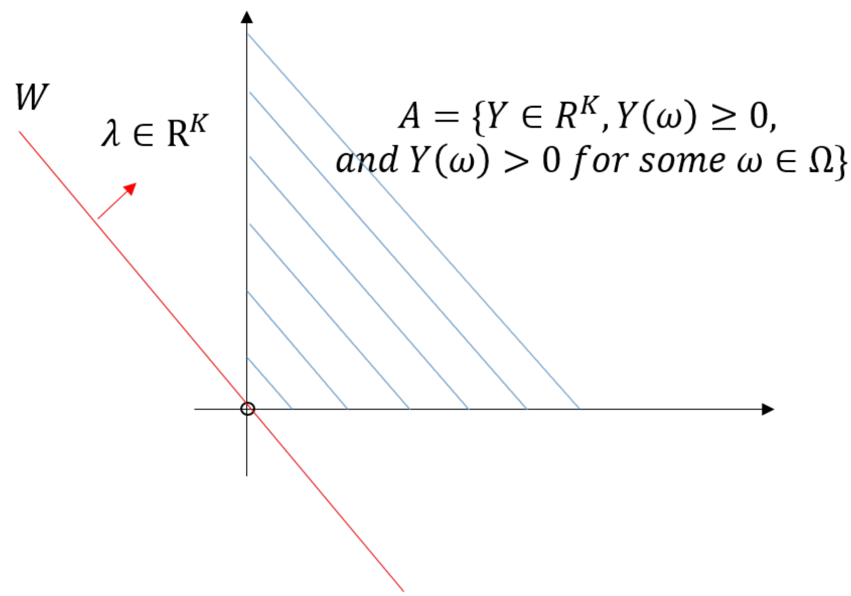


Figure 15.4.1

Remark 15.4.1.

Lemma 15.4.4 (arbitrage connection between multiperiod and single period market model).

- *A multi-period finite-state market model is free of arbitrage if and only if all its constituent single-period finite-state market model is free of arbitrage.*
- *Under the equivalent martingale measure Q associated with a multi-period finite-state market model, for any single period n,*

$$S^*(n) = E_Q[S^*(n+1)|\mathcal{F}_n].$$

Proof. (1)(a) If single-period market model has arbitrage self-financing strategy, we can simply use it in the multi-period market model to get arbitrage. (b) If the □

15.4.3.3 *Assets with dividends*

Theorem 15.4.3 (first fundamental law for assets with dividends). *Let $S_1(t), S_2(t), ..., S_N(t), t = 0, 1, ..., T$ be stochastic price process of assets. Let $D_1(t), D_2(t), ..., D_N(t), t = 0, 1, ..., T$ be the dividend payment of asset i at time t. Then the market model has no arbitrage opportunities if and only if there exists a probability measure Q, known as risk-neutral measure, with $Q(\omega) > 0, \forall \omega$, such that for every i and t,*

$$S_i^*(t) = E_Q[\sum_{s=t+1}^{\tau} \frac{D(s)}{B(s)} + S_i^*(\tau)|\mathcal{F}_t], \forall \tau > t.$$

Explicitly,

$$S_i^*(t) - \sum_{s=t+1}^{\tau} E_Q[\frac{D(s)}{B(s)}|\mathcal{F}_t] = E_Q[S_i^*(\tau)|\mathcal{F}_t].$$

Proof. From the results for non-dividend paying assets(Theorem 15.4.2,Lemma 15.4.4) and $t = \tau$, we have

$$\frac{S(\tau-1)}{B(\tau-1)} = E_Q[\frac{S(\tau)}{B(\tau)} + \frac{D(\tau)}{B(\tau)}|\mathcal{F}_{\tau-1}].$$

Continuing forward, we have

$$\begin{aligned}\frac{S(\tau-2)}{B(\tau-2)} &= E_Q[\frac{S(\tau-1)}{B(\tau-1)} + \frac{D(\tau-1)}{B(\tau-1)}|\mathcal{F}_{\tau-2}] \\ &= E_Q[E_Q[\frac{S(\tau)}{B(\tau)} + \frac{D(\tau)}{B(\tau)}|\mathcal{F}_{\tau-1}] + \frac{D(\tau-1)}{B(\tau-1)}|\mathcal{F}_{\tau-2}] \\ &= E_Q[\frac{S(\tau)}{B(\tau)} + \frac{D(\tau)}{B(\tau)} + \frac{D(\tau-1)}{B(\tau-1)}|\mathcal{F}_{\tau-2}]\end{aligned}$$

Continuing forward to t and we will get the result. □

15.4.3.4 *Application*

Lemma 15.4.5 (martingale properties of self-financing strategies under risk-neutral measure). *Assume no arbitrage opportunity exists in the multi-period finite-state market model. Under the risk-neutral measure Q,*

- *the **discounted gain process** of every self-financing strategy h is martingale.*
- *the **discounted value process** of every self-financing strategy h is martingale.*
- *every self-financing strategy h admits no arbitrage.*

Proof. (1)

$$\begin{aligned} E_Q[G^*(t)|\mathcal{F}_s] &= E_Q[\sum_{u=1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i^*(u)|\mathcal{F}_s] \\ &= E_Q[\sum_{u=1}^{s}\sum_{i=1}^{N} h_i(u)\Delta S_i^*(u)|\mathcal{F}_s] \\ &+ E_Q[\sum_{u=s+1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i^*(u)|\mathcal{F}_s] \\ &= E_Q[\sum_{u=1}^{s}\sum_{i=1}^{N} h_i(u)\Delta S_i^*(u)|\mathcal{F}_s] = G^*(s) \end{aligned}$$

where we use the fact that

$$E_Q[\Delta S_i^*(u)|\mathcal{F}_s] = E_Q[\Delta S_i^*(u+1) - S_i^*(u)|\mathcal{F}_s] = 0, \forall u \geq s.$$

Therefore $G^*(t)$ is a Q-martingale. (2) Note that $V^*(t) = V^*(s) + G^*(t) - G^*(s)$, we have

$$E_Q[V^*(t)|\mathcal{F}_s] = E_Q[V^*(s) + G^*(t) - G^*(s)|\mathcal{F}_s] = V^*(s),$$

where we use $E_Q[G^*(t) - G^*(s)|\mathcal{F}_s] = 0$ from (1). (3) Because (2), h is not an abitrage due to Theorem 15.4.1Lemma 15.4.3. (Lemma 15.4.3 shows that the discounted version V^* satisfies the no-arbitrage condition)

□

Theorem 15.4.4 (no-arbitrage price of any attainable claim). *Assume no arbitrage opportunity exists in the multi-period finite-state market model. Let h be a self-financing strategy replicate an attainable claim Y. Let $V(t)$ be the value process associated with h. Then the **no-arbitrage price process** Y is given by*

$$V_Y(t) = B(t)E_Q[\frac{Y}{B(T)}|\mathcal{F}_t].$$

Proof. Our goal is to prove that when the price process of all attainable claims are defined in this way, **no self-financing strategies involving buying and selling these attainable claims and other assets are arbitrage opportunities.**

Becuase the discounted price proccesses of Y and other assets are martingales, then no arbitrage exists due to Lemma 15.4.3. □

Note 15.4.2 (uniqueness, no-arbitrage price and intrinsic price)**.**

- In the theorem, we give one possible form of price process of attainable claims; we can only say these price processes admit no arbitrage; we cannot ensure its uniqueness.
- We should distinguish the no-arbitrage prices from intrinsic prices, which are the results of market forces equilibrium. See Definition 15.1.1 for more details.

Remark 15.4.2 (pitfalls in 'other possible forms' of no-arbitrage price)**.** It is tempting to define the price process of any attainable claim Y by

$$V(t) = E_P[Y|\mathcal{F}_t],$$

such that $V(t)$ is a martingale because

$$E_P[V(t)|\mathcal{F}_s] = E_P[E_P[Y|\mathcal{F}_t]|\mathcal{F}_s] = V(s).$$

However, such price is not arbitrage free. Let h be the self-financing strategy such that $V_h(T) = Y$. In general,

$$V_h(0) = B(0)E_Q[V_h(T)/B(T)] \neq V_Y(0) = E_P[Y]$$

which violates law of one price and introduce arbitrage opportunities.

15.4.3.5 *Forwards*

Theorem 15.4.5 (forward price under risk-neutral measure). *Consider a security price process S with its dividend process D. Let $F(t,\tau)$ be the forward price process with delivery date τ.*

- *Then under risk-neutral measure Q,*

$$F(t,\tau) = E_Q[\frac{S(\tau)B(t)}{B(\tau)}|\mathcal{F}_t]/DF(t,\tau),$$

where

$$DF(t,\tau) = E_Q[\frac{B(t)}{B(\tau)}|\mathcal{F}_t].$$

- *Under the equivalent martingale measure Q_τ associated with numeraire $DF(t,\tau)$, we have*

$$F(t,\tau) = E_\tau[S(\tau)|\mathcal{F}_t] = E_\tau[F(\tau,\tau)|\mathcal{F}_t],$$

and

$$E_\tau[F(t,\tau)] = F(0,\tau).$$

That is $F(t,\tau)$ is the conditional expectation process of $F(\tau,\tau)$ under measure Q_τ

- *If the bank account process $B(t)$ is deterministic, then*

$$F(t,\tau) = E_Q[S(\tau)|\mathcal{F}_t] = \frac{B(\tau)}{B(t)}S_i(t) - \sum_{s=t+1}^{\tau} D(s)\frac{B(\tau)}{B(s)}|\mathcal{F}_t].$$

- *If the bank account process $B(t)$ is deterministic, then $F(t,\tau)$ is a martingale and*

$$F(t,\tau) = E_Q[S(\tau)|\mathcal{F}_t] = E_Q[F(\tau,\tau)|\mathcal{F}_t],$$

and

$$E_Q[F(t,\tau)] = E_Q[S(\tau)] = F(0,\tau),$$

That is $F(t,\tau)$ is the conditional expectation process of $F(\tau,\tau)$.

Proof. (1) Because the time t value of the forward contract is zero, using pricing formula Theorem 15.4.4, we have

$$E_Q[\frac{B(t)}{B(\tau)}(S(\tau) - F(t,\tau)|\mathcal{F}_t] = 0,$$

which gives

$$F(t)E_Q[\frac{B(t)}{B(\tau)}|\mathcal{F}_t] = E_Q[\frac{B(t)}{B(\tau)}S(\tau)|\mathcal{F}_t].$$

(2) We use the dynamics of $S(t)$ under risk-netrual measure given in Theorem 15.4.3. (3) We use the boundary condition $S(\tau) = F(\tau,\tau)$ and conditional expectation process is martingale(Theorem 15.6.3). □

Theorem 15.4.6 (value of forward contract after initiation). *Let $F(t,T)$ be the forward price of an asset at time t with delivery time T. Consider a forward/futures contract initiated at t_0 with forward price $F(t_0,T)$. Let $V(t)$ be the value of the contract at time t. Denote $DF(t,\tau) = E_Q[\frac{B(t)}{B(\tau)}|\mathcal{F}_t]$.*

- *The **stochastic** value of the forward contract at time t_0+1 is*

$$V(t_0+1) = (F(t_0+1,T) - F(t_0,T))DF(t+1,T)$$

- *The **stochastic** value of the forward contract at time t is*

$$V(t) = (F(t,T) - F(t_0,T))DF(t,T)$$

Proof. (1) Using pricing formula Theorem 15.4.4 for the value $t_0 + 1$, we have

$$\begin{aligned} V(t_0+1) &= E_Q[\frac{B(t_0+1)}{B(T)}(S(T) - F(t_0, T)|\mathcal{F}_{t+1}] \\ &= E_Q[\frac{B(t_0+1)}{B(T)}S(T)|\mathcal{F}_{t+1}] - F(t_0, T)DF(t_0+1, T) \\ &= F(t_0+1, F)DF(t_0+1, T) - F(t_0, T)DF(t_0+1, T) \end{aligned}$$

(2) We can similarly prove other cases. □

15.4.3.6 *Futures*

Theorem 15.4.7 (futures price under risk-neutral measure). *Consider a security price process S with its dividend process D. Let $Fur(t, \tau)$ be the further price process with delivery date τ. Assume marking to market at every discrete time step.*

- *Then under risk-neutral measure Q,*

$$E_Q[\frac{Fur(t,\tau) - Fur(t-1,\tau)}{B(t)}|\mathcal{F}_{t-1}], t = 1, ..., T.$$

- *If $B(t)$ is predictable, i.e., $B(t)$ is measurable with respect to $\mathcal{F}_{t-1}$, then $Fur(t, \tau)$ is a martingale and given by*

$$Fur(t,\tau) = E_Q[S(\tau)|\mathcal{F}_t] = E_Q[Fur(\tau,\tau)|\mathcal{F}_t],$$

  and

$$E_Q[Fur(t,\tau)] = Fur(0,\tau).$$

- *If $B(t)$ is deterministic,*

$$F(t,\tau) = Fur(t,\tau),$$

  that is, forward and futures price process are equivalent.

Proof. (1) Because of the marking to market convention, the time t payoff or value of the futures contract is $Fur(t, \tau) - Fur(t-1, \tau)$. Using pricing formula Theorem 15.4.4, we have

$$V(t-1) = B(t-1)E_Q[\frac{Fur(t,\tau) - Fur(t-1,\tau)}{B(t)}|\mathcal{F}_{t-1}], t = 1, ..., T.$$

Setting $V(t-1) = 0$ due to marking to market, we get the result.

$$F(t)E_Q[\frac{B(t)}{B(\tau)}|\mathcal{F}_t] = E_Q[\frac{B(t)}{B(\tau)}S(\tau)|\mathcal{F}_t].$$

(2) When $B(t)$ is predictable, we have

$$0 = \frac{1}{B(t)} E_Q[Fur(t,\tau) - Fur(t-1,\tau)|\mathcal{F}_{t-1}] = 0.$$

That is,

$$E_Q[Fur(t,\tau)|\mathcal{F}_{t-1}] = Fur(t-1,\tau)$$

Therefore,

$$\begin{aligned} Fur(t,\tau) &= E_Q[Fur(t+1,\tau)|\mathcal{F}_t] \\ &= E_Q[E_Q[Fur(t+2,\tau)|\mathcal{F}_{t+1}]|\mathcal{F}_t] \\ &= E_Q[Fur(t+2,\tau)|\mathcal{F}_t] \\ &\cdots \\ &= E_Q[Fur(\tau,\tau)|\mathcal{F}_t] \end{aligned}$$

(3) When $B(t)$ is deterministic, from Theorem 15.4.5, we have

$$F(t,\tau) = Fur(t,\tau) = E_Q[S(\tau)|\mathcal{F}_t].$$

□

15.4.4 Complete market and second fundamental theorem

15.4.4.1 *Complete market*

Definition 15.4.7 (complete market). *A market is said to be* ***complete*** *if every claim in the market is attainable.*

15.4.4.2 *Second fundamental theorem*

15.4.4.3 *Complete vs. incomplete market*

Note 15.4.3 (pricing in complete vs. incomplete market).

- If the model is complete then we know how to price all the contingent claims.
- If the model is incomplete, then we **only know how to price some of the contingent claims**, namely, all the attainable ones. Let $\mathcal{M}$ denote the set of all the risk-neutral measures. Then

$$p = B(0)E_Q[Y(T)/B(T)], \forall Q \in \mathcal{M},$$

that is, we **get the same price under all risk-neutral measures.**
- For the pricing of non-attainable claims in the incomplete market, we use optimization method to provide lower and upper bounds.

Definition 15.4.8 (no-arbitrage price range for non-attainable claims). *Assume the market has no arbitrage but is incomplete.*

- *The upper bound price is*

$$V_+(X) \triangleq \inf\{B_0 E_Q[Y/B_T] : Y \geq X, Y \text{ attainable}\}.$$

- *The lower bound price is*

$$V_-(X) \triangleq \sup\{B_0 E_Q[Y/B_T] : Y \leq X, Y \text{ attainable}\}.$$

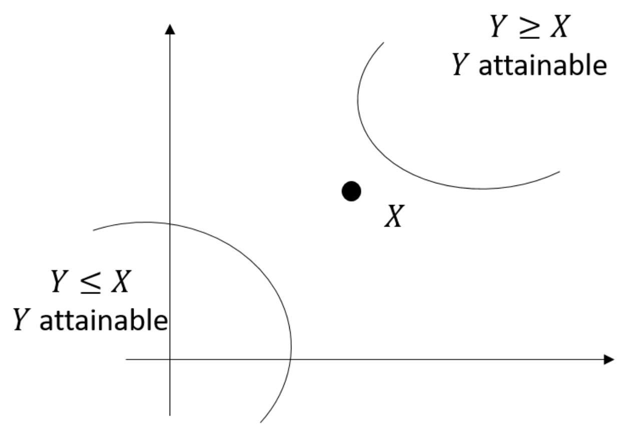


Figure 15.4.2: An illustration of the upper and lower bound price for non-attainable claims.

Remark 15.4.3 (arbitrage justification for lower and upper bound prices)**.**

- If X could be sold for a greater amount p than $V^+(X)$, then we should short X, and use the proceeds to buy Y, which has better payoff than X at time T, i.e., $Y(T) \geq X(T)$. Then at time T sell Y and buy X to cover the short position.
- If X could be bought for a smaller amount p than $V^-(X)$, then we should buy X and short Y. X has better payoff than Y at time T, i.e., $X(T) \geq Y(T)$. Then at time T sell X and buy Y to cover the short position.

Definition 15.4.9 (no-arbitrage price range for non-attainable claims, alternative). *Assume the market has no arbitrage but is incomplete. Let $\mathcal{M}$ denote the set of all the risk-neutral measures.*

- *The upper bound price is*

$$V_+(X) \triangleq \sup\{B_0 E_Q[X/B_T] : Q \in \mathcal{M}\}.$$

- *The lower bound price is*

$$V_-(X) \triangleq \inf\{B_0 E_Q[X/B_T] : Q \in \mathcal{M}\}.$$

- *In particular, if X is attainable, then $V_+(X) = V_-(X)$.*

Note 15.4.4 (equivalence of the two definitions). [4, p. 26]

- Note that the two definitions can be showed equivalent using linear programing duality theorem.
- The alternative definition is easier for calculating lower and upper bound prices.

Example 15.4.1. Consider a contingent claim X with payoff $(30, 20, 10)$. Consider the set of risk-neutral measure are parameterized by $Q = (q, 2-3q, -1+2q)$, where $1/2 < q < 2/3$. Assume zero interest rate. We have

- $E_Q[X] = 30q + 20(2-3q) + 10(-1+2q) = 30 - 10q$.
- $$V_+(X) = \sup_q 30 - 10q = 25.$$
- $$V_-(X) = \inf_q 30 - 10q = 70/3.$$

15.4.5 Extension to infinite states

15.5 Continuous-time market model I: hedging

Assumptions for this section:

- All assets are trade-able
- The risk free rate r are deterministic
- No transaction cost.
- The underlying asset pays no dividends.

- Short selling is permitted.
- The trading of assets can take place continuously in time and amount.

15.5.1 Hedging and market price of risks

15.5.1.1 *Single source of uncertainty*

Lemma 15.5.1 (hedging one source of risk, constraints on parameters of dynamics with common source of risks). *[5, p. 656][6, p. 55] Consider a market with risk-free asset with short rate r. Consider values of two assets as stochastic processes given as:*

$$\frac{df_1}{f_1} = \mu_1 dt + \sigma_1 dz$$

and

$$\frac{df_2}{f_2} = \mu_2 dt + \sigma_2 dz$$

where z is the Wiener process,i.e., the only ***common source of uncertainty****. Under no-arbitrage condition, there exist a λ(called market price of risk) such that*

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2} = \lambda$$

And the stochastic processes f_1 and f_2 will be given as

$$\frac{df_1}{f_1} = (r + \sigma_1\lambda)dt + \sigma_1 dz$$

and

$$\frac{df_2}{f_2} = (r + \sigma_2\lambda)dt + \sigma_2 dz$$

Proof. If we construct the portfolio

$$\Pi = k_1 f_1 + k_2 f_2$$

where $k_1 = \sigma_2 f_2$ and $k_2 = -\sigma_1 f_1$. Then we have

$$d\Pi = k_1 df_1 - k_2 df_2 = (\mu_1\sigma_2 f_1 f_2 - \mu_2\sigma_1 f_1 f_2)dt = r\Pi dt.$$

That is, the portfolio Π is deterministic, and it must have growth rate equal to risk-free rate as required by no-arbitrary and law of one price(Corollary 15.2.1.2). Then, we have

$$\mu_1\sigma_2 + \mu_2\sigma_1 = \sigma_2 r + \sigma_1 r.$$

□

Remark 15.5.1 (interpretation).

- Assets have the common source of uncertainty have dependent growth rate. Because of the common uncertainty, they are correlated and they can be used to hedging each other to reduce risks.
- **If the asset prices do not evolve like this, there exists arbitrage opportunities**. That is, risk-free portfolio does not grow at risk-free rate.
- Also see arbitrage pricing theory in portfolio optimization setup Lemma 11.1.2.

Note 15.5.1 (constraints on parameters of dynamics with common source of risks). Under the no-arbitrage assumption, law of one price(Theorem 15.2.1), (and assumptions about continuous trading, short selling, frictionless market etc,), two Ito processes with single common source of uncertainty have constraints on the drift parameters they have. This lemma says that, they have to satisfy

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2} = \lambda.$$

From mathematical point of view, any $\lambda \in \mathbb{R}$ is reasonable(no arbitrage exists); from economics point of view, any $\lambda > 0$ is reasonable, the large λ reflects investors are more risk-averse and require higher growth rate.

Example 15.5.1. Assume that there are two financial assets, both are dependent on a 90 day interest rate. The first instrument has an expected return 6% per year and a volatility of 20%. For the second instrument a volatility of 30% per year is assumed. Further more, the risk-free rate is 3% per year. The market price of risk for the first instrument is

$$\frac{0.06 - 0.03}{0.2} = 0.15.$$

Then for the second instrument, we estimate its expected return is

$$0.15 \times 0.3 + 0.03 = 0.075.$$

Note 15.5.2 (Derivation of Black-Scholes model and martingale pricing, alternative). indexBlack-Scholes equation Let $V(S(t), t)$ be the value of the derivative as a function of the asset price $S(t)$ and time t. Assume $S(t)$ is governed by

$$dS(t) = \mu S(t)dt + \sigma S(t)dW_t$$

where W_t is the Brownian motion. Then $V(S_t, t)$ is governed by

$$dV = \sigma S \frac{\partial V}{\partial S} dW_t + (\mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t})dt$$

via Ito lemma(Lemma 6.3.1)

Then from Lemma 15.5.1, there exists a λ such that

$$\mu = r + \lambda\sigma$$

$$(\mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t})/V = r + \lambda\sigma S \frac{\partial V}{\partial S}/V$$

Eliminate λ, we get the Black-Scholes equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0.$$

If we change our measure to a new measure Q(risk-neutral measure) such that $\lambda = 0$, then

$$dS(t) = rS(t)dt + \sigma S(t)dW_t$$
$$dV = rVdt + \sigma S \frac{\partial V}{\partial S} dW_t$$

That is, any derivative V will have drift rV no matter its functional form(assume V is a 'nice' function). **Therefore,**

$$e^{-rt}V_t$$

is a martingale, and justify the martingale pricing method

$$V(t) = E_Q[V(T)|\mathcal{F}_t].$$

Note 15.5.3 (When risks cannot be hedged). **Assume** an asset S governed by

$$dS = \mu S dt + \sigma S dW$$

where W is the Brownian motion. Assume the risk associated with W cannot be eliminated(for example, due to the lack of derivatives and assets sharing the same risks).

- If we can estimate μ and σ from the market data, then we can determine market price of risk $\lambda = (\mu - r)/\sigma$.
- In this single asset with unhedgeable risk, any μ value is reasonable, since the model itself will not admit arbitrage. However, for multiple assets with common risks, the drift parameters are constrained if no arbitrage is allowed(Lemma 15.5.1).

Lemma 15.5.2 (hedging one source of risk, constraints on parameters of dynamics with common source of risks, assets with dividends). *[5, p. 656][6, p. 55] Consider a market with risk-free asset with short rate r. Consider values of two assets as stochastic processes given as:*

$$\frac{df_1}{f_1} = \mu_1 dt + \sigma_1 dz$$

and

$$\frac{df_2}{f_2} = \mu_2 dt + \sigma_2 dz$$

where dz is the Wiener process,i.e., the only ***common source of uncertainty****. Further assume that the asset is paying continuous dividends with ratio q_1 and q_2. Under no-arbitrage condition, there exist a λ(called market price of risk) such that*

$$\frac{\mu_1 + q_1 - r}{\sigma_1} = \frac{\mu_2 + q_2 - r}{\sigma_2} = \lambda$$

And the stochastic processes f_1 and f_2 will be given as

$$\frac{df_1}{f_1} = (r - q_1 + \sigma_1 \lambda)dt + \sigma_1 dz$$

and

$$\frac{df_2}{f_2} = (r - q_2 + \sigma_2 \lambda)dt + \sigma_2 dz$$

Proof. (1) The most common proof is that we form a new security f^*, which the same as f except that all income produced by f is reinvested in f. f^* is related to f via

$$f^* = fe^{qt}, df^*/f^* = (\mu + q)dt + \sigma dz.$$

Then we can use analog of Lemma 15.5.1. (2) The most common proof above relies on a special portfolio. Now we can prove most general case: suppose we divide the dividend for two usage, ϕ for reinvestment and $q - \phi$ for money account m, then the total wealth $g = f^* + m$ at time t has dynamics:

$$dg = df^* + dm, df^* = (\mu + \phi q)f^* dt + \sigma f^* dz, dm = rmdt + f(1 - \phi)dt.$$

That is:

$$dg = (\mu + q)f^* + rmdt + \sigma f^* dz$$

From Lemma 15.5.1, we have

$$(\mu + q)f^* + rm = r(m + f^*) + \lambda\sigma f^* \implies (\mu + q - r) = \lambda\sigma.$$

□

Lemma 15.5.3 (linearity properties in no-arbitrage condition). *Consider a market with risk-free asset with short rate r. Consider values of two assets under* ***no arbitrage condition*** *having stochastic processes given by:*

$$\frac{df_1}{f_1} = r + \lambda\sigma_1 dt + \sigma_1 dz$$

and

$$\frac{df_2}{f_2} = r + \lambda\sigma_2 dt + \sigma_2 dz$$

where dz is the Wiener process,i.e., the only ***common source of uncertainty*** *and λ is the market price of risk. Then for arbitrary constant α and β,*

$$\frac{d(\alpha f_1 + \beta f_2)}{\alpha f_1 + \beta f_2} = (r + \lambda\sigma)dt + \sigma dz,$$

where

$$\sigma = \frac{\alpha\sigma_1 f_1 + \beta\sigma_2 f_2}{\alpha f_1 + \beta f_2}.$$

That is, the new asset $\alpha f_1 + \beta f_2$ also satisfies the no arbitrage condition.

Proof. We have

$$\begin{aligned} d\alpha f_1 &= (r + \lambda\sigma_1)\alpha f_1 dt + \sigma_1 \alpha f_1 dz \\ d\beta f_2 &= (r + \lambda\sigma_2)\beta f_2 dt + \sigma_2 \beta f_2 dz \end{aligned}$$

Add together and factor out the common factor $\alpha f_1 + \beta f_2$ can prove the result. □

Remark 15.5.2. This is the restatement of the linear pricing theorem(Theorem 15.2.2).

15.5.1.2 *Multiple sources of uncertainty*

Theorem 15.5.1 (hedging multiple sources of risks, constraints on parameters of dynamics). *[5, p. 659][6, p. 55] Consider a market with risk-free asset with short rate r. Consider n assets as stochastic processes given as:*

$$\frac{df_i}{f_i} = a_i dt + \sum_{j=1}^{m} b_{ij} dz_j, i = 1, 2, ..., n.$$

where $z_j, j = 1, 2, ..., m$ are independent Wiener processes. If there is ***no arbitrage****, then there exists (unnecessarily unique) $\lambda_1, ..., \lambda_m$*

$$a_i - r = \sum_{j=1}^{m} b_{ij}\lambda_j, j = 1, 2..., m.$$

Proof. Denote $A = (a_1, ..., a_n)^T$, $B_{ij} = b_{ij}$, and a portfolio vector $k = (k_1, ..., k_n)^T$. To construct a risk-free portfolio, we have

$$k^T B = 0.$$

This risk-free portfolio should grow at risk-free rate r_f, we have

$$k^T A = k^T r_f.$$

Combine $k^T B = 0, k^T(A - r_f) = 0$, we know that $A - r_f = B\lambda$. (from $k^T B = 0$, we know that $k \in \mathcal{N}(B)$; from $k^T(A - r_f)$, we know that $(A - r_f) \in \mathcal{N}(B)^\perp = \mathcal{R}(B)$). Also see Lemma 11.1.3. □

Remark 15.5.3 (trivial solution for portfolio vector k). In our proof, if B is a tall and thin matrix(no fewer sources of uncertainty than number of assets, for example, 2 stocks with 2 sources of uncertainty; note that when there are more stocks than uncertainty sources, it is guaranteed to have infinitely many solutions for k. see **??**), then k might only have the trivial solution of $k = 0$. However, this does not undermine validity of our conclusion that $(A - r_f) \in \mathcal{N}(B)^\perp = \mathcal{R}(B)$.

Note 15.5.4 (existence and uniqueness of λ, market completeness). We know that the A, B, λ are connected by

$$B\lambda = (A - r_f).$$

Assume A, r_f and B are known, the solutions to λ has the following situations(**??**)

- The linear equation is **inconsistent**, then λ **does not exist. Then there exists arbitrage in the model.**
- The linear equation has one unique solution, then the market is complete and the price is unique under no-arbitrage condition.

- The linear equation has infinitely many solutions, then the market is incomplete and the price is not unique under no-arbitrage condition. The non-unique prices means that there exists other sets of model parameters($A, B, A = r_f + B\lambda$) such that no arbitrage can happen. The unique price situation means that only current model parameters can satisfy the no-arbitrage condition.
- The non-uniqueness situation usually occur when $m \geq n$; i.e. not all the uncertainty can be eliminated. The most simple case is that there is only 1 risky asset in market, then its drift parameter can be any value and the model will not admit arbitrage.

Remark 15.5.4 (interpretation of λ). If $\lambda_i \sigma_i > 0$, then the risk associated with z_i cannot be eliminated, and the investor require a positive excess return for the risk associated with z_i;If $\lambda_i \sigma_i = 0$, then the risk associated with z_i can be eliminated, and the investor would not require compensation for the risk associated with z_i; If $\lambda_i \sigma_i < 0$, then the investor is willing to accept lower return since the risk associated with z_i can help them reduce total risks;

Remark 15.5.5 (market price of risk in CAPM).

- In the CAPM model, where we can only hedge non-systematic risk by portfolio combination, the market price of the risk is
$$\frac{E[r_m] - r_f}{\sigma_m},$$
and the return of **efficient** portfolio is
$$\frac{E[r] - r_f}{\sigma} = \frac{E[r_m] - r_f}{\sigma_m}$$
Investors can only get compensated by taking systematic risks.
- More generally, consider a world of multiple assets described by SDE with multiple sources of uncertainties. **If there exists a portfolio having no risk, then its growth rate must be risk-free rate**.
- Every market price of risk is corresponding to some probability measure. Some value of the market price of risk corresponds to the real world probability measure and the growth rates of security prices that are observed in practice.

Theorem 15.5.2 (hedging multiple sources of risks, constraints on parameters of dynamics, assets with dividends). *[5, p. 659][6, p. 55] Consider a market with risk-free asset with short rate r. Consider n assets as stochastic processes given as:*
$$\frac{df_i}{f_i} = a_j dt + \sum_{i=1}^{n} b_{ij} dz_i, j = 1, 2, ..., n.$$

where $z_i, i = 1, 2, ..., m$ are independent Wiener processes. ***Further assume each asset has a continuous dividend*** *$q_i, i = 1, ..., m$. If there is* ***no arbitrage****, then there exists (unnecessarily unique) $\lambda_1, ..., \lambda_m$*

$$a_i + q_i - r = \sum_{j=1}^{m} \lambda_i b_{ij}, j = 1, 2..., m.$$

Proof. See Lemma 15.5.2. □

15.5.1.3 *Case study on market risk*

Theorem 15.5.3 (Law of one price on market risks). *[7, p. 25] Identify unavoidable risks should have identical expected returns.*

Lemma 15.5.4 (asset dynamics in a world with a few uncorrelated stocks and a riskless bond). *Consider a market with risk-free asset with short rate r. Consider n assets characterized by stochastic processes given as:*

$$\frac{df_i}{f_i} = a_i dt + b_i dz_i, j = 1, 2, ..., n.$$

where $z_i, i = 1, 2, ..., n$ are independent Wiener processes. If there is ***no arbitrage****, then there exists (unnecessarily unique) $\lambda > 0$ such that*

$$a_i - r = b_i \lambda, \forall i.$$

Proof. From Theorem 15.5.1, we know that there exists $\lambda_1, \lambda_2, ..., \lambda_n$ such that

$$a_i - r = b_i \lambda_i, \forall i.$$

The law of one price on market risk(Theorem 15.5.3) requires these λ_i should be equal to each other. If one asset offer a smaller growth rate at the same risk level, the investors will view this asset is undervalued, then the buying force will push up the growth rate.

We further requires $\lambda > 0$ because investors are risk averse(no one is willing to take risks unless being rewarded.) □

Lemma 15.5.5 (asset dynamics in a world with infinitely many uncorrelated stocks and a riskless bond). *Consider a market with risk-free asset with short rate r. Consider infinitely many assets characterized by stochastic processes given by:*

$$\frac{df_i}{f_i} = a_i dt + b_i dz_i, j = 1, 2, ..., n.$$

where $z_i, i = 1, 2, ..., n$ are independent Wiener processes. If there is ***no arbitrage****, then*

$$a_i - r = 0, \forall i.$$

That is, the individual asset dynamics is

$$\frac{df_i}{f_i} = rdt + b_i dz_i, i = 1, 2, ..., n.$$

Or equivalently, ***the market price of idiosyncratic risk is zero.***

Proof. Via linear combination, we can create a portfolio consisting of all the assets such that the risk is zero(due to central limit theorem). This portfolio will have zero risk and therefore requires $\lambda = 0$. Here is the formal proof.

We know that

$$df_i = (r + \lambda b_i) f_i dt + b_i f_i dz_i, b_i > 0$$

then for this special portfolio, we have

$$\begin{aligned} d(\sum_{i=1}^{N} f_i) &= (r + \lambda \frac{\sum_{i=1}^{N} b_i f_i}{\sum_{[i=1]}^{N} f_i}) \sum_{[i=1]}^{N} f_i dt + \sum_{[i=1]}^{N} b_i f_i dz_i \\ &= (r + \lambda \frac{\sum_{i=1}^{N} b_i f_i}{\sum_{[i=1]}^{N} f_i}) \sum_{[i=1]}^{N} f_i dt + \frac{\sum_{i=1}^{N} b_i^2 f_i^2}{\sum_{i=1}^{N} f_i} \sum_{i=1}^{N} f_i dZ \end{aligned}$$

We can show that (assume $\{b_i\}, \{f_i\}$ is unformly bounded away from 0 and ∞)

$$\frac{\sum_{i=1}^{N} b_i^2 f_i^2}{\sum_{i=1}^{N} f_i} \leq \frac{\max\{b_i f_i\}\sqrt{N}}{\min\{f_i\} N} \to 0, \text{ as } N \to \infty$$

and

$$\frac{\sum_{i=1}^{N} b_i f_i}{\sum_{[i=1]}^{N} f_i} \geq \min\{b_i\} > 0.$$

Therefore, as $N \to \infty$, the portfolio has no risk and its growth rate needs to be r. Therefore, $\lambda = 0$. □

Remark 15.5.6 (intuition and interpretation).

- Suppose $\lambda > 0$, then we will just buy the market portfolio, which has no risk but growth rate greater than bond. Then we will borrow infinite money to buy this portfolio and push up r.
- This example clearly shows hedgeable risks will not be rewarded.

Remark 15.5.7 (compared with finite asset case). The finite asset case has a $\lambda > 0$ depending on N(which determines how much can be diversified) and investors risk aversion.

Lemma 15.5.6 (asset dynamics in a world with infinitely many all simultaneously correlated with the entire market, and a riskless bond). *Consider a market with risk-free asset with short rate r. Consider infinitely many assets characterized by stochastic processes given by:*

$$\frac{df_i}{f_i} = a_i dt + \rho_i b_i dW + \sqrt{1-\rho_i} b_i dz_i, b_i > 0 \forall i = 1,2,...,n.$$

where $z_i, i = 1,2,...,n$, and W are independent Wiener processes. Assume there is no arbitrage opportunities in the market. Suppose we form a market portfolio $f_M = \sum_{i=1}^{\infty} f_i$.

It follows that

- *the market portfolio has dynamics*

$$\frac{df_M}{f_M} = (r + \lambda b_M)dt + b_M dW, b_M = \frac{\sum_{i=1}^{\infty} b_i f_i}{\sum_{i=1}^{\infty} f_i}.$$

- *the individual asset has dynamics*

$$\frac{df_i}{f_i} = (r + \lambda \rho_i b_i)dt + \rho_i b_i dW + \sqrt{1-\rho_i} dz_i, b_M = \frac{\sum_{i=1}^{\infty} \rho_i b_i f_i}{\sum_{i=1}^{\infty} f_i}.$$

Proof. From above two lemmas, we can assume the individual asset dynamics is given by

$$\frac{df_i}{f_i} = (r + \lambda \rho b_i + \lambda' \sqrt{1-\rho_i} b_i)dt + \rho_i b_i dW + \sqrt{1-\rho_i} b_i dz_i, \forall i.$$

The market portfolio will require $\lambda' = 0$, and it is given by

$$\frac{df_M}{f_M} = (r + \lambda b_M)dt + b_M dW, b_M = \frac{\sum_{i=1}^{\infty} \rho_i b_i f_i}{\sum_{i=1}^{\infty} f_i}.$$

□

Remark 15.5.8 (alternative derivation of CAPM). Denote $\mu_m = r + \lambda b_M, \mu_i = r + \rho_i \lambda b_i$. Then we have

$$\frac{\mu_i - r}{b_i} = \rho_i \frac{\mu_M - r}{b_M}$$

or equivalently

$$(\mu_i - r) = \beta(\mu_M - r),$$

where

$$\beta = \frac{Cov(df_i, df_M)}{Var[df_M]} = \frac{\rho b_i b_M}{b_M^2} = \frac{\rho_i b_i}{b_M}.$$

This is exactly the CAPM(Theorem 10.1.1).

15.5.1.4 *Asset dynamics when all investors are risk-neutral*

Theorem 15.5.4 (constraints on parameters of dynamics when all investors are risk-neutral). *Consider a market with risk-free asset with short rate r. Consider n assets described by stochastic processes given by:*

$$\frac{df_i}{f_i} = a_i dt + \sum_{i=1}^{n} b_{ij} dz_i, i = 1, 2, ..., n.$$

*where $z_i, i = 1, 2, ..., m$ are independent Wiener processes. If there is **no arbitrage** and **all investors are risk-neutral**, then the asset dynamics is given by*

$$\frac{df_i}{f_i} = r dt + \sum_{i=1}^{n} b_{ij} dz_i, i = 1, 2, ..., n.$$

Proof. In a risk-neutral world, all investors only concern the expected return

$$\frac{dE[f_i]}{E[f_i]} = a_i dt.$$

From the law of one price, $a_i, \forall i$ should equal to the risk-free rate r. Suppose there is an asset grows at a rate greater than r, then all investors will borrow at r and invest this asset, which will push up the risk-free rate. If there is an asset grows at a rate smaller than r, then we should short this asset and invest the money into money market, which has growth rate of r. □

Remark 15.5.9 (zero market price of risk when all investors are risk-neutral). In a risk-neural world, the market price of risk is zero since investors does not require risk premium.

15.5.2 Hedging non-tradable risks

15.5.2.1 *stochastic interest rate*

15.5.2.2 *stochastic volatility*

15.6 Continuous-time market model II: pricing and martingale methods

15.6.1 Trading strategies and arbitrage

15.6.1.1 *Self-financing continuous-time trading strategies*

Definition 15.6.1 (continuous-time trading strategy). *[8, p. 25] A continuous-time trading strategies is a k-dimensional stochastic process θ_t representing the portfolio held. Each component $\theta_t^1, \theta_t^2, ..., \theta_t^n$ are predictable.*

Definition 15.6.2 (discrete-time trading strategy). *A discrete-time trading strategies is a k-dimensional stochastic process θ_t representing the portfolio held:*

- *immediately after trading at time $t-1$ so it is know at time $t-1$.*
- *and immediately before trading at time t.*

Remark 15.6.1. Consider the real axis as the time axis, trading operation only happen at these discrete time points $t_1, t_2, ...$ θ_n represents the portfolio held in the interval (t_{n-1}, t_n). θ_t is generally un-defined at time t since the trading is right occurring.

Definition 15.6.3 (value process and gains process associated with a trading strategy). *[8, p. 24]*

- *The **value process** $V_t(\theta)$, associated with a **discrete-time or continuous-time trading strategy** θ_t, with the price vector S_t, is defined by:*

$$V_t = \sum_{i=1}^{K} \theta_t^i S_t^i.$$

- *The gains process $G_t(\theta)$ associated with a **continuous-time trading strategy** θ_t, with the price vector S_t, is defined by*

$$G_t(\theta) = \sum_{i=1}^{K} \int_0^t \theta_u^i dS_u^i.$$

- *The gains process $G_t(\theta)$ associated with a* ***discrete-time trading strategy*** *θ_t, with the price vector S_t, is defined by*

$$G_t(\theta) = \sum_{i=1}^{K}\sum_{j=1}^{t} \theta_j^i (S_j - S_{j-1}).$$

Definition 15.6.4 (self-financing strategy). *[8, p. 25] A self-financing trading strategy is a trading strategy θ_t where its changes in $V_t(\theta)$ are due to entirely trading gains or loss rather than the injection or withdrawal of cash funds. In particular, a self-financing strategy satisfies:*

$$V_t(\theta) = V_0 + G_t(\theta).$$

Remark 15.6.2 (interpretation). Intuitively, a strategy is self-financing if its value changes only due to changes in the asset prices. In other words, no additional cash inflows or outflows occur after the initial time.

15.6.1.2 *Self-financing trading strategies on discounted assets*

15.6.1.3 *Attainable claims*

Definition 15.6.5 (attainable contingent claim).

- *A contingent claim Y can be defined as a random varible defined on Ω represents the payoff at time T.*
- *A contingent claim Y is said to be* ***attainable*** *if there exists a self-financing trading strategy h whose value at T satisfies $V(T) = Y$.*

Lemma 15.6.1 (vector space of attainable claims of zero initial values). *Consider a sample space $\Omega = \{\omega_1, ..., \omega_K\}$ and a contingent claim Y defined on Ω. The set of possible values $Y(\omega_1), ..., Y(\omega_K)$ of Y can be considered as an element in $\mathbb{R}^K$. Let*

$$\mathcal{G} = \{Y \in \mathbb{R}^K, Y = G(T)\ for\ some\ trading\ strategy\ h\}.$$

Then $\mathcal{G}$ is a vector space.

Proof. Note that the linear structure

$$Y = \sum_{u=1}^{t} h_0(u)\Delta B(u) + \sum_{u=1}^{t}\sum_{i=1}^{N} h_i(u)\Delta S_i(u).$$

If $Y_1, Y_2 \in \mathcal{G}$ and the associated self-financing strategies are h_1, h_2, then $\alpha Y_1 + \beta Y_2$ will have self-financing strategy of $\alpha h_1 + \beta h_2$ □

15.6.1.4 *Arbitrage*

Definition 15.6.6 (arbitrage). *In the multi-period finite-state market model, an arbitrage opportunity exists if there is a self-financing strategy h whose value satisfying*

1. $V(0) = 0$;
2. $V(T) \geq 0$;
3. $Pr(V(T) > 0) > 0$.

More generally, the arbitrage condition can be

1. $V(0) = V_0$;
2. $V(T) \geq V_0$;
3. $Pr(V(T) > V_0) > 0$.

where V_0 is a deterministic constant.

Theorem 15.6.1 (arbitrage opportunity, change of numeraire, and change of measure, recap). *Theorem 15.4.1*

- *A self-financing strategy h is an arbitrage opportunity under a probability measure P* ***if and only if*** *its discounted value V^* satisfying*
  1. $V^*(0) = 0$;
  2. $V^*(T) \geq 0$;
  3. $Pr(V^*(T) > 0) > 0$.
- *A self-financing strategy h is an arbitrage opportunity under a probability measure P* ***if and only if*** *h is an arbitrage opportunity under an* ***equivalent*** *probability measure Q.*

Proof. (1) Note that because $B(t) > 0$ almost surely, then

1. $V(0)/B(0) = 0 \Leftrightarrow V(0) = 0$;
2. $V(T)/B(T) \geq 0 \Leftrightarrow V(0) = 0$;
3. $Pr(V(T)/B(T) > 0) > 0 \Leftrightarrow Pr(V(T) > 0) > 0$ since $V(T)/B(T) > 0 \Leftrightarrow V(T) > 0$.

(2) Let h be an arbitrage opportunity under P. Let $\Omega' \subseteq \Omega$ be a set such that $V(T)(\omega) > 0, \omega \in \Omega'$. Then

$$P(\Omega') > 0 \Leftrightarrow Q(\Omega') > 0,$$

due to the equivalence of P and Q. □

Note 15.6.1 (implication).

- The first part of the theorem simply says that an arbitrage is always an arbitrage no matter which numeraire we are using.

- The second part of the theorem simply says that an arbitrage is always an arbitrage under all equivalent probability measure.
- This theorem gives us the latitude to find the most convenient numeriare and measure to pricing assets. As long as under the chosen measure and numeriare no arbitrage exists, then no arbitrage exists in the real world.

Lemma 15.6.2 (martingale value process is not an arbitrage, recap). Lemma 15.4.2 *Let $V(t)$ be a martingale under a probability measure P with $V(0) = V_0$, where V_0 is a deterministic constant. Then $V(t)$ is not an arbitrage under any equivalent probability measure.*

Proof. If $V'(T) \geq V_0, P(V(T) > V_0) > 0$, we cannot have $E_P[V(t)] = V_0$. Therefore V cannot be an arbitrage. □

Theorem 15.6.2 (self-financing strategy on discounted assets that are martingale).

15.6.2 First fundamental theorems of asset pricing

15.6.2.1 *Harrison-Pliska martingale no-arbitrage theorem*

Theorem 15.6.3 (Harrison-Pliska martingale no-arbitrage theorem). *Consider a financial market with time horizon T and price processes of the risky assets $S_1(t), S_2(t), ..., S_N(t), 0 \leq t \leq T$ and riskless bond B. The riskless bond $B(t) \geq 0$ almost surely. The market model is aribtrage-free under the probability measure P if and only if there exists another probability measure Q such that*

- *P and Q are equivalent probability measures.*
- *the discounted price processes*

$$\frac{S_1(t)}{B(t)}, \frac{S_2(t)}{B(t)}, ..., \frac{S_N(t)}{B(t)},$$

are martingales under Q.

Theorem 15.6.4 (change of numeraire in conditional expectation, extension to other numeriares, recap). Theorem 15.6.15 *Assume there exists a numeraire N and a probability measure Q_N, equivalently to the original probability measure P, such that the* *price of any traded asset X(without intermediate payments) relative to N is a martingale under Q^N, i.e.,*

$$\frac{X_t}{N_t} = E_N[\frac{X_T}{N_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

Let U be an arbitrary numeraire. Then ***there exists*** *a probability measure Q_U, equivalent to the measures P and Q_N, such that the price of any attainable claim*

$$\frac{Y_t}{U_t} = E_U[\frac{Y_T}{U_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

Moreover, the Radon-Nikodyn derivative defining the measure Q_U is given by

$$\frac{dQ_U}{dQ_N} = \frac{U_T N_0}{U_0 N_T}.$$

Remark 15.6.3 (extension to other numeriares).

Remark 15.6.4 (model free implications). The Harrison-Pliska martingale no-arbitrage theorem and its extension to other numeraires does not specify any model dynamics associated with the assets. Therefore, any dynamic model (geometric Brownian motion, jump process, or more general levy process) satisfying the martingale relationship can be used as the model.

15.6.2.2 *First fundamental theorem of asset pricing*

Definition 15.6.7 (risk-neutral measure). *[9, p. 228] A probability measure Q is said to be risk-neutral if*

- *Q and P are equivalent(i.e., P and Q have the same zero probability subsets)*
- *under Q, the discounted stock process $S_i(t)/M(t)$ is a martingale for every $i = 1, 2..., m$*

Theorem 15.6.5 (existence of risk-neutral measure implies no arbitrage, first fundamental theorem of asset pricing). *[9, p. 231]*

- *Let a martingale $M(t)$ under equivalent risk-neutral probability measure Q representing the value of a portfolio with $M(0) = 0$, then it does not admit arbitrage.*
- *If a market admits no arbitrage, then it has an risk-neutral probability measure.*

Proof. (1) Consider the arbitrage ($M(0) = 0, P(M(T) \geq 0) = 1, P(M(T) > 0) > 0$ [10, p. 230]) exists, then $E_Q[M(t)] > 0$ due to the fact that probability measure Q and P are equivalent. This contradicts with the fact that $M(t)$ is a martingale $E[M(t)] = M(0) = 0$.
(2) See Theorem 15.6.7. □

15.6.2.2.1 RISK-NEUTRAL MEASURE I: BASIC CONCEPTS

Definition 15.6.8 (portfolio process and it value). *A portfolio process $\Delta_i(t)$ represents the position of stock i as a function of time. Its value $X(t)$ is governed by the following SDE as:*

$$dX(t) = \sum_{i=1}^{m} \Delta_i dS_i + r(t)(X(t) - \sum_{i=1}^{m} \Delta_i S_i(t))$$

Lemma 15.6.3 (discount portfolio value process is martingale under risk-neutral measure). *[9, p. 230] Let Q be a risk-neutral measure(that is, the discount stock process will be a martingale), then the discounted portfolio value process $D(t)X(t)$ will be a martingale.*

Proof.

$$\begin{aligned} dX(t) &= \sum_{i=1}^{m} \Delta_i dS_i + r(t)(X(t) - \sum_{i=1}^{m} \Delta_i S_i(t)) \\ &= r(t)X(t)dt + \sum_{i=1}^{m} \Delta_i (dS_i(t) - r(t)S_i(t)dt) \\ &= r(t)X(t)dt + \sum_{i=1}^{m} \frac{\Delta_i(t)}{D(t)} d(D(t)S_i(t)) \end{aligned}$$

Then

$$d(D(t)X(t)) = D(t)(dX(t) - R(t)X(t)dt) = \sum_{i=1}^{m} \Delta_i(t) d(D(t)S_i(t)).$$

Since under Q, $D(t)S_i(t)$ is a martingale, $D(t)X(t)$ will also be a martingale. □

Remark 15.6.5 (general pricing procedures). If we can use a portfolio process replicating the payoff of a derivative, then the initial value of the portfolio process is the initial value of the derivative.

Remark 15.6.6 (mathematically interpretation of risk-neutral measure).

- It is tempting to interpret risk-neutral measure from the economical point of view, which, however, can create confusion.
- Mathematically, risk-neutral measure is simply a equivalent measure with some special properties(the most important: the discount portfolio process is a martingale).
- There are infinitely many equivalent measures, but they do not have properties as nice as risk-neutral measure.
- Usually, it is not important to know the explicit form of the risk-neutral measure; however, knowing the dynamic model of assets(e.g., stocks) under such measure is essential.

Remark 15.6.7 (Implications for a martingale)**.** From the properties of martingales(Lemma 5.5.1), we have:

- A martingale is a zero-drift stochastic process. A stochastic process $\theta(t)$ is a martingale if it has the form
$$d\theta = \sigma dW$$
where the variable σ can itself be stochastic or as a function of θ and other stochastic processes, W is the Wiener process.
- The expected value at any future time is equal to its value today:
$$E[\theta(T)] = \theta_0$$
note that $E[]$ here does not condition on any thing.

Remark 15.6.8 (Why cares about martingale)**.**

- Consider the discounted value process $M(t)$. If there exists an **equivalent measure** Q such that under Q, $M(t)$ is a martingale, then $M(t)$ will not admit arbitrage in arbitrary equivalent measure(including the real-world measure).
- Suppose $M(0) = 0$, and let $M_T \triangleq M(T)$ be a random variable defined on $(\Omega, \mathcal{F})$. If there exists a measure P such that $P(M_T \geq 0) = 1$, then that mean $M_T(A), \forall \{A \in \mathcal{F}, P(A) > 0\}$. Then for any other probability measure Q equivalent to P, we must also have $Q(M_T \geq 0) = 1$, since the subsets $B \in \mathcal{F}$ such that $M_T < 0$ have measure 0 in both probability measure P and Q.
- Therefore, the no-arbitrage condition implies there exists a probability measure under which discounted value process is martingale. And if there exists

15.6.2.2.2 Risk-neutral measure II: existence

Theorem 15.6.6 (Sufficient condition (coefficient constraints) for existence of risk-neutral measure). *[9, p. 229] From* **??**, *we know that the SDE for the discount stock process $D(t)S_i(t)$ is*

$$d(D(t)S_i(t)) = D(t)S_i(t)[(\mu_i - r(t))dt + \sum_{j=1}^{d} \sigma_{ij}(t)dW_j(t)].$$

If there exists processes *$\theta_1(t), ..., \theta_d(t)$ satisfying the following,*

$$\mu_i(t) - r(t) = \sum_{j=1}^{d} \sigma_{ij}(t)\theta_j(t), i = 1, ..., m,$$

then there exists a measure, called ***risk-neutral*** *measure, such that the discount stock process under Q becomes:*

$$d(D(t)S_i(t)) = D(t)S_i(t)[\sum_{j=1}^{d} \sigma_{ij}(t)d\hat{W}_j(t)]$$

where $d\hat{W}$ are Brownian motions under Q.

More specifically, Let

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{11} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2d} \\ \vdots & & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & \sigma_{Md} \end{pmatrix}, x = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix}, b = \begin{pmatrix} u_1 - r \\ u_2 - r \\ \vdots \\ u_M - r \end{pmatrix}.$$

If $Ax = b$ has a solution, then there exists a risk-neutral measure; If $Ax = b$ has a unique solution, then there exists a unique risk-neutral measure.

Proof. Use Girsanov theorem(**??**), which says there exists a probability Q such that

$$W(t) + \int_0^t \theta(u)du$$

can be made to be a driftless Brownian motion under Q. Therefore, as long as we can solve $\theta_1(t), ..., \theta_d(t)$ from the linear equation, we can use the Girsanov theorem to find such measure. □

Remark 15.6.9 (Some algebraic conditions)**.**

- First, we always assume $d \leq M$ without loss of generality, since $d > M$ case is equivalent to $d = M$ case(see 6.3.1).
- If $d \leq M$, then there are exactly three possibilities: no solution, unique solution, and infinitely many solutions. See (**??**).

Theorem 15.6.7 (No arbitrage implies the existence of risk-neutral measure). *[9, p. 228] From* **??***, we know that the SDE for the discount stock process $D(t)S_i(t)$ is*

$$d(D(t)S_i(t)) = D(t)S_i(t)[(\mu_i - r(t))dt + \sum_{j=1}^{d} \sigma_{ij}(t)dW_j(t)].$$

If there exists ***no arbitrages*** *in the market model, then* ***there exists processes*** $\theta_1(t), ..., \theta_d(t)$ *satisfying the following,*

$$\mu_i(t) - r(t) = \sum_{j=1}^{d} \sigma_{ij}(t)\theta_j(t), i = 1, ..., m,$$

or equivalently from (Theorem 15.6.6), there exists a measure, called ***risk-neutral*** *measure, such that the discount stock process under Q becomes:*

$$d(D(t)S_i(t)) = D(t)S_i(t)[\sum_{j=1}^{d} \sigma_{ij}(t)d\hat{W}_j(t)]$$

where $d\hat{W}$ are Brownian motions under Q.

Proof. See Theorem 15.5.1. □

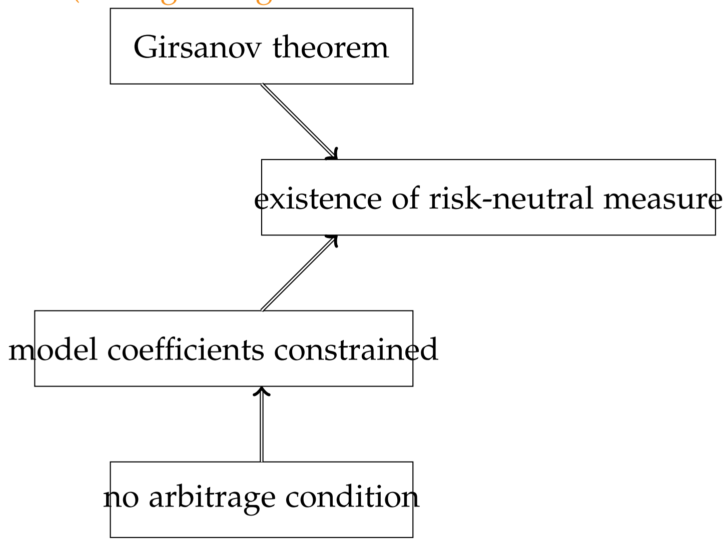
Note 15.6.2 (continuous version of finite-state price)**.**

- Consider a special stock $S(t)$ that pays 1 if and only if an event $A(t) \in \mathcal{F}_t$ occurs, then its prices is

$$E_Q[D(t)S(t)] = \int_{\infty}^{\infty} D(t)I(A(t))dP = D(t)Q(A),$$

  which equals the discounted probability measure.
- Suppose in the market, there are infinite number of such stocks $S_A$that pay 1 if and only if an event $A(t) \in \mathcal{F}(t)$ occurs. Then $Q(A(t)) = S_A(t)/D(t)$; that is, we can calculate the risk-neutral measure from the stock price.
- In both finite-state model and continuous-time version, the no-arbitrage condition guarantees the existence of the risk-neutral measure or state price. See (Theorem 15.3.1 and Theorem 15.6.7).
- For another discussion, see subsubsection 16.2.2.2.

Remark 15.6.10 (the logic diagram for existence of risk neutral measure).



- In the Brownian asset dynamics model, the no-arbitrage condition implies that the model coefficients are constrained (Theorem 15.5.1).
- The constrained coefficients together with Girsanov theorem Theorem 15.6.6 can imply the existence of risk-neutral measure.
- Further, the existence of risk-neutral measure will imply there is no arbitrage opportunitiesTheorem 15.6.5.

15.6.2.3 *Risk-neutral pricing in no-arbitrage markets*

Theorem 15.6.8 (no-arbitrage pricing in no-arbitrage marketcomplete market). *[9, p. 218] Let $D(t)$ be a discount process, and let $V(t)$ be an asset value. Assume the market is free of abitrages. Denote the risk-neutral measure be Q. Then for* ***any attainable contingent claim*** *with payoff $V(T)$ at time T, its time-t* ***no-arbitrage price*** *is given by*

$$D(t)V(t) = \frac{1}{D(t)} E_Q[D(T)V(T)|\mathcal{F}_t].$$

In particular at current time $t = 0$,

$$V(0) = E_Q[D(T)V(T)].$$

Proof. Because the market is free of abitrages, then there exist a portfolio process $X(t)$ such that $X(T) = V(T)$. Let $X(T)$ be a replicate of asset $V(T)$. Based on the lemma (Lemma 15.6.3), then $D(t)X(t)$ is a martingale with Q. Therefore,

$$D(t)X(t) = E_Q[D(T)X(T)|\mathcal{F}_t]$$

If $X(T)$ replicates $V(T)$, then from the Law of one price(Theorem 15.2.1), $V(t)$ will be equal to $X(t)$. □

Example 15.6.1 (European call put price). The European call with strike price K and maturity time T has price C

$$C = E_Q[D(T)\max(S(T)-K,0)];$$

And a put has price P

$$P = E_Q[D(T)\max(K-S(T),0)].$$

And note that under risk-neutral measure, the S_t is following Lemma 15.8.1:

$$dS_t = rS_t dt + \sigma S_t dW_t$$

where W_t is a Wiener process.

Example 15.6.2 (pricing of a special call). Suppose we have a call on stock S_t and matures at T with payoff $\max(S_T^2 - K, 0)$. Then the value at $t = 0$ is

$$C = E_Q[D(T)\max(S_T^2 - K, 0)],$$

where $Y_t = S_t^2$ under risk-neutral measure Q will follow

$$dY_t = (2r+\sigma^2)Y_t dt + (2\sigma)Y_t dW_t.$$

15.6.2.4 *No-arbitrage pricing of futures and forwards*

Corollary 15.6.8.1 (Zero-Coupon Bond price). *[9, p. 240] In a complete market with risk-neutral measure Q, the price for the zero-coupon bond that pays 1 at time T has price at time t as*

$$B(t,T) = \frac{1}{D(t)} E_Q[D(T)|\mathcal{F}_t]$$

In particular at $t = 0$,

$$B(0,T) = E_Q[D(T)].$$

Proof. Use Theorem 15.6.13. Note that $B(T,T) = 1$. □

Remark 15.6.11 (risk-neutral measure is unknown). Note that in this lemma, we do not have the form of risk-neutral measure or how the short-rate dynamics will be under such measure. We just assume it exist, and its form will be discussed in the following sections.

Definition 15.6.9 (forward contract). *[9, p. 241] A forward contract is an agreement to pay a specified delivery price K at a delivery date T for an asset S whose price at t is $S(t)$. Let $F(t,T)$ denote the forward price of this asset at time t, then $F(t,T)$ is the value of K that makes the forward contract having no-arbitrage price zero at at time t.*

Corollary 15.6.8.2 (forward contract pricing). *[9, p. 241] In a complete market with risk-neutral measure Q, the forward price for $S(t)$ is*

$$F(t,T) = \frac{S(t)}{B(t,T)}$$

Proof. Use Theorem 15.6.13.

$$\begin{aligned} 0 &= \frac{1}{D(t)} E_Q[D(T)(S(T) - K)|\mathcal{F}_t] \\ &= \frac{1}{D(t)} E_Q[D(T)S(T)|\mathcal{F}_t] - \frac{K}{D(t)} E_Q[D(T)|\mathcal{F}_t] \\ &= S(t) - KB(t,T) \end{aligned}$$

Therefore,

$$F(t,T) = K = \frac{S(t)}{B(t,T)}$$

□

Remark 15.6.12 (forward pricing vs. future pricing)**.**

- When the interest rate is non-random(even as a function of time), the forward and the future has the same price.
- When the interest rate is random, then the forward and the future price will be different.[9, p. 244]

15.6.3 Second fundamental theorem of asset pricing

15.6.3.1 *Martingale representation theorem*

Theorem 15.6.9 (one dimensional martingale representation theorem). *[11, p. 193][12, p. 49]*

- *Let $W(t), 0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and with natrual filtration $\{\mathcal{F}_t\}$. If $M_t, 0 \leq t \leq T$, is a martingale with respect to the filtration $\{\mathcal{F}_t\}$, then there exists an predictable process [a] h_t such that*

$$M_t = M_0 + \int_0^t h(s)dW(s), 0 \leq t \leq T.$$

- *Let $W(t)$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and with natrual filtration $\{\mathcal{F}_t\}$. Let X be a square integrable random variable measurable with respect to $\mathcal{F}_T$, then there exists a predictable process h which is adapted with respect to $\{\mathcal{F}_t\}$, such that*

$$X = E[X] + \int_0^T h(s)dW(s).$$

a Because Brownian motion has continuous sample path, with respect to the natural filtration of the Brownian motion, predictable process is the same as adapted process

Proof. (2) Define $X(t) = E[X]\mathcal{F}_t]$ and $X(T) = E[X]\mathcal{F}_T]$. From (1), we have

$$\begin{aligned} X_T &= X_0 + \int_0^T h_u dW_u \\ &= E[X]\mathcal{F}_0] + \int_0^T h_u dW_u \\ &= E[X] + \int_0^T h(s)dW(s) \end{aligned}$$

□

Example 15.6.3. [11, p. 193]

- If $X = W_T$, then $h_t = 1$.
- If $X = W_T^2$, then $h_t = 2W_t$.
- If $X = W_T^3$, then $h_t = 3(W_t^2 + T - t)$.
- If $X = \exp(\sigma W_T)$, then $h_t = \sigma \exp(\sigma W_t + \frac{1}{2}\sigma^2(T-t))$.

Theorem 15.6.10 (Multidimensional dimensional martingale representation theorem).

- *Let $W(t), 0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and with natrual filtration $\{\mathcal{F}_t\}$. If $M_t, 0 \leq t \leq T$, is a martingale with respect to the filtration $\{\mathcal{F}_t\}$, then there exists an predictable process [a] h_t such that*

$$M_t = M_0 + \sum_{i=1}^{M} \int_0^t h_i(s)dW_i(s), 0 \leq t \leq T.$$

- *Let $W(t)$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and with natrual filtration $\{\mathcal{F}_t\}$. Let X be a square integrable random variable measurable with respect to $\mathcal{F}_T$, then there exists a predictable process h which is adapted with respect to $\{\mathcal{F}_t\}$, such that*
$$X = E[X] + \int_0^T h(s)dW(s).$$

[a] Because Brownian motion has continuous sample path, with respect to the natural filtration of the Brownian motion, predictable process is the same as adapted process

Proof. Similar to Theorem 15.6.9. □

15.6.3.2 *Second fundamental theorem of asset pricing*

Definition 15.6.10 (complete market). *A market model is complete if **every finite-variance derivative security measurable with respect to** $\mathcal{F}_T$ can be hedged. Or equivalently, a market model is complete if the set of contingent claims measurable with respect to $\mathcal{F}_\infty$ are attainable.*

Theorem 15.6.11 (linear algebra condition for unique risk-neutral measure). ***Assume there exist no arbitrage in the market.*** *Let the SDE of N risky assets be*

$$\frac{df_i}{f_i} = (r_i + \sum_{j=1}^{M} \lambda_j \sigma_{ij})dt + \sum_{j=1}^{M} \sigma_{ij} dw_j$$

where $w_1, ..., w_M$ are M independent Brownian motions. Further, let

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \dots & \sigma_{2M} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \dots & \sigma_{NM} \end{pmatrix}, V = \begin{pmatrix} v_1 + \lambda_1 \\ v_2 + \lambda_2 \\ \vdots \\ \vdots \\ v_M + \lambda_M \end{pmatrix}$$

*Then, **the no arbitrage condition implies there always exists a risk-neutral measure**. Particularly,*

- *When $M > N$ (more risk factors than underlying assets excluding bonds), there exists infinitely many risk-neutral measure Q.*

- *When $M \leq N$ (less risk factors than underlying assets excluding bonds), there are infinitely many or one unique risk-neutral measure Q.*
- *Particularly, there are M independent columns in Σ if and only if there exists one unique risk-neutral measure.*

Under this measure Q, the SDE dynamics is

$$\frac{df_i}{f_i} = r_i dt + \sum_{j=1}^{M} \sigma_{ij} dw_j.$$

Proof. From Girsanov theorem(Theorem 6.7.6), we can construct a new measure Q, such that $dw_i = d\hat{w}_i + v_i, \forall i$. The existence of a risk-neutral measure is equivalent to find a solution to

$$\Sigma V = 0.$$

We have exactly two situations **??**,

- When $M > N$, there are infinitely many solution; that is, there exists infinitely many risk-neutral measure Q.
- When $M \leq N$, there are infinitely many solutions or one unique solution. if there are M independent columns in Σ, there exists one unique solution.

□

Remark 15.6.13 (interpretation).

- Under no arbitrage condition, there always exists risk-neutral measure, which is a restatement of first fundamental theorem of asset pricing(Theorem 15.3.1) that no-arbitrage implies the existence of risk-neutral measure.
- When there exists multiple risk-neutral measures, the pricing will not be unique(since price is obtained by taking expectation with respect to such measure).

Theorem 15.6.12 (Second fundamental theorem of asset pricing, uniqueness of risk-neural measure). *[9, p. 232] Consider a no-arbitrage market model consisting of a money market account $B(t)$ and N risky asset $S(t)$ such that*

$$dB(t) = r(t)B(t)dt$$
$$dS_i(t)/S_i(t) = (r_i + \sum_{j=1}^{M} \lambda_j \sigma_{ij})dt + \sum_{j=1}^{M} \sigma_{ij} dw_j.$$

Then the market is complete if and only if it has a unique risk-neutral measure.

Proof. First note that no-arbitrage condition ensures that the existence of risk-neutral measure(Theorem 15.6.5). (1) Assume the model is complete. Suppose we have two risk-neutral measures $Q_1, Q_2 : \mathcal{F} \to \mathbb{R}$. Consider a special indicator derivative with payoff $I_A, A \in \mathcal{F}$.

$$E_{Q_1}[D(T)I_A] = E_{Q_1}[D(0)I_A] = Q_1(A) = E_{Q_2}[D(T)I_A] = E_{Q_2}[D(0)I_A] = Q_2(A).$$

Since A is arbitrary, then this two measure are equal for all subsets in $\mathcal{F}$. (2) (uniqueness implies completeness) For any derivative with payoff $V(T)$ measurable to the natural filtration generated by $W_1, ..., W_M$ (which is the same natural filtration generated by $W_1^Q, W_2^Q, ..., W_M^Q$, Brownian motions under risk-neutral measure), the martingale representation theorem (Theorem 15.6.10) says that **under risk-neutral measure there exist adaptive processes** $\Lambda_1(t), \Lambda_2(t), ..., \Lambda_M(t)$ such that

$$\frac{V(t)}{B(t)} = \frac{V(0)}{B(0)} + \sum_{i=1}^{M} \int_0^t \Lambda_i(u) dW_i^Q.$$

On the other hand, **if** $V(T)$ **can be replicated via a self-financing strategy** such that

$$\begin{aligned}
\frac{V(t)}{B(t)} &= \frac{V(0)}{B(0)} + \sum_{i=1}^{N} \int_0^t h_i(u) d(S_i(u)/B(u)) \\
&= \frac{V(0)}{B(0)} + \sum_{i=1}^{N} \int_0^t h_i(u) \sum_{j=1}^{M} \frac{\sigma_{ij}(u)}{B(u)} dW_j^Q.
\end{aligned}$$

To solve for $h_1(t), ..., h_N(t)$ for all t, we need to solve

$$\begin{pmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \cdots & \cdots & \sigma_{2N} \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
\sigma_{M1} & \sigma_{N2} & \cdots & \cdots & \sigma_{MN}
\end{pmatrix}
\begin{pmatrix}
S_1(t)h_1(t)/B(t) \\
S_1(t)h_2(t)/B(t) \\
\vdots \\
\vdots \\
S_N(t)h_N(t)/B(t)
\end{pmatrix}
=
\begin{pmatrix}
\Lambda_1 \\
\Lambda_2 \\
\vdots \\
\vdots \\
\Lambda_M
\end{pmatrix}$$

When there exists an unique risk-neutral measure, the linear algebra condition in gives that the matrix Σ has to be full row rank (Theorem 15.6.11) and therefore the linear equation is the consistent and will always have a solution, even though it might not be unique. □

Corollary 15.6.12.1 (completeness of two asset market model). *Consider a market model consisting of a money market account $M(t)$ and a risky asset $S(t)$ such that*

$$\begin{aligned} dB(t) &= r(t)B(t)dt \\ dS(t) &= \mu(t)S(t)dt + \sigma(t)dW(t) \end{aligned}$$

Then any derivative instrument with payoff $V(T)$ measurable to the natural filtration generated by $W(t)$ can be hedged.

Proof. For any derivative with payoff $V(T)$ measurable to the natural filtration generated by $W(t)$, the martingale representation theorem (Theorem 15.6.9) says that **under risk-neutral measure there exist an adaptive process** $\Lambda(t)$ such that

$$\frac{V(t)}{B(t)} = \frac{V(0)}{B(0)} + \int_0^t \Lambda(u)dW^Q.$$

On the other hand, **if $V(T)$ can be replicated via a self-financing strategy** such that

$$\begin{aligned} \frac{V(t)}{B(t)} &= \frac{V(0)}{B(0)} + \int_0^t h(u)d(S(u)/B(u)) \\ &= \frac{V(0)}{B(0)} + \int_0^t h(u)\frac{\sigma(u)}{B(u)}dW^Q. \end{aligned}$$

Therefore, we have

$$h(t) = \Lambda(t)B(t)/(S(t)\sigma(t));$$

that is, we can replicate any derivatives. □

Note 15.6.3 (completeness under other equivalent measures)**.**

- Consider a no-arbitrage market model. If the market model is complete under risk-neutral measure, then it is also complete under other equivalent martingale measure.
- This is because when a derivative can be replicated using a self-financing strategy h, then random payoff of the derivative and the strategy is equal, independent of the measure used.

15.6.3.3 *Pricing in complete markets*

Theorem 15.6.13 (no-arbitrage pricing in complete market). *[9, p. 218] Let $D(t)$ be a discount process, let $V(t)$ be an asset value. Assume the market is complete and let the risk-neutral measure be Q. Then*

$$D(t)V(t) = E_Q[D(T)V(T)|\mathcal{F}_t].$$

In particular at $t = 0$,

$$D(0)V(0) = E_Q[D(T)V(T)]$$

which simplifies as

$$V(0) = E_Q[D(T)V(T)]$$

where $D(0) = 1$.

Remark 15.6.14 (Complete market vs. incomplete market)**.**

15.6.4 Dividend paying underlying assets

15.6.4.1 *General remarks*

Remark 15.6.15 (special issues concerning dividend paying assets)**.**

- If the stock is constantly paying dividends, its price will decrease. From the perspective of investors, its value is decreasing; However, from the perspective of stock holders, their total values are not because the dividends belong to them.
- The general model for a stock paying dividend is given as

$$dS_t = (\mu(t) - a(t))S_t dt + \sigma S_t dW_t$$

- Recall that for non-dividend paying stocks, we require the risk-neutral measure to be such that discount stock process $D(t)S_i(t)$ is a martingale(Definition 15.6.7), and as a consequence, the discount portfolio value process $D(t)X(t)$ is a martingale(Lemma 15.6.3). Therefore, we can use discount portfolio value process to replicate and price other derivatives.
- When the stock is paying dividend, we will just require the risk-neutral measure Q to be such that the discount portfolio value process $D(t)X(t)$ is a martingale. **Under Q, $D(t)S_i(t)$ is usually not a martingale([9, p. 248]); Instead, under Q, the discounted value of the portfolio and its dividend accumulating interest rate will be a martingale.**

Definition 15.6.11 (risk-neutral measure with asset paying dividend). *[9, p. 235] A probability measure Q is said to be risk-neutral if*

- *Q and P are equivalent(i.e., P and Q have the same zero probability subsets)*
- *under Q, the discounted portfolio value process $D(t)X(t)$ is a martingale.*

15.6.4.2 *Continuous-dividend paying stocks*

Lemma 15.6.4 (SDE for continuous dividend paying stocks under risk-neutral measure). *Assume the stock S_t in the real-world probability measure follows*

$$dS_t = (\mu - a)S_t dt + \sigma S_t dW_t$$

where W_t is a Wiener process and a is the dividend rate. Then under the risk-neutral measure, the stock dynamics will be

$$dS_t = [r(t) - a(t)]S_t dt + \sigma(t) S_t dW(t).$$

The solution to the SDE is

$$S(t) = S(0)\exp[\int_0^t \sigma(u) dW_u + \int_0^t (r(t) - a(t) - \frac{1}{2}\sigma^2(u)] du$$

Proof.

$$\begin{aligned} dX(t) &= \Delta(t) dS(t) + \Delta(t) a(t) S(t) dt + r(t)[X(t) - \Delta(t)S(t)]dt \\ &= r(t)X(t)dt + (a(t) - R(t))\Delta(t)S(t)dt + \sigma(t)\Delta(t)S(t)dW(t) \\ &= r(t)X(t)dt + \Delta(t)S(t)\sigma(t)[\theta(t)dt + dW(t)] \end{aligned}$$

where

$$\theta(t) = \frac{\mu(t) - r(t)}{\sigma(t)}.$$

If we define

$$\hat{W}(t) = W(t) + \int_0^t \theta(u) du,$$

then from **??**, there exist a measure Q such that $\hat{W}(t)$ will be a Brownian motion. Note that

$$dS_t = [\mu(t) - a(t)]S_t dt + \sigma(t) S_t dW(t).$$

Plug into

$$dW(t) = d\hat{W}(t) - \frac{\mu(t) - r(t)}{\sigma(t)} dt,$$

then we have

$$dS_t = [r(t) - a(t)]S_t dt + \sigma(t) S_t dW(t).$$

□

Remark 15.6.16 (implication for pricing)**.**

- This distribution of S_t under risk-neutral measure Q is important in pricing options on stocks.
- Note the distribution under risk-neutral measure is different for dividend paying stocks and non-dividend paying stocks.

Remark 15.6.17 (risk-neutral dynamics for stocks and its accumated dividends)**.** Consider we have a money account $M(t)$. Then our stock price S_t and the money account M_t dynamics is given by

$$\begin{aligned} dS_t &= (r(t) - a(t))S_t dt + \sigma(t) S_t dW(t) \\ dM_t &= r(t)M(t)dt + a(t)S_t dt \end{aligned}$$

The total value $S(t) + M(t)$ has the following dynamics given by

$$d(S_t + M_t) = r(t)(S_t + M_t)dt + \sigma(t)S_t dW_t.$$

And the discounted total value $D(t)(S_t + M(t))$ will be martingale under risk-neutral measure.

Theorem 15.6.14 (pricing derivatives of continuous dividend paying stocks). *Let $D(t)$ be a discount process, let $V(t)$ be an derivative of a dividend paying stock S_t. Assume the market is complete and let the risk-neutral measure be Q. Then,*

$$D(t)V(t) = E_Q[D(T)V(T)|\mathcal{F}_t].$$

In particular at $t = 0$,

$$D(0)V(0) = E_Q[D(T)V(T)]$$

which simplifies as

$$V(0) = E_Q[D(T)V(T)]$$

where $D(0) = 1$. ***Note that under the risk-neutral measure*** *Q,*

$$dS_t = [r(t) - a(t)]S_t dt + \sigma(t)S_t dW(t).$$

Proof. Same as Theorem 15.6.13. □

15.6.4.3 *Discrete-dividend paying stocks*

Lemma 15.6.5 (SDE for discrete dividend paying stocks under risk-neutral measure). *[9, p. 240] Consider a stock S_t with discrete dividend payment D_i at $t_i, i = 1, 2, 3, \ldots$. Let current time be t, and $t < t_1 < t_2 < \cdots$. Then its risk-neutral dynamics will be*

$$dS_t = r(t)S_t dt - \sum_{i=1} H(t - t_i)D_i + \sigma(t)S_t dW_t.$$

Proof. Consider we have a money account $M(t)$. Then our stock price S_t and the money account M_t dynamics is given by

$$\begin{aligned} dS_t &= r(t)S_t dt - \sum_{i=1} H(t - t_i)D_i + \sigma(t)S_t dW(t) \\ dM_t &= r(t)M(t)dt + \sum_{i=1} H(t - t_i)D_i \end{aligned}$$

The total value $S(t) + M(t)$ has the following dynamics given by

$$d(S_t + M_t) = r(t)(S_t + M_t)dt + \sigma(t)S_t dW_t.$$

And the discounted total value $D(t)(S_t + M(t))$ will be martingale under risk-neutral measure. □

15.6.5 Numeraire and pricing

15.6.5.1 *No-arbitrage pricing under different numeraire*

Definition 15.6.12 (numeraire). *[8] A numeraire is any positive non-dividend-paying asset.*

Theorem 15.6.15 (change of numeraire in conditional expectation). *[8, p. 27][13, p. 24] Assume there exists a numeraire N and its associated martingale probability measure Q_N, equivalently to the initial measure Q_0, such that the price of any traded asset X(without intermediate payments) relative to N is a martingale under Q^N, i.e.,*

$$\frac{X_t}{N_t} = E_N[\frac{X_T}{N_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

*Let U be an arbitrary numeraire. Then **there exists** a probability measure Q_U, called **martingale measure associated with numeriare** U, equivalent to the initial Q_0, such that*

$$\frac{X_t}{U_t} = E_U[\frac{X_T}{U_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

Moreover, the Radon-Nikodyn derivative defining the measure Q_U is given by

$$\frac{dQ_U}{dQ_N} = \frac{U_T N_0}{U_0 N_T},$$

such that

$$U_t E_U[\frac{X_T}{U_T}|\mathcal{F}_t] = U_t E_N[\frac{X_T}{U_T}\frac{dQ_U}{dQ_N}|\mathcal{F}_t] = E_U[\frac{X_T}{U_T}|\mathcal{F}_t].$$

Proof. From the assumption, we have

$$E_N[\frac{U_T}{N_T}|\mathcal{F}_t] = \frac{U_t}{N_t}.$$

We can define

$$Z = \frac{dQ_U}{dQ_N} = \frac{U_T N_0}{U_0 N_T}$$

such that Z is strictly positive and $E[Z] = 1$.

Then from Bayes theorem for conditional expectation(Theorem 6.7.5), we have

$$E_U[\frac{Y_T}{U_T}|\mathcal{F}_t] = \frac{E_N[\frac{Y_T}{U_T}Z|\mathcal{F}_t]}{E_N[Z|\mathcal{F}_t]}.$$

Note that

$$E_N[\frac{Y_T}{U_T}Z|\mathcal{F}_t] = \frac{N_0}{U_0}E_N[\frac{Y_T}{N_T}|\mathcal{F}_t] = \frac{N_0}{U_0}\frac{Y_t}{N_t},$$

and

$$E_N[Z|\mathcal{F}_t] = \frac{N_t}{U_t}E_N[\frac{U_T}{N_T}|\mathcal{F}_t] = \frac{N_0}{U_0}\frac{U_t}{N_t}.$$

Then

$$E_U[\frac{Y_T}{U_T}|\mathcal{F}_t] = \frac{Y_t}{U_t}.$$

□

Note 15.6.4 (implication)**.**

- The most important aspect of this theorem is that: combining the first fundamental theorem (Theorem 15.6.6) to ensure the existence of risk neutral measure, we can define no-arbitrage price for all attainable claims in a consistent way (i.e., price of a claim is independent of the numeriare used). Therefore, we can choose the type of numeriare most convinient for us.
- The specific consequences of the this new measure on asset dynamics is discussed at Theorem 15.6.18 and Theorem 15.6.19.

Example 15.6.4 (from risk-neutral measure to forward measure). Denote risk-neutral and forward measure by Q and Q_T, we have

$$\frac{dQ}{dQ_T} = \frac{B(T)}{B(t)}\frac{P(t,T)}{P(T,T)}.$$

For an asset X, under risk-neutral measure, we have

$$\begin{aligned} X(t) &= B(t)E_Q[\frac{X(T)}{B(T)}|\mathcal{F}_t] \\ &= B(t)E_T[\frac{X(T)}{B(T)}\frac{dQ}{dQ_T}|\mathcal{F}_t] \\ &= B(t)E_T[\frac{X(T)}{B(T)}\frac{B(T)}{B(t)}\frac{P(t,T)}{P(T,T)}|\mathcal{F}_t] \\ &= P(t,T)E_T[\frac{X(T)}{P(T,T)}|\mathcal{F}_t] \end{aligned}$$

Methodology 15.6.1 (pricing in different numeraire). *Let $N(t)$ be a numeraire. Let E_N denote the expectation with resepct to the martingale measure associated with $N(t)$.*

- *Consider a contingent claim with payoff C unit of domestic currency at future time T, where C is a random variable. Then its price is*

$$V(t) = P(t,T)E_T[\frac{V(T)}{P(T,T)}|\mathcal{F}_t] = P(t,T)\cdot E_T[C|\mathcal{F}_t],$$

  where $P(t,T)$ is the zero-coupon bond numeriare, and E_T denote the expectation with resepct to the martingale measure associated with $P(t,T)$.
- *Consider a contingent claim with payoff $c \cdot N(T)$ at future time T, where c is a constant multiplier. Then its price is*

$$V(t) = N(t)E_N[\frac{C\cdot N(T)}{N(T)}|\mathcal{F}_t] = c\cdot N(t).$$

- *Consider a contingent claim with payoff $C \cdot N(T)$ at future time T, where c is a random variable. Then its price is*

$$V(t) = N(t)\cdot E_N[\frac{c\cdot N(T)}{N(T)}|\mathcal{F}_t] = N(t)\cdot E_N[C|\mathcal{F}_t].$$

Example 15.6.5. Consider a payer swaption (Lemma 19.3.15) with strike K and expiry T. Its payoff is either zero or entering a swap, which can be writen by

$$V(T) = N(T) \cdot (Swap(T) - K)^+,$$

where $Swap(T)$ is the swap rate at future time T, and $N(T)$ is the annuity.

By choosing the annuity as the numeriare, we have its current value given by

$$V(t) = N(t)E_N[\frac{N(T) \cdot (Swap(T) - K)^+}{N(T)}|\mathcal{F}_t] = N(t)E_N[(Swap(T) - K)^+|\mathcal{F}_t].$$

We can evaluate the swaption value if we know the distribution $Swap(T)$ under the martingale measure associated with $N(t)$.

15.6.5.2 *State price and risk-neutral pricing*

Lemma 15.6.6 (state price density and pricing in zero-interest world). *Consider a world with zero interest rate. Define $V(t, \omega_0) \triangleq PV[\delta(\omega - \omega_0)]$, where $PV[\cdot]$ is the present value operator taking an random variable as input and produce a real number as output. It follows that*

- *For an arbitrary payoff $f(\omega)$, a random variable $\Omega \to \mathbb{R}$, we have the decomposition*

$$f(\omega) = \int \delta(\omega - \omega_0) f(\omega_0) d\omega_0.$$

- *No arbitrage condition requires $1 = PV[1]$ and $PV[\cdot]$ is a linear operator.*
- *The present value of an arbitrary payoff $f(\omega)$ is given by*

$$V(t) = PV[f(\omega)] = \int V(t, \omega) f(\omega) d\omega.$$

- $$1 = \int V(t, \omega) d\omega.$$

- *If we define $E_Q[f|\mathcal{F}_t] = \int f(\omega)V(\omega)d\omega$, then*

$$V(t) = E_Q[f|\mathcal{F}_t]$$

.

Proof. (1) Directly from the property of delta function. (2) From law of one price and linear pricing theorem.Theorem 15.2.2Theorem 15.2.1(3)

$$\begin{aligned} V(t) &\triangleq PV[f(\omega)] \\ &= PV[\int \delta(\omega - \omega_0) f(\omega_0) d\omega_0] \\ &= \int PV[\delta(\omega - \omega_0)] f(\omega_0) d\omega_0 \\ &= \int V(t, \omega_0) f(\omega_0) d\omega_0 \end{aligned}$$

□

Definition 15.6.13 (state price density and pricing using money account as numeraire). *A state price density using money account as numeraire is the time t price of the payoff denominated in the money account when the world is realized sample point $\omega \in \Omega$ at time T. We write the state price density as*

$$\frac{V(t, \omega_0)}{B_t} = PV[\frac{\delta(\omega - \omega_0)}{B_T(\omega)}].$$

Theorem 15.6.16. *Assume we have the state price density for all possible states. Given any payoff function $f(S_T), S_T : \Omega \to \mathbb{R}$. We have*

- *The present value of money account B_T(a random variable) is given by*

$$V(t) = B_t.$$

- $\int V(t, \omega) d\omega = 1.$
- $$V(t) = B_t PV[\int \frac{\delta(\omega - \omega_0) f(\omega_0)}{B_T(\omega_)} d\omega_0]$$
- $$V(t) = B_t E_Q[\frac{f(\omega)}{B_T(\omega)} | \mathcal{F}_t]$$

Note 15.6.5 (the fallacy of using real-world probability measure). Let S_T be a random variable. Let $f(S_T)$ be the payoff at time T. Let current time be 0. Further assume zero interest rate. One fallacy of pricing is given by

$$V(0) = E_P[f(S_T)],$$

where P is the real world measure.

- this method is appealing because we can usually estimate the distribution of S_T in the real world measure via historical data.
- However, this method does not take into account the risk-aversion nature of the market, in which market price should be higher than $E_P[f(S_T)]$.

15.6.6 Market model under different martingale measure

15.6.6.1 *Brownian motion in different measures*

Theorem 15.6.17 (changing dynamics between measures). *[8, p. 32] Consider a numeraire U and a numeraire N with its associated measure Q_U and Q_N measure. Assume a n-dimensional diffusion process X_t, **under measure** Q_U, has dynamics*

$$dX_t = \mu_X^U(X_t, t)dt + \sigma_X(X_t, t)dW_t^U,$$

where $\mu_X^U \in \mathbb{R}^n$ and $\sigma_X \in \mathbb{R}^{n\times n}$, $W_t^U \in \mathbb{R}^n$ is a n-dimensional correlated Brownian motion under measure Q_U, and $dW_t^U[dW_t^U]^T = \rho \in \mathbb{R}^{n\times n}$, $\sigma_X = diag(\sigma_{X,1}, \sigma_{X,2}, ..., \sigma_{X,n})$.[a]

Now assume the dynamics of X_t under measure Q_N is given by

$$dX_t = \mu_X^N(X_t, t)dt + \sigma_X(X_t, t)dW_t^N,$$

where $W_t^U \in \mathbb{R}^n$ is a n-dimensional correlated Brownian motion under measure Q_N.

Further assume the numeraire has dynamics, under measure Q_U, given by

$$\begin{aligned} dU_t &= \mu_U^U dt + \sigma_{N,t} dW_t^U \\ dN_t &= \mu_N^U dt + \sigma_{U,t} dW_t^U \end{aligned}$$

where $\sigma_{N,t}, \sigma_{U,t} \in \mathbb{R}^{1\times n}$.

It follows that

- $$\mu_X^N(X_t, t) = \mu_X^U(X_t, t) + \sigma_X(X_t, t)\rho(\frac{\sigma_N}{N_t} - \frac{\sigma_U}{U_t})^T,$$

  or equivalently,

  $$\mu_X^N(X_t, t) = \mu_X^U(X_t, t) + dX_t d(\log(\frac{N_t}{U_t})),$$

- $$dW_t^N = dW_t^U - \rho(\frac{\sigma_N}{N_t} - \frac{\sigma_U}{U_t})^T dt.$$

- *Moreover, if for any X_i/U is a martingale under Q_U, then X_i/N is a martingale under Q_N.*

[a] Note that the superscript denotes different measures and volatilities(the coefficients before Brownian motion) will not change under changing measure.

Proof. Note that we know that $dE_U[X_t] = \mu_X^U dt$. Then we have

$$\begin{aligned}
\mu_X^U dt &= E_U[dX_t] \\
&= E_N[dX_t \frac{dQ_U}{dQ_N}] \\
&= E_N[dX_t \frac{U(t+dt)N(t)}{U(t)N(t+dt)}] \\
&= E_N[dX_t \frac{\frac{U}{N}(t+dt)}{\frac{U}{N}(t)}] \\
&= E_N[dX_t(1+d(\frac{U}{N}))/(U/N)] \\
&= E_N[dX_t] + E_N[dX_t d(\frac{U}{N})/(U/N)] \\
&= E_N[dX_t] + E_N[dX_t d(\log(\frac{U}{N}))] \\
&= \mu_X^N dt + E[\sigma_X dW_t^N((\frac{\sigma_U}{U_t} - \frac{\sigma_N}{N_t})dW_t^N)^T] \\
&= \mu_X^N dt + \sigma_X \rho(\frac{\sigma_U}{U_t} - \frac{\sigma_N}{N_t})^T
\end{aligned}$$

Note that we will get the same result if we use

$$\frac{dQ_U}{dQ_N} = \frac{U(T)N(t)}{U(t)N(T)},$$

since dX_t only correlated with the $[t, t+dt]$ part in $\frac{U(T)}{N(T)}$ in the $O(t^{1/2})$ scale. □

Note 15.6.6 (parameter constraint on drifting parameters).

- Note that under no-arbitrage condition, the drifting parameters $\mu_i, i = 1,2,...,N$ under different measures are not unrelated. There relationship is given by Theorem 15.6.20.
- Also see the following example.

Example 15.6.6. Consider three assets X, U and N. Assume under the measure associated with numeraire associated with U, we have

$$\begin{aligned} dX_t/X_t &= \mu_X^U dt + \sigma_X dW_1^U \\ dU_t/U_t &= \mu_U^U dt + \sigma_U dW_2^U \\ dN_t/N_t &= \mu_N^U dt + \sigma_N dW_3^U \end{aligned}$$

where W_1^U, W_2^U, W_3^U are correlated Brownian motions with $dW_i dW_j = \rho_{ij} dt$.

Note that $\mu_X^U, \mu_U^U, \mu_N^U$ are satisfying

$$\begin{aligned} \mu_X^U - \mu_U^U &= \rho_{XU}\sigma_X\sigma_U - \sigma_U^2 \\ \mu_N^U - \mu_U^U &= \rho_{NU}\sigma_N\sigma_U - \sigma_U^2 \end{aligned}$$

such that X/U and N/U are martingales.

Under the measure Q_N associated with numeraire N, we assume

$$\begin{aligned} dX_t/X_t &= \mu_X^N dt + \sigma_X dW_1^N \\ dU_t/U_t &= \mu_U^N dt + \sigma_U dW_2^N \\ dN_t/N_t &= \mu_N^N dt + \sigma_N dW_3^N \end{aligned}$$

where W_1^N, W_2^N, W_3^N are correlated Brownian motions with $dW_i dW_j = \rho_{ij} dt$.

When changing to the measure Q_N associated with numeraire N, we have

$$\begin{aligned} \mu_X^N &= \mu_X^U + \sigma_X\rho_{XN}\sigma_N - \sigma_X\rho_{XU}\sigma_U \\ \mu_N^N &= \mu_N^U + \sigma_N^2 - \sigma_N\rho_{NU}\sigma_U \end{aligned}$$

Then it can be shown that

$$\begin{aligned} \mu_X^N - \mu_N^N &= \mu_X^U - \mu_N^U + \sigma_X\rho_{XN}\sigma_N - \sigma_X\rho_{XU}\sigma_U - \sigma_N^2 + \sigma_N\rho_{NU}\sigma_U \\ &= \sigma_X\rho_{XN}\sigma_N - \sigma_N^2; \end{aligned}$$

that is, X/N will be a martingale under Q_N.

Lemma 15.6.7. *Consider a SDE under real probability measure given by*

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

Let the SDE of a zero-coupon bond under real probability measure be given by

$$dP(t,T)/P(t,T) = b(t)dt + \Sigma(t,T)dW_t.$$

Let the SDE of an arbitrary asset $S(t)$ under risk-neutral probability measure be given by

$$dS(t,T)/S(t,T) = r(t)dt + \sigma_S(t)dW_t.$$

It follows that

- *Under risk-neutral measure Q,*

$$dX_t = rdt + \sigma dW_t^Q.$$

  where W_t^Q is a Brownian motion under Q.
- *Under risk-neutral measure Q,*

$$dW_s = dW_s^Q - \frac{\mu(s) - r(s)}{\sigma(s)}ds.$$

- *Under forward measure Q_T with respect to $T-$maturity zero-coupon bond,*

$$dX_t = (r + \Sigma^2)dt + \sigma dW_t^T,$$

  where W_t^T is a Brownian motion under Q_T.
- *Under forward measure Q_T with respect to $T-$maturity zero-coupon bond,*

$$dW_s = dW_s^T - \frac{\mu - (r + \Sigma^2(t,T))}{\sigma}ds.$$

Lemma 15.6.8 (changing dynamics from risk-neutral measure to forward measure). *Consider a SDE under risk-neutral probability measure given by*

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

Let the SDE of a zero-coupon bond under real probability measure be given by

$$dP(t,T)/P(t,T) = b(t)dt + \Sigma(t,T)dW_t.$$

It follows that

- *Under forward measure Q_T with respect to $T-$maturity zero-coupon bond,*

$$dX_t = (r + \sigma^2)dt + \sigma dW_t^T,$$

where W_t^T is a Brownian motion under Q_T.

- *Under forward measure Q_T with respect to $T-$maturity zero-coupon bond,*

$$dW_s = dW_s^T - \frac{\mu - (r + \sigma^2)}{\sigma} ds.$$

- *Under forward measure,*

$$\frac{X(t)}{P(t,T)}$$

is a martingale.

15.6.6.2 *SDE approach: Single source of uncertainty*

Theorem 15.6.18 (change of numeraire, single source uncertainty). *[5, p. 661] Consider two Ito process f and g of assets with common source of uncertainty, given by*

$$df = \mu_f f dt + \sigma_f f dz,$$

$$dg = \mu_g g dt + \sigma_g g dz,$$

where z is the Brownian motion under the original measure. Assume g is always positive. It follows that

- *Under no arbitrage condition, there exists an measure Q_g (i.e. set market price of risk to σ_g)such that under such measure, we have*

$$df = (r + \sigma_g \sigma_f) f dt + \sigma_f f d\hat{z},$$

and

$$dg = (r + \sigma_g^2) g dt + \sigma_g g d\hat{z},$$

where r is the risk-free rate and $\hat{z}$ is a Brownian motion under the new measure Q_g.

- *Under the new measure Q_g,*

$$d(\frac{f}{g}) = (\sigma_f - \sigma_g)\frac{f}{g} d\hat{z},$$

that is, the quantity

$$\frac{f}{g}$$

is a martingale under the this measure.

- *Under the new measure Q_g, the Brownian motion z in original measure is given by*

$$dz = d\hat{z} + (\sigma_g - \lambda)dt,$$

where λ is the market price of risk.

Note that if set market price of risk to σ_f, we have

$$d(\frac{f}{g}) = (\sigma_f - \sigma_g)^2 \frac{f}{g} dt + (\sigma_f - \sigma_g)\frac{f}{g} dz.$$

Proof. From the lemma (Lemma 15.5.1), we know that the drift parameters of the two SDE with single common sources are constrained as:

$$\frac{\mu_g - r}{\sigma_g} = \frac{\mu_f - r}{\sigma_f} = \lambda,$$

where λ is unknown function resulting from market forces equilibrium. From Girsanov theorem(Theorem 6.7.6), we can define random quantity

$$Z = \exp(\int_0^T (\sigma_g - \lambda)dz(s) - \frac{1}{2}\int_0^T (\sigma_g - \lambda)^2 ds)$$

to generate a new measure Q_g such that under Q_g

$$dz = d\hat{z} + (\sigma_g - \lambda)dt,$$

Then under Q_g,

$$df = (r + \lambda\sigma_f)f dt + \sigma_f f dz \Leftrightarrow df = (r + \lambda\sigma_f)f dt + \sigma_f f(d\hat{z} + (\sigma_g - \lambda)dt)$$

and

$$dg = (r + \lambda\sigma_g)g dt + \sigma_g g dz \Leftrightarrow dg = (r + \lambda\sigma_g)g dt + \sigma_g g(d\hat{z} + (\sigma_g - \lambda)dt),$$

Arrange terms, we have

$$df/f = (r + \sigma_f\sigma_g)dt + \sigma_f d\hat{z}$$

and

$$dg/g = (r + \sigma_g^2/2)dt + \sigma_g d\hat{z}$$

Or equivalently, we have

$$d(\ln f) = (r + \sigma_f \sigma_g - \sigma_f^2/2)dt + \sigma_f d\hat{z}$$

and

$$d(\ln g) = (r + \sigma_g^2/2)dt + \sigma_g d\hat{z}$$

Therefore, under Q_g

$$d(\ln f - \ln g) = -\frac{(\sigma_f - \sigma_g)^2}{2}dt + (\sigma_f - \sigma_g)d\hat{z}$$

and

$$d(\frac{f}{g}) = (\sigma_f - \sigma_g)\frac{f}{g}d\hat{z}.$$

□

Note 15.6.7. Note that under the measure Q_g, $1/g$ is not a martingale but f/g is.

Remark 15.6.18. It can be showed that Lemma 15.5.1, under the market price of risk σ_g, we have

$$\frac{r + \sigma_g^2 - r}{\sigma_g} = \frac{r + \sigma_g \sigma_f - r}{\sigma_f} = \sigma_g.$$

Example 15.6.7. [14, p. 169] A stock,S_t, follows the Black-Scholes model. A derivative D_t pays $S_T I_{S_T > K}$ at time T. Develop a formula for its current price at time t.

Solution: Use the stock as the numeraire, then under this new measure Q_S,

$$\frac{D(T)}{S(T)}$$

is a martingale. Therefore,

$$\frac{D(t)}{S(t)} = E_S[\frac{D(T)}{S(T)}|\mathcal{F}_t] = E_S[I_{S_T > K}|\mathcal{F}_t].$$

Note that under measure Q_S, the dynamics of $S(t)$ follows

$$dS(t) = (r + \sigma^2)S(t)dt + \sigma S(t)dW(t).$$

This allows us to to obtain

$$\log S(T) = \log S(t) + (r + \frac{1}{2}\sigma^2)(T - t) + \sigma\sqrt{T - t}Z,$$

with Z a standard normal variable. Then

$$D(t) = S(t)P(S_T > K),$$

can be evaluated.

Example 15.6.8. [14, p. 170] A stock,S_t, follows the Black-Scholes model. A derivative D_t pays $S_T \log S_T$ at time T. Develop a formula for its current price at time t.

Solution: Use the stock as the numeraire, then under this new measure Q_S,

$$\frac{D(T)}{S(T)}$$

is a martingale. Therefore,

$$\frac{D(t)}{S(t)} = E_S[\frac{D(T)}{S(T)}|\mathcal{F}_t] = E_S[\log S_T|\mathcal{F}_t].$$

Note that under measure Q_S, the dynamics of $S(t)$ follows

$$dS(t) = (r + \sigma^2)S(t)dt + \sigma S(t)dW(t).$$

This allows us to to obtain

$$\log S(T) = \log S(t) + (r + \frac{1}{2}\sigma^2)(T - t) + \sigma\sqrt{T - t}Z,$$

with Z a standard normal variable. Then

$$E_S[\log S(T)|\mathcal{F}_t] = \log S(t) + (r + \frac{1}{2}\sigma^2)(T - t),$$

and

$$D(t) = S(t)(\log S(t) + (r + \frac{1}{2}\sigma^2)(T - t)).$$

15.6.6.3 *SDE approach: Multiple sources of uncertainty*

Theorem 15.6.19 (change of numeraire, multiple sources of uncertainty). *[5, p. 668] Consider the SDE for assets $f_1, f_2, ..., f_M$ and g driven by n independent Brownian motions $z_1, z_2, ..., z_n$, given by,*

$$\begin{aligned}\frac{df_j}{f_j} &= [r + \sum_{i=1}^n \lambda_i \sigma_{f_j,i}]dt + \sum_{i=1}^n \sigma_{f_j,i} dz_i, j = 1,2,...,M \\ \frac{dg}{g} &= [r + \sum_{i=1}^n \lambda_i \sigma_{g,i}]dt + \sum_{i=1}^n \sigma_{g,i} dz_i\end{aligned}$$

where r is the risk-free rate and $\lambda_1, \lambda_2, ..., \lambda_n$ are market prices of risks. Assume g is always positive. It follows that

- *There exists a new measure Q_g such that*

$$d\hat{z}_i = dz_i + \sigma_{g,i} dt, i = 1,2,...,n;$$

- *Under the new measure Q_g Then*

$$d(\frac{f_j}{g}) = (\frac{f_j}{g})(\sum_{i=1}^n (\sigma_{f_j,i} - \sigma_{g,i}))d\hat{z}_i, j = 1,2,...,M.$$

Proof. From Girsanov theorem(Theorem 6.7.7), we can define random quantity

$$Z = \exp(\int_0^T \theta(s) \cdot dz(s) - \frac{1}{2}\int_0^T \|\theta(s)\|^2 ds),$$

where $\theta_i(s) = (\sigma_{g,i} - \lambda_i), i = 1,2,...,n.$ to generate a new measure Q_g such that under Q_g

$$dz_i = d\hat{z}_i + (\sigma_{g,i} - \lambda_i)dt, i = 1,2,...,n.$$

Therefore, under the measure Q_g, we have

$$\begin{aligned}d\log f_j &= [r + \sum_{i=1}^n (\sigma_{g,i}\sigma_{f_j,i} - \sigma_{f_j,i}^2/2)]dt + \sum_{i=1}^n \sigma_{f_j,i} d\hat{z}_i, j = 1,2,...,M \\ d\log g &= [r + \sum_{i=1}^n (\sigma_{g,i}\sigma_{g,i} - \sigma_{g,i}^2/2)]dt + \sum_{i=1}^n \sigma_{g,i} d\hat{z}_i\end{aligned}$$

so that

$$d\log\frac{f_j}{g} = d(\log f_j - \log g) = [\sum_{i=1}^n (\sigma_{g,i}\sigma_{f_j,i} - \sigma_{g,i}^2/2 - \sigma_{f_j,i}^2/2)] + \sum_{i=1}^n (\sigma_{f_j,i} - \sigma_{g,i})dz_i, j = 1,2,...,M.$$

□

Example 15.6.9. [14, p. 272]Lemma 18.3.12[The Margrabe option] Two stocks,$S_1(t)$ and $S_2(t)$, follows the SDE in real-world measure.

$$dS_1(t) = (r + \lambda_1\sigma_1)S_1(t)dt + \sigma_1 S_1(t)dW_1(t)$$
$$dS_2(t) = (r + \rho\lambda_1\sigma_2 + \lambda_2\sqrt{1-\rho^2}\sigma_2)S_2(t)dt + \rho\sigma_2 S_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t)$$

where W_1 and W_2 are independent Brownian motions.

A derivative D_t pays $\max(S_2(T) - S_1(T), 0)$ at time T. Develop a formula for its current price at time t.

Solution:

Use the stock S_1 as the numeraire, then under this new measure Q_S,

$$\frac{D(T)}{S_1(T)}$$

is a martingale. Therefore,

$$\frac{D(t)}{S_1(t)} = E_{S_1}[\frac{D(T)}{S(T)}|\mathcal{F}_t] = E_{S_1}[\max(\frac{S_2(T)}{S_1(T)} - 1)|\mathcal{F}_t].$$

Note that under measure Q_S (we take $\lambda_1 = \sigma_1, \lambda_2 = 0$ following Theorem 15.6.19), the dynamics of $S_1(t)$ and $S_2(t)$ follows

$$dS_1(t) = (r + \sigma_1^2)S_1(t)dt + \sigma_1 S_1(t)dW_1(t)$$
$$dS_2(t) = (r + \rho\sigma_1\sigma_2)S_2(t)dt + \rho\sigma_2 S_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t)$$
$$d\frac{S_2}{S_1} = \frac{S_2}{S_1}((\rho\sigma_2 - \sigma_1)dW_1 + \sqrt{1-\rho^2}\sigma_2\sigma dW_2)$$

Denote $Y = \frac{S_2}{S_1}$, then Y is a geometric Brownian motion with volatility $\sigma_Y = \sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$.

Then we have $D(t) = S_1(t)E_{S_1}[\max(Y(T) - 1, 0)]$, which can be evaluated.

15.6.6.4 *Market price of risk under different measures*

Theorem 15.6.20 (No arbitrage constraints on parameters of dynamics under different measures). *Consider a market with risk-free asset with short rate r. Consider n assets as stochastic processes given as:*

$$\frac{df_i}{f_i} = a_i dt + \sum_{j=1}^{m} b_{ij} dz_j, i = 1, 2, ..., n.$$

where $z_j, j = 1, 2, ..., m$ are independent Wiener processes.

*Assume there is **no arbitrage** in the market. It follows that*

- *there exists (unnecessarily unique) $\lambda_1, ..., \lambda_m$*

  $$a_i - r = \sum_{j=1}^{m} b_{ij}\lambda_j, j = 1, 2..., m,$$

  such that

  $$\frac{df_i}{f_i} = (r + \sum_{j=1}^{m} \lambda_j b_{ij})dt + \sum_{j=1}^{m} b_{ij} dz_j, i = 1, 2, ..., n.$$

- *under any new equivalent measure Q such that*

  $$dz_i = d\hat{z}_i + \theta_i dt, i = 1, 2, ..., n,$$

  where $\hat{z}_i, i = 1, 2, ..., n$ are independent Brownian motions under Q, there exists $\lambda_1^Q, ..., \lambda_m^Q$

  $$a_i - r = \sum_{j=1}^{m} b_{ij}\lambda_j^Q, j = 1, 2..., m.$$

- *under this measure Q*

  $$\frac{df_i}{f_i} = (r + \sum_{j=1}^{m} \lambda_j^Q b_{ij})dt + \sum_{j=1}^{m} b_{ij} d\hat{z}_j, i = 1, 2, ..., n.$$

Proof. (1) See Theorem 15.5.1. (2)(3) Note that we can use Girsanov theorem(Theorem 6.7.7)to generate a new equivalent measure Q via random quantity

$$Z = \exp(\int_0^T \theta(s) \cdot dz(s) - \frac{1}{2}\int_0^T \|\theta(s)\|^2 ds),$$

where $\theta_i(s) = (\lambda_i^Q - \lambda_i), i = 1, 2, ..., n$. Under Q, we have

$$dz_i = d\hat{z}_i + (\lambda_i^Q - \lambda_i)dt, i = 1, 2, ..., n.$$

plug into

$$\frac{df_i}{f_i} = (r + \sum_{j=1}^{m} \lambda_j b_{ij})dt + \sum_{j=1}^{m} b_{ij} dz_j, i = 1, 2, ..., n,$$

we have

$$\frac{df_i}{f_i} = (r + \sum_{j=1}^{m} \lambda_j^Q b_{ij})dt + \sum_{j=1}^{m} b_{ij} d\hat{z}_j, i = 1, 2, ..., n.$$

□

Note 15.6.8 (implications)**.**

- In Theorem 15.5.1, we prove that under the real world probability measure, the drifting parameters has to satisfy

$$a_i - r = \sum_{j=1}^{m} b_{ij} \lambda_j, j = 1, 2..., m.$$

such that the model admits no arbitrage.
- In this theorem, we prove that under different equivalent measure, or under different numeraire associated measure, such parameter constraint relation always exist.

15.7 Black-Scholes framework

15.7.1 Canonical version of Black-Scholes model

Assumption 15.3 (Assumptions for canonical Black-Scholes model:). *[15, p. 41][7, p. 85][1]*

- *The asset price S follows geometric random walk.*
- *The risk free rate r and volatility σ are known.*
- *No transaction cost.*
- *The underlying asset pays no dividends.*
- ***No arbitrage**.*
- *Short selling is permitted.*
- *The trading of assets can take place continuously in time and amount.*
- *No bid-ask spreads.*

Theorem 15.7.1 (Black-Scholes equation). *[15, p. 41] Let $V(S(t), t)$ be the value of the derivative as a function of the asset price $S(t)$ and time t. Assume $S(t)$* ***under real world measure*** *is governed by*

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where W_t is the Brownian motion. Then V is governed by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0,$$

with final condition $V(S(T), T) = V_T(S(T))$ and boundary condition $V(S,t) = V_a(t)$ on $S = a$ and $V(S,t) = V_b(t)$ on $S = b$.

Proof. Use Ito's lemma(Lemma 6.3.1), we have

$$dV = \sigma S\frac{\partial V}{\partial S} dW_t + (\mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t})dt$$

Construct a portfolio $\Pi = V - \Delta S$, then

$$d\Pi = dV - \Delta dS = \sigma S(\frac{\partial V}{\partial S} - \Delta)dW_t + (\mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \mu\Delta S)dt$$

If we set $\frac{\partial V}{\partial S} = \Delta$, then the value of the portfolio Π will evolutes deterministically, we have

$$d\Pi = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt = r\Pi dt$$

as required by risk free portfolio valuation theorem(Corollary 15.2.1.2). Finally, we have

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

A simpler proof:

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2$$

then

$$d\Pi = dV - \Delta dS = \frac{\partial V}{\partial t}dt + (\frac{\partial V}{PaS} - \Delta)dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2$$

which implies $\Delta = \partial V / \partial S$. And

$$d\Pi = dV - \frac{\partial V}{\partial S}dS = \frac{\partial V}{\partial t}dt + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2 = \frac{\partial V}{\partial t}dt + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 dt = r(V - \frac{\partial V}{\partial S}S)dt.$$

□

Remark 15.7.1 (interpretation and applicability)**.**

- This formula applies to both European call and put.
- **The σ, r, μ can be a function of S_t and t, but they must be a deterministic function** .
- The growth rate of S does not enter into the equation.

Remark 15.7.2 (boundary and final conditions)**.**

- For an European call option with strike price E and expiry date T, the final condition is $C(S,T) = \max(S-E,0)$, and the boundary conditions are $C(0,t) = 0$(when $S = 0$, it will always stays at 0, and the call option is worthless.) and $C(\infty, t) \sim S$(when S is infinitely large, then C has infinite value for any finite E).
- For an European put option with strike price E and expiry date T, the final condition is $P(S,T) = \max(E-S,0)$, and the boundary conditions are $P(0,t) = Ee^{-r(T-t)}$(when $S = 0$, it will always stays at 0, and the put option is worthing the discounted payoff E.) and $P(\infty, t) \sim 0$(when P is infinitely large, then P has 0 value for any finite E. Since it is unwise to buy at an infinitely high price and sell at a finite price E).

Lemma 15.7.1 (martingale pricing under risk-neutral measure). *Consider a derivative V is governed by*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0.$$

Then there exists a measure Q, called ***risk-neutral measure****, under which the underlying asset dynamics is given by*

$$dS = rSdt + \sigma S dW_t$$

where W_t is a Brownian motion.

Moreover, the quantity $e^{-rt}V(t)$ is a martingale under risk-neutral measure Q, which enables martingale pricing formula given by

$$V(t) = E_Q[\exp(-r(T-t))V(S_T)|\mathcal{F}_t].$$

Proof. (1) The existence of such risk-neutral measure is directly from Feyman-Kac theorem (Theorem 15.8.1).

(2) To show $e^{-rt}V(t)$ is a martingale, we want to show it is a driftless SDE under Q.

$$\begin{aligned}
&d(\exp(-rt)V(t,S_t,\sigma_t)) \\
&= -\exp(-rt)rVdt + \exp(-rt)dV \\
&= -\exp(-rt)rVdt + \exp(-rt)[(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \frac{\partial V}{\partial S}dS] \\
&= -\exp(-rt)rVdt + \exp(-rt)[rVdt + \sigma S\frac{\partial V}{\partial S}dW_t] \\
&= \exp(-rt)\sigma S\frac{\partial V}{\partial S}dW_t
\end{aligned}$$

where we use the Black-Scholes equation in the derivation. □

Note 15.7.1 (asset dynamics under real probability and risk-neutral probability). The asset dynamics under risk-neutral probability is given by

$$dS = rSdt + \sigma SdW_t$$

and

$$\begin{aligned}
dV &= (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \frac{\partial V}{\partial S}dS \\
&= rVdt + \frac{\partial V}{\partial S}\sigma SdW_t
\end{aligned}$$

where we use the Black-Scholes equation in the derivation.

Under real probability measure, the dynamics are

$$dS = \mu Sdt + \sigma SdW_t$$

and

$$\begin{aligned} dV &= (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \frac{\partial V}{\partial S} dS \\ &= rV + \frac{\partial V}{\partial S} S(\mu - r)dt + \frac{\partial V}{\partial S}\sigma S dW_t \end{aligned}$$

If we define $\lambda = (\mu - r)/\sigma$, we find that

$$\begin{aligned} dS/S &= (r + \lambda\sigma)dt + \sigma dW_t \\ dV/V &= (r + \lambda \frac{\partial V}{\partial S}\frac{\sigma S}{V})dt + \frac{\partial V}{\partial S}\frac{\sigma S}{V} dW_t \end{aligned}$$

which is consistent with the no-arbitrage condition for single source uncertainty dynamics(**??**).

15.7.2 Fundamental solution and risk-neutral measure

Lemma 15.7.2 (fundamental solution to Black-Scholes equation). *[16, p. 35] Assume constant short rate r. Given Black-Scholes equation*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0,$$

and let $V^\delta(t, S_t, s) > 1$ be the solution for final condition $V_T^\delta(S_T, s) = \delta(S_T - s)$. It follows that

- $$e^{-r(T-t)} = \frac{B(t)}{B(T)} = \int_{-\infty}^{\infty} V^\delta(t, S_t, s) ds$$
- *Define $p^\delta(t, S_t, s) \triangleq e^{r(T-t)} V^\delta(t, S_t, s)$, then*

  $$\int_{-\infty}^{\infty} p^\delta(t, S_t, s) ds = 1.$$

- *For any payoff function $V_T(S_T)$, we have its current value given by*

$$\begin{aligned} V(t) &= \int_{-\infty}^{\infty} V_T(S_T = s) V^\delta(t, S_t, S_T = s) ds \\ &= e^{-r(T-t)} \int_{-\infty}^{\infty} V_T(S_T = s) p^\delta(t, S_t, S_T = s) ds \\ &= e^{-r(T-t)} E_Q[V_T(S_T)|\mathcal{F}_t] \end{aligned}$$

where E_Q denotes taking expectation with respect to p^δ.

Proof. (1)(2) Use the fact that risk-free asset grows at rate r at all world states. (3) use linearity of pricing. □

Remark 15.7.3 (interpretation).

- This lemma shows that the nature of risk-neutral probability measure, which is the state price resulted from market equilibrium based on current market condition(time t and stock price S_t).
- The risk-neutral probability/state price therefore will depends on time t and stock price S_t.

15.7.3 Asset paying dividends

15.7.3.1 *Continuous proportional dividend*

Lemma 15.7.3 (martingale properties of asset paying continuous proportional dividend). *Consider an asset S_t following geometric Brownian motion with dividend rate D. Let Q denote the risk-neutral measure. Then,*

- $$e^{-(r-D)t} S_t = E_Q[e^{-(r-a)T} S_T | \mathcal{F}_t],$$
  that is, the random process $e^{-(r-a)T} S_T$ is a martingale with respect to Q.
- $$e^{-r(T-t)} E_Q[S_T|\mathcal{F}_t] = S_t e^{-D(T-t)}$$
- *(Put-call parity)*
  $$C(K, T, S_t) - P(K, T, S_t) = S_t e^{-D(T-t)} - K e^{-r(T-t)}$$

Proof. (1)(2) recall that under risk-neutral measure, the asset dynamics is given by(Lemma 15.6.4)

$$dS_t/S_t = (r - D)dt + \sigma dW_t.$$

Then use the expectation property of geometric Brownian motion(Corollary 6.3.3.1). (3)

$$\begin{aligned} C - P &= e^{-r(T-t)} E_Q[(S_T - K)^+ - (K - S_T)^+ | \mathcal{F}_t] \\ &= e^{-r(T-t)} E_Q[(S_T - K) | \mathcal{F}_t] \\ &= S_t e^{-D(T-t)} - K e^{-r(T-t)}. \end{aligned}$$

□

Remark 15.7.4 (only valid for deterministic short rate). Only when r is deterministic, we can separate $E_Q[e^{-(r-a)T} S_T | \mathcal{F}_t] = e^{-(r-a)T} E_Q[S_T | \mathcal{F}_t]$; when r is the random, we have to use forward measure method.

Theorem 15.7.2 (Black-Scholes equation with constant dividends). *[15, p. 90] Let $V(S(t), t)$ be the value of the derivative as a function of the asset price $S(t)$ and time t. Assume $S(t)$ is governed by*

$$dS_t = (\mu S_t - D_0 S_t)dt + \sigma S_t dW_t$$

where W_t is the Brownian motion. Then V is governed by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - D_0) S \frac{\partial V}{\partial S} - rV = 0$$

with final condition $V(S(T), T) = V_T(S(T))$ and boundary condition $V(S, t) = V_a(t)$ on $S = a$ and $V(S, t) = V_b(t)$ on $S = b$.

Proof. Use Ito's lemma(Lemma 6.3.1), we have

$$dV = \sigma S \frac{\partial V}{\partial S} dW_t + (\mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t})dt$$

Construct a portfolio $\Pi = V - \Delta S$, and **note that**

$$d\Pi = dV - \Delta(dS + \Delta S dt) = dV - \Delta dS - \Delta D_0 S dt$$

because the dividends are returned back to the stock holders(see Remark 15.6.15). Then we have

$$d\Pi = \sigma S(\frac{\partial V}{\partial S} - \Delta)dW_t + ((\mu - D_0 S)(\frac{\partial V}{\partial S} - \Delta) + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \Delta D_0 S)dt$$

If we set $\frac{\partial V}{\partial S} = \Delta$, then the value of the portfolio Π will evolutes deterministically, we have

$$d\Pi = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - \Delta D_0 S)dt = r\Pi dt = r(V - \Delta S)dt$$

as required by risk free portfolio valuation theorem(Corollary 15.2.1.2). Finally, we have

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}+(r-D_0)S\frac{\partial V}{\partial S}-rV=0$$

□

Remark 15.7.5 (relation to non-dividend paying model). If we make the change of variable as

$$V(S,t)=e^{-D_0(T-t)}V_1(S,t)$$

then the function V_1 satisfies the canonical Black-Scholes equation with r replaced by $r-D_0$; That is, if V satisfies the canonical Black-Scholes equation, then V_1 satisfies the dividend paying Black-Scholes equation.

15.7.3.2 *Discrete cash dividends*

Lemma 15.7.4 (European option). *link*

$$C_t=S_tN(d_1)-$$

15.7.4 Forwards and futures

15.7.4.1 *Forwards contract*

Lemma 15.7.5 (forward price with constant interest rate). *[15, p. 98] Let current time be* 0. *Let* F *be the forward/future price of an asset at* T. *Under martingale/risk-neutral measure* Q *and constant short rate* r*,we have*

$$0=E_Q[\exp(-rT)(S_T-F)]=S_0-F\exp(-rT),$$

or equivalently,

$$F=S_0\exp(rT).$$

If the asset is paying an dividend at rate q*, then*

$$F=S_0\exp((r-q)T).$$

Proof. (1) Note that the payoff of the contract is S_T-F, and the initial price is zero. Using martingale pricing theorem(Theorem 15.6.13), we have $0=E_Q[\exp -rT(S_T-F)]$. Also from the fact that $\exp(-rT)S_T$ is a martingale under risk-neutral measure , we have

$E_Q[\exp -rTS_T] = S_0$. (2) Similarly to (1). Note that $\exp(-(r-q)T)S_T$ is a martingale under risk-neutral measure , we have $E_Q[\exp -rTS_T] = S_0 \exp(qT)$. Therefore, $F = S_0 \exp((r-q)T)$. □

Theorem 15.7.3 (value of forward contract after initiation). *Let $F(t,T)$ be the forward/future price of an asset at time T. Consider a forward/futures contract initiated at t_0 with forward price $F(t_0,T)$. Let $V(t)$ be the value of the contract at time t. Assuming constant short rate r, we have*

- $$\begin{aligned} V(t) &= S(t) - F(t_0,T)\exp(-r(T-t)) \\ &= S(t) - S(t_0)\exp(r(t-t_0))) \end{aligned}$$
- *If the underlying is paying dividend at a rate q, we have*
$$\begin{aligned} V(t) &= S(t)\exp(-q(T-t)) - F(t_0,T)\exp(-r(T-t)) \\ &= S(t)\exp(-q(T-t)) - S(t_0,T)\exp(r(t-t_0))\exp(-q(T-t_0)) \end{aligned}$$

Proof. (1)

$$\exp(-rt)V(t) = E_Q[\exp(-rT)(S_T - F(t_0,T))|\mathcal{F}_t] \implies V(t) = S_t - F(t_0,T)\exp(-r(T-t)).$$

(2) Note that $\exp(-(r-q)T)S_T$ is a martingale under risk-neutral measure , we have $E_Q[\exp -r(T-t)S_T] = S_0\exp(-q(T-t))$. Then

$$\exp(-rt)V(t) = E_Q[\exp(-rT)(S_T - F(t_0,T))|\mathcal{F}_t] \implies V(t) = S_t\exp(-q(T-t)) - F\exp(-r(T-t)).$$

□

Note 15.7.2 (simulating forward contract value). To generate the trajectories of the forward contract value evolution, we can simulate the underlying, S_t, and then use the relation between S_t and $V(t)$ in Theorem 15.7.3.

- As showed in Figure 15.7.1, we simulate the underlying in the real world via model
$$dS_t/S_t = \mu dt + \sigma dW_t,$$
and then evaluate the forward contract via
$$V(t) = S_t - S(t_0)\exp(r(t-t_0)).$$

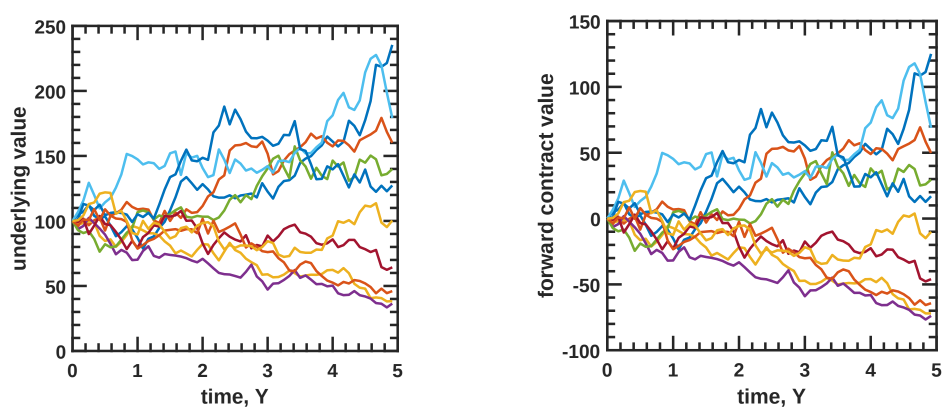
- Note that **we should not simulate the underlying in the risk-neutral measure.** This is because in the risk neutral measure $V(t)$ will have zero expected growth rate,i.e.,

$$V(t) = S(t_0)\exp(r(t-t_0))(M(t)-1),$$

where $M(t)$ is exponential martingale given by

$$M(t) = \exp(-\frac{1}{2}\sigma^2(t-t_0) + \sigma(W(t)-W(t_0))).$$

If a forward contract has zero expected return with risks, then no one wants to enter due to risk aversion even though it costs zero to enter.
- On the other hand, if we simulate the underlying in the real world measure where S_t will have a **different expected growth rate** from r and V_t will grow at a nonzero expected rate(could be positive or negative). Then the market participants will enter either short or long positions based on their estimation of the S_t dynamics.
- Also see [17, p. 145] for a discussion.



(a) Underlying price trajectories. (b) Forward contract value trajectories.

Figure 15.7.1: Demonstration of forward contract value evolution. Underlying simulation parameters: $\mu = 0.06, r = 0.02, \sigma = 0.2, q = 0.0, t_0 = 0$.

Lemma 15.7.6 (forward price with non-constant deterministic interest rate). *Let F be the forward price of an asset. Under the risk-neutral measure Q and non-constant deterministic short rate $r(t)$, we have*

$$0 = E_Q[D(T)(S_T - F)] = S_0 - FD(T),$$

or equivalently,

$$F = S_0/D(T),$$

where $D(t) = \exp(-\int_0^t r(s)ds)$

Proof. Same as the constant short rate case. □

Lemma 15.7.7 (forward price with non-constant deterministic interest rate). *Let F be the forward price of an asset. Under the risk-neutral measure Q and non-constant deterministic short rate $r(t)$, we have*

$$0 = E_Q[D(T)(S_T - F)] = S_0 - FD(T),$$

or equivalently,

$$F = S_0/D(T),$$

where $D(t) = \exp(-\int_0^t r(s)ds)$

Proof. Same as the constant short rate case. □

Lemma 15.7.8 (non-interest-rate asset forward price dynamics with stochastic interest rate). *Let $F(t,T)$ be the current forward price (with maturity T) of an asset S_t uncorrelated with interest rate. Under the risk-neutral measure Q and stochastic short rate $r(t)$, we have*

- $$F(t,T) = \frac{S_t}{P(t,T)}.$$
- ***If assuming independence between*** S_T ***and*** r, *then $F(t,T)$ is martingale under Q; that is*
$$F(t,T) = E_Q[F(T,T)|\mathcal{F}_t].$$

where $P(t,T)$ is the zero-coupon bond price.

Proof. (1)

$$\begin{aligned} 0 &= E_Q[\frac{1}{B(T)}(S_T - F(t,T))|\mathcal{F}_t] \\ &= \frac{S_t}{B(t)} - F(t,T)E_Q[\frac{1}{B(T)}|\mathcal{F}_t] \\ &= \frac{S_t}{B(t)} - F(t,T)P(t,T) \end{aligned}$$

(2)

$$\begin{aligned}0 &= E_Q[\frac{1}{B(T)}(S_T - F(t,T))|\mathcal{F}_t] \\ &= E_Q[\frac{1}{B(T)}\mathcal{F}_t]E_Q[(S_T - F(t,T))|\mathcal{F}_t] \\ \implies E_Q[(S_T - F(t,T))|\mathcal{F}_t] &= 0 \\ E_Q[F(T,T)|\mathcal{F}_t] &= F(t,T)\end{aligned}$$

Note that we use the fact that $(S_T - F(t,T))$ is independent of r(which is usually highly unlikely in real world). □

Lemma 15.7.9 (evolution of forward rate agreement). *Consider a forward contract entered at t_0 with fair rate K. At future time $t_1, t_0 < t_1 < T$, the value of the forward contract is given by*

$$V(t_1) = P(t_1,T)(F(t_0,T) - F(t_1,T)),$$

where $F(t_1,T)$, $F(t_0,T)$ are the fair forward value at t_1 and t_0 and $P(t,T)$ is the T zero coupon bond price at t. ***Note that $P(t_1,T)$, $F(t_1,T)$ and $V(t_1)$ are all random quantities at time t_0.***

Proof. For a contract entered at time t_0, we agree to pay $F(t_0,T)$ at time T to exchange for the asset $S(T)$. At time t_1, a new forward agreement will allow us to pay $F(t_1,T)$ to exchange for the same asset $S(T)$ at time T. The payment $F(t_0,T)$ at time T is worth $P(t_1,T)$ at time t_1, the asset $S(T)$ is worth $P(t_1,T)S(T) = P(t_1,T)F(t_0,T)$. Therefore, the value of the contract at time t_1 is to add the gain of $P(t_1,T)F(t_0,T)$ and subtract the loss $P(t_1,T)F(t_1,T)$, given by

$$V(t_1) = P(t_1,T)(F(t_0,T) - F(t_1,T)).$$

□

15.7.4.2 *Futures contract*

Lemma 15.7.10 (futures price dynamics under risk-neutral measure). *[18, p. 29] Under the risk-neutral measure Q, the underlying asset dynamics is given by*

$$dS_t/S_t = rdt + \sigma dW_t.$$

Consider a futures price $Fur(t)$ on the underlying S_t. It follows that

- *$Fur(t)$ is governed by*

$$\frac{\partial Fur}{\partial t} + \sigma^2 S^2 \frac{\partial^2 Fur}{\partial S^2} + \frac{\partial Fur}{\partial S} rS = 0,$$

with terminal condition $Fur(T, S_T) = S_T$ at maturity T.
- *Assume deterministic short rate $r(t)$, then $Fur(t) = E_Q[S_T|\mathcal{F}_t] = S_t \exp(\int_t^T r(s)ds)$. Moreover, $Fur(t)$ is martingale under Q; that is,*

$$Fur(t;T) = E_Q[S_T|\mathcal{F}_t] = E_Q[Fur(T;T)|\mathcal{F}_t],$$

where T is the maturity time.
- *The dynamics of $Fur(t)$, under risk-neutral measure Q, is given by*

$$dFur(t) = \sigma(t) Fur(t) dW_t.$$

Proof. (1) (a)Consider an investor has a marginal money account M continuous reinvesting in money market and handling the mark-to-market requirement of the futures. Note that the futures contract has no value since the mark-to-market gain/loss will go to the marginal money account.

The dynamics of the money account M is given by

$$dM = rMdt + dFur = rMdt + (\frac{\partial Fur}{\partial t} + \sigma^2 S^2 \frac{\partial^2 Fur}{\partial S^2} + \frac{\partial Fur}{\partial S} rS)dt + \frac{\partial F}{\partial S} \sigma S dW_t.$$

Under risk-neutral measure, the growth rate of any portfolio should be r(Theorem 15.5.4), therefore

$$\frac{\partial Fur}{\partial t} + \sigma^2 S^2 \frac{\partial^2 Fur}{\partial S^2} + \frac{\partial Fur}{\partial S} rS = 0.$$

(b) Another way to derive is to construct riskless portfolio $\Pi = M - \Delta S$, where $\Delta = \partial F/\partial S$. Then

$$d\Pi = (M - \partial F/\partial S S) r dt \implies \frac{\partial Fur}{\partial t} + \sigma^2 S^2 \frac{\partial^2 Fur}{\partial S^2} + \frac{\partial Fur}{\partial S} rS = 0.$$

(2) From Feyman-Kac theorem (Theorem 15.8.1),

$$Fur(t) = E_Q[Fur(T)|\mathcal{F}_t] = E_Q[S_T|\mathcal{F}_t].$$

$$
\begin{aligned}
Fur(t) &= E_Q[S_t \exp(\int_t^T r(s)ds) \exp(\int_t^T \sigma(s)ds - \frac{1}{2}\int_t^T \sigma(s)^2 ds)|\mathcal{F}_t] \\
&= S_t \exp(\int_t^T r(s)ds) E_Q[\exp(\int_t^T \sigma(s)ds - \frac{1}{2}\int_t^T \sigma(s)^2 ds)|\mathcal{F}_t]. \\
&= S_t \exp(\int_t^T r(s)ds)
\end{aligned}
$$

where we use the fact that $\exp(\int_t^T \sigma(s)ds - \frac{1}{2}\int_t^T \sigma(s)^2 ds)$ is a martingale.

(3) (a)

$$
\begin{aligned}
dFur(t) &= d(S_t \exp(\int_t^T r(s)ds)) \\
&= \exp(\int_t^T r(s)ds) dS_t - S_t \exp(\int_t^T r(s)ds) r(t) dt \\
&= \exp(\int_t^T r(s)ds) S_t \sigma(t) dW_t \\
&= Fur(t)\sigma(t) dW_t.
\end{aligned}
$$

(b) From Ito lemma,

$$dFur(t) = (\frac{\partial Fur}{\partial t} + \frac{\partial^2 Fur}{\partial S^2}\sigma^2 S^2 + \frac{\partial Fur}{\partial S} rS_t)dt + \frac{\partial Fur}{\partial S}\sigma S_t dW_t = \frac{\partial Fur}{\partial S}\sigma S_t dW_t = Fur(t)\sigma(t)dW_t$$

where we used (1) to simplify and note that $\frac{\partial Fur}{\partial S} = \exp(\int_t^T r(s)ds)$. □

15.8 Connection between Black-Scholes framework and martingale pricing

15.8.1 Pricing theory via Feyman Kac theorem

Theorem 15.8.1 (Feyman Kac theorem). *Consider the 1D parabolic*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)\frac{\partial^2 V}{\partial S^2} + m(S,t)\frac{\partial V}{\partial S} - rV = 0.$$

The solution is given as

$$V(s,t) = E_Q[e^{\int_s^t r(\tau)d\tau} V(S_t,t)|\mathcal{F}_s]$$

where S_t is a stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS = mdt + \sigma dW_t$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau} V(S_t,t)$$

is a martingale under the measure Q.

Proof.

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= d(e^{\int_0^t r(\tau)d\tau})V(t) + e^{\int_0^t r(\tau)d\tau} dV(t) + d(e^{\int_0^t r(\tau)d\tau})dV \\ &= -e^{\int_0^t -r(\tau)d\tau} r(t)Vdt + e^{-\int_0^t r(\tau)d\tau} dV. \end{aligned}$$

Use the fact that

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}(dS)^2\frac{\partial^2 V}{\partial S^2} = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt.$$

Then plug in dV, we have

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= e^{\int_0^t -r(\tau)d\tau}(-r(t)Vdt + \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt) \\ &= e^{\int_0^t -r(\tau)d\tau}\sigma\frac{\partial V}{\partial S}dW_t \end{aligned}$$

Therefore, $e^{\int_0^t r(\tau)d\tau} V(S_t, t))$ is a martingale. Then we can easily show using martingale property:

$$V(s) = E_Q[e^{-\int_s^t r(\tau)d\tau} V(S_t, t)|\mathcal{F}_s].$$

□

Theorem 15.8.2 (Feyman Kac theorem, alternative). *[9, p. 268] Assume X_t is governed by the following SDE*

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

Suppose we are given a payoff $V(X_T)$ at time T. Define

$$f(x,t) = E[e^{-\int_t^T r(u)du} V(X_T)|X_t = x]$$

for all $t \leq T$, where r is a non-random function. Then $f(x,t)$ is governed by

$$\frac{\partial}{\partial t} f(x,t) + \mu \frac{\partial}{\partial x} f(x,t) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} f(x,t) = rf(x,t)$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Moreover, $E[e^{-\int_0^T r(u)du} V(X_T)|X_t = x]$ is a martingale.

Proof. $e^{-\int_0^t r(u)du} f(x,t)$ is a martingale. Therefore,

$$\begin{aligned} d(e^{-\int_0^t r(u)du} f(x,t)) &= e^{-\int_0^t r(u)du} (f_t + f_x dX + \frac{1}{2} f_{xx} dX dX - rf) \\ &= e^{-\int_0^t r(u)du} (f_t + \mu f_x dt + \frac{1}{2} f_{xx} \sigma^2 - rf) dt + e^{-\int_0^t r(u)du} \sigma f_x dW. \end{aligned}$$

Set the drift term to zero, we get the PDE. □

Remark 15.8.1 (interpretation and financial applications in path-independent derivatives).

- Feyman Kac theorem shows that certain types of parabolic equation can be solve using stochastic differential equation method(by simulating trajectories and take expectations.) **Note that in parabolic differential equation S is not a random variable, S is simply a variable.**
- **Path independence**. Note that the function $V(S_t, t)$ can only be a function of instantaneous value S_t. If V is dependent on the history of S, then we cannot use Feyman Kac theorem.

Remark 15.8.2 (special case of $r = 0$). When $r = 0$, the dynamics of S will not change(i.e. Q will not change), then $V(S_t, t)$ is a martingale. And the parabolic equation becomes Kologorov backward equation.

Theorem 15.8.3 (Kolmogorov backward equation, expectation pricing). *Assume X_t is governed by the following SDE*

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

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t} f(x,t) = \mu \frac{\partial}{\partial x} f(x,t) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} f(x,t)$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t,t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t,t)dt.$$

then we have (use the notation $E[f(X_s,s)|X_t = t] = E_{x,t}[f(X_s,s)]$)

$$\begin{aligned} E_{x,t}[f(X_{t+dt}, t+dt)] &= E_{x,t}[f(x + dX_t, t+dt)] \\ &\approx E_{x,t}[f(x,t) + \partial_x f(x,t)dX_t + \frac{1}{2}\partial_x^2 f(x,t)dX_t^2 + \partial_t f(x,t)dt] \\ &= f(x,t) + \partial_x f(x,t)E_{x,t}[dX_t] + \frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2] + \partial_t f(x,t)dt \\ &= f(x,t) + dt(\partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t)) \end{aligned}$$

$$E_{x,t}[f(X_{t+dt}, t+dt)] = f(x,t) \implies \partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t) = 0$$

□

Remark 15.8.3 (interpretation).

- We can interpret $f(x,t)$ as the price at t and the state is at x. When we cannot hedge the risk, we can price the asset by the expected value of payoff with respect to the real probability. Note that such pricing method does not take into account of the risk-aversion.
- The only difference to Black-Scholes is the extra source decreasing term.

Theorem 15.8.4 (Kolmogorov backward equation with discount). *Assume X_t is governed by the following SDE*

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

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[e^{-\int_t^T r(tau)d\tau} V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t} f(x,t) = \mu \frac{\partial}{\partial x} f(x,t) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} f(x,t) - rf$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t,t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t,t)dt.$$

then we have (use the notation $E[f(X_s,s)|X_t = t] = E_{x,t}[f(X_s,s)]$)

$$\begin{aligned} E_{x,t}[(1-rdt)f(X_{t+dt}, t+dt)] &= E_{x,t}[(1-rdt)f(x+dX_t, t+dt)] \\ &\approx (1-rdt)E_{x,t}[f(x,t)+\partial_x f(x,t)dX_t + \frac{1}{2}\partial_x^2 f(x,t)dX_t^2 + \partial_t f(x,t)dt] \\ &= f(x,t) - rf(x,t)dt + \partial_x f(x,t)E_{x,t}[dX_t] + \frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2] + \partial_t f(x,t)dt \\ &= f(x,t) - rf(x,t)dt + dt(\partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t)) \end{aligned}$$

$$E_{x,t}[(1-rdt)f(X_{t+dt}, t+dt)] = f(x,t)$$

$$\implies \partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t) - rf(x,t) = 0$$

□

Theorem 15.8.5 (Kolmogorov backward equation, multi-dimensional version). *Assume X_1, X_2 is governed by the following SDE*

$$dX_1(t) = \mu_1 dt + \sigma_1 dW_1(t).$$

and

$$dX_2(t) = \mu_2 dt + \sigma_2 dW_2(t).$$

Suppose we are given a payoff $V(X_1(T), X_2(T))$ at time T. Define

$$f(x_1, x_2, t) = E[e^{-\int_t^T r(\tau)d\tau} V(X_1(T), X_2(T)|X_1(t) = x_1, X_2(t) = x_2]$$

for all $t \leq T$. Then $f(x_1, x_2, t)$ is governed by the **Kolmogorov backward equation** *given as*

$$-\frac{\partial}{\partial t} f(x_1, x_2, t) = \mu_1 \frac{\partial}{\partial x_1} f + \mu_2 \frac{\partial}{\partial x_2} f + + \frac{1}{2}\sigma_1^2 \frac{\partial^2}{\partial x_2^2} f + \frac{1}{2}\sigma_2^2 \frac{\partial^2}{\partial x_2^2} f + \sigma_1\sigma_2 \frac{\partial^2}{\partial x_1 \partial x_2} f - rf$$

for $t \leq T$, subject to the final condition $f(x_1, x_2, T) = V(X_1(T), X_2(T))$.

Proof. Similar to 1D case. □

Theorem 15.8.6 (Feyman Kac theorem, multi-dimensional). *Consider the multidimensional parabolic*

$$\frac{\partial V}{\partial t} + \sum_{i=1}^{N} \mu_i \frac{\partial V}{\partial S} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_{ij} \frac{\partial^2 V}{\partial S_i \partial S_j} + -rV = 0$$

where $\gamma_{ij} = \sum_{k=1}^{N} \sigma_{ik}\sigma_{jk}$. The solution is given as

$$V(s,t) = E_Q[e^{\int_s^t r(\tau)d\tau} V(S_t, t) | \mathcal{F}_s]$$

where S_t is a N dimensional stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS_i = \mu_i dt + \sum_{j=1}^{N} \sigma_{i,j} dW_j(t), i = 1, 2, ..., N$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau} V(S_t, t)$$

is a martingale under the measure Q.

Proof. Similar to 1D case. □

15.8.2 Model dynamics under risk-neutral measure

Lemma 15.8.1 (stock dynamics under risk-neutral measure). *Assume the stock S_t in the real-world probability measure follows*

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where W_t is a Wiener process. Then under the risk-neutral measure Q, the stock dynamics will be

$$dS_t = rS_t dt + \sigma S_t d\hat{W}_t,$$

where $\hat{W}_t$ is a Brownian motion under risk-neutral measure Q, no matter what original μ is.

Proof. Based on the definition of risk-neutral measure(Definition 15.6.7), it can be show that

$$d(D(t)S_t) = D(t)dS_t + S_t dD(t) = D(t)(rS_t dt + \sigma S_t dW_t) - rD(t)S_t dt = D(t)\sigma S_t dW_t.$$

□

15.9 Notes on bibliography

For no-arbitrage theory, see [19].

For treatment from economical perspective, see [20] [21].

For martingale methods, see [22].

For PDE methods, see [22][23].

For incomplete markets, see [24].

For the differences between real world measure and risk-neutral measure, see [25].

An excellent book of "P" method for risks and asset allocations, see [26].

For a comprehensive discussion on risks, see [27].

For statistical arbitrage, see [28].

BIBLIOGRAPHY

1. Black, F. & Scholes, M. The pricing of options and corporate liabilities. *Journal of political economy* **81,** 637–654 (1973).
2. Harrison, J. M. & Pliska, S. R. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications* **11,** 215–260 (1981).
3. Stefanica, D. *A primer for the Mathematics of Financial Engineering* (Fe Press, 2008).
4. Pliska, S. *Introduction to Mathematical Finance: Discrete Time Models* (Blackwell: Malden, MA, 1997).
5. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
6. Primbs, J. A. *A Factor Model Approach to Derivative Pricing* (CRC Press, 2016).
7. Derman, E., Miller, M. & Park, D. *The Volatility Smile* ISBN: 9781118959169. https://books.google.com/books?id=WKfBDAAAQBAJ (Wiley, 2016).
8. Brigo, D. & Mercurio, F. *Interest rate models-theory and practice: with smile, inflation and credit* (Springer Science & Business Media, 2007).
9. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
10. Shreve, S. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model* (Springer Science & Business Media, 2012).
11. Chin, E., Nel, D. & Ãlafsson, S. *Problems and Solutions in Mathematical Finance: Equity Derivatives* (John Wiley & Sons, 2017).
12. Oksendal, B. *Stochastic differential equations: an introduction with applications* (Springer Science & Business Media, 2013).
13. McInerney, D. & Zastawniak, T. *Stochastic Interest Rates* ISBN: 9781107002579. https://books.google.com/books?id=8JTwCAAAQBAJ (Cambridge University Press, 2015).
14. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
15. Lalley, S. *The mathematics of financial derivatives: a student introduction* (Chicago University Statistics 390/ Mathematical Finance 345, 2001).
16. Pelsser, A. *Efficient methods for valuing interest rate derivatives* (Springer Science & Business Media, 2000).

17. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
18. Clark, I. J. *Commodity Option Pricing: A Practitioner's Guide* (John Wiley & Sons, 2014).
19. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2004).
20. LeRoy, S. F. & Werner, J. *Principles of financial economics* (Cambridge University Press, 2014).
21. Mishkin, F. S. *The economics of money, banking, and financial markets* (Pearson education, 2007).
22. Pascucci, A. *PDE and martingale methods in option pricing* (Springer Science & Business Media, 2011).
23. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
24. Cernỳ, A. *Mathematical techniques in finance: tools for incomplete markets* (Princeton University Press, 2009).
25. Meucci, A. 'P'Versus' Q': Differences and Commonalities between the Two Areas of Quantitative Finance. *GARP Risk Professional*, 47–50 (2011).
26. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
27. Jorion, P. *Financial Risk Manager Handbook: FRM Part I / Wiley Finance* **pt. 2.** ISBN: 9781118017913. https://books.google.com/books?id=4ceVmGJSNpcC (Wiley, 2010).
28. Leung, T. & Li, X. *Optimal Mean Reversion Trading: Mathematical Analysis and Practical Applications* (World Scientific, 2015).

16

MARKET INSTRUMENT, INFORMATION, AND REPLICATION PRICING

16.1 Overview

Various kinds of financial instruments, including forwards, futures, swaps and options, are actively traded in the market. Their current prices and historical prices reveal critical information that can be employed to give no-arbitrage price for other financial instruments whose prices not observed in the market.

We use previously developed martingale pricing framework to interpret the price information and price other instruments via replication.

In general, forwards, futures and swaps reveal the market implied mean values of a financial random variable, such as interest rate, spot price, forward price, realized volatility, etc.

Vanilla option like European call and put with sufficient number strikes can enable reconstruction of distribution of the underlying at a specified time.

Forward starting products or options on options can further illuminate conditional probabiliy density of underlying asset stochastic process.

16.2 Market implied distributions

16.2.1 Implied mean values

16.2.1.1 *Futures and forwards*

16.2.1.2 *Swaps*

16.2.1.3 *Correlations*

16.2.2 Implied distributions

16.2.2.1 *Implied distribution and risk-neutral measure*

Definition 16.2.1 (implied distribution of underlying asset). *Consider a no-arbitrage market probabilitic model $(\Omega, \mathcal{F}, P)$ such that there exists an equivalent martingale measure, called risk-neutral measure Q. Let current time be 0. Let the asset spot price S_0 be observed in the market.*

- *The implied distribution of an asset value S_t at future time T is the distribution under measure Q* ***conditioning on*** S_0*.*
- *More concretely, given the call/puts market prices $C(K, T, S_0)$ at maturity T. The* ***implied distribution*** *f_{S_T} of S_T is the distribution function such that*

$$C(K, T, S_0) = e^{-rT} \int_0^\infty f_{S_T}(s)(S_T - K)^+ ds, \forall K;$$

  that is, implied distribution is the one that produces market prices of all calls/puts.

Remark 16.2.1 (uniqueness of implied distribution)**.**

- Given a finite set of liquid instruments such as calls/puts, there might exist multiple implied distribution such that the distribuiton of the underlying asset reprice these liquid instruments.
- When calls/puts are actively traded for all strikes $K \in \mathbb{R}^+$, the implied distribution can be uniquely determined by methods in Lemma 16.2.2, Lemma 16.2.4.

Theorem 16.2.1 (no-arbitrage pricing of path-independent options using implied distribution). *Let current time be 0. Let the asset spot price S_0 be observed in the market. Assume we have the implied distribution f_{S_T} for the underlying S_T derived from vanilla* *option market prices. Denote constant risk-free rate by r. It follows that an option with arbitrary payoff $V(S_T)$ at maturity has current value*

$$V_0 = e^{-rT} E_Q[f(S_T)] = e^{-rT} \int_{-\infty}^{\infty} V(s) f_{S_T}(s) ds.$$

Note that in reality, the implied probability for $S_T < 0$ is zero.

Proof. Directly use the definition of implied distribution and expectation under risk-neutral measure. □

16.2.2.2 *Implied distribution and risk-neutral measure*

Definition 16.2.2 (risk-neutral probability,state price). *Consider a probabilistic model $(\Omega, \mathcal{F}, P)$ for the real world. Let Q be an equivalent measure to P and satisfy*

$$Q(A) = V_0, \forall A \in \mathcal{F},$$

where V_0 is the current price of an asset with payoff $1_A e^{rT}$ at maturity T, and r is the constant interest rate.

*Then, we call Q **risk-neutral probability measure**, or (discounted) state price.*

Note 16.2.1 (constraints on risk-neutral probability/state price).

- The risk-neutral probability/state price can be any functional form as long as it is **equivalent**(Definition 15.6.7) to real-world probability; risk-neutral probability is the resulting equilibrium from market forces.
- **(why it should be equivalent?)** Suppose there is a positive price on a payoff event with zero probability to happen. Then the owner of this asset will try to sell it but no one wants to buy it since it cannot occur. As a result, the price will drop to zero. Similarly, suppose there is a zero price for a future payoff greater than 0. This is impossible since buyers will drive the price up.
- Each market participants have their own point of view of the market, their own risk-preference, and their own reserved prices for the assets. Their selling and buying behavior will determine the equilibrium price, or the state price. As long as there is few arbitrageurs, the market will be arbitrage free.
- When market reaches equilibrium and there is no arbitrage opportunities, we should be able to deduce the state price from the existing markets prices of all assets.

Note 16.2.2. [1, p. 177]

- (why is discounted state price) Q_S represents a 'true' state price, then a zero coupon bond has present value given by $e^{-rT} = \int_\Omega dQ_s$, which makes Q_s not satisfying the sum-to-one condition of probability measure. If we define $Q \triangleq e^{rT}Q$, then $\int_\Omega dQ = 1$.
- risk-neutral measure/probability si not the real probability describing likelihood of event occurrence; but it shares important common features of real probability, for example, non-negativeness, the sum of probability of all mutually exclusive event should equal 1.
- (pricing is model independent)We usually assume the real market has no arbitrage opportunities. When the market has prices along all strikes available, the linearity of pricing(Theorem 15.2.2) gives a no-arbitrage price on the spike(delta function) payoff the law of one price(Theorem 15.2.1) plus the convolution theorem enables us to derive a no-arbitrage price for options with arbitrage pay-off profile. **It does not rely on any extra model assumptions**(e.g., Brownian motion model).
- (constraints on risk-neutral probability/state price) The risk-neutral probability/state price can be any functional form as long as it is equivalent to real-world probability; risk-neutral probability is the resulting equilibrium from market forces.

Note 16.2.3 (implied distribution pricing is no-arbitrage pricing; relation to model-based pricing)**.**

- We usually assume the real market has no arbitrage opportunities. When the market has prices along all strikes available, the linearity of pricing(Theorem 15.2.2) gives a no-arbitrage price on the spike(delta function) payoff the law of one price(Theorem 15.2.1) plus the convolution theorem enables us to derive a no-arbitrage price for options with arbitrage pay-off profile. **It does not rely on any extra model assumptions**(e.g., Brownian motion model).
- Implied distribution is not real-world distribution, but is the state price. See 15.6.2,Theorem 15.3.1.
- When we assume an asset model(e.g., Brownian motion), we have to take great effort to find out the model dynamics under risk-neutral measure. Implied distribution pricing is theoretically more general than model based pricing.

Example 16.2.1. Assume that the risk neutral probability distribution for stock XYZ at the end of one year is uniform between 100 and 200. Assume the interest rate is 0.1. What is the value of a security that pays 1 if XYZ is between 140 and 151?

From the risk-neutral pricing theorem we have

$$V = e^{-0.1}\int_{140}^{151} f_{S_T}(x)dx = e^{-0.1}\frac{11}{100}$$

16.2.2.3 *Implied distribution from derivatives*

Lemma 16.2.1 (derivatives of calls as step and Dirac delta function). *[2, p. 34] Let current time be 0. Let the asset spot price S_0 be observed in the market. Let $C(K,T), P(K,T)$ be the call/put price with strike K and maturity T. It follows that*

- $$\frac{dC}{dK} = -H(S_T - K), \frac{dP}{dK} = H(K - S_T).$$
- $$\frac{d^2C}{dK^2} = \frac{d^2P}{dK^2} = \delta(S_T - K).$$
- *Let $D(K,T)$ be a digital call also on S_T with strike K. Then $D(K,T) = -\frac{dC}{dK}$.*

Proof. (informal)

- This is the definition of step function Lemma A.13.1.
- Let $C(x;K)$ denote function $C(x;K) \triangleq (x-K)^+$. Then,

$$\lim_{\epsilon\to 0}\frac{C(x;K-\epsilon) - 2C(x;K) + C(x;K+\epsilon)}{\epsilon^2} = \frac{d^2C}{dK^2} = \delta(x-K),$$

  where $\delta(x-K)$ is the Dirac delta function with spike located at K.
- $C(x;K+\epsilon) - C(x;K)$ is a call spread; long a call spread $C(x;K+\epsilon) - C(x;K)$ and short a call spread $C(x;K) - C(x;K-\epsilon)$ will generate a triangle spike with width 2ϵ and height ϵ and area ϵ^2.

For a review of delta function and step function, see Lemma A.13.1. □

Lemma 16.2.2 (implied distribution from digital call/put). *Let current time be 0. Let the asset spot price S_0 be observed in the market. Assume constant risk-free rate r. Consider market prices of digital call/puts available along a continuum of strikes K with the same maturity T. Then the implied distribution distribution*

$$f_{S_T}(S_T = K) = -e^{rT}\frac{dD(K)}{dK}.$$

where $D(K)$ is the current price of digital call at strike K.

Proof. From the martingale pricing theorem, we have

$$D(K) = e^{-rT} \int_0^\infty H(x-K) f_{S_T}(x) dx,$$

where $H(x-k)$ is the step function. Use the fact that derivative of a step function is Dirac delta(Lemma A.13.1), we have

$$D'(K) = e^{-rT} \int_0^\infty -\delta(x-K) f_{S_T}(x) dx = -e^{-rT} f_{S_T}(K).$$

□

Lemma 16.2.3 (implied distribution from Dirac-delta option). *Let current time be 0. Let the asset spot price S_0 be observed in the market. Consider a special call option on S_T with payoff $\delta(S_T - K)$. If the market prices of $Delta(K)$ are available along a continuum of strike K with same maturity T. Then the implied distribution is given by*

$$f_{S_T}(S_T - K) = e^{rT} Delta(K).$$

Proof.

$$Delta(K) = e^{-rT} E_Q[\delta(S_T - K)] = f_{S_T}(S_T = K).$$

□

Lemma 16.2.4 (implied distribution from European call/put). *Let current time be 0. Let the asset spot price S_0 be observed in the market. Assume constant risk-free rate r. Consider market prices of European call/put available along a continuum of strikes K with the same maturity T. Then the implied distribution distribution*

$$f_{S_T}(S_T = K) = e^{rT} \frac{d^2 C(K)}{dK^2} = e^{rT} \frac{d^2 P(K)}{dK^2},$$

where $C(K), P(K)$ are the prices of call/put at strike K.

Proof. From the martingale pricing formula, we have

$$\begin{aligned} C(K) &= e^{-rT} \int_K^\infty (s-K) f_{S_T}(s) ds \\ &= e^{-rT} \int_K^\infty (s-K) f_{S_T}(s) ds \\ \implies e^{rT} \frac{dC(K)}{dK} &= \int_K^\infty \frac{d(s-K)}{dK} f_{S_T}(s) ds \\ e^{rT} \frac{d^2 C(K)}{dK^2} &= f_{S_T}(K). \end{aligned}$$

The equality for the put is directly from Lemma 18.2.9. □

Note 16.2.4 (implied distribution is state price or risk-neutral measure)**.**

- We usually assume the real market has no arbitrage opportunities. Then no-arbitrage condition can ensure there existence a state price vector or risk-neutral measure. With the state price vector, we can price any payoff by

$$V = E_Q[V(T)],$$

which is from the linearity of pricing(Theorem 15.2.2) and the law of one price(Theorem 15.2.1), and **is irrelevant to any specific asset models**(e.g., Brownian motion model).
- The implied distribution is actually the state price or risk-neutral measure. See 15.6.2,Theorem 15.3.1.
- **(non-uniqueness)** Note that density function maybe only related to derivatives of option prices, denoted here by $C(K)$. Therefore, for different $C(K)$ profiles(e.g., $C(K)$ profiles differ only by a linear function), the density function will be the same.

Note 16.2.5 (features of implied distribution in practice)**.** [2, p. 35]

- The implied distribution usually shows that there exists a long tail on the lower price side, whereas the lognormal distribution stipulates there is a long tail on the higher price side.
- That is, contradicting to the model assumptions(the underlying follows a geometric Brownian motion with constant volatility), the market believes that there is higher likelihood that the market will experience a crash than a large rise.

16.2.2.4 *Implied distribution from derivatives: discrete case*

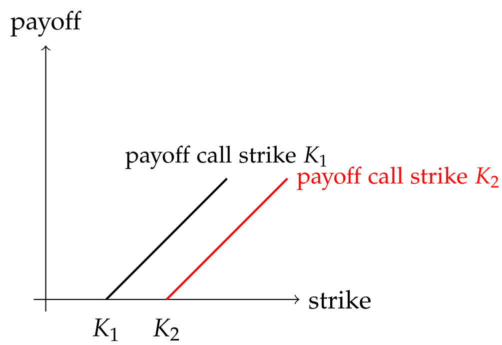
Lemma 16.2.5 (construct implied distribution from discrete call options). *Suppose we have observed market prices $V_0, V_1, V_2, ..., V_n$ for a series of call options on the same underlying S_T with the same expiry T but different strikes $K_0 < K_2 < K_3 < \cdots < K_n$.*

Suppose the implied density of S_T takes piecewise constant, that is, $Pr(S_T \in [K_{i-1}, K_i]) = p_i(K_i - K_{i-1}), i = 1, 2, ..., n$.

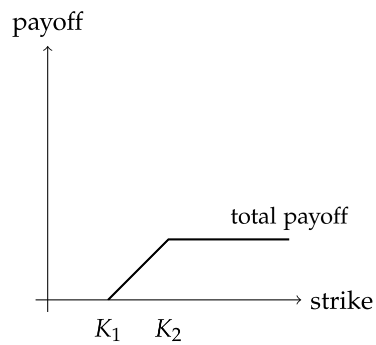
Then the following linear system

$$p_i \frac{K_i - K_{i-1}}{2} + \sum_{j>i}^{n} p_i(K_i - K_{i-1}) = V_i - V_{i-1}, i = 1, 2, ..., n$$

can be used to solve $p_i, i = 1, 2, ..., n$.



(a) payoff of calls at strike K_1, K_2



(b) payoff of $C_1 - C_2$

Figure 16.2.1: Use call option difference to find risk neutral density

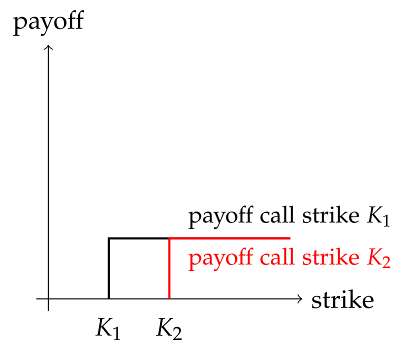
Lemma 16.2.6 (construct risk-neutral density from digital options). *Suppose we have observed market prices* $V_0, V_1, V_2, ..., V_n$ *for a series of call options on the same underlying* S_T *with the same expiry* T *but different strikes* $K_0 < K_2 < K_3 < \cdots < K_n$.

Suppose the implied density of S_T *takes piecewise constant, that is,* $Pr(S_T \in [K_{i-1}, K_i]) = p_i(K_i - K_{i-1}), i = 1, 2, ..., n$.

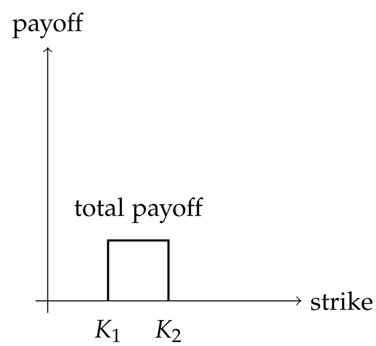
Then the following identities

$$p_i = \frac{V_i - V_{i-1}}{K_i - K_{i-1}}, i = 1, 2, ..., n$$

can be used to solve $p_i, i = 1, 2, ..., n$.



(a) payoff of calls at strike K_1, K_2



(b) payoff of $C_1 - C_2$

Figure 16.2.2: Use digital option difference to find risk neutral density

16.2.3 Application in replicating strategies

16.2.3.1 *Continuous static hedging path-independent derivatives*

Theorem 16.2.2 (payoff function decomposition lemma). *[3] Given a twice-continuously differentiable function $f(x)$(assume $f(x) = 0, \forall x < 0$), it has following decomposition:*

- $$f(x) = \int_0^{\kappa} f(K)\delta(x-K)dK + \int_{\kappa}^{\infty} f(K)\delta(x-K)dK;$$

- $$\begin{aligned} f(x) &= f(K)1(x<K)|_0^\kappa - \int_0^\kappa f'(K)1(x<K)dK \\ &\quad - f(K)1(x\geq K)|_\kappa^\infty + \int_\kappa^\infty f'(K)1(x\geq K)dK \\ &= f(\kappa) - \int_0^\kappa f'(K)1(x<K)dK + \int_\kappa^\infty f'(K)1(x\geq K)dK \end{aligned}$$

- $$f(x) = f(\kappa) + f'(\kappa)(x-\kappa) + \int_0^\kappa f''(K)(K-x)^+dK + \int_\kappa^\infty f''(K)(x-K)^+dK;$$

Proof. (1)straight forward; (2) use the fact(Lemma A.13.1) that

$$\frac{d1(x<K)}{dK} = \delta(K-x) = \delta(x-K)$$

and

$$\frac{d1(x\geq K)}{dK} = -\delta(K-x) = -\delta(x-K).$$

(3) use the step function definition

$$\max\{K-x,0\} = \frac{dH(K-x)}{dK}$$

□

Lemma 16.2.7 (Path-independent option hedging using digital calls). *Assume we have the implied distribution f_{S_T} for the underlying S_T derived from vanilla option market prices. It follows that an option with arbitrary payoff $V(S_T)$ at maturity has current value*

$$V_0 = \int_0^\infty V(s)f_{S_T}(s)ds = V(0) + \int_0^\infty V'(K)D(K)dK.$$

The hedging strategy is: long zero-coupon bonds in quantity $V(0)e^{-rT}$ and long all digital calls struck at $K>0$ in quantities $V'(K)$.

Proof.

$$\begin{aligned}
V_0 &= -\int_0^\infty V(K)D'(K)dK \\
&= -[V(K)D(K)]_0^\infty + \int_0^\infty V'(K)D(K)dK \\
&= V(0)D(0) - \int_0^\infty V'(K)D(K)dK \\
&= V(0) + \int_0^\infty V'(K)D(K)dK
\end{aligned}$$

where we the fact of $D(0) = 1$.

(2) We can also directly use the decomposition lemma Theorem 16.2.2, and set $\kappa = 0$ such that

$$V(S_T) = V(0) + V'(0)S_T + \int_0^\infty V''(K)(S_T - K)^+ dK.$$

□

Lemma 16.2.8 (Path-independent option hedging using European calls). *[2, p. 36][1, p. 188] Assume we have the implied distribution f_{S_T} for the underlying S_T derived from vanilla option market prices. It follows that an option with arbitrary payoff $V(S_T)$ at maturity has current value*

$$V_0 = \int_{-\infty}^\infty V(s)f_{S_T}(s)ds = V(0)e^{-rT} + V'(0)S_0 + \int_0^\infty V''(K)c(K)dK.$$

The hedging strategy *is: long zero-coupon bonds in quantity $V(0)e^{-rT}$, long underlying stock in quantity $V'(0)$, and long all vanilla calls struck at $K > 0$ in quantities $V''(K)$.*

Proof. (1)Let $c(K)$ represents the market prices of calls with strike K and maturity date T.

$$\begin{aligned}
V_0 &= \int_0^\infty V(K)c''(K)dK \\
&= [V(K)c'(K)]_0^\infty - \int_0^\infty V'(K)c'(K)dK \\
&= -V(0)c'(0) + [V'(K)c(K)]_0^\infty + \int_0^\infty V''(K)c(K)dK \\
&= -V(0)c'(0) + V'(0)c(0) + \int_0^\infty V''(K)c(K)dK \\
&= V(0)e^{-rT} + V'(0)S_0 + \int_0^\infty V''(K)c(K)dK
\end{aligned}$$

where $c'(0) = -e^{-rT}$, $c(0) = S_0$.

Consider a portfolio contains a zero-coupon bond of quantity $V(0)$, stock of quantity $V'(0)$, and all vanilla calls struck at $K > 0$ in quantities $V''(K)$. Its current value is given by

$$\begin{aligned} V_0 &= e^{-rT} E_Q[V(0) + S_T + \int_0^\infty V''(K)(S_T - K)^+ dK] \\ &= V(0)e^{-rT} + V'(0)S_0 + \int_0^\infty V''(K)c(K)dK \end{aligned}$$

where E_Q is taken expectation with respect to the implied distribution such that

$$S_0 = e^{-rT} E_Q[S_T], c(K) = e^{-rT} E_Q[(S_T - K)^+].$$

(2) We can also directly use the decomposition lemma Theorem 16.2.2, and set $\kappa = 0$ such that

$$V(S_T) = V(0) + V'(0)S_T + \int_0^\infty V''(K)(S_T - K)^+ dK.$$

□

Lemma 16.2.9 (Path-independent option hedging using European calls and puts). *[2, p. 36][1, p. 188] Assume we have the implied distribution f_{S_T} for the underlying S_T derived from vanilla option market prices. It follows that an option with arbitrary payoff $V(S_T)$ at maturity has current value*

$$V_0 = \int_{-\infty}^\infty V(s) f_{S_T}(s) ds = (V(A) - V'(A)A)e^{-rT} + V'(A)S_0 + \int_0^A V''(K)p(K)dK + \int_A^\infty V''(K)c(K)dK.$$

where A is an arbitrary nonnegative number.

***The hedging strategy** is: long zero-coupon bonds in quantity $(V(A) - V'(A)A)e^{-rT}$, long the underlying stock in quantity $V'(A)$, long all vanilla puts struck at $K < A$ in quantities $V''(K)$, and long all vanilla calls struck at $K > A$ in quantities $V''(K)$.*

Proof. We can also directly use the decomposition lemma Theorem 16.2.2, and set $\kappa = A$. □

Remark 16.2.2 (practical importance in using both calls and puts). In reality, puts with small strikes are much more liquid than calls with small strikes, since these puts serve as insurance of market crash and are in high demand.

Example 16.2.2 (replicating log contract using vanilla options). Lemma 18.2.17

- The final payoff of a log contract, $L_T = \ln(S_T/K)$, can be decomposed as

$$\ln(S_T/K) = \frac{S_T - K}{K} - \int_K^\infty \frac{1}{v^2}(S_T - v)^+ dv - \int_0^K \frac{1}{v^2}(v - S_T)^+ dv.$$

- The replicating strategy is:
  - long $1/K$ unit of forward with strike price K.
  - short $1/v$ unit of put options at strike v where v ranges from 0 to K.
  - short $1/v$ unit of call options at strike v where v ranges from K to ∞.

16.2.3.2 *Discrete static hedging path-independent derivatives*

Lemma 16.2.10 (replicating positively-supported piece-wise linear continuous payoff function via calls). *[1, pp. 41, 78] Consider a piece-wise linear continuous payoff function $f(x)$ satisfying*

- $f(x) = 0, \forall x < 0$;
- *On x axis, $K_0 = 0$, $K_1 < K_2 < ...$ are successive kink points.*
- *$f(x)$ intercept at y axis I.*
- *$\lambda_1, \lambda_2, ...$ are slopes of successive linear pieces.*

Let $C(K)$ denote the current market price of a call option with strike K and same maturity. Then, the current value is given by

$$V(t) = Ie^{-r(T-t)} + \lambda_1(C(K_0) - C(K_1)) + \lambda_2(C(K_1) - C(K_2)) + \ldots.$$

Note that $S_t = C(K_0 = 0)$.

Proof. The final payoff can be constructed using the following functions

$$f(x) = I + \lambda_1((x - K_0)^+ - (x - K_1)^+) + \lambda_2((x - K_1)^+ - (x - K_2)^+) + \ldots,$$

since $f(x) \triangleq (x - K_0)^+ - (x - K_1)^+$ is a linear segment given by

$$f(x) = \begin{cases} 0, x \leq K_0 \\ x - K_0, K_0 < x < K_1 \\ K_1 - K_0, x \geq K_1 \end{cases}.$$

□

Lemma 16.2.11 (replicating positively-supported piece-wise linear continuous payoff function via puts). *Consider a piece-wise linear continuous payoff function $f(x)$ satisfying*

- $f(x) = 0, \forall x < 0$;
- *On x axis, $K_0 = 0$, $0 < K_1 < K_2 <$ ... are successive kink points from right to left.*
- *$f(x)$ intercept at y axis I.*
- *$\lambda_1, \lambda_2, \ldots$ are slopes of successive linear pieces.*

Let $C(K)$ denote the current market price of a call option with strike K and same maturity. Then, the current value is given by

$$V(t) = Ie^{-r(T-t)} - \lambda_1(-P(K_0) + P(K_1) - (K_1 - K_0)e^{-r(T-t)}) - \lambda_2(P(K_1) - P(K_2) - (K_2 - K_1)e^{-r(T-t)})$$

Note that $S_t = C(K_0 = 0)$.

Proof. The final payoff can be constructed using the following functions

$$f(x) = I + \lambda_1(-(K_0 - x)^+ + (K_1 - x)^+) + \lambda_2(-(x - K_1)^+ + (x - K_2)^+) + \ldots,$$

since $g(x) \triangleq -(K_0 - x)^+ + (K_1 - x)^+$ is a linear segment given by

$$g(x) = \begin{cases} K_1 - K_0, x \leq K_0 \\ -(x - K_1), K_0 < x < K_1 \\ 0, x \geq K_2 \end{cases},$$

and $f(x) \triangleq -(K_0 - x)^+ + (K_1 - x)^+ - (K_1 - K_0)$ is a linear segment given by

$$f(x) = \begin{cases} 0, x \leq K_0 \\ -(x - K_0), K_0 < x < K_1 \\ -(K_1 - K_0), x \geq K_1 \end{cases}.$$

□

Remark 16.2.3 (practical implications).

- In practice, we can use this method to approximate arbitrary payoff function using available call options in the market.
- We can further use put-call parity to replace illiquid out-of-money calls by more liquid out-of-money puts.

Theorem 16.2.3 (general discrete replication via Hilbert space approximation). *Suppose we have a continuous payoff function $f(x)$ with the support $[a,b]$. Suppose we have liquid products with continuous payoff function given by $g_1(x), g_2(x), ..., g_n(x)$(the support is also $[a,b]$). Define inner product between two payoff functions as*

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx.$$

It follows that

- *we can approximatively replicate the payoff function via*

$$\hat{f}(x) = \sum_{i=1}^{n} c_i g_i(x),$$

where the vector c is determined by

$$Gc = p,$$

where $G_{ij} = \left\langle g_i, g_j \right\rangle, p_i = \langle g_i, f \rangle$.

- *The price due to the approximate replication is given by*

$$NPV[\hat{f}] = \sum_{i=1}^{n} c_i NPV[g_i].$$

Proof. (1) See Hilbert space approximation theorem(**??**). (2) See linear pricing theorem(Theorem 15.2.2). □

16.2.3.3 *Pricing path-dependent derivatives*

Note 16.2.6 (path-dependent derivative pricing using implied distributions). Suppose we have the market prices of European calls for all strikes K and all maturities T, then we are able to obtain the implied distribution of $S(t)$ for all future t, which will enables us to price path-dependent derivatives(e.g., Asian options).

16.2.4 Implied forward distribution

16.2.4.1 *Forward starting option basics*

Definition 16.2.3 (forward starting call/put option). *[4, p. 602]*

- *Consider two dates $0 < T_0 < T$. A **forward starting call/put option** allows the holder to receive, at time T_0, a call/put option expiring at T, with strike set equal to $S(T_0)K$, for some $K > 0$. That is, the option's life starts at T_0, but the holder pays at time 0 the premium of the option.*
- *We can view a call option has payoff at time T given by*

$$(S(T) - KS(T_0))^+.$$

Lemma 16.2.12 (forward starting option with general distributions). *Consider two dates $0 < T_1 < T_2$ and a forward starting call option payoff at time T given by*

$$(S(T_2) - KS(T_1))^+.$$

Under risk-neutral measure Q, the time 0 value of the forward starting call option is given by

$$\begin{aligned} V(0) &= e^{-rT_2} E_Q[(S(T_2) - KS(T_1))^+] \\ &= e^{-rT_1} E_Q[g(S_1)] \end{aligned}$$

where

$$\begin{aligned} g(s_1) &= e^{-r(T_2-T_1)} E_Q[(S_2 - KS_1)^+ | S_1 = s_1] \\ &= e^{-r(T_2-T_1)} \int_{\mathbb{R}} (s_2 - Ks_1)^+ f_{S_2|S_1}(s_2|s_1) ds_2 \end{aligned}$$

and $f_{S_2|S_1}$ is the conditional pdf of S_2 given S_1.

Proof. Based on the definition, we have

$$\begin{aligned} V(0) &= e^{-rT_2} E_Q[(S(T_2) - KS(T_1))^+] \\ &= e^{-rT_2} \int_{\mathbb{R}^2} (s_2 - Ks_1)^+ f_{S_1,S_2}(s_1, s_2) ds_1 ds_2 \\ &= e^{-rT_1} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-r(T_2-T_1)} (s_2 - Ks_1)^+ f_{S_2|S_1}(s_2|s_1) ds_2 f_{S_1}(s_1) ds_1 \\ &= e^{-rT} \int_{\mathbb{R}} g(s_1) f_{S_1}(s_1) ds_1 \\ &= e^{-rT} E_Q[g(S_1)] \end{aligned}$$

□

Lemma 16.2.13 (forward starting option pricing in Geometric Browian motion model). *Consider two dates $0 < T_0 < T$. A forward starting call option allows the holder to receive, at time T_0, a call option expiring at T, with strike set equal to $S(T_0)K$, for some $K > 0$. Assuming the underlying asset S_t has dynamics given by*

$$dS_t/S_t = rdt + \sigma dW_t.$$

The value at time 0 is given by

$$V(0) = S(0) \cdot c(1, T - T_0, K),$$

where $c(1, T - T_0, K)$ is the call option price with spot 1, strike K, and tenor $T - T_0$ in the Black model given by

$$c(1, T - T_0, K) = N(d_1) - K \exp(-r(T - T_0))N(d_2),$$

where

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{1}{K}) + (r + \sigma^2/2)(T - T_0)] \\ d_2 &= \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{1}{K}) + (r - \sigma^2/2)(T - T_0)]. \end{aligned}$$

Proof. Note that the payoff at terminal time T is given by

$$FS(T) = (S(T) - KS(T_0))^+$$

then its value at T_0 is given by(Lemma 18.2.1)

$$c(S(T_0), T - T_0, KS(T_0)) = N(d_1)S(T_0) - N(d_2)KS(T_0)e^{-r(T-t)},$$

where

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{S(T_0)}{KS(T_0)}) + (r + \sigma^2/2)(T - T_0)] \\ d_2 &= \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{S(T_0)}{KS(T_0)}) + (r - \sigma^2/2)(T - T_0)]. \end{aligned}$$

Note that we can also write

$$c(S(T_0), T - T_0, KS(T_0)) = S(T_0)c(1, T - T_0, K),$$

where $c(1, T - T_0, K)$ is a deterministic quantity.

Then,

$$\begin{aligned}V(0) &= E_Q[\exp(-rT_0)S(T_0)\cdot c(1, T-T_0, K)] \\ &= E_Q[\exp(-rT_0)S(T_0)]c(1, T-T_0, K) \\ &= S(0)c(1, T-T_0, K)\end{aligned}$$

where we use the fact that $\exp(-rt)S(t)$ is a martingale under risk-neutral measure. $\square$

Remark 16.2.4 (strike is stochastic and unknown). Note that usually a forward starting call option has strike set equal to $S(T_0)K$, for some $K > 0$. And $S(T_0)$ is a stochastic unknown quantity.

Remark 16.2.5 (the situation of general strike). For the case of general strike K instead of $KS(T_0)$, we have

$$V(0) = E_Q[\exp(-rT_0)c(S(T_0), T-T_0, KS(T_0))]$$

where

$$\begin{aligned}c(S(T_0), T-T_0, K) &= N(d_1)S(T_0) - N(d_2)Ke^{-r(T-t)} \\ d_1 &= \frac{1}{\sigma\sqrt{T-T_0}}[\ln(\frac{S(T_0)}{K}) + (r+\sigma^2/2)(T-T_0)] \\ d_2 &= \frac{1}{\sigma\sqrt{T-T_0}}[\ln(\frac{S(T_0)}{K}) + (r-\sigma^2/2)(T-T_0)].\end{aligned}$$

Note that we need the distribution of $S(T_0)$ in order to evaluate the expectation; and the integral will require numerical integration.

Remark 16.2.6 (how to extract complete implied transitioning probability from the market).

- If in the market there are actively traded double Dirac delta option with payoff $\delta(S_1 - K_1) \times \delta(S_2 - K)$, where $S_1 \triangleq S(T_1), S_2 \triangleq S(T_2)$. Then we have the joint pdf given by

$$f_{S_1,S_2}(K_1, K_2) = V_0(K_1, K_2)e^{rT_2}.$$

- Such joint pdf can be integrated to get marginal pdf and conditional pdf (i.e., transition pdf $f_{S_2|S_1}$).

Theorem 16.2.4 (pricing arbitrary forward starting derivatives). *Consider two dates $0 < T_1 < T_2$. Consider a derivative on the asset process S_t with payoff at T_2 given by $S_1 \cdot h(\frac{S_2}{S_1})$, where $S_1 \triangleq S(T_1), S_2 \triangleq S(T_2)$, and h is function of the ration S_2/S_1.*

Then the price of this derivative is given by

$$V_0 = E_Q[e^{-rT_2}S_1 \cdot h(\frac{S_2}{S_1})] = \int_{\mathbb{R}} V_F(k)h(k)dk,$$

where $V_F(k)$ is the second derivative of a forward starting call option price at strike k, that is

$$V_F(k) = \frac{\partial}{\partial k^2} e^{-rT_2} E_Q[(S_2 - S_1 \cdot k)^+].$$

16.2.4.2 *Implied forward distribution*

16.2.4.3 *Implied forward volatility*

Definition 16.2.4 (Implied forward volatility defined via European and digital option).

- *Consider two dates $0 < T_0 < T$ and a forward starting call option allows the holder to receive, at time T_0, a call option expiring at T, whose value is given by*

$$(\frac{S_T}{S_{T_0}} - K)^+$$

Let the market price of the forward starting option be C^{mkt}. Then, the implied forward volatility, denoted by $\sigma(T_0, T, K)$, is the value that the following equation hold; that is

$$N(d_1) - \exp(-r(T - T_0))KN(d_2) = C^{mkt},$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{1}{K}) + (r + \sigma^2/2)(T - T_0)]$$
$$d_2 = \frac{1}{\sigma\sqrt{T - T_0}}[\ln(\frac{1}{K}) + (r - \sigma^2/2)(T - T_0)].$$

- *Consider two dates $0 < T_0 < T$ and a forward starting digital call allows the holder to receive, at time T_0, a digital call option expiring at T, whose value is given by*

$$\mathbf{1}_{(\frac{S_T}{S_{T_0}} - K)}$$

Let the market price of the forward starting option be C^{mkt}. Then, the implied forward volatility, denoted by $\sigma(T_0, T, K)$, is the value that the following equation hold; that is

$$\exp(-r(T - T_0))KN(d_2) = C^{mkt},$$

where

$$d_2 = \frac{1}{\sigma\sqrt{T-T_0}}[\ln(\frac{1}{K}) + (r-\sigma^2/2)(T-T_0)].$$

Note 16.2.7 (interpretation of implied forward volatility in the Black-Scholes setting).

- The implied forward volatility is parameterized by three parameters, denoted by $\sigma(T_0, T, K)$.
- Under the geometric Brownian motion, the implied forward volatility determines the implied distribution of S_T/S_{T_0} conditioning on the information at time 0; that is

$$\frac{S_T}{S_{T_0}} = \exp(r(T-T_0) - \frac{1}{2}\sigma(T-T_0)^2 + \sigma\sqrt{T-T_0}Z), Z \sim N(0,1),$$

  where the σ depends on the strike.
- Conditioning on the information at time T_0(that is, S_{T_0} is known), then

Lemma 16.2.14 (application examples for forward implied volatility). *Consider two dates* $0 < T_0 < T$. *Suppose we know the implied volatility* σ_0 *and forward implied volatility* $\sigma_{0,1}$. *Then we can price the following instruments with market consistence.*

- $$(S_T - S_{T_0} - K)^+$$
  $$E[E[(S_T - S_{T_0} - K)^+|cF_{T_0}]|\mathcal{F}_0]$$

Proof. (1) The current value of the instrument can be written by

$$E[E[(S_T - S_{T_0} - K)^+|\mathcal{F}_{T_0}]|\mathcal{F}_0].$$

With forward volatility, we can evaluate

$$\begin{aligned}
&E[(S_T - S_{T_0} - K)^+|cF_{T_0}] \\
&= S_{T_0}E[(\frac{S_T}{S_{T_0}} - 1 - K/S_{T_0})|\mathcal{F}_{T_0}] \\
&= f(S_{T_0})
\end{aligned}$$

since for

$$E[(S_T - S_{T_0} - K)^+|cF_{T_0}]$$

□

16.3 Implied volatility surface

16.3.1 Fundamentals and facts

Definition 16.3.1 (implied volatility). *[4, p. 341] Let $C(K,T;S_t)$ be the current time t market value of a call option with strike K, maturity T and underlying S_t. The implied volatility for European options on the stock S_t is the $\sigma^*(K,T;S_t)$ that the following equality hold:*

$$C_{BS}(S_t,K,r,T,\sigma^*) = C(K,T,S_t)$$

.

Remark 16.3.1 (interpretation).

- The implied volatility reflects the market's view on the volatility of the stock.
- The implied volatility obtained from actively traded options are usually used to price other options.

Definition 16.3.2 (volatility smile, term structure, and surface). *[4, p. 431]*

- *The implied volatility of an option with a certain life as a function of its strike price is known as a volatility smile.*
- *The implied volatility of an option with a certain life as a function of its maturity is known as a volatility term structure*
- *The implied volatility of an option with a certain life as a function of its strike price and its maturity is known as a volatility surface.*

Remark 16.3.2 (interpretation).

- The implied volatility as a function of strike price for equity option market is usually down-sloping and it is also known as **volatility skew**.
- The implied volatility vs the strike price is usually valley-shaped for currency markets.

Lemma 16.3.1 (No arbitrage bounds on European call/put price profile and smiles). *[1, p. 153][2, p. 22] Let $C(K,T;S_t)$ and $P(K,T;S_t)$ denote the current market prices of European calls/puts with strike K and maturity T.*

The bounds on price profile is given by

- *$C \geq Se^{-d\tau} - Ke^{-r\tau}$, where d is the dividend rate, $\tau = T - t$, and r the short rate.*
- $$\frac{\partial C}{\partial K} \leq 0, \frac{\partial P}{\partial K} \geq 0;$$

- $$\frac{\partial^2 C}{\partial K^2} \geq 0, \frac{\partial^2 P}{\partial K^2} \geq 0;$$
- $$\int_0^\infty \frac{\partial^2 C}{\partial K^2} dK = \int_0^\infty \frac{\partial^2 P}{\partial K^2} dK = e^{-rT}.$$

Proof. (1) use the call-put parity(Lemma 15.7.3) and the fact that a put price $P \geq 0$. (2)(3) use the relations between derivatives of call/puts and delta/step function. See Lemma 16.2.1. (4) Use the fact that $\frac{\partial^2 C}{\partial K^2}$ will give a discounted risk-neutral measure that integrate to 1. See Lemma 16.2.4. □

Remark 16.3.3 (bounds are model independent)**.**

- Note that bounds on prices only relies on the assumption that there exists no arbitrage opportunities in the market rather than any model.
- Any market model that violates above should not be used.

16.3.2 Stylized market facts about volatility

Note 16.3.1 (equity market implied volatility feature)**.** [5, pp. 233, 239][1, p. 5][6, p. 486] Consider European call/put options having same maturity dates but different strikes. The volatility skew (Figure 16.3.1) is often observed in the market due to the following reasons:

- Stock investors, as a whole, are worrying more about the falling stock price than the rising stock price; therefore, the protective put strategy(Definition 12.5.5) with out-of-money put is in high demand and push the out-of-money put price up.
- Out-of-money puts have higher implied volatility(that is, higher price than Black model predicts) than in-the-money puts because out-of-money puts resemble an insurance against market crash and risk-averse buyers are willing to pay extra money for insurance.
- Out-of-money puts have higher implied volatility(that is, higher price) than out-of-money calls because buyers/market worry more about the falling prices than the rising prices.

Consider European call/put options having same strike but different maturity dates.

- At volatile period, options of smaller maturities tend to have higher implied volatility.

- At tranquil period, options of larger maturities tend to have higher implied volatility.

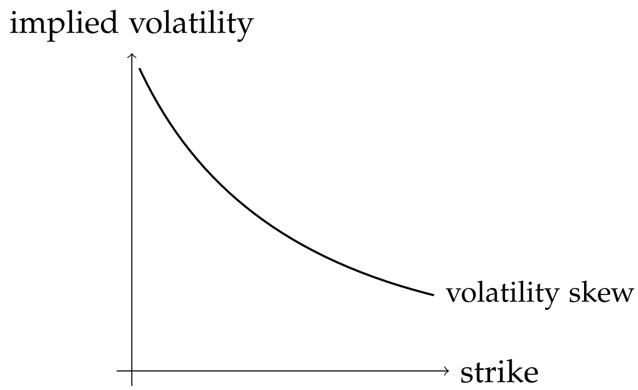


Figure 16.3.1: volatility skew for stock index

Note 16.3.2 (commodity market implied volatility feature). [6, p. 487][5, p. 236] Consider the commodity market. The volatility skew, as showed in Figure 16.3.2, is often observed in the market.

- In commodity markets where end users try to protect themselves against rising prices by either buying **protective calls**(Definition 12.5.6) at higher exercise prices(out-of-money calls) or selling **covered puts**(Definition 12.5.9) at lower exercise prices(out-of-money put).
- As a result, in the commodity market, lower exercise prices have lower implied volatilities, and higher exercise prices have higher implied volatilities.

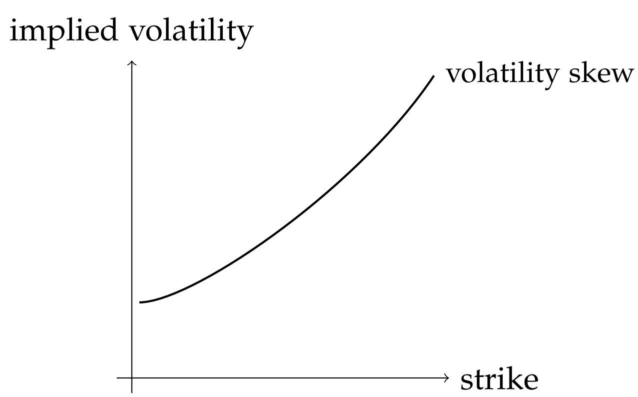


Figure 16.3.2: volatility skew for commodity market

Note 16.3.3 (FX market implied volatility feature). [6, p. 488][5, p. 236] Consider the FX market. The volatility skew, as showed in Figure 16.3.3, is often observed in the market.

- For FX market market participants, when one party is worrying about the rising exchange rate(buy protective call or covered put), there is also an counterparty worrying about the falling exchange rate(buy protective put or covered call).
- As a result, in the FX market, lower strike and higher strike prices tend to have higher implied volatilities, and intermediate strike have lower implied volatility.

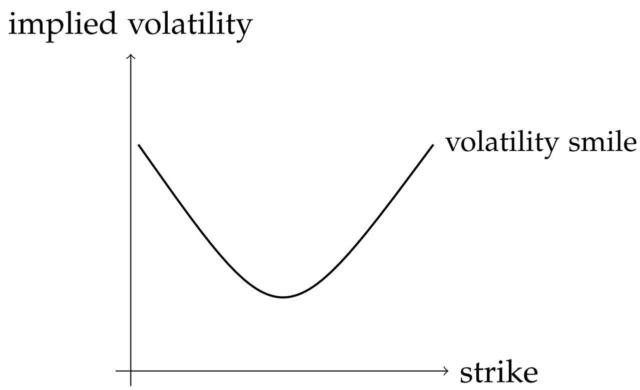


Figure 16.3.3: volatility smile for FX market

Remark 16.3.4 (discussion on smiles in different markets). For smiles in equity, interest rate, foreign exchange, see [1, p. 143].

Note 16.3.4 (geometric Brownian motion with varying volatility can generate arbitrage distributions).

16.3.3 Implied distribution from implied volatility

Theorem 16.3.1 (deriving implied distribution from implied volatility). *[7] Let $\sigma^*(K,T;S_t)$ be the implied volatility such that the market prices of calls $C(K,T;S_t)$ can be parameterized by*

$$C(K,T;S_t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)}$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r + [\sigma^*]^2/2)(T-t)]$$
$$d_2 = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r - [\sigma^*]^2/2)(T-t)].$$

Then, the implied distribution parametrized in log moneyness $\kappa = \ln(K/F(t,T))$ *is given by*

$$f_\kappa(\kappa = k) = e^{rT}\frac{g(k)}{\sqrt{2\pi w(k,T)}}\exp(-\frac{d_-(k)^2}{2}),$$

where

$$w(k,T) = \sigma^2(k,T)(T-t)$$
$$d_\pm(k) = -\frac{k}{\sqrt{w(k,T)}} \pm \frac{\sqrt{w(k,T)}}{2}$$
$$g(k) = 1 - \frac{\kappa}{w}\frac{\partial w}{\partial \kappa} + \frac{1}{4}(\frac{\partial w}{\partial \kappa})^2(\frac{\kappa^2}{w^2} - \frac{1}{w} - \frac{1}{4}) + \frac{1}{2}\frac{\partial^2 w}{\partial \kappa^2}$$

The implied distribution parameterized in strike K *is given by*

$$f_{S_T}(S_T = K) = e^{rT}\frac{g_K(K)}{K\sqrt{2\pi w(k,T)}}\exp(-\frac{d_2(K)^2}{2}).$$

where $g_K(K)$ *is the composite function* $g_K(K) = g(\ln K/F(t,T))$.

Proof. See Lemma 16.2.4. Note that σ^* also has dependence on K. □

16.4 Implied volatility surface parameterization

16.4.1 Static arbitrage

16.4.1.1 *Butterfly arbitrage*

Definition 16.4.1 (free of calendar arbitrage). *Consider an underlying asset with proportional dividends. We say a implied volatility is free of calendar arbitrage if the*

$$C(K_2,)$$

Definition 16.4.2 (free butterfly arbitrage condition). *[7] Denote the volatility surface by $\sigma(K,T)$. Fix T_0, a slice $\sigma(K,T_0)$ is said to be* ***free of butterfly arbitrage*** *if the corresponding risk-neutral density is non-negative.*

Definition 16.4.3 (free of static arbitrage). *[7] A volatility surface $\sigma(K,T)$ is* ***free of static arbitrage*** *if and only if the following conditions are satisfied:*

- *it is free of calendar spread arbitrage.*
- *fix T_0, each time slice $\sigma(K,T_0)$ is free of butterfly arbitrage.*

Lemma 16.4.1 (martingale inequality). *link[7] Let X_t is a martingale, $L \in \mathbb{R}$, and $0 < t_1 < t_2$ two future times. Assume current time is 0. Then*

$$E[(X(t_2) - L)^+] \geq E[(X(t_1) - L)^+].$$

Proof.

$$\begin{aligned} E[(X(t_2) - L)^+ | \mathcal{F}_0] &= E[E[(X(t_2) - L)^+ | \mathcal{F}_1] | \mathcal{F}_0] \\ &\geq E[(E[(X(t_2) - L) | \mathcal{F}_1])^+ | \mathcal{F}_0] \\ &\geq E[(X(t_1) - L)^+ | \mathcal{F}_0] \end{aligned}$$

where we use the iterative conditional expectation property, Jensen's inequality($f : x \to x^+$ is a convex function) and the fact that $X(t)$ is a martingale such that

$$X(t_1) - L = E[(X(t_2) - L) | \mathcal{F}_1].$$

□

Theorem 16.4.1 (calender arbitrage free stocks with proportional dividends). *[7]*
If dividends

Proof. Let $F(t,T)$ denote the forward price at time t with delivery date T, then $F(t,T_2)$ is martingale under the risk neutral measure.

$$\begin{aligned}
C_2(S_0,T_2,K) &= e^{-rT_2}E_Q[(F(T_2,T_2)-K)^+] \\
&\geq e^{-rT_2}E_Q[(F(T_1,T_2)-K)^+] \\
&= e^{-rT_2}E_Q[(F(T_1,T_1)e^{(r-q)(T_2-T_1)}-K)^+] \\
&= e^{-rT_1}E_Q[(F(T_1,T_1)-Ke^{-(r-q)(T_2-T_1)})^+] \\
&= C_1(S_0,T,Ke^{-r(T_2-T_1)})
\end{aligned}$$

Note that the two call options have different strikes but same log-forward strikes; that is

$$\frac{Ke^{-r(T_2-T_1)}}{F(0,T_1)} = \frac{K}{F(0,T_2)}.$$

□

Corollary 16.4.1.1 (calender arbitrage free stocks with proportional dividends and cash dividends). *link*

Proof.

$$\begin{aligned}
& E[(X(t_2)-L)^+] \geq E[(X(t_1)-L)^+] \\
\Leftrightarrow\ & E[(\frac{S(t_2)-D(t_2)}{F(0,t_2)-D(t_2)}-L)^+] \geq E[(\frac{S(t_1)-D(t_1)}{F(0,t_1)-D(t_1)}-L)^+] \\
\Leftrightarrow\ & \frac{1}{F(0,t_2)-D(t_2)}E[(S(t_2)-(D(t_2)+L(F(0,t_2)-D(t_2)))^+] \geq \frac{1}{F(0,t_1)-D(t_1)}E[(S(t_1)-(D(t_1)+L(\\
\Leftrightarrow\ & \frac{C_2(K_2,t_2)\exp(rt_2)}{F(0,t_2)-D(t_2)} \geq \frac{C_2(K_1,t_1)\exp(rt_1)}{F(0,t_1)-D(t_1)}
\end{aligned}$$

where

$$K_1 \triangleq D(t_1)+L(F(0,t_1)-D(t_1)), K_2 \triangleq D(t_2)+L(F(0,t_2)-D(t_2)).$$

□

16.5 Notes on bibliography

For no-arbitrage theory, see [8].

For treatment from economical perspective, see [9] [10].

For martingale methods, see [11].

For PDE methods, see [11][12].

For incomplete markets, see [13].

For the differences between real world measure and risk-neutral measure, see [14].

An excellent book of "P" method for risks and asset allocations, see [15].

For a comprehensive discussion on risks, see [16].

For statistical arbitrage, see [17].

[18].

For stochastic volatility model, see [19], [20], [21].

For empirical evidence of volatility surface, see [22].

BIBLIOGRAPHY

1. Derman, E., Miller, M. & Park, D. *The Volatility Smile* ISBN: 9781118959169. https://books.google.com/books?id=WKfBDAAAQBAJ (Wiley, 2016).
2. Bossu, S. *Advanced Equity Derivatives: Volatility and Correlation* (John Wiley & Sons, 2014).
3. Carr, P. & Madan, D. Towards a theory of volatility trading. *Volatility: New estimation techniques for pricing derivatives*, 417–427 (1998).
4. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
5. Alexander, C. *Market Risk Analysis, Pricing, Hedging and Trading Financial Instruments* ISBN: 9780470772812. https://books.google.com/books?id=M-6UEM18UHkC (Wiley, 2008).
6. Natenberg, S. *Option Volatility and Pricing Workbook, Second Edition* ISBN: 9780071819053. https://books.google.com/books?id=5calngEACAAJ (McGraw-Hill Education, 2014).
7. Gatheral, J. & Jacquier, A. Arbitrage-free SVI volatility surfaces. *Quantitative Finance* **14,** 59–71 (2014).
8. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2004).
9. LeRoy, S. F. & Werner, J. *Principles of financial economics* (Cambridge University Press, 2014).
10. Mishkin, F. S. *The economics of money, banking, and financial markets* (Pearson education, 2007).
11. Pascucci, A. *PDE and martingale methods in option pricing* (Springer Science & Business Media, 2011).
12. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
13. Cernỳ, A. *Mathematical techniques in finance: tools for incomplete markets* (Princeton University Press, 2009).
14. Meucci, A. 'P'Versus' Q': Differences and Commonalities between the Two Areas of Quantitative Finance. *GARP Risk Professional*, 47–50 (2011).
15. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
16. Jorion, P. *Financial Risk Manager Handbook: FRM Part I / Wiley Finance* **pt. 2.** ISBN: 9781118017913. https://books.google.com/books?id=4ceVmGJSNpcC (Wiley, 2010).

17. Leung, T. & Li, X. *Optimal Mean Reversion Trading: Mathematical Analysis and Practical Applications* (World Scientific, 2015).
18. Derman, E., Kani, I. & Goldman, N. C. Implied trinomial trees of the volatility smile. *Journal of Derivatives* (1996).
19. Lewis, A. L. *Option Valuation Under Stochastic Volatility II* (Finance Press, Newport Beach, CA, 2009).
20. Zhu, J. *Applications of Fourier transform to smile modeling: Theory and implementation* (Springer Science & Business Media, 2009).
21. Mikhailov, S. & Nögel, U. *Heston's stochastic volatility model: Implementation, calibration and some extensions* (John Wiley and Sons, 2004).
22. Daglish, T., Hull, J. & Suo, W. Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance* **7**, 507–524 (2007).

17

MODEL CONSTRUCTION METHODS AND PRINCIPLES

17.1 Overview

Model construction for financial market is more art than science. In this chapter, we cover different types models used for pricing, ranging from simple static distribution model to sophisticated stochastic volatility model. We particularly emphasize how

How model parameters should be specified to ensure no-arbitrage opportunities. And how to ensure model-market consistence via calibration procedures. We will discuss the pros and cons of different types of model. We will see that financial market models are excellent illustration of the famous quote "all models are wrong but some are useful".

17.2 Static distribution model

Various distribution for ST can be selected as long as this condition is satisfied: for example, normal, lognormal. One distribution is preferred over other distribution based on the analytical pricing possibility and calibration to market.

Parameters in the assumed distribution were tuned so that the model prices fit the market prices for contracts whose market prices could be observed.

17.2.1 Principles

Remark 17.2.1 (consistence with the no-arbitrage pricing framework). We have showed in (), to have no-arbitrage pricing, the spot price should satisfy

$$\frac{S_t}{M(t)} = E_Q[\frac{S(T)}{M(T)}|\mathcal{F}_t]$$

under risk-neutral measure.

And the forward price should satisfy

$$F(t, T_M) = E_Q[F(T, T_M)|\mathcal{F}_t]$$

under risk-neutral measure.

Remark 17.2.2 (achieve market consistence via calibration). The goal of calibration is to find a set of parameters such that the static distribution is close to implied distribution.

17.2.2 Black lognormal model

Definition 17.2.1. *Let current time be t. In a Black lognormal model, we assume the spot price S_t of an asset has a lognormal distribution under risk-neutral measure at time T, given by*

$$S_T = S_t \exp((r - q - \frac{1}{2}\sigma^2)(T - t) + \sigma\sqrt{T - t}Z), Z \in N(0, 1).$$

Similarly, we assume the forward price $F(t, T_M)$ of an asset has a lognormal distribution under risk-neutral measure at time T, given by

$$F(T, T_M) = F(t, T_M) \exp((-\frac{1}{2}\sigma^2)(T - t) + \sigma\sqrt{T - t}Z), Z \in N(0, 1).$$

Theorem 17.2.1 (Black model European call/put pricing for assets without dividends). *[1, p. 219] Let current time t. Denote the spot by S_t and forward by $F(t,T)$. In the Black-Scholes model, an European call option with strike price K and expiry T is given as:*

$$C(S_t,t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)}$$

where

$$\begin{aligned}
d_1 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r+\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) + \sigma^2/2(T-t)] \\
d_2 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) - \sigma^2/2(T-t)] \\
d_2 &= d_1 - \sigma\sqrt{T-t} \\
N(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
F(t,T) &= S_t \exp(r(T-t))
\end{aligned}$$

We further have

- *The price of a zero strike is S_t.*
- *The price of the put can be derived based on put-call parity $P_t + S_t = C_t + Ke^{-r(T-t)}$ (Lemma 15.2.1) as*

$$P(S_t,t) = N(-d_2)Ke^{-r(T-t)} - N(-d_1)S_t.$$

Proof. We use martingale method.

$$C(S_T,T) = E_Q[e^{-r(T-t)}\max(S_T-K,0)|\mathcal{F}_t].$$

where

$$S(T) = S(t)\exp((r - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma(W(T)-W(t))).$$

Let $Y = (W(T)-W(t))/\sqrt{T-t}$, then $Y \sim N(0,1)$, and we can write

$$S(T) = S(t)\exp((r - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)).$$

Let

$$d_2 = \frac{\ln\frac{S(t)}{K} + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$

such that

$$S(t)\exp((r - \frac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}(-d_2)) = K.$$

Then the for the standard normal Y, We have

$$Pr(S(t)\exp((r-\frac{1}{2}\sigma(s)^2)(T-t)+\sigma\tau Y)) > K) = Pr(Y > -d_2) = Pr(Y < d_2).$$

$$\begin{aligned}
C(S_T, T) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}(S(t)\exp(+\sigma\tau y + (r-\frac{1}{2}\sigma^2)\tau) - K)e^{-\frac{1}{2}y^2}dy \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\sigma\tau y - \frac{1}{2}\sigma^2\tau)e^{-\frac{1}{2}y^2}dy] - [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}e^{-\frac{1}{2}y^2}dy] \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\frac{1}{2}(y-\sigma\sqrt{\tau})^2)dy] - e^{-r\tau}KN(d_2) \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2-\sigma\sqrt{\tau}}^{\infty} S(t)\exp(-\frac{1}{2}z^2)dz] - e^{-r\tau}KN(d_2) \\
&= S(t)N(d_1) - e^{-r\tau}KN(d_2)
\end{aligned}$$

where we used the fact that $-d_1 = -d_2 - \sigma\sqrt{\tau}$. □

Theorem 17.2.2 (Black model European call and put for assets with dividends). *[1, p. 236][2, p. 373] Let current time t. Denote the spot by S_t and forward by $F(t,T)$. In the Black-Scholes model, an European call option with strike price K for an asset paying continuous dividends a is given as:*

$$C(S_t, t) = N(d_1)S_t e^{-a(T-t)} - N(d_2)Ke^{-r(T-t)}$$

where

$$\begin{aligned}
d_1 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-a+\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) + \sigma^2/2(T-t)] \\
d_2 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-a-\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) - \sigma^2/2(T-t)] \\
d_2 &= d_1 - \sigma\sqrt{T-t} \\
N(x) &= \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy \\
F(t,T) &= S_t\exp((r-a)(T-t))
\end{aligned}$$

The price of the put can be derived based on put-call parity

$$P_t + S_t e^{-a(T-t)} = C_t + Ke^{-r(T-t)}$$

(*Lemma 15.2.1*) *as*

$$P(S_t, t) = N(-d_2)Ke^{-r(T-t)} - N(-d)S_t e^{-a(T-t)}.$$

Proof. We use martingale method.

$$C(S_T, T) = E_Q[e^{-r(T-t)} \max(S_T - K, 0)|\mathcal{F}_t].$$

where

$$S(T) = S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma(W(T) - W(t))).$$

Let $Y = (W(T) - W(t))/\sqrt{T-t}$, then $Y \sim N(0,1)$, and we can write

$$S(T) = S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)).$$

Let

$$d_2 = \frac{\ln\frac{S(t)}{K} + (r - a - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$

such that

$$S(t)\exp((r - a - \frac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}(-d_2)) = K.$$

Then the for the standard normal Y, We have

$$Pr(S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)) > K) = Pr(Y > -d_2) = Pr(Y < d_2).$$

$$\begin{aligned}
C(S_T, T) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}(S(t)\exp(+\sigma\tau y + (r - a - \frac{1}{2}\sigma^2)\tau) - K)e^{-\frac{1}{2}y^2}dy \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\sigma\tau y - \frac{1}{2}\sigma^2\tau)e^{-\frac{1}{2}y^2}dy] - [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}e^{-\frac{1}{2}y^2}dy] \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}(y - \sigma\sqrt{\tau})^2)dy] - e^{-r\tau}KN(d_2) \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2-\sigma\sqrt{\tau}}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}z^2)dz] - e^{-r\tau}KN(d_2) \\
&= S(t)\exp(-\tau a)N(d_1) - e^{-r\tau}KN(d_2)
\end{aligned}$$

where we used the fact that $-d_1 = -d_2 - \sigma\sqrt{\tau}$. □

17.2.3 Extended lognormal model

Lemma 17.2.1 (black formula with extended lognormal distribution family). *[3]*

- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$

  where

  $$d_{1,2} = \frac{\log M_1(T) - \log K}{\sigma} \pm \frac{1}{2}\sigma$$

  and

  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

- $$c = \exp(-rT)[(M_1(T) - \tau)N(d_1) - (K - \tau)N(d_2)],$$

  where

  $$d_{1,2} = \frac{\log(M_1(T) - \tau) - \log(K - \tau)}{\sigma} \pm \frac{1}{2}\sigma$$

  and

  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$

  where

  $$d_{1,2} = \frac{\log -M_1(T) - \log -K}{\sigma} \pm \frac{1}{2}\sigma$$

  and

  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$

  where

  $$d_{1,2} = \frac{\log M_1(T) - \log K}{\sigma} \pm \frac{1}{2}\sigma$$

  and

  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

17.2.4 Bachelier normal model

Definition 17.2.2 (Bachelier normal model). *Let current time be t. In a Bachelier normal model, we assume the spot price S_t of an asset has a normal distribution under risk-neutral measure at time T, given by*

$$S_T = S_t + (r-q)(T-t) + \sigma\sqrt{T-t}Z, Z \in N(0,1).$$

Similarly, we assume the forward price $F(t, T_M)$ of an asset has a normal distribution under risk-neutral measure at time T, given by

$$F(T, T_M) = F(t, T_M) + \sigma\sqrt{T-t}Z, Z \in N(0,1).$$

Lemma 17.2.2 (European call/put pricing under normal model). *[4] Under the normal forward price model, the current values of European call and put are given by*

- $$\begin{aligned} C(t) &= \exp(-r(T-t))E_Q[(F(T,T)-K)^+|\mathcal{F}_t] \\ &= \exp(-r(T-t))((F(t,T)-K)N(d_1) + \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}e^{-d_1^2/2}); \end{aligned}$$

- $$\begin{aligned} P(t) &= \exp(-r(T-t))E_Q[(K-F(T,T))^+|\mathcal{F}_t] \\ &= \exp(-r(T-t))((K-F(t,T))N(-d_1) + \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}e^{-d_1^2/2}); \end{aligned}$$

- ***(put call parity)***

$$C(t) - P(t) = \exp(-r(T-t))(F(t,T)-K)$$

where $d_1 = \frac{F-K}{\sigma\sqrt{T-t}}$

Proof. From Feyman-Kac theorem(Theorem 15.8.1), the forward price under risk-neutral measure has dynamics

$$dF = \sigma dW_t.$$

That is $F(T) - F(t) \sim N(0, \sigma^2(T-t))$

$$\begin{aligned} C(t) &= e^{-r(T-t)} E_Q[(F_T - K)^+ | \mathcal{F}_t] \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (F_t + \sigma\sqrt{T-t}x - K)^+ e^{-x^2/2} dx \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi}} \int_{\frac{K-F}{\sigma\sqrt{T-t}}}^{\infty} (F_t + \sigma\sqrt{T-t}x - K) e^{-x^2/2} dx \end{aligned}$$

The rest is straight forward. □

17.2.5 Multivariate lognormal model

Definition 17.2.3 (multivariate lognormal model). *Let current time be t. In a Black lognormal model, we assume the spot prices of assets $S_1, S_2, ..., S_N$ have a multivariate lognormal distribution under risk-neutral measure at time T, given by*

$$\ln S_i(T) = \ln S_t + (r - q - \frac{1}{2}\sigma_i^2)(T-t) + \sigma_i\sqrt{T-t}Z, Z \in N(0,1);$$

$$corr(\ln S_i, \ln S_j) = \rho_{ij}.$$

Similarly, we assume the forward prices of assets $F_1(t, T_M), F_2(t, T_M), ..., F_N(t, T_M)$ have a multivariate lognormal distribution under risk-neutral measure at time T, given by

$$\ln F_i(T, T_M) = \ln F_i(T, T_M) + (-\frac{1}{2}\sigma_i^2)(T-t) + \sigma_i\sqrt{T-t}Z, Z \in N(0,1).$$

$$corr(\ln F_i(T, T_M), \ln F_i(T, T_M)) = \rho_{ij}.$$

17.2.5.1 *Different assets*

17.2.5.2 *One asset at different future times*

17.2.5.3 *Moments*

Lemma 17.2.3 (moments of basket underlying dynamics under risk-neutral measure). *Assume under risk-neutral measure Q that the underlying dynamics are given by*

$$dS_t^{(i)}/S_t^{(i)} = (r-q)dt + \sigma_i dW_i(t), i = 1, 2, ..., n,$$

and $dW_i dW_j = \rho_{ij} dt$. Denote $B(T) = \sum_{i=1}^n w_i S_T^{(i)}$. It following that

- $$B(T) = \sum_{i=1}^{n} F_i w_i \exp(\sigma_i W_i(T) - \frac{1}{2}\sigma_i^2 T),$$
  where $F_i = S^{(i)}(0)\exp((r-q)T)$.
- $$E[B(T)] = \sum_{i=1}^{n} F_i w_i.$$
- $$E[B(T)^2] = \sum_{i=1}^{n}\sum_{j=1}^{n} F_i F_j w_i w_j \exp(\rho_{ij}\sigma_i\sigma_j T).$$
- $$E[B(T)^3] = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} F_i F_j F_k w_i w_j w_k (\exp((\rho_{ij}\sigma_i\sigma_j + \rho_{ik}\sigma_i\sigma_k + \rho_{jk}\sigma_j\sigma_k)T)).$$

Proof. (1) Note that

$$S_T^{(i)} = S_0^{(i)} \exp((r - q - \frac{1}{2}\sigma_i^2 T) + \sigma_i W_i(T)).$$

(2) Note that

$$E[\exp(\sigma_i W_i(T) - \frac{1}{2}\sigma_i^2 T)] = 1.$$

(3)

$$\begin{aligned}
&E[\exp(\sigma_i W_i(T) - \frac{1}{2}\sigma_i^2 T)\exp(\sigma_j W_j(T) - \frac{1}{2}\sigma_j^2 T)] \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T) M(\sigma_i, \sigma_j) \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(\frac{1}{2}(\sigma_i, \sigma_j)\begin{bmatrix} T & \rho T \\ \rho T & T \end{bmatrix}(\sigma_i, \sigma_j)^T) \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(\frac{1}{2}(\sigma_i^2 T + \sigma_j^2 T + 2\rho_{ij}\sigma_i\sigma_j T) \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(\frac{1}{2}(\sigma_i^2 T + \sigma_j^2 T + 2\rho_{ij}\sigma_i\sigma_j T) \\
&= \exp(\rho_{ij}\sigma_i\sigma_j)
\end{aligned}$$

where $M(t_1, t_2)$ is the mgf for the random vector $(W_i(T), W_j(T))$. (4)

$$
\begin{aligned}
&E[\exp(\sigma_i W_i(T) - \frac{1}{2}\sigma_i^2 T)\exp(\sigma_j W_j(T) - \frac{1}{2}\sigma_j^2 T)\exp(\sigma_k W_k(T) - \frac{1}{2}\sigma_k^2 T)] \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(-\frac{1}{2}\sigma_k^2 T) M(\sigma_i, \sigma_j, \sigma_k) \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(-\frac{1}{2}\sigma_k^2 T)\exp(\frac{1}{2}(\sigma_i, \sigma_j, \sigma_k)\begin{bmatrix} T & \rho_{ij}T & \rho_{ik}T \\ \rho_{ij}T & T & \rho_{jk}T \\ \rho_{ik}T & \rho_{jk}T & T \end{bmatrix}(\sigma_i, \sigma_j, \sigma_k)^T) \\
&= \exp(-\frac{1}{2}\sigma_i^2 T)\exp(-\frac{1}{2}\sigma_j^2 T)\exp(-\frac{1}{2}\sigma_k^2 T)\exp(\frac{1}{2}(\sigma_i^2 T + \sigma_j^2 T + \sigma_k^2 T + 2\rho_{ij}\sigma_i\sigma_j T + 2\rho_{ik}\sigma_i\sigma_k T + 2\rho_{jk}\sigma_k\sigma_j T \\
&= \exp(\rho_{ij}\sigma_i\sigma_j)
\end{aligned}
$$

where $M(t_1, t_2, t_3)$ is the mgf for the random vector $(W_i(T), W_j(T), W_k(T))$. □

17.2.5.4 *Moment matching methods*

Lemma 17.2.4 (two-parameter lognormal moment matching). *[5][6, p. 232] Let $T_1, T_2, ..., T_n$ be a set of dates. Consider a set of random variables given by*

$$X_j = S_0 \exp((r - \frac{1}{2}\sigma^2)T_j + \sigma W(T_j)).$$

and the averaging random variable

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

It follows that

- $$
\begin{aligned}
M_1 &= E[X] = \frac{1}{n}S_0 \sum_{i=1}^{n} \exp(rT_i) \\
M_2 &= E[X^2] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i X_j] \\
&= \frac{1}{n^2}S_0^2 \sum_{i=1}^{n} \exp(2rT_i + \sigma^2 T_i) + \frac{2}{n^2}S_0^2 \sum_{i=1}^{n}\sum_{j>i}^{n} \exp(r(T_j - T_i) + 2rT_i + \sigma^2 T_i)
\end{aligned}
$$

- *The random variable $Y = \mu \exp(-\frac{1}{2}\nu^2 + \nu Z)$ has first moment μ and second moment is $\mu^2 \exp(\nu^2)$.*
- *Let $\mu = M_1, \nu^2 = \log(M_2/M_1^2)$, we have*

$$E[Y] = M_1, E[Y^2] = M_2.$$

Proof. (1)(a) Use linearity of expectation. Note that

$$E[\exp(\sigma\sqrt{T_j}Z)] = \exp(\frac{1}{2}\sigma^2 T_j), Z \sim N(0,1).$$

(b) Note that for $j > i$, we have

$$X_i X_j = X_i^2(X_j/X_i) = X_i^2 \exp((r - \frac{1}{2}\sigma^2)(T_j - T_i) + \sigma\sqrt{T_j - T_i}Z), Z \sim N(0,1).$$

Then due to independence

$$\begin{aligned} E[X_i X_j] &= E[X_i^2]E[\exp((r - \frac{1}{2}\sigma^2)(T_j - T_i) + \sigma\sqrt{T_j - T_i}Z)] \\ &= S_0^2 \exp((2r + \sigma^2)T_j)\exp(r(T_j - T_i)) \end{aligned}$$

(2)

$$E[Y] = \mu \exp(-\frac{1}{2}\nu^2)E[\exp(\nu Z)] = \mu \exp(-\frac{1}{2}\nu^2)\exp(\frac{1}{2}\nu^2) = \mu.$$

and

$$E[Y^2] = \mu^2 \exp(-\nu^2)E[\exp(2\nu Z)] = \mu^2 \exp(-\nu^2)\exp(\frac{1}{2}4\nu^2) = \mu^2 \exp(\nu^2).$$

(3) Straight forward. □

Lemma 17.2.5 (European call price using moment matching approximation). *Let $0 < T_1 < T_2 < ... < T_n \leq T$ be a set of dates. Consider a set of random variables given by*

$$X_j = F_0 \exp(-\frac{1}{2}\sigma^2 T_j + \sigma W(T_j)).$$

and the averaging random variable

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Consider a call option with strike K and maturity T such that its final payoff at T is given by

$$V(T) = (X - K)^+$$

. It follows that its value at $t = 0, t < T_1$, approximated by log-normal distribution, is given by

$$V(0) = P(t,T)(M_1 N(d_1) - KN(d_2),$$

where

$$d_{\pm} = \frac{\ln(M_1/K) \pm \frac{1}{2}\nu^2}{\nu},$$

and

$$M_1 = F_0, M_2 = \frac{2}{n^2}\sum_{i=1}^{n}(n-i+1)F_0^2\exp(\sigma^2 T_i), \nu^2 = \log(M_2/M_1^2).$$

Proof. Using lognormal moment matching(Lemma 17.2.4), we know that the distribution of X can be approximated by $Y = M_1\exp(-\frac{1}{2}\nu^2 + \nu Z), Z \sim N(0,1)$.Then we follow the standard procedure of deriving vanilla call option. □

Lemma 17.2.6 (two-parameter lognormal moment matching for weighted averaging lognormal). *[5][6, p. 232] Let $T_1, T_2, ..., T_n$ be a set of dates. Consider a set of random variables given by*

$$X_j = S_0\exp((r - \frac{1}{2}\sigma^2)T_j + \sigma W(T_j)).$$

and the averaging random variable

$$X = \sum_{i=1}^{n} w_i X_i.$$

It follows that

- $$\begin{aligned} M_1 &= E[X] = S_0\sum_{i=1}^{n} w_i\exp(rT_i) \\ M_2 &= E[X^2] = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j E[X_i X_j] \\ &= S_0^2\sum_{i=1}^{n} w_i^2\exp(2rT_i + \sigma^2 T_i) + S_0^2\sum_{i=1}^{n}\sum_{j>i} 2w_i w_j\exp(r(T_j - T_i) + 2rT_i + \sigma^2 T_i) \end{aligned}$$
- *The random variable $Y = \mu\exp(-\frac{1}{2}\nu^2 + \nu Z)$ has first moment μ and second moment is $\mu^2\exp(\nu^2)$.*

- *Let* $\mu = M_1, \nu^2 = \log(M_2/M_1^2)$, *we have*

$$E[Y] = M_1, E[Y^2] = M_2.$$

Proof. See Lemma 17.2.4. □

17.2.6 Multivariate normal model

Definition 17.2.4 (multivariate normal model). *Let current time be t. In a Black lognormal model, we assume the spot prices of assets* $S_1, S_2, ..., S_N$ *have a multivariate lognormal distribution under risk-neutral measure at time* T*, given by*

$$S_i(T) = S_t + (r-q)(T-t) + \sigma_i\sqrt{T-t}Z, Z \in N(0,1);$$

$$corr(S_i, S_j) = \rho_{ij}.$$

Similarly, we assume the forward prices of assets $F_1(t, T_M), F_2(t, T_M), ..., F_N(t, T_M)$ *have a multivariate lognormal distribution under risk-neutral measure at time* T*, given by*

$$F_i(T, T_M) = F_i(T, T_M) + \sigma_i\sqrt{T-t}Z, Z \in N(0,1).$$

$$corr(F_i(T, T_M), F_i(T, T_M)) = \rho_{ij}.$$

Remark 17.2.3 (applications). Multivariate normal model can be used to model the static distribution of averages of interest rate.

17.2.7 General copula approach

17.3 Local volatility method

17.3.1 Fundamentals

Theorem 17.3.1 (Dupire's equation). *[7, p. 245][8, p. 57] Let* $C(K, T; S_0)$ *denote the* ***current market call prices*** *at different strikes K and different maturities T. Further assume an asset* S_t *has risk-neutral SDE given by*

$$dS_t = (r-d)S_t dt + \sigma(S_t, t)S_t dW_t.$$

If the local volatility $\sigma(S,T)$ satisfies

$$C_T = \frac{1}{2}\sigma^2 K^2 C_{KK} - (r-d)KC_K - dC,$$

or equivalently

$$\sigma^2(K,T;S_0) = \frac{C_T + (r-d)KC_K + dC}{\frac{1}{2}K^2 C_{KK}},$$

where

$$C_T = \frac{C(K,T)}{\partial T}, C_K = \frac{C(K,T)}{\partial K}, C_{KK} = \frac{C(K,T)}{\partial K^2},$$

then

$$E_Q[(S_T - K)^+ | S_0] = C(K,T|S_0), \forall K, T.$$

That is, the Black model prediction matches with the observed market price exactly.

Proof. Note that the Fokker-Planck equation(Lemma 8.1.1) governing probability evolution p of S_t under risk-neutral measure is given by

$$p_t = -[(r-d)sp]_s + \frac{1}{2}[(\sigma s)^2 p]_s s.$$

The martingale pricing method gives the call option price with maturity T and strike K given by

$$C = e^{-rT}E_Q[(S_T - K)^+] = e^{-rT}\int_K^\infty (s-K)pds = e^{-rT}\int_K^\infty spds - Ke^{-rT}\int_K^\infty pds.$$

Differentiating C with respect to K gives $C_K = -e^{-rT}\int_K^\infty pds$, then we have $C - KC_K = e^{-rT}\int_K^\infty spds$. Differentiating C with respect to K twice $C_{KK} = e^{-rT}p$. Differentiating C with respect to T gives

$$\begin{aligned}
C_T &= -rC - e^{-rT}\int_K^\infty (s-K)[(r-d)sp]_s ds + \frac{1}{2}e^{-rT}\int_K^\infty (s-K)[(\sigma s)^2 p]_{ss} \\
&= -rC + (r-d)e^{-rT}\int_K^\infty spds + \frac{1}{2}e^{-rT}(\sigma K)^2 p \\
&= -rC + (r-d)(C - KC_K) + \frac{1}{2}\sigma^2 K^2 C_{KK} \\
&= -(r-d)KC_K - dC + \frac{1}{2}\sigma^2 K^2 C_{KK}
\end{aligned}$$

□

Remark 17.3.1 (interpretation).

- Using the local volatility calculated from Dupire's equation and the martingale pricing formula, we have able to reproduce the market prices for calls with different strikes and maturities.
- Note that what we get from Dupire's equation is $\sigma(S_T, T)$. When we want to use $\sigma(S_T, T)$ in Monte Carlo simulation, we replace T by t and K by S_t. The validity of this replacement is the proof(the Fokker-Planck equation part).

Remark 17.3.2 (static smile property of local volatility surface). If the market option price set are the same for different spot price S_0, the local volatility $\sigma(S, T)$ will be the same.

Lemma 17.3.1 (deriving implied distribution from local volatility). *Let $\sigma_{loc}(K, t)$ be the local volatility satisfying Dupire's equation. Then the implied distribution for S_t at a future time t is given by*

$$S(t) = S(0)\exp(\int_0^t [\mu(S_u, u) - \frac{1}{2}\sigma(S_u, u)^2_{loc}]du + \int_0^t \sigma(S_u, u)^*_{loc} dB(u)).$$

where $B(t)$ is a Brownian motion, and μ is the drift at risk-neutral measure.

Proof. The solution to the SDE

$$dS(t) = \mu(t)S(t)dt + \sigma^*(t)S(t)dB(t)$$

is at Lemma 6.3.10. □

Lemma 17.3.2 (connection between local volatilty and implied volatility when there is no strike dependence). *[9, p. 279] If both implied volatility σ^* and local volatility σ_{loc} have no dependence on the strike K. Then,*

$$\frac{1}{T}\int_0^T \sigma^2_{loc}(u)du = \sigma^{*2}(T),$$

that is, implied volatility square is the time-average value of local volatility square.

Proof. When there is no strike dependence, the implied distribution from the local volatility is given by

$$\begin{aligned} S(T) &= S(0)\exp(\int_0^T [\mu(S_u, u) - \frac{1}{2}\sigma(S_u, u)^2_{loc}]du + \int_0^T \sigma(S_u, u)^*_{loc} dB(u)) \\ &= S(0)\exp(\int_0^T \mu(S_u, u)du - \frac{1}{2}\sigma_m^2 T + \sigma_m B(T)) \end{aligned}$$

where we use result in Lemma 6.3.9 to derive

$$\sigma_m^2 = \frac{1}{T}\int_0^T \sigma_{loc}^2(u)du.$$

The distribution of

$$S(0)\exp(\int_0^T \mu(S_u,u)du - \frac{1}{2}\sigma_m^2 T + \sigma_m B(T))$$

with σ_m replaced by σ_{loc} is exactly what we use to derive Black-Scholes equation. □

Lemma 17.3.3 (deriving local volatility from implied volatility). *[9, p. 278][10, p. 13] The local volatility can be derived from implied volatility, given by,*

$$\sigma^2(K,T;S_0) = \frac{C_T + (r-d)KC_K + dC}{\frac{1}{2}K^2 C_{KK}},$$

where

$$\begin{aligned} C_K &= \frac{\partial C_{BS}}{\partial K} + \frac{\partial C_{BS}}{\partial \sigma^*}\frac{\partial \sigma^*}{\partial K} \\ C_T &= \frac{\partial C_{BS}}{\partial T} + \frac{\partial C_{BS}}{\partial \sigma^*}\frac{\partial \sigma^*}{\partial T} \\ C_{KK} &= \frac{\partial^2 C_{BS}}{\partial K^2} + \frac{\partial^2 C_{BS}}{\partial \sigma^* \partial K}\frac{\partial \sigma^*}{\partial K} + (\frac{\partial}{\partial K}(\frac{\partial C_{BS}}{\partial \sigma^*})) + \frac{\partial C_{BS}}{\partial \sigma^*}\frac{\partial^2 \sigma^*}{\partial K^2} \end{aligned}$$

$$\sigma^2(K,T;S_0) = \frac{\frac{\partial\sigma}{\partial T} + \frac{\sigma}{T} + 2K(r-d)\frac{\partial\sigma}{\partial K}}{K^2(\frac{\partial^2\sigma}{\partial K^2} - D\sqrt{T}(\frac{\partial\sigma}{\partial K})^2 + \frac{1}{\sigma}(\frac{1}{K\sqrt{T}} + D\frac{\partial\sigma}{\partial K})^2)},$$

where $D = \frac{1}{\sigma\sqrt{T}}(\ln\frac{S_0}{K} + (r-d+\frac{1}{2}\sigma^2)T)$.

Denote $w = w(T,K) = \sigma^2(T,K)T$, *then*

$$\sigma^2(K,T;S_0) = \frac{\frac{\partial w}{\partial T}}{1 - \frac{\kappa}{w}\frac{\partial w}{\partial \kappa} + \frac{1}{4}(\frac{\partial w}{\partial \kappa})^2(\frac{\kappa^2}{w^2} - \frac{1}{w} - \frac{1}{4}) + \frac{1}{2}\frac{\partial^2 w}{\partial \kappa^2}}$$

where $\kappa = \ln(\frac{K}{F})$, *F is the forward price.*

Proof. Note that the market prices of calls, $C(K,T)$, can be written by

$$C(K,T) = C_{BS}(S_t, K, T, \sigma^*(K,T)),$$

where C_{BS} is the Black-Scholes formula for calls and σ^* is the implied volatility.

Then, we use chain rule. □

17.3.2 constant elasticity of variance model

17.3.3 Applications

Remark 17.3.3 (Local volatility method in binomial tree model). In each node(subsubsection 17.7.2.3), we can set the jumping probability based on the local volatility $\sigma_{loc}(S_t, t)$.

Note 17.3.1 (Monte Carlo with local volatility to price other options). [8, p. 47] We can use local volatility to simulate the dynamics of the underlying asset, given by

$$S(t + \Delta t) = S(t) + v_t S(t)\Delta t + \sigma_{loc}(t, S_t)\epsilon_t \sqrt{\Delta t},$$

where Δt is the integration step, ϵ_t is a standard normal random variable, and v_t is the risk-neutral drift.

17.4 Stochastic volatility model

17.4.1 The Fourier transform method

financial claim	payoff function	payoff transform	k-plane restrictions
European call	$\max(S_T - C, 0)$	$-\frac{C^{ik+1}}{k^2 - ik}$	$\Im k > 1$
European put	$\max(C - S_T, 0)$	$-\frac{C^{ik+1}}{k^2 - ik}$	$\Im k < 0$
covered call	$\min(S_T, C)$	$\frac{C^{ik+1}}{k^2 - ik}$	$0 < \Im k < 1$
delta function	$\delta(S_T - C)$	C^{ik}	none
money market	1	$2\pi\delta(k)$	none

Table 17.4.1: Generalized Fourier transform for common financial claims. ([11, p. 37])

17.4.2 Heston stochastic volatility model

Definition 17.4.1 (Heston stochastic volatility model). *The Heston model for an asset $S(t)$ with stochastic variance $V(t)$ is given by*

$$dS(t) = rS(t)dt + \sqrt{V(t)}S(t)dW_1(t),$$
$$dV(t) = (a - bV(t))dt + \sigma\sqrt{V(t)}dW_2,$$

where the interest rate r is constant, W_1 and W_2 are correlated Brownian motions with $dW_1 dW_2 = \rho dt$.

17.4.3 Black-Scholes model for stochastic volatility

Lemma 17.4.1 (Black-Scholes model with stochastic volatility). *[12, p. 300][9, p. 345][13, p. 882] Assume **under real-world probability measure**, we have dynamics*

$$dS = \mu S dt + \sigma S dW_1$$
$$d\sigma = p(S, \sigma, t)dt + q(S, \sigma, t)dW_2$$

where W_1 and W_2 are two correlated Brownian motions with $dW_1 dW_2 = \rho$.

Then the value of an option on S, denoted by $V(S,\sigma,t)$ is governed by

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2}+rS\frac{\partial V}{\partial S}+(p-\lambda q)\frac{\partial V}{\partial\sigma}-rV=0.$$

Proof. Consider a portfolio

$$\Pi=V-\Delta S-\Delta_1 V_1,$$

where V_1 is another tradable option to hedge volatility risks.

$$\begin{aligned}d\Pi=&(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2})dt\\ &-\Delta_1(\frac{\partial V_1}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V_1}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V_1}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V_1}{\partial\sigma^2})dt\\ &+(\frac{\partial V}{\partial S}-\Delta-\Delta_1\frac{\partial V_1}{\partial S})dS\\ &+(\frac{\partial V}{\partial\sigma}-\Delta_1\frac{\partial V_1}{\partial\sigma})d\sigma\end{aligned}$$

To eliminate all randomness, we choose Δ,Δ_1 such that

$$\frac{\partial V}{\partial S}-\Delta-\Delta_1\frac{\partial V_1}{\partial S}=0,\frac{\partial V}{\partial\sigma}-\Delta_1\frac{\partial V_1}{\partial\sigma}=0.$$

This gives us

$$\Delta_1=\frac{\partial V/\partial\sigma}{\partial V_1/\partial\sigma},\Delta=\frac{\partial V}{\partial S}+\frac{\partial V/\partial\sigma}{\partial V_1/\partial\sigma}\frac{\partial V_1}{\partial S}.$$

Then, we have

$$\begin{aligned}d\Pi=&(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2})dt\\ &-\Delta_1(\frac{\partial V_1}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V_1}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V_1}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V_1}{\partial\sigma^2})dt\\ &=r\Pi dt=r(V-\Delta S-\Delta_1V_1)dt\end{aligned}$$

Rearrange and remove cancel dt, we have

$$\begin{aligned}&\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2}+rS\frac{\partial V}{\partial S}-rV/\frac{\partial V}{\partial\sigma}\\ =&\frac{\partial V_1}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V_1}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V_1}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V_1}{\partial\sigma^2}+rS\frac{\partial V_1}{\partial S}-rV_1/\frac{\partial V_1}{\partial\sigma}\end{aligned}$$

Note that since V_1 is an arbitrary option, therefore, the ratio can only be a function, denoted by $-(p-\lambda q)$, of S, σ, t (without V, V_1). Thus, we have

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma S q\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2}+rS\frac{\partial V}{\partial S}-rV/\frac{\partial V}{\partial\sigma}=-(p-\lambda q).$$

□

Lemma 17.4.2 (risk-neutral pricing in stochastic volatility model). *Assume there exists a risk-neutral measure Q such that the dynamics of the asset and the volatility is given by*

$$\begin{aligned} dS &= rSdt+\sigma SdW_1 \\ d\sigma &= m(S,\sigma,t)dt+q(S,\sigma,t)dW_2 \end{aligned}$$

where W_1 and W_2 are two correlated Brownian motions with $dW_1dW_2=\rho$.

Then for a derivative with payoff $V_T(S_T)$, its current value is given by

$$V(S_t,t)=E_Q[e^{-r(T-t)}V_T(S_T)|\mathcal{F}_t]$$

Proof. We want to show that the dynamics of $e^{-rt}V(t,S_t,\sigma_t)$ **under risk-neutral measure** is driftless SDE.

$$\begin{aligned}
&d(\exp(-rt)V(t,S_t,\sigma_t)\\
&=-\exp(-rt)rV+\exp(-rt)dV\\
&=-\exp(-rt)rV+\exp(-rt)[(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma Sq\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2})dt+\frac{\partial V}{\partial S}dS+\frac{\partial V}{\partial\sigma}d\sigma]\\
&=\frac{\partial V}{\partial S}\sigma SdW_1+\frac{\partial V}{\partial\sigma}qdW_2\\
&=\sqrt{(\frac{\partial V}{\partial S}\sigma S)^2+(\frac{\partial V}{\partial\sigma}q)^2+(2\rho\frac{\partial V}{\partial\sigma}q)\frac{\partial V}{\partial S}\sigma S}dW_3
\end{aligned}$$

where W_3 is a Brownian motion and we use the Black-Scholes equation in the derivation.

□

Remark 17.4.1 (risk-neutral interpretation using Feyman Kac theorem). In Lemma 17.4.1, we show that the value of the derivative is governed by

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}+\rho\sigma S q\frac{\partial^2 V}{\partial S\partial\sigma}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial\sigma^2}+rS\frac{\partial V}{\partial S}+(p-\lambda q)\frac{\partial V}{\partial\sigma}-rV=0.$$

The Feyman-Kac theorem enables us to interpret the underlying dynamics, where we replace $(p-\lambda q)$ by m.

Remark 17.4.2 (how to implement the pricing model).

- We can use Monte Carlo simulation to evaluate the expectation as long as we have the model parameter under risk-neutral measure.
- Even we know the model parameter in the real-probabiliy measure(which can be estimated from the historical data), we cannot use them. The risk-neutral parameter has to be estimated from prices of other derivatives.

Note 17.4.1 (dynamics in real probability measure and risk-neutral measure). The asset dynamics under risk-neutral probability is given by

$$dS = rSdt + \sigma S dW_t$$

and

$$\begin{aligned} dV &= (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \frac{\partial V}{\partial S} dS \\ &= rVdt + \frac{\partial V}{\partial S}\sigma S dW_1 + \frac{\partial V}{\partial \sigma} q dW_2 \end{aligned}$$

where we use the Black-Scholes equation in the derivation.

Under real probability measure, the dynamics are

$$dS = \mu S dt + \sigma S dW_t$$

and

$$\begin{aligned} dV &= (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2} + \rho q \sigma \frac{\partial^2 V}{\partial \sigma^2})dt + \frac{\partial V}{\partial S} dS + \frac{\partial V}{\partial \sigma} d\sigma \\ &= rV + \frac{\partial V}{\partial S} S(\mu - r)dt + \frac{\partial V}{\partial \sigma}(p - m) + \frac{\partial V}{\partial S}\sigma S dW_1 + \frac{\partial V}{\partial \sigma} q dW_2 \end{aligned}$$

If we define $\lambda_1 = (\mu - r)/\sigma$ and $\lambda_2 = (p - m)/q$, we find that

$$\begin{aligned} dS/S &= (r + \lambda\sigma)dt + \sigma dW_t \\ dV/V &= (r + \lambda_1 \frac{\partial V}{\partial S}\frac{\sigma S}{V} + \lambda_2 \frac{\partial V}{\partial \sigma}\frac{q}{V})dt + \frac{\partial V}{\partial S}\frac{\sigma S}{V} dW_1 + \frac{\partial V}{\partial \sigma}\frac{q}{V} dW_2 \end{aligned}$$

which is consistent with the no-arbitrage condition for multiple source uncertainty dynamics(Theorem 15.5.1).

Example 17.4.1 (Heston stochastic volatility model). [1, p. 288] Suppose that under a risk-neutral measure Q, a stock price is governed by

$$dS(t) = rS(t)dt + \sqrt{V(t)}S(t)dW(t),$$

where the interest rate r is constant and volatility $\sqrt{V(t)}$ is governed by

$$dV(t) = (a - bV(t))dt + \sigma\sqrt{V(t)}dW_2,$$

where W_1 and w_2 are correlated Brownian motions with $dW_1 dW_2 = \rho dt$.

Consider a call maturing at $T \geq t$ has price at t given by

$$c(t, S(t), V(t)) = E_Q[e^{-r(T-t)}\max(S(T) - K, 0)|\mathcal{F}_t],$$

then $c(t, s, v)$ will satisfy the following PDE

$$c_t + rsc_s + (a - bv)c_v + \frac{1}{2}s^2 vc_{ss} + \rho\sigma svc_{sv} + \frac{1}{2}\sigma^2 vc_{vv} = rc$$

in the region $0 \leq t \leq T, s \geq 0$ and $v \geq 0$. The function $c(t, s, v)$ also satisfies the boundary condition

$$\begin{aligned}
c(T, s, v) &= \max(s - K, 0), \forall s \geq 0, v \geq 0 \\
c(t, 0, v) &= 0, \forall 0 \leq t \leq T, v \geq 0 \\
c(t, s, 0) &= \max(s - e^{-r(T-t)}K, 0), \forall 0 \leq t \leq T, s \geq 0 \\
\lim_{s\to\infty} \frac{c(t, s, v)}{S - K} &= 1, \forall 0 \leq t \leq T, v \geq 0 \\
\lim_{v\to\infty} c(t, s, v) &= s, \forall 0 \leq t \leq T, s \geq 0
\end{aligned}$$

Remark:
It has practical difficulties in using the pricing formula since $V(t)$ cannot be directly observed from the market.

17.4.4 SARB model

17.4.4.1 *Motivations*

Note 17.4.2. [14]

17.4.4.2 *The model*

Definition 17.4.2 (SABR stochastic volatility model). *The futures prices $F(t, T_m)$ with maturity T_m is given by*

$$\begin{aligned} dF(t, T_m) &= \alpha(t)F^\beta(t, T_m)dW_1 \\ d\alpha(t, T_m) &= \nu\alpha(t, T_m)dW_2 \\ \alpha(0) &= \alpha \\ dW_1 dW_2 &= \rho dt \end{aligned}$$

where $\alpha(t, T_m)$ is the stochastic volatility associated with maturity T_m, β is the CEV constant, ν is the volatility of volatility, α is the initial value of the volatility, and ρ is the correlation coefficient.

Lemma 17.4.3 (European call/option pricing). *[14]*

- *The European call price with strike K and maturity T is given by*

$$V_c(t) = P(t, T)(F(t, T)N(d_1) - KN(d_2))$$

where $F(t, T)$ is the forward price the underlying, and

$$d_{1,2} = \frac{\log(F(t, T)/K) \pm \frac{1}{2}\sigma_B^2(T - t)}{\sigma_B^2\sqrt{T - t}},$$

and the implied volatility $\sigma(F, K)$ is given by

$$\begin{aligned} \sigma_B(F, K) &= \frac{\alpha}{(FK)^{(1-\beta)/2}[1 + \frac{(1-\beta)^2}{24}\log^2(F/K) + \frac{(1-\beta)^4}{1920}\log^4(F/K) + \cdots]} \\ &\quad \cdot \frac{z}{\chi(z)} \cdot \{1 + [\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(FK)^{(1-\beta)/2}} + \frac{2 - 3\rho^2}{24}\nu^2](T - t) + \cdots\} \\ z &= \frac{\nu}{\alpha}(FK)^{(1-\beta)/2}\log(F/K) \\ \chi(z) &= \log\frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho} \end{aligned}$$

- *At-the-money volatility*

$$\sigma_B(F, F) = \frac{\alpha}{(FK)^{(1-\beta)/2}}\{1 + [\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(FK)^{(1-\beta)/2}} + \frac{2 - 3\rho^2}{24}\nu^2](T - t) + \cdots\}$$

- *The put price given by*

$$V_p(t) = V_c(t) + P(t,T)(K - F(t,T)).$$

Note 17.4.3 (choices of parameters for the equity volatility surface)**.**

- At every maturity T, we use four parameter α_0, β, ν and ρ to characterize the volatility smile; usually, β is pre-selected, and the rest is calibrated to implied volatility surface using non-linear optimization.
- For a set of maturity dates $T_1, T_2, ..., T_n$, we use a set of $4n$ parameters to characterize the volatility surface. Usually, β is chosen to be constant independent of maturity date.

Note 17.4.4 (choices of parameters for the interest rate volatility surface)**.** subsection 19.11.1

- At every maturity T and the interest rate or swap rate tenors Δ, we use four parameter α_0, β, ν and ρ to characterize the volatility smile; usually, β is pre-selected, and the rest is calibrated to implied volatility surface using non-linear optimization.
- For a set of maturity dates $T_1, T_2, ..., T_n$ and tenors $\Delta_1, \Delta_2, ..., \Delta_m$, we use a set of $4mn$ parameters to characterize the volatility surface/cube. Usually, β is chosen to be constant independent of maturity date and tenors.

17.4.4.3 *Managing smile risk*

Note 17.4.5 (SABR model and the risks associated with vanilla options)**.**

- Let $BS(f, K, \sigma, T)$ denote the call option prices at time 0. Then the call option price under SABR model can be written as

$$V(f, K, \sigma_B(f, K; \alpha, \beta, \rho, \nu), T).$$

Therefore, market risk is associated with parameters $f, \alpha, \beta, \rho, \mu$(usually β is assumed to be a constant and therefore will be the source of market risk).
- The eliminate the market risks, we want to maintain a portfolio, whose value is denoted by V_P, such that

$$\frac{\partial V_P}{\partial f} = \frac{\partial V_P}{\partial \alpha} = \frac{\partial V_P}{\partial \beta} = \frac{\partial V_P}{\partial \nu} = 0,$$

where

$$\begin{aligned}
\frac{\partial V}{\partial f} &= \frac{\partial BS}{\partial f} + \frac{\partial BS}{\partial \sigma_B}\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \mu)}{\partial f} \\
\frac{\partial V}{\partial \alpha} &= \frac{\partial BS}{\partial \sigma_B}\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \mu)}{\partial f} \\
\frac{\partial V}{\partial \nu} &= \frac{\partial BS}{\partial \sigma_B}\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \mu)}{\partial \nu} \\
\frac{\partial V}{\partial \rho} &= \frac{\partial BS}{\partial \sigma_B}\frac{\partial \sigma_B(K, f; \alpha, \beta, \rho, \mu)}{\partial \rho}
\end{aligned}$$

- To achieve this, we can buy/sell underlying, and other options. Note that α, β, ν are non-tradables; therefore, we need to buy and sell other options to make the portfolio neutral.

17.5 Multi-factor model

17.5.1 Primer on factor model dynamics

Definition 17.5.1 (one factor model for a curve). *Let $F(t,T)$ denote a forward curve. A factor model is given by*

$$dF(t,T) = \sigma(t,T)dW_t.$$

We can interpret this SDE as inifite number SDE parameterized by parameter T sharing the same risk driver W_t.

Remark 17.5.1 (about factor model)**.**

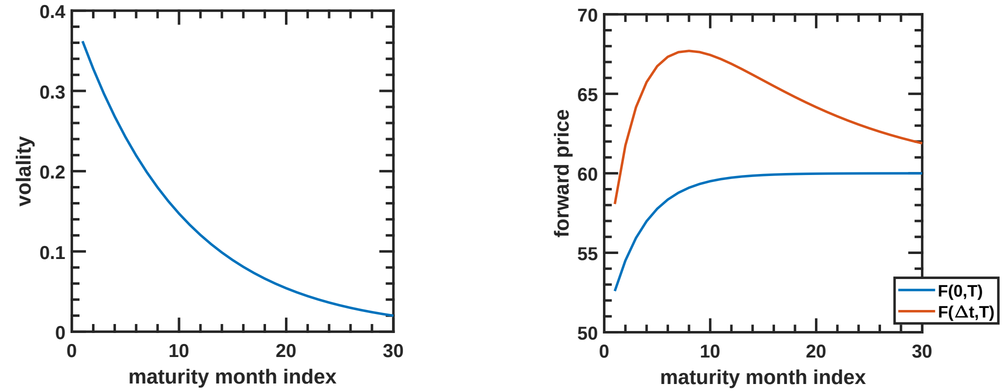
- **Static factor model** represent a large number N of random variables in terms of a small number K of difference random variables called factors.
- **Dynamic factor models** represent a large number N of time series in terms of a small number K of different time series called dynamic factors.

Note 17.5.1 (moving mode with different volatilty structures)**.** The first-order approximation discrete-time form of the one factor model is given by

$$F(t+\Delta t, T) = F(t,T) + \sigma(t,T)\sqrt{\Delta t}Z, Z \in T.$$

By choosing different volatility structure $\sigma(t,T)$, the one-factor model can produce different moving mode in the forward curve.

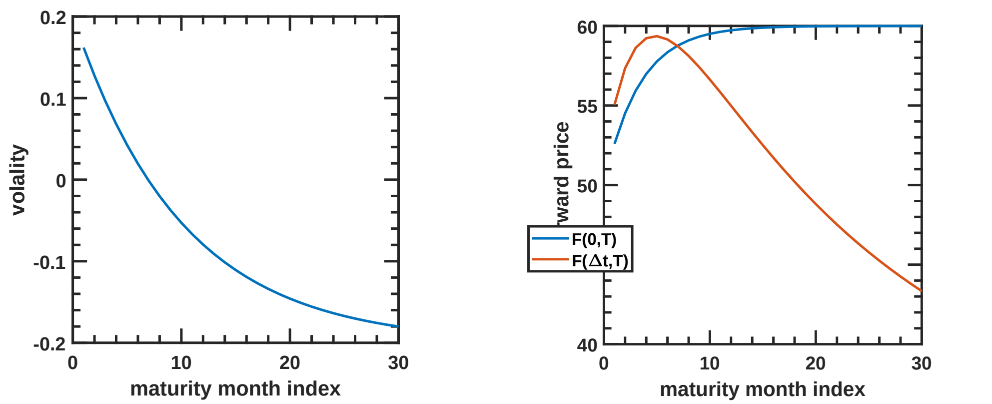
- (**parallel movment**) If we choose $\sigma(t,T) > 0$ for all T, then the curve $F(t+\Delta t, T)$ will be overall above(or below, depends on the realization of Z) the initial curve $F(t,T)$, as showed in Figure 17.5.1.
- (**twisting movement**) If we choose $\sigma(t,T) > 0$ for small T and $\sigma(t,T) < 0$ for large T, then the curve $F(t+\Delta t, T)$ will be above(or below, depends on the realization of Z) the initial curve $F(t,T)$ at the near end and below the initial curve $F(t,T)$ at the far end, as showed in Figure 17.5.2.
- (**bending movement**) If we choose $\sigma(t,T) > 0$ for small and large T, then the curve $F(t+\Delta t, T)$ will be overall above(or below, depends on the realization of Z) the initial curve $F(t,T)$, as showed in Figure 17.5.3.



(a) Volatility term structure $\sigma(t, T)$ as a function of T at time t.

(b) Demonstration of forward curve movement.

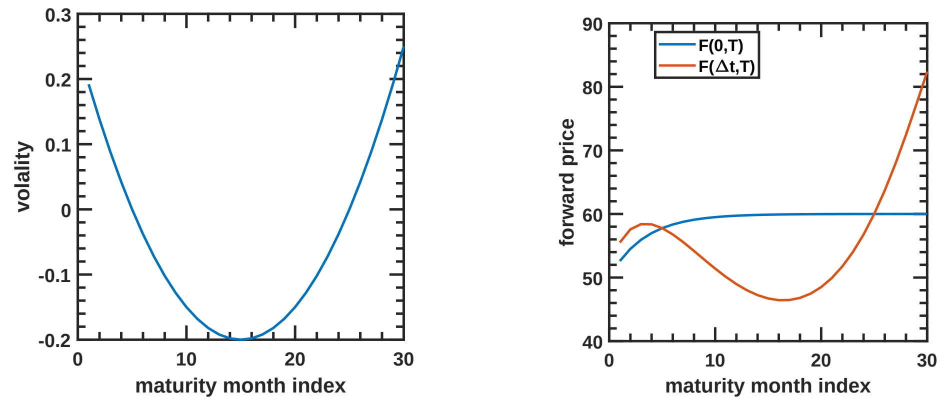
Figure 17.5.1: One factor model producing parallel movement of forward curve



(a) Volatility term structure $\sigma(t, T)$ as a function of T at time t.

(b) Demonstration of forward curve movement.

Figure 17.5.2: One factor model producing twisting movement of forward curve



(a) Volatility term structure $\sigma(t, T)$ as a function of T at time t.

(b) Demonstration of forward curve movement.

Figure 17.5.3: One factor model producing bending movement of forward curve

17.6 The real-world market

17.6.1 Calibration and pricing

17.6.2 The role of model

Remark 17.6.1 (the role of model). [2, p. 440]

- How important is the option-pricing model if traders are prepared to use a different volatility for every option? It can be argued that the BlackâĂŞScholesâĂŞMerton model is no more than a sophisticated interpolation tool used by traders for ensuring that an option is priced consistently with the market prices of other actively traded options.
- If traders stopped using Black-Scholes-Merton and switched to another plausible model, then the volatility surface and the shape of the smile would change, but arguably the dollar prices quoted in the market would not change appreciably. Even delta, if calculated as outlined in the previous section, does not change too much as the model is changed.
- Models have most effect on the pricing of derivatives when similar derivatives do not trade actively in the market. For example, the pricing of many of the nonstandard exotic derivatives we will discuss in later chapters is model-dependent.

17.6.3 Extension to unhedgeable risk factors

Remark 17.6.2. Suppose we are modeling the mortgage related instruments, where the prepayment behavior of the borrower can be a complex function of GDP, inflation, monetary and fiscal policies. In other words, the prepayment can be modeled as a stochastic process driven by a set of external risk factors independents of those risk factors driving the asset prices.

We can still using martingale method and change of measure to make the discounted asset prices between martingale and procced to pricing. The change of measure will not affect the external risk factors and the prepayment stochastic process.

17.7 Tree model

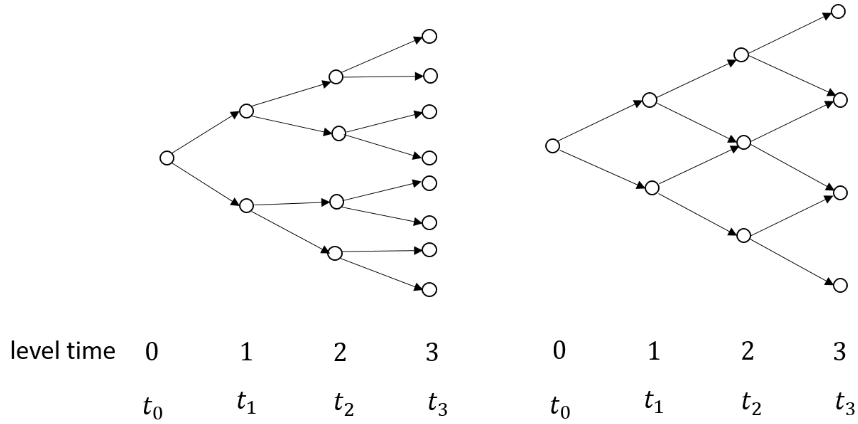


Figure 17.7.1: non-recombining tree (left) vs. recombining tree (right).

17.7.1 Binomial tree market model

Definition 17.7.1 (Binomial tree market model).

- *A risky stock price process $S(t)$, is represented by $S = \{S(t), t = 0, 1, ..., T\}$.*
- *In each time period the stock price either goes up by a factor $u > 1$ with probability p, or goes down by a factor $0 < d < 1$ with probability $1p$. For each t, $S(t) = S(0)u^{n_t}d^{tn_t}$, where n_t represents the number of up moves up to t.*
- *The bank account process B is deterministic with $B(0) = 1$ and a constant interest rate $0 < r < 1$. Hence $B(t) = (1+r)^t$.*
- *The sample space Ω contains $K = 2^T$ different paths. Each sample point ω is a path with $U(\omega)$ up moves, and $T - U(\omega)$ down moves. The underlying probability P is defined by*
$$P(\omega) = p^U(\omega)(1p)^{TU(\omega)}, 0 < p < 1.$$

Lemma 17.7.1. *There exists a **unique** martingale equivalent measure Q such that*

$$Q(\omega) = q^U(\omega)(1-q)^{TU(\omega)}, q = \frac{1+r-d}{u-d}$$

if and only if $d < 1 + r < u$.

Moreover, the binomial tree model is complete.

Proof. Let $\theta_t = n_t - n_{t-1}$ (i.e., $\theta_t = 1$ is a Bernoulli random variable and n_t is the Binomial random variable) Then for every t, the martingale condition gives

$$\frac{S(t)}{B(t)} = S^*(t) = S^*(t-1)(1+r)^{-1}u^{\theta_t}d^{1-\theta_t}.$$

Therefore,

$$\begin{aligned}
&E_Q[S^*(t)|\mathcal{F}_{t-1}] = S^*(t-1) \\
\Leftrightarrow & uQ(\theta_t = 1|n_{t-1}) + d(1 - Q(\theta_t = 1|n_{t-1})) = 1 + r \\
\Leftrightarrow & q = Q(\theta_t = 1|n_{t-1}) = \frac{1+r-d}{u-d}
\end{aligned}$$

Only when $d < 1 + r < u$, we can guarantee $0 < q < 1$.

The tree model is complete because the martingale measure is unique. □

Remark 17.7.1 (interpretation on parameterizing probability measure function)**.**

- For a multi-period finite state model, it may require $K^T - 1$ number of fully parameterize the probability measure funciton. Then we might need multiple existing assets to find an unique martingale measure.
- The tree model using only one parameter, i.e., jump probability to specify the dynamics of the model and only one parameter to parameterize the probability measure function.

Example 17.7.1 (pricing a call using binomial tree)*.* Consider a call with payoff $g(S(T)) = (S(T) - K)^+$. Note that

$$S(t)u^n d^{T-t-n} - K > 0 \Leftrightarrow n > \frac{\log(K/S(t)d^{T-t})}{\log(u/d)}.$$

Let n^* be the smallest n such that the inequality holds.

Then the value of the call at time t is given by

$$\begin{aligned}
V(t) = & \underbrace{\frac{S(t)B(t)}{B(T)} \sum_{n=n^*}^{T-t} \binom{T-t}{n} (uq)^n ((1-q)d)^{T-t-n}}_{B(t)E_Q[S(T)/B(T)\mathbf{1}(S(T)>K)|\mathcal{F}_t]} \\
& - \underbrace{\frac{KB(t)}{B(T)} \sum_{n=n^*}^{T-t} \binom{T-t}{n} (q)^n ((1-q))^{T-t-n}}_{B(t)E_Q[K/B(T)\mathbf{1}(S(T)>K)|\mathcal{F}_t]}
\end{aligned}$$

17.7.2 Binomial models

17.7.2.1 *Basic constructions*

Definition 17.7.2 (binomial tree model). *A binomial tree model for an asset (e.g. stock) dynamics consists of*

- *nodes (i,j): representing date $i = 0,1,...,N$ and states $j = 0,1,...,N$*
- *values on node (i,j) represent the asset price at i and state j.*
- *the binomial tree model is a representation of discrete-time asset stochastic process $S(1), S(2), ..., S(N)$, where $S(i)$ can take values $s_{i,1}, ..., s_{i,i}$.*
- *state transition probability:*

$$P((i+1,s)|(i,j)) = \begin{cases} q, s = j+1 \\ 1-q, s = j \end{cases}$$

- *the sample space Ω consists of all possible 2^N sample paths; We use $\omega_{i,j}$ to denote the set of sample paths that pass through state j at time i.*

17.7.2.2 *State price and general pricing*

Definition 17.7.3 (Arrow-Debreu security). *An* ***Arrow-Debreu security*** *is a security that has a payoff that solely pays 1 at time n if world state $\omega_{n,j}$ is realized. Its price, called,* ***state price*** *at $t = 0$ is denoted as $\lambda_{n,j}$*

With the existence of risk-neutral measure Q, we have

$$\lambda_{n,j} = \frac{1}{(1+r\Delta t)^n} E_Q[\mathbf{1}(S_n = s_{n,j})] = \frac{1}{(1+r\Delta t)^n} E_Q[\mathbf{1}(\omega_{n,j})].$$

Further, the probability of reaching state $s_{n,j}$ from state $s_{0,0}$ is given by

$$Pr_Q(S_n = s_{n,j}) = (1+r\Delta t)^n \lambda_{n,j}.$$

Theorem 17.7.1 (state price recursive relation). *In a N period binomial tree model, where there will be $K+1$ states, then state prices satisfying* ***forward equations****:*

$$\begin{aligned}\lambda_{n+1,j} &= \frac{1}{1+r\Delta t}(Q_n(j;j-1)\lambda_{n,j-1} + Q_n(j;j)\lambda_{n,j}), 0 < j \leq n+1 \text{ (interior)}\\ \lambda_{n+1,0} &= \frac{1}{1+r\Delta t}Q_n(0;0)\lambda_{n,0} \text{ (upper boundary)}\\ \lambda_{n+1,n+2} &= \frac{1}{1+r\Delta t}Q_n(n+2;n+1)\lambda_{n,n+1} \text{ (lower boundary)}\end{aligned}$$

where we have ***boundary condition*** *$\lambda_{0,0} = 1$, $Q_n(j;i)$ denotes the equilibrium measure for transitioning from state i at period n to state j at period $n+1$, and r is the interest rate.*

More over, we have

$$\sum_{j=1}^{n+2} \lambda_{n+1,j} = \frac{1}{1+r\Delta t}\sum_{j=1}^{n+1} \lambda_{n,j}.$$

Proof. (1)Consider an interior node (n,i). By definition,

$$\begin{aligned}\lambda_{n,i} &= \frac{1}{(1+r\Delta t)^n}E_Q[\mathbf{1}(S_n = s_{n,i})]\\ &= \frac{1}{(1+r\Delta t)^n}Pr_Q(S_n = s_{n,i}) \text{ (} Pr_Q \text{ is the measure)}\\ &= \frac{1}{(1+r\Delta t)^n}[Pr_Q((S_n = s_{n,i}) \cap (S_{n-1} = s_{n-1,i}) + Pr_Q((S_n = s_{n,i}) \cap (S_{n-1} = s_{n-1,i-1})]\\ &= \frac{1}{(1+r\Delta t)^n}[Q_{n-1}(i,i)Pr_Q(S_{n-1} = s_{n-1,i}) + Q_{n-1}(i,i-1)Pr_Q(S_{n-1} = s_{n-1,i-1})]\\ &= \frac{1}{(1+r\Delta t)}[\frac{1}{(1+r\Delta t)^{n-1}}Q_{n-1}(i,i)E_Q[\mathbf{1}(S_{n-1} = s_{n-1,i})] + Q_{n-1}(i,i-1)E_Q[\mathbf{1}(S_{n-1} = s_{n-1,i-1})]]\\ &= \frac{1}{1+r\Delta t}(Q_n(i;i)\lambda_{n-1,i} + Q_n(i;i-1)\lambda_{n-1,i-1})\end{aligned}$$

(2)

$$\begin{aligned}\sum_{j=1}^{n+2} \lambda_{n+1,j} &= \frac{1}{(1+r\Delta t)}(\lambda_{n,0}(Q_n(0,0) + Q_n(1,0)) + \lambda_{n,1}(Q_n(1,1) + Q_n(2,1)) + \cdots + \lambda_{n,n+1}(Q_n(n+1,n\\ &= \frac{1}{(1+r\Delta t)}(\lambda_{n,0} + \lambda_{n,1} + \cdots + \lambda_{n,n+1})\end{aligned}$$

□

Theorem 17.7.2 (pricing simple derivatives using state prices). *Suppose $\lambda_{i,j}$ is known for all possible i, j.*

- *for a zero coupon bond that has a payoff 1 at period n, its no-arbitrage price at time 0 is*
$$V(0) = \sum_{j=1}^{n+1} \lambda_{n,j} = \frac{1}{(1 + r\Delta t)^n}.$$
- *for any derivative that has a payoff vector $D(S_n)$ at period n, its arbitrage price at time 0is*
$$V(0) = \sum_{j=1}^{n+1} \lambda_{n,j} D(s_{n,j}).$$

Proof. (1) Using relationship

$$\lambda_{i,j} = \frac{1}{(1 + r\Delta t)^n} E_Q[\mathbf{1}(S_i = s_{i,j})] = \frac{1}{(1 + r\Delta t)^n} E_Q[\mathbf{1}(\omega_{i,j})]$$

(2) Using either linear pricing theorem(Theorem 15.2.2). Or using definition

$$V(0) = \frac{1}{(1 + r\Delta t)^n} E_Q[D(S_n)] = \sum_{j=1}^{n+1} \frac{1}{(1 + r\Delta t)^n} D(s_{n,j}) E_Q[\mathbf{1}(S_n = s_{n,j})].$$

□

Remark 17.7.2 (pricing more complex derivative using state price).

- Pricing using state price usually limit to European style derivatives or American style derivatives with addition dynamical programming.
- For more complex derivative that involves the joint distribution of S at multiple time points (e.g., spread option), the state price method cannot help.

17.7.2.3 *GBM to a Binomial Model*

Given the continuous geometric Brownian motion, we can convert them into equivalent binomial model parameters:

- $R_n = \exp(rT/n)$, where n is the number of periods in binomial model
- $R_n - c_n = \exp((r - c)T/n) \approx 1 + rT/n - cT/n$, r is the short rate and c is the constant dividend paying rate.
- $u_n = \exp(\sigma\sqrt{T/n})$
- $d_n = 1/u_n$
- the risk-neutral probability is calculated as $q = \frac{\exp((r-c)T/n) - d}{u - d}$

Definition 17.7.4 (binomial model). *A binomial lattice model for the stock consists of*

- *nodes (i,j): representing date $i = 0,1,...,n$ and states $j = 0,1,...,n$*
- *values on node (i,j) represent the stock price at i and state j.*
- *state transition probability:*

$$P((i+1,s)|(i,j)) = \begin{cases} q, s = j+1 \\ 1-q, s = j \end{cases}$$

17.7.2.4 *Convergence of Binomial model*

Theorem 17.7.3 (convergence in distribution of the Binomial model). *[15, p. 169]* *Let $r_n = e^{r/n} - 1, d_n = e^{-\sigma/\sqrt{n}}, u_n = e^{\sigma/\sqrt{n}}$ such that the equivalent equivalent martingale measure of the nth binomial model for simulating S is given as*

$$q = \frac{r_n + 1 - d_n}{u_n - d_n}$$

Then the multi-period binomial model for $S_n(t)$ under risk-neutral probability converges in distribution (as $n \to \infty$) to

$$S_t = S_0 \exp(\sigma W_t + (r - \frac{1}{2}\sigma^2)t)$$

where S_t is the solution to the Ito process

$$dS = rSdt + \sigma SdW$$

Remark 17.7.3 (implications). We know that under martingale measure, the geometric Brownian motion of the stock will be

$$dS = rSdt + \sigma SdW.$$

And this theorem shows that if we increase n, the binomial tree simulation is approximating the stock price dynamics under martingale measure.

17.7.3 Implied tree

17.7.4 Multifactor trees

17.7.4.1 *Two factor trees*

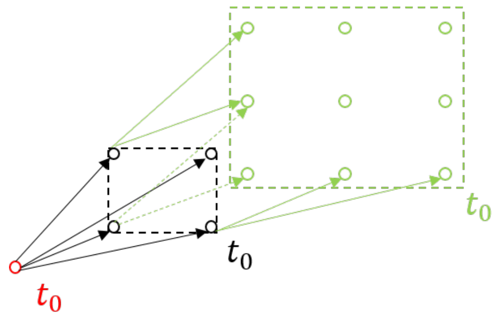


Figure 17.7.2: A three-dimensional tree representation of two-asset dynamics

Remark 17.7.4. A three-dimensional tree might be used to approximate the two-asset dynamics given by

$$\begin{aligned} dS_1/S_1 &= (r-q_1)dt + \sigma_1 dW_1 \\ dS_2/S_2 &= (r-q_2)dt + \sigma_2 dW_2 \\ dW_1 dW_2 &= \rho dt \end{aligned}$$

We can use transformation

$$x_1 = \sigma_2 \ln S_1 + \sigma_1 \ln S_2, x_2 = \sigma_2 \ln S_1 - \sigma_1 \ln S_2,$$
$$S_1 = \exp(\frac{x_1 + x_2}{2\sigma_2}), S_2 = \exp(\frac{x_1 - x_2}{2\sigma_1}),$$

and use the following independent process

$$dx_1 = (\sigma_2(r - q_1 - \sigma_1^2/2) + \sigma_1(r - q_2 - \sigma_2^2/2))dt + \sigma_1\sigma_2\sqrt{2(1+\rho)}dZ_A$$
$$dx_1 = (\sigma_2(r - q_1 - \sigma_1^2/2) - \sigma_1(r - q_2 - \sigma_2^2/2))dt + \sigma_1\sigma_2\sqrt{2(1+\rho)}dZ_B$$

where Z_A, Z_B are indepndent Brownian motions.

17.8 Notes on bibliography

For no-arbitrage theory, see [16].

For treatment from economical perspective, see [17] [18].

For martingale methods, see [19].

For PDE methods, see [19][13].

For incomplete markets, see [20].

For the differences between real world measure and risk-neutral measure, see [21].

An excellent book of "P" method for risks and asset allocations, see [22].

For a comprehensive discussion on risks, see [23].

For statistical arbitrage, see [24].

[25].

For stochastic volatility model, see [11], [26], [27].

BIBLIOGRAPHY

1. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
2. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
3. Borovkova, S., Permana, F. J. & Weide, H. V. A closed form approach to the valuation and hedging of basket and spread options. *Journal of Derivatives* **14,** 8 (2007).
4. Iwasawa, K. Analytical formula for the European normal Black Scholes formula. *Unpublished working paper, viewed* **27** (2013).
5. Brigo, D., Mercurio, F., Rapisarda, F. & Scotti, R. Approximated moment-matching dynamics for basket-options simulation (2001).
6. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
7. Alexander, C. *Market Risk Analysis, Pricing, Hedging and Trading Financial Instruments* ISBN: 9780470772812. https://books.google.com/books?id=M-6UEM18UHkC (Wiley, 2008).
8. Bossu, S. *Advanced Equity Derivatives: Volatility and Correlation* (John Wiley & Sons, 2014).
9. Derman, E., Miller, M. & Park, D. *The Volatility Smile* ISBN: 9781118959169. https://books.google.com/books?id=WKfBDAAAQBAJ (Wiley, 2016).
10. Gatheral, J. *The volatility surface: a practitioner's guide* (John Wiley & Sons, 2011).
11. Lewis, A. L. *Option Valuation Under Stochastic Volatility II* (Finance Press, Newport Beach, CA, 2009).
12. Wilmott, P. *Derivatives: The Theory and Practice of Financial Engineering* ISBN: 9780471983897. https://books.google.com/books?id=9NPuAAAAMAAJ (Wiley, 1998).
13. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
14. Hagan, P. K. & Kumar, D. Managing smile risk. *Wilmott Magazine*, 84–108.
15. Steland, A. *Financial statistics and mathematical finance: methods, models and applications* (John Wiley & Sons, 2012).
16. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2004).
17. LeRoy, S. F. & Werner, J. *Principles of financial economics* (Cambridge University Press, 2014).

18. Mishkin, F. S. *The economics of money, banking, and financial markets* (Pearson education, 2007).
19. Pascucci, A. *PDE and martingale methods in option pricing* (Springer Science & Business Media, 2011).
20. Cernỳ, A. *Mathematical techniques in finance: tools for incomplete markets* (Princeton University Press, 2009).
21. Meucci, A. 'P'Versus' Q': Differences and Commonalities between the Two Areas of Quantitative Finance. *GARP Risk Professional*, 47–50 (2011).
22. Meucci, A. *Risk and asset allocation* (Springer Science & Business Media, 2009).
23. Jorion, P. *Financial Risk Manager Handbook: FRM Part I / Wiley Finance* **pt. 2.** ISBN: 9781118017913. https://books.google.com/books?id=4ceVmGJSNpcC (Wiley, 2010).
24. Leung, T. & Li, X. *Optimal Mean Reversion Trading: Mathematical Analysis and Practical Applications* (World Scientific, 2015).
25. Derman, E., Kani, I. & Goldman, N. C. Implied trinomial trees of the volatility smile. *Journal of Derivatives* (1996).
26. Zhu, J. *Applications of Fourier transform to smile modeling: Theory and implementation* (Springer Science & Business Media, 2009).
27. Mikhailov, S. & Nögel, U. *Heston's stochastic volatility model: Implementation, calibration and some extensions* (John Wiley and Sons, 2004).

18

EQUITY DERIVATIVES

18.1 Dynamic Models of assets

18.1.1 Geometric Brownian motion model

18.1.1.1 *Basics*

Definition 18.1.1 (geometric Brownian motion (GBM)model). *The GBM model for the stock price S_t is given as*

$$dS_t = \mu(t)S_t dt + \sigma(t)S_t dW_t$$

where W_t is Brownian motion.

Lemma 18.1.1 (solutions to geometric SDE, recap). *The solution to the SDE*

$$dX(t) = \mu(t)X(t)dt + \sigma(t)X(t)dB(t)$$

is given as

$$X(t) = X(0)\exp(\int_0^t [\mu(s) - \frac{1}{2}\sigma(s)^2]ds + \int_0^t \sigma(s)dB(s)).$$

Particularly, if $\mu(t)$ and $\sigma(t)$ is time independent, then

$$X(t) = X(0)\exp((\mu - \frac{1}{2}\sigma^2)t + \sigma B(t))$$

Proof. see Lemma 6.3.10. □

Lemma 18.1.2 (probability distribution). *The GBM (with constant coefficients)of the stock price S_t gives that*

$$S(t)/S(0) \sim LN((\mu - \frac{1}{2}\sigma^2)t, \sigma^2)$$

where LN is the lognormal representation.

Proof. See Definition 2.2.6. Note that $(\mu - \frac{1}{2}\sigma^2)t + \sigma B(t) \sim N((\mu - \frac{1}{2}\sigma^2)t, \sigma^2)$. □

Lemma 18.1.3 (martingale properties of the solution).

- *Let $M_t = \exp(\sigma W_t - \frac{1}{2}\sigma^2 t)$, where W_t is the Brownian motion. Then M_t is a martingale with respect the Brownian motion filtration.*

- *The GBM (with constant coefficients)of the stock price S_t is a martingale when $\mu = 0$; If $\mu > 0$, then $E[S_t|\mathcal{F}_u] > S_u, t > u$ (i.e., S_t is a supermartingale); If $\mu < 0$, then $E[S_t|\mathcal{F}_u] < S_u, t > u$ (i.e., S_t is a submartingale);*

Proof. see Lemma 5.5.2. □

Remark 18.1.1. In the real-world probability, stock price usually follow supermartingale; that is, holding stock will lead to increase(greater than risk-free rate) of fortune in the expectation sense. This is because

- The market is incomplete: (1) Not all the stocks have derivatives to hedging. (2) Even we have the derivatives, the risks cannot be completely hedged because of transaction cost, stochastic volatility, and incomplete information.
- stock holders are compensated for taking risks.

Definition 18.1.2 (multiple dimension geometric Brownian motion (GBM)model). *[1, p. 183]*

- *The N dimensional GBM model for the stock price $S_i, i = 1, ..., N$ is given as*

$$dS_i(t) = \mu_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dW_i(t)$$

*where $W_i, i = 1, ..., n$ is **correlated Brownian motion** such that*

$$E[dW_i] = 0, E[dW_i dW_j] = \rho_{ij}dt.$$

The instantaneous covariance structure is given by

$$Cov[dS_i dS_j] = \sigma_i\sigma_j\rho_{ij}dt.$$

- *(multiple factor representation) A m factor geometric Brownian motion model for the stock price $S_i, i = 1, ..., N$ is given by*

$$dS_i(t)/S_i(t) = \mu_i(t)dt + \sum_{i=1}^{m} \sigma_{ij}dW_j(t)$$

*where $W_i, i = 1, ..., m$ is **uncorrelated Brownian motion** such that*

$$E[dW_i] = 0, E[dW_i dW_j] = \delta_{ij}dt.$$

The instantaneous covariance structure is given by

$$Cov[dS_i dS_j] = \sum_{k=1}^{m} \sigma_{im}\sigma_{jm}dt.$$

Remark 18.1.2. The multiple factor model can be used as lower dimensional approximation when actual the instantaneous covariance structure has lower rank.

18.1.1.2 *Model calibration*

Lemma 18.1.4 (parameter estimation for single asset constant coefficient dynamics). *[2, p. 36] In the geometric Brownian motion model of the stock price given by,*

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

the drift μ and volatility σ can be estimated from a sequence of recent prices $\{S_0, S_1, S_2, ..., S_n\}$ with time step Δt, given as

- *Compute returns*
$$x_i = \ln S_i - \ln S_{i-1}, i = 1, 2, ..., n.$$
- *Compute statistics*
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$
- *(MLE)Estimate drift via*
$$\hat{\mu} = \frac{1}{\Delta t}\overline{x} + \frac{S_{xx}}{2\Delta t}.$$
- *(MLE)Estimate variance via*
$$\hat{\sigma}^2 = \frac{S_{xx}}{\Delta t}.$$

Proof. Note that for the SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

we can transform it to the following equivalent form

$$\begin{aligned} X_t &= \ln S_t \\ dX_t &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t. \end{aligned}$$

X_t will have the following discrete-time form

$$X(t + \Delta t) - X(t) = (\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t} Z, Z \in N(0, 1).$$

The rest follows results in maximum likelihood estimation in linear regression (**??**). □

Lemma 18.1.5 (parameter estimation for multiple asset constant coefficient dynamics). *[2, p. 36] In the correlated geometric Brownian motion model of the N stock price given by,*

$$dS^i(t) = \mu_i(t)S^i(t)dt + \sigma_i(t)S^i(t)dW_i(t)$$

where $W_i, i = 1, ..., d$ is correlated Brownian motion such that

$$E[dW_i] = 0, E[dW_i dW_j] = \rho_{ij}dt,$$

the drifts $\mu_i, i = 1, ..., N$ and volatilities $\sigma_i, i = 1, ..., N$ can be estimated from a sequence of recent prices $\{S_0^i, S_1^i, S_2^i, ..., S_n^i\}, i = 1, 2, ..., d$ with time step Δt, given as [3]

- *Compute returns*

$$x_i^j = \ln S_i - \ln S_{i-1}, i = 1, 2, ..., n; j = 1, 2, ..., d;$$

- *Compute statistics*

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^n x_i^j,$$
$$[S_{xx}]_{ij} = \frac{1}{n}\sum_{k=1}^n (x_k^i - \overline{x}_i)(x_k^j - \overline{x}_j).$$

- *Estimate drift via*

$$\hat{\mu}_i = \frac{1}{\Delta t}\sum_{i=1}^n x_i + \frac{[S_{xx}]_{ii}}{2\Delta t}.$$

- *Estimate variance via*

$$\hat{\sigma}_i^2 = \frac{[S_{xx}]_{ii}}{\Delta t}$$

- *Estimate correlation via*

$$\hat{\rho}_{ij} = \frac{[S_{xx}]_{ij}}{\Delta t \hat{\sigma}_i \hat{\sigma}_j}.$$

18.1.2 Models with jumps

18.2 Black-Scholes model: application examples

18.2.1 European call and put

18.2.1.1 *The pricing*

Lemma 18.2.1 (European call/put pricing). *[4, p. 219] Let current time t. Denote the spot by S_t and forward by $F(t,T)$. In the Black-Scholes model, an European call option with strike price K and expiry T is given as:*

$$C(S_t,t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)}$$

where

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r+\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) + \sigma^2/2(T-t)] \\ d_2 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) - \sigma^2/2(T-t)] \\ d_2 &= d_1 - \sigma\sqrt{T-t} \\ N(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ F(t,T) &= S_t \exp(r(T-t)) \end{aligned}$$

We further have

- *The price of a zero strike is S_t.*
- *The price of the put can be derived based on put-call parity $P_t + S_t = C_t + Ke^{-r(T-t)}$ (Lemma 15.2.1) as*

$$P(S_t,t) = N(-d_2)Ke^{-r(T-t)} - N(-d_1)S_t.$$

Proof. See Theorem 17.2.1. □

Remark 18.2.1 (interpretation). The call pricing formula is given as

$$C(S_t,t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)} = (N(d_1) - N(d_2))S_t + N(d_2)(S_t - Ke^{-r(T-t)}),$$

where $N(d_2)$ is the probability of stock price S_T greater than K(i.e. the probability the call will be exercised.). $N(d_1) - N(d_2)$ is the probability of increasing value from current stock.

Remark 18.2.2 (interpretation).

- The option price will increase as the volatility(as we showed in the Greeks).
- The option price depends only on model parameters σ and r of the underlying assets:risk-free bond and stock. One can use the market price of the option is calibrate σ and r.
- One can improve the model by allowing: (1) σ to follow stochastic volatility model; (2) r to follow stochastic short rate model.

Remark 18.2.3 (interpretation from original publication). [5]

- In general, it seems clear that the higher the price of the stock, the greater the value of the option. When the stock price is much greater than the exercise price, the option is almost sure to be exercised. The current value of the option will thus be approximately equal to the price of the stock minus the price of a pure discount bond that matures on the same date as the option, with a face value equal to the striking price of the option.
- On the other hand, if the price of the stock is much less than the exercise price, the option is almost sure to expire without being exercised, so its value will be near zero.
- If the expiration date of the option is very far in the future, then the price of a bond that pays the exercise price on the maturity date will be very low, and the value of the option will be approximately equal to the price of the stock.
- On the other hand, if the expiration date is very near, the value of the option will be approximately equal to the stock price minus the exercise price, or zero, if the stock price is less than the exercise price.

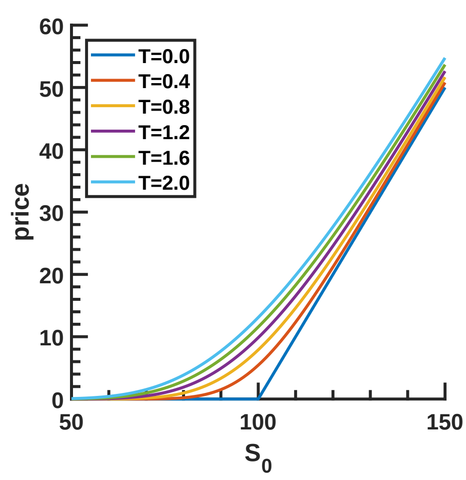


Figure 18.2.1: Price profile of a call option at different spot prices S_0 and different time to maturity T. The strike of the option is $K = 100$.

Lemma 18.2.2 (European call and put with dividends). *[4, p. 236][6, p. 373] Let current time t. Denote the spot by S_t and forward by $F(t,T)$. In the Black-Scholes model, an European call option with strike price K for an asset paying continuous dividends a is given as:*

$$C(S_t,t) = N(d_1)S_t e^{-a(T-t)} - N(d_2)Ke^{-r(T-t)}$$

where

$$\begin{aligned}
d_1 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-a+\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) + \sigma^2/2(T-t)] \\
d_2 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r-a-\sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) - \sigma^2/2(T-t)] \\
d_2 &= d_1 - \sigma\sqrt{T-t} \\
N(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
F(t,T) &= S_t \exp((r-a)(T-t))
\end{aligned}$$

The price of the put can be derived based on put-call parity

$$P_t + S_t e^{-a(T-t)} = C_t + Ke^{-r(T-t)}$$

(Lemma 15.2.1) as

$$P(S_t,t) = N(-d_2)Ke^{-r(T-t)} - N(-d_1)S_t e^{-a(T-t)}.$$

Proof. See Theorem 17.2.2. □

Example 18.2.1. [6, p. 374] Consider an European call option on S&P500 that is two months from maturity. The current value of the index is 930, the exercise price is 900, the risk-free rate is 8% per annum, and the volatility of the index is 20% per annum. Dividend yields is 3% per annum. Then we have $S_0 = 930, K = 900, r = 0.08, \sigma = 0.2$, and $T = 2/12$. And

$$d_1 = \frac{1}{0.2\sqrt{2/12}}[\ln(\frac{930}{900}) + (0.08 - 0.03 + 0.02^2/2)(2/12) = 0.5444;$$

$$d_2 = \frac{1}{0.2\sqrt{2/12}}[\ln(\frac{930}{900}) + (0.08 - 0.03 - 0.02^2/2)(2/12) = 0.4628.$$

$$N(d_1) = 0.7069, N(d_2) = 0.6782;$$

and the call price is

$$c = 930 \times N(d_1)\exp(-q \times 2/12) - 900 \times N(d_2)\exp(-r \times 2/12) = 51.83.$$

Lemma 18.2.3 (implied dividend yield). *[6, p. 376] If we want to get the market perspective on the dividend yield, we can use*

$$q = -\frac{1}{T}\ln\frac{C - P + K\exp(-rT)}{S_0},$$

where S_0 is the spot price, and C/P are the call/put prices with the same strike K and maturity T.

Proof. Use the put-call parity

$$C - P = S_0\exp(-qT) - K\exp(-rT).$$

□

18.2.1.2 *Other relations*

Lemma 18.2.4.
- *Let X_T be a martingale and $Pr(X_T < 0) = 0$. Then for all positive K, T such that*
$$(X_0 - K)^+ < E[(X_T - K)^+] < X_0.$$
- *When the expectation is taken under the risk-neutral measure, we have*
$$(F(0,T) - K)^+ < E_Q[(F(T,T) - K)^+] < F(0,T).$$
$$(S_0\exp((r-a)T) - K)^+ < E_Q[(S_T - K)^+] < S_0\exp((r-a)T).$$
- *The call option price bound*
$$(S_0 - K))^+ < (S_0 - K\exp(-(r-a)T))^+ < \exp(-(r-a)T)E_Q[(S_T - K)^+] = C < S_0.$$

Proof. (1) The function $f(x) = (x - K)^+$ is convex. Jansen inequality gives that

$$E[(X_T - K)^+] > [(E[X_T] - K)^+] = (X_0 - K)^+.$$

And

$$E[(X_T - K)^+] < E[(X_T - K = K)^+] = E[(X_T)^+] = E[X_T] = X_0.$$

(2) (a) Use the fact that $F(t,T)$ is a martingale under risk-neutral measure. (b) use $F(t,T) = S_t\exp((r-a)(T-t))$. (b) The call option price is

$$C = \exp(-(r-a)T)E_Q[(S_T - K)^+],$$

therefore is smaller than spot S_0 and greater than intrinsic $(S_0 - K)^+$. □

Lemma 18.2.5. *Consider the black formula for call and put pricing given by*

$$C(S_t, K, \sigma, t) = N(d_1)S_t e^{-q\tau} - N(d_2)Ke^{-r\tau}, C(F(t,T), K, \sigma, \tau) = N(d_1)F(t,T)e^{-r\tau} - N(d_2)Ke^{-r\tau},$$

$$P(S_t, K, \sigma, t) = -N(-d_1)S_t e^{-q\tau} + N(-d_2)Ke^{-r\tau}, P(F(t,T), K, \sigma, \tau) = -N(-d_1)F(t,T)e^{-r\tau} + N(-d_2)$$

where

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r - a + \sigma^2/2)(T-t)] = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{F(t,T)}{K}) + \sigma^2/2(T-t)] \\ d_2 &= d_1 - \sigma\sqrt{T-t} \\ N(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ F(t,T) &= S_t \exp((r-a)(T-t)) \end{aligned}$$

It follows that

- $$P(S_t, K, \sigma, \tau) = C(-S_t, -K, -\sigma, \tau)$$
- $$C(F(t,T), K, \sigma, \tau) = P(K, F(t,T), \sigma, T)$$

Proof. (1) Note that when we replace S_t, K, σ by $-S_t, -K, -\sigma$, d_1 will change sign. (2) Note that when we replace the order of $F(t,T)$ and K, d_1 will change to $-d_2$. □

18.2.1.3 *Option value after initialization*

Lemma 18.2.6. *Suppose the underlying S_t has real-world dynamics given by*

$$dS_t/S_t = \mu dt + \sigma dW_t.$$

Consider a call option with strike K and expiry T.

- *the value $C(t)$ of a call option in the real world will be*

$$C(t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)}.$$

- *the value $C(t)$ of a call option will have real-world dynamics given by*

$$dC_t = (rC_t + \frac{\partial C_t}{\partial S_t}\lambda\sigma S_t)dt + \frac{\partial C_t}{\partial S_t}\sigma dW_t$$

where λ is the market price of the stock risk given by

$$\lambda = \frac{\mu - r}{\sigma}.$$

Proof. (1) Straight forward from Lemma 18.2.1.(2) Because $C(t) = C(t, S(t))$, Ito lemma (Lemma 6.3.1)gives

$$\begin{aligned}
dC &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S_t}dS_t + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}dS_t dS_t \\
&= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S_t}(\mu S_t dt + \sigma S_t dW_t) + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\sigma^2 S_t^2 dt \\
&= (\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S_t} r S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\sigma^2 S_t^2)dt + \frac{\partial C}{\partial S_t}(\mu - r)S_t dt + \frac{\partial C}{\partial S_t}\sigma S_t dW_t \\
&= (rC_t + \frac{\partial C}{\partial S_t}(\mu - r)S_t)dt + \frac{\partial C}{\partial S_t}\sigma S_t dW_t \\
&= (rC_t + \frac{\partial C_t}{\partial S_t}\lambda\sigma S_t)dt + \frac{\partial C_t}{\partial S_t}\sigma dW_t
\end{aligned}$$

where we used the the Black-Scholes equation

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} + rS_t\frac{\partial C}{\partial S_t} - rC = 0.$$

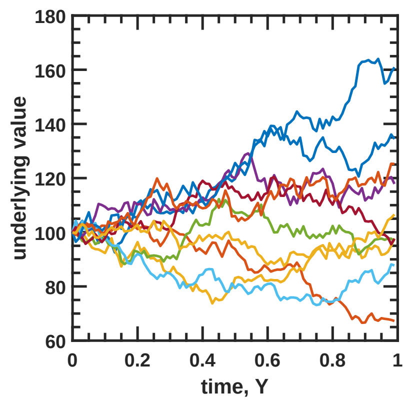
□

Note 18.2.1 (simulating European option value). To generate the trajectories of the forward contract value evolution, we can simulate the underlying, S_t, and then use the relation between S_t and $V(t)$ in Theorem 15.7.3.

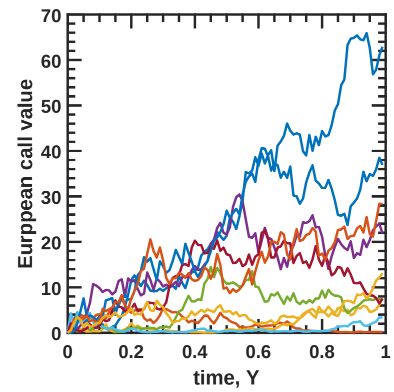
- As showed in Figure 18.2.2, we simulate the underlying in the real world via model
$$dS_t/S_t = \mu dt + \sigma dW_t,$$
and then evaluate the European call contract via
$$V(t) = N(d_1)S_t e^{-q(T-t)} - N(d_2)Ke^{-r(T-t)}.$$
- Note that **we should not simulate the underlying in the risk-neutral measure.** This is because in the risk neutral measure $V(t)$ will have expected growth rate of r. If an option contract has risk-free rate expected return with risks, then no one wants to enter due to risk aversion.
- On the other hand, if we simulate the underlying in the real world measure where S_t will have a different expected growth rate from r. Then the market

participants will enter either short or long positions based on their estimation of the S_t dynamics.

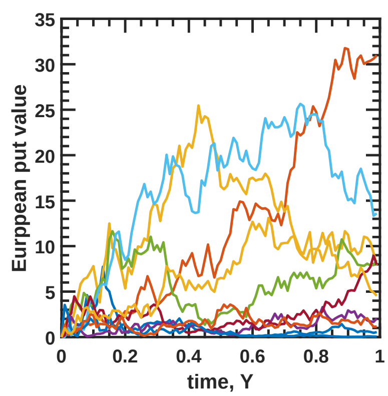
- Also see [7, p. 145] for a discussion.



(a) Underlying price trajectories.



(b) European call value trajectories.



(c) European put value trajectories.

Figure 18.2.2: Demonstration of European option value evolution. Underlying simulation parameters: $\mu = 0.06, r = 0.02, \sigma = 0.2, q = 0.0, t_0 = 0, S(t_0) = 100, K = 100, T = 1Y$.

18.2.1.4 *Profit and loss bounds*

Table 18.2.1: Maxium gain and loss for call and put options

(a) Call option profit and loss bounds

	maximum loss	maximum gain
buyer(long)	premium	unlimited
seller(short)	unlimited	premium

(b) Put option profit and loss bounds

	maximum loss	maximum gain
buyer(long)	premium	strike-premium
seller(short)	strike-premium	premium

18.2.2 Greeks for European call and put

18.2.2.1 *Basics*

Definition 18.2.1 (Greeks).

- ***Delta***:
$$\Delta = \frac{\partial V}{\partial S}$$
- ***Gamma***:
$$\Gamma = \frac{\partial^2 V}{\partial S^2}$$
- ***Theta***:
$$\Theta = \frac{\partial V}{\partial t}$$
- ***Vega***:
$$\nu = \frac{\partial V}{\partial \sigma}$$
- ***Rho***:
$$\rho = \frac{\partial V}{\partial r}$$
- ***Volga***:
$$Volga = \frac{\partial^2 V}{\partial \sigma^2}$$

Lemma 18.2.7 (preparation results). *[8] Define*

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r+\sigma^2/2)\tau], d_2 = d_1 - \sigma\sqrt{\tau}, \tau = T - t.$$

We have

- $$S_t \exp(-q\tau) n(d_1) = K\exp(-r\tau)n(d_2),$$
  $$F(t,T)n(d_1) = Kn(d_2).$$

  where forward rate $F(t,T) = S(t,T)\exp(-(r-q)(T-t)$.
- $$\frac{\partial d_2}{\partial S_t} = \frac{\partial d_1}{\partial S_t}$$
  $$\frac{\partial d_2}{\partial K} = \frac{\partial d_1}{\partial K}$$
  $$\frac{\partial d_2}{\partial \tau} = \frac{\partial d_1}{\partial \tau} - \frac{1}{2}\frac{\sigma}{\sqrt{\tau}}$$
  $$\frac{\partial d_2}{\partial \sigma} = \frac{\partial d_1}{\partial \sigma} - \sqrt{\tau}$$
  $$\frac{\partial d_2}{\partial r} = \frac{\partial d_1}{\partial r}$$

Proof. Note that

$$\begin{aligned}
&\frac{\partial}{\partial S_t} N(d_1) \\
&= N'(d_1)\frac{\partial}{\partial S_t} d_1 \\
&= \frac{1}{\sqrt{2\pi}}\exp(-d_1^2/2)\frac{1}{S_t\sigma\sqrt{\tau}} \\
&= \frac{1}{\sqrt{2\pi\tau}S_t\sigma}\exp(-d_1^2/2)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial S_t}N(d_2)\\
&= N'(d_2)\frac{\partial}{\partial S_t}d_2\\
&= \frac{1}{\sqrt{2\pi}}\exp(-d_2^2/2)\frac{1}{S_t\sigma\sqrt{\tau}}\\
&= \frac{1}{\sqrt{2\pi\tau}S_t\sigma}\exp(-(d_1-\sigma\sqrt{\tau})^2/2)\\
&= \frac{1}{\sqrt{2\pi\tau}S_t\sigma}\exp(-d_1^2/2)\exp(-\sigma^2\tau/2)\exp(\sigma\sqrt{\tau}d_1)\\
&= \frac{1}{\sqrt{2\pi\tau}S_t\sigma}\exp(-d_1^2/2)\frac{S_t}{K}\exp(r\tau)
\end{aligned}$$

□

18.2.2.2 *Delta*

Lemma 18.2.8 (Delta). *Let current time be t. Consider a call/put option with strike K and expiry T. Let the short rate be r.*

- *The Delta for a call option in the Black-Scholes pricing is given as*

$$\Delta_C = \frac{\partial C_t}{\partial S_t} = N(d_1), d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r+\sigma^2/2)\tau], \tau = T-t.$$

  And the Delta for a put option is given as

$$\Delta_P = \frac{\partial P_t}{\partial S_t} = \frac{\partial C_t}{\partial S_t} - 1 = N(d_1) - 1 = -N(-d_1).$$

- *If the underlying asset is paying continuous dividends with rate q, then*

$$\Delta_C = \exp(-q(T-t))N(d_1), \Delta_P = \exp(-q(T-t))(1-N(d_1)) = -\exp(-q(T-t))N(-d_1),$$

  where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r-q+\sigma^2/2)\tau], \tau = T-t.$$

- *The Delta with respect to forward rate $F(t,T) = S(t,T)\exp(-(r-q)(T-t)$ is given by*

$$\Delta_{C,F} = \frac{\partial C_t}{\partial F(t,T)} = \exp(-r\tau)N(d_1), \Delta_{P,F} = \frac{\partial P_t}{\partial F(t,T)} = -\exp(-r\tau)N(-d_1).$$

- *The Delta with respect to strike K is given by*

$$\Delta_{C,K} = \frac{\partial C_t}{\partial K} = \exp(-r\tau)N(d_1), \Delta_{P,K} = \frac{\partial P_t}{\partial F(t,T)} = -\exp(-r\tau)N(-d_1).$$

Proof. (1)(a) direct valuation. Note that

$$\frac{\partial C_t}{\partial S_t} = N(d_1) + S_t \frac{\partial}{\partial S_t} N(d_1) - K\exp(-r\tau)\frac{\partial}{\partial S_t} N(d_2).$$

Note that

$$\begin{aligned} &\frac{\partial}{\partial S_t} N(d_1) \\ &= N'(d_1)\frac{\partial}{\partial S_t} d_1 \\ &= \frac{1}{\sqrt{2\pi}} \exp(-d_1^2/2)\frac{1}{S_t\sigma\sqrt{\tau}} \\ &= \frac{1}{\sqrt{2\pi\tau}S_t\sigma} \exp(-d_1^2/2) \end{aligned}$$

and

$$\begin{aligned} &\frac{\partial}{\partial S_t} N(d_2) \\ &= N'(d_2)\frac{\partial}{\partial S_t} d_2 \\ &= \frac{1}{\sqrt{2\pi}} \exp(-d_2^2/2)\frac{1}{S_t\sigma\sqrt{\tau}} \\ &= \frac{1}{\sqrt{2\pi\tau}S_t\sigma} \exp(-(d_1 - \sigma\sqrt{\tau})^2/2) \\ &= \frac{1}{\sqrt{2\pi\tau}S_t\sigma} \exp(-d_1^2/2)\exp(-\sigma^2\tau/2)\exp(\sigma\sqrt{\tau}d_1) \\ &= \frac{1}{\sqrt{2\pi\tau}S_t\sigma} \exp(-d_1^2/2)\frac{S_t}{K}\exp(r\tau) \end{aligned}$$

Therefore,

$$S_t\frac{\partial}{\partial S_t} N(d_1) - K\exp(-r\tau)\frac{\partial}{\partial S_t} N(d_2) = 0, \frac{\partial C_t}{\partial S_t} = N(d_1).$$

(1)(b) For the put: From put-call parity(Lemma 15.2.1), we have $P_t + S_t = C_t + Ke^{-r(T-t)}$. Take first differential on both sides, we get

$$\frac{\partial P_t}{\partial S_t} + 1 = \frac{\partial C_t}{\partial S_t}.$$

(2) Similar to (1). (3) Use the chain rule that

$$\frac{\partial C_t}{\partial F(t,T)} = \frac{\partial C_t}{\partial F(t,T)} \frac{\partial F(t,T)}{\partial S_t}.$$

(4)(a)

$$\frac{\partial C}{\partial K} = S_t \exp(-q\tau) n(d_1) \frac{\partial d_1}{\partial K} - \exp(-r\tau) N(d_2) - K \exp(-r\tau) n(d_2) \frac{\partial d_2}{\partial K} = -\exp(-r\tau) N(d_2.$$

where we use

$$S_t \exp(-q\tau) n(d_1) \frac{\partial d_1}{\partial K} = K \exp(-r\tau) n(d_2) \frac{\partial d_2}{\partial K}$$

in Lemma 18.2.7. (b) Use put-call parity such that

$$\frac{\partial C}{\partial K} - \frac{\partial P}{\partial K} = -\exp(-r\tau).$$

□

Remark 18.2.4 (different ATM convention). There are different interpretations when we refer to ATM strike.

- ATM-spot: $K = S_0$.
- ATM-forward: $K = F_0$.
- ATM-value-neutral: K such that call value = put value.
- ATM-Delta-neutral: K such that call delta = -put delta. In this case,

$$\Delta_C = -\Delta_P = 0.5 \implies d_1 = 0, K = S_0 \exp((r + \frac{\sigma^2}{2})T).$$

Remark 18.2.5 (interpretation).

- The Delta for a call is always positive, which mean if S_t increases, C_t will increase.
- The Delta for a put is always negative, which mean if S_t increases, P_t will decrease.

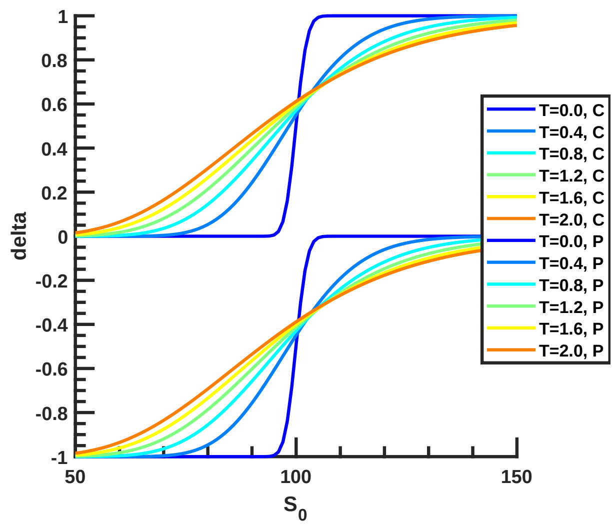


Figure 18.2.3: Delta profile of a call option(upper) and a put option(lower) at different spot prices S_0 and different time to maturity T. The strike of the option is $K = 100$.

18.2.2.3 *Gamma*

Lemma 18.2.9 (Gamma and convexity). *Let current time be t. Consider a call/put option with strike K and expiry T. Let the short rate be r and the dividend rate be q. In the Black-Scholes framework, the Gamma of the call/put option is given as*

$$\Gamma = \frac{\partial^2 C}{\partial S_t^2} = \frac{\partial^2 C}{\partial S_t^2} = \exp(-q(T-t))\frac{N'(d_1)}{S_t \sigma \sqrt{T-t}} = K\exp(-r(T-t))\frac{N'(d_2)}{S_t^2 \sigma \sqrt{T-t}},$$

where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r - q + \sigma^2/2)\tau], \tau = T - t.$$

Moreover, we have

- *C is a convex function of the spot price S_t; that is, $\frac{\partial^2 C}{\partial S^2} \geq 0$.*
- *The Γ for a put option is same as that of the call option due to put-call parity..*

Proof. (1)Since

$$\frac{\partial C}{\partial S} = N(d_1), d_1 = \frac{1}{\sigma\sqrt{T-t}}[\ln(\frac{S_t}{K}) + (r + \sigma^2/2)(T-t)],$$

we have

$$\frac{\partial^2 C}{\partial S^2} = \frac{N'(d_1)}{S\sigma\sqrt{T-t}}.$$

It is easy to see that $\Gamma \geq 0$, therefore, C_t is convex. c (2)For the put: From put-call parity(Lemma 15.2.1), we have $P_t + S_t = C_t + Ke^{-r(T-t)}$. Take twice differential on both sides, we get

$$\frac{\partial^2 P_t}{\partial S_t^2} = \frac{\partial^2 C_t}{\partial S_t^2}.$$

□

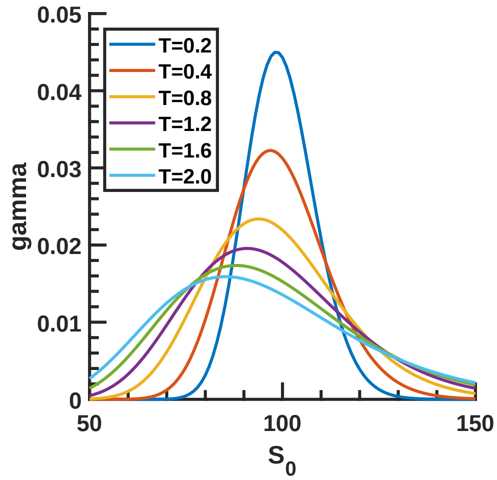


Figure 18.2.4: Gamma profile of a call option(or a put option) at different spot prices S_0 and different time to maturity T. The strike of the option is $K = 100$.

18.2.2.4 *Theta*

Lemma 18.2.10 (Theta). *[6, p. 409]Let current time be t. Consider a call/put option with strike K and expiry T. Let the short rate be r and the dividend rate be q. In the Black-Scholes framework, the Theta of the call/put is given by,*

$$\Theta_C = -\frac{S_t N'(d_1)\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}N(d_2) + qS_t \exp(-q(T-t))N(d_1)$$
$$\Theta_P = -\frac{S_t N'(d_1)\sigma}{2\sqrt{T-t}} + rKe^{-r(T-t)}N(d_2) - qS_t \exp(-q(T-t))N(-d_1)$$

where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r - q + \sigma^2/2)\tau], \tau = T - t$$

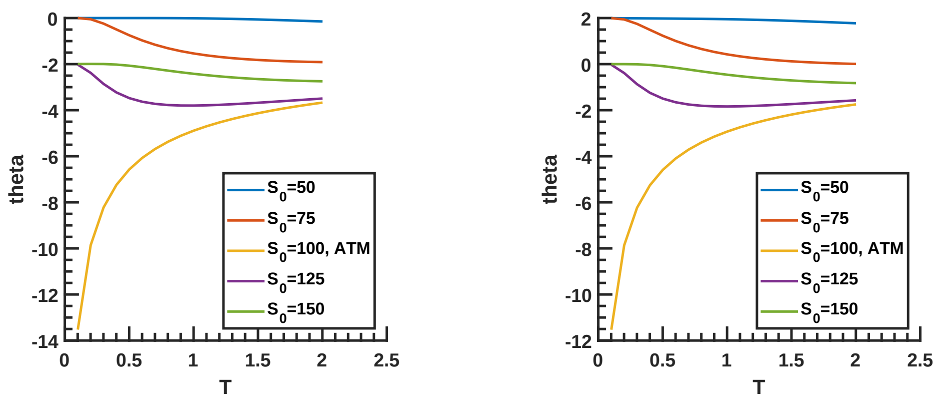
Note that

$$\Theta_C - \Theta_P = -\frac{d}{dt}(Ke^{-r(T-t)}) = -qS_t \exp(-q(T-t)) - rKe^{-r(T-t)}$$

due to put-call parity.

Remark 18.2.6 (example theta and interpretation). As showed in Figure 18.2.5, we examine the Theta for call and put options on a non-dividend-paying underlying. We have the following observation.

- ATM call and put have large theta values, which indicate the dramatic decreasing the of the time value for ATM options approaching expiry.
- Keeping other factors remain fixed, call options always decrease value when approaching expiry.
- For ATM and out-of-money put options, their value will decrease when approaching expiry; For deep-in-the-money put options, put options might increase value when approaching expiry.



(a) Underlying price trajectories. (b) European call value trajectories.

Figure 18.2.5: Theta profile of a call option(a) and a put option(b) at different spot prices S_0 and different time to maturity T. The strike of the option is $K = 100$. The dividend rate $q = 0$.

Remark 18.2.7 (interpretation). Given a Geometric Brownian motion with constant drift μ and volatility σ. Its mean and variance is given by(Corollary 6.3.3.1)

- $E[S(t)] = S(0)e^{rt}$
- $Var[S(t)] = S(0)^2 e^{2rt}(e^{\sigma^2 t} - 1)$

- If $r \geq 0$, then an European call will always decrease in value; Intuitively, geometric Brownian motion has an upward drift , the larger volatility, the more likely it is going to arrive at the higher asset value.
- If $r \geq 0$, then an European put might not lose value. The result depends on the magnitude of r. Intuitively, geometric Brownian motion will spread the underlying and increase the value of put(due to $Gamma > 0$); a large positive drift will increase the likelihood of out-of- money situation.

Corollary 18.2.0.1 (never early exercise an American call in the Black-Scholes world). *Let $C(T;K)$ denote the price (the price calculated in Black-Scholes world)of a European call with the same strike K but a continuum of maturity. Since $\frac{dC}{dT} = -\Theta_C > 0$, calls with more distant exercising date will have more value; that is, never early exercising an American call.*

Note 18.2.2 (what the real world will be). In the real world, the asset dynamics need to have a state-dependent volatility $\sigma(S_t.t)$ in order to match the market prices (see Theorem 17.3.1).

$$\frac{\partial C}{\partial T} = \frac{\partial C_{BS}}{\partial \sigma^*}\frac{\partial \sigma^*}{\partial T}.$$

Since the term structure of implied volatility can be anything, the time-value of call option in real world is uncertain.

18.2.2.5 *Vega and Volga*

Lemma 18.2.11 (Vega). *Let current time be t. Consider a call/put option with strike K and expiry T. Let the short rate be r and the dividend rate be q. In the Black-Scholes framework, the Vega of the call/put is given by,*

$$\nu = \frac{\partial C_t}{\partial \sigma} = \frac{\partial P_t}{\partial \sigma} = S_t e^{-q(T-t)}\sqrt{T-t}N'(d_1) = Ke^{-r(T-t)}\sqrt{T-t}N'(d_2),$$

where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r - q + \sigma^2/2)\tau], \tau = T - t.$$

Remark 18.2.8 (interpretation).

- The Vega is always positive, indicating that the option prices will increase as volatility increases.
- The Vega of all options decrease as approaching expiration since a long-term option is more sensitive to change in the volatility.

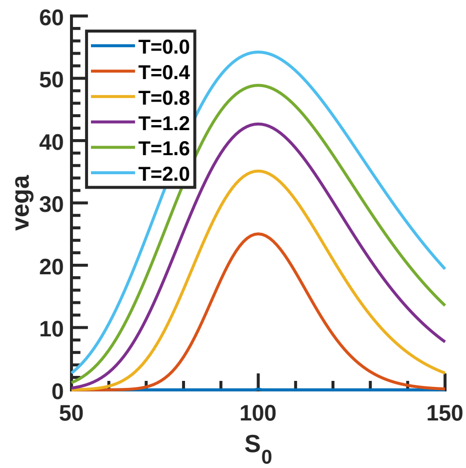


Figure 18.2.6: Vega profile of a call option(or a put option) at different spot prices S_0 and different time to maturity T. The strike of the option is $K = 100$.

Lemma 18.2.12 (Volga). *For European call and put options,*

$$Volga = \frac{\partial^2 C_t}{\partial \sigma^2} = \frac{\partial^2 P_t}{\partial \sigma^2} = \nu \frac{d_1 d_2}{\sigma}$$

Proof. direct calculation. □

Remark 18.2.9 (convexity on volatility)**.**

- For most out-of-money call options, both d_1 and d_2 are negative; for most in-the-money call options, both d_1 and d_2 are positive. So $d_1 d_2 > 0$ in most cases and C_t is a convex function on σ.
- When close at-the-money, we might have $d_1 d_2 < 0$.

18.2.2.6 *Rho*

Lemma 18.2.13 (Rho). *[6, p. 417] Let current time be t. Consider a call/put option with strike K and expiry T. Let the short rate be r and the dividend rate be q. In the Black-Scholes framework, the Rho of the call/put is given by,*

$$Rho_C = \frac{\partial V}{\partial r} = K(T-t)e^{-r(T-t)}N(d_2) > 0, Rho_P = \frac{\partial V}{\partial r} = -K(T-t)e^{-r(T-t)}N(-d_2) < 0.$$

where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}[\ln(\frac{S_t}{K}) + (r - q + \sigma^2/2)\tau], \tau = T - t.$$

Note that

$$Rho_C - Rho_P = K(T-t)\exp(-r(T-t))N(d_2).$$

18.2.2.7 *Additional relationship*

Lemma 18.2.14 (relationship between Delta, Theta, and Gamma). *[6, p. 414] Consider a portfolio $\pi = C - \Delta S$, where $\Delta = \frac{\partial C}{\partial S}$. We have*

- $$\begin{aligned} d\pi &= \frac{\partial \pi}{\partial t}dt + (\frac{\partial C}{\partial S} - \Delta)dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \\ &= \frac{\partial \pi}{\partial t}dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \\ &= r\pi dt; \end{aligned}$$

  that is, the value of the portfolio will grow at a rate r.
- $\frac{\partial \pi}{\partial S} = 0$; *that is, the portfolio π is the delta-neutral.*
- $\frac{\partial^2 \pi}{\partial S^2} = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}$; *that is, the portfolio π is not Gamma-neutral.*
- *For a delta-neutral portfolio π,*

  $$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma = r\pi,$$

  where

  $$\Theta = \frac{\partial \pi}{\partial t}, \Gamma = \frac{\partial^2 \pi}{\partial S^2}$$

Proof.

$$\begin{aligned} d\pi \triangleq d(C - \frac{\partial C}{\partial S_t}S_t) &= (\frac{\partial C}{\partial t}dt + \frac{\partial V}{\partial S_t}dS_t + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}dt) + \frac{\partial C}{\partial S_t}dS_t \\ &= (\frac{\partial C}{\partial t}dt + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}dt) \\ &= rV - rS_t \frac{\partial C}{\partial S_t}S_t = r\pi \end{aligned}$$

where in the last step we use Black-Scholes equation

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0.$$

□

18.2.2.8 *Greeks with implied volatility*

Lemma 18.2.15 (Delta and Gamma with implied volatility). *Let $V(S_t, K, \sigma(K, S_t))$ be the value at time t of an European option of strike K and maturity T on the spot S_t, where $\sigma(K, F_t)$ denotes the implied volatility that make $V(S_t, K, \sigma(K, S_t))$ equal the market price. The Delta and Gamma calculation are given by*

- $$\Delta = \frac{\partial}{\partial S_t} V(S_t, K, \sigma(K, S_t)) = \frac{\partial V}{\partial S_t} + \frac{\partial V}{\partial \sigma}\frac{\sigma}{\partial S_t}.$$
- $$\Gamma = \frac{\partial^2}{\partial S_t^2} V(S_t, K, \sigma(K, S_t)) = \frac{\partial^2 V}{\partial S_t^2} + \frac{\partial^2 V}{\partial S_t \partial \sigma}\frac{\partial \sigma}{\partial S_t} + \frac{\partial^2 V}{\partial \sigma^2}\frac{\partial^2 \sigma}{\partial S_t^2} + \frac{\partial V}{\partial \sigma}\frac{\partial^2 \sigma}{\partial S_t^2}.$$

Proof. Use calculus and chain rule. □

18.2.3 Log contract

Definition 18.2.2 (log contract). *[9, p. 70] A log contract L is a derivative that, at expiration T, pays the value $L_T = \ln(S_T/K)$, where K is the strike.*

Lemma 18.2.16 (value of a log contract). *Assume Black-Scholes world assumptions hold. Then current value of a log contract with strike K is given by*

$$V_0 = \ln(\frac{S_0}{K}) + rT - \frac{1}{2}\sigma^2 T.$$

Proof. In the Black-Scholes world, S_T under risk-neutral measure is given by

$$\ln S_T = \ln S_0 + rT - \frac{1}{2}\sigma^2 T + \sigma W_T.$$

Take expectation, we have

$$E_Q[\ln S_T] = \ln S_0 + rT - \frac{1}{2}\sigma^2 T.$$

Finally,

$$V_0 = E_Q[\frac{\ln S_T}{K}] = \ln S_0/K + rT - \frac{1}{2}\sigma^2 T$$

□

Lemma 18.2.17 (replicating log contract using vanilla options).

- *The final payoff of a log contract,* $L_T = \ln(S_T/K)$*, can be decomposed as*

$$\ln(S_T/K) = \frac{S_T - K}{K} - \int_K^{\infty} \frac{1}{v^2}(S_T - v)^+ dv - \int_0^K \frac{1}{v^2}(v - S_T)^+ dv.$$

- *The replicating strategy is:*
  - *long* $1/K$ *unit of forward with strike price* K*.*
  - *short* $1/v$ *unit of put options at strike* v *where* v *ranges from* 0 *to* K*.*
  - *short* $1/v$ *unit of call options at strike* v *where* v *ranges from* K *to* ∞*.*

Proof. Use the result in Theorem 16.2.2, we have for any twice-continuously differentiable function $f(x)$, we have

$$f(x) = f(\kappa) + f'(\kappa)(x-\kappa) + \int_0^{\kappa} f''(K)(K-x)^+ dK + \int_{\kappa}^{\infty} f''(K)(x-K)^+ dK.$$

Take $x = S_T, \kappa = K$, we will get the result. □

18.2.4 Options on futures and forwards

Lemma 18.2.18 (forward price dynamics at constant interest rate). *[10, p. 101] Assume constant interest rate. The forward price for an asset at* T *determined at* t *is given as*

$$F(t) = S_t \exp(r(T-t)).$$

Assuming S_t*, under real-world probability measure, follows a geometric Brownian motion with drift* μ *and volatility* σ*, then we have*

$$dF_t = (\mu - r)F_t dt + \sigma F_t dW_t.$$

Particularly, under the risk-neutral measure Q *such that* S_t *has a drift* $rS_t dt$*, we have*

$$dF_t = \sigma F_t d\hat{W}_t,$$

where $\hat{W}$ *is the Brownian motion under the measure* Q*.*

Proof. Note that

$$dF_t = -rSe^{r(T-t)}dt + e^{r(T-t)}dS,$$

then we have

$$dF_t = (\mu - r)S\exp(r(T-t))dt + \sigma S\exp(r(T-t))dW_t.$$

Under the measure Q, $dW_t = d\hat{W}_t - (\mu - r)dt$. □

Lemma 18.2.19 (Black-Scholes equation for options on forwards/futures). *[10, p. 101] Let $V(F(t), t)$ be the value of the derivative as a function of the forward/futures price $F(t)$ and time t. Assume $F(t)$ is governed by*

$$dF_t = (\mu - r)F_t dt + \sigma F_t dW_t$$

where W_t is the Brownian motion. Then V is governed by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 F^2 \frac{\partial^2 V}{\partial F^2} - rV = 0$$

with final condition $V(F(T), T) = V_T(F(T))$ and boundary condition $V(S, t) = V_a(t)$ on $S = a$ and $V(S, t) = V_b(t)$ on $S = b$.

Proof. (1) (PDE method) Consider the portfolio $\Pi = V - \Delta F$, (we are hedging by entering forward contracts at different time)then

$$dV = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 F^2)dt + \frac{\partial V}{\partial F}dF.$$

Let $\Delta = \frac{\partial V}{\partial F}$, then

$$d\Pi = dV - \Delta dF = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 F^2)dt = rVdt.$$

Note that the right-hand side is $rVdt$ instead of $r\Pi dt$, since it takes zero money to enter the contract. (2) (martingale method) Note that $D(t)V(F(t), t)$is a martingale under the risk-neutral measure Q. Using martingale pricing theorem(Theorem 15.6.13), we have

$$D(t)V(F(t), t) = E_Q[D(T)V(F(T), T)|\mathcal{F}_t].$$

Because $F(t)$ is following SDE under Q given by

$$dF_t = \sigma F_t d\hat{W}_t,$$

we can derive the associated PDE using Feyman-Kac theorem(Theorem 15.8.1). □

Lemma 18.2.20 (Solutions to Black-Scholes equation for futures and forwards). *[10, p. 101] The call option with strike K and matures at T will have price:*

$$C(F, t) = e^{r(T-t)}(FN(d_1) - EN(d_2)).$$

Proof. Since the Black-Scholes equation is similar to the case of asset paying dividends, we can use results from Theorem 15.6.13. □

18.2.5 Options of multiple assets

Lemma 18.2.21 (options on multiple assets). *[1, p. 183] The option value $V(S_1,...,S_d,t)$ on d assets with Geometric Brownian motion is given as*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i\sigma_j\rho_{ij}S_iS_j\frac{\partial^2 V}{\partial S_i \partial S_j} + \sum_{i=1}^{d}(r-D_i)S_i\frac{\partial V}{\partial S_i} - rV = 0$$

under the risk-neutral measure Q, the asset dynamics become

$$dS_i(t) = (r(t)-D_i)S_i(t)dt + \sigma(t)S_i(t)dW_i(t), i = 1,...,d.$$

We have

$$V(t) = E_Q[V(S_1(T),...,S_d(T)|\mathcal{F}_t].$$

Proof. Let

$$\Pi = V(S_1,...,S_d,t) - \sum_{i=1}^{d}\Delta_i S_i,$$

then

$$d\Pi = \frac{\partial V}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i\sigma_j\rho_{ij}S_iS_j\frac{\partial^2 V}{\partial S_i \partial S_j}dt + \sum_{i=1}^{d}(\frac{\partial V}{\partial S_i} - \Delta_i)dS_i.$$

If we choose $\Delta_i = \frac{\partial V}{\partial S_i}, \forall i$, then the portfolio is risk-free. Then $d\Pi = r\Pi dt$. We have

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i\sigma_j\rho_{ij}S_iS_j\frac{\partial^2 V}{\partial S_i \partial S_j} + \sum_{i=1}^{d}(r-D_i)S_i\frac{\partial V}{\partial S_i} - rV = 0.$$

Then we can use Theorem 15.8.6 to show the rest. □

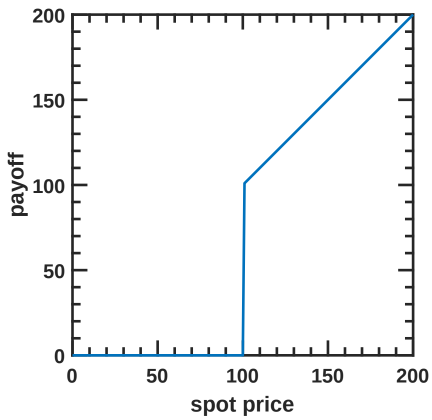
18.3 Exotic Option pricing

18.3.1 Digital option

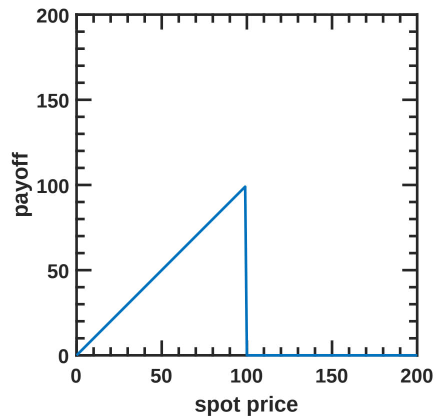
18.3.1.1 *Valuation*

Definition 18.3.1 (cash or nothing digital call option). *A cash-or-nothing digital call or put option on the underlying asset S_t at maturity T with strike K has payoff $1_{S_T>K}$ or $1(S_T - K)$.*

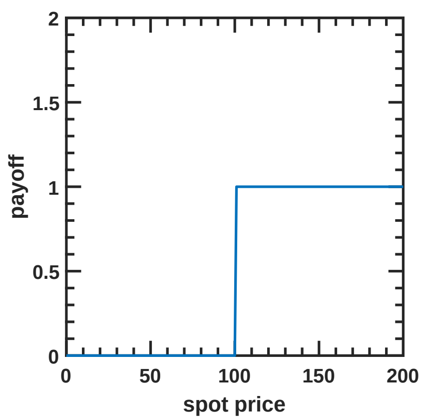
Definition 18.3.2 (asset or nothing digital call option). *An asset-or-nothing digital call or put option on the underlying asset S_t at maturity T with strike K has payoff $S_T 1_{S_T>K}$ or $S_T 1_{S_T-K}$.*



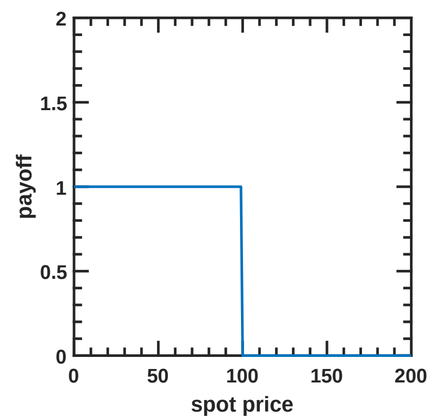
(a) Asset-or-nothing digital call option payoff at maturity. The strike $K = 100$.



(b) Asset-or-nothing digital put option payoff at maturity. The strike $K = 100$.



(c) Cash-or-nothing digital call option payoff at maturity. The strike $K = 100$.



(d) Cash-or-nothing digital put option payoff at maturity. The strike $K = 100$.

Figure 18.3.1: Payoff function for digital options.

Lemma 18.3.1 (put-call parity for digital options). *Assume current time is t and the risk-free rate is constant r. It follows that*

- *The cash-or-nothing digital call and put with the same maturity T and strike K satisfy*

$$C_t + P_t = \exp(-r(T-t)),$$

  where C_t, P_t denote the current values for the call and the put.
- *The asset-or-nothing digital call and put with the same maturity T and strike K satisfy*

$$C_t + P_t = \exp(-D(T-t))S_t,$$

where C_t, P_t, S_t denote the current values for the call, the put and the asset, D is the dividend rate.

Proof. (1) Note that $C_T + P_T$ has payoff of 1 at all possible situations. Therefore

$$C_t + P_t = \exp(-r(T-t))E_Q[C_T + P_T|\mathcal{F}_t] = \exp(-r(T-t)).$$

(2)Note that $C_T + P_T$ has payoff of S_T at all possible situations. Therefore

$$C_t + P_t = \exp(-r(T-t))E_Q[C_T + P_T|\mathcal{F}_t] = \exp(-r(T-t))E_Q[S_T|\mathcal{F}_t] = \exp(-D(T-t))S_t.$$

□

Lemma 18.3.2 (cash or nothing digital call/put pricing). *Assume that the underlying asset S_t is following*

$$dS_t/S_t = \mu dt + \sigma dW_t.$$

The risk-neutral price of a digital call option is

$$V(t) = e^{-r(T-t)}P_Q(S_T > K) = e^{-r(T-t)}N(d),$$

where

$$d = (\log(S_t/K) + (r - \frac{\sigma^2}{2})(T-t))/\sigma\sqrt{T-t}.$$

or equivalently,

$$d = (\log(F_t/K) - \frac{\sigma^2}{2}(T-t))/\sigma\sqrt{T-t}$$

where F_t is the forward price.

Similarly, the risk-neutral price of a digital put option is

$$V(t) = e^{-r(T-t)}P_Q(S_T < K) = e^{-r(T-t)}N(-d).$$

Proof. (1) Under the risk-neutral measure, the S_T is given by

$$S_T = S_t \exp((r - \frac{\sigma^2}{2})(T-t) + \sigma(W_T - W_t)).$$

Then

$$\begin{aligned}
P(S_T > K) &= P(S_t \exp((r - \frac{\sigma^2}{2})(T-t) + \sigma(W_T - W_t)) > K) \\
&= P(S_t \exp((r - \frac{\sigma^2}{2})(T-t) + \sigma\sqrt{T-t}Z) > K) \\
&= P(Z > -(\log(S/K) + (r - \frac{\sigma^2}{2})(T-t))/\sigma\sqrt{T-t}) \\
&= P(Z > -d) = N(d)
\end{aligned}$$

where

$$d = (\log(S_t/K) + (r - \frac{\sigma^2}{2})(T-t))/\sigma\sqrt{T-t}.$$

(2)

$$\begin{aligned}P(S_T < K) &= P(S_t \exp((r - \frac{\sigma^2}{2})(T-t) + \sigma(W_T - W_t)) < K) \\ &= P(S_t \exp((r - \frac{\sigma^2}{2})(T-t) + \sigma\sqrt{T-t}Z) < K) \\ &= P(Z < -(\log(S/K) + (r - \frac{\sigma^2}{2})(T-t))/\sigma\sqrt{T-t}) \\ &= P(Z < -d) = N(-d)\end{aligned}$$

□

Lemma 18.3.3 (asset or nothing digital call/put pricing). *Assume that the underlying asset S_t under risk neutral measure is following*

$$dS_t/S_t = (r-a)dt + \sigma dW_t,$$

where r is the short rate and a is the dividend rate. The risk-neutral price of a digital call option is

$$V_C(t) = e^{-r(T-t)}E_Q[S_T 1_{(S_T>K)}] = e^{-a(T-t)}S(t)N(d),$$

where

$$d = (\log(S_t/K) + (r - a - \frac{\sigma^2}{2})(T-t))/\sigma\sqrt{T-t}.$$

or equivalently,

$$d = (\log(F_t/K) + \frac{\sigma^2}{2}(T-t))/\sigma\sqrt{T-t}$$

where F_t is the forward price.

Similarly, the risk-neutral price of a digital put option(using put-call parity Lemma 18.3.1) is

$$V_P(t) = (\exp(-a(T-t))S_t - V_C(t)) = e^{-a(T-t)}S(t)N(-d).$$

Proof. (1) Under the risk-neutral measure, the S_T is given by

$$S_T = S_t \exp((r - a - \frac{\sigma^2}{2})(T-t) + \sigma(W_T - W_t)).$$

Let $Y = (W(T) - W(t))/\sqrt{T-t}$, then $Y \sim N(0,1)$, and we can write

$$S(T) = S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)\tau + \sigma\tau Y)),$$

where $\tau = T - t$.

Let

$$d_2 = \frac{\ln \frac{S(t)}{K} + (r - a - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$

such that

$$S(t)\exp((r - a - \frac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}(-d_2)) = K.$$

Then the for the standard normal Y, We have

$$Pr(S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)) > K) = Pr(Y > -d_2) = Pr(Y < d_2).$$

Then

$$\begin{aligned}
V(t) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}(S(t)\exp(+\sigma\tau y + (r - a - \frac{1}{2}\sigma^2)\tau))e^{-\frac{1}{2}y^2}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\sigma\tau y - \frac{1}{2}\sigma^2\tau)e^{-\frac{1}{2}y^2}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}(y - \sigma\sqrt{\tau})^2)dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-d_2-\sigma\sqrt{\tau}}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}z^2)dz \\
&= S(t)\exp(-\tau a)N(d)
\end{aligned}$$

where we used the fact that $-d = -d_2 - \sigma\sqrt{\tau}$. (2)

$$V_P(t) = (\exp(-a(T-t))S_t - V_C(t)) = e^{-a(T-t)}S(t)(1 - N(d)) = e^{-a(T-t)}S(t)N(-d).$$

□

We use martingale method.

$$C(S_T, T) = E_Q[e^{-r(T-t)}\max(S_T - K, 0)|\mathcal{F}_t].$$

where

$$S(T) = S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma(W(T) - W(t))).$$

Let $Y = (W(T) - W(t))/\sqrt{T-t}$, then $Y \sim N(0,1)$, and we can write

$$S(T) = S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)).$$

Let
$$d_2 = \frac{\ln \frac{S(t)}{K} + (r - a - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$
such that
$$S(t)\exp((r - a - \frac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}(-d_2)) = K.$$
Then the for the standard normal Y, We have
$$Pr(S(t)\exp((r - a - \frac{1}{2}\sigma(s)^2)(T-t) + \sigma\tau Y)) > K) = Pr(Y > -d_2) = Pr(Y < d_2).$$

$$\begin{aligned}
C(S_T, T) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}(S(t)\exp(+\sigma\tau y + (r - a - \frac{1}{2}\sigma^2)\tau) - K)e^{-\frac{1}{2}y^2}dy \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\sigma\tau y - \frac{1}{2}\sigma^2\tau)e^{-\frac{1}{2}y^2}dy] - [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} e^{-r\tau}e^{-\frac{1}{2}y^2}dy] \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}(y - \sigma\sqrt{\tau})^2)dy] - e^{-r\tau}KN(d_2) \\
&= [\frac{1}{\sqrt{2\pi}}\int_{-d_2-\sigma\sqrt{\tau}}^{\infty} S(t)\exp(-\tau a)\exp(-\frac{1}{2}z^2)dz] - e^{-r\tau}KN(d_2) \\
&= S(t)\exp(-\tau a)N(d_1) - e^{-r\tau}KN(d_2)
\end{aligned}$$
where we used the fact that $-d_1 = -d_2 - \sigma\sqrt{\tau}$.

Lemma 18.3.4 (hedging digital call option, connection to European call). *[11, p. 2]* *Let $D(K,T)$ be the price of a digital call option. Let $C(K,T)$ be the price of an European call option. We have*

- $$D = -\frac{\partial C}{\partial K}.$$
- *To (approximately) hedge a short position of D, we can long a call at strike K and short a call at strike $K+\epsilon$ in quantity $1/\epsilon$.*

*Note that **these results are model independent**.*

Proof. We can use the result at Lemma 16.2.1. If we use a Black-Scholes model(which satisfies the no-arbitrage condition), then
$$C(S_t, t, T, K) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)},$$
and we can show
$$-\frac{PaC}{\partial K} = N(d_2)e^{-r(T-t)} = D(K,T).$$
□

Remark 18.3.1 (theoretical importance of digital option). Digital options have important theoretical consequence in pricing exotic option and volatility model. See section 18.5.

18.3.1.2 *Risk analysis*

Lemma 18.3.5 (greeks for cash-or-nothing digital options). *Let the current time be t. Consider a cash-or-nothing digital call/put option with strike K and expiry T.*

- *(Delta)*
$$\Delta = \frac{\partial C_t}{\partial S_t} = \frac{\exp(-r(T-t))}{\sigma S_t \sqrt{T-t}} N'(d_2).$$
- *(Gamma)*
$$\Delta = \frac{\partial^2 C_t}{\partial S_t^2} = -\frac{\exp(-r(T-t))d_1}{\sigma S_t (T-t)} N'(d_2).$$
- *(Theta)*
$$\Delta = \frac{\partial C_t}{\partial \tau} = r\exp(-r\tau)N(d_2) + \exp(-r\tau)N'(d_2)(\frac{d_1}{2\tau} - \frac{r}{\sigma\sqrt{\tau}}).$$
- *(Vega)*
$$\Delta = \frac{\partial C_t}{\partial \sigma} = -\frac{\exp(-r(T-t))d_1}{\sigma} N'(d_2).$$

where
$$d_1 = \frac{\ln(S_t/K) + (r - \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, d_2 = d_1 + \sigma\sqrt{T-t}.$$

Proof. (1) Note that $C_t = \exp(-r(T-t))N(d_2)$. Then
$$\frac{\partial C_t}{\partial S_t} = \exp(-r(T-t))N'(d_2)\frac{\partial d_2}{\partial S_t} = \exp(-r(T-t))N'(d_2)\frac{1}{S_t\sigma\sqrt{T-t}}.$$
□

18.3.2 Asian option

18.3.2.1 *Canonical approach*

Definition 18.3.3 (Asian option). *An Asian call option with strike K and maturity T has payoff given by*
$$V(T) = \max(\frac{1}{T}\int_0^T S(u)du - K, 0),$$

where $S(t)$ is the stochastic process for the underlying asset.

Remark 18.3.2 (business needs). An Asian option can be used when managing the risk exposure(e.g., stock price) for a period of time. For example, a company needs to buy a stock in the next 12 months, with each month a fixed quantity. Then the company has the exposure to the stock price in the next 12 months.

Lemma 18.3.6 (price equation). *[4, pp. 279, 322] The Asian call price $v(t,x,y)$ satisfies*

$$v_t(t,x,y) + rxv_x(t,x,y) + xv_y(t,x,y) + \frac{1}{2}\sigma^2x^2v_{xx}(t,x,y) = rv(t,x,y), 0 \leq t \leq T, x \geq 0, y \in \mathbb{R},$$

with boundary conditions

$$\begin{aligned} v(t,0,y) &= e^{-r(T-t)}(\max(\frac{y}{T} - K, 0)) \\ \lim_{y\to-\infty} v(t,x,y) &= 0, 0 \leq t \leq T, x \geq 0 \\ v(T,x,y) &= \max(\frac{y}{T} - K, 0), x \geq 0, y \in \mathbb{R}. \end{aligned}$$

Proof. Define $Y(t) = \int_0^T S(u)du$ such that $dY(t) = Y(t)dt$. Under risk-neutral measure Q,

$$e^{-rt}V(t) = E_Q[e^{-rT}\max(\frac{1}{T}Y(T) - K, 0)|\mathcal{F}_t]$$

is a martingale.

$$\begin{aligned} d(e^{-rt}v(t,x,y)) &= e^{-rt}(-rvdt + v_tdt + v_xdX + v_ydY + \frac{1}{2}v_{xx}dXdX + v_{xy}dXdY + \frac{1}{2}v_{yy}dYdY) \\ &= e^{-rt}(-rvdt + v_tdt + v_xrxdt + v_yxdt + \frac{1}{2}v_{xx}\sigma^2x^2dt) + e^{-rt}v_x\sigma xdW(t) \end{aligned}$$

set the drift to zero, we get the PDE. □

Remark 18.3.3 (interpretation). Under risk-neutral measure Q, the dynamics of $(Y(t), S(t))$ is given by

$$\begin{aligned} dS(t) &= rS(t)dt + \sigma S(t)dW(t) \\ dY(t) &= S(t)dt. \end{aligned}$$

In terms of quadratic variation properties, we have

$$dYdY = 0, dYdS = 0$$

Remark 18.3.4 (how to hedge).

18.3.2.2 *put-call parity*

Lemma 18.3.7 (put-call parity for Asian options). *Consider an Asian call with expiration date t_N and payoff given by*

$$\max(A(t_N) - K, 0),$$

where

$$A(t_m) = \frac{1}{m+1}\sum_{i=0}^{m} S(t_i).$$

Similarly, consider an Asian put with payoff given by

$$\max(K - A(t_N), 0).$$

It follows that

- *at maturity t_N, we have*

$$C(t_N) - P(t_N) = A(t_N) - K.$$

- *at any earlier date $t < t_0$, we have*

$$C(t) - P(t) = \frac{1}{N+1}\sum_{i=0}^{N} \exp(-r(t_N - t_i))S(t) - K\exp(-r(t_N - t)).$$

Proof. (1) straight forward. (2)

$$C(t_N) - P(t_N) = \frac{1}{N+1}\sum_{i=0}^{N} S(t_i) - K$$

$$E[\exp(-r(t_N - t))C(t_N)|\mathcal{F}_t] - E[\exp(-r(t_N - t))P(t_N)|\mathcal{F}_t]] = \frac{1}{N+1}\sum_{i=0}^{N} E[\exp(-r(t_N - t))S(t_i)|\mathcal{F}_t] -$$

$$C(t) - P(t) = \frac{1}{N+1}\sum_{i=0}^{N} E[\exp(-r(t_i - t - t_i + t_N))S(t$$

$$C(t) - P(t) = \frac{1}{N+1}\sum_{i=0}^{N} \exp(-r(t_N - t_i))S(t) - K\exp$$

□

18.3.2.3 *Approximation via geometric average*

Lemma 18.3.8. *Suppose*

$$S_t = S_0 \exp(vt + \sigma W_t), v = r - \frac{1}{2}\sigma^2.$$

Then the geometric average

$$G_m = (S_0 \times S_1 \times \cdots \times S_m)^{1/(m+1)},$$

where $S_i = S(ih), h = T/m$, has log-normal distribution given by

$$G_m = S_0 \exp(v\frac{T}{2} + \sigma\sqrt{\frac{2m+1}{6(m+1)}}W_T).$$

If $m \to \infty$, we have

$$G_\infty = S_0 \exp(v\frac{T}{2} + \sigma\sqrt{\frac{1}{3}}W_T).$$

Proof. (1) The coefficient for the drift is given by

$$\frac{1}{m+1}(h + 2h + \cdots + mh) = \frac{mh}{2} = \frac{T}{2}.$$

(2) Note that

$$\begin{aligned}
&(W_0 + W_1 + W_2 + \cdots + W_m) \\
=&m(W_1 - W_0) + (m-1)(W_2 - W_1) + (m-2)(W_3 - W_2) + \cdots + (W_m - W_{m-1}) \\
=&\sqrt{m^2 + (m-1)^2 + (m-2)^2 + \cdots + 1^2)}Z, Z \sim (0, T/m) \\
=&\sqrt{\frac{m(m+1)(2m+1)}{6}}Z \\
=&\sqrt{\frac{m(m+1)(2m+1)}{6}}\frac{\sqrt{W_T}}{\sqrt{m}}
\end{aligned}$$

□

Remark 18.3.5 (alternative derivation). To derive the coefficient for Brownian motion term, we can consider the fact(Theorem 6.2.5) of

$$\frac{1}{T}\int_0^T W_u du = \frac{1}{\sqrt{3}}W_T.$$

18.3.2.4 *Moment matching method*

18.3.3 Basket option

18.3.3.1 *Basics*

Definition 18.3.4 (basket option). *A basket call option with strike K, maturity T, and n underlying $S_t^{(1)}, S_t^{(2)}, ..., , S_t^{(n)}$ has payoff*

$$V_T = (\sum_{i=1}^{n} w_i S_T^{(i)} - K)^+,$$

where $w_i > 0, w_i \in \mathbb{R}$ are the weights.

Note 18.3.1 (business needs: the price of basket option vs. a portfolio of vanilla options)**.** Consider a basket option with payoff at maturity T given by

$$V_T^B = (\sum_{i=1}^{n} \frac{1}{n} S_T^{(i)} - K)^+;$$

and a portfolio of call options with payoff at maturity T given by

$$V_T^P = \sum_{i=1}^{n} \frac{1}{n} (S_T^{(i)} - K)^+.$$

Also we assume $S_0^{(i)}, i = 1, 2, ..., n$ are close to K.

Then in general, the basket's option value will be much lower than the value of vanilla option portfolio due to the following arguments:

- If each asset $S^{(i)}$ are independent from each other, then we can view $S^{(1)}, ..., S^{(n)}$ as iid random sample of S. Then the term $\sum_{i=1}^{n} \frac{1}{n} S_T^{(i)}$ will have a $1/\sqrt{n}$ variance of S.
- If each asset $S^{(i)}$ are perfectly correlated with each other, then we can view $S^{(1)}, ..., S^{(n)}$ as exact copy of S. Then the term $\sum_{i=1}^{n} \frac{1}{n} S_T^{(i)}$ is the same as S. In this case, the basket option will have similar price to the portfolio of options.

Therefore, a company usually prefers basket option when managing the risk of a set of assets due to its lower cost.

18.3.3.2 *Moment matching method for pricing*

Lemma 18.3.9 (black formula with extended lognormal distribution family). *[12]*

- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$
  where
  $$d_{1,2} = \frac{\log M_1(T) - \log K}{\sigma} \pm \frac{1}{2}\sigma$$
  and
  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$
- $$c = \exp(-rT)[(M_1(T) - \tau)N(d_1) - (K - \tau)N(d_2)],$$
  where
  $$d_{1,2} = \frac{\log(M_1(T) - \tau) - \log(K - \tau)}{\sigma} \pm \frac{1}{2}\sigma$$
  and
  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$
- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$
  where
  $$d_{1,2} = \frac{\log -M_1(T) - \log -K}{\sigma} \pm \frac{1}{2}\sigma$$
  and
  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$
- $$c = \exp(-rT)[M_1(T)N(d_1) - KN(d_2)],$$
  where
  $$d_{1,2} = \frac{\log M_1(T) - \log K}{\sigma} \pm \frac{1}{2}\sigma$$
  and
  $$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

Lemma 18.3.10 (log-normal approximate pricing for vanilla call and put). *Assume under risk-neutral measure Q that the underlying dynamics are given by*

$$dS_t^{(i)}/S_t^{(i)} = (r-q)dt + \sigma_i dW_i(t), i = 1, 2, ..., n,$$

and $dW_i dW_j = \rho_{ij} dt$. *Denote* $B(T) = \sum_{i=1}^n w_i S_T^{(i)}$. *It follows that*

- *The random variable* $B(T)$ *can be approximated by* Y *by matching first two moments defined by*

  $$Y = M_1 \exp(-\frac{1}{2}\sigma^2 + \sigma Z), Z \in N(0,1),$$

  where

  $$M_1 = \sum_{i=1}^n w_i F_i, M_2 = \sum_{i=1}^n \sum_{j=1}^n F_i F_j w_i w_j \exp(\rho_{ij}\sigma_i\sigma_j T), F_i = S^{(i)}(0)\exp((r-q)T).$$

- $$C = \exp(-(r-q)T)(M_1(T)N(d_1) - KN(d_2)).$$

- $$P = \exp(-(r-q)T)(-M_1(T)N(-d_1) + KN(-d_2)).$$

where

$$d_{1,2} = \frac{\log M_1(T) - \log K}{\sigma} \pm \frac{1}{2}\sigma$$

and

$$\sigma = \sqrt{\log(\frac{M_2(T)}{M_1(T)^2})}.$$

Proof. (1) The moments of $B(T)$ is discussed in Lemma 17.2.3. The log-normal approximation is discussed in Lemma 2.2.17. □

18.3.3.3 *Copula method for pricing*

Lemma 18.3.11 (copula method for pricing basket option). *Consider a basket option with payoff at maturity T given by*

$$V_T^B = (\sum_{i=1}^n \frac{1}{n} S_T^{(i)} - K)^+;$$

and a portfolio of call options with payoff at maturity T given by

$$V_T^P = \sum_{i=1}^{n} \frac{1}{n}(S_T^{(i)} - K)^+.$$

We make the following assumptions:

- *We can construct the implied cdf $F_i$for each $S_T^{(i)}, i = 1, 2, ..., n$.(Lemma 16.2.5).*
- *The joint distribution of $S_T^{(1)}, S_T^{(2)}, ..., S_T^{(n)}$ has Gaussian copula with correlation matrix Σ.*[a]

*Then we can use the following simulation method(**??**) to generate one pricing sample*

- *First generate $(Y_1, Y_2, ..., Y_n) \sim MN(0, \Sigma)$.*
- *Return $S_T^{(1)} = F_1^{-1}(\phi(Y_1)), S_T^{(2)} = F_2^{-1}(\phi(Y_2)), ..., S_T^{(n)} = F_n^{-1}(\phi(Y_n))$,, where ϕ is standard normal cdf.*

a the correlation matrix can be estimated from historical data.

18.3.4 Exchange(Margrabe) option

Definition 18.3.5 (exchange option). *[6, p. 611] Let U_t and V_t be the price processes of two assets. An exchange option with maturity T has payoff at T given by*

$$\max(V_T - U_T, 0).$$

Lemma 18.3.12 (price of exchange option). *[6, p. 612] Consider an exchange option on two assets S_1 and S_2 with maturity T. Assume S_1 and S_2 are following Geometric Brownian motion in the real-world given by*

$$dS_1(t) = (r - q_1 + \lambda_1\sigma_1)S_1(t)dt + \sigma_1 S_1(t)dW_1(t)$$
$$dS_2(t) = (r - q_2 + \rho\lambda_1\sigma_2 + \lambda_2\sqrt{1-\rho^2}\sigma_2)S_2(t)dt + \rho\sigma_2 S_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t)$$

where W_1 and W_2 are independent Brownian motions.

Then the value of the exchange option at time 0 is given by

$$V_0 = S_1(0)\exp(-q_1 T)N(d_1) - S_2(0)\exp(-q_2 T)N(d_2)$$

where

$$d_1 = \frac{\ln(S_1(0)/S_2(0) + (q_2 - q_1 + \hat{\sigma}^2/2)}{\hat{\sigma}\sqrt{T}}, d_2 = d_1 - \hat{\sigma}\sqrt{T},$$

and

$$\hat{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Proof. Two stocks,$S_1(t)$ and $S_2(t)$, follows the SDE in real-world measure.

$$dS_1(t) = (r + \lambda_1\sigma_1)S_1(t)dt + \sigma_1 S_1(t)dW_1(t)$$
$$dS_2(t) = (r + \rho\lambda_1\sigma_2 + \lambda_2\sqrt{1-\rho^2}\sigma_2)S_2(t)dt + \rho\sigma_2 S_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t)$$

where W_1 and W_2 are independent Brownian motions.

A derivative D_t pays $\max(S_2(T) - S_1(T), 0)$ at time T. Develop a formula for its current price at time t.

Solution:

Use the stock S_1 as the numeraire, then under this new measure Q_S,

$$\frac{D(T)}{S_1(T)}$$

is a martingale. Therefore,

$$\frac{D(t)}{S_1(t)} = E_{S_1}[\frac{D(T)}{S(T)}|\mathcal{F}_t] = E_{S_1}[\max(\frac{S_2(T)}{S_1(T)} - 1)|\mathcal{F}_t].$$

Note that under measure Q_S (we take $\lambda_1 = \sigma_1, \lambda_2 = 0$ following Theorem 15.6.19), the dynamics of $S_1(t)$ and $S_2(t)$ follows

$$\begin{aligned} dS_1(t) &= (r + \sigma_1^2)S_1(t)dt + \sigma_1 S_1(t)dW_1(t) \\ dS_2(t) &= (r + \rho\sigma_1\sigma_2)S_2(t)dt + \ \rho\sigma_2 S_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t) \\ d\frac{S_2}{S_1} &= \frac{S_2}{S_1}((\rho\sigma_2 - \sigma_1)dW_1 + \sqrt{1-\rho^2}\sigma_2\sigma dW_2) \end{aligned}$$

Denote $Y = \frac{S_2}{S_1}$, then Y is a geometric Brownian motion with volatility $\sigma_Y = \sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$.

Then we have $D(t) = S_1(t)E_{S_1}[\max(Y(T) - 1, 0)]$, which can be evaluated. $\square$

Remark 18.3.6 (effects of correlation). Note that negative correlation will increase the value of an exchange option. Consider the following example in Figure 18.3.2, where $S_1(0) = 120, \sigma_1 = 0.3, S_2(0) = 100, \sigma_2 = 0.4, r = 1\%$ and correlation ranges from -1 to 1.

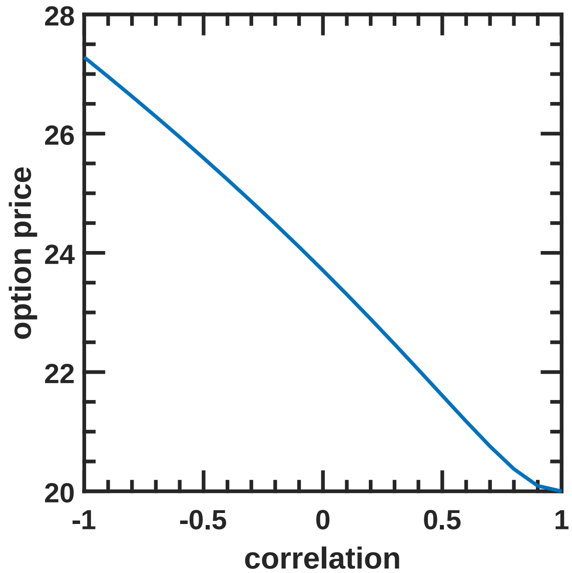


Figure 18.3.2: Correlation effect on exchange option price.

18.3.5 Cliquet option

Definition 18.3.6 (Cliquet option). *Let $T_0 < T_1 < ... < T_n$ be a set of dates. A Cliquet option seller with unit notional pays the buyer at the maturity T_s*

$$\max(\sum_{i=1}^{n} \max(0, \min(Cap, \frac{S_i - S_{i-1}}{S_{i-1}})), Floor),$$

where S_i is the underlying price at T_i,Cap is the local cap, and $Floor$ is the global floor.

Lemma 18.3.13.

18.3.6 Chooser option

Definition 18.3.7.

Lemma 18.3.14 (price). *[6, p. 604] Let current time T_0 be zero. A chooser option has right at T_1 to choose a pair of call and put options with strike K and maturity $T_2, T_2 > T_1$. Let*

r denote the risk-free interest rate and q the continuous dividend rate. It follows that the choose option has payoff at T_1 given by

$$\max(c, p),$$

where c and p denote the price at T_1 of a call and a put with strike K and maturity T_2. The payoff can be decomposed as

- *A call option with strike price K and maturity T_2;*
- $\exp(-q(T_2 - T_1))$ *unit of put option with strike price* $K\exp(-(r-q)(T_2 - T_1))$ *and maturity T_1.*

Proof. Let c and p denote the price at T_1 of a call and a put with strike K and maturity T_2. The payoff at T_1 is given by

$$\begin{aligned}
\max(c, p) &= \max(c, c + K\exp(-r(T_2 - T_1)) - S_1\exp(-q(T_2 - T_1)) \\
&= c + \exp(-q(T_2 - T_1))\max(0, K\exp(-(r-q)(T_2 - T_1)) - S_1)
\end{aligned}$$

where use the put-call parity Lemma 18.2.2 given by

$$c - p = S_1\exp(-q(T_2 - T_1)) - K\exp(-r(T_2 - T_1)).$$

□

Lemma 18.3.15 (price of general chooser option). *[6, p. 612] For general chooser options, we can have the following decompositions*

$$\begin{aligned}
\min(U_T, V_T) &= V_T - \max(V_T - U_T, 0) \\
\max(U_T, V_T) &= U_T + \max(V_T - U_T, 0).
\end{aligned}$$

That is, we can decompose it into an asset and an exchange option(Lemma 18.3.12).

Proof. (1)

$$\begin{aligned}
\min(U_T, V_T) &= \min(U_T - V_T, V_T - V_T) + V_T \\
&= \min(U_T - V_T, V_T - V_T) + V_T \\
&= V_T - \max(V_T - U_T, 0)
\end{aligned}$$

(2)

$$\max(U_T, V_T) = \max(U_T - U_T, V_T - U_T) + U_T = U_T + \max(V_T - U_T, 0)$$

□

18.3.7 Barrier option

Theorem 18.3.1 (PDE pricing). *[13, p. 205]*

$$\frac{\partial C(S,t)}{\partial t} + rS\frac{\partial C(S,t)}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C(S,t)}{\partial S^2} = rC$$

with final condition $C(S,T) = f(S)$, with boundary condition $C(B,t) = 0$.

Theorem 18.3.2 (risk-neutral pricing). *[13, p. 216]*

$$\frac{\partial C(S,t)}{\partial t} + rS\frac{\partial C(S,t)}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C(S,t)}{\partial S^2} = rC$$

with final condition $C(S,T) = f(S)$, with boundary condition $C(B,t) = 0$.

18.3.8 Early exercise options

18.3.8.1 *Foundations*

Theorem 18.3.3 (the pricing principle of early exercising options).

- *Consider a Bermudian put option with strike K and a set of exercising dates $t_0, t_1, ..., t_M$, with t_M be the final maturity date.*
  Then its current value is
- *Consider an American put option with strike K and final maturity date T.*
  Then its current value is

Lemma 18.3.16 (the recursive structure for early exercising put option). *Consider a put option with strike K and a set of exercising dates $t_0, t_1, ..., t_M$, with t_M be the final maturity date. Let r be the interest rate, and $\tau_i = \tau_{i+1} - \tau_i$. Let $V(S_{t_i}, t_i)$ denote the future value of the put option at time t_i. We have*

- *At final maturity date,*
$$V(S_{t_M}, t_M) = \max(K - S_{t_M}, 0).$$
- *At previous exercising dates $t_i < t_M$,*
$$V(S_{t_M}, t_M) = \max(\max(K - S_{t_i}, 0), \exp(-r\tau_i)E_Q[V(S_{t_{i+1}}, t_{i+1})|S_{t_i}].$$

- *The current value of the put option is given by*

$$P = V(S_{t_0}, t_0),$$

where S_{t_0} is the spot value observed in the market.

Remark 18.3.7 (exercising value and continuing value)**.**

18.3.8.2 *Perpetual American put*

Definition 18.3.8 (perpetual American put).

Lemma 18.3.17.

$$V(x; L)$$

Remark 18.3.8 (time independence of the exercising policy)**.** Note that from optimal control point of view, our exercising policy is time-independent due to the infinite horizon of the problem. For finite horizon American put, the exercising policy will be time-dependent.

18.3.8.3 *American call: non-dividend paying underlying*

Theorem 18.3.4 (submartingale property for non-dividend-paying underlying). *[4, p. 361] Let $h(x)$ be a non-negative, convex function of $x \geq 0$ satisfying $h(0) = 0$. It follows that*

- *The discounted intrinsic value $h(S(t))/M(t)$ of the American derivative securities(i.e., having early exercising feature) that pays $h(S(t))$ upon exercise under risk-neutral measure is a submartingale, i.e.,*

$$h(S(u))/M(u) \leq E_Q[h(S(t))/M(t)|\mathcal{F}_u], 0 \leq u \leq t.$$

- *The price of the American derivative security expiring at time T and having intrinsic value $h(S(t)), 0 \leq t \leq T$, is the same as the price of the European derivative security paying $h(S(T))$ at expiration T; that is*

$$V^{Amer.}(0) = \max_{0 \leq u \leq T} E_Q[\frac{h(S_u)}{M(u)}] = E_Q[\frac{h(S_T)}{M(T)}] = V^{Eur.}(0).$$

- *For an American derivative security expiring at time T with payoff $(S_T - K)^+$ on the non-dividend paying asset S_t, its price is the same the European option.*

Proof. (1)First note that because $h(x)$ is convex, we have for $0 \leq \lambda \leq 1$ and $0 \leq x_1 \leq x_2$, we have

$$h((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)h(x_1) + \lambda h(x_2).$$

Take $x_1 = 0$, we have

$$h(\lambda x) \leq \lambda h(x), \forall x \geq 0, 0 \leq \lambda \leq 1.(*)$$

For $0 \leq u \leq t$, we have $0 \leq 1/M(t) \leq 1$, the equation $(*)$ implies that

$$E_Q[h(S(t))/M(t)|\mathcal{F}_u] \geq E_Q[h(S(t)/M(t))|\mathcal{F}_u].$$

Using Jensen's inequality, we further have

$$E_Q[h(S(t)/M(t))|\mathcal{F}_u] \geq h(E_Q[S(t)/M(t)|\mathcal{F}_u]) = h(S(u)/M(u)),$$

where we use the martingale property $E_Q[S(t)/M(t)|\mathcal{F}_u] = S(u)/M(u)$. Eventually, we have

$$E_Q[h(S(t))/M(t)|\mathcal{F}_u] \geq h(S(u)/M(u)).$$

(2) Use the fact from (1) such that

$$E_Q[\frac{h(S_u)}{M(u)}] \leq E_Q[E_Q[\frac{h(S_T)}{M(T)}|\mathcal{F}_u]] = E_Q[\frac{h(S_T)}{M(T)}].$$

Therefore,

$$V^{Amer.}(0) = \max_{0 \leq u \leq T} E_Q[\frac{h(S_u)}{M(u)}] = E_Q[\frac{h(S_T)}{M(T)}] = V^{Eur.}(0)$$

(3) Directly from (2). □

Remark 18.3.9 (interpretation on model-free interpretation).

- Note that our lemma only replies on the martingale property under risk-neutral measure $E_Q[S(t)/M(t)|\mathcal{F}_u] = S(u)/M(u)$, which will hold in no-arbitrage market.
- If we assume the underlying S_t has the following risk-neutral dynamics given by $dS_t/S_t = r(t)dt + \sigma(t)dW_t$, then even if $\sigma(t)$ is state dependent, we still have the submartingale property.
- For an American derivative security expiring at time T with payoff $(S_T - K)^+$ on the non-dividend paying asset S_t, there are three factors contributing to the submartingale property of the discounted process $\frac{(S(t)-K)^+}{M(t)}$:
  - $\frac{S(t)}{M(t)}$ is a martingale, thus submartingale.
  - $-\frac{K}{M(t)}$ is a submartingale because of the negative sign and the increasing nature of $M(t)$.
  - The above two make $\frac{(S(t)-K)}{M(t)}$ a submartingale, and the fact $(S(t)-K) < (S(t)-K)^+$ make $\frac{(S(t)-K)^+}{M(t)}$ a submartingale.

Remark 18.3.10 (compare with dividend paying underlying and American put).

- For an American derivative security expiring at time T with payoff $(S_T - K)^+$ on the non-dividend paying asset S_t, there are three factors contributing to the submartingale property of the discounted process $\frac{(S(t)-K)^+}{M(t)}$:
  - $\frac{S(t)}{M(t)}$ is a martingale, thus submartingale.
  - $-\frac{K}{M(t)}$ is a submartingale because of the negative sign and the increasing nature of $M(t)$.
  - The above two make $\frac{(S(t)-K)}{M(t)}$ a submartingale, and the fact $(S(t) - K) < (S(t) - K)^+$ make $\frac{(S(t)-K)^+}{M(t)}$ a submartingale.
- For a dividend-paying underlying asset S_t, $\frac{S(t)}{M(t)}$ is a supermartingale and $-\frac{K}{M(t)}$ is a submartingale; therefore, $\frac{(S(t)-K)}{M(t)}$ is not necessarily a submartingale.
- For an American put expiring at time T with payoff $(K - S_T)^+$, $\frac{K}{M(t)}$ is a supermartingale and $\frac{S(t)}{M(t)}$ is a martingale; therefore, $\frac{(K-S(t))}{M(t)}$ is not a submartingale.

18.3.8.4 *Approximate least square approach*

Lemma 18.3.18 (approximate least square for early exercise put option). *Consider a put option with strike K and a set of exercising dates $t_0, t_1, ..., t_M$, with t_M be the final maturity date. Let r be the interest rate, and $\tau_i = \tau_{i+1} - \tau_i$.*

- *Define the following parameters and notations*
  - *number of sample paths N,*
  - *Q basis functions, denoted by $f_i(x), i = 1, 2, ..., Q$ for regression.*
  - *Denote $V(S_{t_i}, t_i)$ as the future value of the put option at time t_i.*
- *Use Monte Carlo to simulate N trajectories for asset value process $S(t)$. Use $S_{t_j}^{(i)}$ to denote the i trajectory value at time t_j.*
- *At time t_M, we evaluate the cash flow for each trajectory, denoted by*

  $$y_i = \max(K - S_{t_M}^{(i)}, 0),$$

  and approximate the future value function by

  $$\hat{V}(S_{t_M}, t_M) = \sum_{i=1}^{Q} \alpha_i(t_M) f_i(S_{t_M}),$$

where $\alpha_i(t_M)$ is the coefficients by regressing

$$(y_i, f_1(S_{t_M}^{(i)}), f_2(S_{t_M}^{(i)}), ..., f_Q(S_{t_M}^{(i)})), i = 1, 2, ..., N.$$

- *At earlier times $t_j < t_M$, we evaluate the cash flow for each trajectory i, denoted by*

$$y_i = \max(\max(K - S_{t_j}^{(i)}, 0), \exp(-r\tau_j)\hat{V}(S_{t_{j+1}}^{(i)}, t_{j+1}),$$

and approximate the future value function by

$$\hat{V}(S_{t_j}, t_j) = \sum_{i=1}^{Q} \alpha_i(t_j) f_i(S_{t_j}),$$

where $\alpha_i(t_j)$ is the coefficients by regressing

$$(y_i, f_1(S_{t_j}^{(i)}), f_2(S_{t_j}^{(i)}), ..., f_Q(S_{t_j}^{(i)})), i = 1, 2, ..., N.$$

- *The current value of the put option is given by*

$$P = \max(\max(K - S_{t_0}, 0), \frac{1}{N}\sum_{i=1}^{n} \exp(-r\tau_0) V(S_{t_1}, t_1)),$$

where S_{t_0} is the spot value observed in the market.

Remark 18.3.11 (interpretation).

- The choice of basis function can be polynomials given by $1, x, x^2, ...$
- When there are multiple underlyings, basis functions need to include forms like $x_1 x_2$ to account for the interactions between underlyings.

18.3.9 Compound option

Definition 18.3.9 (compound option). *Suppose the compound option has strike K and maturity date T while the underlying option has strike K^* and maturity date $T^* > T$.*

- *a call-on-call option gives holder the right buy a call option with strike K^* and maturity T^* at time T; it has payoff at T given by*

$$V_T(S_T) = \max\{0, C_{BS}(S_T, K^*, T^*) - K\}.$$

- *a put on call option has payoff at T*

$$V_T(S_T) = \max\{0, K - C_{BS}(S_T, K^*, T^*)\}.$$

- *a call on put option has payoff at T*

$$V_T(S_T) = \max\{0, P_{BS}(S_T, K^*, T^*) - K\}.$$

- *a put on put option has payoff at T*

$$V_T(S_T) = \max\{0, K - P_{BS}(S_T, K^*, T^*)\}.$$

where

$$\begin{aligned} C_{BS}(S_T, K^*, T) &= N(d_1)S_T - N(d_2)Ke^{-r(T^*-T)}, \\ P_{BS}(S_T, K^*, T) &= -N(-d_1)S_T + N(-d_2)Ke^{-r(T^*-T)}, \end{aligned}$$

where

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T^*-T}}[\ln(\frac{S_T}{K}) + (r+\sigma^2/2)(T^*-T)] \\ d_2 &= d_1 - \sigma\sqrt{T^*-T} \\ N(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \end{aligned}$$

Remark 18.3.12 (business application). Consider a company bids to complete a large project in one year. If they win the bid, they will need financing for $100 million for 3 years. They could buy a three-year interest rate cap beginning the date of the contract but this could be very expensive if they do not win the contract.

In this situation,the company could buy a compoun call option on a three-year interest cap, giving their right to buy the three-year cap at the time of bidding. If they win the contract, they then exercise the option for the interest rate cap at the predetermined premium. And if they do not win the contract, they can let the option expire. The advantage is a lower initial outlay and reduced risk.

Lemma 18.3.19 (compound option put-call parity).

- $$C_{CallonCall}(S_0, T_1, K_1; T_2, K_2) - P_{PutonCall}(S_0, T_1, K_1; T_2, K_2) = C(S_0; T_2, K_2) - K_1 \exp(-rT_1)$$

- $$C_{CallonPut}(S_0, T_1, K_1; T_2, K_2) - P_{PutonPut}(S_0, T_1, K_1; T_2, K_2) = P(X; T_2, K_2) - K_1 \exp(-rT_1)$$

Proof. Straightforward. □

Lemma 18.3.20. *Let $f(x)$ be the PDF of $N(\mu, \sigma^2)$. Then*

$$\int_a^\infty f(x)N(bx+c)dx = \int_a^\infty f(x)\int_{-\infty}^{bx+c} f(y)dydx = N_2(\frac{\mu - a}{\sigma}, \frac{b\mu + c}{\sqrt{1+b^2\sigma^2}}; \rho),$$

where $N_2(u, v; \rho)$ is the joint cmf for bivariate standard normal variables (U, V) with correlation

$$\rho = b\sigma/\sqrt{1+b^2\sigma^2}.$$

Proof. In the original formulation $X = simN(\mu, \sigma^2), Y \sim N(0,1)$, X and Y are independent. Introduce

$$U = -X, V = -bX + Y,$$

the above integral becomes

$$Pr(X \geq a, Y \leq bX + c) = Pr(U \leq -a, V \leq c),$$

Further note that we can express U, V via

$$U = \sigma Z_1 - \mu, V = b\sigma Z_2 - b\mu, (Z_1, Z_2) \sim MN(0, I; \rho), \rho = \frac{b\sigma}{\sqrt{b^2\sigma^2 + 1}}.$$

Then

$$Pr(U \leq -a, V \leq c) = Pr(Z_1 \leq \frac{\mu - a}{\sigma}, Z_2 \leq \frac{b\mu + c}{\sqrt{1+b^2\sigma^2}}).$$

□

Lemma 18.3.21. *Consider a call on call compound option has strike K and maturity date T while the underlying option has strike K^* and maturity date $T^* > T$. Then its price is given by*

$$V(t) = e^{-r(T-t)}E_Q[V_T(S_T)|\mathcal{F}_t] = e^{-r(T-t)}\int_{S^*}^\infty (C_{BS}(x, K, T) - K)f_{S_T}(x)dx,$$

where S^* is the break-even asset price such that $C_{BS}(S^*, K^*, T^*) = K$.

The above integral can be further reduce to

$$e^{-r(T-t)}\int_{S^*}^{\infty}(C_{BS}(x,K,T)-K)f_{S_T}(x)dx = e^{-r(T-t)}\int_{\ln(S^*/S_t)}^{\infty}(C_{BS}(S_te^x,K,T)-K)f(x)dx.$$

where $f(x)$ is the pdf of $N(\mu,\sigma_S^2)$.

The evaluation is given by

$$V(t) = S_tN_2(D_1,D_1^*;\rho) - K^*e^{-r(T^*-t)}N_2(D_2,D_2^*;\rho) - Ke^{-r(T-t)}N(D_2^*).$$

where

$$D_1^* = D_2^* = \frac{\mu-a}{\sigma}, D_1 = D_2 = \frac{b\mu+c}{\sqrt{1+b^2\sigma^2}},$$

and

$$\mu = (r+\sigma_S^2/2)(T-t), \sigma = \sigma_S\sqrt{T-t}, \rho = \sqrt{(T-t)/(T^*-t)},$$
$$a = \ln(S^*/S_t), b = 1/(\sigma_S\sqrt{T^*-T})$$

18.3.10 Variance swap

18.3.10.1 *Basics*

Definition 18.3.10 (variance swap). *[11, p. 61] The payoff of a long position on variance swap on an underlying S_t with strike K_{Var} and maturity T is given by*

$$V_T = N\times(\sigma_R^2 - K_{Var}^2),$$

where N is the notional amount, σ_R^2 is the annualized volatility of daily log-return between current $t=0$ and $t=T$ under zero-mean assumption:

$$\sigma_R^2 = \frac{252}{T-2}\sum_{i=0}^{T-1}\ln^2(S_{i+1}/S_i)^2.$$

Remark 18.3.13 (other variance products). [14, p. 119]

- **volatility swap**, where two parties exchange future realized volatility with a fixed value.
- **CBOE VIX futures**, which enables investors to bet the short-term implied volatility at a future date.
- **Options on realized volatility**, options on VIX index.

Lemma 18.3.22 (the fair value of a variance swap via risk-neutral pricing). *Let $\sigma^2(t)$ be a stochastic process modeling the realized variance via*

$$dS_t/S_t = rdt + \sigma(t)dW_t,$$

under risk-neutral measure Q. Further assume constant interest rate r. It follows that

- *The current value at $t = 0$ is given by*

$$V_0 = E_Q[\exp(-rT)(\sigma_R^2 - K_{Var})],$$

where

$$\sigma_R^2 = \frac{1}{T}\int_0^T \sigma^2(t)dt.$$

- *The fair strike that makes $V_0 = 0$ is given by*

$$K_{Var} = E_Q[\sigma_R^2] = \frac{1}{T}E_Q[\int_0^T \sigma^2(t)dt].$$

Proof. (1) Note that $\sigma_R^2 = \frac{1}{T}\int_0^T \sigma^2(t)dt$ is the continuous-time version of realized variance starting from $t = 0$ to T. (2)

$$V_0 = E_Q[\exp(-rT)(\sigma_R^2 - K_{Var})] = \exp(-rT)E_Q[(\sigma_R^2 - K_{Var})] = 0 \implies K_{Var} = \sigma_R^2.$$

□

18.3.10.2 *Pricing via replication*

Lemma 18.3.23 (continuous replicating realized variance). *Let $\sigma^2(t)$ be a stochastic process modeling the realized variance via*

$$dS_t/S_t = rdt + \sigma(t)dW_t,$$

under risk-neutral measure Q. Further assume constant interest rate r. It follows that

- *the fair strike that makes $V_0 = 0$ is given by*

$$K_{Var} = \frac{2}{T}(rT - E_Q[\ln\frac{S_T}{S_0}]).$$

- *the fair strike equals the value of the following portfolios:*
  - *cash value $2r$.*
  - *a short position of $\frac{2\exp(rT)}{T}\frac{1}{S_0}$ unit of forward with strike price S_0.*

- *long* $\frac{2\exp(rT)}{T}\frac{1}{v}$ *unit of put options at strike* v *where* v *ranges from* 0 *to* S_0.
- *long* $\frac{2\exp(rT)}{T}\frac{1}{v}$ *unit of call options at strike* v *where* v *ranges from* S_0 *to* ∞.

Proof. (1)Under the dynamic model for S_t, we have

$$\begin{aligned} d(\ln S_t) &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dZ_t \\ \implies \sigma^2 dt &= \frac{2dS_t}{S_t} - d(\ln S_t) \\ K_{Var} = \frac{1}{T}\int_0^T \sigma^2 dt &= \frac{2}{T}[\int_0^T \frac{dS_t}{S_t} - d(\ln S_t)] \\ &= \frac{2}{T}[\int_0^T \frac{dS_t}{S_t} - (\ln S_T - \ln S_0)]. \end{aligned}$$

Now note

$$E[\frac{dS_t}{S_t}] = E[rdt + \sigma(t)dW_t] = rdt \implies \int_0^T \frac{dS_t}{S_t} = rT.$$

Therefore,

$$K_{Var} = \frac{2}{T}(rT - E_Q[\ln \frac{S_T}{S_0}])$$

(2) the details on replicating the payoff $\ln(S_T/S_0)$ is at Lemma 18.2.17.

□

Note 18.3.2 (practical discrete replicating strategy for the log part).

18.3.10.3 *Pricing via Heston stochastic model*

Lemma 18.3.24. *Assume the underlying asset* S_t *and the variance* V_t*, under risk-neutral measure, is governed by Heston stochastic model*

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{V(t)}S(t)dW_1(t), \\ dV(t) &= k(\theta - V(t))dt + \eta\sqrt{V(t)}dW_2, \end{aligned}$$

where the interest rate r *is constant,* W_1 *and* W_2 *are correlated Brownian motions with* $dW_1 dW_2 = \rho dt$.

Let the current time be 0. Consider a variance swap with maturity T*. It follows that*

- *The expected variance is given by*

$$E[V(t)^2] = (V(0) - \theta)\exp(-kt) + \theta.$$

- *The fair strike at time 0 is given by*

$$K_{var} = \frac{1}{T} E_Q[\int_0^T \sigma^2(t)dt] = \frac{1}{kT}(1 - \exp(-kT))(V(0) - \theta) + \theta.$$

- *The value of a variance swap at time $t, 0 \leq t \leq T$, is given by*

$$P(t) = N\exp(-r(T-t))[E_{net}(V) - K_{var}],$$

where

$$E_{net}(V) = \frac{1}{T}(tV_R + \frac{1}{k}(1 - \exp(-k(T-t)(V(0) - \theta)) + (T-t)\theta),$$

and V_R is the realized variance up to time t.

Proof. (1)

$$\begin{aligned}
dV(s) &= k(\theta - V(s))ds + \eta\sqrt{V(s)}dW_s \\
\int_0^t dV(s) &= \int_0^t k(\theta - V(s))ds + \int_0^t \eta\sqrt{V(s)}dW_s \\
V(t) - V(0) &= k\theta t - k\int_0^t V(s)ds + \int_0^t \eta\sqrt{V(s)}dW_s \\
E[V(t) - V(0)] &= E[k\theta t] - E[k\int_0^t V(s)ds] + E[\int_0^t \eta\sqrt{V(s)}dW_s] \\
E[V(t)] - V(0) &= k\theta t - E[k\int_0^t V(s)ds] + 0 \\
E[V(t)] - V(0) &= k\theta t - kE[\int_0^t E[V(s)]ds]
\end{aligned}$$

where we use the property of Ito integral such that $E[\int_0^t \eta\sqrt{V(s)}dW_s] = 0$.

Denote $\mu(t) = E[V(t)]$; therefore, we have

$$\mu(t) - V(0) = k\theta t - k\int_0^t \mu(s)ds.$$

Take derivative, we have

$$\mu(t)' = k(\theta - \mu(t)),$$

which has solution

$$\mu(t) = \theta + \exp(-kt)(V(0) - \theta).$$

(2)

$$\begin{aligned}
K_{var} &= \frac{1}{T} E_Q[\int_0^T \sigma^2(t) dt] \\
&= \frac{1}{T} \int_0^T E_Q[\sigma^2(t)] dt \\
&= \frac{1}{T} \int_0^T (\theta + \exp(-kt)(V(0) - \theta)) dt \\
&= \frac{1}{T} (\theta T + \frac{1}{kT}(1 - \exp(-kT))(V(0) - \theta)) \\
&= \frac{1}{kT}(1 - \exp(-kT))(V(0) - \theta) + \theta
\end{aligned}$$

(3) TOdo □

18.3.11 Autocallables

18.3.11.1 *single asset autocallables*

Definition 18.3.11. *Payoff Description Consider an autocallable note based on a single asset S, a structure which pays coupons depending on the underlying's performance reaching two triggers H and B, and has a payoff defined as follows: at each observation date t_i, $(i = 1...n)$ we have*

$$Coupon(t_i) = Notional \times C \times \mathbf{1}_{Ret(t_i) \geq B} \times \mathbf{1}_{M(t_i) < H}$$

The notional redemption can be at any observation date, not necessarily at maturity.

$$Redemption(t_i) = Notional \times \mathbf{1}_{Ret(t_i) \geq B} \times \mathbf{1}_{M(t_i) < H}.$$

18.3.11.2 *Multi-asset autocallables*

Definition 18.3.12. *Assume that we start with a basket composed of n assets S1, S2,..., Sn then a worst-of autocallable note based on this basket has the following payoff. At each observation date ti : Coupon(ti) = Notional ÃŮ C ÃŮ 1WRet(ti)âĽěB ÃŮ 1(maxj=1,...,iâĹŠ1(WRet(tj)))<H H and B are respectively the autocall and coupon triggers.*

the holder receives back 100product autocalls. Otherwise, the notional is paid at maturity. Redemption(ti) = Notional ÃŮ 1WRet(ti)âĽěH ÃŮ 1(maxj=1,...,iâĹŠ1(WRet(tj)))<H Consider an autocallable note based on a single asset S, a structure which pays coupons

depending on the underlyingâĂŹs performance reaching two triggers H and B, and has a payoff

18.4 Numerical Methods for Option Pricing

18.4.1 Monte Carlo methods

Example 18.4.1 (Monte Carlo for Asian options). [3, p. 118] An Asian option uses the average price of the underlaying over some prescribed period of time in determining the final payoff. We can use following algorithms for option pricing:

- Generate underlying stock trajectories using Monte Carlo under risk-neutral measure.
- Averaging the stock prices and calculate the associated payoff for each trajectory.
- Averaging the payoffs and discount to current value, which gives the price of the option.

Definition 18.4.1 (barrier options). *[3, p. 121] A barrier option is a type of path-dependent option where the payoff is determined by whether or not the price of the underlying asset price crosses a certain level(i.e., the barrier). There are two general types of barrier options, âĂŸinâĂŹ and âĂŸoutâĂŹ options. In knock-out options, the contract is canceled if the barrier is crossed throughout the whole life. Knock-in options on the other hand are activated only if the barrier is crossed.*

Remark 18.4.1. Barrier options are path-dependent exotic options, its payoff depends on its whole path of prices of the underlying. The payoff of vanilla option only depends on the the underlying asset price at expiration.

Remark 18.4.2 (Monte Carlo for American option). Regular Monte Carlo method cannot be used to price American option, since whether to exercise at a earlier time point depends on the current value of the option, which depends on future trajectories.

18.5 Volatility and correlation modeling

18.5.1 Hedging with implied volatility

18.5.1.1 *Hedging principle*

18.5.1.2 *Patterns of volatility surface dynamics*

Definition 18.5.1 (sticky strike rule). *[9, p. 310] The* ***sticky strike*** *rule assumes that an option with a fixed strike will have the same implied volatility when the underlying price has infinitesimal change.*

Let t be the current time. Let $\sigma(S_t, K)$ denote the implied volatility. Then sticky strike says that $\sigma(S_t, K) = f(K)$, for some function f.

One possible linear approximation $\sigma(S_t, K)$ is given by

$$\sigma(S_t, K) = \sigma_0 - \beta(K - S_0),$$

where $\sigma_0, S_0 \in \mathbb{R}$ and β is a constant determining the slope of the skew.

Definition 18.5.2 (sticky moneyness rule). *[9, p. 311] The* ***sticky strike*** *rule assumes that an option with a fixed strike will have the same implied volatility with anther option with the same moneyness (K/S) when the underlying price has infinitesimal change.*

Let t be the current time. Let $\sigma(S_t, K)$ denote the implied volatility. Then sticky strike says that $\sigma(S_t, K) = f(K/S_t)$, for some function f.

One possible linear approximation $\sigma(S_t, K)$ is given by

$$\sigma(S_t, K) = \sigma_0 - \beta(K/S_t - \alpha),$$

where $\sigma_0, \alpha \in \mathbb{R}$ and β is a constant determining the slope of the skew.

Remark 18.5.1 (implementation with sticky strike and sticky moneyness rule). Suppose we have an implied volatility skew parameterized by moneyness $\sigma(K/S_t)$. Suppose now we want to find out the implied volatility we should use for underlying $S_t + \Delta S$.

- For the sticky strike rule, we should use the original implied volatility.
- For the sticky moneyness rule, we should use $\sigma(K/(S_t + \Delta S))$.

Definition 18.5.3 (sticky delta rule). *[9, p. 312]*

18.5.2 Shifted lognormal model

Definition 18.5.4 (shifted-Lognormal model). *[15] Denote forward rate by $F(t,T)$. In the Shifted-Lognormal model model, $F(t,T)$, under risk-neutral measure, is assumed to satisfy*

$$F(t,T) = X(t) + \theta$$
$$dX(t) = \sigma X(t) dW_t$$

where θ, σ are constants, $\sigma > 0$, and $W(t)$ is a standard Brownian motion. Or equivalently

$$dF(t,T) = \sigma(F(t,T) - \theta) dW_t.$$

Remark 18.5.2. Let θ be such that

$$C_{BS}(S_0 - \theta, K - \theta)$$

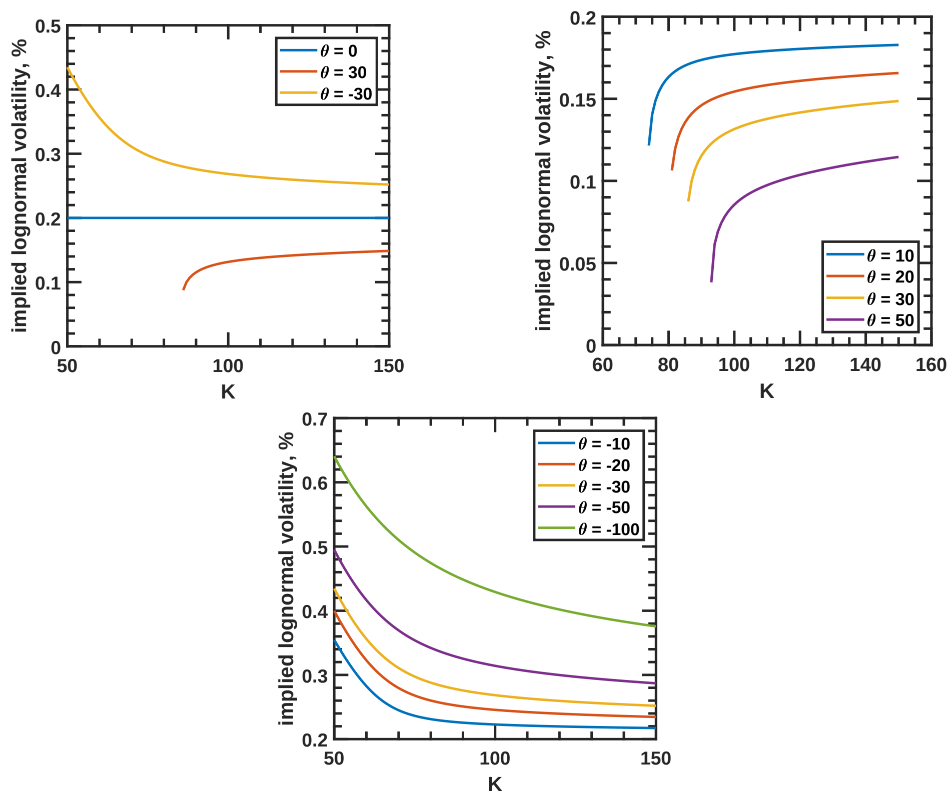


Figure 18.5.1: Implied lognormal volatility at different shift parameter in the shifted-anti-lognormal model. $F(0,T) = 100, r = 0.02, \sigma = 0.2, T = 1$.

Definition 18.5.5 (shifted-anti-lognormal model). *[15] Denote forward rate by $F(t,T)$. In the Shifted-Lognormal model model, $F(t,T)$, under risk-neutral measure, is assumed to satisfy*

$$dF(t,T) = \sigma(F(t,T) - \theta)dW_t, 0 < S_0 < \theta, \sigma < 0.$$

where θ, σ are constants, and $W(t)$ is a standard Brownian motion.

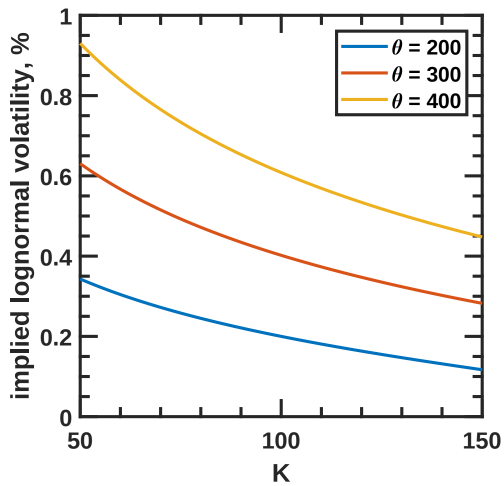


Figure 18.5.2: Implied lognormal volatility at different shift parameter in the shifted-anti-lognormal model. $F(0,T) = 100, r = 0.02, \sigma = 0.2, T = 1$.

18.5.3 Normal model

18.5.3.1 *The model*

Lemma 18.5.1 (Black-Scholes equation for normal underlying asset dynamics). *Let $V(F(t),t)$ be the value of the derivative as a function of an asset's futures price $F(t)$. Assume $F(t) \triangleq F(t,T)$* ***under real world measure*** *is governed by*

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

where W_t is the Brownian motion. Then V is governed by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial F^2} - rV = 0,$$

with final condition $V(F(T),T) = V_T(F(T))$.

Proof. Use Ito's lemma(Lemma 6.3.1), we have

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial F}dF + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dF)^2$$

Construct a portfolio $\Pi = V - \Delta F$, and take $\Delta = \frac{\partial V}{\partial F}$, then

$$d\Pi = dV - \Delta dF = \frac{\partial V}{\partial t}dt + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial F^2}dt = r\pi dt = rV.$$

Note that we use the no-arbitrage condition that risk-free asset grows at rate r and entering futures contract costs zero. □

Lemma 18.5.2 (Greeks for European options). *[16] Let current time be t. Consider a call/put option with strike K and maturity T. Assume the forward price of the underlying follows the normal model. It follows that*

- ***Delta***

$$\Delta_{C,F} = \frac{\partial C}{\partial F} = \exp(-r(T-t))N(d_1),$$
$$\Delta_{P,F} = \frac{\partial P}{\partial F} = -\exp(-r(T-t))N(-d_1).$$
$$\Delta_{C,K} = \frac{\partial C}{\partial K} = -\exp(-r(T-t))N(d_1).$$
$$\Delta_{P,K} = \frac{\partial P}{\partial K} = \exp(-r(T-t))N(-d_1).$$

- ***Gamma***

$$Gamma_{P,F} = Gamma_{C,F} = \frac{\partial^2 C}{\partial F^2} = \frac{\exp(-r(T-t))}{\sigma\sqrt{T-t}}\frac{1}{\sqrt{2\pi}}\exp(-d_1^2/2).$$

- ***Vega***

$$Vega_P = Vega_C = \frac{\partial C}{\partial \sigma} = \exp(-r(T-t))\frac{\sqrt{T-t}}{\sqrt{2\pi}}\exp(-d_1^2/2).$$

Proof. Use the pricing formula and the put-call parity given by

$$C(t) = \exp(-r(T-t))((F(t,T)-K)N(d_1) + \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}e^{-d_1^2/2});$$

and

$$C(t) - P(t) = \exp(-r(T-t))(F(t,T)-K).$$

(1)(a)

$$\begin{aligned}\frac{\partial C}{\partial F(t,T)} &= \exp(-r(T-t))N(d_1) + \exp(-r(T-t))((F(t,T)-K)n(d_1)\frac{\partial d_1}{\partial F(t,T)} - \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}d_1 e^{-d_1^2/2}\frac{\partial d_1}{\partial F(t,T)} \\ &= \exp(-r(T-t))N(d_1) + \exp(-r(T-t))\frac{\partial d_1}{\partial F(t,T)}((F(t,T)-K)n(d_1) - \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}d_1 e^{-d_1^2/2}) \\ &= \exp(-r(T-t))N(d_1).\end{aligned}$$

(b)Use put-call parity,

$$\frac{\partial C}{\partial F(t,T)} - \frac{\partial P}{\partial F(t,T)} = \exp(-r(T-t)).$$

(c) Use the symmetry between $F(t,T)$ and K, we have

$$\frac{\partial C}{\partial K} = -\frac{\partial C}{\partial F(t,T)}$$

(d)Use put-call parity,

$$\frac{\partial P}{\partial K} - \frac{\partial P}{\partial K} = -\exp(-r(T-t)).$$

□

18.5.3.2 *The implied volatility*

Lemma 18.5.3 (downward sloping implied lognormal volatility). link *In the normal forward price model, the slope of the implied volatility curve $\sigma(K)$ in the neighborhood of the S_T is downward sloping.*

Proof. In the normal model,

$$\frac{\partial C_N}{\partial K} = -N(\frac{F(0,T)-K}{\sigma\sqrt{T}})$$

In the lognormal model

$$\frac{\partial C_{BS}}{\partial K} = \frac{\partial C_{BC}}{\partial K} + \frac{\partial C}{\partial \sigma}\frac{\partial \sigma}{\partial K},$$

On the hand

$$C_N = C_{BS}(K, \sigma_I(K)).$$

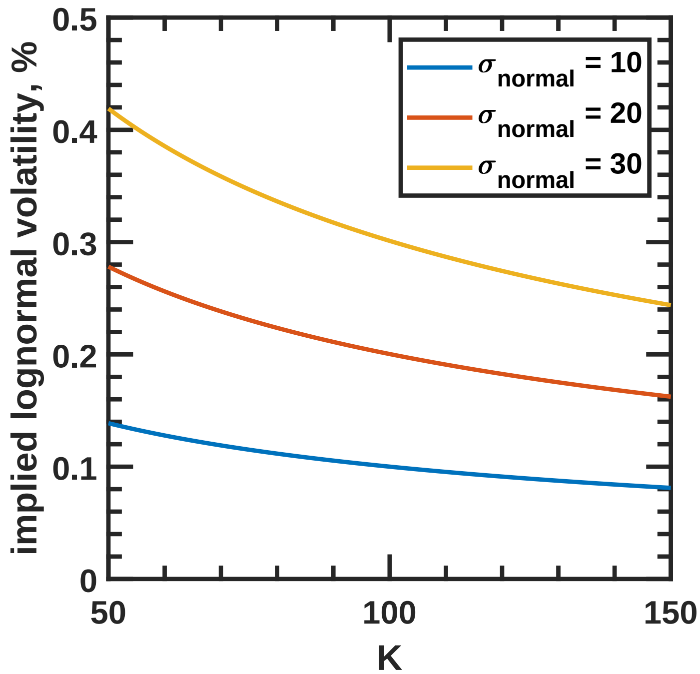


Figure 18.5.3: Normal model implied lognormal volatility at different normal volatility levels. The parameters are $T = 1, r = 0.02, F(0,T) = 100$.

Therefore

$$\begin{aligned}
\frac{dC_N}{dK} &= \frac{dC_{BS}}{dK} = \frac{\partial C_{BC}}{\partial K} + \frac{\partial C}{\partial \sigma}\frac{\partial \sigma}{\partial K} \\
\implies \frac{\partial \sigma}{\partial K} &= \frac{1}{\frac{\partial C}{\partial \sigma}}(\frac{dC_N}{dK} - \frac{\partial C_{BS}}{\partial K}) \\
\frac{\partial \sigma}{\partial K} &= \frac{1}{\frac{\partial C}{\partial \sigma}}(-N(\frac{F(0,T)-K}{\sigma\sqrt{T}}N(d_2)) \\
&\leq 0 \\
C_N &= C_{BS}(K, \sigma_I(K)).
\end{aligned}$$

□

Remark 18.5.3 (implied lognormal volatility). As showed in Figure 18.5.3, normal model generally generate downward sloping implied lognormal volatility. The larger the normal volatility, the steeper the slope.

18.6 Hedging and pricing

18.6.1 Static hedging

Remark 18.6.1. The general principles of static hedging is discussed at subsubsection 16.2.3.1.

18.6.2 Dynamic hedging fundamentals

Note 18.6.1 (general remarks on dynamic hedging). [9, p. 17]

- (**static vs. dynamic**)static hedging only relies on the final payoff function and the law of one price, whereas dynamic hedging relies on the additional assumptions on the asset dynamic model and the accuracy of model parameters.
- (**practical difficulties**): The efficacy of dynamic hedging rests on the correctness of the assumed underlying asset dynamics, the abilities to continuous trading to re-balance the portfolio. However, bid-ask spread, illiquidity, and market impact due to trade can affect the the hedging strategy.

18.6.2.1 *Delta hedging in Black-Scholes world*

Lemma 18.6.1 (delta-hedging in continuous time). *Consider a portfolio continuously maintained at position $\pi(t) = -C(S_t, t) + \frac{\partial C}{\partial S_t} S_t$. Then the portfolio π is risk-less and grows at a rate of r. In particular, We have*

- $$\begin{aligned} d\pi &= \frac{\partial \pi}{\partial t} dt + (\frac{\partial C}{\partial S} - \Delta) dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \\ &= \frac{\partial \pi}{\partial t} dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \\ &= r\pi dt; \end{aligned}$$

  that is, the value of the portfolio will grow at a rate r.
- $\frac{\partial \pi}{\partial S} = 0$; *that is, the portfolio π is the delta-neutral.*
- $\frac{\partial^2 \pi}{\partial S^2} = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}$; *that is, the portfolio π is not Gamma-neutral.*

Proof.

$$\begin{aligned} d\pi \triangleq d(C - \frac{\partial C}{\partial S_t}S_t) &= (\frac{\partial C}{\partial t}dt + \frac{\partial V}{\partial S_t}dS_t + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}dt) + \frac{\partial C}{\partial S_t}dS_t \\ &= (\frac{\partial C}{\partial t}dt + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}dt) \\ &= rV - rS_t\frac{\partial C}{\partial S_t}S_t = r\pi \end{aligned}$$

where in the last step we use Black-Scholes equation

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0.$$

□

Note 18.6.2 (interpretation and how the market works). Assume we are in the Black-Scholes world. Consider a bank sells a call option to a counter-party at price $C_0 = e^{-rT}E_Q[(S_T - K)]$ and then the bank performs continuous hedging on the short position. Any remaining cash is invested in the money market to receive interest and any cash borrowed is funded at the same interest rate.

- The buyer (we assume he borrow money to buy)of a call option at maturity date T will have a profit-and-loss given by $-C_0e^{rT} + (S_T - K)^+$, which S_T is a random number has the **real world probability distribution**.
- When the bank is performing continuously hedging, it is equivalent that the bank invest the money C_0 into the money market. At the maturity date T, eventually the bank will have profit-and-loss given by

$$C_0 - (S_T - K)^+,$$

  where S_T is has the **risk-neutral probability distribution**. The expected profit-and-loss of the bank is zero.
- If the bank does not do any hedging, the bank will have an expected profit-and-loss

$$C_0e^{rT} - E[(S_T - K)^+] < C_0e^{rT} - E_Q[(S_T - K)^+] = 0;$$

  That is, the bank will lose money. Here we assume the stock in the real probability measure will have a larger growth rate than r since it is risky asset(otherwise it is undervalued and the buyers will push the price up).
- If the bank sells at fair price, then the bank will have zero profit-and-loss because we assume the financial market is perfectly competitive. In reality(it is not perfectly competitive), the bank will charge a higher price, for example

between $e^{-rT}E_Q[(S_T - K)]$ and $e^{-rT}E[(S_T - K)]$; that is, a price higher than cost $e^{-rT}E_Q[(S_T - K)]$ and lower than the buyer's reserve price $e^{-rT}E[(S_T - K)]$.

Definition 18.6.1 (Discrete-time Delta hedging). *[14, p. 84][6, p. 408]*

- *The option is hedged at evenly spaced discrete time points, $t_0, t_1, ..., t_n$, such that $\delta t \triangleq t_i - t_{i-1}$ and t_n representing the expiration time of the option.*
- *$C_i = C(S_i, t_i)$ denote the price of the option at time t_i when the stock price is S_i.*
- *$\Delta_i = \Delta(S_i, t_i)$ denote the number of shares of the stock S that we short at the start each period i.*
- *Any remaining cash is invested in the money market to receive interest and any cash borrowed is funded at the same interest rate.*

We have the following hedging actions:

- *At the beginning t_0, we short Δ_0 shares of stock, and we have position $C_0 - \Delta_0 S_0$ and cash $\Delta_0 S_0$.*
- *At time t_1, we short $\Delta_1 - \Delta_0$ stock, and we have position $C_1 - \Delta_1 S_1$ and cash $\Delta_0 S_0 e^{r\delta t} + (\Delta_1 - \Delta_0) S_1$.*
- *At time t_2, we short $\Delta_2 - \Delta_1$ stock, and we have position $C_2 - \Delta_2 S_2$ and cash $\Delta_0 S_0 e^{2r\delta t} + (\Delta_1 - \Delta_0) S_1 (\Delta_1 - \Delta_0) S_1 e^{r\delta t} + (\Delta_2 - \Delta_1) S_2$.*
- ...
- *At time t_n, we short $\Delta_n - \Delta_{n-1}$ stock, and we have position $C_n - \Delta_n S_n$ and cash $\Delta_0 S_0 e^{rn\delta t} + \cdots + (\Delta_n - \Delta_{n-1}) S_n$.*

*Given a call option, we want to hedge its price C_t change due to the price fluctuation of S_t, then the **first-order** hedging strategy is to for any call in the short position, we long Δ unit of S_t. Then the portfolio has value $V_t = -C_t + \Delta S_t$ and*

$$\frac{\partial V_t}{\partial S_t} \approx \Delta - \Delta = 0$$

which suggests the portfolio is neutral to the S_t changes.

Caution!

- *Note that when there are large change in S_t, we might need the Γ as well.*
- *When $t \to T$, then $\Delta \to \infty$, indicating that when close to maturity, it is very difficult to do do Delta hedging.*
- *Theoretically, after we short a call $C_t(S_t)$ at spot price S_t at $t = 0$, we have continuously buy or sell stock to keep a portfolio*

$$V_t = -C_t$$

Remark 18.6.2 (why we do not hedge time factor?).

Lemma 18.6.2 (asymptotic of discrete-time hedging). *At maturity t_n, we have total asset valued at*

$$C_n - \Delta_n S_n + \Delta_0 S_0 e^{rn\delta t} + \cdots + (\Delta_n - \Delta_{n-1})S_n$$

.

Proof. At maturity t_n, we have position $C_n - \Delta_n S_n$ and cash $\Delta_0 S_0 e^{rn\delta t} + \cdots + (\Delta_n - \Delta_{n-1})S_n$. □

Lemma 18.6.3 (discrete-time hedging error). *[9, p. 111][6, p. 430]*

- *the hedging error for a single step of time interval δt is approximately*

$$\delta HE \approx \frac{1}{2}\Gamma\sigma^2 S^2 (Z^2 - 1),$$

where $\Gamma \triangleq \partial^2 C/\partial S^2$

- *Over n steps to expiration, the total error is*

$$HE \approx \sum_{i=1}^{n} \frac{1}{2}\Gamma_i \sigma_i^2 S_i^2 (Z_i^2 - 1)\delta t$$

- $$E[HE] = 0, Var[HE] = E[\sum_{i=1}^{n} \frac{1}{2}\Gamma_i^2 \sigma_i^4 S_i^4 \delta t^2].$$

- *$Var[HE] \propto O(1/n)$*

Proof. (1)(2) The Talyor expansion on the delta-neutral profile $\pi =\triangleq C - \Delta S$ is given by

$$\pi(t+\delta t) \approx \pi(t) + \frac{\partial \pi}{\partial S}\delta S + \frac{\partial \pi}{\partial t}\delta t + \frac{1}{2}\Gamma(\delta S)^2 = \pi(t) + \frac{\partial \pi}{\partial t}\delta t + \frac{1}{2}\Gamma(\delta S)^2.$$

The single step hedging error is given by

$$\begin{aligned}
\delta HE &= \pi(t+\delta t) - \pi e^{r\delta t} \\
&\approx \pi + \frac{\partial \pi}{\partial t}\delta t + \frac{1}{2}\Gamma(\delta S)^2 - \pi(t) - \pi(t) r\delta t \\
&= \pi(t) + \frac{\partial \pi}{\partial t}\delta t + \frac{1}{2}\Gamma(\delta S)^2 - \pi(t) - \pi(t) r\delta t \\
&= \frac{1}{2}\Gamma(\delta S)^2 - \frac{1}{2}\sigma^2 S^2 \Gamma \delta t \\
&= \frac{1}{2}\Gamma\sigma^2 S^2 Z^2 \delta t - \frac{1}{2}\sigma^2 S^2 \Gamma \delta t \\
&= \frac{1}{2}\Gamma\sigma^2 S^2 (Z^2 - 1)\delta t
\end{aligned}$$

To estimate $(\delta S)^2$, we use the following

$$
\begin{aligned}
S_{t+\delta t} &= S_t \exp((\mu - \frac{\sigma^2}{2})\delta t + \sigma\sqrt{\delta t} Z) \\
\implies \frac{(\delta S)^2}{S_t^2} &= \frac{(S_{t+\delta t} - S_t)^2}{S_t^2} \\
&\approx ((\mu - \frac{\sigma^2}{2})\delta t + \sigma\sqrt{\delta t} Z)^2 \\
&= \sigma^2 Z^2 \delta t + O((\delta t)^2)
\end{aligned}
$$

We also the relation with Greeks for a continuously delta-hedged portfolio (Lemma 18.2.14):

$$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma = r\pi.$$

(3) Note that Z^2 is $\chi^2(1)$ random variable with mean 1 and variance 2. See Lemma 2.2.36. (4) Use $\delta t = (T-t)/n$.

□

18.6.2.2 *Practical issues: which volatility to use?*

Lemma 18.6.4 (P&L of delta-hedged options). *[14, p. 86][11, p. 60]*

- $$P\&L_{\Delta t} \approx \Delta \Delta S + \frac{1}{2}\Gamma(\Delta S)^2 + \Theta \Delta t + \nu \Delta \sigma + \rho \Delta r.$$

where $\Theta = \frac{\partial V_t}{\partial t}$ *and* $\Gamma = \frac{\partial^2 V}{\partial S^2}$.

$$\Theta + \frac{1}{2}\sigma^2 S_t^t \Gamma = rV_t,$$

Use approximation $\Delta \approx -\frac{1}{2}\Gamma\sigma^2 S^2$. *If the option is delta-hedged at N regular intervals of length* Δt *until maturity, the cumulative P&L proxy equation is given by*

$$Cumu.P\&L \approx \frac{1}{2}\sum_{t=0}^{N-1} \Gamma_t S_t^2 [(\frac{\Delta S_t}{S_t})^2 - (\sigma^* \sqrt{\Delta t})^2].$$

If the option is delta-hedged at continuous time, the cumulative P&L proxy equation is given by

$$Cumu.P\&L = \frac{1}{2}\int_0^T e^{r(T-t)}\Gamma_t S_t^2[\sigma_t^2 - \sigma^*]dt.$$

18.6.3 Discrete time hedging

Lemma 18.6.5 (price of derivatives under discrete time hedging). *The price of a derivative resulted from discrete-time hedging strategy is given by*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r_d S\frac{\partial V}{\partial S} - r_d V = 0$$

where r_d is the market expected growth rate for a discrete-time delta-hedge portfolio growth rate.

Proof.

$$\begin{aligned} \delta\pi &= \frac{\partial V}{\partial t}\delta t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 Z^2 \delta t, Z \sim N(0,1) \\ \implies E[\delta\pi] &= \frac{\partial V}{\partial t}\delta t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 \delta t = r_d(V - \frac{\partial V}{\partial S}S)\delta t \end{aligned}$$

□

Remark 18.6.3 (prices of European derivatives). Let $C_{Blk}(r, S_t, K, T, \sigma)$ denote the price of a call when short rate is r.

- A European call will have price $C_{Blk}(r_d, S_t, K, T, \sigma)$. Since $\partial C_{Blk}/\partial r > 0$ (Lemma 18.2.13) and $r_d > r$, a European call will have higher price than Black-Scholes model price.
- When a bank sells a call, the bank should sell at a higher price since discrete-time delta-hedging contains extra risk.

Remark 18.6.4 (optimal delta-hedging). The amount of stock used to hedged can be further optimized to minimized the variance of growth rate of a delta-hedging portfolio.See [1, p. 773].

18.6.4 Hedging with transaction cost

Lemma 18.6.6 (governing equation of derivative value when there is transaction cost). *[9, p. 125][1, p. 785] Assume the delta-hedging is performed at the interval δt. Let c denote the transaction cost of buying or selling per unit of S. Then*

- *the transaction cost to maintain delta-hedging at each interval δt is*

$$c\left|\frac{\partial^2 V}{\partial S^2}\eta\right|\sigma S^2\sqrt{\delta t}, \eta \sim N(0,1).$$

- *the expected value of transaction cost is*

$$c\left|\frac{\partial^2 V}{\partial S^2}\right|\sqrt{\frac{2}{\pi}}\sigma S^2\sqrt{\delta t}$$

- *the governing equation for value of the derivative is*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} - c\sigma S^2\sqrt{\frac{2}{\pi\delta t}}\left|\frac{\partial^2 V}{\partial S^2}\right| + r_d S\frac{\partial V}{\partial S} - r_d V = 0$$

where r_d is the market expected growth rate for a discrete-time delta-hedge portfolio growth rate.

Proof. (1) To maintain delta-neutral position, the number of underlaying assets we should keep is $\Delta = \partial V/\partial S$. Therefore, when we rebalance our portfolio, the number of the underlying we should buy/sell is

$$\partial V/\partial S(S+\delta S) - \partial V/\partial S(S) = \partial^2 aV/\partial S^2\delta S = \partial^2 aV/\partial S^2\sigma S\eta\sqrt{\delta t}.$$

Then, the transaction cost will be

$$cS(\partial V/\partial S(S+\delta S) - \partial V/\partial S(S)) = c\left|\frac{\partial^2 V}{\partial S^2}\right|\sigma S^2\eta\sqrt{\delta t}$$

(2) note that $E[|\eta|] = \sqrt{2/\pi}$. (3) Consider the portfolio $\pi = V - \partial V/\partial SS$, we have

$$\begin{aligned}E[\delta\pi] &= E[\frac{\partial V}{\partial t}\delta t + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\delta t - c\left|\frac{\partial^2 V}{\partial S^2}\eta\right|\sigma S^2\sqrt{\delta t}] \\ &= (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + c\left|\frac{\partial^2 V}{\partial S^2}\right|\sqrt{\frac{2}{\pi\delta t}}\sigma S^2)\delta t = r_d\pi\delta t\end{aligned}$$

□

Note 18.6.3 (price of a European derivatives). Let $C_{Blk}(r, S_t, K, T, \sigma)$ denote the price of a call when short rate is r and volatility of the stock is σ. Then,

- the long position of a European call ($\partial^2 V/\partial S^2 > 0$) has value
$$V = C_{Blk}(r_d, S_t, K, T, \hat{\sigma}), \hat{\sigma}^2 = \sigma^2 - 2c\sigma\sqrt{\frac{2}{\pi\delta t}}.$$
Since $\frac{\partial C_{Blk}}{\partial r} > 0$, $r_d \geq r$ increases the value of a long position. However, the smaller volatility decrease the value of a long position.
- the short position of a European call($\partial^2 V/\partial S^2 < 0$) has value
$$V = C_{Blk}(r_d, S_t, K, T, \hat{\sigma}), \hat{\sigma}^2 = \sigma^2 + 2c\sigma\sqrt{\frac{2}{\pi\delta t}}.$$
Since $\frac{\partial C_{Blk}}{\partial r} > 0$, $r_d \geq r$ decreases the value of a short position. However, the larger volatility increases the value of a short position.
- If $r_d = r$, when a bank sells the call, the bank should charge more than the Black-Scholes' model price in order to cover the hedging cost.
- If $r_d = r$, when an investor long the call, the investor should pay less since the hedging cost will reduce the profitability of a call.

Remark 18.6.5 (breaking the linear pricing law). [1, p. 787]

- We can see that the governing PDE is a nonlinear parabolic PDE. Therefore, the long position of a call is not equal to the negative price of the short position of a call.
- In the real world, the existence of transaction cost will also break the linear pricing law(Theorem 15.2.2).

18.7 Notes on bibliography

For practical investment strategy using options, see [17].

For advanced treatment of options, see [18].

For volatility and correlation modeling, see [11][9][19].

Practical option books:[20][21][22][23].

For dynamic hedging, see [24][25].

For equity option quotes data, see NYSE.

BIBLIOGRAPHY

1. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
2. Remillard, B. *Statistical methods for financial engineering* (CRC Press, 2013).
3. Shonkwiler, R. W. *Finance with Monte Carlo* (Springer, 2013).
4. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
5. Black, F. & Scholes, M. The pricing of options and corporate liabilities. *Journal of political economy* **81,** 637–654 (1973).
6. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
7. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
8. Yu, X. & Xie, X. On Derivations of Black-Scholes Greek Letters. *Research Journal of Finance and Accounting* **4,** 80–85 (2013).
9. Derman, E., Miller, M. & Park, D. *The Volatility Smile* ISBN: 9781118959169. https://books.google.com/books?id=WKfBDAAAQBAJ (Wiley, 2016).
10. Lalley, S. *The mathematics of financial derivatives: a student introduction* (Chicago University Statistics 390/ Mathematical Finance 345, 2001).
11. Bossu, S. *Advanced Equity Derivatives: Volatility and Correlation* (John Wiley & Sons, 2014).
12. Borovkova, S., Permana, F. J. & Weide, H. V. A closed form approach to the valuation and hedging of basket and spread options. *Journal of Derivatives* **14,** 8 (2007).
13. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
14. Bossu, S. & Henrotte, P. *An Introduction to Equity Derivatives: Theory and Practice* ISBN: 9781119961857. https://books.google.com/books?id=SgG8cg0uAEgC (Wiley, 2012).
15. Lee, R. & Wang, D. Displaced lognormal volatility skews: analysis and applications to stochastic volatility simulations. *Annals of Finance* **8,** 159–181 (2012).
16. Iwasawa, K. Analytical formula for the European normal Black Scholes formula. *Unpublished working paper, viewed* **27** (2013).
17. McMillan, L. G. *Options as a strategic investment* (Penguin, 2002).

18. Guyon, J. & Henry-Labordère, P. *Nonlinear option pricing* (CRC Press, 2013).
19. Rebonato, R. *Volatility and correlation: the perfect hedger and the fox* (John Wiley & Sons, 2005).
20. Shover, L. *Trading Options in Turbulent Markets: Master Uncertainty Through Active Volatility Management* (John Wiley & Sons, 2012).
21. Haug, E. G. *The complete guide to option pricing formulas* (McGraw-Hill Companies, 2007).
22. Haug, E. G. *Derivatives models on models* (John Wiley & Sons, 2013).
23. Bennett, C. *Trading Volatility: Trading Volatility, Correlation, Term Structure and Skew* (CreateSpace Independent Publishing Platform, 2014).
24. Passarelli, D. *Trading options Greeks: How time, volatility, and other pricing factors drive profits* (John Wiley & Sons, 2012).
25. Taleb, N. *Dynamic hedging: managing vanilla and exotic options* (John Wiley & Sons, 1997).

19

INTEREST RATE MODELS & DERIVATIVES

Overview

Modeling interest rate plays a critical role in pricing and risk management for interest rate derivatives and fixed income products.

19.1 Interest rate concepts

19.1.1 Basic definitions

19.1.1.1 *Zero-coupon bond*

Definition 19.1.1 (zero-coupon bond). *A **zero-coupon bond** is a bond that pays its face value (say, $1)at the maturity time T. The set of prices of zero-coupon bonds for various time-horizons/maturities is known as the **zero-coupon curve**.*

Remark 19.1.1 (importance of zero-coupon bond)**.**

- Zero-coupon bonds are not actively traded within the interbank market; however, they are important because interbank interest rates such as LIBOR and swap rates can be defined in terms of zero-coupon bonds.

Definition 19.1.2 (yield of zero-coupon bond). *The yield of the zero-coupon bond with maturity T is the equivalent **constant interest rate** such that the price of the zero-coupon bond invested at time 0 and accumulated at this interest would grow to the face value at T.*

Remark 19.1.2. We can view zero-coupon bond as a derivative on the interest rate that its value today depends on its face value and the interest rate dynamics.

Definition 19.1.3 (price of zero-coupon bonds). *We denote $P(t,T)$ as the price at time t of a zero-coupon bond that matures at time T(pays $1). where we have some practical constraints and issues:*

1. *$t \leq T$(before maturity)*
2. *t is not necessarily the time the bond is issuing.*
3. *$P(t,t) = 1, P(t,T) \leq 1, \forall T \geq t$, assuming that the interest rate is always non-negative.*

Remark 19.1.3 (interpretation)**.** The bond price is a function of two parameters t and T.

Remark 19.1.4 (caution!)**.** [1, p. 5] Let $t < S < T$. It is tempting to use no-arbitrage argument to show

$$P(t,T) = P(t,S)P(S,T).$$

However, such equality will not hold since $P(S,T)$ is a stochastic quantity whereas $P(t,T)$ and $P(t,S)$ are deterministic quantities.

19.1.1.2 *Spot rates*

Definition 19.1.4 (simply compounded spot rate). *[1, p. 2]*

- *The **simply compounded spot rate** at time t for maturity T is defined as the annualized rate of return from holding a zero-coupon bond from time t until maturity T. It is denoted by $L(t, T)$.*
- *The simply compounded spot rate for tenor 1M, 3M, 6M, 12M usually can be directly observed from the market, e.g., annual rate for CDs.*

Theorem 19.1.1 (no-arbitrage relation between spot rate and zero-coupon bond). *Let current time be t. Assume the market is free ofarbitrage opportunities.*

- *At current time t, the (default-free) zero-coupon bond price then can be expressed in terms of the spot rate as*

$$P(t,T) = \frac{1}{1+(T-t)L(t,T)}.$$

Or equivalently,

$$L(t,T) = \frac{1-P(t,T)}{(T-t)P(t,T)}.$$

- *At future time S, no arbitrage condition requires that*

$$P(S,T) = \frac{1}{1+(T-S)L(S,T)}.$$

- *At future time S, no arbitrage condition requires that*

$$1+(T-S)L(S,T) = \frac{M(T)}{M(S)}.$$

Proof. (1) If zero-coupon bond is not priced this way, then we can long or short the zero-coupon bond at current time t to make arbitrages. (2)(3) If zero-coupon bond or money market account are not priced this way, then we can long or short the zero-coupon bond at time S to make arbitrages. □

Remark 19.1.5 (interpretation of equality for stochastic quantities/processes)**.**

- Let current time be t. For future time $S > t$, we can view $L(S,T)$ and $P(S,T)$ and $M(S)/M(T)$ as stochastic process indexed by S.
- The arbitrage strategy is applied to each realization path, therefore, on each sample path $\omega \in \Omega$, we have

$$\frac{1}{1+(T-S)L(S,T)(\omega)} = P(S,T)(\omega) = \frac{M(T)(\omega)}{M(S)(\omega)}.$$

Definition 19.1.5 (continuously compounded spot rate, yield, zero curve). *Because of the no-arbitrage relationship between sport rate and zero-coupon bonds (Theorem 19.1.1), we can similarly define*

- *The **continuously compounded spot rate** is the annualized logarithmic rate of return from holding the bond from time t until maturity T. It is denoted by $R(t,T)$, and related to zero-coupon bond as*

$$R(t,T) = -\frac{\ln P(t,T)}{T-t}.$$

- *$R(t,T)$ is also called **yield to maturity**. The graph of $R(t,T)$ versus maturity T is known as the **yield curve or zero curve**.*
- *The zero-coupon bond price then can be expressed in terms of the spot rate as*

$$P(t,T) = \exp(-R(t,T)(T-t)).$$

Remark 19.1.6 (the parameter of yield curve)**.** We usually denote yield curve by $Y(T) = R(t=0,T)$. Note that in yield curve, we can only vary the maturity rather than the current time t. Yield curves can be directly observed from the market

Remark 19.1.7 (relations of simply and continuously compounded spot rate)**.**

$$\lim_{T\to t} R(t,T) = \lim_{T\to t} L(t,T).$$

That is, a short holding period, the two spot rates are the same.

19.1.1.3 *LIBOR payment*

Theorem 19.1.2 (current no-arbitrage value of LIBOR payment). *[1, p. 6] Let $S < T$. Consider the future LIBOR-based payment $(T-S)L(S,T)$ at time T. Its arbitrage-free value at time $t < T$ is $P(t,S) - P(t,T), T > S$. That is*

$$P(t,S) = P(t,T) = E_Q[\frac{M(t)(T-S)L(S,T)}{M(T)}|\mathcal{F}_t].$$

Proof. (replicating method) We use the following strategy to replicate the payoff at time T.

- At time t, we buy an $S-$bond and sell a $T-$bond.
- At time S, the long position in S-bond matures to yield one dollar. Then use this income to buy an amount of $1/P(S,T)$ of $T-$bonds.
- At time T, the net position is $1/P(S-T) - 1$, which equals $(T-S)L(S,T)$ based on the definition of $L(S,T)$

□

19.1.1.4 *Forward zero-coupon bonds*

Definition 19.1.6 (forward (zero-coupon) bond contract, forward (zero–coupon)bond price). *[1, p. 7] Let $t < S < T$. Let the current time be t.*

- *A **forward (zero-coupon) bond contract** is a contract between two parties to buy or sell a zero-coupon bond, which matures at time T, at a future time S with a price agreed upon at time t today.*
- *The **forward bond price** is the agreed price such that the contract has zero value.*
- *The forward bond price*
$$FB(t; S, T) = \frac{P(t,T)}{P(t,S)},$$
which is the price we agree on time t that we need to pay at time S to get the $T-$bond.

Lemma 19.1.1 (forward zero-coupon bond price). *Let $t < S < T$. Let the current time be t. In a no-arbitrage market, the forward zero-coupon bond is given by*

- $$FB(t; S, T) = P(t,T)(1 + (S-t)L(t,S))$$
- $$FB(t; S, T) = \frac{P(t,T)}{P(t,S)}$$

Proof. (1) We can replicate the forward contract to deliver a $T-$maturity bond at time S in the following way:

- Borrow $P(t,T)$ at time t; buy zero-coupon bond at time t with price $P(t,T)$.
- At time S, we delivery the bond and get payment $FB(t; S, T)$. Also, we pay the loan with a total of $P(t,T)(1 + (S-t)L(t,S))$.

Because our initial capital is zero, then our final payoff $+FB(t; S, T - P(t,T)(1 + (S - t)L(t,S))$ should also be zero. (2) Use the relation between spot rate and zero-coupon bond (Theorem 19.1.1). □

Remark 19.1.8. Let A be the amount of money we pay at time S to get the $T-$bond, the case flow at time 0 is
$$V(0) = -AP(t,S) + P(t,T).$$
where $AP(t,S)$ is the present value of the future cash A at time S, and $P(t,T)$ is the present value of the future cash 1 at time T. Set $V(0) = 0$, we have
$$A = \frac{P(t,T)}{P(t,S)}.$$

For positive interest rate, we have $FP(t;S,T) \leq 1$.

Remark 19.1.9 (forward rate vs. forward bond price).

- $$FP(t;S,T) = \frac{P(t,T)}{P(t,S)} = \frac{1}{1+(T-S)F(t;S,T)}.$$
  We can think of the forward rate as the simply compounded rate of return over the time interval $[S,T]$ implied by the forward bond price.
- Recall that the zero-coupon bond price can be expressed as
  $$P(t,T) = \frac{1}{1+(T-t)L(t,T)}.$$

Lemma 19.1.2 (zero-coupon bond decomposition). *Let $T_0 < T_1 < \cdots < T_n$ be a set of dates. Then*

$$\begin{aligned} P(T_0,T_n) &= FP(T_0;T_0,T_1)FP(T_0;T_1,T_2)\cdots FP(T_0;T_{n-1},T_n) \\ &= \prod_{i=1}^{n} \frac{1}{1+(T_i-T_{i-1})F(T_0,T_{i-1},T_i)} \\ &= \prod_{i=1}^{n} \exp(-(T_i-T_{i-1})R(T_0;T_{i-1},T_i)) \end{aligned}$$

Proof. Note the definitions:

$$FP(t;S,T) = \frac{P(t,T)}{P(t,S)} = \exp(-R(t;S,T)(T-S)),$$

and

$$FP(t;S,T) = \frac{P(t,T)}{P(t,S)} = \frac{1}{1+(T-S)F(t;S,T)}.$$

□

19.1.1.5 *Forward rates*

Definition 19.1.7 (forward rate agreement). *[1, p. 5]*

- *Let $t < S < T$. A unit notional* ***forward rate agreement(FRA)*** *is a contract entered into at time t, when the issuer agrees to pay the holder at time T the LIBOR $L(S,T)$ in exchange for a fixed rate K applied to unit notional amount. The value of the payoff at time T is given by*
  $$(T-S)(K-L(S,T)).$$

- *The value of K is selected such that the value of FRA at time t is zero. Such K is called* ***forward rate****, denoted by* $F(t, S, T)$.

Remark 19.1.10 (using forward rate agreement to lock future rates).

- Suppose that party A use a forward rate agreement to lock the future rate on period $[S, T]$ at current time t.
- As showed in Figure 19.1.1, party A can enter the forward rate agreement with party B at time t, and then invest in the money market at time S.
- At time T, party A can exchange with party B the cash flow to receive fixed rate payments.

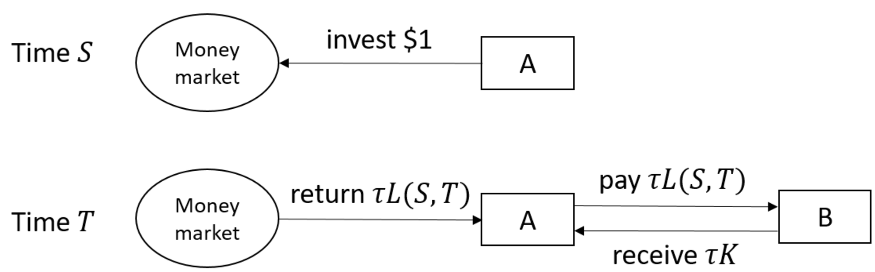


Figure 19.1.1: Using forward rate contract to lock future rate.

Remark 19.1.11 (interpretation, spot vs forward rate).

- Spot rate is the current market rate of return if we invest in money market for a period of time.
- Forward rate implied from forward rate agreement is future market rate of return if we invest in money market for a future period of time.

Theorem 19.1.3 (no-arbitrage forward rate). *Let* $t < T_1 < T_2$. *Let current time be* t. *In a no-arbitrage market, it is required that*

$$F(t, T_1, T_2) = \frac{1}{T_2 - T_1}(\frac{P(t, T_2)}{P(t, T_1)} - 1),$$

Further more, the time t *value of LIBOR payment* $(T_2 - T_1)L(T_1, T_2)$ *is given by* $(T_2 - T_1)F(t_1; T_1, T_2)P(t_1, T_2)$.

Proof. (1)(2)Note that the current value of a LIBOR payment via replication method (Theorem 19.1.2) is given by

$$P(t, T_1) - P(t, T_2),$$

which should equal the current value of the fixed rate payment, i.e.,

$$(T_2 - T_1)F(t, T_2, T_1)P(t, T_2).$$

$\square$

Example 19.1.1. For two annually compounded spot rate $L(0,1) = 4\%$ and $L(0,2) = 5\%$, maturing in one and two years respectively, then the one-year to two-year forward rate $F(0;1,2)$ can be computed as follows:

$$F(0;1,2) = (P(0,1)/P(0,2) - 1)$$

where

$$P(0,1)(1 + L(0,1)) = 1, P(0,2)(1 + 2L(0,2)) = 1.$$

Theorem 19.1.4 (forward rate agreement value). *[1, p. 6] Consider a forward rate agreement(FRA), whose interest accrual period is $[S,T]$, has payoff at time T given by*

$$(T - S)(K - L(S,T)).$$

- ***(fair value)*** *It value at current time t is given as*

$$V(t) = P(t,T)(T - S)K - P(t,S) + P(t,T).$$

- ***(fair rate)*** *If we set*

$$K \triangleq F(t;S,T) = \frac{P(t,S) - P(t,T)}{(T - S)P(t,T)} = \frac{1}{T - S}(\frac{P(t,S)}{P(t,T)} - 1),$$

the FRA has value 0 at time t. Or equivalently,

$$\frac{P(t,T)}{P(t,S)} = \frac{1}{1 + (T - S)F(t,S,T)}.$$

Note that by definition $F(S;S,T) = L(S,T)$.

- ***(value evolution)*** *Consider a forward contract entered at t_0 with fair rate K. At future time $t_1, t_0 < t_1 < T$, the value of the forward contract is given by*

$$V(t_1) = P(t_1,T)\tau(F(t_0;S,T) - F(t_1;S,T)),$$

where $F(t_1;S,T)$ is the fair forward rate at t_1, $\tau = T - S$. ***Note that** $P(t_1,T)$, $F(t_1;S,T)$ **and** $V(t_1)$ **are all random quantities at time** t_0.*

Proof. (1)(2)Use the current value of LIBOR payment formula Theorem 19.1.2. (3) (a) (replication method) Because time t_1 of the floating payment $(T-S)F(t_1;S,T)P(t_1,T)$

(b) (martingale method) Note that at any time $t<T$, the fair value forward rate is related to the LIBOR payment via

$$\begin{aligned}
&V(t) = P(t,T)E_T[\tau F(t;S,T) - \tau L(S,T)|\mathcal{F}_t] = 0 \\
&\implies P(t,T)\tau F(t;S,T) = P(t,T)E_T[\tau L(S,T)|\mathcal{F}_t] = PV_t[\tau L(S,T)]
\end{aligned}$$

where $PV_t[\tau L(S,T)]$ denotes the t-value of the future T-LIBOR payment $\tau L(S,T)$.

For a contract entered at time t_0, the fixed payment has value $P(t_1,T)\tau F(t_0;S,T)$, the LIBOR payment has value

$$PV_{t_1}[\tau L(S,T)] = P(t_1,T)E_{t_1}[\tau L(S,T)|\mathcal{F}_{t_1}] = P(t_1,T)\tau F(t_1;S,T).$$

Therefore, the value of the forward contract is given by

$$V(t_1) = P(t_1,T)\tau F(t_0;S,T) - P(t_1,T)\tau F(t_1;S,T).$$

□

Definition 19.1.8 (continuously compounded forward rate). *The continuously compounded forward rate at time t, which applies between future times T and $S(t\leq S<T)$ is defined as*

$$R(t;S,T) \triangleq -\frac{1}{T-S}\log\frac{P(t,T)}{P(t,S)} = -\frac{1}{T-S}\log(FP(t;S,T)),$$

such that

$$FP(t;S,T) = \frac{P(t,T)}{P(t,S)} = \exp(-R(t;S,T)(T-S)).$$

Lemma 19.1.3 (no-arbitrage relationship for forward rates decomposition). *Let $t\leq T_1<T_2<T_3$. Then in a no-arbitrage market, it is required that*

$$(T_3-T_1)F(t;T_1,T_3)+1 = [(T_2-T_1)F(t;T_1,T_2)+1][(T_3-T_2)F(t;T_2,T_3)+1].$$

Proof. A forward contract locking the rate in period $[T_1,T_3]$ can be exactly replicated by two forward contracts locking the rate in the period $[T_1,T_2]$ and $[T_2,T_3]$. □

19.1.2 Instantaneous rates

Definition 19.1.9 (instantaneous forward rate).

- *The instantaneous forward rate at time t is*

$$f(t,T) \triangleq \lim_{S\to T} R(t;T,S) = \lim_{S\to T} F(t;T,S) = -\frac{\partial}{\partial T}\log P(t,T)$$

- *We also have (integration based on (1))*

$$P(t,T) = \exp(-\int_t^T f(t,u)du).$$

The dependence of $f(t,T)$ on the maturity T is known as the ***term structure of forward rate curve*** *or* ***forward rate curve***[a] *at time t.*

[a] Note that forward rate curve and forward curve usually refers to different curves. Forward curve is associated with a tenor.

Definition 19.1.10 (short rate, risk-free rate). *The* ***short rate*** *is defined as*

$$r(t) = \lim_{T\to t} R(t,T) = \lim_{T\to t} L(t,T) = R(t,t) = f(t,t).$$

Short rate is what we usually refer to as risk-free rate.

Note 19.1.1 (Caution! non-constant short rate cannot determine zero-coupon bond price)**.** Note that

$$\begin{aligned} P(t,T) &= \exp(-R(t,T)(T-t)) = \exp(-\int_t^T f(t,s)ds) \\ &= E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] \neq \exp(-\int_t^T r(s)ds) \end{aligned}$$

Note that $\exp(-\int_t^T f(t,s)ds)$ is a scalar whereas $\exp(-\int_t^T r(s)ds)$ is a random variable.

However, under risk-neutral measure, we have

$$P(t,T) = g(t,r(t),T) = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t].$$

Lemma 19.1.4 (relationship between yield, zero curve and forward curve). *For a given current time t, each of the curves $P(t,T), f(t,T), R(t,T), F(t,T_1,T_2)$ have the following relationships:*

- $$P(t,T) = \exp(-R(t,T)(T-t)) = \exp(-\int_t^T f(t,s)ds).$$
- $$F(t,T_1,T_2) = \frac{1}{T_2-T_1}(\frac{P(t,T_2)}{P(t,T_1)} - 1) = \frac{1}{T_2-T_1}(\exp(\int_{T_1}^{T_2} f(t,s)ds) - 1).$$

Proof. (1) Based on the definition of yield, zero curve and instantaneous forward rate. (2) Note that from Theorem 19.1.4, we have

$$F(t,T_1,T_2) = \frac{1}{T_2-T_1}(\frac{P(t,T_2)}{P(t,T_1)} - 1).$$

□

Definition 19.1.11 (cash account, money market account). *The **cash account** is a special security that earns short rate interest, we denote the value by $B(t)$ with $B(0)=1$. In continuous time model, we have*

$$B(t) = \int_0^t e^{r(s)}ds$$

In discrete time model, we have

$$B(n) = \prod_{i=1}^n (1+r_i)$$

Remark 19.1.12 (interpretation)**.**

- Note that $B(t)$ is a **random variable** because the short rate $r(t)$ is random.
- The cash account/money market account can be thought of as the amount earned by starting with a unit dollar at time 0 and continuously reinvesting it at the short rate over infinitesimal time interval $[t, t+\delta t]$

Remark 19.1.13 (short rate contains less information than zero-coupon curve)**.**

- Short rate alone cannot determine the price of the zero-coupon bond.
- Money account and zero-coupon bond price have non-trivial relationship.

19.1.3 Coupon-bearing bonds

Definition 19.1.12 (fixed-coupon bond). *[1, p. 10] A* ***fixed-coupon bond*** *is a financial instrument that pays the holder deterministic(known at time $t \leq T_0$) amounts $c_1, ..., c_n$, referred to as* ***coupon rate****, at times $T_1, ..., T_n, T_0 < T_1 < \cdots < T_n$. At maturity time T_n, the holder receives the face value N and the final coupon c_n.*

Lemma 19.1.5 (present value of fixed coupon bonds). *The present value for a coupon-bearing bond is given by*

$$B_{fixed}(t) = \sum_{i=1}^{n} c_i P(t, T_i) + NP(t, T_n).$$

where c_i is the coupon rate and N is the face value.

Proof. Note that the present no-arbitrage price for a fixed payment c_i at time T_i is $c_i P(t, T_i)$. □

Definition 19.1.13 (floating rate note). *[1, p. 10] Let $0 \leq T_0 < T_1 < \cdots < T_n$ be a set of dates. A* ***floating rate note*** *is a financial instrument that pays the holder stochastic(unknown at time $t \leq T_0$) amounts $\tau_1 L(T_0, T_1), ..., \tau_n L(T_{n-1}, T_n) \times N$ at times $T_1, ..., T_n, T_0 < T_1 < \cdots < T_n$. At maturity time T_n, the holder receives the face value N and the final coupon $\tau_n L(T_{n-1}, T_n) \times N$.*

Remark 19.1.14 (difference between fixed coupon bonds and floating rate note). At time $t \leq T_0$, the coupon payment at future times is known for coupon-bearing bonds, whereas unknown for floating rate note(since $L(T_{i-1}, T_i)$ is unknown at t).

Lemma 19.1.6 (valuation of a floating-rate note). *Let $0 \leq T_0 < T_1 < \cdots < T_n$ be a set of dates. Consider a flating rate note that pays float coupon $\tau_i L(T_{i-1}, T_i) \times N$ at $T_i, i = 1, 2, ..., n$ and pays face value N at T_n. Let current time be $t_0, t_0 \leq T_0$. It follows that*

- *The present value for a floating-rate note is given by*

$$B_{floating}(t_0) = NP(t, T_0).$$

  Particularly if $t_0 = T_0$, then

$$B_{floating}(T_0) = N.$$

- *(value evolution)* ***Assume interest rate curve remain constant.***

- *Consider a future time t. Immediately after coupon payment date $T_1, T_2, ..., T_{n-1}$, the value of the note is given by*

$$B_{floating}(T_i) = N.$$

- *For $T_{i-1} < t < T_i$, the value is*

$$B_{floating}(t) = N \times DF(t, T_i)(1 + \tau_i F(t_0, T_{i-1}, T_i)).$$

Proof. (1)Use the result that the present value of $\tau L(T_{i-1}, T_i) = P(t, T_{i-1}) - P(t, T_i)$.

$$B_{floating}(t) = N \sum_{i=1}^{n} (P(t, T_{i-1}) - P(t, T_i)) + NP(t, T_n) = NP(t, T_0)$$

(2) Straight forward. □

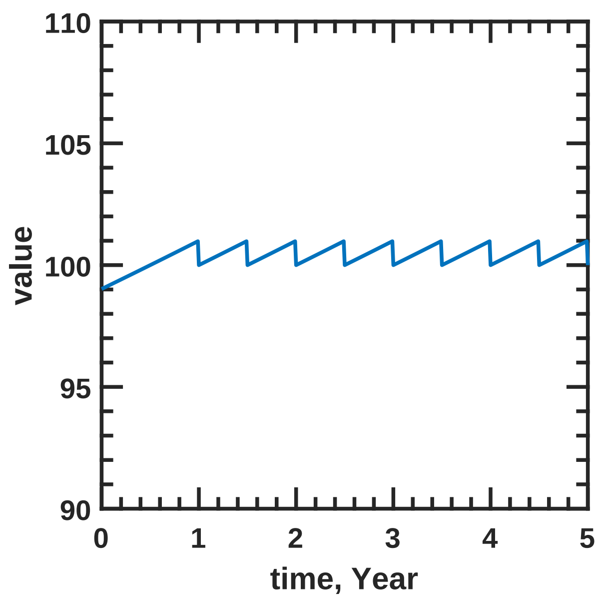


Figure 19.1.2: Value evoluation of the floating rate note with maturity 5Y. Coupon payment dates are 1Y, 1.5Y, ..., 5Y. The interest rate curve is flat with effective annual rate of 2%.

19.1.4 Interest rate swap

19.1.4.1 *The swap contract*

Definition 19.1.14 (interest rate swap). *[1, p. 12]*

- *An **interest rate swap** is an OTC instrument in which two counterparties exchange a set of payments at a fixed rate of interest for a set of payments at a floating rate, typically the LIBOR.*
- *If the holder is paying a floating rate and receiving the fixed rate, the swap is said to be a **receiver swap**. If the holder is receiving a floating rate and paying the fixed rate, the swap is said to be a **payer swap**.*
- *Consider a unit notional amount $N = 1$ and a set of dates $T_0 < T_1 < ... < T_n$ with $\tau_i = T_i - T_{i-1}, i = 1, ..., n$. At time T_i, the holder of a payer swap pays a fixed amount $\tau_i K$, where K is the **swap rate**, in exchange for a floating payment of $\tau_i L(T_{i-1}, T_i)$, where $L(T_{i-1}, T_i)$ is the spot LIBOR rate(a random quantity) in future time interval T_{i-1} and T_i.*
- *The payment dates $T_1, ..., T_n$ are **settlement dates**, and dates $T_0, ..., T_{n-1}$ are **reset dates.** The first reset data T_0 is called the **start date** of the swap.*
- *If current time $t < T_0$, the swap agreement is called a **forward-starting** swap; If current time $t = T_0$, the swap agreement is called **spot-starting** swap.*

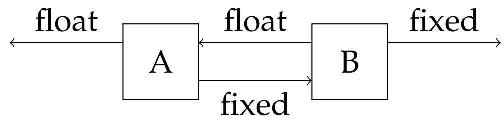
Example 19.1.2 (application of forward starting swap). For example(see Figure 19.1.3a), company B takes a loan for $400 million at a fixed interest rate and company A takes a loan for $400 million at a floating interest rate. Company B expects that the rate six months in future will decline and therefore wants to convert its fixed rate into a floating one. But because the rate changes are not expected to happen right away, the company just wants to lock in the swap rate for later. On the other hand, company A believes that interest rates will increase six months in the future. It does not want to convert into a fixed rate loan right away but wants to protect itself by locking in the rate now. The two companies may enter into forward starting interest rate swap to hedge their risk.

Remark 19.1.15 (application of swaps in debt issuance). During a public offering of corporate bonds to the public, investors usually prefers floating rate bonds, whereas corporates prefer issuing fixed rate bonds. The corporate will enter a swap with a bank, in which the corporate will pay fixed coupon to the bond and pay investors the float rate coupon received from the bank.

Remark 19.1.16 (application of swaps). link

- **Portfolio management.** Interest rate swaps allow portfolio managers to adjust interest rate exposure to the yield curve shape change (parallel move, twisting, steepening, etc, but not the implied volatility of yield curves). By increasing or decreasing interest rate exposure in various parts of the yield curve using swaps, managers can either ramp-up or neutralize their exposure to changes in the shape of the curve,

(a) payment flow after entering the swap



(b) payment flow after entering the swap

Figure 19.1.3: Use swap to exchange payment flow

and can also express views on credit spreads. Swaps can also act as substitutes for other, less liquid fixed income instruments. Moreover, long-dated interest rate swaps can increase the duration of a portfolio, making them an effective tool in Liability Driven Investing, where managers aim to match the duration of assets with that of long-term liabilities.

- **Speculation.** Because swaps require little capital up front, they give fixed income traders a way to speculate on movements in interest rates while potentially avoiding the cost of long and short positions in Treasuries. For example, to speculate that five-year rates will fall using cash in the Treasury market, a trader must invest cash or borrowed capital to buy a five-year Treasury note. Instead, the trader could 'receive' fixed in a five-year swap transaction, which offers a similar speculative bet on falling rates, but does not require significant capital up front.
- **Corporate debt issuance.** During a public offering of corporate/municipal bonds to the public, investors usually prefers floating rate bonds, whereas corporates/municipalities prefer issuing fixed rate bonds. The corporate/municipalities will enter a swap with a bank, in which the corporate/municipalities will pay fixed coupon to the bond and pay investors the float rate coupon received from the bank.
- **Corporate finance.** Firms with floating rate liabilities, such as loans linked to LIBOR, can enter into swaps where they pay fixed and receive floating, as noted earlier. Companies might also set up swaps to pay floating and receive fixed as a hedge against falling interest rates. Companies might also set up swaps pay floating and receive fixed to reduce interest rate exposure if floating rates more closely match their assets or income stream.
- **Risk management.** Banks and other financial institutions are involved in a huge number of transactions involving loans, derivatives contracts and other investments. The bulk of fixed and floating interest rate exposures typically cancel each other out, but any remaining interest rate risk can be offset with interest rate swaps.
- **Rate-locks on bond issuance.** When corporations decide to issue fixed-rate bonds, they usually lock in the current interest rate by entering into swap contracts. That

gives them time to go out and find investors for the bonds. Once they actually sell the bonds, they exit the swap contracts. If rates have gone up since the decision to sell bonds, the swap contracts will be worth more, offsetting the increased financing cost.

Remark 19.1.17 (relation to coupon-bearing bonds). An interest rate swap can be viewed as longing a fixed coupon bond and shorting a floating-rate note.

19.1.4.2 *Valuation of a swap and swap rate*

Lemma 19.1.7 (value of an interest rate swap). *Let the current time be $t \leq T_0$. consider a swap with unit notional.*

- *The current value of the fixed leg is*

$$V_{fixed}(t) = K \sum_{i=1}^{n} \tau_i P(t, T_i).$$

- *The current value of the floating leg is*

$$V_{float}(t) = \sum_{i=1}^{n} (P(t, T_{i-1}) - P(t, T_i)) = P(t, T_0) - P(t, T_n).$$

  Particularly for spot-starting swaps such that $t = T_0$,

$$V_{float}(t) = 1 - P(t, T_n).$$

- *The current value of a payer swap contract is*

$$V(t, K) = V_{float}(t) - V_{fixed}(t) = P(t, T_0) - P(t, T_n) - K \sum_{i=1}^{n} \tau_i P(t, T_i).$$

Proof. (1) Straight forward from cash flow discounting. (2) Note that from Theorem 19.1.2, we have $\tau_i L(T_{i-1}, T_i) = P(t, T_{i-1}) - P(t, T_i)$. For future payment $K\tau_i$, its present value at time t is $K\tau_i P(t, T_i)$. □

Definition 19.1.15 (forward swap rate). *[1, p. 13] Let the current time be t. The* ***forward swap rate****, denote by $S(t; T_0, T_n)$, associated with a swap exchanging cash flows with specified frequency from T_0 to T_n ($t \leq T_0$), is given by*

$$S(t; T_0, T_n) = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)},$$

such that the time t value of the forward swap zero.

Note that

- *The quantity $\Delta = T_n - T_0$ is known as* ***swap tenor****.*
- *The forward swap rate $S(t; T_0, T_n)$ is a known number since we know the current term structure $P(t,T), T \in \mathbb{R}^+$. For any future time $s > t$, $S(s; T_0, \Delta)$ is an unknown random variable since $P(s,T), T \in \mathbb{R}^+$ is random.*

Remark 19.1.18 (application of forward swap rate in calibration)**.** In reality, swap rates are traded benchmark securities. Since we have expressed the swap rate in terms of the zero-coupon bond prices, we can determine the zero-coupon curve from market swap rates.

Lemma 19.1.8 (swap rate as the sum of weighted forward rate). *Consider a swap with underlying tenor $\Delta = T_n - T_0$. Let the current time be t.*

- *The forward swap rate given by*
$$S(t; T_0, T_n)(t) = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)},$$
can also be written as
$$S(t; T_0, T_n) = \sum_{i=1}^{n} w_i(t) F(t; T_{i-1}, T_i)$$
where
$$w_i(t) = \frac{\tau_i P(t, T_i)}{\sum_{j=1}^{n} P(t, T_j)}.$$
- *If this is a* ***continuous swap*** *where the exchange of cash flow occurs within $[T_0, T_1]$. Then*
$$S_c(t; T_0, T_n) = \frac{\int_{T_0}^{T_n} f(t,s) P(t,s) ds}{\int_{T_0}^{T_n} P(t,s) ds},$$
where $f(t,s)$ is the instantaneous forward rate observed on current time t.

Proof. (1) Note that the present value of a interest rate swap is given by

$$\begin{aligned} V(t,K) &= \sum_{i=1}^{n} PV(\tau_i L(T_{i-1}, T_i)) - K \sum_{i=1}^{n} \tau_i P(t, T_i) \\ &= \sum_{i=1}^{n} \tau_i F(t; T_{i-1}, T_i) P(t, T_i) - K \sum_{i=1}^{n} \tau_i P(t, T_i), \end{aligned}$$

where we use the result in Theorem 19.1.2. Set the present value to zero and we also get the result. (2) For a continuous swap, the fixed leg has value

$$K\int_{T_0}^{T_n} P(t,s)ds,$$

and the floating leg has

$$\int_{T_0}^{T_n} f(t,s)P(t,s)ds,$$

where we can view the floating leg is paying $f(t,s)$ in every time point. $\square$

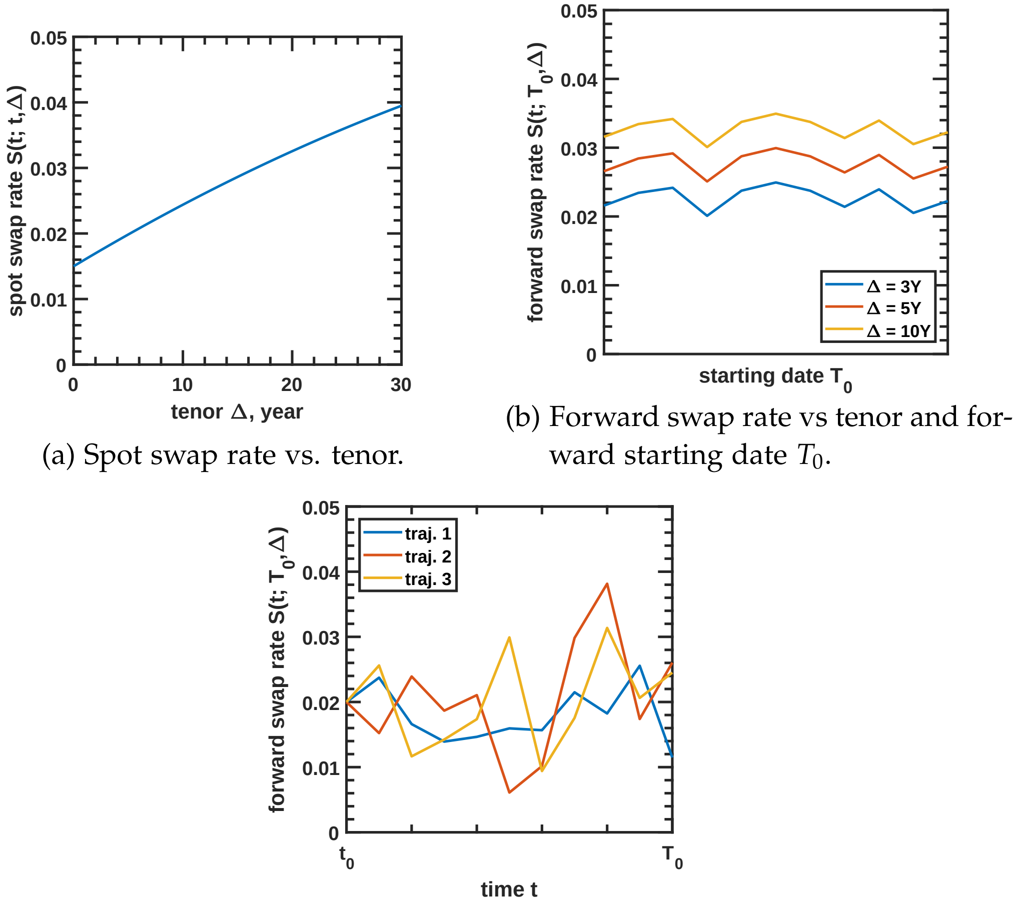
Lemma 19.1.9 (mark-to-market value of a swap contract). *Consider an interest rate payer swap starting at T_0 with tenor Δ entering at $t_0 \leq T_0$.*

- *During future time $t, t_0 \leq t \leq T_0$, the value of the swap contract $V(t)$*

19.1.4.3 *Understand swap curves*

Note 19.1.2 (different swap rate curve). The swap rate $S(t;T_0,T_N), \Delta = T_N - T_0$ is characterized by three parameters. Varying these parameters can yield different curves, as showed in Figure 19.1.4.

- Let the current time be t. The "**spot swap curve**" refers to an x-y chart of par swap rates $S(t;t,t+\Delta)$ plotted against their tenor Δ.
- If we fix current time t and tenor Δ, we can get a swap rate $S(t;T_0,T_0+\Delta)$ vs. starting date T_0. This curve is like the forward curve of libor rates that can be observed at the current market.
- Consider a swap contract with fixed starting date T_0 and tenor Δ. Let current time be t_0. For any time $t_0 \leq t \leq T_0$, $S(t;T_0,T_0+\Delta)$ is an unknown random quantity.



(a) Spot swap rate vs. tenor.

(b) Forward swap rate vs tenor and forward starting date T_0.

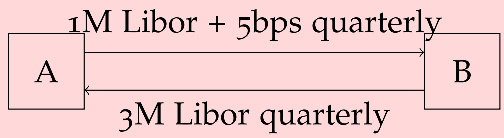
(c) The evolution of forward swap rate $S(t; T_0, \Delta)$ in different realizations.

Figure 19.1.4: Swap rate concept demo.

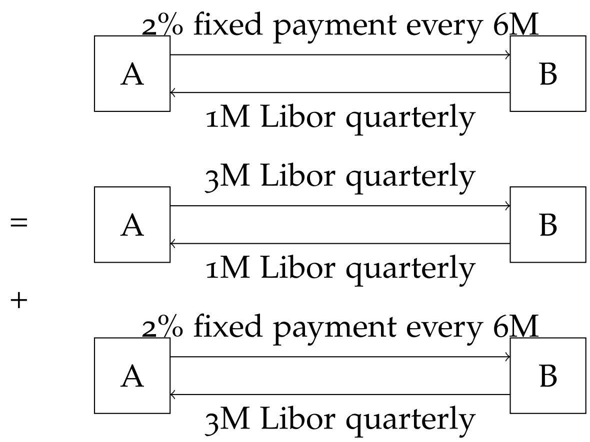
19.1.4.4 *Basis swap*

Definition 19.1.16 (basis swap). *[2, p. 28] A basis swap is an interest rate swap consisting of two floating legs, which are tied to different indices. To make a swap have zero value at starting date, a spread is added to one of the floating leg.*

Example 19.1.3. Consider a 2-year basis swap in the following figure that one counterparty pays 1-month LIBOR plus 5 basis point, and the other pays 3-month LIBOR. The two party will exchange the cash flow quarterly money for two years.



Remark 19.1.19 (use basis swap to construct other fixed-float swaps). We can use basis swap to construct new fixed-float swaps that are not available on the market. As we show in the following figure, we can use a 3M-1M basis swap and a regular 3M-fixed swap to construct a 1M-fixed swap.



19.1.5 Other common swaps and indices

19.1.5.1 *CMS and CMS index*

Definition 19.1.17 (constant-maturity swap, CMS). *[1, p. 133][3, p. 557] Take a unit notional amount and a set of dates $t < T_0 < T_1 < \cdots < T_n < \cdots < T_{m+n}$ with accrual periods $\tau_i = T_i - T_{i-1}$. A **constant-maturity swap(CMS)** involves a series of payments of an amount $\tau_{i+1} S_{i,i+m}(T_i)$, where $S_{i,i+m}(T_i)$ is the forward swap rate setting at at T_i and maturing at T_{i+m}, made in exchange for a fixed amount $\tau_{i+1}K$. These payments occur at times $T_{i+1}, i = 0, ..., n-1$.*

Remark 19.1.20 (CMS swap vs. ordinary interest rate swap).

- In the ordinary interest rate swap, the future payment for a period is a quantity related to LIBOR rate $L(T_i, T_{i+1})$(which is a random quantity, but its expected value under the T_{i+1} forward measure is $F(t, T_i, T_{i+1})$).
- In the CMS swap,the future payment for a period is a quantity related to the swap rate of a constant tenor swap(which is a random quantity, but its expected value under the annuity measure is the current swap rate.)

- A CMS swap is like a swap on swap rate, while an ordinary interest rate swap is a swap on LIBOR interest rate.

Remark 19.1.21 (business application). link The Constant Maturity Swap is employed by two types of users: 1. Investors or institutions attempting to hedge or exploit the yield curve while seeking the flexibility that the swap will provide and 2. Investors or institutions seeking to maintain a constant liability duration or constant asset.

The advantages of a constant maturity swap are:

It maintains a constant duration The user can determine âĂIJconstant maturityâĂİ as any point on the yield curve It can be booked the same way as an interest rate swap The disadvantages of a constant maturity swap are:

It requires documentation from the International Swaps and Derivatives Association (ISDA) It has the potential for unlimited losses.

Comparison betweee

-

Remark 19.1.22 (comparison between cap, swap, and CMS swap in hedging). Cap and swaps are usualy indexed to LIBOR 3M, LIBOR 1M, BMA etc., i.e., short-term interest rates.

- A cap with long life (e.g., the cap consisting of 40 caplets with a total coverage of 10 Year) can hedge long-term exposure to short-term interest rate risk. A 10Y cap indexed to LIBOR 3M can hedge 3M interest rate risk for 10 years.
- Similarly, a swap with long life can also hedge long-term exposure to short-term interst rate risk. A 10Y swap indexed to LIBOR 3M can hedge 3M interest rate risk for 10 years.
- A N-year-life CMS swap indexed to M-year can hedge N year exposure to M-year interest rate risk.

19.1.5.2 *BMA swap and BMA index*

Definition 19.1.18 (BMA swap and BMA index).

- *A Bond Market Association (BMA) Swap is a type of swap arrangement in which two parties agree to exchange interest rates on debt obligations, where the floating rate is based on the U.S. SIFMA Municipal Swap Index. One of the parties involved will swap a fixed interest rate for a floating rate, while the other party will swap a floating rate for a fixed rate.*
- *US BOND MARKET ASSOCIATION MUNICIPAL SWAP INDEX (BMA INDEX) This index is produced weekly, reflecting the average rate of issues of tax-exempt*

variable-rate debt, and serves as a benchmark floating rate in municipal swap transactions. The BMA index is usually 65%-70% of its taxable equivalent, the 1-month LIBOR.

19.1.6 Negative interest rates

Remark 19.1.23 (why central bank set interest rate). link

- Central banks in many developed countries have set key interest rates below zero, and as a result, portions of the yield curves in these countries have dropped to negative levels.
- Although individuals are not paying banks to hold their money, negative interest rates imposed by a central bank effectively mean commercial banks are required to pay for holding excess reserves with the central bank. For example, if the deposit rate were âĂŠ1%, for every $10 million held with the central bank, the commercial bank would have a balance of around $9.9 million at the end of a year.
- In general, the purpose of negative interest rate policy is to boost the economy. The theory is that commercial banks will be dissuaded from maintaining large balances with the central bank and will instead lend money to businesses and consumers who will, in turn, spend the money. The increase in lending and spending is likely to boost economic activity, leading to growth and inflation.
- For Switzerland, Denmark and Sweden, the rationale for lowering policy rates below zero had more to do with their currencies and the associated exchange rates than credit creation. The objective was to put downward pressure on the currency in order to stimulate trade by making exports cheaper and imports more expensive.

Remark 19.1.24 (why investors invest in negative yield bonds).

- **Currency speculation**: An investor may make an investment into a negative-yield bond as a proxy for a currency investment if they think the currency they will be repaid in will appreciate in value in an amount that exceeds the negative yield of the bond.
- **Speculation rolling yield:** Even though the yield curve is partial negative, or completely negative, as long as the yield curve has upward sloping shape, the investors can generate positive gain by buy long-dated bonds, and sell them before maturity.
- **Speculating further decreasing yield:** Investors can also generate positive return by speculating when the yield further decreases.
- **Regulatory Requirement:** Some investors have to compulsorily hold Govt bonds. These would be banks, central banks, insurance cos, etc. They do not have an option. They are required to purchase certain types of relatively safe assets, regardless of their yield.

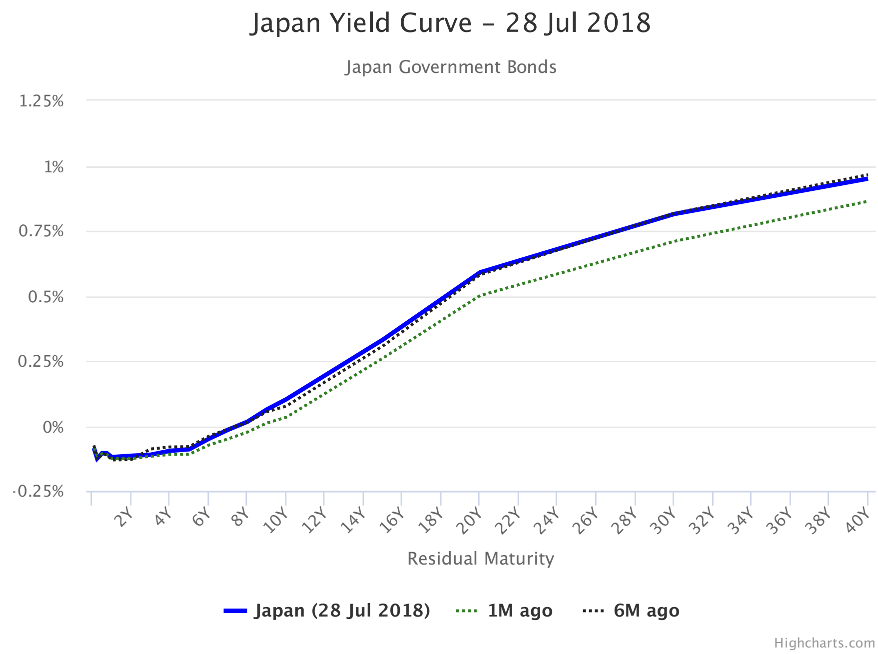


Figure 19.1.5: Japan government bond yield curve as of 2018-7-28. Data source: http://www.worldgovernmentbonds.com/country/japan/

- **Risk-averse institutional investosr:** Then there are investors who think the world will fall apart and feel the safest to keep their money with the Govt. They are ok to lose 0.5% of their capital, then to lose a lot more elsewhere. Consider an institution that has to be risk averse - when they have a lot of money they cannot stuff cash under a mattress and have rats eat it, nor you trust a bank deposit enough (banks default) and other bonds are too risky.

19.2 Common interest rate derivatives

19.2.1 Options on bonds

Definition 19.2.1 (vanilla call option on zero-coupon bond). *A call option is the right to buy a bond at time S which matures at $T > S$. The call option has strike K and expiry $S > t$. The payoff of the call option at expiry S is*

$$\max(P(S,T) - K, 0).$$

Remark 19.2.1. The goal is to determine the price of the call option at $t < S < T$.

19.2.2 Caps and floors

Definition 19.2.2 (cap). *[1, p. 33] A **cap** is a portfolio of n call options on LIBOR. Consider a unit notational amount and a set of dates $T_0 < T_1 < \cdots < T_n$ with $\tau_i = T_i - T_{i-1}, i = 1, 2, ..., n$. At time T_i, the holder of an interest rate cap receives $\tau_i \max(L(T_{i-1}, T_i) - K, 0)$, where K is the **cap rate**. Each of these n call options are known as **caplets**.*

The ith caplet is a European call option with expiry T_i written on the spot LIBOR rate $L(T_{i-1}, T_i)$(a random quantity) with payoff

$$(T_i - T_{i-1}) \max(L(T_{i-1}, T_i) - K, 0).$$

Definition 19.2.3 (floor). *A **floor** is a portfolio of n put options on LIBOR. Consider a unit notational amount and a set of dates $T_0 < T_1 < \cdots < T_n$ with $\tau_i = T_i - T_{i-1}, i = 1, 2, ..., n$. At time T_i, the holder of an interest rate floor receives $\tau_i \max(K - L(T_{i-1}, T_i), 0)$, where K is the **floor rate**. Each of these n put options are known as **floorlets**.*

The ith floorlet is a European call option with expiry T_i written on the spot LIBOR rate $L(T_{i-1}, T_i)$ with payoff

$$(T_i - T_{i-1}) \max(K - L(T_{i-1}, T_i), 0).$$

Remark 19.2.2 (interpretation applications of caps and floors). [2, p. 135]

- A caplet is a call on rates. Its payoff on a 3M LIBOR is given by

$$(L(T, T + 3M) - K)^+.$$

  The buyer of a cap is protecting against rates moving higher.

- A floorlet is a put on rates. Its payoff on a 3M LIBOR is given by
$$(K - L(T, T + 3M))^+.$$
The buyer of a floorlet is protecting against rates moving higher.
- Caps can be used by borrowers to hedge against increasing interest rate; floors can be used by lenders to hedge against decreasing interest rate.
- For example, a borrower who is paying the LIBOR rate on a loan can protect himself against a rise in rates by buying a cap at 2.5%. If the interest rate exceeds 2.5%, the extra money used to repay the loan is compensated by the increasing value of the cap.Therefore, the interest payments are effectively "capped" at 2.5% from the borrowers' point of view.
- Mortgage servicers are exposed to prepayment risk when interest rate is decreasing. Therefore, mortgage servicer can purchase floors on 10Y CMS to hedge the prepayment risk.

19.2.3 Swaption

19.2.3.1 *Basic concepts*

Definition 19.2.4 (swaption). *[1, p. 35] Consider an interest rate swap with payment dates $T_1, T_2, ..., T_n$ and reset dates $T_0, T_1, ..., T_{n-1}$.*

- *A **payer(receiver) swaption on spot-starting swaps** gives the holder the right to enter into an interest rate swap with maturity date $T_s = T_0$. Note that T_0 is the first reset date of the underlying swap. T_0 is also the **expiry date** of the swaption.*
- *A **payer(receiver) swaption on forward-starting swaps** gives the holder the right to enter into an interest rate swap with maturity date $T_s < T_0$. Note that T_0 is the first reset date of the underlying swap. T_0 is also the **expiry date** of the swaption.*

Remark 19.2.3 (summary and interpretation).

- A payer swaption is an option to enter into a payer swap (the swap is paying fixed rate) at a later date.
- A receiver swaption is an option to enter into a receiver swap (the swap is receiving fixed rate) at a later date.
- A payer swaption is like a call option on forward swap rate.
- A receiver swaption is like a put option on forward swap rate.

Remark 19.2.4 (business application of swaption). Swaptions are used to hedge issuance of bonds or to hedge call features in bonds

Remark 19.2.5 (comparison between swaptions and cap/floor). [2, p. 178]

- A buyer of a cap is buying a series of options. For example, the buyer of a 5-year cap on 3-month LIBOR is buying a series of 20 different call options.
- Instead, a buyer of a payer swaption, on the other hand, is buying a single option. For example, a 5 into 10 payer swaption will protect the buyer from an increase in 10-year swap rates, five years from now.
- The choice of which option is more appropriate depends on the needs of the particular investor. If, for example, the investor manages a bank portfolio that has a liability tied to 3-month LIBOR resetting quarterly over the next five years, then it would seem more appropriate for the bank to purchase a 5-year cap on 3-month LIBOR instead of purchasing a swaption. If, however, the investor manages a mortgage portfolio that has exposure to 10-year swap rates increasing, say, five years from now, then it would appear more appropriate for the investor to buy a 5 into 10 payer swaption rather than a cap. (Note that the market does not have call options on 10 year spot rate.)

Definition 19.2.5 (Bermudan interest rate swaption). *[1, p. 59] Consider a unit notional amount and a set of dates $0 < T_0 < T_1 < ... < T_n$. The holder of* ***a payer (or receiver) Bermudan swaption*** *with strike K has the right to enter a payer (or receiver) interest rate swap at any time T_k for $k = 0, ..., l, l < n$. The underlying swap has reset dates $T_0, ..., T_{n-1}$ and settlement dates $T_1, ..., T_n$ and fair swap rate K.*

19.2.3.2 *Swaption market basics*

19.2.4 Cancellable swap

Definition 19.2.6 (Cancellable swap).

- *An European cancelable payer swap is a payer swap in which the fixed rate payer has the right to terminate the trade at a given date during the swap period prior to its maturity. A Bermudian cancelable payer swap is a payer swap in which the fixed rate payer has the right to terminate the trade at a set prespecified dates prior to its maturity.*
- *More formally, let $T_S = T_0 < T_1, ..., T_n < T_E$ be a set of dates for resetting and coupon payments. A cancellable swap is paying a fixed rate K and receiving a float rate over period T_S and T_E and has cancellable date $T_{C,i}, T_S \leq T_{C_i} \leq T_E, i = 1, 2, ..., K$.*
-

Example 19.2.1. [2, p. 233] Consider a 5-year swap in which the fixed rate payer has the right to cancel the swap after one year. If the fixed rate payer does cancel the swap after one year, then both counterparties will make final coupon payments at the end of the first year, and no further cash flows will be made in the swap. If the fixed rate payer does not cancel the trade at the end of the first year, then the swap will remain outstanding for the remaining four years.

Lemma 19.2.1 (static replication of an European cancellable swaps).

- *Consider an European cancelable swap with start date, end date, cancelable date and fixed rate given by T_S, T_E, T_C, K. Its payoff at cancellation date T_C is*

$$\max(V(T_C, T_C, T_E, K), 0),$$

where $V(T_C, T_C, T_E, K)$ is the time-T_C value of a payer swap, with strike K, starting at T_C and ending at T_E.
- *A long position in cancellable swap can be replicated by a long position payer swap and a long position of receiver European swapation.*
Particularly, at the cancellation date T_C, the payoff of the cancellable swap is

$$\max(V(T_C, T_C, T_E, K), 0),$$

and the replicating portfolio has payoff

$$\underbrace{V(T_C, T_C, T_E, K)}_{\text{swap value}} + \underbrace{\max(-V(T_C, T_C, T_E, K), 0)}_{\text{receiver swaption}} = \max(V(T_C, T_C, T_E, K), 0).$$

19.2.5 Eurodollar futures

Definition 19.2.7 (Eurodollar futures). *[4, p. 171][5, p. 141]CMEgroup*

- *Eurodollar futures are cash-settled futures contracts with final futures price based on three-month(or one-month) LIBOR at the expiration date:*

$$P_{futures}(T_0; T_0, T_1) = 100(1 - L(T_0, T_1)),$$

where $L(T_0, T_1)$ is the Libor rate at T_0 for tenor $T_1 - T_0$(usually $T_1 = T_0 + 0.25$ or $T_1 = T_0 + 1/12$, hence the tenor is 3-month or 1-month).See Table 19.2.1.
- *(**quoting convention**) At $t < T_0$, the futures price is quoted as*

$$P_{futures}(t; T_0, T_1) = 100(1 - Fur(t, T_0, T_1))$$

where $Fur(t, T_0, T_1)$ is the futures rate. Therefore, the implied forward interest rate is

$$Fur(t, T_0, T_1) = 1 - \frac{P_{futures}}{100}.$$

- *For example, a price of 95.00 implies an 3M forward libor interest rate of 5%.*
- *A Eurodollar futures contract usually has notational amount 250000 such that the change of one basis point in Fur will lead to change of 25$.*

Table 19.2.1: 3-month Eurodollar futures contract schedule

contract month	**delivery/maturity/expiry date**
Mar 2018	19-Mar-2018
Apr 2018	16-Apr-2018
May 2018	14-May-2018
...	...

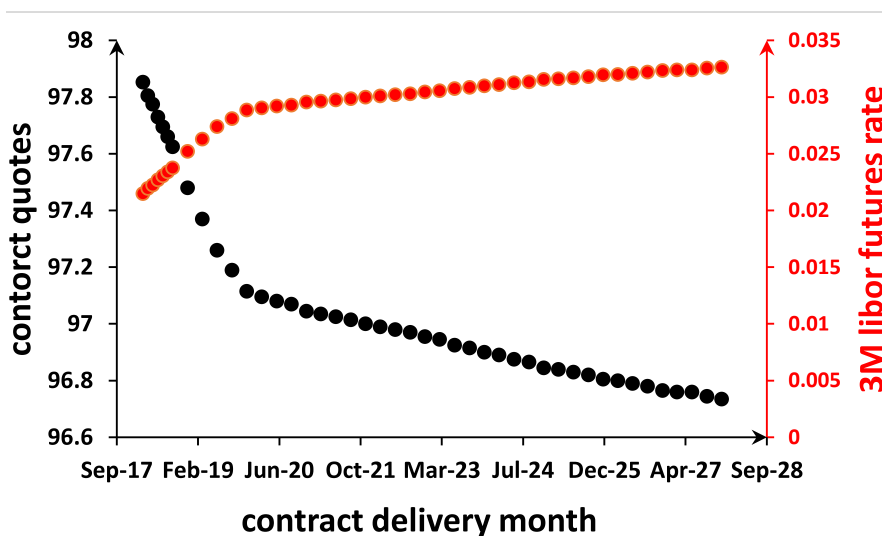


Figure 19.2.1: 3M libor futures rate curve as of Mar,12,2018. Data source http://www.cmegroup.com/trading/interest-rates/stir/eurodollar.html

Example 19.2.2. For example, suppose an investor buys a single three-month Mar2018 contract at 95.00 (implied settlement LIBOR of 5.00%) on Feb, 1, 2018. Then starting from Feb-1-2018 to the delivery/maturity date 19-Mar-2018, there will be cash flow exchange between the investor and the CME exchange. The amount of the exchange is dependent on market 3M libor rate and the contract specified libor rate(here is 5%).

- if at the close of business on Feb-1-2018, the contract price has risen to 95.01 (implying a LIBOR decrease to 4.99%), US$25 will be paid into the investor's margin account;
- if at the close of business on that day, the contract price has fallen to 94.99 (implying a LIBOR increase to 5.01%), US$25 will be deducted from the investor's margin account.
- The process will continue until the end of Mar-19-2018.

19.3 Martingale pricing framework

19.3.1 Change of numeraire

19.3.1.1 *Principles*

Definition 19.3.1 (numeraire). *[1, p. 23] A numeraire is defined as any traded asset that pays no dividends and whose price $A(t)$ is positive at any time $t \geq 0$.*

Theorem 19.3.1 (equivalent martingale measure result, change of numeraire). *[5, p. 661] Consider a market having risk-free asset with short rate r and the market price of risk being σ_g associated with only source of uncertainty. Then under such measure, we have*

$$df = (r + \sigma_g\sigma_f)f dt + \sigma_f f dz$$

and

$$dg = (r + \sigma_g^2)g dt + \sigma_g g dz$$

Then

$$d(\frac{f}{g}) = (\sigma_f - \sigma_g)\frac{f}{g}dz,$$

that is, the quantity

$$\frac{f}{g}$$

is a martingale under the measure associated with market price of risk σ_g.

Proof.

$$d(\ln f) = (r + \sigma_f\sigma_g - \sigma_f^2/2)dt + \sigma_f dz$$

and

$$d(\ln g) = (r + +\sigma_f^2/2)dt + \sigma_f dz$$

Therefore

$$d(\ln f - \ln g) = -\frac{(\sigma_f - \sigma_g)^2}{2}dt + (\sigma_f - \sigma_g)dz$$

and

$$d(\frac{f}{g}) = (\sigma_f - \sigma_g)\frac{f}{g}dz$$

□

Remark 19.3.1. It can be showed that Lemma 15.5.1, under the market price of risk σ_g, we have

$$\frac{r+\sigma_g^2-r}{\sigma_g}=\frac{r+\sigma_g\sigma_f-r}{\sigma_f}=\sigma_g.$$

Corollary 19.3.1.1. *[5, p. 661]*

$$\frac{f_0}{g_0}=E_g(\frac{f_T}{g_T})$$

or

$$f_0=g_0E_g(\frac{f_T}{g_T})$$

where E_g denotes the expected value in a world that is forward risk-neutral with respect to g(i.e., the market price of risk is σ_g).

Remark 19.3.2 (how to do evaluation). Under the market price of σ_g, f_t and g_t are governed by the following SDEs given as

$$df=(r+\sigma_g\sigma_f)fdt+\sigma_f fdz$$

and

$$dg=(r+\sigma_g^2)gdt+\sigma_g gdz$$

or equivalently,

$$d(\frac{f}{g})=(\sigma_f-\sigma_g)\frac{f}{g}dz$$

We can solve to get the f_T/g_T and g_T.

19.3.1.2 *Choice of numeraire*

19.3.1.2.1 MONEY ACCOUNT AS NUMERAIRE

Corollary 19.3.1.2 (Money market account as the numeraire).

$$dB=rBdt,B(t)=B(0)\exp(\int_0^t r(s)ds)$$

where r ***can be stochastic****. Then*

$$f_0=B_0E_B(\frac{f_T}{B_T})=B_0E_B[\frac{f_T}{B_0\exp(\int_0^T rdt)}]=E_B[\exp(-\int_0^T rdt)f_T]$$

where E_B denotes the expected value in the traditional risk-neutral with respect to B(i.e., the market price of risk is $\sigma_g = 0$). If the short-term interest rate r is constant, then

$$f_0 = e^{-rT} E_B(f_T).$$

Remark 19.3.3 (how to evaluate expectation). Assume interest rate is not stochastic. Under the market price of $\sigma_B = 0$, f_t is governed by the following SDE given as

$$df = rf dt + \sigma_f f dz.$$

We can solve to get the B_T

Remark 19.3.4 (stochastic interest rate issue, theoretical method).

- Propose some short rate model on r and asset dynamics model on f(the dynamics of f should be under risk-neutral measure).
- Evaluate the expectation using **the joint distribution** of $\exp(-\int_0^T r(t)dt)$ and f_T.
- If the short rate dynamics and the asset dynamics do not have the common sources of uncertainty, then

  $$E_B[\exp(-\int_0^T r(t)ds) f_T] = E_B[\exp(-\int_0^T r(t)ds)] E_B[f_T] = P(0,T) E_B[f_T].$$

  However, this is impossible since f_t will have drift term including r under risk-neutral measure. This is also the motivation of using zero-coupon bond as the numeraire.

Remark 19.3.5 (stochastic interest rate issue, simulation method).

- Simulate the short term rate r trajectory using some assumed short rate model.
- For each simulated r trajectory:
  - calculate the expected payoff using $E_B[f_T]$
  - discount the expected payoff using $\exp(-\int_0^T rdt) E_B[f_T]$
- Average over many different trajectories of interest rate.

Lemma 19.3.1 (futures contract pricing under stochastic interest rate). *Assume the futures marginal account is cleared in continuous time. Then an asset $S(t)$ has a futures contract price given by*

$$Fu(t) = E_B[S_T | \mathcal{F}_t]$$

where the expectation E_B is taken with respect to the risk-neutral measure.

19.3.1.2.2 ZERO-COUPON BOND AS NUMERAIRE

Lemma 19.3.2 (pricing using Annuity as numeraire). *Assume there exists a risk-neutral measure such that the price of any traded asset X(without intermediate payments) relative to the money account $B(t)$ is a martingale under Q, i.e.,*

$$\frac{X_t}{B_t} = E_N[\frac{X_T}{B_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

Let $P(0,T)$ the new numeraire. Then ***there exists*** *a probability measure Q_T, equivalent to Q, such that the price of any attainable claim*

$$\frac{Y_t}{P(t,T)} = E_T[\frac{Y_T}{P(T,T)}|\mathcal{F}_t] = E_T[Y_T|\mathcal{F}_t], 0 \leq t \leq T.$$

Proof. Directly from Theorem 15.6.15. □

Remark 19.3.6 (how to evaluate the expectation and the dynamics of $P(t,T)$)**.** To evaluate the expectation, we usually first propose a dynamics model for $P(t,T)$, then use Theorem 19.3.1 to find out the dynamics of Y_T under this forward measure.

Theorem 19.3.2 (forward price of general assets). *[5, p. 663][1, p. 27] Consider* ***any*** *variable S that is not an interesting rate. A forward contract on S with maturity T is defined as a contract that pays off $S(T) - K$ at time T, where $S(T)$ is the value θ at time T. Therefore*

$$F = E_T[S(T)].$$

Or equivalently,

$$S(t) = P(t,T)F \Leftrightarrow F = \frac{S(t)}{P(t,T)} = E_T[\frac{S(T)}{P(T,T)}] = E_T[S(T)]$$

Proof. Define f_0 as the value of this forward contract at $t = 0$. Then

$$f_0 = P(0,T)(E_T[S(T)] - K),$$

where E_T is taking expectation with respect to the forward risk neutral with respect to $P(t,T)$. Note that the forward price F of $S(T)$ is the value of K such that $f_0 = 0$. Therefore

$$F = E_T[\theta_T].$$

Note that $P(t,T)F$ means the value of F discounted to today's value. And under forward measure

$$\frac{S(T)}{P(t,T)}$$

is a martingale(Theorem 15.6.18).

□

Remark 19.3.7 (interpretation).

- The forward price of arbitrary asset is a martingale under forward measure with respect to $P(t,T)$. In other words, the forward price can be regarded as an unbiased estimation of $X(T)$, where the expectation is taken **under the forward measure**.
- Suppose we work under the risk-neutral measure Q, the forward price $F(t,T)$ has to satisfy

$$V(t) = B(t)E_Q[\frac{X(T)-F(t,T)}{B(T)}|\mathcal{F}_t] = 0$$

where $B(t)$ is the cash account. Because now $B(T)$ is a random variable, and evaluate $E_Q[\frac{X(T)-F(t,T)}{B(T)}|\mathcal{F}_t]$ requires the joint distribution of $X(T)$ and $B(T)$, which can be difficult to calculate. Therefore, under forward measure, we greatly simplify the calculation.

Remark 19.3.8 (constant interest rate case,forward price of stock). Assume the short rate r is constant, then the dynamics of $P(t,T)$ is

$$dP(t,T) = rP(t,T)dt.$$

Under the forward measure with respect to $P(t,T)$, the dynamics of stock is

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

such that $E_T[S_T] = e^{rT}S_0 = F$.

19.3.1.2.3 Annuity as numeraire

Definition 19.3.2 (swap anuity). *[1, p. 12]*

*Given a set of dates $T_0 < T_1 < ... < T_n$ with accrual periods $\tau_i = T_i - T_{i-1}$. The **swap annuity** is defined as*

$$A_{0,n}(t) = \sum_{i=1}^{n} \tau_i P(t,T_i).$$

Note that $A_{0,n}(t)$ is a stochastic process.

Lemma 19.3.3 (pricing using Annuity as numeraire). *Assume there exists a risk-netural measure such that the price of any traded asset X(without intermediate payments) relative to the money account $B(t)$ is a martingale under Q, i.e.,*

$$\frac{X_t}{B_t} = E_N[\frac{X_T}{B_T}|\mathcal{F}_t], 0 \leq t \leq T.$$

Let $A_{0,n}(t)$ the new numeraire. Then ***there exists*** *a probability measure Q_A, equivalent to Q, such that the price of any attainable claim*

$$\frac{Y_t}{A_{0,n}(t)} = E_U[\frac{Y_T}{A_{0,n}(T)}|\mathcal{F}_t], 0 \leq t \leq T.$$

Proof. Directly from Theorem 15.6.15. □

Remark 19.3.9 (when to use it). The lemma will be useful to price derivatives with a payoff given by $A_{0,n}(T)V(T)$. Then we have

$$V(t) = A_{0,n}(t)E_A[V(T)|\mathcal{F}_t].$$

An example is at Lemma 19.10.8.

19.3.2 Martingale properties under forward measure

19.3.2.1 *Forward price under forward measure*

Theorem 19.3.3 (No-arbitrage forward price for general assets under stochastic interest rate). *[4, p. 112] Consider an asset S_t, its forward price at maturity date T is given by*

- $$F(t,T) = \frac{S_t}{P(t,T)}$$

  for zero dividend assets.

- $$F(t,T) = \frac{S_t \exp(-q(T-t))}{P(t,T)}$$

  for assets with continuous dividend rate q.

- $$F(t,T) = \frac{S_t - \sum_{i=1}^n q_i P(t,T_i)}{P(t,T)}$$

for assets with discrete dividends $\{q_i\}$ at time $\{T_i\}$; or equivalently,

$$F(t,T) = S_t + \underbrace{P(t,T)((T-t)F(t,t,T)}_{\textit{funding cost}} - \underbrace{\sum_{i=1}^{n} q_i(1+F(t,T_i,T)(T-T_i)}_{\textbf{\textit{benefits}}},$$

where $F(t,T_i,T)$ is the forward rate with tenor $[T_i,T]$.

Proof. (1) The seller of the contract will take the following strategy:

- enter the contract, which costs zero, as the seller party.
- short $S_t/P(t,T)$ units of $T-$maturity zero-coupon bond.
- long 1 unit of the asset at price S_t.
- At time T, sell the counterparty of the asset at forward price F and repay the short of the zero-coupon bond.

Since the initial cost/position is zero, the payoff for the seller at T is

$$0 = F - S_t/P(t,T)P(T,T) \implies F = \frac{S_t}{P(t,T)}.$$

(2)The seller of the contract will take the following strategy:

- enter the contract, which costs zero, as the seller party.
- short $S_t\exp(-q(T-t))/P(t,T)$ units of $T-$maturity zero-coupon bond.
- long $\exp(-q(T-t))$ unit of the asset at price S_t.
- the dividends from the stock will be reinvested to buy more shares.
- At time T, sell the counterparty of the asset at forward price F and repay the short of the zero-coupon bond.

Let ϕ be the number of shares, then

$$d\phi(t) = (q\phi(t)S_t)/S_t dt = qdt \implies \phi(t) = \exp(qt)\phi(0).$$

Then the total value of the asset at time T is given by

$$\phi(T)S_T = \phi(0)e^{q(T-t)}S_T = S_T.$$

We will sell S_T at price F at time T. Since the initial cost is zero, the payoff for the seller is

$$0 = F - S_t\exp(-q(T-t))/P(t,T) \implies F = S_t\exp(-q(T-t))/P(t,T).$$

(3)(a)The seller of the contract will take the following strategy:

- enter the contract, which costs zero, as the seller party.
- short q_i units of T_i-maturity zero coupon bonds, for $i = 1,2,...,n$.

- short $\frac{S_t - \sum_{i=1}^n q_i P(t,T_i)}{P(t,T)}$ units of $T-$maturity zero-coupon bond.
- long 1 unit of the asset at price S_t.
- The received dividends are used to repay the short of zero-coupon bonds at $T_i, \forall i = 1,2,...,n$.
- At time T, sell the counterparty of the asset at forward price F and repay the short of the zero-coupon bond.

Since the initial cost is zero, the payoff for the seller is

$$0 = F - \frac{S_t - \sum_{i=1}^n q_i P(t,T_i)}{P(t,T)} \implies F = \frac{S_t - \sum_{i=1}^n q_i P(t,T_i)}{P(t,T)}.$$

(b) The funding cost(borrow money to buy and therefore need to pay interest) is given by

$$P(t,T)((T-t)F(t,t,T).$$

The benefit are coupons paid at T_i and reinvest at forward rate locked at $F(t,T_i,T)$. □

Theorem 19.3.4 (forward price of general assets under forward measure is martingale). *[5, p. 663][1, p. 27] Consider* ***any*** *financial asset (no matter paying dividend or not) S that is not an interesting rate. A forward contract on S with delivery date T is defined as a contract that pays off $S(T) - K$ at time T, where $S(T)$ is the asset value at time T. Let $F(t,T)$ denote the forward price at time t. Then*

$$F(t,T) = E_T[S(T)|\mathcal{F}_t].$$

Proof. (1) From Theorem 19.3.3, we note that no-arbitrage condition requires that

$$F(t,T) = \frac{S_t - \sum_{i=1}^n q_i P(t,T_i)}{P(t,T)}$$

for assets with discrete dividends $\{q_i\}$ at time $\{T_i\}$.

Therefore, $F(t,T)$ is a martingale under T forward measure.

(2) (another method for non-dividend paying asset) Define V_t as the value of this forward contract at t. Then

$$V_t = P(t,T)(E_T[(S(T)-K)|\mathcal{F}_t],$$

where E_T is taking expectation with respect to the forward measure associated with $P(t,T)$. Note that the forward price $F(t,T)$ of $S(T)$ is the value of K such that $V_t = 0$. Therefore

$$F(t,T) = E_T[S(T)|\mathcal{F}_t].$$

□

Remark 19.3.10 (interpretation)**.**

- The forward price of arbitrary asset is a martingale under forward measure with respect to $P(t,T)$. In other words, the forward price can be regarded as an unbiased estimation of $S(T)$, where the expectation is taken **under the forward measure**.
- Suppose we work under the risk-neutral measure Q, the forward price $F(t,T)$ has to satisfy
$$V(t) = B(t)E_Q[\frac{X(T) - F(t,T)}{B(T)}|\mathcal{F}_t] = 0$$
where $B(t)$ is the cash account. Because now $B(T)$ is a random variable, and evaluate $E_Q[\frac{X(T)-F(t,T)}{B(T)}|\mathcal{F}_t]$ requires the joint distribution of $X(T)$ and $B(T)$, which can be difficult to calculate. Therefore, under forward measure, we greatly simplify the calculation.

Example 19.3.1 (constant interest rate case, forward price of stock). Assume the short rate r is constant, then the dynamics of $P(t,T)$ is
$$dP(t,T) = rP(t,T)dt.$$
Under the forward measure with respect to $P(t,T)$, the dynamics of stock is
$$dS_t = rS_t dt + \sigma S_t dW_t,$$
such that $E_T[S_T] = e^{rT}S_0 = F$.

19.3.2.2 *Forward bond and forward rate*

Lemma 19.3.4 (forward bond and forward rate under forward measure). *[1, p. 29] Let $t \leq S \leq T$. We have*

- *The forward bond price FP is a martingale under forward measure with respect to $P(t,S)$; that is,*
$$FP(t;S,T) \triangleq \frac{P(t,T)}{P(t,S)} = E_S[P(S,T)|\mathcal{F}_t].$$
- *The simply compounded forward rate $F(t;S,T)$ is a martingale under forward measure with respect to $P(t,T)$; that is,*
$$F(t;S,T) \triangleq \frac{1}{T-S}(\frac{P(t,S)}{P(t,T)} - 1) = E_T[F(t';S,T)|\mathcal{F}_t], t \leq t' \leq S.$$
- *In particular, if $t = S$, then $F(S;S,T) = L(S,T)$, we have*

$$F(t;S,T) = E_T[L(S,T)|\mathcal{F}_t].$$

Proof. (1) Under forward measure $P(t,S)$, the quantity $FP(t;S,T) \triangleq \frac{P(t,T)}{P(t,S)}$ is a martingale(Theorem 19.3.5). (2) Note that $\frac{P(t,S)}{P(t,T)}$ is martingale under forward measure with respect to $P(t,T)$. Adding and a constant -1 and multiplying a constant will not change its martingale nature. □

19.3.3 Dynamics under forward measure

19.3.3.1 *Dynamics under forward measure*

Lemma 19.3.5 (changing dynamics from risk-neutral measure to forward measure). *Assume under risk-neutral measure, we have*

$$\begin{aligned} dS_t/S_t &= rdt + \boldsymbol{\sigma}_S^T d\boldsymbol{W}(t) \\ dP(t,T)/P(t,T) &= rdt + \boldsymbol{\sigma}_P^T d\boldsymbol{W}(t) \end{aligned}$$

where $\boldsymbol{\sigma}_S$ and $\boldsymbol{\sigma}_P$ are volatility vectors, and $\boldsymbol{W}(t)$ is the Brownian motion vector.

Then the forward measure Q_T is generated via

$$Z = \frac{dQ_P}{dQ} = \exp(\int_0^t \boldsymbol{\sigma}_P^T d\boldsymbol{W}_s - \int_0^t \frac{1}{2}\boldsymbol{\sigma}_P^T \boldsymbol{\sigma}_P ds)$$

such that under which

- $$d\boldsymbol{W}_t = d\boldsymbol{W}_t^T + \boldsymbol{\sigma}_P dt.$$
- $$d(\frac{S_t}{P_t}) = \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_P)^T d\boldsymbol{W}_t$$

Proof. Using Ito rule, we have

$$\begin{aligned} d(\frac{S_t}{P_t}) &= \frac{dS_t}{P_t} - \frac{S_t dP_t}{P_t^2} - \frac{dS_t dP_t}{P_t^2} + \frac{dP_t^2}{P_t^3} \\ &= \frac{S_t}{P_t}(rdt + \boldsymbol{\sigma}_S^T d\boldsymbol{W}_t - rdt - \boldsymbol{\sigma}_P^T d\boldsymbol{W}_t) - \boldsymbol{\sigma}_S^T \boldsymbol{\sigma}_T dt + \boldsymbol{\sigma}_P^T \boldsymbol{\sigma}_T dt) \\ &= \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_P)^T (d\boldsymbol{W}_t - \boldsymbol{\sigma}_P dt). \end{aligned}$$

We can define a new measure via

$$Z = \frac{dQ_T}{dQ} = \exp(\int_0^t \boldsymbol{sigma}_T^T d\mathbf{W}_s - \int_0^t \frac{1}{2}\boldsymbol{\sigma}_T^T\boldsymbol{\sigma}_T ds)$$

such that

$$d\mathbf{W}_t = d\mathbf{W}_t^T + \boldsymbol{\sigma}_T dt.$$

Then under the forward measure Q_T,

$$d(\frac{S_t}{P_t}) = \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_T)^T d\mathbf{W}_t.$$

Also see Theorem 15.6.19. □

Lemma 19.3.6 (changing dynamics from real probability measure to forward measure). *Assume under real probability measure, we have*

$$\begin{aligned} dS_t/S_t &= (r + \lambda^T\boldsymbol{\sigma}_S)dt + \boldsymbol{\sigma}_S^T d\mathbf{W}(t) \\ dP(t,T)/P(t,T) &= (r + \lambda^T\boldsymbol{\sigma}_P)dt + \boldsymbol{\sigma}_P^T d\mathbf{W}(t) \end{aligned}$$

where $\boldsymbol{\sigma}_S$ and $\boldsymbol{\sigma}_P$ are volatility vectors, and $\mathbf{W}(t)$ is the Brownian motion vector, λ is the vector of market price of risks.

Then the forward measure Q_T is generated via

$$Z = \frac{dQ_P}{dQ} = \exp(\int_0^t (\boldsymbol{\sigma}_P - \lambda)^T d\mathbf{W}_s - \int_0^t \frac{1}{2}(\boldsymbol{\sigma}_P - \lambda)^T(\boldsymbol{\sigma}_P - \lambda)ds)$$

such that under which

- $$d\mathbf{W}_t = d\mathbf{W}_t^T + (\boldsymbol{\sigma}_P - \lambda)dt.$$
- $$d(\frac{S_t}{P_t}) = \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_P)^T d\mathbf{W}_t$$

Proof. Using Ito rule, we have

$$\begin{aligned} d(\frac{S_t}{P_t}) &= \frac{dS_t}{P_t} - \frac{S_t dP_t}{P_t^2} - \frac{dS_t dP_t}{P_t^2} + \frac{dP_t^2}{P_t^3} \\ &= \frac{S_t}{P_t}(rdt + \lambda^T\boldsymbol{\sigma}_S dt + \boldsymbol{\sigma}_S^T d\mathbf{W}_t - rdt - \lambda^T\boldsymbol{\sigma}_P dt - \boldsymbol{\sigma}_P^T d\mathbf{W}_t) - \boldsymbol{\sigma}_S^T\boldsymbol{\sigma}_T dt + \boldsymbol{\sigma}_P^T\boldsymbol{\sigma}_T dt) \\ &= \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_P)^T(d\mathbf{W}_t - (\boldsymbol{\sigma}_P - \lambda)dt). \end{aligned}$$

We can define a new measure via

$$Z = \frac{dQ_T}{dQ} = \exp(\int_0^t \boldsymbol{sigma}_T^T d\boldsymbol{W}_s - \int_0^t \frac{1}{2}\boldsymbol{\sigma}_T^T \boldsymbol{\sigma}_T ds)$$

such that

$$d\boldsymbol{W}_t = d\boldsymbol{W}_t^T + \boldsymbol{\sigma}_T dt.$$

Then under the forward measure Q_T,

$$d(\frac{S_t}{P_t}) = \frac{S_t}{P_t}(\boldsymbol{\sigma}_S - \boldsymbol{\sigma}_T)^T d\boldsymbol{W}_t.$$

Also see Theorem 15.6.19. □

Theorem 19.3.5 (forward bond dynamics under forward measure). *Assume $P(t,T)$ obeys the SDE*

$$dP(t,T) = P(t,T)r(t)dt + P(t,T)\sigma(t,T)dW(t)$$

where $W(t)$ is the standard Brownian motion under the risk-neutral measure Q, and $r(t)$ is the instantaneous short rate. Then,

- *the forward bond price $FP(t;S,T) = \frac{P(t,T)}{P(t,S)}, S \leq T$ has SDE, under measure Q, given by:*

$$dFP/FP = (\sigma(t,T) - \sigma(t,S))dW(t) + \sigma(t,S)(\sigma(t,S) - \sigma(t,T))dt.$$

- *The ratio $\triangleq \frac{P(t,S)}{P(t,T)}, S \leq T$ has SDE, under measure Q, given by:*

$$d\tilde{FP}/\tilde{FP} = (\sigma(t,S) - \sigma(t,T))dW(t) + \sigma(t,T)(\sigma(t,T) - \sigma(t,S))dt.$$

- *Under forward measure with respect to $P(t,S)$.The forward bond price $FP(t;S,T)$ has SDE*

$$dFP/FP = (\sigma(t,T) - \sigma(t,S))dW^S(t).$$

Moreover, $FP(t;S,T)$ will follow a geometric Brownian motion (Lemma 6.3.10) given by

$$FP(t';S,T) = FP(t;S,T)\exp(\int_t^{t'} \sigma'(s)dW^S(s) - \frac{1}{2}\int_t^{t'} \sigma'(s)^2 ds),$$

where $\sigma'(s) = (\sigma(u,T) - \sigma(u,S))$.
- *Use the fact that $FP(S;S,T) = P(S,T)$, we have*

$$P(S,T) = FP(S;S,T) = FP(t;S,T)\exp(\int_t^S \sigma'(s)dW^S(s) - \frac{1}{2}\int_t^S \sigma'(s)^2 ds)$$

- *In the forward measure with respect to $P(t,S)$, the Brownian motion under Q is changed to*
$$dW^S(t) = dW(t) - \sigma(t,S)dt.$$

Proof. (1)(2) Use Ito rule. Particularly, use $X = \log P(t,T), Y = \log P(t,S)$ such that $dX - dY = d(\log(P(t,T)/P(t,S))$ to simplify calculation. (3) Under forward measure of $P(t,S)$, the drift term $\sigma(t,S)(\sigma(t,S) - \sigma(t,T))$ will add $\sigma(t,S)(\sigma(t,T) - \sigma(t,S))$ based on Theorem 19.3.1Theorem 15.6.18. □

Remark 19.3.11 (caution! forward rate is not martingale in risk-neutral measure)**.** Note that we have shown that for assets other than interest rate derivative, the forward price, under independence assumption, will follow a martingale under risk-neutral measure(see Lemma 15.7.8). For interest rate derivatives, due to the correlation between forward rate the money market numeraire, forward rate is not a martingale under risk-neutral measure.

Remark 19.3.12 (Assumptions on bond dynamics under risk-neutral measure)**.** The validity of this assumption is discussed at Lemma 19.5.1.

Remark 19.3.13 (applications and analog is stocks)**.**

- One of most important result in this theorem is, under forward measure, it gives the distribution of the price at future time S of a zero-coupon bond maturing at T, given by
$$P(S,T) = FP(t;S,T)\exp(\int_t^S \sigma'(s)dW^S(s) - \frac{1}{2}\int_t^S \sigma'(s)^2 ds).$$
- The analog for stock price M prediction under risk-neutral measure Q gives
$$M(S) = M(t)\exp(\int_t^S \sigma_M(s)dW^Q(s) - \frac{1}{2}\int_t^S \sigma_M(s)^2 ds).$$
- The prediction of $P(S,T)$ is useful to price Vanilla bond options.

Remark 19.3.14 (understand $P(S,T)$ and $L(S,T)$)**.** Note that $P(S,T)$ and $L(S,T)$ are random variables; but they **are not $\mathcal{F}_t, t < S$ measurable**. On the contrary, $FP(t;S,T)$ and $F(t;S,T)$ are $\mathcal{F}_t$ measurable, and they are the prediction on $P(S,T)$ and $L(S,T)$ based on the information $\mathcal{F}_t$, under the forward measure.

Example 19.3.2 (application examples). Use the fact that $F(t;S,T)$ is a martingale under forward measure with respect to $P(t,T)$.

- An forward rate agreement with rate K has payoff at future time T
$$(T-S)(K - L(S,T)).$$

Note that its current value $V(t)$ satisfies

$$\frac{V(t)}{P(t,T)} = E_T[\frac{(T-S)(K-L(S,T))}{P(T,T)}|\mathcal{F}_t] = (T-S)K - (T-S)F(t;S,T).$$

- An interest rate swap has present value

$$PS(t) = \sum_{i=1}^{n} P(t,T_i)E_{T_i}[\tau_i(L(T_{i-1},T_i)-K)|\mathcal{F}_t] = \sum_{i=1}^{n} P(t,T_i)\tau_i(F(t;T_{i-1},T_i)-K).$$

Remark 19.3.15 (drift term of bond SDE). The SDE for zero-coupon bond may have different drift terms instead of $P(t,T)r(t)dt$ in the real probability measure; however, in complete market with no-arbitrage assumption, there exists an unique risk-neutral measure under which the bond SDE will have drift term $P(t,T)r(t)$.

19.3.3.2 *State price interpretation of forward measure*

Lemma 19.3.7 (fundamental solution to term structure equation). *[6, p. 35] Assume the short rate r under risk-neutral measure is governed by*

$$dr = m(r,t)dt + \sigma(r,t)dz.$$

Given the term-structure equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} + mS\frac{\partial V}{\partial S} - rV = 0,$$

and let $V^\delta(t,r_t,s) > 1$ be the solution for final condition $V_T^\delta(r_T,s) = \delta(r_T - s)$. It follows that

- $$P(t,T,r_t) = \int_{-\infty}^{\infty} V^\delta(t,r_t,s)ds$$

- *Define $p^\delta(t,r_t,s) \triangleq V^\delta(t,r_t,s)/P(t,T,r_t)$, then*

$$\int_{-\infty}^{\infty} p^\delta(t,r_t,s)ds = 1.$$

- *For any payoff function $V_T(r_T)$, we have its current value given by*

$$\begin{aligned} V(t) &= \int_{-\infty}^{\infty} V_T(r_T = s)V^{\delta}(t, r_t, r_T = s)ds \\ &= P(t,T,r_t)\int_{-\infty}^{\infty} V_T(S_T = s)p^{\delta}(t, r_t, r_T = s)ds \\ &= P(t,T,r_t)E_T[V_T(S_T)|\mathcal{F}_t] \end{aligned}$$

where E_T denotes taking expectation with respect to p^{δ}.

Proof. (1)(2) Use the fact that a zero-coupon bond pays 1 at all world states. (3) use linearity of pricing. □

Remark 19.3.16 (interpretation)**.**

- This lemma shows that the nature of forward probability measure, which is the state price(expressed in unit of $P(t,T,r_t)$) resulted from market equilibrium based on current market condition(time t and current short rate r_t).
- The forward probability/state price therefore will depends on time t and short rate r_t.

19.3.4 Applications in pricing

19.3.4.1 *LIBOR payment*

Theorem 19.3.6 (current no-arbitrage value of LIBOR payment, recap). *[1, p. 6] Let $S < T$. Consider the future LIBOR-based payment $(T-S)L(S,T)$ at time T. Its arbitrage-free value at time $t < T$ is $P(t,S) - P(t,T), T > S$. That is*

$$\begin{aligned} V(t) &= E_Q[\frac{M(t)(T-S)L(S,T)}{M(T)}|\mathcal{F}_t] \\ &= P(t,S) - P(t,T) \end{aligned}$$

Proof. (martingale method under forward measure Lemma 19.3.4)

$$\begin{aligned}
V(t) &= P(t,T)E_T[\delta L(S,T)|\mathcal{F}_t] \\
&= P(t,T)E_T[\delta F(S;S,T)|\mathcal{F}_t] \\
&= P(t,T)\delta F(t;S,T) \\
&= P(t,T)(\frac{P(t,S)}{P(t,T)} - 1) \\
&= P(t,S) - P(t,T)
\end{aligned}$$

(martingale method under risk-neutral measure)

$$\begin{aligned}
V(t) &= M(t)E_Q[\frac{(T-S)L(S,T)}{M(T)}|\mathcal{F}_t] \\
&= E_Q[\frac{M(t)}{M(T)}(\frac{M(T)}{M(S)} - 1)|\mathcal{F}_t] \\
&= E_Q[\frac{M(t)}{M(S)} - \frac{M(t)}{M(T)}|\mathcal{F}_t] \\
&= P(t,S) - P(t,T)
\end{aligned}$$

where we use no-arbitrage relation between LIBOR and money account (Theorem 19.1.1).

See also the replication method (Theorem 19.1.2). □

Lemma 19.3.8 (current value of LIBOR payment risk-neutral Method). *[1, p. 6] Let $t < S < T$. It follows that*

- *Time t value of $(T-S)L(S,T)$ is given by*

$$V(t) = B(t)E_Q[\frac{(T-S)L(S,T)}{B(T)}|\mathcal{F}_t] = P(t,S) - P(t,T).$$

- $$E_T[L(S,T)|\mathcal{F}_t] = F(t,S,T)$$

Proof. (3) Note that under risk-neutral measure

$$\frac{P(t,T)}{B(t)} = E_Q[\frac{P(T,T)}{B(T)}|\mathcal{F}_t].$$

(4) From risk-neutral pricing and the results from (1)(2)(3), we have

$$\begin{aligned}V(t) &= B(t)E_Q[\frac{(T-S)L(S,T)}{B(T)}|\mathcal{F}_t] \\ &= E_Q[\frac{B(t)}{B(T)}(\frac{B(T)}{B(S)}-1)|\mathcal{F}_t] \\ &= E_Q[\frac{B(t)}{B(S)}-\frac{B(t)}{B(T)}|\mathcal{F}_t] \\ &= P(t,S)-P(t,T)\end{aligned}$$

(4)

$$\begin{aligned}E_T[L(S,T)|\mathcal{F}_t] &= E_T[\frac{1}{T-S}(\frac{1}{P(S,T)}-1)|\mathcal{F}_t] \\ &= E_T[\frac{1}{T-S}(\frac{P(S,S)}{P(S,T)}-1)|\mathcal{F}_t] \\ &= \frac{1}{T-S}(\frac{P(t,S)}{P(t,T)}-1) \\ &= F(t,S,T)\end{aligned}$$

where we use the fact that $\frac{P(t,S)}{P(t,T)}$ is a martingale under T forward measure. □

Remark 19.3.17 (interpretation)**.**

- (interpret $L(S,T)$) The quantity $(T-S)L(S,T)$ is stochastic and not $\mathcal{F}_t$ measurable. Note that from definition,

$$P(S,T) = \frac{1}{1+(T-S)L(S,T)}.$$

  And $P(t,S)-P(t,T)$ is the market price at time t, which is $\mathcal{F}_t$ measurable(or known at time t).
- (LABOR-in-arrears) If the payment is made at time S, see Lemma 19.13.1.
- **(caution!)** note that $1/P(S,T)$ is not a martingale under T forward measure, but $\frac{P(t,S)}{P(t,T)}$ is .

19.3.4.2 *Bonds*

19.3.4.3 *Forward rate contract*

19.3.4.4 *Forward bond contract*

19.3.4.5 *Interest rate swaps*

Lemma 19.3.9 (the evolution of an interest rate swap value). *Let the current time be t_0. Consider a payer swap starting at T_0 with cash flow at $T_1, T_2, ..., T_n$ given by*

$$\tau_i(L(T_{i-1}, T_i) - K), i = 1, 2, ..., n,$$

such that when $K = S(t_0; T_0, \Delta)$, the value of swap at current time t_0 is zero. It follows that

- *At future time $t_1, t_0 \leq t_1 \leq T_0$, the value of the contract is given by*

$$V(t_1) = A(t_1, T_0, T_n)(S(t_1; T_0, T_n) - S(t_0; T_0, T_n)),$$

where $S(t_1; T_0, \Delta)$ is the fair forward starting rate at t_1, and $A(t, T_0, T_n)$ is the annuity price given by

$$A(t, T_0, T_n) = \sum_{i=1}^{n} \tau_i P(t_i, T_n).$$

- *At future time $t_1, t_1 > T_0$, the value of the contract is given by*

$$V(t_1) = A_{\alpha(t_1),n}(t_1, T_{\alpha(t_1)}, T_n)(S(t_1; T_{\alpha(t_1)}, T_n) - S(t_0; T_0, T_n)),$$

where $\alpha(t_1)$ is the minimum integer such that $T_\alpha \geq t_1$, $S(t_1, T_{\alpha(t_1)}, T_n)$ is the fair forward rate at t_1 for a swap with payment spanning $T_{\alpha(t_1)}, ..., T_n$, $\tau = T - S$, $A_{\alpha(t_1,n)}(t, T_{\alpha(t_1)}, T_n)$ is the annuity price given by

$$A(t, T_{\alpha(t_1)}, T_n) = \sum_{i=\alpha(t_1)}^{n} \tau_i P(t_1, T_n).$$

At future time $t_1, t_0 < t_1 < T_n$, the

Note that** $A(t, T_{\alpha(t_1)}, T_n)$**,** $S(t_1, T_{\alpha(t_1)}, T_n)$ **and** $V(t_1)$ **are all random quantities at time** t_0**.

In particular, $V(t_0) = V(T_n) = 0$.

Proof. Note that at any time $t < T_n$, the fair value swap rate is related to the LIBOR payments via

$$V(t) = \sum_{i=\alpha(t)}^{n} P(t, T_i)\tau_i E_{T_i}[S_{\alpha(t),n}(t) - L(T_{i-1}, T_i)|\mathcal{F}_t] = 0$$

$$\implies A_{\alpha(t),n}(t) S_{\alpha(t),n}(t) = \sum_{i=\alpha(t)}^{n} P(t, T_i)\tau_i E_{T_i}[L(T_{i-1}, T_i)|\mathcal{F}_t]$$

For a contract entered at time t_0, the fixed leg has t_1-value $A_{\alpha(t_1),n}(t_1)S_{\alpha(t_0),n}(t)$, the floating leg has value

$$\sum_{i=\alpha(t_1)}^{n} P(t_1,T_i)\tau_i E_{T_i}[L(T_{i-1},T_i)|\mathcal{F}_t] = A_{\alpha(t_1),n}(t)S_{\alpha(t_1),n}(t).$$

Therefore, the value of the forward contract is given by

$$V(t_1) = A(t_1,T_{\alpha(t_1)},T_n)(S(t_1,T_{\alpha(t_1)},T_n) - S(t_0,T_0,T_n)).$$

□

Remark 19.3.18 (reduction to forward contract value evolution). Note that forward contract value evolution(Theorem 19.1.4) is given by

$$V(t_1) = P(t_1,T)\tau(F(t_0;S,T) - F(t_1;S,T)),$$

which is the same as one-period interest rate swap with annuity $A(t) = \tau P(t,T)$.

19.3.4.6 *Vanilla bond options*

Lemma 19.3.10 (Black's model for call option on zero-coupon bond). *[1, p. 32] The value at time t of a European call option maturing at S with strike K on a bond maturing at $T > S$ is given by*

$$V(t) = P(t,S)E_S[\max(P(S,T)-K,0)|\mathcal{F}_t],$$

where the expectation is taken with respect to forward measure of $P(t,S)$. If zero-coupon bonds follows the SDE

$$dP(t,T) = r(t)P(t,T)dt + \sigma(t,T)P(t,T)dW(t)$$

with deterministic volatility $\sigma(t,T)$, then

$$V(t) \triangleq BC(t;S,T,K) = P(t,T)N(d_+) - KP(t,S)N(d_-),$$

where

$$d_+ = \frac{\log\frac{FB(t;S,T)}{K} + \frac{1}{2}v(t,S)^2}{v(t,S)} = \frac{\log\frac{P(t,T)}{KP(t,S)} + \frac{1}{2}v(t,S)^2}{v(t,S)}, d_- = d_+ - v(t,S).$$

and

$$v(t,S)^2 = \int_t^S (\sigma(u,T)-\sigma(u,S))^2 dt$$

Proof. Under the forward measure with respect to $P(t,S)$, we have

$$\begin{aligned}\frac{V(t)}{P(t,S)} &= E_S[\frac{\max(P(S,T)-K,0)}{P(S,S)}|\mathcal{F}_t] \\ &= E_S[\max(P(S,T)-K,0)] \\ &= E_S[FP(S;S,T)1_{P(S,T)>K}] - E_S[K1_{P(S,T)>K}] \\ &= FP(t;S,T)N(d_+) - KN(d_-)\end{aligned}$$

Note that $FB(t;S,T)$ is martingale under S forward measure. Then probability distribution of $P(S,T)$ under such forward measure Theorem 19.3.5 will be such that

$$P(S,T) = FB(S;S,T) = FB(t;S,T)\exp(\int_t^S \sigma'(s)dW^S(s) - \frac{1}{2}\int_t^S \sigma'(s)^2 ds),$$

where $\sigma'(s) = (\sigma(u,T) - \sigma(u,S))$. Note that it is easy to see that $\ln \frac{P(S,T)}{FP(t;S,T)}$ has mean $\frac{1}{2}\int_t^S \sigma'(s)^2 ds$ and variance $\int_t^S \sigma'(s)^2 dt$.

Note that $FP(t;S,T)$ is $\mathcal{F}_t$ measurable; therefore, we can calculate the distribution $P(S,T)$.

□

Lemma 19.3.11 (put-call parity and put option pricing). *Let $BC(t;S,T,K)$ and $BP(t;S,T,K)$ denote the time t value of the call and put option (expiration date S and strike K) on a bond matures at time T, then*

$$BC(t;S,T,K) - BP(t;S,T,K) = P(t,T) - KP(t,S).$$

Therefore, the put option price is given by

$$BP(t;S,T,K) = KP(t,S)N(-d_-) - P(t,T)N(-d_+).$$

Proof. We the change of numeraire and forward measure with respect to $P(t,S)$, we haves

$$\begin{aligned}\max(P(S,T)-K,0) - \max(K-P(S,T),0) &= P(S,T) - K \\ E_S[\max(P(S,T)-K,0)|\mathcal{F}_t] - E_S[\max(K-P(S,T),0)|\mathcal{F}_t] &= E_S[P(S,T)|\mathcal{F}_t] - E_S[K|\mathcal{F}_t] \\ (BC(t;S,T,K) - BP(t;S,T,K))/P(t,S) &= P(t,T)/P(t,S) - K \\ BC(t;S,T,K) - BP(t;S,T,K) &= P(t,T) - KP(t,S)\end{aligned}$$

□

19.3.4.7 *Caps and floors*

Lemma 19.3.12 (Black's model for European caps). *[5, p. 680][3, p. 199] Consider a caplet with payoff at t_{k+1} given by*

$$(L(t_k, t_{k+1}) - K)^+$$

*where K is the **cap rate**. Assume under the forward measure associated with numeriare $P(t, t_{k+1})$, the forward rate $F(t, t_k, t_{k+1})$ has dynamics*

$$dF = \tilde{\sigma}_F(t) F dW_t.$$

Then the value of a caplet is given by

$$\delta_k P(t, t_{k+1})[F_k N(d_1) - KN(d_2)],$$

where

$$d_1 = \frac{\ln(F_k/K) + \sigma_F^2/2}{\sigma_F}, d_2 = d_1 - \sigma_F,$$

$\delta_k = t_{k+1} - t_k$, $F_k = F(t; t_k, t_{k+1}), \sigma_F^2 = \int_t^{t_k} \tilde{\sigma}_F(t)^2 dt.$

The cap consisting of n caplets has price given by

$$\sum_{k=1}^{n} \delta_k P(t, t_{k+1})[F_k N(d_1) - KN(d_2)].$$

Proof. Under the forward measure associated with numeriare $P(t, t_{k+1})$ We have pricing formula

$$V(t) = P(t, t_{k+1}) E_{t_{k+1}}[(L(t_k, t_{k+1}) - K)^+ | \mathcal{F}_t],$$

Note that

$$L(t_k, t_{k+1}) = F(t_k; t_k, t_{k+1}) = F(t; t_k, t_{k+1}) \exp(-\frac{1}{2}\sigma_F^2 + \sigma_F Z),$$

where $Z \sim N(0,1), \sigma_F^2 = \int_t^{t_k} \tilde{\sigma}_F(t)^2 dt.$ □

Lemma 19.3.13 (Pricing caplets/floorlets using Vanilla options on bonds). *[1, p. 34] Let $BC(t; T_{i-1}, T_i, \frac{1}{1+\tau_i K}), BP(t; T_{i-1}, T_i, \frac{1}{1+\tau_i K})$ denote a call/put option (expiry T_{i-1}) on a zero-coupon bond (maturity T_i) with strike $\frac{1}{1+\tau_i K}$.*

Then

$$Caplet(t) = (1 + \tau_i K) BP(t; T_{i-1}, T_i, \frac{1}{1 + \tau_i K})$$

$$Flrlet(t) = (1 + \tau_i K)BC(t; T_{i-1}, T_i, \frac{1}{1 + \tau_i K})$$

Proof. The payoff at T_{i-1} for a caplet with strike K and maturity T_{i-1} is given by

$$\tau_i P(T_{i-1}, T_i)(L(T_{i-1}, T_i) - K)^+$$

From the definitions,

$$P(T_{i-1}, T_i) = \frac{1}{1 + \tau L(T_{i-1}, T_i)} \Leftrightarrow L(T_{i-1}, T_i) = \frac{1 - P(T_{i-1}, T_i)}{\tau P(T_{i-1}, T_i)}$$

We have

$$(1 + \tau_i K)(\frac{1}{1 + \tau_i K} - P(T_{i-1}, T_i))^+.$$

This is equivalent to a vanilla put option with strick $\frac{1}{1+\tau_i K}$. □

Lemma 19.3.14 (put-call parity for caps and floors). *[5, p. 680] Let V_{cap}, V_{floor} and V_{swap} denote the prices of a cap, a floor and a swap. Let the cap and floor have the same strike price, R_K. The swap is an agreement to receive LIBOR and pay a fixed rate of R_K with no exchange of payment on the first reset date.All three instruments have the same life and the same frequency of payments. We have*

$$V_{cap} = V_{floor} + V_{swap}.$$

Proof. Consider we long a cap and short a floor. The cap provides a cash flow of $LIBOR - R_K$ when $LIBOR > R_K$; the short floor provides a cash flow of $-(R_K - LIBOR)$ when $LIBOR < R_K$. Therefore, the position amounts to a swap with cash flow $LIBOR - R_K$. □

19.3.4.8 *Swaption*

Lemma 19.3.15 (value of a swaption). *[1, p. 36] Consider an interest rate swap with payment dates $T_1, T_2, ..., T_n$ and reset dates $T_0, T_1, ..., T_{n-1}$. Let K be the strike and $T_s \leq T_0$ be the maturity of the swaption. It follows that*

- *The payer swaption payoff at maturity date T_s is*

$$PSwpt(T_s, T_0, T_n) = [PS(T_s)]^+ = A(T_s, T_0, T_n)(S(T_s, T_0, T_n) - K)^+,$$

where $A(t, T_0, T_n)$ is the swap annuity given by

$$A(t, T_0, T_n) = \sum_{i=1}^{n} \tau_i P(t, T_i).$$

- *The payer swaption value at time $t < T_s$ is given by*

$$V(t) = A(t, T_0, T_n) E_A[(S_{0,n}(T_s) - K)^+ | \mathcal{F}_t],$$

where E_A is the expectation taken with respect to the martingale measure associated with $A(t, T_0, T_n)$.

- *If K is set to $S(t, T_0, T_n), t < T_s$, then*

$$V(t) = A(t, T_0, T_n) E_A[(S(T_s, T_0, T_n) - S(t, T_0, T_n))^+ | \mathcal{F}_t].$$

Note that the forward swap rate $S(t, T_0, T_n)$ is a stochastic process.

Proof. (1) Under the forward measure of $T_1, T_2, ..., T_n$, we have

$$\begin{aligned}
PS(T_s) &= \sum_{i=1}^{n} P(T_s, T_i) E_{T_i}[\tau_i (L(T_{i-1}, T_i) - K) \mathcal{F}_{T_s}] \\
&= \sum_{i=1}^{n} P(T_0, T_i) \tau_i (F(T_0; T_{i-1}, T_i) - K) \\
&= P(T_s, T_0) - P(T_s, T_n) - K \sum_{i=1}^{n} \tau_i P(T_s, T_i) \\
&= (S_{0,n}(T_s) - K) \sum_{i=1}^{n} \tau_i P(T_s, T_i) \\
&= (S_{0,n}(T_s) - K) A_{0,n}(T_s)
\end{aligned}$$

(2) Under the annuity measure, we have

$$V(t) = A_{0,n}(t) E_A[\frac{A_{0,n}(T_s)(S_{0,n}(T_s) - K)^+}{A_{0,n}(T_s)} | \mathcal{F}_t] = A_{0,n}(t) E_A[(S_{0,n}(T_s) - K)^+ | \mathcal{F}_t],$$

□

Remark 19.3.19 (model dependent valuation). Note that the valuation of $E_A[(S_{0,n}(T_s) - K)^+ | \mathcal{F}_t]$ depends on the model dynamics of $S_{0,n}$. See Lemma 19.10.10 for an example.

Lemma 19.3.16 (put-call parity for swaption). *Consider a payer swaption and a receiver swaption with the same strike K and maturity T_s. It follows that*

- *the payoff at maturity date T_s satisfies*

$$[PS(T_s)]^+ - [-PS(T_0)]^+ = PS(T_s);$$

- *the value at $t_0 < t < T_s$ satisfies*

$$PSwpt_{0,n}(t) - RSwpt_{0,n}(t) = PS(t) = A_{0,n}(t_1)(S_{0,n}(t) - S_{0,n}(t_0)),$$

where $S_{0,n}(t_0), t_0 \leq t$ *is the fair value swap rate at* t_0 *when the swap contract was entered.*

- $$PSwpt_{0,n}(t_0) = RSwpt_{0,n}(t_0).$$

Proof. (1) Straight forward. (2) Using annuity as the measure, we have

$$[PS(T_s)]^+ - [-PS(T_0)]^+ = PS(T_s)$$

$$A(t)E_A[\frac{[PS(T_s)]^+ - [-PS(T_0)]^+}{A(T_s)}|\mathcal{F}_t] = A(t)E_A[\frac{PS(T_s)}{A(T_s)}|\mathcal{F}_t] = PS(t) = A_{0,n}(t)(S_{0,n}(t) - S_{0,n}(t_0))$$

where we use the value of interest rate swap(Lemma 19.3.9). (3) Note that $PS(t_0) = 0$. □

19.3.4.9 *Equity option pricing in stochastic interest rate*

Lemma 19.3.17 (European equity options pricing under stochastic interest rate). *[5, p. 666] Assume the stock price follows geometric Brownian motion and the price of zero-coupon-bond* $P(0,T)$ *is known.*

Define F_0 *and* F_T *as the forward price of the asset at time* 0 *and* T *for a contract maturing at time* T.

The European call option price is given by

$$c = P(0,T)E_T[\max(S_T - K, 0)] = P(0,T)E_T[\max(F_T - K, 0)].$$

where E_T *denotes the expectations in a world that is forward risk neutral with respect to* $P(t,T)$. *Assume that* F_t *is lognormal distribution* ***in this forward measure*** *with variance* $\sigma_F^2 T$. *Then*

$$c = P(0,T)[F_0 N(d_1) - KN(d_2)],$$

where

$$d_1 = \frac{\ln(F_0/K) + \sigma_F^2 T/2}{\sigma_F\sqrt{T}}, d_2 = d_1 - \sigma_F T.$$

Proof. From the definition of forward price $F_T = S_T$. Under forward measure with respect of $P(t,T)$,

$$F_t \triangleq \frac{S(t)}{P(t,T)}$$

is a martingale(Theorem 19.3.4).

Therefore, $F_T = F_0 \exp(-\frac{1}{2}\sigma_F^2 T + \sigma_F\sqrt{T}Z)$, where $Z \in N(0,1)$

Use the distribution of F_T under forward measure, we can calculate

$$P(0,T)E_T[\max(F_T - K, 0)]$$

.

Note that the $F_0 = S_0/P(0,T)$ and σ_F can be calculated using Theorem 15.6.19. $\square$

Note 19.3.1 (uncertainty increase the value of options). The stochastic nature of the interest rate will make $\sigma_F > \sigma_S$, thus generally increasing the value of the options. See the following example.

Example 19.3.3. link Consider the interest rate model and the stock price model (under risk-neutral measure) as

$$dr_t = (\theta_t - ar_t)dt + \sigma_0 dW_t^1$$
$$dS_t = S_t(r_t dt + \sigma(\rho dW_t^1 + \sqrt{1-\rho^2})dW_t^2)$$

where W_t^1, W_t^2 are two independent Brownian motions.Then,

The zero-coupon bond price $P(t,T)$ is given as

$$dP(t,T) = P(t,T)(r_t dt - \sigma_0 D(t,T)dW_t),$$

where

$$D(t,T) = \int_t^T e^{-a(s-t)}ds$$

- Under the forward risk-neutral of $P(t,T)$, we have

$$dP(t,T) = P(t,T)(r_t - \sigma_0^2 D(t,T)^2 dW_t^1)dt - \sigma_0 D(t,T)dW_t^1),$$

and

$$dS_t = S_t((r_t - \rho\sigma\sigma_0 D(t,T))dt + \sigma(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2)$$

- Let $F(t,T) = S_t/P(t,T)$, then $F(t,T)$ is a martingale under the forward risk-neutral of $P(t,T)$, given as

$$dF(t,T) = F(t,T)((\sigma\rho + \sigma_0 D(t,T))dW_t^1 + \sigma\sqrt{1-\rho^2}dW_t^2),$$

which is a martingale.
Then, we can evaluate the distribution of $F(T,T)$ and the expectation $E_T[\max(F(T,T)-K,0)]$.

19.4 Market yield curve analysis and construction

19.4.1 Concepts and facts

Note 19.4.1 (yield curve concept and market observations). link Let the current time be t and the maturity be T. A yield curve observed at the market today is usually denoted by $y(T)$. It means that if we invest 1$ in the money market for T years, we will get $(1+y)^T$ or $\exp(yT)$ after T years.

As showed in Figure 19.4.1, the market yield curves usually have three types of shapes:

- **(normal yield curve)** A normal or up-sloped yield curve indicates yields on longer-term bonds may continue to rise, responding to periods of economic expansion. When investors expect longer-maturity bond yields to become even higher in the future, many would temporarily park their funds in shorter-term securities in hopes of purchasing longer-term bonds later for higher yields. In a rising interest rate environment, it is risky to have investments tied up in longer-term bonds when their value has yet to decline as a result of higher yields over time. The increasing temporary demand for shorter-term securities pushes their yields even lower, setting in motion a steeper up-sloped normal yield curve.
- **(inverted yield curve)** An inverted or down-sloped yield curve suggests yields on longer-term bonds may continue to fall, corresponding to periods of economic recession. When investors expect longer-maturity bond yields to become even lower in the future, many would purchase longer-maturity bonds to lock in yields before they decrease further. The increasing onset of demand for longer-maturity bonds and the lack of demand for shorter-term securities lead to higher prices but lower yields on longer-maturity bonds, and lower prices but higher yields on shorter-term securities, further inverting a down-sloped yield curve.
- **(flat(humped) yield curve)** A flat yield curve may arise from normal or inverted yield curve, depending on changing economic conditions. When the economy is transitioning from expansion to slower development and even recession, yields on longer-maturity bonds tend to fall and yields on shorter-term securities likely rise, inverting a normal yield curve into a flat yield curve. When the economy is transitioning from recession to recovery and potential expansion, yields on longer-maturity bonds are set to rise and yields on shorter-maturity securities are sure to fall, tilting an inverted yield curve toward a flat yield curve.

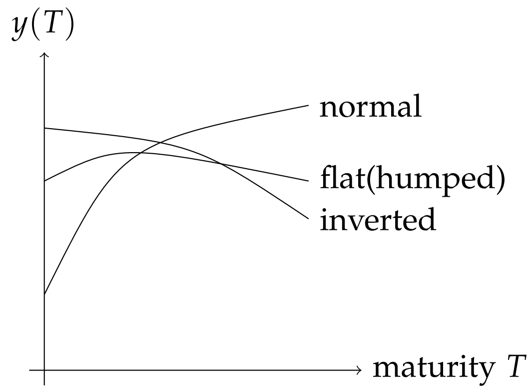


Figure 19.4.1: Different types of market observed yield curve shape

Note 19.4.2 (equivalence of yield curve, zero-coupon curve, and (instantaneous)forward curve, recap of Lemma 19.1.4).

- Given the zero-coupon curve $P(t,T)$, we can calculate the continuously compounding yield curve

$$y(t,T) = R(t,T) \triangleq -\frac{\log P(t,T)}{T-t},$$

and the instantaneous forward curve

$$f(t,T) \triangleq -\frac{\partial}{\partial T}\log P(t,T).$$

- The yield curve and instantaneous forward rate curve are connected by

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T) = \frac{\partial}{\partial T}(R(t,T)(T-t)) = R(t,T) + (T-t)\frac{\partial}{\partial T}R(t,T).$$

  - If $f(t,T) = f_0$(i.e., it is flat), then $R(t,T) = f(t,T) = 0$.
  - If $R(t,T)$ is monotonically increasing, then $f(t,T)$ is above $R(t,T)$.
  - If $R(t,T)$ is monotonically decreasing, then $f(t,T)$ is below $R(t,T)$.
- The instantaneous forward rate curve can be used to calculate the forward rate of any tenor (T_1,T_2) via

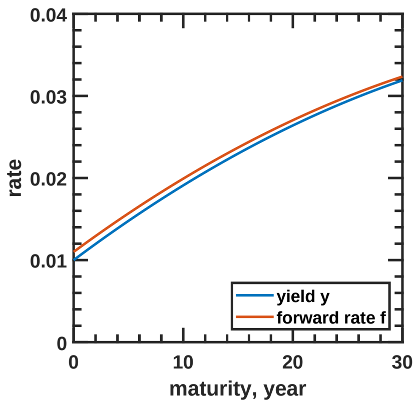
$$F(t,T_1,T_2) = \frac{1}{T_2-T_1}(\exp(\int_{T_1}^{T_2} f(t,s)ds) - 1).$$

- Similarly, given by the yield curve $R(t,T)$, we have

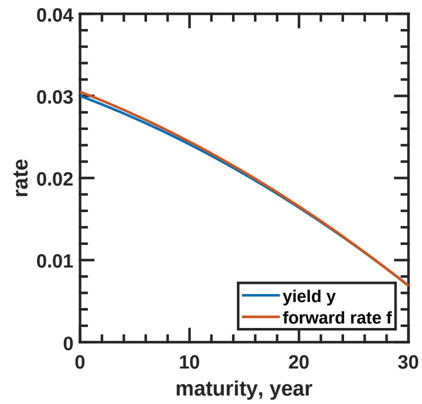
$$P(t,T) = \exp(-R(t,T)(T-t)).$$

Remark 19.4.1. recall that

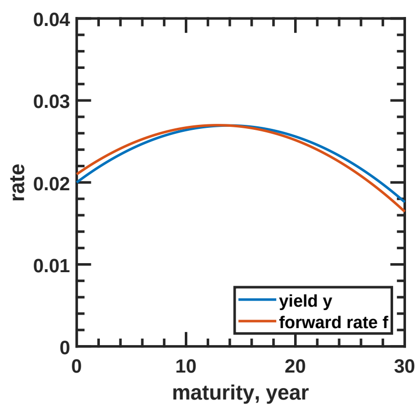
$$P(t,T) = \exp(-R(t,T)(T-t)), P(t,T) = \exp(-\int_t^T f(t,s)ds)$$



(a) **Normal** yield curve and forward rate curve.



(b) **Inverted** yield curve and forward rate curve.



(c) **Humped** yield curve and forward rate curve.

Figure 19.4.2: Yield curves and forward rate curves.

19.4.2 Market common instrument data

Table 19.4.1: USD LIBOR spot rate as of March-27-2018

USD LIBOR overnight	1.69750%
USD LIBOR 1week	1.73125%
USD LIBOR 1month	1.87688%
USD LIBOR 2 months	1.99438%
USD LIBOR 3 months	2.30200%
USD LIBOR 6 months	2.45299%
USD LIBOR 12 months	2.67138%

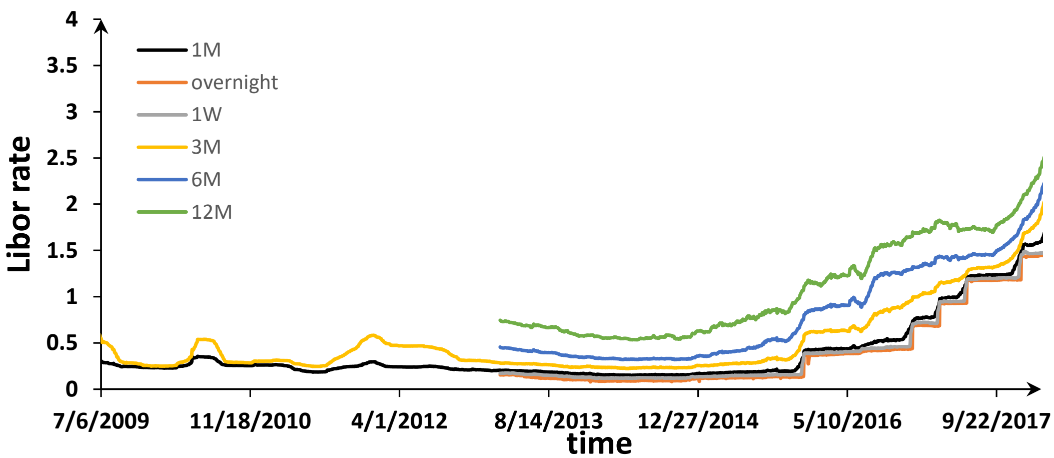


Figure 19.4.3: Historical Libor spot rate with different tenors. Data source https://fred.stlouisfed.org

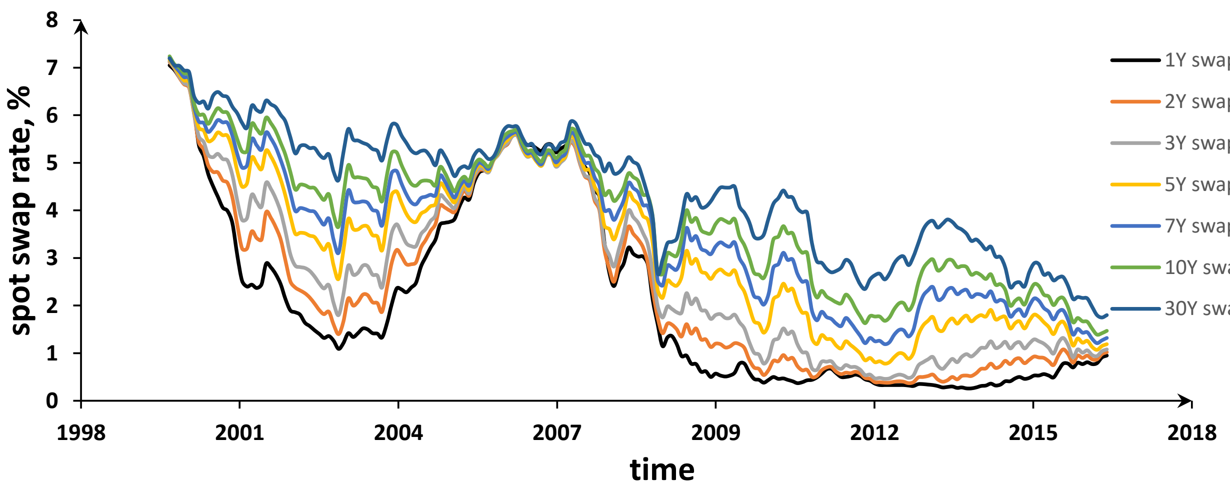


Figure 19.4.4: Historical swap spot rate with different tenors. Data source https://fred.stlouisfed.org

19.4.3 Calibration: instantaneous forward rate

Remark 19.4.2 (how the real world calibrate a zero-coupon curve)**.** [5, p. 164]

- Usually LIBOR rates only available for maturities up to 12 months.
- Eurodollar futures can be used to produce zero-coupon curve up to 5 years.
- Interest rate swap can be then used to extend to 30 years.

Lemma 19.4.1 (forward rate calibration from market). *[1, p. 15]*

Setup:

- *We are given a sequence of co-initial interest rate swaps starting at the spot date t.* ***They have coincident reset dates but different maturity dates.***
- *Let $T_0, ..., T_{n_i} - 1$ denote reset dates for swap i and let T_{n_i} denote its maturity date.*
- *Let the length of the swaps, i.e $(T_{n_i} - T_0)$ form an increasing sequence $n_1 < n_2 < \cdots$.*
- *Assume the piecewise constant interpolation of the instantaneous forward rate*

$$f(t,T) = \begin{cases} f_1, t \leq T \leq T_{n_1} \\ f_{i+1}, T_{n_i} \leq T \leq T_{n_{i+1}}, i = 1,2,... \end{cases}.$$

Procedures to find f_1:

- *For $i = 1$, and $j = 0, ..., n_1$, we have*

$$P(t, T_j) = \exp(-\int_t^{T_j} f(t,u)du) = \exp(-f_1(T_j - t)).$$

From the definition of forward swap rate(Definition 19.1.14), we have

$$\exp(-f_1(T_0 - t)) - \exp(-f_1(T_{n_1} - t)) = S_{0,n_1}(t)\sum_{j=1}^{n_1} \tau_j \exp(-f_1(T_j - t)),$$

where $S_{0,n_1}(t)$ denote the forward swap rate.
- *Therefore, we can solve f_1 using root-find algorithms.*

Procedures to find f_2:

- *For $i = 1$, and $j = n_1, n_1 + 1, ..., n_2$, we have*

$$P(t, T_j) = \exp(-\int_t^{T_j} f(t,u)du) = P(t, T_{n_1})\exp(-f_2(T_j - T_{n_1})).$$

From the definition of forward swap rate(Definition 19.1.14), we have

$$\begin{aligned}
P(t, T_0) - P(t, T_{n_m}) &= S_{0,n_m}(t)\sum_{j=1}^{n_m} \tau_j P(t, T_j) \\
\exp(-f_1(T_0 - t)) - P(t, T_{n_{m-1}})\exp(-f_m(T_{n_m} - T_{n_{m-1}})) &= S_{0,n_m}(t)(\sum_{j=1}^{n_{m-1}} \tau_j P(t, T_j) \\
&+ \sum_{j=n_{m-1}+1}^{n_m} \tau_j P(t, T_j)) \\
P(t, T_{n_{m-1}})\exp(-f_m(T_{n_m} - T_{n_{m-1}})) &= S_{0,n_m}(t)(\sum_{j=1}^{n_{m-1}} \tau_j P(t, T_j) \\
&+ P(t, T_{n_{m-1}})\sum_{j=n_{m-1}+1}^{n_m} \tau_j \exp(-f_m(T_j - T
\end{aligned}$$

where $S_{0,n_2}(t)$ denote the forward swap rate.
- *Therefore, we can solve f_2 using root-find algorithms.*

Procedures to find f_m***:***

- *For* $i = m$*, and* $j = n_m + 1, ..., n_{m+1}$*, we have*

$$P(t, T_j) = \exp(-\int_t^{T_j} f(t, u) du) = P(t, T_{n_{m-1}}) \exp(-f_2(T_j - T_m)).$$

From the definition of forward swap rate(Definition 19.1.14), we have

$$\begin{aligned} P(t, T_0) - P(t, T_{n_2}) &= S_{0,n_2}(t) \sum_{j=1}^{n_2} \tau_j P(t, T_j) \\ \exp(-f_1(T_0 - t)) - P(t, T_{n_1}) \exp(-f_2(T_{n_2} - T_{n_1})) &= S_{0,n_2}(t)(\sum_{j=1}^{n_1} \tau_j P(t, T_j) \\ &+ \sum_{j=n_1+1}^{n_2} \tau_j P(t, T_j)) \\ P(t, T_{n_1}) \exp(-f_2(T_{n_2} - T_{n_1})) &= S_{0,n_2}(t)(\sum_{j=1}^{n_1} \tau_j P(t, T_j) + P(t, T_{n_1}) \\ &\sum_{j=n_1+1}^{n_2} \tau_j \exp(-f_2(T_j - T_{n_1}) \end{aligned}$$

where $S_{0,n_2}(t)$ *denote the forward swap rate.*

- *Therefore, we can solve* f_2 *using root-find algorithms.*

19.4.4 Calibration: zero-coupon bond curve

Lemma 19.4.2 (calculate bond price from swap rate bootstrapping). *[1, p. 15] Suppose we have swap rate with maturity dates* $n_1 = 1, n_2 = 2, n_3 = 3, ...$

Then use

$$P(t, T_{n_i}) = \frac{P(t, T_0) - \sum_{j=1}^{n_i - 1} r_i \tau_j P(t, T_j)}{1 + \tau_{n_i} r_i}.$$

to calculate $P(t, n_i)$*.* ***We assume*** $P(t, T_0) = 1$ ***since*** t, T_0 ***are very close***

- ***Procedures to find*** $P(t, 1), n_1 = 1$***:***

$$P(t, 1) = \frac{P(t, T_0)}{1 + \tau_{n_1} S_{0,n_1}(t)}.$$

- ***Procedures to find*** $P(t,2), n_2 = 2$**:**

$$P(t,2) = \frac{P(t,T_0) + \sum_{j=1}^{n_2-1} S_{0,n_2}(t)\tau_j P(t,T_j)}{1 + \tau_{n_2} S_{0,n_2}(t)}.$$

- ***Procedures to find*** $P(t,k), n_k = k$**:**

$$P(t,k) = \frac{P(t,T_0) + \sum_{j=1}^{n_k-1} S_{0,n_k}(t)\tau_j P(t,T_j)}{1 + \tau_{n_2} S_{0,n_k}(t)}.$$

Table 19.4.2: OIS curve calibration instruments

Product/quote name	time range
CME Fed Fund Futures	0Y - 2Y
3M LIBOR/Fed fund rate swap	2Y - 50Y
3M LIBOR/SA fixed swap	2Y - 60Y

Table 19.4.3: LIBOR 3M forward curve calibration instruments

Product/quote name	time range
USD money market 3M deposit rate	0 - 3M
CME 3M Eurodollar futures	3M - 3Y
3M LIBOR/SA fixed swap	4Y - 50Y

Table 19.4.4: LIBOR 1M forward curve calibration instruments

Product/quote name	time range
USD money market 1M deposit rate	0 - 1M
USD FRA 1M LIBOR	0 - 3M
3M LIBOR/1M LIBOR basis swap	3M - 30Y

Table 19.4.5: LIBOR 6M forward curve calibration instruments

Product/quote name	time range
USD money market 1M deposit rate	0 - 6M
3M LIBOR/6M LIBOR basis swap	6M - 30Y

19.4.5 Correlation structure from market data

Lemma 19.4.3 (correlation structure in market forward curves). *[4, p. 100] Suppose we have forward curve $f(t, T_i), i = 1, 2, ..., M$ for M maturities(we can obtain these curve from calibration Lemma 19.4.1). Denote the sample points in these curves as $f_{n,i} \triangleq f(n\Delta\tau, T_i + n\Delta)$, where $\Delta\tau = 1/12$ represents the interval of one month. Suppose there are N such observations.*

Define the difference quantity

$$\Delta f_{n,i} = f((n+1)\Delta, T_i + (n+1)\Delta) - f(n\Delta, T_i + n\Delta), \forall i = 1, 2, ..., M.$$

We can calculate the following statistical quantities:

- ***(mean drift)**:*

$$\mu(T_i) = \overline{\Delta f_i} = \frac{1}{N} \sum_{n=0}^{N-1} \Delta f_{n,i}.$$

- ***($M \times M$ covariance matrix)**:*

$$C_{ij} = \frac{1}{N} \sum_{n=0}^{N-1} (\Delta f_{n,i} - \overline{\Delta f_i})(\Delta f_{n,j} - \overline{\Delta f_j}).$$

- ***spectral decomposition of covariance matrix***

$$C = \sum_{k=1}^{M} \lambda_k \Delta\tau \boldsymbol{v_k} \boldsymbol{v_k}^T,$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$.

- ***reconstructing forward curve parallel evolution dynamics:***

$$\boldsymbol{f}_{n+1} - \boldsymbol{f}_n = \sum_{k=1}^{M} \sqrt{\lambda_k \Delta\tau} \boldsymbol{v}_k \eta_{k,n},$$

where $\eta_{k,n} \sim N(0,1)$, *and*

$$f_n = \begin{bmatrix} f_{n,1} \\ f_{n,1} \\ \vdots \\ f_{n,M} \end{bmatrix}, \overline{f} = \begin{bmatrix} \overline{f_1} \\ \overline{f_2} \\ \vdots \\ \overline{f_M} \end{bmatrix}.$$

Remark 19.4.3 (observable market data)**.**

- At the beginning, the bond issuer just issue bonds of different maturities. The market prices of these bonds will enable us to construct $f(0,T), \forall T > 0$ via Lemma 19.4.1.
- As time goes on, the prices at different t will enable us to trace out $f(t,T), \forall T > 0$.
- The bond issuer will continue to issue new bonds. The market prices of these newly issued bond will enable us to trace out

$$f(\tau, T+\tau), \forall T, \tau > 0.$$

Lemma 19.4.4 (constructing HJM model from market data). *[4, p. 103]*

- ***Parallel shifting dynamics in real probability measure.***

$$df(t,t+T) = \boldsymbol{\sigma}^T d\boldsymbol{W}_t + \mu(T)dt.$$

where $\boldsymbol{\sigma} \in \mathbb{R}^M, \boldsymbol{\sigma}(T) = (v_1(T)\sqrt{\lambda_1}, v_2(T)\sqrt{\lambda_2}, ..., v_M(T)\sqrt{\lambda_M})$
- ***HJM formulation in real probability measure.***

$$df(t,T) = \boldsymbol{\sigma}^T(T-t)d\boldsymbol{W}_t + \mu(T-t)dt.$$

where $\boldsymbol{\sigma} \in \mathbb{R}^M, \boldsymbol{\sigma}(T) = (v_1(T)\sqrt{\lambda_1}, v_2(T)\sqrt{\lambda_2}, ..., v_M(T)\sqrt{\lambda_M})$
- ***HJM formulation in risk-neutral probability measure.***

$$df(t,T) = \boldsymbol{\sigma}^T d\boldsymbol{W}_t + \mu(T)dt.$$

where

$$\mu(t,T) = \boldsymbol{\sigma}^T(T-t)\int_t^T \boldsymbol{\sigma}(s-t)ds$$

Remark 19.4.4 (stationarity assumption)**.** Note that here the correlation structure extraction assumes that the forward curve dynamics is stationary.

19.4.6 A linear algebra approach

19.5 Short rate models: principles and examples

The short-rate model we model the evolution of r_t, as a stochastic process under a risk-neutral measure 0. then the price at time t of a zero-coupon bond maturing at time T with a payoff of 1 is given by where F is the natural filtration for the process. In such framework, the zero-coupon bond and variance interest rate (LIBOR, par swap rate) connected with zero-coupon bond can be derived.

In this section, we will discuss the principles governing a short rate model. Different favors of short rate model including single factor and multiple factors models, will be addressed in the following sections.

In this section, we discuss one class of short rate model, known as equilibrium models. This class of short rate models usually have two to three model parameters; therefore cannot calibrate to the market observed yield curve and admit arbitrage.

We study this class of models is not for practial application but to gain understanding on how short rate models work.

19.5.1 Principles of short-rate model

19.5.1.1 *Understand risk-neutral measure: martingale pricing*

Theorem 19.5.1 (Black model for interest rate derivatives). *[7, p. 513][6, p. 25] Consider the short rate follows a SDE under* ***real-world*** *probability measure, given as*

$$dr = u(r,t)dt + \sigma(r,t)dW_t.$$

Then

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r} + (u - \lambda\sigma)\frac{\partial V}{\partial r} - rV = 0.$$

There exists an equivalent measure Q, ***also known as risk-neutral measure****, under which*

$$dr = (u - \lambda(r,t)\sigma)dt + \sigma dW_t$$

and

$$e^{-\int_0^T r(\tau)d\tau}V(T,r_T)$$

is a martingale under Q.

The pricing formula under risk-neutral measure is given by

$$V(t, r_t) = E_Q[e^{-\int_t^T r(s)ds} V(T, r_T)]$$

Proof. Consider two zero-coupon bonds maturing at T_1 and T_2 with values denoted by $V_1(r, t, T_1)$ and $V_2(r, t, T_2)$. Then

$$\Pi = V_1 - \Delta V_2$$

and

$$d\Pi = \frac{\partial V_1}{\partial t}dt + \frac{\partial V_1}{\partial r}dr + \frac{1}{2}\sigma^2\frac{\partial V_1}{\partial r^2}dr - \Delta(\frac{\partial V_2}{\partial t}dt + \frac{\partial V_2}{\partial r}dr + \frac{1}{2}\sigma^2\frac{\partial^2 V_2}{\partial r^2}dt)$$

then choose

$$\Delta = \frac{\partial V_1}{\partial r} / \frac{\partial V_2}{\partial r}$$

will make $d\Pi$ deterministic.

Then

$$\frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial V_1}{\partial r^2} - rV_1}{\frac{\partial V_1}{\partial r}} = \frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V_2}{\partial r^2} - rV_2}{\frac{\partial V_2}{\partial r}}$$

Let

$$\frac{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} - rV}{\frac{\partial V}{\partial r}} = a(r, t).$$

If we write

$$a(r, t) = \sigma(r, t)\lambda(r, t) - u(r, t)$$

for a given $u(r, t)$ and non-zero $\sigma(r, t)$. Then

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r} + (u - \lambda\sigma)\frac{\partial V}{\partial r} - rV = 0.$$

From the Feyman Kac theorem(Theorem 15.8.1), there exists an equivalent measure Q, also known as risk-neutral measure, under which

$$dr = (u - \lambda(r, t)\sigma)dt + \sigma dW_t$$

and

$$e^{\int_0^T r(\tau)d\tau} P(T)$$

is a martingale under Q. □

Remark 19.5.1. It can be showed that $\lambda(r, t)$ is a proper form; that is

$$\lambda(r, t) = (a(r, t) + u(r, t))/\sigma(r, t)$$

Note 19.5.1 (interpretation).

- In stock option pricing, we can hedge option by traded the underlying stochastic asset. For bond option pricing, the underlying stochastic quantity, i.e., interest rate, is not tradable. Therefore, we have to use bonds of different maturities to hedge the risks.
- **Knowing real-world interest rate dynamics is insufficient for pricing**. We also need to know to λ in order to price. This is because by hedging risks can change the price.
- **market risk has to be fitted from market**. Since λ usually has to be fitted from the market, an alternative method is directly assume the short rate dynamics under risk-neutral measure is given by

$$dr = mdt + \sigma dW_t.$$

  and then fit m from the market. This is the approach in Theorem 19.5.2.
- **Practical difficulties in maringlate pricing formula even we know short rate dynamics under** Q:
  - r_t is not observed, therefore r_T is difficult to compute.
  - **For complex derivatives,** the expectation of $E_Q[e^{-\int_t^T r(s)ds} V(T, r_T)]$ is difficult to evaluate since the joint distribution between $e^{-\int_t^T r(s)ds}$ and $V(T, r_T)$ is unknown.(even though we might be able to evaluate $E_Q[V(T, r_T)]$ and $E_Q[V(T, r_T)]$ separately.)
  - The method is tractable only for simple derivatives, for example, zero coupon bond where $V(T, r_T) = 1$.
- **Monte carlo simulation:** we can certainly use Monte carlo

Remark 19.5.2. Note that similar method has been used in Lemma 17.4.1.

Lemma 19.5.1 (interest rate derivative dynamics under risk-neutral measure). *Consider the short rate follows a SDE under* ***risk-neutral*** *probability measure, given by*

$$dr = m(r, t)dt + \sigma(r, t)dW_t.$$

Let V be a derivative on short rate r. Under risk-neutral measure, the dynamics of V is given by

$$dV = rVdt + \frac{\partial V}{\partial r}\sigma dW_t.$$

Proof. From Ito lemma, we have

$$\begin{aligned} dV &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial r}dr + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}(dr)^2 \\ &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial r}dr + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}\sigma^2 dt \end{aligned}$$

From Theorem 19.5.1, we know that $V(r,t)$ has to satisfy

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r} + (u - \lambda\sigma)\frac{\partial V}{\partial r} - rV = 0.$$

Then the dynamics of V is equivalent to

$$dV = rV + \frac{\partial V}{\partial r}\sigma dW_t.$$

□

19.5.1.2 *Term structure equation*

Theorem 19.5.2 (term structure equation). *[5, p. 707][8, p. 274] Let the short rate r under risk-neutral measure Q follows the general model*

$$dr = m(r,t)dt + s(r,t)dz,$$

where z is a Brownian motion. Then

- *Any derivative with value $f(r(t),t)$ dependent on r and t is governed by*

$$df = (\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial r} + \frac{1}{2}s^2\frac{\partial f}{\partial r^2})dt + s\frac{\partial f}{\partial r}dz.$$

- *Under the risk-neutral measure Q, we have*

$$\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial r} + \frac{1}{2}s^2\frac{\partial f}{\partial r^2} = rf,$$

  with final condition $f(r(T),T) = V(r)$.

Proof. (1) Use Ito's lemma.(Lemma 6.3.1).(2) Use Feyman-Kac theorem Theorem 15.8.1. See more details in Theorem 19.5.1. □

Corollary 19.5.2.1 (term-structure equation of a zero-coupon bond). *[8, p. 274][1, p. 43] Given the stochastic short rate model under the risk-neutral measure Q*

$$dr = m(r,t)dt + s(r,t)dz.$$

It follows that

- *Further assume the price $P(t,T)$ of a zero-coupon bond matures at T is given by $P(t,r(t),T) = f(r(t),t)$. Then $f(r(t),t)$ governed by*

$$\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial r} + \frac{1}{2}s^2\frac{\partial f}{\partial r^2} = rf$$

  with boundary condition $f(r,T) = P(T,r(T),T) = 1, \forall r$.
- *$P(t,T)$ has the following SDE representation:*

$$dP(t,T) = r(t)P(t,T)dt + s(r,t)\frac{\partial P}{\partial r}dW_t.$$

Proof. (1) Direct consequence of Theorem 19.5.2. (2) Using Ito rule,

$$\begin{aligned} dP(t,T) &= \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial r}dr + \frac{1}{2}\frac{\partial^2 P(t,T)}{\partial r \partial r}(dr)^2 \\ &= \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial r}(mdt + sdW_t) + \frac{1}{2}\frac{\partial^2 P(t,T)}{\partial r \partial r}s^2 dt \\ &= (\frac{\partial P}{\partial t} + \frac{\partial P}{\partial r}m + \frac{1}{2}\frac{\partial^2 P(t,T)}{\partial r \partial r}s^2)dt + \frac{\partial P}{\partial r}sdW_t \\ &= r(t)P(t,T)dt + \frac{\partial P}{\partial r}sdW_t \end{aligned}$$

where we use the term structure equation from (1).

□

Remark 19.5.3. The solution will gives the bond price $P(t,r(t),T)$ as a function of short rate at $r(t)$.

Example 19.5.1 (Hull-White interest rate model and zero-coupon bond pricing). [8, p. 274] In the Hull-White model, the evolution of the interest rate is given by

$$dr(t) = (a(t) - b(t)r(t))dt + \sigma(t)dW(t),$$

where $a(t), b(t)$ and $\sigma(t)$ are deterministic function of time. The governing equation for $P(t,T;r(t)) = f(r(t),t)$ is given by

$$\frac{\partial f}{\partial t} + (a(t) - b(t)r)\frac{\partial f}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial f}{\partial r^2} = rf,$$

with final condition $f(r,T) = 1, \forall r$.

Suppose this PDE has solution

$$P(t,r(t),T) = f(r(t),t) = \exp(-rC(t,T) - A(t,T)),$$

then

$$\begin{aligned} f_t(t,r) &= (-rC'(t,T) - A'(t,T))f(t,r) \\ f_r(t,r) &= -C(t,T)f(t,r), \\ f_{rr}(t,r) &= C^2(t,T)f(t,r) \end{aligned}$$

where $C'(t,T) = \frac{\partial}{\partial t}C(t,T)$, $A'(t,T) = \frac{\partial}{\partial t}A(t,T)$.

Substitute into the term structure function we get

$$[(-C'(t,T) + b(t)C(t,T) - 1)r - A'(t,T) - a(t)C(t,T) + \frac{1}{2}\sigma^2(t)C^2(t,T)]f(t,r) = 0.$$

Because this equation holds for all r, then we must have

$$\begin{aligned} -C'(t,T) + b(t)C(t,T) - 1 = 0 \\ A'(t,T) = -a(t)C(t,T) + \frac{1}{2}\sigma^2(t)C^2(t,T). \end{aligned}$$

Solve $C(t,T)$ and $A(t,T)$ with boundary condition $C(T,T) = A(T,T) = 0$, we have

$$\begin{aligned} C(t,T) &= \int_t^T e^{(-\int_t^s b(v)dv)}ds \\ A(t,T) &= \int_t^T (a(s)C(s,T) - \frac{1}{2}\sigma^2(s)C^2(s,T))ds. \end{aligned}$$

Then, the yield curve is given as

$$Y(t,T) = -\frac{1}{T-t}\log P(t,r(t),T) = \frac{1}{T-t}(rC(t,T) + A(t,T)).$$

That is, $Y(t,T)$ is an **affine function** of r. More deeply, since $r(t)$ is the random variable, $Y(t,T)$ is also a random variable.

Example 19.5.2 (CIR interest rate model and zero-coupon bond pricing). [8, p. 275] In the CIR model, the evolution of the interest rate is given by

$$dr(t) = (a - b(t)r(t))dt + \sigma\sqrt{r(t)}dW(t),$$

where a, b and σ are constants. The governing equation for $P(t,T;r(t)) = f(r(t),t)$ is given by

$$\frac{\partial f}{\partial t} + (a - br)\frac{\partial f}{\partial r} + \frac{1}{2}\sigma^2 r\frac{\partial f}{\partial r^2} = rf,$$

with final condition $f(r,T) = 1, \forall r$.

Suppose this PDE has solution

$$P(t,r(t),T) = f(r(t),t) = \exp(-rC(t,T) - A(t,T)),$$

then

$$\begin{aligned}
f_t(t,r) &= (-rC'(t,T) - A'(t,T))f(t,r) \\
f_r(t,r) &= -C(t,T)f(t,r), \\
f_{rr}(t,r) &= C^2(t,T)f(t,r)
\end{aligned}$$

where $C'(t,T) = \frac{\partial}{\partial t}C(t,T)$, $A'(t,T) = \frac{\partial}{\partial t}A(t,T)$.

Substitute into the term structure function we get

$$[(-C'(t,T) + bC(t,T) - 1 + \frac{1}{2}\sigma^2(t)C^2(t,T))r - A'(t,T) - a(t)C(t,T)]f(t,r) = 0.$$

Because this equation holds for all r, then we must have

$$\begin{aligned}
-C'(t,T) + b(t)C(t,T) - 1 + \frac{1}{2}\sigma^2(t)C^2(t,T) = 0 \\
A'(t,T) = -a(t)C(t,T).
\end{aligned}$$

Solve $C(t,T)$ and $A(t,T)$ with boundary condition $C(T,T) = A(T,T) = 0$, we have

$$\begin{aligned} C(t,T) &= \frac{sinh(\gamma(T-t))}{\gamma cosh(\gamma(T-t)) + \frac{1}{2}bsinh(\gamma(T-t))} \\ A(t,T) &= -\frac{2a}{\sigma^2}\ln(\frac{\gamma\exp(\frac{1}{2}b(T-t))}{\gamma cosh(\gamma(T-t)) + \frac{1}{2}bsinh(\gamma(T-t))}) \\ \gamma &= \frac{1}{2}\sqrt{b^2+2\sigma^2}. \end{aligned}$$

Remark 19.5.4 (practical difficulties). In the short model method, we can find the yield curve and bond price to be a function of short rate $r(t)$. However, the short rate $r(t)$ cannot be directly observed in the market.

19.5.1.3 *Instruments related to short rates*

Definition 19.5.1 (money market account). *Let the cash account be*

$$M(t) = M(0)\exp(\int_0^t r(s)ds).$$

Lemma 19.5.2 (zero-coupon bond pricing). *Suppose the risk-free rate $r(t)$ is stochastic. Assume no-arbitrage condition holds and the market is complete.*

- *There exists a measure Q, equivalent to P, under which, for each T, the discounted price process $P(t,T)/B(t)$ is a martingale for all $t : 0 < t < T$.*
- $$P(t,T) = B(t)E_Q[\exp(-\int_0^T r(s)ds)P(T,T)|\mathcal{F}_t] = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t].$$

Proof. Use the fundamental theorem **??**, $P(t,T)/B(t)$ is a martingale. Also note that we use the fact that $P(T,T) = 1, B(t)$ is $\mathcal{F}_t$ measurable. □

Remark 19.5.5 (no-arbitrage restrictions for short rate model under risk-neutral measure).

- Unlike dynamic models for tradable assets (e.g., stocks), the dynamic model should satisfy that discounted asset prices are martingales under the equivalent martingale measure. There is almost no restriction on coefficients in a typical short rate model under risk-neutral measure given by

$$dr(t) = \alpha(t)dt + \sigma(t)dW_t.$$

- The discounted zero-coupon bond price is given by

$$\frac{P(t,T)}{M(t)} = E_Q[\exp(-\int_0^T r(s)ds)|\mathcal{F}_t],$$

which is a martingale under risk-neutral measure since $\exp(-\int_0^T r(s)ds)$ is a fixed random variable (using conditional expectation and Theorem 5.5.1).

Lemma 19.5.3 (SDE of bond price under risk-neutral measure). *Assume, under real-world measure, the SDE for the price of zero-coupon bond $P(t,T)$ is given by*

$$dP(t,T) = \mu(t,P)P(t,T)dt + \sigma(t,P)P(t,T)dW(t)$$

*where $W(t)$ is the Brownian motion. Let $r(t)$ be the (stochastic)short rate. We further **assume there is no arbitrage opportunities**. Then there exists a risk-neutral measure(i.e. use money account as the numeraire) such that*

$$dP(t,T) = r(t)P(t,T)dt + \sigma(t,P)P(t,T)d\hat{W}(t)$$

where $\hat{W}(t)$ is the Brownian motion under the risk-neutral measure.

Proof. Under no-arbitrage condition(Theorem 15.5.1), we can write the dynamics of bond price and money account as

$$\begin{aligned} dP(t,T) &= (r + \lambda_p\sigma(t,P))P(t,T)dt + \sigma(t,P)P(t,T)dW(t) \\ dB(t) &= rB(t,T)dt \end{aligned}$$

When using money account as numeraire, we should set $\lambda_P = 0$(Theorem 15.6.19). Therefore, under the risk-neutral measure, we have

$$dP(t,T) = r(t)P(t,T)dt + \sigma(t,P)P(t,T)d\hat{W}(t)$$

where $\hat{W}(t)$ is the Brownian motion under the risk-neutral measure. □

19.5.2 Merton model

19.5.2.1 *Basics*

Definition 19.5.2 (Merton model). *[1, p. 44] In Merton model, the SDE for the short rate **under risk-neutral measure** is*

$$dr(t) = \alpha dt + \sigma dW(t),$$

where α and σ are constants, and $W(t)$ is a Brownian motion under the risk-neutral measure Q.

Lemma 19.5.4 (zero-coupon bond price in Merton model). *[1, p. 45] Consider the short rate under risk neutral measure is evolving under the Merton model with drift α and volatility σ. We have*

- *The solution for the short rate is given by*
$$r(s) = r(t) + \alpha(s-t) + \sigma(W(s) - W(t)).$$
And
$$r(s) \sim N(r(t) + \alpha(s-t), \sigma^2(s-t)), s > t.$$
-
$$\int_t^T r(s) = r(t)(T-t) + \frac{1}{2}\alpha(T-t)^2 + \sigma \int_t^T (T-s)dW(s),$$
which is the Gaussian process with mean
$$m(t) = r(t)(T-t) + \frac{1}{2}\alpha(T-t)^2$$
and variance
$$s^2 = \frac{1}{3}\sigma^2(T-t)^3$$
under risk-neutral measure.
- *The zero-coupon bond price in the Merton model is given as*
$$P(t,T) = \exp((-r(t)(T-t) - \frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3),$$
and $P(T,T) = 1$.
Define $A(t,T) = \exp(-\frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3), B(t,T) = T-t$. Then we can write
$$P(t,T) = A(t,T)\exp(-B(t,T)r(t)).$$
- *The yield curve $y(t,T)$ in the Merton model is given by*
$$y(t,T) \triangleq -\frac{\log P(t,T)}{T-t} = r(t) + \frac{1}{2}\alpha(T-t) - \frac{1}{6}\sigma^2(T-t)^2.$$
-
$$dP(t,T) = r(t)P(t,T)dt - (T-t)P(t,T)\sigma dW(t)$$

Proof. (1) See Lemma 6.3.11.

(2) Since $r(t)$ is $\mathcal{F}_t$ measurable and X is independent of $\mathcal{F}_t$, We have

$$\begin{aligned} P(t,T) &= E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] \\ &= E_Q[\exp(X)] \\ &= M_X(-1) = \exp(-m + \frac{1}{2}s^2) \end{aligned}$$

where $X(t)$ is the Gaussian random variable $\exp(-\int_0^T r(s)ds)$, and M_X is the moment generating function of X (3) straight forward. (4) Use

$$dP = \frac{\partial P}{\partial t} + \frac{\partial P}{\partial r}dr + \frac{\partial^2 P}{\partial r^2}drdr$$

and

$$\begin{aligned} \frac{\partial P}{\partial t} &= (r + \alpha(T-r) - \frac{1}{2}\sigma^2(T-t)^2)P \\ \frac{\partial P}{\partial r} &= -(T-t)P \\ \frac{\partial^2 P}{\partial r^2} &= (T-t)^2 P \end{aligned}$$

□

Note 19.5.2 (equivalence to PDE method). It can be showed that that solution

$$P(t,r,T) = \exp(-r(T-t) - \frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3)$$

also satisfy the term-structure equation

$$\frac{\partial P}{\partial t} + \alpha\frac{\partial P}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial r^2} = rP,$$

given by Theorem 19.5.2. Note that

$$\begin{aligned} \frac{\partial P}{\partial t} &= (r + \alpha(T-r) - \frac{1}{2}\sigma^2(T-t)^2)P \\ \frac{\partial P}{\partial r} &= -(T-t)P \\ \frac{\partial^2 P}{\partial r^2} &= (T-t)^2 P \end{aligned}$$

Then
$$\frac{\partial P}{\partial t} + \alpha\frac{\partial P}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial r^2} = rP.$$

Remark 19.5.6 (risk-neutral process is not a real world process)**.** Note that the model is to model how the short rate will move under risk-neutral probability measure instead of the real-world probability measure. Why we do this is discussed at Theorem 19.5.1.

Remark 19.5.7 (shortcomings of Merton model)**.**

- One problem is that short rate can be negative.
- The model lacks mean-reversion property.
- Not Arbitrage free.

19.5.2.2 *Yield curve in Merton model*

Remark 19.5.8 (construction of the yield curve and model calibration)**.** Note that the yield curve can be constructed by
$$R(t,T) = -\frac{\log P(t,T)}{T-t} = (r(t) + \frac{1}{2}\alpha(T-t) - \frac{1}{6}\sigma^2(T-t)^2).$$

We have the following interpretation

- we can calibrate the yield curve $R(0,T)$ to market yield curve and use this model to predict future yield curve.
- larger volatility in the interest rate will decreases yield; particularly, it will decrease the value of long-term yield more than the short-term.
- **parallel shift property contradicts market observation:** because α,σ is constant, shift t and T at the same time will not change yield curve. However, usually in the market $R(t+\tau, T+\tau)$ will change(that is, the price of newly issue bonds with the same maturities will be different).
- **lack of flexibility:** this model only has two parameters and usually it is difficult to fit the market yield curve. Mathematically,
$$\frac{dP(t+s,T+s)}{ds} = \frac{\partial P}{\partial t}ds + \frac{\partial P}{\partial T}ds = 0.$$
- **possible arbitrage opportunities:** We usually assume the market contains no arbitrages. Since the model usually cannot fit to the current market term structure, it may contains arbitrage opportunities.

Remark 19.5.9 (shape control of yield curve)**.** Let the current time be 0. A yield curve $y(0,T)$ is characterized by three parameters: $r(0),\alpha,\sigma$. As showed in Figure 19.5.1,

- $r(0)$ controls the overall level.

- α controls the up slopping.
- σ controls the bending down. The larger volatility will tend to bend down the yield curve.

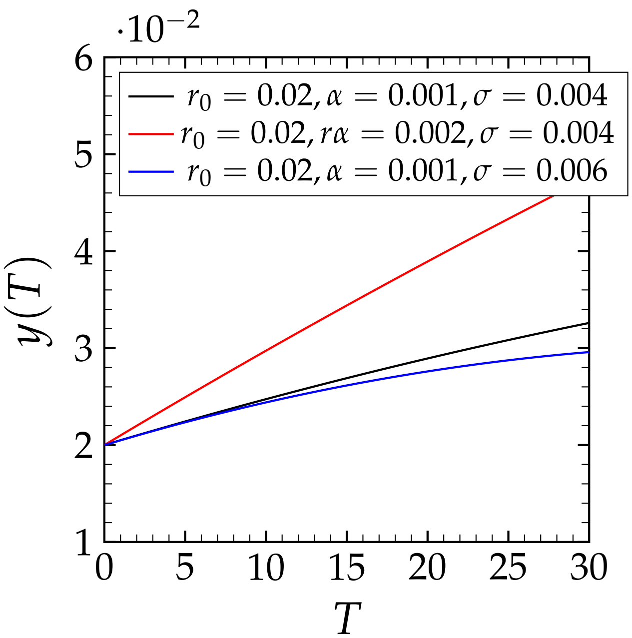


Figure 19.5.1: Current yield curve resulting from Merton's model with different parameters.

Lemma 19.5.5 (evolution of yield curves). *Consider the short rate under risk neutral measure is evolving under the Merton model with drift α and volatility σ.*

- ***(evolution of whole yield curve)*** *Let the current time be oïijŇ then the future yield curve $y(\tau, T+\tau)$ is given by*

$$\begin{aligned} y(\tau, T+\tau) &= r(\tau) + \frac{1}{2}\alpha T + \frac{1}{6}\sigma^2 T^2 \\ &= \alpha\tau + \sigma W(\tau) + y(0, T) \end{aligned}$$

  where $y(0, T)$ is the current yield curve.
- ***evolution of yield associated with a fixed maturity date*** *Let the current time be oïijŇ then the future yield $y(\tau, T)$ is given by*

$$\begin{aligned} y(\tau, T) &= r(\tau) + \frac{1}{2}(T-\tau) + \frac{1}{6}\sigma^2 (T-\tau)^2 \\ &= r(0) + \alpha\tau + \sigma W(\tau) + \frac{1}{2}(T-\tau) + \frac{1}{6}\sigma^2 (T-\tau)^2 \end{aligned}$$

Particularly, $y(T,T) = r(0) + \alpha T + \sigma W(T)$.

Proof. □

Methodology 19.5.1 (simulate the future term structure in Merton model). *Suppose*

- *current time is 0.*
- *we are given the current term structure* $y(0,T), T \in \mathbb{R}^+$.
- *we are given the Merton model*

$$dr = \alpha dt + \sigma dW_t.$$

- *we are given the initial short rate* $r(0)$.

Then we can generate a ***sample term structure*** *in future time* $t > 0$ *via the following procedure:*

- *simulate* $r(t)$ *by drawing from the normal distribution*

$$r(t) \sim N(r(0) + \alpha t, \sigma^2 t).$$

- *the sample yield curve is given by*

$$y(t,T) = r(t) + \frac{1}{2}(T-t) + \frac{1}{6}\sigma^2(T-t)^2.$$

Remark 19.5.10. As showed in Figure 19.5.2, the initial yield curve is given by

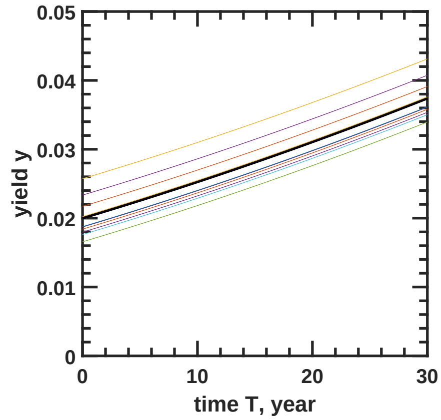
$$y(0,T) = r_0 + \frac{1}{2}\alpha T + \frac{1}{6}\sigma^2 T^2,$$

where $r_0 = 0.02, \alpha = 0.001, \sigma = 0.004$.

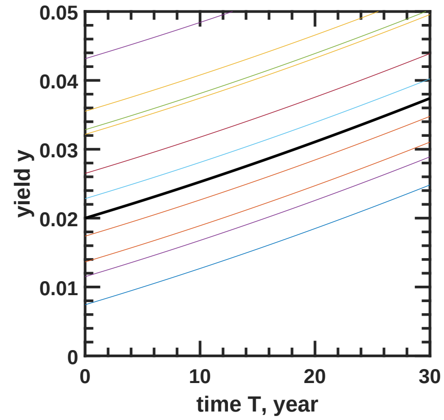
- Figure A shows the different realization of the yield curve $y(1, T+1)$ after 1 year. We can see that only parallel dynamics of yield curve is allowed in the Merton model. There is no mean-reversion in these yield curve realizations..
- Figure B shows the different realization of the yield curve $y(1, T+1)$ after 10 year. Compared to the 1 year's realizations, the yield curve can move far away from the initial yield curve.
- Figure C and D shows the 5Y and 10Y zero rate evolution realizations in the Merton model. These trajectories can also be used to calculate the price trajectories of a zero coupon bond maturing in 5Y and 10Y via

$$P(t,T) = \exp(-y(t,T)(T-t)),$$

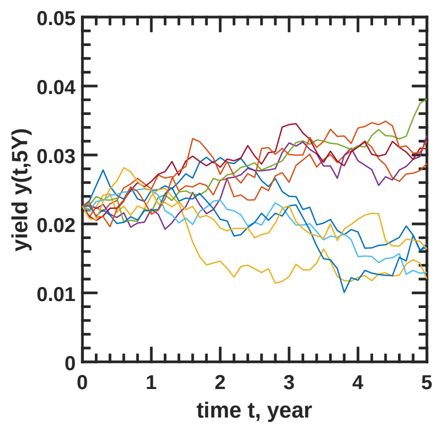
as we showed in Figure E and F.



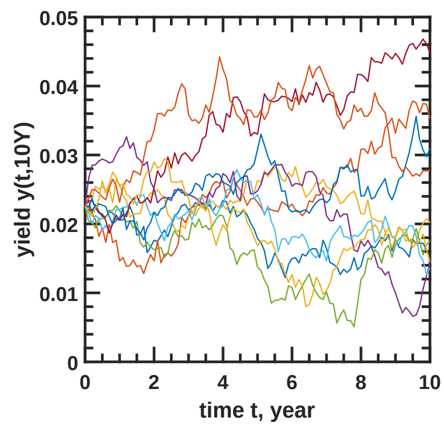
(a) Realizations of yield curve after 1Y (thin lines) from the initial yield curve(thick line).



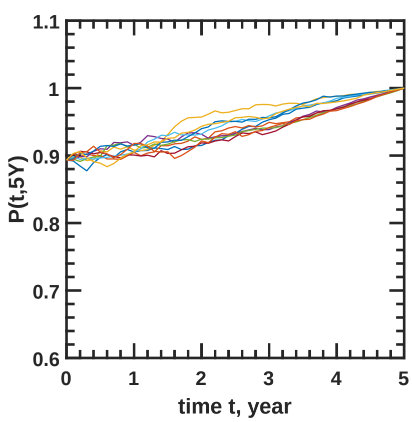
(b) Realizations of yield curve after 5Y (thin lines) from the initial yield curve(thick line).



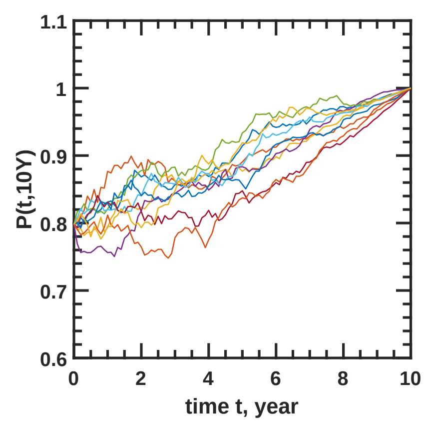
(c) Evolution of yield $y(t, 5Y)$ from the initial value $y(0, 5Y)$.



(d) Evolution of yield $y(t, 10Y)$ from the initial value $y(0, 10Y)$.



(e) 5Y zero-coupon bond price trajectories.



(f) 10Y zero-coupon bond price trajectories.

Figure 19.5.2: Yield curve dynamics in Merton model

19.5.2.3 *Instantaneous forward rate dynamics*

Lemma 19.5.6 (instantaneous forward rate dynamics in Merton model). *Consider the short rate under risk neutral measure is evolving under the Merton model with drift α and volatility σ.*

- *The instantaneous forward rate is given by*

$$f(t,T) = r(t) + \alpha(T-t)^2 - \frac{1}{2}\sigma^2(T-t)^2.$$

- *The dynamics of instantaneous forward rate is given by*

$$df(t,T) = \sigma^2(T-t)dt + \sigma dW_t$$

Proof. (1) Note that in the Merton model, we have zero-coupon bond price given by

$$P(t,T) = \exp((-r(T-t) - \frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3),$$

and the instantaneous forward rate is given by

$$f(t,T) = -\frac{\ln P(t,T)}{dT} = r(t) + \alpha(T-t) - \frac{1}{2}\sigma^2(T-t)^2.$$

(2)

$$\begin{aligned} df(t,T) &= dr(t) - \alpha dt + \sigma^2(T-t)dt \\ &= \alpha dt + \sigma dW_t - \alpha dt + \sigma^2(T-t)dt \\ &= \sigma^2(T-t)dt + \sigma dW_t \end{aligned}$$

□

19.5.2.4 *Derivative pricing*

Lemma 19.5.7 (zero-coupon bond price and forward price SDE in Merton model). *Given the short rate SDE under risk-neutral measure*

$$dr(t) = \alpha dt + \sigma dW(t),$$

the zero-coupon bond price $P(t,T)$ satisfies SDE

$$dP(t,T) = P(t,T)(r(t) + \alpha(T-t) - \frac{\sigma^2}{2}(T-t)^2)dt - P(t,T)(T-t)\sigma dW(t).$$

Moreover, the forward bond price $FP(t;S,T) \triangleq \frac{P(t,T)}{P(t,S)}$ *satisfies SDE*

$$dFP(t;S,T) = FP(t;S,T)\sigma(T-S)dW^S(t),$$

where W^S *is a Brownian motion under forward measure with respect to* $P(t,S)$*.*

Proof. (1) From Ito lemma, we have

$$dP(t,r,T) = \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial r}dr + \frac{1}{2}\frac{\partial^2 P}{\partial r^2}drdr.$$

Note that

$$\begin{aligned}\frac{\partial P}{\partial t} &= P(t,T)(r(t) + \alpha(T-t) - \frac{\sigma^2}{2}(T-t)^2) \\ \frac{\partial P}{\partial r}dr &= -P(t,T)(T-t)dr \\ \frac{\partial^2 P}{\partial r^2}drdr &= P(t,T)(T-t)^2\sigma^2 dt\end{aligned}$$

then

$$dP(t,T) = P(t,T)r(t)dt - P(t,T)(T-t)\sigma dW(t).$$

(2) use Theorem 19.3.1, we have

$$dFP(t;S,T) = FP(t;S,T)(\sigma_{P(t,T)} - \sigma_{P(t,S)})dW^S(t).$$

□

Lemma 19.5.8 (Price of European call option on zero-coupon bond). *The value at time* t *of a European call option maturing at* S *with strike* K *on a bond maturing at* $T > S$ *is given by*

$$V(t) = P(t,S)E_S[\max(P(S,T)-K,0)|\mathcal{F}_t],$$

where the expectation is taken with respect to forward measure of $P(t,S)$*. It can be showed that*

$$V(t) = PC(t;S,T,K) = P(t,T)N(d_+) - KP(t,S)N(d_-),$$

where

$$d_+ = \frac{\log\frac{FP(t;S,T)}{K} + \frac{1}{2}v(t,S)^2}{v(t,S)} = \frac{\log\frac{B(t,T)}{KB(t,S)} + \frac{1}{2}v(t,S)^2}{v(t,S)}, d_- = d_+ - v(t,S).$$

and

$$FP(t;S,T) \triangleq P(t,T)/P(t,S)$$

$$v(t,S)^2 = \int_t^S (\sigma(T-S))^2 dt = (\sigma(T-S))^2(S-t)$$

Proof. Note that our bond dynamics is given by

$$dP(t,T) = P(t,T)r(t)dt - P(t,T)(T-t)\sigma dW(t).$$

Then, we can use Lemma 19.3.10 □

Remark 19.5.11 (general discussion). For a general discussion on pricing bond options based on bond dynamics, see Lemma 19.3.10.

19.5.3 Vasicek model

19.5.3.1 *Basics*

Definition 19.5.3 (Vasicek model). *[5, p. 684][1, p. 45] In Vasicek model, the SDE for r under risk-neutral measure Q*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW_t$$

where W_t is the Brownian motion under risk-neutral measure, θ, α, σ are constants.

Remark 19.5.12 (general characteristics).

- The model has mean reversion to θ.
- One drawback is that negative risk-free rate can occur.

Lemma 19.5.9 (short rate solution in the Vasicek model). *Assume the short rate, under risk-neutral measure, following Vasicek SDE*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW(t).$$

Let $0 \leq t \leq s \leq T$. It follows that

- *The short rate solution in the Vasicek SDE is given by*

$$r(s) = r(t)\exp(-\alpha(s-t)) + \frac{\theta}{\alpha}(1-\exp(-\alpha(s-t))) + \sigma\int_t^s \exp(-\alpha(s-u))dW(u).$$

- ***Conditioning on** $\mathcal{F}_t$, we have*

$$E[r(s)|\mathcal{F}_t] = r(t)\exp(-\alpha(s-t)) + \frac{\theta}{\alpha}(1-\exp(-\alpha(s-t))),$$

$$Var[r(s)|\mathcal{F}_t] = \frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha(s-t))).,$$
$$r(s)|\mathcal{F}_t \sim N(E[r(s)|\mathcal{F}_t], Var[r(s)|\mathcal{F}_t]).$$

Proof. See the OU process results in section 6.4. □

Lemma 19.5.10 (zero-coupon bond price in Vasicek model). *[1, p. 48] Assume the short rate, under risk-neutral measure, following SDE*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW(t).$$

Denote

$$D(t,T) = \int_t^T e^{-\alpha(s-t)} ds.$$

We have

- *Let $s \geq t$, we have*

$$r(s) = r(t)\exp(-\alpha(s-t)) + \frac{\theta}{\alpha}(1 - \exp(-\alpha(s-t))) + \sigma \int_t^s \exp(-\alpha(s-u))dW(u).$$

- *Let $I(t,T) \triangleq \int_t^T r(s)ds$. Then conditioning on $\mathcal{F}_t$, $I(t,T)$ has Gaussian distribution with mean $M(t,T)$ and variance $V(t,T)$ given by*

$$M(t,T) = r(t)\frac{1-\exp(-\alpha(T-t))}{\alpha} + \frac{\theta}{\alpha}(T-t) - \frac{\theta}{\alpha}\frac{1-\exp(-\alpha(T-t))}{\alpha}$$

$$V(t,T) = \frac{\sigma^2}{\alpha^2}(T - t + \frac{2}{\alpha}\exp(-\alpha(T-t)) - \frac{1}{2\alpha}\exp(-2\alpha(T-t)) - \frac{3}{2\alpha})$$

- *The zero-coupon bond price is given by*

$$P(t,T) = E_Q[\exp(-I(t,T))|\mathcal{F}_t] = \exp(-M(t,T) + \frac{1}{2}V(t,T)).$$

Define

$$B(t,T) = \frac{1-\exp(-\alpha(T-t))}{\alpha},$$

$$A(t,T) = \exp((\frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2})(B(t,T) - (T-t)) - \frac{\sigma^2}{4\alpha}B^2(t,T)),$$

then

$$P(t,T) = A(t,T)\exp(-r(t)B(t,T)).$$

- *The yield curve is given by*

$$y(t,T) = -\frac{\ln P(t,T)}{T-t} = r(t)\frac{B(t,T)}{T-t} - \frac{\ln A(t,T)}{T-t}.$$

Proof. (1) From above lemma. (2)(a)Note that given the observation $r(t)$ at t, we have

$$r(s) = r(t)\exp(-\alpha(s-t)) + \frac{\theta}{\alpha}(1-\exp(-\alpha(s-t))) + \sigma\int_t^s \exp(-\alpha(s-u))dW(u).$$

Therefore,

$$\begin{aligned}
&\int_t^T r(s)ds \\
&= \int_t^T r(t)e^{-\alpha(s-t)}ds + \frac{\theta}{\alpha}\int_t^T (1-e^{-\alpha(s-t))})ds + \int_t^T\int_t^s \sigma e^{-\sigma(s-u)}dW(u)ds \\
&= r(t)\frac{1-e^{-\alpha(T-t)}}{\alpha} + \frac{\theta}{\alpha}(T-t) - \frac{\theta}{\alpha}\frac{1-e^{-\alpha(T-t)}}{\alpha} + \sigma\int_t^T\int_s^T e^{-\alpha(s-u)}dudW(s) \\
&= r(t)\frac{1-e^{-\alpha(T-t)}}{\alpha} + \frac{\theta}{\alpha}(T-t) - \frac{\theta}{\alpha}\frac{1-e^{-\alpha(T-t)}}{\alpha} + \int_t^T \frac{\sigma}{\alpha}(1-e^{-\alpha(T-s)})dW(s)
\end{aligned}$$

where we changed the order of integration. From this, we note that

$$E[\int_t^T x_1(u)du] = r(t)\frac{1-\exp(-\alpha(T-t))}{\alpha} + \frac{\theta}{\alpha}(T-t) - \frac{\theta}{\alpha}\frac{1-\exp(-\alpha(T-t))}{\alpha}.$$

(b) For $Var[\int_t^T r(s)ds]$, we have

$$\begin{aligned}
&Var[\int_t^T r(s)ds \\
&= E[\int_t^T \frac{\sigma}{\alpha}(1-\exp(-\alpha(T-s)))dW(s)\int_t^T \frac{\sigma}{\alpha}(1-\exp(-\alpha(T-s)))dW(s)] \\
&= \frac{\sigma^2}{\alpha^2}(\int_t^T ds + \int_t^T \exp(-2\alpha(T-s))ds - 2\int_t^T \exp(-\alpha(T-s))ds) \\
&= \frac{\sigma^2}{\alpha^2}(T-t+\frac{2}{\alpha}\exp(-\alpha(T-t)) - \frac{1}{2\alpha}\exp(-2\alpha(T-t)) - \frac{3}{2\alpha}).
\end{aligned}$$

(3) If we define $B(t,T) = \frac{1-exp(-\alpha(T-t))}{\alpha}$,then

$$M(t,T) = r(t)B(t,T) + \frac{\theta}{\alpha}(T-t) - \frac{\theta}{\alpha}B(t,T),$$

$$V(t,T) = \frac{\sigma^2}{\alpha^2}(T-t) - \frac{\sigma^2}{\alpha^2}B(t,T) - \frac{\sigma^2}{2k}B^2(t,T).$$

□

Remark 19.5.13 (applications to price zero-coupon bond). Use Vasicek model, the stochastic short rate model is given as

$$dr = a(b-r)dt + \sigma dz,$$

the price $P(t,T)$ of a zero-coupon bond matures at T is governed by

$$\frac{\partial P}{\partial t} + a(b-r)\frac{\partial P}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial P}{\partial r^2} = rP$$

with boundary condition $P(T,T) = 1$.

Remark 19.5.14 (model calibration to market). Suppose we are given prices of N zero-coupon bond with different maturities at $t = 0$. Then, we can use nonlinear least square to solve model parameters θ, α, σ such that the error between the theoretical prediction and market observation is minimized.

Note that because we only have 3 parameters, it is usually difficult to match with the term structure of the market.

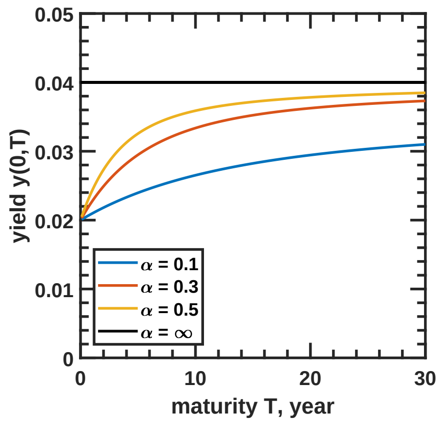
19.5.3.2 *Yield curve dynamics*

Remark 19.5.15 (yield curve shape). From Figure 19.5.3, we can see that

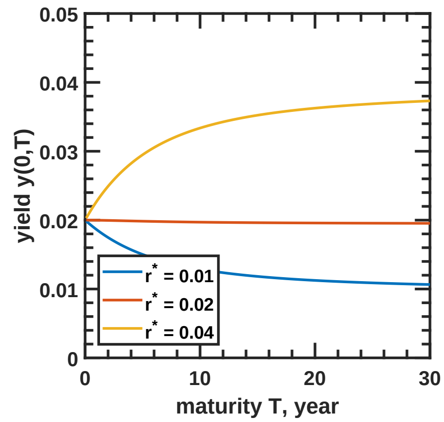
- (reversion speed effect) If $\alpha \to \infty$(but θ/α remains to be a constant), we can see that

$$B(t,T) = 0, A(t,T) = \exp(-\frac{\theta}{\alpha})(T-t)) = \exp(-r^*(T-t)), y(t,T) = r^*.$$

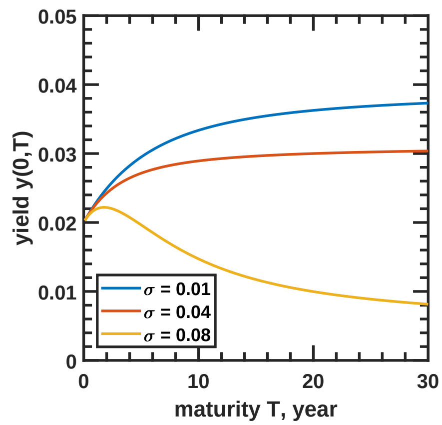
  If reversion speed α increases, then the yield curve will more resemble the equilibrium flat yield curve given by $y(t,T)* = r^*$.
- (mean level effect) The mean level will tend to pull the far-end of the yield curve toward the mean level.
- (volatility level effect) The volatility level tend to pull the far-end of the yield curve down.



(a) Yield curve shape vs. reversion speed parameter.



(b) Yield curve shape vs. mean level $r^* = \frac{\theta}{\alpha}$.



(c) Yield curve shape vs. volatility.

Figure 19.5.3: Control yield curve shape via Vasicek model parameters.

Lemma 19.5.11 (evolution of yield curves). *Let the current time be oïijŇ Consider the short rate under risk neutral measure is evolving under the Vasicek model with drift α and volatility σ such that*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW(t).$$

The yield curve is given by

$$y(0,T) = r(0)\frac{B(0,T)}{T} - \frac{\ln A(0,T)}{T}.$$

where

$$B(t,T) = \frac{1 - \exp(-\alpha(T-t))}{\alpha},$$

$$A(t,T) = \exp((\frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2})(B(t,T) - (T-t)) - \frac{\sigma^2}{4\alpha}B^2(t,T))),$$

- ***(evolution of whole yield curve)*** *The future yield curve $y(\tau, T+\tau)$ is given by*

$$\begin{aligned} y(\tau, T+\tau) &= r(\tau)\frac{B(\tau, T+\tau)}{T} - \frac{\ln A(\tau, T+\tau)}{T} \\ &= r(\tau)\frac{B(0,T)}{T} - \frac{\ln A(0,T)}{T}. \end{aligned}$$

- ***evolution of yield associated with a fixed maturity date*** *The future yield $y(\tau, T)$ is given by*

$$y(\tau, T) = r(\tau)r(\tau)\frac{B(\tau, T+\tau)}{T} - \frac{\ln A(\tau, T+\tau)}{T}$$

where

$$r(\tau) = r(0)\exp(-\alpha\tau) + \frac{\theta}{\alpha}(1 - \exp(-\alpha\tau)) + \sigma\int_0^\tau \exp(-\alpha(\tau-u))dW(u).$$

Methodology 19.5.2 (simulate the future term structure in Vasicek model). *Suppose*

- *current time is 0.*
- *we are given the current term structure $y(0,T), T \in \mathbb{R}^+$.*
- *we are given a Vasicek model*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW_t$$

- *we are given the initial short rate $r(0)$*

Then we can generate a ***sample term structure*** *in future time $t > 0$ via the following procedure:*

- *simulate $r(t)$ by drawing from the normal distribution*

$$r(t) \sim N(m(t), v^2(t)),$$

where

$$m(t) = r(0)\exp(-\alpha t) + \frac{\theta}{\alpha}(1 - \exp(-\alpha t)),$$

$$v^2(t) = \frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha)).$$

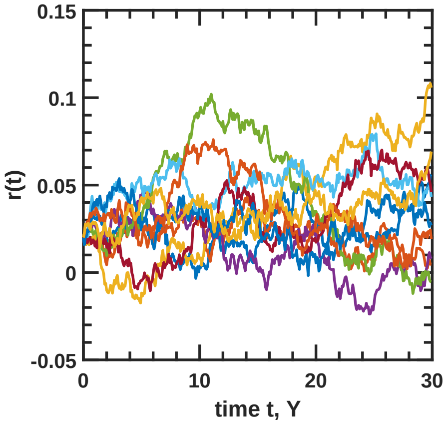
- *the sample yield curve is given by*

$$y(t,T) = r(t)\frac{B(t,T)}{T-t} - \frac{\ln A(t,T)}{T-t}$$

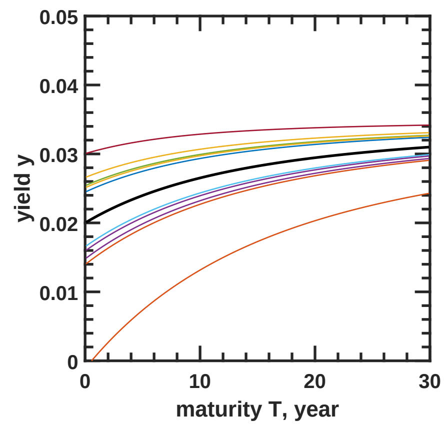
where

$$B(t,T) = \frac{1-\exp(-\alpha(T-t))}{\alpha},$$

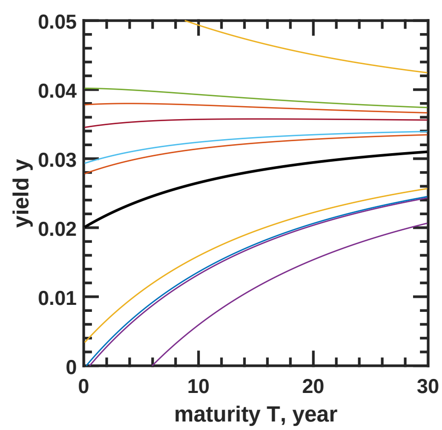
$$A(t,T) = \exp((\frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2})(B(t,T)-(T-t)) - \frac{\sigma^2}{4\alpha}B^2(t,T))).$$



(a) A. Realizations of short rate trajectories.



(b) A. Realizations of yield curve after 1Y (thin lines) from the initial yield curve(thick line).



(c) B. Realizations of yield curve after 5Y (thin lines) from the initial yield curve(thick line).

Figure 19.5.4: Yield curve dynamics in Vasicek model

19.5.3.3 *Instantaneous forward rate dynamics*

Lemma 19.5.12 (instantaneous forward rate dynamics in Vasicek model). *Consider the short rate under risk neutral measure is evolving under the Vasicek model given by*

$$dr(t) = (\theta - kr(t))dt + \sigma dW_t.$$

It follows that

- *The instantaneous forward rate is given by*

$$f(t,T) = r(t)\exp(-k(T-t)) + \frac{\theta}{k}(1-\exp(-k(T-t))) - \frac{\sigma^2}{2k^2}(1-\exp(-k(T-t)))^2.$$

- *The dynamics of instantaneous forward rate is given by*

$$df(t,T) = \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt + \sigma\exp(-k(T-t))dW_t.$$

Proof. (1) Note that in the Vasicek model, we have zero-coupon bond price given by

$$P(t,T) = \exp((-r(T-t) - \frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3),$$

and the instantaneous forward rate is given by

$$f(t,T) = -\frac{\ln P(t,T)}{dT} = r(t) + \alpha(T-t) - \frac{1}{2}\sigma^2(T-t)^2.$$

(2)

$$\begin{aligned}
df(t,T) &= \exp(-k(T-t))dr(t) + br_t\exp(-k(T-t))dt - \theta\exp(-k(T-t)) \\
&+ \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt \\
&= \exp(-k(T-t))((\theta - \alpha r(t))dt + \sigma dW_t) + br_t\exp(-k(T-t))dt - \theta\exp(-k(T-t)) \\
&+ \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt \\
&= \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt + \sigma\exp(-k(T-t))dW_t
\end{aligned}$$

□

19.5.3.4 *Derivative pricing*

Lemma 19.5.13 (zero-coupon bond price and forward price SDE in Vasicek model). *Given the short rate SDE under risk-neutral measure*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW(t),$$

we have

- *the zero-coupon bond price $P(t, r(t), T)$ satisfies*

$$\frac{\partial P}{\partial t} + (\theta(t) - \alpha r(t))\frac{\partial P}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial r^2} = rP.$$

- *the zero-coupon bond price $P(t, T)$ satisfies SDE*

$$dP(t,T) = P(t,T)r(t)dt - P(t,T)D(t,T)\sigma dW(t).$$

  where

$$D(t,T) = \int_t^T e^{-\alpha(s-t)}ds.$$

- *Moreover, the forward bond price $FP(t; S, T) \triangleq \frac{P(t,T)}{P(t,S)}$ satisfies SDE*

$$dFP(t;S,T) = FP(t;S,T)\sigma(D(t,S) - D(t,T))dW^S(t),$$

  where W^S is a Brownian motion under forward measure with respect to $P(t,S)$.

Proof. (1) Use Theorem 19.5.2. (2) From Ito lemma, we have

$$dP(t,r,T) = \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial r}dr + \frac{1}{2}\frac{\partial^2 P}{\partial r^2}drdr.$$

and (1). (3) use Theorem 19.3.1, we have

$$dFP(t;S,T) = FP(t;S,T)(\sigma_{P(t,T)} - \sigma_{P(t,S)})dW^S(t).$$

□

Lemma 19.5.14 (European call option price). *The value at time t of a European call option maturing at S with strike K on a bond maturing at $T > S$ is given by*

$$V(t) = P(t,S)E_S[\max(P(S,T) - K, 0)|\mathcal{F}_t],$$

where the expectation is taken with respect to forward measure of $P(t,S)$. It can be showed that

$$V(t) = BC(t;S,T,K) = P(t,T)N(d_+) - KP(t,S)N(d_-),$$

where

$$d_+ = \frac{\log \frac{FP(t;S,T)}{K} + \frac{1}{2}v(t,S)^2}{v(t,S)} = \frac{\log \frac{B(t,T)}{KB(t,S)} + \frac{1}{2}v(t,S)^2}{v(t,S)}, d_- = d_+ - v(t,S).$$

and

$$v(t,S)^2 = \int_t^S (\sigma(D(t,S) - D(t,T)))^2 dt$$

Proof. Under forward measure Q_S,

$$V(t) = P(t,S)E_S[(P(S,T) - K)^+ | \mathcal{F}_t] = P(t,S)E_S[(FP_S(S) - K)^+ | \mathcal{F}_t],$$

where $FP_S(t) = P(t,T)/P(t,S)$.

Note that the forward bond price dynamics $FP_S(t)$ under Q_S is given by

$$dFP_S(t) = P(t,T)(r(t))dt - FP_S(t)\sigma(D(t,S) - D(t,T))dW^S(t),$$

which is geometric Brownian motion. The rest is routine. □

19.5.4 Cox-Ingersoll-Ross model

19.5.4.1 *The model*

Definition 19.5.4 (Cox-Ingersoll-Ross model). *[5, p. 684][3, p. 64] In Cox-Ingersoll-Ross model, the short rate dynamics under risk-neutral measure is given by*

$$dr(t) = k(\mu - r(t))dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0$$

where W_t is the Brownian motion, r_0, k, μ, σ are constant model parameters, and $2k\mu > \sigma^2$ is imposed to ensure that the origin is inaccessible.

Lemma 19.5.15 (basic statistical properties of $r(t)$).

- $$r_{t+T} = \frac{Y}{2c},$$

  where Y is a non-central Chi-squared distribution with $4k\mu/\sigma^2$ degrees of freedom and non-centrality parameter $2cr_t \exp(-kT)$, and

  $$c = \frac{2k}{(1 - \exp(-kT))\sigma^2}.$$

- $$E[r_t|r_0] = r_0 \exp(-kt) + \mu(1-\exp(-kt))$$
- $$Var[r_t|r_0] = r_0\frac{\sigma^2}{k}(\exp(-kt)-\exp(-2kt)) + \frac{\mu\sigma^2}{2k}(1-\exp(-kt))^2.$$
- *The variance has the following properties:*
  - *Over short time scales, i.e., $t \ll 1/k$, the variance of r grows linearly as*
  $$Var[r_t] \approx r_0\sigma^2 t,\ for\ t \ll 1/k.$$
  - *Over long time scales, i.e., the variances approaches a constant:*
  $$Var[r_t] = \mu\frac{\sigma^2}{2k},\ for\ t \gg 1/k.$$

Proof. (2) Note that $r(t)$ has solution given by

$$r(t) = e^{-kt}r(0) - \mu(e^{-kt}-1) + \int_0^t e^{-k(t-s)}\sigma\sqrt{r(s)}dW_s.$$

Therefore,

$$E[r(t)] = e^{-kt}r(0) - \mu(e^{-kt}-1),$$

and

$$\begin{aligned}
Var[r(t)] &= Var[\int_0^t e^{-k(t-s)}\sigma\sqrt{r(s)}dW_s] \\
&= \sigma^2 e^{-2kt} Var[\int_0^t e^{ks}\sqrt{r(s)}dW_s] \\
&= \sigma^2 e^{-2kt} E[(\int_0^t e^{ks}\sqrt{r(s)}dW_s)^2] \\
&= \sigma^2 e^{-2kt}\int_0^t e^{2ks}E[r(s)]ds \\
&= \sigma^2 e^{-2kt}\int_0^t e^{2ks}(e^{-ks}r(0) - \mu(e^{-ks}-1))ds \\
&= \sigma^2 e^{-2kt}\int_0^t e^{ks}r(0) - \mu(e^{ks}-e^{2ks}))ds \\
&= r_0\frac{\sigma^2}{k}(\exp(-kt)-\exp(-2kt)) + \frac{\mu\sigma^2}{2k}(1-\exp(-kt))^2
\end{aligned}$$

(3)(a) Use first order approximation $\exp(-kt) \approx 1-kt$. (b) Use $\exp(-kt) \approx 0$. □

Lemma 19.5.16 (zero-coupon bond price in CIR model). *[3, p. 64] In the CIR model with parameter k, μ, σ, the price at time t of a zero-coupon bond with maturity T is*

$$P(t,T) = A(t,T)\exp(-B(t,T)r(t)),$$

where

$$\begin{aligned} A(t,T) &= [\frac{2h\exp((k+h)(T-t)/2)}{2h+(k+h)(\exp((T-t)h)-1)}]^{2k\mu/\sigma^2}, \\ B(t,T) &= \frac{2(\exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)} \\ h &= \sqrt{k^2+2\sigma^2} \end{aligned}$$

19.5.4.2 *Yield curve in CIR model*

Lemma 19.5.17 (yield curve in CIR model). *[3, p. 64] In the CIR model with parameter k, μ, σ, the yield curve at time t is*

$$y(t,T) = -\frac{1}{T-t}\ln A(t,T) + \frac{1}{T-t}B(t,T)r(t)),$$

where

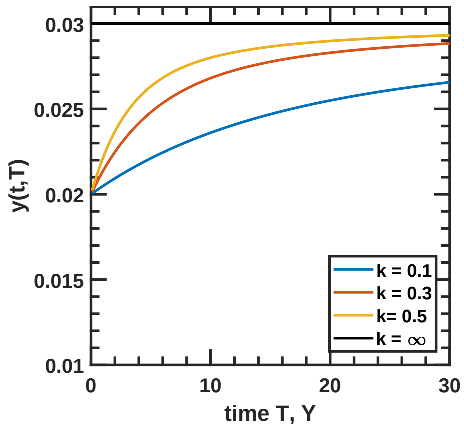
$$\begin{aligned} \ln A(t,T) &= \frac{2k\mu}{\sigma^2}\ln\frac{2h\exp((k+h)(T-t)/2)}{2h+(k+h)(\exp((T-t)h)-1)}, \\ B(t,T) &= \frac{2(\exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)} \\ h &= \sqrt{k^2+2\sigma^2}. \end{aligned}$$

Remark 19.5.16 (yield curve shape). From Figure 19.5.3, we can see that

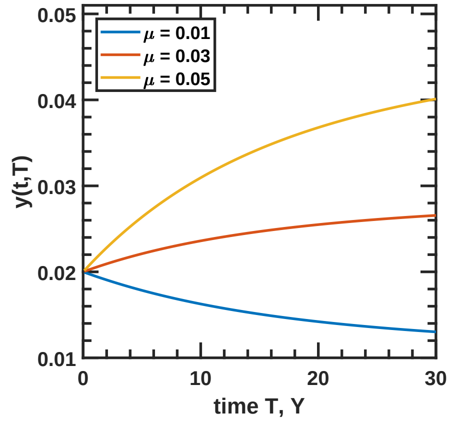
- (reversion speed effect) If $\alpha \to \infty$(but θ/α remains to be a constant), we can see that

$$B(t,T) = 0, A(t,T) = \exp(-\frac{\theta}{\alpha})(T-t)) = \exp(-r^*(T-t)), y(t,T) = r^*.$$

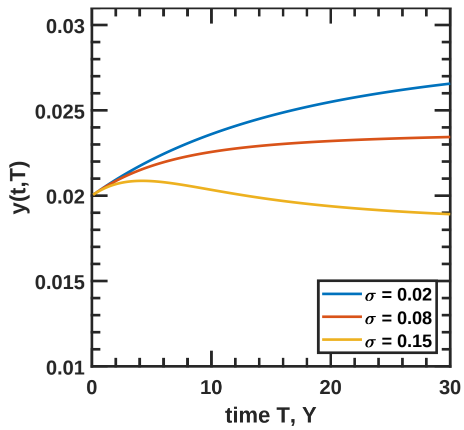
  If reversion speed α increases, then the yield curve will more resemble the equilibrium flat yield curve given by $y(t,T)* = r^*$.
- (mean level effect) The mean level will tend to pull the far-end of the yield curve toward the mean level.
- (volatility level effect) The volatility level tend to pull the far-end of the yield curve down.



(a) Yield curve shape vs. reversion speed.



(b) Yield curve shape vs. mean level μ.



(c) Yield curve shape vs. volatility.

Figure 19.5.5: Control yield curve shape via CIR model parameters.

Lemma 19.5.18 (evolution of yield curves). *Let the current time be oïijŇConsider the short rate under risk neutral measure is evolving under the CIR model with parameter* (μ, k, σ),

$$dr(t) = k(\mu - r(t))dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0,$$

where W_t *is the Brownian motion,* r_0, k, μ, σ *are constant model parameters. The yield curve is given by*

$$y(0,T) = r(0)\frac{B(0,T)}{T} - \frac{\ln A(0,T)}{T}.$$

where

$$A(t,T) = [\frac{2h\exp((k+h)(T-t)/2)}{2h+(k+h)(\exp((T-t)h)-1)}]^{2k\mu/\sigma^2},$$
$$B(t,T) = \frac{2(\exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)}$$
$$h = \sqrt{k^2+2\sigma^2}$$

- ***(evolution of whole yield curve)*** *The future yield curve $y(\tau, T+\tau)$ is given by*

$$y(\tau, T+\tau) = r(\tau)\frac{B(\tau, T+\tau)}{T} - \frac{\ln A(\tau, T+\tau)}{T}$$
$$= r(\tau)\frac{B(0,T)}{T} - \frac{\ln A(0,T)}{T}.$$

- ***evolution of yield associated with a fixed maturity date*** *The future yield $y(\tau, T)$ is given by*

$$y(\tau, T) = r(\tau)r(\tau)\frac{B(\tau, T+\tau)}{T} - \frac{\ln A(\tau, T+\tau)}{T}$$

Proof. □

Methodology 19.5.3 (simulate the future term structure in CIR model). *Suppose*

- *current time is 0.*
- *we are given the current term structure $y(0,T), T \in \mathbb{R}^+$.*
- *we are given a CIR model*

$$dr(t) = k(\mu - r(t))dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0.$$

- *we are given the initial short rate $r(0)$*

Then we can generate a ***sample term structure*** *in future time $t > 0$ via the following procedure:*

- *simulate $r(t)$ by Monte carlo simulation:*

$$r(t+dt) = r(t) + k(\mu - r(t))dt + \sigma\sqrt{r(t)dt}Z, Z \sim N(0,1).$$

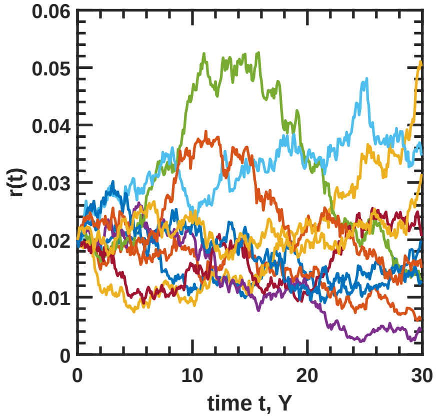
- *the sample yield curve is given by*

$$y(t,T) = r(t)\frac{B(t,T)}{T-t} - \frac{\ln A(t,T)}{T-t}$$

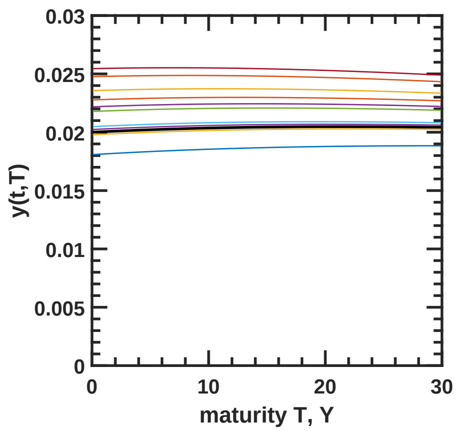
where

$$B(t,T) = \frac{1-\exp(-\alpha(T-t))}{\alpha},$$

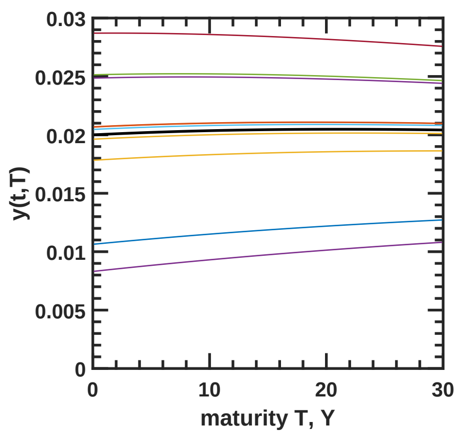
$$A(t,T) = \exp((\frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2})(B(t,T) - (T-t)) - \frac{\sigma^2}{4\alpha}B^2(t,T))).$$



(a) A. Realizations of short rate trajectories.



(b) A. Realizations of yield curve after 1Y (thin lines) from the initial yield curve(thick line).



(c) B. Realizations of yield curve after 5Y (thin lines) from the initial yield curve(thick line).

Figure 19.5.6: Yield curve dynamics in Merton model

19.5.4.3 *Instantaneous forward rate dynamics*

Lemma 19.5.19 (instantaneous forward rate dynamics in CIR model). *Consider a CIR model under risk-neutral measure given by*

$$dr(t) = k(\mu - r(t))dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0.$$

Let current time be t.

- *The instantaneous forward rate given by the model is*

$$f(t,T) = \frac{2k\mu(\exp((T-t)h-1))}{2h+(k+h)(\exp((T-t)h)-1)} + r_0\frac{4h^2\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))^2},$$

where $h = \sqrt{k^2+2\sigma^2}$.

Proof. Note that

$$\begin{aligned}
&\frac{\partial}{\partial T}B(t,T) \\
&= \frac{\partial}{\partial T}\frac{2(exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)} \\
&= \frac{(2h+(k+h)(\exp(h(T-t))-1)2h\exp(h(T-t)) - 2(\exp(h(T-t))-1)(k+h)h\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))^2} \\
&= \frac{4h^2\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))^2}
\end{aligned}$$

Note that

$$\ln A(t,T) = \frac{2k\mu}{\sigma^2}(\ln 2h + \frac{(k+h)(T-t)}{2} - \ln(2h+(k+h)(\exp((T-t)h)-1).$$

Then

$$\begin{aligned}
&\frac{\partial}{\partial T}\ln A(t,T) \\
&= \frac{2k\mu}{\sigma^2}(\frac{k+h}{2} - \frac{(k+h)h\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))} \\
&= \frac{2k\mu}{\sigma^2}(\frac{k+h}{2} - \frac{(k+h)h\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))} \\
&= \frac{2k\mu}{\sigma^2}(\frac{(k+h)h + \frac{(k+h)^2}{2}(\exp((T-t)h)-1) - (k+h)h\exp((T-t)h)}{(2h+(k+h)(\exp((T-t)h)-1))} \\
&= \frac{2k\mu}{\sigma^2}\frac{(\exp((T-t)h)-1)(k^2-h^2)/2}{(2h+(k+h)(\exp((T-t)h)-1))} \\
&= \frac{2k\mu}{\sigma^2}\frac{(\exp((T-t)h)-1)(-\sigma^2)}{(2h+(k+h)(\exp((T-t)h)-1))} \\
&= \frac{(\exp((T-t)h)-1)(-2k\mu)}{(2h+(k+h)(\exp((T-t)h)-1))}
\end{aligned}$$

□

19.5.4.4 *Applications*

Lemma 19.5.20 (zero-coupon bond price in CIR model). *[3, p. 66] The price at time t of a zero-coupon bond with maturity T is*

$$P(t,T) = E_Q[\exp(-\int_t^T r(s)ds)] = A(t,T)\exp(-B(t,T)r(t),$$

where

$$\begin{aligned}
A(t,T) &= [\frac{2h\exp((k+h)(T-t)/2)}{2h+(k+h)(\exp((T-t)h)-1)}] \\
B(t,T) &= [\frac{2(\exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)}] \\
h &= \sqrt{k^2+2\sigma^2}.
\end{aligned}$$

Under risk-neutral measure Q, the bond dynamics is given by

$$dP(t,T) = r(t)P(t,T)dt - B(t,T)P(t,T)\sigma\sqrt{r(t)}dW(t).$$

19.5.5 Model extension using deterministic shift

19.5.5.1 *Principles*

Theorem 19.5.3. *[3, p. 95] Let current time be 0. Let x_t be a (integrable) stochastic process and let the short rate process $r(t)$ be*

$$r_t = x_t + \phi(t),$$

where $\phi(t)$ is deterministic-shift function

It follows that

- *The zero-coupon bond price is given by*

$$P(t,T) = E_Q[\exp(-\int_t^T r_s ds)|\mathcal{F}_t] = \exp(-\int_t^T \phi(s)ds)F(t,T,x_t).$$

- *Let the current market term structure be represented by $f^*(0,T)$. Then the term structure produced by the model will exactly match with the current market term structure, i.e.,*

$$-\frac{\partial \ln P(0,T)}{\partial T} = f^*(0,T)$$

  if

$$\begin{aligned}\phi(T) &= f^*(0,T) + \frac{\partial \ln F(0,t,x_0)}{\partial T} \\ &= f^*(0,T) - f(0,T), f(0,T) \triangleq -\frac{\partial \ln F(0,t,x_0)}{\partial T}.\end{aligned}$$

- *With the choice of ϕ in (2), we have*

$$P(t,T) = \frac{P^*(0,T)F(0,t,x_0)}{P^*(0,t)F(0,T,x_0)}F(t,T,x_t),$$

  where $P^*(t,T)$ is the current market term structure.

Proof. (1)

$$\begin{aligned}
P(t,T) &= E_Q[\exp(-\int_t^T r_s ds)|\mathcal{F}_t] \\
&= E_Q[\exp(-\int_t^T \phi(s) + x_s ds)|\mathcal{F}_t] \\
&= \exp(-\int_t^T \phi(s)ds)E_Q[\exp(-\int_t^T x_s ds)|\mathcal{F}_t] \\
&= \exp(-\int_t^T \phi(s)ds)F(t,T,x_t).
\end{aligned}$$

(2)

$$\begin{aligned}
f^*(0,T) &= -\frac{\partial \ln P(0,T)}{\partial T} \\
&= \phi(T) - \frac{\partial \ln F(0,T,x_0)}{\partial T} \\
\implies \phi(T) &= f^*(0,T) + \frac{\partial \ln F(0,T,x_0)}{\partial T}.
\end{aligned}$$

(3)

$$\begin{aligned}
&\exp(-\int_t^T \phi(s)ds) \\
&= \exp(-\int_t^T f^*(0,s) + \frac{\partial \ln F(0,s,x_0)}{\partial s}ds) \\
&= \exp(-\int_t^T f^*(0,s)ds)\exp(-\int_t^T \frac{\partial \ln F(0,s,x_0)}{\partial s}ds) \\
&= \frac{P^*(0,T)}{P^*(0,t)}\exp(-\int_t^T d\ln F(0,s,x_0)) \\
&= \frac{P^*(0,T)}{P^*(0,t)}\exp(-\ln F(0,t,x_0) + \ln F(0,T,x_0)) \\
&= \frac{P^*(0,T)F(0,t,x_0)}{P^*(0,Tt)F(0,T,x_0)}
\end{aligned}$$

where we use the fact that

$$\exp(-\int_t^T f^*(0,s)ds) = \frac{\exp(-\int_0^T f^*(0,s)ds)}{\exp(-\int_0^t f^*(0,s)ds)} = \frac{P^*(0,T)}{P^*(0,t)}.$$

Eventually,

$$P(t,T) = \exp(-\int_t^T \phi(s)ds)F(t,T,x_t) = \frac{P^*(0,T)F(0,t,x_0)}{P^*(0,t)F(0,T,x_0)}F(t,T,x_t).$$

□

Remark 19.5.17 (implication for calibration)**.**

- The deterministic shift extension enables the perfect match to the initial term structure.
- Suppose the stochastic process is given by

$$dx_t = \mu(t)dt + \sigma(t)dW_t.$$

  We can choose $\phi(t)$ to fit the initial term structure and choose $\sigma(t), \mu(t)$ to match the volatility term structure.

19.5.5.2 *Extended Vasicek model*

Definition 19.5.5 (extended Vasicek model). *The extended CIR model under risk-neutral model is given by*

$$\begin{aligned} dx(t) &= -kx(t)dt + \sigma dW_t, x(0) = 0. \\ r(t) &= x(t) + \alpha(t). \end{aligned}$$

Lemma 19.5.21. *Consider the short rate $r(t)$ under risk neutral measure is evolving under the following model given by*

$$\begin{aligned} dx(t) &= -kx(t)dt + \sigma dW_t, x(0) = 0. \\ r(t) &= x(t) + \alpha(t) \end{aligned}$$

Let the current time be 0. It follows that

- *If the current market term structure is given by $f^M(0,t)$, then choose*

$$\alpha(t) = f^M(0,t) + \frac{\sigma^2}{2k^2}(1-\exp(-kt))^2,$$

  can match the market term structure.
- *With the choice of ϕ in (1), we have*

$$P(t,T) = \frac{P^*(0,T)A(0,t)\exp(-B(0,t)x_0)}{P^*(0,t)A(0,T)\exp(-B(0,T)x_0)}A(t,T)\exp(-B(t,T)x_t),$$

  where $P^*(0,T)$ is the current market term structure, and

$$B(t,T) = \frac{1-\exp(-k(T-t))}{\alpha},$$

$$A(t,T) = \exp((-\frac{\sigma^2}{2k^2})(B(t,T) - (T-t)) - \frac{\sigma^2}{4k}B^2(t,T))),$$

Proof. (1)Note that the forward rate associated with the dynamics of

$$dx(t) = -kx(t)dt + \sigma dW_t, x(0) = 0$$

is given by(Lemma 19.5.12)

$$f(0,T) = -\frac{\sigma^2}{2k^2}(1 - \exp(-k(T-t)))^2.$$

Then we use the principle (Theorem 19.5.3), we have

$$\alpha(t) = f^M(0,t) - f(0,t).$$

(2) Note that the zero-coupon bond price associated with the dynamics of

$$dx(t) = -kx(t)dt + \sigma dW_t, x(0) = 0$$

is given by(Lemma 19.5.10)

$$P(t,T) = A(t,T)\exp(-B(t,T)x(t)),$$

where

$$B(t,T) = \frac{1 - \exp(-k(T-t))}{\alpha},$$

$$A(t,T) = \exp((-\frac{\sigma^2}{2k^2})(B(t,T) - (T-t)) - \frac{\sigma^2}{4k}B^2(t,T))).$$

□

19.5.5.3 *Extended CIR model*

Definition 19.5.6 (extended CIR model). *The extended CIR model under risk-neutral model is given by*

$$\begin{aligned} dx(t) &= k(\mu - x(t))dt + \sigma\sqrt{x(t)}dW(t), x(0) = x_0 \\ r_t &= x_t + \phi(t). \end{aligned}$$

Lemma 19.5.22 (extended CIR model properties). *[3, p. 95] Let current time be 0. Let the original model has risk-neutral dynamics be*

$$dx(t) = k(\mu - x(t))dt + \sigma\sqrt{x(t)}dW(t), x(0) = x_0.$$

Denote

$$F(t,T,x_t) = E_Q[\exp(-\int_t^T x_s ds)|\mathcal{F}_t].$$

The deterministic-shift extension consists of defining the short rate by

$$r_t = x_t + \phi(t).$$

It follows that

- *Let the current market term structure be represented by $f^*(0,T)$. Then the term structure produced by the model will exactly match with the current market term structure, i.e.,*
$$-\frac{\partial \ln P(0,T)}{\partial T} = f^*(0,T)$$
if
$$\begin{aligned}\phi(T) &= f^*(0,T) - f(0,T)\\ &= f^*(0,T) - \frac{2k\mu(\exp(th-1))}{2h+(k+h)(\exp(th)-1)} - r_0\frac{4h^2\exp(th)}{(2h+(k+h)(\exp(th)-1))^2},\\ h &= \sqrt{k^2+2\sigma^2}.\end{aligned}$$
- *With the choice of ϕ in (1), we have*
$$P(t,T) = \frac{P^*(0,T)A(0,t)\exp(-B(0,t)x_0)}{P^*(0,t)A(0,T)\exp(-B(0,T)x_0)}A(t,T)\exp(-B(t,T)x_t),$$
where $P^*(0,T)$ is the current market term structure, and
$$\begin{aligned}A(t,T) &= [\frac{2h\exp((k+h)(T-t)/2)}{2h+(k+h)(\exp((T-t)h)-1)}]\\ B(t,T) &= [\frac{2(\exp((T-t)h)-1)}{2h+(k+h)(\exp((T-t)h)-1)}]\\ h &= \sqrt{k^2+2\sigma^2}.\end{aligned}$$

Proof. (1) Note that the forward rate derived from CIR model(Lemma 19.5.19) is given by

$$f(0,T) = \frac{2k\mu(\exp(Th-1))}{2h+(k+h)(\exp(Th)-1)} + r_0\frac{4h^2\exp(Th)}{(2h+(k+h)(\exp(Th)-1))^2}, h = \sqrt{k^2+2\sigma^2}$$

□

19.5.6 Ho-Lee model

Definition 19.5.7 (Ho-Lee model). *[5, p. 689] The Ho-Lee model for short rate, under the risk-neutral measure Q, is given by*

$$dr = \theta(t)dt + \sigma dW_t$$

where W_t is the Brownian motion, and σ is a constant.

Lemma 19.5.23 (zero-coupon bond price in Ho-Lee model). *[6, p. 26] With the Ho-Lee short rate model under the risk-neutral measure Q, given by*

$$dr = \theta(t)dt + \sigma dW_t,$$

the zero-coupon bond price is given by

$$P(t,T) = \exp(-m(t) + \frac{1}{2}s^2)$$

where

$$m(t) = r(t)(T-t) + \int_t^T \theta(u)(T-u)du$$

and variance

$$s^2 = \frac{1}{3}\sigma^2(T-t)^3$$

under risk-neutral measure.

Define $A(t,T) = \exp(-\frac{1}{2}s^2 - \int_t^T \theta(u)(T-u)du), B(t,T) = T-t$. Then we can write

$$P(t,T) = A(t,T)\exp(-B(t,T)r(t)).$$

The zero-coupon bond dynamics is given by

$$dP(t,T) = r(t)P(t,T)dt - \sigma(T-t)P(t,T)dW_t$$

Proof. (1) We have

$$r(s) = r(t) + \int_t^s \theta(u)du + \sigma(W(s) - W(t)),$$

then

$$\begin{aligned}\int_t^T r(s)ds &= r(t)(T-t) + \int_t^T (\int_t^s \theta(u)du)ds + \sigma\int_t^T (T-s)dW(s)\\ &= r(t)(T-t) + \int_t^T \theta(u)(T-u)du + \sigma\int_t^T (T-s)dW(s)\end{aligned}$$

where we exchange the integral. Note that

$$d((T-s)W(s)) = -W(s)ds + (T-s)dW(s),$$

then

$$[(T-s)W(s)]|_t^T = -\int_t^T W(s)ds + \int_t^T (T-s)dW(s).$$

(2) similar to Lemma 19.5.4. (3) Take differential on $P(t,T)$ and use the fact that $P(t,T)$ has to satisfy term structure function such that drift is r. See Lemma 19.5.1. □

Lemma 19.5.24 (fitting θ to initial term structure). *[6, p. 27] Suppose we have market prices of zero-coupon bond $P(t,T), T > 0$, then*

$$\theta(T) = -\frac{\partial^2}{\partial T^2}\log(P(t,T)) + \sigma^2 T$$

will enable Ho-Lee model to fit current term structure exactly.

Proof.

$$\begin{aligned}
\log P &= r(t)(T-t) + \int_t^T \theta(u)(T-u)du - \frac{1}{6}\sigma^2(T-t)^3 \\
\implies \frac{\partial \log P}{\partial T} &= -r_0 - \int_t^T \theta(s)ds + \frac{1}{2}\sigma^2 T^2 \\
&= -r_0 - \int_t^T \theta(s)ds + \frac{1}{2}\sigma^2 T^2 \\
\implies \frac{\partial^2 \log P}{\partial T^2} &= -\theta(T) + \sigma^2 T
\end{aligned}$$

□

Remark 19.5.18 (how to fit σ ?)**.** Once we fit θ to initial term structure, the only variable undetermined is σ, which can be determined from market prices of other vanilla bond options.

Remark 19.5.19 (interpretation)**.**

- The Ho-Lee model is the simplest no-arbitrage short-rate model.
- The value of $\theta(t)$ will be calibrated to make match the Zero-Coupon Bond prices, which depends on σ.

19.5.7 Hull-White model

19.5.7.1 *Fundamentals*

Definition 19.5.8 (Hull-White model). *[5, p. 691] In Hull-White model, the SDE for r under risk-neutral measure is given*

$$dr = (\theta(t) - \alpha r(t))dt + \sigma(t)dW_t$$

where W_t is the Brownian motion under risk-neutral measure, and α is constant, and the coefficients $\theta(t)$ and $\sigma(t)$ are time-dependent variables.

Lemma 19.5.25 (zero-coupon bond pricing in Hull-White model). *[1, p. 50] Assume the short rate r follows the Hull-White model under risk-neutral measure Q. We have* [a]

- $$r(s) = r(t)e^{-\alpha(s-t)} + \int_t^s \theta(u)e^{-\alpha(s-u)}du + \int_t^s \sigma(u)e^{-\alpha(s-u)}dW(u)$$

  In particular, $r(s)$ is a Gaussian process with mean

  $$m(s) = r(t)e^{-\alpha(s-t)} + \int_t^s \theta(u)e^{-\alpha(s-u)}du$$

  and variance

  $$v^2 = \int_t^s \sigma(u)^2 e^{-2\alpha(s-u)}du.$$

- *Let $D(t,T) \triangleq \int_t^T e^{-\alpha(s-t)}ds$, then*

  $$-\int_t^T r(s)ds = -D(t,T)r(t) - \int_t^T \theta(u)D(u,T)du - \int_t^T \sigma(u)D(u,T)dW(u).$$

- *The zero-coupon bond price*

  $$P(t,T) = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] = \exp(m(t) + \frac{1}{2}s^2(t)).$$

  where

  $$m(t) = -D(t,T)r(t) - \int_t^T \theta(u)D(u,T)du$$

  and

  $$s^2(t) = \int_t^T \sigma(u)^2 D(u,T)^2 du.$$

Define $A(t,T) = \exp(\frac{1}{2}s(t)^2 - \int_t^T \theta(u)D(u,T)du), B(t,T) = D(t,T)$. *Then we can write*

$$P(t,T) = A(t,T)\exp(-B(t,T)r(t)).$$

- *The zero-coupon bond has the dynamics*

$$dP(t,T) = P(t,T)(r(t)dt - \sigma(t)B(t,T)dW_t).$$

[a] Note that if current time is t, then $r(t)$ is observable and $P(t,T)$ is a deterministic value for each T; if current time is 0, then $r(t)$ is unknown and $P(t,T)$ is a random variable quantity for each T.

Proof. (1) Use **??**. (2)

$$-\int_t^T r(s)ds = -\int_t^T r(t)e^{-\alpha(s-t)}ds - \int_t^T\int_t^s \theta(u)e^{-\alpha(s-u)}duds - \int_t^T\int_t^s \sigma(u)e^{-\alpha(s-u)}dW(u)ds$$

We then use change of variable in integral given by

$$\begin{aligned}\int_t^T\int_t^s \theta(u)e^{-\alpha(s-u)}duds &= \int_t^T\int_u^T \theta(u)e^{-\alpha(s-u)}dsdu \\ &= \int_t^T\int_u^T \theta(u)e^{-\alpha(s-u)}dsdu \\ &= \int_t^T\int_u^T \theta(u)D(u,T)du\end{aligned}$$

Similarly,

$$\int_t^T\int_t^s \sigma(u)e^{-\alpha(s-u)}dW(u)ds = \; = \int_t^T\int_u^T \theta(u)D(u,T)dW(u)$$

(3) Note that $-\int_t^T r(s)ds$ is a Gaussian process with mean

$$m = -D(t,T)r(t) - \int_t^T \theta(u)D(u,T)du$$

and variance

$$s^2 = \int_t^s \sigma(u)^2 D(u,T)^2 du.$$

Use the property of log normal variableLemma 2.2.14. □

Lemma 19.5.26 (zero-coupon bond price fitting to the current term structure and its distribution). *[1, p. 54]*

- *The Hull-White short rate model fitting to the current(current time is 0) term structure(characterized by $f(0,t), t \geq 0$) is given by*

$$\theta(t) = \frac{\partial f(0,t)}{\partial t} + \alpha f(0,t) + \int_0^t \sigma^2 e^{-2\alpha(t-s)} ds.$$

- *The future zero-coupon bond price, a random quantity, based on information on time 0 is given by*

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp(-A(t,T) - r(t)D(t,T)),$$

where

$$r(t) = r(0)e^{-\alpha t} + \int_0^t \theta(u)D(0,t)du + \int_0^t \sigma(u)D(0,t)dW(u),$$

$$A(t,T) = -f(0,t)D(t,T) + \frac{1}{2}D^2(t,T)\int_0^t \sigma^2(u)\exp(-2\alpha(t-u))du,$$

$$D(t,T) = \int_t^T \exp(-\alpha(T-s))ds.$$

- *The future yield curve $y(t,T)$, a random quantity, based on information on time 0 is given by*

$$y(t,T) = -\frac{\ln P(t,T)}{T-t} = -\frac{1}{T-t}(\ln \frac{P(0,T)}{P(0,t)} - A(t,T) - r(t)D(t,T)).$$

- *Under risk-neutral measure $P(t,T)$ is lognormal distribution; $\ln P(t,T)$ has variance given by*

$$D^2(t,T)Var[r(t)] = D^2(t,T)\int_0^t \sigma^2(u)\exp(-2\alpha(t-u))du.$$

Proof. (1) See Lemma 19.9.5.(2) See reference. (3)(4) Note that $r(t)$ is the Gaussian random variable(Lemma 19.5.25) and then $Var[\ln P(t,T)] = Var[r(t)D(t,T)] = D(t,T)^2 Var[r(t)]$. □

19.5.7.2 *Yield curve dynamics*

19.5.7.3 *Derivative pricing*

[1, p. 53]

Lemma 19.5.27 (Pricing European call option on zero-coupon bond). *The value at time t of a European call option maturing at S with strike K on a bond maturing at T > S is given by*

$$V(t) = P(t,S)E_S[\max(P(S,T)-K,0)|\mathcal{F}_t],$$

where the expectation is taken with respect to forward measure of $P(t,S)$. It can be showed that

$$V(t) = BC(t;S,T,K) = P(t,T)N(d_+) - KP(t,S)N(d_-),$$

where

$$d_+ = \frac{\log \frac{FP(t;S,T)}{K} + \frac{1}{2}v(t,S)^2}{v(t,S)} = \frac{\log \frac{B(t,T)}{KB(t,S)} + \frac{1}{2}v(t,S)^2}{v(t,S)}, d_- = d_+ - v(t,S).$$

and

$$v(t,S)^2 = \int_t^S (\sigma(D(t,S)-D(t,T)))^2 dt$$

Proof. Under forward measure Q_S,

$$V(t) = P(t,S)E_S[(P(S,T)-K)^+|\mathcal{F}_t] = P(t,S)E_S[(FP_S(S)-K)^+|\mathcal{F}_t],$$

where $FP_S(t) = P(t,T)/P(t,S)$.

Note that the forward bond price dynamics $FP_S(t)$ under Q_S is given by

$$dFP_S(t) = P(t,T)(r(t))dt - FP_S(t)\sigma(D(t,S)-D(t,T))dW^S(t),$$

which is geometric Brownian motion. The rest is routine.

□

Lemma 19.5.28 (Swaption pricing). *[1, p. 59] Consider an call option with strike K and expiry T_0 on a payer swap with unit notional, settlement dates $T_1, ..., T_n$ and reset dates $T_0, ..., T_{n-1}$.Denote $\tau_i = T_i - T_{i-1}$. Let the current time be $t < T_0$.*

- *The swapiton's payoff at time T_0 can be written by*

$$(1 - K\sum_{i=1}^n \tau_i P(T_0,T_i) - P(T_0,T_n))^+$$

- *Let $\tilde{r}$ be such that*

$$K\sum_{i=1}^n \tau_i F(T_0,T_i;\tilde{r}) + F(T_0,T_n;\tilde{r}) = 1,$$

where[a]

$$F(T_0,T_i;r) = A(T_0,T_i)\exp(-B(T_0,T_i)r).$$

Then

$$(1-K\sum_{i=1}^{n}\tau_i P(T_0,T_i)-P(T_0,T_n))^+ = K\sum_{i=1}^{n}\tau_i(P_i-P(T_0,T_i))^+ + (P_n-P(T_0,T_n))^+,$$

where we use the notation

$$P_i = F(T_0,T_i;\tilde{r}).$$

- *The current value of the swaption at time t is given by*

$$V(t) = K\sum_{i=1}^{n}\tau_i BP(t,T_0,T_i,P_i) + BP(t,T_0,T_i,P_n)$$

where $BP(t,T_0,T_i,P_i)$ is the time t put option price of the bond $P(T_0,T_i)$ with strike P_i.

[a] note that $F(T_0,T_i;r)$ is a decreasing function on r, and $K\sum_{i=1}^{n}\tau_i F(T_0,T_i;\tilde{r}) + F(T_0,T_n;\tilde{r})$ is also a decreasing function on r; then we must have a unique root.

Proof. (1) Note that the value of a payer interest rate swap(Lemma 19.1.7) is given by

$$V(T_0,K) = P(T_0,T_0) - P(T_0,T_n) - K\sum_{i=1}^{n}\tau_i P(T_0,T_i).$$

(2)

$$\begin{aligned}
&(1-K\sum_{i=1}^{n}\tau_i P(T_0,T_i)-P(T_0,T_n))^+\\
&= (K\sum_{i=1}^{n}\tau_i P_i + P_n - K\sum_{i=1}^{n}\tau_i P(T_0,T_i) - P(T_0,T_n))^+\\
&= (K\sum_{i=1}^{n}\tau_i(P_i-P(T_0,T_i)) + (P_n-P(T_0,T_n))^+\\
&= K\sum_{i=1}^{n}\tau_i(P_i-P(T_0,T_i))^+ + (P_n-P(T_0,T_n))^+
\end{aligned}$$

The reason that the last step holds is because

- If $r(T_0) > \tilde{r}$, then

$$(K\sum_{i=1}^{n}\tau_i(P_i-P(T_0,T_i)) + (P_n-P(T_0,T_n)))^+ = 0,$$

since $K\sum_{i=1}^{n}\tau_i P_i + P_n < 1$.

- If $r(T_0) < \tilde{r}$, then

$$\begin{aligned}
&(K\sum_{i=1}^{n}\tau_i(P_i - P(T_0,T_i)) + (P_n - P(T_0,T_n)))^+ \\
&= K\sum_{i=1}^{n}\tau_i(P_i - P(T_0,T_i)) + (P_n - P(T_0,T_n)) \\
&= 1 - K\sum_{i=1}^{n}\tau_i P(T_0,T_i) - P(T_0,T_n)
\end{aligned}$$

(3) Since we already decompose the payoff to a linear combination of put options on bond, then the current value of the swaption is also the linear combination of the current value of these put options on bonds. □

Lemma 19.5.29 (Bermudan swaption pricing). *Consider an Bermudan swaption option with strike K and exercising date $0 < T_0, T_1, ..., T_{n-1}$ on a payer swap with unit notional, settlement dates $T_1, ..., T_n$ and reset dates $T_0, ..., T_{n-1}$. Denote $\tau_i = T_i - T_{i-1}$. Let the current time be $0 < T_0$. Denote the Bermudan swaption value at time T_i by $Berm(T_i)$.*

- *If the Bermudan swaption has not been exercised before T_i, then the (undiscounted)exercise value at T_i is*
$$Ex(T_i) = (PS(T_i, K))^+,$$
where $PS(T_i, K)$ denotes the time T_i value(a random quantity) of the swap with fixed rate K, settlement dates $T_{i+1}, ..., T_n$ and reset dates $T_i, ..., T_{n-1}$.
- *If the Bermudan swaption has not been exercised before T_i, then the (undiscounted)continuation value at T_i is*
$$C(T_i) = P(T_i, T_{i+1})E_{T_{i+1}}(Berm(T_{i+1})|\mathcal{F}_{T_i}).$$
In particular, $C(T_{n-1}) = 0$.
- *The value of the Bermudan swaption is*
$$Berm(T_i) = \max(Ex(T_i), C(T_i)).$$

Proof. This is just the Bellman's principle. Note that the continuation value at time T_i is equivalent to a Bermudan swaption starting at T_{i+1} and then discount it to T_i. □

Methodology 19.5.4 (backward induction to calculate Bermudan swaption value). *Consider an Bermudan swaption option with strike K and exercising date $0 < T_0, T_1, ..., T_{n-1}$ on a payer swap with unit notional, settlement dates $T_1, ..., T_n$ and reset dates $T_0, ..., T_{n-1}$. Denote $\tau_i = T_i - T_{i-1}$. Let the current time be $0 < T_0$. Denote the Bermudan swaption value at time T_i by $Berm(T_i)$. $PS(T_i, K)$ denotes the time T_i value(a* *random quantity) of the swap with fixed rate K, settlement dates $T_{i+1}, ..., T_n$ and reset dates $T_i, ..., T_{n-1}$.*

- *We start with T_{n-1}, where*

$$Berm(T_{n-1}) = \max(Ex(T_{n-1}), C(T_{n-1})) = (PS(T_{n-1}, K))^+,$$

*where $PS(T_{n-1}, K)$ denotes the time T_{n-1} value(a random quantity) of **a one-period payer swap** with fixed rate K, settlement dates T_n and reset dates T_{n-1}.*
- *Back to T_{n-2}, we calculate*

$$C(T_{n-2}) = P(T_{n-2}, T_{n-1})E_{T_{n-1}}[Berm(T_{n-1})|\mathcal{F}_{T_{n-2}}].$$

$$Ex(T_{n-2}) = (PS(T_{n-2}, K))^+,$$

$$Berm(T_{n-2}) = \max(Ex(T_{n-2}), C(T_{n-2})).$$

- *Continue the backward induction process to T_0, then we calculate the current time value*

$$Berm(0) = P(0, T_0)E_{T_0}[Berm(T_{n-1})|\mathcal{F}_{T_{n-2}}]$$

Remark 19.5.20 (analytical vs. numerical method).

- Note that it is generally impossible to evaluate the value of a Bermudan swaption analytically; this is because $Berm(T_i), C(T_i), Ex(T_i)$ are all random variables as a function of the stochastic short rate $r(T_i)$ at T_i.
- A simulation pricing approach will be:
  - Simulate N short rate trajectories $r(t), t \in [0, T_{n-1}]$.
  - On each trajectory $r^{(i)}(t)$, calculate $Berm^{(i)}(0)$ using the backward induction method.
  - Aggregate and take average to get the estimate of $Berm(0)$.

19.5.7.4 *Monte carlo simulation*

Methodology 19.5.5 (simulate the future term structure in Hull-White model). *Suppose*

- *current time is 0.*
- *we are given the current term structure $y(0, T), T \in \mathbb{R}^+$(or $P(0, T), T \in \mathbb{R}^+$, or $f(0, t), \mathbb{R}^+$).*
- *we are given a calibrated Hull-White model*

$$dr = (\theta(t) - \alpha r(t))dt + \sigma(t)dW_t.$$

- *we are given the initial short rate* $r(0)$*(which can be derived from* $r(0) = f(0,0)$*).*

Then we can generate a ***sample term structure*** *in future time* $t > 0$ *via the following procedure:*

- *simulate* $r(t)$ *by drawing from the normal distribution*
$$r(t) \sim N(m(t), v^2(t)),$$
where
$$m(t) = r(0)e^{-\alpha t} + \int_0^t \theta(u)e^{-\alpha(t-u)}du,$$
$$v^2 = \int_0^t \sigma(u)^2 e^{-2\alpha(t-u)}du.$$
- *the sample yield curve is given by*
$$y(t,T) = -\frac{\ln P(t,T)}{T-t} = -\frac{1}{T-t}(\ln\frac{P(0,T)}{P(0,t)} - A(t,T) - r(t)D(t,T)),$$
where
$$A(t,T) = -f(0,t)D(t,T) + \frac{1}{2}D^2(t,T)\int_0^t \sigma^2(u)\exp(-2\alpha(t-u))du,$$
$$D(t,T) = \int_t^T \exp(-\alpha(T-s))ds.$$

Proof. See Lemma 19.5.26. □

19.5.7.5 *Trinomial tree model for short rate*

Definition 19.5.9 (trinomial tree model). *[3, p. 78][5, p. 724] Let* $r^*(t) = f(0,t) + \frac{\sigma^2}{2a^2}(1-e^{-at})^2$ *denote the instantaneous equilibrium value of the Hull-White model. Let* $x(t) \triangleq r(t) - r^*$ *be the deviation from the equilibrium value.*

A trinomial tree model for x *consists of*

- *nodes* (i,j)*: representing date* $t = i\Delta t$*, and states* $x = j\Delta x$*, where usually we choose* $\Delta x = \sigma\sqrt{3\Delta t}$*.*

- *state transition probability:*

$$P((i+1,s)|(i,j)) = \begin{cases} p_u, s = k+1 \\ p_m, s = k \\ p_d, s = k-1 \end{cases}$$

where $k = round(M_{i,j}/\Delta r)$.

Lemma 19.5.30. *In the trinomial tree model, we have*

- $$E[x(t_{i+1})|x(t_i) = x_{i,j}] \approx x_{i,j}\exp(-a\Delta t) \triangleq M_{i,j}$$
$$Var[x(t_{i+1})|x(t_i) = x_{i,j}] \approx \frac{\sigma^2}{2a}(1 - e^{-2a\Delta t}) \triangleq V_i^2$$

- *state transition probability:*

$$P((i+1,s)|(i,j)) = \begin{cases} p_u = \frac{1}{6} + \frac{\eta_{j,k}^2}{6V_j^2} + \frac{\eta_{j,k}}{2\sqrt{3V_i}}, s = k+1 \\ p_m = \frac{2}{3} - \frac{\eta_{j,k}^2}{3V_i^2}, s = k \\ p_d = \frac{1}{6} + \frac{\eta_{j,k}^2}{6V_i^2} - \frac{\eta_{j,k}}{2\sqrt{3V_i}}, s = k-1 \end{cases}$$

- *The state transition probability satisfies*

$$p_u + p_m + p_d = 1$$
$$E[x_{i+1,k}|x_{i,j}] = M_{i,j}$$
$$Var[x_{i+1,k}|x_{i,j}] = V_i^2$$

Proof. (1) Note that SDE for x is given by

$$dx = -axdt + \sigma dW_t.$$

(2)(3) Directly verify. □

19.5.7.6 *Hull-White model calibration*

19.5.8 Two factor Gaussian model

19.5.8.1 *The model*

Definition 19.5.10 (two-factor Gaussian additive model). *The two-factor OU process is given by the following SDE*

$$\begin{aligned} r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\ dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

Lemma 19.5.31 (zero-coupon bond price in two-factor Gaussian model). *Consider a two factor Gaussian additive short rate model in Definition 19.5.10.*

It follows that

- $$\begin{aligned} P(t,T) &\triangleq E[\exp(-\int_t^T r(s)ds)] \\ &= F_1(t,T,x_1(t))F_2(t,T,x_2(t))\exp(U(t,T))\exp(-\int_t^T \psi(u)du). \end{aligned}$$

where

$$\begin{aligned} F_i(t,T,x_1(t)) &\triangleq E[\exp(-\int_t^T x_1(s)ds)] \\ &= \exp(-\frac{1-\exp(-a_1(T-t))}{a_1}x_1(t) + \frac{1}{2}(\frac{\sigma^2}{a_1^2}(T-t+ \\ &\frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}))), i = 1,2,. \end{aligned}$$

and

$$\begin{aligned} U(t,T) = \frac{\rho\sigma_1\sigma_2}{a_1 a_2}(T-t+\frac{\exp(-a_1(T-t))-1}{a_1} + \frac{\exp(-a_2(T-t))-1}{a_2} \\ + \frac{\exp(-(a_1+a_2)(T-t))-1}{a_1+a_2}). \end{aligned}$$

- *If we choose $\psi(T), T \in \mathbb{R}^+$ to match the current term structure given by $f^M(0,T)$, then*

$$\psi(T) = f^M(0,T) + \frac{\sigma_1^2}{2a_1^2}(1-\exp(-a_1T))^2 + \frac{\sigma_1^2}{2a_2^2}(1-\exp(-a_2T))^2 + \rho\frac{\sigma_1\sigma_2}{a_1a_2}(1-\exp(-a_1T))(1-\exp(-a_2T))$$

- *The zero-coupon bond price, which agrees with the current term structure, is given by*

$$\exp(-\int_t^T \psi(u)du) = \frac{P^M(0,T)}{P^M(0,t)}\exp(-\frac{1}{2}(V(0,T)-V(0,t)).$$

- *With the choice of ϕ in (2), we have*

$$P(t,T) = \frac{P^*(0,T)F(0,t,x_{10},x_{20})}{P^*(0,t)F(0,T,x_{10},x_{20})}F(t,T,x_1(t),x_2(t)),$$

where $P^*(0,T)$ is the current market term structure and

$$F(t,T,x_1(t),x_2(t)) = F_1(t,T,x_1(t))F_2(t,T,x_2(t))\exp(U(t,T))$$

Proof. Use Lemma 6.4.8 and Theorem 19.5.3. □

Lemma 19.5.32 (bond option). *[1, p. 65]*

$$v(0,S) = \frac{\sigma_1^2}{2a_1^3}(1-\exp(-a_1(T-S)))^2(1-\exp(-2a_1S)) + \frac{\sigma_2^2}{2a_2^3}(1-\exp(-a_2(T-S)))^2(1-\exp(-2a_2
+ \frac{2\rho\sigma_1\sigma_2}{a_1a_2(a_1+a_2)}((1-\exp(-a_1(T-S)))(1-\exp(-a_2(T-S)))(1-\exp(-(a_1+a_2)S)))$$

19.5.8.2 *Yield curve dynamics*

19.5.8.3 *Instantaneous forward rate dynamics*

Lemma 19.5.33 (instantaneous forward rate dynamics in two factor Gaussian model). **??**109]privault2012elementary *Consider the short rate under risk neutral measure is evolving under the Vasicek model given by*

$$dr(t) = (\theta - \alpha r(t))dt + \sigma dW_t.$$

It follows that

- *The instantaneous forward rate is given by*

$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T} \quad = \psi(T) + f_1(t,T) + f_2(t,T) - \rho\frac{\sigma_1\sigma_2}{a_1a_2}(1-\exp(-a_1(T-t)))(1-\exp(-$$

where

$$f_1(t,T) = x_1(t)\exp(-a_1(T-t)) - \frac{\sigma_1^2}{2a_1^2}(1-\exp(-a_1(T-t))^2)$$

and

$$f_2(t,T) = x_2(t)\exp(-a_2(T-t)) - \frac{\sigma_1^2}{2a_2^2}(1-\exp(-a_2(T-t))^2)$$

- *The dynamics of instantaneous forward rate is given by*

$$d_tf(t,T) = d_tf_1 + d_tf_2 - \rho\sigma_1\sigma_2\exp(-a_1(T-t))\exp(-a_2(T-t))dt$$

where

$$d_tf_1(t,T) = \frac{\sigma_1^2}{a_1}(1-\exp(-a_1(T-t)))\exp(-a_1(T-t))dt + \sigma\exp(-a_1(T-t))dW_1$$

and

$$d_tf_2(t,T) = \frac{\sigma_2^2}{a_2}(1-\exp(-a_2(T-t)))\exp(-a_2(T-t))dt + \sigma\exp(-a_2(T-t))dW_2$$

- *We can write the forward rate dynamics in a more compact form given by*

$$d_tf(t,T) = \alpha(t,T)dt + \sigma_f(t,T)dW$$

where the forward rate volatility is given by

$$\sigma_f(t,T) = \sqrt{\sigma_1^2\exp(-a_1(T-t)) + \sigma_2^2\exp(-a_2(T-t)) + \sigma_2^2 + 2\rho\sigma_1\sigma_2\exp(-a_1(T-t))\exp(-a_2}$$

and the forward rate drift is given by

$$\alpha(t,T) = \sigma_f(t,T)\int_t^T \sigma_f(t,u)du.$$

Proof. (1) Note that in the two factor Gaussian model, we have zero-coupon bond price given by(Definition 19.5.10)

$$P(t,T) = F_1(t,T,x_1(t))F_2(t,T,x_2(t))\exp(U(t,T))\exp(-\int_t^T \psi(u)du).$$

We can easily verify

$$f_1(t,T) = -\frac{\ln F(t,T,x_1(t))}{dT}, f_2(t,T) = -\frac{\ln F(t,T,x_2(t))}{dT}$$

using the results in Vasicek model(Lemma 19.5.12). and the instantaneous forward rate is given by

$$f(t,T) = -\frac{\ln P(t,T)}{dT} = r(t) + \alpha(T-t) - \frac{1}{2}\sigma^2(T-t)^2.$$

(2)

$$\begin{aligned}
df_t(t,T) &= \exp(-k(T-t))dr(t) + br_t\exp(-k(T-t))dt - \theta\exp(-k(T-t)) \\
&+ \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt \\
&= \exp(-k(T-t))((\theta-\alpha r(t))dt + \sigma dW_t) + br_t\exp(-k(T-t))dt - \theta\exp(-k(T-t)) \\
&+ \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt \\
&= \frac{\sigma^2}{k}(1-\exp(-k(T-t)))\exp(-k(T-t))dt + \sigma\exp(-k(T-t))dW_t
\end{aligned}$$

□

Remark 19.5.21 (humped forward volatility). [3, p. 149]

- A humped volatility structure similar to what is commonly observed in the market for the caplets volatility, may be only reproduced for negative values of correlation ρ.
- If ρ is positive, then all terms in σ_f is decreasing and the hump is impossible.

19.5.8.4 *Correlation problems in factor models*

Note 19.5.3 (correlation in one-factor affine model). [9, p. 103] Consider a general one-factor model such that the zero-coupon bond price can be written by

$$P(t,T) = A(t,T)\exp(-B(t,T)r(t)).$$

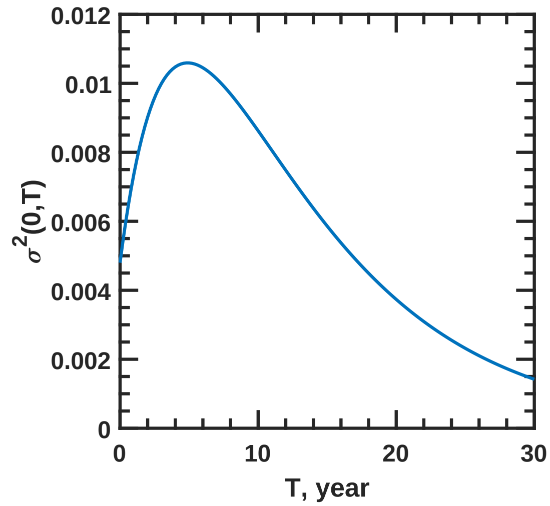


Figure 19.5.7: Humped shaped forward volatility produced with parameters: sig1 = 0.12; sig2 = 0.1; rho = -1; k1 = 0.05; k2 = 0.2;

- The zero-coupon bonds with maturities T_1 and T_2 are related by

$$\frac{P(t,T_2)}{P(t,T_1)} = \frac{A(t,T_2)}{A(t,T_1)}\exp(-r_t(B(t,T_1) - B(t,T_2)).$$

- Take the log on both sides, we have

$$\begin{aligned}
\ln P(t,T_2) &= \ln P(t,T_1) + \ln A(t,T_2) - \ln A(t,T_1) - r_t(B(t,T_2) - B(t,T_1)) \\
&= \ln P(t,T_1) + \ln A(t,T_2) - \ln A(t,T_1) + \frac{\ln P(t,T_1) - \ln A(t,T_1)}{B(t,T_1)}(B(t,T_2) - B(t,T_1)) \\
&= \frac{B(t,T_2)}{B(t,T_1)}\ln P(t,T_1) + \ln A(t,T_2) - \ln A(t,T_1) - \frac{\ln A(t,T_1)}{B(t,T_1)}(B(t,T_1) - B(t,T_2))
\end{aligned}$$

- We can see the $\ln P(t,T_2)$ and $\ln P(t,T_1)$ are linear deterministic function of each other; therefore from **??**

$$corr(\ln P(t,T_1), \ln P(t,T_2)) = \pm 1.$$

19.5.8.5 *Two factor Hull-White model*

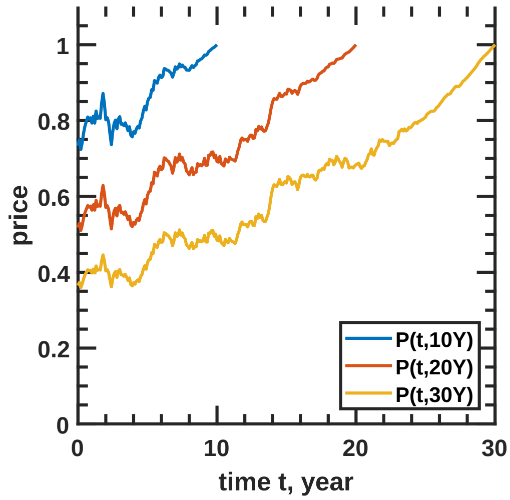


Figure 19.5.8: One realization of zero-coupon bond price evolution until maturity. It is clear that the price movement of bonds(10Y, 20Y, and 30Y) are similar to each other.

Definition 19.5.11. *[1, p. 62] In the two-factor Hull-White model, we assume,* ***under risk-neutral measure****, the short rate dynamics is given by*

$$\begin{aligned} dr(t) &= (\theta(t) + u(t) - \alpha r(t))dt + \delta dW(t) \\ du(t) &= -\beta u(t)dt + \epsilon dZ(t) \end{aligned}$$

where $E[dW(t)dZ(t)] = \rho dt$.

Remark 19.5.22. The two factor Hull-White model is better at capturing the correlation of rates of different maturities.

19.5.8.6 *Multi-factor Gaussian model*

19.5.8.7 *The model*

Definition 19.5.12 (multi-factor Gaussian additive model). *The two-factor OU process is given by the following SDE*

$$\begin{aligned} r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\ dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \\ &\cdots\cdots \\ dx_n(t) &= -a_n x_2(t)dt + \sigma_n dW_n(t), x_n(0) = x_{n0} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

Lemma 19.5.34 (zero-coupon bond price in multi-factor Gaussian additive model). *Consider a multi-factor Gaussian additive short rate model in Definition 19.5.12.*

It follows that

- $$\begin{aligned} P(t,T) &\triangleq E[\exp(-\int_t^T r(s)ds)] \\ &= \prod_{i=1}^n F_i(t,T,x_i(t))\exp(U(t,T))\exp(-\int_t^T \psi(u)du). \end{aligned}$$

  where

  $$\begin{aligned} F_i(t,T,x_1(t)) &\triangleq E[\exp(-\int_t^T x_1(s)ds)] \\ &= \exp(-\frac{1-\exp(-a_1(T-t))}{a_1}x_1(t) + \frac{1}{2}(\frac{\sigma^2}{a_1^2}(T-t+ \\ &\frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}))), i = 1,2,...,n. \end{aligned}$$

  and

  $$\begin{aligned} U(t,T) = \sum_{1\leq i<j\leq n} \frac{\rho_{ij}\sigma_i\sigma_j}{a_i a_j}(T-t+\frac{\exp(-a_i(T-t))-1}{a_i} + \frac{\exp(-a_j(T-t))-1}{a_j} \\ + \frac{\exp(-(a_i+a_j)(T-t))-1}{a_i+a_j}). \end{aligned}$$

- *If we choose $\psi(T), T \in \mathbb{R}^+$ to match the current term structure given by $f^M(0,T)$, then*

  $$\begin{aligned} \psi(T) &= f^M(0,T) + \sum_{i=1}^n \frac{\sigma_i^2}{2a_i^2}(1-\exp(-a_iT))^2 \\ &+ \sum_{1\leq i<j\leq n} \rho_{ij}\frac{\sigma_i\sigma_j}{a_i a_j}(1-\exp(-a_iT))(1-\exp(-a_jT)) \end{aligned}$$

- *The zero-coupon bond price, which agrees with the current term structure, is given by*

$$\exp(-\int_t^T \psi(u)du) = \frac{P^M(0,T)}{P^M(0,t)}\exp(-\frac{1}{2}(V(0,T)-V(0,t)).$$

- *With the choice of ϕ in (2), we have*

$$P(t,T) = \frac{P^*(0,T)F(0,t,x_{10},x_{20},...,x_{n0})}{P^*(0,t)F(0,T,x_{10},x_{20},x_{n0})}F(t,T,x_1(t),x_2(t),...,x_n(t)),$$

where $P^*(0,T)$ is the current market term structure and

$$F(t,T,x_1(t),x_2(t),...,x_n(t)) = \prod_{i=1}^n F_i(t,T,x_i(t))\exp(U(t,T))$$

Proof. See Lemma 19.5.31. □

19.5.9 Log-normal type model

19.6 Short-rate model: affine term structure models

19.6.1 General theorem

Definition 19.6.1 (affine term structure short rate model). *[10, p. 512]*

- *Consider a time-homogeneous risk-neutral multidiemnsional dynamical system given by*

$$dx(t) = \mu(x(t))dt + \sigma(x(t))dW(t),$$

where $x(t) = (x_1(t), x_2(t), ..., x_d(t))^T$ has state space $D \subseteq \mathbb{R}^d$; $W(t)$ is a d dimensional independent Brownian motion.
And the short rate $r(t)$ is associated with state $x(t)$ via

$$r(t) = F(x(t)).$$

An affine term structure short rate model is such that $\mu, \sigma\sigma^T$ and F are affine functions of x.

- *Equivalently, an affine term structure short rate model has the following representation:*

$$dx(t) = (a - bx(t))dt + \Sigma \begin{pmatrix} \sqrt{v_1(x(t))} & 0 & \cdots & 0 \\ 0 & \sqrt{v_1(x(t))} & \vdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{v_1(x(t))} \end{pmatrix} dW(t),$$

where $a \in \mathbb{R}^d, b \in \mathbb{R}^{d\times d}$, *and*

$$v_i(x) = \alpha_i + \beta_i^T x, \alpha_i \in \mathbb{R}, \beta_i \in \mathbb{R}^d, i = 1, 2, ..., d$$

And the short rate $r(t)$ *is associated with state* $x(t)$ *via*

$$r(t) = \xi_0 + \xi^T x(t) = \xi_0 + \sum_{i=1}^{d} \xi_i^T x(t).$$

An affine term structure model is such that $\mu, \sigma\sigma^T$ *and* F *are affine functions of* x.

Remark 19.6.1 (parameter constraint). [10, p. 512] Note that there are constrains on the possible choices of μ, σ, F or $\alpha, \beta, a, b, \xi_0, \xi_1$ in order for the SDE to have a solution. It is discussed in the reference.

Theorem 19.6.1 (soluiton to affine term structure short rate model). *[10, p. 515] The zero coupon bond solution in the affine term structure model in Definition 19.6.1 is given by*

$$P(t, T, x) = \exp(A(T-t) - B(T-t)^T x) = \exp(A(T-t) - \sum_{i=1}^{d} B_i(T-t)x_i),$$

where the real-valued function $A(\tau)$ *and the vector-valued function* $B(\tau)$ *satisfy the system of Riccati ODE equations given by*

$$\frac{d}{d\tau}B(\tau) = -b^T B(\tau) - \frac{1}{2}\beta^T diag(\Sigma^T B(\tau))\Sigma^T B(\tau) + \xi,$$
$$\frac{d}{d\tau}A(\tau) = -a^T B(\tau) + \frac{1}{2}\alpha^T diag(\Sigma^T B(\tau))\Sigma^T B(\tau) - \xi_0,$$

with initial conditions $A(0) = 0, B(0) = 0$, where $\alpha = (\alpha_1, ..., \alpha_d)^T$ and ith row of matrix β is given by $\beta_i^T, i = 1, ..., d$. The ODE system can be written component-wise as

$$\frac{d}{dr}B_i(\tau) = -\sum_{j=1}^{d} b_{j,i}B_j(\tau) - \frac{1}{2}\sum_{k=1}^{d}(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau))^2 + \xi_i, i = 1, 2, ..., d,$$
$$\frac{d}{dr}A(\tau) = -\sum_{j=1}^{d} a_jB_j(\tau) + \frac{1}{2}\sum_{k=1}^{d}\alpha_k(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau))^2 - \xi_0$$

Proof. □

Remark 19.6.2 (fitting initial term strucctture). To fit the initial term structure using affine term structure model, we can use deterministic shift extension in Theorem 19.5.3.

19.6.2 One factor affine term structure model

Lemma 19.6.1 (solution to one-factor affine term structure model). *The zero coupon bond solution in the affine term structure model in Definition 19.6.1 is given by*

$$P(t, T, x) = \exp(A(T-t) - B(T-t)^T x) = \exp(A(T-t) - \sum_{i=1}^{d} B_i(T-t)x_i),$$

where the real-valued function $A(\tau)$ and the vector-valued function $B(\tau)$ satisfy the system of Riccati ODE equations given by

$$\frac{d}{d\tau}B(\tau) = -b^T B(\tau) - \frac{1}{2}\beta^T diag(\Sigma^T B(\tau))\Sigma^T B(\tau) + \xi,$$
$$\frac{d}{d\tau}A(\tau) = -a^T B(\tau) + \frac{1}{2}\alpha^T diag(\Sigma^T B(\tau))\Sigma^T B(\tau) - \xi_0,$$

with initial conditions $A(0) = 0, B(0) = 0$, where $\alpha = (\alpha_1, ..., \alpha_d)^T$ and ith row of matrix β is given by $\beta_i^T, i = 1, ..., d$. The ODE system can be written component-wise as

$$\frac{d}{dr}B_i(\tau) = -\sum_{j=1}^{d} b_{j,i}B_j(\tau) - \frac{1}{2}\sum_{k=1}^{d}(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau))^2 + \xi_i, i = 1, 2, ..., d,$$
$$\frac{d}{dr}A(\tau) = -\sum_{j=1}^{d} a_jB_j(\tau) + \frac{1}{2}\sum_{k=1}^{d}\alpha_k(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau))^2 - \xi_0$$

Proof.

$$\frac{\partial P(t,T)}{\partial t} = \frac{A'}{A}P(t,T) - rB'P(t,T)$$
$$\frac{\partial P(t,T)}{\partial r} = -B(t,T)P(t,T)$$
$$\frac{\partial^2 P(t,T)}{\partial r \partial r} = B^2(t,T)P(t,T)$$

□

[11]

Lemma 19.6.2. *Consider a one-factor affine term structure short-rate model such that the zero-coupon bond price is given by*

$$P(t,T) \triangleq E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] = \exp(A(t,T) - B(t,T)r(t)),$$

with short rate model given by

$$dr(t) = \alpha(r(t),t)dt + \sigma(r(t),t)dW_t.$$

It follows that

- *Under risk-neutral measure, the SDE for $P(t,T)$ is lognormal and given by*
$$dP(t,T)/P(t,T) = r(t)dt - \sigma(t)B(t,T)dW_t^Q.$$
- *A forward bond with forward delivery date T_F and bond maturity date $T_P, T_F \leq T_P$, with its price denoted by $FB(t,T_F,T_P) = P(t,T_F)/P(t,T_P)$, has dynamics under the T_F forward measure given by*
$$dFB(t,T_F,T_P)/FB(t,T_F,T_P) = \sigma(t)(B(t,T_P) - B(t,T_F))dW^{T_P}.$$
- *The log variance of $FB(T_F,T_P)$, or $P(T_F,T_P)$ is given by*
$$Var[\ln FB(T_F,T_P)|\mathcal{F}_t] = B^2(T_F,T_P)Var[r(t_F)|\mathcal{F}_t].$$

Proof. (1) Note that
$$\frac{\partial P(t,T)}{\partial r} = -B(t,T)P(t,T).$$
Using the term structure function theory (Theorem 19.5.2), we have

$$dP(t,T) = r(t)P(t,T) + \sigma\frac{\partial P(t,T)}{\partial r}dW_t = r(t)P(t,T) - \sigma B(t,T)P(t,T).$$

(2)

(3) Use the fact that

$$\ln FB(T_F, T_P) = -\ln P(T_F, T_P) = -\ln A(T_F, T_P) + B(T_F, T_P) r(T_F).$$

□

19.6.3 Approximate swaption pricing

Lemma 19.6.3 (swaption approximation pricing). *[12] Denote by $S(t, T_0, T_n)$ the swap rate prevailing at time t for the swap starting at T_0 and ending on T_n, with $t < T_0 < T_n$. The swap rate is given by*

$$S(t, T_0, T_N) = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)} = S(t, T_0, T_N) = \frac{P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)} (\frac{P(t, T_0)}{P(t, T_n)} - 1).$$

Define the approximator $\tilde{S}(t, T_0, T_N)$ by

$$\tilde{S}(t, T_0, T_N) = \frac{P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)} (\frac{P(t, T_0)}{P(t, T_n)} - 1).$$

•
$$\begin{aligned}\frac{d\tilde{S}(t, T_0, T_n)}{\tilde{S}(t, T_0, T_n)} &= \frac{1}{\tilde{S}(t, T_0, T_n)} \frac{P(0, T_n)}{\sum_{i=1}^{n} \tau_i P(0, T_i)} d(\frac{P(t, T_0)}{P(t, T_n)}) \\ &= \frac{S(0, T_0, T_n) P(0, T_n) P(t, T_n)}{\tilde{S}(t, T_0, T_n)[P(0, T_0) - P(0, T_n)] P(t, T_n)} \sigma(t)(B(t, T_n) - B(t, T_0)) dW^{T_n}(t) \\ &= drift\ + \frac{P(0, T_n)}{(P(0, T_0) - P(0, T_n))} \sigma(t)(B(t, T_n) - B(t, T_0)) dW^A(t)\end{aligned}$$

•
$$V_{swap}(T_0, T_n) = (\frac{P(0, T_0)}{P(0, T_0) - P(0, T_n)})^2 V_p(0, T_0, T_n)$$

Methodology 19.6.1 (Hull-White model calibration).

$$\begin{aligned}\sigma(t) &= \sigma_i, t \in [t_i, t_{i+1}), i = 0, ..., N-2 \\ \sigma(t) &= \sigma_{N-1}, t \geq t_{N-1}\end{aligned}$$

19.6.4 Multi-factor affine term structure model

19.7 Short-rate model: quadratic Gaussian model

19.7.1 The model

Definition 19.7.1 (quadratic Gaussian model). *[13][10, p. 520] Under the risk-neutral measure Q, define a multidimensional Gaussian process*

$$dz(t) = \sigma(t)dW(t),$$

where $z(t)$ is N-dimensional state vector, $\sigma(t)$ is a deterministic $N \times N$ matrix-valued function such that $\sigma(t)\sigma(t)^T$ is nonsingular, and $W(t)$ is an N-dimensional standard independent Brownian motion.

*An N **dimensional quadratic Gaussian model** is defined by specifying that the short rate is a quadratic form of the Gaussian vector $z(t)$ given by*

$$r(t) = z(t)^T\gamma(t)z(t) + \beta(t)^T z(t) + \alpha(t),$$

where $\gamma(t)$ is an $N \times N$ symmetric matrix, and $\beta(t)$ is an N dimensional vector, and $\alpha(t)$ is a scalar deterministic function used to fit the initial yield curve.

19.7.2 One-factor Quadratic Gaussian model

Definition 19.7.2 (One factor Quadratic Gaussian model). *[10, p. 443] A One-factor Quadratic Gaussian model for the short rate $r(t)$ under the risk-neutral measure is given by*

$$r(t) = \alpha(t) + \beta(t)y(t) + \gamma(t)y(t)^2,$$

where

$$dy(t) = -k(t)y(t)dt + \sigma(t)dW(t), y(0) = 0.$$

Remark 19.7.1 (reduction to two-factor linear Gaussian model)**.** [10, p. 443] If we denote $u(t) = y(t)^2$, then we can see that $r(t)$ is a linear function of the state vector $(y(t), u(t))$, which follows the SDEs

$$\begin{aligned} dy(t) &= -k(t)y(t)dt + \sigma(t)dW_1(t), y(0) = 0 \\ du(t) &= \sigma(t)^2 - 2k(t)u(t)dt + 2\sigma(t)y(t)dW_2(t), u(0) = 0 \end{aligned}$$

19.8 Tree implementation of short rate model

19.8.1 Hull-White Tree

19.8.1.1 *Basic setup*

Definition 19.8.1 (Hull White trinomial tree model). *A binomial tree model for an asset (e.g. stock) dynamics consists of*

- *nodes (i,j): representing date $i = 0, 1, ..., N$ and states $j = 0, 1, ..., N$*
- *values on node (i,j) represent the asset price at i and state j.*
- *the binomial tree model is a representation of discrete-time asset stochastic process $S(1), S(2), ..., S(N)$, where $S(i)$ can take values $s_{i,1}, ..., s_{i,i}$.*
- *state transition probability:*

$$P((i+1,s)|(i,j)) = \begin{cases} q, s = j+1 \\ 1-q, s = j \end{cases}$$

- *the sample space Ω consists of all possible 2^N sample paths; We use $\omega_{i,j}$ to denote the set of sample paths that pass through state j at time i.*

19.8.1.2 *State price*

Definition 19.8.2 (Arrow-Debreu security). *An **Arrow-Debreu security** is a security that has a payoff that solely pays 1 at time n if world state $\omega_{n,j}$ is realized. Its price, called, **state price** at $t = 0$ is denoted as $\lambda_{n,j}$*

With the existence of risk-neutral measure Q, we have

$$\lambda_{n,j} = E_Q[\sum_{\omega \in \Omega_{n,j}} \prod_{i=1}^{n-1} \frac{1}{1 + R_i(\omega)\Delta t}].$$

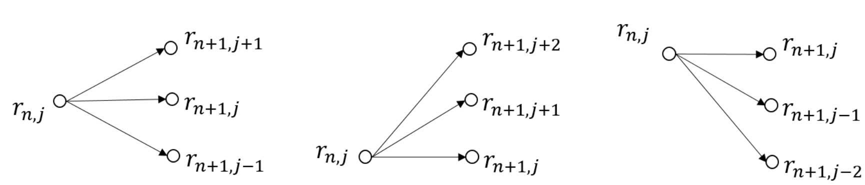


Figure 19.8.1

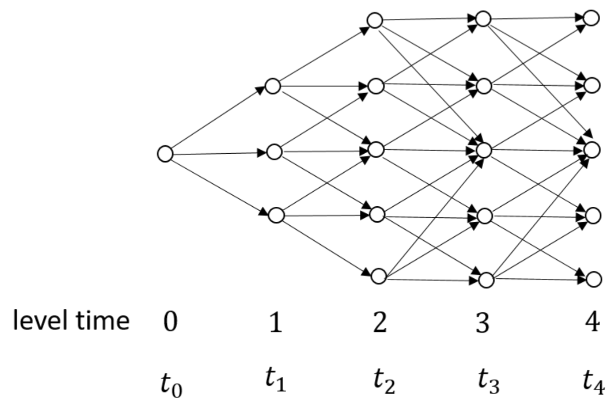


Figure 19.8.2: Hull White Tree example.

Further, the probability of reaching state $s_{n,j}$ from state $s_{0,0}$ is given by

$$Pr_Q(S_n = s_{n,j}) = (1 + r\Delta t)^n \lambda_{n,j}.$$

Theorem 19.8.1 (forward equation pricing). *In a K period binomial lattice model, where there will be $K+1$ states, then state prices satisfying **forward equations**:*

$$\lambda_{n+1,i} = \sum_{j\in Pa(i)} \frac{1}{1 + r_{n,j}\Delta t} Q_k(i;j)\lambda_{k,s-1}$$

*where $Pa(i)$ denote the parent nodes of (n,i) and we have **boundary condition** $\lambda_{0,0} = 1$, $Q_n(s';s)$ denotes the risk-neutral measure for transitioning probability from state s at period n to state s' at period $n+1$, and r_{ij} is the interest rate at period i and state j.*

Proof. By definition,

$$\begin{aligned}
\lambda_{n+1,i} &= E_Q[\sum_{\omega\in\Omega_{n+1,i}}\prod_{k=1}^{n}\frac{1}{1+R_k(\omega)\Delta t}] \\
&= E_Q[\sum_{j\in Pa(i)}\sum_{\omega\in\Omega_{n+1,i}\cap\Omega_{n,j}}\prod_{k=1}^{n}\frac{1}{1+R_k(\omega)\Delta t}] \\
&= E_Q[\sum_{j\in Pa(i)}\sum_{\omega\in\Omega_{n,j}}\prod_{k=1}^{n-1}\frac{1}{1+R_k(\omega)\Delta t}\frac{1}{1+r_{n,j}\Delta t}Q_n(i,j)] \\
&= \sum_{j\in Pa(i)}E_Q[\sum_{\omega\in\Omega_{n,j}}\prod_{k=1}^{n-1}\frac{1}{1+R_k(\omega)\Delta t}]\frac{1}{1+r_{n,j}\Delta t}Q_n(i,j) \\
&= \sum_{j\in Pa(i)}\lambda_{n,j}\frac{1}{1+r_{i,j}\Delta t}Q_n(i,j)
\end{aligned}$$

□

Lemma 19.8.1 (pricing using Hull-White Tree). *Suppose in a Hull-White tree the state price $\lambda_{i,j}$ is known for all possible i,j.*

- *For a zero coupon bond that has a payoff 1 at period n, its no-arbitrage price at time 0 is*
$$V(0)=\sum_{j=1}^{n+1}\lambda_{n,j}=\frac{1}{(1+r\Delta t)^n}.$$
- *For a deterministic cash flow C at period n, its no-arbitrage price at time 0 is*
$$V(0)=\sum_{j=1}^{n+1}C\lambda_{n,j}=\frac{1}{(1+r\Delta t)^n}.$$

19.8.2 Black-Karasinski tree

19.8.3 Black-Derman-Toy tree

19.8.4 State price

19.9 Forward rate model: HJM framework

The HeathâĂŤJarrowâĂŤMorton (HJM)[**heath1992bond**] framework is a general framework to model the evolution of interest rate curve by specifying the dynamics of instantaneous forward rate curve. The key to these techniques is the recognition that the drifts of the no-arbitrage evolution of certain variables can be expressed as functions of their volatilities and the correlations among themselves. In other words, no drift estimation is needed.

When the volatility and drift of the instantaneous forward rate are assumed to be deterministic, this is known as the Gaussian HeathâĂŤJarrowâĂŤMorton (HJM) model of forward rates. We can see that short rate model is a dynamical reduction of the HJM model.

19.9.1 Single factor HJM framework

19.9.1.1 *Principles*

Theorem 19.9.1 (HJM framework for one factor forward rate dynamics). *[1, p. 69]* *Assume that the forward rate $f(t,T)$,* ***under risk-neutral measure*** *Q,is given*

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

where $W(t)$ is a Brownian motion, and $\sigma(t,T)$ is adapted to filtration $\mathcal{F}_t$ for each $T > 0$, and

$$\int_0^T \int_0^T |\alpha(s,u)|\, dsdu < \infty, \int_0^T (\int_0^T |\sigma(s,u)|^2\, ds)^{1/2} du < \infty$$

almost surely. Then,

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du.$$

As a consequence,

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T)(\int_s^T \sigma(s,u)du)dt + \int_0^t \sigma(s,T)dW(s).$$

Furthermore, the zero-coupon bond price $P(t,T)$ satisfies SDE

$$dP(t,T) = r(t)P(t,T)dt + \Sigma(t,T)P(t,T)dW(t)$$

with log-volatility

$$\Sigma(t,T) = -\int_t^T \sigma(t,u)du.$$

Proof. Under risk-neutral measure Q, the bond price is related to forward rate as

$$P(t,T) = \exp(-\int_t^T f(t,u)du).$$

Use

$$\begin{aligned} d(\int_t^T f(t,u)du) &= -f(t,t)dt + \int_t^T (\alpha(t,u)dt + \sigma(t,u)dW(t))du \\ &= -r(t)dt + (\int_t^T \alpha(t,u)du)dt + (\int_t^T \sigma(t,u)du)dW(t) \end{aligned}$$

Then

$$\begin{aligned} dP(t,T) &= d(\exp(-\int_t^T f(t,u)du)) \\ &= P(t,T)d(\int_t^T f(t,u)du) \\ &= (r(t) - \int_t^T \alpha(t,u)du)du)^2)P(t,T)dt - (\int_t^T \sigma(t,u)du)P(t,T)dW(t) \\ &+ \frac{1}{2}(\int_t^T \sigma(t,u)du)dW(t)(\int_t^T \sigma(t,u)du)dW(t) \\ &= (r(t) - \int_t^T \alpha(t,u)du)du)^2)P(t,T)dt - (\int_t^T \sigma(t,u)du)P(t,T)dW(t) \\ &+ \frac{1}{2}(\int_t^T \sigma(t,u)du)(\int_t^T \sigma(t,u)du)dt \end{aligned}$$

Because under risk-neutral measure Q, $P(t,T)$ will have drift $r(t)$(Lemma 19.5.3). Therefore,

$$r(t) - \int_t^T \alpha(t,u)du + \frac{1}{2}(\int_t^T \sigma(t,u)du)^2 = r(t),$$

Then

$$\int_t^T \alpha(t,u)du = \frac{1}{2}(\int_t^T \sigma(t,u)du)^2.$$

Differentiating both sides with respect to T, we have

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du.$$

□

Remark 19.9.1 (current term structure always matched). From the HJM result, we have the forward curve evolution function as

$$f_{predict}(t,T) = f_{observed}(0,T) + \int_0^t \alpha(s,T)(\int_s^T \sigma(s,u)du)dt + \int_0^t \sigma(s,T)dW(s).$$

If we set $t = 0$, then we have

$$f_{predict}(t,T) = f_{observed}(0,T);$$

that is, our prediction on current term structure is consistent with the observation.

Remark 19.9.2 (calculating the short rate dynamics). We can obtain the short rate using the definition

$$r(t) \triangleq f(t,t)$$

to obtain

$$r(t) = f(0,t) + \int_0^t \alpha(s,t)(\int_s^t \sigma(s,u)du)dt + \int_0^t \sigma(s,t)dW(s).$$

Example 19.9.1 (parallel shift arbitrage). Suppose current forward rate curve is given by $f(0,T), T > 0$. At a future time $t > 0$, the forward curve is predicted(using model) to be **deterministically** given by $f(t,T) = f(0,T) + \epsilon, \epsilon > 0$. Then the parallel shift of the forward rate is known to adopt arbitrage([14]).

From the HJM no-arbitrage condition, it is immediately clear that for a deterministic dynamics(i.e. volatility is zero), the drift has to be 0; that is, the forward curve should remain constant.

Lemma 19.9.1 (correlation structure in one-factor HJM forward model). *Let $T_1, T_2, ..., T_N$ be two maturity dates. Then*

- *The N dimensional stochastic process $(f(t,T_1), f(t,T_2), ..., f(t,T_N))$ has each component being a Gaussian process, but might not be joint Gaussian*

- *The pair correlation is given by*

$$Cov[f(t,T_1), f(t,T_2)] = \int_0^t \sigma(s,T_1)\sigma(s,T_2)ds.$$

Proof. (1)The one-factor HJM forward rate model, for all $T > 0$,

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t$$

is a state-independent linear SDE, thus Gaussian process(Lemma 6.3.9). Consider $\sigma(t,T_1) = \sigma, \sigma(t,T_2) = \sigma$, then $f(t,T_1) = f(t,T_2)$ such that $(f(t,T_1), f(t,T_2))$ is not a joint normal(their covariance matrix is singular). (2) Using Ito isometry. □

Lemma 19.9.2 (forward rate dynamics under forward measure). *[4, p. 116] The forward rate dynamics under $T-$maturity forward measure Q_T is given by*

$$df(t,T) = \sigma(t,T)dW^T(t),$$

where $W^T(t)$ is a Brownian motion under Q_T.

Proof. Under risk-neutral measure Q, the forward rate dynamics (Theorem 19.9.1)is given by

$$df(t,T) = -\sigma(t,T)\Sigma(t,T)dt + \sigma(t,T)dW(t),$$

where

$$\Sigma(t,T) = -\int_t^T \sigma(t,u)du.$$

Under the forward measure Q_T,

$$dW(t) = dW(t)^T + \Sigma(t,T)dt,$$

where $W^T(t)$ is a Brownian motion under forward measure.

Then

$$df(t,T) = -\sigma(t,T)\Sigma(t,T)dt + \sigma(t,T)(dW^T(t) + \Sigma(t,T)dt) = \sigma(t,T)dW^T(t).$$

□

Note 19.9.1 (single-factor model as a reduced form model). Note that we can also write the single-factor model by

$$df(t,T) + \alpha(t,T)dt + \sigma(t,T)dZ^T(t),$$

where we have a set(infinitely many) of Brownian motions $Z^T(t)$ indexed by the maturity T. By requiring $f(t,T)$ is driven by 1 factors, we are requiring

$$Rank(dZ^T[dZ^T]^T) = 1.$$

19.9.1.2 *Gaussian HJM model*

Lemma 19.9.3 (Reduce to Ho-Lee model). *[1, p. 71] Let $\sigma(t,T) = \sigma$ for all $t \leq T$. Then*

$$\alpha(t,T) = \sigma \int_t^T \sigma du = \sigma^2(T-t),$$

and

$$\begin{aligned} f(t,T) &= f(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T)dW(s) \\ &= f(0,T) + \frac{1}{2}\sigma^2 t(2T-t) + \sigma W(t). \end{aligned}$$

Then the short rate is given by

$$r(t) = f(t,t) = f(0,t) + \frac{1}{2}\sigma^2 t^2 + \sigma W(t)$$

and

$$dr(t) = (\frac{\partial f(0,t)}{\partial t} + \sigma^2 t)dt + \sigma dW(t).$$

By Ho-Lee model zero coupon bond price Lemma 19.5.23, we can calculate the zero-coupon bond price as

$$P(t,T) = \exp(-(T-t)(r(t) - f(0,t)) - \int_t^T f(0,s)ds - \frac{1}{2}\sigma^2 t(T-t)^2).$$

Moreover, it can be showed that this prediction is consistent with current term structure, showed by

$$P(0,T) = \exp(-\int_0^T f(0,s)ds).$$

Lemma 19.9.4 (covariance in parallel shift forward rate dynamics). *Consider the Ho-Lee model for forward rate, which is given by*

$$df(t,T) = \sigma^2(T-t)dt + \sigma^2 dW_t,$$

and

$$f(t,T) = f(0,T) + \frac{1}{2}\sigma^2 t(2T-t) + \sigma W(t).$$

Then for given t_0 and T_0,

- *$f(t_0,T_0)$ and $f(t_0+s, T_0+s), s \geq 0$ have the covariance given by $\sigma^2(t_0)$, even though they might have different means.*
- *$f(t_0,T_0)$ and $f(t_0+s, T_0+s), s \geq 0$ have the correlation given by*

$$\rho = \frac{t_0}{\sqrt{t_0(t_0+s)}}$$

- *$f(t_0,T_0)$ and $f(t_1,T_1)$ have the covariance given by $\sigma^2 \min(t_0,t_1)$, even though they might have different means.*

Proof. (1)(2)We can show that

$$f(t_0,T_0) = f(0,T_0) + \frac{1}{2}\sigma^2 t_0(2T_0 - t_0) + \sigma W(t_0).$$

and

$$f(t_0+s, T_0+s) = f(0,T_0) + \frac{1}{2}\sigma^2 (t_0+s)(2T_0 - t_0 + s) + \sigma W(t_0).$$

It is easy to see that they might have different means, but the covariance is given by $\sigma^2 t_0$, independent of s.

(2) similar to (1). The demean portion of $f(t_0,T_0)$ and $f(t_1,T_1)$ are given by $\sigma W(t_0)$ and $\sigma W(t_1)$.

□

Remark 19.9.3 (slow decaying correlation and non-stationary covariance)**.**

- Note that the time series $G(s) = f(t_0+s, T_0+s)$ has a slowly decaying correlation, whereas in Hull white model, the correlation is exponentially decaying.
- The time series $G(s) = f(t_0+s, T_0+s)$ has a non-stationary covariance structure. For example, $Cov(G(s_1), G(0_2))$ is not solely dependent on the difference $s_1 - s_2$.

Lemma 19.9.5 (Reduce to Hull-White model). *[1, p. 71] Take $\sigma(t,T) = \sigma(t)e^{-\alpha(T-t)}$, where $\sigma(t)$ is a deterministic function and α is a constant.*

Then, by Theorem 19.9.1, we have the drift term given by

$$\begin{aligned}\alpha(t,T) &= \sigma(t,T)\int_t^T \sigma(t,u)du \\ &= \sigma^2(t)e^{-\alpha(T-t)}D(t,T),\end{aligned}$$

where $D(t,T)$ *is given by* $D(t,T) \triangleq \int_t^T e^{-\alpha(s-t)}ds$

It follows that

- *The forward curve is*

$$\begin{aligned}f(t,T) &= f(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T)dW(s) \\ &= f(0,T) + \int_0^t \sigma^2(s)e^{-\alpha(T-s)}D(s,T)ds + \int_0^t \sigma(s)e^{-\alpha(T-s)}dW(s).\end{aligned}$$

- *The short rate is*

$$r(t) = f(t,t) = f(0,t) + \int_0^t \sigma^2(s)e^{-\alpha(t-s)}D(s,T)ds + \int_0^t \sigma(s)e^{-\alpha(t-s)}dW(s).$$

- *The short rate model(matching the current term structure)satisfies SDE*

$$dr(t) = (\theta(t) - \alpha r(t))dt + \sigma(t)dW(t),$$

with

$$\theta(t) = \frac{\partial f(0,t)}{\partial t} + \alpha f(0,t) + \int_0^t \sigma(s)^2 e^{-2\alpha(t-s)}ds.$$

Proof. (1)(2) are straight forward. (3) Note that

$$\begin{aligned}dr(t) &= \frac{\partial f(0,t)}{\partial t}dt + \int_0^t \sigma^2(s)\frac{d}{dt}(e^{-\alpha(t-s)D(s,t)})ds + \sigma(t)dW(t) + \int_0^t \sigma\frac{d}{dt}(e^{-\alpha(t,s)})dW(s)) \\ &= \frac{\partial f(0,t)}{\partial t}dt + \sigma(t)dW(t) - \alpha\int_0^t \sigma(s)e^{-\alpha(T-s)}dW(s) \\ &+ \int_0^t \sigma^2(s)(-\alpha e^{-\alpha(t-s)}D(s,t) + e^{-2\alpha(t-s)})ds\end{aligned}$$

where we use

$$\frac{d}{dt}D(s,t) = e^{-\alpha(t-s)}.$$

Further we use the relation

$$\alpha f(0,t) - \alpha r(t) = -\alpha\int_0^t \sigma^2(s)e^{-\alpha(t-s)}D(s,T)ds - \alpha\int_0^t \sigma(s)e^{-\alpha(t-s)}dW(s)$$

to get to the result. □

Lemma 19.9.6 (covariance in parallel shift forward rate dynamics). *Consider the Hull-White model for forward rate, which is given by*

$$f(t,T) = f(0,T) + \int_0^t \sigma^2(s)e^{-\alpha(T-s)}D(s,T)ds + \int_0^t \sigma(s)e^{-\alpha(T-s)}dW(s).$$

Then for given t_0 and T_0,

- *$f(t_0,T_0)$ and $f(t_0+s,T_0+s), s \geq 0$ have the covariance given by*

$$\exp(-\alpha s)\int_0^{t_0} \sigma^2(u)\exp(-2\alpha(T_0-u)du),$$

  even though they might have different means.
- *$f(t_0,T_0)$ and $f(t_0+s,T_0+s), s \geq 0$ have the correlation given by*

$$\rho = \frac{t_0}{\sqrt{t_0(t_0+s)}}$$

- *$f(t_0,T_0)$ and $f(t_1,T_1)$ have the covariance given by $\sigma^2 \min(t_0,t_1)$, even though they might have different means.*

Proof. (1)(2)We can show that

$$f(t_0,T_0) = f(0,T_0) + \frac{1}{2}\sigma^2 t_0(2T_0 - t_0) + \sigma W(t_0).$$

and

$$f(t_0+s,T_0+s) = f(0,T_0) + \frac{1}{2}\sigma^2(t_0+s)(2T_0 - t_0 + s) + \sigma W(t_0).$$

It is easy to see that they might have different means, but the covariance is given by $\sigma^2 t_0$, independent of s.

$$\begin{aligned}
\rho &= \frac{\exp(-\alpha s)\int_0^{t_0}\sigma^2(u)\exp(-2\alpha(T_0-u)du)}{\sqrt{\int_0^{t_0}\sigma^2(u)\exp(-2\alpha(T_0-u)du)\int_0^{t_0+s}\sigma^2(u)\exp(-2\alpha(T_0+s-u)du)}} \\
&= \frac{\exp(-\alpha s)\int_0^{t_0}\sigma^2(u)\exp(-2\alpha(T_0-u)du)}{\sqrt{\int_0^{t_0}\sigma^2(u)\exp(-2\alpha(T_0-u)du)\exp(-2\alpha s)\int_0^{t_0+s}\sigma^2(u)\exp(-2\alpha(T_0-u)du)}} \\
&= \frac{1}{1+} \\
&= \frac{1}{\sqrt{\int_{t_0}^{t_0+s}\sigma^2(u)\exp(-2\alpha(T_0-u)du)}} \\
&= \frac{1}{\sqrt{\int_0^{s}\sigma^2(u)\exp(-2\alpha(T_0-t_0-u)du)}} \\
&= \frac{1}{\sqrt{\exp(-2\alpha(T_0-t_0))\int_0^{s}\sigma^2(u)\exp(2\alpha u)du)}}
\end{aligned}$$

(3) similar to (1). The demean portion of $f(t_0, T_0)$ and $f(t_1, T_1)$ are given by $\sigma W(t_0)$ and $\sigma W(t_1)$.

□

Remark 19.9.4 (slow decaying correlation and stationary correlation, non-stationary covariance).

- Note that the time series $G(s) = f(t_0 + s, T_0 + s)$ has a slowly decaying correlation, whereas in Hull white model, the correlation is exponentially decaying.
- The time series $G(s) = f(t_0 + s, T_0 + s)$ has a stationary covariance structure. Note that $Cov(G(s_1), G(0_2))$ is not solely dependent on the difference $s_1 - s_2$.

19.9.1.3 *Option pricing in Gaussian HJM models*

Lemma 19.9.7 (Pricing call option on zero coupon bond). *Given the volatility structure $\sigma(t,u)$ in HJM model, the zero coupon bond dynamics under risk-neutral measure Q is given by*

$$dP(t,T) = r(t)P(t,T)dt + \Sigma(t,T)P(t,T)dW_t,$$

where $\Sigma(t,T) = -\int_t^T \sigma(t,u)du$.

The value at time t of a European call option maturing at S with strike K on a bond maturing at $T > S$ is given by

$$V(t) \triangleq BC(t; S, T, K) = P(t,T)N(d_+) - KP(t,S)N(d_-),$$

where

$$d_+ = \frac{\log \frac{FP(t;S,T)}{K} + \frac{1}{2}v(t,S)^2}{v(t,S)} = \frac{\log \frac{P(t,T)}{KP(t,S)} + \frac{1}{2}v(t,S)^2}{v(t,S)}, d_- = d_+ - v(t,S).$$

and

$$v(t,S)^2 = \int_t^S (\Sigma(u,T) - \Sigma(u,S))^2 dt$$

Proof. See Theorem 19.9.1 and Lemma 19.3.10. □

19.9.2 Multi-factor HJM framework

Theorem 19.9.2. *[1, p. 69] Assume that the forward rate $f(t,T)$, under risk-neutral measure Q,is given*

$$df(t,T) = \alpha(t,T)dt + \sum_{i=1}^{n} \sigma_i(t,T)dW_i(t),$$

where $W_1(t), ..., W_n$ are independent Brownian motions, and $\sigma_1(t,T), ..., \sigma_n(t,T)$ are adapted to filtration $\mathcal{F}_t$ for each $T > 0$, and

$$\int_0^T \int_0^T |\alpha(s,u)|\, dsdu < \infty, \int_0^T (\int_0^T |\sigma_i(s,u)|^2 ds)^{1/2} du < \infty, \forall i$$

almost surely. Then,

$$\alpha(t,T) = \sum_{i=1}^{n} (\sigma_i(t,T) \int_t^T \sigma_i(t,u))du.$$

As a consequence,

$$f(t,T) = f(0,T) + \int_0^t \sum_{i=1}^{n} (\alpha_i(s,T)(\int_s^T \sigma_i(s,u)du))dt + \int_0^t \sum_{i=1}^{n} \sigma_i(s,T)dW_i(s).$$

Furthermore, the zero-coupon bond price $P(t,T)$ satisfies SDE

$$dP(t,T) = r(t)P(t,T)dt + \sum_{i=1}^{n} \Sigma_i(t,T)P(t,T)dW_i(t)$$

with log-volatility

$$\Sigma(t,T) = -\int_t^T \sigma_i(t,u)du.$$

Proof. We use the notation $\sigma = (\sigma_1, \sigma_2, ..., \sigma_n)^T, dW_t = (dW_1, dW_2, ..., dW_n)^T$. Under risk-neutral measure Q, the bond price is related to forward rate as

$$P(t,T) = \exp(-\int_t^T f(t,u)du).$$

Use

$$\begin{aligned} d(\int_t^T f(t,u)du) &= -f(t,t)dt + \int_t^T (\alpha(t,u)dt + \sigma^T dW(t))du \\ &= -r(t)dt + (\int_t^T \alpha(t,u)du)dt + (\int_t^T \sigma^T du)dW(t) \end{aligned}$$

Then

$$\begin{aligned} dP(t,T) &= d(\exp(-\int_t^T f(t,u)du)) \\ &= P(t,T)d(\int_t^T f(t,u)du) \\ &= (r(t) - \int_t^T \alpha(t,u)du)du)^2)P(t,T)dt - (\int_t^T \sigma^T(t,u)du)P(t,T)dW(t) \\ &+ \frac{1}{2}(\int_t^T \sigma^T(t,u)du)dW(t)[(\int_t^T \sigma^T(t,u)du)dW(t)]^T \\ &= (r(t) - \int_t^T \alpha(t,u)du)du)^2)P(t,T)dt - (\int_t^T \sigma^T(t,u)du)P(t,T)dW(t) \\ &+ \frac{1}{2}(\int_t^T \sigma^T(t,u)du)(\int_t^T \sigma(t,u)du)dt \end{aligned}$$

Because under risk-neutral measure Q, $P(t,T)$ will have drift $r(t)$(Lemma 19.5.3). Therefore,

$$r(t) - \int_t^T \alpha(t,u)du + \frac{1}{2}\left\|(\int_t^T \sigma(t,u)du)\right\|^2 = r(t),$$

Then
$$\int_t^T \alpha(t,u)du = \frac{1}{2}\left\|(\int_t^T \sigma(t,u)du)\right\|^2.$$
Differentiating both sides with respect to T, we have
$$\alpha(t,T) = \sigma(t,T)^T \int_t^T \sigma(t,u)du.$$
□

Lemma 19.9.8 (correlation structure in one-factor HJM forward model). *Let $T_1, T_2, ..., T_N$ be two maturity dates. Then*

- *The pair correlation is given by*
$$Cov[f(t,T_1), f(t,T_2)] = \sum_{i=1}^{n} \int_0^t \sigma_i(s,T_1)\sigma_i(s,T_2)ds.$$
- *The instantaneous correlation is given by*
$$Cov[df(t,T_1), df(t,T_2)] = \sum_{i=1}^{n} \sigma_i(t,T_1)\sigma_i(t,T_2)dt.$$

Proof. Using Ito isometry. □

Note 19.9.2 (multifactor model as a reduced form model). Note that we can also write the multifactor model by
$$df(t,T) + \alpha(t,T)dt + \sigma(t,T)dZ^T(t),$$
where we have a set(infinitely many) of Brownian motions $Z^T(t)$ indexed by the maturity T. By requiring $f(t,T)$ is driven by n factors, we are requiring
$$Rank(dZ^T[dZ^T]^T) = n.$$

19.9.3 Forward rate model calibration

Lemma 19.9.9 (calibrating a linear interpolating forward curve). *[4, p. 96] Suppose we are given by M yield, denoted by $R(0,T_1), R(0,T_2), ..., R(0,T_M)$, for M maturity dates, where $0 < T_1 < ... < T_M$.*

Further assume the forward curve is taking the following function form:

$$f(0,T) = R(0,T_1), \forall 0 \leq T < T_1$$
$$f(0,T) = f(0,T_{i-1}) + \alpha_i(T - T_{i-1}), \forall T_{i-1} \leq T < T_i, \forall i = 2,...,M$$

The calibrating strategy is:

- *Calculating zero-coupon price $P(t,T_i), \forall i = 1,2,...,M$.*
- *Solve $\alpha_i, i = 2,3,...,M$ from the following equation:*

$$\begin{aligned} P(0,T_i) &= P(0,T_{i-1}) \exp(-\int_{T_{i-1}}^{T_i} f(0,s)ds) \\ &= P(0,T_{i-1}) \exp(-\int_{T_{i-1}}^{T_i} f(0,T_{i-1}) + \alpha_i(s - T_{i-1})ds) \\ &= P(0,T_{i-1}) \exp(-f(0,T_{i-1})(T_i - T_{i-1}) + \frac{1}{2}\alpha_i(T_i - T_{i-1})^2) \end{aligned}$$

Eventually, we have a linear interpolating $f(0,T), 0 \leq T \leq T_M$.

Remark 19.9.5 (solving non-smoothness issue). With the calculate $T_i, f(0,T_i)$ pairs, we can use cubic spline interpolation (subsection A.16.1)to remove non-smoothness.

19.9.4 Connection to short-rate model

19.9.4.1 *From short-rate model to forward-rate model*

Theorem 19.9.3. *[4, p. 139] Consider an arbitrage-free short-rate model given by*

$$dr_t = a(t,r_t)dt + b(t,r_t)dW_t.$$

Then it can be converted an HJM model with forward-rate volatility given by

$$\sigma(t,T) = \rho(r_t,t)\frac{\partial^2 g}{\partial x \partial T}(r_t,t,T)$$

where

$$g(x,t,T) = -\ln P(t,T) = -\ln E_Q[\exp(-\int_t^T r_s ds)|r_t = x],$$

Proof. Note that
$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T} = \frac{\partial g}{\partial T}(r_t,t,T).$$

Hence, the process for the instantaneous forward rate is

$$\begin{aligned} f(t,T) &= \frac{\partial f}{\partial x}dr_t + (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2})dt \\ &= \rho\frac{\partial^2 g}{\partial x \partial T}dW_t + \ drift\ term \end{aligned}$$

□

Example 19.9.2 (Hull-White model to HJM model). Consider the Hull-White model given by
$$dr_t = \kappa(\theta_t - r_t)dt + \sigma dW_t.$$
Then
$$g(r_t,t,T) = -\ln P(t,T) = B(t,T)r_t - \ln A(t,T),$$
where
$$B(t,T) = \frac{1-\exp(-\kappa(T-t))}{\kappa}.$$
Then

$$\begin{aligned} \frac{\partial^2 g}{\partial r_t \partial T}(r_t,t,T) &= \frac{\partial B(t,T)}{\partial T} = \exp(-\kappa(T-t)) \\ \sigma(t,T) &= \sigma\frac{\partial^2 g}{\partial x \partial T}(r_t,t,T) = \sigma\exp(-\kappa(T-t)) \end{aligned}$$

Example 19.9.3 (Ho-Lee model). The Ho-Lee model for short rate, under the risk-neutral measure Q, is given by

$$dr = \theta(t)dt + \sigma dW_t$$

where W_t is the Brownian motion, and σ is a constant.

19.9.4.2 *Criterion for a Markovian model*

Note 19.9.3. [4, p. 138] Note that the short-rate model is connected to the forward-rate model via
$$dr_t = \frac{\partial f(t,T)}{\partial t}|_{T=t}dt + \sigma^T(t,t)dW_t,$$

where $\sigma(t,t) \in \mathbb{R}^n$ and W_t is n-dimensional Brownian motion, and

$$\frac{\partial f(t,T)}{\partial t}|_{T=t}$$

Theorem 19.9.4. *[4, p. 139] Suppose that the forward-rate volatility satisfies*

$$\frac{\partial \sigma_i(t,T)}{\partial T} = -\kappa_i(T)\sigma_i(t,T), i = 1,2,...,n, (*)$$

for some deterministic function $\kappa_i(T)$. Then

$$\begin{aligned} d\psi_i(t) &= (\sigma_i^2(t,t) - 2\kappa_i(t)\psi_i(t))dt \\ d\chi_i(t) &= (\psi_i(t) - \kappa_i(t)\chi_i(t))dt + \sigma_i(t,t)dW_i(t), \end{aligned}$$

for $i = 1,...,n$.

Proof. For $\psi_i(t)$, there is

$$\begin{aligned} d\psi_i(t) &= \sigma_i^2(t,t)dt + (\int_0^t 2\sigma_i(s,t))dt \\ &= \sigma_i^2(t,t)dt - 2\kappa_i(t)(\int_0^t \sigma_i^2(s,t)ds)dt \\ &= (\sigma_i^2(t,t) - 2\kappa_i(t)\psi_i(t))dt, \end{aligned}$$

while for $\chi_i(t)$, we have noticing that $b_i(t,t) = 0$,

$$d\chi_i(t) = (\int_0^t \frac{\partial b_i(s,t)}{\partial t}ds + \frac{\sigma_i(s,t)}{\partial t}dW_i(s)) + \sigma_i(t,t)dW_i(t) \ (**).$$

Because of equation $(*)$ and

$$\begin{aligned} \frac{\partial b_i(s,t)}{\partial t} &= \frac{\sigma_i(s,t)}{\partial t}\int_s^t \sigma_i(s,u)du + \sigma_i^2(s,t) \\ &= -\kappa_i(t)b_i(s,t) + \sigma_i^2(s,t) \end{aligned}$$

we can write Equation $(**)$ as

$$\begin{aligned} d\chi_i(t) = &-\kappa_i(t)(\int_0^t b_i(s,t)ds + \sigma_i(s,t)dW_i(s))dt \\ &+ (\int_0^t \sigma_i^2(s,t)ds)dt + \sigma_i(t,t)dW_i(t) \\ &= (\psi_i(t) - \kappa_i(t)\chi_i(t))dt + \sigma_i(t,t)dW_i(t). \end{aligned}$$

□

Remark 19.9.6 (interpretation).

- The connection is established by the price of the zero-coupon bonds resulted from the two types of models.
- An arbitrage-free HJM model is completed by determined by its forward volatility(Theorem 19.9.1).

Theorem 19.9.5. *[4, p. 142] Suppose that the short-rate volatility, $\sigma(t,t)$, is a deterministic function of time. Then a necessary condition for the short rate to be Markovian is*

$$\frac{\sigma_i(t,T)}{\partial T} = -\kappa_i(T)\sigma_i(t,T),$$

for some scalar function, $\kappa_i(T), 1 \leq i \leq n$.

Proof. For the short rate to be Markovian, we require that $r_T - r_t$ depends only on r_t and $\{dW_s, s \in (t,T), dW_s \in \mathbb{R}^n\}$. In fact, we have

$$\begin{aligned} r_T - r_t &= f(0,T) - f(0,t) + \int_0^T \sigma^T(s,T)\Sigma(s,T)ds \\ &- \int_0^t \sigma^T(s,t)\Sigma(s,t)ds + \int_0^T \sigma^T(s,T)dW_s - \int_0^t \sigma^T(s,t)dW_s \\ &= f(0,T) - f(0,t) + \int_0^T \sigma^T(s,T)\Sigma(s,T)ds \\ &- \int_0^t \sigma^T(s,t)\Sigma(s,t)ds + \int_t^T \sigma^T(s,T)dW_s + \int_0^t (\sigma(s,T) - \sigma(s,t))^T dW_s \end{aligned}$$

The last term in the above equation cannot depend on $\{dW_s, s \in (t,T), dW_s \in \mathbb{R}^n\}$, so it can depend only on r_t. Because

$$\int_0^t \sigma^T(s,t)dW_s = r_t + deterministic\ functions,$$

we conclude that

$$\int_0^t \sigma^T(s,t)dW_s$$

is also a deterministic function of r_t. Hence, we have correlation relationship

$$Corr[\int_0^t \sigma^T(s,t)dW_s, \int_0^t \sigma^T(s,t)dW_s] = 1.$$

The last equality can be rewritten into

$$E_Q[\int_0^t \sigma^T(s,T)dW_s \times \int_0^t \sigma^T(s,t)dW_s] = (E_Q[(\int_0^t \sigma^T(s,T)dW_s)^2])^{1/2}(E_Q[(\int_0^t \sigma^T(s,t)dW_s)^2])^{1/2}.$$

By Ito isometry, the above implies that

$$\left|\int_0^t \sigma^T(s,T)\sigma(s,t)ds\right| = (\int_0^t \|\sigma(s,T)\|^2 ds)^{1/2}(\int_0^t \|\sigma(s,t)\|^2 ds)^{1/2},$$

that is, the equality is achieved in the Cauchy equality, and the equality holds if and only if

$$\sigma(s,t) = \alpha(t,T)\sigma(s,T), 0 \leq s \leq t,$$

for some deterministic scalar function α.

Similarly, we also have

$$\sigma(s,t) = \alpha(t,T')\sigma(s,T'), 0 \leq s \leq t,$$

for some other T'. Assume that $\sigma_i(s,t) \neq 0$, we then have

$$\frac{\sigma_i(s,T)}{\sigma_i(s,T')} = \frac{\alpha(t,T')}{\alpha(t,T)} = \frac{\alpha(0,T')}{\alpha(0,T)}, i = 1,2,...,n.$$

Making $T' = s$, we have then proved that $\sigma_i(s,T)$ can be factorized as

$$\sigma_i(s,T) = x_i(s)y_i(T).$$

By differentiating the above equation with respect to T, we obtain

$$\frac{\partial \sigma_i(s,T)}{\partial T} = x_i(s)y_i(T)\frac{\partial \ln y_i(T)}{\partial T}.$$

Denote $\frac{\partial \ln y_i(T)}{\partial T}$ by $-\kappa_i(T)$, we complete the proof. □

Theorem 19.9.6 (zero coupon bond price). *[4, p. 140]*

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp(-\int_t^T (\sum_{i=1}^n \int_0^t b_i(s,u)ds + \sigma_i(s,u)dW_i(s))du)$$

Proof. (1) For the first term in the exponent, we have

□

19.10 LIBOR market models

The LIBOR market model, also known as the BGM Model (Brace Gatarek Musiela Model, in reference to the names of some of the inventors) is a financial model of interest rates.[1] It is used for pricing interest rate derivatives, especially exotic derivatives like Bermudan swaptions, ratchet caps and floors, target redemption notes, autocaps, zero coupon swaptions, constant maturity swaps and spread options, among many others.

19.10.1 LIBOR market model

19.10.1.1 *The model*

Definition 19.10.1 (LIBOR market model). *[1, p. 84] Consider a set of dates $0 \leq T_0 < T_1 < \cdots < T_n$ with accrual periods $\tau_i = T_i - T_{i-1}$. Denote the forward rate*

$$F_i(t) \triangleq F(t; T_{i-1}, T_i) = \frac{1}{\tau_i}(\frac{P(t, T_{i-1})}{P(t, T_i)} - 1).$$

In the LIBOR model, we assume the forward rate $F_i(t)$ under the forward measure of T_j-maturity zero-coupon bond are given by

$$dF_i(t) = \mu_i^j(F_i(t), t)dt + C_i(F_i(t), t)dZ_i^j(t),$$

*where $C_i(F_i(t), t)$ is **the instantaneous volatilities** of the forward rates $F_i(t)$ for each $i = 1, 2, ..., n$, $dZ_i dZ_j = \rho_{i,j} dt$ and Z_i^j is a Brownian motion under forward measure Q_{T_j}.*

Note 19.10.1 (relation to HJM framework)**.**

- The dynamics of $F(t; T_1, T_2)$ is the **approximate** average of $f(t, T)$ over period of T_1 and T_2.

$$\begin{aligned} F(t; T_1, T_2) &= \frac{1}{T_2 - T_1}(\frac{P(t, T_1)}{P(t, T_2)} - 1) \\ &= \frac{1}{T_2 - T_1}(\frac{\exp(-\int_t^{T_1} f(t, s)ds)}{\exp(-\int_t^{T_2} f(t, s)ds)} - 1) \\ &= \frac{1}{T_2 - T_1}(\exp(\int_{T_1}^{T_2} f(t, s)ds) - 1) \\ &\approx \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} f(t, s)ds \end{aligned}$$

- If $T_2 \to T_1$, then $F(t; T_1, T_2) = f(t, T_1)$.

Remark 19.10.1 (choice of volatility structure). The volatility function $C_j(F_j(t), t)$ can take different forms. For example,

- (normal model)$C_j(F_j(t), t) = \sigma_j(t)$.
- (CEV model)$C_j(F_j(t), t) = \sigma_j(t)$.
- (lognormal model)$C_j(F_j(t), t) = \sigma_j(t)$.
- (shifted normal model)$C_j(F_j(t), t) = \sigma_j(t)$.

where the function $\sigma_j(t)$ is deterministic, $0 \leq \beta_j \leq 1, \delta_j \geq 0$.

19.10.1.2 *Drifts under different measures*

Lemma 19.10.1 (drifts under different forward measure). *[1, p. 89] The drift terms under different Q_{T_i} is given by*

$$\mu_i^j = \begin{cases} -\sum_{k=i+1}^{j} \frac{\tau_k \rho_{k,i} C_i(t) C_k(t) F_i(t) F_k(t)}{1+\tau_k F_k(t)}, & j > i \\ 0, & i{=}j \\ \sum_{k=j+1}^{i} \frac{\tau_k \rho_{k,i} C_i(t) C_k(t) F_i(t) F_k(t)}{1+\tau_k F_k(t)}, & i > j \end{cases}$$

In particular, if $C_i(F_i(t), t) = \sigma_i(t) F_i(t)$, we have The drift $\mu_i^j(t)$ of the forward rate $F_i(t)$ are given by

$$\mu_i^j = \begin{cases} -\sum_{k=i+1}^{j} \frac{\tau_k \rho_{k,i} \sigma_i(t) \sigma_k(t)}{1+\tau_k F_k(t)}, & j > i \\ 0, & i{=}j \\ \sum_{k=j+1}^{i} \frac{\tau_k \rho_{k,i} \sigma_i(t) \sigma_k(t)}{1+\tau_k F_k(t)}, & i > j \end{cases}$$

Proof. The drift μ_i under Q_{T_i} is zero since F_i^i is a martingale. μ_i under measure $Q_{T_j}, j > i$ is given by(Theorem 15.6.17)

$$\begin{aligned}
\mu_i^j dt &= dF_i d\log(\frac{P(t,T_j)}{P(t,T_i)}) \\
&= -dF_i d\log(\frac{P(t,T_i)}{P(t,T_j)}) \\
&= -dF_i d\log(\prod_{k=i+1}^{j}(1+\tau_k F_k)) \\
&= -dF_i \sum_{k=i+1}^{j} \frac{dF_k}{1+\tau_k F_k} \\
&= -\sum_{k=i+1}^{j} \frac{\tau_k \rho_{k,i} C_i(t) C_k(t)}{1+\tau_k F_k(t)} dt
\end{aligned}$$

where we use the relation

$$\frac{P(t,T_i)}{P(t,T_j)} = \prod_{k=i+1}^{j}(1+\delta_k F_k).$$

□

Lemma 19.10.2 (drifts under terminal measure). *[1, p. 90] The measure associated with the zero-coupon bond $P(t,T_n)$ maturing at time T_n is called **terminal measure**. Under terminal measure, the forward rate dynamics are given by*

$$dF_i(t) = \mu_i^n(t)dt + C_i(F_i(t),t)dZ_i^n(t), \forall i,$$

where

$$\mu_i^n = -\sum_{k=i+1}^{n} \frac{\tau_k \rho_{k,i} C_i(t) C_k(t)}{1+\tau_k F_k(t)},$$

and $Z_1^n(t),...,Z_n^n(t)$ are correlated Brownian motions under measure Q_{T_n} such that $dZ_i^n(t)dZ_j^n = \rho_{i,j}(t)$.

In particular, if $C_i(F_i(t),t) = \sigma_i(t)F_i(t)$, we have

$$dF_i(t) = \mu_i^n(t)dt + \sigma_i(t)F_i(t)dZ_i^n(t), \forall i,$$

where

$$\mu_i^n = -\sum_{k=i+1}^{n} \frac{\tau_k \rho_{k,i} \sigma_i(t)\sigma_k(t) F_i(t) F_k(t)}{1+\tau_k F_k(t)}.$$

Proof. Directly from Lemma 19.10.1. □

Definition 19.10.2 (discrete money market account, spot measure). *For any $t \in [0, T_n]$, let $\alpha(t)$ denote the index of the next reset date at time t, that is $\alpha(t) = \min\{j : t \leq T_j, j = 0, ..., n\}$. The **discrete money market account** is defined as*

$$L(t) = P(t, T_{\alpha(t)}) \prod_{k=0}^{\alpha(t)} \frac{1}{P(T_{k-1}, T_k)}.$$

*The measure associated with discrete money market account numeraire is called **spot measure**.*

Remark 19.10.2 (intuition before discrete money account)**.**

- At time 0, we begin by investing one dollar to buy an amount $P(0, T_0)^{-1}$ of the T_0 bonds.
- At time T_0, the bonds will be worth $P(0, T_0)^{-1}$, which we reinvest to buy an amount $P(0, T_0)^{-1} P(T_0, T_1)^{-1}$ of T_1 bonds.
- We continue in this way, reinvesting all the proceeds at each date T_{i-1} into zero-coupon bonds maturing at the next date T_i, for $i = 1, 2, ..., n$.

Lemma 19.10.3 (drifts under spot measure). *[1, p. 90] The measure associated with the zero-coupon bond $P(t, T_n)$ maturing at time T_n is called **terminal measure**. Under terminal measure, the forward rate dynamics are given by*

$$dF_i(t) = \mu_i^n(t) F_i(t) dt + C_i(F_i(t), t) dZ_i^n(t), \forall i,$$

where

$$\mu_i^L(t) = \sum_{k=\alpha(t)+1}^{i} \frac{\tau_k \rho_{k,i} C_i(t) C_k(t)}{1 + \tau_k F_k(t)},$$

and $Z_1^L(t), ..., Z_n^L(t)$ are correlated Brownian motions under measure Q_{T_n} such that $dZ_i^L(t) dZ_j^L = \rho_{i,j}(t)$.

In particular, if $C_i(F_i(t), t) = \sigma_i(t) F_i(t)$, we have

$$dF_i(t) = \mu_i^L(t) dt + \sigma_i(t) F_i(t) dZ_i^n(t), \forall i,$$

where

$$\mu_i^L = \sum_{k=\alpha(t)+1}^{i} \frac{\tau_k \rho_{k,i} \sigma_i(t) \sigma_k(t) F_i(t) F_k(t)}{1 + \tau_k F_k(t)}.$$

Proof. The drift μ_i under Q_{T_i} is zero since F_i^i is a martingale. μ_i under measure $Q_{T_j}, j > i$ is given by(Theorem 15.6.17)

$$\begin{aligned}
\mu_i^j dt &= dF_i d\log(\frac{L(t)}{P(t,T_i)}) \\
&= dF_i d\log(\frac{P(t,T_{\alpha(t)})}{P(t,T_i)}) \\
&= dF_i d\log(\prod_{k=\alpha(t)+1}^{i}(1+\tau_k F_k)) \\
&= dF_i d\log(\prod_{k=\alpha(t)+1}^{i}(1+\tau_k F_k)\prod_{k=0}^{\alpha(t)}(1+\tau_k L(T_{k-1},T_k))) \\
&= dF_i d\log(\prod_{k=\alpha(t)+1}^{i}(1+\tau_k F_k)) \\
&= dF_i \sum_{k=\alpha(t)+1}^{i}\frac{dF_k}{1+\tau_k F_k} \\
&= \sum_{k=\alpha(t)+1}^{i}\frac{\tau_k\rho_{k,i}C_i(t)C_k(t)}{1+\tau_k F_k(t)}dt
\end{aligned}$$

where we use the relation

$$\frac{P(t,T_i)}{P(t,T_j)} = \prod_{k=i+1}^{j}(1+\delta_k F_k),$$

and $L(T_{k-1},T_k)$ is known at time t.

□

19.10.1.3 *Pricing applications*

Lemma 19.10.4 (zero-coupon bond price). *[1, p. 125] Let $t < T_0 < T_1 < \cdots < T_n$ be a set of dates. Then*

$$\begin{aligned}
P(T_0,T_n) &= FP(T_0;T_0,T_1)FP(T_0;T_1,T_2)\cdots FP(T_0;T_{n-1},T_n) \\
&= \prod_{i=1}^{n}\frac{1}{1+(T_i-T_{i-1})F(T_0,T_{i-1},T_i)}
\end{aligned}$$

That is, at time t, the random variable $P(T_0,T_n)$ can be expressed as a product of functions involving random variables $F(T_0;T_{i-1},T_i)$.

Proof. See Lemma 19.1.2 □

Remark 19.10.3 (evaluation using Monte Carlo simulation). Suppose we want to generate a random sample of $P(T_0, T_n)$, we can obtain this sample by generating random samples of $F(T_0, T_{i-1}, T_i)$ by simulating forward rate dynamics.

Lemma 19.10.5 (Black's formula for caplet in the LIBOR market model). *Consider a caplet having payoff at T_i given by*

$$\tau_i(L(T_{i-1}, T_i) - K)^+ = \tau_i(F(T_{i-1}; T_{i-1}, T_i) - K)^+,$$

where $\tau_i = T_i - T_{i-1}$.

The price of this caplet at time $t < T_{i-1}$ is given by

$$Cpl_i(t) = \tau_i P(t, T_i)(F(t; T_{i-1}, T_i)N(d_+) - KN(d_-)),$$

where

$$d_\pm = \frac{\ln F(t; T_{i-1}, T_i)/K \pm \frac{1}{2}\overline{\sigma}^2}{\overline{\sigma}}, \overline{\sigma}^2 = \int_t^{T_{i-1}} \sigma_i(s)^2 ds$$

Proof. Under the forward measure Q_{T_i},

$$\begin{aligned} Cpl_i(t) &= P(t, T_i)\tau_i E_{T_i}[F_i(T_{i-1}) - K)^+|\mathcal{F}_t] \\ &= P(t, T_i)E_{T_i}[F_i(T_{i-1})\mathbf{1}_{F_i(T_{i-1})>K}|\mathcal{F}_t] - P(t, T_i)KE_{T_i}[\mathbf{1}_{F_i(T_{i-1})}|\mathcal{F}_t] \end{aligned}$$

In the LIBOR market model, under the forward measure Q_{T_i}, we have

$$dF_i(t) = \sigma_i(t)F_i dW_t,$$

therefore

$$F_i(T_{i-1}) = F_i(t)\exp\left(\int_t^{T_{i-1}} \sigma_i(s)dW_s - \frac{1}{2}\int_t^{T_{i-1}} \sigma_i^2(s)ds\right).$$

The rest is just routine. □

Remark 19.10.4 (caplet/cap pricing does not depend on the correlation of different forward rates). [3, p. 221] It can be seen that the prices of caplets/caps do not depend on the correlation matrix of the Brownian motions.Therefore, caplet/caps cannot be used to calibrate correlation.

Lemma 19.10.6 (pricing arbitrary interest rate European derivatives). *Consider a European derivative security with expiry date T_j and payoff of the form $h(F_1(T_j), F_2(T_j), ..., F_n(T_j))$, where*

- $F_i(T_j) = F(T_j; T_{i-1}, T_i), \forall i = j+1, ..., n.$

- $F_i(T_j) = F(T_{i-1}; T_{i-1}, T_i) = L(T_{i-1}, T_i), \forall i = 1, ..., j.$

The price at time 0 of this option is given by

$$P(0, T_n)E_{P_{T_n}}[\frac{h(F_1(T_j), ..., F_n(T_j))}{P(T_j, T_n)}]$$

Proof. Use change of numeraire technique (**??**), we have

$$\frac{dP_{T_j}}{dP_{T_n}} = \frac{P(T_j, T_j)}{P(0, T_j)} \frac{P(0, T_n)}{P(T_j, T_n)}.$$

Then,

$$\begin{aligned}
&P(0, T_i)E_{P_{T_i}}[h(F_1(T_j), ..., F_n(T_j))] \\
&= P(0, T_i)E_{P_{T_n}}[h(F_1(T_j), ..., F_n(T_j))\frac{dP_{T_i}}{dP_{T_n}}] \\
&= P(t, T_i)E_{P_{T_n}}[h(F_1(T_j), ..., F_n(T_j))\frac{P(T_j, T_j)}{P(0, T_j)} \frac{P(0, T_n)}{P(T_j, T_n)}] \\
&= P(0, T_n)E_{P_{T_n}}[\frac{h(F_1(T_j), ..., F_n(T_j))}{P(T_j, T_n)}]
\end{aligned}$$

□

19.10.2 LIBOR swap market model

19.10.2.1 *The model*

Lemma 19.10.7 (forward swap rate dynamics under forward swap measure). *Assume* $P(t, T)$ *obeys the SDE*

$$dP(t, T)/P(t, T) = (r(t) + \lambda\sigma(t, T))dt + P(t, T)\sigma(t, T)dW(t)$$

where $W(t)$ *is the standard Brownian motion under the real world probability measure, and* $r(t)$ *is the instantaneous short rate.*

Given a set of dates $T_0 < T_1 < ... < T_n$ *with accrual periods* $\tau_i = T_i - T_{i-1}$*. The* ***swap annuity*** *is defined as*

$$A_{0,n}(t) = \sum_{i=1}^{n} \tau_i P(t, T_i).$$

*The **forward swap rate**, denote by $S_{0,n}(t)$, is defined by,*

$$S_{0,n}(t) = \frac{P(t,T_0) - P(t,T_n)}{A_{0,n}(t)}.$$

Then,

- *The swap annuity $A_{0,n}(t)$ has SDE given by*

$$dA_{0,n}(t)/A_{0,n}(t) = (r + \lambda\sigma_A)dt + \sigma_A dW_t.$$

- *The quantity $P(t,T_0) - P(t,T_n)$ has SDE given by*

$$d(P(t,T_0) - P(t,T_n))/(P(t,T_0) - P(t,T_n)) = (r + \lambda\sigma_P)dt + \sigma_P dW_t.$$

- *Under forward measure Q_A with respect to $A_{0,n}(t)$(i.e. set $\lambda = \sigma_A$), the dynamics of $S_{0,n}(t)$ is given by*

$$dS_{0,n}(t) = S_{0,n}(t)(\sigma_P - \sigma_A)dW^S(t).$$

where W^S is a Brownian motion under the measure Q_A.

Proof. (1)(2)Using linearity in no-arbitrage condition(Lemma 15.5.3). (3) use Theorem 15.6.18. ☐

Lemma 19.10.8 (Black's formula for swaptions). *Assume the forward swap rate $S_{0,n}(t)$ under the forward swap measure(using swap annuity $A_{0,n}(t)$ as the numeraire) has dynamics*

$$dS_{0,n}(t) = \sigma_{0,n}(t)S_{0,n}(t)dW(t),$$

where $\sigma_{0,n}(t)$ is a deterministic time-dependent volatility and $W(t)$ is a Brownian motion under forward swap measure.

The payoff at maturity date T_0 is

$$A_{0,n}(T_0)(S_{0,n}(T_0) - K)^+,$$

where K is the strike.

The current price is given by

$$\begin{aligned} PSwpt_{0,n}(t) &= A_{0,n}(t)E_A[(S_{0,n} - K)^+ | \mathcal{F}_t] \\ &= A_{0,n}(t)(S_{0,n}(t)N(d_+) - KN(d_-)) \end{aligned}$$

where

$$d_{\pm} = \frac{\ln(S_{0,n}(t)/K) \pm \frac{1}{2}\sigma^2}{\sigma}, \sigma^2 = \int_t^{T_0} \sigma_{0,n}(s)^2 ds$$

Proof. Using the change of numeraire method, under forward measure Q_A, we have(Lemma 19.3.3)

$$\frac{V(t)}{A_{0,n}(t)} = E_A[\frac{V(T_0)}{A_{0,n}(t)}|\mathcal{F}_t] = E_A[\frac{A_{0,n}(T_0)(S_{0,n}(T_0)-K)^+}{A_{0,n}(t)}|\mathcal{F}_t] = E_A[(S_{0,n}(T_0)-K)^+|\mathcal{F}_t]$$

Note that under forward measure Q_A, the distribution $S_{0,n}(t)$ is log-normal and the rest is straight-forward. □

Remark 19.10.5 (how to choose numeraire)**.**

- If the final payoff is $(S_{0,t}(T_0)-K)^+$, and we can use forward measure Q_T associated with $P(t,T_0)$. Then
$$V(t) = P(t,T_0)E_T[(S_{0,t}(T_0)-K)^+|\mathcal{F}_t],$$
which can be easily evaluated if we know the distribution of $S_{0,t}(T_0)$ under measure Q_T.(see the example following Theorem 15.6.18).
- If the final payoff is
$$A_{0,n}(T_0)(S_{0,n}(T_0)-K)^+,$$
then it is wise to use forward swap measure.

19.10.2.2 *LMM approximation for swap rate volatility*

Lemma 19.10.9 (swap rate volatility approximation). *[1, p. 110]*

Under the forward measure Q_A with respect to the swap rate, we have

- $S_{0,n}(t)$ *is a martingale, and*
$$dS_{0,n}(t) = \sigma_{0,n}(t)S_{0,n}(t)dW^A(t),$$
- $$w_i(s) = \frac{\tau_i P(s,T_i)}{A_{0,n}(s)}$$
is a martingale.
- $$w_i(s)F_i(s) = \frac{P(s,T_{i-1}) - P(s,T_i)}{A_{0,n}(s)}.$$
is a martingale.

Moreover,

$$\begin{aligned}\sigma_{0,n}(s)^2 &= \frac{dS_{0,n}(s)}{S_{0,n}(s)}\frac{dS_{0,n}(s)}{S_{0,n}(s)} \\ &\approx \frac{\sum_{i=1}^n \sum_{j=1}^n w_i(s)F_i(s)w_j(s)F_j(s)}{S_{0,n}(s)^2}\sigma_i(s)\sigma_j(s)\rho_{i,j}ds \\ &\approx \frac{\sum_{i=1}^n \sum_{j=1}^n w_i(t)F_i(t)w_j(t)F_j(t)}{S_{0,n}(t)^2}\sigma_i(s)\sigma_j(s)\rho_{i,j}ds, t < s\end{aligned}$$

Proof. (1)(2)(3) directly from Lemma 19.10.7. (4) From Lemma 19.1.8, we have

$$\begin{aligned}S_{0,n}(t) &= \sum_{i=1}^n w_i(t)F_i(t) \\ dS_{0,n}(s) &= \sum_{i=1}^n w_i(s)dF_i(s) + \sum_{i=1}^n F_i(s)dw_i(s) + drift \\ &= \sum_{i=1}^n w_i(s)F_i(s)\sigma_i(s)dZ_i^A(s) + \sum_{i=1}^n F_i(s)dw_i(s) \\ &\approx \sum_{i=1}^n w_i(s)F_i(s)\sigma_i(s)dZ_i^A(s)\end{aligned}$$

where $drift$ represents terms of $O(ds)$ from $dF_i(s)dw_i(s)$, Z_i^A is a Brownian motion at measure Q_A. In the approximation, we neglects $\sum_{i=1}^n F_i(s)dw_i(s)$. $drift$ term disappear because $S_{0,n}(t)$ is a martingale under measure Q_A.

Note that in the last step, we approximate $w_i(s)F_i(s)$ and $S_{0,n}(s)$ by their previous value since they are martingales. □

Lemma 19.10.10 (Black's formula for swaptions with LMM approximate volatility). *Following Lemma 19.10.8, the price for a derivative with payoff*

$$A_{0,n}(T_0)(S_{0,n}(T_0) - K)^+$$

at T_0 is given by

$$PSwpt_{0,n}(t) = A_{0,n}(t)(S_{0,n}(t)N(d_+) - KN(d_-)).$$

where

$$d_{\pm} = \frac{\ln(S_{0,n}(t)/K) \pm \frac{1}{2}\sigma^2}{\sigma},$$

$$\sigma^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{w_i(t)F_i(t)w_j(t)F_j(t)}{S_{0,n}(t)^2}\int_t^{T_0} \sigma_i(s)\sigma_j(s)\rho_{i,j}ds$$

Proof. Use results from Lemma 19.10.9, we have

$$\sigma_{0,n}(s)^2 ds \approx \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i(t)F_i(t)w_j(t)F_j(t)}{S_{0,n}(t)^2}\sigma_i(s)\sigma_j(s)\rho_{i,j}ds, t < s.$$

□

19.10.3 Calibration and implementation

19.10.3.1 *Volatility calibration*

Example 19.10.1. [1, p. 121]

- Assume the instantaneous volatility takes the following time-homogeneous parametric form

$$\sigma_i(t;\alpha) = (a + b(T_{i-1} - t))e^{-c(T_{i-1}-t)} + d,$$

  with $\alpha = \{a, b, c, d\}$. Denote

$$\nu_i(\alpha) = \int_t^{T_{i-1}} \sigma_i(s;\alpha)ds.$$

- Denote σ_i^*as the implied caplet volatility such that

$$Cpl_i^{mkt}(t) = Cpl_i(t;\sigma_i^*),$$

  where

$$Cpl_i(t,\overline{\sigma}) = \tau_i P(t,T_i)(F(t;T_{i-1},T_i)N(d_+) - KN(d_-),$$

  and

$$d_{\pm} = \frac{\ln F(t;T_{i-1},T_i)/K \pm \frac{1}{2}\overline{\sigma}^2}{\overline{\sigma}}.$$

- The parameter α for the volatility is solved from optimization problem

$$\min_{\alpha} \sum_{i=1}^{n} ((\sigma_i^*)^2 - v_i(\alpha)^2)^2.$$

- Suppose we already construct the forward curve(Lemma 19.4.1), then

$$F(t, T_i, T_{i+1}) = \frac{1}{T_{i+1} - T_i}(\exp(\int_{T_i}^{T_{i+1} f(t,s)ds} - 1)).$$

19.10.3.2 *Correlation calibration*

Lemma 19.10.11 (Black's formula for swaptions with LMM approximate volatility). *Following Lemma 19.10.8, the price for a derivative with payoff*

$$A_{0,n}(T_0)(S_{0,n}(T_0) - K)^+$$

at T_0 is given by

$$PSwpt_{0,n}(t) = A_{0,n}(t)(S_{0,n}(t)N(d_+) - KN(d_-)).$$

where

$$\begin{aligned} d_{\pm} &= \frac{\ln(S_{0,n}(t)/K) \pm \frac{1}{2}\sigma^2}{\sigma}, \\ \sigma^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{w_i(t)F_i(t)w_j(t)F_j(t)}{S_{0,n}(t)^2} \int_t^{T_0} \sigma_i(s)\sigma_j(s)\rho_{i,j}ds \end{aligned}$$

Proof. See Lemma 19.10.10. □

Note 19.10.2 (calibration the correlation structure)**.**

- The correlation of historical forward curve is a good approximation to the correlation in the forward measure since correlation invariance property(Corollary 6.7.7.2).
- The Black's formula for swaptions with LMM approximate volatility also enables the usage of swaption market prices to calibrate correlation.
- Good references on this topic are [15, p. 125][4, p. 212].

Remark 19.10.6 (intuitions behind calibrating correlation using swaption)**.**

- Note that swap rate can be expressed as weighted sum of forward rate(Lemma 19.1.8)
$$S_{0,n}(t) = \sum_{i=1}^{n} w_i(t) F(t; T_{i-1}, T_i)$$
where
$$w_i(t) = \frac{\tau_i P(t, T_i)}{\sum_{j=1}^{n} P(t, T_j)}.$$
- The valuation of the swaption requires the variance of random variable $\sum_{i=1}^{n} w_i F_{t_{ex}; T_{i-1}, T_i}$(If we assume each F is log-normal and we lognormal distribution to approximate $\sum_{i=1}^{n} w_i F_{t_{ex}; T_{i-1}, T_i}$). The calculation of the variance will involves terms in the covariance matrix(See also the moment matching approach(Lemma 17.2.6)).

19.11 Volatility modeling

19.11.1 Implied Black volatility

19.11.1.1 *Cap and floor volatility*

Definition 19.11.1 (implied Black cap and floor volatility).

- *Consider a unit notational amount and a set of dates $T_0 < T_1 < \cdots < T_n$ with $\tau_i = T_i - T_{i-1}, i = 1, 2, ..., n$. At time T_i, the holder of an interest rate cap receives $\tau_i \max(L(T_{i-1}, T_i) - K, 0)$, where K is the* ***cap rate****. Each of these n call options are known as* ***caplets****.*
  The ith caplet is a European call option with expiry T_i written on the spot LIBOR rate $L(T_{i-1}, T_i)$(a random quantity) with payoff

$$(T_i - T_{i-1}) \max(L(T_{i-1}, T_i) - K, 0).$$

- *Consider the Black formula for caplets*

$$Cpl_i^B(t;\sigma) = \tau_i P(t, T_i)(F(t; T_{i-1}, T_i)N(d_+) - KN(d_-))$$

where

$$d_+ = \frac{\ln \frac{F(t;T_{i-1},T_i)}{K} + \frac{1}{2}\sigma^2(T_{i-1} - t)}{\sigma\sqrt{T_{i-1} - t}}, d_- = d_+ - \sigma\sqrt{T_{i-1} - t}.$$

- *The implied Black spot volatility $\hat{\sigma}_i^{caplet}$ of the ith caplet is defined as the unique solution to the equation*

$$Cpl_i^m kt(t) = Cpl_i^{Black}(t; \hat{\sigma}_i^{caplet}).$$

- *The implied Black flat volatility $\hat{\sigma}^{cap}$ of the ith caplet is defined as the unique solution to the equation*

$$Cap^m kt(t) = \sum_{i=1}^{n} Cpl_i^{Black}(t; \hat{\sigma}_n^{cap}).$$

Remark 19.11.1 (the parameters in the implied cap/floor volatility). The implied cap/floor volatility usually can be characterized by three parameter swaption expiry T_s, strike K, and the interest rate accrual period length $\Delta = T_i - T_{i-1}$.

Note 19.11.1 (volatility market observations). [5, p. 681]

- (observation) Usually in the market, volatility term structure like Figure 19.11.1 is usually observed. Particularly, there is a 'hump' at about the 2 to 3year point.

- (explanation) One possible explanation is as follows. Rates at the short end of the zero curve are controlled by central banks. By contrast, 2- and 3-year interest rates are determined to a large extent by the activities of traders. These traders may be overreacting to the changes observed in the short rate and causing the volatility of these rates to be higher than the volatlity of short rates. For maturities beyond 2 to 3 years, the mean reversion of interest rates.

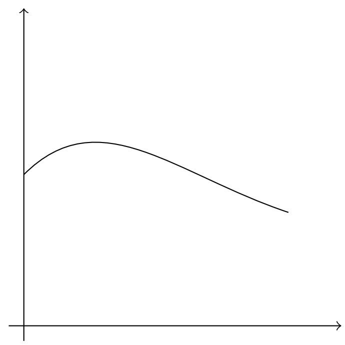


Figure 19.11.1: The cap/floor implied volatility hump

19.11.1.2 *Swaption volatility*

Definition 19.11.2. *Consider an interest rate swap with payment dates $T_1, T_2, ..., T_n$ and reset dates $T_0, T_1, ..., T_{n-1}$. Consider a* ***payer(receiver) swaption on spot-starting swaps*** *gives the holder the right to enter into an interest rate swap with maturity date $T_s = T_0$ such that the payoff at T_s is The payoff at maturity date T_0 is*

$$A(T_0, T_0, T_n)(S(T_0, T_0, T_n) - K)^+,$$

where K is the strike, $S_{0,n}(T_0)$ is the fair swap rate at T_0, and

$$A(t, T_0, T_n) = \sum_{i=1}^{n} (T_i - T_{i-1}) P(t, T_i).$$

The Black formula(Lemma 19.10.8) for the value of this swaption at current time $t < T_s$ is given by

$$\begin{aligned} PSwpt^B_{0,n}(t;\sigma) &= A(t, T_0, T_n) E_A[(S_{0,n} - K)^+ | \mathcal{F}_t] \\ &= A(t, T_0, T_n)(S(t, T_0, T_n)N(d_+) - KN(d_-)) \end{aligned}$$

where

$$d_{\pm} = \frac{\ln(S_{0,n}(t)/K) \pm \frac{1}{2}\sigma^2}{\sigma}.$$

Denote the market price of the swaption is given by $PSwpt^M_{0,n}$*. Then* ***the implied swaption volatility****, denoted by* $\sigma^{imp}_{SWap}(K, T_s, \Delta)$*, is the volatility that make the following equation hold:*

$$PSwpt^B_{0,n}(t; \sigma^{imp}_{SW}(K, T_s, \Delta)) = PSwpt^M_{0,n},$$

where we use $\Delta = T_n - T_0$ *to denote the underlying swap total length.*

Note 19.11.2 (the parameters in the implied swaption volatility)**.**

- Usually, the underlying swaps have standard swap dates $T_0, T_1, ..., T_n$; therefore, the swap tensor can be simply represented by its total length $\Delta = T_n - T_0$, which usually takes 1Y, 2Y, 3Y, 5Y, 10Y, and 30Y.
- The option expiry ranges from 0D, 7D, 14D, 1M, 2M, 3M, 6M, 12M, ..., 1Y, 2Y, ..., 20Y, 40Y.
- With the underlying standard swap, the implied swaption volatility usually can be characterized by three parameter swaption expiry T_s, strike K, and the underlying swap length Δ.

19.11.1.3 *Comparisons between swaption and cap/floor implied volatility*

Note 19.11.3 (comparisons between swaption and cap/floor implied volatility)**.** Denote $\sigma^{imp}_{CF}(T_C, K, \Delta_{CF})$ and $\sigma^{imp}_{SW}(T_S, K, \Delta_{SW})$ as the cap/floor implied volatility and the swaption implied volatility respectively.

- Δ_{CF} usually takes values of 1M, 3M, 6M, 1Y; Δ_{SW} usually takes values of 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 15Y, 20Y, 30Y.
- Δ_{CF} represents the volatility of short-tensor interest rate(1M to 1Y) at the expiry T_{CF}; Δ_{CF} represents the volatility of short-tensor interest rate(1M to 1Y) averaged from T_0 to T_n at the expiry T_{SW} since swap rate is 'kind of' average interest rate(Lemma 19.1.8) over a future period.

Expiry \Tenor	0.25Y	1Y	2Y	3Y	4Y	5Y	7Y	10Y	15Y	20Y	25Y	30Y	40Y
0D													
7D													
14D													
1M													
2M													
3M													
6M													
9M													
1Y													
1.5Y													
2Y													
3Y													
...													
10Y													
15Y													
20Y													
25Y													
30Y													
40Y													

Table 19.11.1: An ATM volatility table from market traded swaptions.

Expiry \Tenor	1M	3M	6M	12M
0D				
7D				
14D				
1M				
2M				
3M				
6M				
9M				
1Y				
1.5Y				
2Y				
3Y				
...				
10Y				
15Y				
20Y				
25Y				
30Y				
40Y				

Table 19.11.2: An ATM volatility table from market traded caplets and floorlets.

19.11.2 Constant Elasticity of Variance (CEV) model

Definition 19.11.3 (Constant Elasticity of Variance(CEV) model). *[3, p. 456] Denote a forward LIBOR rate by $F_i(t)$. In the Shifted-Lognormal model model, $F_i(t)$, under forward measure of T_i-maturity zero-coupon bond, is assumed to satisfy*

$$dF_i(t) = F_i(t)^{\beta}\sigma_i(t)dW_t$$

where $\beta \in [0,1]$ is a real constant, $\sigma(t)$ is a deterministic function of time, and $W(t)$ is a standard Brownian motion.

19.11.3 Shifted-Lognormal model

Definition 19.11.4 (Shifted-Lognormal model). *[3, p. 454] Denote a forward LIBOR rate by $F_i(t)$. In the Shifted-Lognormal model model, $F_i(t)$, under forward measure of T_i-maturity zero-coupon bond, is assumed to satisfy*

$$\begin{aligned} F_i(t) &= X_i(t) - \alpha \\ dX_i(t) &= \beta(t)X_j(t)dW_t \end{aligned}$$

where α is a real constant, $\beta(t)$ is a deterministic function of time, and $W(t)$ is a standard Brownian motion.

Note 19.11.4 (relation between shift-lognormal model and CEV model)**.** Note that the Taylor expansion of $F_i(t)^{\beta}$ gives

$$\begin{aligned} F_i(t)^{\beta} &= F_i(0)^{\beta} + \beta F_i(0)^{\beta-1}(F_i(t) - F_i(0)) + O((F_i(t) - F_i(0)^2) \\ &= F_i(0)^{\beta-1}(F_i(t) + (1-\beta)F_i(0)) + O((F_i(t) - F_i(0)^2) \\ &\approx F_i(0)^{\beta-1}(F_i(t) + (1-\beta)F_i(0)) \end{aligned}$$

Therefore,$dF_i(t) = F_i(t)^{\beta}\sigma_i(t)dW_t$ can be approximately written by

$$dF_i(t) = \sigma_i(t)F_i(0)^{\beta-1}(F_i(t) + (1-\beta)F_i(0))dW_t,$$

or equivalently,

$$d(F_i(t) + \alpha) = \sigma_i^D(t)(F_i(t) + \alpha)dW_t,$$

where

$$\alpha = \frac{(1-\beta)F_i(0)}{\beta}, \sigma_i^D(t) = \beta F_i(0)^{\beta-1}\sigma_i(t).$$

Lemma 19.11.1 (solution to Shifted-Lognormal model). *[3, p. 454] In the Shifted-Lognormal model, it follows that*

- *The solution of X_i is given by*

$$X_i(T) = X_i(t)\exp(-\frac{1}{2}\int_t^T \beta^2(u)du + \int_t^T \beta(u)dW(u))$$

- *The solution of F_i is given by*

$$F_i(T) = -\alpha + (F_i(t)+\alpha)\exp(-\frac{1}{2}\int_t^T \beta^2(u)du + \int_t^T \beta(u)dW(u))$$

Proof. Straight forward. □

Lemma 19.11.2 (pricing caplet). *[1, p. 151]*

The current value of a caplet with strike K and maturity T_{i-1} is given by

$$V(t) = P(t,T_i)\tau_i E_{Q_{T_i}}[(F_i(T_{i-1})-K)^+|\mathcal{F}_t] = P(t,T_i)E_{Q_{T_i}}[(X_i(T_{i-1})+\alpha-K)^+|\mathcal{F}_t]$$

Specifically,

$$\begin{aligned} V(t) &= Cpl(K+\alpha, F_i(t)+\alpha, \Sigma(t,T) \\ &= \tau_i P(t,T_i)((F_i(t)+\alpha)N(d_+) - (K+\alpha)N(d_-)) \end{aligned}$$

where

$$d_\pm = \frac{\ln\frac{F_i(t)+\alpha}{K+\alpha} \pm \frac{1}{2}\Sigma(t,T)}{\sqrt{\Sigma(t,T)}}$$

and

$$\Sigma(t,T)^2 = \int_t^T \beta^2(u)du.$$

Proof. Straight forward. □

Lemma 19.11.3. *The at-the-money(ATM) implied volatility for a caplet with strike K and maturity T_{i-1} is given by*

$$\sigma_{im}(K) = \frac{\sigma F_i(t)}{F_i(t)+\alpha},$$

Remark 19.11.2 (parameterization of the volatility surface). Consider a single maturity date T_{i-1} and a single tenor τ_i, the volatility smile(implied Black volatility as a function of K) is given by

19.11.4 SABR model

Definition 19.11.5 (SABR model). *[1][16] Denote a forward LIBOR rate by $F_i(t)$. In the SABR model, $F_i(t)$, under forward measure of T_i-maturity zero-coupon bond, is assumed to satisfy*

$$\begin{aligned} dF_i(t) &= F_i(t)^\beta \sigma_i(t) dW(t) \\ d\sigma_i(t) &= \nu \sigma_i(t) dZ(t) \\ \sigma_i(0) &= \sigma_i^0 \end{aligned}$$

where $W(t)$ and $Z(t)$ are Brownian motions under the forward measure Q_{T_i} such that $dW(t)dZ(t) = \rho dt$, ν is the constant volatility of $\sigma_i(t)$. The parameter ranges are $\nu \geq 0, 0 \leq \beta < 1, -1 < \rho < 1$.

Lemma 19.11.4 (European call/option pricing).

- *The European call price with strike K and maturity T is given by*

$$V_c(t) = P(t,T)(F(t,T)N(d_1) - KN(d_2))$$

where $F(t,T)$ is the forward price the underlying, and

$$d_{1,2} = \frac{\log(F(t,T)/K) \pm \frac{1}{2}\sigma_B^2(T-t)}{\sigma_B^2\sqrt{T-t}},$$

and the implied volatility $\sigma(F,K)$ is given by

$$\begin{aligned} \sigma_B(F,K) &= \frac{\alpha}{(FK)^{(1-\beta)/2}[1 + \frac{(1-\beta)^2}{24}\log^2(F/K) + \frac{(1-\beta)^4}{1920}\log^4(F/K) + \cdots]} \\ &\cdot \frac{z}{\chi(z)} \cdot \{1 + [\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(FK)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2](T-t) + \cdots\} \\ z &= \frac{\nu}{\alpha}(FK)^{(1-\beta)/2}\log(F/K) \\ \chi(z) &= \log\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho} \end{aligned}$$

- *At-the-money volatility*

$$\sigma_B(F,F) = \frac{\alpha}{(FK)^{(1-\beta)/2}}\{1+[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FK)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\nu\alpha}{(FK)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2](T-t)+\cdots\}$$

- *The put price given by*

$$V_p(t) = V_c(t) + P(t,T)(K-F(t,T)).$$

Remark 19.11.3 (choices of parameter for the volatility cube).

- At every maturity T, we use four parameter α_0, β, ν and ρ to characterize the volatility smile.
- For a set of maturity dates $T_1, T_2, ..., T_n$ and different tenors $\tau_1, \tau_2, ..., \tau_m$, we use a set of $4mn$ parameters to characterize the volatility surface. Usually, β is chosen to be constant independent of maturity date and tenors.

Note 19.11.5 (how parameters affects the shape of parameterized volatility surface). [1]

- An increase in the initial volatility σ_i^0 is an approximate upwards parallel shift in the Black implied volatility smile.
- A change in β from 1 to 0 will the cause the slope of smile to steepen.
- The slope of the smile can also be affected by ρ; therefore, we usually fix β and vary ρ.
- The ν parameter introduces curvature of the smile. The larger ν gives larger curvature.

19.12 Convexity adjustment

19.12.1 Principles of convexity adjustment

19.12.2 Convexity adjustment for interest rate products

19.12.2.1 *Convexity adjustment for LIBOR-in-arrears*

Lemma 19.12.1 (convexity adjustment for LIBOR-in-arrears). *Assume the forward rate $F(t) \triangleq F(t;S,T)$ has log-normal dynamics under forward measure Q_T, given by*

$$dF(t) = \tilde{\sigma}(t)F(t)dW_t.$$

Let $t < S < T, \tau = T - S$. Then,

$$\begin{aligned} E_S[L(S,T)|\mathcal{F}_t] &= F(t;S,T) + \frac{\tau F^2}{1+\tau F}(\exp(\sigma^2) - 1) \\ &\approx F(t;S,T) + \frac{\tau F^2}{1+\tau F}\sigma^2 \end{aligned}$$

where

$$\sigma^2 = \int_t^S \tilde{\sigma}^2(s)ds.$$

Proof. See Lemma 19.13.4. □

19.12.2.2 *Convexity adjustment for LIBOR rate*

Lemma 19.12.2.

- $$E_{T_p}[L(S,T)] = E_T[L(S,T)\frac{P(S,T_p)P(0,T)}{P(S,T)P(0,T_p)}] = \frac{P(0,T)}{P(0,T_p)}E_T[L(S,T)\frac{P(S,T_p)}{P(S,T)}].$$

- *Assume linear swap rate model such that*

$$\frac{P(S,T_p)}{P(S,T)} = \alpha + \beta(T_p)L(S,T),$$

where α is a constant, and $\beta(T_p)$ is a function of T_p. Then

$$\beta(T_p) = \frac{P(0,T_p)/P(0,T) - \alpha}{F(0,S,T)}, \alpha = 1.$$

Proof. Using the martingale property under martingale measure associated with $P(t,T)$, we have

$$\frac{P(0,T_p)}{P(0,T)} = E_T[\frac{P(S,T_p)}{P(S,T)}] = E_T[\alpha + \beta(T_p)L(S,T)] = \alpha + \beta(T_p)F(0,S,T).$$

Then we can get the result for $\beta(T_p)$. For α, we have

$$1 = \frac{P(S,T)}{P(S,T)} = \alpha + \beta(T)L(T,T) = \alpha.$$

□

19.12.2.3 *Convexity adjustment for CMS rate*

Lemma 19.12.3.

-
$$E_{T_p}[S(T_0)] = E_A[S(T_0)\frac{P(T_0,T_p)A(0)}{A(T_0)P(0,T_p)}] = \frac{A(0)}{P(0,T_p)}E_A[S(T_0)\frac{P(T_0,T_p)}{A(T_0)}].$$

- *Assume linear swap rate model such that*

$$\frac{P(T_0,T_p)}{A(T_0)} = \alpha + \beta(T_p)S(T_0),$$

where α is a constant, and $\beta(T_p)$ is a function of T_p. Then

$$\beta(T_p) = \frac{P(0,T_p)/A(0) - \alpha}{S(0)}, \alpha = \frac{1}{\sum_{i=1}^n \tau_i}.$$

Proof. Using the martingale property under martingale measure associated with $A(t)$, we have

$$\frac{P(0,T_p)}{A(0)} = E_A[\frac{P(T_0,T_p)}{A(T_0)}] = E_A[\alpha + \beta(T_p)S(T_p)] = \alpha + \beta(T_p)S(0).$$

Then we can get the result for $\beta(T_p)$. For α, we have

$$\begin{aligned} 1 &= \frac{\sum_{i=1}^n \tau_i P(T_0, T_i)}{A(T_0)} \\ &= \sum_{i=1}^n \tau_i(\alpha + \beta(T_i)S(T_0)) \\ &= \sum_{i=1}^n \tau_i\alpha + \sum_{i=1}^n \tau_i\beta(T_i)S(T_0) \\ &= \sum_{i=1}^n \tau_i\alpha + \sum_{i=1}^n \tau_i S(T_0)\frac{P(0,T_i)/A(0) - \alpha}{S(0)} \\ &= \sum_{i=1}^n \tau_i\alpha + (1 - \sum_{i=1}^n \tau_i\alpha)\frac{S(T_0)}{S(0)} \\ \implies (1 - \sum_{i=1}^n \tau_i\alpha)(1 - \frac{S(T_0)}{S(0)}) = 0 \end{aligned}$$

□

19.12.3 Convexity adjustment for stocks

[4, p. 259]

19.13 Exotic interest rate derivatives

19.13.1 LIBOR-in-arrears and in-arrears swap

19.13.1.1 *LIBOR-in-arrears*

Definition 19.13.1 (LIBOR-in-arrears). *A LIBOR-in-arrears contract is a contract that pays amount $\tau L(S,T), S < T$ at time S, where $\tau = T - S$.*

Lemma 19.13.1 (pricing LIBOR-in-arrears). *[1, p. 130] Consider a LIBOR-in-arrears contracts pays $\tau L(S,T)$ at time S.*

Then it time-t value is given by

$$Lia(t) = \tau P(t,S)E_S[L(S,T)|\mathcal{F}_t].$$

where E_S is the expectation taken with respect to forward measure of $P(t,S)$. or equivalently,

$$\begin{aligned} Lia(t) &= \tau P(t,T)E_T[L(S,T)|\mathcal{F}_t] + \tau^2 P(0,S)E_T[L(S,T)^2|\mathcal{F}_t] \\ &= \tau P(t,T)F(t;S,T) + \tau^2 P(0,S)E_T[L(S,T)^2|\mathcal{F}_t] \end{aligned}$$

where E_T is the expectation taken with respect to forward measure of $P(t,T)$.

Proof. Define

$$\frac{dQ_S}{dQ_T} = \frac{P(t,T)P(S,S)}{P(t,S)P(S,T)} = \frac{P(t,T)}{P(t,S)}(1+\tau L(S,T)).$$

Use change of numeraire technique(**??**), we have

$$\begin{aligned} P(t,S)E_S[L(S,T)|\mathcal{F}_t] &= P(t,S)E_T[L(S,T)\frac{dQ_S}{dQ_T}|\mathcal{F}_t] \\ &= P(t,S)E_T[L(S,T)\frac{P(t,T)}{P(t,S)}(1+\tau L(S,T))|\mathcal{F}_t] \\ &= \tau P(t,T)E_T[L(S,T)|\mathcal{F}_t] + \tau^2 P(0,S)E_T[L(S,T)^2|\mathcal{F}_t] \end{aligned}$$

Then we use the fact that $F(t;S,T) = E_T[L(S,T)|\mathcal{F}_t]$ in Lemma 19.3.4. □

Lemma 19.13.2 (evaluating LIBOR-in-arrears using caplets). *[1, p. 132] It follows that*

- $$L(S,T)^2 = 2\int_0^{L(S,T)} (L(S,T)-K)dK = 2\int_0^\infty (L(S,T)-K)^+ dK.$$

- $$\tau^2 P(0,S)E_T[L(S,T)^2|\mathcal{F}_t] = 2\tau\int_0^\infty \tau P(0,T)E_T[(L(S,T)-K)^+|\mathcal{F}_t]dK = 2\tau\int_0^\infty Cpl(t,K,S,T)dK$$

that is, we can use the caplet prices to price $L(S,T)^2$.

Lemma 19.13.3 (evaluating LIBOR-in-arrears using log-normal model). *[1, p. 132] Assume the forward rate $F(t) \triangleq F(t;S,T)$ has log-normal dynamics under forward measure Q_T, given by*

$$dF(t) = \tilde{\sigma}(t)F(t)dW_t.$$

Then

$$\begin{aligned} E_T[L(S,T)^2|\mathcal{F}_t] &= F(t)^2\exp(-\sigma^2)E_T[\exp(2\sigma Z)|\mathcal{F}_t], Z\sim N(0,1) \\ &= F(t)^2\exp(\sigma^2) \end{aligned}$$

where

$$\sigma^2 = \int_t^S \tilde{\sigma}^2(s)ds.$$

Proof.

$$\begin{aligned} dF(t) &= \tilde{\sigma}(t)F(t)dW_t \\ \implies L(S,T)^2 = F(S)^2 &= [F(t)\exp(-\frac{1}{2}\sigma^2+\sigma Z)]^2, Z\sim N(0,1) \\ &= F(t)^2\exp(-\sigma^2+2\sigma Z) \\ \implies E_T[L(S,T)^2] &= F(t)^2\exp(-\sigma^2)E_T[\exp(2\sigma Z)] \\ &= F(t)^2\exp(-\sigma^2)\exp(2\sigma^2) \\ &= F(t)^2\exp(\sigma^2) \end{aligned}$$

where

$$\sigma^2 = \int_t^S \tilde{\sigma}^2(s)ds.$$

and we use properties of lognormal random variable(Lemma 2.2.14). □

Lemma 19.13.4 (convexity adjustment for LIBOR-in-arrears). *Assume the forward rate $F(t) \triangleq F(t;S,T)$ has log-normal dynamics under forward measure Q_T, given by*

$$dF(t) = \tilde{\sigma}(t)F(t)dW_t.$$

Let $t < S < T, \tau = T - S$. Then,

$$\begin{aligned} E_S[L(S,T)|\mathcal{F}_t] &= F(t;S,T) + \frac{\tau F^2}{1+\tau F}(\exp(\sigma^2) - 1) \\ &\approx F(t;S,T) + \frac{\tau F^2}{1+\tau F}\sigma^2 \end{aligned}$$

where

$$\sigma^2 = \int_t^S \tilde{\sigma}^2(s)ds.$$

Proof. Define

$$\frac{dQ_S}{dQ_T} = \frac{P(t,T)P(S,S)}{P(t,S)P(S,T)} = \frac{P(t,T)}{P(t,S)}(1+\tau L(S,T)).$$

Use change of numeraire technique(**??**), we have

$$\begin{aligned} E_S[L(S,T)|\mathcal{F}_t] &= E_T[L(S,T)\frac{dQ_S}{dQ_T}|\mathcal{F}_t] \\ &= E_T[L(S,T)\frac{P(t,T)}{P(t,S)}(1+\tau L(S,T))|\mathcal{F}_t] \\ &= \frac{P(t,T)}{P(t,S)}(E_T[L(S,T)|\mathcal{F}_t] + \tau E_T[L(S,T)^2|\mathcal{F}_t]) \\ &= \frac{1}{1+\tau F(t;S,T)}(F(t;S,T) + \tau F(t;S,T)^2 \exp(\sigma^2)) \end{aligned}$$

Then

$$\Delta \triangleq E_S[L(S,T)|\mathcal{F}_t] - F(t;S,T) =$$

Then we use the fact that $F(t;S,T) = E_T[L(S,T)|\mathcal{F}_t]$ in Lemma 19.3.4. □

19.13.1.2 *In-arrears swap*

Definition 19.13.2 (In-arrears swap). *[1, p. 132] The holder of a payer in-arrears swap pays a fixed amount $\tau_i K$ in exchange for a floating payment $\tau_i L(T_{i-1}, T_i)$ at time $T_{i-1}, i = 1, ..., n$ rather than at time T_i.*

Lemma 19.13.5 (pricing In-arrears swap). *[1, p. 132] The current value at time t of an in-arrears swap is*

$$PSia(t) = \sum_{i=1}^n \tau_i P(t, T_{i-1}) E_{T_{i-1}}(L(T_{i-1}, T_i) - K) = \sum_{i=1}^n Lia_i(t) - K\sum_{i=1}^n \tau_i P(0, T_{i-1}).$$

where $Lia_i(t)$ is the current value at time t for a LIBOR-in-arrears payment of $\tau L(T_{i-1}, T_i)$.

Proof. Straight forward. □

19.13.2 Constant-maturity swaps and related products

19.13.2.1 *Constant-maturity swaps*

Lemma 19.13.6 (pricing constant-maturity swap). *The current time t value of a CMS contract is*

$$CMS(t) = \sum_{i=0}^{n-1} \tau_{i+1} P(0, T_{i+1}) E_{P_{T_{i+1}}}[(S_{i,i+m}(T_i) - K)|\mathcal{F}_t].$$

For each term we have,

$$P(0, T_{i+1}) E_{P_{T_{i+1}}}[S_{i,i+m}(T_i)|\mathcal{F}_t] = A_{i,i+m}(T_{i+1}) E_{A_{i,i+m}}[S_{i,i+m}(T_i) \frac{P(T_i, T_{i+m})}{A_{i,i+m}(T_i)}|\mathcal{F}_t].$$

If we assume $P(T_i, T_{i+m}) = (\alpha + \beta S_{i,i+m}(T_i)) A_{i,i+m}(T_i)$, then

$$E_{A_{i,i+m}}[S_{i,i+m}(T_i) \frac{P(T_i, T_{i+m})}{A_{i,i+m}(T_i)}|\mathcal{F}_t] = \alpha S_{i,i+m}(t) + \beta E_{A_{i,i+m}}[S_{i,i+m}(T_i)^2|\mathcal{F}_t].$$

Note that $S_{i,i+m}(t)$ has log-normal dynamics under forward swap measure $Q_{A_{i,i+m}}$(Lemma 19.10.7).

Proof. Note that $S_{i,i+m}(t)$ has nice property under measure $Q_{A_{i,i+m}}$ instead of $Q_{P_{T_{i+1}}}$.

Use change of numeraire technique (**??**), we have

$$\frac{dP_{T_{i+1}}}{dP_{A_{i,i+m}}} = \frac{P(T_{i+1}, T_{i+1})}{P(t, T_{i+1})} \frac{A_{i,i+m}(t)}{A_{i,i+m}(T_i)}.$$

and

$$E_{P_{T_{i+1}}}[\cdots|\mathcal{F}_t] = E_{A_{i,i+m}}[\cdots \frac{dP_{T_{i+1}}}{dP_{A_{i,i+m}}}|\mathcal{F}_t]$$

□

Remark 19.13.1 (determination of the linear swap approximation parameters).

-

19.13.2.2 *CMS caps, floors and spread options*

Definition 19.13.3. *[1, p. 136] A CMS cap(or floor) is an agreement where the holder receives cash payment on a set of predefined dates depending on the spot swap rate. A CMS cap consists of a series CMS caplets.*

A CMS caplet with unit notional paying at time T_{i+1} on an m-period swap rate setting at time T_i. Note that the swap rate is associated with an swap usually has final payment much grater than T_{i+1}. The payoff at time T_{i+1} is

$$\tau_{i+1}(S_{i,i+m}(T_i) - K)^+,$$

where $\tau_{i+1} = T_{i+1} - T_i$, K is the strike rate.

Remark 19.13.2 (calibration of swap rate model)**.** Note that CMS caps/floors can be used to calibrate the volatility in the log-normal model of swap rates.

19.13.3 Ratchet floater

Definition 19.13.4 (ratchet floater). *[1, p. 138] Take a unit notional amount and a set of dates, $0 \leq T_0 \leq T_1 \cdots T_n$ with acrual periods $\tau_i = T_i - T_{i-1}$.*

- *At time T_i, the holder receives a payment $\tau_i(L(T_{i-1}, T_i) + X)$, where X is a preassigned constant spread, and pays coupon c_i.*
- *The first coupon $c_1 = \tau_1(L(T_0, T_1) + Y)$; the coupon $c_i, i = 2, ..., n$ satisfies*

$$c_i = c_{i-1} + \min((\tau_i(L(T_{i-1}, T_i) + Y) - c_{i-1})^+, \alpha),$$

where $\alpha > 0$ is preassigned constant.

Lemma 19.13.7. *[1, p. 137] The current time t value of a ratchet floater is given by*

$$V(t) = \sum_{i=1}^{n} P(t, T_i) E_{P_{T_i}}[\tau_i(F(T_{i-1}; T_{i-1}, T_i) + X) - c_i | \mathcal{F}_t],$$

where $E_{P_{T_i}}$ is taking expectation with respect to the forward measure of $P(t, T_i)$.

Or equivalently,

$$V(t) = P(t, T_n) \sum_{i=1}^{n} E_{P_{T_n}}[\frac{\tau_i(F(T_{i-1}; T_{i-1}, T_i) + X) - c_i}{P(T_i, T_n)} | \mathcal{F}_t].$$

Proof. Use change of numeraire technique (**??**), we have

$$\frac{dP_{T_i}}{dP_{T_n}} = \frac{P(T_i,T_i)}{P(t,T_i)}\frac{P(t,T_n)}{P(T_i,T_n)}.$$

Then,

$$\begin{aligned}
&P(t,T_i)E_{P_{T_i}}[\tau_i(F(T_{i-1};T_{i-1},T_i)+X)-c_i|\mathcal{F}_t] \\
&= P(t,T_i)E_{P_{T_n}}[(\tau_i(F(T_{i-1};T_{i-1},T_i)+X)-c_i)\frac{dP_{T_i}}{dP_{T_n}}|\mathcal{F}_t] \\
&= P(t,T_i)E_{P_{T_n}}[(\tau_i(F(T_{i-1};T_{i-1},T_i)+X)-c_i)\frac{P(T_i,T_i)}{P(t,T_i)}\frac{P(t,T_n)}{P(T_i,T_n)}|\mathcal{F}_t] \\
&= P(t,T_n)\sum_{i=1}^{n} E_{P_{T_n}}[\frac{\tau_i(F(T_{i-1};T_{i-1},T_i)+X)-c_i}{P(T_i,T_n)}|\mathcal{F}_t]
\end{aligned}$$

□

Remark 19.13.3 (Monte Carlo evaluation)**.**

- The sample point of $P(T_i,T_n)$ under Q_{T_n} is discussed at Lemma 19.10.4.

19.13.4 Interest rate futures

19.13.4.1 *Fundamentals*

Lemma 19.13.8 (dynamics of futures under risk neutral measure). *[5, p. 392] Under risk-neutral measure Q,*

- *$Fur(t;T_0,T_1)$ is a martingale such that*

$$Fur(t;T_0,T_1) = E[Fur(s;T_0,T_1)|\mathcal{F}_t], s > t$$

- $$Fur(t;T_0,T_1) = E_Q[L(T_0,T_1)|\mathcal{F}_t]$$

- *In the Brownian motion representation, the futures rate has risk-neutral dynamics given by*

$$dFur(t;S,T) = \sigma(t;S,T)Fur(t;S,T)$$

19.13.4.2 *Convexity adjustment*

Lemma 19.13.9 (convexity adjustment for futures rate). *Let* $Fur(t;S,T) = E_Q[L(S,T)|\mathcal{F}_t]$ *be the futures rate. Let* $F(t;S,T) = E_T[L(S,T)|\mathcal{F}_t]$ *be the forward rate. Denote the convexity adjustment by* $\Delta = Fur(t;S,T) - F(t;S,T)$*, then*

$$\Delta = \frac{1}{\tau}(E_Q[\frac{P(S,S)}{P(S,T)}|\mathcal{F}_t] - \frac{P(t,S)}{P(t,T)}]),$$

or equivalently,

$$\Delta = E_Q[L(S,T)(1 - \frac{B(t)}{B(T)}\frac{1}{P(t,T)})|\mathcal{F}_t].$$

Or equivalently,

$$\Delta = -\frac{1}{P(t,T)}cov(L(S,T)\frac{B(t)}{B(T)}).$$

Proof. (1)

$$\begin{aligned}
\Delta &= E_Q[L(S,T)|\mathcal{F}_t] - E_T[L(S,T)|\mathcal{F}_t] \\
&= E_Q[\frac{1}{\tau}(\frac{P(S,S)}{P(S,T)} - 1)|\mathcal{F}_t] - E_T[\frac{1}{\tau}(\frac{P(S,S)}{P(S,T)} - 1)|\mathcal{F}_t] \\
&= E_Q[\frac{1}{\tau}(\frac{P(S,S)}{P(S,T)} - 1)|\mathcal{F}_t] - \frac{1}{\tau}(\frac{P(t,S)}{P(t,T)} - 1) \\
&= \frac{1}{\tau}(E_Q[\frac{P(S,S)}{P(S,T)|\mathcal{F}_t} - \frac{P(t,S)}{P(t,T)}])
\end{aligned}$$

(2)

$$\begin{aligned}
\Delta &= E_Q[L(S,T)|\mathcal{F}_t] - E_T[L(S,T)|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - E_Q[L(S,T)\frac{dQ_T}{dQ}|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - E_Q[L(S,T)\frac{P(T,T)B(t)}{P(t,T)B(T)}|\mathcal{F}_t] \\
&= E_Q[L(S,T)(1 - \frac{B(t)}{B(T)}\frac{1}{P(t,T)})|\mathcal{F}_t]
\end{aligned}$$

(3) Note that

$$
\begin{aligned}
\Delta &= E_Q[L(S,T)|\mathcal{F}_t] - E_T[L(S,T)|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - E_Q[L(S,T)\frac{dQ_T}{dQ}|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - E_Q[L(S,T)\frac{P(T,T)B(t)}{P(t,T)B(T)}|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - Cov(L(S,T)\frac{P(T,T)B(t)}{P(t,T)B(T)}|\mathcal{F}_t) - E_Q[L(S,T)|\mathcal{F}_t]E_Q[\frac{P(T,T)B(t)}{P(t,T)B(T)}|\mathcal{F}_t] \\
&= E_Q[L(S,T)|\mathcal{F}_t] - Cov(L(S,T)\frac{P(T,T)B(t)}{P(t,T)B(T)}|\mathcal{F}_t) - E_Q[L(S,T)|\mathcal{F}_t] \\
&= -Cov(L(S,T)\frac{B(t)}{P(t,T)B(T)}|\mathcal{F}_t)
\end{aligned}
$$

□

Remark 19.13.4 (futures rate usually greater than forward rate). Note that because $L(S,T)$ and $1/B(T)$ are negatively correlated, therefore $\Delta > 0$; that is, futures rate is greater than forward rate.

Lemma 19.13.10 (interest rate futures convexity adjustment in the Gaussian HJM framework). *[4, p. 262] In the HJM framework(Theorem 19.9.1), the forward rate $f(t,T)$, **under risk-neutral measure** Q,is given*

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

where $W(t)$ is a Brownian motion,

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du.$$

Furthermore, the zero-coupon bond price $P(t,T)$ under Q satisfies SDE

$$dP(t,T) = r(t)P(t,T)dt + \Sigma(t,T)P(t,T)dW(t)$$

with log-volatility

$$\Sigma(t,T) = -\int_t^T \sigma(t,u)du.$$

It follows that

- $$\frac{P(S,S)}{P(S,T)} = \frac{P(t,S)}{P(t,T)}\exp(\int_t^S (\Sigma(u,S)-\Sigma(u,T)dW_u - \int_t^S \frac{1}{2}(\Sigma(u,S)-\Sigma(u,T))^2 du)$$
$$\exp(\int_t^S \Sigma(u,S)(\Sigma(u,S)-\Sigma(u,T))du)$$

- $$E_Q[\frac{P(S,S)}{P(S,T)}\mathcal{F}_t] = \frac{P(t,S)}{P(t,T)}\exp(\int_t^S \Sigma(u,S)(\Sigma(u,S)-\Sigma(u,T))du)$$

- *The convexity adjustment* $\Delta = Fur(t;S,T) - F(t;S,T)$ *is given by*
$$\begin{aligned}\Delta &= \frac{1}{\tau}\frac{P(t,S)}{P(t,T)}(\exp(\int_t^S \Sigma(u,S)(\Sigma(u,S)-\Sigma(u,T))du)-1) > 0\\ &= \frac{\tau F(t;S,T)+1}{\tau}(\exp(\int_t^S \Sigma(u,S)(\Sigma(u,S)-\Sigma(u,T))du)-1)\end{aligned}$$

Proof. (1)

From Theorem 19.3.5, we have The ratio $\triangleq \frac{P(t,S)}{P(t,T)}, S \leq T$ has SDE, under measure Q, given by:

$$d\tilde{FP}/\tilde{FP} = (\Sigma(t,S)-\Sigma(t,T))dW(t) + \Sigma(t,T)(\Sigma(t,T)-\Sigma(t,S))dt.$$

(2) note that

$$\exp(\int_t^S (\Sigma(u,S)-\Sigma(u,T)dW_u - \int_t^S \frac{1}{2}(\Sigma(u,S)-\Sigma(u,T))^2 du)$$

is the martingale with mean value 1 under measure Q.

(3) Use Lemma 19.13.9. □

Remark 19.13.5. For more discussion on convexity adjustment, see [4][3, p. 559].

Lemma 19.13.11 (interest rate futures convexity adjustment in the Ho-Lee model). *[5, p. 144][4, p. 264]*

The Ho-Lee short rate model(Lemma 19.5.23) under the risk-neutral measure Q gives zero-coupon bond price as

$$dP(t,T)/P(t,T) = r(t)dt - \sigma(T-t)dW_t.$$

The convexity adjustment $\Delta = Fur(t; S, T) - F(t; S, T)$ *is given by*

$$\Delta = \frac{1}{\tau}\frac{P(t,S)}{P(t,T)}(\exp(\int_t^S \Sigma(u,S)(\Sigma(u,S) - \Sigma(u,T))du) - 1).$$

Proof. (1)

From Theorem 19.3.5, we have The ratio $\triangleq \frac{P(t,S)}{P(t,T)}, S \leq T$ has SDE, under measure Q, given by:

$$d\tilde{FP}/\tilde{FP} = (\Sigma(t,S) - \Sigma(t,T))dW(t) + \Sigma(t,T)(\Sigma(t,T) - \Sigma(t,S))dt.$$

(2) note that

$$\exp(\int_t^S (\Sigma(u,S) - \Sigma(u,T)dW_u - \int_t^S \frac{1}{2}(\Sigma(u,S) - \Sigma(u,T))^2 du)$$

is the martingale with mean value 1 under measure Q.

(3) Use Lemma 19.13.9. □

19.13.4.3 *Interest rate futures options*

Definition 19.13.5 (European call option on futures). *An European call option on Eurodollar futures obtains the right to enter into a Eurodollar futures at the specified strike price K.*

A call option has maturity T earlier than the delivery date T_1 *of the futures. The total payoff of a call option is* $\max(Fur(T; T_1, T_2) - K, 0)$

Example 19.13.1. [5, p. 385] It is February and the futures prices for the June Eurodollar contract is 93.82(corresponding to a 3-month Eurodollar interest rate of 6.18% per annum). The price of a call option on the contract with a strike price of 94.00 is quoted as 10 basis point.

Suppose in June, the Eurodollar futures price increases to 94.78, then the long position of the call will enable the investor to get a futures price of 94, therefore gaining profit.

Lemma 19.13.12 (pricing call option on Eurodollar futures). *The price of call option maturing at T with payoff $(Fur(T)-K)^+$ is given by*

$$\begin{aligned} V(t) &= E_Q[\exp(-\int_t^T r(s)ds)(Fur(T)-K)^+|\mathcal{F}_t] \\ &= P(t,T)E_T[(Fur(T)-K)^+|\mathcal{F}_t] \\ &= P(t,T)E_T[(F(T)+\Delta(T)-K)^+|\mathcal{F}_t] \end{aligned}$$

where $F(t)$ the forward price and $\Delta(T) = \triangleq Fur(T) - F(t)$ is the convexity adjustment.

19.13.4.4 *Interest rate futures as hedging instrument*

19.13.5 Callable bonds

Definition 19.13.6 (callable zero-coupon bond). *A **callable zero-coupon bond** is a zero-coupon bond that allows the issuer the right, but not the obligation, to buy back from the bond holders at **pre-specified prices** on the **pre-specified call dates**.*

Remark 19.13.6 (understand how callability benefits issuers/borrowers).

- Suppose a company issue a callable zero-coupon bond at price 0.1 with maturity five years later. The embedded call option has call date four year later with call price 0.8.
- Suppose the interest rate drops, and therefore the bond price drops during the next four years. Particularly at the call date, the price of 1-year tenor zero-coupon bond is at 0.9.
- Then the company can issue/sell 1 year zero-coupon bond at 0.9 and then buy back the previous bond at 0.8. In the process, the company gain 0.1.
- In summary, the embedded call option gives the issuer the right to refinance the debt when there is a lower interest rate in the market.

Lemma 19.13.13. *Consider a callable zero-coupon bond with bond maturity date at T. Let $T_c < T$ be the maturity date of the embedded call option and let K be the strike, then the value at time t is given by*

$$V(t) = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] - E_Q[\exp(-\int_t^{T_c} r(s)ds)(P(T_c,T)-K)^+|\mathcal{F}_t].$$

In other words,

$$price\ of\ callable\ bond = price\ of\ normal\ bond - price\ of\ embedded\ option.$$

Proof. We can analyze the discounted cash flow to the bond holders at time T and time T_c:

$$\begin{aligned}
V(t) &= E_Q[\frac{B(t)}{B(T)}\mathbf{1}_{P(T_c,T)<K}|\mathcal{F}_t] + E_Q[\frac{B(t)}{B(T_c)}K\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\frac{B(t)}{B(T)}|\mathcal{F}_t] - E_Q[\frac{B(t)}{B(T)}\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] + E_Q[\frac{B(t)}{B(T_c)}K\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\frac{B(t)}{B(T)}|\mathcal{F}_t] - E_Q[\frac{B(t)}{B(T_c)}\frac{B(T_c)}{B(T)}\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] + E_Q[\frac{B(t)}{B(T_c)}K\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\frac{B(t)}{B(T)}|\mathcal{F}_t] - E_Q[\frac{B(t)}{B(T_c)}E[\frac{B(T_c)}{B(T)}|\mathcal{F}_{T_c}]\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] + E_Q[\frac{B(t)}{B(T_c)}K\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\frac{B(t)}{B(T)}|\mathcal{F}_t] - E_Q[\frac{B(t)}{B(T_c)}P(T_c,T)\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] + E_Q[\frac{B(t)}{B(T_c)}K\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\frac{B(t)}{B(T)}|\mathcal{F}_t] - E_Q[\frac{B(t)}{B(T_c)}(P(T_c,T)-K)\mathbf{1}_{P(T_c,T)>K}|\mathcal{F}_t] \\
&= E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t] - E_Q[\exp(-\int_t^{T_c} r(s)ds)(P(T_c,T)-K)^+|\mathcal{F}_t]
\end{aligned}$$

□

19.13.6 Bermudian option

19.14 Hedging and risk management

19.14.1 Swaps

19.14.1.1 *DV01*

Definition 19.14.1 (DV01). *DV01 is the dollar value change when the rate (either floating rate or fixed rate) change by 1 basis point.*

Lemma 19.14.1 (DV01 for floating and fixed leg). *Consider a set of dates $T_0^f < T_1^f < T_2^f < \cdots < T_n^f$. Let the float leg payment be $\tau_i^f L(T_{i-1}^f, T_i^f) \cdot N_i$ at T_i. Let current time be 0. Then the fixed leg DV01 is defined as*

$$DV01_{float} = 10^{-4}\sum_{i=1}^{N}\tau_i^f N_i DF(0, T_i^f).$$

Consider a set of dates $T_0^g < T_1^g < T_2^g < \cdots < T_n^g$. Let the fixed leg payment be $\tau_i^g K \cdot M_i$ at T_i^g. Let current time be 0. Then the fixed leg DV01 is defined as

$$DV01_{fixed} = 10^{-4} \sum_{i=1}^{N} \tau_i^g M_i DF(0, T_i^g),$$

Proof. (1) By definition

$$DV01_{float} = \sum_{i=1}^{N} \tau_i^f (L(T_{i-1}^f, T_i^f) + 1bp - L(T_{i-1}^f, T_i^f)) N_i DF(0, T_i^f).$$

(2) Similar to (1). □

19.14.2 Zero curve risk

Remark 19.14.1 (handling zero-curve market risk). The zero-curve's change will change the value of instruments that have a dependence on it. The impact on the value can be evaluated in the following ways:

- Calculate the impact of a 1-basis-point parallel shift in the zero curve. This is sometimes termed a DV01.
- Calculate the impact of small changes in the quotes for each of the instruments used to construct the zero curve.
- Divide the zero curve (or the forward curve) into a number of sections (or buckets). Calculate the impact of shifting the rates in one bucket by 1 basis point, keeping the rest of the initial term structure unchanged.
- Carry out a principal components analysis as outlined as follows. Calculate a delta with respect to the changes in each of the first few factors. The first delta then measures the impact of a small, approximately parallel, shift in the zero curve; the second delta measures the impact of a small twist in the zero curve; and so on.

Lemma 19.14.2 (principal component analysis of zero-curve market risk).

19.15 Approximate pricing methods

19.15.0.1 *Equivalent cash flow*

Lemma 19.15.1 (equivalent cash flow for LIBOR payment single curve framework). *Consider single curve framework. Let $t < T_1 < T_2, \tau = T_2 - T_1$. Then a LIBOR payment $\tau L(T_1, T_2)$ has an equivalent cash flow of 1\$ at time T_1 and -1\$ at time T_2.*

By equivalence, we mean that the two stochastic processes defined by

$$V_1(t) = E_Q[\frac{M(t)}{M(T_2)}\tau L(T_1, T_2)|\mathcal{F}_t],$$

and

$$V_2(t) = E_Q[\frac{M(t)}{M(T_1)} - \frac{M(t)}{M(T_2)}|\mathcal{F}_t].$$

are path-wise equal.

Proof.

$$\begin{aligned} V_1(t) &= E_Q[\frac{M(t)}{M(T_2)}\tau L(T_1, T_2)|\mathcal{F}_t] \\ &= E_Q[\frac{M(t)}{M(T_2)}(\frac{M(T_2)}{M(T_1)} - 1)|\mathcal{F}_t] \\ &= E_Q[\frac{M(t)}{M(T_1)} - \frac{M(t)}{M(T_2)}|\mathcal{F}_t] \\ &= V_2(t) \end{aligned}$$

where use the relation $(1 + \tau L(T_1, T_2)) = M(T_2)/M(T_1)$ holds due to Theorem 19.1.1. □

Lemma 19.15.2 (equivalent cash flow for LIBOR payment multi-curve framework). *Consider multi-curve framework. Let $t < T_1 < T_2, \tau = T_2 - T_1$. Assume static spread exists between OIS curve and LIBOR forward curve. Then a LIBOR payment $\tau L(T_1, T_2)$ has an equivalent cash flow of 1\$ at time T_1, -1\$ at time T_2, and $\tau(F_L(t, T_1, T_2) - F_{OIS}(t, T_1, T_2))$ for all $t < T_1$.*

By equivalence, we mean that the two stochastic processes defined by

$$V_1(t) = E_Q[\frac{M(t)}{M(T_2)}\tau L(T_1, T_2)|\mathcal{F}_t],$$

and

$$V_2(t) = E_Q[\frac{M(t)}{M(T_1)} - \frac{M(t)}{M(T_2)} + \frac{M(t)(\tau s}{M(T_2)}|\mathcal{F}_t].$$

are path-wise equal, where $s = F_L(t, T_1, T_2) - F_{OIS}(t, T_1, T_2)$.

Proof.

$$\begin{aligned}
V_1(t) &= E_Q[\frac{M(t)}{M(T_2)}\tau L(T_1, T_2)|\mathcal{F}_t] \\
&= E_Q[\frac{M(t)}{M(T_2)}\tau(L_{OIS}(T_1, T_2) + s)|\mathcal{F}_t] \\
&= E_Q[\frac{M(t)}{M(T_2)}(\frac{M(T_2)}{M(T_1)} - 1 + s\tau)|\mathcal{F}_t] \\
&= E_Q[\frac{M(t)}{M(T_1)} - \frac{M(t)}{M(T_2)} + \frac{M(t)(\tau s}{M(T_2)}|\mathcal{F}_t] \\
&= V_2(t)
\end{aligned}$$

where use the relation $(1 + \tau L_{OIS}(T_1, T_2)) = M(T_2)/M(T_1)$ holds due to Theorem 19.1.1.

□

Equivalent cash flows are useful for pricing various options in term structure models. The leg can be expressed, exactly or approximately, as a sum of simple cash paymens all on or after the input asOfDate. The

cash flows have the characateristics that:

The NPV of the cash flows, when calculated using the ycDiscount, ? is exactly equal to the NPV of the 304 leg calculated with the input asOfDate, stublype, ycDiscount, and ? ycForecast.

The dynamics of the cashflows under changes of ycDiscount is the best approximation to the dynamics of 306 the leg, assuming that the spread between ycForecast and ycDiscount is static.

19.16 Inflation linked derivatives

19.16.1 Markets and products

The primary purpose of inflation derivatives is the transfer of inflation risk. For example, real estate companies may want to shed some of their natural exposure to inflation risk, while pension funds may want to cover their natural liabilities to this risk. In their simplest form, inflation derivatives provide an efficient way to transfer inflation risk. But their flexibility also allows them to replicate in derivative form the inflation risks embedded in other instruments such as standard cash instruments (that is, inflationlinked bonds). For example, as we will see later, an inflation swap can be theoretically replicated using a portfolio of a zero-coupon inflation-linked bond and a zero-coupon nominal bond.

19.16.2 Pricing Models

19.16.2.1 *Basic concepts*

Definition 19.16.1 (inflation index, nominal value, real value).
- *We use inflation index $I(t)$ to represent the price level or inflation level.*
- *Let $V^n(t)$ denote the nominal value; that is the value measured in the unit of money. Let $V^r(t)$ denote the real value; that is the value measured in terms common goods and services.*
- *The nomial value and real value are related by $V^n(t) = I(t)V^r(t)$.*

Definition 19.16.2 (inflation-linked zero-coupon bond).

- *An inflation-linked zero-coupon bond (ILZCB)is a bond that has a single payment $I(T)$ at maturity T, where $I(t)$ is the inflation index.*
- *At maturity T, the bond has real value of 1 given by*

$$\frac{I(T)}{I(T)} = 1.$$

- *The current time t value of an*

19.16.2.2 *Martingale pricing framework*

Definition 19.16.3 (money account numeraire). *Let $r(t)$ be the real short rate and $n(t)$ be the nomial short rate.*

- *The nominal money account $M^n(t)$ is defined by*

$$M^n(t) = \exp(\int_0^t n(s)ds).$$

  The martingale measure associated with $M^n(t)$ is called nominal risk-neutral measure, denoted by Q^n.
- *The real money account $M^r(t)$ is defined by*

$$M^r(t) = \exp(\int_0^t r(s)ds).$$

  The martingale measure associated with $M^r(t)$ is denoted by Q^r.

Theorem 19.16.1 (martingale property of inflation index under nominal risk-neutral measure). *Let current time be $t, t \leq T$.*

- $$I(t)\frac{M^r(t)}{M^n(t)} = E_{Q^n}[\frac{I(T)M^r(T)}{M^n(T)}|\mathcal{F}_t].$$
- *If the interest rates are constants given by r, n, we have*

$$E_{Q^d}[I(T)|\mathcal{F}_t] = \exp((n-r)(T-t))I(t).$$

- *If we assume $S(t)$ is governed by geometric SDE, then under domestic risk-neutral measure, we have representation*

$$dI(t)/I(t) = (n(t) - r(t))dt + \sigma(t)dW^Q(t).$$

Proof. Under the nomial risk-neutral measure, the nominal value of the real money market account $M^r(T)I(T)$ will have martingale property given by

$$I(t)\frac{M^r(t)}{M^n(t)} = E_{Q^n}[\frac{I(T)M^r(T)}{M^n(T)}|\mathcal{F}_t].$$

□

19.16.2.3 *Dynamic models*

Theorem 19.16.2 (no-arbitrage Hull-White short rate model). *Consider the following dynamic model under the nominal risk-neutral measure*

$$\begin{aligned} dn(t) &= (\theta_n(t) - k_n n(t))dt + \sigma_n(t)dW_n^n(t) \\ dr(t) &= (\theta_r(t) - \rho_{r,I}\sigma_I\sigma_r - k_r r(t))dt + \sigma_r(t)dW_r^n(t) \\ dI(t) &= I(t)(n(t) - r(t))dt + \sigma_I I(t)dW_I^n(t) \end{aligned}$$

Then,

- *the quantities* $\frac{I(t)M^r(t)}{M^n(t)}$ *and* $\frac{I(t)P^r(t,T)}{M^n(t)}$ *are martingales under nominal risk-neutral measure.*
- *under the real risk-neutral measure, the real short rate has dynamics*

$$dr(t) = (\theta_r(t) - k_r r(t))dt + \sigma_r(t)dW_r(t)^r$$

where W_r^r *is a Brownian motion under real risk-neutral measure, and*

$$dW_r^r = dW_r - \rho_{r,I}\sigma_I\sigma_r dt.$$

Proof. (1)(a) To show $\frac{I(t)M^r(t)}{M^n(t)}$ is a martingale, we have

$$\begin{aligned} & d(\ln \frac{I(t)M^r(t)}{M^n(t)}) \\ &= d\ln I(t) + d\ln M^r(t) - d\ln M^n(t) \\ &= (n(t) - r(t))dt - \frac{1}{2}\sigma_I^2 dt + \sigma_I dW_I + r(t)dt - n(t)dt \\ &= -\frac{1}{2}\sigma_I^2 dt + \sigma_I dW_I \end{aligned}$$

$$\implies d\frac{I(t)M^r(t)}{M^n(t)} = \sigma_I dW_I$$

(b) Note that $P^r(t)$ will have the following dynamics (Theorem 22.4.1)

$$dP^r(t,T)/P^r(t,T) = (r(t) - \rho_{r,I}\sigma_I\sigma_r)dt - \sigma_r B(t,T)dW_r^n).$$

Therefore

$$
\begin{aligned}
& d(\ln \frac{I(t)P^r(t,T)}{M^n(t)}) \\
& = d\ln I(t) + d\ln P^r(t,T) - d\ln M^n(t) \\
& = (n(t) - r(t))dt - \frac{1}{2}\sigma_I^2 dt + \sigma_I dW_I + (r(t) - \rho_{r,I}\sigma_I\sigma_r)dt - \sigma_r B(t,T)dW_r - \frac{1}{2}\sigma_r^2 B^2(t,T)dt - n(t)dt \\
& = -\frac{1}{2}\sigma_I^2 dt - \frac{1}{2}\sigma_r^2 B^2(t,T)dt - \rho_{r,I}\sigma_I\sigma_r dt + \sigma_I dW_I - \sigma_r B(t,T)dW_r \\
\implies & d\frac{I(t)M^r(t)}{M^n(t)} = \sigma_I dW_I - \sigma_r B(t,T)dW_r
\end{aligned}
$$

(2) To find out the effects on dynamics when changing measure, we have (Theorem 15.6.17) □

19.16.3 Pricing examples

Lemma 19.16.1 (pricing simple inflation linked cash flows).

- *Consider an inflation linked zero-coupon bond with payment $I(T)$ at maturity T. Its current time t nominal value is defined by*

$$P^r(t,T) = M^n(t)E_{Q^n}[\frac{I(T)}{M^n(T)}|\mathcal{F}_t].$$

- *Consider a payment at time T given by*

$$N(\frac{I(T)}{I_0}) - 1),$$

where I_0 is a known reference inflation index. Its current time t nominal value is defined by

$$V(t) = N(\frac{I(t)}{I_0}P^r(t,T) - P^n(t,T)),$$

where $P^n(t,T)$ is the nominal price of a zero-coupon bond.

19.17 Notes on bibliography

Classical books on interest rate theory includes [3][1][17][10][18][19].

[11]

For comprehensive treatment on the LIBOR market, see [15].

For practical topics, e.g., calibration, hedging, and correlation modeling, see [20].

For convexity correction, see [21][22].

For yield curve construction, see [23].

For multi-curve framework, see[24][25].

Problems and solutions in mathematical finance: interest rates and inflation indexed derivatives

For inflation indexed derivatives, see [26].

BIBLIOGRAPHY

1. McInerney, D. & Zastawniak, T. *Stochastic Interest Rates* ISBN: 9781107002579. https://books.google.com/books?id=8JTwCAAAQBAJ (Cambridge University Press, 2015).
2. Corb, H. *Interest Rate Swaps and Other Derivatives* ISBN: 9780231530361. https://books.google.com/books?id=Gc0W0Y1CsDgC (Columbia University Press, 2012).
3. Brigo, D. & Mercurio, F. *Interest rate models-theory and practice: with smile, inflation and credit* (Springer Science & Business Media, 2007).
4. Wu, L. *Interest rate modeling: Theory and practice* (CRC Press, 2009).
5. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
6. Pelsser, A. *Efficient methods for valuing interest rate derivatives* (Springer Science & Business Media, 2000).
7. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
8. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
9. Privault, N. *An elementary introduction to stochastic interest rate modeling* (World scientific, 2012).
10. Andersen, L. & Piterbarg, V. *Interest Rate Modeling, Volume II Interest Rate Modeling* **v. 2.** ISBN: 9780984422111. https://books.google.com/books?id=tgmucQAACAAJ (Atlantic Financial Press, 2010).
11. Filipovic, D. *Term-Structure Models: A Graduate Course* ISBN: 9783540680154. https://books.google.com/books?id=KqcSh6CavaAC (Springer Berlin Heidelberg, 2009).
12. Gurrieri, S., Nakabayashi, M. & Wong, T. Calibration methods of Hull-White model (2009).
13. Piterbarg, V. Rates squared. *Risk* **22,** 100 (2009).
14. Cairns, A. J. *Interest rate models: an introduction* (Princeton University Press, 2004).
15. Rebonato, R., McKay, K. & White, R. *The SABR/LIBOR Market Model: Pricing, calibration and hedging for complex interest-rate derivatives* (John Wiley & Sons, 2011).
16. Hagan, P. K. & Kumar, D. Managing smile risk. *Wilmott Magazine*, 84–108.
17. Andersen, L. & Piterbarg, V. *Interest Rate Modeling, Volume I Interest Rate Modeling* **v. 1.** ISBN: 9780984422104. https://books.google.com/books?id=oI2fcQAACAAJ (Atlantic Financial Press, 2010).

18. Andersen, L. & Piterbarg, V. *Interest Rate Modeling, Volume III Interest Rate Modeling* **v. 3.** ISBN: 9780984422128. https://books.google.com/books?id=RxHVcQAACAAJ (Atlantic Financial Press, 2010).

19. Jarrow, R. A. *Modeling fixed-income securities and interest rate options* (Stanford University Press, 2002).

20. Rebonato, R. *Volatility and correlation: the perfect hedger and the fox* (John Wiley & Sons, 2005).

21. Pelsser, A. Mathematical foundation of convexity correction. *Quantitative Finance* **3,** 59–65 (2003).

22. Hagan, P. S. Convexity conundrums: Pricing CMS swaps, caps, and floors. *The Best of Wilmott*, 305 (2005).

23. Methods for constructing a yield curve. *WILMOTT Magazine, May*, 70–81 (2008).

24. Grbac, Z. & Runggaldier, W. J. *Interest rate modeling: post-crisis challenges and approaches* (Springer, 2015).

25. Henrard, M. *Interest rate modelling in the multi-curve framework: Foundations, evolution and implementation* (Springer, 2014).

26. Kerkhof, J. Inflation derivatives explained (2005).

20

INTEREST RATE MODELING IN MULTI-CURVE ENVIRONMENT

20.1 Foundations

20.2 Multiple curve framework modeling

20.2.1 Basics

20.2.1.1 *Concepts*

Remark 20.2.1 (LIBOR vs. OIS rate).

- London Interbank Offered Rate (LIBOR) is the average interest rate that banks charge each other for short-term, unsecured loans. The rate for different lending durations - from overnight to one-year - are published daily.
- LIBOR is designed to provide banks around the world with an accurate picture of short term funding cost for banks. Each day, several of the world's leading banks report what it would cost them to borrow from other lenders on the London interbank market. LIBOR is the average of these responses, it represents the equilibrium rates resulted from market forces.
- OIS rate
- Why OIS rate is free from credit risk Because the parties in a basic interest rate swap don't exchange principal, but rather the difference of the two interest streams, credit risk isn't a major factor in determining the OIS rate. During normal economic times, it's not a major influence on LIBOR, either. But we now know that this dynamic changes during times of turmoil, when different lenders begin to worry about each other's solvency.

Remark 20.2.2 (choice of LIBOR and OIS curve for discounting). [1, p. 209]

Finance theory leads to the conclusion that we should always use the best proxy available for the risk-free rate when discounting in situations where riskless portfolios have been set up. Arguably the OIS zero curve is as close to risk-free as we can get. It should therefore be used for discounting regardless of whether the transaction is collateralized

Following the credit crisis, most banks have changed their risk-free discount rates for collateralized transactions from LIBOR to what are known as overnight indexed swap (OIS) rates (see next section). But for non-collateralized transactions they continue to use LIBOR, or an even higher discount rate. (See Section 2.5 for a discussion of collateralization.) This reflects a belief that the discount rate used by a bank for a derivative should represent its average funding costs, not a true risk-free rate. The average funding costs for a non-collateralized derivative is considered to be at least as high as LIBOR.

For credit risk free portfolios, we should use OIS and account for the credit risk using addition CVA/DVA calculation. For cash instruments (e.g., bonds) where CVA/DVA is usually not calculated or applied during pricing, we should use LIBOR curve for discounting.

Can we prove the equivalence of CVA and discounting with spread

cite this âĂŸâĂŸLIBOR vs. OIS: The Derivatives Discounting Dilemma,âĂŹâĂŹ Journal of Investment Management

20.2.1.2 *Different curves*

Definition 20.2.1 (OIS rate payment, LIBOR payment).

Definition 20.2.2 (money market account in multi-curve environment).

Definition 20.2.3 (LIBOR zero curve, OIS zero curve, spread curve).

- *A **LIBOR forward curve** of fixed tenor Δ is denoted by*

$$F_L(t; T, T+\Delta).$$

- *The **OIS forward curve** is the riskless(no credit risk) and is denoted by*

$$F_{OIS}(t; T, T+\Delta).$$

- *The difference between a LIBOR forward curve and OIS forward curve is called OIS-LIBOR spread, denoted by $s = F_L(t; T, T+\Delta) - F_{OIS}(t; T, T+\Delta))$.*

Definition 20.2.4 (LIBOR forward curve, OIS forward curve, spread curve).

- *A **LIBOR forward curve** of fixed tenor Δ is denoted by*

$$F_L(t;T,T+\Delta).$$

- *The **OIS forward curve** is the riskless(no credit risk) and is denoted by*

$$F_{OIS}(t;T,T+\Delta).$$

- *The difference between a LIBOR forward curve and OIS forward curve is called OIS-LIBOR spread, denoted by $s = F_L(t;T,T+\Delta) - F_{OIS}(t;T,T+\Delta))$.*

Note 20.2.1.

- Before the 2007 credit crisis, LIBOR forward curves of different tenors approximately equals (by market) the OIS curve. LIBOR forwards rates of different tenors are **deterministically** related by Lemma 19.1.3: Let $t \leq T_1 < T_2 < T_3$. Then,

$$(T_3 - T_1)F(t;T_1,T_3) + 1 = [(T_2 - T_1)F(t;T_1,T_2) + 1][(T_3 - T_2)F(t;T_2,T_3) + 1].$$

  Therefore, the set of LIBOR forward curves and OIS curve can simply be represented by a single instantaneous forward curve.
- After the crisis, such relation does not hold since the market believes the longer term debts are associated with more credit risks.
- For example, a 6-month Libor forward curve has a spread over a 3-month Libro forward curve.

Lemma 20.2.1.

- *discounting using OIS curve*

$$NPV(t) = P(t,T)\tau L(S,T), \tau = T - S$$

.

20.2.1.3 *Construction of multiple curves*

Remark 20.2.3 (Calculating forward rates using OIS vs. not using OIS). [1, p. 205] Suppose that the 1-year LIBOR rate is 5% and the 2-year LIBOR-for-fixed swap rate with annual payments is 6%. Both rates are annually compounded.

- Suppose a bank uses LIBOR rates for discounting. The fixed leg of the swap has value

$$V_{fixed} = 6 \times DF(0,1Y) + 6 \times DF(0,2Y).$$

The floating leg of the swap has value

$$V_{float} = 5 \times DF(0,1Y) + \times DF(0,2Y)E_{T=2Y}[L(1Y,2Y)] = 5 \times DF(0,1Y) + DF(0,2Y) \times F(0,1Y,2Y)$$

Further use the fact that

$$DF(0,1Y) = \frac{1}{1+0.045}, F(0,1Y,2Y) = \frac{1}{2Y-1Y}(\frac{DF(0,2Y)}{DF(0,1Y)} - 1),$$

we can solve

$$F(0,1Y,2Y) = 7.0707\%.$$

- Suppose a bank uses OIS zero curve for discounting; the 1- and 2-year OIS zero rates are 4.5% and 5.5% with annual compounding. Then

$$DF(0,1Y) = \frac{1}{1+0.045}, DF(0,2Y) = \frac{1}{1+0.055}.$$

Similar to (1), we can obtain

$$F_{OIS}(0,1Y,2Y) = 7.0651\%.$$

20.2.1.4 *Instrument pricing using multiple curves*

Theorem 20.2.1 (fundamental relationship). *Let current time be $t, t < S < T$.*

- *Define OIS-LIBOR forward rate*

$$F(t;S,T) \triangleq E_T[L(S,T)|\mathcal{F}_t],$$

where $L(S,T)$ is LIBOR rate with accrual period from S to T, E_T is taking expectation with respect to the martingale measure associated with numeraire of OIS zero coupon bond $P(t,T)$.

- *The NPV of LIBOR payment $(T-S)L(S,T)$ is given by*

$$M(t)E_Q[\frac{(T-S)L(S,T)}{M(T)}|\mathcal{F}_t] = P(t,T)F(t;S,T)(T-S),$$

where Q is the risk-neutral measure associated with the numeraire of OIS money account $M(t)$.

Proof. Using Theorem 15.6.15, we have

$$
\begin{aligned}
&M(t)E_Q[\frac{(T-S)L(S,T)}{M(T)}|\mathcal{F}_t] \\
=&M(t)E_Q[\frac{(T-S)L(S,T)}{M(T)}\frac{M(T)P(t,T)}{M(t)P(T,T)}|\mathcal{F}_t] \\
=&P(t,T)E_Q[(T-S)L(S,T)|\mathcal{F}_t] \\
=&P(t,T)F(t;S,T)(T-S)
\end{aligned}
$$

□

Lemma 20.2.2 (interest rate swap rate using OIS curve). *Consider a set of dates $t < T_0 < T_1 < \cdots < T_n$, and denote $\tau_i = T_i - T_{i-1}$.*

- *The floating leg of an interest rate swap has NPV*

$$V_{flt} = \sum_{i=1}^{n} P(t,T_i)\tau_i F(t;T_{i-1},T_i),$$

  where $F(t;T_{i-1},T_i)$ is the OIS-LIBOR forward rate.
- *The fixed leg of an interest rate swap has NPV*

$$V_{fxd} = \sum_{i=1}^{n} P(t,T_i)\tau_i K.$$

- *The interest rate swap using OIS discounting, called OIS swap rate, is given by*

$$S_{0,n}(t) = \frac{\sum_{i=1}^{n}\tau_i F(t;T_{i-1},T_i)P(t,T_i)}{\sum_{i=1}^{n}\tau_i P(t,T_i)}.$$

- *$S_{0,n}(t)$ is a martingale under annuity measure associated with the default-free annuity*

$$A_{0,n}(t) = \sum_{i=1}^{n}\tau_i P(t,T_i).$$

Remark 20.2.4. OIS swap rate compared with LIBOR swap rate is given by

$$
\begin{aligned}
s^{OIS} &= \frac{\sum_{i=1}^{n}\tau_i F(t;T_{i-1},T_i)P(t,T_i)}{\sum_{i=1}^{n}\tau_i P(t,T_i)}, \\
s^{LIBOR} &= \frac{\sum_{i=1}^{n}\tau_i F(t;T_{i-1},T_i)P_L(t,T_i)}{\sum_{i=1}^{n}\tau_i P_L(t,T_i)} = \frac{\sum_{i=1}^{n}\tau_i F(t;T_{i-1},T_i)P(t,T_i)\rho_i}{\sum_{i=1}^{n}\tau_i P(t,T_i)\rho_i},
\end{aligned}
$$

where $\rho_i = P_L(t,T_i)/P(t,T_i)$.

Lemma 20.2.3 (Black model for cap/floor pricing with OIS discounting). *Consider a set of dates $t < T_0 < T_1 < \cdots < T_n$, and denote $\tau_i = T_i - T_{i-1}$. Let T_0 be the expiry time of a swaption associated with a swap. Assume $S(T_0, T_0, T_n)$ has log-normal distribution with volatility $\sigma_S\sqrt{T_0 - t}$ under annual measure. The NPV of a swaption with payoff*

$$A(T_0, T_0, T_n)(S_{0,n}(T_0) - K)^+$$

is given by

$$V_i(t, \sigma_i) = \tau_i P(t, T_i)(F(t; T_{i-1}, T_i)N(d_+) - KN(d_-)),$$

where

$$d_\pm = \frac{\ln(F(t; T_{i-1}, T_i)/K) \pm \frac{1}{2}\sigma_i^2 T_{i-1}}{\sigma\sqrt{T_{i-1}}}.$$

The

*The parameter σ_i that make $V(t) = V^{mkt}(t)$ is called **OIS-Cap-Floor implied volatility**.*

Proof. Using Theorem 15.6.15, we have

$$\begin{aligned}
&M(t)E_Q[\frac{A_{0,n}(T_0)(S_{0,n}(T_0) - K)^+}{M(T_0)}|\mathcal{F}_t] \\
&= M(t)E_Q[\frac{A_{0,n}(T_0)(S_{0,n}(T_0) - K)^+}{M(T)}\frac{M(T)A_{0,n}(t)}{M(t)A_{0,n}(T_0)}|\mathcal{F}_t] \\
&= A_{0,n}(t)E_A[(S_{0,n}(T_0) - K)^+|\mathcal{F}_t] \\
&= A_{0,n}(t)(S_{0,n}(t)N(d_+) - KN(d_-))
\end{aligned}$$

□

Lemma 20.2.4 (Black model for swaption pricing with OIS discounting). *Consider a set of dates $t < T_0 < T_1 < \cdots < T_n$, and denote $\tau_i = T_i - T_{i-1}$. Let T_0 be the expiry time of a swaption associated with a swap. Assume $S_{0,n}(T_0)$ has log-normal distribution with volatility $\sigma_S\sqrt{T_0 - t}$ under annual measure. The NPV of a swaption with payoff*

$$A(T_0, T_0, T_n)(S_{0,n}(T_0) - K)^+$$

is given by

$$V(t) = A(t, T_0, T_n)(S_{0,n}(t)N(d_+) - KN(d_-)),$$

where

$$d_\pm = \frac{\ln(S_{0,n}(t)/K) \pm \frac{1}{2}\sigma^2}{\sigma}.$$

The parameter σ_S that make $V(t) = V^{mkt}(t)$ is called ***OIS-swaption implied volatility****.*

Proof. Using Theorem 15.6.15, we have

$$\begin{aligned}
&M(t)E_Q[\frac{A_{0,n}(T_0)(S_{0,n}(T_0)-K)^+}{M(T_0)}|\mathcal{F}_t] \\
&= M(t)E_Q[\frac{A_{0,n}(T_0)(S_{0,n}(T_0)-K)^+}{M(T)}\frac{M(T)A_{0,n}(t)}{M(t)A_{0,n}(T_0)}|\mathcal{F}_t] \\
&= A_{0,n}(t)E_A[(S_{0,n}(T_0)-K)^+|\mathcal{F}_t] \\
&= A_{0,n}(t)(S_{0,n}(t)N(d_+)-KN(d_-))
\end{aligned}$$

□

Remark 20.2.5. The OIS swapation pricing is different from single-curve swaption pricing because $A(t, T_0, T_n)$ here is constructed using OIS discounting.

20.2.1.5 *Curve construction*

20.3 Notes on bibliography

For multi-curve framework, see[2][3].

BIBLIOGRAPHY

1. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
2. Grbac, Z. & Runggaldier, W. J. *Interest rate modeling: post-crisis challenges and approaches* (Springer, 2015).
3. Henrard, M. *Interest rate modelling in the multi-curve framework: Foundations, evolution and implementation* (Springer, 2014).

21

FIXED INCOME ANALYTICS & STRATEGIES

Overview

A fixed income security is an investment that provides a return in the form of fixed periodic payments and the eventual return of principal at maturity. Unlike a variable-income security, where payments change based on some underlying measure such as short-term interest rates, the payments of a fixed-income security are known in advance.

The most common type of fixed income instruments are bonds issued by corporates, government, and banks. Fixed income instruments are heavily used by issuers to raise funding for business expansion, government expenditure.

Fixed income instruments analytics have evolve from simple discounting from sophisticated interest rate and credit term structure modeling

We start this chapter with time value of money calculation. We will cover defaultable bond, and mortgage back securities

21.1 Bond markets and products

21.1.1 Vanilla bonds

21.1.1.1 *Fixed rate bonds*

Definition 21.1.1 (bond). *The bond is a debt security, under which the issuer owes the holders a debt and is obliged to pay them fixed amont of interest (termed, the coupon) at specified coupon dates and to repay the principal at the maturity date.*

Definition 21.1.2 (coupon rates: stated coupon rate and effective annual rate).

21.1.1.2 *Treasury bonds*

Definition 21.1.3 (Treasury bonds).

- *On-the-run Treasury bonds are the most recently issued U.S. Treasury bonds or notes of a particular maturity.*
- *Off-the-run Treasury bonds refer to Treasury securities that have been issued before the most recent issue and are still outstanding.*
- *A Treasury transitions from onthe-run to off-the-run once a newer set of Treasuries is released for sale.*
- *Because on-the-run issues are the most liquid, they typically trade at a slight premium and therefore yield a little less than their off-the-run counterparts.*

21.1.2 Floating rate note

Definition 21.1.4. *Floating rate notes (FRNs, floaters) have a variable coupon that is linked to a reference rate of interest, such as Libor or Euribor. For example, the coupon may be defined as three-month USD LIBOR + 0.20%. The coupon rate is resetted periodically, typically every one or three months.*

Remark 21.1.1 (interest rate risk characteristics of floating rate note)**.**

- Floating rate note usually has little exposure to interest rate risk. Consider a sudden rise in market interest rates happened just before one of the ïňĆoaterâĂŹs reset dates, the holder of the ïňĆoater would temporarily be receiving a coupon based on a below-market rate. However, at the next reset date, ïňĆoaterâĂŹs coupon would adjust to prevailing market rates. Such interest rate risk is also called reset risk.

21.1.3 Bonds with options

Definition 21.1.5 (callable bond). *[1, p. 34] A **callable bond** gives the issuer the right to redeem all or part of the bond before the specif ed maturity date.*

The available exercise styles include American call, European call, and Bermudian style call.

Remark 21.1.2 (motivation and intuition of callable bond). [1, p. 34]

- Te primary reason why issuers choose to issue callable bonds rather than non-callable bonds is to protect themselves against a decline in interest rates. If market interest rates fall or credit quality of the issuer improves, the issuer of a callable bond has the right to replace an old, expensive bond issue with a new, cheaper bond issue.
- For example, assume that the market interest rate was 6% at the time of issuance and that a company issued a bond with yield 7%. Now assume that the market interest rate has fallen to 4% and that the companyâĂŹs creditworthiness has not changed; it can still issue at the market interest rate plus 100 bps. If the original bond is callable, the company can redeem it and replace it with a new bond paying 5% annually.
- Callable bonds present investors with a higher level of reinvestment risk than non-callable bonds; that is, if the bonds are called, bondholders have to reinvest funds in a lower interest rate environment. For this reason, callable bonds have to offer at a higher yield and sell at a lower price than otherwise similar non-callable bonds.

Example 21.1.1. [1, p. 36] Consider a 30 Y callable bond(face value 100) issued on Aug 2012 at a price of 98.195. The bond is callable according to the following schedule.

Year	Call Price	Year	Call Price
2022	103.870	2028	101.548
2023	103.485	2029	101.161
2024	103	2030	100.774
2025	102.709	2031	100.387
2026	102.322	$\geq$2032	100
2027	101.955		

We can see that

- The call protection is upto 2022.
- The bond is sold at discount. The call price is always greater than the par in order to protect a decline of interest rate(in such case, the bond price can exceed 100).

Definition 21.1.6 (putable bond). *[1, p. 36] A* ***putable bond*** *gives the bondholders the right to sell the bond back to the issuer at a pre-determined price on specified dates.*

Remark 21.1.3 (motivation and intuition of putable bond)**.** [1, p. 36]

- Putable bonds are beneficial for the bondholder by guaranteeing a pre-specified selling price at the redemption dates.
- If interest rates rise after this issue date, thus depressing the bondâĂŹs price, the bondholders can put the bond back to the issuer and get cash. This cash can be reinvested in bonds that offer higher interest.
- The price of a putable bond will be higher than the price of an otherwise similar bond issued without the put provision. Similarly, the yield on a bond with a put provision will be lower than the yield on an otherwise similar non-putable bond.

Definition 21.1.7 (convertible bond). *[1, p. 37] A* ***convertible bond*** *is a hybrid security with both debt and equity features. It gives the bondholder the right to exchange the bond for a specified number of common shares in the issuing company. Thus, a convertible bond can be viewed as the combination of a straight bond (option-free bond) plus an embedded equity call option.*

Remark 21.1.4 (motivation and intuition of convertible bond)**.** [1, p. 37]

- From the investor's perspective, a convertible bond offers several advantages relative to a non-convertible bond.
  - it gives the bondholder the ability to convert into equity in case of share price appreciation. At the same time, the bondholder receives downside protection; if the share price does not appreciate, the convertible bond offers the comfort of regular coupon payments and the promise of principal repayment at maturity.
  - Even if the share price decline, the price of a convertible bond cannot fall below the price of the straight bond because it contains optionality.
- Because the conversion provision is valuable to bondholders, the price of a convertible bond is higher than the price of an otherwise similar bond without the conversion provision. Similarly, the yield on a convertible bond is lower than the yield on an otherwise similar non-convertible bond.
- From the issuer's perspective, convertible bonds offer two main advantages.
  - The first is reduced interest expense. Issuers are usually able to offer below-market coupon rates because of investors' attraction to the conversion feature.

- The second advantage is the elimination of debt if the conversion option is exercised. But the conversion option is dilutive to existing shareholders

21.2 Cash flow mathematics: growth and discounting

21.2.1 Compounding convention and future values

Definition 21.2.1. *[2, p. 592][3, p. 9] Given a annualized interest rate R, the interest rate calculation depends on the compounding convention.*

- *If the compounding convention is continuous, then unit notional in n years will grow to*
$$\exp(Rn).$$
- *If the compounding convention is once per year, then unit notional in n years will grow to*
$$(1+R)^n.$$
- *If the compounding convention is m times per year, then unit notional in n years will grow to*
$$(1+R/m)^{mn}.$$

Lemma 21.2.1 (conversion). *[2, p. 592] Let R_c be the annual continuous compounding rate. Let R_m be the annual compounding rate with frequency m. Then we have*

$$R_c = m\ln(1+\frac{R_m}{m}),$$

and

$$R_m = m(\exp(R_c/m)-1).$$

Proof. Consider unit notional, we have

$$\exp(R_c n) = (1+R_m/m)^{mn}.$$

□

Example 21.2.1. [3, p. 9] Consider a bank offers a CD with a two year maturity($N = 2$), a stated annual interest rate of 8 percent compounded quarterly ($m = 4$), and a feature allowing reinvestment of the interest at the same interest rate. Let the notional be 1$, The value of the CD at maturity is given by

$$V = (1+\frac{r}{m})^{mN} = (1.02)^8 = 1.171659.$$

21.2.2 Present value of cash flow

Definition 21.2.2 (present value of a single cash flow in future). *[3, p. 15]*

- *Given a future cash flow FV that is to be received in N periods and an interest rate per period of r, its present value is given by*

$$PV = FV\frac{1}{(1+r)^N}.$$

- *Suppose the compounding frequency is m per year, and the annual interest rate is r. Then a future cash flow FV N years later has present value given by*

$$PV = FV\frac{1}{(1+\frac{r}{m})^{Nm}}.$$

Remark 21.2.1 (present value calculation depends on the convention). Note that different compounding convention can lead to different present value calculation.

Lemma 21.2.2 (present value of annuity and perpetuity). *Consider an annuity with N payments of A occurs on period indexed from $t = 1$ to $t = N$. Let current time be $t = 0$. Then the present value of the annuity is given by*

$$\begin{aligned} PV &= \frac{A}{(1+r)} + \frac{A}{(1+r)^2} + \cdots + \frac{A}{(1+r)^N} \\ &= \frac{A}{r}(1 - \frac{1}{(1+r)^N}) \end{aligned}$$

Particularly for a perpetuity(in which $N \to \infty$, its present value is given by

$$PV = \frac{A}{r}.$$

Proof. Note that from the summation formula for geometric series

$$\begin{aligned} PV &= \frac{A}{(1+r)} + \frac{A}{(1+r)^2} + \cdots + \frac{A}{(1+r)^N} \\ &= \frac{A}{1+r}(\frac{1-\frac{1}{(1+r)^N}}{1-\frac{1}{1+r}}) \\ &= \frac{A}{1+r}(1-\frac{1}{(1+r)^N})\frac{1+r}{r} \\ &= \frac{A}{r}(1 - \frac{1}{(1+r)^N}) \end{aligned}$$

□

21.2.3 Different interest rates

Definition 21.2.3 (repurchase agreement, repo, repo rate). *[4, p. 171][5, p. 79] A repurchase agreement(repo) is an agreement by an owner of securities(usually bonds) to sell the securities to investors(or other counter-parties) at one price and buy them back at another price(usually higher) in a future date.*

*Usually in repo, it requires collateral for borrowing. Therefore, the interest rate implied by the repo price is riskless interest rate,called **repo rate**.*

Definition 21.2.4 (Treasury rate). *[2, p. 178] Treasury rates are the interest rates an investor earns on Treasury bills and Treasury bonds; These are the instruments used by a government to borrow in its own money.*

Remark 21.2.2 (more about Treasury rates)**.** [2, p. 178]

- It is usually assumed that there is no chance that a government will default on a debt denominated its own currency; therefore Treasury rates are usually regarded as risk-free rates.
- In practice, they are regarded as artificially low because:
  - the amount of capital a bank is required to hold to support an investment in Treasury bills and bonds is substantially lower than the capital requirement for a similar investment.
  - In the US, Treasury instruments are given a favorable tax treatment compared with other fixed income investments.

Definition 21.2.5 (overnight indexed swap(OIS), OIS rate). *[5, p. 202] An **overnight indexed swap** is a swap where a fixed rate for a period(e.g., 1 month or 3 months) is exchanged for the geometric average of the overnight unsecured borrowing rate between financial institutions in the US.*

*The fixed rate in an OIS is referred to as the **OIS rate**.*

Remark 21.2.3 (LIBOR, Red Fund rate, repo rate)**.** [5, pp. 79, 201]

- Both LIBOR and fed fund rate are unsecured borrowing rates. On average, overnight LiBOR rate has been about 6 basis point higher than the fed fund rate.
- The repo rate is usually slightly below fed fund rate.
- Interest rate implied from US Treasury bill is considered artificially low. It is the lowest.

- After the 2007 crisis, OIS rate is used for discounting when valuing collateralized derivatives; LIBOR rate is used for discounting when valuing non-collateralized derivatives. OIS rate now is viewed as risk-free rate and it is lower than LIBOR.

21.3 Bond valuation

21.3.1 Valuation via single rate discounting

21.3.1.1 *Basics*

Definition 21.3.1 (market discount rate, required yield, required rate of return). *[1, p. 92] The **market discount rate** is used in the time-value-of-money calculation to obtain the present value. The market discount rate is the rate of return required by investors or market as a whole given the risk of the investment in the bond. It is also called the **required yield, or the required rate of return.***

Definition 21.3.2 (valuation of bond using single rate model). *Let current time be* 0*. Let* $t_1, t_2, ..., t_n$ *be future times when fixed cash flow* $c_1, c_2, ..., c_n$ *of the bond occur . Let* y *be the market required yield for the bond observed in the market. Then*

- *the price for a continuously-compounding bond is*

$$P = \sum_{i=1}^{n} c_i e^{-y t_i}.$$

- *the price* y *for an annually simple compounding bond is*

$$P = \sum_{i=1}^{n} c_i \frac{1}{(1+y)^{t_i}}.$$

- *the price for a bond with simple compounding* m *times per year is*

$$P = \sum_{i=1}^{n} c_i \frac{1}{(1+y/m)^{m t_i}}.$$

Remark 21.3.1 (Issues with single discount rate model).

- **(Not taking into account term structure of interest rate)** Note that the single discount rate model is problematic since does not take into account term structure of interest rate; that is, we are implicitly assuming the forward rate is flat (or equivalently, the short rate is constant.). In practice, we would expect different interest rates for different investment horizons.
- **(Pricing depends on compounding convention)**. Note that the compounding convention is not in the bond's contract parameters.

Definition 21.3.3 (implied bond yield-to-maturity). *Let current time be* 0. *Let* $t_1, t_2, ..., t_n$ *be future times when fixed cash flow* $c_1, c_2, ..., c_n$ *of the bond occur .* ***Let*** P ***be the current price of the bond observed on the market****. Then*

- *the* ***yield-to-maturity*** *for a continuously-compounding bond is a number* y *such that*

$$P = \sum_{i=1}^{n} c_i e^{-y t_i}.$$

- *the* ***yield-to-maturity*** *for an annually simple compounding bond is a number* y *such that*

$$P = \sum_{i=1}^{n} c_i \frac{1}{(1+y)^{t_i}}.$$

- *the* ***yield-to-maturity*** *for a bond with simple compounding* m *times per year is a number* y *such that*

$$P = \sum_{i=1}^{n} c_i \frac{1}{(1+y/m)^{m t_i}}.$$

The bond yield-to-maturity is the implied market discount rate and depends on the bond coupon payment frequency.

Example 21.3.1. [1, p. 97] Consider a bond with a 4Y tenor, a par face 100, a coupon rate 3.5%, and 4 annual coupon payments. Suppose we observe the market price to be 103.75. Then

$$103.75 = \frac{3.5}{1+y} + \frac{3.5}{1+y} + \frac{3.5}{1+y} + \frac{3.5}{1+y} \implies y = 0.02503.$$

Note 21.3.1 (caution! bond yield-to-maturity vs. yield curve). [6, p. 102]

- Only for zero-coupon bond, its yield-to-maturity for different maturities is equivalent to the yield curve.
- For any other coupon bonds, their yield-to-maturity for different maturities is usually larger than the yield curve; the difference is due to the intermediate cash flow.

Remark 21.3.2 (bond yield vs. zero-coupon bond price). For a bond with single payment c occurring at T, we have

$$y = \frac{1}{T} \ln \frac{P(0,T)}{c},$$

where $P(0,T)$ is the zero-coupon bond(maturing at T) price at time 0.

Note 21.3.2 (interpret yield-to-maturity). [1, p. 96] Te yield-to-maturity is the rate of return on the bond to an investor given three critical assumptions:

- The investor holds the bond to maturity.If the bond holder does not hold the bond to maturity, and sell the bond before it matures, then the return might be negative since the bond can depreciates due to credit worsening,interest rate rises, etc.
- The issuer does not default on any of the payments.
- The investor is able to reinvest coupon payments at that same yield. For example(see below table), consider a 2Y coupon bond with coupon payment at 1Y. If we reinvest the coupon at return rate y, then at $T = 2Y$, we will get a cash flow $C(1+y) + N + C$, whose present value is

$$PV = \frac{1}{(1+y)^2}(C(1+y) + N + C) = \frac{C}{1+y} + \frac{N+C}{(1+y)^2} = P.$$

	action	cash flow
T=0	buy the coupon bond at price P	-P
T=1	receive the coupon and reinvest the coupon	+C - C
T=2	receive the final coupon, face value and the proceeds of the reinvestment	C(1+y) + N + C

However, in reality, not all the investors will hold the bond to maturity and reinvest the coupons(even reinvest the coupon at the same rate might not be possible since the rate is changing). Therefore, the yield-to-maturity does not necessarily reflect the rate of return in investing a bond.

Theorem 21.3.1 (No arbitrage condition for bond yield-to-maturity). *Assuming*

- *Bonds are defaultable.*
- *All investors will reinvest them coupons.*
- *The forward rate is flat such that investor can reinvest at the same rate.*
- *The market is free of arbitrage.*

- *For all default-free bonds(for any coupon schedule it has) maturing at T, its yield-to-maturity must equal the market discount rate or required yield corresponding to that maturity T.*
- *If holding and reinvesting until maturity, then investing in any coupon-bearing bond is equivalent to investing in zero-coupon bond with the same maturity. In other*

words, given the same initial cash, buying zero-coupon bond and the coupon-bearing bond will have the same rate of return. [a]

[a] Note that coupon-bearing bond will be more expensively than the zero-coupon bond.

Proof. (1) Consider two default-free bonds have the same maturity date T. The first bond has the market required yield, whereas the second bond has the higher yield-to-maturity than the required yield. Then we can short the first bond(or borrow money at the market required yield) and long and hold the second bond to maturity. At maturity, the cash flow we get from the second bond(including reinvesting coupon received) will payoff the borrowed money and gain extra money. See also 21.3.2. (2)

□

Lemma 21.3.1 (bond price and coupon rate relationship). *Consider a coupon bond with coupon rate c,maturity T and face value N. Let the market required yield be $y(T)$. Denote the bond price by P. Then*

- *If $c = y(T)$, then $P = N$. That is, the bond is* ***sold at par.***
- *If $c < y(T)$, then $P < N$. That is, the bond is* ***sold at discount.***
- *If $c > y(T)$, then $P > N$. That is, the bond is* ***sold at premium.***

Proof. □

Example 21.3.2. [1, p. 93]

- Consider a 5 Y bond. The annual coupon rate is 4%, the yield is 6%, and the face value is 100. The bond price is sold at 91.175 because

$$\frac{4}{1.06} + \frac{4}{1.06^2} + \frac{4}{1.06^3} + \frac{4}{1.06^4} + \frac{104}{1.06^5} = 91.575.$$

- Consider a 5 Y bond. The annual coupon rate is 8%, the yield is 8%, and the face value is 100. The bond price is sold at 91.175 because

$$\frac{8}{1.06} + \frac{8}{1.06^2} + \frac{8}{1.06^3} + \frac{8}{1.06^4} + \frac{108}{1.06^5} = 108.425.$$

- Consider a 5 Y bond. The annual coupon rate is 6%, the yield is 6%, and the face value is 100. The bond price is sold at 100 because

$$\frac{6}{1.06} + \frac{6}{1.06^2} + \frac{6}{1.06^3} + \frac{6}{1.06^4} + \frac{106}{1.06^5} = 100.$$

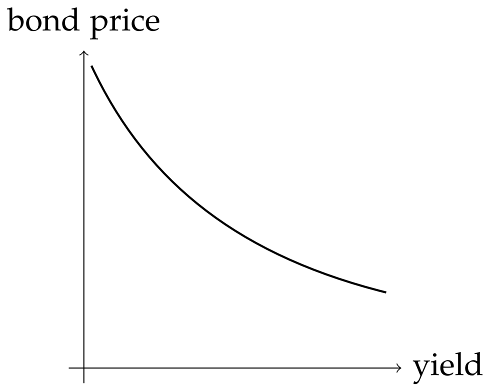


Figure 21.3.1: general price yield relationship. Note its convexity.

21.3.1.2 *Qualitative treatment on risks*

Remark 21.3.3.

- An investor who holds a fixed-rate bond to maturity will earn an annualized rate of return equal to the YTM of the bond when purchased.
- An investor who sells a bond prior to maturity will earn a rate of return equal to the YTM at purchase if the YTM at sale has not changed since purchase.
- If the market YTM for the bond, our assumed reinvestment rate, increases (decreases) after the bond is purchased but before the first coupon date, abuy-and-hold investorâĂŹs realized return will be higher (lower) than the YTM of the bond when purchased.
- If the market YTM for the bond, our assumed reinvestment rate, increases after the bond is purchased but before the first coupon date, a bond investor will earn a rate of return that is lower than the YTM at bond purchase if the bond is held for a short period.
- If the market YTM for the bond, our assumed reinvestment rate, decreases after the bond is purchased but before the first coupon date, a bond investor will earn a rate of return that is lower than the YTM at bond purchase if the bond is held for a long period

21.3.2 Valuation via zero-coupon bond curve

Definition 21.3.4 (valuation of bond using zero-coupon bond curve). *Let current time be* 0. *Let* $t_1, t_2, ..., t_n$ *be future times when fixed cash flow* $c_1, c_2, ..., c_n$ *of the bond* *occur. Let $DF(0,t)$ be the market observed zero-coupon bond curve. Then the price for a continuously-compounding bond is*

$$P = \sum_{i=1}^{n} c_i DF(0, t_i).$$

21.3.3 Dirty and clean price convention

Definition 21.3.5 (dirty and clean price, accrued interest). *Consider a set of coupon payment dates $t_0 < t_1 < ... < t_N$. Let current time be $t_S, t_0 \leq t_S < t_1$, and the bond coupon rate be C.*

- ***Accrued interest*** *is defined as*

$$AI(t_S) = \begin{cases} (t_S - t_0) \times C, t_0 \leq t_S < t_1 \\ 0, t_S = t_1 \ (after\ coupon\ payment) \end{cases}$$

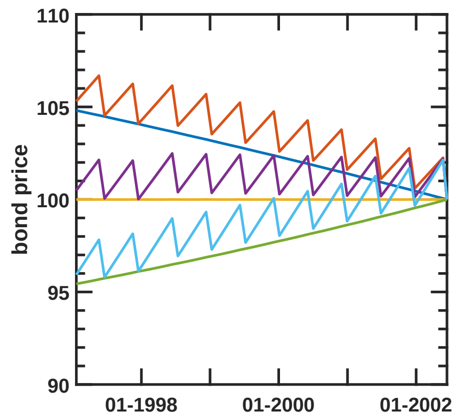
- ***Dirty price*** *of a bond is the price of a bond including any interest that has accrued since issue of the most recent coupon payment.*
  - *In the single rate discounting model,*

$$PV_{dirty} =$$

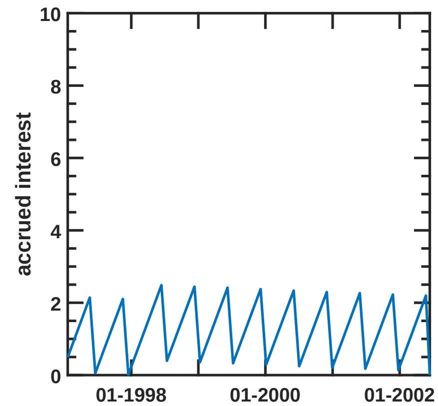
  - *In the curve discounting model,*
- ***Clean price*** *of a bond is the price of a bond excluding the accrued interest.* ***Clean prices are usually quoted.***
- $$dirty\ price \ = \ clean\ price \ + \ accrued\ interest$$

Remark 21.3.4 (understand dirty price and clean price)**.**

- Dirty price is the NPV of future cash flows; it is the fair market price.
- Clean price is not the NPV of future cash flows; it is not the fair market price.
- Clean price is solely for quoting purpose. Dirty price is suitable for quoting due to its jigsaw-like pattern (see Figure 21.3.2).
- Dirty price is always greater than clean price.



(a) Bond dirty (with saw-tooth pattern) and clean (smooth) prices trajectories for bonds sold at premium, at par and at discount. The bond coupon rate is fixed and the market required yield is varied.



(b) Bond accured interst as a function of time. Immediately after coupon payment date, it will drop to zero. Accured interst only depends on day count and coupon rate

Figure 21.3.2: Bond clean price, dirty price, and accrued interest.

Example 21.3.3. [1, p. 107] A 6% German corporate bond is priced for settlement on 18 June 2015. The bond makes semiannual coupon payments on 19 March and 19 September of each year and matures on 19 September 2026. The corporate bond uses the 30/360 day-count convention for accrued interest. Suppose the stated annual yield-to-maturity is 5.8%.

Note that there are 23 periodic payments starting from Sep-19-2015 to Sep-19-2026. Also there are 89 days between Mar-19 and Jun-18, or a semi-year fraction of 89/180.

Therefore, the dirty price is

$$PV_{dirty} = (1.029)^{89/180} \times (\frac{3}{1.029} + \frac{3}{1.029^2}) + \cdots + \frac{103}{1.029^{23}} = 103.109;$$

and the accrued interest is

$$AI = 3 \times \frac{89}{180} = 1.4833.$$

21.3.4 Discussion on different approaches

21.3.5 Forward bond price

Theorem 21.3.2 (No-arbitrage forward price for bonds).

- *Consider a zero-coupon coupon has maturity T and spot price $Z(t,T)$. Its forward price with delivery date τ is given by*

$$FB(t,T) = \frac{Z(t,T)}{Z(t,\tau)};$$

if in multi-curve environment, then

$$FB(t,T) = Z(t,T)(1 + F(t,t,\tau)(\tau - t)),$$

where $F(t,t,\tau)$ is the spot LIBOR rate with tenor $[t,\tau]$.
- *Consider a bond with coupon payment c_i at $T_i, i = 1,2,...,N$. The maturity date is $T = T_N$. Let S_t be its* ***spot dirty price*** *its* ***forward dirty price*** *with delivery date $\tau < T$ is given by*

$$F_t = \frac{S_t - \sum_{i=1}^{k} c_i Z(t,T_i)}{Z(t,\tau)},$$

where $T_1 < T_2 < \cdots < T_k < \tau$; If in multi-curve environment,

$$F_t = S_t + \underbrace{S_t(\tau - t)F(t,t,\tau)}_{\text{funding cost}} - \underbrace{\sum_{i=1}^{k} c_i(1 + F(t,T_i,\tau)(\tau - T_i)}_{\textbf{benefits}},$$

where $F(t,T_i,\tau)$ is the forward rate with tenor $[T_i,\tau]$.
- *Let AI denote the accrued interest. Let S_t, F_t denote the clean spot and forward prices. They are given by*

$$F_t = \frac{S_t + AI(t) - \sum_{i=1}^{k} c_i Z(t,T_i)}{Z(t,\tau)} - AI(\tau),$$

and

$$F_t = S_t + AI(t) + S_t(\tau - t)F(t,t,\tau) - \sum_{i=1}^{k} c_i(1 + F(t,T_i,\tau)(\tau - T_i) - AI(\tau).$$

Proof. See Theorem 19.3.3. □

21.4 Bond risks

21.4.1 Time path of bond price

21.4.1.1 *Time path with yield curve unchanged*

Lemma 21.4.1.

- *Consider a zero-coupon bond with fixed maturity T. Assume constant short rate r. Then its time path is given by*

$$P(t;T) = \exp(-r(T-t)).$$

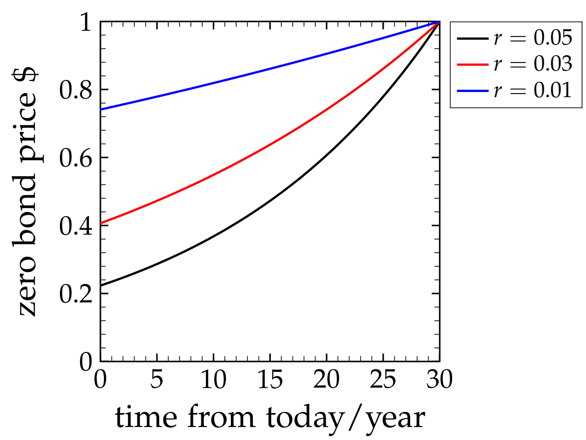


Figure 21.4.1: 30 Y zero-coupon bond time path at different continuous compounding rate r

21.4.1.2 *Time path with changing yield curve*

Methodology 21.4.1 (simulate the future term structure in Hull-White model).

- *Suppose*
  - *current time is 0.*
  - *we are given the current term structure $y(0,T), T \in \mathbb{R}^+$(or $P(0,T), T \in \mathbb{R}^+$, or $f(0,t), \mathbb{R}^+$).*
  - *we are given a calibrated Hull-White model*

$$dr = (\theta(t) - \alpha r(t))dt + \sigma(t)dW_t.$$

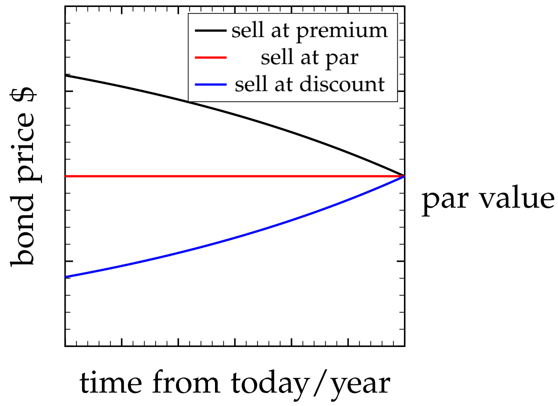


Figure 21.4.2: 30 Y zero-coupon bond time path at different continuous compounding rate r

- *we are given the initial short rate $r(0)$(which can be derived from $r(0) = f(0,0)$).*
- *Then we can generate a* ***sample term structure*** *$y(t,T)$ and the associated* ***sample bond price*** *$V(t)$ in future time $t > 0$ via the following procedure:*
  - *simulate $r(t)$ by drawing from the normal distribution*

$$r(t) \sim N(m(t), v^2(t)),$$

where

$$m(t) = r(0)e^{-\alpha t} + \int_0^t \theta(u)e^{-\alpha(t-u)}du,$$

$$v^2 = \int_0^t \sigma(u)^2 e^{-2\alpha(t-u)}du.$$

  - *the sample yield curve is given by*

$$y(t,T) = -\frac{\ln P(t,T)}{T-t} = -\frac{1}{T-t}(\ln \frac{P(0,T)}{P(0,t)} - A(t,T) - r(t)D(t,T)),$$

where

$$A(t,T) = -f(0,t)D(t,T) + \frac{1}{2}D^2(t,T)\int_0^t \sigma^2(u)\exp(-2\alpha(t-u))du,$$

$$D(t,T) = \int_t^T \exp(-\alpha(T-s))ds.$$

- *calculate the sample bond price $V(t)$ via*

$$V(t) = \sum_{i=1}^{m} \mathbf{1}(t_i > t) c_i DF(t, t_i),$$

where

$$DF(t, t_i) = \exp(-y(t, t_i)(t_i - t)).$$

- *Let $V^{(i)}(t), i = 1, 2, ..., N$ denote the N generated samples. Then we can use these samples to get the distribution of $V(t)$. Note that $V(t)$ is a random variable.*

Proof. (1) For the future term structure generation method via Hull-White model, see 19.5.5. (2) Note that from risk-neutral pricing theory

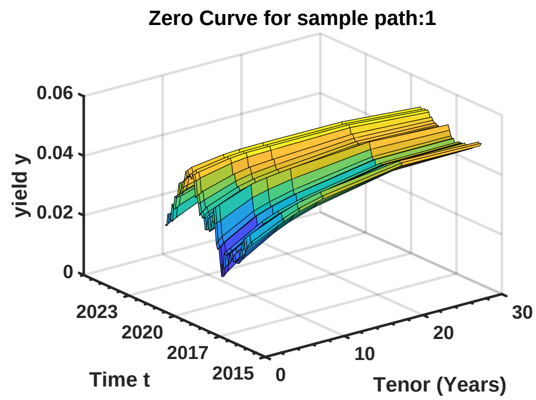
$$\begin{aligned} V(t) &= E_Q[\sum_{i=1}^{m} \mathbf{1}(t_i > t) c_i \exp(-\int_t^{t_i} r(s)ds) | \mathcal{F}_t] \\ &= \sum_{i=1}^{m} \mathbf{1}(t_i > t) c_i E_Q[\exp(-\int_t^{t_i} r(s)ds) | \mathcal{F}_t] \\ &= \sum_{i=1}^{m} \mathbf{1}(t_i > t) c_i DF(t, t_i) \end{aligned}$$

where by definition

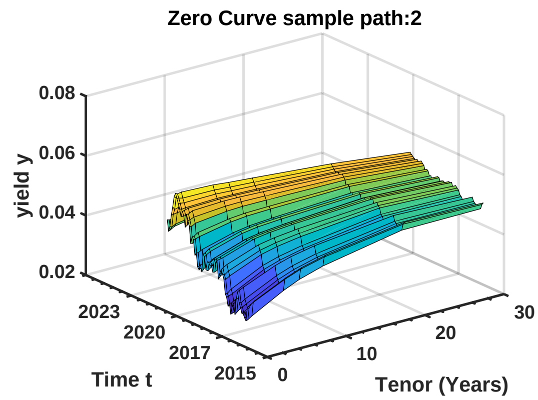
$$DF(t, t_i) = \exp(-y(t, t_i)(t_i - t)) = E_Q[\exp(-\int_t^{t_i} r(s)ds) | \mathcal{F}_t].$$

is a random variable. □

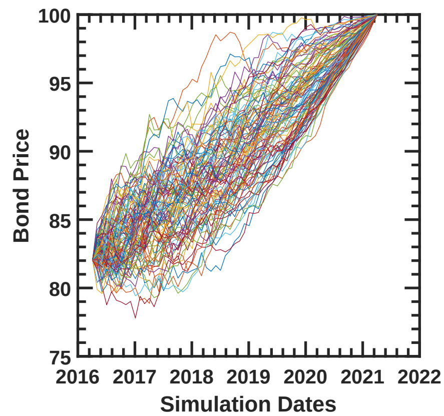
Example 21.4.1. As showed in Figure 21.4.3c, we use HW model to simulate the yield curve evolution, and then obtain the time path of 5Y bond with evolving yield curves.



(a) The future yield curve in one simulation scenario.



(b) The future yield curve in another simulation scenario.



(c) Time path of a 5Y bond with simulating yield curve.

Figure 21.4.3: Bond dynamics.

21.4.2 Risk analytics in single rate model

21.4.2.1 *Duration*

Definition 21.4.1 (duration, Macaulay's duration). *[5, p. 95][2, p. 182] Given a bond P and the bond yield y such that*

$$P = \sum_{i=1}^{n} c_i e^{-y t_i}$$

where $c_i, i = 1, ..., n$ are cash flows at time $t_i, i = 1, ..., n$.

- *The **duration** of the bond is defined as*

$$D = -\frac{dP}{dy}\frac{1}{P} = \frac{\sum_{i=1}^{n} c_i t_i e^{-yt_i}}{\sum_{i=1}^{n} c_i e^{-yt_i}}.$$

- *An alternative definition of duration is*

$$D = \sum_{i=1}^{n} t_i(\frac{v_i}{P}),$$

where $v_i = c_i e^{-t_i y}$;that is, duration can be viewed as a weighted average of the times when payments are made.

- *The **dollar duration**of the bond is defined as*

$$D = -\frac{dP}{dy} = \sum_{i=1}^{n} c_i t_i e^{-yt_i}.$$

Definition 21.4.2 (modified duration). *[2, p. 182] Given a bond P and the bond yield y such that*

$$P = \sum_{i=1}^{n} c_i (1 + \frac{y}{m})^{-mt_i}$$

where $c_i, i = 1, ..., n$ are cash flows at time $t_i, i = 1, ..., n$.

- *The **duration** of the bond is defined as*

$$D = -\frac{dP}{dy}\frac{1}{P} = \frac{\sum_{i=1}^{n} c_i t_i (1 + \frac{y}{m})^{-mt_i}(1 + \frac{y}{m})^{-1}}{P}.$$

- *An **modified duration** is defined as*

$$D_m = \sum_{i=1}^{n} t_i(\frac{v_i}{P}),$$

where $v_i = c_i(1 + \frac{y}{m})^{-mt_i}$;that is, duration can be viewed as a weighted average of the times when payments are made.

- *The modified duration and duration are connected by $D_m = D(1 + y/m)$.*

Remark 21.4.1. When m approaches ∞, duration and modified duration are equivalent.

Lemma 21.4.2. *Intuitively, **duration** measures how long, on average, a bondholder must wait to receive cash payments.*

- *A zero-coupon bond maturing in T periods has a duration of T;*

- *A coupon bond maturing in T periods has duration T since some payments are received before.*
- *The more coupon received before maturing, the smaller the duration.*

Lemma 21.4.3 (basic properties of duration).

- *Given y and T, duration is higher when C is lower.*
- *Given C and T, duration is lower when y is higher.*
- *Given C and y, duration is generally higher when T is higher (except for some deep discount coupon bonds).*
- *The duration of a portfolio P containing M bonds is*

$$D_P = \sum_{i=1}^{M} w_i D_i,$$

where w_i is the percentage weight of bond i in P.

Example 21.4.2 (weighted summation property). Consider the following three bonds with properties listed in the table.

tenor	**yield,%**	**duration**	**convexity**
2Y	7.71	1.78	0.041
5Y	8.35	3.96	0.195
10Y	8.84	6.54	0.568

Then a portfolio of \$5.4M of 2-year note and \$4.6M of the 10 year note will have duration

$$\frac{5.4 \times 1.78 + 4.6 \times 6.54}{10} = 3.96.$$

Lemma 21.4.4.

- *zero coupon bond*

$$D = T.$$

- *Flat perpetuity*

$$D = \frac{1}{y} + 1.$$

- *Flat annuity*

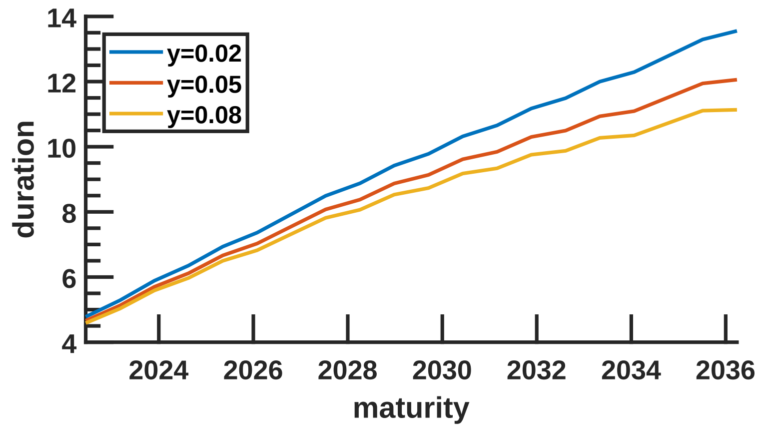
$$D = \frac{1}{y} + 1 - \frac{T}{(1+y)^T - 1}.$$

- *Coupon bond*

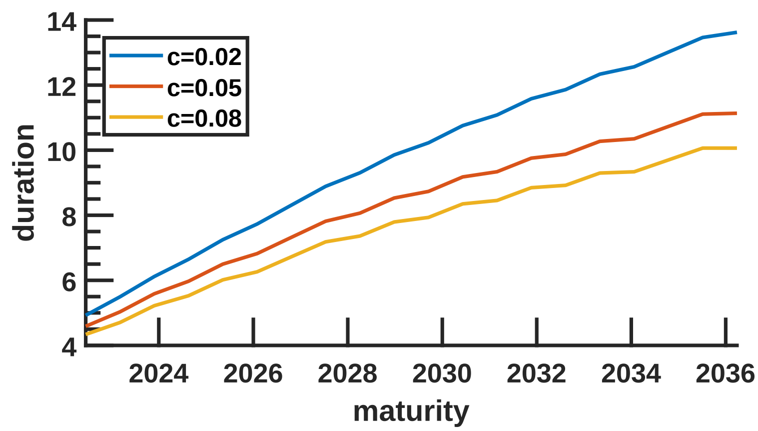
$$D = \frac{1}{y} + 1 - \frac{(1+y) + T(c-y)}{c[(1+y)^T - 1] + y}.$$

- *Coupon bond selling at par*

$$D = \frac{1+y}{y}(1 - \frac{1}{(1+y)^T}).$$



(a) Effects of yield and maturity on duration of a bond



(b) Effects of coupon rate and maturity on duration of a bond

Figure 21.4.4: Effects of coupon rate, yield, and maturity on the duration of a bond

21.4.2.2 *Convexity*

Definition 21.4.3 (convexity). *[5, p. 95] Given a bond P and the bond yield y such that*

$$P = \sum_{i=1}^{n} c_i e^{-yt_i}$$

where $c_i, i = 1, ..., n$ are cash flows at time $t_i, i = 1, ..., n$.

- *The **convexity** of the bond is defined as*

$$C = \frac{d^2 P}{dy^2} \frac{1}{P}.$$

- *The **dollar convexity** of the bond is defined as*

$$C = \frac{d^2 P}{dy^2}.$$

Lemma 21.4.5 (basic properties of convexity).

- *Convexity is always positive.*
- *A zero-coupon bond maturing in T periods has a duration of T;*
- *Given y and T, convexity is higher when C is lower.*
- *Given y and D, convexity is lower when C is lower.*
- *The convexity of a portfolio P containing M bonds is*

$$C_P = \sum_{i=1}^{M} w_i C_i,$$

where w_i is the percentage weight of bond i in P.

Example 21.4.3 (weighted summation property). Consider the following three bonds with properties listed in the table.

tenor	**yield,%**	**duration**	**convexity**
2Y	7.71	1.78	0.041
5Y	8.35	3.96	0.195
10Y	8.84	6.54	0.568

Then a portfolio of \$5.4M of 2-year note and \$4.6M of the 10 year note will have convexity

$$\frac{5.4 \times 0.041 + 4.6 \times 0.568}{10} = 0.283.$$

21.4.2.3 *Duration based hedging*

Lemma 21.4.6 (bond price changes due to yield change). *A change in yield Δy will lead to approximate change of ΔP given by ,*

$$\Delta P = \frac{dP}{dy}\Delta y + \frac{1}{2}\frac{d^2P}{dy^2}\Delta y^2$$

or equivalently,

$$\frac{\Delta P}{P} = -D\Delta y + \frac{1}{2}C\Delta y^2$$

Proof. Taylor expansion. □

Note 21.4.1 (first-order neutral hedging). We can long a bond and short a

Note 21.4.2. Assumptions in duration based hedging:

- the yield curve is flat; that is, $R(t, T) = const$ for all maturities T and a fixed t.
- the yield curve is flat at each point in time; that is, $R(t, T) = const$ for all maturities T and a fixed t.
- the change in yield is small.

(Note that theoretically the price of a interest rate product will depends on the whole yield curve, instead of a single yield value assumed here).

Common hedging instrument includes

- Bonds
- Interest rate swap
- Interest rate futures
- Interest rate options

The strategy is the hold a dynamic portfolio contains the original portfolio and the hedging instruments such that the dynamic portfolio has zero duration.

Lemma 21.4.7 (hedging yield curve: parallel shift vs. non-parallel shift). *Consider a portfolio consisting of two bonds P_1 and P_2 such that $V = n_1 P_1 + n_2 P_2$. Denote the yield of the two bonds by y_1 and y_2, and their durations by D_1 and D_2.*

Define $\Delta V = V(y) - V(\Delta y)$, where y here is the whole yield curve. It follows that

- *Suppose yield curve will only shift in a parallel manner, then to the first order accuracy, if*
$$n_2 = -n_1 \frac{D_1 P_1}{D_2 P_2},$$
then
$$\Delta V = 0.$$
- *If the yield curve shifts in a non-parallel manner such that $\Delta y_1 \neq \Delta y_2$, then choosing*
$$n_2 = -n_1 \frac{D_1 P_1}{D_2 P_2} \frac{Cov(\Delta y_1, \Delta y_2)}{Var[\Delta y_2]}.$$
will minimize $Var[\Delta V]$.

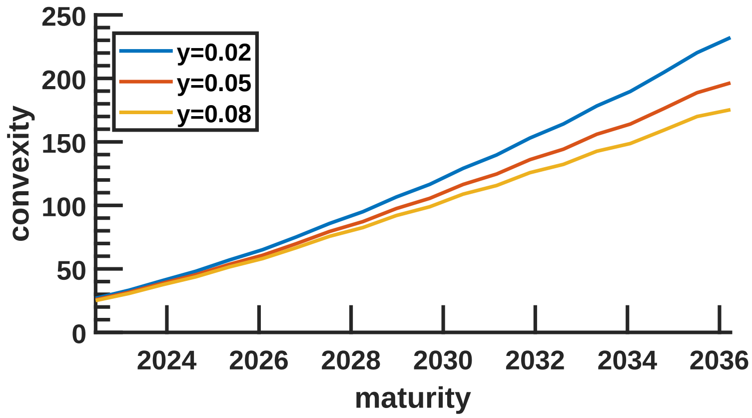
Proof. (1) Note that
$$\Delta V = -n_1 P_1 D_1 \Delta y_1 - n_1 P_1 D_1 \Delta y_1,$$
where Δy_1 denotes the infinitesimal change of the yield curve occurs on y_1.

(2) Note that
$$Var[\Delta V] = (n_1 D_1 P_1)^2 Var[\Delta y_1] + (n_2 D_2 P_2)^2 Var[\Delta y_2] + 2 n_1 n_2 D_1 D_2 P_1 P_2 Cov(\Delta y_1, \Delta y_2).$$

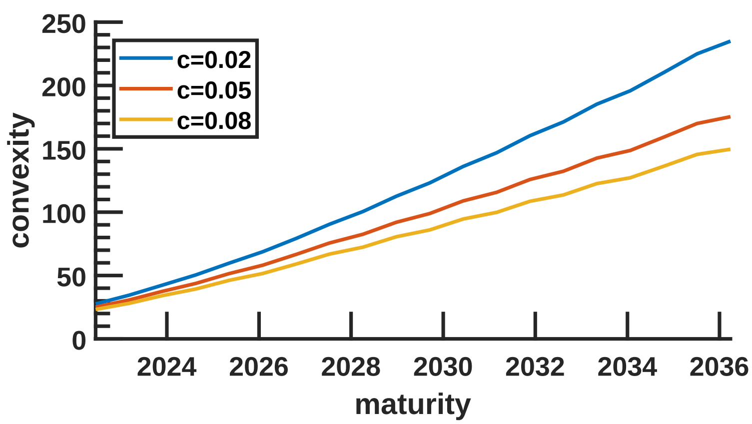
Set the first derivative with respect to n_2 to zero, we have
$$\frac{\partial Var[\Delta V}{\partial n_2} = 2 n_2 (D_2 P_2)^2 Var[\Delta y_2] + 2 n_1 D_1 D_2 P_1 P_2 Cov(\Delta y_1, \Delta y_2) = 0,$$

We have
$$n_2 = -n_1 \frac{D_1 P_1}{D_2 P_2} \frac{Cov(\Delta y_1, \Delta y_2)}{Var[\Delta y_2]}.$$

□



(a) Effects of yield and maturity on duration of a bond



(b) Effects of coupon rate and maturity on duration of a bond

Figure 21.4.5: Effects of coupon rate, yield, and maturity on the duration of a bond

21.4.3 Yield curve dynamics modeling

21.4.3.1 *PCA analysis*

Definition 21.4.4 (PCA analysis of curve daily change). *Let swap rate curve(or yield curve) be represented by a vector $y \in \mathbb{R}^N$, where N is the number of different tenors. Suppose we have $M+1$ historical daily observations of curve, denoted by $y_0, y_1, y_2, ..., y_M$. The historical daily change therefore can be constructed via $\Delta y_i = y_i - y_{i-1}$.*

It follows that

- *the covariance matrix Σ of the daily change can be obtained via*

$$\Sigma = \frac{1}{M}\sum_{i=1}^{N}(\Delta y_i - \overline{\Delta y})(\Delta y_i - \overline{\Delta y})^T.$$

- *The PCA factor of the daily change can be obtained via eigendecomposition*

$$\Sigma = V\Lambda V^T.$$

Example 21.4.4. [2, p. 193]

- We analyze **daily spot swap rates** with maturities of 1 year, 2 years, 3 years, 4 years, 5 years, 7 years, 10 years, and 30 years observing between 2000 and 2011. Table 21.4.1 shows the PCA factors and the associated eigenvectors. Figure 21.4.6 plots the first three dominating PCA factors.
- The first eigenvector/factor represents the parallel shift mode; the second eigenvector represents the steepening model; the third eigenvector represents the bending mode.

Note 21.4.3 (interpret eigenvalue as the variance of the factor score). In the PCA framework, we can decompose the daily change of swap rate Δy(we assume there are N tenors, then $\Delta y \in \mathbb{R}^N$)as

$$\Delta y = \sum_{i=1}^{N} v_i s_i,$$

where $v_i \in \mathbb{R}^N$ and $s_i \in \mathbb{R}$ is known as the **factor score**.

The variance of the score i can be calculated via

$$\begin{aligned} Var[s_i] &= Var[\Delta y \cdot v_i] \\ &= v_i^T Cov[\Delta y] v_i \\ &= v_i^T (\sum_{i=1}^{N} \lambda_i v_i v_i^T) v_i \\ &= \lambda_i \end{aligned}$$

Table 21.4.1: Eigenvectors and eigenvalues for swap rate daily change

(a) Eigenvectors for swap rate daily change

	PC1	PC2	PC3	⋯	PC8
1Y	0.216	-0.501	0.627	⋯	-0.034
2Y	0.331	-0.429	0.129	⋯	0.236
3Y	0.372	-0.267	-0.157	⋯	-0.564
4Y	0.392	-0.110	-0.256	⋯	0.512
5Y	0.404	0.019	-0.355	⋯	-0.327
7Y	0.394	0.194	-0.195	⋯	0.422
10Y	0.376	0.371	0.068	⋯	-0.279
30Y	0.305	0.554	0.575	⋯	0.032

(b) Eigenvalues for swap rate daily change

	PC1	**PC2**	**PC3**	⋯	**PC8**
eigenvalue	4.77	2.08	1.29	⋯	-0.034

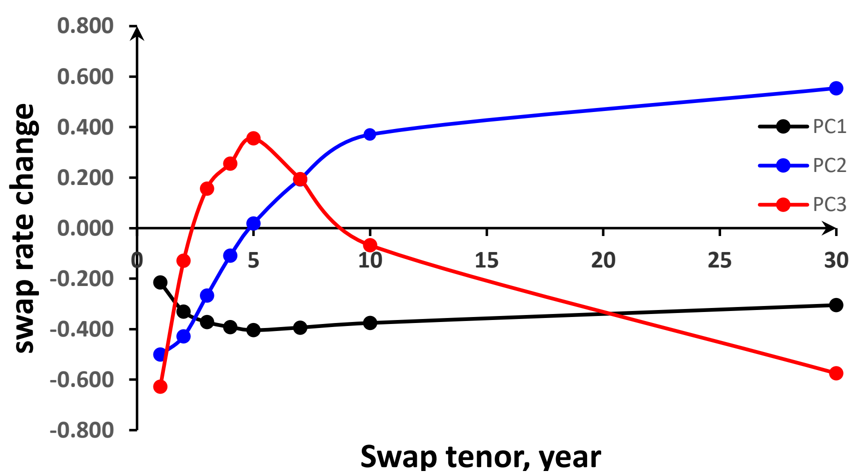


Figure 21.4.6: Demonstration of first three dominating PCA factor in the swap rate curve daily change.

21.4.3.2 *A PCA factor simulator*

Definition 21.4.5. *Let swap rate curve(or yield curve) be represented by a vector $y \in \mathbb{R}^N$, where N is the number of different tenors. Denote today's curve by y_0. And assume the **daily change** of the curve can be decomposed via PCA by*

$$\Delta y \approx v_1 s_1 + v_2 s_2 + v_3 s_3,$$

where $v_1, v_2, v_3 \in \mathbb{R}^N$ are three dominating PCA factors and s_1, s_2, s_3 are the associated factor scores.

- *Then we can generate the daily change scenario i via*

$$\Delta y_i = \lambda_1 v_1 Z_1 + \lambda_2 v_2 Z_2 + \lambda_3 v_3 Z_3,$$

  where Z_1, Z_2, Z_3 are independent standard normal.
- *If daily change of the curve is independent of each other, then we can generate scenario of curve in the long horizon T via*

$$y_T = y_0 + \sum_{i=1}^{T} \Delta y_i.$$

21.4.4 Risk analytics in curve model

21.4.4.1 *Statistical approximation hedging*

21.4.5 Callable bonds risk analytics

When interest rates fall, most bond(without embedded options) prices rise. But callable bond prices actually fall when rates fall; this is because the likelihood of being called increases.

If youâĂŹre considering a callable bond, youâĂŹll want to look at two important factors.

- What do you expect to happen to interest rates between now and the call date? If you think rates will rise or hold steady, you donâĂŹt have to worry about the bond being called.
- But if you think rates may fall, youâĂŹll want to be paid for the additional risk in a callable bond.

Buying a callable bond is like writing a call option, you get the premium if it is not exercised.

21.5 Defaultable bond anaysis

21.5.1 Spread concept

Definition 21.5.1 (yield spread). *[7]*

- *The **yield spread** is the difference between the yield-to-maturity of the credit risky bond and the yield-to-maturity of an on-the-run treasury benchmark bond with similar but not necessarily identical maturity.*
- *Let y_D denote the yield-to-maturity of defaultable bond; let y_B denote the yield-to-maturity of a referenced on-the-run treasury benchmark bond, then*

$$yield\ spread = y_D - y_B.$$

Definition 21.5.2 (I-spread). *[7]*

- *The **Interpolated Spread or I-spread** is the difference between the yield to maturity of the bond and the linearly interpolated yield to the same maturity on an appropriate reference curve.*
- *Let the time-to-maturity of the defaultable bond be T_D. Let y_D denote the yield-to-maturity of defaultable bond; let y_1 and y_2 be the yield-to-maturity of two government bonds with time-to-maturity T_1 and T_2 such that $T_1 \leq T_D \leq T_2$. Then*

$$ISpread = y_D - [y_1 + \frac{y_2 - y_1}{T_2 - T_1}(T_D - T_1)].$$

Definition 21.5.3 (option adjusted spread). *[7]*

- *The **Option Adjusted Spread**(OAS) is the parallel shift to the LIBOR zero rate curve required in order that the adjusted curve reprices the bond*
- *The OAS is sometimes referred to as the Zero-Volatility Spread (ZVS) or Z-Spread.*
- *Let current time be zero. Let LIBOR zero-rate curve be represented by $y(T)$, where T is the maturity.*
  Consider a bond with coupon C at $T_1, ..., T_n$ and principal N at T_n, and semi-anual compounding. The OAS, denoted by s, is the constant such that

$$p^{full} = \sum_{i=1}^{n} \frac{C}{(1 + \frac{y(T_i)+s}{2})^{2T_i}} + \frac{N}{(1 + \frac{y(T_n)+s}{2})^{2T_n}}.$$

Remark 21.5.1. [7]

- The OAS reflects a parallel shift of the spread against LIBOR. Only the spreads are bumped rather than the whole yield. As a consequence, the OAS takes into account the shape of the term structure of LIBOR rates.
- The OAS assumes that cashflows can be reinvested at LIBOR+OAS. As a result, future expectations about interest rates are taken into account. However there still remains reinvestment risk as it is not possible to lock in this forward rate today.

21.5.2 Spread approach

21.5.3 Bond price via credit term structure

Lemma 21.5.1. *[8] Consider a bond with notional cash flow $CF^p(t_i)$ and interest cash flow $CF^i(t_i)$ at $t_i, i = 1, 2, ..., N$. Let τ be the random default time. Let R_p be the par value F^p recovery rate at default and R_c be the accured interest AI at default. Then the fair value of a defaultable bond at current time t is given by*

$$\begin{aligned} PV(t) = &\sum_{t_i > t}^{t_N = T} (CF^p(t_i) + CF^i(t_i)) \cdot E_Q[df(t, t_i) \cdot \mathbf{1}_{t_i < \tau}] \\ &+ \int_t^T E_Q[(R_p \cdot F^p(\tau) + R_c \cdot AI(\tau)) \cdot df(t, \tau) \cdot \mathbf{1}_{u < \tau \leq u + du} du] \end{aligned}$$

21.6 Bond strategies

21.6.1 stationary yield curve strategy

Definition 21.6.1 (riding the yield curve).

- *If the current term structure slopes up and it is expected to remain unchanged, then buying longer term bonds produces higher returns over time.*
- *If the yield curve shifts upward(even if its slope remains the same), investing in long term bonds will be inferior.*

Example 21.6.1.

- Suppose a 5 year zero-coupon bond has a yield of 6% (so a price of \$747.26) and a 10 year zero-coupon bond has a yield of 7.5% (i.e. a price of \$485.19).
- Buying the 5 year zero-coupon bond locks in a return of 6%. If the yield curve remains unchanged, in 5 years the 10 year zero-coupon bond will sell for \$747.26.
- Eventually, buying the 10 year zero-coupon bond and holding it for 5 years generates a return of $(747.26/485.19)^{1/5} - 1 = 9.02\%$.
- If the yield curve shifts upward in 5 years, then the 10 year zero-coupon bond can only be sold at a lower price.

21.6.2 Yield curve parallel movement strategy

Definition 21.6.2 (barbell strategy). *[9, p. 478]*

- *In the barbell strategy, we duplicate the duration of an existing bond using a portfolio with one shorter maturity bond and one longer maturity bond in such a way as to increase convexity.*
- *When there is a parallel shift in the yield curve, the barbell portfolio can outperform the intermediate tenor due to extra gain in the convexity.*

Example 21.6.2. Consider the following three bonds with properties listed in the table.

tenor	yield,%	duration	convexity
2Y	7.71	1.78	0.041
5Y	8.35	3.96	0.195
10Y	8.84	6.54	0.568

We have

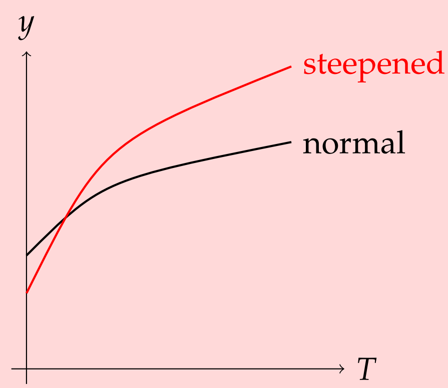
- a barbell portfolio of \$5.4M of 2-year note and \$4.6M of the 10 year note will have yield, convexity and duration given by

$$y_P = \frac{5.4 \times 7.71 + 4.6 \times 8.84}{10} = 8.23$$

$$C_P = \frac{5.4 \times 0.041 + 4.6 \times 0.568}{10} = 0.283,$$

$$D_P = \frac{5.4 \times 1.78 + 4.6 \times 6.54}{10} = 3.96.$$

- Assume a parallel shift in the yield curve, the barbell portfolio can outperform the 5-year note as long as the curve **shifts up or down significantly** due to larger convexity. This is because the barbell portfolio has lower yield than the 5 year bond; this is known as **convexity cost**[9, p. 480].
- However, if the yield curve steepens the barbell can substantially underperform the 5-year note, since the 10 Y bond will lose larger value.



21.7 Bond portfolio management

21.7.1 Bond strategy classification

Definition 21.7.1 (bond portfolio strategy). *[10]*

- ***Pure bond indexing (or full replication approach).*** *The goal here is to produce a portfolio that is a perfect match to the benchmark portfolio. The pure bond indexing approach attempts to duplicate the index by owning all the bonds in the index in the same percentage as the index.*
- ***Enhanced indexing by matching primary risk factors.*** *The goal here is to produce a portfolio match the primary index risk factors of a benchmark portfolio and achieve a higher return than under full replication.*
- ***Enhanced indexing by small risk factor mismatches.****The goal here is to not only produce a portfolio match the primary index risk factors of a benchmark portfolio, but also allows the manager to tilt the portfolio in favor of any of the other risk factors like relative value in certain sectors, quality, term structure, and so on, in order to achieve a higher return.*
- ***Active management by larger risk factor mismatches.*** *The manager makes deliberately larger mismatches on the primary risk factors to actively pursuing opportunities in the market to increase the return.*
- ***Full-blown active management.*** *Full-blown active management involves the possibility of aggressive mismatches on duration, sector weights, and other factors.*

21.8 Mortgage backed security

21.8.1 Mortgage basics

21.8.1.1 *Fixed rate mortgage mechanics*

Lemma 21.8.1 (fixed rate mortgage payment calculation). *[6, p. 584] Denote the original principal amount or loan balance as $B(0)$, f as payment frequency each year, y as the mortgage rate, MP as the periodic mortgage payment.*

- *The eqution to solve X for a T year mortgage is given by*

$$\begin{aligned} MB(0) &= X \cdot \sum_{n=1}^{f\cdot T} \frac{1}{(1+y/f)^{f\cdot n}} \\ &= X\frac{f}{y}(1 - \frac{1}{(1+y/f)^{f\cdot T}}). \end{aligned}$$

- *Let $MB(n)$ denote the principal amount outstanding after n mortgage payments. The interest component in the $n+1$ mortgage payment is*

$$I(n+1) = MB(n) \cdot \frac{y}{f}.$$

  The principal component in the $n+1$ mortgage payment is

$$P(n+1) = X - MB(n) \cdot \frac{y}{f}.$$

  And

$$B(n+1) = B(n) - P(n+1)$$

Example 21.8.1. [6, p. 584]

- A homeowner might borrow \$100,000 from a bank at 4% mortgate rate and agree to make payments of \$477.42 every month for 30 years. The mortgage rate and the monthly payment are related by the following equation:

$$477.42 \cdot \sum_{n=1}^{360} \frac{1}{(1+0.04/12)^n} = 100000.$$

- The interest component in the first payment is

$$100000 \cdot \frac{0.04}{12} = 333.33.$$

- The interest component in the first payment is

$$477.42 - 333.33 = 144.08.$$

- The outstanding loan balance after the first payment is

$$100000 - 144.08 = 99855.92.$$

Lemma 21.8.2 (analytics of mortgage payment).

- $$MP = MB(0)(\frac{z(1+z)^n}{(1+z)^n - 1})$$

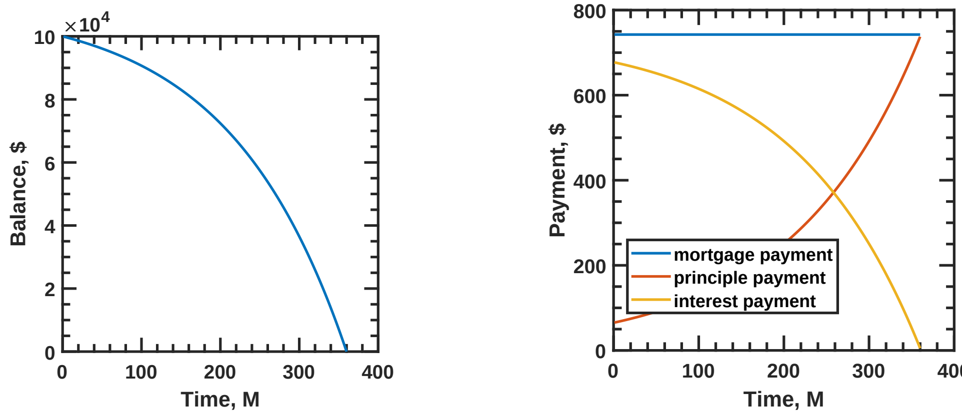
- $$MB(t) = MB(0)(\frac{(1+z)^n - (1+z)^t}{(1+z)^n - 1}).$$

- $$I(t) = MB(t) \cdot z = MB(0)(\frac{z(1+z)^n - (1+z)^{t-1}}{(1+z)^n - 1}), P(t) = MP - I(t) = MB(0)(\frac{z(1+z)^{t-1}}{(1+z)^n - 1}).$$

Proof. Note that

$$\begin{aligned} MB(t) &= \frac{MP}{1+z} + \frac{MP}{(1+z)^2} + \cdots + \frac{MP}{(1+z)^{n-t}} \\ &= MP(\frac{1}{1+z} \frac{(\frac{1}{1+z})^{n-t} - 1}{\frac{1}{1+z} - 1}) \\ &= MP\frac{1}{z}(\frac{1}{(1+z)^{n-t}} - 1) \\ &= MB(0)(\frac{(1+z)^n - (1+z)^t}{(1+z)^n - 1}) \end{aligned}$$

□



(a) The future yield curve in one simulation scenario. (b) The future yield curve in another simulation scenario.

Figure 21.8.1: Bond dynamics.

21.8.1.2 *Prepayment modeling*

Definition 21.8.1.

- *The* ***single monthly mortality rate*** *at month* n*, denoted by* SMM_n*, is the percentage of* $B(n)$ *that is prepaid during month* n*.*
- *The constant prepayment rate* CPR_n *is the annualized* SMM_n *given by*

$$CPR_n = 1 - (1 - SMM_n)^{12}$$

$$CPR_n = 1 - (1 - SMM)$$

Lemma 21.8.3. *The SMM is often annualized to a constant prepayment rate or conditional prepayment rate (CPR).*

Definition 21.8.2 (public securities association (PSA) model). *The model assumes that the prepayment rate starts at 0.2% CPR in the first month and then rises 0.2% CPR per month until month 30; after that, the prepayment rate levels out at 6% CPR. We have*

$$PSA = \frac{CPR(t)}{\min(t, 30M) \times 0.2},$$

and

$$CPR(t) = \begin{cases} 0.2\% \times t, t \leq 30M \\ 6\%, t \geq 31M \end{cases}.$$

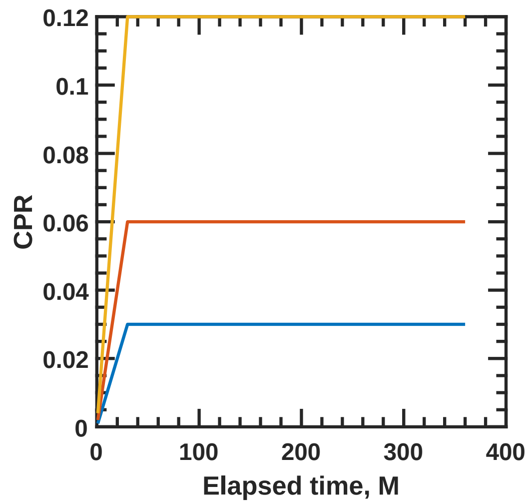


Figure 21.8.2

Definition 21.8.3 (Basic relationship among prepayment model). *[11, p. 334] Denote*

- *z monthly gross mortgage rate; $z = c + s$, where c is the monthly coupon rate going to the investors after after the monthly service fee charged.*
- *$\overline{MB}(t)$ as the loan outstanding at month t after considering scheduled principle payment and prepayment from first month to month $t-1$; $MB(t)$ as the loan outstanding at month t after considering scheduled principle payment and excluding any prepayments.*
- *$SMM(t)$ the single monthly mortality rate at month t;*
- *$P(t)$ monthly scheduled principle payment for month t, **excluding any prepayments**.*
- *$\overline{P}(t)$ projected monthly principle payment for month t, including both scheduled principle payment and prepayment; $\overline{P}(t)$ projected monthly principle payment for month t, including both scheduled principle payment and prepayment.*
- *$\overline{SP}(t)$ projected scheduled monthly principle payment for month t, resulting from projected.*
- *$\overline{MP}(t)$, projected monthly mortgage payment for month t after considering the scheduled principle payment and prepayment up to month $t-1$ and resulting refinancing.*
- *$\overline{CF}(t)$ projected cash flow for month t.*

Then

- $$\overline{MP}(t+1) = \overline{MB}(t)\frac{1}{\frac{1}{1+z} + \frac{1}{(1+z)^2} + \cdots \frac{1}{(1+z)^{N-t}}}.$$
- $$\overline{I}(t+1) = \overline{MB}(t)\cdot z, \overline{SP}(t+1) = \overline{MP}(t+1) - \overline{I}(t+1).$$
- $$\overline{PR}(t+1) \triangleq (\overline{MB}(t) - \overline{SP}(t+1))\cdot SMM(t+1).$$
- $$\overline{P}(t+1) = \overline{SP}(t+1) + \overline{PR}(t+1).$$
- $$\overline{MB}(t+1) = \overline{MB}(t) - \overline{P}(t+1).$$
- $$\overline{CF}(t) = \overline{MP}(t) + \overline{PR}(t) - \overline{S}(t)$$
- *The total undiscounted interest expenses/incomes for the borrower/lender is*
$$\sum_{i=1}^{N} \overline{I}(t).$$

Lemma 21.8.4 (analytics in prepayment model). *[11, p. 334] Denote*

- $\overline{I}(t)$ *projected monthly interest payment after considering the scheduled principle payment and prepayment up to month* $t-1$ *and resulting refinancing;* $I(t)$ *scheduled interest payment for month* t.
- $b(t) = \prod_{n=1}^{t}(1 - SMM(n))$

It follows that

- $$\overline{MP}(t) = \overline{MB}(t-1)(\frac{z(1+z)^{N-t+1}}{(1+z)^{N-t+1}-1})$$
- $$\overline{MB}(t) = MB(t)\cdot b(t), \overline{MP}(t) = MP\cdot b(t-1), \overline{I}(t) = I(t)\cdot b(t-1), \overline{SP}(t) = P(t)\cdot b(t-1).$$

Proof. (1) Using definition and summation
$$\frac{\overline{MP}(t+1)}{\overline{MB}(t)} = \frac{MP(t+1)}{MB(t)} = \frac{1}{\frac{1}{1+z} + \frac{1}{(1+z)^2} + \cdots \frac{1}{(1+z)^{N-t}}}.$$

(2)

$$\begin{aligned}
\overline{MB}(t+1) &= \overline{MB}(t) - \overline{P}(t+1) \\
&= \overline{MB}(t) - (\overline{SP}(t+1) + (\overline{MB}(t) - \overline{SP}(t+1)) \cdot SMM(t+1)) \\
&= (\overline{MB}(t) + \overline{SP}(t+1))(1 - SMM(t+1))
\end{aligned}$$

For $t = 0$, we have

$$\begin{aligned}
\overline{MB}(1) &= (\overline{MB}(0) + \overline{SP}(1))(1 - SMM(1)) \\
&= (MB(0) + P(1))(1 - SMM(1)) \\
&= MB(1) \cdot b(1)
\end{aligned}$$

where we use the definition $\overline{MB}(0) = MB(0), \overline{SP}(1) = P(1)$. Therefore,

$$\overline{MP}(2) = MP \cdot b(1), \overline{I}(2) = I(2) \cdot b(1), \overline{SP}(2) = P(2) \cdot b(1).$$

For $t = 1$, we have

$$\begin{aligned}
\overline{MB}(2) &= (\overline{MB}(1) + \overline{SP}(2))(1 - SMM(2)) \\
&= (MB(0) + P(1))(1 - SMM(1))(1 - SMM(2)) \\
&= MB(2)b(2)
\end{aligned}$$

Therefore,

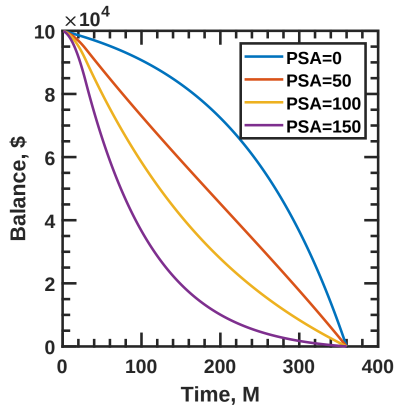
$$\overline{MP}(3) = MP \cdot b(2), \overline{I}(3) = I(3) \cdot b(2), \overline{SP}(3) = P(3) \cdot b(2).$$

Continue the process and we will get the result.

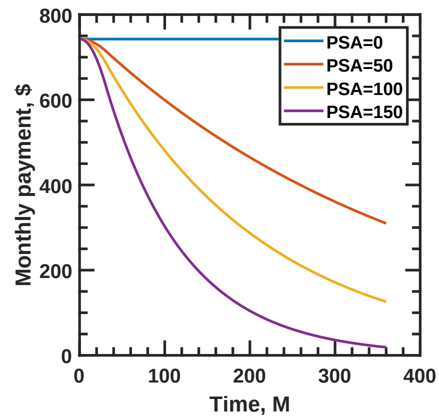
□

Remark 21.8.1 (understand prepayment risk).

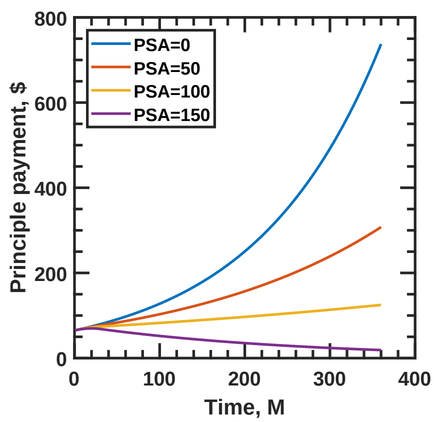
- From Figure Figure 21.8.3, we can see that as prepayment speed increase, the loan balance will decrease faster than the scheduled.
- The higher the prepayment rate, the lower the total interest expenses/incomes.
- The prepayment model predicts that the prepayment amount will first increase for the first 30M and then decrease. The increase is due to the increasing CRP for the first 30M; the decreasing is because the remaining loan balance is decreasing.



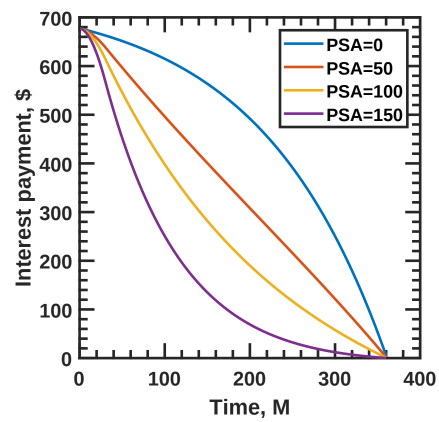
(a) Expected loan balance dynamics after considering prepayment at different speeds.



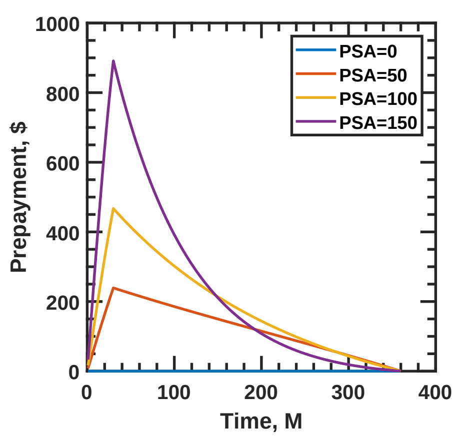
(b) Scheduled monthly mortgage payments (excluding prepayment at this month) at different prepayment speeds.



(c) Expected monthly principle payments after considering prepayment at different speeds.



(d) Expected monthly interest payments after considering prepayment at different speeds.



(e) Expected monthly prepayments after considering prepayment at different speeds.

Figure 21.8.3: Balance and Cash flow with prepayment for a mortgage.

21.8.1.3 *Cash flow yield and spreads*

Definition 21.8.4 (cash flow yield). *Consider a mortgate with N monthly payments. Let $\overline{CF}(t)$ denote the projected cash flow for month t with prepayment considered. The **monthly cash flow yield** y is the root of the following cash flow discounting equation:*

$$0 = \frac{\overline{CF}(1)}{(1+y)} + \frac{\overline{CF}(2)}{(1+y)^2} + \cdots \frac{\overline{CF}(N)}{(1+y)^N}.$$

Definition 21.8.5 (option). *Consider a mortgate with N monthly payments. Let $\overline{CF}(t)$ denote the projected cash flow for month t with prepayment considered. The **monthly cash flow yield** y is the root of the following cash flow discounting equation:*

$$0 = \frac{\overline{CF}(1)}{(1+y)} + \frac{\overline{CF}(2)}{(1+y)^2} + \cdots \frac{\overline{CF}(N)}{(1+y)^N}.$$

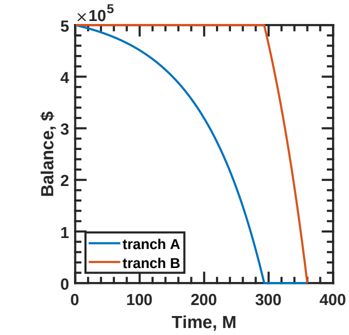
21.8.1.4 *Risk analytics*

21.8.2 Collateralized mortgage obligation (CMO)

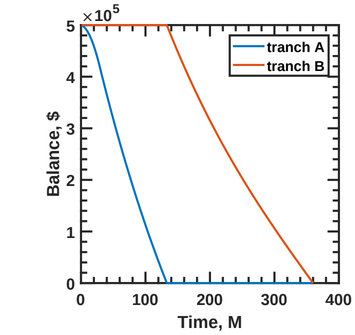
21.8.2.1 *Sequential CMO*

Definition 21.8.6.

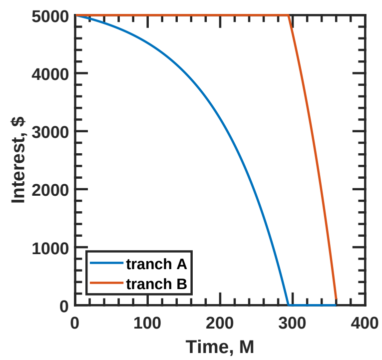
- *In the sequential CMO, the cash flow from the mortgage pass-through will be further re-distributed across tranch A, B, C, ...*
- *The total principal payment, including both scheduled and unexpected principal payment, will first go to tranch A until its balance reaches zero; then the principal payment will go to tranch B.*
- *Because the loan balance in tranch A will decrease first, investors on tranch A will receive smaller interest payments.*



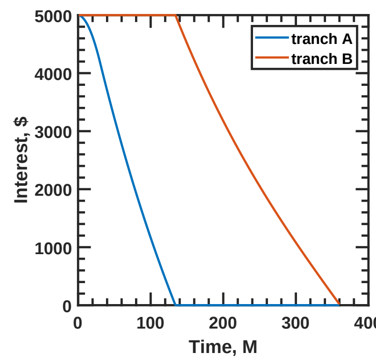
(a) Expected loan balance dynamics for tranch A and tranch B without prepayment risk.



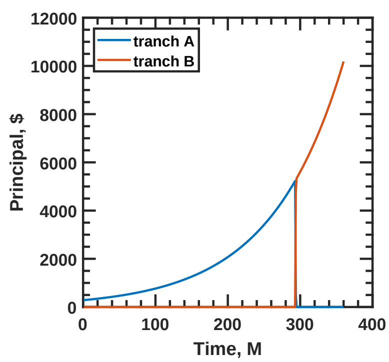
(b) Expected loan balance dynamics for tranch A and tranch B with prepayment risk.



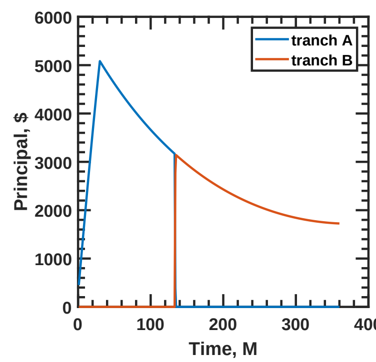
(c) Expected interest payment for tranch A and tranch B without prepayment risk.



(d) Expected interest payment for tranch A and tranch B with prepayment risk.



(e) Expected principal payment for tranch A and tranch B without prepayment risk.



(f) Expected principal payment for tranch A and tranch B with prepayment risk.

Figure 21.8.4: Balance and Cash flow with for a sequential CMO with two tranches: monthly coupon rate 0.01, initial loan balance 5MM, prepayment speed PSA = 100%.

21.9 Notes on bibliography

Major references are .

[1][9][6]. For mortgage modeling, see [12].

For bond credit spreads, see [7].

BIBLIOGRAPHY

1. Pinto, J., Petitt, B. & Pirie, W. *Fixed Income Analysis 3rd Edition* ISBN: 9781118999509. https://books.google.com/books?id=rOgbBgAAQBAJ (Wiley, 2015).
2. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
3. DeFusco, R. A., McLeavey, D. W., Pinto, J. E., Anson, M. J. & Runkle, D. E. *Quantitative investment analysis* (John Wiley & Sons, 2015).
4. Wu, L. *Interest rate modeling: Theory and practice* (CRC Press, 2009).
5. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
6. Tuckman, B. & Serrat, A. *Fixed income securities: tools for today's markets* (John Wiley & Sons, 2011).
7. O'Kane, D. & Sen, S. Credit spreads explained. *Journal of Credit Risk* **1,** 61–78 (2005).
8. Berd, A. M., Mashal, R. & Wang, P. Defining, estimating and using credit term structures. Part 1: Consistent Valuation Measures (2009).
9. Fabozzi, F. *Bond Markets, Analysis, and Strategies* ISBN: 9780133797138. https://books.google.com/books?id=a7WgBwAAQBAJ (Pearson Education, 2015).
10. Maginn, J. L., Tuttle, D. L., McLeavey, D. W., Pinto, J. E., *et al. Managing investment portfolios: a dynamic process* (John Wiley & Sons, 2007).
11. Fabozzi, F. *Fixed income mathematics, 3rd Ed* (McGraw-hill, 1993).
12. Davidson, A. & Levin, A. *Mortgage Valuation Models: Embedded Options, Risk, and Uncertainty* (Oxford University Press, 2014).

22

FOREIGN EXCHANGE MODELING

22.1 Overview

The foreign exchange (FX) market and its currency option market are probably the world's largest market. The major market participants include investment institutions, banks, international corporations, and private investors.

FX market shares many similarities with equity market since foreign currency can be viewed as risky assets with additional FX market risk factors, e.g., FX spot rate, FX spot rate volatility etc. Additionally, FX market can be perceived from both perspective of domestic investor and foreign investors; therefore, some symmetric properties can be employed or have to be satisfied.

This chapter will start with basics of FX market (spot rate, forward rate, interest rate parity, and option delta conventions); then we move to the martingale pricing framework addressing the general pricing theorem in the FX market; We Finally, we extending single-currency short-rate and forward rate modeling framework to cross-country setting as the preparation for credit valuation adjustment in

22.2 The FX basics and market

22.2.1 FX market

22.2.1.1 *Spot market*

Definition 22.2.1 (FX spot rate).

- *The first currency of a currency pair is called the base currency, and the second currency is called the quote currency. The currency pair shows how much of the quote currency is needed to purchase one unit of the base currency.*
- *The **FX spot rate** S_t or $S(t)$, represents the number of units of domestic currency needed to buy one nit of foreign current at time t. Essentially, **spot rate** is the price of foreign currency in the unit of domestic currency.*
- *The FX spot rate is usually denoted by symbol **FOR-DOM** or **FOR/DOM**. For example, EUR-USD = 1.39 means that one EUR is worth 1.39 USD. The EUR-USD = 1.39 quote is equivalent to USD-EUR = 0.7194.*

Remark 22.2.1 (commonly traded currency pairs)**.** In the FX market, the four most popular traded currency pairs are:

- EUR/USD (euro/dollar)
- USD/JPY (U.S. dollar/Japanese yen)
- GBP/USD (British pound/dollar)
- USD/CHF (U.S. dollar/Swiss franc)

The three less popular commodity pairs are:

- AUD/USD (Australian dollar/U.S. dollar)
- USD/CAD (U.S. dollar/Canadian dollar)
- NZD/USD (New Zealand dollar/U.S. dollar)

These currency pairs, along with their various combinations (such as EUR/JPY, GBP/JPY and EUR/GBP) account for the majority of all speculative trading in FX.

Definition 22.2.2 (FX forward rate).

- *The **FX forward rate** $F(t,T)$, is the exchange rte between the domestic currency and the foreign currency at some future point of time T as observed at the present time $t(t < T)$.*
- *The FX forward rate is usually derived from forward/futures contract traded on the market(**??**).*

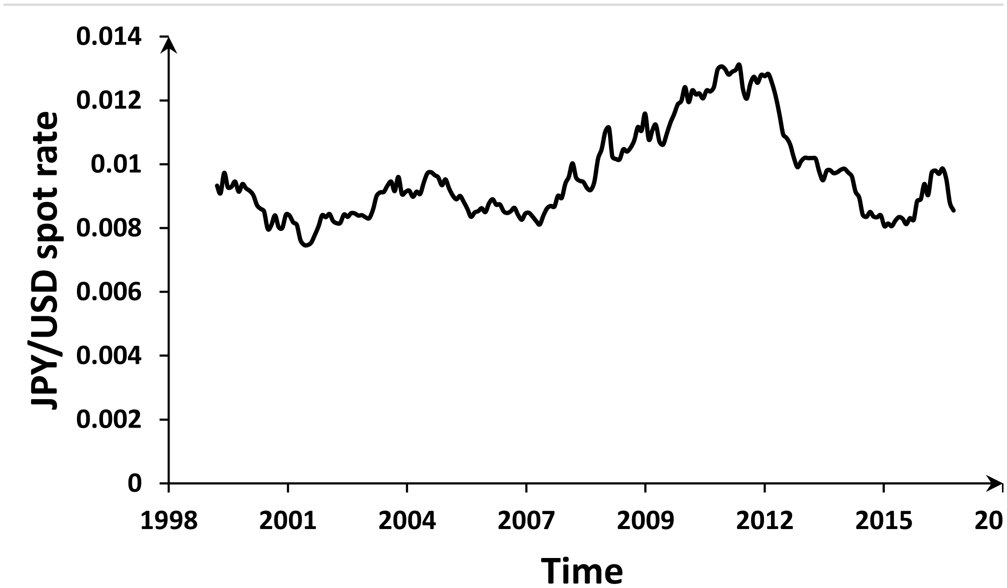


Figure 22.2.1: FX JPY/USD historical quote time series.

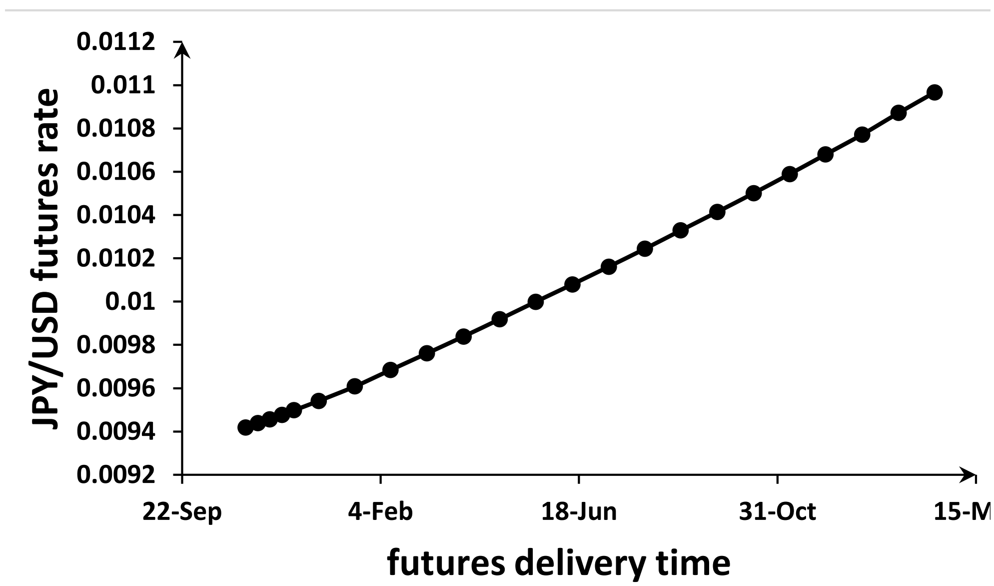


Figure 22.2.2: FX JPY/USD forward rate curve observed on Mar,11,2018

Table 22.2.1: FX futures and European options on FX futures. data source: CME group http://www.cmegroup.com/trading/fx/g10/japanese-yen.html

contracts	futures settlement date	futures delivery date	option expiry date
MAR18	19-Mar-2018	22-Mar-2018	09-Mar-2018
APR18	16-Apr-2018	18-Apr-2018	06-Apr-2018
MAY18	14-May-2018	16-May-2018	04-May-2018
JUN18	18-Jun-2018	20-Jun-2018	08-Jun-2018
...	...	...	...

22.2.1.2 *Forward market*

22.2.1.3 *FX trading PnL accounting*

Remark 22.2.2.
- In the FX trading, domestic investors usually will calculate the profit&loss(PnL) of their position in domestic currency. Economically, domestic currency is the currency investors will ultimately hold and use.
- However, a domestic investors can also calculate PnL using foreign currency if he wants to ultimately emigrate to foreign country and use foreign currency.
- Consider the holding of domestic money market account. Investors will not have FX exposure based on domestic currency accounting; However, they have have FX exposure based on foreign currency accounting.

22.2.2 The integrated market

22.2.2.1 *The setup of the integrated market*

22.2.2.2 *Fundamental relationships in the integrated market*

Theorem 22.2.1 (no-arbitrage relation for asset spot prices denominated in different currencies). *Suppose there is a market where an asset is quoted in both domestic currency and foreign currency. Let $V^f(t)$ denote its foreign currency denominated price at time t. Let $V^d(t)$ denote its domestic currency denominated price at time t. Let $S(t)$ denote the spot rate. Then*

$$S(t)V^f(t) = V^d(t).$$

Let current time be zero. For time $t > 0$, the equality means for all sample path $\omega \in \Omega$, we have

$$S(t,\omega)V^f(t,\omega) = V^d(t,\omega).$$

Proof. If such relation does not hold (suppose $S(t)V^f(t) > V^d(t)$), we can short 1 units of asset using foreign quote and get $S(t)V^f(t)$ domestic currency; buy 1 unit of the asset using domestic quote and cover the short position. There is a riskless profit of $V^d(t) - S(t)V^f(t) - V^d(t)$. □

Theorem 22.2.2 (no-arbitrage relation for asset spot prices denominated in different currencies). *Suppose at current time t there is a market where an asset is quoted in both domestic currency and foreign currency. Let $F^f(t,T)$ denote its foreign currency denominated forward price at time t. Let $V^d(t)$ denote its domestic currency denominated forward price at time t. Let $F^X(t,T)$ denote the forward FX rate. Then for all $t \leq T$, we have*

$$F^X(t,T)F^f(t,T) = F^d(t,T).$$

Let current time be zero. For time $t > 0$, the equality means for all sample path $\omega \in \Omega$, we have

$$F^X(t,T,\omega)F^f(t,T,\omega) = F^d(t,T,\omega).$$

Proof. If such relation does not hold (suppose $F^X(t,T)F^f(t,T) > F^d(t,T)$), we can short 1 unit of forward asset contract using foreign quote, enter the FX forward contract to buy $F^X(t,T)$ units of domestic currency at time T, and enter 1 unit of forward asset contract using domestic quote.

At time T, there is a riskless profit of

$$\underbrace{F^X(t,T)F^f(t,T)}_{\text{gain from short forward asset contract}} - \underbrace{F^d(t,T)}_{payforthelongforwardassetcontract}.$$

□

22.2.3 Volatility skew

22.2.3.1 *Delta types*

Definition 22.2.3 (spot delta). *[1]*

- *In the Black model, the sensitivity of the European option price (denominated in domestic currency) with respect to the spot rate S_t is given by(Lemma 22.6.7)*

$$\Delta_S(K,\sigma,\phi) \triangleq \frac{\partial v}{\partial S_t} = \phi \exp(-r_f(T-t))N(\phi d_+),$$

where $\phi = 1$ for call, $\phi = -1$ for put,

$$v(K,\sigma,S_t,\phi) = \phi(e^{-r^f\tau}S_tN(d_+) - e^{-r\tau}KN(d_-)),$$

and

$$d_\pm = \frac{1}{\sigma\sqrt{\tau}}[\log(\frac{S_t}{K}) + (r^d - r^f \pm \frac{1}{2}\sigma^2)\tau], \tau = T-t.$$

- *And we have put-call delta parity*

$$\Delta_S(K,\sigma,+1) - \Delta_S(K,\sigma,-1) = \exp(-r_f(T-t)).$$

Remark 22.2.3 (comparison with hedging equity option)**.**

- In equity markets, one would buy Δ_S units of the stock, using $\Delta_S \cdot S(t)$ units of currency, to hedge a short vanilla option position.
- In FX markets, this is equivalent to buying Δ_S times the foreign notional NïijŇ in the mean time, selling of $\Delta_S NS_t$ units of domestic currency.

Definition 22.2.4 (forward delta). *[1]*

- *In the black model, the sensitivity of the European option price (denominated in domestic currency) with respect to the forward rate $F(t,T)$ is given by(Lemma 22.6.7)*

$$\Delta_F(K,\sigma,\phi) \triangleq \frac{\partial v}{\partial F(t,T)} = \frac{\partial v}{\partial S_t}\frac{\partial S_t}{\partial F(t,T)} = \phi N(\phi d_+),$$

where $\phi = 1$ for call, $\phi = -1$ for put,

$$v(K,\sigma,S_t,\phi) = \phi(e^{-r\tau}(F(t,T)N(d_+) - KN(d_-)),$$

and

$$d_\pm = \frac{1}{\sigma\sqrt{\tau}}[\log(\frac{F(t,T)}{K}) \pm \frac{1}{2}\sigma^2)\tau], \tau = T-t.$$

- *And we have put-call delta parity*

$$\Delta_F(K,\sigma,+1) - \Delta_F(K,\sigma,-1) = 1.$$

Remark 22.2.4. In FX markets, to hedge a short option, we need to long futures with foreign currency notion $\Delta_F \times N$, where N is the foreign currency notional of the option.

Definition 22.2.5 (premium-adjusted spot delta). *[1]*

- *In the Black model, the sensitivity of the European option price (denominated in domestic currency) with respect to the forward rate $F(t,T)$ is given by(Lemma 22.6.7)*

$$\Delta_{S,pa}(K,\sigma,\phi) \triangleq \frac{\partial v/S_t}{\partial S_t} = \Delta_S - \frac{v}{S_t}(FOM\ to\ buy) = \phi\exp(-r_f\tau)\frac{K}{F(t,T)}N(\phi d_-),$$

where $\phi = 1$ for call, $\phi = -1$ for put,

$$v(K,\sigma,S_t,\phi) = \phi(e^{-r\tau}(F(t,T)N(d_+) - KN(d_-)),$$

and

$$d_\pm = \frac{1}{\sigma\sqrt{\tau}}[\log(\frac{F(t,T)}{K}) \pm \frac{1}{2}\sigma^2)\tau], \tau = T-t.$$

- *And we have put-call delta parity*

$$\Delta_{S,pa}(K,\sigma,+1) - \Delta_{S,pa}(K,\sigma,-1) = \exp(-r_f\tau)\frac{K}{f}.$$

Remark 22.2.5. Suppose the value of an option with a notional of 1,000 EUR was calculated as 74 EUR. Assuming a short position with a delta of 60% means, that buying 600 EUR is necessary to hedge. However the final hedge quantity will be 526 EUR which is the delta quantity reduced by the received premium in EUR. Consequently, the premium-adjusted delta would be 52.63%. T

Lemma 22.2.1 (neutrality of a portfolio).

Remark 22.2.6.

- In the un-adjusted spot delta, we assume the option is quoted in domestic currency and its payoff is in foreign currency.
- The definition of the premium-adjusted spot delta takes care of the correction induced by payment of the premium in foreign currency, which is the amount by which the delta hedge in foreign currency has to be corrected.

Definition 22.2.6 (premium-adjusted forward delta). *[1]*

- *In the black model, the sensitivity of the European option with respect to the forward rate $F(t,T)$ is given by(Lemma 22.6.7)*

$$\Delta_F(K,\sigma,\phi) \triangleq \frac{\partial v}{\partial F(t,T)} = \frac{\partial v}{\partial S_t}\frac{\partial S_t}{\partial F(t,T)} = \phi N(\phi d_+),$$

where $\phi = 1$ *for call,* $\phi = -1$ *for put,*

$$v(K, \sigma, S_t, \phi) = \phi(e^{-r\tau}(F(t,T)N(d_+) - KN(d_-)),$$

and

$$d_\pm = \frac{1}{\sigma\sqrt{\tau}}[\log(\frac{F(t,T)}{K}) \pm \frac{1}{2}\sigma^2)\tau], \tau = T - t.$$

- *And we have put-call delta parity*

$$\Delta_F(K, \sigma, +1) - \Delta_F(K, \sigma, -1) = 1.$$

22.2.3.2 *conversion Delta to strike*

Methodology 22.2.1. *[1]*

- *From*

$$\Delta_F(K, \phi) = \phi N(\phi d_+) = \phi N(\phi \frac{\ln(F(t,T)/K) + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}}),$$

we can solve for the strike

$$K = F(t,T)\exp(-\phi\sigma\sqrt{\tau}N^{-1}(\phi\Delta_f) + \frac{1}{2}\sigma^2\tau).$$

- *From*

$$\Delta_{F,pa}(K, \phi) = \phi\frac{K}{F(t,T)}N(\phi d_-) = \phi\frac{K}{f}N(\phi\frac{\ln(F(t,T)/K) - \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}}),$$

we need to numerically solve for the strike.

22.2.3.3 *Market convention*

Definition 22.2.7 (risk-reversal).

- ***ATM quote*** *is usually volatility of the call option with strike such that call spot/forward delta is equal to negative put spot/forward delta.*
- ***Risk reversal quote*** *is usually given by.*

$$R_{25} = \sigma_{C,25} - \sigma_{P,25}$$

that is for a given maturity, the 25 risk reversal is the vol of the 25 delta call less the vol of the 25 delta put. The 25 delta put is the put whose strike has been chosen such that the delta is -25%.

- ***Strangle quote*** *is usually given by*

$$\sigma_{25-S-Q} = \frac{\sigma_{25C} + \sigma_{25P}}{2} - \sigma_{ATM}.$$

- ***Market strangle quote*** *is usually given by.*

$$\sigma_{25-S-M} = \sigma_{ATM} + \sigma_{25-S-Q}.$$

Note 22.2.1 (market convention of quoting delta). [2, p. 848]

- In the FX market, instead of quoting an option price by specifying its expiry, strike, and volatility, denoted by (T, K, σ), it is also common to quote its expiry, delta and volatility, denoted by (T, Δ, σ).
- For example, consider a six-month 25 delta call has a volatility of 13%.
- We can solve for the strike from delta formula

$$\exp(-D(T-t))N(d_1) = 0.25,$$

where
$$d_1 = \frac{\ln(S_t/K) + (r - D + \frac{1}{2}0.13^2)(T-t)}{0.13\sqrt{T-t}}.$$
Then we can solve for the strike K.

Maturity	$\cdots$	$35\Delta_P$	$40\Delta_P$	$45\Delta_P$	ATM	$45\Delta_C$	$40\Delta_C$	$35\Delta_C$	$\cdots$
1W	$\cdots$	15.5	15.3	15.1	15	15.2	15.4	15.6	$\cdots$
1M	$\cdots$	20.8	20.4	20.1	20	20.2	20.5	20.9	$\cdots$

Table 22.2.2: FX implied volatility for call or put options with different Δ and different time-to-maturities(1 week and 1 month). 45Δ

22.2.4 Dynamical model for spot rate

Unlike the dynamic models for other asset classes, e.g., equity, interest rate, commodity etc, the choice of dynamical model for FX spot rate has to possess some symmetric properties inherent to the FX market. For example, any dynamics model for spot rate S_t should guarantee that the dynamics of $1/S_t$ belong to the same type of model. Further, let S_1 stand for USD/JPY, S_2 stand for EUR/USD, S_3 stand for GBP/USD, then the dynamics of $S_4 \triangleq S_1 \cdot S_2$, representing EUR/JPY, and $S_5 = S_2/S_3$, representing EUR/GBP, should belong to the same type of model.

Lognormal types of model, including constan-volatility lognormal model and local volatility model, naturally satisfy these symmetric requirements. However, dynamical models, like normal model, shifted lognormal model, stochastic volatility model, do not satisfy these requirements and cannot be used to model FX spot rates.

22.3 Martingale pricing framework

22.3.1 Numeraires and martingales

Definition 22.3.1 (different numeraire). *Let M be the domestic money account, M^f be the foreign money account and $S(t)$ the exchange rate. Let $P^d(t,T), P^f(t,T)$ be zero-coupon bond prices in domestic currency and in foreign currencies such that*

$$P^d(t,T) = M(t)E_{Q^d}[\frac{1}{M(T)}|\mathcal{F}_t], P^f(t,T) = M^f(t)E_{Q^f}[\frac{1}{M^f(T)}|\mathcal{F}_t].$$

Then

- ***domestic (risk-neutral) measure*** *is the measure associated with doemstic money account numeraire $M(t)$. We denote expectation with respect to this measure as $E_{Q^d}[\cdot]$.*
- ***foreign (risk-neutral) measure*** *is the measure associated with foreign money account numeraire $S(t)M^f(t)$. We denote expectation with respect to this measure as $E_{Q^f}[\cdot]$.*
- ***domestic T-forward measure*** *is the measure associated with doemstic T maturity zero-coupon bond numeraire $P(t,T)$. We denote expectation with respect to this measure as $E_{T^d}[\cdot]$.*
- ***foreign T-forward measure*** *is the measure associated with foreign T maturity zero-coupon bond numeraire $S(t)P^f(t)$. We denote expectation with respect to this measure as $E_{T^f}[\cdot]$.*
- ***foreign-denomenated foreign risk-neutral measure*** *is the measure associated with foreign money account numeraire $M^f(t)$. We denote expectation with respect to this measure as $E_{Q^{f/f}}[\cdot]$.*
- ***foreign-denomenated foreign T-forward measure*** *is the measure associated with foreign T maturity zero-coupon bond numeraire $P^f(t)$. We denote expectation with respect to this measure as $E_{T^{f/f}}[\cdot]$.*

Theorem 22.3.1 (martingale properties associated with different measures). *Let X be an asset price denomiated in domestic currency. Let Y be an asset price denomiated in foreign currency. It follows that*

- $\frac{X(t)}{M(t)}$ *and* $\frac{S(t)Y(t)}{M(t)}$ *are martingales under domestic (risk-neutral) measure.*
- $\frac{X(t)}{M^f(t)S(t)}$ *and* $\frac{S(t)Y(t)}{M^f(t)S(t)}$ *are martingales under foreign (risk-neutral) measure.*
- $\frac{X(t)}{P^d(t,T)}$ *and* $\frac{S(t)Y(t)}{P^d(t,T)}$ *are martingales under domestic T forward measure.*
- $\frac{X(t)}{P^f(t,T)S(t)}$ *and* $\frac{S(t)Y(t)}{S(t)P^f(t,T)}$ *are martingales under foreign T forward measure.*

- $\frac{X(t)/S(t)}{M^f(t)}$ *and* $\frac{Y(t)}{M^f(t)}$ *are martingales under foreign-denomenated foreign (risk-neutral) measure.*
- $\frac{X(t)/S(t)}{P^f(t,T)}$ *and* $\frac{Y(t)}{P^f(t,T)}$ *are martingales under foreign-denomenated foreign T forward measure.*
- ***(numeriare invariance)*** *Let Z and W be two assets denominated in the same currency, then denominating in an arbitrary currency,* $\frac{Z}{W}$ *is a martingale under W numeriaire denominated in the same currency.*

Proof. See Theorem 15.6.15. □

Note 22.3.1 (equivalence of foreign (risk-neutral) measure and foreign-denominated foreign (risk-neutral) measure)**.** From above, we can see that foreign (risk-neutral) measure and foreign-denominated foreign (risk-neutral) measure are equivalent; Under either of them, $\frac{X(t)}{M^f(t)S(t)}$ and $\frac{Y(t)}{M^f(t)}$ are martingales.

22.3.2 Change of numeraire

Lemma 22.3.1 (cross-country change of numeraire). *[3] Let M be the domestic money account, M^f be the foreign money account and S the exchange rate.*

- *The Radon-Nikodym derivative to change measure from domestic money account M to the foreign money account M^f is given by*

$$Z(T) = \frac{dQ^f}{dQ^d} = \frac{S(T)M^f(T)}{S(t)M^f(t)}\frac{M(t)}{M(T)}$$

- $$E_{Q^d}[Z_T|\mathcal{F}_t] = 1$$

- $$E_{Q^f}[X] = E_{Q^d}[X\frac{dQ^f}{dQ^d}]$$

Proof. (3) Note that $S(T)M^f(T)/M(T)$ is martingale under Q^d, we have

$$E_{Q^d}[Z_T|\mathcal{F}_t] = E_{Q^d}[\frac{S(T)M^f(T)}{S(t)M^f(t)}\frac{M(t)}{M(T)}|\mathcal{F}_t] = \frac{S(T)M^f(T)}{S(t)M^f(t)}\frac{M(t)}{M(T)} = 1.$$

□

Theorem 22.3.2 (change of measure for dynamic models). *Let $X_1, X_2, ..., X_n$ be domestic assets having domestic risk-neutral dynamics*

$$dX_i/X_i = r^d dt + \sigma_{X,i} dW_i^d.$$

Let $Y_1, Y_2, ..., Y_n$ be foreign assets having foreign risk-neutral dynamics

$$dY_i/Y_i = r^f dt + \sigma_{Y,i} dZ_i^f.$$

It follows that

- *Under domestic measure Q^d,*

$$dS/S = (r^d - r^f)dt + \sigma_S dB^d,$$

  where B^d is the Brownian motion under Q^d; under foreign measure Q^f,

$$d(1/S)/(1/S) = (r^f - r^d)dt - \sigma_S dB^f,$$

  where B^d is the Brownian motion under Q^d;
- *Under domestic measure, the foreign asset dynamics is given by*

$$dY_i/Y_i = (r^f - \rho_{B,Z_i}\sigma_S\sigma_{Y,i})dt + \sigma_{Y,i} dZ_i^d.$$

- *Under foreign measure, the domestic asset dynamics is given by*

$$dX_i/X_i = (r^d + \rho_{B,Zi}\sigma_S\sigma_{Y,i})dt + \sigma_{Y,i} dZ_i^f,$$

  where $dB^d dZ_i^f = \rho_{B,Z_i} dt$
- *Under domestic measure,*

$$d(SY_i)/(SY_i) = r^d dt + \sigma_{Y,i} dZ_i^d + \sigma_B dB^f.$$

- *Under foreign measure,*

$$d(X_i/S)/(X_i/S) = r^f dt + \sigma_{Y,i} dZ_i^d - \sigma_B dB^f.$$

Proof. Similar to the proof of Theorem 15.6.17. Let $N(t) = M^f(t)S(t)$ denote the foreign money account numeraire. Then

$$\begin{aligned}
E_{Q^f}[\frac{dX_t}{X_t}] &= E_{Q^d}[\frac{dX_t}{X_t}\frac{dQ^f}{dQ^d}] \\
&= E_{Q^d}[\frac{dX_t}{X_t}\frac{S(T)M^f(T)}{S(t)M^f(t)}\frac{M(t)}{M(T)}] \\
&= E_{Q^d}[\frac{dX_t}{X_t}(1 + d\ln(N(t)/M(t))] \\
&= E_{Q^d}[\frac{dX_t}{X_t}(1 + d\ln(N(t)/M(t))] \\
&= rdt + \rho_{XS}\sigma_X\sigma_S
\end{aligned}$$

We can similarly prove other facts.

(4) can be directly obtained from (1) and (2) using Ito process product rule. (5) can be directly obtained from (3) and (4) using Ito process product rule. □

Lemma 22.3.2 (cross-country change of numeraire for zero-coupon bonds). *[4, p. 700] Let $P^d(t,T), P^f(t,T)$ be zero-coupon bond prices in domestic currency and in foreign currencies such that*

$$P^d(t,T) = M(t)E_{Q^d}[\frac{1}{M(T)}|\mathcal{F}_t], P^f(t,T) = M^f(t)E_{Q^f}[\frac{1}{M^f(T)}|\mathcal{F}_t].$$

Let M be the domestic money account, M^f be the foreign money account and S the exchange rate.

- *The $S(T)P^f(t,T)$ is the foreign T-zero-coupon-bond numeraire. The Radon-Nikodym derivative to change measure from domestic T-zero-coupon-bond numeraire $P^d(t,T)$ to the foreign T-zero-coupon-bond numeraire $P^f(t,T)$ is given by*

$$Z(T) = \frac{dS_{T^d}}{dS_{T^f}} = \frac{S(T)P^f(T,T)}{S(t)P^f(t,T)}\frac{P^d(t,T)}{P^d(T,T)} = \frac{S(T)}{S(t)P^f(t,T)}P^d(t,T) = \frac{S(T)}{F(t,T)}$$

- $$1/Z(T) = \frac{dS_{T^f}}{dS_{T^d}} = \frac{F(t,T)}{S(T)}$$

- *(Todo!)*
$$E_{T^f}[Z_T|\mathcal{F}_t] = 1$$

- $$E_{T^f}[X] = E_{T^d}[X\frac{dS_{T^f}}{dS_{T^d}}]$$

- *Support $S(T)$ is a martingale under measure T^d and $\frac{dQ^f}{dQ^d}$ has log-normal process with volatility Σ_Z and correlation ρ with $S(t)$. Then*

$$E_{T^f}[S(T)|\mathcal{F}_t] = S(t)\exp(\rho\sigma_Z\sigma_S(T-t))$$

Proof. (1)(2) Use Theorem 15.6.15 . □

Lemma 22.3.3 (cross-country change of numeraire for zero-coupon bonds and money market account). *[4, p. 700] Let $P^d(t,T), P^f(t,T)$ be zero-coupon bond prices in domestic currency and in foreign currencies such that*

$$P^d(t,T) = M(t)E_{Q^d}[\frac{1}{M(T)}|\mathcal{F}_t], P^f(t,T) = M^f(t)E_{Q^f}[\frac{1}{M^f(T)}|\mathcal{F}_t].$$

Let M be the domestic money account, M^f be the foreign money account and S the exchange rate. The $S(T)P^f(t,T)$ is the foreign T-zero-coupon-bond numeraire.

It follows that

- $$M^f(t)E_{Q^f}[\frac{V(T)}{M^f(T)}|\mathcal{F}_t] = P^f(t,T)E_{T^f}[V(T)|\mathcal{F}_t].$$
- $$M(t)E_{Q^d}[\frac{V(T)}{M(T)}|\mathcal{F}_t] = P^d(t,T)E_{T^d}[V(T)|\mathcal{F}_t].$$
- *The Radon-Nikodym derivative to change measure from domestic T-zero-coupon-bond numeraire $P^d(t,T)$ to the foreign money market account numeraire $S(t)M^f(t)$is given by*
$$Z(T) = \frac{dS_{T_d}}{dQ^f} = \frac{P^d(T,T)}{P^d(t,T)}\frac{S(t)M^f(t)}{S(T)M^f(T)}.$$
- *The Radon-Nikodym derivative to change measure from foreign T-zero-coupon-bond numeraire $P^f(t,T)S(t)$ to the domestic money market account numeraire $M(t)$ is given by*
$$Z(T) = \frac{dS_{T_f}}{dQ^d} = \frac{S(T)P^f(T,T)}{S(t)P^f(t,T)}\frac{M(t)}{M(T)}.$$

Proof. The key theorem we are using is Theorem 15.6.15 (1) Note that

$$\frac{dS_{T^f}}{dQ^f} = \frac{S(T)P^f(T,T)}{S(t)P^f(t,T)}\frac{S(t)M^f(t)}{S(T)M^f(T)} = \frac{P^f(T,T)}{P^f(t,T)}\frac{M^f(t)}{M^f(T)},$$

then

$$\begin{aligned}
M^f(t)E_{Q^f}[\frac{V(T)}{M^f(T)}|\mathcal{F}_t] &= M^f(t)E_{T^f}[\frac{V(T)}{M^f(T)}\frac{dQ^f}{dS_{T^f}}|\mathcal{F}_t] \\
&= M^f(t)E_{T^f}[\frac{V(T)}{M^f(T)}\frac{dQ^f}{dS_{T^f}}|\mathcal{F}_t] \\
&= M^f(t)E_{T^f}[\frac{V(T)}{M^f(T)}\frac{P^f(t,T)}{P^f(T,T)}\frac{M^f(T)}{M^f(t)}|\mathcal{F}_t] \\
&= P^f(t,T)E_{T^f}[\frac{V(T)}{P^f(T,T)}|\mathcal{F}_t] \\
&= P^f(t,T)E_{T^f}[V(T)|\mathcal{F}_t]
\end{aligned}$$

(2) Similar to (1), note that

$$\frac{dS_{T^d}}{dQ^d} = \frac{P^d(T,T)}{P^d(t,T)}\frac{M^d(t)}{M^d(T)}.$$

(3)(4) straight forward. □

22.3.3 FX rate dynamics under different measure

22.3.3.1 *Two countries*

Theorem 22.3.3 (martingale property of spot rates under domestic measure). *Let current time be $t, t \leq T$.*

- $$S(t)\frac{M^f(t)}{M(t)} = E_{Q^d}[\frac{S(T)M^f(T)}{M(T)}|\mathcal{F}_t].$$
- *If the interest rates are constants given by r^d, r^f, we have*
  $$E_{Q^d}[S(T)|\mathcal{F}_t] = \exp((r^d - r^f)(T-t))S(t).$$
- *If we assume $S(t)$ is governed by geometric SDE, then under domestic risk-neutral measure, we have representation*
  $$dS(t)/S(t) = (r^d - r^f)dt + \sigma(t)dW^Q(t).$$

Proof. Under the domestic risk-neutral measure, the domestic value of the foreign money market account $M^f(T)S(T)$ will have martingale property

$$S(t)\frac{M^f(t)}{M(t)} = E_{Q^d}[\frac{S(T)M^f(T)}{M(T)}|\mathcal{F}_t].$$

□

Theorem 22.3.4 (martingale property of spot rates under domestic measure). *Let current time be $t, t \leq T$.*

- $$\frac{M(t)}{S(t)M^f(t)} = E_{Q^f}[\frac{M(T)}{M^f(T)S(T)}|\mathcal{F}_t].$$
- *If the interest rates are constants given by r^d, r^f, we have*
$$E_{Q^f}[\frac{1}{S(T)}|\mathcal{F}_t] = \exp((r^f - r^d)(T-t))\frac{1}{S(t)}.$$
- *If we assume $S(t)$ is governed by geometric SDE has representation under domestic risk-neutral measure given by*
$$dS(t)/S(t) = (r^d - r^f)dt + \sigma(t)dW^d(t).$$
If we assume $1/S(t)$ is governed by geometric SDE, then under foreign risk-neutral measure, we have representation
$$d(1/S(t))/(1/S(t)) = (r^f - r^d)dt - \sigma(t)dW^f(t).$$

Proof. (1) Under the domestic risk-neutral measure, the domestic value of the foreign money market account $M^f(T)S(T)$ will have martingale property

$$S(t)\frac{M^f(t)}{M(t)} = E_{Q^d}[\frac{S(T)M^f(T)}{M(T)}|\mathcal{F}_t].$$

(3) Using Ito quotient rule, we have

$$d(1/S(t))/(1/S(t)) = (r^f - r^d + \sigma^2(t))dt - \sigma(t)dW^d(t).$$

to do □

22.3.3.2 *Multiple countries*

Notation:

$$S^{A/B} \triangleq \text{number of units of currency } A \text{ required to purchase } 1 \text{ unit of currency } B.$$

Lemma 22.3.4 (triangle relationship).

- $$X_t^{A/B} = X_t^{A/C} \times X_t^{C/B}.$$
- *If*
$$dX_t^{A/C}/X_t^{A/C} = \mu^{A/C}dt + \sigma^{A/C}dW_t^{A/C}$$
$$dX_t^{C/B}/X_t^{C/B} = \mu^{C/B}dt + \sigma^{C/B}dW_t^{C/B},$$
then
$$dX_t^{A/B}/X_t^{A/B} = (\mu^{A/C} + \mu^{C/B} + \rho\sigma^{A/C}\sigma^{C/B})dt + \sigma^{A/C}dW_t^{A/C} + \sigma^{C/B}dW_t^{C/B}$$

Proof. (1) Use no-arbitrage argument. (2) Use **??**. □

Remark 22.3.1 (interpretation).

- In practice, exchange rates do not necessarily follow geometric Brownian motion; so above triangle relationship might not be useful.
- In theory, if geometric Brownian motion actually describes exchange rate dynamics, then two known exchange rate dynamics should fully determine the third.

22.3.4 Asset dynamics under different measure

22.3.4.1 *Foreign money money account under domestic measure*

Definition 22.3.2 (basic concepts).

- ***Domestic money account*** *denominated in domestic currency:*
$$M(t) = \exp(\int_0^t r(u)du).$$
- ***Foreign money account*** *denominated in foreign currency:*
$$M^f(t) = \exp(\int_0^t r^f(u)du).$$
- *The* ***exchange rate*** $S(t)$ *gives units on domestic currency per unit of foreign currency.*
$$dS(t) = \gamma(t)S(t)dt + \sigma_2(t)S(t)(\rho(t)dW_1(t) + \sqrt{1-\rho^2}dW_2),$$

where W_1 and W_2 are independent Brownian motions.

- *$M^f(t)S(t)$ gives the value of the foreign money account denominated in domestic currency.*

Lemma 22.3.5 (governing SDE for foreign money market account). *Assume exchange rate $S(t)$, under real-world measure, is given by*

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

And let $r^f(t)$ be the deterministic foreign interest rate. It follows that

- *Then the dynamics for the foreign money market account under real-world measure is*

$$d(M^f(t)S(t))/(M^f(t)S(t)) = (r^f(t) + \mu(t))dt + \sigma(t)dW(t).$$

- *Then the dynamics for the foreign money market account under domestic risk-neutral measure is*

$$d(M^f(t)S(t))/(M^f(t)S(t)) = r^d(t)dt + \sigma(t)dW^Q(t).$$

Therefore, investing in foreign money account becomes risky due to the risk from exchange rate.

Proof. (1) Ito rule.

$$\begin{aligned} d(M^f(t)S(t)) &= SdM^f + M^f dS + dM^f dS \\ &= SM^f rdt + M^f(\gamma(t)S(t)dt + \sigma_2(t)S(t)dW(t)) + 0 \end{aligned}$$

(2) In the domestic risk-neutral measure, the geometric SDE for $S(t)$ is given by (Theorem 22.3.3)

$$dS(t)/S(t) = (r^d - r^f)dt + \sigma(t)dW^Q(t).$$

□

Remark 22.3.2 (understand exchange rate).

- If we view the foreign currency as an asset, then its current price is S, and its price process **under domestic risk-neutral measure** is given by

$$dS(t) = S(t)[(r(t) - r^f(t)dt + \sigma_2\rho(t)d\hat{W}_1(t) + \sigma_2(t)\sqrt{1-\rho^2(t)}d\hat{W}_2].$$

- Moreover, we view it as a dividend-paying asset. Because holding one unit of foreign currency can generate continuous dividends by investing in the foreign money market.

22.3.4.2 *Asset dynamics for two countries*

Theorem 22.3.5 (no-arbitrage governing SDE for asset dynamics under domestic measure). *Assume domestic assets $X_1^d, X_2^d, ..., X_M^d$, exchange rate S, domestic money account M, foreign money account M^f, and foreign asset $Y_1, Y_2, ..., Y_M$, in the domestic measure, follow the below dynamics*

$$\begin{aligned}
dS_t/S_t &= (r^d - r^f)dt + \sigma_S dB^d \\
dX_i^d/X_i^d &= r^d dt + \sigma_{X,i} dW_1^d, i = 1,2,...,M \\
dY_i^f/Y_i^f &= r^f - \rho_{S,Y_i}\sigma_S\sigma_{Y,i}dt + \sigma_{Y,i}dZ_i^d, i = 1,2,....,N \\
dS_tY_i^f/S_tY_i^f &= r^d dt + \sigma_{Y,i}dZ_i^d + \sigma_S dB^d, i = 1,2,....,N \\
dM/M &= r^d dt \\
dM^f/M^f &= r^f dt
\end{aligned}$$

where

- $dW_i^d dW_j^d = \sigma_{ij}^X dt$
- $dZ_i^d dZ_j^d = \sigma_{ij}^Y dt$
- W_i^d*s and* Z_i^d*s are independent*
- $dW_i^d dB^d = \rho_{X,S}dt$, $dZ_i^d dB^d = \rho_{Y,S}dt$.

It follows that

- $X_i^d(t)/M(t), i = 1,2,...,M, S(t)Y_i^f(t)/M(t), j = 1,...,N$, $S(t)M^f(t)$ *are martingales under domestic risk-neutral measure.*
- $X_i^d(t)/S(t)M^f(t), i = 1,2,...,M, Y_i^f(t)/M^f(t), j = 1,...,N$, $M(t)/S(t)$ *are martingales under foreign risk-neutral measure. And they adopt dynamics*

$$\begin{aligned}
d(1/S_t)/(1/S_t) &= (r^f - r^d)dt - \sigma_S dB^f \\
dX_i^d/X_i^d &= (r^d + \rho_{S,X_i}\sigma_S\sigma_{X,i})dt + \sigma_{X,i}dW_1^f, i = 1,2,...,M \\
dY_i^f/Y_i^f &= r^f dt + \sigma_{Y,i}dZ_i^f, i = 1,2,....,N \\
dX_i^d/S_t/X_i^d/S_t &= r^f dt + \sigma_{X,i}dZ_i^f - \sigma_S dB^f, i = 1,2,....,N \\
dM/M &= r^d dt \\
dM^f/M^f &= r^f dt
\end{aligned}$$

  where
  - $dW_i^f dW_j^f = \sigma_{ij}^X dt$
  - $dZ_i^f dZ_j^f = \sigma_{ij}^Y dt$

– W_i^f s and Z_i^f s are independent
– $dW_i^f dB^f = \rho_{X,S}dt$, $dZ_i^f dB^f = \rho_{Y,S}dt$.

* *Therefore, for both domestic and foreign investors, there do not exist self-financing strategies that are arbitrages.*

Proof. □

22.3.4.3 *Asset dynamics for multiple countries*

Lemma 22.3.6 (extension to multiple assets of multiple countries).

22.3.4.4 *Discussion on existence of risk-neutral measure*

Lemma 22.3.7 (existence of domestic and foreign measures in the geometric SDE framework). *Assume domestic asset S_1, exchange rate S, domestic money account M, foreign money account M^f(measured in foreign currency), and foreign asset S_2(measured in foreign currency), in the real-world measure, follow the below dynamics*

$$\begin{aligned} dS_1^d/S_1^d &= \mu_1 dt + \sigma_1 dW_1 \\ dS_2^f/S_2^f &= \mu_2 dt + \sigma_2 dW_2 \\ dX/X &= \gamma dt + \sigma_X dW_3 \\ dM/M &= rdt \\ dM^f/M^f &= r^f dt \end{aligned}$$

where $dW_i dW_j = \rho_{ij}dt$.

It follows that

* *The foreign assets S_2^f, M^f have value dynamics in domestic currency given by*

$$\begin{aligned} d(XS_2^f)/(XS_2^f) &= (\mu_2 + \gamma + \rho_{23}\sigma_2\sigma_X)dt + \sigma_2 dW_2 + \sigma_X dW_3 \\ d(XM^f)/(XM^f) &= (r^f + \gamma)dt + \sigma_X dW_3 \end{aligned}$$

* *If there exists $\theta_1, \theta_2, \theta_2$ such that*

$$\begin{aligned} \sigma_1\theta_1 &= \mu_1 - r \\ \sigma_2\theta_2 + \sigma_X\theta_3 &= \mu_2 + \gamma - r + \rho_{23}\sigma_2\sigma_X \\ \sigma_X\theta_3 &= r^f - r + \gamma \end{aligned}$$

then there exist a domestic risk-neutral measure Q^d such that under which

$$\begin{aligned}\sigma_2 dW_2 &= \sigma_2 dW_2^d - (r^f - r + \gamma + \rho_{23}\sigma_2\sigma_X)dt \\ \sigma_X dW_3 &= \sigma_X dW_3^d - (\mu_2 - r^f)dt \\ \sigma_2 dW_2 + \sigma_X dW_3 &= \sigma_2 dW_2^d + \sigma_X dW_3^d - (\mu_2 + \gamma + \rho_{23}\sigma_2\sigma_3 - r)dt\end{aligned}$$

and

$$\begin{aligned}dS_1/S_1 &= rdt + \sigma_1 dW_1^d \\ d(S_2)/(S_2) &= (r^f - \rho_{23}\sigma_2\sigma_X)dt + \sigma_2 dW_2^d + \sigma_X dW_3^d \\ d(XS_2)/(XS_2) &= rdt + \sigma_2 dW_2^d + \sigma_X dW_3^d \\ d(XM^f)/(XM^f) &= rdt + \sigma_X dW_3^d \\ dS/S &= (r - r^f)dt + \sigma_X dW_3^d \\ d(1/X)/(1/X) &= (r^f - r + \sigma_X^2)dt - \sigma_X dW_3^d\end{aligned}$$

- *In the foreign measure (the existence is guaranteed by Theorem 15.6.15), we have*

$$\begin{aligned}dW_1^f &= dW_1^d - \rho_{13}\sigma_X dt \\ dW_2^f &= dW_2^d - \rho_{23}\sigma_X dt \\ dW_3^f &= dW_3^d - \sigma_X dt\end{aligned}$$

and

$$\begin{aligned}d(1/X)/(1/X) &= (r^f - r)dt - \sigma_X dW_3^f \\ dS_1/S_1 &= (r + \rho_{13}\sigma_1\sigma_X)dt + \sigma_1 dW_1^f \\ d(S_1/X)/(S_1/X) &= (r^f)dt + \sigma_1 dW_1^f \\ d(S_2)/(S_2) &= (r^f)dt + \sigma_2 dW_2^f \\ d(M^f)/(M^f) &= r^f dt \\ d(M/X)/(M/X) &= r^f dt - \sigma_X dW_3^f \\ dX/X &= (r - r^f + \sigma_X^2)dt + \sigma_X dW_3^f\end{aligned}$$

- *Under domestic measure, assets denominated in domestic currency S_1, SS_2, SM^f, all have drift r; under foreign measure, assets denominated in foreign currency $S_1/S, S_2, M/S$, all have drift r^f.*

Proof. Use the Brownian motion under different measure(Theorem 15.6.17), we have

$$\begin{bmatrix} dW_1^f \\ dW_2^f \\ dW_3^f \end{bmatrix} = \begin{bmatrix} dW_1^d \\ dW_2^d \\ dW_3^d \end{bmatrix} + \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \sigma_S \end{bmatrix}.$$

□

Remark 22.3.3 (asset interpretation).

- M/X is the value of domestic money account measured in foreign currency.
- S_1/X is the value of domestic asset measured in foreign currency.
- $M^f S$ is the value of foreign money account measured in domestic currency.
- $S_2 X$ is the value of foreign asset measured in domestic currency.

Remark 22.3.4 (inconsistence and arbitrage). Note that if there are no solutions of θ_1, θ_2 due to inconsistence, then there are arbitrage opportunities(15.5.4).

Lemma 22.3.8 (existence of domestic and foreign measures in the geometric SDE framework, multiple assets). *Assume domestic asset $X_1, X_2, ..., X_M$, exchange rate S, domestic money account M, foreign money account M^f, and foreign asset $Y_1, Y_2, ..., Y_M$, in the real-world measure, follow the below dynamics*

$$\begin{aligned} dX_i/X_i &= \mu_{X,i}dt + \sigma_{X,i}dW_1, i = 1,2,...,M \\ dY_i/Y_i &= \mu_{Y,i}dt + \sigma_{Y,i}dZ_i, i = 1,2,....,N \\ dS/S &= \gamma dt + \sigma_S dB \\ dM/M &= rdt \\ dM^f/M^f &= r^f dt \end{aligned}$$

where $dW_i dW_j = \Sigma_{ij}^X dt$, $dZ_i dZ_j = \Sigma_{ij}^Y dt$.

It follows that

- $$\begin{aligned} d(SY_i)/(SY_i) &= (\mu_{Y,i} + \gamma + \rho_{SY_i}\sigma_{Y_i}\sigma_S)dt + \sigma_{Y_i}dZ_i + \sigma_S dB \\ d(SM^f)/(SM^f) &= (r^f + \gamma)dt + \sigma_S dB \end{aligned}$$

- *If there exists $\theta_{X,i}, \theta_{Y,i}, \theta_S$ such that*

$$\begin{aligned}\sigma_1\theta_{X_i} &= \mu_{X_i} - r\\ \sigma_2\theta_{Y_i} + \sigma_S\theta_S &= \mu_{Y_i} + \gamma - r\\ \sigma_S\theta_S &= r^f - r + \gamma\end{aligned}$$

then there exist a domestic risk-neutral measure Q^d such that under which

$$\begin{aligned}\sigma_{Y,i}dZ_i &= \sigma_{Y,i}dZ_i^d - (r^f - r + \gamma)dt\\ \sigma_S dB &= \sigma_S dB^d - (\mu_2 - r^f)dt\\ \sigma_2 dW_2 + \sigma_S dW_3 &= \sigma_2 dW_2^d + \sigma_S dW_3^d - (\mu_2 + \gamma + \rho_{23}\sigma_2\sigma_3 - r)dt\end{aligned}$$

and

$$\begin{aligned}d(SS_2)/(SS_2) &= rdt + \sigma_2 dW_2^d + \sigma_S dW_3^d\\ d(SM^f)/(SM^f) &= rdt + \sigma_S dW_3^d\\ dS/S &= (r - r^f)dt + \sigma_S dW_3^d\end{aligned}$$

Remark 22.3.5 (induced drift when change of measure for correlated Brownian motion)**.** Corollary 6.7.7.1

22.3.5 Martingale pricing method

22.3.5.1 *General principle*

Theorem 22.3.6 (two equivalent methods to evaluate domestic value of foreign cash flow). *Consider a random foreign cash flow $V^f(T)$ at T. Let $V^f(t)$ denote its foreign value at time $t < T$. Let $V^d(t)$ denote its domestic value at time $t < T$. Then its t-value in domestic currency can be evaluated in the following two equivalent ways:*

- *(convert to domestic quote and then discount)*

$$V^d(t) = E_{Q^d}[\frac{M(t)S(T)V^f(T)}{M(T)}|\mathcal{F}_t].$$

- *(discount and then covert to domestic quote)*

$$\begin{aligned}V^d(t) &= S(t)V^f(t)\\ &= S(t)E_{Q^f}[\frac{M^f(t)V^f(T)}{M^f(T)}|\mathcal{F}_t]\end{aligned}$$

- *That is, we can either view $S(T)V^f(T)$ as a domestic asset or value $V^f(T)$ in foreign currency and then convert it to domestic value.*

Proof. Use Lemma 22.3.1 we can show that

$$\begin{aligned}
E_{Q^f}[\frac{S(t)V^f(T)M^f(t)}{M^f(T)}|\mathcal{F}_t] &= E_{Q^d}[\frac{S(t)V(T)M^f(t)}{M^f(T)}\frac{dQ^f}{dQ^d}|\mathcal{F}_t] \\
&= E_{Q^d}[\frac{S(t)V(T)M^f(t)}{M^f(T)}\frac{S(T)M^f(T)}{S(t)M^f(t)}\frac{M(t)}{M(T)}|\mathcal{F}_t] \\
&= E_{Q^d}[\frac{M(t)S(T)V(T)}{M(T)}|\mathcal{F}_t]
\end{aligned}$$

□

Theorem 22.3.7 (martingale pricing in foreign exchange models). *Consider an asset with payoff $V^d(T)$ at maturity T, denominated by domestic currency. Assume the existence of risk-neutral measure Q^d and Q^f. Let $V^d(t)$ denote its price denominated in domestic value at time $t < T$. It follows that*

- $$\frac{V^d(t)}{M(t)} = E_{Q^d}[\frac{V^d(T)}{M(T)}|\mathcal{F}_t].$$
- $$\frac{V^d(t)}{M^f(t)S(t)} = E_{Q^f}[\frac{V(T)}{M^f(T)Q^d(T)}|\mathcal{F}_t].$$
- *Consider a random foreign cash flow $V^f(T)$ at T. Let $V^f(t)$ denote its price denominated in foreign currency at time $t < T$. We have*

  $$V^f(t) = E_{Q^f}[\frac{M^f(t)V^f(T)}{M^f(T)}|\mathcal{F}_t].$$

  and

  $$V^d(t) = V^f(t)S(t).$$

Proof. Directly from Theorem 15.6.15, Theorem 22.3.2, and Theorem 22.3.6. Note that (3) is directly implied by (2) via

$$\frac{V^f(t)S(t)}{M^f(t)S(t)} = E_{Q^f}[\frac{V^d(T)S(T)}{M^f(T)S(T)}|\mathcal{F}_t]$$
$$\frac{V^f(t)}{M^f(t)} = E_{Q^f}[\frac{V^d(T)}{M^f(T)}|\mathcal{F}_t]$$
$$V^f(t) = M^f(t)E_{Q^f}[\frac{V^f(T)}{M^f(T)}|\mathcal{F}_t]$$

□

Theorem 22.3.8 (martingale pricing in foreign exchange models with stochastic interest rate). *Consider an asset with payoff $V(T)$ at maturity T, denominated by domestic currency. Assume the existence of risk-neutral measure Q^d and Q^f. It follows that*

- $$V(t) = P^d(t,T)E_{T^d}[V(T)|\mathcal{F}_t].$$
- $$V(t) = S(t)P^d(t,T)E_{T^f}[\frac{V(T)}{S(T)}|\mathcal{F}_t].$$
- *Consider a random foreign cash flow $V^f(T)$ at T. Let $V^f(t)$ denote its foreign value at time $t < T$. Let $V^d(t)$ denote its domestic value at time $t < T$. We have*

$$V^d(t) = S(t)P^f(t,T)E_{T^f}[V^f(T)|\mathcal{F}_t] = S(t)V^f(t)$$
$$V^f(t) = P^f(t,T)E_{T^f}[V^f(T)|\mathcal{F}_t]$$

Proof. (1) To change the domestic risk-netural measure to domestic forward measure, we have (1) To change the foreign risk-netural measure to foreign forward measure, we have

□

22.4 Cross-country interest rate modeling

22.4.1 Basic concepts

Definition 22.4.1 (zero coupon bond).

- $P^d(t,T)$ *value at* t*, denominated in domestic currency, of a zero-coupon bond paying 1 unit domestic currency at maturity* T*.*
- $P^f(t,T)$ *value at* t*, denominated in foreign currency, of a zero-coupon bond paying 1 unit foreign currency at maturity* T*.*

Lemma 22.4.1 (zero-coupon bond dynamics under domestic and foreign measure).

- $$P^d(t,T) = M(t)E_{Q^d}[\frac{1}{M(T)}|\mathcal{F}_t].$$
  $$P^f(t,T) = M^f(t)E_{Q^f}[\frac{1}{M^f(T)}|\mathcal{F}_t].$$
  $$P^f(t,T) = S(t)M^f(t)E_{Q^d}[\frac{1}{M^f(T)S(T)}|\mathcal{F}_t].$$
- *Under domestic measure* Q^d*,*
  $$dP^d(t,T)/P^d(t,T) = r^d dt + \sigma_d dW_t^d$$
  $$dP^f(t,T)/P^f(t,T) = (r^f - \rho_{Sd}\sigma_S\sigma_d)dt + \sigma_d dY_t^d$$
- *Under foreign measure* Q^f*,*
  $$dP^d(t,T)/P^d(t,T) = rdt + \sigma_d dW_t^d$$
  $$dP^f(t,T)/P^f(t,T) = r^f dt + \sigma_d dY_t^f$$

Proof. See Theorem 22.3.2. □

22.4.2 Cross-country short rate dynamics

22.4.2.1 *Principles*

Theorem 22.4.1 (foreign short rate dynamics under domestic measure). *[5] Let the foreign short rate model, under foreign measure, have dynamics given by*

$$dr_t^f = mdt + \sigma_r dW_r^f,$$

such that the foreign zero coupon bond price dynamics is given by (Theorem 19.5.2)

$$dP^f(t,T)/P^f(t,T) = r^f dt + \sigma_r \frac{\partial P^f}{dr^f} dW_r^f.$$

It follows that

- *Under the domestic risk-neutral measure,*

$$dP^f(t,T)/P^f(t,T) = (r^f - \rho\sigma_r \frac{\partial P^f}{dr^f}\sigma_S)dt + \sigma_r \frac{\partial P^f}{dr^f} dW_t^d,$$

where W_t^d is a Brownian motion under risk-neutral measure, and under domestic risk-neutral measure,

$$dW_r^f = dW_r^d - \rho\sigma_S.$$

- *Under the domestic risk-neutral measure,*

$$dr_t^f = (m - \rho\sigma_r\sigma_S)dt + \sigma_r dW_r^d,$$

Proof. (1) Under the domestic risk-neutral measure, the FX rate has representation

$$dS(t)/S(t) = (r^d - r^f)dt + \sigma_S dW_S^d.$$

Then

$$d(P^f(t,T)S(t)/M^d(t))/(P^f(t,T)S(t)/M^d(t)) = \sigma_r \frac{\partial P^f}{dr^f} dW_r^f + \sigma_S dW_S^d + \rho\sigma_r \frac{\partial P^f}{dr^f}\sigma_S dt,$$

In order for $P^f(t,T)S(t)/M^d(t)$ to be a martingale under domestic risk-neutral measure, we need to have

$$dW_r^f = dW_r^d - \rho\sigma_S dt.$$

(we can also derive this using Theorem 15.6.17.)

(2) directly from (1). □

22.4.2.2 *Hull-White model*

Lemma 22.4.2 (cross-currency Hull-White model under domestic measure). *Let there be $N+1$ countries. Assume the domestic short rate dynamics, under the domestic risk-neutral measure Q^d, is given by Hull-White model:*

$$dr_0 = (\theta_0(t) - k_0(t)r_0(t))dt + \sigma_0(t)dW_0(t),$$

where the subscript 0 denote domestic and

$$\theta_0(t) = \frac{\partial f_0(0,t)}{\partial t} + \kappa_0(t)f_0(0,t) + \phi_0(t)$$
$$\phi_0(t) = \int_0^t \sigma_0^2(u)\exp(-2\beta_0(u,t)dt)du$$
$$\beta_0(t,T) = \int_t^T k_0(u)du$$

Similarly, let the short rate model in these N countries, under their associated foreign measure, have dynamics given by

$$dr_t^{f,i} = (\theta_i - k_i(t)r_i(t))dt + \sigma_i dW_i(t).$$

Assume the exchange rate $S_i(t)$ of the ith foreign currency with respect to domestic currency, under the domestic risk-neutral measure Q^d, is given by

$$dS_i/S_i = (r_0(t) - r_i(t))dt + \xi_i(t)dZ_i, i = 1,2,...,N.$$

Then the short rate dynamics in those foreign countries under ***domestic measure*** *are given by*

$$dr_t^{f,i} = (\theta_i - k_i(t)r_i(t) - \rho_{Si,ri}\sigma_i(t)\xi_i(t))dt + \sigma_i dW_i(t), i = 1,2,...,N,$$

where $E[dW_i(t)dZ_i] = \rho_{Si,ri}dt$.

Proof. See Theorem 22.4.1. □

22.4.3 Cross-country HJM framework

Lemma 22.4.3 (foreign forward rate dynamics under domestic measure). *Let the foreign forward rate model, under foreign measure, have dynamics given by*

$$df^f(t,T) = \alpha(t,T)dt + \sigma(t,T)dW(t),$$

where $W(t)$ is a Brownian motion and

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du.$$

The foreign zero coupon bond price dynamics is given by

$$dP^f(t,T)/P^f(t,T) = r^f(t)dt + \Sigma(t,T)dW(t)$$

with

$$\Sigma(t,T) = -\int_t^T \sigma(t,u)du.$$

It follows that

- *Under the domestic measure,*

$$dP^f(t,T)/P^f(t,T) = (r^f - \rho\sigma_S\Sigma(t,T))dt + \Sigma(t,T)dW_t^d.$$

- *Under the domestic measure,*

$$df_t^f = (\alpha(t,T) - \rho\sigma(t,T)\sigma_S)dt + \sigma(t,T)dW_t^d,$$

Proof. See the change of measure for dynamic models(Theorem 22.3.2). □

22.5 Black-Scholes PDE method

Theorem 22.5.1 (Black-Scholes equation for foreign assets). *[2, p. 189] Assume an foreign asset S_t has price process(denominated in foreign currency)*

$$dS = \mu_S S dt + \sigma_S S dW_1.$$

Assume the exchange rate $X(t)$ is governed by

$$dX = \mu_X X dt + \sigma_X S dW_2.$$

Let r^f and r denote short rates in foreign money account and domestic money account.

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma_S^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma_S\sigma_X SX \frac{\partial^2 V}{\partial S \partial X} + \frac{1}{2}\sigma_X^2 X^2 \frac{\partial^2 V}{\partial X^2} +$$
$$X(r - r^f)\frac{\partial V}{\partial S} + S(r^f - \rho\sigma_S\sigma_X)\frac{\partial V}{\partial S} - rV = 0.$$

Proof. We consider a portfolio

$$\Pi = V(S, X, t) - \Delta_X S M^f - \Delta_S X S,$$

where $\Delta_X M^f X$ represents buying Δ_X units of foreign currency investing in foreign money market, and $\Delta_X M^f X$ represents buying Δ_S units of foreign asset.

$$\begin{aligned} d\Pi =& (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma_X^2 X^2 \frac{\partial^2 V}{\partial X^2} + \rho\sigma_S\sigma_X SX \frac{\partial^2 V}{\partial X \partial S} + \frac{1}{2}\sigma_S^2 S^2 \frac{\partial^2 V}{\partial S^2} \\ & - \rho\sigma_S\sigma_X \Delta_X \Delta_S - r^f \Delta_X M^f X)dt + (\frac{\partial V}{\partial S} - \Delta_X M^f - \Delta_S S)dX \\ & + (\frac{\partial V}{\partial S} - \Delta_S X)dS \end{aligned}$$

where we use $d(\Delta_S M^f X) = \Delta_X X M^f r^f dt + \Delta_S M^f dX$ and $d(\Delta_S S X) = \Delta_S S dX + \Delta_S X dS + \Delta_S dX dS$.

To eliminate risks, we choose

$$\Delta_X M^f = \frac{\partial V}{\partial S} - \frac{S}{X}\frac{\partial V}{\partial S}, \Delta_S = \frac{1}{X}\frac{\partial V}{\partial S},$$

and set growth rate for $d\Pi = r\Pi dt = r(V - \Delta_S M^f X - \Delta_S S X)dt$, we have

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma_S^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma_S\sigma_X SX \frac{\partial^2 V}{\partial S \partial X} + \frac{1}{2}\sigma_X^2 X^2 \frac{\partial^2 V}{\partial X^2} +$$
$$X(r - r^f)\frac{\partial V}{\partial S} + S(r^f - \rho\sigma_S\sigma_X)\frac{\partial V}{\partial S} - rV = 0.$$

□

Note 22.5.1 (connection via Feyman-Kac theorem). From Theorem 15.8.6, we know that there exists a domestic measure such that S and X are governed by

$$\begin{aligned} dS &= S(r - r^f)dt + \sigma_S S dW_1 \\ dX &= X(r^f - \rho\sigma_S\sigma_S)dt + \sigma_X S dW_2 \end{aligned}$$

where W_1 and W_2 are correlated Brownian motions with $dW_1 dW_2 = \rho dt$. which is the same as [6, p. 276] as martingale method in domestic risk-neutral measure.

The stock price dominated in domestic currency is given by

$$\begin{aligned} d(XS) &= SdS + SdS + dSdS \\ &= XS(rdt + (\sigma_S dW_1 + \sigma_S dW_2)_) \end{aligned}$$

which has drift r.

22.6 Pricing examples

22.6.1 FX forward contract and FX forward rate

Definition 22.6.1 (FX forward contract and FX forward rate).

- ***FX forward contracts*** *are transactions in which two party agree to exchange a specified amount of different currencies at some future date, with the exchange rate K being set at the time the contract is entered into.*
- *The payoff denominated in domestic currency at the delivery date T for a domestic investor purchasing foreign currency via a forward contract is given by*

$$(\underbrace{N \cdot S(T)}_{\text{domestic value of N foreign currency}} - \underbrace{K \cdot N}_{\text{demestic currency paid}}).$$

- *The fixed rate K such that forward contract has zero value is called forward FX rate.*
- *Forward FX rate can be viewed as the forward price of 1 unit foreign currency.*

Remark 22.6.1 (symmetric relationship on forwards).

- To long CCY1/CCY2 forward contract (buy CCY1 by selling CCY2) is equivalent to short CCY2/CCY1 (sell CCY2 and receive CCY1) forward contract.
- To long CCY2/CCY1 forward contract is equivalent to short CCY1/CCY2 forward contract.

Theorem 22.6.1 (no arbitrage constraint on forward FX rate). *Let current time be t and let the forward delivery date be T. Let $S(t)$ be the spot rate and $F(t,T)$ be the FX forward rate. Denote zero-coupon bond prices in domestic currency and in foreign currency are given by*

$$P^d(t,T) = M(t)E_{Q^d}[\frac{1}{M(T)}|\mathcal{F}_t], P^f(t,T) = M^f(t)E_{Q^f}[\frac{1}{M^f(T)}|\mathcal{F}_t].$$

- *The no-arbitrage condition requires that*

$$F(t,T) = S(t)\frac{P^f(t,T)}{P^d(t,T)}.$$

- *(martingale property) The FX forward rate $F(t,T)$ is a martingale under the martingale measure associated with domestic zero-coupon bond $P^d(t,T)$. And it satisfy*

$$F(t',T) = E_{T^d}[S(T)|\mathcal{F}_{t'}] = E_{T^d}[F(T,T)|\mathcal{F}_{t'}],$$

and

$$E[F(t',T)|\mathcal{F}_t] = E[S(T)|\mathcal{F}_t], t \leq t' < T.$$

- *Assume constant interest rate.*[a] *Under domestic risk-neutral measure,*

$$F(t,T) = E_{Q^d}[S(T)|\mathcal{F}_t] = e^{(r+r^f)(T-t)}S(t),$$

and $F(t,T)$ is a martingale.

[a] the case of deterministic interest rate can be similarly derived.

Proof. (1)(a) (no-arbitrage replication method) We can replicate the forward contract payoff in the following way:

- At time t, borrow $X(t)P^f(t,T)$ amount of domestic currency and buy $1/P^f(t,T)$ unit of foreign currrency; the $1/P^f(t,T)$ unit of foreign currency will buy 1 unit of foreign zero-coupon bond.
- At time T, get 1 unit of foreign currency from foreign money market; at the same time pay for the principle and the interest cost in total $X(t)P^f(t,T)/P^d(t,T)$.

Therefore, the forward FX rate (i.e., forward price of 1 unit foreign currency) is given by

$$F(t,T) = X(t)P^f(t,T)/P^d(t,T)$$

. (b) (martingale method)

$$\begin{aligned}
&M(t)E_{Q^d}[(S(T) - F(t,T))/M(T)|\mathcal{F}_t] = 0 \\
\implies &M(t)E_{Q^d}[S(T)/M(T)|\mathcal{F}_t] = M(t)F(t,T)E_{Q^d}[1/M(T)] = F(t,T)P^d(t,T)
\end{aligned}$$

Use change of measure(Theorem 15.6.15), we have

$$\begin{aligned}
E_{Q^d}[\frac{M(t)S(T)}{M(T)}|\mathcal{F}_t] &= E_{Q^f}[\frac{M(t)S(T)}{M(T)}\frac{dQ^d}{dQ^f}|\mathcal{F}_t] \\
&= E_{Q^f}[\frac{M(t)S(T)}{M(T)}\frac{M(T)}{M(t)}\frac{M^f(t)S(t)}{M^f(T)S(T)}|\mathcal{F}_t] \\
&= E_{Q^f}[\frac{M^f(t)S(t)}{M^f(T)}|\mathcal{F}_t] \\
&= S(t)E_{Q^f}[\frac{M^f(t)}{M^f(T)}|\mathcal{F}_t] \\
&= S(t)P^f(t,T)
\end{aligned}$$

(2)

$$
\begin{aligned}
E_{Q^d}[M(t)(S(T)-F(t,T))/M(T)|\mathcal{F}_t] &= 0\\
P^d(t,T)E_{T^d}[S(T)-F(t,T)|\mathcal{F}_t] &= 0\\
\implies F(t,T) = E_{T^d}[S(T)|\mathcal{F}_t]&.
\end{aligned}
$$

(3) Use the result from **??**. (1)

$$E_{Q^d}[e^{-rT}(S(T)-F)|\mathcal{F}_t] = 0 \implies E_{Q^d}[S(T)|\mathcal{F}_t] = F(t,T)$$

Then we use Theorem 22.3.3 to get

$$F(t,T) = e^{(r+r^f)(T-t)}S(t)$$

□

Remark 22.6.2 (from the perspective of foreign investors).

- Under foreign risk-neutral measure, the forward price F at time t (measured in foreign currency) of one unit of domestic currency, to be delivered at time T, is determined by

$$E_{Q^f}[e^{r^f T}(\frac{1}{S(T)} - F^f(t,T)|\mathcal{F}_t] = 0,$$

or equivalently,

$$F^f(t,T) = E_{Q^f}[\frac{1}{S(T)}|\mathcal{F}_t] = e^{(r^f-r)(T-t)}\frac{1}{S(t)},$$

where $F^f(t,T)$ is a martingale under foreign risk-neutral measure.
- forward rate is martingale under forward measure

Note 22.6.1 (triangle relationship between forward exchange, domestic zero-curve, and foreign zero-curve). Note that $F(t,T), P^d(t,T), P^f(t,T), S(t)$ are not independent. Their relation is given by

$$F(t,T) = S(t)\frac{P^f(t,T)}{P^d(t,T)}.$$

Lemma 22.6.1 (value of a forward contract). *Let current time be t. Consider a unit notional forward contract starting at $t_0 \leq t$, with forward exchange rate $F(t_0,T)$ and*

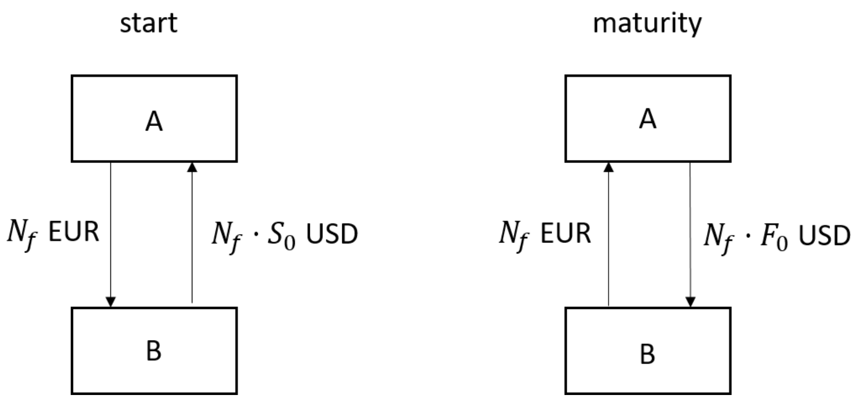


Figure 22.6.1: An illustration of FX swap. At the begining, two parties, A and B, will exchange equivalent value of currencies. At the maturity of the FX swap, the two parties will exchange the currencies with amount set by the FX forward rate.

delivery date T. Its value(long position) denominated in domestic currency at time t is given by

$$V(t) = P^d(t,T)(F(t,T) - F(t_0,T)) = S(t)P^d(t,T) - P^d(t,T)F(t_0,T).$$

Proof.

$$\begin{aligned} V(t) &= P^d(t,T)E_{T^d}[(S(T) - F(t_0,T))|\mathcal{F}_t] \\ &= P^d(t,T)E_{T^d}[(F(T,T) - F(t_0,T))|\mathcal{F}_t] \\ &= P^d(t,T)(F(t,T) - F(t_0,T)) \end{aligned}$$

where we used the fact that forward rate is a martingale under forward measureTheorem 22.6.1. □

22.6.2 FX swaps

Definition 22.6.2 (FX swap).
- *An FX swap agreement is a contract in which one party borrows one currency from, and simultaneously lends another to, the second party. Each party uses the repayment obligation to its counterparty as collateral and the amount of repayment is fixed at the FX forward rate as of the start of the contract. Thus, FX swaps can be viewed as FX risk-free collateralised borrowing/lending.*
- *An FX swap is then a regular spot FX trade combined with a forward trade.*

22.6.3 Value of cash flows

Lemma 22.6.2 (domestic discounted value of fixed foreign cash flow). *Consider fixed foreign cash flow at future times $T_1, T_2, ..., T_n$ given by $V_1, V_2, ..., V_n$. Then at $t < T_1$ the domestic value is given by*

$$V(t) = S(t) \sum_{i=1}^{n} P^f(t, T_i) V_i = V_i F(t, T_i) P^d(t, T_i),$$

where $F(t, T_i)$ is the forward exchange rate with maturity date T_i.

Proof. Note that we use the triangle relationship(Theorem 22.6.1)

$$F(t, T) = S(t) \frac{P^f(t, T)}{P^d(t, T)}.$$

□

Lemma 22.6.3 (domestic value random foreign cash flow). *Consider random foreign cash flow at future times $T_1, T_2, ..., T_n$ given by $V^f(T_1), V^f(T_2), ..., V^f(T_n)$. Then at $t < T_1$ the domestic value is given by*

$$V^d(t) = S(t) \sum_{i=1}^{n} V_i^f(t),$$

where

$$V_i^f(t) = M^f(t) E_{Q^f}[\frac{V(T_i)}{M^f(T)} | \mathcal{F}_t].$$

Other equivalent formulation includes:

- $$V^d(t) = \sum_{i=1}^{n} P^d(t, T_i) \sum_{i=1}^{n} E_{T_i^d}[S(T_i) V(T_i) | \mathcal{F}_t].$$

- $$V^d(t) = M(t) \sum_{i=1}^{n} E_{Q^d}[S(T_i) V(T_i) / M(T_i) | \mathcal{F}_t].$$

Proof. (1) Note that $V_i^f(t) = M^f(t) E_{Q^f}[\frac{V(T_i)}{M^f(T)} | \mathcal{F}_t]$ is the present value dominated in foreign currency for foreign cash flow $V(T_i)$ based on the martingale pricing method(Theorem 22.3.7). (2)(3) Directly from the equivalent approach for evaluating domestic values of foreign cash flow. □

22.6.4 Cross currency swaps

Definition 22.6.3 (fixed-fixed cross-currency swap). *[7, p. 210]Given a set of dates $T_0, T_1, ..., T_m$ and denote $\tau_i = T_i - T_{i-1}$. A typical fixed-fixed cross currency swap between two counterparties domestic A and foreign B involves three phases:*

- *Initial exchange:*
  - *At T_0, B pays N^f in foreign currency to A.*
  - *At T_0, A pays N in domestic currency to B.*
- *Running period:*
  - *At $T_i, i = 1, 2, ..., M$, B pays fixed $N\tau_i K_i$ in domestic currency to A.*
  - *At $T_i, i = 1, 2, ..., M$, A pays fixed $N^f \tau_i K_i^f$ in foreign currency to B.*
- *Final exchange:*
  - *At T_M, B pays fixed N in domestic currency to A.*
  - *At T_M, A pays fixed N^f in foreign currency to B.*

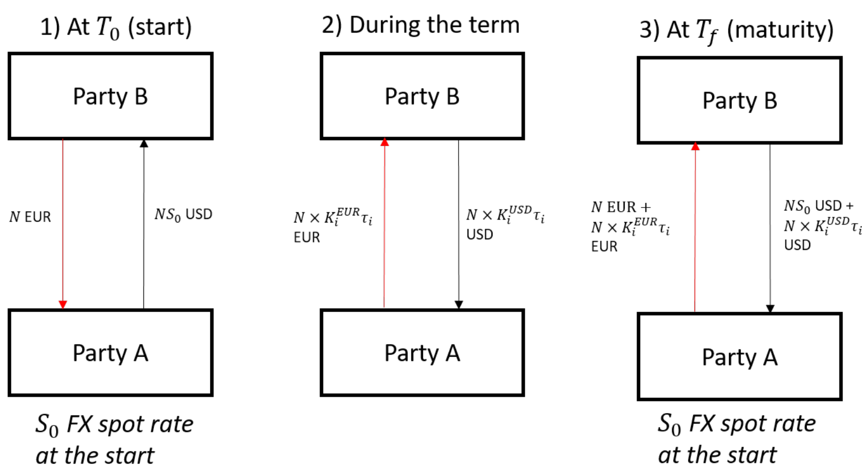


Figure 22.6.2: Cross currency fixed-fixed swap.

Remark 22.6.3 (application of cross currency swap). link

- A cross currency swap allows two parties to borrow foreign currency by issuing domestic bonds (borrowing foreign currency sometimes can be more difficult for companies than borrowing domestic money.)
- Suppose the British Petroleum Company needs \$150 million US dollars to finance its new refining facility in the U.S. Also, suppose that the Piper Shoe Company, a US company, plans to issue 100 million pounds to set up its distribution center in London. To meet each other's needs, suppose that both companies go to a swap bank that sets up the following agreements:

- **Initial stage:** The British Petroleum Company will issue 5-year 100 million pound bonds with 7.5% interest rate. It will then deliver the 100 million pound to the swap bank who will pass it on to the U.S. Piper Company to finance the construction of its British distribution center. The Piper Company will issue 5-year 150 million dollar bonds paying 10% interest. The Piper Company will then pass the 150 million dollars to swap bank that will pass it on to the British Petroleum Company who will use the funds to finance the construction of its U.S. refinery.
- **Running periods:** The British company, with its U.S. asset (refinery), will pay the 10% interest on $150 million to the swap bank who will pass it on to the American company so it can pay its U.S. bondholders. The American company, with its British asset , will pay the 7.5% interest on 100 million pounds, to the swap bank who will pass it on to the British company so it can pay its British bondholders.
- **Final exchange:** At maturity, the British company will pay $150 million to the swap bank who will pass it on to the American company so it can pay its U.S. bondholders. At maturity, the American company will pay 100 million pound to the swap bank who will pass it on to the British company so it can pay its British bondholders.

Lemma 22.6.4 (valuing fixed-fixed cross-currency swap). *Consider a fixed-fixed cross-currency swap where two parties domestic A and foreign B exchange cash flow at a set of dates* $T_0, T_1, ..., T_m$*(denote* $\tau_i = T_i - T_{i-1}$*). Then from the perspective of A, we have*

- *the initial exchange has value denominated in domestic currency*

$$V_{init}(t) = \underbrace{-NP^d(t, T_0)}_{pay\ to\ B} + \underbrace{N^f P^f(t, T_0)S(t)}_{receive\ from\ B}.$$

- *the running period has value denominated in domestic currency*

$$V_{run}(t) = \underbrace{-\sum_{i=1}^{M} N^f \tau_i K_i P^f(t, T_i)S(t)}_{pay\ to\ B} + \underbrace{\sum_{i=1}^{M} N\tau_i K_i P^d(t, T_i)}_{receive\ from\ B}.$$

- *the final exchange has value denominated in domestic currency*

$$V_{final}(t) = \underbrace{-N^f P^f(t, T_M)S(t)}_{pay\ to\ B} + \underbrace{NP^d(t, T_M)}_{receive\ from\ B}.$$

- *the current value of the swap at time t is*

$$\begin{aligned} V(t) &= V_{init}(t) + V_{run}(t) + V_{final}(t) \\ &= N(-P^d(t,T_0) + \sum_{i=1}^{M} N\tau_i K_i P^d(t,T_i) + P^d(t,T_M)) \\ &+ N^f S(t)(-P^f(t,T_0) + \sum_{i=1}^{M} N\tau_i K_i P^f(t,T_i) + P^f(t,T_M)) \end{aligned}$$

Proof. Straight forward application of cash discounting(Lemma 22.6.2). □

Remark 22.6.4 (other equivalent evaluation method). Note that under no-arbitrage condition, there are always two equivalent ways to evaluate the domestic value of foreign cash flow(Theorem 22.3.6). For example,

- For the initial exchange, we can also write

$$V_{init}(t) = -NP^d(t,T_0) + N^f P^f(t,T_0)S(t) = -NP(t,T_0) + N^f F(t,T_0)P^d(t,T_0).$$

where $F(t,T)$ is the forward exchange rate and we have used the triangle relation **(??)**

$$F(t,T) = S(t)\frac{P^f(t,T)}{P^d(t,T)}.$$

- Similarly, for the running period, we can also write

$$\begin{aligned} V_{run}(t) &= -\sum_{i=1}^{M} N^f \tau_i K_i P^f(t,T_i)S(t) + \sum_{i=1}^{M} N\tau_i K_i P(t,T_0) \\ &= -\sum_{i=1}^{M} N^f \tau_i K_i P^d(t,T_i)F(t,T_i) + \sum_{i=1}^{M} N\tau_i K_i P(t,T_0) \end{aligned}$$

22.6.5 Cross-currency basis swaps

Definition 22.6.4 (cross-currency basis swap). *A cross-currency basis swap is a floating/floating swap where two parties borrow from and simultaneously lend to each other an equivalent amount of money denominated in two different currencies for a predefined period of time. Given a set of dates $T_0, T_1, ..., T_m$ and denote $\tau_i = T_i - T_{i-1}$. A typical float-float cross currency basis swap between two counterparties A and B involves three phases:*

- *Initial exchange:*
  - *At T_0, B pays $N^f = N/S_0$ in foreign currency to A, where S_0 is FX spot rate at T_0.*

- At T_0, A pays N in domestic currency to B.
- *Running period:*
  - At $T_i, i = 1, 2, ..., M$, B pays fixed $N\tau_i L^d(T_{i-1}, T_i)$ in domestic currency to A.
  - At $T_i, i = 1, 2, ..., M$, A pays fixed $N^f \tau_i L^f(T_{i-1}, T_i)$ in foreign currency to B.
- *Final exchange:*
  - At T_M, B pays fixed N in domestic currency to B.
  - At T_M, A pays fixed N^f in foreign currency to B.

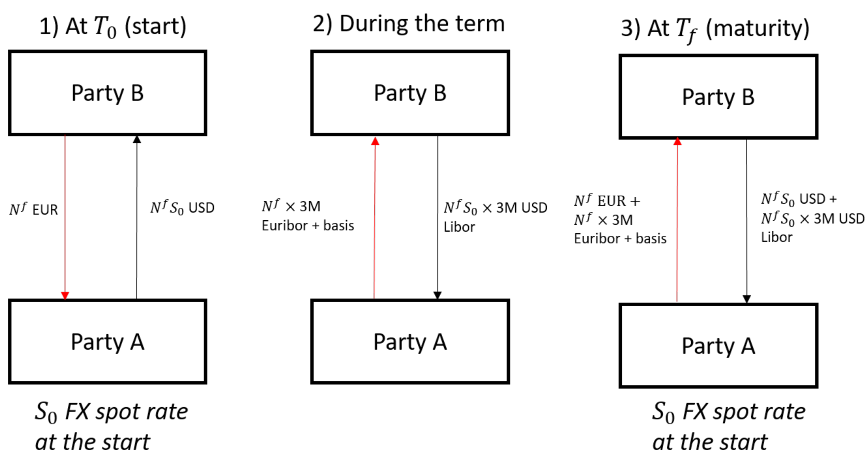


Figure 22.6.3: Cross currency basis swap.

Remark 22.6.5 (understand the cash flow in cross-currency basis swap)**.** Let A be the domestic party. The cash flows from A's point of view are

- For undiscounted domestic currency leg(cash flow in domestic currency)
  - At T_0, $C_d(T_0) = -S(T_0)N_f$
  - At $T_i, i = 1, 2, ..., M$, $C_d(T_i) = L^d(T_{i-1}, T_i)\tau_i S(T_0)N^f$.
  - At T_M, $C_d(T_M) = S(T_0)N^f$.
- For undiscounted foreign currency leg
  - At T_0, $C_f(T_0) = N_f$
  - At $T_i, i = 1, 2, ..., M$, $C_f(T_i) = -L^f(T_{i-1}, T_i)\tau_i N^f$
  - At T_M, $C_f(T_M) = -N_f$

Lemma 22.6.5 (valuing float-float cross-currency swap). *Consider a fixed-fixed cross-currency swap where two parties domestic A and foreign B exchange cash flow at a set of dates $T_0, T_1, ..., T_M$(denote $\tau_i = T_i - T_{i-1}$). Then from the perspective of A, we have*

- *the initial exchange has value denominated in domestic currency*

$$V_{init}(t) = \underbrace{-NP^d(t,T_0)}_{pay\ to\ B} + \underbrace{N^f P^f(t,T_0)S(t)}_{receive\ from\ B}.$$

- *the running period has value denominated in domestic currency*

$$\begin{aligned} V_{run}(t) &= \underbrace{-\sum_{i=1}^{M} N^f S(t) P^f(t,T_i)\tau_i E_{T^f}[L^f(T_{i-1},T_i)|\mathcal{F}_t]}_{pay\ to\ B} + \underbrace{\sum_{i=1}^{M} N\tau_i P^d(t,T) E_{T^d}[L^d(T_{i-1},T_i)]}_{receive\ from\ B} \\ &= -\sum_{i=1}^{M} N^f S(t) P^f(t,T_i)\tau_i F^f(t,T_{i-1},T_i) + \sum_{i=1}^{M} N\tau_i P^d(t,T) F^d(,tT_{i-1},T_i) \end{aligned}$$

- *the final exchange has value denominated in domestic currency*

$$V_{final}(t) = \underbrace{-N^f P^f(t,T_M)S(t)}_{pay\ to\ B} + \underbrace{NP^d(t,T_M)}_{receive\ from\ B}.$$

- *the current value of the swap at time t is*

$$\begin{aligned} V(t) &= V_{init}(t) + V_{run}(t) + V_{final}(t) \\ &= \underbrace{N(-P^d(t,T_0) + \sum_{i=1}^{M} N\tau_i F^d(t,T_{i-1},T_i)P^d(t,T_i) + P^d(t,T_M))}_{domestic\ leg} \\ &\quad + \underbrace{N^f S(t)(-P^f(t,T_0) + \sum_{i=1}^{M} N\tau_i F^f(t,T_{i-1},T_i)P^f(t,T_i) + P^f(t,T_M))}_{foreign\ leg} \end{aligned}$$

Proof. Straight forward application of cash discounting(Lemma 22.6.2). □

Definition 22.6.5 (mark-to-market(MtM) cross-currency basis swap). *A mark-to-market(MtM) cross-currency basis swap is a floating/floating swap where two parties borrow from and simultaneously lend to each other an equivalent amount of money denominated in two different currencies for a predefined period of time.*

The major currency has a notional resettable leg, while the minor currency

Given a set of dates $T_0, T_1, ..., T_m$ *and denote* $\tau_i = T_i - T_{i-1}$. *Assume the domenstic currency has a notional resettable leg, a typical MtM cross currency swap between two counterparties foreign A and domestic B involves three phases:*

- *Initial exchange:*
  - *At T_0, B pays N^f in foreign currency to A,*
  - *At T_0, A pays $N^f \times S_0$ in domestic currency to B, where S_0 is FX spot rate at T_0.*
- *Running period:*
  - *At $T_i, i = 1, 2, ..., M$, B pays $N^f S_{i-1}\tau_i L_{dom}(T_i, T_{i+1})$ in domestic currency to A. B pays $N^f \times S_{i-1}$ domesitc currency to A, and B receives $N^f \times S_i$ domesitc currency from B.*
  - *At $T_i, i = 1, 2, ..., M$, A pays fixed $N\tau_i L_{for}(T_i, T_{i+1})$ in foreign currency to B.*
- *Final exchange:*
  - *At T_M, B pays $N^f \times S_{M-1}$ in domestic currency to A.*
  - *At T_M, A pays fixed N^f in foreign currency to B.*

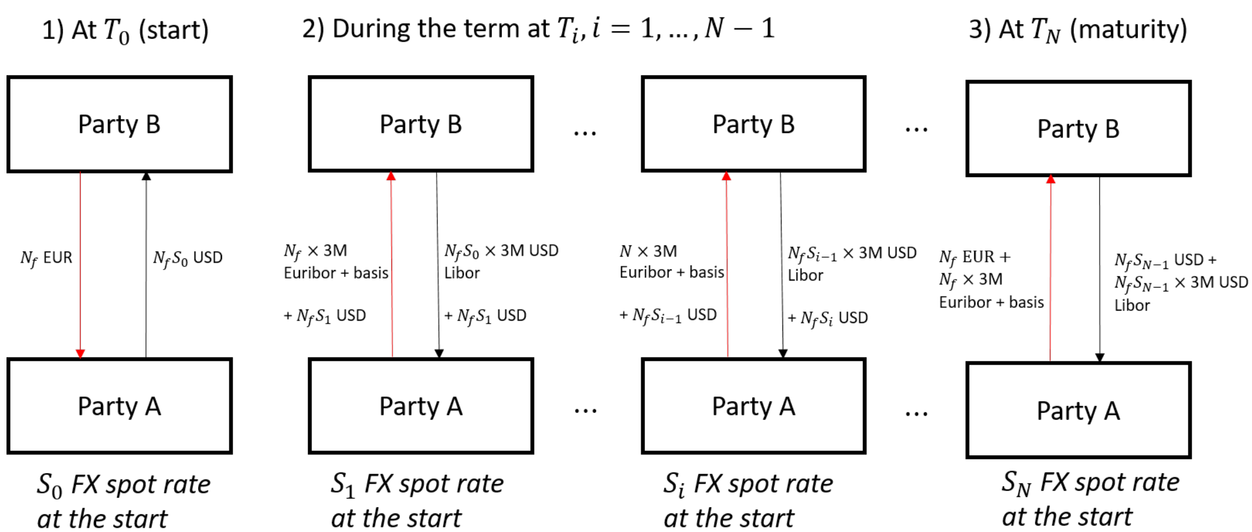


Figure 22.6.4: MtM Cross currency basis swap.

Remark 22.6.6 (understand the cash flow in MtM cross-currency basis swap). Let A be the domestic party. The cash flows from A's point of view are

- For undiscounted domestic currency resettable leg(cash flow in domestic currency)
  - At T_0, $C_d(T_0) = -S(T_0)N_f$
  - At $T_i, i = 1, 2, ..., M$, $C_d(T_i) = L^d(T_{i-1}, T_i)\tau_i S(T_{i-1})N^f + (S(T_{i-1}) - S(T_i))N^f$
  - At T_M, $C_d(T_M) = S(T_M)N^f$.
- For undiscounted foreign currency leg
  - At T_0, $C_f(T_0) = N_f$.
  - At $T_i, i = 1, 2, ..., M$, $C_f(T_i) = -L^f(T_{i-1}, T_i)\tau_i N^f$.
  - At T_M, $C_f(T_M) = -N_f$.

Note that for the resettable leg cash flow, we can decompose into N single period swaps over T_{i-1}, T_i, each one with discounted cash flow given by

$$\underbrace{-S(T_i)N_f}_{occurs\ at\ T_{i-1}} + \underbrace{L^d(T_{i-1}, T_i)\delta_i S(T_{i-1})N_f + S(T_{i-1})N_f}_{occurs\ at\ T_i}.$$

Lemma 22.6.6 (valuing MtM float-float cross-currency swap). *Consider a cross-currency swap with $N^f = 1$.*

- *For the domestic party, the present value of a one-period leg going between T_{i-1} and T_i(see Remark 22.6.6), $i = 0, 1, 2, ..., M$, is given by*

$$V_i^d(t) = \underbrace{P^d(t, T_{i-1})E_{T^d}[-N^f S(T_{i-1})|\mathcal{F}_t]}_{pay\ out\ new\ notional\ at\ T_{i-1}} + \underbrace{P^d(t, T_i)E_{T^d}[S(T_{i-1})|\mathcal{F}_t]}_{receive\ notional\ at\ T_i} + \underbrace{P^d(t, T_i)E_{T^d}[S(T_{i-1})L^d(T_{i-}}_{receive\ interest\ at\ T_i}$$

- *If **assuming the independence between foreign exchange rate S and interest rate**, we have*

$$V_i^d(t) = -P^d(t, T_{i-1})N^f F_X(t, T_{i-1}) + P^d(t, T_i)F_X(t, T_{i-1}) + P^d(t, T_i)F_X(t, T_{i-1})F^d(t, T_{i-1}, T_i)\tau_i$$

where $F_X(t, T)$ is the forword exchange rate, $F^d(t, T_i, T_{i+1})$ is the domestic forward Libor rate.

- *The present value of the domestic leg is*

$$\begin{aligned} V^d(t) &= \sum_{i=0}^{N} V_n \\ &= -N^f(P^d(t, T_0)F_X(t, T_0) - P(t, T_N)F_X(t, T_{N-1}) \\ &- \sum_{i=1}^{N-1} P^d(t, T_i)(F_X(t, T_i) - F_X(t, T_{i-1}) \\ &+ \sum_{i=1}^{N} P^d(t, T_i)F_X(t, T_{i-1})F^d(t, T_{i-1}, T_i)\tau_i \end{aligned}$$

- *The present value of the foreign leg is*

$$V^f(t) = N^f S(t)(-P^f(t, T_0) + \sum_{i=1}^{M} N\tau_i F^f(t, T_{i-1}, T_i)P^f(t, T_i) + P^f(t, T_M)).$$

- *The present value of the swap is*

$$V(t) = V^d(t) + V^f(t).$$

Proof. (1) direct cash flow discounting Remark 22.6.6. (2) With independence assumption, we have

$$P^d(t,T_i)E_{T^d}[S(T_{i-1})|\mathcal{F}_t] = E_{Q^d}[S(T_{i-1})\frac{M(t)}{M(T_i)}|\mathcal{F}_t] = E_{Q^d}[S(T_{i-1})|\mathcal{F}_t]E_{Q^d}[\frac{M(t)}{M(T_i)}|\mathcal{F}_t] = F_X(t,T_{i-1})P^d(t,T$$

where use the martingale property of foreign exchange rate in **??**. Similarly,

$$\begin{aligned}
&P^d(t,T_i)E_{T^d}[S(T_{i-1})L^d(T_{i-1},T_i)\tau_i|\mathcal{F}_t] \\
&= P^d(t,T_i)E_{T^d}[S(T_{i-1})|\mathcal{F}_t]P^d(t,T_i)E_{T^d}L^d(T_{i-1},T_i)\tau_i|\mathcal{F}_t] \\
&= P^d(t,T_i)F_X(t,T_{i-1})F^d(t,T_{i-1},T_i).
\end{aligned}$$

□

22.6.6 Options on spot and forward exchange rate

22.6.6.1 *Basic concepts*

Definition 22.6.6 (European call/put option).

- *A **call option on the exchange rate** on $S(t)$ with N notional is a financial contract with payoff $N(S(T)-K)^+$ in domestic currency, where K is the strike parameter, at the maturity date T.*
- *Or equivalently, a call option on exchange rate gives the buyer the right to use K units of domestic currency to buy 1 unit of foreign currency whose value is $S(T)$. Therefore, when the $S(T)>K$, the option buyer will buy foreign currency at price K, which is lower than its market price $S(T)$.*
- *Similarly, a **put option on the exchange rate** on $S(t)$ with N notional gives the holder to right to sell N units of foreign currency at K domestic value. It has payoff given by $N(K-S(T))^+$.*

Remark 22.6.7 (premium and payment conventions)**.**

- The premium of option can be paid in either domestic currency or foreign currency.
- From the perspective of domestic investors, the notional of the option contract is foreign currency.
- The payoff of the option can be in either domestic currency or foreign currency.

Remark 22.6.8 (symmetric relationship on options)**.**

- A CCY1/CCY2 call option (right to buy CCY1 by selling CCY2) is equivalent to a CCY2/CCY1 put option (right to sell CCY2 and receive CCY1).
- The relation between CCY1/CCY2 call and CCY1/CCY2 is connected by put-call parity.

22.6.6.2 *Pricing*

Lemma 22.6.7 (European option on spot exchange rate in Black model with payoff and premium paid in domestic currency). *[8, p. 390][4, p. 377] Let current time be t. Assume constant domestic interest rate r, constant foreign interest rate r^f, and constant volatility σ_S. Consider a call or put option on the FX spot rate S_t with strike K and expiry T whose payoff in domestic currency is $(S(T)-K)^+$. At time t, the value(in domestic currency) is*

$$C(t,S(t)) = e^{-r(T-t)}E_{Q^d}[(S(T)-K)^+|\mathcal{F}_t],$$

*where the expectation is taken with respect to domestic risk-neutral measure Q^d. In particular, given by the fact that(**??**), under Q^d, the dynamics of S is*

$$dS(t) = (r-r^f)S(t)dt + \sigma_S S(T)d\hat{W},$$

then current value is given by

$$C(t,S(t)) = e^{-r^f\tau}S(t)N(d_+) - e^{-r\tau}KN(d_-),$$

$$P(t,S(t)) = e^{-r^f\tau}S(t)N(d_+) - e^{-r\tau}KN(d_-),$$

where

$$d_\pm = \frac{1}{\sigma_S\sqrt{\tau}}[\log(\frac{S(t)}{K}) + (r-r^f \pm \frac{1}{2}\sigma_S^2)\tau], \tau = T-t.$$

Proof. Use martingale pricing at Theorem 22.3.7 such that

$$C(t,S(t)) = \exp(-r_d(T-t))E_{Q^d}[(S(T)-K)^+|\mathcal{F}_t].$$

In the Black model, we assume the dynamics of $S(t)$ under domestic risk-neutral measure will follow

$$dS(t)/S(t) = (r_d - r_f)dt + \sigma_S dW_t.$$

Then we use the result from equity option(Lemma 18.2.2). □

Remark 22.6.9 (price in terms of forward exchange rate)**.** [4, p. 378] Using the forward exchange rate expression given by(**??**) $F(t) = S(t)\exp((r-r_f)(T-t)$, we have

$$C = \exp(-r\tau)(F(t)N(d_1) - KN(d_2))$$
$$P = \exp(-r\tau)(-F(t)N(-d_1) + KN(-d_2))$$

where

$$d_{1,2} = \frac{1}{\sigma_S\sqrt{\tau}}[\log(\frac{F(t)}{K}) + (\pm\frac{1}{2}\sigma_S^2)\tau], \tau = T-t.$$

Remark 22.6.10 (value of the option if premium is in foreign currency)**.** The value of the option at current time t will be worth$\frac{V(t)}{S_t}$ in the unit of foreign currency.

Remark 22.6.11 (business needs)**.** [9, p. 113]

Consider the company Microsoft.

- In some countries (e.g., Europe, Japan, and Australia), it bills in the local currency and converts its net revenue to U.S. dollars monthly. For these currencies, Microsoft has a exposure to exchange rate movements.
- In other countries (e.g., those in Latin America, Eastern Europe, and Southeast Asia), it bills in U.S. dollars. Suppose the U.S. dollar appreciates against the currency of a country in which it is billing in U.S. dollars. People in the country will find Microsoft's products more expensive. As a result, Microsoft is likely to reduce its (U.S. dollar) price in the country or face a decline in sales. Microsoft therefore also has a foreign exchange exposure-both when it bills in U.S. dollars and in foreign currency.
- Microsoft therefore will use options to manage the foreign exchange rate risk.

22.6.6.3 *Put-call duality*

Lemma 22.6.8 (put-call duality). *[8, p. 391] Consider a call on FX rate $S(T)$ with strike K and expiry T. Its price denominated in domestic currency equals its price of a K units of put on $1/S(T)$ with strike $1/K$ dominated in foreign currency. That is,*

$$\frac{1}{S_t}v^d_{call}(S_t, K, T, r_d, r_t) = Kv^d_{put}(\frac{1}{S}, \frac{1}{K}, T).$$

Proof. Change of measure approach:

$$\begin{aligned}
& C(S(T),K) \textit{ in domestic currency} \\
=& E_{Q^d}[\frac{M(t)}{M(T)}(S(T)-K)^+|\mathcal{F}_t] \\
=& E_{Q^f}[\frac{M(t)}{M(T)}(S(T)-K)^+\frac{dQ^d}{dQ^f}|\mathcal{F}_t] \\
=& E_{Q^f}[\frac{M(t)}{M(T)}(S(T)-K)^+\frac{S(t)M(T)M^f(t)}{M^f(T)S(T)M(t)}|\mathcal{F}_t] \\
=& E_{Q^f}[\frac{M^f(t)S(t)}{M^f(T)S(T)}(S(T)-K)^+|\mathcal{F}_t] \\
=& E_{Q^f}[\frac{M^f(t)S(t)}{M^f(T)}K(\frac{1}{K}-\frac{1}{S(T)})^+|\mathcal{F}_t] \\
=& S(t)KE_{Q^f}[\frac{M^f(t)}{M^f(T)}K(\frac{1}{K}-\frac{1}{S(T)})^+|\mathcal{F}_t] \\
=& K\times P(\frac{1}{S(T)},\frac{1}{K}) \textit{ dominated in foreign currency}
\end{aligned}$$

□

Remark 22.6.12 (interpretation)**.**

- An option to buy one unit of EUR and sell K units of USD (which is a call) is equivalent to an option to sell K units of USD and buy one unit of EUR (which is a put).
- For a domestic (USD) investor, the payoff of the call in USD is

  $$(S_T-K)^+;$$

  for a foreign (EUR) investor, the payoff of the put in EUR is

  $$(1-KS'_T)^+, S^T = 1/S_T.$$

- Let

  $$V^d = e^{-r_dT}E_{Q^d}[(S_T-K)^+], V^f = e^{-r_fT}E_{Q^f}[(1/K-1/S_T)^+],$$

  then $V^d = S_0K\cdot V^f$ based on Theorem 22.3.6.

22.6.6.4 *Hedging*

Remark 22.6.13 (Delta hedging for exchange rate option)**.**

- Consider hedging a long position of call with payoff $C_T = (S_T - K)^+$. Let $\Delta = \partial C_t / \partial S_t$. We can short Δ units of foreign currency and receive payment in $\Delta \cdot S_t$ of domestic currency.
- If S_t increases, our call option will increase value in the unit of domestic currency; however, to close our short position in the foreign currency, we need more domestic currency than before to buy(therefore we are losing money).

22.6.7 Forwards on foreign assets

22.6.7.1 *Forwards with domestic forward price*

Definition 22.6.7 (Forwards with domestic forward price).

- *Let S_t^f denote the foreign price of a foreign asset. A forward contract is an agreement to buy a foreign asset with domestic price K^d at future time T.*
- *The payoff in domestic currency at time T is given by*

$$V^d(T) = (S_T^f X_T - K^d),$$

where X_T is the FX rate at time T.

Lemma 22.6.9 (Forward price for forwards with domestic forward price). *Let S_t^f denote the foreign price of a foreign asset. Let X_t denote FX spot rate. Consider a forward with maturity T and domestic price K^d. Let current time be t.*

- *The current value of the forward contract is*

$$\begin{aligned} V^d(t) &= E_{Q^d}[\frac{M(t)(X_T S_T^f - K^d)}{M(T)} | \mathcal{F}_t] \\ &= P^d(t,T)(E_{T^d}[X_T S_T^f] - K^d) \end{aligned}$$

- *The forward price make the forward contract have zero value is given by*

$$F^d(t,T) = E_{T^d}[X_T S_T^f].$$

- *Assume deterministic interest rate, we have*

$$F^d(t,T) = E_{Q^d}[X_T S_T^f] = X_t S_t^f \exp(r^d(T-t)).$$

Proof. □

22.6.7.2 *Forwards with foreign forward price*

Definition 22.6.8 (Forwards with foreign forward price).

- *Let S_t^f denote the foreign price of a foreign asset. A forward contract is an agreement to buy a foreign asset with foreign price K^f at future time T.*
- *The payoff in domestic currency at time T is given by*

$$V^d(T) = (S_T^f X_T - K^d),$$

where X_T is the FX rate at time T.

Lemma 22.6.10 (Forward price for forwards with domestic forward price). *Let S_t^f denote the foreign price of a foreign asset. Let X_t denote FX spot rate. Consider a forward with maturity T and domestic price K^d. Let current time be t.*

- *The current value of the forward contract is*

$$\begin{aligned} V^d(t) &= X_t E_{Q^f}[\frac{M^f(t)(S_T^f - K^f)}{M^f(T)}|\mathcal{F}_t] \\ &= X_t P^f(t,T)(E_{T^f}[X_T S_T^f] - K^f) \end{aligned}$$

- *The forward price make the forward contract have zero value is given by*

$$F^f(t,T) = E_{T^f}[S_T^f].$$

- *Assume deterministic interest rate, we have*

$$F^f(t,T) = E_{Q^f}[S_T^f] = S_t^f \exp(r^f(T-t)).$$

Proof. □

22.6.8 Options on foreign assets

22.6.8.1 *Basic principles*

Lemma 22.6.11 (option on foreign asset paying domestic currency). *Let the foreign asset S_t^f, denominated in foreign currency, have dynamics given by*

$$dS_t^f / S_t^f = \mu dt + \sigma_S dW_1.$$

Let the exchange rate X_t have dynamics given by

$$dX_t/X_t = \gamma dt + \sigma_X dW_2, dW_1 dW_2 = \rho dt.$$

It follows that

- *Under domestic measure Q^d, S_t and X_t have dynamics given by*

$$d(S_t)/(S_t) = (r^f - \rho\sigma_S\sigma_S)dt + \sigma_S dW_1^d$$
$$d(X_t)/(X_t) = (r^d - r^f)dt + \sigma_X dW_2^d$$

- *Consider an option will payoff $V^d(S_T^f, X_T^f)$ in domestic currency, then its current value is*
$$V^d(t) = M(t)E_{Q^d}[\frac{V^d(S_T)}{M(T)}|\mathcal{F}_t].$$
- *Consider an option will payoff $V^f(S_T^f, X_T^f)$ in foreign currency, then its current domestic value is*
$$V^d(t) = M(t)E_{Q^d}[\frac{V^f(S_T, X_T)X_T^f}{M(T)}|\mathcal{F}_t],$$
or
$$V^d(t) = S(t)V^f(t), V^f = M^f(t)E_{Q^f}[\frac{V^f(S_T, X_T)}{M^f(T)}|\mathcal{F}_t].$$

Proof. From Lemma 22.3.7. □

Remark 22.6.14. Note that $S_t X_t$ is a martingale under domestic measure (Lemma 22.3.7), but S_t is not.

22.6.8.2 *European options with domestic strike*

Definition 22.6.9 (European options with domestic strike).

- *Let S_t^f denote the foreign price of a foreign asset. An European call option with domestic strike K^d gives the holder the right to buy the foreign asset at future time T with domestic price K^d.*
- *The payoff in domestic currency at time T is given by*

$$V^d(T) = (S_T^f X_T - K^d)^+,$$

where X_T is the FX rate at time T.

Lemma 22.6.12 (European options with domestic strike pricing). *Let S_t^f denote the foreign price of a foreign asset. Let X_t denote FX spot rate. Consider an European call with domestic strike K^d and maturity T Assume X_tS_t has domestic risk-neutral dynamics (Theorem 22.3.5) given by*

$$dX_tS_t^f/X_tS_t^f = (r^d - q)dt + \overline{\sigma}_S dW_t,$$

where $\overline{\sigma}_S^2 = \sigma_X^2 + 2\rho\sigma_S\sigma_X + \sigma_S^2$, σ_S is the volatility of asset S_t under foreign risk-neutral measure. then

$$V_t = BS(S_t^f X_t, r^d, q, K^d, T, \overline{\sigma}_S),$$

where $BS(S_t^f X_t, r^d, q, K^d, T, \overline{\sigma}_S)$ is the Black-Scholes formula with spot $S_t^f X_t$, drift r^d, and dividend rate q.

Proof. We can view $X_tS_t^f$ as a domestic asset with drift $r^d - q$, where $q = r^d - r^f - \rho\sigma_S\sigma_X$ is the dividend rate. □

22.6.8.3 *European options with foreign strike*

Definition 22.6.10 (European options with foreign strike).

- *Let S_t^f denote the foreign price of a foreign asset. An European call option with foreign strike K^f gives the holder the right to buy the foreign asset at future time T with foreign price K^f.*
- *The payoff in domestic currency at time T is given by*

  $$V^d(T) = X_T(S_T^f - K^f)^+,$$

  where X_T is the FX rate at time T.

Lemma 22.6.13 (European options with foreign strike pricing). *Let S_t^f denote the foreign price of a foreign asset. Let X_t denote FX spot rate. Consider an European call with foreign strike K^f and maturity T Assume S_t has foreign risk-neutral dynamics (Theorem 22.3.5) given by*

$$dS_t^f/S_t^f = (r^f - q)dt + \sigma_S dW_t,$$

where σ_S is the volatility of asset S_t under foreign risk-neutral measure. then

$$V_t^d = X_t BS(S_t, r^f, q, K^f, T, \sigma_S),$$

where $BS(S_t^f, r^f, q, K^f, T, \sigma_S)$ is the Black-Scholes formula with spot S_t^f, drift r^f, and dividend rate q.

Proof. From Theorem 22.3.6, we have

$$\begin{aligned}
V^d(t) &= X(t)V^f(t) \\
&= X(t)E_{Q^f}[\frac{M^f(t)V^f(T)}{M^f(T)}|\mathcal{F}_t] \\
&= X(t)E_{Q^f}[\frac{M^f(t)(S_T^f - K^f)^+}{M^f(T)}|\mathcal{F}_t] \\
&= X_t BS(S_t^f, r^f, q, K^f, T, \sigma_S)
\end{aligned}$$

□

22.6.9 Quanto securities

22.6.9.1 *Quanto forwards*

Definition 22.6.11 (quanto forward). *Let S_t^f denote the foreign price of a foreign asset. From the perspective of domestic investor, a Quanto forward is agreement to purchase the foreign asset at a fixed domestic price at a fixed exchange rate $\overline{X}$.*

The fixed domestic price chosen such that the price of the contract is zero at the time of entry called quanto-forward price.

Definition 22.6.12 (quanto-forward curve). *A quanto-forward curve observed at current time t on foreign asset S_t^f is defined by*

$$F^d(t,T) = E_{Q^d}[S(T)|\mathcal{F}_t].$$

Lemma 22.6.14 (quanto-forward price and quanto-forward curve). *[10, p. 118] Consider a quanto-forward contract on foreign asset S_t^f with delivery date T and fixed FX rate $\overline{X}$. Let current time be t. Then the quanto-forward price is given by*

$$F^d(t,T;\overline{X}) = \overline{X}E_{Q^d}[S_T|\mathcal{F}_t].$$

Assume S_T^f has domestic risk-neutral dynamics (Theorem 22.3.5) given by

$$dS_t^f/S_t^f = (r^f - \rho\sigma_S\sigma_X)dt + \sigma_S dW_t,$$

then

$$F^d(t,T) = S_t^f \exp((r_f - \rho\sigma_X\sigma_S)(T-t)) = F^f(t,T)\exp(-\rho\sigma_X\sigma_S(T-t)),$$

where $F^f(t,T)$ is the forward curve on foreign asset denominated in foreign currency given by $F^f(t,T) = S_t^f \exp(r_f(T-t))$

Proof. In the domestic measure, we have

$$0 = E_{Q^d}[e^{-r^d(T-t)}(\overline{X}S_T - F^d(0,T))];$$

therefore

$$F^d(t,T) = E_{Q^d}[\overline{X}S_T] = \overline{X}E_{Q^d}[S_T|\mathcal{F}_t].$$

□

22.6.9.2 *Quanto options*

Definition 22.6.13 (European quanto option with domestic strike).

- *Let S_t^f denote the foreign price of a foreign asset. An European call quanto-option with domestic strike K^d gives the holder the right to buy the foreign asset at future time T with domestic price K and fixed FX rate $\overline{X}$.*
- *The payoff in domestic currency at time T is given by*

$$V^d(T) = (S_T^f\overline{X} - K^d)^+.$$

Lemma 22.6.15 (quanto option pricing). *Let S_t^f denote the foreign price of a foreign asset. Consider an European call quanto-option with domestic strike K^d, maturity T, and fixed FX rate $\overline{X}$. Assume S_T has domestic risk-neutral dynamics (Theorem 22.3.5) given by*

$$dS_t/S_t = (r^f - \rho\sigma_S\sigma_X)dt + \sigma_S dW_t,$$

where σ_S is the volatility of asset S_t under foreign risk-neutral measure. then

$$V_t = BS(S_t\overline{X}, r^d, q, K^d, T, \sigma_S),$$

where $BS(S_t\overline{X}, r^d, q, K^d, T, \sigma_S)$ is the Black-Scholes formula with spot $S_t\overline{X}$, drift r^d and dividend rate $q = r^d - r^f - \rho\sigma_S\sigma_X$.

Proof. We can view S_t as a domestic asset with drift $r^d - q$, where $q = r^d - r^f - \rho\sigma_S\sigma_X$ is the dividend rate. $\square$

22.7 Volatility modeling

see [11].

22.8 Notes on bibliography

Major references are [8][2][11][12, ch 14]. [13][1][14].

For Quanto securities, see **??**.

BIBLIOGRAPHY

1. Reiswich, D. & Uwe, W. FX volatility smile construction. *Wilmott* **2012,** 58–69 (2012).
2. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
3. Pelsser, A. Mathematical foundation of convexity correction. *Quantitative Finance* **3,** 59–65 (2003).
4. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
5. Grzelak, L. A. & Oosterlee, C. W. On cross-currency models with stochastic volatility and correlated interest rates. *Applied Mathematical Finance* **19,** 1–35 (2012).
6. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
7. Brigo, D., Morini, M. & Pallavicini, A. *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes* ISBN: 9780470661789. https://books.google.com/books?id=Y-X9zqVcOFkC (Wiley, 2013).
8. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
9. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
10. Back, K. *A course in derivative securities: Introduction to theory and computation* (Springer Science & Business Media, 2006).
11. Clark, I. J. *Foreign exchange option pricing: a practitioner's guide* (John Wiley & Sons, 2011).
12. Musiela, M. & Rutkowski, M. *Martingale methods in financial modelling* (Springer Science & Business Media, 2006).
13. Reiswich, D. & Wystup, U. A guide to FX options quoting conventions. *Journal of Derivatives* **18,** 58 (2010).
14. Castagna, A. *Fx options and smile risk* (John Wiley & Sons, 2010).

23

CREDIT RISK MODELING

23.1 Overview

Credit risk can be defined as the possibility that a contractual counterparty does not meet its obligations stated in the contract, therefore causing the creditor a financial loss. Credit default swap is Credit derivatives are financial instruments that essentially provide insurance against the credit deterioration and default of the counterparty.

We start this chapter with a basic probabilistic model on default. We then incorporate the credit default probability description into the risk-neutral pricing framework.

We also introduce the Merton's structure model and credit migration model.

23.2 Default term structure modeling

23.2.1 Concepts of default

Definition 23.2.1 (default concepts).

- *The **default indicator process** among a time interval* $[0, T]$ *is defined by*
$$N(u) = \mathbf{1}_{(\tau < u)},$$
where the stopping time τ *is the default event.*
- *The **hazard rate function*** $\lambda(t)$ *associated with jump process* $N(u)$ *describes the rate of a jump occurs at the next instant given that the life has survived to time* t.
- *The **hazard function** is defined as* $\Lambda(t) = \int_0^t \lambda(s)ds$ *is the accumulated default intensity.*
- $e^{-\Lambda} = P(N(t) = 0)$ *is the probability that there is no default in the interval* $(0, t)$.
- ***Survival probability*** $S(t, u)$ *is defined as*
$$S(t, u) = P(\tau > u | \tau > t) = \exp(-\int_t^u \lambda(s)ds).$$
Note that $\lambda(u) = -\frac{\partial \ln S(t,u)}{\partial u}$.
- ***Default probability*** $D(t, u)$ *is defined as*
$$D(t, u) = P(\tau < u | \tau > t) = 1 - S(t, u)$$
- ***Survival indicator function*** $I(t)$ *is defined as*
$$I(t) = \mathbf{1}_{\tau > t}.$$
Note that $S(t, u) = E[I(u) | I(t) = 0]$.

Definition 23.2.2 (default loss). *[1, p. 6] Consider a set of obligors, indexed by* $i = 1, ..., I$. *Define*

- $N_i(t)$ *: Default indicator process.* $N_i(t) = 1$ *if obligor* i *has defaulted by time* t*, and o otherwise.*
- $E_i(t)$ *: Exposure process.* $E_i(t)$ *is the amount we would lose if obligor* i *defaulted at time* t *with zero recovery. Also called exposure at default.*
- $L_i(t)$*: Loss given default of obligor* i *at time* t. $L_i(t) \leq 1$ *since a fraction of* $E_i(t)$ *may be recovered in bankruptcy proceedings. The **recovery rate** is given by* $R_i(t) = 1 - L_i(t)$.

*Then the **default loss** of obligor i*

$$D_i(T) \triangleq N_i(T) \times E_i(\tau_i) \times L_i(\tau_i),$$

where T is some time horizon, say 1 year, and τ_i is the time of default.

23.2.2 Probability characterization of default

Definition 23.2.3. *Let τ be a random variable be the default time.*

- *Let F_τ be the **default cumulative density function** such that*

$$F_\tau(t) \triangleq Pr(\tau < t), F(0) = 0$$

- *Let $S : \mathbb{R} \to [0,1]$ be the **survival probability** such that*

$$S(t) = Pr(\tau > t) = 1 - F(t), S(0) = 1.$$

- *Let f_τ be the **default probability density function** such that*

$$f_\tau(t) \triangleq \lim_{\Delta \to 0} \frac{Pr(t < \tau < t + \Delta)}{\Delta} = \frac{dF_\tau(t)}{dt} = -\frac{dS_\tau(t)}{dt}.$$

- *Let $h(t)$ be the **conditional instantaneous default probability** or **hazard rate** such that*

$$h(t) = \lim_{\Delta \to 0} \frac{Pr(t < \tau < t + \Delta t | \tau > t)}{\Delta t}.$$

Lemma 23.2.1 (basic properties of default probabilities in the hazard curve model). *Let the current time be 0. Assume the default time τ has cdf given by*

$$F_\tau(t) = 1 - \exp(-\int_0^t h(s)ds),$$

where $h(t)$ is the hazard rate function. It follows that

- *(recover hazard rate function)*

$$h(t) = \frac{1}{S(t)}\frac{dF_\tau(t)}{dt} = \frac{f_\tau(t)}{S(t)} = -\frac{S'(t)}{S(t)}.$$

- *(calculation of survival probability)*

$$Pr(\tau > t) = S(t) = \exp(-\int_0^t h(s)ds).$$

If $h(s) = h_0$ is flat, then

$$Pr(\tau > t) = S(t) = \exp(-h_0 t).$$

- *(calculation of default probability density)*

$$f_\tau(t) = S(t)h(t) = h(t)\exp(-\int_0^t h(s)ds).$$

- ***Prior default probability*** *at future time interval $[T_1, T_2]$ is given by*

$$\begin{aligned} Pr(T_1 < \tau < T_2) &= \int_{T_1}^{T_2} f(s)ds \\ &= S(T_1) - S(T_2) \\ &= S(T_1)(1 - \exp(-\int_{T_1}^{T_2} h(s)ds)) \end{aligned}$$

- ***Conditional default probability*** *at future time interval $[T_1, T_2]$($T_0 < T_1$) is given by*

$$Pr(T_1 < \tau < T_2 | \tau > T_0) = \frac{S(T_1) - S(T_2)}{S(T_0)}.$$

- *Conditional survival probability at future time T_1 given survival till T_0 is given by*

$$Pr(T_1 < \tau | \tau > T_0) = \frac{S(T_1)}{S(T_0)}.$$

Proof. (1) From definition, we have

$$h(t) = \lim_{\Delta \to 0} \frac{Pr(t < \tau < t + \Delta t | \tau > t)}{\Delta t} = \lim_{\Delta \to 0} \frac{1}{\Delta t} \frac{Pr(t < \tau < t + \Delta t)}{Pr(t > \tau)} = \frac{f(t)}{S(t)}.$$

(2) From (1), since $S' = -Sh, S(0) = 1$, we can solve

$$S(t) = S(0)\exp(-\int_0^t h(s)ds) = \exp(-\int_0^t h(s)ds).$$

(3) From (1), we have $f(t) = S(t)h(t)$. (3) Note that

$$\int_{T_1}^{T_2} f_\tau(s)ds = \int_{T_1}^{T_2} h(s)S(s)ds = -\int_{T_1}^{T_2} S'(s)ds = S(T_1) - S(T_2).$$

(4) Note that

$$P(T_1 < \tau < T_2 | \tau > T_0) = \frac{P(T_1 < \tau < T_2)}{P(\tau > T_0)} = \frac{S(T_1) - S(T_2)}{S(T_0)}.$$

□

23.2.3 Structure models

23.2.3.1 *Canonical Merton model*

Definition 23.2.4 (Merton's model for firm debt). *[2, p. 48][3, p. 48] Assume under risk-neutral measure, the firm value(i.e. total asset) V is governed by SDE*

$$dV_t / V_t = rdt + \sigma dW_t,$$

where r is the risk-free rate and σ is the volatility.

Assume the firm has a debt L matures at time T.

At time T, the repayment of the debt is

$$\min\{L_T, V(T)\}.$$

The firm's equity value S is given by

$$S \triangleq V - L$$

Lemma 23.2.2 (governing PDE for debt value). *[4, p. 640] Assume the asset value V of the firm under real probability measure is governed by*

$$dV = \mu V dt + \sigma V dW_t.$$

Then the current value of the debt L is governed by PDE

$$\frac{\partial L}{\partial t} + \frac{1}{2}\sigma^2 V^2 \frac{\partial^2 L}{\partial V^2} + rV\frac{\partial L}{\partial V} - rV = 0,$$

with final condition $L(V, T) = \min(L_T, V_T)$.

Proof. □

Lemma 23.2.3 (analytical results in Merton's model).

- *The default probability(**in risk-neutral measure**) is given by $P(V(T) < L)$ is given by*
$$P(N(0,1) \leq \frac{\log(T/V(0) - (r - \frac{\sigma^2}{2}))}{\sigma\sqrt{T}}).$$
- *The bond/debt value is given by*
$$\begin{aligned} L(t) &= E_Q[e^{-r(T-t)} \min\{V_T, L_T\} | \mathcal{F}_t] \\ &= E_Q[e^{-r(T-t)}[L_T - (L_T - V_T)^+] | \mathcal{F}_t] \\ &= e^{-r(T-t)} L_T - Put(t, T, V_t, L_T) \end{aligned}$$
- *The firms equity value*
$$S_t \triangleq V_t - L(t) = Call(t, T, V_t, L_T).$$

Proof. (3) Use put-call parity
$$Call(t, T, V_t, L_T) - Put(t, T, V_t, L_T) = V_t - e^{-r(T-t)} L_T$$
□

Remark 23.2.1. The debt holder's position is equivalent to long bond and short put, which will decrease value at increasing volatility.

The equity holder's position is equivalent to long call, which increase value at increasing volatility.

Lemma 23.2.4 (Default probability in real world). *Assume under real world probability measure, the firm value V is governed by SDE*
$$dV_t/V_t = \mu dt + \sigma dW_t.$$

Assume the firm will default its debt when its value $V(t)$ first reach $K(t)$ and the debt maturity date T. Let $P(V,t)$ denote the default probability if the firm has value v at time t. Then the governing equation for P is
$$\frac{\partial P}{\partial t} + \mu \frac{\partial P}{\partial V} + \frac{1}{2}\sigma^2 \frac{\partial^2 P}{\partial V^2} = 0,$$
with boundary condition $P(V(t) = K(t), t \leq T) = 1, P(x, T) = 0.$

Proof. See Lemma 6.1.15. □

Note 23.2.1 (Estimating default probability from Merton model). [5, p. 553] Suppose a firm has one zero-coupon bond outstanding and that the bond matures at time T. Define

- V_0: value of the company's assets today
- V_T: value of the company's assets at time T
- E_0: value of the company's equity today
- E_T: value of the company's equity at time T
- D: debt repayment due at time T
- σ_V: volatility of assets
- σ_E: volatility of equity

The firm's equity at time T is **assumed** to be

$$E_T = \max(V_T - D, 0),$$

and therefore under Merton's model, the current value of the equity is given by

$$E_0 = V_0 N(d_1) - D\exp(-rT)N(d_2),$$

where

$$d_{1,2} = \frac{\ln(V_0/D) \pm (r + \sigma_V^2/2)T}{\sigma_V\sqrt{T}}.$$

It follows that

- the risk-neutral probability that the firm will default on the debt is $N(-d_2)$.
- to evaluate d_2, we require V_0 and σ_V. Assume E_0 is observable for publicly traded firms and σ_E can be estimated from historical stock price data. We have two simultaneous equation to determine V_0 and σ_V given by

$$\begin{aligned} E_0 &= V_0 N(d_1) - D\exp(-rT)N(d_2) \\ \sigma_E E_0 &= \frac{\partial E}{\partial V}\sigma_V V_0 = N(d_1)\sigma_V V_0 \end{aligned}$$

23.2.3.2 *Connection to reduced form model*

23.3 Risk-neutral pricing framework

23.3.1 Risk-neutral pricing fundamentals

23.3.1.1 *Principle*

Lemma 23.3.1. *Consider an asset with payoff $V(T)$ at maturity T. Let τ be the default time. Then the current value of the asset is*

$$V(t) = E_Q[\exp(-\int_t^T r(s)ds)V(T)\mathbf{1}_{\tau>T} + \exp(-\int_t^\tau r(s)ds)V(\tau)R(\tau)\mathbf{1}_{t<\tau<T}],$$

where R is the recovery rate.4

Proof. Note that $\exp(-\int_t^\tau r(s)ds)V(\tau)R(\tau)\mathbf{1}_{t<\tau<T}$ denotes the recovered value at the default time. □

23.3.1.2 *Defaultable zero-coupon bond*

Definition 23.3.1 (zero coupon bond). *[6]*

- *The* ***default-free zero-coupon bonds*** *maturing at T have price at $t < T$ given by*

$$P(t,T).$$

- *The* ***defaultable zero-recovery zero-coupon bonds*** *maturing at T have price at $t < T$ given by*

$$P_0^d(t,T).$$

- *The* ***defaultable zero-coupon bonds*** *maturing at T have price at $t < T$ given by*

$$P^d(t,T).$$

- *The* ***credit spread*** *is defined by*

$$s(t,T) = \frac{1}{T-t}\ln(\frac{1}{P^d(t,T)}) - \frac{1}{T-t}\ln(\frac{1}{P(t,T)}).$$

Lemma 23.3.2 (defaultable zero-coupon bond pricing). *[3, p. 62][6, p. 54]*

- *Under risk-neutral measure Q, the default-free zero-coupon bonds have prices given by*
$$P(t,T) = E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t].$$
- *Let τ denote the random default time. Under risk-neutral measure Q, the defaultable zero-recovery zero-coupon bonds have prices given by*
$$P_0^d(t,T) = E_Q[\exp(-\int_t^T r(s)ds)\mathbf{1}_{\tau>T}|\mathcal{F}_t].$$
- *Let τ denote the random default time. Under risk-neutral measure Q, the defaultable zero-recovery bonds have prices given by*
$$P_0^d(t,T) = E_Q[\exp(-\int_t^T r(s)ds)(\mathbf{1}_{\tau>T} + R\mathbf{1}_{\tau<T})|\mathcal{F}_t],$$
where R is the recovery rate at maturity.

Proof. $\tau > T$ indicates no default occurs before maturity. □

Lemma 23.3.3 (defaultable zero-coupon bond pricing with independence and deterministic assumption). *[6, p. 54] Assume the default process is characterized by the deterministic intensity parameter $\lambda(t)$ independent of stochastic process $r(t)$. Let τ denote the random default time. It follows that*

- *The defaultable zero-recovery zero-coupon bonds have prices given by*
$$\begin{aligned} P_0^d(t,T) &= E_Q[\exp(-\int_t^T r(s)ds)\mathbf{1}_{\tau>T}|\mathcal{F}_t] \\ &= E_Q[\exp(-\int_t^T r(s)+\lambda(s)ds)|\mathcal{F}_t] \\ &= P(t,T)\exp(-\int_t^T \lambda(s)ds) \\ &= P(t,T)S(t,T). \end{aligned}$$
- *Then the defaultable zero-coupon bond has price given by*
$$P^d(t,T) = P(t,T)E_Q[\mathbf{1}_{\tau>T} + R\mathbf{1}_{\tau<T}|\mathcal{F}_t] = P(t,T)(S(t,T) + R(1-S(t,T))).$$
where R is the external given recovery rate paid at T.

- *If $\lambda(t)$ and $r(t)$ are constant, then*

$$P_0^d(t,T) = P^d(t,T)\exp(-\lambda(T-t)).$$

Proof. (1)

$$\begin{aligned} P_0^d(t,T) &= E_Q[\exp(-\int_t^T r(s)ds)\mathbf{1}_{\tau>T}|\mathcal{F}_t] \\ &= P(t,T)E_Q[\mathbf{1}_{\tau>T}|\mathcal{F}_t] \\ &= P(t,T)S(t,T) = P(t,T)\exp(-\int_t^T \lambda(s)ds), \end{aligned}$$

where $S(t,T) = Q(\tau > T|\tau > t) = \exp(-\int_t^T \lambda(s)ds)$.

□

Definition 23.3.2 (implied survival probability, implied default term structure). *[6, p. 54] Denote the default time by τ.*

- *The **implied survival probability** between from t to T as seen from t(i.e. based on information accumulated up to t)*

$$S^*(t,T) = \frac{P_0^d(t,T)}{P(t,T)}.$$

- *The **implied default probability** over $[t,T]$ is given by*

$$D^*(t,T) = 1 - S^*(t,T).$$

- *The **implied density of default time** is*

$$Q(\tau \in [T, T+dt]|\mathcal{F}_t) = -\frac{\partial S^*(t,T)}{\partial T}dt.$$

- *The **implied hazard rate of default** at time T is*

$$h^*(T) = -\frac{\partial \ln S(t,T)}{\partial T}$$

Lemma 23.3.4 (decomposition of defaultable zero-coupon bond). *[6, p. 63] Let $t = T_0 < T_1 < T_2 < \cdots < T_n$ be a set of dates. Let $\delta_i = T_{i+1} - T_i$.*

- $$P(t,T_k) = \prod_{i=1}^{k} \frac{1}{1+\delta_{i-1}F(t,T_{i-1},T_i)}$$

- $$P_0^d(t,T_k) = P(t,T_k)S(t,T_k) = P(t,T_k)\prod_{i=1}^{k}\frac{1}{1+\delta_{i-1}H(t,T_{i-1},T_i)}$$

Proof. (1) see the zero-coupon bond decomposition (Lemma 19.1.2). (2) Use the that

$$S_c(t;T_1,T_2) = \frac{S(t,T_2)}{S(t,T_1)},$$

and

$$S_c(t;T_1,T_2) = \frac{1}{1+(T_2-T_1)H(t;T_1,T_2)}.$$

□

Lemma 23.3.5 (current value of recovery at default). *[6, pp. 60, 118][7, pp. 48, 52] Let $t = T_0 < T_1 < T_2 < \cdots < T_n$ be a set of dates. Let $\delta_i = T_{i+1} - T_i$. Then*

- *the current value of \$1 payoff at T_{k+1} if and only if a default occurs in $(T_k, T_{k+1}]$.*

$$e(t,T_k,T_{k+1}) = \delta_k H(t,T_k,T_{k+1})P_0^d(t,T_{k+1}).$$

- *Assume the default process is characterized by an deterministic intensity parameter $\lambda(t)$ independent of stochastic process $r(t)$. We have*

$$\begin{aligned} e(t,T_k,T_{k+1}) &= E_Q[\exp(-\int_t^{T_{k+1}} r(s)ds)\mathbf{1}_{T_k<\tau<T_{k+1}}|\mathcal{F}_t] \\ &= P(0,T_{k+1})(\exp(-\int_t^{T_k}\lambda(s)ds) - \exp(-\int_t^{T_{k+1}}\lambda(s)ds)) \end{aligned}$$

- ***(alternative, payment at default)*** *the current value of \$1 payoff at τ if and only if a default occurs in $(T_k, T_{k+1}]$.*

$$e_1(t,T_k,T_{k+1}) = E_Q[\exp(-\int_t^{\tau} r(s)ds)\mathbf{1}_{T_k<\tau<T_{k+1}})|\mathcal{F}_t].$$

In particular, if assume the default is characterized by the hazard rate $h(t,s)$, then

$$\begin{aligned} e_1(t,T_k,T_{k+1}) &= E_Q[\int_{T_k}^{T_{k+1}} \exp(-\int_t^{\tau} r(s)ds)S(t,\tau)h(t,\tau)d\tau|\mathcal{F}_t] \\ &= E_Q[\int_{T_k}^{T_{k+1}} \exp(-\int_t^{\tau} r(s)+h(t,s)ds)h(t,\tau)d\tau|\mathcal{F}_t] \end{aligned}$$

- ***(alternative, random payment at default)****the current value of a random payoff $\Phi(\tau)$ at τ if and only if a default occurs in $(T_k, T_{k+1}]$.*

$$e_2(t, T_k, T_{k+1}) = E_Q[\exp(-\int_t^\tau r(s)ds\Phi(\tau)\mathbf{1}_{T_k<\tau<T_{k+1}})|\mathcal{F}_t],$$

Proof. (1)Note that the probability of default at $(T_k, T_{k+1}]$ is given by

$$Pr(T_k < \tau < T_{k+1}|\mathcal{F}_t) = S(t, T_k) - S(t, T_{k+1}) = S(t, T_k)(1 - S_c(t, T_k, T_{k+1})).$$

Then the current value is

$$\begin{aligned} e(t, T_k, T_{k+1}) &= P(t, T_{k+1})S(t, T_k)(1 - S_c(t, T_k, T_{k+1})) \\ &= P(t, T_{k+1})\delta_k H(t, t_k, t_{k+1})S(t, T_{k+1}) \\ &= \delta_k H(t, T_k, T_{k+1})P_0^d(t, T_{k+1}) \end{aligned}$$

(2)Note that the probability of default at $(T_k, T_{k+1}]$ is given by

$$Pr(T_k < \tau < T_{k+1}|\mathcal{F}_t) = S(t, T_k) - S(t, T_{k+1}) = (\exp(-\int_t^{T_k}\lambda(s)ds) - \exp(-\int_t^{T_{k+1}}\lambda(s)ds)).$$

(3) Note that the conditional instantaneous default probability at τ is given by

$$S(t, \tau)h(t, \tau)d\tau.$$

□

Remark 23.3.1. The specific choice of e depends on the contract and the specific scenario.

23.3.2 Risk neutral vs. real default probability

Note 23.3.1. [8, p. 228]

- If one is interested in estimating the economic capital and risk charges, one should use real-world probability; If the objective is to price and hedge credit-related securities, we need to use the risk-neutral probability.
- The real-world probability is estimated from historical information; the risk-neutral probability can be implied from existing market price of credit product.

Example 23.3.1.

- Consider a year bond with face value 100 and 0.07 coupon at maturity. The one-year risk-free rate is 0.05. Assume the real default probability PD = 0.01, and recovery at default is 0.5.
- Discounting the expected payoff using the real default probability is given by

$$\frac{107 \times 0.99 + 0.5 \times 100 \times 0.01}{1 + 0.05} = 101.36.$$

- This price is higher for the buyers since it does not account for the risk aversion for the buyers. Taking expectation with respect to real probability is mistake since it assumes buyer are not risk averse.
- Suppose the market price is 100. Denote PD^* as the implied default probability or risk-neutral probability. We have

$$\frac{107 \times (1 - PD^*) + 0.5 \times 100 \times PD^*}{1 + 0.05} = 100,$$

gives $PD^* = 0.0351$. Such implied probability then can be used to price other credit product related to the same firm by taking expectation.

23.3.3 Survival curve construction

23.3.3.1 *Survival curve construction from CDS spread*

Note 23.3.2 (bootstrap method to construct hazard curve from CDS spread)**.** Note that(**??**)

$$s^{CDS} = (1 - \pi) \frac{\sum_{i=1}^{n} \delta_{i-1} H(t, T_{i-1}, T_i) P_0^d(t, T_i)}{\sum_{i=1}^{n} \delta_{i-1} P_0^d(t, T_i)},$$

where

$$H(t, T, T + \Delta t) = \frac{1}{\Delta t}(\exp(\int_T^{T+\Delta t} h(t,s)ds) - 1).$$

By assuming piece-wise linear form of $h(t,s), s \geq t$ and using market data of s^{CDS} for various maturities and tenors, we can solve the $h(t,s)$.

Lemma 23.3.6. *[7, p. 117] Let $z(t,T)$ be **flat default rate** such that*

$$S(t,T) = \exp(-z(t,T)(T-t)).$$

We have

- $$S(t,T) = \exp(-\int_t^T h(t,s)ds)$$
- $$h(t,T) = \frac{\partial}{\partial T}(z(t,T)(T-t))$$
- $$h(t,T) = -\frac{\partial}{\partial T}\ln S(t,T))$$

Proof. Note that

$$-z(t,T)(T-t) = -\ln S(t,T) = \int_t^T h(t,s)ds.$$

□

Remark 23.3.2 (analogs in interest rate modeling).

- $z(t,T)$ here resembles the yield, $h(t,T)$ here resembles the forward rate.
- The construction of $h(t,T)$ from $z(t,T)$ is the same in 19.4.2

23.3.3.2 *Survival curve construction from bonds*

Lemma 23.3.7 (survival curve construction from defaultable zero-coupon bond). *Suppose we are given defaultable zero coupon bond price $P^d(t,T_1), P^d(t,T_2), ..., P^d(t,T_N)$ on a set of maturities $T_1, T_2, ..., T_N$. Further assume the recovery rate are given by $R_1, R_2, ..., R_N$ and default behavior is independent of exposure. Then the survival probability on $T_1, T_2, ..., T_N$ are given by*

$$S(t,T_i) = (\frac{P^d(t,T_i)}{P(t,T_i)} - R_i)/(1-R_i),$$

where $P(t,T_i)$ is the non-defaultable zero coupon bond price.

Proof. Note that from defaultable zero-coupon bond pricing(Lemma 23.3.3), we have

$$P^d(t,T_i) = P(t,T_i)(S(t,T_i) + R_i(1-S(t,T_i))).$$

Rearrange and we get

$$S(t, T_i) = (\frac{P^d(t, T_i)}{P(t, T_i)} - R_i)/(1 - R_i).$$

$\square$

23.4 Risk neutral pricing examples

23.4.1 Defaultable fixed-coupon bond

Definition 23.4.1 (defaultable fixed-coupon bond). *[6, p. 65] Let $t < T_0 < T_1 < T_2 < \cdots < T_n$ be a set of dates. Let $\delta_i = T_{i+1} - T_i$. A defaultable fixed-coupon bond has coupon payments c_i at $T_i, i = 1, ..., n$. At maturity T_n, there is also a principal \$1 payment. If a default occurs at $(T_i, T_{i+1}]$, then the payment at T_{i+1} is π.*

Lemma 23.4.1 (price of a defaultable fixed-coupon bond). *[6, p. 65] The current price of a defaultable fixed-coupon bond is*

$$\begin{aligned} C(t) = & \sum_{i=1}^{n} \delta_{i-1} F(t, T_{i-1}, T_i) P_0^d(0, T_i) \ (coupons) \\ & + P_0^d(t, T_n) \ (principal) \\ & + \pi \sum_{i=1}^{n} e(t, T_{i-1}, T_i) \ (recovery) \end{aligned}$$

Proof. Straight forward. □

23.4.2 Defaultable floater

Definition 23.4.2 (defaultable floater). *[6, p. 65] Let $t < T_0 < T_1 < T_2 < \cdots < T_n$ be a set of dates. Let $\delta_i = T_{i+1} - T_i$. A defaultable fixed-coupon bond has coupon payments*

$$c_i = \delta_{i-1}(L(T_{i-1}, T_i) + s^{par})$$

at $T_i, i = 1, ..., n$. At maturity T_n, there is also a principal \$1 payment. If a default occurs at $(T_i, T_{i+1}]$, then the payment at T_{i+1} is π.

Lemma 23.4.2 (price of a defaultable floater bond). *[6, p. 66] The current price of a defaultable floater is*

$$\begin{aligned}C(t) &= \sum_{i=1}^{n} c_i P_0^d(0, T_i) \ (defaultable\ LIBOR\ payment) \\ &+ s^{par} \sum_{i=1}^{n} \delta_{i-1} P_0^d(0, T_i) \ (coupon\ spread) \\ &+ P_0^d(t, T_n) \ (principal) \\ &+ \pi \sum_{i=1}^{n} e(t, T_{i-1}, T_i) \ (recovery)\end{aligned}$$

Proof. Note that the current value of LIBOR payment $\tau L(S, T), \tau = T - S$ with default can be expressed as

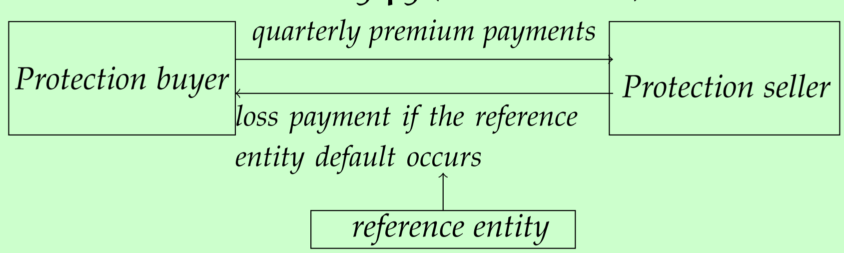
$$\begin{aligned}V(t) &= E_Q[\exp(-\int_t^T r(s)ds)\tau L(S, T)\mathbf{1}_{\tau > T}|\mathcal{F}_t] \\ &= E_Q[\exp(-\int_t^T r(s)ds)\tau L(S, T)|\mathcal{F}_t]E_Q[\mathbf{1}_{\tau > T}|\mathcal{F}_t] \\ &= P(t, T)\delta F(t, S, T)S(t, T) \\ &= \delta F(t, S, T)P_0^d(t, T).\end{aligned}$$

where we use result in Theorem 19.1.2. □

23.4.3 Credit default swaps

23.4.3.1 *Business of CDS*

Definition 23.4.3 (CDS contract).



CDS convention

- When long CDS, we mean buying protection; we pay premium in exchange for loss payment.
- When short CDS, we mean selling protection; we receive premium in exchange for receipt of loss payment.

Remark 23.4.1 (How to use CDS). link CDS can be used by investor for speculation, hedging, and arbitrage.

- **speculation** CDS allow investors to speculate on changes in CDS spreads of single names or of market indices such as the North American CDX index or the European iTraxx index. For a long position, the widening spread will increase the CDS value.
- **hedging**. CDS are often used to manage the risk of default that arises from holding debt. A bank may hedge its risk that a borrower may default on a loan by buying protection. By offloading a particular credit risk, a bank is not required to hold as much capital in reserve against the risk of default.
- **arbitrage**. A company's stock price and its CDS spread should exhibit negative correlation; i.e., if the outlook for a company improves then its share price should go up and its CDS spread should tighten. For example, if a company has announced some bad news and its share price has dropped by 25%, but its CDS spread has remained unchanged, then an investor can long CDS and wait for the spread to widen and increase the value of the CDS.

23.4.3.2 *standard CDS swap*

Definition 23.4.4 (premium leg coupon payment convention). *Consider a set of coupon payment dates $t_0 < t_1 < \cdots < t_N$ and the protection starting date t_S. t_S satisfies $t_0 \leq t_S < t_1$.*

- *(initial coupon payment convention) Suppose the reference entity defaults after t_1. Even the protection date t_s starts between coupon payment dates t_0, t_1, the contract still requires the buyer to pay the full coupon $(t_1 - t_0) \times C$ instead of a partial coupon of $(t_1 - t_S) \times C$.*
- *Note that only when default $\tau \geq t_i$, the protection buyer will pay the i coupon(protection fee) for protection period i. If default time $\tau < t_{i-1}$, the buyer does not pay the i coupon for protection period i. If default time $t_{i-1} \leq \tau < t_i$, the buyer only pay the partial i coupon for protection period from t_{i-1} to τ.*
- *Details of coupon payment relation and default time are given by*

Default time τ	*PV of full premium payment*	*PV of accrued premium payment*
$t_0 \leq \tau < t_1$	0	$C \times \Delta(t_0, \tau) \times DF(\tau)$
$t_1 \leq \tau < t_2$	$C \times \Delta_1 \times DF(t_1)$	$C \times \Delta(t_0, \tau) \times DF(\tau)$
$t_2 \leq \tau < t_3$	$C \times \Delta_2 \times DF(t_2)$	$C \times \Delta(t_2, \tau) \times DF(\tau)$
$\cdots$	$\cdots$	$\cdots$
$t_{n-1} \leq \tau < t_n$	$C \times \Delta_{n-1} \times DF(t_{n-1})$	$C \times \Delta(t_{n-1}, \tau) \times DF(\tau)$
$t_n \leq \tau$	$C \times \Delta_n \times DF(t_n)$	0

where $\Delta_i = t_i - t_{i-1}$*, and* $\Delta(t_i, \tau)$ *is the year fraction for period* $[t_i, \tau]$*, and* $DF(t)$ *is the discount factor from* t_S *to* t*.*

Lemma 23.4.3 (present value of premium leg). [9]*Consider a set of coupon payment dates* $t_0 < t_1 < \cdots < t_N = T$ *and the protection starting date and the valuation date be* $t_S, t_0 < t_S < t_1$.

- *The* ***present value of the premium leg excluding the partial coupon*** *is given by*

$$\begin{aligned} PV_{fullCoupon}(t_S) &= C \times E_Q[\sum_{i=1}^N \Delta_i df(t_S, t_i)\mathbf{1}(\tau > t_i)] \\ &= C \times \sum_{i=1}^N \Delta_i DF(t_S, t_i) Q(t_i) \end{aligned}$$

  where

$$df(t_S, t_i) = \exp(-\int_{t_0}^{t_i} r(s)ds), DF(t_S, t_i) = E_Q[df(t_S, t_i)|\mathcal{F}_S], Q(t_i) = E[\mathbf{1}(\tau > t_i)] = Pr(\tau > t_i).$$

- *The* ***present value of the partial coupon****, which pays at default time, is given by*

$$\begin{aligned} PV_{partialCoupon}(t_S) &= C \times E_Q[\sum_{i=1}^N (\tau - t_{i-1}) df(t_S, \tau)\mathbf{1}(t_{i-1} \leq \tau \leq t_i)] \\ &= -C \times \sum_{i=1}^N (\int_{t_{i-1}}^{t_i} DF(t_S, s)(s - t_{i-1})dQ(s) \end{aligned}$$

- *The full present **clean value** of the premium leg is given by*

$$\begin{aligned} PV_{premium,clean} &= PV_{fullCoupon} + PV_{partialCoupon} \\ &= C \times \sum_{i=1}^{N} (\Delta_i DF(t_S, t_i) Q(t_i) - \int_{t_{i-1}}^{t_i} DF(t_0, s)(s - t_{i-1}) dQ(s)). \end{aligned}$$

*In particular, we define **RPV01** by*

$$RPV01 = \sum_{i=1}^{N} (\Delta_i DF(t_S, t_i) Q(t_i) - \int_{t_{i-1}}^{t_i} DF(t_0, s)(s - t_{i-1}) dQ(s)).$$

- *If the hazard curve $h(t)$ is piece-wise constant with kinks at $t_1, ..., t_N$, i.e,*

$$h(t) = \begin{cases} \lambda_1, t_S < t < t_1 \\ \lambda_2, t_1 < t < t_2 \\ \cdots \\ \lambda_n, t_{n-1} < t < t_n \end{cases}$$

And further assume the partial coupon is paid at the next immediate coupon date (instead of default time) then

$$PV_{premium,clean} = C \sum_{i=1}^{N} DF(t_S, t_i) [\frac{Q(t_{i-1}) - Q(t_i)}{\lambda_i}]$$

and

$$RPV01 = \sum_{i=1}^{N} DF(t_S, t_i) [\frac{Q(t_{i-1}) - Q(t_i)}{\lambda_i}].$$

Note that we call this clean value because it does not follow the contract spefication.

Proof. (1) straight forward from the cash flow in premium leg (see Remark 23.4.2). Denote $DF_i = DF(t_S, t_i), df_i = df(t_S, t_i), \Delta_i = t_i - t_{i-1}, \Delta_1 = t_1 - t_S$.. Let F_τ, f_τ be the cmf and pdf of default time τ. Let $Q_i = Pr(\tau > t_i) = \int_{t_i}^{\infty} f_\tau(t)dt$. We have

$$\begin{aligned}
PV_{prem.clean} &= \sum_{i=1}^{N} E_Q[C \cdot \Delta_i \cdot \mathbf{1}_{\tau > t_i} \cdot df_i] + E_Q[C \cdot (\tau - t_{i-1}) \cdot \mathbf{1}_{t_{i-1} \leq \tau \leq t_i} \cdot df_i] \\
&= C \sum_{i=1}^{N} df_i [\int_{t_i}^{\infty} \Delta_i f_\tau(t)dt + \int_{t_{i-1}}^{t_i} (t - t_{i-1}) f_\tau(t)dt] \\
&= C \sum_{i=1}^{N} DF_i [\Delta_i PND_i + \int_{t_{i-1}}^{t_i} (t - t_{i-1}) dF_\tau(t)] \\
&= C \sum_{i=1}^{N} DF_i [\Delta_i PND_i + (t_i - t_{i-1}) F(t_i) - \int_{t_{i-1}}^{t_i} F_\tau(t)dt] \\
&= C \sum_{i=1}^{N} DF_i [\Delta_i - \int_{t_{i-1}}^{t_i} F_\tau(t)dt] \\
&= C \sum_{i=1}^{N} DF_i [\int_{t_{i-1}}^{t_i} 1dt - \int_{t_{i-1}}^{t_i} F_\tau(t)dt] \\
&= C \sum_{i=1}^{N} DF_i [\int_{t_{i-1}}^{t_i} (1 - F_\tau(t))dt] \\
&= C \sum_{i=1}^{N} DF_i [\int_{t_{i-1}}^{t_i} \exp(-\int_{t_0}^{t} \lambda(s)ds)dt] \\
&= C \sum_{i=1}^{N} DF_i [\frac{Q_{i-1} - Q_i}{\lambda_i}].
\end{aligned}$$

□

Remark 23.4.2 (understand the coupon cash flow in the premium leg). Let $t_1 < t_2 < \cdots < t_n$ denote the coupon payment dates. Let τ denote the default time.

- Let t_0 be the current time. The discounted cash flow is

$$\begin{aligned}
CL = &\mathbf{1}(t_1 \leq \tau < t_2) C \Delta_1 DF(t_0, t_1) \\
&+ \mathbf{1}(t_2 \leq \tau < t_3) C(\Delta_1 DF(t_0, t_1) + \Delta_2 DF(t_0, t_2)) \\
&+ \mathbf{1}(t_3 \leq \tau < t_4) C(\Delta_1 DF(t_0, t_1) + \Delta_2 DF(t_0, t_2) + \Delta_3 DF(t_0, t_3)) \\
&+ \cdots \\
&+ \mathbf{1}(t_n \leq \tau) C(\sum_{i=1}^{n} \Delta_i DF(t_0, t_i))
\end{aligned}$$

- The cash flow can be rewritten by

$$\begin{aligned}
CL = & C\Delta_1 DF(t_0, t_1)(\mathbf{1}(t_1 \leq \tau < t_2) + \mathbf{1}(t_1 \leq \tau < t_2) + \cdots + \mathbf{1}(t_n \leq \tau)) \\
& + C\Delta_2 DF(t_0, t_2)(\mathbf{1}(t_2 \leq \tau < t_3) + \mathbf{1}(t_3 \leq \tau < t_4) + \cdots + \mathbf{1}(t_n \leq \tau)) \\
& + \cdots \\
& + C\Delta_n DF(t_0, t_n)\mathbf{1}(t_n \leq \tau) \\
& = C\Delta_1 DF(t_0, t_1)\mathbf{1}(t_1 \leq \tau) \\
& + C\Delta_2 DF(t_0, t_2)\mathbf{1}(t_2 \leq \tau) \\
& + \cdots \\
& + C\Delta_n DF(t_0, t_n)\mathbf{1}(t_n \leq \tau)
\end{aligned}$$

where we used the equality

$$(\mathbf{1}(t_1 \leq \tau < t_2) + \mathbf{1}(t_1 \leq \tau < t_2) + \cdots + \mathbf{1}(t_n \leq \tau)) = \mathbf{1}(t_1 \leq \tau).$$

Definition 23.4.5 (present value of a CDS contract). *Consider the perspective of a protection buyer.*

- *Let t be the current date. Let t_S be the protection starting date and $t \leq t_S$. Let t_1 be the next coupon date and t_0 be the previous coupon payment date. Then* ***accrued premium***[a]*, denoted by AP, is given by*

$$AP = C \times (t_S - t_0) \times DF(t, t_S)Q(t, t_S),$$

where In the case where the current date is the protection starting date, i.e., $t = t_s$, we have

$$AP = C \times (t_S - t_0),$$

where C is the premium coupon rate and $(t_S - t_0)$ is the year fraction of the no-protection period.
- *The* ***present clean value*** *of a CDS contract is*

$$V_{clean}(t) = PV_{default}(t) - PV_{premium,clean}(t).$$

- *The* ***present dirty value*** *of a CDS contract is*

$$\begin{aligned}
V(t) &= PV_{default}(t) - PV_{premium,dirty} \\
&= PV_{default}(t) - PV_{premium,clean} - AP.
\end{aligned}$$

[a] AP is the positive amount the protection seller should compensate the protection buyer for the no-protection between t_0 and t_S

Remark 23.4.3 (understand dirty price and clean price).
- Dirty value is calculated based on the cash flow specificed by the contract; **Dirty value the fair market value of the cash flows**.
- Clean value calculation is not completely following the contract because at the first coupon date (t_1), only fraction of coupon $(t_1 - t_S) \times C$ is paid instead of the full coupon $(t_1 - t_0) \times C$, which is specified in the contract. Therefore, clean value is not the fair value.
- From the perspective of the buyer, dirty value is smaller than clean value.

23.4.3.3 *Continuous-time CDS pricing*

Lemma 23.4.4 (price of a continuous-time CDS). *Consider a CDS with protection period from 0 to time T. Denote the CDS spread by S^{CDS}, which is the premium paid per unit time by the protection buyer. Assuming deterministic hazard rate function $h(t)$, recovery rate R, and short rate function $r(t)$. We have*

- *The value of the fixed/premium leg at current time $t = 0$ is*

$$V_{fixed}(0) = s^{CDS} \int_0^T \exp(-\int_0^u (h(s) + r(s))ds)du.$$

- *The value of the floating/default leg (paying (1-R) if default is*

$$V_{float}(0) = (1 - R) \int_0^T h(u) \exp(-\int_0^u (h(s) + r(s))ds)du.$$

- $$s^{CDS} = (1 - R) \frac{\int_0^T h(u) \exp(-\int_0^u (h(s) + r(s))ds)du}{\int_0^T \exp(-\int_0^u (h(s) + r(s))ds)du}$$

Proof. (1) We use $0 = T_0 < T_1 < \cdots < T_n = T$ to partition the interval $[0, T]$, we have

$$\begin{aligned}
V_{fixed}(0; n) &= S^{CDS} \sum_{i=1}^n (T_i - T_{i-1}) \exp(-\int_0^{T_i} r(s)ds) Pr(\tau > T_i) \\
&= S^{CDS} \sum_{i=1}^n (T_i - T_{i-1}) \exp(-\int_0^{T_i} r(s)ds) \exp(-\int_0^{T_i} h(s)ds \\
&= S^{CDS} \sum_{i=1}^n (T_i - T_{i-1}) \exp(-\int_0^{T_i} (r(s) + h(s))ds)
\end{aligned}$$

where we the default probability characterization in Lemma 23.2.1 such that

$$Pr(\tau > T_i) = \exp(-\int_0^{T_i} h(s)ds).$$

As we refine the partition $n \to \infty$, we have

$$V_{fixed}(0;\infty) = V_{fixed}(0) = s^{CDS} \int_0^T \exp(-\int_0^u (h(s)+r(s))ds)du.$$

(2)We use $0 = T_0 < T_1 < \cdots < T_n = T$ to partition the interval $[0,T]$, we have

$$\begin{aligned}
V_{float}(0;n) &= (1-R)\sum_{i=1}^n \exp(-\int_0^{T_i} r(s)ds)Pr(T_{i-1} < \tau < T_i) \\
&= (1-R)\sum_{i=1}^n \exp(-\int_0^{T_i} r(s)ds)S(T_{i-1})(1-\exp(-\int_{T_{i-1}}^{T_i} h(s)ds)) \\
&= (1-R)\sum_{i=1}^n \exp(-\int_0^{T_i} r(s)ds)\exp(-\int_0^{T_{i-1}} h(s)ds)(1-\exp(-\int_{T_{i-1}}^{T_i} h(s)ds) \\
&= (1-R)\sum_{i=1}^n \exp(-\int_0^{T_i} (r(s)+h(s))ds)(1-\exp(-\int_{T_{i-1}}^{T_i} h(s)ds) \\
&\approx (1-R)\sum_{i=1}^n \exp(-\int_0^{T_i} (r(s)+h(s))ds)h(T_i)(T_i - T_{i-1})
\end{aligned}$$

where we the default probability characterization in Lemma 23.2.1 such that

$$Pr(T_1 < \tau < T_2) = S(T_1)(1-\exp(-\int_{T_1}^{T_2} h(s)ds)).$$

As we refine the partition $n \to \infty$, we have

$$V_{float}(0;\infty) = V_{float}(0) = (1-R)\int_0^T h(u)\exp(-\int_0^u (h(s)+r(s))ds)du.$$

□

Lemma 23.4.5 (credit spread triangle). *Consider a CDS with protection period from 0 to time T. Denote the CDS spread by S^{CDS}, which is the premium paid per unit time by the protection buyer. Assuming deterministic hazard rate function $h(t)$, recovery rate R, and short rate function $r(t)$.* ***If** $h(t), r(t)$ **are constant functions**, we have*

$$h = \frac{S^{CDS}}{(1-R)}.$$

Proof. (1) Note that for constant hazard rate and short rate, we have

$$\begin{aligned}&\int_0^T h(u)\exp(-\int_0^u (h(s)+r(s))ds)du\\&=\int_0^T h\exp(-(h+r)u)du\\&=\frac{h}{h+r}(1-\exp(-(h+r)T))\end{aligned}$$

and

$$\begin{aligned}&\int_0^T \exp(-\int_0^u (h(s)+r(s))ds)du\\&=\int_0^T \exp(-(h+r)u)du\\&=\frac{1}{h+r}(1-\exp(-(h+r)T))\end{aligned}$$

therefore,

$$s^{CDS}=(1-R)\frac{\int_0^T h(u)\exp(-\int_0^u (h(s)+r(s))ds)du}{\int_0^T \exp(-\int_0^u (h(s)+r(s))ds)du}=(1-R)h.$$

□

Remark 23.4.4 (Another interpretation for nonconstant function in infinitesimal time intervals)**.**

- The protection buyer pays(on expectation sense) $S^{CDS}\delta t(1-h\delta t)$ during the interval $[t,t+\delta t]$, where $(1-h\delta t)$ is the probability of not defaulting conditioned on no default occurs before t. The protection seller pays(on expectation sense) $h\delta t(1-R)$ during the interval $[t,t+\delta t]$,where $(h\delta t)$ is the probability of defaulting conditioned on no default occurs before t.
  Equaling the two payment and we will get the result.
- Note that we get the equation under the assumption that the discounted cash flow in an infinitesimal time interval should equal; Usually, we require the discounted cash flow in the whole protection period to be equal.

Lemma 23.4.6 (value evolution for a continuous-time CDS). *Consider a CDS with protection period from 0 to time T. Let current time be $t, t\in[0,T]$. Let C be the contract spread.*

- *The current value of CDS contract is given by*

$$\begin{aligned} V(t) &= V_{float}(t) - V_{fixed}(t) \\ &= (1-R)\int_t^T h(u)\exp(-\int_t^u (h(s)+r(s))ds)du - C\int_t^T \exp(-\int_t^u (h(s)+r(s))ds)du. \end{aligned}$$

- *Assume the hazard rate and short rate are flat, i.e.,* $h(t) = h_0, r(t) = r_0$. *Then the value of the floating/default leg (paying (1-R) if default is*

$$\begin{aligned} V(t) &= ((1-R)h_0 - C)\int_t^T \exp(-\int_t^u (h_0+r_0)ds)du \\ &= ((1-R)h_0 - C)\frac{1-\exp(-(r_0+h_0)(T-t))}{r_0+h_0} \end{aligned}$$

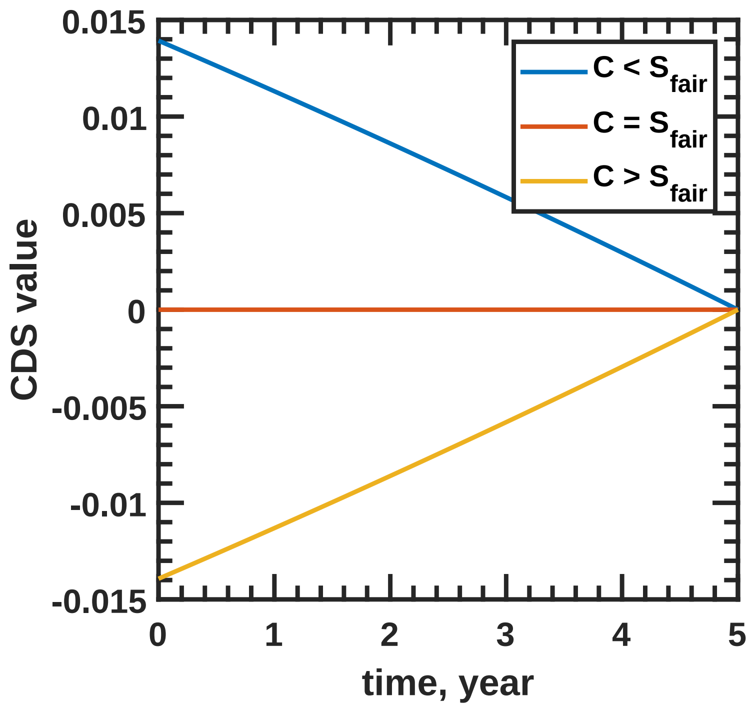


Figure 23.4.1

23.4.3.4 *Monte Carlo method for CDS pricing*

Methodology 23.4.1. *Suppose we are given by the cdf of default time* τ, *denoted by* F_τ.

We can use the following procedure to generate

- *generate a sample* $\tau^{(i)}$ *using* F_τ.

- *evaluate the premium leg value* $V^{(i)}_{premium}$ *and default leg value* $V^{(i)}_{protection}$
- *generate a sample value* $V^{(i)} = V^{(i)}_{protection} - V^{(i)}_{premium}$.
- $$\hat{V}_{clean} = \frac{1}{N}\sum_{i=1}^{N} V^{(i)}.$$

23.4.3.5 *Risk management of CDS*

Remark 23.4.5 (risk factors of a CDS swap).

Lemma 23.4.7 (risk due to cds spread).

Lemma 23.4.8 (risk due to interest rate, qualitative result). *Consider a long position in CDS such that is value is default payment minus premium payment(see Figure 23.4.2).*

- *If the reference entity has relatively low probability to default, then the value of CDS will increase when interest rate increases.*
- *If the reference entity has relatively high probability to default, then the value of CDS will decrease when interest rate increases.*

Proof. (informal) Note that when interest rate increases, the default payment and the premium payment will both decrease since discount factors are decreasing. We can view default payment and premium as certain form of bonds and we compare their dollar duration. (1) for entity with low probability to default, the dollar duration of default payment is small since it is very likely not to receive any default payment; on the other hand, the dollar duration for the premium payment is large, since it is very likely the CDS buyer has to finish all the payments. (2) for entity with high probability to default, the dollar duration of default payment is big since it is very likely to receive the default payment; on the other hand, the dollar duration for the premium payment is small, since it is very likely the CDS buyer will stop premium payment when default occurs half-way. □

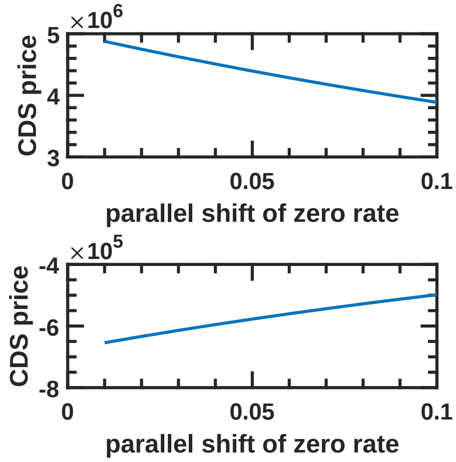


Figure 23.4.2: Effect of interest rate parallel shifting up on CDS price. (top) high probability of default case; (below) low probability of default case.

Lemma 23.4.9 (CDS time risk).

- *In general, for increasing hazard rate curve, the clean price of a CDS long position will increase when approaching maturity; for decreasing hazard rate curve, the clean price of a CDS long position will decrease when approaching maturity; For perfectly flat hazard curve, the clean price*
- *Assume zero interest rate and flat hazard curve, the time path for a continuously CDS is given by*

Proof. (1)(informal) the long position of a CDS will receive default payment and pay out premium. When approaching maturity, the premium payment is decreasing, □

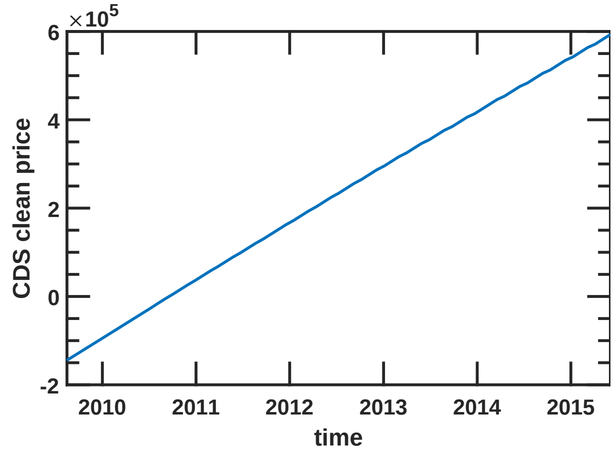


Figure 23.4.3: CDS clean price(buy protection) time path for a CDS contract maturing on 20-Sep-2015.

23.4.4 Amortizing credit default swap

23.4.5 CDS option pricing

Definition 23.4.6 (CDS option). *A credit default swap option, or CDS option, is also known as a credit default swaption. It is an option on a credit default swap. A CDS option gives its holder the right, but not the obligation, to buy (call) or sell (put) protection on a specified reference entity for a specified future time period for a certain spread. The option is knocked out if the reference entity defaults during the life of the option. This knock-out feature marks the fundamental difference between a CDS option and a vanilla option. Most commonly traded CDS options are European style options.*

23.4.6 Default digital swaps

Lemma 23.4.10 (price of a credit default swap). *[6, p. 66] Denote the spread by S^{dig}.*

- *The value of the fixed/premium leg is*

$$V_{fixed}(t) = s^{CDS}\sum_{i=1}^{n}\delta_{i-1}P_0^d(t,T_i)$$

- *The value of the floating/default leg (paying 1\$) if defaults in (T_i, T_{i+1})) is*

$$V_{float}(t) = \sum_{i=1}^{n} e(t, T_{i-1}, T_i) = \sum_{i=1}^{n} \delta_{i-1} H(t, T_{i-1}, T_i) P_0^d(t, T_i)$$

- $$s^{dig} = \frac{\sum_{i=1}^{n} \delta_{i-1} H(t, T_{i-1}, T_i) P_0^d(t, T_i)}{\sum_{i=1}^{n} \delta_{i-1} P_0^d(t, T_i)}$$

Proof. Straight forward. □

23.4.7 Asset swap packages

Definition 23.4.7 (asset swap package). *An asset swap is a combination of a defaultable bond with a fixed for floating interest rate swap. The bond coupon is swapped into a floating rate based on LIBOR plus a spread. This spread is known as* ***asset swap spread****.*

Example 23.4.1. [6, p. 13] Suppose A and B enter the asset swap contract. At each coupon date $T_i, i = 1, ..., N$.

- B pays to A: c, the amount of the fixed coupons of a bond issued by company C. The bond is defaultable.
- A pays to B: Libor and the asset swap spread.

23.4.8 Total return swaps

Definition 23.4.8. *[6, p. 13] In a total return swap, A and B agree to exchange all cash flows that arise from two different investments. Usually one of these two investments is defaultable investment, and the other is default free.*

Example 23.4.2. [6, p. 13] Suppose A and B enter a total return swap.

A pays to B at regular payment dates $T_i, i \leq N$:

- The coupon c of the bond issued by company C, which is defaultable.
- The price appreciation of the bond, $(C(T_i) - C(T_{i-1}))^+$ since last payment.
- The principal repayment of the bond C(at the final payment date).
- The recovery value of the bond(if there was a default).

B pays to A at the same regular payment dates $T_i, i \leq N$:

- A regular fee of Libor plus a spread
- The price depreciation of the bond, $(-C(T_i) + C(T_{i-1}))^+$ since last payment.
- The par value of the bond(if there was a default in the meantime)

23.5 Estimating default term structures

23.5.1 Estimating from CDS

Remark 23.5.1.
- Estimating default term structures from CDS requires the input recovery rate;
- Market implied recovery rates are usually unavaiable due to the lack of liquid instruments traded on the recovery rate;
- If there exists digial CDS, default term structure can be estimated without knowing the recovery rate.

23.6 PDE methods

23.6.1 Non-hedgeable default

Lemma 23.6.1. *[4, pp. 652, 656] Assume the short rate has dynamics in the real probability measure*

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

and default-free zero-coupon bond $Z(r,t;T)$(e.g. treasury bill) is governed by (Theorem 19.5.1)

$$\frac{\partial Z}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 Z}{\partial r^2} + m\frac{\partial Z}{\partial r} - rV = 0.$$

where m is the drift of the r in the risk-neural measure. Consider a defaultable zero-coupon bond(e.g. corporate bond) that pays 1 at maturity T. Assume the default event follows a Poisson process ***independent of*** *r with rate parameter h and has recovery rate $R \in [0,1]$. Then, the current value of the defaultable zero-coupon bond is governed by*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m\frac{\partial V}{\partial r} - (r_d + (1-R)h)V = 0,$$

with final condition $V(r,T;T) = 1$. In here $r_d(h,\lambda)$ is the required expected growth rate of the market for a portfolio without market risk but subjects to default risk.

Proof. Construct a portfolio

$$\Pi = V - \Delta Z.$$

Based on the definition of Poisson process (**??**), during time interval $[t, t+dt]$, there is $(1-hdt)$ probability that no jump will happen, therefore

$$d\Pi = ((\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2})dt + \frac{\partial V}{\partial r}dr) - \Delta((\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2})dt + \frac{\partial V}{\partial r}dr).$$

When a jump(i.e. default) occurs with probability hdt, we have

$$d\Pi = -(1-R)V + O((dt)^{1/2}),$$

where $O((dt)^{1/2})$ represents variations due to Brownian motion.

If we choose $\Delta = \frac{\partial V}{\partial r} / \frac{\partial Z}{\partial r}$ and require the expected growth rate to be r_d, we have

$$\begin{aligned}
&E[d\Pi] = r\Pi dt \\
&\implies (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m\frac{\partial V}{\partial r})dt(1-hdt) + hdt(-(1-R)V + O((dt)^{1/2})) = rVdt \\
&\implies \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m\frac{\partial V}{\partial r} - (r_d + (1-R)h)V = 0.
\end{aligned}$$

□

Lemma 23.6.2 (price of a zero-recovery defaultable zero-coupon bond). *Assume the current value of a defaultable zero-recovery zero-coupon bond is governed by*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m\frac{\partial V}{\partial r} - (r_d + h)V = 0,$$

with final condition $V(r, T; T) = 1$.

Then there exists a risk-neutral measure Q such that there exists a measure Q such that

- $$V(t, r) = E_Q[\exp(\int_t^T r_d(s) + h(s)ds|\mathcal{F}_t] = P(t, T)E_Q[\exp(\int_t^T r_d(s) - r(s) + h(s)ds|\mathcal{F}_t].$$
- *The dynamics of r under Q is*
  $$dr = mdt + \sigma dW_t,$$
  where W_t is a Brownian motion.
- *Under risk-neutral measure Q, the default Poisson process is characterized by intensity parameter $r_d - r + h$.*

Remark 23.6.1 (jump probability in real-world measure or risk-neutral measure)**.** In this risk-neutral world, the jump/default intensity parameter increases from h in the real world to $r_d - r + h$ to account for the risk aversion.

23.6.2 Hedgeable default

See [10]

23.6.3 Hedging the default

Lemma 23.6.3. *[4, p. 661] Assume the short rate has dynamics in the real probability measure*

$$dr = \mu dt + \sigma_r dW_t.$$

Consider a risky bond(experience both default and interest rate risk) $V(r, t)$. Then its value is governed by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma_r^2 \frac{\partial^2 V}{\partial r^2} + (\mu - \lambda_r \sigma_r)\frac{\partial V}{\partial r} - (r + \lambda_d(r,t))V = 0,$$

with final condition $V(r_T, T) = 1$.

Proof. We introduce another risky bond V_1into the portfolio to help hedging the default. To do this we assume that default in one bond automatically leads to the default of the other. Consider the portfolio

$$\Pi = V - \Delta Z - \Delta_1 V_1,$$

where Δ is the riskless bond.

Let p be the default probability. We have

$$dV = (\frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial r}dr + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}(dr)^2)(1 - pdt) + ((1-R))Vpdt,$$

$$dZ = (\frac{\partial Z}{\partial t}dt + \frac{\partial Z}{\partial r}dr + \frac{1}{2}\frac{\partial^2 Z}{\partial r^2}(dr)^2),$$

and

$$dV_1 = (\frac{\partial V_1}{\partial t}dt + \frac{\partial V_1}{\partial r}dr + \frac{1}{2}\frac{\partial^2 V_1}{\partial r^2}(dr)^2)(1 - pdt) + ((1-R))V_1 pdt.$$

If we choose

$$\Delta_1 = \frac{V}{V_1}, \Delta = \frac{V_1\frac{\partial V}{\partial r} - V\frac{\partial V_1}{\partial r}}{V_1\frac{\partial Z}{\partial r}},$$

we can eliminate both the default risk and the interest rate risk

□

23.7 Default correlation modeling

23.7.1 Sampling correlated default time

Remark 23.7.1. Let τ be the random default time with cdf F_τ. Then we can sample τ in the following two ways(see **??**):

- Draw a random number U with uniform distribution $U([0,1])$, then

$$\tau = F_\tau^{-1}(u).$$

  Or equivalently, $\tau < t$ if and only if $U < F_\tau(t)$,i.e.,

$$Pr(\tau < t) = Pr(u < F_\tau(t)) = F_\tau(t).$$

- Draw a random number $Z \sim N(0,1)$, then

$$\tau = F_\tau^{-1}(\phi(u)),$$

  where ϕ is the cdf for a standard normal distribution. Or equivalently, $\tau < t$ if and only if $Z < \phi^{-1}(F_\tau(t))$,i.e.,

$$Pr(\tau < t) = Pr(Z < \phi^{-1}(F_\tau(t)) = F_\tau(t).$$

Lemma 23.7.1 (generation of dependent default time). *Let $T_1, T_2, ..., T_n$ denote the random default time for n parties. Assume the hazard curve for each party is given by $h_i(t), t \geq 0$ such that the marginal cdf of default time is given by*

$$F_i(t) = Pr(T_i \leq t) = 1 - \exp(-\int_0^t h(s)ds).$$

Further assume the copula associated with the joint cdf is Gaussian copula with correlation matrix R.

Consider the following random number generating process

- *Simulate $Y_1, Y_2, ..., Y_n$ from $MN(0,R)$.*
- *Obtain $T_1, T_2, ..., T_n$ using $T_i = F_i^{-1}(\phi(Y_i))$, where ϕ is the cdf of a standard normal.*

$T_1, T_2, ..., T_n$ will follow the joint cdf F.

Proof. Directly from **??**. □

Note 23.7.1 (first default time random sample from simulation). In each simulation we generate one sample of default times $t_1, t_2, ..., t_n$. The first-to-default time is simply $t = \min(t_1, t_2, ..., t_n)$.

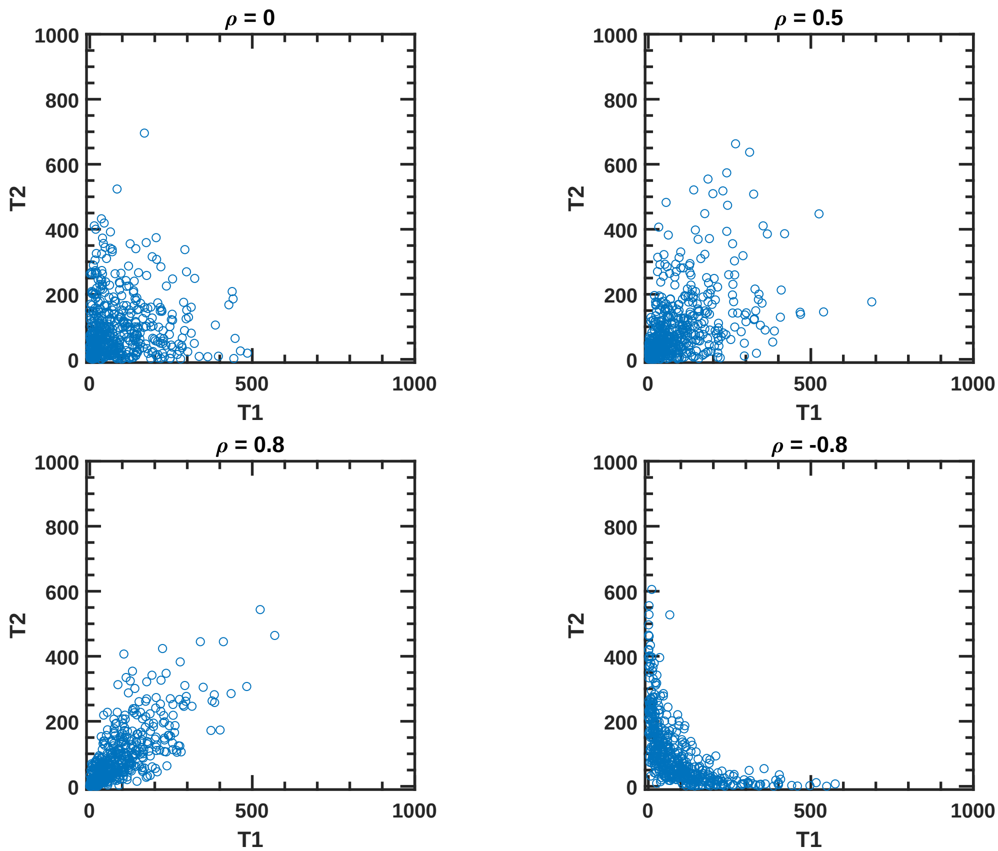


Figure 23.7.1: Generated correlated default time via Gaussian copula with different correlations. The hazard rate for both parties is $h(t) = 0.01$.

23.7.2 First-to-default modeling and valuation

23.7.2.1 *General principles*

Lemma 23.7.2 (first-to-default probability and contract valuation under independence assumption). *Let $T_1, T_2, ..., T_n$ denote the* ***independent*** *random default time for n*

parties. Assume the hazard curve for each party is given by $h_i(t), t \geq 0$ such that the marginal cdf of default time is given by

$$F_i(t) = Pr(T_i \leq t) = 1 - \exp(-\int_0^t h(s)ds).$$

It follows that

- *The first-to-default time $T = \min\{T_1, T_2, ..., T_n\}$ has cdf and density given by*

$$F_T(t) = 1 - \exp(-\int_0^t \sum_{i=1}^n h_i(s)ds), f_T(t) = (\sum_{i=1}^n h_i(s)) \exp(-\int_0^t \sum_{i=1}^n h_i(s)ds).$$

- *Consider a contract maturing at T_c which pays 1 if the first default among n parties occur between 0 and T_c. The value of the contract at time 0 is*

$$V(0) = \int_0^{T_c} P(0,u) f_T(t) dt.$$

where $P(0,u)$ is the zero coupon bond price at 0 with maturity u.

Proof. (1) From **??**, we know that

$$F_T(t) = 1 - \prod_{i=1}^n (1 - F_i(t)) = 1 - \prod_{i=1}^n \exp(-\int_0^t h_i(s)ds) = 1 - \exp(-\int_0^t \sum_{i=1}^n h_i(s)ds).$$

(2) Straight forward. □

23.7.2.2 *Multivariate Gaussian copula approximation*

Lemma 23.7.3 (Multivariate Gaussian copula approximation for joint default modeling). *Let $T_1, T_2, ..., T_n$ denote the random default time for n parties. Assume marginal cdf of default time is given by $F_i(t)$. Further* ***assume the copula associated with the joint cdf*** *is Gaussian copula with correlation matrix R.*

- *Then $T_1, T_2, ..., T_n$ will follow the joint cdf F given by*

$$F_{T_1,T_2,...,T_n}(t_1, t_2, ..., t_n) = \Phi(\phi^{-1}(F(T_1)), \phi^{-1}(F(T_2)), ..., \phi^{-1}(F(T_n))).$$

- *Let $T = \min(T_1, T_2, ..., T_n)$ be the first default time. Then*

$$F_T(t) = 1 - F_{T_1,T_2,...,T_n}(t, t, ..., t)$$

Proof. (1) See **??**. (2)

$$\begin{aligned} 1 - F_T(t) &= Pr(T > t) \\ &= Pr(T_1 > t, T_2 > t, ..., T_n > t) \\ &= \end{aligned}$$

□

23.7.2.3 *Gaussian one-factor model*

Lemma 23.7.4 (Multivariate Gaussian copula approximation for Bernoulli default modeling). *Consider n parties that will default in the next period with unconditional probability p_i. Define a new proxy random variable X_i by*

$$X_i = a_i F + \sqrt{1 - a_i^2} Z_i, i = 1, 2, ..., n,$$

where F is a common factor affecting defaults for all firms and Z_i is a factor affecting only firm i. F and Z_i are independent standard normal variables. It follows that

- $$Pr(X_i \text{ will default}) \triangleq p_i = Pr(X_i < \phi^{-1}(p_i)),$$
  where ϕ is the cdf of the standard normal variable.
- *The* ***conditional default probability*** *of firm i before t conditioning on the observation of F is given by*
  $$Pr(X_i \text{ will default} | F = f) \triangleq Pr(T_i < t | F = f) = \phi(\frac{\phi^{-1}(Q_i(t)) - a_i f}{\sqrt{1 - a_i^2}}),$$
- *The* ***unconditional probability of no default*** *is given by*
  $$Pr(\text{no default}) \triangleq \int_{-\infty}^{\infty} \prod_{i=1}^{n} (1 - Pr(X_i \text{ will default} | F = f)) g(f) df$$
  where $g(f) = \frac{1}{\sqrt{2\pi}} \exp(-f^2/2)$.

Proof. (1) Note that $X_i \sim N(0, 1)$. Therefore,

$$Pr(X_i < \phi^{-1}(p_i)) = \phi(\phi^{-1}(p_i)) = p_i.$$

(2)

$$\begin{aligned} Pr(X_i < \phi^{-1}(p_i)) &= Pr(a_i F + \sqrt{1-a_i^2} Z_i < \phi^{-1}(p_i)) \\ &= Pr(Z_i < \frac{\phi^{-1}(p_i) - a_i F}{\sqrt{1-a_i^2}}) \end{aligned}$$

(3) straight forward. □

Lemma 23.7.5 (Multivariate Gaussian copula approximation for joint default modeling with single factor). *[5, p. 563] Let $T_1, T_2, ..., T_n$ denote the random default time for n parties. Assume marginal cdf of default time is given by $Q_i(t)$. Define a new proxy random variable $X_i = \phi^{-1}(Q_i(T_i)), i = 1, 2, ..., n$, and* ***assume*** *$X_i$ can be modeled by*

$$X_i = a_i F + \sqrt{1-a_i^2} Z_i, i = 1, 2, ..., n,$$

where F is a common factor affecting defaults for all firms and Z_i is a factor affecting only firm i. F and Z_i are independent standard normal variables. It follows that

- *The* ***conditional default probability*** *of firm i before t conditioning on the observation of F is given by*

$$Q_i(t|F = f) \triangleq Pr(T_i < t|F = f) = \phi(\frac{\phi^{-1}(Q_i(t)) - a_i f}{\sqrt{1-a_i^2}}),$$

  where ϕ is the standard normal cdf.
- *The* ***unconditional default probability*** *is given by*

$$Q_i(t) = \int_{-\infty}^{\infty} \phi(\frac{\phi^{-1}(Q_i(T)) - a_i f}{\sqrt{1-a_i^2}}) g(f) df,$$

  where $g(f) = \frac{1}{\sqrt{2\pi}} \exp(-f^2/2)$.
- *Conditioning on the observation of F, the conditional probability of having m defaults among N firms is given by*

$$Pr(X = m|F = f) = \binom{N}{m} (p(y))^m (1 - p(y))^{N-m}.$$

and the unconditional probability is given by

$$Pr(X = m) = \int_{-\infty}^{\infty} \binom{N}{m} (p(y))^m (1 - p(y))^{N-m} g(f) df.$$

- *Let* $T = \min(T_1, T_2, ..., T_n)$. *Then the* ***conditional and unconditional first-to-default probabilities*** *are given by*

$$\begin{aligned} Q_T(t|F = f) &\triangleq Pr(T < t|F = f) \\ &= 1 - (1 - Q_i(t|F = f))^n \\ Q_T(t) &\triangleq Pr(T < t) \\ &= 1 - \int_{-\infty}^{\infty} (1 - Q_i(t|F = f))^n g(f) df. \end{aligned}$$

Proof. (1)

$$\begin{aligned} Pr(T_i < t) &= Pr(T_i < t) \\ &= Pr(Q_i(T_i) < Q_i(t)) \\ &= Pr(\phi^{-1}(Q_i(T_i)) < \phi^{-1}(Q_i(t))) \\ &= Pr(X_i < \phi^{-1}(Q_i(t))) \\ &= Pr(a_i F + \sqrt{1 - a_i^2} Z_i < \phi^{-1}(Q_i(t))) \\ &= Pr(Z_i < \frac{\phi^{-1}(Q_i(t)) - a_i F}{\sqrt{1 - a_i^2}}) \\ \implies Q_i(T|F = f) \triangleq Pr(T_i < t|F) &= Pr(Z_i < \frac{\phi^{-1}(Q_i(T)) - a_i F}{\sqrt{1 - a_i^2}} |F = f) \\ &= \phi(\frac{\phi^{-1}(Q_i(t)) - a_i f}{\sqrt{1 - a_i^2}}) \end{aligned}$$

(2) Use the fact that

$$Pr(T_i < t, F = f) = Pr(T_i < t|F = f) f_F(f),$$

and then marginalize out F. (3) Note that when conditioning on the F, the default behavior among firms are independent. Therefore the number of defaulting firms can be characterized by binomial distribution. (4) Note that when conditioning on the F, the default time T_is are independent from each other. Then we the first-to-default probability calculation under independence assumption result(see Lemma 23.7.2). □

Remark 23.7.2 (interpret proxy random variable and its representation)**.** Note that we define a proxy random variable X_i by $X_i = \phi^{-1}(Q_i(T_i))$. We can see that

- X_i is a standard normal variable because

$$Pr(X_i < x) = Pr(\phi^{-1}(Q_i(T_i)) < x) = Pr(Q_i(T_i) < \phi(x)) = \phi(x),$$

  where we used the fact that $Q_i(T_i)$ is a uniform random variable(**??**).
- We represent $X_i = a_i F + \sqrt{1-a_i^2} Z_i$ is consistent. Because $a_i F + \sqrt{1-a_i^2} Z_i \sim N(0,1)$.

Remark 23.7.3 (the correlation structure for the Gaussian copula implied by the factor model)**.** Consider a one factor model given by

$$V_i = a_i Y + \sqrt{1-a_i^2} Z_i, i = 1,2,...,n,$$

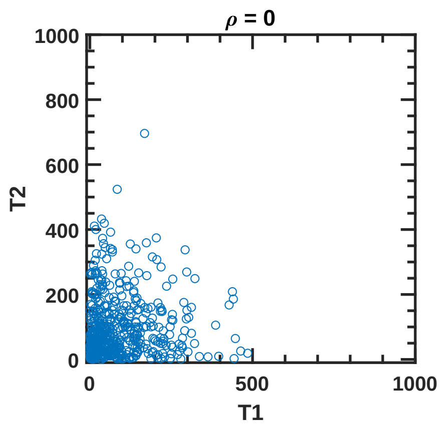
where $(F, Z_1, Z_2, ..., Z_n)$ are independent standard normal variables.

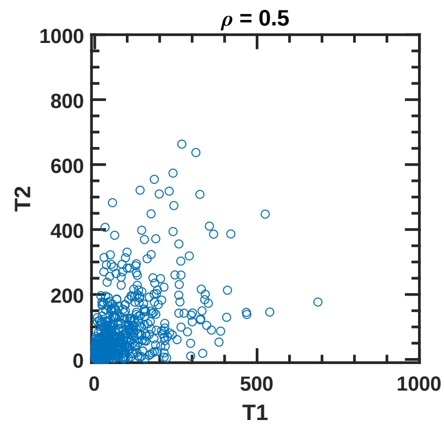
Then the correlation structure implied by the factor model is given by(**??**)

$$\begin{aligned}
Cov &= \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix} [a_1, a_2, ..., a_n] + \begin{bmatrix} 1-a_1^2 & & & \\ & 1-a_2^2 & & \\ & & \ddots & \\ & & & 1-a_n^2 \end{bmatrix} \\
&= \begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & a_2 a_2 & \cdots & a_2 a_n \\ \vdots & \cdots & \ddots & \vdots \\ a_n a_1 & a_n a_2 & \cdots & a_n a_n \end{bmatrix} \begin{bmatrix} 1-a_1^2 & & & \\ & 1-a_2^2 & & \\ & & \ddots & \\ & & & 1-a_n^2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & 1 & \cdots & a_2 a_n \\ \vdots & \cdots & \ddots & \vdots \\ a_n a_1 & a_n a_2 & \cdots & 1 \end{bmatrix}
\end{aligned}$$

(a) Calculation of first default probability(thick solid black) from individual default probability of 10 reference names using Gaussian one factor model with $\rho = 0.5$.

(b) First default probability as a funciton of correlation of the underlying names.





Remark 23.7.4 (the correlation structure for the Gaussian copula implied by the factor model). Consider a one factor model given by

$$V_i = \sqrt{\rho} Y + \sqrt{1-\rho} Z_i, i = 1, 2, ..., n,$$

where $(F, Z_1, Z_2, ..., Z_n)$ are independent standard normal variables.

Then the correlation structure implied by the factor model is given by(**??**)

$$Cov = \begin{bmatrix} \sqrt{\rho} \\ \sqrt{\rho} \\ \cdots \\ \sqrt{\rho} \end{bmatrix} [\sqrt{\rho}, \sqrt{\rho}, ..., \sqrt{\rho}] + \begin{bmatrix} 1-\rho & & & \\ & 1-\rho & & \\ & & \ddots & \\ & & & 1-\rho \end{bmatrix} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \cdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

23.7.2.4 *Gaussian multi-factor model*

Lemma 23.7.6 (Multivariate Gaussian copula approximation for joint default modeling with multiple factors). *Let $T_1, T_2, ..., T_n$ denote the random default time for n par-*

ties. Assume marginal cdf of default time is given by $Q_i(t)$. Define a new proxy random variable $X_i = \phi^{-1}(Q_i(T_i)), i = 1, 2, ..., n$, and ***assume*** *$X_i$ can be modeled by*

$$X_i = \sum_{j=1}^{m} a_{ij} F_j + \sqrt{1 - \sum_{j=1}^{m} a_{ij}^2} Z_i, i = 1, 2, ..., n,$$

where $F_1, F_2, ..., F_m$ are common factors affecting defaults for all firms and Z_i is a factor affecting only firm i. $F_1, F_2, ..., F_m$ and Z_i are mutually independent standard normal variables. It follows that

- *The* ***conditional default probability*** *of firm i before t conditioning on the observation of $F_1, F_2, ..., F_m$ is given by*

  $$Q_i(t|F_1 = f_1, ..., F_m = f_m) \triangleq Pr(T_i < t|F_1 = f_1, ..., F_m = f_m) = \phi(\frac{\phi^{-1}(Q_i(t)) - \sum_{j=1}^{m} a_{ij} f_j}{\sqrt{1 - \sum_{j=1}^{m} a_{ij}^2}}),$$

  where ϕ is the standard normal cdf.
- *The* ***unconditional default probability*** *is given by*

  $$Q_i(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(\frac{\phi^{-1}(Q_i(t)) - \sum_{j=1}^{m} a_{ij} f_j}{\sqrt{1 - \sum_{j=1}^{m} a_{ij}^2}}) g(f_1)...g(f_m) df_1...df_m,$$

  where $g(f) = \frac{1}{\sqrt{2\pi}} \exp(-f^2/2)$.
- *Let $T = \min(T_1, T_2, ..., T_n)$. Then the* ***conditional and unconditional first-to-default probabilities*** *are given by*

  $$\begin{aligned} Q_T(t|F_1 = f_1, ..., F_m = f_m) &\triangleq Pr(T < t|F_1 = f_1, ..., F_m = f_m) \\ &= 1 - (1 - Q_i(t|F_1 = f_1, ..., F_m = f_m))^n \\ Q_T(t) &\triangleq Pr(T < t) \\ &= 1 - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (1 - Q_i(t|F_1 = f_1, ..., F_m = f_m))^n g(f_1)...g(f_m) df_1...d \end{aligned}$$

Proof. See the single factor model(Lemma 23.7.5). □

23.7.3 Structural model approximation

[6, p. 305][5, p. 244]

23.7.3.1 *Factor model*

Lemma 23.7.7 (one-factor default model). *[6, p. 308] Assume the asset values of the firms are driven by one common driving factor Y such that the firm's asset is modeled by*

$$V_i = \sqrt{\rho}Y + \sqrt{1-\rho}Z_i, i = 1, 2, ..., N$$

where $Y \sim N(0, \sigma_Y^2)$, $\epsilon_i \sim N(0, w_i^2)$, *and* $(Y, Z_1, Z_2, ..., Z_N)$ *are independent.*

- *If a firm will default when its asset value is below K, then the default probability conditioning on the observation of factor value Y is given by*

$$p(y) = \Phi(\frac{K - \sqrt{\rho}y}{\sqrt{1-\rho}w_i}),$$

  where Φ *is the cdf for standard normal.*
- *The probability of having m defaults among N firms is given by*

$$Pr(X = m | F = f) = \binom{N}{m}(p(y))^m(1-p(y))^{N-m}.$$

Proof. (1)

$$\begin{aligned} p(y) &\triangleq Pr(V_i < K | Y = y) \\ &= Pr(\sqrt{\rho}Y + \sqrt{1-\rho}Z_i < K | Y = y) \\ &= Pr(Z_i < \frac{K - \sqrt{\rho}y}{\sqrt{1-\rho}} | Y = y) \\ &= \Phi(\frac{K - \sqrt{\rho}y}{\sqrt{1-\rho}w_i}) \end{aligned}$$

(2) Note that when conditioning on the Y, the default behavior among firms are independent. Therefore the number of defaulting firms can be characterized by binomial distribution. □

Remark 23.7.5 (interpretation). [6, p. 307] The systematic risk factor Y can be viewed as an indicator of the state of the business cycle, such as GDP, unemployment rate etc; and the idiosyncratic factor ϵ_n as a firm-specific effects factor such as the management strategy of the firm or the innovation of firm.

Lemma 23.7.8 (large uniform portfolio approximation).

Example 23.7.1. Consider a one factor model given by

$$V_i = \sqrt{\rho}Y + \sqrt{1-\rho}Z_i, i = 1,2,...,n,$$

where $(F, Z_1, Z_2, ..., Z_n)$ are independent standard normal variables.

Then the covariance structure implied by the factor model is given by

$$\begin{aligned}
Cov &= \begin{bmatrix} \sqrt{\rho} \\ \sqrt{\rho} \\ \cdots \\ \sqrt{\rho} \end{bmatrix} [\sqrt{\rho}, \sqrt{\rho}, ..., \sqrt{\rho}] + \begin{bmatrix} 1-\rho & & & \\ & 1-\rho & & \\ & & \ddots & \\ & & & 1-\rho \end{bmatrix} \\
&= \begin{bmatrix} \rho & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \vdots & \cdots & \ddots & \vdots \\ \rho & \rho & \cdots & \rho \end{bmatrix} \begin{bmatrix} 1-\rho & & & \\ & 1-\rho & & \\ & & \ddots & \\ & & & 1-\rho \end{bmatrix} \\
&= \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \cdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}
\end{aligned}$$

which is also a valid correlation matrix.

Lemma 23.7.9 (multiple factor default model). *[6, p. 312] Assume the asset values of the firms are driven by M driving factors $Y_j, j \leq M$ such that the firm's asset is modeled by*

$$V_i = \sum_{j=1}^{M} \beta_j Y_j + \epsilon_i, i = 1,2,...,N$$

where $Y = (Y_1, Y_2, ..., Y_N) \sim MN(0, \Omega_Y)$, $\epsilon_i \sim N(0, w_i^2)$, and $(Y, Z_1, Z_2, ..., Z_N)$ are independent. It follows that

- *If a firm will default when its asset value is below K, then the default probability conditioning on the observation of factor value $Y_1, Y_2, ..., Y_n$ is given by*

$$p(y) = \Phi(\frac{K - \sum_{i=1}^{M} \beta_i y_i}{w_i}),$$

where Φ is the cdf for standard normal.
- *The probability of having m defaults among N firms is given by*

$$Pr(X = m|Y = y) = \binom{N}{m}(p(y))^m(1 - p(y))^{N-m}.$$

Proof. (1)

$$\begin{aligned}
p(y) &\triangleq Pr(V_i < K|Y = y) \\
&= Pr(\sum_{i=1}^{M} \beta_i y_i + Z_i < K|Y = y) \\
&= Pr(Z_i < K - \sum_{i=1}^{M} \beta_i y_i|Y = y) \\
&= \Phi(\frac{K - \sum_{i=1}^{M} \beta_i y_i}{w_i}
\end{aligned}$$

(2) Note that when conditioning on the Y, the default behavior among firms are independent. Therefore the number of defaulting firms can be characterized by binomial distribution. □

23.7.3.2 *Asset classes model*

Definition 23.7.1. *[6, p. 313]Assume that the portfolio consists of two classes of obligors C_1 and C_2. There are N_1 obligors of class C_1 and N_2 obligors of class C_2. We assume the asset value follows*

$$\begin{aligned}
V_{n_1} &= \frac{1}{\sqrt{\beta_{11}^2 + \beta_{12}^2 + 1}}(Y_1\beta_{11} + Y_2\beta_{12} + \epsilon_{n_1}), \forall n_1 \in C_1 \\
V_{n_2} &= \frac{1}{\sqrt{\beta_{21}^2 + \beta_{22}^2 + 1}}(Y_1\beta_{21} + Y_2\beta_{22} + \epsilon_{n_2}), \forall n_2 \in C_2
\end{aligned}$$

where two factors Y_1 and Y_2 and the noises ϵ_{n_1} and ϵ_{n_2} are independently standard normal variables.

Lemma 23.7.10. *[6, p. 313] Conisder a two-asset-classes model. It follows that*

- *The obligors' assets within one class are correlated with a correlation coefficient of ρ_1 and ρ_2 respectively, where*

$$\rho_1 = \frac{\beta_{11}^2 + \beta_{12}^2}{1 + \beta_{11}^2 + \beta_{12}^2}$$
$$\rho_2 = \frac{\beta_{21}^2 + \beta_{22}^2}{1 + \beta_{21}^2 + \beta_{22}^2}$$

- *The correlation of two obligors of different classes is*

$$\rho_{12} = \frac{\beta_{11}\beta_{21} + \beta_{12}\beta_{22}}{\sqrt{1 + \beta_{11}^2 + \beta_{12}^2}\sqrt{1 + \beta_{21}^2 + \beta_{22}^2}}$$

Proof. (1) Note that $Var[V_{n_1}] = Var[V_{n_2}] = 1$. Therefore

$$\begin{aligned}
\rho_1 &= E[\frac{1}{\sqrt{\beta_{11}^2 + \beta_{12}^2 + 1}}(Y_1\beta_{11} + Y_2\beta_{12} + \epsilon_{n_1})\frac{1}{\sqrt{\beta_{11}^2 + \beta_{12}^2 + 1}}(Y_1\beta_{11} + Y_2\beta_{12} + \epsilon_{n_1'})] \\
&= \frac{1}{1 + \beta_{11}^2 + \beta_{12}^2}(\beta_{11}^2 Var[Y_1] + \beta_{12}^2 Var[Y_2]) \\
&= \frac{\beta_{11}^2 + \beta_{12}^2}{1 + \beta_{11}^2 + \beta_{12}^2}.
\end{aligned}$$

Similarly,we can derive ρ_2. (2) Note that

$$\begin{aligned}
\rho_{12} &= E[\frac{1}{\sqrt{\beta_{11}^2 + \beta_{12}^2 + 1}}(Y_1\beta_{11} + Y_2\beta_{12} + \epsilon_{n_1})\frac{1}{\sqrt{\beta_{21}^2 + \beta_{22}^2 + 1}}(Y_1\beta_{21} + Y_2\beta_{22} + \epsilon_{n_2})] \\
&= \frac{1}{\sqrt{\beta_{11}^2 + \beta_{12}^2 + 1}}\frac{1}{\sqrt{\beta_{21}^2 + \beta_{22}^2 + 1}}(\beta_{11}\beta_{21} Var[Y_1] + \beta_{12}\beta_{22} Var[Y_2]) \\
&= \frac{\beta_{11}\beta_{21} + \beta_{12}\beta_{22}}{\sqrt{1 + \beta_{11}^2 + \beta_{12}^2}\sqrt{1 + \beta_{21}^2 + \beta_{22}^2}}.
\end{aligned}$$

□

23.7.4 Application examples

23.7.4.1 *CDO*

Definition 23.7.2. *[5, p. 583] The originator of a synthetic CDO chooses a portfolio of companies and a maturity (e.g., 5 years) for the structure. It sells CDS protection on each company in the portfolio with the CDS maturities equaling the maturity of the structure. The synthetic CDO principal is the total of the notional principals underlying the CDSs. The originator has cash inflows equal to the the CDS spreads and cash outflows when companies in the portfolio default. Tranches are formed and the cash inflows and outflows are distributed to tranches. The rules for determining the cash inflows and outflows of tranches are more straightforward for a synthetic CDO than for a cash CDO.*

*Suppose that there are only three tranches:**equity, mezzanine, and senior**.*

- *The equity tranche is responsible for the payouts on the CDSs until they reach 5of the synthetic CDO principal. It earns a spread of 1,000 basis points per year on the outstanding tranche principal.*
- *The mezzanine tranche is responsible for payouts in excess of 5% up to a maximum of 20% of the synthetic CDO principal. It earns a spread of 100 basis points per year on the outstanding tranche principal.*
- *The senior tranche is responsible for payouts in excess of 20%. It earns a spread of 10 basis points per year on the outstanding tranche principal.*

Example 23.7.2. [5, p. 583] To understand how the synthetic CDO would work, suppose that its principal is \$100 million. The equity, mezzanine, and senior tranche principals are \$5 million, \$15 million, and \$80 million, respectively. The tranches initially earn the specified spreads on these notional principals. Suppose that after 1 year defaults by companies in the portfolio lead to payouts of \$2 million on the CDSs. The equity tranche holders are responsible for these payouts. The equity tranche principal reduces to \$3 million and its spread (1,000 basis points) is then earned on \$3 million instead of \$5 million. If, later during the life of the CDO, there are further payouts of \$4 million on the CDSs, the cumulative of the payments required by the equity tranche is \$5 million, so that its outstanding principal becomes zero. The mezzanine tranche holders have to pay \$1 million. This reduces their outstanding principal to \$14 million.

Remark 23.7.6 (how correlation affects cash flow). [5, p. 585] The valuation of a tranche of a synthetic CDO is similarly dependent on default correlation. If the correlation is low, the junior equity tranche is very risky and the senior tranches are very safe. As the default correlation increases, the junior tranches become less risky and the senior tranches become more risky. In the limit where the default correlation is perfect and the recovery rate is zero, the tranches are equally risky.

23.7.4.2 *Basket default swap*

[11, p. 171]

23.7.4.3 *CDX*

[11, p. 172]

Definition 23.7.3.

- *Credit default swap index (CDX)*

Remark 23.7.7 (commonn CDS indexes). link The major tradable indexes include: CDX, ABX, CMBX, and LCDX. The CDX indexes are broken out between investment grade (IG), high yield (HY), high volatility (HVOL), crossover (XO) and emerging market (EM). For example, the CDX.NA.HY is an index based on a basket of North American (NA) single-name high-yield CDSs. The crossover index includes names that are split rated, meaning they are rated "investment grade" by one agency and "below investment grade" by another.

The ABX and CMBX are baskets of CDSs on two securitized products: asset-backed securities (ABS) and commercial mortgage-backed securities (CMBS). The ABX is based on ABS home equity loans, and the CMBX âĂŞ on CMBS. So, the ticker ABX.HE.AA, for example, denotes an index that is based on a basket of 20 ABS Home Equity (HE) CDSs whose reference obligations are 'AA'-rated bonds. There are five separate ABX indexes for ratings ranging from "AAA" to "BBB-." The CMBX also has the same breakdown of five indexes by ratings, but is based on a basket of 25 CDSs, which reference CMBS securities. (For related insight, see "Behind the Scenes of Your Mortgage.")

The LCDX is a credit-derivative index with a basket made up of single-name, loan-only CDSs. The loans referred to are leveraged loans. The basket is made up of 100 names. Although a bank loan is considered secured debt, the names that usually trade in the leveraged loan market are lower-quality credits. (If they could issue in the normal IG markets, they would.) Therefore, the LCDX index is used mostly by those looking for exposure to high-yield debt.

All of the above indexes are issued by the CDS Index Company and administered by Markit. For these indexes to work, they must have sufficient liquidity. Therefore, the issuer has commitments from the largest dealers (large investment banks) to provide liquidity in the market.

23.8 Credit Rating models

23.8.1 The model

Definition 23.8.1 (setup). *[6, p. 228]*

- *There are K rating classes. The set of rating classes is denoted by $S = \{1, 2, ..., K\}$. S is called **state space**.*
- *The rating process of a firm is a stochastic process $\{R(t)\}_{t\in\mathbb{R}_+}$ such that $R(t)$ is a random variable taking value in S.*
- *A transition matrix $Q(T_1, T_2) \in \mathbb{R}^{K\times K}$ such that*

$$Q_{ij}(T_1, T_2) = P(R(T_2) = j | R(T_1) = i), \forall i, j \in S.$$

- *Time homogeneity: $Q(T_2, T_1) = Q(T_2 - T_1)$.*
- *Generator.*

23.8.2 Pricing applications

23.8.2.1 *Defaultable zero coupon bond*

(notations)

- The price of a zero-coupon bond is denoted by $P^d(t, T, R(t))$.
- The payoff at maturity is given by

$$P^d(T, T, R(T) = k_T) = Roc(k_T),$$

where $Roc(k_T) \in [0, 1]$ is the recovery rate for a bond with rating k_T.

Lemma 23.8.1. *[6, p. 239] Consider a bond issued by a company with initial credit rating of $R(0) = k_0$. The bond matures at T. The time 0 value of the bond is*

$$P^d(0, T, k_0) = E_Q[\exp(-\int_0^T r(s)ds)P^d(T, T, R(T))|R(0) = k_0].$$

Assuming the credit migrating process is independent of risk-free rate dynamics, we have

$$P^d(0, T, k_0) = P(0, T)\sum_{k=1}^{K} Q_{k_0,k}(0, T)P^d(T, T, R(T) = k).$$

where $Q_{k_0,k}(0,T)$ denotes the risk-neutral transition probability

$$Q_{k_0,k}(0,T) = P(R(T) = k|R(0) = k_0).$$

Proof. Note that $E_Q[\exp(-\int_0^T r(s)ds)P^d(T,T,R(T))|R(0) = k_0] = E_Q[\exp(-\int_0^T r(s)ds)]E_Q[P^d(T,T,R(T))|R(0) = k_0]$. □

23.8.2.2 *Credit rating and interest rate derivatives*

Lemma 23.8.2 (governing equation for credit rating and interest rate derivative). *[6, p. 241] Assuming under risk neutral measure Q, the short rate dynamics is given by*

$$dr = \mu_r dt + \sigma_r dW_t,$$

and the continuous time Markov chain model for the credit rating is given by

$$\frac{dQ}{dt} = \Lambda Q,$$

where $Q \in \mathbb{R}^K$ and $Q \in \mathbb{R}^{K\times K}$.

Let $\boldsymbol{F} = (F_1(r,t), F_2(r,t), ..., F_K(r,t)$, where $F_k(r,t)$ is the value of asset at time t when the risk-free rate is $r(t)$ and the credit rating at time t is k.

Then the governing equation for $\boldsymbol{F}$ is given by

$$0 = \frac{\partial}{\partial t}\boldsymbol{F} + \frac{1}{2}\sigma_r^2\frac{\partial^2}{\partial r^2}\boldsymbol{F} + \mu_r\frac{\partial}{\partial r}\boldsymbol{F} + \Lambda\boldsymbol{F} - r\boldsymbol{F},$$

with final condition $\boldsymbol{F}(r,T) = (F_1(r(T)), F_2(r(T)), ..., F_K(r(T)))$.

Proof.

$$\begin{aligned} dF_R(t,r) &= \frac{\partial}{\partial t}F_R dt + \frac{1}{2}\sigma_r^2\frac{\partial^2}{\partial r^2}F_R dt + \Delta_{in} - \Delta_{out} \\ &= \frac{\partial}{\partial t}F_R dt + \frac{1}{2}\sigma_r^2\frac{\partial^2}{\partial r^2}F_R dt + \mu_r\frac{\partial}{\partial r}F_R dt + \sigma_r\frac{\partial}{\partial r}F_R dW_t + \Delta_{in} - \Delta_{out} \\ \implies E[dF_R(t,r)] &= \frac{\partial}{\partial t}F_R dt + \frac{1}{2}\sigma_r^2\frac{\partial^2}{\partial r^2}F_R dt + \mu_r\frac{\partial}{\partial r}F_R dt + \sum_k \Lambda_{R,k} dt F_k(r,t) = rF_R dt \end{aligned}$$

where $\Delta_{in} - \Delta_{out}$ is the random quantity of net flowing into the credit R class within time interval dt, and $E_Q[\Delta_{in} - \Delta_{out}] = \sum_k \Lambda_{R,k} dt F_k(r,t)$. We also use the fact that in risk-neutral measure, value of assets will grow at short rate r. □

Remark 23.8.1 (evaluation via Monte Carol simulation). Under risk-neutral measure, the price at time 0 is given by

$$V(0, r(0), R(0)) = E_Q[\exp(-\int_0^T r(s)ds)F(T, r(T), R(T)|r(0), R(0)],$$

by generating trajectories of $r(t)$ and $R(t)$(using discrete-time approximation to continuous-time Markov chain) with initial condition $r(0)$ and $R(0)$ and then taking average.

23.9 Stochastic hazard rate models

23.9.1 Characterizing default with stochastic hazard rate

23.9.1.1 *The reduced form model*

Remark 23.9.1 (general remarks on term structure). Here by term structure, we mean the probability characterization of default will change due to the upcoming information.

Definition 23.9.1 (conditional survival probability, forward hazard rate).

- *Denote $S(t,T)$ as the* ***conditional survival probability*** *from t to T conditioned on $\mathcal{F}_t$, the information up to time t. It is given by*

$$S(t,T) = Pr(\tau > T | \tau > t, \mathcal{F}_t).$$

- *Denote $f(t,s)$ as the* ***forward hazard rate*** *from s to $s+ds$ conditioned on $\mathcal{F}_t, t < s$. It is given by*

$$f(t,T) = \lim_{\Delta T \to 0} \frac{1}{\Delta T} Pr(T < \tau < T + \Delta T | \tau > t, \tau > T, \mathcal{F}_t).$$

Note 23.9.1 (understand the stochastic nature of conditional survival probability). Let the current time be 0.

- $S(t,T)$ and $f(t,T)$ are both random quantity because they are conditioned on future information set $\mathcal{F}_t$.
- We can represent the previous stochastic hazard rate $h(t)$ as $f(t,t)$ in our more generalized framework.
- If $h(t)$ is deterministic, then

$$h(t) = f(s,t), \forall s.$$

Lemma 23.9.1 (default probability in the stochastic intensity framework). *Let τ be the default time. Let $h(t)$ be the stochastic hazard rate. Let current time be 0. It follows that*

- *Default time cumulative distribution function*

$$F_\tau(t) = Pr(\tau < t) = 1 - E[\exp(-\int_t^T h(s)ds)].$$

- *Survival probability*

$$S(t,T) = Pr(\tau > T|\tau > t, \mathcal{F}_t) \neq \frac{S(T)}{S(t)}, S(t,t) = 1.$$

- $$S(t,T) = E[\exp(-\int_t^T h(s)ds)] = \exp(-\int_t^T f(t,s)ds)$$

- $$f(t,T) = -\frac{\partial \ln S(t,T)}{\partial T}.$$

- *Default time density is given by*

$$f_\tau(t) = E[h(t)\exp(-\int_0^t h(s)ds)]$$

Proof. similar to the above lemma.

Conditioning on the hazard rate trajectory from 0 to T, we have

$$Pr(\tau > T|h(s), 0 \leq s \leq T) = \exp(-\int_0^T h(s)ds).$$

Therefore

$$Pr(\tau > T) = \int_{h\in H} Pr(\lambda > T|h(s), 0 \leq s \leq T)f(h)dh = E[\exp(-\int_0^T h(s)ds)].$$

□

Remark 23.9.2. Note that

$$S(t,T) \neq \frac{S(T)}{S(t)}.$$

This is because

$$S(t,T) = E[\exp(-\int_t^T h(t)ds], S(t) = E[\exp(-\int_0^t h(s)ds)], S(T) = E[\exp(-\int_0^T h(s)ds)].$$

Remark 23.9.3.

- $S(t,T) \neq \frac{S(T)}{S(t)}$ is because the information $\mathcal{F}_t$ will generally 'change' the distribution of τ.
- If h is deterministic, then $h(t,s) = h(s)$ is because of the information $\mathcal{F}_t$ does not give more information on h.
- If h is stochastic, then $h(t,s) \neq h(s)$ is because of the information $\mathcal{F}_t$ does not give more information on h.

Definition 23.9.2 (forward conditional probability). *[6, p. 56] Denote the default time by τ. Let current time t.*

- *The **conditional survival probability** between from T_1 to T_2 as seen from t(i.e. based on information accumulated up to t)*

$$S_c(t, T_1, T_2) = Pr(\tau > T_2 | \tau > T_1, \mathcal{F}_t) = \frac{S(t, T_2)}{S(t, T_1)}.$$

- *The **conditional probability of default** between from T_1 to T_2 as seen from t is*

$$\begin{aligned} & Pr(T_1 < \tau < T_2 | \tau > T_1, \mathcal{F}_t) \\ =& S(t, T_1) - S(t, T_2) \\ =& \exp(-\int_{T_1}^{T_2} f(t, s) ds) S(t, T_1) \end{aligned}$$

.

- *The **discrete hazard rate of default** over $[T, T + \Delta t]$ as seen from t is*

$$H(t, T, T + \Delta t) = \frac{1}{\Delta t}(\frac{S(t, T)}{S(t, T + \Delta t)} - 1) = \frac{1}{\Delta t}(\frac{1}{S_c(t, T, T + \Delta t)} - 1)$$

- *The **forward hazard rate of default** at time T is*

$$h(t, T) = \lim_{\Delta t \to 0} H(t, T, T + \Delta t) = -\frac{\partial \ln S(t, T)}{\partial T}$$

  Note that $S(t, T) = \exp(-\int_t^T h(t, s) ds)$.

- $$H(t, T, T + \Delta t) = \frac{1}{\Delta t}(\exp(\int_T^{T + \Delta t} h(t, s) ds) - 1).$$

Remark 23.9.4 (the term structure of default)**.**

- To incorporate the effect of information affects hazard rate, we use $h(t, s)$ to parameterize the hazard rate.
- resembles the instantaneous forward rate.
- $S(t, T_1, T_2) = \exp(-\int_{T_1}^{T_2} h(t, s) ds)$ can have different values at different t. This describes how information affects the briefs on the survival probability in the future time interval.

Note 23.9.2 (analog between survival probability and zero-coupon bonds)**.**

- $$S(T,T) = 1, P(T,T) = 1.$$
- $$S(t,T) = \exp(-\int_t^T h(t,s)ds), P(t,T) = \exp(-\int_t^T f(t,s)ds).$$
- $$S_c(t,T_1,T_2) = \frac{S(t,T_2)}{S(t,T_1)}, FP(t,T_1,T_2) = \frac{P(t,T_2)}{P(t,T_1)}$$
- $$S_c(t,T_1,T_2) = \frac{1}{1+(T_2-T_1)H(t,T_1,T_2)}, FP(t,T_1,T_2) = \frac{1}{1+(T_2-T_1)F(t;T_1,T_2)}.$$
- $$S_c(t,T_1,T_2) = \exp(-\int_{T_1}^{T_2} h(t,s)ds), FP(t,T_1,T_2) = \exp(-\int_{T_1}^{T_2} f(t,s)ds).$$
- $$S(t,T_1) = S_c(t,t,T_1), P(t,T_1) = FP(t,t,T_1).$$
- $S(T_1,T_2)$ and $P(T_1,T_2)$ are both random variable.

23.9.1.2 *Motivation for stochastic hazard rate model*

Remark 23.9.5 (motivation).
- In modeling wrong way risk we need a stochastic hazard rate to introduce correlation between asset value and default probability.
- CVA risk modeling requires the calculation of the CVA in future time, which requires in the hazard curve in the future.
- Calculation of CVA on credit product requires the calculation of future values of credit products; the future values of the credit products require the hazard curve in the future.

23.9.2 Model zoo for survival curve dynamics

23.9.2.1 *Non-mean-reverting Gaussian hazard rate model*

Definition 23.9.3 (non-mean-reverting Gaussian hazard rate model). *In the non-mean-reverting Gaussian hazard rate model, we assume that, under risk-neutral measure, the stochastic hazard rate $h(t)$ is governed by*

$$d\lambda(t) = \mu(t)dt + \sigma dW$$

.

Lemma 23.9.2 (survival curve). *Assume the dynamics of the stochastic hazard rate λ under risk-neutral measure is given by*

$$d\lambda(t) = \mu(t)dt + \sigma dW$$

. Let current time be 0. Let $S(t,T)$ denote the conditional survival probability(Definition 23.9.1). Then

- *Stochastic default rate(Definition 23.9.1) is given by*

$$\lambda(s) = \lambda(t) + \int_t^s \mu(s)ds + \int_s^t \sigma dW_s.$$

- *Conditional survival probability is given by*

$$\begin{aligned} S(t,T) &= E[\exp(-\int_t^T \lambda(s))|\mathcal{F}_t] \\ &= \exp((-\lambda(t)(T-t) - \int_t^T \mu(s)(T-s)ds + \frac{1}{6}\sigma^2(T-t)^3). \end{aligned}$$

If $\mu(s) = \mu_0$, then

$$S(t,T) = \exp((-\lambda(t)(T-t) - \frac{1}{2}\mu_0(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3).$$

Proof. (1) directly integrate the SDE. (2) See Lemma 6.3.11. □

Lemma 23.9.3 (evolution of survival curves). *Consider the stochastic instantaneous default rate under risk neutral measure is evolving under the model given by*

$$d\lambda(t) = \mu(t)dt + \sigma dW.$$

- ***(evolution of whole survival curve)*** *Let the current time be 0, then the future survival curve $S(\tau, T+\tau)$ is given by*

$$S(\tau, T+\tau) = \exp((-\lambda(0)T - \int_{\tau}^{T+\tau} \mu(s)(T+\tau-s)ds + \frac{1}{6}\sigma^2 T^3)$$

where $S(0,T)$ is the current yield curve.

- ***evolution of survival probability associated with a survival date** T. Let the current time be 0ïijŇ then the future survival probability $S(\tau, T)$ at time T conditioned on the information of t is given by*

$$S(\tau, T) = \exp((-\lambda(\tau)(T-\tau) - \int_{\tau}^{T} \mu(s)(T-s)ds + \frac{1}{6}\sigma^2 (T-\tau)^3)$$

Proof. □

Methodology 23.9.1 (simulate the future survival curve in non-mean-reverting Gaussian model). *Suppose*

- *current time is 0.*
- *we are given the current survival probability $S(0,T), T \in \mathbb{R}^+$.*
- *we are given the hazard rate model*

$$d\lambda(t) = \mu(t)dt + \sigma dW$$

.

- *we are given the initial hazard rate $\lambda(0)$.*

*Then we can generate a **sample survival probability surve** in future time $t > 0$ via the following procedure:*

- *simulate $\lambda(t)ds$ by drawing from the normal distribution*

$$\lambda(t)ds \sim N(\lambda(0) + \int_0^t \mu(s)ds, \sigma^2 t).$$

- *the sample survival curve is given by*

$$S(t,T) = \exp((-\lambda(t)(T-t) - \int_t^T \mu(s)(T-s)ds + \frac{1}{6}\sigma^2(T-t)^3).$$

Proof. We use the fact(Lemma 6.3.11) that

$$\int_0^t \lambda(s)ds = \lambda(0) + \int_0^t \mu(s)(T-s)ds + \sigma \int_0^t (t-s)dW_s$$

has normal distribution of

$$N(\lambda(0) + \int_0^t \mu(s)(T-s)ds, \sigma^2 \int_0^t (t-s)^2 ds).$$

□

23.9.2.2 *Extended CIR hazard rate model*

Definition 23.9.4 (extended CIR model). *The extended CIR model under risk-neutral model is given by*

$$\begin{aligned} dx(t) &= -kx(t)dt + \sigma\sqrt{x(t)}dW_t, x(0) = 0. \\ h(t) &= x(t) + \alpha(t). \end{aligned}$$

Lemma 23.9.4. *Consider the stochastic hazard rate $h(t)$ under risk neutral measure is evolving under the following model given by*

$$\begin{aligned} dx(t) &= -kx(t)dt + \sigma dW_t, x(0) = 0. \\ h(t) &= x(t) + \alpha(t) \end{aligned}$$

Let the current time be 0. It follows that

- *If the current market term structure is given by $f^M(0,t) = -\frac{\partial \ln S^M(0,T)}{T}$, then choose*

$$\alpha(t) = f^M(0,t) + \frac{\sigma^2}{2k^2}(1-\exp(-kt))^2,$$

  can match the market term structure.
- *With the choice of ϕ in (1), we have*

$$S(t,T) = \frac{S^M(0,T)A(0,t)\exp(-B(0,t)x_0)}{S^M(0,t)A(0,T)\exp(-B(0,T)x_0)}A(t,T)\exp(-B(t,T)x_t),$$

  where $S^M(0,T)$ is the current market survival probability term structure, and

$$B(t,T) = \frac{1-\exp(-k(T-t))}{\alpha},$$

$$A(t,T) = \exp((-\frac{\sigma^2}{2k^2})(B(t,T)-(T-t)) - \frac{\sigma^2}{4k}B^2(t,T))),$$

23.9.3 Non-mean-reverting Gaussian short rate and hazard rate model

23.9.3.1 *The model*

Definition 23.9.5 (Gaussian model). *[7, p. 49] Assume the dynamics of short rate $r(t)$ and hazard rate λ under risk-neutral measure are given by*

$$\begin{aligned} dr &= \theta(t)dt + \sigma_r dW_r \\ d\lambda &= \mu(t)dt + \sigma_\lambda dW_\lambda \end{aligned}$$

where $E[dW_r dW_\lambda] = \rho dt$.

Lemma 23.9.5 (mean and variance of integral of sum of Gaussian process, recap). *Consider the dynamics of short rate $r(t)$ and hazard rate λ are given by*

$$\begin{aligned} dr &= \theta(t)dt + \sigma_r dW_r \\ d\lambda &= \mu(t) + \sigma_\lambda dW_\lambda, E[dW_r dW_\lambda] = \rho dt \end{aligned}$$

Further define

$$I(t,T) = \int_t^T r(s) + \lambda(s) ds.$$

- *$I(t,T)$ is a Gaussian random variable.*
- *It has mean and variance given by*

$$M(t,T) = (r(t) + \lambda(t))(T-t) + \int_t^T (T-s)(\theta(s) + \mu(s))ds,$$

$$V(t,T) = (\sigma_r^2 + 2\rho\sigma_r\sigma_\lambda + \sigma_\lambda^2)\frac{(T-t)^3}{3}.$$

Proof. See Lemma 6.3.12. □

Lemma 23.9.6 (price of zero-recovery zero-coupon bond). *[7, p. 49] The price of a zero-recovery zero-coupon bond is given by(Lemma 23.3.2)*

$$P_0^d(t,T) = E_Q[\exp(-\int_t^T (r(s) + \lambda(s))ds)].$$

Assume the r and λ follow the Gaussian model, then

$$P_0^d(t,T) = \exp(-(r(t)+\lambda(t))(T-t) - \int_t^T (T-s)(\theta(s)+\mu(s))ds + \sigma\frac{(T-t)^3}{6}),$$

where $\sigma^2 = \sigma_r^2 + 2\rho\sigma_r\sigma_\lambda + \sigma_\lambda^2$.

Proof. Let $m(t) = \mu(s) + \lambda(s)$, then

$$dm(t) = (\theta(t) + \mu(t))dt + \sigma dW.$$

To evaluate

$$E_Q[\exp(-\int_t^T m(s)ds)|\mathcal{F}_t],$$

we can use the result in Lemma 19.5.23. □

23.9.3.2 *Price dynamics*

23.9.4 Mean-reversion Gaussian short rate and hazard rate model

Definition 23.9.6 (Gaussian model). *[7, p. 49][6, p. 167] Assume the dynamics of short rate $r(t)$ and hazard rate $\lambda(t)$ under risk-neutral measure are given by*

$$\begin{aligned} dr &= k_1(\theta(t) - r)dt + \sigma_r dW_r \\ d\lambda &= k_2(\mu(t) - \lambda)dt + \sigma_\lambda dW_\lambda \end{aligned}$$

where $E[dW_r dW_\lambda] = \rho dt$.

23.10 No-arbitrage condition for collaterallized assets

Definition 23.10.1 (fully collaterallized contract). *[12] Let A and B be two party entering into a contract. Let $V(t)$ be the value of the contract to A. We say the contract is fully collaterallized if the following conditions are satisfied:*

- *If $V(t) > 0$ for A, then party B will post collateral worth $V(t)$ to A.*
- *Let $c(t)$ be the contractually specified collateral rate $c(t)$ on $V(t)$. When B post a collateral worth $V(t)$, A will pay B the cost $V(t)c(t)$.*

Note 23.10.1 (cash flow in the fully collaterallized contract)**.** [12] Assume A buys some collaterallized asset from B.

- At the beginning time t. A pays amount $V(t)$ to B, and B posts a collateral worth $V(t)$ to A.
- At time $t + dt$, A pays the collateral rate $c(t)V(t)dt$ to B, B adjusts the collateral value to $V(t + dt)$.
- The net cash flow for A at $t + dt$ is

$$V(t + dt) - V(t) - c(t)V(t)dt = dV(t) - c(t)V(t)dt.$$

- At any future time, the termination of the contract will not cost extra loss to either parties.

Lemma 23.10.1 (no-arbitrage condition for risk-free collaterallized asset dynamics). *Consider an fully collaterallized asset. If this asset is risk-free, then no-arbitrage condition implies*

- *the collateral rate $c(t)$ equals risk-free rate.*
- *the net value growth rate of this asset is 0.*

Proof. Suppose A and B enter a contract with value $V(t) > 0$ with respect to A. Assume $V(t)$ is risk-free, therefore its value will grow at a rate of r. B should post a collateral of value $V(t)$ to A, and A should pay B a collateral rate c. If $c < r$, then B will terminate the contract since B is not willing to be the counterparty of A to lose money. Similar analysis applies to $c > r$. □

Remark 23.10.1 (compared with non-collaterallized risk-free asset dynamics)**.**

- If not considering the collateral fee, holding a risk-free asset will have the same growth rate of r, which is the same as the no-collateral world(Corollary 15.2.1.2).
- The payment of the collateral fee makes the net value growth zero.

Lemma 23.10.2 (no-arbitrage condition for fully collaterallized asset dynamics, single source of uncertainty). *Consider two assets V_1 and V_2 are both collaterallized with rate $c(t)$. Assume in the real-world measure, the asset prices follow*

$$\begin{aligned} dV_1/V_1 &= \mu_1 dt + \sigma_1 dW(t) \\ dV_2/V_2 &= \mu_1 dt + \sigma_1 dW(t). \end{aligned}$$

If there is no arbitrage in the market model, then

- *there exists a function $\lambda(t)$ such that*

$$\frac{\mu_1(t) - c(t)}{\sigma_1(t)} = \frac{\mu_2(t) - c(t)}{\sigma_2(t)} = \lambda(t).$$

- *there exists a measure Q, equivalent to the real-world measure P, such that under which*

$$\begin{aligned} dV_1/V_1 &= c(t)dt + \sigma_1 dW^Q(t) \\ dV_2/V_2 &= c(t)dt + \sigma_1 dW^Q(t). \end{aligned}$$

where $W^Q(t)$ is a Brownian motion under Q.

Proof. (1)Consider a portfolio of $\sigma_2(t)V_2(t)$ of asset 1 and $-\sigma_1(t)V_1$ of asset 2. The change of value of this portfolio is

$$\begin{aligned} d\pi(t) &= \sigma_2(t)V_2(t)(dV_1(t) - c(t)V_1(t)dt) - \sigma_1 V_1(t)(dV_2(t) - c(t)V_2 dt) \\ &= \sigma_2(t)V_1(t)V_2(t)(\mu_1(t) - c(t))dt - \sigma_1(t)V_1(t)V_2(t)(\mu_2(t) - c(t))dt \\ &= 0 \\ \implies & \sigma_2(t)(\mu_1(t) - c(t)) = \sigma_1(t)(\mu_2(t) - c(t)) \end{aligned}$$

note that we use the cash flow analysis for collaterallized assets(23.10.1) (2) Use Girsanov theorem. There exists a measure Q such that $dW = dW^Q - \lambda dt$. □

23.11 Binomial lattice method

Definition 23.11.1 (defaultable bonds). *A defaultable bond is characterized by coupon rate c, face value F, and recovery value R(random fraction of face value recovered on default).*

Definition 23.11.2. *A binomial lattice model for short rate consists of*

- *nodes (i,j): representing date $i = 0,1,...,n$ and states $j = 0,1,...,n$*
- *values on node (i,j) represent short rate $r_{i,j}$*
- *1-step default probability $h(t)$ for bond defaults in $[t, t+1)$*
- *state transition probability from non-default state:*

$$P((i+1,s,\eta)|(i,j,0)) = \begin{cases} p_u h_{i,j}, s = j+1, \eta = 0 \\ p_u(1-h_{i,j}), s = j+1, \eta = 0 \\ p_d h_{i,j}, s = j, \eta = \\ p_d(1-h_{i,j}), s = j, \eta = \\ 0, otherwise \end{cases}$$

  here $h_{i,j}$ reflects the state dependent default probability.
- *state transition probability from default state:*

$$P((i+1,s,1)|(i,j,1)) = \begin{cases} p_u, s = j+1 \\ p_d, s = j \\ 0, otherwise \end{cases}$$

23.12 Notes on bibliography

Major references are [7][6][4].

For default correlation modeling, see [13].

For CDS valuation, see [9][14].

For estimation of default term structure, see [15][16][17][18]

CDS spread data, Markit

BIBLIOGRAPHY

1. Pereira, J. P. *Credit Risk* (Universidade Nova de Lisboa, 2016).
2. Brigo, D., Morini, M. & Pallavicini, A. *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes* ISBN: 9780470661789. https://books.google.com/books?id=Y-X9zqVcOFkC (Wiley, 2013).
3. Ammann, M. *Credit risk valuation: methods, models, and applications* (Springer Science & Business Media, 2013).
4. Wilmott, P. *Paul Wilmott on quantitative finance* (John Wiley & Sons, 2013).
5. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
6. Schönbucher, P. J. *Credit derivatives pricing models: models, pricing and implementation* (John Wiley & Sons, 2003).
7. O'Kane, D. *Modelling single-name and multi-name credit derivatives* (John Wiley & Sons, 2011).
8. Bluhm, C., Overbeck, L. & Wagner, C. *Introduction to credit risk modeling* (Crc Press, 2016).
9. White, R. The pricing and risk management of credit default swaps, with a focus on the isda model. *OpenGamma Quantitative Research* **16** (2013).
10. Deng, G., Mallett, J. & McCann, C. Modeling autocallable structured products. *Journal of Derivatives & Hedge Funds* **17,** 326–340 (2011).
11. Roncalli, T. Lecture Notes on Risk Management & Financial Regulation (2016).
12. Piterbarg, V. *Collateral, Funding, and Discounting* (Barclays).
13. Li, D. X. On default correlation: A copula function approach (1999).
14. O'Kane, D. & Turnbull, S. Valuation of credit default swaps. *Lehman Brothers quantitative credit research quarterly* **2003** (2003).
15. Berd, A., Mashal, R. & Wang, P. Estimating implied default probabilities from credit bond prices. *Fixed income-quantitative credit research, Lehman Brothers* **7** (2003).
16. Berd, A. M., Mashal, R. & Wang, P. Defining, estimating and using credit term structures. Part 1: Consistent Valuation Measures (2009).
17. Berd, A. M., Mashal, R. & Wang, P. Defining, estimating and using credit term structures. Part 2: Consistent Risk Measures (2004).

18. Berd, A. M., Mashal, R. & Wang, P. Defining, estimating and using credit term structures. Part 3: Consistent CDS-Bond Basis (2004).

24

COMMODITY MODELING

24.1 Commodity market basics

24.1.1 Overview

Definition 24.1.1 (common commodities). *Common commodities include:*

- *Agricultural products - corn, soybean, wheat*
- *Energy products - WTI crude oil, Brent crude oil, natural gas*
- *Precious metals - gold, silver, platinum*
- *Industrial metals - copper, aluminum, tin*
- *Soft commodities - coffee, cocoa, sugar*

Note 24.1.1 (general comments)**.**

- Spot market generally does not exist since commodities need to be transported and stored.
- Major markets are futures, including electricity, crude oil, heating oil, natural gas.
- Contracts may be for physical delivery or financially settlement.

24.1.2 Forward curve basics

24.1.2.1 *Forward curve*

Definition 24.1.2 (futures contract and forward curve for commodities). *[1, p. 71]*

- *Futures contracts on commodities have delivery dates or maturiteis ranging from 1W to 3Y.*
- *The future contract with the nearest delivery date is called prompt forward contract.*
- *Let current time be t. Denote the forward/futures price associated with maturity T contract as $F(t,T)$. The forward curve prevailing at date t for a given commodity is a graphical representation of the set $\{F(t,T), T \geq t\}$ of forward prices for different maturities T.*

Definition 24.1.3 (contango, backwardation). *[1, p. 73]*

- *Most commodities forward curve have volatile short-ends, and quasi-stable long-ends. One important explanation is that the time period prior to delivery allows for the production process to adjust for shocks in supply and demand.*
- *Long-end is determined by marginal cost of production.*

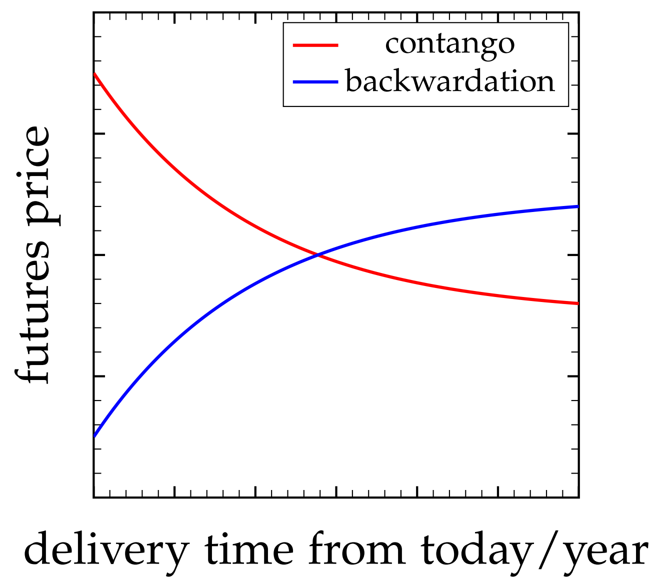


Figure 24.1.1: contango and backwardation forward curves

- *Short-end governed by short termed supply and demand:*
  - *When there are excesses supply of commodity, curve is upward sloping(**contango**). From the spot-forward relation(Lemma 24.1.1) $F(t,T) = S(t)\exp((r-y)(T-t))$, when $r-y \geq 0$, we can observe contango.*
  - *When there are shortages of commodity, curve is downward sloping(**backwardation**).From the spot-forward relation(Lemma 24.1.1) $F(t,T) = S(t)\exp((r-y)(T-t))$, when $r-y < 0$, that is we have large convenience yield due to the shortage of supply, we can observe backwardation.*

See Figure 24.1.1 for contango and backwardation.

Definition 24.1.4 (commodity swap).

- *Swaps entail exchanging a fixed payment stream for a floating payment stream.*
- *Floating stream is typically linked to a commodity spot price or price index*
- *The floating leg can be viewed as a series of forward contract.*
- *The fixed leg can be viewed as a series of coupon payments*
- *Fixed leg payments determined such that both legs have equal value.*

Definition 24.1.5 (commodity options).

- *European Call options on commodity forward contracts*
- *Asian Call options on commodity forward contracts*

- *Floating strike call options are call options with the strike level set at a future date based on some price index. Both the commodity forward price and the index price are uncertain.*
- *A calendar spread option is an option to exchange a T_2 maturity forward contract for a T_1 maturity forward contract at a cost K at time T, where $T_1, T_2 > T$. The payoff at time T is*
$$V(T) = (F(T, T_1) - F(T, T_2) - K)^+.$$
- *Intercommodity spread options are options based on the difference in two commodities. The payoff at maturity T is given by*

Definition 24.1.6 (convenience yield). *[1, p. 24]*

- ***Convenience yield** is a benefit accrues to the owner of the physical commodity but not to the holder of a forward contract. Analogously, the dividend yield is paid to the owner of a stock but not the holder of a derivative contract written on the stock.*
- *The convenience yield is expressed as rate y such that the benefit of holding the commodity worth spot $S(t)$ will equal to $S(t)ydt$ over the interval $(t, t + dt)$.*
- *(economical interpretation) The benefit of holding the physical commodity will have a productive value since they allow us to meet the unexpected demand.*
- *The convenience yield is defined as the positive gain attached to the physical commodity minus the cost of storage.*

24.1.2.2 *Relationship between forward and spot*

Lemma 24.1.1 (spot-forward relationship for a storable commodity). *[1, pp. 37, 44]*

- *Under the assumption of constant interest rate, the spot-forward relation is given by*
$$F(t, T) = S(t)\exp((r - y)(T - t))$$
where $y = y_1 - c$, y_1 is the benefit from the holding the physical commodity and c is the storage cost.
- *Under the assumption of stochastic interest rate, the spot-forward relation is given by*
$$F(t, T) = S(t)\exp((-y)(T - t))/P(t, T),$$
where $y = y_1 - c$, y_1 is the benefit from the holding the physical commodity and c is the storage cost, and $P(t, T)$ is the T zero-coupon bond price at time t.

Note 24.1.2 (interpretation). [1, p. 37] Based on the spot-forward relation, we can approximately write

$$F(t,T) \approx S(t)(1 + r(T-t) + c(T-t) - y_1(T-t)).$$

We can interpret the difference between spot and the forward as:

- $S(t)r(T-t)$ is the funding cost of financing the purchase of S.
- $S(t)c(T-t)$ is the storage cost of holding the physical commodity.
- $S(t)y_1(T-t)$ is the benefit of holding the physical commodity.

Note 24.1.3 (how storage fee assets prices). [2, p. 878] Assume zero interest rate. Assume further that today's spot price of gas is $S_0 = 30$ and the storage cost is 1 per year. Then the no-arbitrage price for the natural gas will be $S_1 = 31$ one year later.

The following strategy ensures that the price is free of arbitrage.

$F_T < 31$			$F_T > 31$		
T=0	short S_0; enter futures to buy F_T	+30	T=0	long S_0 and store it, enter futures to sell at F_T	-30
T=1	settle the futures, receive payment for storage fee, and return S_T	+1-F_T	T=1	settle the futures, and pay the storage fee	+F_T - 1
payoff		$31 - F_T$	payoff		$F_T - 31$

24.1.3 Oil market

Remark 24.1.1 (general remarks). [3, p. 157]

- The most commonly encountered oils by far are West Texas Intermediate (WTI) and Brent. West Texas Intermediate trades on both the New York Metals Exchange (NYMEX) and the Intercontinental Exchange (ICE), whereas Brent trades predominantly on the Intercontinental Exchange (ICE).
- WTI is a light sweet North American crude oil; Brent crude is a European crude oil.
- Futures on WTI/Brent for a sequence of calendar monthsTable 24.1.1.
- WTI futures contracts on NYMEX are for physical delivery; WTI contracts also trade on the ICE, but these are cash settled.

- Options on WTI futures are also traded on NYMEX, as showed in Table 24.1.2. These options expire three business days before the expiry for the underlying futures contract.

  [3, p. 162]

Table 24.1.1: Cal 12 WTI-NYMEX strip of futures contract

contract	expiry date	cash date	first notice	delivery date
JAN12	20-Dec-11	21-Dec-11	22-Dec-11	Jan-12
FEB12	20-Jan-12	23-Jan-12	24-Jan-12	Feb-12
MAR12	21-Feb-12	22-Feb-12	23-Feb-12	Mar-12
APR12	20-Mar-12	21-Mar-12	22-Mar-12	Apr-12
...	...	...	...	...

Table 24.1.2: Cal 12 WTI-NYMEX strip of options on futures contract

contracts	option expiry date	futures expiry date	cash date
JAN12	15-Dec-11	20-Dec-11	22-Dec-11
FEB12	17-Jan-12	20-Jan-12	24-Jan-12
MAR12	15-Feb-12	21-Feb-12	23-Feb-12
APR12	15-Mar-12	20-Mar-12	22-Mar-12
...	...	...	...

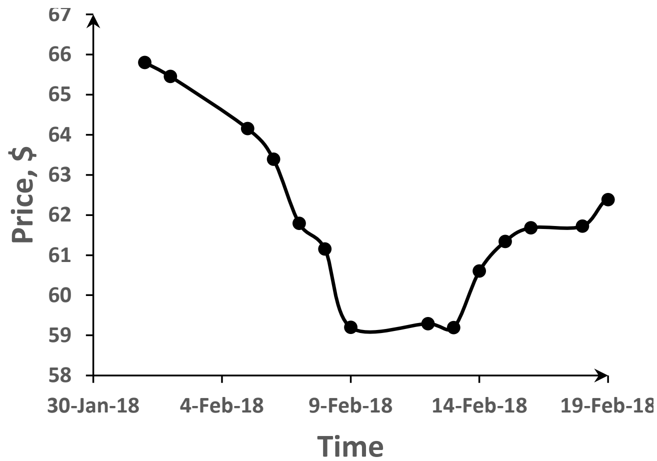


Figure 24.1.2: WTI APR18 contract historical prices starting from Feb,1,2018 to Mar,2,2018

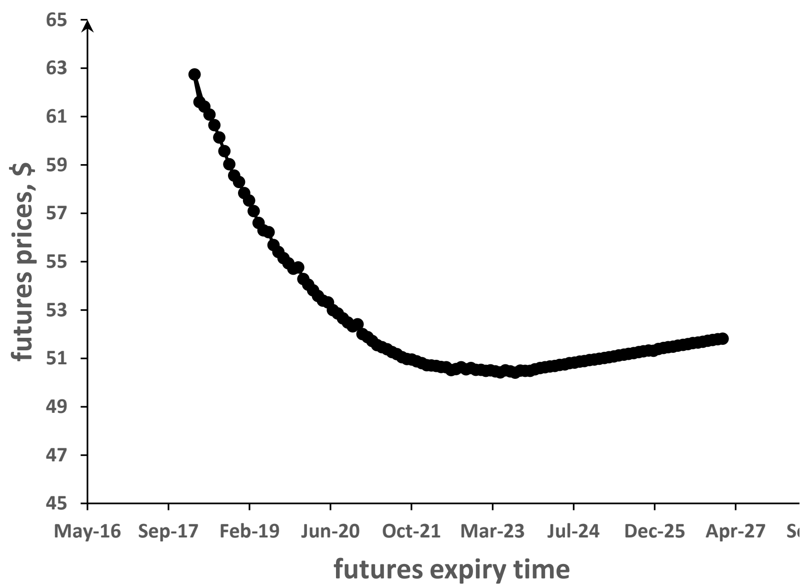


Figure 24.1.3: WTI forward curve observed on Mar,2,2018

24.1.4 Natural gas market

Remark 24.1.2 (general remarks).

- Natural gas contracts are traded in various geographical regions. The most liquid contract is the Henry Hub contract on the NYMEX for delivery of natural gas to the pipeline interconnector at **Henry Hub**.

- Henry Hub futures contracts traded on NYMEX(see Table 24.1.3) are for physical delivery of natural gas containing 10,000 MMBtu of extractable energy, into the Henry Hub pipeline complex operated by Sabine Hub Services at Erath, Louisiana.
- Natural gas futures contract also trades on the ICE, but is cash settled and is for a smaller contract size of 2500 MMBtu.
- Options on Henry Hub futures contracts trade on NYMEX, the expiry date being the business day before the futures contract expiry date.

Table 24.1.3: Cal 12 NG-NYMEX strip of futures contract

contract	expiry date	cash date	delivery date
JAN12	28-Dec-11	29-Dec-11	Jan-12
FEB12	27-Jan-12	30-Jan-12	Feb-12
MAR12	27-Feb-12	28-Feb-12	Mar-12
APR12	28-Mar-12	29-Mar-12	Apr-12
...	...	...	...

Definition 24.1.7 (seasonality).

- *Seasonality occurs in many commodities(crude oil is the main exception)*
- *If the storage capacity exceeds the wavelength, then no humps*
- *As showed in Figure 24.1.4, natural gas has large hump in winter, small hump in summer for the forward curve; however, the prompt futures price does not display significant seasonality(Figure 24.1.5).*
- *Gasoline has large hump in summer*
- *Electricity has humps in winter and summer, negative hump on weekends, and intro-day structure.*

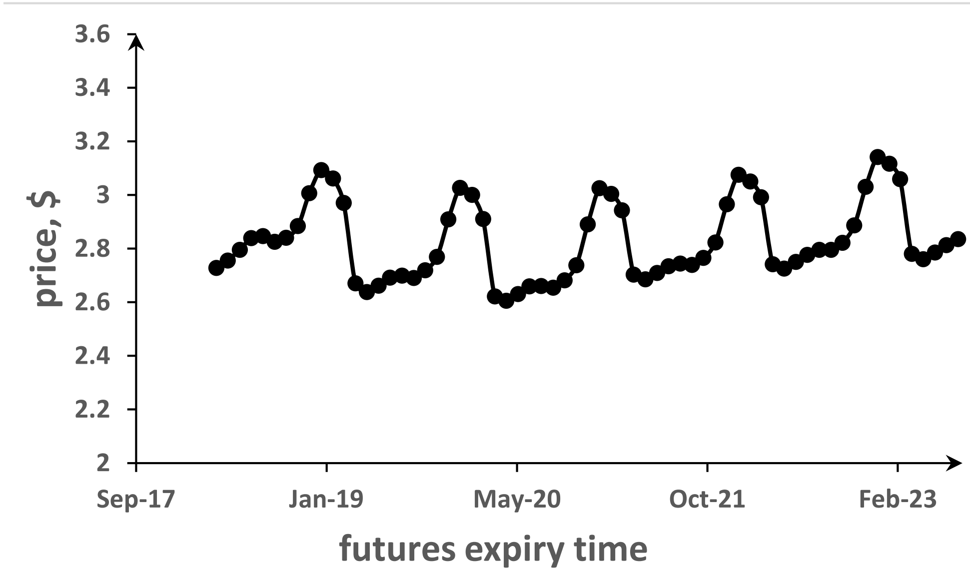


Figure 24.1.4: NG forward curve observed on Mar,2,2018

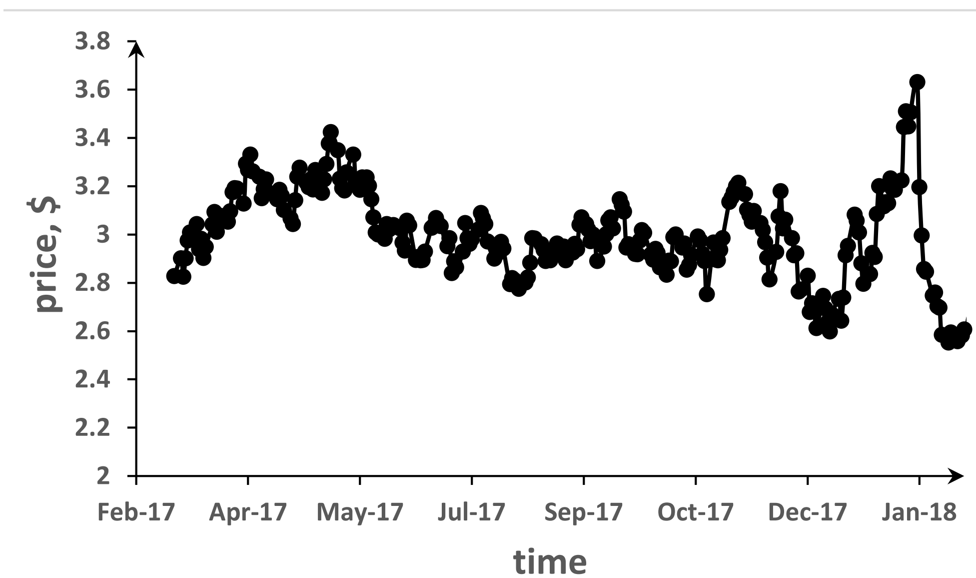


Figure 24.1.5: NG APR18 contract historical prices starting from Mar,2,2017 to Mar,2,2018

24.1.5 Base metal market

Remark 24.1.3 (general remarks). Futures contracts for base metals are primarily traded on the London Metal Exchange maturity dates for futures contracts on the LME.

Futures contracts on COMEX are subject to cash clearing and daily margining

ot every futures contract has an option traded against it. The futures contracts which do have options available to trade against them are known as the monthly value dates, these being those futures with maturity dates on the third Wednesday of the month.

Table 24.1.4: Base metal market in LME and COMEX

metal	maximum tenor	
	futures	options
London Metal Exchange (LME)		
copper	123M	63M
aluminium high grade	123M	63M
aluminium alloy	27M	27M
NASAAC	27M	27M
zinc special high grade	63M	63M
nickel	63M	63M
lead	63M	63M
tin	15M	15M
New York Mercantile Exchange (COMEX)		
copper high grade	60M	24M

Table 24.1.5: LME base metal futures delivery date specifications

	copper aluminium	lead nickel zinc	NASAAC aluminium alloy	tin	
cash/spot	T+2	T+2	T+2	T+2	
0-3M	daily	daily	daily	daily	
3M-6M	weekly	weekly	weekly	weekly	every Wednesday
7-15M	monthly	monthly	monthly	monthly	3rd Wednesday
16-27M	monthly	monthly	monthly	-	3rd Wednesday
28-63M	monthly	monthly	-	-	3rd Wednesday
64-123M	monthly	-	-	-	3rd Wednesday

Table 24.1.6: Option expiry on futures in LME

option expiry	underlying future maturity	underlying future
Wed 05-Jun-2013	Wed 19-Jun-2013	JUN13
Wed 03-Jul-2013	Wed 17-Jul-2013	JUL13
Wed 07-Aug-2013	Wed 21-Aug-2013	AUG13
Wed 04-Sep-2013	Wed 18-Sep-2013	SEP13
...	...	....

24.1.6 Basis risk

Definition 24.1.8 (basis risk). *link*

Basis is the difference between the price of an energy commodity in one market and the price of an energy commodity in different market.

- *The different 'market' can be a different location, also known as* ***locational basis****, a different product or quality which we can be referred to as* ***product or quality basis*** *or a different tenor or time frame, which we refer to as* ***calendar basis****.*

Definition 24.1.9 (location risk). *link*

> ***Locational basis risk*** *is the risk that you encounter when you hedge with a contract that doesn't have the same or similar delivery point as the risk you are seeking to hedge.*
> - *As an example, if a US Gulf Coast oil producer decides to hedge their crude oil price risk with NYMEX WTI futures (which are deliverable in Cushing, Oklahoma), the producer is exposed to the locational basis risk between Cushing and their local market price (i.e. LLS âĂŞ light Louisiana sweet). To quantify this example, the August NYMEX WTI crude oil swap closed yesterday at $46.61 while the August LLS crude oil basis swap (LLS-WTI) closed yesterday at $1.54 which means that the forward market for August LLS swaps is trading at a $1.54 premium to August WTI swaps.*

Definition 24.1.10 (product or quality basis risk). *link*

> ***Product or quality basis risk*** *is the risk that you encounter when you hedge with a contract that isn't the same product or quality as the product that you are seeking to hedge.*
> - *As an example, jet fuel is often hedged with crude oil, gasoil or ultra-low sulfur diesel fuel. While jet fuel, gasoil and ULSD are similar and highly correlated they are not one in the same. As a result, if one chooses to hedge jet fuel with crude oil, gasoil, ULSD or another product, they are often exposing themselves to significant product or quality basis risk, in addition to locational basis risk.*

Definition 24.1.11 (calendar basis risk). *link*

> *Calendar basis risk, also known as calendar spread risk, is the risk that arises from hedging with a contract that doesn't expire, settle or mature on the same date as the underlying exposure.*
> - *As an example, a large consumer (i.e. a vehicle fleet) of gasoline might decide to hedge their exposure to gasoline price by purchasing NYMEX RBOB gasoline futures. In this example, the consumer is exposed to calendar basis risk as NYMEX gasoline futures expire on the last day of the month prior to the delivery month i.e. the August RBOB gasoline futures contract expired on July 29, the last trading day of the month.*

24.2 Spot models

24.2.1 Schwartz one factor model

Definition 24.2.1 (Schwartz one factor model). *[4] The Schwartz one factor model is given by*

$$dS_t/S_t = k(\theta - \ln S_t)dt + \sigma dW_t,$$

where W_t is a Wiener process under the real-world measure. Or equivalently,

$$\begin{aligned} X_t &= \ln S_t \\ dX_t &= k(\alpha - X)dt + \sigma dW_t \\ \alpha &= \theta - \frac{\sigma^2}{2k}. \end{aligned}$$

Lemma 24.2.1 (solution to the Schwartz one factor model). *Consider the Schwartz one factor model, under risk-neutral measure, given by*

$$\begin{aligned} X_t &= \ln S_t \\ dX_t &= k(\alpha - X)dt + \sigma dW_t \\ \alpha &= \theta - \frac{\sigma^2}{2k}. \end{aligned}$$

It follows taht

- $$E_Q[X(T)|\mathcal{F}_t] = \exp(-k(T-t))X(t) + (1 - \exp(-k(T-t)))\alpha.$$
  $$Var_Q[X(T)|\mathcal{F}_t] =$$

- $$S(T) = S(t)\exp(\theta' + (\ln(S_t) - \theta')\exp(-k(T-t)) + \sigma\int_t^T \exp(-k(T-u))dW_u),$$

  where $\theta' = \theta - \frac{1}{2}\sigma^2$.

- *The mean, variance and invariant distribution of the log-spot price under real-world measure are*

$$\begin{aligned}
E[\ln(S(T)/S(t))|\mathcal{F}_t] &= \theta' + (\ln(S_t) - \theta')\exp(-k(T-t)) \\
Var[\ln(S(T)/S(t))|\mathcal{F}_t] &= \frac{\sigma^2}{2k}(1-\exp(-2k(T-t))) \\
E[\ln(S(\infty)/S(t))|\mathcal{F}_t] &= \theta' \\
Var[\ln(S(\infty)/S(t))|\mathcal{F}_t] &= \frac{\sigma^2}{2k}
\end{aligned}$$

- *The forward price is given by*

$$F(t,T) = E_Q[S(T)|\mathcal{F}_t] = \exp(E_Q[X(T)|\mathcal{F}_t] + \frac{1}{2}Var[X(T)|\mathcal{F}_t])$$

Proof. (1) Use the results from OU process(Lemma 6.4.1). (2)(3) straight forward. Use the fact that

$$S(T) = \exp(X(T)) \Leftrightarrow E[S(T)] = \exp(E[X(T)] + \frac{1}{2}Var[X(T)]).$$

□

Remark 24.2.1 (model calibration and model weakness).

- The model calibration involves selecting the model parameter k, θ, σ such that the model derived forward curve can match the market observed forward curve.
- The weakness of the model is that it is hard to match with the market observed forward curve.

24.3 Forward price model

24.3.1 Multi-factor model

Lemma 24.3.1 (No-arbitrage contraint for futures price).

24.3.1.1 *The model*

Definition 24.3.1 (Black model). *[1, p. 73]*

- *(single factor model)The Black forward price single factor model assumes that under risk-neutral measure*
$$\frac{dF(t,T)}{dt} = \sigma(t,T)dW_t$$
where W_t is a Brownian motion under risk-neutral measure Q.
- *(multi-factor model) The Black forward price multi-factor model assumes that under risk-neutral measure*
$$\frac{dF(t,T)}{dt} = \sum_{k=1}^{K} \sigma_k(t,T)dW_k(t),$$
where $W_1, W_2, ..., W_K$ are ***independent*** *Brownian motion under risk-neutral measure Q such that $dW_i dW_j = \rho_{ij} dt$.*

Remark 24.3.1 (choices of parameters).

- The covariance structure[1, p. 73] can be estimated from principle component analysis of historical forward price curves(Lemma 19.4.3).
- Volatility functions often assumed to be deterministic
- The **incremental stationarity** of the $\ln F(t,T)$ with fixed time-to-maturity $T-t$ can be realized by assuming
$$\sigma_k(t,T) = \sigma_k(T-t).$$

Lemma 24.3.2 (basic statistical properties of forward prices). *Consider forward price model*
$$\frac{dF(t,T)}{dt} = \sum_{k=1}^{K} \sigma_k(t,T)dW_k(t),$$
where $W_1, W_2, ..., W_K$ are independent Brownian motion under risk-neutral measure Q such that $dW_i dW_j = \rho_{ij} dt$. It follows that

- *For any $s \geq t$ and fixed T*

$$E[F(s,T)|\mathcal{F}_t] = F(t,T).$$

That is, $F(t,T)$ is a martingale with fixed T.

- *For any $s \geq t \geq t_0$, the convariance structure is given by*

$$E[F(t,T_1)F(s,T_2)] = F(t_0,T)F(t_0,T)\exp(\Sigma_{ij}(t)),$$

where

$$\Sigma_{ij}(t) = \int_{t_0}^{t} \sum_{k=1}^{K} \sigma_k(u,T_i)\sigma_k(u,T_j)du$$

Proof. Directly from the multi-dimensional geometric SDE(Lemma 6.3.15) □

24.3.1.2 *Interprete the covariance structure*

Definition 24.3.2 (rolling futures price). *[5, p. 15] Define*

$$\begin{aligned}\ln f(t;\tau_j) &\triangleq \ln F(t,t+\tau_j) \\ &\approx \frac{(t+\tau_j-T_j)\ln F(t,T_{j+1}) + (T_{j+1}-t-\tau_j)\ln F(t,T_j))}{T_{j+1}-T_j}, T_j < t+\tau_j < T_{j+1}.\end{aligned}$$

*We say $f(t;\tau_j)$ is **the rolling futures price** at time t for contracts maturing at $\tau_j + t$.*

Remark 24.3.2 (the dynamics of full forward curve via interpolation).

- Note that given the dynamics of finite number of points $F(t,T_1), F(t,T_2), ..., F(t,T_n)$ on the full forward curve $F(t,T)$. We can approximate $F(t,\tau)$ by weighted sum of nearby points.
- If $\sigma_i(t,T_i)$ is state independent, then $F(t,T_i)$ will be a Gaussian process; the interpolated dynamics $F(t,\tau)$ will also be a Gaussian process(Theorem 5.3.2).

Lemma 24.3.3 (rolling futures prices dynamics and covariance structure). *[5, p. 15]*

- *The rolling futures price dynamics for N futures expiring at $\tau_1, \tau_2, ..., \tau_N$ are given by*

$$d\ln f(t;\tau_j) = v(t;\tau_j)dt + \sum_{k=1}^{M} \sigma_k(\tau_j)dW_k(t), j = 1,2,...,N$$

$$v_j(t) = [\mu(t;\tau_j) - \frac{1}{2}\sum_{k=1}^{M}\sigma_k^2(\tau_j)] + \frac{\partial \ln F(t,t+\tau_j)}{\partial T_j}$$

where $W_1, W_2, ..., W_M$ are M independent Brownian motions.

- *The vector form is given by*

$$d \ln f = vdt + \Sigma dW, \Sigma \in \mathbb{R}^{N \times M}$$

where

$$f(t) = \begin{bmatrix} f(t;\tau_1) \\ f(t;\tau_2) \\ \vdots \\ f(t;\tau_N) \end{bmatrix}, v(t) = \begin{bmatrix} v(t;\tau_1) \\ v(t;\tau_2) \\ \vdots \\ v(t;\tau_N) \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1(\tau_1) & \sigma_2(\tau_1)) & \cdots & \sigma_M(\tau_1)) \\ \sigma_1(\tau_2) & \sigma_2(\tau_2) & \cdots & \sigma_M(\tau_2) \\ \vdots & \ddots & & \vdots \\ \sigma_1(\tau_N)) & \sigma_2(\tau_N)) & \cdots & \sigma_M(\tau_N)) \end{bmatrix}, dW = \begin{bmatrix} dW_1 \\ dW_2 \\ \vdots \\ dW_N \end{bmatrix}$$

- *The covariance matrix for vector $d \ln f - vdt$ is given by*

$$Cov(d \ln f - vdt, d \ln f - vdt) = \Sigma\Sigma^T dt.$$

Specifically,

$$Cov(d \ln f(t;\tau_i) - v_i dt, d \ln f(t;\tau_j) - v_j dt) = \sum_{k=1}^{M} \sigma_k(\tau_i)\sigma_k(\tau_j)dt = C(\tau_i, \tau_j)dt$$

is a constant.

- *An alternative vector form is given by*

$$d \ln f = vdt + DdZ,$$

or equivalently

$$d \ln f(t;\tau_i) = vdt + D_i dZ_i, i = 1, 2, ..., N$$

where $D = diag(D_1, D_2, ..., D_N), D_i = \sqrt{(\Sigma\Sigma^T)_{ii}}$, and $dZdZ^T = \rho dt$ is correlated Brownian motion, $\rho = D^{-1/2}\Sigma\Sigma^T D^{-1/2}$is the correlation matrix associated with covariance matrix $\Sigma\Sigma^T$.

- *An PCA representation is given by*

$$d \ln f = vdt + \sqrt{\Lambda} U dW,$$

where $\Sigma\Sigma^T = U\Lambda U^T, U \in \mathbb{R}^{N \times M}, \Lambda \in \mathbb{R}^{M \times M}$.

Proof. To show the equivalence between the two forms, we have

$$DdZdZ^T D = D^{1/2}\rho D^{1/2} dt = \Sigma\Sigma^T dt = \Sigma dW dW^T \Sigma^T.$$

□

Remark 24.3.3. Note that $\frac{\partial \ln F(t,t+\tau_j)}{\partial T_j}$ is deterministic quantity since $F(t, t+\tau_j)$ is known at time t.

24.3.2 Calibration and implied volatility surfaces

24.3.2.1 *Implied volatility surface*

Remark 24.3.4. In the commodity futures market, each futures contract usually only has standard options traded on an expiry date close to the futures contract delivery date and at multiple strikes. This impose several challenges:

- It is difficult to obtain market implied dynamics for each futures contract since all its options expire on one date.
- It is difficult to obtain market implied correlation for different futures contracts.

As a consequence, it requires a dynamical model with correlation structure to approximately price derivatives like Asian option, swaption, and spread options.

Definition 24.3.3 (moneyness, implied volatility).

- *The moneyness level for an European call/put maturing at T on the futures $F(t, T_1)$ at current time t is defined as*
$$\mathcal{M} \triangleq F(t, T)/K,$$
where K is the strike and $T < T_1$.
- *The implied volatility is defined as $\sigma^*(\mathcal{M}, T)$ such that*
$$V^{mkt}(t) = Blk(K, T, F(t, T_1), \sigma^*(\mathcal{M}, T)),$$
where $V^{mkt}(t)$ is the market observed price.
- *The implied volatility surface is the $\sigma^*(\mathcal{M}, T)$ as a function of $\mathcal{M}$ and T. Usually, T extends to longest maturity date in the market.*

Definition 24.3.4 (augmented futures price lognormal model to fit market data). *Let $F(t, T_1), F(t, T_2), ..., F(t, T_N)$ be the futures prices maturing at $T_1, T_2, ..., T_n$. If we assume the rolling futures price dynamics follow*
$$\ln f(t; \tau_i) = v_i dt + D(\tau_i) dZ_i, i = 1, 2, ..., N, dZdZ^T = \rho dt,$$

then the augmented futures prices for a given by moneyness $\mathcal{M}$ is defined by

$$dF(t, T_i) = g(t, \mathcal{M}) D(T_i - t) dZ_i, i = 1, 2, ..., N.$$

*Note that the function D is fully determined from covariant structure analysis*24.3.1.

Lemma 24.3.4 (bootstrap method to obtain g).

- *Let τ_0 be the closest maturity date of a call/put option with moneyness $\mathcal{M}$ on a futures $F(t, T_0), T_0 > \tau_0$. Then*

$$\int_t^{\tau_0} g(s, \mathcal{M})^2 D(T_0 - s)^2 ds = \sigma^*(\tau_0, \mathcal{M})^2(\tau_0 - t).$$

- *Let τ_1 be the closest maturity date of a call/put option with moneyness $\mathcal{M}$ on a another futures $F(t, T_1), T_1 > \tau_1 > T_0 > \tau_0$. Then*

$$\begin{aligned} \int_t^{\tau_1} g(s, \mathcal{M})^2 D(T_1 - s)^2 ds &= \sigma^*(\tau_1, \mathcal{M})^2(\tau_1 - t) \\ \int_t^{\tau_0} g(s, \mathcal{M})^2 D(T_0 - s)^2 ds + \int_{\tau_0}^{\tau_1} g(s, \mathcal{M})^2 (D(T_1 - s)^2 - D(T_0 - s)^2) ds &= \sigma^*(\tau_1, \mathcal{M})^2(\tau_1 - t) \\ \sigma^*(\tau_0, \mathcal{M})^2(\tau_0 - t) + \int_{\tau_0}^{\tau_1} g(s, \mathcal{M})^2 (D(T_1 - s)^2 - D(T_0 - s)^2) ds &= \sigma^*(\tau_1, \mathcal{M})^2(\tau_1 - t) \end{aligned}$$

- *Continuing this process we are able to obtain the function $g(t, \mathcal{M})$ where t extends to longest maturity date.*

24.3.2.2 *Covariance structure*

Note 24.3.1 (covariant structure analysis)**.** Consider there are $K+1$ consecutive sample points for each rolling future prices $f(t, \tau_j), j = 1, 2, ..., N$. Denote the de-trend data matrix

$$X = [x_1 \; x_2 \; \cdots \; x_N],$$

where

$$x_i = \begin{bmatrix} \ln f_i(t_2)/f_i(t_1) - \hat{\mu}_i \\ \ln f_i(t_2)/f_i(t_1) - \hat{\mu}_i \\ \vdots \\ \ln f_i(t_2)/f_i(t_1) - \hat{\mu}_i \end{bmatrix}$$

where $f_i(t_j) \triangleq f(t_j;\tau_i) \triangleq F(t_j, t_j+\tau_i)$, $\hat{\mu}_i = \frac{1}{K}\ln f_i(t_{M+1})/f_i(t_1)$.

The covariance matrix can be estimated by

$$\Sigma\Sigma^T = \frac{1}{M-1}XX^T$$

Note that the diagonal terms $D_i = D(\tau_i) = (\Sigma\Sigma^T)_{ii}$ can be parameterized as a function of tenors of different lengths.

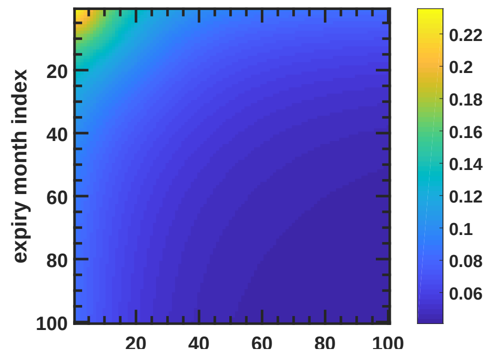
24.3.2.3 *The covariance structure data*

Remark 24.3.5.

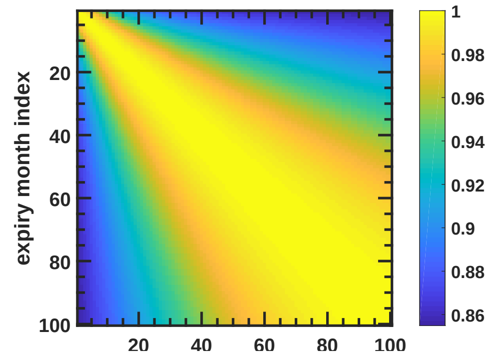
- When we assume the **incremental stationarity** of the $\ln F(t,T)$ with fixed time-to-maturity $T-t$, we can parameterize

$$Cov(d\ln F(t_1,T_1), d\ln F(t_2,T_2)) = C(T_1-t_1, T_2-t_1).$$

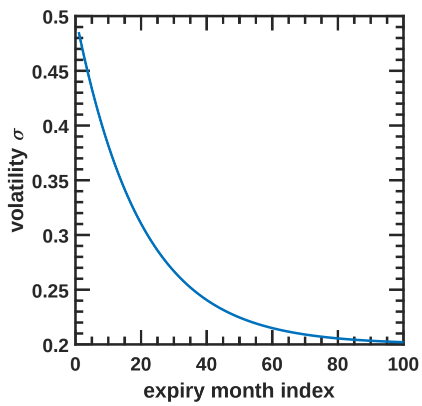
  One example of such function $C(s_1,s_2)$ is showed in Figure 24.3.1(a)(b).
- Usually the matrix $C(s_1,s_2)$ are rank deficient. The eigenvalues showed in (c) and eigenvectors showed in (d) of $C(s_1,s_2)$ can give the factor representation of $d\ln F$.



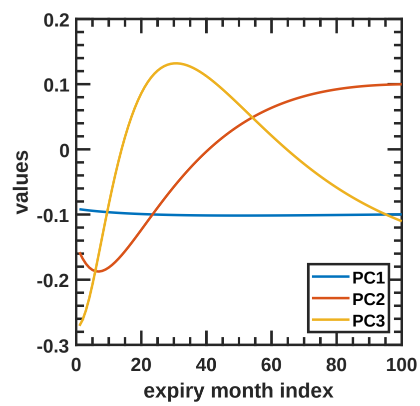
(a) Forward price covariance.



(b) Forward price correlation.



(c) Volatility term structure.



(d) Principal component factors.

Figure 24.3.1: The covariance structure of the forward curves.

24.4 Pricing examples

24.4.1 European options on futures

Lemma 24.4.1 (Options on futures). *[6, p. 101] Assume the futures price $F(t, T_1)$ has the risk-neutral dynamics of*

$$dF(t, T_1)/F(t, T_1) = \sigma(t) dW_t.$$

Let current time be t and the constant short rate be r The call/put option with strike K and matures at $T, T < T_1$ will have price

$$C(t) = e^{r(T-t)}(F(t, T_1)N(d_1) - KN(d_2).$$

$$P(t) = -e^{r(T-t)}(F(t, T_1)N(-d_1) + KN(-d_2).$$

where

$$d_1 = \frac{1}{\Sigma\sqrt{T-t}}[\ln(\frac{F(t, T_1)}{K}) + \Sigma^2/2(T-t)]$$
$$d_2 = d_1 - \Sigma\sqrt{T-t}$$

$$\Sigma^2 = \frac{1}{T-t}\int_t^T \sigma^2(u) du.$$

Proof. Note that $F(T, T_1)$ is the log-normal random variable $LN(F(t, T_0), \Sigma^2(T-t)$. The pricing derivation is same as Lemma 18.2.1. □

24.4.2 Asian options

Lemma 24.4.2 (distribution of arithmetic average of single futures). *Let $t_0 \leq t_1 < t_2 < \cdots < t_m \leq T$ be a set of dates. Let*

$$X = \sum_{i=1}^{m} w_i F(t_i, T).$$

Then

- $E_Q[X|\mathcal{F}_0] = \sum_{i=1}^m w_i E_Q[F(t_i, T)|\mathcal{F}_0] = \sum_{i=1}^m w_i F(t_0, T).$

- $$\begin{aligned} E[X^2] &= \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j E_Q[F(t_i,T)F(t_j,T)|t_0] \\ &= \sum_{i=1}^{m}\sum_{j=i+1}^{m} 2w_i w_j F^2(t_0,T)\exp(\int_{t_0}^{t_i} g(s)g(s)D(T-s)D(T-s)ds) \end{aligned}$$

Proof. (1) Use linearity of expectation and the fact that $F(t,T)$ is a martingale in measure Q. (2)

$$\begin{aligned} F(t_i,T) &= F(t_0,T)\exp(-\frac{1}{2}\int_{t_0}^{t_i} g^2(s)D^2(T-s)ds + \int_{t_0}^{t_i} g(s)D(T-s)dW(s)) \\ \implies F(t_i,T)F(t_j,T) &= F(t_0,T)^2\exp(-\frac{1}{2}\int_{t_0}^{t_i} g^2(s)D^2(T-s)ds + \int_{t_0}^{t_1} g(s)D(T-s)dW(s)) \\ &\exp(-\frac{1}{2}\int_{t_0}^{t_j} g^2(s)D^2(T-s)ds + \int_{t_0}^{t_j} g(s)D(T-s)dW(s)) \\ &= F(t_0,T)^2\exp(-\int_{t_0}^{t_i} g^2(s)D^2(T-s)ds + 2\int_{t_0}^{t_1} g(s)D(T-s)dW(s)) \\ &\exp(-\frac{1}{2}\int_{t_i}^{t_j-t_i} g^2(s)D^2(T-s)ds + \int_{t_i}^{t_j-t_i} g(s)D(T-s)dW(s)) \\ &= F(t_0,T)^2\exp(-\int_{t_0}^{t_i} g^2(s)D^2(T-s)ds + 2\int_{t_0}^{t_1} g(s)D(T-s)dW(s)) \\ E[F(t_i,T)F(t_j,T)] &= F(t_0,T)^2\exp(\int_{t_0}^{t_i} g^2(s)D^2(T-s)ds) \end{aligned}$$

□

Lemma 24.4.3 (distribution of arithmetic average of multiple futures). *Let $t_0 \leq t_1 < T_1 < t_2 < T_2 < \cdots < t_m < T_m$ be a set of dates. Let*

$$X = \sum_{i=1}^{m} w_i F(t_i, T_i).$$

Then

- $E_Q[X|\mathcal{F}_0] = \sum_{i=1}^{m} w_i E_Q[F(t_i,T)|\mathcal{F}_0] = \sum_{i=1}^{m} w_i F(t_0,T_i)$.

- $$\begin{aligned} E[X^2] &= \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j E_Q[F(t_i,T)F(t_j,T)|t_0] \\ &= \sum_{i=1}^{m}\sum_{j=i+1}^{m} 2w_i w_j F(t_0,T_i)F(t_0,T_j)\exp(\int_{t_0}^{t_i} g(s)g(s)D(T_i-s)D(T_j-s)\rho_{ij}ds) \end{aligned}$$

Proof. (1) Use linearity of expectation and the fact that $F(t,T)$ is a martingale in measure Q. (2)

$$\begin{aligned} F(t_i,T_i) &= F(t_0,T_i)\exp(-\frac{1}{2}\int_{t_0}^{t_i} g^2(s)D^2(T_i-s)ds + \int_{t_0}^{t_i} g(s)D(T_i-s)dZ_i(s)) \\ \implies F(t_i,T)F(t_j,T) &= F(t_0,T)^2\exp(-\frac{1}{2}\int_{t_0}^{t_i} g^2(s)D^2(T_i-s)ds + \int_{t_0}^{t_1} g(s)D(T_i-s)dZ_i(s)) \\ &\exp(-\frac{1}{2}\int_{t_0}^{t_j} g^2(s)D^2(T_j-s)ds + \int_{t_0}^{t_j} g(s)D(T_j-s)dZ_j(s)) \\ E[F(t_i,T_i)F(t_j,T_j)] &= F(t_0,T_i)F(t_0,T_j)\exp(\int_{t_0}^{t_i} g(s)^2 D(T_i-s)D(T_j-s)\rho_{ij}ds) \end{aligned}$$

To get $E[F(t_i,T_i)F(t_j,T_j)]$, we use the fact(Lemma 2.2.14) that If $X_1 \in N(\mu_1,\sigma_1^2)$, $X_2 \in N(\mu_2,\sigma_2^2)$, then

$$E[\exp(X_1+X_2)] = \exp(E[X_1]+E[X_2]+\frac{1}{2}Var[X]+\frac{1}{2}Var[X_2]+Cov(X_1,X_2)).$$

□

24.4.3 Commodity swaps

A single futures contract enables the holder to get exposure to the futures price change on a single delivery date. Considering the highly volatile nature of commodity prices, exposure to futures prices averaging on multiple dates might be more desired. Commodity swap meets such needs, and can be viewed as a collection of multiple futures on the same commodity.

Definition 24.4.1 (commondity swap). *[3, p. 42]*

- *A **commodity swap** involves exchanging floating payments indexed to commodity prices against a fixed known price K, with the cash settlement either at the end of the swap or on a regular interval basis.*
- *The floating payment is usually indexed to spot price or prompt futures price.*

Lemma 24.4.4 (commodity swap value and fair swap rate). *[3, p. 42] Let $t_0 < t_1 < t_2 < \cdots < t_n$ be a set of dates. Let $\mathcal{T}(t_i)$ be a function map t_i to the nearest futures contract maturity date. Let t_0 be current time.*

- *Consider a swap exchanging a floating leg with a fixed leg. The floating leg is paying $\frac{1}{n}\sum_{i=1}^n F(t_i, \mathcal{T}(t_i))$ at time T_{stl} with weight $T_{stl} < t_1$; the fixed leg is paying amount K at time T_{stl}. We assume $t_0 < T_{stl}$.*
  - *The value of the swap at time t_0 is*

$$\begin{aligned} V_0 = \exp(-r(T_{stl} - t_0))E_Q[\frac{1}{n}\sum_{i=1}^n F(t_i, \mathcal{T}(t_i))|\mathcal{F}_0] \\ - \exp(-r(T_{stl} - t_0))K = \exp(-r(T_{stl} - t_0))[\frac{1}{n}\sum_{i=1}^n F(t_0, \mathcal{T}(t_i)) - K]. \end{aligned}$$

  - *The fair swap rate is*

$$V_0 = 0 \implies K_{swap} = \frac{1}{n}\sum_{i=1}^n F(t_0, \mathcal{T}(t_i))$$

-
  - *Consider a swap exchanging floating legs $F(t_i, \mathcal{T}(t_i))$ with a fixed leg amount K at time $T_{stl,i}, i = 1, 2, ..., n$. Then the value of swap at time t_0 is given by*

$$V_0 = E_Q[\sum_{i=1}^n \exp(-r(T_{stl,i} - t_0))V_i] = \sum_{i=1}^n \exp(-r(T_{stl,i} - t_0))[F(t_0, \mathcal{T}(t_i)) - nK],$$

    where

$$V_i = F(t_i, \mathcal{T}(t_i)) - K.$$

  - *The fair swap rate is*

$$K_{swap} = \frac{\sum_{i=1}^n \exp(-r(T_{stl,i} - t_0))F(t_0, \mathcal{T}(t_i))}{\sum_{i=1}^n \exp(-r(T_{stl,i} - t_0))}.$$

Proof. (1) Note that under log-normal model, the futures price is a martingale; that is,

$$F(t_0, \mathcal{T}(t_i)) = E[F(t_i, \mathcal{T}(t_i))|\mathcal{F}_0].$$

To get the fair swap rate, we have

$$V_0 = 0 \implies K_{swap} = \frac{1}{n}\sum_{i=1}^{n} F(t_0, \mathcal{T}(t_i))$$

(2) Similar to (1). □

Lemma 24.4.5 (Greeks associated with a commodity swap). *Let $t_0 < t_1 < t_2 < \cdots < t_n$ be a set of dates. Let $\mathcal{T}(t_i)$ be a function map t_i to the nearest futures contract maturity date. Let t_0 be current time. Consider a swap exchanging floating legs $F(t_i, \mathcal{T}(t_i))$ with a fixed leg amount K at time $T_{stl,i}, i = 1, 2, ..., n$.*

- *Then the value of swap at time t_0 is given by*

$$V_0 = E_Q[\sum_{i=1}^{n} \exp(-r(T_{stl,i} - t_0))V_i] = \sum_{i=1}^{n} \exp(-r(T_{stl,i} - t_0))[F(t_0, \mathcal{T}(t_i)) - nK],$$

where

$$V_i = F(t_i, \mathcal{T}(t_i)) - K.$$

- *The swap has exposure the forward price curve; that is,*

$$\frac{\partial V_0}{\partial F(t_0, \mathcal{T}(t_i))} = \exp(-r(T_{stl,i} - t_0)).$$

- *The swap has zero Gamma and vega risk; that is,*

$$\frac{\partial^2 V_0}{\partial F(t_0, \mathcal{T}(t_i))^2} = 0, \frac{\partial V_0}{\partial \sigma} = 0.$$

Proof. Straight forward from Lemma 24.4.4. □

24.4.4 Spread options

24.4.4.1 *Basics*

Spread options are very popular in all commodity markets. Common spreads include crackspread, which is the difference between the prices of refined products and the price of crude oil input, sparkspread, which is the difference between the price of electricity (output) and the price of the corresponding quantity of primary fuel (input).

Spread options can be used by oil refineries, power plant, etc to hedge the risks associated with the input and ouput commondiites prices, hence locking the profit.

Calendar spread options are options on the prices spread of futures with two different maturities on the same commodity. They can be used by companies providing stroage services, such as gas storage, water reservoirs, for valuation and hedging purpose.

In summary, we can define spread options as follows:

Definition 24.4.2 (spread option on futures prices).

- *A spread is the difference between two commodity prices (either spot price or futures price).*
- *A spread on two futures prices $F(T, T_1)$ and $F(T, T_2), T_1 \neq T_2$ is called calendar spread.*
- *A spread option on two futures prices $F_1(t, T_1), F_2(t, T_2)$ with strike K and maturity $T < T_1, T < T_2$ has payoff*

$$V(T) = \max(F(T, T_2) - F(T, T_1) - K, 0).$$

24.4.4.2 *Pricing*

Lemma 24.4.6 (pricing when strike is zero). *Consider a spread option on two futures prices $F(t, T_1), F(t, T_2)$ with strike 0, maturity $T < T_1, T < T_2$, and payoff*

$$V(T) = \max(F(T, T_2) - F(T, T_1), 0).$$

Assume the following risk-neutral model for $F_1(t) \triangleq F(t, T_1)$, $F_2(t) \triangleq F(t, T_2)$,

$$\begin{aligned} dF_1(t) &= \sigma_1 F_1(t) dW_1(t) \\ dF_2(t) &= \rho\sigma_2 F_2(t) dW_1(t) + \sqrt{1-\rho^2}\sigma_2 F_2(t) dW_2(t) \end{aligned}$$

*where W_1 and W_2 are **independent** Brownian motions.*

Then the value of the exchange option at time 0 is given by

$$V_0 = F_1(0)N(d_1) - F_2(0)N(d_2)$$

where

$$d_1 = \frac{\ln(F_1(0)/F_2(0) + (\hat{\sigma}^2/2)}{\hat{\sigma}\sqrt{T}}, d_2 = d_1 - \hat{\sigma}\sqrt{T},$$

and

$$\hat{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Proof. Use the F_1 as the numeraire, then under this new measure Q_F,

$$\frac{V(T)}{S_1(T)}$$

is a martingale. Therefore,

$$\frac{V(0)}{F_1(t)} = E_F[\frac{V(T)}{F_1(T)}|\mathcal{F}_t] = E_F[\max(\frac{F_2(T)}{F_1(T)} - 1)|\mathcal{F}_t].$$

Note that under measure Q_F Theorem 15.6.19), the dynamics of $F_1(t)$ and $F_2(t)$ follows

$$\begin{aligned}
dF_1(t) &= (\sigma_1^2)F_1(t)dt + \sigma_1 F_1(t)dW_1(t) \\
dF_2(t) &= (r + \rho\sigma_1\sigma_2)F_2(t)dt + \rho\sigma_2 F_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 S_2(t)dW_2(t) \\
d\frac{F_2}{F_1} &= \frac{F_2}{F_1}((\rho\sigma_2 - \sigma_1)dW_1 + \sqrt{1-\rho^2}\sigma_2\sigma dW_2)
\end{aligned}$$

Denote $Y = \frac{F_2}{F_1}$, then Y is a geometric Brownian motion with volatility $\sigma_Y = \sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$.

Then we have $V(t) = F_1(t)E_F[\max(Y(T) - 1, 0)]$, which can be evaluated. □

Lemma 24.4.7 (pricing with Kirk approximation). *Consider a spread option on two futures prices $F(t, T_1), F(t, T_2)$ with strike K, maturity $T < T_1, T < T_2$, and payoff*

$$V(T) = \max(F(T, T_2) - F(T, T_1), 0).$$

Assume the following risk-neutral model for $F_1(t) \triangleq F(t, T_1)$, $F_2(t) \triangleq F(t, T_2)$,

$$\begin{aligned}
dF_1(t) &= \sigma_1 F_1(t)dW_1(t) \\
dF_2(t) &= \rho\sigma_2 F_2(t)dW_1(t) + \sqrt{1-\rho^2}\sigma_2 F_2(t)dW_2(t)
\end{aligned}$$

*where W_1 and W_2 are **independent** Brownian motions.*

If $K \ll F_1(T)$, then the approximate value of the exchange option at time 0 is given by

$$V_0 = (F_1(0) + K)N(d_1) - F_2(0)N(d_2)$$

where

$$d_1 = \frac{\ln(F_1(0) + K/F_2(0) + (\hat{\sigma}^2/2)}{\hat{\sigma}\sqrt{T}}, d_2 = d_1 - \hat{\sigma}\sqrt{T},$$

and

$$\hat{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Proof. In the Kirk approximation, we assume

$$d(F_1 + K)/(F_1 + K) = \sigma_1 dW_1.$$

View $F_1 + K$ as a new random variable, then we can use the previous result in Lemma 24.4.6. □

24.4.5 Commodity quanto securities

Quanto securities are quite common in energy commodity markets. For example, an Chinese customer planning to buy oil from International Petroleum Exchange in London will have exposure to both oil price and currency risks. Quanto securities, e.g., quanto forwards and quanto options, usually can help hedge the currency risks.

The pricing methods of quanto securities can be found in subsection 22.6.9.

24.5 Notes on bibliography

Major references are .

[7][3]

[8][9][1]

For multi-factor model, see [10].

Problems and solutions in mathematical finance: commodity and foreign exchange

Market data sources: investing.com

BIBLIOGRAPHY

1. Geman, H. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy* (John Wiley & Sons, 2009).
2. Roncoroni, A., Fusai, G. & Cummins, M. *Handbook of multi-commodity markets and products: structuring, trading and risk management* (John Wiley & Sons, 2015).
3. Clark, I. J. *Commodity Option Pricing: A Practitioner's Guide* (John Wiley & Sons, 2014).
4. Schwartz, E. S. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The journal of finance* **52,** 923–973 (1997).
5. Ellefsen, P. E. & Sclavounos, P. D. Multi-Factor Model of Correlated Commodity-Forward Curves for Crude Oil and Shipping Markets (2009).
6. Lalley, S. *The mathematics of financial derivatives: a student introduction* (Chicago University Statistics 390/ Mathematical Finance 345, 2001).
7. Burger, M., Graeber, B. & Schindlmayr, G. *Managing energy risk: An integrated view on power and other energy markets* (John Wiley & Sons, 2008).
8. Jewson, S. & Brix, A. *Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations* (Cambridge University Press, 2005).
9. Parodi, P. *Pricing in general insurance* (CRC Press, 2014).
10. Bukkapatanam, V. & Doraiswamy, S. T. Multi factor models for the commodities futures curve: Forecasting and pricing. *Futures* **15,** 20.

Part IV

MATHEMATICAL RISK MANAGEMENT

25

FINANCIAL RISK MANAGEMENT

25.1 Risk concept and Risk measure

25.1.1 Different types of risks

Definition 25.1.1 (financial risk). *Financial risks are those that arise from exposure to financial markets. Examples areïijŽ*

- ***Credit risk.*** *This is the uncertainty about whether the counterparty to a transaction will fulfill its contractual obligations.*
- ***Liquidity risk.*** *This is the risk of loss when selling an asset at a time when market conditions make the sales price less than the underlying fair value of the asset.*
- ***Market risk.*** *This is the uncertainty about market prices of assets (stocks, commodities, and currencies) and interest rates.*

Definition 25.1.2 (non-financial risk). ***Non-financial risks*** *arise from the operations of the organization and from sources external to the organization. Examples are:*

- ***Operational risk.*** *This is the risk that human error or faulty organizational processes will result in losses.*
- ***Solvency risk.*** *This is the risk that the organization will be unable to continue to operate because it has run out of cash.*
- ***Regulatory risk.*** *This is the risk that the regulatory environment will change, imposing costs on the firm or restricting its activities.*
- ***Governmental or political risk (including tax risk)****. This is the risk that political actions outside a specific regulatory framework, such as increases in tax rates, will impose significant costs on an organization.*
- ***Legal risk.*** *This is the uncertainty about the organization's exposure to future legal action.*
- ***Model risk.*** *This is the risk that asset valuations based on the organization's analytical models are incorrect.*
- ***Tail risk.*** *This is the risk that extreme events (those in the tails of the distribution of outcomes) are more likely than the organization's analysis indicates, especially from incorrectly concluding that the distribution of outcomes is normal.*
- ***Accounting risk.*** *This is the risk that the organization's accounting policies and estimates are judged to be incorrect.*

25.1.2 Risk measures

Definition 25.1.3 (risk measure). *[1, p. 111] Given a portfolio w, a risk measure is function $R(w)$ mapping to a number in $\mathbb{R}$.*

Definition 25.1.4 (example risk measures). *[1, p. 112]*

- ***volatility of the loss***
$$R(w) \triangleq \sigma(L(w)).$$
- ***standard-deviation based measure***
$$R(w) \triangleq E[L(w)] + c \cdot \sigma(L(w)).$$
- ***value-at-risk***
$$R(w) \triangleq VaR_\alpha(w) = \inf\{l : Pr(L(w) \leq l) \geq \alpha\}.$$
- ***expected shortfall***
$$R(w) \triangleq ES_\alpha(w) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(w)du.$$
Note that
$$ES_\alpha(w) = E[L(w)|L(w) \geq VaR_\alpha(w)].$$

Definition 25.1.5 (coherent risk measure). *[1, p. 111] A risk measure $R(w)$ is said to be coherent if it satisfies*

- ***sub-additivity**:*
$$R(w_1 + w_2) \leq R(w_1) + R(w_2);$$
that is the risk of two portfolios should be less than adding the risk of the two separate portfolios.
- ***homogeneity**:*
$$R(\lambda w) = \lambda R(w), \forall \lambda \geq 0;$$
that is leveraging the portfolio increases the risk measure in the same magnitude.
- ***monotonicity**: if $w_1 \prec w_2$, then $R(w_1) \geq R(w_2)$; that is, if portfolio w_2 has a better return than portfolio w_1 under all scenarios, risk measure $R(w_1)$ should be higher than risk measure $R(w_2)$.*
- ***translation invariance**: If $m \in \mathbb{R}$, then $R(w+m) = R(w) - m$; that is, adding a cash position of amount m to the portfolio reduces the risk by m.*

Remark 25.1.1 (interpretation on translation invariance). [1, p. 111]

- Consider a portfolio w has **loss** distribution $R(w) = -\mu + Y$, where Y is a zero mean random variable and $-\mu$ is the mean value of the loss(μ is the mean value of the gain).
- When adding a risk-free cash position C in the portfolio, we want to have $R(w + C) = -(\mu + C) + Y = R(w) - C$..

Definition 25.1.6 (convexity property). *[1, p. 111] The risk measure $R(w)$ is said to be convex if*

$$R(\lambda w_1 + (1-\lambda)w_2) \leq \lambda R(w_1) + (1-\lambda)R(w_2), \forall \lambda \in [0,1].$$

Lemma 25.1.1. *Convexity implies sub-additivity and homogeneity but not the converse.*

Proof. The convexity implies

$$R(\lambda w_1 + (1-\lambda)w_2) \leq \lambda R(w_1) + (1-\lambda)R(w_2).$$

Take $\lambda = 1$ we get homogeneity; take $\lambda = 1/2$ we get sub-additivity. □

25.2 Value at risk

25.2.1 Foundations

Definition 25.2.1 (value at risk). *[1, p. 65] The -value-at-risk(VaR) $VaR_\alpha(w;h)$ or $VaR_\alpha(L)$ is defined as the potential loss which the portfolio w can suffer for a given confidence level α and a fixed holding period h. Three parameters are needed for the computation VaR:*

- *the **holding period** h, which indicates the time period to calculate the loss;*
- *the **confidence level** α, which gives the probability that the loss is lower than the VaR(or equivalently, there is $1-\alpha$ probability that the loss is greater than the VaR).*
- *the **portfolio** w, which gives the allocation in terms of risky assets and is related to risk factors.*

Remark 25.2.1. Typical values of α are 90%, 95%, 99%.

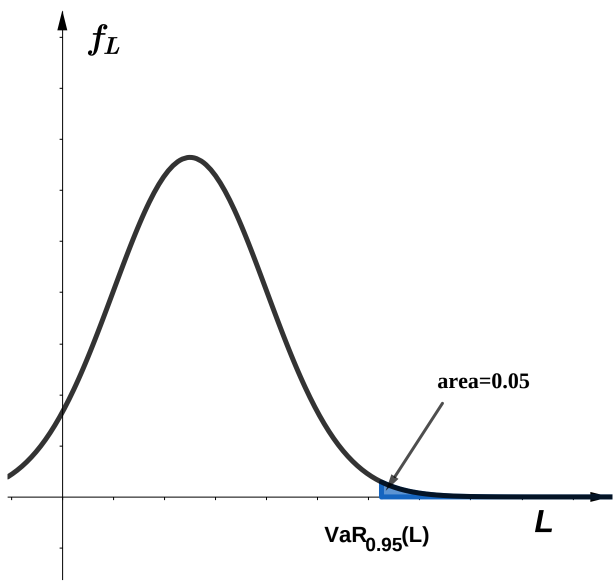


Figure 25.2.1: VaR demo. Given the loss L distribution, we can determine VaR.

Definition 25.2.2 (VaR calculation). *[1, p. 65] Let $P_t(w)$ be the mark-to-market value of the portfolio w at time t. We define the loss of the portfolio(which is a random variable) at future time $t+h$ as*

$$L(w) = P_t(w) - P_{t+h}(w).$$

Then $VaR_\alpha(w;h)$ is given by

$$VaR_\alpha(w;h) = F_L^{-1}(\alpha),$$

where F_L is the CDF of the random loss L. Often, we write $VaR_\alpha(w;h)$ as $VaR_\alpha(L)$.

Remark 25.2.2.

- VaR is a 'tail' risk measure
- VaR_p is increasing in p

Definition 25.2.3 (expected shortfall, conditional value at risk). *The conditional value at risk $CVaR_p(L)$ of random variable L at the confidence level $\alpha \in (0,1)$ is defined as*

$$\begin{aligned} ES = CVaR_\alpha(L) &= E[L|L \geq VaR_\alpha(L)] \\ &= \frac{\int_{VaR_\alpha(L)}^{\infty} x f_L(x)dx}{P(L \geq VaR_\alpha(L))} \\ &= \frac{1}{1-\alpha}\int_{VaR_\alpha(L)}^{\infty} x f_L(x)dx \\ &= \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_u(L)du \end{aligned}$$

Remark 25.2.3 (the transformation details). We use the following procedure to do the transformation

$$\begin{aligned} ES &= \frac{1}{1-\alpha}\int_{VaR_\alpha(L)}^{\infty} x f_L(x)dx \\ &= \frac{1}{1-\alpha}\int_{VaR_\alpha(L)}^{\infty} x dF_L(x) \\ &= \frac{1}{1-\alpha}\int_{\alpha}^{1} F_L^{-1}(u)du \\ &= \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_u(L)du \end{aligned}$$

where we used the variable transformation for integrals(**??**).

Lemma 25.2.1 (VaR basic properties). *VaR satisfies **homogeneity, monotonicity, and translational invariance** but fails sub-additivity; therefore, **VaR is not a coherent risk measure**. Specifically,*

- *For any $\lambda \geq 0$, $VaR_\alpha(\lambda L) = \lambda VaR_\alpha(\lambda L)$.*
- *For random variables L_1 and L_2 such that $L_1(\omega) < L_2(\omega) \forall \omega \in \Omega$, we have $VaR_\alpha(L_1) < VaR_\alpha(L_2)$.*

- *For any $C \in \mathbb{R}$, $VaR_\alpha(L+C) = VaR_\alpha(L) + C$.*

Proof. (1)(homogeneity) For any $\lambda \geq 0$, we have

$$VaR_\alpha(\lambda L) = F^{-1}_{\lambda L}(\alpha) = \lambda F^{-1}_L(\alpha) = VaR_\alpha(L),$$

where we use the scaling property of cdf Lemma 1.4.8. (2)(monotonicity) Consider L_1 and L_2 for the losses of two portfolios and assume $L_1(\omega) < L_2(\omega) \forall \omega \in \Omega$. Then given a fixed α, we have

$$VaR_\alpha(L_1) < VaR_\alpha(L_2).$$

(3)(translational invariance)Consider a constant loss $C \in \mathbb{R}$. We have

$$VaR_\alpha(L+C) = F^{-1}_{L+C}(\alpha) = C + F^{-1}_L(\alpha) = C + VaR_\alpha,$$

where we use the translational property of cdf Lemma 1.4.8. □

Lemma 25.2.2 (ES properties). *Expected shortfall satisfies **sub-additivity, homogeneity, monotonicity, and translational invariance**; Expect shortfall is a coherent risk measure. Specifically,*

- *For random variables L_1 and L_2, $ES(L_1 + L_2) \leq ES(L_1) + ES(L_2)$.*
- *For any $\lambda \geq 0$, $ES(\lambda L) = \lambda ES(\lambda L)$.*
- *For random variables L_1 and L_2 such that $L_1(\omega) < L_2(\omega) \forall \omega \in \Omega$, we have $ES(L_1) < ES(L_2)$.*
- *For any $C \in \mathbb{R}$, $ES(L+C) = ES(L) + C$.*

Proof. Note that $ES = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du$. Using the VaR properties of homogeneity, monotonicity, and translational invariance, we have (1) (homogeneity) For any $\lambda \geq 0$, we have

$$ES(\lambda L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(\lambda L)du = \frac{1}{1-\alpha}\int_\alpha^1 \lambda VaR_u(L)du = \lambda ES(L),$$

(2) (monotonicity) Consider L_1 and L_2 for the losses of two portfolios and assume $L_1(\omega) < L_2(\omega) \forall \omega \in \Omega$. Then given a fixed α, we have

$$ES(L+C) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L_1)du < \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L_2)du = ES(L_2).$$

(3) (translational invariance)Consider a constant loss $C \in \mathbb{R}$. We have

$$ES(L+C) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L+C)du = \frac{1}{1-\alpha}\int_\alpha^1 (VaR_u(L) + C)du = ES(L) + C.$$

(4) (subadditivity)

□

25.2.2 Analytical VaR and ES

25.2.2.1 *Loss with normal distribution*

Lemma 25.2.3 (VaR with normal distribution of loss). *[1, p. 73] Assume $L(w) \sim N(\mu, \sigma^2)$(usually $\mu < 0$). Then*

$$VaR_\alpha(w; h) = \mu + \Phi^{-1}(\alpha)\sigma,$$

where $\Phi(x)$ is the cdf for the standard normal distribution.

Proof. Denote $F_L(x) : \mathbb{R} \to [0, 1]$ as the cdf of random variable $L(w)$. Then $Var_\alpha(w; h) = F_L^{-1}(\alpha)$. We have

$$\begin{aligned}
Pr(L \leq F_L^{-1}(\alpha)) &= \alpha \\
Pr(\mu + \sigma Z \leq F_L^{-1}(\alpha)) &= \alpha, Z \sim N(0,1) \\
Pr(Z \leq \frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= \alpha \\
\Phi(\frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= \alpha \\
\Phi^{-1} \circ \Phi(\frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= \Phi^{-1}(\alpha) \\
\frac{F_L^{-1}(\alpha) - \mu}{\sigma} &= \Phi^{-1}(\alpha) \\
F_L^{-1}(\alpha) &= \mu + \Phi^{-1}(\alpha)\sigma
\end{aligned}$$

□

Corollary 25.2.0.1 (extensions for VaR). *[1, p. 76] Let $R_{t+h} \in \mathbb{R}^n$ be the vector of asset returns in a horizon h. We note that the loss for a portfolio characterized by weight vector $W_t \in \mathbb{R}^n$ is given by*

$$L(W) = -W_t^T R_{t+h}.$$

- *Assume $R_{t+h} \sim MN(\mu, \Sigma)$. Then VaR with confidence level α is given by*

$$VaR_\alpha(W; h) = -W_t^T \mu + \Phi^{-1}(\alpha)\sqrt{W_t^T \Sigma W_t}.$$

- *Assume* $R_{t+h} = B\mathcal{F}_t + \epsilon_t, B \in \mathbb{R}^{n\times m}, \mathcal{F}_t \sim MN(\mu, \Omega), \epsilon_t \sim MN(0, D)$, $\mathcal{F}_t, \epsilon_t$ *are m-dimension random vectors independent to each other. Then VaR with confidence level* α *is given by*

$$VaR_\alpha(W;h) = -W_t^T B\mu + \Phi^{-1}(\alpha)\sqrt{W_t^T(B\Omega B^T + D)W_t}.$$

Proof. (1)Note that $L(w) \sim N(W^T\mu, W^T\Sigma W)$. (2) Note that

$$E[L(w)] = E[-W_t^T R_{t+h}] = W^T B\mu, Cov[L(w)] = W_t^T(B\Omega B^T + D)W_t.$$

□

Lemma 25.2.4 (ES with normal distribution of loss). *[2, p. 264] Assume* $L(w) \sim N(\mu, \sigma^2)$*(usually* $\mu < 0$*). Then*

$$ES_\alpha(w;h) = \mu + \sigma\frac{\exp(-y^2/2)}{\sqrt{2\pi}(1-\alpha)}, y = \Phi^{-1}(\alpha),$$

where $\Phi(x)$ *is the cdf for the standard normal distribution.*

Proof. Note that from definition of ES, we have

$$\begin{aligned}
ES &= \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du \\
&= \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(\mu+\sigma Z)du, Z \sim N(0,1) \\
&= \frac{1}{1-\alpha}\int_\alpha^1 \sigma VaR_u(Z) + \mu du \\
&= \mu + \sigma\frac{1}{1-\alpha}\int_\alpha^1 VaR_u(Z)du \\
&= \mu + \sigma\frac{1}{1-\alpha}\int_{VaR_\alpha(Z)}^\infty x f_Z(x)dx \\
&= \mu + \sigma\frac{1}{1-\alpha}\int_y^\infty \frac{x}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})dx, y = VaR_\alpha(Z) \\
&= \mu + \sigma\frac{1}{(1-\alpha)\sqrt{2\pi}}\exp(-\frac{y^2}{2})dx
\end{aligned}$$

where we use the homogeneity and translational invariance of VaR(Lemma 25.2.1). □

Corollary 25.2.0.2 (extensions for ES). *[1, p. 76] Let $R_{t+h} \in \mathbb{R}^n$ be the vector of asset returns in a horizon h. We note that the loss for a portfolio characterized by weight vector $W_t \in \mathbb{R}^n$ is given by*

$$L(W) = -W_t^T R_{t+h}.$$

- *Assume $R_{t+h} \sim MN(\mu, \Sigma)$. Then ES with confidence level α is given by*

$$ES = -W_t^T \mu + \sqrt{W_t^T \Sigma W_t} \frac{\exp(-y^2/2)}{\sqrt{2\pi}(1-\alpha)}, y = \Phi^{-1}(\alpha),$$

  where $\Phi(x)$ is the cdf for the standard normal distribution.
- *Assume $R_{t+h} = B\mathcal{F}_t + \epsilon_t, B \in \mathbb{R}^{n\times m}, \mathcal{F}_t \sim MN(\mu, \Omega), \epsilon_t \sim MN(0, D)$, $\mathcal{F}_t, \epsilon_t$ are m-dimension random vectors independent to each other. Then ES with confidence level α is given by*

$$ES = -W_t^T B\mu + \sqrt{W_t^T (B\Omega B^T + D) W_t} \frac{\exp(-y^2/2)}{\sqrt{2\pi}(1-\alpha)}, y = \Phi^{-1}(\alpha),$$

  where $\Phi(x)$ is the cdf for the standard normal distribution.

Proof. (1)Note that $L(w) \sim N(W^T\mu, W^T\Sigma W)$. (2) Note that

$$E[L(w)] = E[-W_t^T R_{t+h}] = W^T B\mu, Cov[L(w)] = W_t^T (B\Omega B^T + D) W_t.$$

□

25.2.2.2 *Loss with t distribution*

Lemma 25.2.5 (VaR with t distribution of loss). *[1, p. 73] Assume $L(w) \sim t_v(\mu, \sigma^2), v > 2$(usually $\mu < 0$). Let $L_1, L_2, ..., L_n$ denote the random sample of L. Then*

$$VaR_\alpha(w; h) = \mu + T^{-1}(\alpha)\sigma,$$

where $T(x)$ is the cdf for the standard normal distribution.

If σ is estimated using mothod of moments, we have

$$VaR_\alpha(w; h) = \mu + T^{-1}(\alpha)\hat{\sigma}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (L_i - \bar{L})^2 \frac{v-2}{v}$$

.

Proof. (1) Note that $L = \mu + \sigma Z, Z \sim t_v(0,1)$. Then $Var_\alpha(w;h) = F_L^{-1}(\alpha)$. We have

$$\begin{aligned}
Pr(L \leq F_L^{-1}(\alpha)) &= \alpha \\
Pr(\mu + \sigma Z \leq F_L^{-1}(\alpha)) &= \alpha, Z \sim t_v(0,1) \\
Pr(Z \leq \frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= \alpha \\
T(\frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= \alpha \\
T^{-1} \circ T(\frac{F_L^{-1}(\alpha) - \mu}{\sigma}) &= T^{-1}(\alpha) \\
\frac{F_L^{-1}(\alpha) - \mu}{\sigma} &= T^{-1}(\alpha) \\
F_L^{-1}(\alpha) &= \mu + T^{-1}(\alpha)\sigma
\end{aligned}$$

(2) Use method of moments estimator in Lemma 3.2.2. □

25.2.2.3 *Loss with correlated normal distribution*

25.2.2.4 *Loss with approximate normal distribution*

Lemma 25.2.6 (Application of Cornish-Fisher expansion for approximate normal distribution of loss). *[3, p. 170] Suppose that the random loss L for a given horizon h has mean μ and variance σ^2. Then*

$$VaR_\alpha(L) \approx \mu + \sigma z_\alpha^{cf}$$

where

$$z_\alpha^{cf} = q_\alpha + \frac{(q_\alpha^2 - 1)S(L)}{6} + \frac{(q_\alpha^3 - 3q_\alpha)K(L)}{24} - \frac{(2q_\alpha^3 - 5q_\alpha)S^2(L)}{36},$$

where $S(L)$ is skewness, $K(L)$ is kurtosis, z_α^{cf} is the Cornish-Fisher approximate quantile value for the confidence level α, and q_α is the quantile value for the standard normal distribution with confidence level α.

Proof. Directly from Theorem 2.3.1. □

25.2.2.5 *Loss with polynomial tails*

Lemma 25.2.7. *[4, p. 566] Assume the loss L of a portfolio have a polynomial right tail such that*

$$f_L(y) \sim Ay^{-(\beta+1)}, y \to \infty,$$

where $A, \beta > 0$ are constants. β is known as ***tail index****. It follows that*

-
$$Pr(L > y) = \int_y^\infty f_L(u)du = \frac{A}{a}y^{-\beta}, y \to \infty.$$

-
$$VaR(\alpha) \triangleq F_L^{-1}(\alpha) = [\frac{(1-\alpha)\beta}{A}]^{-1/\beta}.$$

- *Consider two confidence levels α_0, α_1, we have*

$$\frac{VaR(\alpha_1)}{Var(\alpha_0)} = (\frac{(1-\alpha_1)}{1-\alpha_0})^{1/\beta}.$$

That is, if we know the tail index α, and a VaR at α_0, then we can get $VaR(\alpha_1)$.

Proof. (1)

$$\begin{aligned} Pr(L > y) &= \int_y^\infty f_L(u)du \\ &= \int_y^\infty Au^{-(\beta+1)}du \\ &= -\frac{A}{\beta}u^{-\beta}|_y^\infty \\ &= \frac{A}{\beta}y^{-\beta} \end{aligned}$$

(2)

$$\begin{aligned} F_L(L < y) = 1 - \frac{A}{\beta}y^{-\beta} = \alpha \implies \frac{A}{\beta}y^{-\beta} &= 1 - \alpha \\ y^{-\beta} &= \frac{(1-\alpha)\beta}{A} \\ VaR(\alpha) = y &= [\frac{(1-\alpha)\beta}{A}]^{-1/\beta} \end{aligned}$$

(3)

$$\frac{VaR(\alpha_0)}{VaR(\alpha_1)} = (\frac{(1-\alpha_0)}{1-\alpha_1})^{-1/\beta} = (\frac{(1-\alpha_1)}{1-\alpha_0})^{1/\beta}.$$

□

Lemma 25.2.8 (regression estimating of the tail index). *Assume the loss L of a portfolio have a polynomial right tail such that*

$$f_L(y) \sim Ay^{-(\beta+1)}, y \to \infty.$$

We know that

$$\log(Pr(L \geq y) = \log(A/\beta) - \beta \log y.$$

If $L_{(1)}, L_{(2)}, ..., L_{(n)}$ are the order statistics of losses sorted in descending order, then we can construct a set of paired data for regression $(k/n, L_k), k > m$, and use the following the regression model

$$\log(k/n) = \log(A/\beta) - \beta \log(L_{(k)}),$$

to get the slope β.

25.2.3 Historical simulation approach

Remark 25.2.4 (general approach).

- Historical simulation them involves the day-to-day changes in the values of market variables that occurred in the past in a direct way to estimate the probability distribution of the change in the value of the current portfolio between today to tomorrow.
- Historical simulation approach is the most popular non-parametric approach.

Definition 25.2.4 (scenario, market variable simulation). *[2, pp. 278, 282]*

- *Define v_i as the value of a market variable on Day i and suppose that today is day $n(i < n)$. The i scenario in the historical simulation approach assumes that the value of the market variable tomorrow will be*

$$\textit{market value under i scenario} = v_n \frac{v_i}{v_{i-1}}.$$

-

Remark 25.2.5 (weighting scenario). [2, p. 285] Suppose we have observations/scenarios for n day-to-day changes ordered in $i = 1, 2, ..., n$.

- in ordinary historical simulation, each scenario is given by $1/n$.
- in weighting historical simulation, each scenario is given by

$$\frac{\lambda^{n-i}(1-\lambda)}{1-\lambda^n}, 0 < \lambda < 1.$$

- in volatility weighting historical simulation, each scenario is given by

$$\textit{market value under i scenario} = v_n \frac{v_{i-1} + (v_i - v_{i-1})\sigma_{n+1}/\sigma_i}{v_{i-1}},$$

  where σ_{n+1} denotes the volatility of market variable between today and tomorrow.

Lemma 25.2.9 (standard deviation of percentile estimation). *[2, p. 282] Suppose that the q-percentile of the distribution is estimated as x. The standard error of the estimate is*

$$\frac{1}{f(x)}\sqrt{\frac{(1-q)q}{n}},$$

where n is the number of observations and $f(x)$ is the estimate of the probability density function of the loss evaluated at x.

25.2.3.1 *Delta-gamma approximation*

Lemma 25.2.10 (delta-gamma approximation for VaR). *[2, p. 289] Suppose an instrument depends on several market variables, $S_i, i \leq i \leq k$.*

- *the price change of the instrument is given by*

$$\Delta P = \sum_{i=1}^{k} \delta_i \Delta S_i + \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{1}{2}\gamma_{ij}\Delta S_i \Delta S_j,$$

  where δ_i and γ_{ij} are defined by

$$\delta_i = \frac{\partial P}{\partial S_i}, \gamma_{ij} = \frac{\partial^2 P}{\partial S_i \partial S_j},$$

  and ΔS_i is the market variable change.
- *Suppose for each market variable S_i, we have n historical scenarios indexed by $S_i^{(j)}, j = 1, 2, ..., n$. Then the price change under scenario m is given by*

$$\Delta P^{(m)} = \sum_{i=1}^{k} \delta_i \Delta S_i^{(m)} + \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{1}{2}\gamma_{ij} S_i^{(m)} \Delta S_i^{(m)}.$$

- *The VaR with α percentile is the α percentile among the n losses given by $-\Delta P_1, -\Delta P_2, ..., -\Delta P_n$.*

Proof. Directly from the Taylor expansion and definition of VaR. □

Example 25.2.1 (the price change of a call option). [1, p. 104] Consider a call option price, denoted by C_t. Its price change is due to the market risk can be approximated by

$$C_{t+h} - C_t \approx Delta(S_{t+h} - S_t) + \frac{1}{2} Gamma(S_{t+h} - S_t)^2 + \frac{\partial C_t}{\partial t} h + Vega(\sigma_{t+h}^{imp} - \sigma_t^{imp}).$$

Lemma 25.2.11 (ladder gamma). *Suppose a function V depends on N factors denoted by $F_1, F_2, ..., F_N$. Define*

$$D_{u,i} = \frac{V(F_{j\neq i} + \delta, F_i + 2\delta) - V(F_{j\neq i} + \delta, F_i)}{(2\delta)},$$

and

$$D_{d,i} = \frac{V(F_{j\neq i} - \delta, F_i - 2\delta) - V(F_{j\neq i} - \delta, F_i)}{(-2\delta)}.$$

It follows that

- $$\frac{D_{u,i} - D_{d,i}}{2\delta} = \sum_{j\neq i} \frac{\partial^2 V}{\partial F_j \partial F_i} + \frac{\partial^2 V}{\partial F_i^2}.$$
- $$\sum_{i=1}^{N} \frac{D_{u,i} - D_{d,i}}{2\delta} = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 V}{\partial F_j \partial F_i}$$

Proof. Use Taylor expansion, we have

$$\begin{aligned} V(F_{j\neq i} + \delta, F_i + 2\delta) = V(F) &+ \sum_{j\neq i} \frac{\partial V}{\partial F_j}\delta + \frac{\partial V}{\partial F_i} 2\delta + \frac{1}{2}\sum_{j\neq i} \frac{\partial^2 V}{\partial F_j^2}\delta^2 \\ &+ \sum_{j\neq i}\sum_{k\neq i,j} \frac{\partial^2 V}{\partial F_j \partial F_k}\delta^2 + \sum_{j\neq i} \frac{\partial^2 V}{\partial F_j \partial F_k}\delta \cdot 2\delta + \frac{1}{2}\frac{\partial^2 V}{\partial F_i^2}(2\delta)^2 \end{aligned}$$

$$\begin{aligned} V(F_{j\neq i} - \delta, F_i - 2\delta) = V(F) &- \sum_{j\neq i} \frac{\partial V}{\partial F_j}\delta - \frac{\partial V}{\partial F_i} 2\delta + \frac{1}{2}\sum_{j\neq i} \frac{\partial^2 V}{\partial F_j^2}\delta^2 \\ &+ \sum_{j\neq i}\sum_{k\neq i,j} \frac{\partial^2 V}{\partial F_j \partial F_k}\delta^2 + \sum_{j\neq i} \frac{\partial^2 V}{\partial F_j \partial F_k}\delta \cdot 2\delta + \frac{1}{2}\frac{\partial^2 V}{\partial F_i^2}(2\delta)^2 \end{aligned}$$

$$V(F_{j\neq i}+\delta, F_i) = V(F) + \sum_{j\neq i}\frac{\partial V}{\partial F_j}\delta + \frac{1}{2}\sum_{j\neq i}\frac{\partial^2 V}{\partial F_j^2}\delta^2 + \sum_{j\neq i}\sum_{k\neq i,j}\frac{\partial^2 V}{\partial F_j \partial F_k}\delta^2$$
$$V(F_{j\neq i}-\delta, F_i) = V(F) - \sum_{j\neq i}\frac{\partial V}{\partial F_j}\delta + \frac{1}{2}\sum_{j\neq i}\frac{\partial^2 V}{\partial F_j^2}\delta^2 + \sum_{j\neq i}\sum_{k\neq i,j}\frac{\partial^2 V}{\partial F_j \partial F_k}\delta^2$$

Then

$$\begin{aligned}
&\frac{D_{u,i}-D_{d,i}}{2\delta} \\
&= \frac{V(F_{j\neq i}+\delta, F_i+2\delta) - V(F_{j\neq i}+\delta, F_i)}{(2\delta)^2} + \frac{V(F_{j\neq i}-\delta, F_i-2\delta) - V(F_{j\neq i}-\delta, F_i)}{(2\delta)^2} \\
&= \frac{2\sum_{j\neq i}\frac{\partial^2 V}{\partial F_j \partial F_k}\delta\cdot 2\delta + \frac{\partial^2 V}{\partial F_i^2}(2\delta)^2}{(2\delta)^2} \\
&= \sum_{j\neq i}\frac{\partial^2 V}{\partial F_j \partial F_i} + \frac{\partial^2 V}{\partial F_i^2}
\end{aligned}$$

□

25.2.4 VaR calculation via extreme value theory

Definition 25.2.5 (conditional tail cdf). *[2, p. 290] Let L be the random loss with cdf F_L. The conditional tail cdf of L with loss threshold parameter u is defined by*

$$F_L(y;u) \triangleq Pr(L < u+y | L > u) = \frac{Pr(u < L < u+y)}{Pr(L > u)} = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

Remark 25.2.6. Note that, for a fixed parameter u, $F_L(y;u)$ is a qualified cdf since it is nondecreasing with respect to y and $F_L(\infty;u) = 1$.

Theorem 25.2.1 (approximate parametric tail distribution from Extreme value theory). *[2, p. 290]*

- *Let $u \in \mathbb{R}$ denote the loss threshold parameter. When u is large[a],*

$$Pr(L < u+y | L > u) \approx G(y;\xi,\beta) = 1 - [1+\xi\frac{y}{\beta}]^{-1/\xi},$$

where ξ is the shape parameter and β is the scale parameter.

- *The associated density function for the excess loss $Y = L - u$ is given by*

$$g(y;\xi,\beta) = \frac{1}{\beta}(1+\frac{\xi y}{\beta})^{-1/\xi-1}$$

- *The resulting parametric distribution for loss with threshold parameter u is given by*

$$\begin{aligned} Pr(L<x) &= 1-(1-G(x-u;\xi,\beta))Pr(L>u) \\ &\approx 1-(1-G(x-u;\xi,\beta))\frac{N_u}{N} \\ &= 1-\frac{N_u}{N}[1+\xi\frac{x-u}{\beta}]^{-1/\xi} \end{aligned}$$

where N is the total sample number and N_u is the number of samples with loss greater than u.

[a] by large, we usually mean at least 0.95 percentile of L

Proof. (1) Todo; (2) take derivative of G with respect to y. (3) Note that

$$\begin{aligned} Pr(L<x) &= 1-Pr(L>x) \\ &= 1-Pr(L>x|L>u)Pr(L>u) \\ &= 1-Pr(L-u+u>x|L>u)Pr(L>u) \\ &= 1-Pr(L-u>x-u|L>u)Pr(L>u) \\ &= 1-(1-G(x-u;\xi,\beta))Pr(L>u) \\ &\approx 1-(1-G(x-u;\xi,\beta))\frac{N_u}{N} \end{aligned}$$

□

Note 25.2.1 (interpretation).

- Note that $G(y;\xi,\beta)$ is a qualified cdf since it is nondecreasing with respect to y and $G(\infty;\xi,\beta) = 1$.
- **(Connection to power law)** If we set $u = \beta/\xi$, then

$$Pr(L>x) = \frac{N_u}{N}[\frac{\xi}{\beta}]^{-1/\xi},$$

which is an example of power law distribution(Definition 2.2.34).

Lemma 25.2.12 (maximum likelihood parameter estimation for parametric tail distribution). *[2, p. 290] For a given loss threshold parameter u, the likelihood function for parameter ξ, β with N_u observed losses above u, denoted by $l_1, l_2, ..., l_{N_u}$, is given by*

$$L(\xi, \beta) \prod_{i=1}^{N_u} \frac{1}{\beta}(1 + \frac{\xi(l_i - u)}{\beta})^{-1/\xi - 1}.$$

Proof. From extreme value theory(Theorem 25.2.1), the excess loss denoted by $l_1 - u, l_2 - u, ..., l_{N_u} - u$ will have density function given by

$$g(y; \xi, \beta) = \frac{1}{\beta}(1 + \frac{\xi y}{\beta})^{-1/\xi - 1}.$$

□

Theorem 25.2.2 (calculation of VaR and ES). *[2, p. 290]*

Suppose we have parameter u, β, ξ characterizing the tail distribution of the loss L. Let N be the total sample number and N_u be the number of samples with loss greater than u. Then

- *The VaR with confidence level q is given by*

$$VaR_L(q) = F_L^{-1}(q) = u + \frac{\beta}{\xi}(\frac{N}{N_u}(1-q)^{-\xi} - 1).$$

- *The ES with confidence level q is given by*

$$ES = \frac{VaR + \beta - \xi u}{1 - \xi}.$$

Proof. (1)From tail distribution (Theorem 25.2.1), we know that

$$Pr(L < x) = 1 - \frac{N_u}{N}[1 + \xi\frac{x - u}{\beta}]^{-1/\xi}.$$

With confidence q, we solve

$$Pr(L < VaR) = 1 - \frac{N_u}{N}[1 + \xi\frac{VaR - u}{\beta}]^{-1/\xi} = q,$$

and get

$$VaR_L(q) = F_L^{-1}(q) = u + \frac{\beta}{\xi}(\frac{N}{N_u}(1-q)^{-\xi} - 1).$$

(2) From the definition of ES(Definition 25.2.3), we have

$$\begin{aligned}
ES &= \frac{1}{1-q}\int_q^1 VaR(s)ds \\
&= u - \frac{\beta}{\xi} + \frac{\beta}{\xi}(\frac{N}{N_u}(1-q)^{-\xi})/(1-\xi) \\
&= VaR_L(q)/(1-\xi) + u(1-1/(1-\xi)) - \frac{\beta}{\xi}(1-1/(1-\xi)) \\
&= \frac{VaR+\beta-\xi u}{1-\xi}
\end{aligned}$$

□

25.2.5 Back-testing

Lemma 25.2.13 (VaR back-testing using binomial model with independence assumption). *[2, p. 271] Suppose that the confidence level for a one-day VaR is α. Then the VaR model predicts that the probability of the VaR being exceeded on any given day is*

$$p = 1X100.$$

Suppose that we look at a total of n days and we observe that the VaR level is exceeded on m of the days.

We consider the following hypothesis testing

- H_0 *: The probability of an exception on any given day is p.*
- H_1*: The probability of an exception on any given day is greater than p.*

We have the following decision rules: if

$$\sum_{k=m}^{n}\binom{n}{k}p^k(1-p)^{n-k} < 5\%$$

we reject H_0 and accept H_1.

Proof. We assume each day's loss exceeding VaR is independent Bernoulli random variable, then the number of exceeding days is following Binomial distribution.

By calculating the probability

$$\sum_{k=m}^{n}\binom{n}{k}p^k(1-p)^{n-k} < 5\%,$$

we are actually using p-value method to examine the probability that m out of n days' losses exceeding VaR. $\square$

25.3 VaR for different assets

25.3.1 Risk drivers for different assets

25.3.1.1 *Interest rate assets*

spot interest rate with different tenor, forward interest rate with different tensors and different maturities, volatility surface(ATM, skew, correlation) of forward interest rate with different tensors and different maturities

25.3.1.2 *Credit derivatives*

hazard curve, interest rate

25.3.1.3 *Commodity assets*

ATM implied volatility, implied volatility skew, correlations among forward contracts at different maturities, interest rate, price of forward contract at different maturities

25.3.1.4 *FX*

ATM implied volatility, implied volatility skew, correlations among underlying spots, domestic interest rate, foreign interest rate, FX spot rate

25.3.1.5 *Equity*

ATM implied volatility, implied volatility skew, correlations among underlying spots, interest rate,

25.3.2 Credit portfolios

Lemma 25.3.1 (large uniform portfolio approximation for default modeling). *[5, p. 564] Consider a bank with a very large portfolio of similar loans. As an approximation, lLet $T_1, T_2, ..., T_n$ denote the random default time for n parties. Assume the marginal cdfs* *of default time are all given by* $Q_i(t) = Q(t), i = 1, 2, ..., n$. *Let* $x_i = \phi^{-1}(Q(t)), i = 1, 2, ..., n$, *and assume* x_i *follows*

$$x_i = \sqrt{\rho}F + \sqrt{1-\rho}Z_i, i = 1, 2, ..., n,$$

where F is a common factor affecting defaults for all firms and Z_i *is a factor affecting only firm i. F and* Z_i *are independent standard normal variables. It follows that*

- *Conditioning on the value* $F = f$, *the probability of default for each loan before T years is*

$$Pr(T_i < t | F = f) = \phi(\frac{\phi^{-1}(Q(t)) - \sqrt{\rho}f}{\sqrt{1-\rho}}).$$

- ***the percentage of defaults*** *among n loans with confidence level* α *before t years on this large portfolio will be*

$$V(\alpha, t) = \phi(\frac{\phi^{-1}(Q(t)) - \sqrt{\rho}\phi^{-1}(1-\alpha)}{\sqrt{1-\rho}}),$$

where ϕ *is the standard normal cdf.*
- *The VaR for the loan size L before T years is given by*

$$VaR(\alpha) = L(1-R)V(\alpha, t).$$

Proof. (1) See Lemma 23.7.4. (2) Note that conditioning on F, each loan's default is independent from others. In the large number approximiation, the default percentage conditioning on F is also $Pr(T_i < t | F = f)$. Because $Pr(T_i < t | F = f)$ is the decreasing function of f, therefore we take f to be $1 - \alpha$ percentile. (3) Note that $L(1-R)$ is the loss at default for each loan. □

Example 25.3.1. [5, p. 565] Suppose that a bank has a large pool of outstanding loans. The 1-year marginal probability of default is average 0.02%. The average correlation among the loans is estimated to be 0.1. Then the percentage of defaults among all loans with confidence level 0.999 before 1 years is given by

$$V(0.999, t) = \phi(\frac{\phi^{-1}(0.02) - \sqrt{0.1}\phi^{-1}(1-0.999)}{\sqrt{1-0.1}}) = 0.128.$$

25.4 Notes on bibliography

The major reference are [1][6] [3].

BIBLIOGRAPHY

1. Roncalli, T. Lecture Notes on Risk Management & Financial Regulation (2016).
2. Hull, J. *Risk management and financial institutions,+ Web Site* (John Wiley & Sons, 2012).
3. Alexander, C. *Market Risk Analysis, Value at Risk Models* ISBN: 9780470745076. https://books.google.com/books?id=j5l82vMfcbQC (Wiley, 2009).
4. Ruppert, D. *Statistics and data analysis for financial engineering, 2ed* (Springer, 2015).
5. Hull, J. *Options, Futures and Other Derivatives 9 edition* (Pearson Education, 2009).
6. Jorion, P. *Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial Risk* ISBN: 9780071736923. https://books.google.com/books?id=nnblKhI7KP8C (McGraw-Hill Education, 2006).

26

XVA AND COUNTERPARTY RISK MODELING

26.1 Overview

Counterparty credit risk rises when the counterparty to a financial contract will default prior to the expiration of the contract and will not make all the payments required by the contract. CVA is to put a price for the loss due to the counterparty credit risk.

26.2 Basic concepts and remarks

26.2.1 Concepts

26.2.1.1 *Counterparty credit risk*

Definition 26.2.1 (counterparty credit risk). *[1]*

- *Counterparty credit risk is the risk that the counterparty to a financial contract will default prior to the expiration of the contract and will not make all the payments required by the contract.*
- *Only the contracts privately negotiated between counterparties- over-the-counter (OTC) derivatives and security financing transactions (SFT)-are subject to counterparty risk.*
- *Exchange-traded derivatives are not affected by counterparty risk, because the exchange guarantees the cash flows promised by the derivative to the counterparties.*

Definition 26.2.2 (approaches to mitigate counterparty risk).

- *netting: contracts which are assets are used to pay some contracts which are debts in advance of other debts.*
- *collaterlization: cash and other assets of firm used as collateral are used to pay corresponding contracts in advance of other debts.*

26.2.1.2 *Exposure and netting*

Remark 26.2.1 (how business work). [1] Suppose a counterparty risk occurs, the loss to the bank is calculated based on the following two situations:

- If the contract value is negative for the bank at the time of default, the bank has a net loss of zero. We assume the bank closes out the position by paying the defaulting counterparty the market value of the contract; and then the bank enters into a similar contract with another counterparty and receives the market value of the contract; Therefore the bank has zero net loss.
- If the contract value is positive for the bank at the time of default, we assume the bank closes out the position, but receives nothing from the defaulting counterparty; and then the bank enters into a similar contract with another counterparty and pays the market value of the contract; therefore the bank has a net loss equal to the the contract's market value.

Definition 26.2.3 (uncollateralized exposure at contract-level and counterparty-level with collateral). *[1]*

- *Let $V_i(t)$ be the value of a contract i between two parties(our own party and the counterparty) at time t. The* ***contract-level exposure*** *of our own party at time t is defined as*
$$E_i(t) = \max\{V_i(t), 0\}.$$
- *Suppose there are n trades/contracts between the two parties and all contracts cannot be netted. Then, the* ***counterparty-level exposure*** *of our own party at time t is defined as*
$$E(t) = \sum_{i=1}^{n} \max\{V_i(t), 0\}.$$
- *Suppose there are n trades/contracts between the two parties and all contracts* ***can be partitioned into*** K ***netting groups*** $NA_1, ..., NA_K$*. Then, the counterparty-level exposure of our own party at time t is defined as*
$$E(t) = \sum_{i=1}^{K} \max\{\sum_{j \in NA_i} V_j(t), 0\}.$$

Note that exposure $E(t)$ are random quantities unknown at present time.

Remark 26.2.2 (netting can reduce exposure). Note that
$$\sum_{i=1}^{K} \max\{\sum_{j \in NA_i} V_j(t), 0\} \leq \sum_{i=1}^{K} \sum_{j \in NA_i} \max\{V_j(t), 0\},$$
that is, netting can reduce exposure.

Example 26.2.1. For example, if a counterparty has an option written by the bank worthing 50, and in another contract between the counterparty and the bank, the contract has value of 80 in favor of the the bank, then a bank has exposure of 30 under netting agreement.

26.2.1.3 *Margin agreement and collateral*

Definition 26.2.4 (collteral). *[2, p. 38]*

- *In loans and lending transactions, collateral is an asset of the borrower that is transferred to the lender if the borrower defaults.*

- *For example, in a mortgage the typical collateral is the real estate property being acquired with the help of the loan.*

Definition 26.2.5 (margin agreeement concept). *[3, p. 63][1]*

- *A **margin agreement** is a legally binding contract that requires one or both counterparties to post collateral when the uncollateralized exposure exceeds a threshold and to post additional collateral if this excess grows larger.*
- ***Threshold** H: define the mark-to-market level above which collateral has be to posted. When the exposure E is above H, the collateral should be posted is $E - H$. If $E < H$, no collateral is required.*
- ***Independent amount** I: defines an amount of extra collateral that must be posted irrespective of the exposure.*
- ***Minimum transfer amount** MTA defines the minimum amount of collateral that can be called at a time.*
- ***Effective Threshold** TH defines the maximum amount of loss if counterparty defaults, given by
$$TH = H + MTA - I.$$*

Definition 26.2.6 (marginal period of risk).

- *In a marginal agreement, the collateral is monitored and called for with a fixed frequency. The period between two marginal calls is known as **call period.***
- *The time interval from the last exchange of collateral to the counterparty's default is known as **marginal period of risk**.*

Definition 26.2.7 (collaterallized exposure). *[1]*

- *The **collateral amount** $C(t)$ of the counterparty posts to us is determined by the uncollaterallized exposure at time $t - s$ against a threshold value H, given by*
$$C(t) = \max\{E(t-s) - H, 0\},$$
where s is the marginal period of risk.
- *The **collaterallized exposure** is given by*
$$E_C(t) = \max\{E(t) - C(t), 0\},$$

where

$$E(t) = \sum_{i=1}^{K} \max\{\sum_{j \in NA_i} V_j(t), 0\}.$$

Remark 26.2.3 (counterparty risk under collateral agreement)**.**

- Because of the existence of the threshold H, collateral agreements do not eliminate all counterparty risk;
- The maximum exposure under collateral agreement is given by

$$E(t) - E(t-s) + H,$$

  that is, the movement of exposure between $t-s$ and t plus the threshold value H.
- If the marginal period of risk is zero, then the maximum exposure under collateral agreement is given by H.

26.2.1.4 *Aggregation*

The aggregation is to add trade level values into a counterparty level value according to collateral rules and netting rules

Trade values within a netting group can be netted. Different collateral rules apply to netted trade values within each collateral group.

For example, for a counterparty with 8 trades with 2 netting groups and 2 collateral groups in each netting group.

- Netting group 1
  - Collateral group 1
    * Trade 1
    * Trade 2
  - Collateral group 2
    * Trade 3
    * Trade 4
- Netting group 2
  - Collateral group 3
    * Trade 5
    * Trade 6
  - Collateral group 4
    * Trade 7
    * Trade 8

Let V_{ij} be the value of the j-th trade in the i-th collateral group. Let $H_{cpty,i} > 0$ and $H_{self,i} < 0$ be the thresholds of counterparty and ourself, respectively.

The collateral amount in the i-th collateral group is

$$C_i = \underbrace{(\sum_j V_{i,j} - H_{cpty,i})^+}_{from\ counterparty\ ,\geq 0} + \underbrace{(\sum_j V_{i,j} - H_{self,i})^-}_{from\ ourself\ ,\leq 0}.$$

where $(x)^+ = \max(x,0), (x)^- = \min(x,0)$.

The aggregated collateralized value is

$$V_C = \sum_i (V_i - C_i) = \sum_i (\sum_j V_{i,j} - C_i).$$

The aggregated collateralized positive and negative exposures on the counterparty level are

$$Col\ Pos\ Expo = V_C^+ = [\sum_i (\sum_j V_{i,j} - C_i)]^+$$
$$Col\ Neg\ Expo = V_C^- = [\sum_i (\sum_j V_{i,j} - C_i)]^-$$

26.2.2 Business and purpose

26.2.2.1 *Motivation*

Remark 26.2.4 (motivation for calculating counterparty credit exposure). [4, p. 10]For financial institutions, calculating counterparty credit exposure has the following purpose:

- Setting limits on the amount of business allowed with a particular counterparty.
- Dynamic hedging of counterparty risk, by buying credit protection on the counterparty. This in effect allows one to trade away counterparty credit risk.
- Computation of risk weighted assets and capital requirements.
- Obtaining insight about prices of complex transactions in potential future scenarios. For example, while counterparty risk is concerned with measuring how high the value of a transaction can go (and therefore how much a counterpartywould owe), there are similarities between this and computing Value at Risk, or stress testing, where one would be interested in how much the value of a transaction could drop.

26.2.2.2 *Characteristics for different asset classes*

Remark 26.2.5.

- Equity derivatives tend to have small exposure to counterparty default risk since most trades are often short-dated and cash collateralized.
- Interest derivatives (swaps, caps/floors) tend to have large exposure to counterparty default risk since most trades are often long-dated.
- Cash instruments (e.g., bonds) usually do not need to calculate CVA/DVA, because their prices already contain the valuation adjustments.

26.2.2.3 *The business*

Remark 26.2.6 (how CVA trading desk make money)**.**

26.3 CVA/DVA model

26.3.1 Risk neural pricing framework

Definition 26.3.1 (credit valuation adjustment (CVA), debit valuation adjustment (DVA)).

- *Consider a contract between our own party and a counterparty and let V denote its default value. The adjustment of current value $V(t)$ due to possible default event of the counterparty(occurring particularly when $V > 0$) is called* ***credit valuation adjustment(CVA)****. The adjustment of current value $V(t)$ due to possible default event of our own party (occurring particularly when $V < 0$) is called* ***debit valuation adjustment(DVA)****.*
- *(complete formulation) bilateral CVA: both parties are considered risky and face exposures depending on the value of the positions they hold against each other.*

$$\begin{aligned} V(t) = & \underbrace{E_Q[\frac{M(t)}{M(T)}V(T)\mathbf{1}(\tau_{CP} > T)]}_{value\ when\ no\ default} \\ & + \underbrace{(1-R_{CP})E_Q[\frac{M(t)}{M(T)}V(T)\mathbf{1}(t \leq \tau_{CP} \leq T, \tau_{CP} \leq \tau_{SP})]}_{CVA,\ value\ lost\ due\ to\ counterparty\ default} \\ & + \underbrace{(1-R_{SP})E_Q[\frac{M(t)}{M(T)}V(T)\mathbf{1}(t \leq \tau_{SP} \leq T, \tau_{CP} \geq \tau_{SP})]}_{DVA,\ counterparty\ value\ lost\ due\ to\ counterpartydefault} \end{aligned}$$

- *unilateral CVA: one party (the investor) is considered default-free and only the exposure to the counterparty matters.*

$$CVA(t) = 1 - R_{CP})E_Q[\frac{M(t)}{M(T)}V(T)\mathbf{1}(t \leq \tau_{CP} \leq T)]$$

Theorem 26.3.1 (general risk-neutral pricing framework for CVA/DVA). *[1] Let $V(u), u < T$ denote the default-free contract(maturing at T) value at time u with respect to our own party(represented using subscript SP), which is generally given by*

$$V(u) = B(t)E_Q[\frac{V(T)}{B(T)}|\mathcal{F}_u].$$

Let τ denote the default time. Assume the independence among V and τ. Let current time be t.

- $$\begin{aligned} CVA(t) &= (1 - R_{CP}) E_Q[DF(t,\tau) \cdot V^+(\tau) \cdot \mathbf{1}_{t \leq \tau_{CP} \leq T}] \\ &= (1 - R_{CP}) \int_t^T EPE(\tau) dF_{CP}(u) \end{aligned}$$

  where $DF(t,\tau)$ is the (stocastic) discount factor $DF(t,\tau) = \exp(-\int_t^\tau r(s)ds)$ discounting cash flow from τ to t,$V^+(u) = \max(0, V(u))$, $F_{CP}(u) \triangleq Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event. The $EPE(u)$ is the discounted positive exposure of the trades at time u given by

  $$EPE(u) = M(t) E_Q[\frac{V^+(u)}{M(\tau)} | \mathcal{F}_t].$$

- $$\begin{aligned} DVA(t) &= -(1 - R_{SP}) E_Q[DF(t,\tau) \cdot V^-(\tau) \cdot \mathbf{1}_{t \leq \tau_{SP} \leq T}] \\ &= -(1 - R_{SP}) \int_t^T ENE(u) dF_{SP}(u) \end{aligned}$$

  where $V^-(u) = \max(0, -V(u))$, $F_{SP}(u) \triangleq Pr(\tau_{SP} \leq u)$ is the marginal distribution function for our own party default time τ_{SP}, R_{SP} is the deterministic recovery rate constant from the default event. The $ENE(u)$ is the discounted positive exposure of the trades at time u given by

  $$ENE(u) = M(t) E_Q[\frac{V^-(u)}{M(\tau)} | \mathcal{F}_t].$$

Remark 26.3.1 (practical evaluation strategy)**.**

- The risk-neutral default probability can be derived either from credit default swaps(if they are liquid) or from the credit spreads in the bonds issued by the counterparty. Specifically, there are three types of counterparties:
  - **Liquid counterparties** refer to counterparties whose name is liquidly traded in the CDS market(see 23.3.2).
  - **Semi-liquid counterparties** refer to the counterparties whose bonds are liquidly traded ly the bond market therefore a default probability can be inferred from the bond credit spread(see Lemma 23.3.7).
  - **illiquid counterparties** refers to counterparties that do not fall into the previous two categories.

- If we have the default hazard curve, we can evaluate $dF_{SP}(u)$.
- If we have the calibrated dynamic model for V under risk-neutral measure(or other convenient measure), we can evaluate $EPE(t), NPE(t)$.

Remark 26.3.2 (exposure and margin agreement). In the risk-neutral pricing framework, the EPE and NPE calculation is dependent on the netting, collateralization, and margin agreement.

- If there is margin agreement, then V^+, V^- should be the collaterialized exposure given in Definition 26.2.7.
- If there is no margin agreement, then V^+, V^- should be the uncollaterialized exposure given in Definition 26.2.3.

Note 26.3.1 (Calculation of CVA using conditional default probabilities). Let τ_1 and τ_2 denote the default times for our party and the counterparty.

- In the calculation of CVA, the essential part is calculate the exposure $V^+(t)$ in time t and then weighted with the default probability $Pr(\tau_2 = t)$ of the counterparty.
- An alternative way to calculate the default probability is using

$$Pr(\tau_2 = t, \tau_1 > \tau_2),$$

  because if our party default first, then we will not have the loss due to counterparty default.
- If τ_1 and τ_2 are independent, then

$$Pr(\tau_2 = t, \tau_1 > \tau_2) = f_{\tau_2}(t)S_2(t),$$

  where f is the marginal density and S is the survival probability.
- In general, the conditional default approach will have a smaller CVA since $S_2(t) < 1$.

26.3.2 Analytical approach

Lemma 26.3.1 (normal trade value model with exposure independent of counterparty credit quality). [5] *Assume the value $V_i(t)$ of trade $i, i = 1, 2, ..., N$ at each future time t, under risk-neutral Q, is given by*

$$V_i(t) = \mu_i(t) + \sigma_i(t)Z_i, Z_i \sim N(0, 1).$$

where $Z_1, Z_2, ..., Z_n$ are joint normal with $E[Z_iZ_j] = \rho_{ij}$. Assume the exposure is independent of counterparty's credit quality.

- *The portfolio consists of $V(t) = \sum_{i=1}^N V_i(t)$ is a normal random variable with mean and variance given by*

$$\mu(t) = \sum_{i=1}^N \mu_i, \sigma^2(t) = \sum_{i=1}^N \sum_{j=1}^N \rho_{ij}\sigma_i(t)\sigma_j(t)$$

- *In presence of margin agreement with instanstaneous collateral and threshod H, the expected exposure is given by*

$$\begin{aligned} e(t) &\triangleq E_Q[V(t)\mathbf{1}(0 < V(t) < H)] + H \times E_Q[\mathbf{1}(V(t) > H)] \\ &= \mu(t)[N(\frac{\mu(t)}{\sigma(t)}) - N(\frac{\mu(t) - H}{\sigma(t)})] \\ &+ \sigma(t)[\phi(\frac{\mu(t)}{\sigma(t)}) - \phi(\frac{\mu(t) - H}{\sigma(t)})] + HN(\frac{\mu(t) - H}{\sigma(t)}) \end{aligned}$$

  where $N(x)$ and $\phi(x)$ are the cdf and pdf of standard normal distribution.
- *If there is no margin agreement, we have the expected exposure given by*

$$\begin{aligned} e(t) &\triangleq E_Q[V(t)\mathbf{1}(V(t) > 0)] \\ &= \mu(t)[N(\frac{\mu(t)}{\sigma(t)})] + \sigma(t)[\phi(\frac{\mu(t)}{\sigma(t)})] \end{aligned}$$

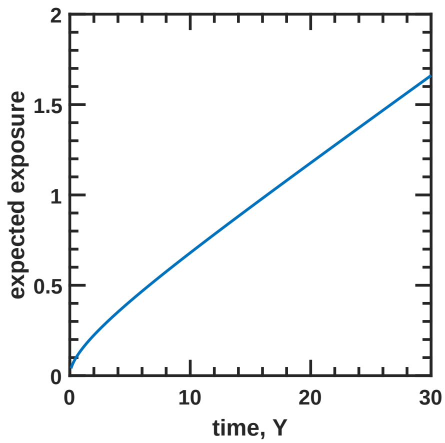
- *The CVA is given by*

$$CVA = (1 - R_{CP}) \int_0^T DF(0,t)e(t)dF_\tau(t).$$

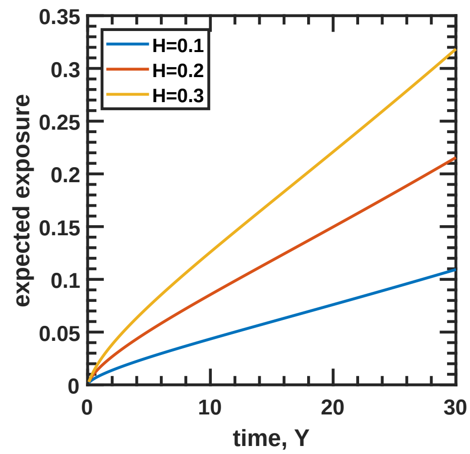
Proof. (1) straight forward. (2)

$$\begin{aligned} e(t) &= E_Q[V(t)\mathbf{1}(0 < V(t) < H)] + H \times E_Q[\mathbf{1}(V(t) > H)] \\ &= E_Q[\mu(t) + \sigma(t)Z\mathbf{1}(0 < \mu(t) + \sigma(t)Z < H)] + H \times E_Q[\mathbf{1}(\mu(t) + \sigma(t)Z > H)] \\ &= \int_{-\frac{\mu(t)}{\sigma(t)}}^{\frac{H-\mu(t)}{\sigma(t)}} [\mu(t) + \sigma(t)x]\phi(x)dx + H\int_{\frac{H-\mu(t)}{\sigma(t)}}^{\infty} \phi(x)dx \\ &= \mu(t)[N(\frac{\mu(t)}{\sigma(t)}) - N(\frac{\mu(t) - H}{\sigma(t)})] \\ &+ \sigma(t)[\phi(\frac{\mu(t)}{\sigma(t)}) - \phi(\frac{\mu(t) - H}{\sigma(t)})] + HN(\frac{\mu(t) - H}{\sigma(t)}) \end{aligned}$$

(3) Use (2) and take $H \to \infty$. (4) The definition of CVA. □



(a) Expected positive exposure without margin agreement.



(b) Expected positive positive exposure with instantaneous collateral and different threshold H.

Figure 26.3.1: Analytical exposure calculation with normal approximation. The trade value is characterized by $V(t) = 0.05t + 0.3\sqrt{t}Z, Z \in N(0,1)$.

26.3.3 Monte carlo appraoch

Methodology 26.3.1 (single trade level discrete evaluation via Monte Carlo).

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T]$.*
- *Generate M market scenarios (interest rate, FX, equity, futures, etc.) for each time point T_i.*
- *For each simulation time point T_i and scarario j: calculate the trade value $V^{(j)}(T_i)$, the collateral $C^{(j)}(T_i)$, and the collaterialized exposure $E_C^{(j)}(T_i) = \max(V^{(j)}(T_i) - C^{(j)}(T_i), 0)$.*
- *Compuate the discounted expected positive exposure at T_i by*

$$E\hat{P}E(T_i) = \frac{1}{M}\sum_{j=1}^{M} DF^{(j)}(t, T_i)E_C^{(j)}(T_i)$$

- *The CVA calculation can be approximated using the following discrete-time form:*

$$CVA(t) = [(1 - R_{CP})\sum_{i=1}^{N} E\hat{P}E(T_i)Pr(T_{i-1} \leq \tau_{CP} \leq T_i).$$

Remark 26.3.3 (interpretation). With the simulated market risk factors, we can evaluate the trade value based on instrument types.

- Consider an European option on the stock S_t. By generating a sample S_{T_i}, we can use Black-Scholes to evaluate $V(T_i; S_{T_i})$.
- Consider a forward rate agreement with maturity date T, whose value depends on the zero rate curve. By generating a zero rate curve from T_i to T, we can evaluate the contract value at T_i.
- Consider an interest rate swap, whose value depends on the zero rate curve. By generating a zero rate curve from T_i to the last payment date, we can evaluate the contract value at T_i.

Methodology 26.3.2 (Counterparty level discrete evaluation via Monte Carlo). *Consider a couterparty associated with K trades with trade value denoted by $V_1, V_2, ..., V_K$ with H netting groups.*

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T]$.*
- *Generate M market scenarios (interest rate, FX, equity, futures, etc.) for each time point T_i.*
- *For each simulation time point T_i and scarario j:*
  - *For each trade $k = 1, 2, ..., K$, calculate the trade value $V_k^{(j)}(T_i)$.*
  - *Calculate the uncollaterillized exposure via*

$$E^{(j)}(T_i) = \sum_{i=1}^{H} \max\{\sum_{h \in NA_i} V_h^{(j)}(T_i), 0\}.$$

  - *Calculate the the collateral $C^{(j)}(T_i)$, and the collaterialized exposure $E_C^{(j)}(T_i) = \max(E^{(j)}(T_i) - C^{(j)}(T_i), 0)$.*
- *Compuate the discounted expected positive exposure at T_i by*

$$E\hat{P}E(T_i) = \frac{1}{M} \sum_{j=1}^{M} DF^{(j)}(t, T_i) E_C^{(j)}(T_i)$$

- *The CVA calculation can be approximated using the following discrete-time form:*

$$CVA(t) = [(1 - R_{CP}) \sum_{i=1}^{N} E\hat{P}E(T_i) Pr(T_{i-1} \leq \tau_{CP} \leq T_i).$$

26.4 Single trade CVA calculation

26.4.1 Basic products

26.4.1.1 *Forward agreement*

Lemma 26.4.1 (CVA for single forward rate agreement). *[2, p. 98] Let current time be t. Let $C(t;s,K)$ denote the time t price of a call option with maturity s and strike K such that*

$$C(t;s,K) = E_Q[\frac{M(t)}{M(s)}(S(s)-K)^+|\mathcal{F}_t].$$

Consider a forward agreement with expiry T initialized at time t with strike price $F(t,T)$. Denote the underlying spot by S_t. It follows that

- *The evoluation of the forward agreement value(a random variable) is given by*

$$V(s) = S_s - F(t,T)\exp(-r(T-s)).$$

- *The CVA (without marginal agreement)for the forward agreement is given by*

$$CVA(t) = (1-R_{CP})\int_t^T C(t;s,F(t,T)\exp(-r(T-s)))dF_{CP}(s)$$

  where $F_{CP}(u) \triangleq Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event.

Proof. (1) For the forward price evoluation, see Theorem 15.7.3. (2) From the definition of CVA, we have

$$\begin{aligned} CVA(t) &= E_Q[DF(t,\tau)\cdot V^+(\tau)\cdot(1-R_{CP})\mathbf{1}_{t\leq\tau_{CP}\leq T}] \\ &= \int_t^T [(1-R_{CP})PV_t[V^+(s)]]dF_{CP}(s)du \\ &= (1-R_{CP})\int_t^T C(t;s,F(t,T)\exp(-r(T-s)))dF_{CP}(s) \end{aligned}$$

□

26.4.1.2 *Option*

Lemma 26.4.2 (CVA/DVA for vanilla option). *Let current time be 0. Consider a call option with payoff $V(T) = (S_T - K)^+$ with maturity date T, where S_T is the underlying and K is the strike. Assume constant interest rate r. Then, the CVA for this call option is given by*

- *The discounted mean positive exposure is $DPE(t) = V(0)$; the discounted mean negative exposure is 0.*
- $$CVA = (1 - R_{CP})V(0)F_{CP}(T).$$
- $DVA = 0$

where $V(0)$ is the current value of the call option, $F_{CP}(u) \triangleq Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event.

Proof. (1) From the risk-neutral pricing framework(Theorem 26.3.1), we have

$$\begin{aligned} DPE(t) &= \exp(-rt)E_Q[E_Q[V(t)|\mathcal{F}_u]|\mathcal{F}_0] \\ &= \exp(-rt)E_Q[V(t)|\mathcal{F}_0] \\ &= E_Q[\exp(-rt)V(t)|\mathcal{F}_0] \\ &= V(0) \end{aligned}$$

(2) From the risk-neutral pricing framework(Theorem 26.3.1), we have

$$\begin{aligned} CVA &= (1 - R_{CP})\int_0^T \exp(-ru)E_Q[\exp(-r(T-u))E_Q[V(T)|\mathcal{F}_u]|\mathcal{F}_0]\frac{dF_{CP}(u)}{du}du \\ &= (1 - R_{CP})\int_0^T \exp(-ru)E_Q[\exp(-r(T-u))V(T)|\mathcal{F}_0]\frac{dF_{CP}(u)}{du}du \\ &= (1 - R_{CP})\int_0^T \exp(-rT)E_Q[V(T)|\mathcal{F}_0]\frac{dF_{CP}(u)}{du}du \\ &= (1 - R_{CP})V(0)\int_0^T \frac{dF_{CP}(u)}{du}du \\ &= (1 - R_{CP})V(0)F_{CP}(T) \end{aligned}$$

(2) DVA is zero is because the negative exposure is always zero. □

26.4.2 Interesst rate products

26.4.2.1 *Interest rate swap*

Lemma 26.4.3 (Analytical approach for CVA for single interest rate swap without collateralization). *[2, p. 98] Take a unit notional amount and a set of dates $t < T_0 < T_1 < \cdots < T_n < \cdots < T_{m+n}$ with accrual periods $\tau_i = T_i - T_{i-1}$. Let $PS(t; s, T_{\alpha(s)}, T_n, K)$ denote the the interest rate swaption price at time t with maturity s,strike K and a underlying swap spanning payment dates $T_{\alpha(s)}, T_{\alpha(s)+1}, ..., T_n$, where $\alpha(s)$ is the minimum integer such that $T_\alpha \geq s$.* [a] .

Then, the CVA for an interest rate swap is given by

$$CVA(t) = (1 - R_{CP}) \int_t^T PS(t; u, T_{\alpha(u)}, T_n, K) dF_{CP}(u),$$

where $F_{CP}(u) \triangleq Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event.

If the default hazard curve $h(t) = h_0$ is flat, then

$$CVA(t) = (1 - R_{CP}) \int_t^T PS(t; u, T_{\alpha(u)}, T_n, K) h_0 S(t, u) du,$$

where $S(t, u) = \exp(-\int_t^u h_0 ds) = \exp(-h_0(u - t))$.

a a swaption's maturity date should be before the first reset date of the underlying swap

Proof. From the value evolution of interest rate (Lemma 19.3.9), the swaption value is given by

$$PS(t; s, T_{\alpha(s)}, T_n, K) = A_{\alpha(s),n}(t) E_{A_{\alpha(s),n}}[(S_{\alpha(s),n}(s) - K)^+ | \mathcal{F}_t],$$

where $K = S_{0,n}(t)$.

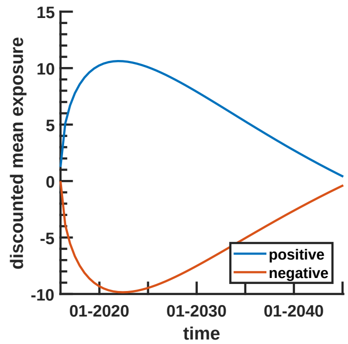
See the default probability properties, Lemma 23.2.1. □

Remark 26.4.1 (interpretation).

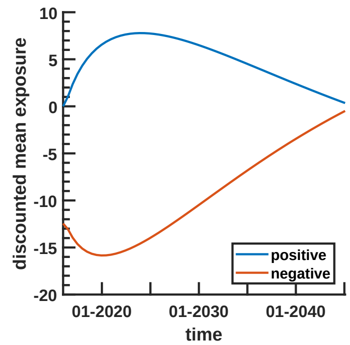
- There is only one interest rate, but we decompose its exposure calculation from current time t to its final payment date into infinitely many swaptions.
- The rate K is the fair swap rate $S(t)$ set at the time swap contract is entered; that is, $K = S(t)$. Because $S(t)$ will change with t, the value $V(t)$ of the contract will change with t.
- There are two points where $V(t)$ is known to be zero. At time t where both parties enter the contract and set the par swap rate; at time T_b, the last payment date, when there is no further cash flow exchange any more.

Example 26.4.1. Figure 26.4.1 shows the analytical calculation result of the expected exposure for an interest rate swap. We have the following observations:

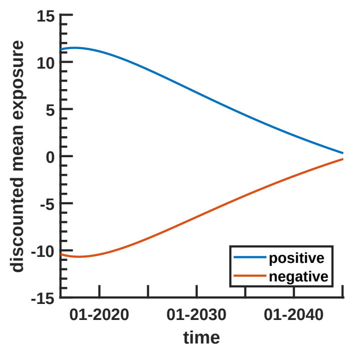
- All the exposure will converge to zero at the end of the swap.
- For spot-starting swaps with par rate, the initial exposures are zero; for spot-starting swaps with unfair rate, the exposure for one party will be zero whereas the exposure for anther party will be the present value of the swap.
- For forward-starting swaps with par rate, the initial exposures are non-zero buy symmetric; for forward-starting swaps with unfair rate, the exposures for the two parties will be unsymmetric and the difference of the exposure will be the present value of the swap.



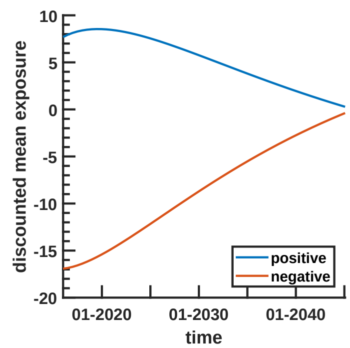
(a) Expected positive and negative exposure for a spot-starting swap with fair strike.



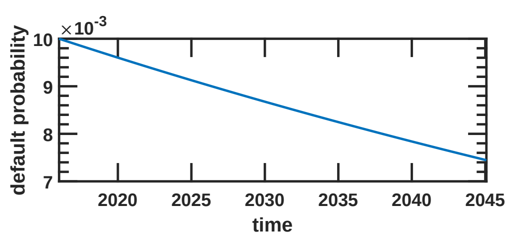
(b) Expected positive and negative exposure for a spot-starting swap with unfair strike.



(c) Expected positive and negative exposure for a forward-starting swap with fair strike.



(d) Expected positive and negative exposure for a forward-starting swap with unfair strike.



(e) Cumulative default probability.

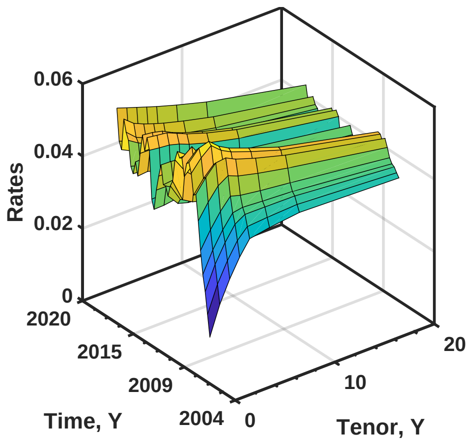
Figure 26.4.1: Analytical exposure calculation of a swap with underlying tenor from 2018 to 2048.

Methodology 26.4.1 (MC simulation method for a single swap CVA). *Suppose*

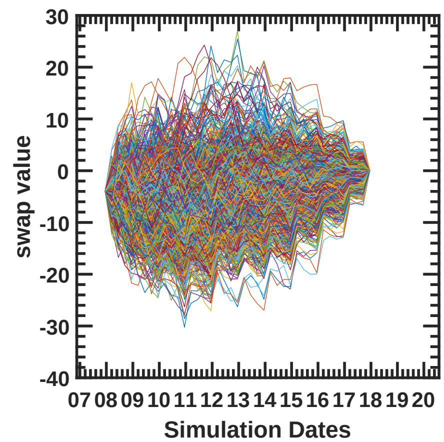
- *The current time be t and the ending time of the underlying swap be T_f.*
- *We are given a calibrated short rate model.*

We can use the following procedure to calculate the CVA of the underlying swap

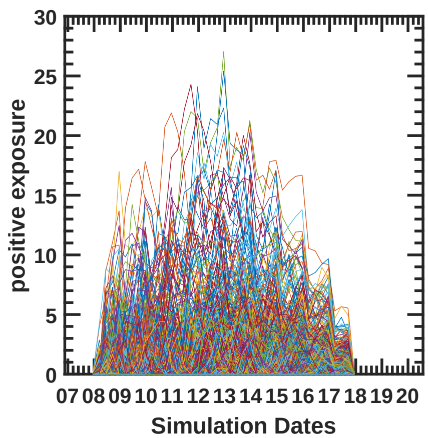
- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T_f]$.*
- *Simulate M short rate trajectories $r^{(j)}(t), j = 1, 2, ..., M$ at $T_i, i = 1, 2, ..., N$ such that we can evaluate the future realization of the zero curve $P(T_i, T), T > T_i$ via*
$$P(T_i, T) = A(T_i, T) \exp(-B(T_i, T) r^{(j)}(T_i)).$$
- *The zero curves enables us to evaluate the swap value at future time $V^{(j)}(T_i), j = 1, 2, ..., M$ and its positive exposure $[V^+]^{(j)}(T_i)$.*
- *The mean positive exposure at time T_i can be estimated by*
$$MPE(T_i) = \frac{1}{M}\sum_{j=1}^{M}[V^+]^{(j)}(T_i).$$
- *The CVA calculation can be approximated using the following discrete-time form:*
$$CVA(t) = [(1 - R_{CP})\sum_{i=1}^{N} DF(t, T_i) MPE(T_i) Pr(T_{i-1} \leq \tau_{CP} \leq T_i) du.$$



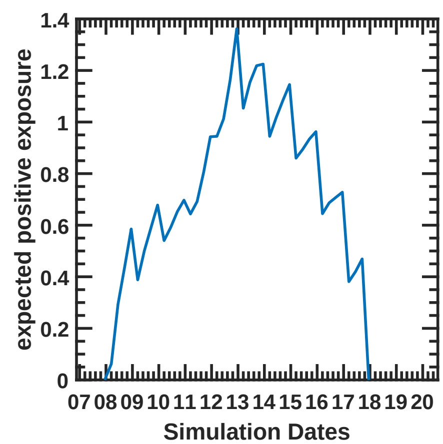
(a) Evoluation of yield curve(one scenario).



(b) Swap value trajectories.



(c) Postive exposure trajectories.



(d) Expected positive exposure until maturity.

Figure 26.4.2: Simulation exposure calculation of a swap with underlying tenor from 2008 to 2018.

Lemma 26.4.4 (Analytical approximation for CVA for single interest rate swap with collateralization). *[2, p. 98] Take a unit notional amount and a set of dates* $t < T_0 < T_1 < \cdots < T_n < \cdots < T_{m+n}$ *with accrual periods* $\tau_i = T_i - T_{i-1}$. *Let* $PS(t; s, T_{\alpha(s)}, T_n, K)$ *denote the the interest rate swaption price at time t with maturity s,strike K and a underlying swap spanning payment dates* $T_{\alpha(s)}, T_{\alpha(s)+1}, ..., T_n$, *where* $\alpha(s)$ *is the minimum integer such that* $T_\alpha \geq s$. [a].

It follows that

- *Let $V(t)$ be the value (a random variable) of the swap at time t. Let the collateral amount be $V(t-\delta t)$, where δt is the marginal period of risk. Then the undiscounted expected exposure at time t is given by*

$$EE(t) \triangleq E_A[V(t) - V(t-\delta t)|\mathcal{F}_t] \approx 0.$$

- $$EPE(t) \triangleq E_A[V(t) - V(t-\delta t)^+] \approx A(0)\sigma_S(0,t,T_n)\sqrt{\delta t}\frac{1}{2\pi}.$$

Then, the CVA for an interest rate swap is given by

$$CVA(t) = (1-R_{CP})\int_t^T PS(t;u,T_{\alpha(u)},T_n,K)dF_{CP}(u),$$

where $F_{CP}(u) \triangleq Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event.

If the default hazard curve $h(t) = h_0$ is flat, then

$$CVA(t) = (1-R_{CP})\int_t^T PS(t;u,T_{\alpha(u)},T_n,K)h_0 S(t,u)du,$$

where $S(t,u) = \exp(-\int_t^u h_0 ds) = \exp(-h_0(u-t))$.

[a] a swaption's maturity date should be before the first reset date of the underlying swap

Proof. From the value evolution of interest rate (Lemma 19.3.9), the swaption value is given by

$$PS(t;s,T_{\alpha(s)},T_n,K) = A_{\alpha(s),n}(t)E_{A_{\alpha(s),n}}[(S_{\alpha(s),n}(s)-K)^+|\mathcal{F}_t],$$

where $K = S_{0,n}(t)$.

See the default probability properties, Lemma 23.2.1. □

26.4.2.2 *Bond*

Lemma 26.4.5. *Consider a zero-coupon bond maturing at time T. Assume independence between the interest rate and the issuer's credit quality. Then the CVA for this bond is*

$$CVA = P(0,T)(1-R_{CP})Pr(\tau_{CP} < T),$$

where $P(0,T)$ is the default-free zero-coupon bond price, $Pr(\tau_{CP} \leq u)$ is the marginal distribution function for the counterparty default time τ_{CP}, R_{CP} is the deterministic recovery rate constant from the default event. That is, the bond is worth $P(0,T) - CVA$.

Proof.

$$\begin{aligned}
CVA &= (1-R)\int_0^T E_Q[\exp(-\int_0^t r(s)ds)E_Q[\exp(-\int_0^t r(s)ds)|\mathcal{F}_t]]dF_\tau(t) \\
&= (1-R)\int_0^T E_Q[\exp(-\int_0^t r(s)ds)E_Q[\exp(-\int_t^T r(s)ds)|\mathcal{F}_t]]dF_\tau(t) \\
&= (1-R)\int_0^T E_Q[E_Q[\exp(-\int_0^t r(s)ds)\exp(-\int_t^T r(s)ds)|\mathcal{F}_t]]dF_\tau(t) \\
&= (1-R)\int_0^T E_Q[\exp(-\int_0^T r(s)ds)]dF_\tau(t) \\
&= (1-R)P(0,T)\int_0^T dF_\tau(t) \\
&= (1-R)P(0,T)Pr(\tau_{CP} \leq T)
\end{aligned}$$

□

26.4.3 Credit products

26.4.3.1 *CDS*

Methodology 26.4.2 (CVA/DVA for CDS). *Assume we have*

- *the market observed non-deterministic survival curves for our party $S_1(T)$ and the counterparty $S_2(T)$, and the reference entity $S_3(T)$.*
- *a calibrated stochastic hazard rate model for the reference entity*

$$dh(t) = \mu_1(t)dt + \sigma_1(t)dW_1(t),$$

such that the survival probability for the reference party is given by

$$S_h(t,T) = A_h(t,T)\exp(-B_h(t,T)h(t)).$$

- *a short rate model*

$$dr(t) = \mu_r(t)dt + \sigma_r(t)dW_2(t)$$

such that the discount curve $DF(t,T)$ is given by

$$DF(t,T) = A_r(t,T)\exp(-B_r(t,T)r(t)).$$

Let the current time be 0. Let the CDS contract maturity time be T_n. We discrete the time horizon by $0 \leq t_1 \leq t_2 < \cdots < T_n$.

The Monte Carlo procedure to calculate the CVA is given by

- *We simulate N the hazard rate $h(t)$ trajectories and short rate trajectories r from $t = 0$ to $t = T$.*
- *For each trajectory j, at time t_i, we evaluate the a sampled survival curve $S_h^{(j)}(t_i, T), T > t_i$ using the sampled value $h(t_i)^{(j)}$ and the sample discount factor curve $DF^j(t_i, T), T > t_i$ using the sampled value $r(t_i)^{(j)}$.*
- *The sampled survival curve $S_r^{(j)}(t_i, T), T > t_i$ and discount factor curve $DF^j(t_i, T), T > t_i$ allow us to evaluate the a sample CDS value $V^{(j)}$ at time t.*
- *The positive exposure at time t is then given by $\frac{1}{N}\sum_{j=1}^{N} \max(V^{(j)}, 0)$.*

26.4.4 Foreign exchange

26.4.4.1 *Cross-currency swap*

Methodology 26.4.3 (MC simulation method for a cross-currency swap CVA). *[2, p. 206] Suppose*

- *The current time be t and the ending time of the underlying swap be T_f.*
- *We are given a calibrated (risk neutral)short rate model for domestic interest rate r_d and foreign interest rate r_f(under its own risk-neutral measures).*
- *We are given a FX spot rate at current time.*

We can use the following procedure to calculate the CVA of the underlying swap

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T_f]$.*
- *Simulate M short rate and FX spot rate trajectories $r_d^{(j)}(t), r_f^{(j)}(t), Q^{(j)}(t), j = 1, 2, ..., M$ at $T_i, i = 1, 2, ..., N$ such that we can evaluate the future realization of the zero curve $P^f(T_i, T), P^d(T_i, T), T > T_i$ via*

$$P^f(T_i, T) = A^f(T_i, T)\exp(-B^f(T_i, T)r_f^{(j)}(T_i)), P^d(T_i, T) = A^d(T_i, T)\exp(-B^d(T_i, T)r_d^{(j)}(T_i)).$$

- *The zero curves and FX spot rate trajectories enable us to evaluate the swap value at future time $V^{(j)}(T_i), j = 1, 2, ..., M$ and its positive exposure $[V^+]^{(j)}(T_i)$ using Lemma 22.6.5.*
- *The mean positive exposure at time T_i can be estimated by*

$$MPE(T_i) = \frac{1}{M}\sum_{j=1}^{M}[V^+]^{(j)}(T_i).$$

- *The CVA calculation can be approximated using the following discrete-time form:*

$$CVA(t) = [(1 - R_{CP}) \sum_{i=1}^{N} DF(t, T_i) MPE(T_i) Pr(T_{i-1} \leq \tau_{CP} \leq T_i) du.$$

Methodology 26.4.4 (MC simulation method for a mark-to-market cross-currency swap CVA). *Suppose*

- *The current time be t and the ending time of the underlying swap be T_f.*
- *We are given a calibrated (risk neutral)short rate model for domestic interest rate r_d, and foreign interest rate r_f, and the exchange rate $Q(t)$. All models are given by Theorem 22.4.1.*

We can use the following procedure to calculate the CVA of the underlying swap

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T_f]$.*
- *Simulate M short rate and FX spot rate trajectories $r_d^{(j)}(t), r_f^{(j)}(t), Q^{(j)}(t), j = 1, 2, ..., M$ at $T_i, i = 1, 2, ..., N$ such that we can evaluate the future realization of the zero curves and FX forward rate agreement $P^f(T_i, T), P^d(T_i, T), T > T_i$ via*

$$\begin{aligned}
P^f(T_i, T) &= A^f(T_i, T) \exp(-B^f(T_i, T) r_f^{(j)}(T_i)) \\
P^d(T_i, T) &= A^d(T_i, T) \exp(-B^d(T_i, T) r_d^{(j)}(T_i)) \\
F_X(T_i, T) &= Q(T) P^f(T_i, T) / P^d(T_i, T).
\end{aligned}$$

- *The zero curves and FX forward curve trajectories enable us to evaluate the swap value at future time $V^{(j)}(T_i), j = 1, 2, ..., M$ and its positive exposure $[V^+]^{(j)}(T_i)$ using Lemma 22.6.6.*
- *The mean positive exposure at time T_i can be estimated by*

$$MPE(T_i) = \frac{1}{M} \sum_{j=1}^{M} [V^+]^{(j)}(T_i).$$

- *The CVA calculation can be approximated using the following discrete-time form:*

$$CVA(t) = [(1 - R_{CP}) \sum_{i=1}^{N} DF(t, T_i) MPE(T_i) Pr(T_{i-1} \leq \tau_{CP} \leq T_i) du.$$

26.5 Wrong/right way risk model

26.5.1 Concepts

Definition 26.5.1 (wrong-way risk, right-way risk). *[3, p. 309]*

- ***Wrong-way risk, WWR*** *indicates a detrimental relationship between exposure and default probability that actually reduces counterparty risk. WWR is the risk that occurs exposure to a counterparty is adversely correlated with the credit quality of that counterparty. It arises when default risk and credit exposure increase together.*
- ***Right-way risk*** *indicates a beneficial relationship between exposure and default probability that actually reduces counterparty risk. It occurs when the exposure decreases as the default probability increases.*

Definition 26.5.2 (specific WWR and general WWR).

- ***Specific WWR*** *arises due to counterparty specific factors: a rating downgrade, poor earnings or litigation.*
- ***General WWR*** *occurs when the trade position is affected by macroeconomic factors: interest rates, inflation, political tension in a particular region.*

Example 26.5.1 (collaterlized loan). Bank A enters into a collateralized loan with Bank B (the counterparty). The collateral that Bank B provides to A can be of different nature.

- bonds issued by Bank B (specific **WWR**)
- bonds issued by a different issuer belonging to a similar industry, or the same country or geographical region (general **WWR**). This kind of risk is both difficult to detect in the trading book, hard to measure and complex to resolve.

Example 26.5.2 (right way risk).

- An airline usually protects itself against a rise in fuel prices by entering into long oil derivative contracts.
- A company would normally issue calls and not puts on its stock.
- (put option). Buying a put option on a stock where the underlying company is more likely to default when its stock value drops(and the put option becomes more valuable).

26.5.2 Analytical modeling method

Lemma 26.5.1 (copula and normal approximation approach for CVA with WWR). *[5] Define the proxy standard random variable(**??**) $Y = \phi^{-1}(F_\tau(\tau))$ for the counterparty default time τ. Assume the trade values $V_i, i = 1, 2, ..., N$ are given by $V_i(t) = \mu_i(t) + \sigma_i(t)X_i, X_i \sim N(0,1)$.*

Assume $(X_1, X_2, ..., X_N, Y)$ are jointly multivariate Gaussian with zero mean and covariance/correlation Σ.

It follows that

- *Conditioned on the default time $\tau = t$, or equivalently $Y = y(t) = \phi^{-1}(F_\tau(t))$, $X = (X_1, X_2, ..., X_n)$ are jointly multivariate Gaussian with mean and covariance*

$$\mu_{X|Y} = \Sigma_{XY}y(t), \Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{XY}^T.$$

- *Conditioning on $Y = y(t) = \phi^{-1}(F_\tau(t))$, the portfolio value $V = \sum_{i=1}^N V_i$ is a Gaussian random variable with mean and variance given by*

$$\mu_{V|Y}(t) = \mathbf{1}^T\mu_{X|Y}, \sigma_{V|Y}^2(t) = \mathbf{1}^T\Sigma_{X|Y}\mathbf{1}.$$

- *The CVA is given by*

$$CVA = (1 - R_{CP})\int_0^T DF(0,t)e(t|y(t))dF_\tau(t),$$

where $e(t|y(t)))$ is the conditional exposure conditioning on $Y = y = \phi^{-1}(F_\tau(t))$, given by

$$\begin{aligned} e(t; y(t)) = \mu_{V|Y}(t)[\phi(\frac{\mu_{V|Y}(t)}{\sigma_{V|Y}(t)}) - \phi(\frac{\mu_{V|Y}(t) - H}{\sigma_{V|Y}(t)})] \\ + \sigma_{V|Y}(t)[f(\frac{\mu_{V|Y}(t)}{\sigma_{V|Y}(t)}) - f(\frac{\mu_{V|Y}(t) - H}{\sigma_{V|Y}(t)})] + H\phi(\frac{\mu_{V|Y}(t) - H}{\sigma_{V|Y}(t)}) \end{aligned}$$

where f is the pdf of the normal distribution.

Proof. (1)(2) Use the conditional and affine property of multivariate Gaussian distribution (Theorem 2.2.2). (3)(4) Use Lemma 26.3.1. □

Remark 26.5.1 (copula approach with factor representation). [5] In the copula approach, if we assume $(X_1, X_2, ..., X_N, Y)$ have the following factor representation:

$$\begin{aligned}
X_1 &= b_{11}F_1 + b_{12}F_2 + \sqrt{1 - b_{11}^2 - b_{12}^2}\epsilon_1 \\
X_2 &= b_{21}F_1 + b_{22}F_2 + \sqrt{1 - b_{21}^2 - b_{22}^2}\epsilon_2 \\
&\cdots\cdots \\
X_n &= b_{n1}F_1 + b_{n2}F_2 + \sqrt{1 - b_{n1}^2 - b_{n2}^2}\epsilon_n. \\
Y &= b_{Y1}F_1 + \sqrt{1 - b_{Y1}^2}\epsilon_Y
\end{aligned}$$

where $F_1, F_2, \epsilon_1, ..., \epsilon_N, \epsilon_Y$ are independent standard normal random variables.

Then we can calculate the conditional mean and covariance matrix via

$$\Sigma_{XY} = [b_{11}b_{Y1}\ b_{21}b_{Y1}\ \cdots\ b_{N1}b_{Y1}]^T,$$

$$\Sigma_{XX} = BB^T, B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \cdots & \cdots \\ b_{N1} & b_{N2} \end{bmatrix}.$$

26.5.3 Monte carlo method

26.5.3.1 *Stochastic intensity model*

Lemma 26.5.2 (default probability in the stochastic intensity framework). *Let τ be the default time. Let $h(t)$ be the stochastic hazard rate. Let current time be 0. It follows that*

- *Default time cumulative distribution function*

$$F_\tau(t) = Pr(\tau < t) = 1 - E[\exp(-\int_t^T h(s)ds)].$$

- *Survivial probability*

$$S(0,t) = 1 - F_\tau(t) = 1 - E[\exp(-\int_t^T h(s)ds)]$$

- *Default time density*

$$f_\tau(t) = E[h(t)\exp(-\int_0^t h(s)ds)]$$

Proof. See Lemma 23.9.1. □

Example 26.5.3 (examples of stochastic intensity model).

- (non-mean-reverting Gaussian hazard rate model) In the non-mean-reverting Gaussian hazard rate model, we assume that, under risk-neutral measure, the stochastic hazard rate $h(t)$ is governed by

$$d\lambda(t) = \mu(t)dt + \sigma dW$$

.

- (extended CIR model) The extended CIR model under risk-neutral model is given by

$$\begin{aligned} dx(t) &= -kx(t)dt + \sigma\sqrt{x}dW_t, x(0) = 0. \\ h(t) &= x(t) + \alpha(t). \end{aligned}$$

26.5.3.2 *Methods*

Methodology 26.5.1 (single trade level discrete evaluation via Monte Carlo).

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T]$.*
- *Jointly simulate M market scenarios (interest rate, FX, equity, futures, etc.) and intensity $\lambda(t)$ for each time point T_i.*
- *For each simulation time point T_i and scarario j: calculate the trade value $V^{(j)}(T_i)$, the collateral $C^{(j)}(T_i)$, and the collaterialized exposure $E_C^{(j)}(T_i) = \max(V^{(j)}(T_i) - C^{(j)}(T_i), 0)$.*
- *Compute the unconditional default probability based on $\lambda^{(j)}(t)$ given by*

$$Pr^{(j)}(T_{i-1} \leq \tau_{CP} \leq T_i) = \exp(-\int_t^{T_{i-1}} \lambda^{(j)}(s)ds) - \exp(-\int_t^{T_i} \lambda^{(j)}(s)ds).$$

- *Compuate the CVA along the trajectory j at T_i using the following discrete-time form given by*

$$CVA^{(j)}(t)[(1-R_{CP})\sum_{i=1}^{N} EPE^{(j)}(T_i)Pr^{(j)}(T_{i-1}\leq \tau_{CP}\leq T_i),$$

where

$$EPE^{(j)}(T_i) = \frac{1}{M}\sum_{j=1}^{M} DF^{(j)}(t,T_i)E_C^{(j)}(T_i).$$

- *The CVA calculation can be approximated*

$$CVA(t) = \frac{1}{M}\sum_{j=1}^{M} CVA^{(j)}(t)[(1-R_{CP})\sum_{i=1}^{N} E\hat{P}E(T_i)Pr(T_{i-1}\leq \tau_{CP}\leq T_i).$$

Remark 26.5.2 (comparion with CVA without WWR). In the CVA model without WWR, the default probability is a deterministic quantity; whereas without WWR, it has different realization for different market scenarios.

Example 26.5.4 (market risk factor and default intensity joint model).

- (interest rate) the dynamics of short rate $r(t)$ and hazard rate $\lambda(t)$ under risk-neutral measure are given by

$$\begin{aligned} dr &= k_1(\theta(t)-r)dt + \sigma_r dW_r \\ d\lambda &= k_2(\mu(t)-\lambda)dt + \sigma_\lambda dW_\lambda \end{aligned}$$

where $E[dW_r dW_\lambda] = \rho dt$.
- (credit, e.g., CDS) Let the stochastic hazard rate for our party $h_1(t)$, he counterparty $h_2(t)$, and the reference entity $h_3(t)$ be given by

$$dh(t) = \mu_1(t)dt + \sigma_1(t)dW_1(t)$$

- (Hull-White CVA model[6])

$$\lambda(t) = \exp(bV_t + a),$$

where V_t is the **stochastic** exposure at time t

Methodology 26.5.2 (Counterparty level discrete evaluation via Monte Carlo). *Consider a couterparty associated with K trades with trade value denoted by $V_1, V_2, ..., V_K$ with H netting groups.*

- *Let $t = T_0 < T_1 < T_2 < \cdots < T_N$ be the partition time interval of $[t, T]$.*
- *Jointly simulate M market scenarios (interest rate, FX, equity, futures, etc.) and intensity $\lambda(t)$ for each time point T_i.*
- *For each simulation time point T_i and scarario j:*
  - *For each trade $k = 1, 2, ..., K$, calculate the trade value $V_k^{(j)}(T_i)$.*
  - *Calculate the uncollaterillized exposure via*

$$E^{(j)}(T_i) = \sum_{i=1}^{H} \max\{\sum_{h \in NA_i} V_h^{(j)}(T_i), 0\}.$$

  - *Calculate the the collateral $C^{(j)}(T_i)$, and the collaterialized exposure $E_C^{(j)}(T_i) = \max(E^{(j)}(T_i) - C^{(j)}(T_i), 0)$.*
- *Compute the unconditional default probability based on $\lambda^{(j)}(t)$ given by*

$$Pr^{(j)}(T_{i-1} \leq \tau_{CP} \leq T_i) = \exp(-\int_t^{T_{i-1}} \lambda^{(j)}(s)ds) - \exp(-\int_t^{T_i} \lambda^{(j)}(s)ds).$$

- *Compuate the CVA along the trajectory j at T_i using the following discrete-time form given by*

$$CVA^{(j)}(t)[(1 - R_{CP}) \sum_{i=1}^{N} EPE^{(j)}(T_i) Pr^{(j)}(T_{i-1} \leq \tau_{CP} \leq T_i),$$

  where

$$EPE^{(j)}(T_i) = \frac{1}{M} \sum_{j=1}^{M} DF^{(j)}(t, T_i) E_C^{(j)}(T_i).$$

- *The CVA calculation can be approximated*

$$CVA(t) = \frac{1}{M} \sum_{j=1}^{M} CVA^{(j)}(t)[(1 - R_{CP}) \sum_{i=1}^{N} E\hat{P}E(T_i) Pr(T_{i-1} \leq \tau_{CP} \leq T_i).$$

26.6 Potential future exposure

26.6.1 Concepts

Definition 26.6.1 (future exposure). *[3, p. 127]*

- *The* ***positive exposure(PF)*** *is a random variable, whose support is on $[0, +\infty)$, on a future date t, which is used to quantify the loss(after considering netting and collateral agreement etc) at the future date if the counterparty defaults.*
- *The* ***expected positive exposure(EE)*** *is the average exposure on a future date t. The curve $EE(t)$ as a function of future time t, is called* ***expected exposure profile****.*

Definition 26.6.2 (potential future exposure). *[3, p. 127]*

- ***Potential (positive) future exposure(PFE)*** *is the maximum amount of positive exposure expected to occur on a future date with a high degree of statistical confidence. For example, the 95% PFE is the level of potential exposure that is exceeded with with 5% probability.*
- *The curve $PFE(t)$ as a function of future time t, is called* ***potential exposure profile****. The peak of $PFE(t)$ is referred to as* ***maximum potential future exposure****.*

Remark 26.6.1 (PFE vs. VaR). [3, p. 125]

- VaR is usually used to quantify the worst-scenario loss in a **short time horizon**,like 1 day or 10 days.
- PFE is also used to quantify the worst-scenario loss, however, usually in a **long time horizon**, like months to years.

26.6.2 Analytical potential future exposure

Lemma 26.6.1 (PFE with normal exposure dynamics). *Assume the positive exposure $E(t)$ is governed by normal model given by*

$$dE(t) = \mu dt + \sigma dW(t),$$

where $W(t)$ is a Brownian motion. Further assume current time is $t = 0$ and denote $E_0 = E(t = 0)$.

It follows that

- *The potential (positive) future exposure with confidence level α at a future date t is given by*
$$PFE(t) = E_0(\mu t + \phi^{-1}(\alpha)\sigma\sqrt{t}),$$
where $\phi(x)$ is the cdf of standard normal. In particular if $\alpha = 97.5\%$,
$$PFE(t) = E_0(\mu t + 1.96\sigma\sqrt{t}).$$

Proof. From the solution to the linear arithmetic SDE(Lemma 6.3.9), we have
$$E(t) = E(0)(\mu t + \sigma\sqrt{t}Z), Z \sim N(0,1).$$
To get the PFE, we have
$$\begin{aligned}
Pr(E(t) < PFE(t)) &= \alpha \\
Pr(E(0)(\mu t + \sigma\sqrt{t}Z) < PFE(t)) &= \alpha, Z \sim N(0,1) \\
Pr(Z < (PFE(t)/E(0) - \mu t)/\sigma\sqrt{t}) &= \alpha \\
\implies (PFE(t)/E(0) - \mu t)/\sigma\sqrt{t} &= \phi^{-1}(\alpha) \\
\implies PFE(t) &= E_0(\mu t + \phi^{-1}(\alpha)\sigma\sqrt{t})
\end{aligned}$$
□

Lemma 26.6.2 (PFE with lognormal exposure dynamics). *Assume the positive exposure $E(t)$ is governed by lognormal model given by*
$$dE(t) = \mu E(t)dt + \sigma E(t)dW(t),$$
where $W(t)$ is a Brownian motion. Further assume current time is $t = 0$ and denote $E_0 = E(t = 0)$.

It follows that

- *The potential (positive) future exposure with confidence level α at a future date t is given by*
$$PFE(t) = E_0 \exp(\mu t - \frac{1}{2}\sigma^2 t + \phi^{-1}(\alpha)\sigma\sqrt{t}),$$
where $\phi(x)$ is the cdf of standard normal. In particular if $\alpha = 97.5\%$,
$$PFE(t) = E_0 \exp(\mu t - \frac{1}{2}\sigma^2 t + 1.96\sigma\sqrt{t}).$$

Proof. From the solution to the linear geometric SDE(Lemma 6.3.10), we have
$$E(t) = E(0)\exp(\mu t - \frac{1}{2}\sigma^2 t + \sigma W(t)).$$

To get the PFE, we have

$$
\begin{aligned}
Pr(E(t) < PFE(t)) &= \alpha \\
Pr(E(0)\exp(\mu t - \frac{1}{2}\sigma^2 t + \sigma\sqrt{t}Z) < PFE(t)) &= \alpha, Z \sim N(0,1) \\
Pr(Z < (\ln(PFE(t)/E(0)) - \mu t + \frac{1}{2}\sigma^2 t)/\sigma\sqrt{t}) &= \alpha \\
\implies (\ln(PFE(t)/E(0)) - \mu t + \frac{1}{2}\sigma^2 t)/\sigma\sqrt{t} &= \phi^{-1}(\alpha) \\
\implies PFE(t) &= E_0 \exp(\mu t - \frac{1}{2}\sigma^2 t + \phi^{-1}(\alpha)\sigma\sqrt{t})
\end{aligned}
$$

□

26.7 Notes on bibliography

Major references are [3][2][5][7][8][9].

BIBLIOGRAPHY

1. Zhu, S. H. & Pykhtin, M. A guide to modeling counterparty credit risk (2007).
2. Brigo, D., Morini, M. & Pallavicini, A. *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes* ISBN: 9780470661789. https://books.google.com/books?id=Y-X9zqVcOFkC (Wiley, 2013).
3. Gregory, J. *Counterparty credit risk and credit value adjustment: A continuing challenge for global financial markets* (John Wiley & Sons, 2012).
4. Cesari, G. *et al. Modelling, pricing, and hedging counterparty credit exposure: A technical guide* (Springer Science & Business Media, 2009).
5. Pykhtin, M. & Rosen, D. Pricing counterparty risk at the trade level and CVA allocations (2010).
6. Hull, J. & White, A. CVA AND WRONG WAY RISK. *Financial Analysts Journal* **68,** 58–69 (2012).
7. Ghamami, S. & Goldberg, L. Stochastic intensity models of wrong way risk: wrong way CVA need not exceed independent CVA (2014).
8. Pykhtin, M. Modeling credit exposure for collateralized counterparties. *The Journal of Credit risk* **5,** 3 (2009).
9. Li, M. & Mercurio, F. Jumping with default: wrong-way-risk modeling for credit valuation adjustment (2016).

A

COMMON NOTATIONS AND FACTS

A.1 Some common spaces

The metric space $(\mathbb{R}^n, d_2)$ is the set $\mathbb{R}^n$ with metric $d_2(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

[1, p. 122]The metric space l^2 is the set of all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that $\sum_{i=1} \infty x_i^2 < \infty$,i.e., $\sum_{i=1} \infty x_i^2$ converges. The metric is usually defined as

$$d_2(\{x_n\}, \{y_n\}) = \sqrt{\sum_{k=1}^{\infty}(x_i - y_i)^2}$$

The metric space $l^p, 1 \leq p < \infty$,is the set of all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that $\sum_{i=1} \infty |x_i|^p < \infty$,i.e., $\sum_{i=1} \infty |x_i|^p$ converges. The metric is usually defined as

$$d_p(\{x_n\}, \{y_n\}) = \sqrt[p]{\sum_{k=1}^{\infty}(x_i - y_i)^p}$$

The metric space l^∞, is the set all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that every x_i is bounded. The metric is defined as

$$d_\infty(\{x_n\}, \{y_n\}) = \sup_n |x_n - y_n|$$

[2, p. 75].The metric space $C[a,b] = (C[a,b], d_\infty)$ denote the set of real-valued(or complex valued) functions defined on the interval $[a,b]$. The metric d_∞ is given as

$$d_\infty(x,y) = \sup_t |x(t) - y(t)|$$

Remark A.1.1. **Caution!** Sometimes $C[a,b]$ refers to only continuous functions.[3, p. 23]

The metric space $(C[a,b], d_p)$ denote the set of real-valued(or complex valued) functions defined on the interval $[a,b]$. The metric d_p is given as

$$d_p(x,y) = [\int_a^b |x(t) - y(t)|^p \, dt]^{1/p}$$

where $1 \leq p < \infty$.[2, p. 75].

The vector space $\mathcal{L}(V,W)$ usually denotes the set of all linear operators from V into W.

A.1.1 Notations on continuously differentiable functions

- C^0 refers to continuous function

- C^1 refers to functions having continuous first derivatives, also called continuously differentiable functions.
- C^2 refers to functions having continuous second derivatives
- C^∞ refers to smooth functions

A.2 Different modes of continuity

Chain of inclusions for functions over a closed and bounded subset of the real line

$$continuouslyDifferntiable \subseteq LipschitzContinouous \subseteq UniformlyContinuous$$

Remark A.2.1.

- Continuously differentiable on a closed interval indicates the derivative is bounded $f' \leq M$, then we have
$$|f(x) - f(y)| = f'(s)|x - y| \leq M|x - y|$$
hence Lipschitz continuous.
- $f(x) = |x|$ is Lipschitz continuous but is not differentiable everywhere except at $x = 0$, therefore it is not continously differentiable.
- Lipschitz continuous $\rightarrow$ continous:
$$|f(x) - f(y)| \leq L|x - y| \rightarrow 0$$
as $|x - y| \rightarrow 0$

Lemma A.2.1 (differentiable implies continuous). *If f is differentiable on $[a, b]$, then it is continuous on $[a, b]$.*

Proof:

$$\lim_{y \to x} f(y) - f(x) = \lim_{y \to x} (y - x)(f(y) - f(x))/(y - x) =$$
$$\lim_{y \to x} (y - x) \lim_{y \to x} (f(y) - f(x))/(y - x) = 0$$

where we have use the property that if two limits exist then they can multiply.[1, p. 42].

Remark A.2.2. This lemma indicates that a function differentiable everywhere will be continous everywhere.

Lemma A.2.2 (differentiable everywhere NOT implies continuously differentiable). *A function is differentiable everywhere NOT implies it is continuously differentiable function.*

The standard example is

$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x}), x \neq 0 \\ 0, x = 0 \end{cases}$$

This function can be differentiated every where and $f'(0) = 0$, but $\lim_{x \to 0} f'(x)$ does not exist. See link.

A.2.1 continuity vs. uniform continuity

Definition A.2.1. *A function $f : X \to Y$ is* uniformly continuous *if for every $\epsilon > 0$ there exist a $\delta > 0$ such that for every $x, x_0 \in X$,*

$$\rho(x, x_0) \leq \delta \Rightarrow \rho(f(x), f(x_0)) < \epsilon$$

Theorem A.2.1. *[1, p. 154] If f is a continuous function from a compact metric space M_1 into a metric space M_2, then f is uniformly continuous on M_1.*

Corollary A.2.1.1. *[1, p. 154] If f is a continuous real-value function on a closed and bounded subset X of $\mathbb{R}^n$, then f is uniformly continuous on X.*

Example A.2.1. The function $f(x) = x^2$ is continuous but not uniformly continuous on the interval $(0, \infty)$.

Lemma A.2.3 (sufficient condition). *Let $S = \mathbb{R}$. if f is global Lipschitz continuous, i.e.*

$$\left|f(x_1) - f(x_2)\right| < M|x_1 - x_2|$$

$\forall x_1, x_2 \in S$, then f is uniformly continuous.

Proof: $\left|f(x_1) - f(x_2)\right| < M|x_1 - x_2| \to 0$

A.3 Exchanges of limits

A.3.1 Overall remark

Remark A.3.1.

- Usually, the necessary conditions for exchanging limits is difficult to find, therefore only sufficient conditions are given.
- Many operations are in nature taking limits, for example, summing infinite terms is taking limits on partial sums; integrals is taking limits on both summation and partitions; derivative is taking limits on quotient expressions.

A.3.2 exchange limits with infinite summations

Let $\lim_{m\to\infty}\sum_{n=1}^{\infty} f(m,n) = \sum_{n=1}^{\infty}\lim_{m\to\infty} f(m,n)$ Based on dominated convergence, if there is a $g(n)$ such that $f(m,n) < g(n), \forall m$ and $\sum_{n=1}^{\infty} g(n)$ exists, then we can exchange.

To use the dominated convergence theorem in Lebesgue integral, we can define a simple function $s_n$on $[0,\infty]$ take $f(m,n)$ on the interval $[m-1,m)$. Then the integral of s_n with respect to Legesbue measure on real lime will give the $\int_{[0,\infty)} s_n d\mu = \sum_{m=1}^{\infty} f(m,n)$

Theorem A.3.1. *[1, pp. 94, 373]Let $a_{m,n}$ be non-negative and $\sum_m^{\infty}\sum_n^{\infty} a_{m,n}$ exists, then*

$$\sum_m^{\infty}\sum_n^{\infty} a_{m,n} = \sum_n^{\infty}\sum_m^{\infty} a_{m,n}$$

Corollary A.3.1.1. *Let $a_{m,n}$ be increasing on both m,n and $\lim_{m\to\infty}\lim_{n\to\infty} a_{m,n}$ exists, then*

$$\lim_{m\to\infty}\lim_{n\to\infty} a_{m,n} = \lim_{n\to\infty}\lim_{m\to\infty} a_{m,n}$$

Proof: by constructing partial sums.

A.3.3 Exchange limits with integration and differentiation

Theorem A.3.2. *[1, p. 249] Let α be a function of bounded variation on $[a,b]$ and let f_n be a sequence of functions in $\mathcal{R}_\alpha[a,b]$ which converges uniformly to a function f. Then $f \in \mathcal{R}_\alpha[a,b]$ and*

$$\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b \lim_{n\to\infty} f d\alpha$$

Theorem A.3.3. *[1, p. 249] Let $\{f_n\}$ be a sequence of differentiable functions on (a,b). Suppose that*

- *f_n' is continuous on (a,b)*
- *$\{f_n\}$ converges pointwise to f*
- *$\{f_n'\}$ converges uniformly*

then f is differentiable on (a,b) and f_n' converges uniformly to f'.

A.3.4 Exchange differentiation with integration

Theorem A.3.4. *Let $f(x,y)$ be continuous on $[a,b] \times [c,d]$. Then*

$$\phi(y) = \int_a^b f(x,y)dx$$

defined above is continuous function on $[c,d]$

Proof: for any $\epsilon > 0$, there exist δ, such that

$$|\phi(y) - \phi(y')| \leq \int_a^b |f(x,y) - f(x,y')|\, dx \leq \epsilon(b-a) \forall |y - y'| < \delta$$

where we have the fact of $f(x,y) - f(x,y)$ is bounded (since continuous function on a compace set is uniformly continuous and will have maximum and minimum) which shows $\phi(y)$is uniformly continuous.

Theorem A.3.5. *Let f and f_y be continuous on $[a,b] \times [c,d]$. Then ϕ is differentiable and*

$$\phi_y = \int_a^b f_y(x,y)dx$$

Proof:

$$\frac{\phi(y+h) - \phi(y)}{h} = \frac{1}{h}\int_a^b f(x,y+h) - f(x,y)dx = \int_a^b f_y(x,z)$$

due to Taylor theorem, where $z \in [y, y+h]$. Then

$$\left|\frac{\phi(y+h)-\phi(y)}{h} - \int_a^b f_y(x,y)dx\right| \leq \int_a^b \left|f_y(x,z) - f_y(x,y)\right| dx$$

Because f_y is continuous on compact set, then it is uniformly continuous. Therefore given $\epsilon > 0$, there exists δ such that

$$\left|f_y(x,y') - f_y(x,y)\right| < \epsilon/(b-a), \forall \left|y - y'\right| < \delta$$

Taking $h < \delta$, we have

$$\left|\frac{\phi(y+h)-\phi(y)}{h} - \int_a^b f_y(x,y)dx\right| < \epsilon.$$

Take the limit on h and we get the result.

A.3.5 Exchange limit and function evaluations

Lemma A.3.1. *Let $\{x_n\}$ be a sequence with limit x, let f be a continuous function*

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x)$$

Proof: from the definition of continuous function.

A.4 useful inequalities

Lemma A.4.1 (arithmetic-geometric mean inequality). *For $x_1, ..., X_n \geq 0$, we have*

$$(x_1 x_2 ... x_n)^{1/n} \leq \sum_{i=1}^{n} x_i / n.$$

Specifically,

$$\frac{x_1 + x_2}{x} \geq \sqrt{x_1 x_2}$$

Proof. use $y = \ln(x)$ and concavity of $\ln(x)$ □

A.4.1 Gronwall's inequality

see [4]

A.4.2 Inequality for norms

Lemma A.4.2. *[5]For L^p normed space, we have*

$$\|x\|_2 \leq \|x\|_1$$

where

$$\|x\|_2 = (\int_{-\infty}^{\infty} |f(x)|^2 \, d\mu(x))^{0.5}$$

and

$$\|x\|_1 = \int_{-\infty}^{\infty} |f(x)| \, d\mu(x)$$

Proof: for finite dimensional normed space cases: we need to prove

$$\sqrt{x_1^2 + x_2^2 + ... + x_n^2} \leq |x_1| + |x_2| + ... + |x_n|$$

By squaring both sides, we can get the result. For continuous case, TODO

Theorem A.4.1. *[5]For L^p normed space, we have*

$$\|x\|_q \leq \|x\|_q$$

whenever $p \leq q$ *where*

$$\|x\|_q = (\int_{-\infty}^{\infty} |f(x)|^q \, d\mu(x))^{1/q}$$

Proof: todo

Remark A.4.1. For complete description on L^p norms, see [5]

A.4.3 Young's inequality for product

Lemma A.4.3. *If* $a, b \geq 0$*, and* $p, q > 1, 1/p + 1/q = 1$*, then*

$$ab \leq a^p/p + b^q/q$$

Proof:

$$\log(a^p/p + b^q/q) \geq \log(a^p)/p + \log(a^q)/q = \log(a) + \log(b) = \log(ab)$$

where we use the fact of log is concave.

A.5 Basic logic for proof

[6, p. 60]The negation of

for any $\epsilon > 0$, there exist $N > 0$, such that for all $n > N$, we have $\|a_n - a\| < \epsilon$

is

there exist an $\epsilon > 0$, such that for every $N > 0$, such that for all $n > N$, we have $\|a_n - a\| > \epsilon$.

[6, p. 60]The negation of

for any $\epsilon > 0$, there exist $\delta > 0$, such that for all $\|x - x_0\| < \delta$, we have $\|f(x) - f(x_0)\| < \epsilon$

is

there exist an $\epsilon > 0$, such that for every $\delta > 0$, such that for all $\|x - x_0\| < \delta$, we have $\|f(x) - f(x_0)\| > \epsilon$.

A.6 Common series summation

Lemma A.6.1. *[7, p. 1]*

- $$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$
- $$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$
- $$\sum_{k=1}^{n} k^3 = [\frac{n(n+1)}{2}]^2$$

Lemma A.6.2. *[7, p. 1]* *Assume* $q \neq 1$.

- $$\sum_{k=1}^{n} aq^{k-1} = a\frac{q^n - 1}{q-1}$$
- $$\sum_{k=0}^{n-1} kq^k = \frac{(n-1)q^n}{q-1} + \frac{(q-q^n)}{(q-1)^2}$$
- $$\sum_{k=0}^{n-1} (n-1-k)q^k = -(n-1)\frac{1}{q-1} - \frac{(q-q^n)}{(q-1)^2}$$

Proof. (3) use (1)(2), we have

$$\sum_{k=0}^{n-1} (n-1-k)q^k = (n-1)\frac{q^n-1}{q-1} - \frac{(n-1)q^n}{q-1} - \frac{(q-q^n)}{(q-1)^2}$$

□

A.7 Some common limits

Lemma A.7.1 (Stirling approximation). *link*

For positive integer n,

$$\ln n! = n\ln n - n + O(\ln n).$$

- $$\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq en^{n+1/2}e^{-n}, \forall n > 0.$$

Lemma A.7.2 (common limits summary).

- $$\lim_{n\to\infty} \frac{\ln n}{n} = 0.$$
- $$\lim_{n\to\infty} \frac{x^n}{n!} = 0, \forall x \in \mathbb{R}.$$
- $$\lim_{n\to\infty} \frac{n!}{n^n} = 0.$$
- $$\lim_{n\to\infty} M^{1/n} = 1$$

  for any $M > 0$.
- $$\lim_{n\to\infty} \frac{\ln n!}{n} = \infty, \lim_{n\to\infty} (n!)^{1/n} = \infty.$$

Proof. (2)see **??** and **??**.(3)**??**.(4) **??**. (5) (a)Use Stirling approximation $\ln n! = n\ln n - n + O(\ln n)$ and $\ln n!/n = n - 1 + O(\ln n/n) \to \infty$.(b) Note that $(n!)^{1/n} = \exp(\ln(n!)^{1/n}) = \exp(\frac{\ln n!}{n})$. □

Note A.7.1. A helpful and general summary, as $n \to \infty$

$$\ln n \ll n^r (r > 0) \ll a^n (a > 1) \ll n! \ll n^n.$$

Lemma A.7.3 (property of e). *Define*

$$\lim_{n\to\infty} (1 + 1/n)^n = e$$

and then

$$\lim_{n\to\infty}(1 + x/n)^n = e^x$$

for any real x.

Proof.

$$\lim_{n\to\infty}((1 + x/n)^{n/x})^x = e^x$$

use the fact the $f(y) = y^x$ is continuous, such that function evaluation and limit can be exchanged. □

A.8 Useful properties of matrix

Remark A.8.1 (references). The most important reference for this section is the "The Matrix Cookbook".[8]

A.8.1 Matrix derivatives

For $A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n$, we have:

$$\frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$$
$$\frac{\partial Ax}{\partial x} = A$$
$$\frac{\partial BAx}{\partial x} = BA$$
$$\frac{\partial x^T Ax}{\partial x} = (A + A^T)x$$
$$\frac{\partial x^T Ax}{\partial x} = 2A$$

Lemma A.8.1. *If $f(x) = g(Ax), A \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n$ for some differentiable function $g(y)$, then*

$$\nabla f = A^T \nabla g$$

In particularly, $a \in \mathbb{R}^n$, then

$$\nabla a^T Ax = A^T x$$

A.8.2 Matrix inversion lemma

Lemma A.8.2 (matrix inversion lemma). *[9, p. 120]*

- $(E - FH^{-1}G)^{-1} = E^{-1} + E^{-1}F(H - GE^{-1}F)^{-1}GE^{-1}$
- $(E - FH^{-1}G)^{-1}FH^{-1} = E^{-1}F(H - GE^{-1}F)^{-1}$

Proof: (1) can be verified (2) expand the parenthesis using (1).

A.8.3 Block matrix

Lemma A.8.3. *Given an $(m \times p)$ matrix A with q row partions and s colun partitions and a $(p \times n)$ matrix B with s row partions and r colun partitions,*

$$A = \begin{matrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q1} & A_{q2} & \cdots & A_{qs} \end{matrix}, B = \begin{matrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{sr} \end{matrix},$$

then the matrix product

$$C = AB$$

can be formed blockwise, giving C as an $(m \times n)$ matrix with q row partitions and r column partitions. In particular,

$$C_{\alpha\beta} = \sum_{\gamma=1}^{s} A_{\alpha\gamma} B_{\gamma\beta}.$$

Lemma A.8.4 (sum of vector product to matrix product). *Consider column vectors $x_1, x_2, ..., x_N \in \mathbb{R}^d$ and column vectors $x_1, x_2, ..., x_N \in \mathbb{R}^d$. It follows that*

- $$\sum_{i=1}^{N} x_i^T y_i = X_C^T Y_C,$$

  where $X_C \in \mathbb{R}^{Nd}$ is a vector stacking all the $x_1, ..., x_N$ (similarly Y_C).
- $$\sum_{i=1}^{N} x_i y_i^T = X_R^T Y_R,$$

  where $X_R \in \mathbb{R}^{N \times d}$ is a matrix stacking all the $x_1^T, ..., x_N^T$ (similarly Y_R).

A.8.4 Matrix trace

Lemma A.8.5.

- $\|A\|_F^2 = Tr(AA^T)$

Lemma A.8.6 (matrix trace).

- *(linearity)$Tr(aA + bB) = aTr(A) + bTr(B)$*
- *(commutative) $Tr(AB) = Tr(BA)$*
- *(invariance under transposition) $Tr(A) = Tr(A^T)$*

- *(cyclic rule)* $Tr(ABCD) = Tr(DABC) = Tr(CDAB) = Tr(BCDA)$ *or* $Tr(ABC) = Tr(CAB) = Tr(BCA)$

Proof. (1)(2)(3) can be proved directly from definition. (4) We can group three elements together and commute with the fourth. For example, we can group (ABC) together and commute with D to prove the first equality. □

Corollary A.8.0.1.

- *(invariance under similar transformation)* $Tr(PAP^{-1}) = Tr(A)$
- $Tr(X^TY) = Tr(XY^T) = Tr(Y^TX) = Tr(XY^T)$

Proof. (1) Use cyclic rule. (2) Use invariance under transposition and commutative rule. □

A.8.5 Matrix elementary operator

Lemma A.8.7 (elementary operator matrix). *Left multiply a matrix A by an elementary matrix is equivalent to performing row operation. The elementary operator matrix is given by*

- *(Interchange row i and j) For example, exchange row 2 and row 3:*

$$R_1 = \begin{bmatrix} 1 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 1 \end{bmatrix}$$

- *(Multiply row i by s) For example*

$$R_2 = \begin{bmatrix} 1 & & & \\ & s & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

- *(Add s times row i to row j) For example, add s times row 2 to row 3*

$$R_3 = \begin{bmatrix} 1 & & & \\ & 1 & 0 & \\ & s & 1 & \\ & & & 1 \end{bmatrix}.$$

Note that $R_3 = R_1R_2 \neq R_2R_1$.

Lemma A.8.8 (elementary operator matrix). *Right multiply a matrix A by an elementary matrix is equivalent to performing row operation. The elementary operator matrix is given by*

- *(Interchange column i and j) For example, exchange row 2 and row 3:*

$$C_1 = \begin{bmatrix} 1 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 1 \end{bmatrix}$$

- *(Multiply column i by s) For example*

$$C_2 = \begin{bmatrix} 1 & & & \\ & s & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

- *(Add s times column i to column j) For example, add s times column 2 to column 3*

$$C_3 = \begin{bmatrix} 1 & & & \\ & 1 & s & \\ & & 1 & \\ & & & 1 \end{bmatrix}.$$

Note that $C_3 = C_1C_2 \neq C_2C_1$.

A.8.6 Matrix determinant

Lemma A.8.9 (properties of determinant).

- *All elementary operator matrix has determinant 1.*
- *For matrix* $A \in \mathbb{R}^{n\times n}$,
$$det(kA) = k^n det(A).$$
- $det(AB) = det(A)det(B)$.
- *All elementary operation on a matrix will not change its determinant.*

A.9 Numerical integration

Definition A.9.1 (Newton-Cotes Formula). *Suppose we want to evaluate $\int_a^b f(x)dx$. We can evaluate $f(x)$ at $n+1$ equally spacing points $x_i = a + i(b-a)/n$, and then we approximate $f(x)$ by n degree of Lagrange polynomial and do the integral. Specifically, we have*

$$\int_a^b f(x)dx \approx \int_a^b L(x)dx = \int_a^b (\sum_{i=0}^n f(x_i)l_i(x)) = \sum_{i=1}^n f(x_i)\int_a^b l_i(x)dx = \sum_{i=1}^n f(x_i)w_i$$

where L is the Lagrange polynomial of degree n, and $l_i(x), i = 0, ..., n$ is the (n+1) Lagrange polynomial basis, given as **??**.

Example A.9.1. Consider we use degree 1 Lagrange polynomial to approximate $f(x)$, then

$$L(x) = f(a)\frac{x-a}{b-a} + f(b)\frac{x-b}{a-b}$$

where $l_0(x) = \frac{x-a}{b-a}$ and $l_1(x) = \frac{x-b}{a-b}$. Then

$$w_0 = \int_a^b l_0(x)dx = \frac{1}{2}, w_1 = \int_a^b l_1(x)dx = \frac{1}{2}.$$

Table A.9.1: Closed Newton-Cotes Formula

Notation: $\int_a^b f(x)dx$, $f_i = f(x_i), x_i = a + i(b-a)/n$			
Degree	Name	Formula	Error term
1	Trapezoid rule	$\frac{b-a}{2}(f_0 + f_1)$	$-\frac{(b-a)^3}{12}f^{(2)}(\eta)$
2	Simpson's rule	$\frac{b-a}{6}(f_0 + 4f_1 + f_2)$	$-\frac{(b-a)^5}{2880}f^{(4)}(\eta)$
3	Simpson's 3/8 rule	$\frac{b-a}{8}(f_0 + 3f_1 + 3f_2 + f_3)$	$-\frac{(b-a)^3}{6480}f^{(5)}(\eta)$

Remark A.9.1 (Error analysis). For detailed error analysis, see [10, p. 252].

Remark A.9.2 (how to use). Usually, given the integral $\int_a^b f(x)dx$, we will first divide into smaller intervals and do the numerical integral on each interval and add them up. For example

$$\int_a^b f(x)dx = \int_a^{a+h} f(x)dx + \int_{a+h}^{a+2h} f(x)dx ... + \int_{b-h}^b f(x)dx.$$

Lemma A.9.1 (Trapezoid rule and the error bound). *Given the integral* $\int_a^b f(x)dx$*, we have*

$$\int_a^b f(x)dx \approx \frac{b-a}{n}(\frac{f(a)}{2} + \sum_{k=1}^{n-1}(f(a+k\frac{b-a}{n})) + \frac{f(b)}{2})$$

where we divide $b-a$ *into* n *subintervals, and we do the Trapezoid rule on each subinterval. And the error bound is*

$$\frac{(b-a)^3}{12n^2} \max_{x\in[a,b]} f^{(2)}(x)$$

Proof. Note that on each subinterval, the error is $-\frac{(b-a/n)^3}{12}f^{(2)}(\eta)$. Sum up n terms, and we have upper bound

$$\frac{(b-a)^3}{12n^3} n \max_{x\in[a,b]} f^{(2)}(x)$$

□

Lemma A.9.2 (Midpoint rule and the error bound). *Given the integral* $\int_a^b f(x)dx$*, we have*

$$\int_a^b f(x)dx \approx \frac{b-a}{n}(\sum_{k=1}^{n}(f(a+(k-0.5)\frac{b-a}{n})))$$

where we divide $b-a$ *into* n *subintervals, and we do the Trapezoid rule on each subinterval. And the error bound is*

$$\frac{(b-a)^3}{n^2}K$$

A.9.1 Gaussian quadrature

$$\int_a^b w(x)f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

which is exact when f is a polynomial.

Remark A.9.3. In Newton-Cotes formulas, we fix nodes and try to find suitable weights; in Gaussian quadrature, we use a weighted sum of function values at specified points within the domain of integration.

A.10 vector calculus

Lemma A.10.1 (divergence theorem).

$$\iiint_V (\nabla \cdot \mathbf{F}) dV = \oiint_{S(V)} \mathbf{F} \cdot \hat{\mathbf{n}} dS$$

$$\iiint_V (\nabla \times \mathbf{F}) dV = \oiint_{S(V)} \hat{\mathbf{n}} \times \mathbf{F} dS$$

$$\iiint_V (\nabla f) dV = \oiint_{S(V)} \hat{\mathbf{n}} f dS$$

Lemma A.10.2 (Lapacian product rule). *Given functions $u : \mathbb{R}^n \to \mathbb{R}$ and $v : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\nabla^2(uv) = u\nabla v + 2\nabla u \cdot \nabla v + v\nabla^2 u.$$

Proof. Directly use product rule. □

A.11 Pigeon Hole principle

Lemma A.11.1.

- *If there are n holes and the number of pigeons* $N \geq n+1$*, then at least 2 pigeons have to share one of the holes.*
- *If there are n and the number of pigeons* $N \geq mn+1$*, then at least* $m+1$ *pigeons have to share one of the holes.*

Example A.11.1. Problem : Prove that in any party there will be at least two persons who have shaken hands with equal number of people.

Solution : Let us say there are n persons attending the party. Obviously, this problem makes sense only when n $\geq$ 2. If no two person have shaken hands with equal number of people then there handshake count must differ at least by 1. So the possible choices for hand shake count would be 0,1,...,n-1. There are exactly n choices and n people. But the catch is that if there exist a person with n-1 handshake count, there can't be a person with 0 handshake count. Thus reducing the possible choices to n-1. Now, due to pigeon hole principle, we have that at least two person will have the same number of handshake count.

A.12 Numerical linear algebra computation complexity

Note A.12.1. [11, p. 606]

- For a $m \times n$ matrix multiplying a n dimensional vector, mn.
- For a $n \times n$ matrix multiplying a $n \times n$ matrix, n^3(without optimization).
- For a $n \times n$ matrix, LU decomposition $2n^3/3$ (for symmetric matrix $n^3/3$).
- For a $m \times n$ matrix, Cholesky decomposition $4m^2n/3$ (for square matrix $4n^3/3$).
- For a $m \times n$ matrix, QR decomposition $4m^2n/3$ (for square matrix $4n^3/3$).

Note A.12.2 (solving triangular linear system). Let L be a $n \times n$ lower triangle matrix, the forward substitution algorithm for solving

$$Ly = d,$$

is given by

```
y(1) = d(1) / L(1,1);
for i=2:n
y(i) = (d(i) - L(i,1:i-1)* y(1:i-1))/L(i,i)
end
```

This algorithm has complexity of $O(n^2)$.

Let U be a $n \times n$ upper triangle matrix, the backward substitution algorithm for solving

$$Ux = d,$$

is given by

```
x(n) = d(n)/U(n,n);
for i = n - 1: -1 :1
x(i) = (d(i) - U(i,i + 1:n)*x(i + 1:n) )/U(i,i)
end
```

This algorithm has complexity of $O(n^2)$.

A.13 Distributions

Lemma A.13.1. *[12, p. 579] Let K be an externally given parameter. We have*

- $$\int_{-\infty}^{\infty} \delta(x)dx = 1, x\delta(x) = 0, \int_{-\infty}^{\infty} f(x)\delta(x-K)dx = f(K).$$
- $\delta(g(x)) = \sum_i \frac{\delta(x-x_i)}{|g'(x_i)|}$, *where* $x_i, i = 1, 2, \ldots$ *are the zeros of the function* $g(x)$.
- $\delta(\lambda x) = \frac{\delta(x)}{|\lambda|}, \delta(x-K) = \delta(K-x)$.
- *(step function definition)*
$$H(x) \triangleq \frac{d}{dx}\max\{x, 0\}, H(x-A) \triangleq \frac{d}{dx}\max\{x-A, 0\}$$
- $H(x-K) + H(K-x) = 1$.
- $$\frac{dH(x-K)}{dx} = \delta(x-K), \frac{dH(K-x)}{dx} = -\delta(x-K).$$

Proof. Use $H(x-K) + H(K-x) = 1$ to prove $\frac{dH(K-x)}{dx}$. □

A.14 Common integrals

Lemma A.14.1.

- $$\int_0^\infty e^{-ax^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}, \int_{-\infty}^\infty e^{-ax^2}dx = \sqrt{\frac{\pi}{a}}$$
- $$\int_0^\infty xe^{-ax^2}dx = \frac{1}{2a}, \int_{-\infty}^\infty xe^{-ax^2}dx = 0$$
- $$\int_0^\infty x^2e^{-ax^2}dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}}$$
- $$\int_0^\infty x^ne^{-ax}dx = \frac{n!}{a^{n+1}}$$
- $$\int_0^\infty x^me^{-ax^2} = \frac{\Gamma((m+1)/2)}{2a^{(m+1)/2}}$$

A.15 Nonlinear root finding

A.15.1 Bisection method

Definition A.15.1.

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$.
- ***(Initial input)***: *Initial guess of* l_0 *and* r_0 *such that*

$$f(l_0) < 0, f(r_0) > 0; \text{ or } f(l_0) > 0, f(r_0) < 0.$$

- ***Repeat (i) is the iteration index***:
  - *Let* $m = \frac{r_i + l_i}{2}$.
  - *If* $f(l_i)f(m) < 0$, *then* $l_{i+1} = l_i, r_{i+1} = m$.
  - *If* $f(l_i)f(m) > 0$*(then we must have* $f(r_i)f(m) < 0$*), then* $l_{i+1} = m, r_{i+1} = r_i$.
  - *If* $f(l_i)f(m) = 0$, *then* m *is the root.*

A.15.2 Newton method

Definition A.15.2.

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$. f *is assumed to be differentiable.*
- ***(Initial input)***: *Initial guess of* x_0.
- ***Repeat****(i is the iteration index):*
  - *Let* $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$.
  - *If* $f(x_{i+1}) = 0$, *then* x_{i+1} *is the root.*

A.15.3 Secant method

Definition A.15.3.

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$. f *is assumed to be differentiable.*
- ***(Initial input)***: *Initial guess of* x_0, x_1.
- ***Repeat****(i is the iteration index):*
  - *Let*

$$x_{i+1} = x_i - f(x_i)\frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})}$$

.

- *If $f(x_{i+1}) = 0$, then x_{i+1} is the root.*

Remark A.15.1 (derivation)**.** Starting with initial guesses x_0, x_1, we construct a first order approximation of $f(x)$ via

$$y = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1).$$

And we solve the root for the first-order approximation problem via

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1) = 0 \implies x_2 = x_1 - f(x_1)\frac{x_1 - x_0}{f(x_1) - f(x_0)}.$$

Then we continue the process with x_1, x_2.

Remark A.15.2 (convergence property)**.**

- There is no guarantee on the global convergence to the root of f.
- Only when the initial values x_0 and x_1 are sufficiently close to the root, the iterates x_n will converge to the root.

A.16 Interpolation

A.16.1 cubic interpolation

Definition A.16.1 (the cubic spine line functional form). *[13]*

- *Suppose* $x_1, x_2, ..., x_n$ *and* $y_1, y_2, ..., y_n$ *are known.*
- *A cubic spine line is given by*

$$y(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, x_i \leq x \leq x_{i+1}, i = 1, 2, ..., n - 1.$$

- *There are* $4n - 4$ *unknowns.*
- *Note that*

$$\begin{aligned} y'(x) &= b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2, x_i < x < x_{i+1} \\ y''(x) &= 2c_i + 6d_i(x - x_i), x_i < x < x_{i+1} \\ y'''(x) &= 6d_i, x_i < x < x_{i+1} \end{aligned}$$

Definition A.16.2 (natural cubic spline condition). *[13]* *Let* $h_i = x_{i+1} - x_i$

- ***(spline line passing data points)****: for* $i = 1, 2, ..., n - 1$, $a_i = y_i$*;* $a_{n-1} + b_{n-1}h_{n-1} + c_{n-1}h_{n-1}^2 + d_{n-1}h_{n-1}^3 = y_n$.
- ***(interpolating function is continuous)****; that is,*

$$\lim_{x \to x_i -} y(x) = \lim_{x \to x_i +} y(x) \implies a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 = a_{i+1}, \forall i = 1, 2, ..., n - 2.$$

- ***(interpolating function is differentiable)****; note that the interpolating function is differentiable on interval, therefore we require that,*

$$\lim_{x \to x_i -} y'(x) = \lim_{x \to x_i +} y'(x) \implies b_i + 2c_i h_i + 3d_i h_i^2 = b_{i+1}, \forall i = 1, 2, ..., n - 2$$

  :
- ***(interpolating function is twice differentiable and the second derivative at each endpoint is 0)****; that is,*

$$\lim_{x \to x_i -} y''(x) = \lim_{x \to x_i +} y''(x) \implies c_i + 3d_i h_i = c_{i+1}, \forall i = 1, 2, ..., n - 2,$$

  and $y''(x_1+) = y''(x_n-) = 0$.
- *these* $4n - 4$ *equations will solve the* $4n - 4$ *unknowns.*

A.17 Finance resources

A.17.1 Book resources

A.17.1.1 *General advice*

The website of Mark Joshi, see link.

Recommended books on quantitative finance, see link.

General advice for becoming a quant, see link.

A.17.1.2 *Quantitative investment*

Factor Investing and Asset Allocation: A Business Cycle Perspective

Quantitative Equity Investing: Techniques and Strategies

Asset Management: A Systematic Approach to Factor Investing

Beyond Smart Beta: Index Investment Strategies for Active Portfolio Management

Your Complete Guide to Factor-Based Investing: The Way Smart Money Invests Today

Quantitative Investment Portfolio Analytics In R: An Introduction To R For Modeling Portfolio Risk and Return

A Quantitative Primer on Investments with R

A.17.1.3 *Softwares*

Pyfolio: pyfolio is a Python library for performance and risk analysis of financial portfolios developed by Quantopian Inc. It works well with the Zipline open source backtesting library.

Zipline: Zipline is a Pythonic algorithmic trading library. It is an event-driven system for backtesting. Zipline is currently used in production as the backtesting and live-trading engine powering Quantopian âĂŞ a free, community-centered, hosted platform for building and executing trading strategies.

PyAlgoTrade: PyAlgoTrade is a Python Algorithmic Trading Library with focus on backtesting and support for paper-trading and live-trading. LetâĂŹs say you have an idea for a trading strategy and youâĂŹd like to evaluate it with historical data and see how it behaves. PyAlgoTrade allows you to do so with minimal effort.

backTrader: https://github.com/backtrader/backtrader

A.17.1.4 *Commodity*

Natural gas trading in north America

Trading Natural Gas: Cash, Futures, Options and Swaps

Valuation and Risk Management in Energy Markets

Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging

Managing Energy Risk: An Integrated View on Power and Other Energy Markets

A.17.2 People to follow

Stefan Andreer (linkin)

Cristian Homescu (linkedin)

Xuan Che (linkedin)

BIBLIOGRAPHY

1. Johnsonbaugh, R. & Pfaffenberger, W. Foundations of mathematical analysis (2010).
2. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
3. Luenberger, D. G. *Optimization by vector space methods* (John Wiley & Sons, 1969).
4. Wikipedia. *Gronwall's inequality — Wikipedia, The Free Encyclopedia* [Online; accessed 10-July-2016]. 2016. https://en.wikipedia.org/w/index.php?title=Gr%C3%B6nwall%27s_inequality&oldid=717747531.
5. Wikipedia. *Lp space — Wikipedia, The Free Encyclopedia* [Online; accessed 15-August-2016]. 2016. https://en.wikipedia.org/w/index.php?title=Lp_space&oldid=733853566.
6. Hammack, R. H. *Book of Proof* (Richard Hammack, 2013).
7. Gradshteyn, I. S. & Ryzhik, I. M. *Table of integrals, series, and products* (Academic press, 2014).
8. Petersen, K. & Pedersen, M. The matrix cookbook, nov 2012. *URL http://www2. imm. dtu. dk/pubdb/p. php* **3274.**
9. Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT press, 2012).
10. Atkinson, K. E. *An introduction to numerical analysis* (John Wiley & Sons, 1989).
11. Nocedal, J. & Wright, S. *Numerical optimization* (Springer Science & Business Media, 2006).
12. Prosperetti, A. Advanced mathematics for applications. *SIAM Review Vol. 55, Issue 2 (June 2013)* **55,** 403 (2013).
13. Hagan, P. S. & West, G. Interpolation methods for curve construction. *Applied Mathematical Finance* **13,** 89–129 (2006).

ALPHABETICAL INDEX