Math 622 - Lecture notes II

TABLE OF CONTENTS

CHAPTER 1 Jump process models

1.1 Motivations

Previously, most of the models of the stock process we have encountered are continuous, i.e the stock price is not supposed to have jumps. Very quickly we see this assumption is restrictive: when the stock pays dividend, the stock price has a downward jump corresponding to the amount of divident payout. However, the dividend payment can be covered within the continuous framework without introducing any new ideas, essentially because the dividend payment times are **deterministic**.

Dividend payments are not the only phenomenons that cause the stock price to have jumps, obviously. In reality, we quickly observe many instances where stock price jumps, and the most important characteristic of these jumps is that they happen at **random** times. Being able to model stock prices that incorporate jumps at random (or more precisly, stopping times) and learning how to price financial products based on these models are the main focus of this Chapter.

1.2 Overview of price modeling in continuous time

Let $\{\mathcal{F}(t);\, t \geq 0\}$ be a filtration modeling the accumulation of market information available to investors as time progresses. A simple paradigm guides the construction of models for an asset price, $\{S(t);\, t \geq 0\}$, that is a continuous function of time:

$$\frac{\mathrm{d}S(t)}{S(t)} = \alpha(t)\,\mathrm{d}t + dM(t), \tag{1.1}$$

where α is a process adapted to $\{\mathcal{F}(t);\, t \geq 0\}$, M is a martingale with respect to $\{\mathcal{F}(t);\, t \geq 0\}$, and $dS(t) = S(t+\mathrm{d}t) - S(t)$ denotes the price increment at t for an infinitesimally small positive increment of time, $\mathrm{d}t$. This is a formal equation, because $\mathrm{d}t$ is not precisely defined. Intuitively, (1.1) says that if $\mathrm{d}t$ is replaced by a small finite time, the left- and right-hand sides are approximately equal and that the approximation is better the smaller $\mathrm{d}t$ is.

This modeling framework is entirely natural. Because M is a martingale,

$$E[dM(t)\big|\mathcal{F}(t)] = E[M(t+\mathrm{d}t) - M(t)\big|\mathcal{F}(t)] = 0.$$

Therefore

$$E\Big[\frac{\mathrm{d}S(t)}{S(t)}\Big|\mathcal{F}(t)\Big] = \alpha(t)\,\mathrm{d}t.$$

This means that $\alpha(t)\,\mathrm{d}t$ is the expected infinitesimal return on owning a share of the asset over the period $[t, t+\mathrm{d}t]$, conditional on the market history at time t. Therefore, from equation (1.1), $dM(t)$ is the fluctuation of the return about this conditional mean. Essentially, M is *the source of the random fluctuations in the price; we say that the noise M drives the evolution of S*. Using paradigm (1.1), we can break down price modeling into the separate problems of modeling α and M.

1.3 Models based on Brownian motion, a review

From the perspective of equation (1.1), the theory of stochastic integrals with respect to Brownian motion is a mechanism for producing a large and flexible class of martingales to use for M. Let us recall in broad outline, how this theory goes. We start with a continuous process W, namely Brownian motion, which is not just a martingale, but a process with independent and stationary increments. It is assumed that W is adapted to $\{\mathcal{F}(t);\, t \geq 0\}$ and that future increments $W(t+h)-W(t)$, $h \geq 0$ are independent of $\mathcal{F}(t)$. We then define $\int_0^t \alpha(s)\,\mathrm{d}W(s)$ for processes α adapted to $\{\mathcal{F}(t);\, t \geq 0\}$ and satisfying $E[\int_0^T \alpha^2(s)\,\mathrm{d}s] < \infty$ for all T. The definition proceeds in two steps. First, we define the integral if α has the form:

$$\alpha(t) = \sum_{i=1}^{t} \alpha_i \mathbf{1}_{(t_{i-1},t_i]}(t)$$

where $t_0 < t_1 < \cdots < t_n$ and α_i is $\mathcal{F}(t_{i-1})$-measurable for each i. The definition is:

$$\int_0^t \alpha(s)\,\mathrm{d}W(s) := \sum_{i=1}^{n} \alpha_i [W(t_i \wedge t) - W(t_{i-1} \wedge t)]$$

This is the accumumlated return, up to time t, from betting amount α_i on the increment of W over every interval $[t_{i-1} \wedge t, t_i \wedge t]$. We then show this integral satisfies the Itô isometry,

$$E\Big[\Big(\int_0^t \alpha(s)\,\mathrm{d}W(s)\Big)^2\Big] = E\Big[\int_0^t \alpha^2(s)\,\mathrm{d}s\Big],$$

use this isometry to extend the definition to more general, adapted integrands, and obtain in this way a large family of martingales of the form,

$$M(t) := \int_0^t \sigma(s)\,\mathrm{d}W(s),$$

where $\sigma(t)$ is adapted to $\{\mathcal{F}(t);\, t \geq 0\}$ and $E[\int_0^T \sigma^2(s)\, \mathrm{d}s] < \infty$ for all T.

When $\int_0^t \sigma(s)\, \mathrm{d}W(s)$ is used for M in (1.1), the price model becomes:

$$dS(t) = \alpha(t)S(t)\, \mathrm{d}t + S(t)\, \sigma(t)\mathrm{d}W(t). \tag{1.2}$$

Here we have expressed $dM(t)$ as $\sigma(t)\mathrm{d}W(t)$; formally it consists of a Gaussian term $dW(t)$, with mean zero and variance $\mathrm{d}t$, *independent* of the past, times a volatility factor $\sigma(t)$ that is known at time t. For this class of price equations, the job of modeling reduces to choosing α and σ. (Pricing derivatives requires taking expectations with respect to a risk-neutral measure, and we found that this measure does not depend on α. Therefore, for pricing we really only need to model volatility.)

1.4 The problem of modeling jumps

As we mentioned in the Introduction, the main constraint of model (1.2) above is continuity; $W(t)$, $M(t) = \int_0^t \sigma(s)\, \mathrm{d}W(s)$, and, consequently, the price $S(t)$ solving (1.2) are all **continuous functions of** t with probability one.

Of course, in reality *prices move in steps of discrete size.* So long as these steps are small, continuous models of form (1.2) should be okay, if returns over small intervals are approximately Gaussian. But occasionally, prices take large, sudden and unexpected jumps, such as a market shock, and returns might not be Gaussian.

Therefore, one would like to allow jumps in price models. This will not only incorporate the phenomenon of sudden large jumps, but will also offer a richer family of models for fitting the empirically observed, statistical behavior of prices.

Price models with jumps can be obtained by introducing jumps into the noise, M, in equation (1.1). To proceed we need to know **how to define martingales with jumps** and we need **a theory to interpret and solve stochastic differential equations with jump terms.**

The strategy for carrying this out parallels stochastic integration theory for Brownian motion. One starts with *independent increment* processes, $X(t)$, that are martingales and that have jumps. The simplest example is *the compensated Poisson process.* Then one defines stochastic integrals, $\int_0^t \sigma(s)\, \mathrm{d}X(s)$, and **establishes conditions on** $\sigma(\cdot)$ **so that these integrals are martingales.** This produces a large class of martingales with jumps to use as the driving noise in price equations. Finally, one **extends Itô calculus to stochastic integrals with jumps.** This calculus can then be used to analyze derivatives based on the new price models.

The compensated Poisson process is derived from the Poisson process. To understand it, we start out with the Poisson process.

Remark 1.4.1. *As mentioned in the Introduction, we can also have stock price jumps in the case of dividend payments. In this case the stock will be modeled as followed:*

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t) + S(t-)dJ(t),$$

where if we let $0 < t_1 < \cdots < t_n < T$ be the dividend payment days and $\alpha_i, i = 1, \cdots, n$ be the dividend percentage (that is at time t_i the dividend paid is $\alpha_i S(t_i-)$) then

$$J(t) = \sum_{i=1}^{n} -\alpha_i \mathbf{1}_{\{t \geq t_i\}}.$$

(See also Shreve's Section 5.5.3)

The point is here $J(t)$ is NOT a martingale, nor can it be made into a martingale by being compensated as a compensated Poisson process. Thus the dividend payment stock's model does not fall into the theory of jump processes discussed in this Chapter as far as the martingale aspect is concerned. However, the mathematical tools developed here can be used to analyze the jump part in the dividend paying stock in a similar way as we present later in this Chapter.

1.5 The most basic model of jump processes: Poisson process

1.5.1 Heuristics about Poisson process

We think of Poisson process as followed: suppose that we have an alarm clock that will ring after a *random* time τ, where τ is exponentially distributed with some mean $\frac{1}{\lambda}$. We keep account of the value of the Poisson process at any time t by the notation $N(t)$. At time 0, we set the alarm clock and set $N(0) = 0$. When the alarm rings, we increase the value of N by 1, that is we set $N(\tau) = 1$ and repeat the whole process (i.e. we reset the alarm clock and increase the value of N by 1 the next time the clock rings). The resulting process $N(t)$ is then a Poisson process with rate λ. We observe that the larger λ is, the clock would be likely to ring sooner and the more jumps would likely happen in a given time interval [0,T]. It is also clear that $N(t)$ is constant in between the "ring" times.

1.5.2 Formal mathematical definition

a. τ (as a R.V.) is said to be exponentially distributed with rate λ if it has the density

$$f(t) = \lambda e^{-\lambda t}\mathbf{1}_{(t \geq 0)}.$$

It follows that $E(\tau) = \frac{1}{\lambda}$ and $Var(\tau) = \frac{1}{\lambda^2}$. An important property of exponential random variable is the *memoryless property*:

$$\mathbb{P}(\tau > t + s | \tau > s) = \mathbb{P}(\tau > t).$$

b. Let $\tau_i, i = 1, 2, ...$ be a sequence of i.i.d. Exponential(λ). Let $S_k := \sum_{i=1}^{k} \tau_i$. The Poisson process $N(t)$ with rate λ is defined as:

$$N(t) = \sum_{i=1}^{\infty} \mathbf{1}_{(t \geq S_i)}.$$

τ_i is called the *inter-arrival time.* It is the wait time from the $(i-1)^{th}$ jump to the i^{th} jump. S_i is called the *arrival time.* It is the time of the i^{th} jump.

1.5.3 Important basic properties

a. Distribution: $N(t)$ is has distribution Poisson(λt), that is

$$\mathbb{P}(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Proof. Let $S_n = \sum_{i=1}^{n} \tau_i$ be the arrival time, then

$$\begin{aligned} \mathbb{P}(N(t) = k) &= \mathbb{P}(S_{k+1} > t, S_k \leq t) \\ &= \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_{k+1} > t, S_k > t) = \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t). \end{aligned}$$

From Shreve's Lemma 11.2.1, S_n has Gamma(λ, n) distribution. That is, it has the density:

$$g_n(s) = \frac{(\lambda s)^{n-1}}{(n-1)!}\lambda e^{-\lambda s}, s \geq 0.$$

It is a straight forward matter of integration now to verify that

$$\mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

The integration can be tedious, however. Another way to verify it is as followed: Denote $f(t) := \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t)$ and note that $f(t)$ satisfies the following ODE:

$$\begin{aligned} f'(t) &= g_k(t) - g_{k+1}(t) = \frac{(\lambda t)^{k-1}}{(k-1)!}\lambda e^{-\lambda t} - \frac{(\lambda t)^k}{k!}\lambda e^{-\lambda t} \\ f(0) &= 0. \end{aligned}$$

It is clear that $f(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ is the *unique* solution to the above ODE. The verification is complete. ■

b. $N(t)$ has independent increment. That is if we denote $\mathcal{F}_t$ to be the filtration generated by $N(s), 0 \leq s \leq t$ then for all $t \leq t_1 < t_2$, $N(t_2) - N(t_1)$ is independent of $\mathcal{F}_t$.

Heuristic reason: Let $0 \leq s < t$. Clearly $N(t) - N(s)$ counts the number of jumps starting from time s. Given all the information up to time s, what is the distribution of the first jump time after s? That is, we want to compute $\mathbb{P}(S_{N(s)+1} \geq t|\mathcal{F}_s)$, where S_n is the arrival time as defined in Shreve (11.2.4). Note that since $N(s)$ represents the number of jumps up to time s, $S_{N(s)+1}$ is exactly the time of *the first jump after time s.*

But this is the same as computing $\mathbb{P}(\tau_{N(s)+1} \geq t - S_{N(s)}|\tau_{N(s)+1} \geq s - S_{N(s)})$. Note that $S_{N(s)}$ here represents the time of the *last jump before time s,* and $\tau_{N(s)+1}$ is *the wait time between the last jump before time s and the first jump after time s.* So $\mathbb{P}(\tau_{N(s)+1} \geq t - S_{N(s)}|\tau_{N(s)+1} \geq s - S_{N(s)})$ asks for the probability that we have to wait until after time t for the first jump after time s, given that we know we have waited up until time s since the last jump before s, which has the same content as $\mathbb{P}(S_{N(s)+1} \geq t|\mathcal{F}_s)$.

Note also that $\mathbb{P}(\tau_{N(s)+1} \geq t - S_{N(s)}|\tau_{N(s)+1} \geq s - S_{N(s)}) = \mathbb{P}(\tau_{N(s)+1} \geq t - s + s - S_{N(s)}|\tau_{N(s)+1} \geq s - S_{N(s)})$. Since $\mathcal{F}_s$ is given, $N(s)$ should be looked at as a constant here. But from the memoryless property of $\tau_{N(s)+1}$, we get

$$\mathbb{P}(\tau_{N(s)+1} \geq t - s + s - \tau_{N(s)}|\tau_{N(s)+1} \geq s - \tau_{N(s)}) = \mathbb{P}(\tau_{N(s)+1} \geq t - s).$$

That is, the first jump time after s can be looked at as an exponential clock starts at time s, hence independent of the past information. Using the independence of interarrival times, it is clear now that the increments of $N(t)$ after time s is independent of the information up to time s. ■

c. $N(t)$ has stationary increment. More specifically, $N(t) - N(s)$ has distribution Possion($\lambda(t-s)$).

Heuristic reason: It follows from the same arguments of part b. ■

1.6 Generalizations of Poisson process

1.6.1 Compound Poisson process

The Poisson process we introduced has the satisfactory property that it jumps at *random* times. However, each of the jump is by definition of length 1, which is rather restrictive. It is desirable in terms of being realistic to have random jumps in our model. To that end, we proceed as followed.

Let $N(t)$ be a Poisson process with rate λ and let $Y_0 = 0, Y_i, i = 1, 2, ...$ be i.i.d.(and also independent of N(t)) with $E(Y_i) = \mu$. Define

$$Q(t) = \sum_{i=0}^{N(t)} Y_i,$$

then $Q(t)$ is called a compound Poisson process. Similar to a Poisson process, $Q(t)$ also has the basic properties of *independent and stationary increments.* We *do not know* the specific distribution of $Q(t) - Q(s)$ (it depends on the distribution of Y_i's , of course), but we do know that $E(Q(t) - Q(s)) = \mu\lambda(t - s)$.

1.6.2 Pure jump process

Poisson process and compound Poisson process are examples of pure jump processes.

Definition 1.6.1. *A stochastic process* $\{J(t); t \geq 0\}$ *is called a* **pure jump process** *if its sample paths are* **right-continuous** *and* **piecewise-constant***. (Recall that this entails that each sample path of* $J(t)$ *admits only a finite number of jumps in any finite time interval.) A pure jump process* N *is called a* **counting process** *if* $N(0) = 0$ *and it its jumps* **all have magnitude 1;** *hence it can only increase, and is always integer-valued. If* N *is a counting process,* $N(t)$ *counts the number of jumps in the time interval* $[0, t]$.

Remark 1.6.2. *In stochastic integration theory, the definition of a pure jump process is more general than the one here and allows infinite numbers of jumps in finite time intervals.*

Remark 1.6.3. *By right continuity, a pure jump process* $J(t)$ *CANNOT jump at time 0, which is always the conventional time that we start observing the process.*

1.6.3 Levy process

So far the three processes that we have encountered: Brownian motion, Poisson and compound Poisson processes have these three properties in common:

- Its value at time 0 is 0 : $X(0) = 0$.
- It has càdlàgpath.
- It has stationary and independent increments.

A process $X(t)$ is said to be a Levy process *starting at 0* if it satisfies these three properties (clearly if we change the first property to $X(0) = x$ then we would get a Levy process starting at x). Brownian motion is an example of a continuous Levy process and Poisson process is an example of a pure jump Levy process. Indeed, Brownian motion, compound Poisson process and pure jump process may be thought as “building blocks" of a Levy process (See Levy-Ito decomposition on Wikipedia, for example).

A rather simple but important property of Levy process is as followed: If $X_1, X_2, ...X_n$ are *independent* Levy process then $\sum_{i=1}^{n} X_i$ is a Levy process. In particular, if we consider $S(t) = X(t) + Q(t)$, where $X(t)$ is a Geometric Brownian motion with the drift μ and volatility σ *constant*), $Q(t)$ a compound Poisson process then $S(t)$ is a Levy process.

1.7 Martingale property

In Math finance, we always require the discounted underlying to be a martingale, so that no arbitrage can happen. As mentioned, the Levy process is intimately connected to our stock models, so it's natural to first study the martingale property of Levy processes.

1.7.1 Levy process

Let $X(t)$ be a Levy process and $\mathcal{F}(t)$ its filtration. If $E(X(1)) = \mu$ then it can be shown that $E(X(t)) = \mu t$. Similarly, if $Var(X(1)) = \sigma^2$ then it can be shown that $Var(X(t)) = \sigma^2 t$. Since X has independent increment, one can check that

$$\begin{aligned} Y(t) &= X(t) - \mu t; \\ Z(t) &= (X(t) - \mu t)^2 - \sigma^2 t \end{aligned}$$

are martingales with respect to $\mathcal{F}(t)$.

1.7.2 Brownian motion

Let $W(t)$ be a Brownian motion and $\mathcal{F}(t)$ its filtration. Then $W(t)$ and $W^2(t) - t$ are martingales w.r.t. $\mathcal{F}(t)$. More importantly, we have the following exponential martingale associated with Brownian motion:

$$Z(t) = e^{\sigma W_t - \frac{1}{2}\sigma^2 t}.$$

1.7.3 Poisson process

Let $N(t)$ be a Poisson process and $\mathcal{F}(t)$ its filtration. Then $N(t) - \lambda t$ (called a *compensated Poisson process*) and $(N(t) - \lambda t)^2 - \lambda t$ are martingales w.r.t. $\mathcal{F}(t)$. We also have the following exponential martingale associated with $N(t)$:

$$Z(t) = \exp\big(iuN(t) - \lambda t(e^{iu} - 1)\big), \forall u \in \mathbb{R}.$$

1.7.4 Compound Poisson process

Let $Q(t)$ be a compound Poisson process and $\mathcal{F}(t)$ its filtration. That is

$$Q(t) = \sum_{i=0}^{N_t} Y_i,$$

where $N(t)$ is a Possion(λ) process. Let $E(Y_1) = \mu$ and $Var(Y_1) = \sigma^2$. One can check that

$$E(Q_t) = \lambda \mu t$$

and

$$Var(Q_t) = \lambda t(\sigma^2 + \mu^2).$$

Then

$$Q(t) - \mu\lambda t$$

(called a *compensated compound Poisson process*) and

$$(Q(t) - \mu\lambda t)^2 - \lambda t(\sigma^2 + \mu^2)$$

are both martingales w.r.t. $\mathcal{F}(t)$.

Let $\phi(u) := \mathbb{E}(e^{iuY_1})$ be the characteristic function of Y_i. Then we also have the following exponential martingale associated with $Q(t)$:

$$Z(t) = \exp\big(iuQ(t) - \lambda t(\phi(u) - 1)\big), \forall u \in \mathbb{R}.$$

1.8 Lebesgue-Stieltjes integral

1.8.1 Motivation

Now that we have introduced Poisson process, it is easy to see how to incorprate jumps into the current Black-Scholes stock model. Specifically, let $X(t)$ be a geometric Brownian motion:

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

and $N(t)$ a Poisson process. Then defining the stock process as $S(t) := X(t) + N(t)$ already gives us a stock price that jumps at random times, and in between the jumps behave as a geometric Brownian motion.

Let Δ_t represent the number of shares of $S(t)$ we hold at time t. As you might remember from the previous material, we need to know how to evaluate the integral $\int_0^t \Delta_s dS_s$, since it is connected with the value of a portfolio that has S as a component. It is reasonable to expect that

$$\int_0^t \Delta_s dS_s = \int_0^t \Delta_s dX_s + \int_0^t \Delta_s dN_s,$$

and we already know how to evaluate $\int_0^t \Delta_s dX_s$ from the chapter on Ito integral. It remains to define $\int_0^t \Delta_s dN_s$.

For each event ω, the path $N_t(\omega)$ (as a function of t) belongs to a special class of functions called *the functions of bounded variation*. For this reason, $\int_0^t \Delta_s dN_s$ is defined via the concept of Lebesgue-Stieltjes integral of classical analysis. It still has some subtleties, however, mostly due to the facts that $N(t)$ has *jumps*, so the regularity (left or right continuity) of the integrand Δ_t affects the value of the integral. For this reason, we will review some basic aspects of the Lebesgue-Stieltjes integral with respect to càdlàgintegrator in the next section.

1.8.2 Mathematical preliminaries; right-continuous functions with left limits

1. *Limits, continuity and jumps.* Let $f(t)$ be a function defined for $t \geq 0$. The right limit of f at $t \geq 0$ is

$$f(t+) := \lim_{s \downarrow t} f(s), \quad \text{assuming it exists.}$$

The left limit of f at $t > 0$ is

$$f(t-) := \lim_{s \uparrow t} f(s), \quad \text{assuming it exists.}$$

As a convention, we set $f(0-) = f(0)$. The jump of f at t is the difference of these limits and is denoted

$$\triangle f(t) = f(t+) - f(t-)$$

A function f is said to be right-continuous with left limits if $f(t) = f(t+)$ for all t and if $f(t-)$ exists for all t. Such functions are sometimes called *càdlàg* functions in the literature. It is worth knowing this term so we shall use it. It is an acronym of the French phrase meaning 'right-continuous with left limits': *continu à droite, limites à gauche*.

If we are to allow jumps in price models, we need a convention for *what the price is at the exact time of the jump.* Our convention shall be that all prices are càdlàg functions. Hence $S(t) = S(t+)$ is the price that the asset jumps to at time t, $S(t-)$ is the price immediately before the jump, and $\triangle S(t) = S(t) - S(t-)$ is the size of the jump.

The same convention will be imposed on the driving martingale M in our price models, since it is the martingale M that causes the jump. Likewise, in the theory of stochastic integration that we develop, *stochastic integrals will have càdlàg paths.*

Along with these conventions we need to re-interpret the intuitive meaning of $dS(t)$. You should think of this as the increment $S(t + \mathrm{d}t) - S(t-)$; Of course, this coincides with $S(t + \mathrm{d}t) - S(t)$ at times t at which $S(t)$ is continuous; if not, *it captures the jump at time* t. Note that, as usual, the identity, $dS(t) = S(t + \mathrm{d}t) - S(t-)$, does not have a strict meaning, since $\mathrm{d}t$ does not have a strict meaning, but it is a correct guide to thinking about the movements of S over small time intervals.

2. *Important facts to know about càdlàg functions.* Let f be càdlàg:

(i) The function $t \to f(t-)$ is left-continuous.

(ii) The set of points at which f is *not* continuous is either finite or countably infinite.

(iii) Since $f(t)$ and $f(t-)$ differ only at points at which f is not continuous and there are only countably many such points, $\int_0^T f(t-)g(t)\,\mathrm{d}t = \int_0^T f(t)g(t)\,\mathrm{d}t$, for any $T > 0$ and g.

1.8.3 Lebesgue-Stieltjes integrals for increasing, right-continuous integrators

A. Let $G(t)$, $t \geq 0$, be a function that can be written in the form

$$G(t) = A_1(t) - A_2(t), \text{ where } A_1 \text{ and } A_2 \text{ are } \textit{increasing and right-continuous.} \tag{1.3}$$

Since A_1 and A_2 are increasing, $A_i(t-) = \lim_{s\uparrow t} A_i(s)$ exists automatically for all $t > 0$, and for each $i = 1, 2$. Hence A_1, and A_2, and, therefore, G, are all càdlàg.

It is easy to see that if $G_1(t)$ and $G_2(t)$ are both functions with the property (1.3), then so is any linear combination $\alpha G_1(t) + \beta G_2(t)$.

A function G that can be written as the difference of two increasing function has the special property of being a function of *bounded variation*. You will learn what this means in a homework exercise. Conversely, any function of bounded variation may be written as the difference of increasing functions. Therefore, we shall summarize the condition (1.2) by saying G is a càdlàg function of bounded variation.

Examples of functions of the form (1.3).

(i) G is a right-continuous, piecewise constant functions of the form

$$G(t) = \sum_0^n a_k \mathbf{1}_{[t_k, t_{k+1})}(t)$$

where $0 = t_0 < t_1 < \cdots < t_n$, and $t_{n+1} = \infty$.

(ii) $G(t)$ is a differentiable function, $G(t) = G(0) + \int_0^t g(s)\,\mathrm{d}s$.

(iii) $G(t) = \int_0^t g_1(s)\,\mathrm{d}s + G_2(t)$, where G_2 is right-continuous and piecewise constant, as in (i).

Explanation : (i) The function $\mathbf{1}_{[a,\infty)}(t)$ is right-continuous and increasing. Therefore $\mathbf{1}_{[a,b)}(t) = \mathbf{1}_{[a,\infty)}(t) - \mathbf{1}_{[b,\infty)}(t)$ is a function of form (1.3). Because G as defined in (i) is a linear combination of functions of the form (1.3), it also has this form.

As for (ii), observe that

$$G(t) = G(0) + \int_0^t g(s)\,ds = G(0) + \int_0^t |g(s)|\mathbf{1}_{\{g(s)\geq 0\}}\,\mathrm{d}s - \int_0^t |g(s)|\mathbf{1}_{\{g(s)<0\}}\,\mathrm{d}s$$

decomposes G into the difference of two continuous, increasing functions.

(iii) is a consequence of (i) and (ii). ◇

B. If G is a càdlàg function of bounded variation, there is a natural way to define integrals, which we shall denote,

$$\int_0^t H(s)\,\mathrm{d}G(s),$$

built on the increments of G. In the mathematical literature, these are called Lebesgue-Stieltjes integrals.

1.8.4 Lebesgue-Stieltjes integral for left-continuous, piecewise constant integrands

A function $H(t)$, $t \geq 0$ is left-continuous and piecewise constant if it has the form:

$$H(s) = \sum_{i=1}^{n} c_i \mathbf{1}_{(t_{i-1},t_i]}(s) + c_{n+1}\mathbf{1}_{(t_n,\infty)}(s) \tag{1.4}$$

where $0 = t_0 < t_1 < \cdots < t_n$. For convenience let $t_{n+1} = \infty$. For this H, define

$$\int_{(0,t]} H(s)\,\mathrm{d}G(s) := \sum_{i=1}^{n+1} c_i \Big[G(t_i \wedge t) - G(t_{i-1} \wedge t)\Big] \tag{1.5}$$

Usually, we will write the integral in (1.5) as $\int_0^t H(s)\,\mathrm{d}G(s)$.

Motivation and comments.

1. This definition should be no surprise; the idea is to multiply the value of H on each interval $(t_{i-1}, t_i]$ by the increment of G over that interval. If $G(s) = s$, then

$$\int_{(0,t]} H(s)\,\mathrm{d}G(s) = \sum_{i=1}^{n+1} c_i \Big[t_i \wedge t - t_{i-1} \wedge t\Big] = \int_0^t H(s)\,\mathrm{d}s,$$

which is the usual Riemann (or Lebesgue) integral. If instead we replaced G by a Brownian motion W, we would get the Itô integral of H.

2. The fact that we used intervals of the form $(t_{i-1}, t_i]$ in the definition of H in (1.5), so that H is left-continuous, **is not an accident and is tied up with the assumption that G is right-continuous**.

First, we model G as being right-continuous according to our intuition that a shock cannot be predictable (you can observe the behavior of the stock up until the time of the shock - the jump time - but you will not be able to tell the value of the stock after the jump based on your observation).

Second, the result in Example 1 below also works well with our intuition: the change in the portfolio value after the shock is the change in the stock price ($\Delta G(\tau)$) multiplied with the number of shares we hold at the time of the shock ($H(\tau)$). Note that if we use a right continuous integrand $H(s)$, we will NOT get a similar result. This explains the choice of left continuous integrand for our basic building block of Lebesgue-Stieltjes integral w.r.t. right continuous integrator.

Example 1. Consider the simplest example, where G is piecewise constant with a single jump at time τ:

$$G(t) = \begin{cases} a_0, & \text{if } 0 \leq t < \tau; \\ a_1, & \text{if } t \geq \tau, \end{cases} \tag{1.6}$$

where $a_0 \neq a_1$. We will show that

$$\int_{(0,t]} H(s)\,\mathrm{d}G(s) = \begin{cases} H(\tau)\triangle G(\tau), & \text{if } t \geq \tau; \\ 0, & \text{if } t < \tau. \end{cases} \qquad \diamond \tag{1.7}$$

Let H be given as in (1.4). Observe that

$$G(t_k \wedge t) - G(t_{k-1} \wedge t) = \begin{cases} a_1 - a_0 = \triangle G(\tau), & \text{if } t_{k-1} \wedge t < \tau \leq t_k \wedge t; \\ 0, & \text{otherwise.} \end{cases}$$

Also, notice that if $t_{k-1} < \tau \leq t_k$, then $H(\tau) = c_k$, since H has the constant value c_k on $(t_{k-1}, t_k]$. Thus each term in the sum on the right-hand side of (1.5) is zero, except if $t_{k-1} \wedge t < \tau \leq t_k \wedge t$, and for this k, $c_k\Big[G(t_k \wedge t) - G(t_{k-1} \wedge t)\Big] = H(\tau)\triangle G(\tau)$. If $t < \tau$, there is no k such that $t_{k-1} \wedge t < \tau \leq t_k \wedge t$, and so the sum in (1.5) is zero. If $t \geq \tau$, the sum contains one non-zero term, which we have shown equals $H(\tau)\triangle G(\tau)$. This proves (1.7).

Example 2. Let $G(t) = G(0) + \int_0^t g(s)\,\mathrm{d}s$, that is, G is differentiable and $G'(t) = g(t)$. Let H be as in (1.4). Then

$$\begin{aligned} \int_{(0,t]} H(s)\,\mathrm{d}G(s) &= \sum_{i=1}^{n+1} c_i \Big[G(t_i \wedge t) - G(t_{i-1}) \wedge t)\Big] \\ &= \sum_{i=1}^{n+1} c_i \cdot \int_{t_{i-1}\wedge t}^{t_i \wedge t} g(s)\,\mathrm{d}s \\ &= \int_0^t \Big[\sum_{i=1}^{n+1} c_i \mathbf{1}_{(t_{i-1},t_i]}(s)\Big]\, g(s)\,\mathrm{d}s \\ &= \int_0^t H(s)g(s)\,\mathrm{d}s, \qquad \diamond \end{aligned} \tag{1.8}$$

1.8.5 Lebesgue-Stieltjes integral for Borel measurable integrands

The following theorem states that definition (1.5) can be extended in a unique way to a large class of integrands. The proof requires tools of measure theory beyond the scope of this course. It is only important for you to know what the theorem says. This will usually be enough for you to understand what is going on if you encounter Lebesgue-Stieltjes integrals when reading mathematical finance literature.

Theorem 1. *There is a* **unique** *way to assign to each bounded, Borel measurable function H and bounded variation function G, an integral $\int_0^t H(s)\,\mathrm{d}G(s)$ for $t > 0$, with the following properties:*

(i) $\int_0^t H(s)\,\mathrm{d}G(s)$ *is defined by (1.5) when* H *has the form given in (1.4);*

(ii) (linearity) $\int_0^t [a_1H_1(s)+a_2H_2(s)]\,\mathrm{d}G(s) = a_1\int_0^t H_1(s)\,\mathrm{d}G(s) + a_2\int_0^t H_2(s)\,\mathrm{d}G(s)$*; and*

$$\int_0^t H(s)\,\mathrm{d}\big[aG_1(s)+bG_2(s)\big] = a\int_0^t H(s)\,\mathrm{d}G_1(s) + b\int_0^t H(s)\,\mathrm{d}G_2(s);$$

(iii) (exchange of limit and integral) Assume $H(s) = \lim_{n\to\infty} H_n(s)$ *for all* s*, where, for some* $K<\infty$, $\big|H_n(s)\big| \le K$ *for all* n *and* $s \ge 0$*. Then*

$$\int_0^t H(s)\,\mathrm{d}G(s) = \lim_{n\to\infty}\int_0^t H_n(s)\,\mathrm{d}G(s).$$

Unfortunately, Theorem 1 only assures us that the Lebesgue-Stieltjes integral can be defined in a meaningful way; it does not directly say how to compute one. Fortunately, in most situations we shall encounter, the Lebesgue-Stieltjes integral can be reduced to familiar and easy-to-handle objects.

$G(t)$ is continuously differentiable

When $G(t) = G(0) + \int_0^t g(s)\,ds$, and H is *any* bounded, Borel function

$$\int_0^t H(s)\,\mathrm{d}G(s) = \int_0^t H(s)g(s)\,\mathrm{d}s. \tag{1.9}$$

We saw this is true in Example 2, when H is a left-continuous, piecewise constant function. This shows that the right-hand side of (1.9) satisfies property (i) of Theorem 1. It also satisfies the properties in (ii), as one can show by direct calculation, and it satisfies property (iii) because of the properties of the Legesgue integral (full explanation omitted!). Thus $\displaystyle\int_0^t H(s)g(s)\,\mathrm{d}s$ must coincide with $\displaystyle\int_0^t H(s)\,\mathrm{d}G(s)$, because, by Theorem 1, the latter integral is uniquely determined by properties (i)—(iii).

$G(t)$ is a pure jump function

Let G be piecewise-constant of the form

$$G(t) = a_0\mathbf{1}_{[0,\tau_1)}(t) + a_1\mathbf{1}_{[\tau_1,\tau_2)}(t) + \cdots + a_n\mathbf{1}_{[\tau_{n-1},\tau_n)} + \cdots$$

Thus G is constant except for jumps at the points $\tau_1 < \tau_2 < \cdots$. Notice that G is defined so as to be càdlàg.

In this case, also for *any* bounded, Borel function H we have

$$\int_0^t H(s)\,\mathrm{d}G(s) = \sum_{j;\tau_j \le t} H(\tau_j)\triangle G(\tau_j). \tag{1.10}$$

To emphasize, the sum is over the jump times of G that occur at time t or before. Since $\triangle G(s) = 0$ if s is not equal to any jump time, it is convenient to write this formula as

$$\int_0^t H(s)\,\mathrm{d}G(s) = \sum_{0<s\le t} H(s)\triangle G(s). \tag{1.11}$$

This result is derived from Theorem 1 by showing that the expression on the right-hand side of (1.10) satisfies properties (i)—(iii) of the Theorem. For simplicity, consider the case when G jumps only at a finite number of times. Properties (ii) and (iii) are easy to verify directly. In example 1, we have verified property (i) for the case when G has a single jump and H is a left-continuous, piecewise constant function. Thus, Theorem 1 implies that (1.10) is true if G has just one jump. If G has multiple jump times, we can write $G(s) = G_1(s) + G_2(s) + \cdots + G_n(t)$, where each G_i is a piecewise-constant, càdlàg function jumping at only one time, and use property (ii) of Theorem (1) to deduce (1.10).

Combination of the above two cases

We will encounter the case $G(t) = \int_0^t g_1(s)\,\mathrm{d}s + G_2(s)$, where G_2 is piecewise constant, càdlàg, as in (1.8.5). Then, by property (ii) of Theorem 1,

$$\int_0^t H(s)\,\mathrm{d}G(s) = \int_0^t H(s)g_1(s)\,\mathrm{d}s + \sum_{0<s\le t} H(s)\triangle G_2(s).$$

However, notice that $\triangle G(s) = \triangle G_2(s)$ for all s because the integral term in G is continuous. Therefore we can rewrite the formula as

$$\int_0^t H(s)\,\mathrm{d}G(s) = \int_0^t H(s)g_1(s)\,\mathrm{d}s + \sum_{0<s\le t} H(s)\triangle G(s). \tag{1.12}$$

If the jumps of G occur at times $0 < \tau_1 < \tau_2 < \ldots$, then $G'(s) = g_1(s)$ exists at any time s not equal to a jump time. Thus,

$$\int_0^t H(s)\,\mathrm{d}G(s) = \sum_i \int_{\tau_{i-1}\wedge t}^{\tau_i\wedge t} H(s)G'(s)\,\mathrm{d}s + \sum_{s\le t} H(s)\triangle G(s). \tag{1.13}$$

This is the easiest form to use.

Because of item (iii) on page 4, if G is differentiable and $G'(s) = g(s)$, and if H is cádlàg, $\int_0^t H(s-)\,\mathrm{d}G(s) = \int_0^t H(s-)g(s)\,\mathrm{d}s = \int_0^t H(s)g(s)\,\mathrm{d}s = \int_0^t H(s)\,\mathrm{d}G(s)$. But if G has jumps, the two integrals may not agree, as we shall see by example later.

Remark Let $\{W(t)(\omega);\, t \geq 0, \omega \in \Omega\}$ be a Brownian motion. The definition of stochastic integrals with respect to W was a fairly complicated affair. We did need such a complex definition? Why not just define $\int_0^t H(s)\,\mathrm{d}W(s)$ by applying Theorem 1 to $W(\cdot)(\omega)$ for each ω and be done with it? Or did we use a complicated definition only because of some clandestine conspiracy to make the lives of math finance students miserable? Solemn vows of secrecy forbid me from answering the last question, but the first two are easy to answer. We cannot apply Theorem 1 to Brownian motion because, with probability one, the paths of Brownian motion *are not functions of bounded variation.* This is due to the fact that Brownian motion has non-trivial quadratic variation. You will get to explore this point in a homework exercise. It is absolutely essential to your understanding of stochastic integration.

1.8.6 Stochastic Integration

Let $\{X(t)(\omega);\, t \geq 0, \omega \in \Omega\}$ be a stochastic process defined on a probability space $(\Omega, I\!\!P)$. Assume that for every ω, $X(t,\omega)$ is a bounded variation function, as a function of t. Let $\{\alpha(t)(\omega);\, t \geq 0, \omega \in \Omega\}$ be another stochastic process. Then,

$$\int_0^t \alpha(s)(\omega)\,\mathrm{d}X(s)(\omega)$$

will always represent the Lebesgue-Stieltjes integral of the process α with respect to the process X. Usually, we suppress the explicit dependence on ω and simply write the integral as $\int_0^t \alpha(s)\,\mathrm{d}X(s)$.

1.9 Stochastic integration w.r.t. semi-martingales

1.9.1 Definition and examples

Let $X(t) = \int_0^t \gamma(s)dW_s + A(t)$, where $W(t)$ is a Brownian motion with respect to a filtration $\mathcal{F}(t)$, $\gamma(t) \in \mathcal{F}(t)$ be such that $\int_0^t \phi(s)dW_s$ is defined and $A(t) \in \mathcal{F}(t)$ a process of bounded variation. $X(t)$ is called a semi-martingale w.r.t. $\mathcal{F}(t)$.

Definition 1.9.1. *Let $\phi(t) \in \mathcal{F}(t)$ be so that $\int_0^t \phi(s)\gamma(s)dWs$ and $\int_0^t \phi(s)dA(s)$ are defined. Then we define*

$$\int_0^t \phi(s)dX(s) := \int_0^t \phi(s)\gamma(s)dWs + \int_0^t \phi(s)dA(s).$$

It is important to note here that $\int_0^t \phi(s)\gamma(s)dWs$ is an Ito integral, which is not defined path-wise (since $W(t)$ has infinite variation) and $\int_0^t \phi(s)dA(s)$ is a Lebesgue-Stieltjes integral, which is defined pathwise using the definition of Section (1.8).

Example 1.9.2. *(i) Let $X(t)$ be a compensated compound Poisson process, i.e. $X(t) = Q(t) - \lambda\mu t$ where $Q(t)$ is a compound Poisson process. Let S_k be the jump times of $Q(t)$. Then*

$$\int_0^t \phi(s)dX(s) = \sum_i \phi(S_i)Y_i\mathbf{1}_{(S_i \leq t)} - \int_0^t \lambda\mu\phi(s)ds.$$

(ii) Let $X(t) = W(t) + J(t)$, where $J(t)$ is a pure jump procress. Then

$$\int_0^t \phi(s)dX(s) = \int_0^t \phi(s)dWs + \sum_{0<s\leq t} \phi(s)\Delta J(s).$$

We understand the term $\sum_{0<s\leq t} \phi(s)\Delta J(s)$ as followed: for each event ω, let $0 < t_1(\omega) < t_2(\omega) < \ldots < t_{n(\omega)}(\omega) \leq t$ be the jump times of $J(t)$. (The fact that there are finitely many jumps in $[0,t]$ and there is no jump at $t = 0$ come from the definition of pure jump process). Also note that the number of jumps in $[0,t], n(\omega)$ is random. Then

$$\int_0^t \phi(s)dJ(s)(\omega) = \sum_{0<s\leq t} \phi(s)\Delta J(s) = \sum_{i=1}^{n(\omega)} \phi(t_i)[J(t_i) - J(t_i-)](\omega).$$

1.9.2 Martingale properties

Suppose we model our stock as

$$S(t) = \sigma W(t) + X(t),$$

where $W(t), X(t) \in \mathcal{F}(t)$ are independent, $W(t)$ is a Brownian motion and $X(t) = Q(t) - \lambda\mu t$ is a compensated compound Poisson process. Then $S(t)$ is a martingale. It is important for us then that if we denote $\phi(t)$ as the number of shares of S we hold at time t , $\int_0^t \phi(r)dS_r$ **is a martingale.**

From Ito integration, we know that if ϕ is an adapted process, then $\int_0^t \phi(s)dW(s)$ is a martingale. So it remains to ask if $\int_0^t \phi(s)dX(s)$ is also a martingale. However, this is not always the case. See Shreve's examples 11.4.4 and 11.4.6.

A suffficient condition for the stochastic integral w.r.t. a jump process (that is also a martingale) to be a martingale is that the integrand is left-continuous (and of course adapted). This is stated in Shreve's theorem 11.4.5. More generally, one can use a predictable integrand (a process that is the limit of a sequence of left-continuous processes) and the stochastic integral w.r.t. a jump martingale will still be a martingale.

In Shreve's example 11.4.6, the following process is considered:

$$\begin{aligned} X(t) &= \int_0^t \mathbf{1}_{[0,S_1]}(s)d(N(s) - \lambda s) \\ &= \int_0^t \mathbf{1}_{[0,S_1]}(s)dN(s) - \int_0^t \mathbf{1}_{[0,S_1]}(s)\lambda ds, \end{aligned}$$

where S_1 is the first jump time of $N(t)$. Note that the integrand here is left continuous.

For $t < S_1$, the integrand $\mathbf{1}_{[0,S_1]}(s) = 1$. On the other hand, $dN(s) = 0$ if $t < S_1$. Thus $X(t) = -\lambda t$.

For $t = S_1$, $\int_0^t \mathbf{1}_{[0,S_1]}(s)dN(s) = 1\Delta N(S_1) = 1$ while $\int_0^t \mathbf{1}_{[0,S_1]}(s)\lambda ds = \lambda S_1$. Thus $X(t) = 1 - \lambda S_1$.

For $s > S_1$, $\mathbf{1}_{[0,S_1]}(s) = 0$ thus $X(t) = 1 - \lambda S_1, t \geq S_1$.

We conclude that

$$\begin{aligned} X(t) &= -\lambda t \mathbf{1}_{(t<S_1)} + (1 - \lambda S_1)\mathbf{1}_{(t \geq S_1)} \\ &= N(t \wedge S_1) - \lambda(t \wedge S_1). \end{aligned}$$

Here we can use the fact that a stopped martingale is a martingale to conclude that $X(t)$ is a martingale since $N(t) - \lambda t$ is a martingale and the above formula showed that $X(t)$ is a stopped martingale.

Using a similar argument, we have

$$Y(t) = \int_0^t \mathbf{1}_{[0,S_1)}(s)d(N(s) - \lambda s) = -\lambda(t \wedge S_1).$$

Heuristically, $\mathbb{P}(S_1 > 0) = 1$ therefore, for $s < t$,

$$\begin{aligned} \mathbb{P}\big(-\lambda(t \wedge S_1) \leq -\lambda(s \wedge S_1)\big) &= 1; \\ \mathbb{P}\big(-\lambda(t \wedge S_1) < -\lambda(s \wedge S_1)\big) &> 0. \end{aligned}$$

Therefore $\mathbb{E}(-\lambda(t \wedge S_1)) < \mathbb{E}(-\lambda(s \wedge S_1))$ and $Y(t)$ is not a martingale. A rigorous proof is provided in Shreve's.

1.10 Ito's formula for jump processes

1.10.1 Ito's formula for one jump process

The most general jump process we will consider in this chapter has the following form:

$$X(t) = X(0) + \int_0^t \alpha(s)ds + \int_0^t \gamma(s)dWs + J(t),$$

where $J(t)$ is a pure jump process (Discussed in Section (1.6.2)). We also denote by $X^c(t)$ the continuous part of X, that is

$$X^c(t) = X(0) + \int_0^t \alpha(s)ds + \int_0^t \gamma(s)dWs.$$

Given a function $f \in C^2$, we would like to obtain a formula for $df(X(t))$. We have the following observations:

(i) If $X(t) = X^c(t)$, i.e. if X has no jump then we have the classical Ito's formula:

$$df(X(t)) = f'(X(t))dXt + \frac{1}{2}f''(X(t))\gamma^2(t)dt.$$

(ii) If $X(t) = J(t)$, then $f(X(t))$ is also a pure jump process. Moreover,

$$f(X(t)) = f(X(0)) + \sum_{0<s\leq t} f(X(s)) - f(X(s-)).$$

(iii) In general when $X(t) = X^c(t) + J(t)$, intuitively we should have $df(X(t))$ following the classical Ito's formula in between the jumps of X and $\Delta f(X(t)) = f(X(t)) - f(X(t-)$ at the jump points of X.

This leads to the following Ito's formula (see Shreve's theorem 11.5.1)

$$\begin{aligned} f(X(t)) &= f(X(0)) + \int_0^t f'(X(s))dX^c(s) + \int_0^t \frac{1}{2}f''(X(s))\gamma^2(s)ds \\ &\quad + \sum_{0<s\leq t} f(X(s)) - f(X(s-)). \end{aligned}$$

Remark 1.10.1. *In general, the above Ito's formula does NOT have a differential form, i.e.* $df(X(t)) = \cdots$. *The reason is generally we cannot express* $\Delta f(X(s)) = f(X(s)) - f(X(s-))$ *in terms of some derivative of* f *and* $\Delta X(s) = X(s) - X(s-)$. *In some special case, for example when* $X(t)$ *is a pure jump process then we may have a differential form for* $df(X(t))$*, but this is not guaranteed. See also the discussion in* (1.11.2).

1.10.2 Ito formula for multiple jump processes

Following similar argument to the one dimensional Ito formula for jump process, we can derive the multi-dimensional Ito formula for jump processes. Here we give the version for two processes. The formula for higher dimension follows the same pattern.

Theorem 1.10.2. *Let X^1, X^2 be two jump processes:*

$$X^i(t) = X^i(0) + \int_0^t \alpha^i(s)ds + \int_0^t \gamma^i(s)dWs + J^i(t), i = 1, 2.$$

Let $f(t, x_1, x_2)$ be a twice differentiable in its spatial variables. Then

$$\begin{aligned} f(t, X_t^1, X_t^2) &= f(t, X_0^1, X_0^2) + \int_0^t \sum_{i=1}^2 f_{x_i}(s, X_s^1, X_s^2)d(X^i)_s^c \\ &+ \frac{1}{2}\int_0^t \sum_{i=1}^2 f_{x_ix_i}(s, X_s^1, X_s^2)(\gamma^i)_s^2 ds \\ &+ \int_0^t f_{x_1x_2}(s, X_s^1, X_s^2)\gamma_s^1\gamma_s^2 ds \\ &+ \sum_{0<s\leq t} f(s, X_s^1, X_s^2) - f(s, X_{s-}^1, X_{s-}^2). \end{aligned}$$

Corollary 1.10.3. *Let X^1, X^2 be two jump processes:*

$$X^i(t) = X^i(0) + \int_0^t \alpha^i(s)ds + \int_0^t \gamma^i(s)dWs + J^i(t), i = 1, 2.$$

Then

$$\begin{aligned} X_t^1X_t^2 &= X_0^1X_0^2 + \int_0^t X_s^1 d(X^2)_s^c + \int_0^t X_s^2 d(X^1)_s^c \\ &+ \int_0^t \gamma_s^1\gamma_s^2 ds + \sum_{0<s\leq t} X_s^1X_s^2 - X_{s-}^1X_{s-}^2. \end{aligned}$$

Remark 1.10.4. *If each X^i is driven by a different Brownian motion W^i and they are independent then the cross variation term in Theorem (1.10.2) $\int_0^t f_{x_1x_2}(s, X_s^1, X_s^2)\gamma_s^1\gamma_s^2 ds$ will disappear, as well as the cross variation term $\int_0^t \gamma_s^1\gamma_s^2 ds$ in Corollary (1.10.3).*

1.11 Models of stock price with jumps

1.11.1 Stock models

Recall that before we model the dynamics of a stock $S(t)$ as followed:

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

Also observe that the most important property of $W(t)$ we used in pricing financial models with $S(t)$ is that it is a $martingale$. This motivates us to replace $W(t)$ with a general martingale with jumps. That is, we let

$$M(t) = \int_0^t \alpha(s)ds + \int_0^t \gamma(s)dWs + J(t),$$

where $J(t)$ is a pure jump process and $J(t) + \int_0^t \alpha(s)ds$ is a martingale. Consider the following model for $S(t)$:

$$S(t) = S(0) + \int_0^t \mu(s)S(s-)ds + \int_0^t S(s-)dM(s). \tag{1.14}$$

Intuitively, the reason we use $S(s-)$ in the RHS is so that at the jump of $M(t)$, we have

$$S(t) - S(t-) = S(t-)\Delta J(t). \tag{1.15}$$

If we think of $\Delta J(t)$ as representing an external shock, then this says the jump in the stock price is its value immediately before the shock occurs multiply with the size of the shock, which makes sense.

Mathematically, using $S(s-)$ in the RHS has the benefit of guaranteeing $\int_0^t S(s-)dM(s)$ to be a martingale under proper conditions (see the discussion in Section 7.2). Either way, it should be noted that we can equivalently write (1.14) as

$$S(t) = S(0) + \int_0^t (\mu(s) + \alpha(s))S(s)ds + \int_0^t S(s)\gamma(s)dW(s) + \sum_{0<s\leq t} S(s-)\Delta J(s) \tag{1.16}$$

That is, we only use $S(s-)$ in conjuction with the jumps in $J(s)$.

Relation (1.15) has another important implication for the jumps of $J(t)$:

$$S(t) = S(t-)(1 + \Delta J(t)).$$

Since we want to use $S(t)$ as a stock price, $S(t) \geq 0$ implies we need to restrict $\Delta J(t) > -1$.

Similar to the classical Black-Scholes model, we have an explicit formula for $S(t)$ satisfying (1.14) or (1.16):

$$S(t) = S(0)\exp\Big[\int_0^t [\mu(s)+\alpha(s)-\frac{1}{2}\gamma^2(s)]ds + \int_0^t \gamma(s)dWs\Big]\prod_{0<s\le t}(1+\Delta J(s)) \quad (1.17)$$

Example 1.11.1. *Geometric Poisson process: If we let $M(t) = \sigma(N(t)-\lambda t)$ then*

$$S(t) = S(0) + \int_0^t S(s-)dMs = S(0)e^{-\sigma\lambda t}\prod_{0<s\le t}(1+\sigma\Delta N(s)) = S(0)e^{-\sigma\lambda t}(1+\sigma)^{N(t)},$$

since we observe that $1+\sigma\Delta N(s) = 1+\sigma$ at all jump points of $N(t)$ and there are exactly $N(t)$ jumps at time t. Also note that since σ is the jump size of the pure jump process $\sigma N(t)$, we require $\sigma > -1$ as in the discussion above.

1.11.2 Some general remarks

Let $W(t)$ be a BM and $N(t)$ be a Poisson process. Observe that

$$X^1(t) = 1 + \int_0^t \sigma X^1(s)ds$$
$$X^2(t) = 1 + \int_0^t \sigma X^2(s)dW(s)$$
$$X^3(t) = 1 + \int_0^t \sigma X^3(s-)dN(s)$$

(note the $X^3(s-)$ in the last equation) have solutions

$$\begin{aligned} X^1(t) &= e^{\sigma t} \\ X^2(t) &= e^{\sigma W(t)-\frac{1}{2}\sigma^2 t} \\ X^3(t) &= (1+\sigma)^{N(t)}, \end{aligned}$$

where the solution for X^1 follows from classical calculus, X^2 from "classical" Ito's formula and X^3 from the calculus for jump processes (see also the discussion about Geometric Poisson process). The point to observe here is that *three very similar differential equations give three distinctly different answers depending on different integrators.*

Also observe that if we apply Ito's formula for jump processes to the $f(N(t)) = (1+\sigma)^{N(t)}$, we get

$$X^3(t) = f(N(t)) = \sum_{s\le t}(1+\sigma)^{N(s)} - (1+\sigma)^{N(s-)}. \quad (1.18)$$

This at first glance does not look like the “differential" form

$$\begin{aligned} dX^3(t) &= \sigma X^3(t-)dN(t) \\ X^3(0) &= 1. \end{aligned} \tag{1.19}$$

However, we observe from (1.18) that $\Delta X^3(s) = (1+\sigma)^{N(s)} - (1+\sigma)^{N(s-)}$. Moreover, at the jump point of N

$$\begin{aligned} \Delta X^3(s) &= (1+\sigma)^{N(s)} - (1+\sigma)^{N(s-)} = (1+\sigma)^{N(s)} - (1+\sigma)^{N(s)-1} \\ &= \sigma(1+\sigma)^{N(s)-1} = \sigma X^3(s-) \\ &= \sigma X^3(s-)\Delta N(s). \end{aligned}$$

Now the agreement between (1.18) and (1.19) are clear. The point here is that it is not immediate to derive “differential" form from the explicit formula of a jump process. Indeed such differential form is not always possible. The fact that $N(t)$ is a counting process (having jump of size 1) is central to the reason why the formula X^3 is nice, as well as that we could re-derive the differential form of $X^3(t)$ from its explicit formula. Replacing $N(t)$ with a general jump process (having arbitrary jump size) in the differential equation for X^3, and you will see that we no longer can easily derive such nice formula anymore.

CHAPTER 2 Change of measure for jump processes

2.1 Motivation

One of the fundamental concept in Math Finance I regarding the Black-Scholes model is the following: Suppose $S(t)$ satisfies

$$S(t) = S(0) + \int_0^t \mu(u)S(u)du + \int_0^t S(u)dW(u),$$

under the objective probability $\mathbb{P}$. Unless $\mu(u) = r$, the interest rate, (which we supposed to be a constant for simplicty) $e^{-rt}S(t)$ *is not a martingale under* $\mathbb{P}$, and thus we cannot price financial product under $\mathbb{P}$. We need to find another measure $\mathbb{Q}$, the risk neutral measure, so that $e^{-rt}S(t)$ is a martingale udner $\mathbb{Q}$. The key idea is that under $\mathbb{Q}$, it must be the case that $\widetilde{W}(t) := \int_0^t (\mu(u) - r)du + W(t)$ is a Brownian motion. So that

$$S(t) = S(0) + \int_0^t rS(u)du + \int_0^t S(u)d\widetilde{W}(u)$$

has the right distribution under $\mathbb{Q}$.

Intuitively, the measure $\mathbb{Q}$ is chosen so that we can "modify the drift" of $W(t)$ and still have the new process $\widetilde{W}(t)$ being a Brownian motion; which results in modifying the drift of $S(t)$ to the desirable drift(in this case, r).

Now suppose $S(t)$ satisfies

$$S(t) = S(0) + \int_0^t \mu S(u)du + \int_0^t S(u-)dM(u),$$

under some objective probability measure $\mathbb{P}$, where $M(t) = N(t) - \lambda t$ is a compensated Poisson process with rate λ under $\mathbb{P}$. Again, we would like that

$$S(t) = S(0) + \int_0^t rS(u)du + \int_0^t S(u-)d\widetilde{M}(u),$$

where $\widetilde{M}(t) := M(t) - (r - \mu)t$ is a martingale under a probability measure $\mathbb{Q}$. Again, since $M(t) = N(t) - \lambda t$, it is clear that $\widetilde{M}(t)$ is a martingale if $N(t)$ becomes a Poisson process with rate $\lambda + (r - \mu)$ under $\mathbb{Q}$. This note discusses how to choose such a measure $\mathbb{Q}$ for various choices of jump martingales M.

2.2 Review of change of measure, Girsanov's theorem

2.2.1 The change of measure kernel $Z(t)$

Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$; $\mathcal{F}(t), 0 \le t \le T$ a filtration with $\mathcal{F}(T) = \mathcal{F}$. If we define another probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}(T))$ via the relation

$$d\mathbb{Q} = Z(T)d\mathbb{P},$$

for some random variable $Z(T)$, that is for all $Y \in \mathcal{F}(T)$

$$\mathbb{E}^{\mathbb{Q}}(Y) := \mathbb{E}^{\mathbb{P}}(Z(T)Y),$$

it must be that $\mathbb{P}(Z(T) \ge 0) = 1$ *and* $\mathbb{E}^{\mathbb{P}}(Z(T)) = 1$.

2.2.2 Restriction of $\mathbb{Q}$ to a smaller sigma algebra $\mathcal{F}(t)$

Let $\mathcal{F}(t), 0 \le t \le T$ be a filtration associated with a probability space $(\Omega, \mathbb{P}, \mathcal{F}(T))$. If $Z(t)$ is *a* $\mathbb{P}$ *martingale*, $Z(T)$ satisfies the conditions in (i), then for all $Y \in \mathcal{F}(t)$

$$\mathbb{E}^{\mathbb{Q}}(Y) = \mathbb{E}^{\mathbb{P}}(Z(t)Y).$$

(See Shreve's Lemma 5.2.1.) Note that this is *not a definition* but a result that follows from the definition in (i) and the fact that Z is a martingale.

2.2.3 Conditional expectation in change of measure

Let Y be $\mathcal{F}_T$ measureable, we have for $t \le T$

$$\mathbb{E}^{\mathbb{Q}}(Y|\mathcal{F}(t)) = \frac{\mathbb{E}^{\mathbb{P}}(Z_T Y|\mathcal{F}_t)}{\mathbb{E}^{\mathbb{P}}(Z_T|\mathcal{F}_t)}.$$

In particular, if $Z(t)$ *is a* $\mathbb{P}$*-martingale* then combining the results above we have for $s \le t$ and $Y \in \mathcal{F}_t$

$$\mathbb{E}^{\mathbb{Q}}(Y|\mathcal{F}_s) = \frac{\mathbb{E}^{\mathbb{P}}(Z_t Y|\mathcal{F}_s)}{Z_s}.$$

2.2.4 Condition for a process to be a martingale under the new measure

Theorem 2.2.1. *Let $X(t)$ be a $\mathcal{F}(t)$ adapted process, $Z(t)$ a $\mathbb{P}$-martingale then $X(t)Z(t)$ is a $\mathbb{P}$ martingale* **if and only if** *$X(t)$ is a $\mathbb{Q}$-martingale.*

Application: Let $X(t)$ be the "Brownian motion with drift" $\widetilde{W}(t)$ or the process $\widetilde{M}(t)$ in section I. Recall that we want $\widetilde{W}(t)$ (or $\widetilde{M}(t)$) be a martingale under $\mathbb{Q}$. This statement gives a sufficient condition for this to happen.

2.3 Some remarks about Girsanov theorem

2.3.1 Characterization of Brownian motion

We all know the Levy's characterization of Brownian motion: continuous martingale with quadratic variation on $[0, t]$ equals to t. There is an equivalent characterization:

Theorem 2.3.1. *Let $X(t)$ be a continuous process such that $X(0) = 0$. Then $X(t)$ is a Brownian motion w.r.t a filtration $\mathcal{F}(t)$ if and only if for all $u \in \mathbb{R}$,*

$$\mathcal{E}^u(X)(t) := e^{uX_t - \frac{1}{2}u^2 t}$$

is a martingale w.r.t $\mathcal{F}_t$.

Proof. Let $X(t)$ be a Brownian motion. It is routine to show that

$$e^{uX(t) - \frac{1}{2}u^2 t}$$

is a martingale.

The converse can be argued heuristically as followed. Suppose that $e^{uX_t - \frac{1}{2}u^2 t}$ is a martingale for all $u \in \mathbb{R}$. Then by definition for $s < t$

$$\mathbb{E}(e^{u(X_t - X_s)} \Big| \mathcal{F}_s) = e^{\frac{1}{2}u^2(t-s)}.$$

Since *for all* $u \in \mathbb{R}$, the RHS is independent of $\mathcal{F}(s)$, $X(t) - X(s)$ is independent of $\mathcal{F}(s)$ (see explaination in the remark below). Hence it has **independent increments**.

Moreover, from the same calculation, the moment generating function of $X(t) - X(s)$ is that of a Normal(0, $t - s$). Hence it has **stationary increments**, and the increments has Normal(0, $t - s$) distribution. Thus $X(t)$ is a Brownian motion.

Remark 2.3.2. *We clarifiy the reason why $X(t) - X(s)$ is independent of $\mathcal{F}(s)$. We make the following claim: if $E(e^{uX}|\mathcal{F}) = E(e^{uX})$ for all $u \in \mathbb{R}$ then X is independent of $\mathcal{F}$.*

This claim is true in turn because of the following result, known as Kac's theorem for characteristic functions: if

$$E(e^{uX+vY}) = E(e^{uX})E(e^{vY}), \forall u, v \in \mathbb{R},$$

then X, Y are independent. A reference can be found in Thereom 1.1.16 of the textbook Levy's processes and Stochastic Calculus by David Applebaum.

If we accept this result, then we see that or all $Y \in \mathcal{F}$,

$$E(e^{uX+vY}) = E\big(E(e^{uX+vY}|\mathcal{F})\big) = E\big(e^{vY}E(e^{uX}|\mathcal{F})\big) = E(e^{uX})E(e^{vY}).$$

Hence X is indepedent of Y for all $Y \in \mathcal{F}$. Hence X is independent of $\mathcal{F}$.

2.3.2 Choice of $Z(t)$ in Girsanov Theorem

Suppose that $W(t)$ is a $\mathbb{P}$ Brownian motion. We want to find $\mathbb{Q}$ via

$$d\mathbb{Q} = Z(T)d\mathbb{P}$$

so that

$$\widetilde{W}(t) := W(t) + \alpha t$$

is a $\mathbb{Q}-$Brownian motion. From the characterization of Brownian motion from the exponential martingale above, we need

$$\mathcal{E}^u(\widetilde{W})(t) = e^{u\widetilde{W}_t - \frac{1}{2}u^2 t}$$

to be a $\mathbb{Q}$ martingale. Observe that

$$\mathcal{E}^u(\widetilde{W})(t) = e^{u\widetilde{W}(t) - \frac{1}{2}u^2 t} = e^{uW(t) - \frac{1}{2}\left(u^2 - 2u\alpha\right)t}.$$

By Theorem (2.2.1), in order for $\mathcal{E}^u(\widetilde{W})(t)$ to be a $\mathbb{Q}$ martingale, we need to choose the change of measure kernel $Z(t)$ so that both $Z(t)$ **and** $\mathcal{E}^u(\widetilde{W})(t)Z(t)$ are $\mathbb{P}$-martingales. Since

$$u^2 - 2u\alpha = (u-\alpha)^2 - \alpha^2,$$

and clearly

$$e^{(u-\alpha)W(t) - \frac{1}{2}(u-\alpha)^2 t}$$

is a $\mathbb{P}$-martingale, we may guess the choice for $Z(t)$ is

$$Z(t) = e^{-\alpha W(t) - \frac{1}{2}\alpha^2 t},$$

which is clearly also a $\mathbb{P}$-martingale.

This intuition also suggests that if we want $\widetilde{W}(t) = W(t) + \int_0^t \alpha(u)du$ to be a $\mathbb{Q}$ Brownian motion, the choice of $Z(t)$ is

$$Z(t) = e^{-\int_0^t \alpha(u)dW(u) - \frac{1}{2}\int_0^t \alpha(u)^2 du},$$

even though the verification now is slightly more involved.

2.4 Change of measure for Poisson processes

2.4.1 Poisson process characterization

Theorem 2.4.1. *A càdlàgprocess $N(t)$, $N(0)$ = 0, is a Poisson process with rate λ w.r.t $\mathcal{F}(t)$ if and only if for all $u \in \mathbb{R}$*

$$\exp\big(uN(t) - \lambda t(e^u - 1)\big)$$

is a martingale w.r.t $\mathcal{F}(t)$.

The proof for this theorem is similar to the proof for the characterization of Brownian motion.

2.4.2 Choice of $Z(t)$

Suppose $N(t)$ is a Poisson process with rate λ under $\mathbb{P}$. We want to find $\mathbb{Q}$ via the change of measure formula

$$d\mathbb{Q} = Z(T)d\mathbb{P}$$

so that $N(t)$ has rate $\widetilde{\lambda}$ under $\mathbb{Q}$. By the characterization of Poisson process, we want

$$\exp\big(uN(t) - \widetilde{\lambda} t(e^u - 1)\big)$$

to be a $\mathbb{Q}$-martingale.

Again, by Theorem (2.2.1), we need to choose $Z(t)$ so that both

$$Z(t) \text{ and } e^{uN(t)-\widetilde{\lambda}t(e^u-1)}Z(t)$$

are $\mathbb{P}$-martingales.

Now the choice of such $Z(t)$ may not be immediately obvious, even though we can use the same reverse engineer idea as we did above with the Brownian motion. Instead, a more natural idea here is to choose a *general exponential martingale* $Z(t)$ associated with $N(t)$ with *an undetermined coefficient.* We perform the change of measure with $Z(t)$ and *expect* that $N(t)$ will remain a Poisson process with different rate under this change of measure. We then observe what rate $N(t)$ will actually be under the new measure using the exponential martingale characterization of a Poisson process. Then we can determine the precise coefficient to achieve the desired rate of $N(t)$ under the new measure.

More specifically, we let

$$Z(t) = e^{aN(t)-\lambda t(e^a-1)},$$

where a is our undetermined coefficient. Then clearly $Z(t)$ is a $\mathbb{P}$-martingale.

Suppose that $N(t)$ remains a Poisson process under $\mathbb{Q}$ and its rate is $\widetilde{\lambda}$. Then by Theorem (2.2.1) and the exponential martingale characterization of Poisson processes we must have

$$e^{uN(t)-\widetilde{\lambda}t(e^u-1)}e^{aN(t)-\lambda t(e^a-1)} = e^{(u+a)N(t)-\widetilde{\lambda}t(e^u-1)-\lambda t(e^a-1)}$$

is a $\mathbb{P}$-martingale.

But since we know

$$e^{(u+a)N(t)-\lambda t(e^{u+a}-1)}$$

is a $\mathbb{P}$-martingale we must have

$$\widetilde{\lambda}(e^u-1)+\lambda(e^a-1) = \lambda(e^{u+a}-1). \tag{2.1}$$

Note that the above equation has to be true $\forall u \in \mathbb{R}$. In particular if we choose $u=-a$ then the RHS equals 0. Thus

$$\widetilde{\lambda} = \lambda\frac{1-e^a}{e^{-a}-1} = \lambda\frac{e^a(1-e^a)}{1-e^a} = \lambda e^a.$$

Plug this in we indeed verify the equation (2.1) for all u.

Thus our conclusion is that if we choose

$$Z(t) = e^{aN(t)-\lambda t(e^a-1)},$$

then $N(t)$ is a Poisson process with rate λe^a under $\mathbb{Q}$. Now if we desire

$$\lambda e^a = \tilde{\lambda},$$

for some pre-given $\tilde{\lambda}$ then clearly $a = \log\left(\frac{\tilde{\lambda}}{\lambda}\right)$ and also

$$Z(t) = \exp\left[\log\left(\frac{\widetilde{\lambda}}{\lambda}\right)N(t) + (\lambda-\widetilde{\lambda})t\right].$$

We have arrived at the following theorem

Theorem 2.4.2. *Let $N(t)$ be a Poisson process with rate λ under a probabilty $\mathbb{P}$ and $\mathcal{F}(t)$ a filtration for $N(t)$. Let $\widetilde{\lambda}$ be given. Define*

$$\begin{aligned} Z(t) &:= \exp\left[\log\left(\frac{\widetilde{\lambda}}{\lambda}\right)N(t) + (\lambda-\widetilde{\lambda})t\right] \\ &= e^{(\lambda-\widetilde{\lambda})t}\left(\frac{\widetilde{\lambda}}{\lambda}\right)^{N(t)}, 0 \le t \le T. \end{aligned}$$

Also define

$$d\mathbb{Q} = Z(T)d\mathbb{P} \text{ on } \mathcal{F}(T).$$

Then $Z(t)$ is a $\mathbb{P}$ martingale and under $\mathbb{Q}$, $N(t)$ is a Poisson process with rate $\widetilde{\lambda}$.

2.5 Change of measure for compound Poisson with discrete jump distribution

Let $Q(t)$ be a compound Poisson process with rate λ. That is

$$Q(t) = \sum_{i=1}^{N(t)} Y_i,$$

$N(t)$ has rate λ under a probabilty $\mathbb{P}$ and $\mathcal{F}(t)$ a filtration for $Q(t)$. Recall that each jump of $Q(t)$ has identical distribution Y_i.

Here we assume that Y_1 (hence all Y_i's) takes values $y_1, y_2, ..., y_M$ with probability

$$\mathbb{P}(Y_1 = y_m) = p_m, 1 \leq m \leq M$$

that is Y_1 has discrete distribution.

We want to change the intensity of $Q(t)$ as well as the distribution of Y_i (that is to change p_m) via the change of measure. For any $\widetilde{\lambda} > 0$ and $\widetilde{p}_m \in (0, 1), \sum_{m=1}^{M} \widetilde{p}_m = 1$ we find a probabilty $\mathbb{Q}$ so that under $\mathbb{Q}$, $Q(t)$ is a compound Poisson process with rate $\widetilde{\lambda}$ and Y_i has distribution

$$\mathbb{Q}(Y_1 = y_m) = \widetilde{p}_m, 1 \leq m \leq M.$$

Before we proceed, we need to mention an important result about decomposing a compound Poisson process with discrete jumps into a sum of Poisson processes.

2.5.1 Summing and Decomposing Compound Poisson processes

Summing compound Poisson processes

Compound Poisson processes can be combined and decomposed in fascinating ways. Shreve treats these in Theorem 11.3.3, page 471, and Corollary 11.3.3, page 473, for the special case when $Y_1, \ldots$ are discrete random variables. We will state more general versions of these properties here, but without a proof: Shreve gives a proof for his special case.

Theorem 11.3.3 says in essence that one can build a compound Poisson process by bringing in jumps of different sizes at different Poisson rates. For example let N_1 and N_2 be two independent, Poisson processes with respective rates λ_1 and λ_2. Then $y_1 N_1(t)$ is a very simple compound Poisson process in which jumps of size y_1 arrive in a Poisson stream of rate λ_1, and $y_2 N_1(t)$ is a very simple compound Poisson process in which jumps of size y_2 arrive in a Poisson stream of rate λ_2.

Let

$$Q(t) = y_1 N_1(t) + y_2 N_2(t).$$

This is a pure jump process whose jumps are either of size y_1 or y_2. The total number of jumps by time t is clearly $N_1(t)+N_2(t)$. Let Y_k denote the size of the kth jump of Q. Then, by definition,

$$Q(t) = \sum_{k=1}^{N_1(t)+N_2(t)} Y_k.$$

Then Theorem 11.3.3 says

(i) $N_1 + N_2$ is a Poisson process with rate $\lambda_1 + \lambda_2$;

(ii) $Y_1, Y_2, \ldots$ are independent and identically distributed with

$$I\!P\Big(Y_i = y_1\Big) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \text{ and } I\!P\Big(Y_i = y_2\Big) = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

As a consequence, Q is a compound Poisson process. Theorem 11.3.3 extends this idea to summing Poisson streams of more than two possible jump sizes. Note that, as a consequence of statement (i) above, *the independent Poisson processes N_1 and N_2 never jump at the same time*.

Theorem 11.3.3 is actually a special case of a much more general theorem: the sum of any finite number of compound Poisson processes is a compound Poisson process.

We give a heuristic reasoning about Theorem 11.3.3. First observe that if $N(t) = N_1(t) + N_2(t)$ then $N(t)$ would jump at the jump time of N_1 or N_2, whichever arrives first. That is $N(t)$ jumps at the minimum of the jump times of N_1 and N_2. Now let $\tau_i, i = 1, 2$ are independent exponential(λ_i) random variables and $\tau = \min(\tau_1, \tau_2)$ then

$$P(\tau \geq t) = P(\tau_1 \geq t)P(\tau_2 \geq t) = e^{-(\lambda_1+\lambda_2)t}.$$

That is τ is an exponential$(\lambda_1+\lambda_2)$ random variable. This gives the intuition about N_1+N_2 being a Poisson process with rate $\lambda_1 + \lambda_2$. The rigorous proof would use the exponential martingale characterization of Poisson processes mentioned above.

Second, $N(t)$ would jump with size y_1 if the jump time of N_1 arrives before the jump time of N_2 and vice versa. We have

$$\begin{aligned} P(\tau_1 < \tau_2) &= \int_0^\infty P(\tau_1 < \tau_2 | \tau_2 = t)\lambda_2 e^{-\lambda_2 t} dt \\ &= \int_0^\infty (1 - e^{-\lambda_1 t})\lambda_2 e^{-\lambda_2 t} dt \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

Similarly

$$P(\tau_1 < \tau_2) = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

This explains the distribution of Y_i, the jump size of N_t.

Decomposing compound Poisson processes

One can also go in the opposite direction and decompose a compound Poisson process into a sum of independent Poisson processes bringing different sized jumps at different rates. We shall state this very generally. In fact, we will start not with a compound Poisson process, but just with some Lévy process.

Let X be a Lévy process. Let A be a subset of $I\!R$ that avoids a neighborhood of 0 in the sense that for some $\epsilon > 0$, $A \bigcap (-\epsilon, \epsilon) = \emptyset$. Let

$$N_A(t) := \sum_{s \le t} \mathbf{1}_{\{\triangle X(s) \in A\}}$$

be the number of jumps of X with values in A that occur by time t. Let

$$X_A(t) := \sum_{s \le t} \triangle X(s) \mathbf{1}_{\{\triangle X(s) \in A\}},$$

$X_A(t)$ will be well defined whenever $N_A(t)$ is finite. The process X_A is the accumulated sum of all the jumps of X with values in the set A. It might be that X never has a jump with values in A (that is, $N_A(t) = 0$ for all $t \ge 0$). But if it does, let $Y_1^A, Y_2^A, \ldots,$ be the first, second, third, etc. jump values of $\{X_A(t);\, t \ge 0\}$. By definition of the terms so far,

$$X_A(t) = \sum_{k=1}^{N_A(t)} Y_k^A.$$

Theorem 2. *(a) Let $X(t)$ be a Lévy process. Then $\{X_A(t);\, t \ge 0\}$ is a compound Poisson process; in other words, N_A is a Poisson process and $Y_1^A, Y_2^A, \ldots$ is a sequence of independent, identically distributed random variables independent of N_A. In addition, $X(t) - X_A(t)$ is a Lévy process and is independent of $X_A(t)$. Thus, $X(t) = X(t) - X_A(t) + X_A(t)$ represents X as the sum of two independent Lévy processes, the first of which has no jumps with values in A, and the second of which only has jumps with values in A.*

(b) Let $\epsilon > 0$ and let $A_1, \ldots, A_n$ be disjoint subsets of $(-\infty, \infty) - (-\epsilon, \epsilon)$. Then $X_{A_1}(\cdot), \ldots, X_{A_n}(\cdot)$ are independent *compound Poisson processes that are all independent of $X(\cdot) - \Big(X_{A_1} + \cdots + X_{A_n}(\cdot)\Big)$.*

The take-home message of this theorem is that the accumulated jumps of a Lévy process into disjoint sets bounded away from 0 are independent, compound Poisson processes. Thus, a Lévy process with jumps has a very rich structure which aggregates the influence of many, independently occurring Poisson streams. Corollary 11.3.4 is a special case of this theorem for compound Poisson process that admit only a finite number of possible jumps sizes.

Theorem 2 is at the heart of a result, due to Lèvy and Khinchine, that characterizes the most general Lèvy process. It says that if X is a Lèvy process, there is a decomposition of the form $X(t) = \mu t + \sigma W(t) + Y(t)$, where W is a Brownian motion independent independent of Y, and where Y is a limit of a sequence of processes $Z_n(t) + m_n t$, with Z_n being a compound Poisson process for each n.

Lastly we give a heuristic reasoning for decomposing a compound Poisson process with discrete jumps. Let

$$Q(t) = \sum_{i=0}^{N(t)} Y_i,$$

where $N(t)$ is a Poisson(λ) process and $P(Y_1 = y_m) = p_m, m = 1, \cdots, M$.

Now if we let

$$Q_m(t) = \sum_{i=0}^{N(t)} Y_i \mathbf{1}_{\{Y_i = y_m\}},$$

then observe that $Q_m(t)$ has independent and stationary increments. That is it is a Levy process. Moreover, one can check that

$$Ee^{u\frac{Q_m(t)}{y_m}} = e^{\lambda p_m t(e^u - 1)}.$$

That is $N_m(t) := \frac{Q_m(t)}{y_m}$ is a Poisson (λp_m) process. Lastly, it is clear from the definition that for $n \neq m$, $N_m(t)$ and $N_n(t)$ *do not jump at the same time.* From an exercise in Homework 3, you'll see that this implies $N_m(t)$ and $N_n(t)$ are independent. This gives the decomposition of $Q(t)$ as

$$Q(t) = \sum_{m=1}^{M} y_m N_m(t),$$

where $N_m(t)$ are independent Poisson processes with rates λp_m.

2.5.2 Change of measure for multiple independent Poisson processes

Lemma 2.5.1. *Let $N_m, m = 1, \cdots, M$ be independent Poisson processes with rates $\lambda_m, m = 1, \cdots, M$. Let $\widetilde{\lambda}_m, m = 1, 2, ..., M$ be given. Define*

$$Z_m(t) := e^{(\lambda_m - \widetilde{\lambda}_m)t} \left(\frac{\widetilde{\lambda}_m}{\lambda_m}\right)^{N_m(t)}$$

$$Z(t) := \prod_{m=1}^{M} Z_m(t)$$

and

$$d\mathbb{Q} = Z(T)d\mathbb{P} \text{ on } \mathcal{F}(T).$$

Then N_m's are independent Poisson processes with rate $\widetilde{\lambda}_i$ under $\mathbb{Q}$.

You will be asked to explore the proof of this Lemma in Homework 2 for the case $M = 2$. The proof for general M is similar.

2.5.3 Change of measure for compound Poisson process with discrete jumps

The decomposition of a compound Poisson process into multiple independent Poisson processes and Lemma (2.5.1) lead to the following result: (Shreve's Lemma 11.6.4, Theorem 11.6.5)

Theorem 2.5.2. *Let*

$$Q(t) = \sum_{i=1}^{N(t)} Y_i,$$

$N(t)$ has rate λ under a probabilty $\mathbb{P}$ and Y_1 takes values $y_1, y_2, ..., y_M$ with probability

$$\mathbb{P}(Y_1 = y_m) = p_m, 1 \leq m \leq M$$

that is Y_1 has discrete distribution.

Let $\widetilde{\lambda}_m, m = 1, 2, ..., M$ be given. Define $Z(t)$ as in Lemma (2.5.1). That is

$$Z(t) \ := \ \prod_{m=1}^{M} e^{(\lambda_m - \widetilde{\lambda}_m)t} \left(\frac{\widetilde{\lambda}_m}{\lambda_m}\right)^{N_m(t)} .$$

Then Z(t) is a $\mathbb{P}$ martingale. Moreover, under $\mathbb{Q}$, $Q(t)$ is a compound Poisson process with rate $\widetilde{\lambda}$ and $P^Q(Y_i = y_m) = \widetilde{p}_m$, where

$$\widetilde{\lambda} \ = \ \sum_{m=1}^{M} \widetilde{\lambda}_m$$

$$\widetilde{p}_m \ = \ \frac{\widetilde{\lambda}_m}{\widetilde{\lambda}}.$$

Remark 2.5.3. *We mentioned at the beginning of this section that we can choose $\widetilde{\lambda}$ and $\widetilde{p}_m$, while Theorem (2.5.2) says we can choose $\widetilde{\lambda}_m$. The difference is artificial. Indeed, given $\widetilde{\lambda}_m$ we can define $\widetilde{\lambda}$ and $\widetilde{p}_m$ as in Theorem (2.5.2). But conversely, we can start out with $\widetilde{\lambda}$ and $\widetilde{p}_m$ and define $\widetilde{\lambda}_m := \widetilde{p}_m \widetilde{\lambda}$. It's up to you and the problem you're dealing with to decide which are the given variables to work with.*

2.5.4 Compound Poisson with continuous jump distribution

Let $Q(t)$ be a compound Poisson process with rate λ under a probabilty $\mathbb{P}$ and $\mathcal{F}(t)$ a filtration for $Q(t)$. Here we assume Y_i has continuous distribution with density function f.

We want to change the intensity of $Q(t)$ as well as the distribution of Y_i (that is the density f) via the change of measure. For any density function $\widetilde{f}$ and $\widetilde{\lambda}$, we find a probabilty $\mathbb{Q}$ so that under $\mathbb{Q}$, $Q(t)$ is a compound Poisson process with rate $\widetilde{\lambda}$ and Y_i has continuous distribution with densitry $\widetilde{f}$.

2.5.5 A rewrite of $Z(t)$ in Theorem (2.5.2)

There is yet another way to write the process $Z(t)$ in Theorem (2.5.2). Note that

$$\begin{aligned} Z(t) &= \prod_{m=1}^{M} e^{(\lambda_m - \widetilde{\lambda}_m)t} \left(\frac{\widetilde{\lambda}_m}{\lambda_m} \right)^{N_m(t)} \\ &= e^{(\lambda - \widetilde{\lambda})t} \prod_{m=1}^{M} \left(\frac{\widetilde{\lambda}\widetilde{p}_m}{\lambda p_m} \right)^{N_m(t)} \\ &= e^{(\lambda - \widetilde{\lambda})t} \frac{\widetilde{\lambda}^{N(t)} \prod_{m=1}^{M} \widetilde{p}_m^{N_m(t)}}{\lambda^{N(t)} \prod_{m=1}^{M} p_m^{N_m(t)}}. \end{aligned}$$

By rearranging terms,

$$\prod_{m=1}^{M} p_m^{N_m(t)} = \prod_{i=1}^{N(t)} p(Y_i),$$

where we define

$$p(Y_i) := p_m \text{ if } Y_i = y_m, m = 1, \cdots, M.$$

To see this equality, note that for each m, there are $N_m(t)$ terms of p_m on the LHS. By definition, for each event ω, the Y_i random variables take on values y_m exactly $N_m(t)$ times. Thus there are also $N_m(t)$ terms of p_m on the RHS.

Similarly we have,

$$\prod_{m=1}^{M} \widetilde{p}_m^{N_m(t)} = \prod_{i=1}^{N(t)} \widetilde{p}(Y_i).$$

Thus

$$
\begin{aligned}
Z(t) &= e^{(\lambda-\widetilde{\lambda})t}\frac{\widetilde{\lambda}^{N(t)}\prod_{i=1}^{N(t)}\widetilde{p}(Y_i)}{\lambda^{N(t)}\prod_{i=1}^{N(t)}p(Y_i)} \\
&= e^{(\lambda-\widetilde{\lambda})t}\prod_{i=1}^{N(t)}\frac{\widetilde{\lambda}\widetilde{p}(Y_i)}{\lambda p(Y_i)}.
\end{aligned}
$$

2.5.6 Change of measure for compound Poisson with continuous jump distribution

The above observation suggests the following choice of $Z(t)$ when Y_i has continuous distribution.

Definition 2.5.4. *Fix $T > 0$. Let $\widetilde{\lambda} > 0$ and a density function $\widetilde{f}$ be given. Define*

$$
Z(t) := e^{(\lambda-\widetilde{\lambda})t}\prod_{i=1}^{N(t)}\frac{\widetilde{\lambda}\widetilde{f}(Y_i)}{\lambda f(Y_i)}. \tag{2.2}
$$

Also define

$$
d\mathbb{Q} = Z(T)d\mathbb{P} \text{ on } \mathcal{F}(T).
$$

Remark 2.5.5. *Since the density function f can be 0, to avoid dividing by 0, we assume $\widetilde{f}(y) = 0$ whenever $f(y) = 0$.*

We have the important results: (Shreve's Lemma 11.6.6, Theorem 11.6.7)

Theorem 2.5.6. *Z(t) defined in (2.2) is a $\mathbb{P}$ martingale (w.r.t. $\mathcal{F}(t)$). Under $\mathbb{Q}$, $Q(t)$ is a compound Poisson process with rate $\widetilde{\lambda}$ and Y_i has continuous distribution with density $\widetilde{f}$.*

Proof.

The proof of this Theorem relies on the following exponential martingale characterization of a compound Poisson process:

Let $\phi(u) := \mathbb{E}(e^{uY})$ be the moment generating function of a random variable Y. Then $Q(t)$ is a compound Poisson process with jump rate λ and i.i.d. jump size Y_i with moment generating function $\phi(u)$ if and only if

$$
Z(t) = \exp\big(uQ(t) - \lambda t(\phi(u) - 1)\big)
$$

is a martingale $\forall u \in \mathbb{R}$.

The details are left to the readers.

2.6 Change of measure for compound Poisson process and Brownian motion

We now consider the case when we have both a compound Poisson process $Q(t)$ and a Brownian motion $W(t)$. We want to find a change of measure kernel $Z(t)$ that would change the rate and the jump distribution of $Q(t)$ and the drift of $W(t)$. First we discuss an easier case when $Q(t)$ is just a Poisson process.

2.6.1 Change of measure for Poisson process and Brownian motion

We first describe an exponential martingale characterization result for Poisson process and Brownian motion.

Lemma 2.6.1. *$N(t)$ is a Poisson process with rate λ and $W(t)$ is a Brownian motion adapted to a filtration $\mathcal{F}(t)$ and they are independent if and only if*

$$e^{u_1 W_t - \frac{1}{2}u_1^2 t + u_2 N_t - \lambda t(e^{u_2}-1)}$$

is a $\mathcal{F}(t)$-martingale for all $u_1, u_2 \in \mathbb{R}$.

The proof of the Lemma follows a similar idea as the proof of the exponential martingale characterization of a Brownian motion or a Poisson process described above. It is clear that when the martingale condition holds then W_t is a Brownian Motion and $N(t)$ is a Poisson process since we can choose $u_1 = 0$ or $u_2 = 0$. The independence follows from the Kac's theorem for characteristic function mentioned in (2.3.2) since the martingale condition being true also implies that

$$E\Big(e^{u_1 W_t + u_2 N_t}\Big) = E\Big(e^{u_1 W_t}\Big)E\Big(e^{u_2 N_t}\Big), \forall u_1, u_2 \in \mathbb{R}.$$

An interesting thing to note is that if $W(t)$ is a Brownian motion and $N(t)$ is a Poisson process adapted to the same filtration $\mathcal{F}(t)$ then they are **automatically** independent. To see this, apply Ito's formula to $e^{u_1 W_t - \frac{1}{2}u_1^2 t + u_2 N_t - \lambda t(e^{u_2}-1)}$ to conclude that it is a martingale. Then we can invoke Kac's theorem to show independence.

With the above characterization, the following change of measure result is automatic upon our previous discussion on the change of measure for Brownian motion and Poisson process.

Theorem 2.6.2. *Let $N(t)$ be a Poisson process with rate λ and $W(t)$ is a Brownian motion under $\mathbb{P}$. Let $\widetilde{\lambda} > 0$ be given. Define*

$$\begin{aligned} Z_1(t) &:= \exp\Big[-\int_0^t \theta(u)dW(u) - \frac{1}{2}\int_0^t \theta^2(u)du\Big]; \\ Z_2(t) &:= e^{(\lambda-\widetilde{\lambda})t}\left(\frac{\widetilde{\lambda}}{\lambda}\right)^{N(t)}; \\ Z(t) &:= Z_1(t)Z_2(t). \end{aligned}$$

Also define

$$d\mathbb{Q} = Z(T)d\mathbb{P} \text{ on } \mathcal{F}(T).$$

Then $\widetilde{W}(t) = W(t) + \int_0^t \theta(u)du$ *is a Brownian motion and* $N(t)$ *is a Poisson process with rate* $\widetilde{\lambda}$ *and they are independent under* $\mathbb{Q}$*.*

2.6.2 Compound Poisson process and Brownian motion

Let $Q(t)$ be a compound Poisson process with rate λ and $W(t)$ a Brownian motion defined on the *same* probabilty space $(\mathbb{P}, \Omega, \mathcal{F})$ and $\mathcal{F}(t)$ a filtration for $Q(t), W(t)$. Here we also assume Y_i has continuous distribution with density function f.

We want to change the intensity of $Q(t)$, the distribution of Y_i (that is the density f) and the drift of $W(t)$ via the change of measure. More speficially, given a function $\theta(u)$, constant $\widetilde{\lambda} > 0$ and density function $\widetilde{f}$, we find the probabilty measure $\mathbb{Q}$ such that under $\mathbb{Q}$, $Q(t)$ is compound Poisson with rate $\widetilde{\lambda}$, $Y(i)$ has density $\widetilde{f}$ and $\widetilde{W}(t) := \int_0^t \theta(u)du + W(t)$ is a Brownian motion. Here we also assume that $\widetilde{f}(y) = 0$ when $f(y) = 0$.

Remark 2.6.3. *Before we proceed, we note that necessarily in this case* $W(t)$ *and* $Q(t)$ *are independent as remarked above (see also Corollary 11.4.9 and Exercise 11.6 in Shreve's).*

Definition 2.6.4. *Fix* $T > 0$*. Let* $\widetilde{\lambda} > 0$ *and a density function* $\widetilde{f}$ *be given. Define*

$$\begin{aligned}
Z_1(t) &:= \exp\Big[-\int_0^t \theta(u)dW(u) - \frac{1}{2}\int_0^t \theta^2(u)du\Big];\\
Z_2(t) &:= e^{(\lambda-\widetilde{\lambda})t}\prod_{i=1}^{N(t)}\left(\frac{\widetilde{\lambda}\widetilde{f}(Y_i)}{\lambda f(Y_i)}\right), 0 \le t \le T;\\
Z(t) &:= Z_1(t)Z_2(t).
\end{aligned}$$

Also define

$$d\mathbb{Q} = Z(T)d\mathbb{P} \text{ on } \mathcal{F}(T).$$

Remark 2.6.5. *Note that* $Z_1(t)$ *is the usual change of measure kernel given by the Girsanov's theorem in Section 5.2. This together with the result in Section* (2.5.4) *and Remark* (2.6.3)*, it is no surprise that* $Z(t)$ *has such form.*

We have the important results: (Shreve's Lemma 11.6.8, Theorem 11.6.9)

Theorem 2.6.6. *Z(t) is a* $\mathbb{P}$ *martingale (w.r.t.* $\mathcal{F}(t)$*). Under* $\mathbb{Q}$*,* $Q(t)$ *is a compound Poisson process with rate* $\widetilde{\lambda}$*,* Y_i *has continuous distribution with density* $\widetilde{f}$*,* $\widetilde{W(t)} = \int_0^t \theta(u)du + W(t)$ *is a Brownian motion. Moreover,* $Q(t)$ *and* $\widetilde{W(t)}$ *are independent under* $\mathbb{Q}$*.*

Remark 2.6.7. *Note that we have the parallel between the independence between $Q(t)$ and $W(t)$ under $\mathbb{P}$ and the independence between $Q(t)$ and $\widetilde{W}(t)$ under $\mathbb{Q}$. This is important since we do not have any restriction on $\theta(t)$. Indeed $\theta(t)$ can be equal to $Q(t)$ and the independence structure still holds.*

Remark 2.6.8. *Even though the theorem in Shreve is stated for Y_i having continuous distribution, it is easy to see that a similar result still holds if Y_i has discrete distribution. In this case, under $\mathbb{Q}$, Y_i would also have discrete distribution with a probability distribution $\widetilde{p}$ (see Section (2.5)). The change of measure kernel $Z_1(t)$ is the same,*

$$Z_2(t) \quad := \quad e^{(\lambda-\widetilde{\lambda})t} \prod_{i=1}^{N(t)} \frac{\widetilde{\lambda}\widetilde{p}(Y_i)}{\lambda p(Y_i)}, 0 \leq t \leq T;$$

and $Z(t) = Z_1(t) Z_2(t)$.

CHAPTER 3 Pricing European call option in jump diffusion models

3.1 Pricing via risk neutral expectation

3.1.1 Model with Poisson noise

Change of measure

Let N_t be a Poisson process with rate λ. Supose we model the stock price as

$$dS_t = \alpha S_t dt + \sigma S_{t-} dM(t),$$

where $M(t) = N_t - \lambda t$ is a $\mathbb{P}$-martingale. Note that here the only random source of S_t is from the jump process N_t.

From section 9 of lecture note 1, we also have

$$S_t = S(0)\exp[(\alpha - \lambda\sigma)t + \log(1+\sigma)N_t].$$

Let $r > 0$ be the interest rate. We want to find $\mathbb{Q}$ such that $e^{-rt}S_t$ is a $\mathbb{Q}$ martingale. If that is the case, since

$$dS_t = rS_t dt + \sigma S_{t-}(dN_t - [\lambda - \frac{\alpha - r}{\sigma}]dt)$$

clearly we need N_t to be a Poisson process with rate $\widetilde{\lambda} = \lambda - \frac{\alpha-r}{\sigma}$. Since $\widetilde{\lambda}$ must be positive, a necessary condition (which implies no arbitrage for the model of S_t) is

$$\lambda - \frac{\alpha - r}{\sigma} > 0.$$

We then define

$$\begin{aligned} d\mathbb{Q} &= Z(T)d\mathbb{P}; \\ Z(t) &= \exp[\log\left(\frac{\widetilde{\lambda}}{\lambda}\right) N_t - (\widetilde{\lambda} - \lambda)t)]. \end{aligned}$$

Note that under $\mathbb{Q}$, we write the dynamics of S_t as

$$dS_t = rS_t dt + \sigma S_{t-} d\widetilde{M}(t),$$

where $\widetilde{M}(t) = N_t - \widetilde{\lambda}t$ is a $\mathbb{Q}$-martingale, which is equivalent to

$$S_t = S(0)\exp[(r - \widetilde{\lambda}\sigma)t + \log(1+\sigma)N_t].$$

Pricing of European call

Let $V(t)$ denote the risk-neutral price of a European Call paying $V(T) = (S_T - K)^+$ at time T. Then by the risk neutral pricing formula, we have

$$V(t) = \mathbb{E}^{\mathbb{Q}}\big[e^{-r(T-t)}(S_T - K)^+|\mathcal{F}(t)\big].$$

It remains to find an expression for $V(t)$. Clearly

$$S_T = S_t \exp[(r - \widetilde{\lambda}\sigma)(T - t) + \log(1 + \sigma)(N_T - N_t)].$$

So by the Independence Lemma, (Shreve's Lemma (2.3.4)), we only need to evaluate

$$c(t, x) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\big(xe^{(r-\widetilde{\lambda}\sigma)(T-t)+\log(1+\sigma)(N_T-N_t)} - K\big)^+\Big],$$

then we have $V(t) = c(t, S_t)$.

Since $N_T - N_t$ has distribution Poisson($\widetilde{\lambda}(T - t)$) under $\mathbb{Q}$, $c(t, x)$ has the expression as an infinite sum, see Shreve's formula (11.7.3). We won't reproduce it here.

3.1.2 Model with compound Poisson noise

Change of measure

Suppose now that

$$dS_t = \alpha S_t dt + \sigma S_{t-}dM(t),$$

where $M(t) = Q(t) - mt$ is a compensated compound Poisson process under $\mathbb{P}$. Under the risk neutral probability $\mathbb{Q}$,

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_{t-}d\tilde{M}(t) \\ &= (r - \sigma\tilde{m})S_t dt + \sigma S_{t-}dQ_t. \end{aligned}$$

So clearly we need

(i) $Q(t)$ to be a compound Poisson process under $\mathbb{Q}$ with $\mathbb{E}^{\mathbb{Q}}(Q(1)) = \tilde{m}$.

(ii) $r - \sigma\tilde{m} = \alpha - \sigma m$.

Note that (ii) gives an equation for $\tilde{m}$. If $Q(t) = \sum_{i=1}^{N_t} Y_i$ and under $\mathbb{Q}$, N_t is a Poisson process with rate $\tilde{\lambda}$ and $\mathbb{E}(Y_i) = \tilde{\mu}$ then

$$\tilde{m} = \tilde{\lambda}\tilde{\mu}.$$

So (ii) also gives an equation for $\tilde{\lambda}$ and $\widetilde{f}$, the distribution of Y_i under $\mathbb{Q}$. From the change of measure sections, we have seen how to choose $Z(T)$ such that the conditions (i) and (ii) are satisfied. Note that this choice may not be unique, as generally equation (ii) has more than 1 unknowns. However, there is also a restriction on the solution $\tilde{\lambda} > 0$. So a simple application of linear algebra result to conclude that there are infinitely many choices of risk neutral measures is not correct.

Pricing of European call option

Observe that

$$dS_t = (r - \sigma\tilde{m})S_t dt + \sigma S_{t-} dQ_t$$

has the solution

$$\begin{aligned} S_t &= S(0)e^{(r-\sigma\tilde{m})t} \prod_{0<s\le t} (1 + \sigma\Delta Q(s)) \\ &= S(0)e^{(r-\sigma\tilde{m})t} \prod_{i=1}^{N_t}(1 + \sigma Y_i). \end{aligned}$$

Also for $t < T$

$$S_T = S_t e^{(r-\sigma\tilde{m})(T-t)} \prod_{i=N_t+1}^{N_T} (1 + \sigma Y_i).$$

Observe *the important fact* that $\prod_{i=N_t+1}^{N_T}(1 + \sigma Y_i)$ *is independent of* $\mathcal{F}(t)$, where $\mathcal{F}(t)$ is a filtration for $Q(t)$. We give an explanation in the next subsection.

Thus $V(t)$, the risk-neutral price of a European Call paying $V(T) = (S_T - K)^+$ at time T for this model is

$$\begin{aligned} V(t) &= \mathbb{E}^{\mathbb{Q}}\big[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}(t)\big] \\ &= c(t, S_t), \end{aligned}$$

where

$$c(t,x) := e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[[xe^{(r-\sigma\tilde{m})(T-t)} \prod_{i=N_t+1}^{N_T} (1 + \sigma Y_i) - K]^+\Big].$$

Since Y_i are independent of $N_T - N_t$, again we can condition on $N_T - N_t = j, j = 1, 2, ...$ to get

$$c(t,x) = e^{-r(T-t)} \sum_{j=0}^{\infty} \kappa(j,x) e^{-\tilde{\lambda}(T-t)} \frac{[\tilde{\lambda}(T-t)]^j}{j!},$$

where

$$\kappa(j,x) = \mathbb{E}^{\mathbb{Q}}\Big[\big(xe^{(r-\sigma\tilde{m})(T-t)} \prod_{i=1}^{j}(1 + \sigma Y_i) - K\big)^+\Big].$$

The independence of $\prod_{i=N_t+1}^{N_T}(1+\sigma Y_i)$ from $\mathcal{F}(t)$

In the derivation above, we claim that $\prod_{i=N_t+1}^{N_T}(1+\sigma Y_i)$ from $\mathcal{F}(t)$. This is not totally obvious since the terms involve N_t, which is in $\mathcal{F}(t)$. However, intuitively you can see that this is believable because there are $N_T - N_t$ terms in the product, which consists of Y_i's. Both $N_T - N_t$ and Y_i's are independent of $\mathcal{F}(t)$.

Rigorously, we use the result mentioned in Lecture 2. That is if $E(e^{uX}|\mathcal{F}) = E(e^{uX})$ for all $u \in \mathbb{R}$ then X is independent of $\mathcal{F}$. We verify that this is the case here. That is we want to show

$$E\Big(e^{u\prod_{i=N_t+1}^{N_T}(1+\sigma Y_i)}|\mathcal{F}_t\Big) = E\Big(e^{u\prod_{i=N_t+1}^{N_T}(1+\sigma Y_i)}\Big), \forall u \in \mathbb{R}.$$

Observe that the above expression would be complicated to handle. But we can simplify it by noting that we can instead show the independence of

$$\log\Big(\prod_{i=N_t+1}^{N_T}(1+\sigma Y_i)\Big) = \sum_{i=N_t+1}^{N_T}\log(1+\sigma Y_i)$$

with $\mathcal{F}_t$. Moreover,

$$e^{u\sum_{i=N_t+1}^{N_T}\log(1+\sigma Y_i)} = \prod_{i=N_t+1}^{N_T} e^{u\log(1+\sigma Y_i)}.$$

Thus we see that we can just prove this general claim for our purpose: let $X_1, X_2, \cdots$ be i.i.d and be independent of $N_t, t > 0$. Then

$$E\Big(\prod_{i=N_t+1}^{N_T} X_i|\mathcal{F}_t\Big) = E\Big(\prod_{i=N_t+1}^{N_T} X_i\Big) = E\Big(\prod_{i=1}^{N_T-N_t} X_i\Big).$$

This indeed will be the statement we'll prove for the rest of this proof. We have

$$E\Big(\prod_{i=N_t+1}^{N_T} X_i|\mathcal{F}(t)\Big) = E\Big(\prod_{i=1}^{N_T-N_t} X_{i+N_t}|\mathcal{F}(t)\Big).$$

Since $N_T - N_t$ is independent of $\mathcal{F}(t)$, by the Independence lemma,

$$\mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_{i+N_t}|\mathcal{F}(t)\big] = f(N_t),$$

where

$$f(k) = \mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_{i+k}\big].$$

We'll be done if we can show

$$f(k) = f(0) = \mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_i\big].$$

Note that

$$\begin{aligned}
\mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_{i+k}\big] &= \sum_j \mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_{i+k} | N_T - N_t = j\big] P(N_T - N_t = j) \\
&= \sum_j \mathbb{E}\big[\prod_{i=1}^{j} X_{i+k} | N_T - N_t = j\big] P(N_T - N_t = j) \\
&= \sum_j \mathbb{E}\big[\prod_{i=1}^{j} X_{i+k}\big] P(N_T - N_t = j) \\
&= \sum_j \Big\{\mathbb{E}\big[X_1\big]\Big\}^j P(N_T - N_t = j),
\end{aligned}$$

where the third equality is because of the independence of X_i's and $N_T - N_t$ and the fourth equality is because of the identical distribution of X_i's.

Using the same conditioning technique, we can also show

$$\mathbb{E}\big[\prod_{i=1}^{N_T-N_t} X_i\big] = \sum_j \Big\{\mathbb{E}\big[X_1\big]\Big\}^j P(N_T - N_t = j).$$

Thus $f(k) = f(0)$ as required.

3.1.3 Model with Brownian motion and compound Poisson noise

Change of measure

Suppose now that

$$dS_t = \alpha S_t dt + \sigma S_{t-} dW(t) + S_{t-} dM(t),$$

where $M(t) = Q(t) - mt$ is a compensated compound Poisson process under $\mathbb{P}$. Under the risk neutral probability $\mathbb{Q}$,

$$\begin{aligned}
dS_t &= rS_t dt + \sigma S_{t-} d\widetilde{W}(t) + S_{t-} d\tilde{M}(t) \\
&= (r - \tilde{m})S_t dt + \sigma S_{t-} d\widetilde{W}(t) + S_{t-} dQ_t,
\end{aligned}$$

where $\widetilde{W}(t) := W(t) + \theta t$ is a $\mathbb{Q}$ Brownian motion and $Q(t)$ is compound Poisson with $\mathbb{E}^{\mathbb{Q}}(Q(1)) = \tilde{m}$.

Thus the equation that θ and $\tilde{m}$ have to satisfy is

$$r + \sigma\theta - \tilde{m} = \alpha - m.$$

Solving this equation for θ and $\tilde{m}$ and use the change of measure result discussed above, we can find $\mathbb{Q}$ such that $e^{-rt}S_t$ is a $\mathbb{Q}$ - martingale.

Pricing of European call

Observe that

$$dS_t \;=\; (r - \tilde{m})S_t dt + \sigma S_{t-} d\widetilde{W}(t) + S_{t-} dQ_t,$$

has the solution

$$S_t = S(0)\exp\left[(r - \tilde{m} - \frac{1}{2}\sigma^2)t + \sigma\tilde{W}(t)\right]\prod_{i=1}^{N_t}(1 + Y_i).$$

Hence for $t < T$,

$$S_T = S_t \exp\left[(r - \tilde{m} - \frac{1}{2}\sigma^2)(T - t) + \sigma(\tilde{W}(T) - \tilde{W}(t))\right]\prod_{i=N_t+1}^{N_T}(1 + Y_i),$$

where we have the *independence* of $\tilde{W}(T) - \tilde{W}(t)$ and $\prod_{i=N_t+1}^{N_T}(1 + Y_i)$ with respect to $\mathcal{F}(t)$ and *also with respect to each other*.

Thus $V(t)$, the risk-neutral price of a European Call paying $V(T) = (S_T - K)^+$ at time T for this model is

$$\begin{aligned} V(t) &= \mathbb{E}^{\mathbb{Q}}\big[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}(t)\big] \\ &= c(t, S_t), \end{aligned}$$

where

$$c(t,x) := e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\Big(xe^{(r-\tilde{m}-\frac{1}{2}\sigma^2)(T-t)+\sigma(\tilde{W}(T)-\tilde{W}(t))}\prod_{i=N_t+1}^{N_T}(1 + Y_i) - K\Big)^+\Big].$$

To find an expression for $c(t,x)$, we first condition on $\prod_{i=N_t+1}^{N_T}(1 + Y_i)$ and use the independence lemma to define a function $\kappa(t,x)$ as

$$\kappa(t,x) := e^{-rt}\mathbb{E}^{\mathbb{Q}}\Big[\Big(xe^{(r-\frac{1}{2}\sigma^2)t+\sigma\sqrt{t}Y} - K\Big)^+\Big],$$

where Y has standard normal distribution. Note that we have an explicit expression for $\kappa(t,x)$ from the Black-Scholes formula. Then

$$c(t,x) = \mathbb{E}^{\mathbb{Q}}\big[\kappa(T-t, xe^{-\tilde{m}(T-t)}\prod_{i=N_t+1}^{N_T}(1+Y_i))\big].$$

Now again conditioning on $N_T - N_t = j$ and using the independence between $Y_i's$ and $N_T - N_t$ we have

$$c(t,x) = \sum_{j=0}^{\infty} e^{-\tilde{\lambda}(T-t)}\frac{(\tilde{\lambda}(T-t))^j}{j!}\mathbb{E}^{\mathbb{Q}}\big[\kappa(T-t, xe^{-\tilde{m}(T-t)}\prod_{i=1}^{j}(1+Y_i))\big].$$

3.2 Pricing via partial differential difference equations

3.2.1 Heuristic

Suppose S_t satisfies

$$dS_t = \alpha S_t dt + \sigma S_t dM(t),$$

where $M(t) = N_t - \lambda t$ is a compensated Poisson process under $\mathbb{P}$.

From the change of measure section, we learned that under the risk neutral measure $\mathbb{Q}$, S_t has the dynamic:

$$dS_t = (r - \tilde{\lambda}\sigma)S_t dt + \sigma S_{t-} dN_t,$$

where $\tilde{\lambda} = \lambda - \frac{\alpha - r}{\sigma}$ and N is a Poisson process with rate $\tilde{\lambda}$ under $\mathbb{Q}$.

The call option price $V(t)$, where $V(T) = (S_T - K)^+$ can be written as

$$\begin{aligned} V(t) &= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}(t)\Big] \\ &= c(t, S_t), \end{aligned}$$

where

$$c(t,x) := e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[(xe^{(r-\tilde{\lambda}\sigma)(T-t)+\log(1+\sigma)(N_T - N_t)} - K)^+\Big].$$

As in the Black-Scholes model, we want to derive an equation that $c(t,x)$ satisfies. *The key principle* here is to apply Ito's formula to $e^{-rt}c(t,S_t)$ to achieve

$$de^{-rt}c(t,S_t) = f(t, c(t,S_t))dt + \text{ something } dM(t),$$

where $M(t)$ *is a* $\mathbb{Q}$*-martingale*. Then the equation that we look for is

$$f(t, c(t,S_t)) = 0.$$

The reason is that $e^{-rt}c(t,S_t)$ is a $\mathbb{Q}$-martingale by definition. Therefore, its drift has to be 0.

3.2.2 Model with Poisson noise

Suppose S_t satisfies

$$dS_t = \alpha S_t dt + \sigma S_t dM(t),$$

where $M(t) = N_t - \lambda t$ is a compensated Poisson process under $\mathbb{P}$.

Apply Ito's formula to $e^{-rt}c(t, S_t)$, recognizing there is no Brownian motion component, we have

$$\begin{aligned}
e^{-rt}c(t,S_t) &= \int_0^t -re^{-ru}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial t}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial x}c(u,S_u)dS^c(u) \\
&\quad + \sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u-,S_{u-})] \\
&= \int_0^t e^{-ru}\Big[-rc(u,S_u) + \frac{\partial}{\partial t}c(u,S_u) + \frac{\partial}{\partial x}c(t,S_u)(r-\tilde{\lambda}\sigma)S_u\Big]du \\
&\quad + \sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u,S_{u-})].
\end{aligned}$$

We need to rewrite $\sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u,S_{u-})]$ as it is *not in differential form*. Two key observations will help us here:

(i) $S_u = (1+\sigma\Delta N(u))S_{u-} = (1+\sigma)S_{u-}$.

(ii) $c(u, S_u)$ jumps at the same points as S_u, which in turn jumps at the same points as $N(u)$. Again keep in mind that $\Delta N(u) = 1$.

Thus

$$\begin{aligned}
\sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u,S_{u-})] &= \sum_{0<u\leq t} e^{-ru}[c(u,S_{u-}(1+\sigma)) - c(u,S_{u-})] \\
= \int_0^t e^{-ru}[c(u,S_{u-}(1+\sigma)) - c(u,S_{u-})]dN(u),
\end{aligned}$$

where the first equality uses observations (i) and second equality uses observation (ii).

Putting all these together gives

$$\begin{aligned}
e^{-rt}c(t,S_t) &= \int_0^t e^{-ru}\Big[-rc(u,S_u) + \frac{\partial}{\partial t}c(u,S_u) + \frac{\partial}{\partial x}c(t,S_u)(r-\tilde{\lambda}\sigma)S_u\Big]du \\
&\quad + \int_0^t e^{-ru}[c(u,S_{u-}(1+\sigma)) - c(u,S_{u-})]dN(u).
\end{aligned}$$

The last thing to do is to change $dN(u)$ to $dM(u)$ for some martingale M. This is easy:

we only need to subtract and add $\tilde{\lambda} du$ to $dN(u)$. So finally

$$\begin{aligned}
e^{-rt}c(t,S_t) &= \int_0^t e^{-ru}\Big[-rc(u,S_u)+\frac{\partial}{\partial t}c(u,S_u)+\frac{\partial}{\partial x}c(t,S_u)(r-\tilde{\lambda}\sigma)S_u \\
&\quad +[c(u,S_{u-}(1+\sigma))-c(u,S_{u-})]\tilde{\lambda}\Big]du \\
&\quad +\int_0^t e^{-ru}[c(u,S_{u-}(1+\sigma))-c(u,S_{u-})]dM(u) \\
&= \int_0^t e^{-ru}\Big[-rc(u,S_u)+\frac{\partial}{\partial t}c(u,S_u)+\frac{\partial}{\partial x}c(t,S_u)(r-\tilde{\lambda}\sigma)S_u \\
&\quad +[c(u,S_u(1+\sigma))-c(u,S_u)]\tilde{\lambda}\Big]du \\
&\quad +\int_0^t e^{-ru}[c(u,S_{u-}(1+\sigma))-c(u-,S_{u-})]dM(u),
\end{aligned}$$

where in the second equality we uses the fact that we are integrating with respect to du so using S_{u-} or S_u gives the same result.

Now apply the principle in Section (3.2.1) we get

Theorem 3.2.1. *The call option price $c(t,x)$ in the model of this section satisfies the differential difference equation*

$$\begin{aligned}
-rc(t,x) &+ \frac{\partial}{\partial t}c(t,x)+(r-\tilde{\lambda}\sigma)x\frac{\partial}{\partial x}c(t,x) \\
&+ \tilde{\lambda}[c(t,x(1+\sigma))-c(t,x)]=0, 0\le t<T, x>0 \\
c(T,x) &= (x-K)^+, x>0.
\end{aligned}$$

3.2.3 Model with compound Poisson noise

Suppose S_t has the dynamic:

$$dS_t=(r-\tilde{m}\sigma)S_t dt+\sigma S_{t-}dQ_t,$$

where $Q(t)$ is a compound Poisson process with rate $\mathbb{E}^{\mathbb{Q}}(Q(1))=1$. We also assume that $Q(t)=\sum_{i=1}^{N_t} Y_i$ where each Y_i takes discrete distribution with values $y_1, y_2, ..., y_m$.

Following the same procedure as the above section, apply Ito's formula to $e^{-rt}c(t,S_t)$

gives

$$\begin{aligned} e^{-rt}c(t,S_t) &= \int_0^t -re^{-ru}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial t}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial x}c(u,S_u)dS^c(u) \\ &\quad + \sum_{0<u\le t} e^{-ru}[c(u,S_u) - c(u-,S_{u-})] \\ &= \int_0^t e^{-ru}\Big[-rc(u,S_u) + \frac{\partial}{\partial t}c(u,S_u) + \frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m}\sigma)S_u\Big]du \\ &\quad + \sum_{0<u\le t} e^{-ru}[c(u,S_u) - c(u,S_{u-})]. \end{aligned}$$

Now by the Poisson process decomposition, we can write

$$Q(t) = \sum_{i=1}^m y_i N_i(t),$$

where each $N_i(t)$ is a Poisson process with rate $\tilde{\lambda}_i, i = 1, ..., m$ under $\mathbb{Q}$. An important fact here is that since N_i's are independent, *they do not jump at the same time*. So at all jump point of Q:

$$1 + \sigma\Delta Q(t) = 1 + \sigma y_i \Delta N_i(t), \text{ for some } i.$$

Thus we have,

$$\begin{aligned} \sum_{0<u\le t} e^{-ru}[c(u,S_u) - c(u,S_{u-})] &= \sum_{0<u\le N_t} e^{-ru}[c(u,S_{u-}(1+\sigma\Delta Q_u)) - c(u,S_{u-})] \\ = \sum_{i=1}^m \Big[\sum_{0<u\le t} e^{-ru}[c(u,S_{u-}(1+\sigma y_i)) - c(u,S_{u-})]\Delta N_i(u)\Big] \\ = \sum_{i=1}^m \Big[\int_0^t e^{-ru}[c(u,S_{u-}(1+\sigma y_i)) - c(u,S_{u-})]dN_i(u)\Big]. \end{aligned}$$

So

$$\begin{aligned} e^{-rt}c(t,S_t) &= \int_0^t e^{-ru}\Big[-rc(u,S_u) + \frac{\partial}{\partial t}c(u,S_u) + \frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m}\sigma)S_u \\ &\quad + \sum_{i=1}^m [c(u,S_u(1+\sigma y_i)) - c(u,S_u)]\tilde{\lambda}_i\Big]du \\ &\quad + \int_0^t e^{-ru}[c(u,S_u) - c(u,S_{u-})]dM(u), \end{aligned}$$

where

$$M(t) = \sum_{i=1}^{m} N_i(t) - \tilde{\lambda}_i t$$

is a $\mathbb{Q}$-martingale.

Setting the dt part to be 0 gives the following:

Theorem 3.2.2. *The call option price $c(t,x)$ in the model of this section satisfies the differential difference equation*

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x) + (r - \tilde{m}\sigma)x\frac{\partial}{\partial x}c(t,x) \\ &+ \sum_{i=1}^{m}[c(t, x(1+\sigma y_i)) - c(t,x)]\tilde{\lambda}_i = 0, 0 \leq t < T, x > 0 \\ c(T,x) &= (x-K)^+, x > 0. \end{aligned}$$

3.2.4 Model with Brownian motion and compound Poisson noise

Suppose S_t has the dynamic:

$$dSt = (r - \tilde{m})S_t dt + S_{t-}dQ_t + \sigma S_t d\tilde{W}(t),$$

where $Q(t)$ is a compound Poisson process with rate $\mathbb{E}^{\mathbb{Q}}(Q(1)) = 1$ and $\tilde{W}(t)$ is a $\mathbb{Q}$ Brownian motion. We also assume that $Q(t) = \sum_{i=1}^{N_t} Y_i$ where each Y_i takes discrete distribution with values $y_1, y_2, ..., y_m$.

Following the same procedure as the above section, apply Ito's formula to $e^{-rt}c(t, S_t)$ gives

$$\begin{aligned} e^{-rt}c(t,S_t) &= \int_0^t -re^{-ru}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial t}c(u,S_u)du + e^{-ru}\frac{\partial}{\partial x}c(u,S_u)dS^c(u) \\ &\quad + \frac{1}{2}e^{-ru}\frac{\partial^2}{\partial x^2}c(u,S_u)\sigma^2 S^2(u)du + \sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u-,S_{u-})] \\ &= \int_0^t e^{-ru}\Big[-rc(u,S_u) + \frac{\partial}{\partial t}c(u,S_u) + \frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m})S_u \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}c(u,S_u)\sigma^2 S^2(u)\Big]du \\ &\quad + \int_0^t e^{-ru}\frac{\partial}{\partial x}c(t,S_u)S_u d\widetilde{W}(u) + \sum_{0<u\leq t} e^{-ru}[c(u,S_u) - c(u,S_{u-})]. \end{aligned}$$

Follow the same exact analysis for $\sum_{0<u\leq t} e^{-ru}[c(u,S_u)-c(u,S_{u-})]$ as in section (3.2.3) we have

$$\begin{aligned} e^{-rt}c(t,S_t) &= \int_0^t e^{-ru}\Big[-rc(u,S_u)+\frac{\partial}{\partial t}c(u,S_u)+\frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m})S_u \\ &\quad +\frac{1}{2}\frac{\partial^2}{\partial x^2}c(u,S_u)\sigma^2S^2(u)+\sum_{i=1}^{m}[c(u,S_u(1+y_i))-c(u,S_u)]\tilde{\lambda}_i\Big]du \\ &\quad +\int_0^t e^{-ru}\frac{\partial}{\partial x}c(t,S_u)S_u d\widetilde{W}(u)+\int_0^t e^{-ru}[c(u,S_u)-c(u,S_{u-})]dM(u), \end{aligned}$$

where

$$M(t)=\sum_{i=1}^{m}N_i(t)-\tilde{\lambda}_i t$$

is a $\mathbb{Q}$-martingale.

Setting the dt part to be 0 gives the following:

Theorem 3.2.3. *The call option price $c(t,x)$ in the model of this section satisfies the differential difference equation*

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x)+(r-\tilde{m})x\frac{\partial}{\partial x}c(t,x)+\frac{1}{2}\frac{\partial^2}{\partial x^2}c(t,x)\sigma^2x^2 \\ &+ \sum_{i=1}^{m}[c(t,x(1+y_i))-c(t,x)]\tilde{\lambda}_i=0, 0\leq t<T, x>0; \\ c(T,x) &= (x-K)^+, x>0. \end{aligned}$$

3.2.5 A unifying approach via Levy measure

Note that all of the above derivations rely on the decomposition of a compound Poisson process with discrete jumps into sums of individual Poisson processes. This technique obviously does not work when we have a compound Poisson process with continuous jump distribution. The way to handle this situation is via the concept of the Levy measure. It will also help us write one single type of equation, called Partial Integro-Differential Equation (PIDE), for all types of our noise, as long as they are compound Poisson process plus a Brownian motion. For a more detailed treament of Levy process with application to finance, see e.g. [2].

Definition 3.2.4. *Let L_t be a Levy process and $A \in \mathcal{B}(\mathbb{R})$ be a Borel measurable subset of the real line. Define*

$$\mu_t^L(A) \ := \ \sum_{0<s\leq t} \mathbf{1}_{\Delta L_s \in A};$$

that is, $\mu_t^L(A)$ counts the number of jumps of L, up to time t, that have size in the set A. We call $\mu_t^L(\cdot)$ the Poisson random measure associated with the Levy process L_t. Note that for a fixed A, $\mu_t^L(A)$ is a counting process. Also define

$$\nu(A) := \mathbb{E}(\mu_1^L(A)).$$

We say ν is the Levy measure associated with the Levy process L_t.

Remark 3.2.5. *In the above definition, usually one would require that the point 0 is "far away" from the set A, that is $0 \notin \overline{A}$. This is because a Levy process L_t can have infinitely many small jumps close to 0, which in turn may make $\mu_t^L(A)$ to be infinite if $0 \in \overline{A}$. However, in the cases we're dealing with, namely upto compound Poisson process, this will not happen. The number of jumps of compound Poisson process in any finite time interval $[0, t]$ will always remain finite. So we do not have to include this restriction in the set A, for ease of introduction to the material.*

Observe that for a fixed A, μ_t^L has independent and stationary increment, which is inherited from the Levy process L_t. Therefore, $\mu_t^L(A)$ is a Poisson process with rate

$$\lambda^A = \nu(A) = \mathbb{E}(\mu_1^L(A)).$$

In other words, the Levy measure ν measures *the expected number of jumps of L_t of a certain height in a time interval of length 1*. The height is determined by what values of the set A you plug in to the measure ν. We list what ν is for the processes we were familiar with in this chapter.

1. Poisson process with rate λ:

$$\nu(dx) = \lambda\delta_1(dx).$$

2. Compound Poisson process with rate λ and discrete jumps $y_1, \cdots, y_M$:

$$\nu(dx) = \lambda \sum_{m=1}^{M} p_m \delta_{y_m}(dx).$$

3. Compound Poisson process with rate λ and continuous jump distribution $f_Y(x)$:

$$\nu(dx) = \lambda f_Y(dx).$$

Remark 3.2.6. *Note that in all of the above examples,*

$$\lambda = \int_{\mathbb{R}} \nu(dx) = \nu(\mathbb{R}).$$

This indeed will be the case for all compound Poisson processes: they have finite Levy measure and the rate is equal to the Levy measure of the real line.

Integrating with respect to the random Poisson measure

For a Levy process L_t with Levy measure ν, adapted to a filtration $\mathcal{F}_t$. We define

$$\int_0^t \int_A f(s,x)\mu^L(ds,dx) := \sum_{0<s\le t} f(s,\Delta L_s)\mathbf{1}_{\Delta L_s \in A}.$$

That is, the integral $\int_0^t \int_A f(s,x)\mu^L(ds,dx)$ is a pure jump process that jumps at the same time as L, with the jump size $f(s,\Delta L_s)$ if the jump of L happens at time s.

What will be important for us is the following martingale result:

Theorem 3.2.7. *Let $f(s,x,\omega)$ be a process with left continuous with right limit paths adapted to the filtration $\mathcal{F}_t$ satisfying certain integrability conditions. Then*

$$\int_0^t \int_A f(s,x,\omega)[\mu^L(ds,dx) - \nu(dx)ds]$$

is a $\mathcal{F}_t$-martingale.

Proof. The proof starts by approximating $f(t,x,\omega)$ by simple processes of the form $\sum_{k=1}^m \xi_k(t)\phi_k(x)$, where $\xi_k(t)$ are $\mathcal{F}_t$ measurable processes and ϕ_k are deterministic functions of x. We prove the martingale property for these simple processes and prove the general result by a convergence argument. For details see [1].

Ito's formula for jump processes, random Poisson measure version

Let X_t be a process of the form

$$X(t) = X_0 + \int_0^t \alpha(s)ds + \int_0^t \gamma(s)dWs + J(t),$$

where $J(t)$ is a compound Poisson process. Let f be a $C^{1,2}$ function. Then

$$\begin{aligned} f(t, X(t)) &= f(0, X_0) + \int_0^t f_t(s, X_s) ds \\ &+ \int_0^t f_x(s, X_s) dX^c(s) + \int_0^t \frac{1}{2} f_{xx}(s, X_s)\gamma^2(s) ds \\ &\quad + \sum_{0<s\le t} f(s, X_s) - f(s, X_{s-}) \\ &= f(0, X_0) + \int_0^t f_t(s, X_s) ds \\ &+ \int_0^t f_x(s, X_s) dX^c(s) + \int_0^t \frac{1}{2} f_{xx}(s, X_s)\gamma^2(s) ds \\ &\quad + \int_0^t \int_{\mathbb{R}} \Big[f(s, X_{s-} + x) - f(s, X_{s-}) \Big] \mu^J(ds, dx). \end{aligned}$$

The reason for the re-writing in the random Poisson measure version is clear: we want to use the martingale result mentioned in the previous section. The equality

$$\sum_{0<s\le t} f(s, X_s) - f(s, X_{s-}) = \int_0^t \int_{\mathbb{R}} \Big[f(s, X_{s-} + x) - f(s, X_{s-}) \Big] \mu^J(ds, dx)$$

comes from the fact that the jumps of X comes from the jumps of J, and $\Delta X_s = \Delta J_s$ at all jump times s.

PIDE for Euro call option with compound Poisson process and Brownian motion noise

Now suppose S_t has the dynamic:

$$dSt = rS_t dt + \sigma S_t d\tilde{W}(t) + \gamma S_{t-} d(Q_t - \tilde{\mu} t),$$

where we added a volatility component γ in the compound Poisson part for generality, even though this is not strictly necessary as it can be incoporated into the jumps of Q. This is the parameter σ in the previous sections (3.2.2), (3.2.3).

Recall that applying the Ito's formula, we have

$$\begin{aligned} e^{-rt} c(t, S_t) &= \int_0^t e^{-ru} \Big[-rc(u, S_u) + \frac{\partial}{\partial t} c(u, S_u) + \frac{\partial}{\partial x} c(t, S_u)(r - \tilde{m}) S_u \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} c(u, S_u)\sigma^2 S^2(u) \Big] du \\ &\quad + \int_0^t e^{-ru} \frac{\partial}{\partial x} c(t, S_u) S_u d\widetilde{W}(u) + \sum_{0<u\le t} e^{-ru} [c(u, S_u) - c(u, S_{u-})]. \end{aligned}$$

Rewriting the term $\sum_{0<u\le t} e^{-ru}[c(u,S_u)-c(u,S_{u-})$ using random Poisson measure we have

$$\begin{aligned}\sum_{0<u\le t} e^{-ru}[c(u,S_u)-c(u,S_{u-}) &= \sum_{0<u\le t} e^{-ru}[c(u,S_{u-}(1+\gamma\Delta Q_t))-c(u,S_{u-})] \\ &= \int_0^t\int_{\mathbb{R}} e^{-ru}[c(u,S_{u-}(1+\gamma x))-c(u,S_{u-})]\mu^Q(du,dx).\end{aligned}$$

Thus applying the martingale result, we have

$$\begin{aligned} e^{-rt}c(t,S_t) &= \int_0^t e^{-ru}\Big[-rc(u,S_u)+\frac{\partial}{\partial t}c(u,S_u)+\frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m})S_u \\ &\quad +\frac{1}{2}\frac{\partial^2}{\partial x^2}c(u,S_u)\sigma^2S^2(u)\Big]du+\int_0^t e^{-ru} \\ &\quad \frac{\partial}{\partial x}c(t,S_u)S_u d\widetilde{W}(u)+\int_0^t\int_{\mathbb{R}} e^{-ru}[c(u,S_{u-}(1+\gamma x))-c(u,S_{u-})]\mu^Q(du,dx) \\ &= \int_0^t e^{-ru}\Big[-rc(u,S_u)+\frac{\partial}{\partial t}c(u,S_u)+\frac{\partial}{\partial x}c(t,S_u)(r-\tilde{m})S_u \\ &\quad +\frac{1}{2}\frac{\partial^2}{\partial x^2}c(u,S_u)\sigma^2S^2(u)+\int_{\mathbb{R}}[c(u,S_{u-}(1+\gamma x))-c(u,S_{u-})]\nu(dx)\Big]du \\ &\quad +\int_0^t e^{-ru}\frac{\partial}{\partial x}c(t,S_u)S_u d\widetilde{W}(u) \\ &\quad +\int_0^t\int_{\mathbb{R}} e^{-ru}[c(u,S_{u-}(1+\gamma x))-c(u,S_{u-})](\mu^Q(du,dx)-\nu(dx)du).\end{aligned}$$

Therefore, $c(t,x)$ satisfies the PIDE

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x)+(r-\tilde{m})x\frac{\partial}{\partial x}c(t,x)+\frac{1}{2}\frac{\partial^2}{\partial x^2}c(t,x)\sigma^2x^2 \\ &+ \int_{\mathbb{R}}[c(t,x(1+\gamma z))-c(t,x)]\nu(dz)=0, 0\le t<T, x>0; \\ c(T,x) &= (x-K)^+, x>0.\end{aligned}$$

In particular we have:
(i) If Q is a Poisson $(\tilde{\lambda})$ process then $\nu(dz)=\tilde{\lambda}\delta_1(dz)$. Thus the PIDE becomes

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x)+(r-\tilde{m})x\frac{\partial}{\partial x}c(t,x)+\frac{1}{2}\frac{\partial^2}{\partial x^2}c(t,x)\sigma^2x^2 \\ &+ \tilde{\lambda}[c(t,x(1+\gamma))-c(t,x)]=0.\end{aligned}$$

(ii) If Q is a compound Poisson with discrete jumps then $\nu(dz) = \sum_{m=1}^{M} \tilde{\lambda}_m \delta_{ym}(dz)$. Thus the PIDE becomes

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x) + (r-\tilde{m})x\frac{\partial}{\partial x}c(t,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}c(t,x)\sigma^2 x^2 \\ &+ \sum_{m=1}^{M} \tilde{\lambda}_m [c(t, x(1+\gamma y_m)) - c(t,x)] = 0. \end{aligned}$$

(iii) If Q is a compound Poisson with continuous jump then $\nu(dz) = \lambda f(z)dz$. Thus the PIDE becomes

$$\begin{aligned} -rc(t,x) &+ \frac{\partial}{\partial t}c(t,x) + (r-\tilde{m})x\frac{\partial}{\partial x}c(t,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}c(t,x)\sigma^2 x^2 \\ &+ \int_{\mathbb{R}} [c(t, x(1+\gamma z) - c(t,x)]\lambda f(z)dz = 0. \end{aligned}$$

You should verify that for cases (i) and (ii) the results are exactly as what we got before in sections (3.2.2), (3.2.3).

CHAPTER 4 PDE for knock out barrier option

4.1 Introduction

In Chapter 7, we consider the risk neutral price for various exotic options:

(i) Knock out Barrier option:

$$V_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \le b\}}.$$

(ii) Lookback option:

$$V_T = \max_{[0,T]} S_t - S(T).$$

(iii) Asian option:

$$V_T = \Big(\frac{1}{T}\int_0^T S_t dt - K\Big)^+$$

The risk-neutral price $V(t)$ in all of these cases can be expressed as

$$V(t) = \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)}V_T|\mathcal{F}(t)\Big].$$

To analyze V_t further, it is tempting to write $V(t) = v(t, S(t))$ for some function $v(t, x)$ and start deriving what equation $v(t, x)$ has to satisfy. However, this is incorrect.

Recall that the basis for us to say there exists such a function $v(t, x)$ is because of the Indepndence lemma, which in turns rely on the fact that we can write

$$S_T = S_t \times (\text{ something independent of } \mathcal{F}(t))$$

and we were working with European option, which only depends on S_T.

That is not the case here: all these three exotic options are *path dependent*, i.e. the expression for V_T involves the values of $S_t, 0 \le t \le T$, not just S_T. So apriori, it is not clear that we can find such a $v(t, x)$. Indeed, for the Lookback and Asian option, we will see that the correct function to deal with is $v(t, x, y)$, not $v(t, x)$, where we need to add another component $Y(t)$ to $S(t)$ so that the joint process $S(t), Y(t)$ have the necessary Markov property.

For the Knockout Barrier option, *the key idea is to analyze the behavior of S_t upto the first time it hits the barrier.* This time, as you may have known, is a *stopping time* with respect to the filtration generated by S_t. So we will begin by reviewing stopping time and its properties. We will then show how we can derive the PDE for the price V_t of a Knockout Barrier option using stopping time.

4.2 Stopping times

4.2.1 Motivation

In financial math, very often and quite naturally, we study random decisions, such as when to exercise your right to buy an option (American call option), or when to accept an offer for the house you are selling (imagine you're putting your house on a market and offer comes in for how much the buyer is willing to pay for the house, which is random). These decisions involve a random time (the time you decide to take action). The time is random because obviously it depends on the path of the stock's price, or of the offers, which are random.

However, there is a common important feature in both cases here: *your decision of when to take action cannot depend on future information*. Mathematically, if we denote $\mathcal{F}(t)$ as the stream of information available to you at time t, and the random time when you take action is τ, then we require:

$$\{\tau \leq t\} \in \mathcal{F}(t).$$

The event $\{\tau \leq t\}$ means you have taken action on or before time t. The event being $\in \mathcal{F}(t)$ then means your decision of taking action on or before time t entirely depends on the information up to time t, i.e. it does not involve future information. Such τ is called a stopping time and it is an important concept to study.

4.2.2 Some preliminary

Discrete vs continuous time

We can model time in 2 ways. Discrete: consider time $n = 0, 1, 2, ..., N$ where N is our terminal time. Continuous: consider time $t \in [0, T]$, where T is our terminal time. Stopping times are defined in both contexts. Generally speaking, discrete time is "easier" to analyze (don't take this statement too literally). The models we will study in Chapter 7,8 are in continuous time. Generally, most of the statements about stopping times have similar versions in both discrete and continuous times. But when one works in continuous time, it is good to pay attention because there will be subtleties that are not present in discrete time.

Filtration, sigma-algebra and the flow of information

We denote $\mathcal{F}(t), t \in [0,T]$ to be the filtration in the time interval $[0,T]$, which represents the information we have available up to time t. We require:

(i) Each $\mathcal{F}(t)$ is a sigma-algebra.

(ii) If $s < t$ then $\mathcal{F}(s) \subseteq \mathcal{F}(t)$.

Condition (i) is about the closure property of $\mathcal{F}(t)$: if $A_i, i = 1, 2, ...$ is a countable sequence of events (meaning the number of events can potentially be infinite) in $\mathcal{F}(t)$, then A_i^c (not A_i), $\cup_{i=1}^{\infty} A_i$ (some of A_i has happened), $\cap_{i=1}^{\infty} A_i$ (all of A_i have happened) are also in $\mathcal{F}(t)$. We also require $\Omega, \Omega \in \mathcal{F}(t)$.

Condition (ii) is about the flow of information, intuitively at the present time t we must also have knowledge of the information of the past up to time s as well.

Sometimes we have $\mathcal{F}(0) = \{\Omega, \Omega\}$. This means any event at time 0 is deterministic. In terms of a random process, this means the process starts out at a deterministic point x, instead of having a random initial distribution.

We can also consider $\mathcal{F}(n), n = 0, 1, ..., N$ as the discrete analog of continuous time filtration. The requirements are the same.

Stopping time definition

Definition 4.2.1. *Let τ be a random variable taking values in $[0,T]$ (resp. $\{0, 1, ..., N\}$). We say τ is a stopping time with respect to $\mathcal{F}(t)$ (resp. $\mathcal{F}(n)$) if for all $t \in [0,T]$ (resp. for all $n = 0, 1, ..., N$)*

$$\begin{aligned} \{\tau \le t\} &\in \mathcal{F}(t) \\ (\textit{ resp. } \{\tau \le n\} &\in \mathcal{F}(n)). \end{aligned}$$

Remark 4.2.2. *Note that the notion of a stopping time is tied to a filtration (similar to the notion of a martingale). It could happen that τ is a stopping time with respect to a filtration $\mathcal{F}(t)$ but not a stopping time with respect to another, smaller filtration $\mathcal{G}(t) \subseteq \mathcal{F}(t)$.*

First important difference between discrete and continuous time

Consider the discrete time. Since if τ is a $\mathcal{F}(n)$ stopping time then $\{\tau < n\} = \{\tau \le n-1\} \in \mathcal{F}(n-1) \subseteq \mathcal{F}(n)$, we have

$$\{\tau \ge n\} = \{\tau < n\}^c \in \mathcal{F}(n)$$

Hence

$$\{\tau = n\} = \{\tau \le n\} \cap \{\tau \ge n\} \in \mathcal{F}(n).$$

Conversely if $\{\tau = n\} \in \mathcal{F}(n)$ *for all* n then $\{\tau \leq n\} = \cup_{i=0}^{n}\{\tau = i\} \in \mathcal{F}(n)$, for all n as well. So we can use either conditions: $\{\tau = n\} \in \mathcal{F}(n)$ or $\{\tau \leq n\} \in \mathcal{F}(n)$ as definition for stopping time in discrete time.

Now consider the continuous time. By the property of stopping time listed below, it is also true that $\{\tau < t\} \in \mathcal{F}(t)$. So $\{\tau \geq t\} = \{\tau < t\}^c \in \mathcal{F}(t)$. Therefore, if τ is a stopping time then

$$\{\tau = t\} = \{\tau \leq t\} \cap \{\tau \geq t\} \in \mathcal{F}(t).$$

However, *it is NOT true that if* $\{\tau = t\} \in \mathcal{F}(t)$ *for all* t *then* $\{\tau \leq t\} \in \mathcal{F}(t)$. The reason is because in continuous time, we need to write

$$\{\tau \leq t\} = \cup_{0 \leq s \leq t}\{\tau = s\},$$

and the RHS involves an *uncountable* union of events, which doesn't have to be contained in the sigma algebra. This explains the choice of using $\{\tau \leq t\} \in \mathcal{F}(t)$ as the definition for continuous time.

Some properties of stopping time

Lemma 4.2.3. *Let* τ_1, τ_2 *be stopping times with respect to* $\mathcal{F}(t)$. *Then*
(i) $\{\tau_1 < t\} \in \mathcal{F}(t), \forall 0 \leq t \leq T$;
(ii) $\min(\tau_1, \tau_2)$ *and* $\max(\tau_1, \tau_2)$ *are stopping times with respect to* $\mathcal{F}(t)$.

Property (i) follows from the fact that

$$\{\tau_1 < t\} = \cup_{n=1}^{\infty}\{\tau_1 \leq t - \frac{1}{n}\},$$

and $\{\tau_1 \leq t - \frac{1}{n}\} \in \mathcal{F}(t - \frac{1}{n}) \subseteq \mathcal{F}(t), \forall n$. Property ii is left as homework exercise.

4.2.3 Some important examples

Example 4.2.4. *Jump time of a Poisson process*
Let $N(t)$ *be a Poisson process. Then*

$$\tau_k := \inf\{t \geq 0 : N(t) = k\}$$

are stopping times with respect to $\mathcal{F}^N(t)$.

Reason: $\{\tau_k \leq t\}$ means the k^{th} jump happened at or before t. But that is the same as at time t, $N(t) \geq k$. Thus

$$\{\tau_k \leq t\} = \{N(t) \geq k\} \in \mathcal{F}(t).$$

Example 4.2.5. *First hitting time to a point of Brownian motion*

Let $b > 0$ be fixed. Define

$$T_b := \inf\{t \geq 0 : W(t) = b\}$$

to be the first time $W(t)$ hits the level b. Note also the convention that $\inf \Omega = \infty$, that is if $W(t)$ never hits b then we set $T_b = \infty$. Then T_b is a stopping time with respect to $\mathcal{F}^W(t)$.

The reasoning here is more complicated. Note that $\{T_b \leq t\}$ means $W(.)$ has hit b at or before time t. But we cannot infer any property of $W(t)$ (say $W(t) \geq b$ based on this information) because W is *not monotone*.

It is better to look at the complement: $\{T_b > t\}$ which means $W(.)$ has NOT hit b at or before t, which since $W(.)$ starts at 0 at time 0 is equivalent to $W(s) < b, 0 < s < t$, the information of which intuitively belongs to $\mathcal{F}(t)$. But this is not rigorous, since again there are uncountably many points s in $[0, t]$.

To fix this, we note that a continuous function is uniquely determined by its values on the rationals, which is countable. Combine these facts we can write

$$\begin{aligned} \{T_b > t\} &= \{W(s) < b, 0 \leq s \leq t\} = \cup_{i=1}^{n} \{W(s) \leq b - \frac{1}{n}, 0 \leq s \leq t\} \\ &= \cup_{i=1}^{n} \{W(s) \leq b - \frac{1}{n}, s \in [0, t] \cap \mathbb{Q}\} \\ &= \cup_{i=1}^{n} \cap_{s \in \mathbb{Q}} \{W(s) \leq b - \frac{1}{n}\}, \end{aligned}$$

and it follows that $\{T_b > t\} \in \mathcal{F}(t)$. Note the subtle fact here that we need to transition from $W(s) < b$ to $W(s) \leq b - \frac{1}{n}$ for some n. The reason is this: if $W(s) < b$ for all s rationals, we can only conclude that $W(s) \leq b$ for all s. But $W(s) \leq b$ for all s rational if and only if $W(s) \leq b$ for all s.

We did not use any special property of Brownian motion besides the fact that it has continuous paths. So

Example 4.2.6. *First hitting time to a point of a continuous process*

Let $b > 0$ be fixed. Let $X(t)$ be a process starting at 0 with continuous paths. Define

$$T_b := \inf\{t \geq 0 : X(t) = b\}$$

to be the first time $X(t)$ hits the level b. Then T_b is a stopping time with respect to $\mathcal{F}^X(t)$.

Example 4.2.7. *Non example: last hitting time*

Let $b > 0$ be fixed. Let $X(t)$ be a process starting at 0 with continuous paths. Define

$$T_b := \sup\{t \geq 0 : X(t) = b\}$$

to be the last time $X(t)$ hits the level b. Then T_b may NOT be a stopping time with respect to $\mathcal{F}^X(t)$.

The reason is this: $\{T_b \leq t\}$ means the last time $X(t)$ hits b is at or before time t. But it is impossible to know whether $X(t)$ will hit b again unless we observe the future paths of $X(t)$, which is forbidden for a stopping time definition. There is an exception: if we know that $X(t)$ is monotone, then once it hits b it will not hit b again. But this is probably the only exception.

Example 4.2.8. *First hitting time to an open set of a continuous process*
Let $b > 0$ be fixed. Let $X(t)$ be a process starting at 0 with continuous paths. Define

$$S_b := \inf\{t \geq 0 : X(t) > b\}$$

to be the first time $X(t)$ hits the open set (b, ∞). Then S_b may NOT be a stopping time with respect to $\mathcal{F}^X(t)$.

The reason is very subtle here. It is tempting to write

$$\begin{aligned}\{S_b > t\} &= \{X_s \leq b, 0 \leq s \leq t\} = \{X_s \leq b, s \in \mathbb{Q}\} \\ &= \cap_{s\in\mathbb{Q}}\{X_s \leq b\},\end{aligned}$$

therefore $\{S_b > t\} \in \mathcal{F}(t)$ and S_b is a stopping time. What happens is the first equality is incorrect, and it is because of the definition of infimum. It could be the case that at time t, $X(t) = b$ and immediately after t, X crosses over b. Then in this case $S_b = t$ and the event we describe is still in the RHS of the above equation. In other words,

$$\{S_b \geq t\} \;=\; \{X_s \leq b, 0 \leq s \leq t\}$$

and we don't have the right inequality to work with here. But note the fact that S is almost a stopping time. We call it an optional time here.

Remark 4.2.9. *Another useful way to think of the above situation is to imagine 2 possible paths of $X(s)$: one path ω hits b at time t and crosses over. The other ω'* **follows the exact same path up to time** *t, hits b at time t and immediately reflects down, and let's say never comes back to level b. Then $S_b(\omega) = t$ and $S_b(\omega') = \infty$. Since the two paths are* **the same up to time** *t, it is impossible to tell the event $S_b = t$ by observing $\mathcal{F}(t)$. So S_b cannot be a stopping time. This can be used as a useful, even though non-rigorous criterion to determine whether a random time is a stopping time.*

4.2.4 Strong Markov property of Brownian motion

It is a well-known fact of Brownian motion that it has independent and stationary increment: if $t > s$ then $W(t) - W(s)$ is independent of $\mathcal{F}(s)$ and has distribution $N(0, t-s)$. In particular, this implies that $W(t) - W(s)$ is a Brownian motion independent of $\mathcal{F}(s)$.

What is interesting is if we replace s by a stopping time, all of these results still hold, except for the technical issue of defining what $\mathcal{F}(\tau)$ means. For our purpose, it is enough to think of $\mathcal{F}(\tau)$ as the sigma algebra containing all information before time τ and we have the following:

Theorem 4.2.10. *Strong Markov property*

Let W be a Brownian motion and $\mathcal{F}(t)$ a filtration for W. Let τ be a $\mathcal{F}(t)$ stopping time. Then $W(\tau + u) - W(\tau), u \geq 0$ is a Brownian motion independent of all the information in the filtration $\mathcal{F}(t)$ before time τ.

This theorem is called the Strong Markov property because it implies that the Markov property of Brownian motion can be applied to a stopping time as well. Indeed, if we accept, in addition to the strong Markov property, the fact that $W(\tau) \in \mathcal{F}(\tau)$ then by the Independence Lemma:

$$E[f\big(W(\tau + u)\big)|\mathcal{F}(\tau)] = g(W(\tau)),$$

where

$$g(x) = E[f\big(x + W_u\big)].$$

4.2.5 An important result in the case of Brownian motion

Let W_t be a Brownian motion starting at 0. Let $b > 0$ and define

$$\begin{aligned} T_b &:= \inf\{t \geq 0 : W(t) = b\} \\ S_b &:= \inf\{t \geq 0 : W(t) > b\}. \end{aligned}$$

Then it is clear that $T_b \leq S_b$. We have also remarked above that T_b is a stopping time with respect to the Brownian filtration while S_b is only an optional time. A very interesting result here is that even though these times are different in nature, the probability of the event that they differ is 0. That is

Lemma 4.2.11. *Let $W(t)$ be a Brownian motion, then*

$$\mathbb{P}(T_b = S_b) = 1.$$

Remark 4.2.12. *The above Lemma says that S_b is equal to T_b up to sets of measure 0. Therefore, if we include sets of measure 0 in $\mathcal{F}(t)$, for all t, a procedure called augmentation of filtration, then S_b is a stopping time with respect to the augmented filtration.*

Proof. We present the idea of the proof of this result here. For complete details, see e.g. [**?**] problem 7.19.

The proof of the Lemma (4.2.11) depends on two other important results about the path of Brownian and the Brownian filtration. They are as followed:

a. Blumenthal 0-1 law : Let $\mathcal{F}_t$ be a filtration generated by a Brownian motion. Let E be an event in the sigma algebra $\mathcal{F}_{0+}$. Then $P(E) = 0$ or $P(E) = 1$.

Remark: $E \in \mathcal{F}_{0+}$ means that we can have the information of E by observing infinitesimally into the future beyond the time 0, but not necessarily exactly at time 0.

b. Infinite crossing property: Let W_t be a Brownian motion starting at 0. Then for **any** $\varepsilon > 0$,

$$P(W_t \text{ crosses 0 infinitely often in the time interval } [0,\varepsilon]) = 1.$$

We will take the 0-1 law as a fact. The infinite crossing property can be explained using the 0-1 law as followed. Define

$$\begin{aligned} T_0^+ &:= \inf\{t \geq 0 : W_t > 0\} \\ T_0^- &:= \inf\{t \geq 0 : W_t < 0\}. \end{aligned}$$

Then arguing as we did before, we can show the events $\{T_0^+ = 0\}$ and $\{T_0^- = 0\}$ are in $\mathcal{F}_{0+}$. Then by Blumenthal 0-1 law, $P(T_0^+ = 0) = 0$ or $P(T_0^+ = 0) = 1$, similarly for T_0^-. By symmetry of the distribution of Brownian motion ($-W_t$ is a Brownian motion iff W_t is a Brownian motion) we also have

$$P(T_0^+ = 0) = P(T_0^- = 0).$$

Therefore, they must both be 0 or both be 1. Now suppose that both $P(T_0^+ = 0) = P(T_0^- = 0) = 0$. That must mean with positive probability we can find an $\varepsilon > 0$ so that $W_t = 0$ identically on $[0,\varepsilon]$. But this is impossible since this implies that with positive probability, the quadratic variation of W_t on $[0,\varepsilon]$ is equal to 0. Thus we must conclude

$$P(T_0^+ = 0) = P(T_0^- = 0) = 1.$$

That is with probability 1, W_t crosses 0 infinitely often in the time interval $[0,\varepsilon]$.

We will now show $\mathbb{P}(T_b = S_b) = 1$ using these two facts and the strong Markov property of W_t. Since T_b is a stopping time, $\widetilde{W}_t^b := W_t - W_{T_b}$ is a Brownian motion for $t \geq T_b$.

Suppose that in contrary to the conclusion of the Lemma, $P(S_b > T_b) > 0$. Then with positive probability, there is a time interval (namely on $[T_b, S_b]$) so that $\widetilde{W}_t^b \leq 0$ on $[T_b, S_b]$. That is $\widetilde{W}^b$ does not cross 0 on the time interval $[T_b, S_b]$. But this contradicts fact b) we mentioned above. This establishes the Lemma.

4.2.6 Stopped processes

Definition 4.2.13. *Given a stochastic X and a random time τ, we define the stopped process X at time τ as*

$$\begin{aligned} X(t\wedge\tau(\omega))(\omega) &:= X(t)(\omega), t\le\tau(\omega)\\ &:= X_{\tau(\omega)}(\omega), t\ge T(\omega). \end{aligned}$$

When τ is a stopping time and X is a martingale then the stopped process is also a martingale via the following theorem:

Theorem 4.2.14. *Let $M(t)$ be a martingale with respect to $\mathcal{F}(t)$ with càdlàgpaths. Let τ be a stopping time with respect to $\mathcal{F}(t)$. Then $M(t\wedge\tau)$ is also a martingale with respect to $\mathcal{F}(t)$.*

This theorem has a discrete time analog:

Theorem 4.2.15. *Let $M(n)$ be a martingale with respect to $\mathcal{F}(n)$ and τ a $\mathcal{F}(n)$ stopping time. Then $X(t\wedge n)$ is also a martingale with respect to $\mathcal{F}(n)$.*

In particular, in the continuous time, when M is a stochastic integral against Brownian motion, then the stopped processed $M(t\wedge\tau)$ is also a martingale when τ is a stopping time. But in this case, we also have an interesting representation of the stopped stochastic integral via the following theorem.

Theorem 4.2.16. *Let $\mathcal{F}(t)$ be a filtration and $W(t)$ a $\mathcal{F}(t)$ Brownian motion. Let α be an adapted process to $\mathcal{F}(t)$ such that $\int_0^t \alpha(s)dW(s)$ is well-defined. Let τ be a $\mathcal{F}(t)$ stopping time. Denote $M(t) := \int_0^t \alpha(s)dW(s)$. Then $M(t\wedge\tau)$ is a $\mathcal{F}(t)$ martingale. Moreover,*

$$M(t\wedge\tau) = \int_0^{t\wedge\tau} \alpha(s)dW(s) = \int_0^t \mathbf{1}_{[0,\tau)}(s)dW(s).$$

The following corollary is an immediate consequence of the above theorem:

Corollary 4.2.17. *Let S_t have the dynamics:*

$$dS_t = \alpha_t dt + \sigma_t dW_t.$$

Then for a stopping time τ

$$\begin{aligned} S_{t\wedge\tau} &= S_0 + \int_0^{t\wedge\tau} \alpha(s)ds + \int_0^{t\wedge\tau} \sigma_s dW(s)\\ &= S_0 + \int_0^t \mathbf{1}_{[0,\tau)}(s)\alpha_s ds + \int_0^t \mathbf{1}_{[0,\tau)}(s)\sigma_s dW(s). \end{aligned}$$

4.2.7 Generalization of Lemma (4.2.11) to Ito processes

Lemma (4.2.11) can be generalized to general Ito process: process that can be written as a Rieman integral plus an Ito integral. The intuition here is that the Ito integral has path property similar to that of Brownian motion: very irregular. On the other hand, the Rieman integral has a differentiable ("regular") path. So when the process $X(t)$ hits b, the effect of the stochastic integral part would win out and cause the process to enter b as in the presence of only a Brownian motion.

Theorem 4.2.18. *Let*

$$X(t) = X(0) + \int_0^t \alpha(s)ds + \int_0^t \sigma(s)dW(s),$$

and suppose that $\mathbb{P}(\sigma(t) \neq 0) = 1$ *for all* t.
Define

$$\begin{aligned} T_b &:= \inf\{t : X(t) = b\} \\ S_b &:= \inf\{t : X(t) > b\}. \end{aligned}$$

Then $\mathbb{P}(T_b = S_b) = 1$.

4.3 Knock-out Barrier option

4.3.1 The goal

Let $S(t)$ satisfies

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

Consider the Knock-out Barrier option with barrier b and strike price K:

$$V_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \leq b\}}.$$

Note: Necessarily we require $K < b$ and $S(0) < b$ so that $\mathbb{P}(V_T > 0) > 0$.
The risk neutral price $V(t)$ can be written as:

$$V(t) = E\Big[e^{-r(T-t)}(S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \leq b\}} | \mathcal{F}(t)\Big].$$

Our goal is to find a function $v(t,x)$ so that $V_t = v(t, S_t)$ and then apply Ito's formula to $e^{-rt}v(t, S_t)$ to find a PDE that $v(t,x)$ satisfies. It is not immediately clear that this can be achieved, as the expression of V_t above involves the term $\mathbf{1}_{\{\max_{[0,T]} S_t \leq b\}}$. Indeed as we shall see there is no such function $v(t,x)$ so that the equality

$$V_t = v(t, S_t)$$

holds true *for all* $0 \le t \le T$.

However, we observe that *before* S_t hits b, it is believable that the option value V_t should be just a function of (t, S_t) (think about the factors that you would use to value V_t in a real life situation before the stock hits the barrier). *After* S_t hits b, the option value V_t stays constant, namely it takes value 0. That is, the option value V_t should be a function of (t, S_t) *upto* the random time T_b, the first time S_t hits b. In other words, we are looking to find a function $v(t, x)$ so that

$$V_t = v(t \wedge T_b, S_{t\wedge T_b}), t \in [0, T].$$

This is what we will establish rigorously in several steps in the following section. This equality will also help us establish a PDE for $v(t, x)$ since by the result of the section (4.2.6), we can apply Ito's formula to $v(t \wedge T_b, S_{t\wedge T_b})$. Note, however, that here we are investigating the dynamics of $v(t, S_t)$ on the time interval $[0, T_b]$. Thus our PDE will not have the usual domain as the one in classical Black-Scholes PDE.

4.3.2 The steps

We proceed to establish

$$V_t = v(t \wedge T_b, S_{t\wedge T_b}),$$

for some function $v(t, x)$ through several steps.

(i) Write $\mathbf{1}_{\{\max_{[0,T]} S_t \le b\}}$ in terms of $S_u, 0 \le u \le t$ and $S_u, t \le u \le T$.

The reason is we want to apply the Independence Lemma (or quote the Markov property of $S(t)$), so heuristically we want to "separate the past and the future". We already know how to do this with S_T. So we apply the same principle to the new term $\mathbf{1}_{\{\max_{[0,T]} S_t \le b\}}$.

This is accomplished as followed:

$$\mathbf{1}_{\{\max_{[0,T]} S_t \le b\}} = \mathbf{1}_{\{\max_{[0,t]} S_u \le b\}} \mathbf{1}_{\{\max_{[t,T]} S_u \le b\}}.$$

It is easy to see why the equality is true: the maximum of the whole path does not exceed b if and only if its maximum on each time interval does not exceed b.

(ii) Recognizing that $\mathbf{1}_{\{\max_{[0,t]} S_u \le b\}} \in \mathcal{F}(t)$, so it can be factored out of $E(.|\mathcal{F}(t))$.

(iii) Define

$$\begin{aligned} \tau_b &:= \inf\{t \ge 0 : S(t) > b\} \wedge T \\ T_b &:= \inf\{t \ge 0 : S(t) = b\} \wedge T \end{aligned}$$

Recall that $P(T_b = \tau_b) = 1$. And so with probability 1:

$$\{\max_{[0,t]} S_u \le b\} = \{\tau_b \ge t\} = \{T_b \ge t\}.$$

The change from τ_b to T_b might seem unimportant and non-intuitive. But it is to apply the optinal stopping theorem for martingale, see the section on the derivation of the PDE below.

(iv) Combine (ii) and (iii) we get

$$V(t) = \mathbf{1}_{T_b \geq t} E\Big[e^{-r(T-t)}(S_T - K)^+ \mathbf{1}_{\{\max_{[t,T]} S_u \leq b\}} | \mathcal{F}(t)\Big].$$

(v) Since

$$S(T) = S(t) e^{(r - \frac{1}{2}\sigma^2)(T-t) + \sigma(W(T) - W(t))}$$

and

$$\max_{[t,T]} S_u = S_t \max_{[t,T]} e^{(r-\frac{1}{2}\sigma^2)(u-t)+\sigma(W(u)-W(t))},$$

note that $\max_{[t,T]} e^{(r-\frac{1}{2}\sigma^2)(u-t)+\sigma(W(u)-W(t))}$ is independent of $\mathcal{F}(t)$, by the Independence Lemma, we get

$$E\Big[e^{-r(T-t)}(S_T - K)^+ \mathbf{1}_{\{\max_{[t,T]} S_u \leq b\}} | \mathcal{F}(t)\Big] = v(t, S(t)).$$

where

$$\begin{aligned} v(t,x) \quad := \quad & E\Big[e^{-r(T-t)}\big(xe^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(W(T)-W(t))} - K\big)^+ \\ & \times \mathbf{1}_{\{x \max_{[t,T]} e^{(r-\frac{1}{2}\sigma^2)(u-t)+\sigma(W(u)-W(t))} \leq b\}}\Big]. \end{aligned}$$

(vi) (*Crucial point*)

$$V(t) = \mathbf{1}_{T_b \geq t} v(t, S(t)) = v(t, S(t \wedge T_b)).$$

Indeed if $T_b \geq t$ then LHS $= v(t, S(t))$ and $t \wedge T_b = t$ so the RHS $= v(t, S(t))$ and the equality is true.

If $T_b < t$ then LHS = 0. $t \wedge T_b = T_b$ so that $S(t \wedge T_b) = b$. Moreover, with probability 1:

$$b \max_{[t,T]} e^{r(u-t)+\sigma(W(u)-W(t))} > b$$

Indeed, if we denote $X(u) := r(u-t) + \sigma(W(u) - W(t)), u \in [t,T]$ then $X(t) = 0$ and by property of Brownian motion,

$$P(X(u) \leq 0, \forall u \in [t,T]) = 0.$$

So there must exist $u \in (t,T]$, $X(u) > 0$ and at that point u, $be^{X(u)} > b$. Thus

$$\begin{aligned} v(t, S(t\wedge T_b)) = v(t,b) &= \Big[e^{-r(T-t)}\big(be^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(W(T)-W(t))} - K\big)^+ \\ &\quad \mathbf{1}_{\{b\max_{[t,T]} e^{(r-\frac{1}{2}\sigma^2)(u-t)+\sigma(W(u)-W(t))} \le b\}}\Big] = 0, \end{aligned}$$

and so the RHS = 0 as well.

(vii) From the above, we see that the function $v(t,x)$ satisfies $v(t,b) = 0$ for all t. Therefore, it follows that

$$v(\tau, b) = 0,$$

for all stopping time τ taking values in $[0,T]$. From which we derive that

$$v(t\wedge T_b, S_{t\wedge T_b}) = v(t, S_{t\wedge T_b}).$$

Indeed, for $t < T_b$ the equalities are clear. For $t > T_b$, then $v(T_b, S_{T_b}) = v(T_b, b) = 0 = v(t,b)$ so the equalities are also true in this case.

Therefore,

$$V_t = v(t, S(t\wedge T_b)) = v(t\wedge T_b, S(t\wedge T_b)).$$

4.3.3 Derivation of the PDE

Derivation

We have

$$\begin{aligned} S(t\wedge T_b) &= S(0) + \int_0^{t\wedge T_b} rS(u)du + \int_0^{t\wedge T_b} \sigma S(u)dW(u) \\ &= S(0) + \int_0^t \mathbf{1}_{[0,T_b)} rS(u)du + \int_0^t \mathbf{1}_{[0,T_b)}\sigma S(u)dW(u). \end{aligned}$$

Apply Ito's formula to $e^{-rt}v(t, S_{t\wedge T_b})$ (where we look at $v(t, S_{t\wedge T_b})$ as a deterministic function of t and the stopped process $S_{t\wedge T_b}$), we have

$$\begin{aligned} e^{-rt}V_t &= e^{-rt}v(t, S_{t\wedge T_b}) = v(0,S_0) + \int_0^t e^{-ru}\big[-rv + v_t + \mathbf{1}_{[0,T_b)}(u) rS_u v_x \\ &\quad + \frac{1}{2}\mathbf{1}_{[0,T_b)}(u)\sigma^2 S_u^2 v_{xx}\big]du + \int_0^t \mathbf{1}_{[0,T_b)}(u)e^{-ru}\sigma S(u)v_x dW_u, \end{aligned}$$

where for all functions v we understood as $v(t,S_t)$, similarly for v_t, v_x, v_{xx}.

Note that this is where the importance of using T_b instead of τ_b is. The stochastic integral

$$\int_0^t e^{-ru}\mathbf{1}_{[0,T_b)}\sigma S(u)v_x dW_u = \int_0^{t\wedge T_b} e^{-ru}\sigma S(u)v_x dW_u$$

is a martingale since T_b is a stopping time. If we use τ_b here we cannot make the same conclusion for τ_b is not a stopping time.

We do not want to set the dt term equal to 0 yet, because in the dt integral above, some terms include $\mathbf{1}_{[0,T_b)}$ (from the stopped process S_t) and some don't.

But this is easy to fix, since $e^{-rt}V_t$ being a martingale implies $e^{-r(t\wedge T_b)}V_{t\wedge T_b}$ is also a martingale. And it's easily seen that

$$\begin{aligned} e^{-r(t\wedge T_b)}V_{t\wedge T_b} &= v(0,S_0) + \int_0^t \mathbf{1}_{[0,T_b)}(u)e^{-ru}\big[-rv + v_t + rS_u v_x \\ &\quad + \frac{1}{2}\sigma^2 S_u^2 v_{xx}\big]du + \int_0^t \mathbf{1}_{[0,T_b)}e^{-ru}\sigma S(u)v_x dW_u, \end{aligned}$$

Therefore we conclude

$$\mathbf{1}_{[0,T_b)}(u)\big[-rv(u,S_u) + v_t(u,S_u) + rS_u v_x(u,S_u) + \frac{1}{2}\sigma^2 S_u^2 v_{xx}(u,S_u)\big] = 0.$$

Domain of the PDE

The equality

$$\mathbf{1}_{[0,T_b)}(u)\big[-rv(u,S_u) + v_t(u,S_u) + rS_u v_x(u,S_u) + \frac{1}{2}\sigma^2 S_u^2 v_{xx}(u,S_u)\big] = 0$$

does NOT permit us to conclude

$$-rv(t,x) + v_t(t,x) + rS_u v_x(t,x) + \frac{1}{2}\sigma^2 S_u^2 v_{xx}(t,x)\big] = 0,$$

for all t, x.

The reason is we can only cancel out the term $\mathbf{1}_{[0,T_b)}(u)$ when it is NOT zero, which is the same as when $0 < S_u \le b$.

Thus the domain for our PDE is $[0,T] \times [0,b]$, which is *different* from the domain we used to work on for European call option: $[0,T] \times [0,\infty)$. One of the effect is that we will have *boundary conditions* for our PDE at $x = 0$ and $x = b$.

Moreover, note that $v(t,0) = 0$ since if $S(t)$ ever hits 0 it will stay there. $v(t,B) = 0$ was explaind in step (vi). These are the boundary conditions for v. We also have the terminal condition $v(T,x) = (x-K)^+$ as usual.

Thus, the PDE that v must satisfy is:

$$\begin{aligned} v_t - rv + rxv_x + \frac{1}{2}x^2\sigma^2 v_{xx} &= 0, 0 \leq t < T, 0 < x < b \\ v(t,0) = v(t,b) &= 0 \\ v(T,x) &= (x-K)^+. \end{aligned}$$

CHAPTER 5 Pricing the Knock out Barrier option and Look back option via Expectation

5.1 Overview

In this lecture, we'll see how we can evaluate the expression

$$V_t = E^Q(e^{-r(T-t)}V_T \Big| \mathcal{F}_t),$$

where

$$V_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \le B\}}$$

for knock-out barrier option, or

$$V_T = \max_{[0,T]} S_t - S_T,$$

for look-back option.

It is clear that to compute V_t in these expressions, we need to know the distribution of $\max_{[0,T]} S_t$. But since

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t},$$

the distribution of $\max_{[0,T]} S_t$ is closely related to the distribution of M_t, the running max of the Brownian motion:

$$M_t := \max_{u\in[0,t]} W_u.$$

Instead of computing the distribution of M_t by it self, we will see that it is easier to compute the joint distribution of M_t, W_t. The key for us to derive this joint distribution is via the reflection principle, which says a reflected Brownian motion is also a Brownian motion. Using this principle and a probability identity, we will derive the joint distribution of M_t, W_t. From that, we can derive the distribution of M_t as a marginal distribution. Finally, we'll see how we can apply this knowledge to evaluate V_t in the two expressions above.

5.2 The reflection principle

5.2.1 Definition

Let W_t be a Brownian motion w.r.t a filtration $\mathcal{F}(t)$ and τ a $\mathcal{F}(t)$ stopping time. We define

$$\begin{aligned} B^\tau &:= W_t, t \le \tau \\ &:= W(\tau) - [W_t - W(\tau)], t > \tau. \end{aligned}$$

That is B^τ is the same as W_t up to the random time τ and after time τ is obtained by reflecting W_t around the horizontal line $y = W(\tau)$. We say B^τ is a reflected Brownian motion at τ.

5.2.2 The reflection principle

Theorem 5.2.1. *The B^τ defined above is a $\mathcal{F}(t)$ Brownian motion.*

In words, the reflection principel says a refleted Brownian motion is a Brownian motion.

The heuristics of why the Theorem is true is

(i) The strong Markov property: $W_t - W(\tau)$ is a Brownian motion independent of $\mathcal{F}(\tau)$ and

(ii) The negative of a Brownian motion is also a Brownian motion. Thus before t, B^τ is a Brownian motion, after τ it is also a Brownian motion (although starting at $W(\tau)$ instead of at 0). The key is how to show when we go across τ the Brownian motion property is still preserved and we achieve that by Levy's characterization of Brownian motion.

Proof. Define

$$\begin{aligned} a(t) &= 1, t \le \tau \\ &= -1, t > \tau. \end{aligned}$$

That is

$$\begin{aligned} a(t) &= \mathbf{1}_{t \le \tau} - \mathbf{1}_{t > \tau} \\ &= \mathbf{1}_{t \le \tau} - (1 - \mathbf{1}_{t \le \tau}) \\ &= 2\mathbf{1}_{t \le \tau} - 1. \end{aligned}$$

It is easy then to see $a(t) \in \mathcal{F}(t), \forall t$ since τ is a stopping time. It is also bounded, hence is in L^2. Thus we can consider $\int_0^t a(s) dW_s$. We have

$$\begin{aligned} \int_0^t a(s) dW_s &= \int_0^t 2\mathbf{1}_{s \le \tau} dW_s - W_t \\ &= \int_0^t 2\mathbf{1}_{[0,\tau)}(s) dW_s - W_t \\ &= 2W(t \wedge \tau) - W_t = B^\tau(t). \end{aligned}$$

(Just consider what happens when $\tau \leq t$ and $\tau > t$.)

Thus $B^\tau(t)$ is a martingale. Moreover, its quadratic variation is:

$$\langle B^\tau \rangle_t = \int_0^t \alpha^2(s) ds = t,$$

since $\alpha(s)$ is either 1 or -1. Thus by Levy's characterization, B^τ is a Brownian motion.

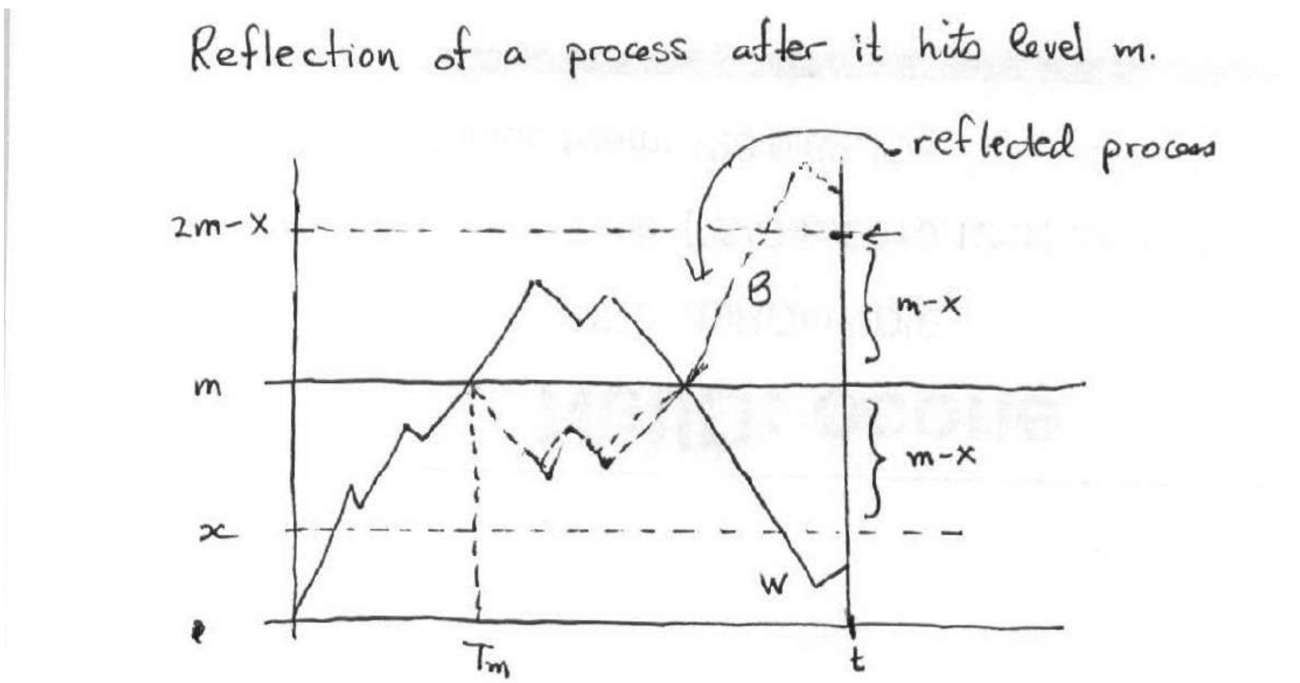
5.2.3 An important identity

Let W_t be a Brownian motion and $M_t := \max_{[0,t]} W_s$ its running maximum. The reflection principle helps us obtain the joint density between W_t and M_t through the following important identity: for $w \leq m, m \geq 0$

$$\Big\{ M_t > m, W_t < w \Big\} = \Big\{ B_t > 2m - w \Big\},$$

where $B_t := B^{\tau_m}(t)$ is the Brownian motion obtained by reflecting W_t at time τ_m, the first hitting time of W_t to level m:

$$\tau_m := \inf\{t \geq 0 : W_t = m\}.$$



Remark 5.2.2. *Our goal with the identity is to use it to derive the joint density $f_t(m, w)$ of M_t, W_t, therefore we are only interested in considering $m \geq w$ and $m \geq 0$ because we always have $M_t \geq W_t$ and $M_t \geq W(0) = 0$.*

Proof. Proof of the identity

(i) Suppose $M_t > m$ and $W_t < w$. Then $M_t > m$ implies $\tau_m < t$ and hence

$$\begin{aligned} B_t &= 2W(\tau_m) - W_t \\ &= 2m - W_t > 2m - w. \end{aligned}$$

(ii) Suppose $B_t > 2m - w$. Then $B_t > m$ because $w \le m$. So it cannot be the case that $B_t = W_t$ since that would imply $W_t > m$ and thus $\tau_m < t$, a contradiction to $B_t = W_t$ only when $t < \tau_m$. Thus $B_t = 2m - W_t$ and $\tau_m < t$ which implies $M_t > m$. Moreover,

$$B_t = 2m - W_t > 2m - w$$

implies $W_t < w$ and we are done.

5.2.4 Joint distribution of W_t and M_t

From the identiy above and the reflection principle (which implies B_t is a Brownian motion) we have

$$P(M_t > m, W_t < w) = P(B_t > 2m - w) = \int_{2m-w}^{\infty} \frac{e^{\frac{-x^2}{2t}}}{\sqrt{2\pi t}} dx.$$

If $f_t(m, w)$ is the joint density of (M_t, W_t) then

$$P(M_t > m, W_t < w) = \int_{-\infty}^{w} \int_{m}^{\infty} f_t(z, x) dz dx = \int_{2m-w}^{\infty} \frac{e^{\frac{-x^2}{2t}}}{\sqrt{2\pi t}} dx.$$

Thus by the Fundamental Theorem of Calculus, we get

$$\begin{aligned} f_t(m, w) &= -\frac{\partial^2}{\partial m \partial w} P(M_t > m, W_t < w) \\ &= \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}} \mathbf{1}_{m \ge 0, w \le m}. \end{aligned}$$

This has the following useful consequence: let $Z_t = 2M_t - W_t$. Then the joint density of (M_t, Z_t) is

$$g_t(m, z) = \frac{2z}{t\sqrt{2\pi t}} e^{-z^2/2t} \mathbf{1}_{\{m>0, z>m\}} = -2 \frac{d}{dz} \frac{e^{-z^2/2t}}{\sqrt{2\pi t}} \mathbf{1}_{\{m>0, z>m\}}. \tag{5.1}$$

5.3 A useful function in evaluation of Barrier and Lookback options

5.3.1 Introduction

When computing the price of Knockout Barrier and Lookback Options, you'll see that because of the structure of the stock price, we'll usually end up computing an expression of the form

$$E\Big[\mathbf{1}_{\{W_s\geq k\}}\mathbf{1}_{\{M_s>b\}}e^{\alpha W_s+\beta M_s}\Big],$$

where α, β, k, b are general parameters that we can plug in depending on the option we're dealing with. Since this expression appears often in this context, we'll denote it by $H_s(\alpha, \beta, k, b)$, as a function of the unspecified parameters at a time s. That is

$$H_s(\alpha,\beta,k,b) := E\Big[\mathbf{1}_{\{W_s\geq k\}}\mathbf{1}_{\{M_s>b\}}e^{\alpha W_s+\beta M_s}\Big].$$

In the following sub-sections, we'll see how we can compute explicitly $H_s(\alpha, \beta, k, b)$ in some special case.

5.3.2 $H_s(\alpha, 0, k, b)$ when $0 \leq b \leq k$

Since $M_s \geq W_s$ we have if $W_s \geq k$ then $M_s \geq W_s \geq k \geq b$.

Thus

$$\Big\{W_s \geq k\Big\} \cap \Big\{M_s \geq b\Big\} = \Big\{W_s \geq k\Big\}.$$

In other words,

$$\mathbf{1}_{\{W_s\geq k\}}\mathbf{1}_{\{M_s>b\}} = \mathbf{1}_{\{W_s\geq k\}}.$$

So

$$H_s(\alpha,0,k,b) = E\Big[\mathbf{1}_{\{W_s\geq k\}}e^{\alpha W_s}\Big] = e^{s\frac{\alpha^2}{2}}N\big(\frac{s\alpha-k}{\sqrt{s}}\big). \tag{5.2}$$

5.3.3 $H_s(\alpha, 0, k, b)$ when $k < b$

Theorem 5.3.1. *If $k < b$,*

$$H_s(\alpha,0,k,b) = e^{s\frac{\alpha^2}{2}}\Big\{N\big(\frac{s\alpha-b}{\sqrt{s}}\big) + e^{2\alpha b}\Big[N\big(\frac{-s\alpha-b}{\sqrt{s}}\big) - N\big(\frac{-s\alpha-2b+k}{\sqrt{s}}\big)\Big]\Big\}.$$

Proof. Since $k < b$,

$$E\Big[\mathbf{1}_{\{W_s \geq k\}}\mathbf{1}_{\{M_s > b\}}e^{\alpha W_s}\Big] = E\Big[\mathbf{1}_{\{W_s \geq b\}}\mathbf{1}_{\{M_s > b\}}e^{\alpha W_s}\Big] + E\Big[\mathbf{1}_{\{k \leq W_s < b\}}\mathbf{1}_{\{M_s > b\}}e^{\alpha W_s}\Big].$$

Now

$$E\Big[\mathbf{1}_{\{W_s \geq b\}}\mathbf{1}_{\{M_s > b\}}e^{\alpha W_s}\Big] = H_s(\alpha, 0, b, b),$$

and we have found the expression for $H_s(\alpha, 0, b, b)$ in Section 2.1. As for the 2nd term, observe that

$$\Big\{k < W_s < b, M_s > b\Big\} = \Big\{W_s < b, M_s > b\Big\} \cap \Big\{k < W_s, M_s > b\Big\}.$$

We have showed that

$$\Big\{W_s < b, M_s > b\Big\} = \Big\{B^{\tau_b}(s) > b\Big\},$$

where B^{τ_b} is again W_t reflected at τ_b, the first hitting time of W_t to level b.

We claim that

$$\Big\{k < W_s, M_s > b\Big\} = \Big\{M_s > b, B^{\tau_b}(s) < 2b - k\Big\}.$$

(This is left as part of the homework).

Thus noting that $B^{\tau_b}(s) > b$ implies $M_s > b$ we get

$$\begin{aligned}\Big\{k < W_s < b, M_s > b\Big\} &= \Big\{B^{\tau_b}(s) > b\Big\} \cap \Big\{M_s > b, B^{\tau_b}(s) < 2b - k\Big\} \\ &= \Big\{b < B^{\tau_b}(s) < 2b - k\Big\}.\end{aligned}$$

We leave it as the other part of the homework to use this and (5.2) to complete the proof.

5.3.4 $H_s(\alpha, \beta, -\infty, b)$

Theorem 5.3.2.

$$\begin{aligned}H_s(\alpha, \beta, -\infty, b) &= \frac{\beta + \alpha}{\beta + 2\alpha} 2e^{\frac{(\alpha+\beta)^2}{2}s} N\big(\frac{(\alpha + \beta)s - b}{\sqrt{s}}\big) \\ &+ \frac{2\alpha}{\beta + 2\alpha} e^{\frac{\alpha^2}{2}s} e^{b(\beta + 2\alpha)} N\big(-\frac{\alpha s + b}{\sqrt{s}}\big).\end{aligned}$$

Proof.

Let $Z_s = 2M_s - W_s$, so that $W_s = 2M_s - Z_s$. We will rewrite the expectation in the definition of $H_s(\alpha, \beta, -\infty, b)$ in terms of $Z(s)$ and M_s and use the joint density for these two random variables, which we stated above in (5.1). Thus,

$$\begin{aligned} H_s(\alpha, \beta, -\infty, b) &= E\left[\mathbf{1}_{\{M_s \geq b\}} e^{\alpha W(s) + \beta M_s}\right] = E\left[\mathbf{1}_{\{M_s \geq b\}} e^{-\alpha Z(s) + (\beta + 2\alpha) M_s}\right] \\ &= \int_b^\infty e^{(\beta+2\alpha)m} \int_m^\infty e^{-\alpha z} \left[-2 \frac{d}{dz} \frac{e^{-z^2/2s}}{\sqrt{2\pi s}}\right] dz\, dm. \end{aligned} \quad (5.3)$$

By integration by parts, and then application of the formula

$$E\left[e^{aX} \mathbf{1}_{\{X \geq c\}}\right] = \int_c^\infty e^{ax} e^{-x^2/(2s)} \frac{dx}{\sqrt{2\pi s}} = e^{(a^2/2)s} N\left(\frac{as - c}{\sqrt{s}}\right), \quad (5.4)$$

where X has Normal$(0, s)$ distribution, the inner integral is

$$2e^{-\alpha m} \frac{e^{-m^2/2s}}{\sqrt{2\pi s}} - 2\alpha \int_m^\infty e^{-\alpha z} \frac{e^{-z^2/2s}}{\sqrt{2\pi s}}\, dz = 2e^{-\alpha m} \frac{e^{-m^2/2s}}{\sqrt{2\pi s}} - 2\alpha e^{\frac{\alpha^2}{2} s} N\left(\frac{-\alpha s - m}{\sqrt{s}}\right)$$

Thus,

$$\begin{aligned} E\left[\mathbf{1}_{\{M_s \geq b\}} e^{\alpha W(s) + \beta M_s}\right] &= 2 \int_b^\infty e^{(\beta+\alpha)m} \frac{e^{-m^2/2s}}{\sqrt{2\pi s}}\, dm \\ &\quad -2\alpha e^{\frac{\alpha^2}{2} s} \int_b^\infty e^{(\beta+2\alpha)m} N\left(\frac{-\alpha s - m}{\sqrt{s}}\right) dm \end{aligned} \quad (5.5)$$

By applying (5.4) again, the first term is $2e^{\frac{(\beta+2\alpha)^2}{2} s} N\left(\frac{(\beta+2\alpha)s - b}{\sqrt{s}}\right)$. By integrating by parts and applying (5.4) yet again, the second term is

$$\frac{2\alpha}{\beta + 2\alpha}\left[e^{((\alpha+\beta)^2/2)s} N\left(\frac{(\alpha+\beta)s - b}{\sqrt{s}}\right) - e^{(\alpha^2/2)s} e^{b(2\alpha+\beta)} N\left(\frac{-\alpha s - b}{\sqrt{s}}\right)\right]$$

By substituting these results in (5.5) one obtains the result.

5.4 Pricing Knock-out Barrier option via expectation

Let S_t satisfies

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

Consider the Knock-out Barrier option with barrier b and strike price K:

$$V_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \leq B\}}.$$

The risk-neutral price V_t can be expressed as

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} V_T \Big| \mathcal{F}(t)\Big] \\ &= \mathbf{1}_{\{\max_{u\in[0,t]} S_u \leq B\}} \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \mathbf{1}_{\{\max_{u\in[t,T]} S_u \leq B\}} \Big| S_t\Big]. \end{aligned}$$

To obtain an explicit formula for V_t, we need to evaluate

$$\begin{aligned} &\mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \mathbf{1}_{\{\max_{u\in[t,T]} S_u \leq B\}} \Big| S_t\Big] \\ &= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \Big| S_t\Big] - \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \mathbf{1}_{\{\max_{u\in[t,T]} S_u > B\}} \Big| S_t\Big] \end{aligned} \tag{5.6}$$

Since $\mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \Big| S_t\Big]$ is already given by Black-Scholes formula, we only need to evaluate

$$\begin{aligned} w(t,x) &:= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K)^+ \mathbf{1}_{\{\max_{u\in[t,T]} S_u > B\}} \Big| S_t = x\Big] \\ &:= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)} (S_T - K) \mathbf{1}_{\{S_T \geq K\}} \mathbf{1}_{\{\max_{u\in[t,T]} S_u > B\}} \Big| S_t = x\Big] \end{aligned} \tag{5.7}$$

Remark 5.4.1. *The split in Equation (5.6) is to allow us to write $w(t,x)$ in the form of $H_s(\alpha, \beta, k, b)$ as we will see later.*

5.4.1 Step 1: A first rewrite of $w(t,x)$

Denote

$$\alpha := \frac{r - \frac{1}{2}\sigma^2}{\sigma}.$$

Then for $s \geq t$

$$S(s) = S_t \exp\Big[\sigma\big\{W_s - W_t + \alpha(s-t)\big\}\Big].$$

The term inside the exponential (modulo the σ) is just a Brownian motion with drift starting at time t. So we denote it by a new name to reflect this fact:

$$\begin{aligned} \widehat{W}(u) &:= W(t+u) - W_t + \alpha u, u \geq 0 \\ \widehat{M}(u) &:= \max_{s\in[0,u]} \widehat{W}_s. \end{aligned}$$

Note that for $s \geq t$

$$\max_{u\in[t,s]} S_u = S_t e^{\sigma \widehat{M}(s-t)}.$$

Then for $s \geq t$ we have

$$\begin{aligned} S(s) &= S_t e^{\sigma \widehat{W}_{s-t}} \\ \mathbf{1}_{S(s)\geq K} &= \mathbf{1}_{\widehat{W}_{s-t}\geq \frac{\log(K/S_t)}{\sigma}} \\ \mathbf{1}_{\max_{u\in[t,s]} S_u > B} &= \mathbf{1}_{\widehat{M}(s-t) > \frac{\log(B/S_t)}{\sigma}} \end{aligned}$$

Then substituting this into Equation (5.7), replacing $S_t = x$ gives

$$\begin{aligned} w(t,x) &= e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\big(xe^{\sigma\widehat{W}_{T-t}} - K\big)\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}\Big] \\ &= xe^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[e^{\sigma\widehat{W}_{T-t}}\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}\Big] \\ &\quad -Ke^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}\Big]. \end{aligned} \tag{5.8}$$

Remark 5.4.2. *Note that the expression in (5.8) involves the distribution of a Brownian motion with drift and its running maximum. Studying Brownian motion with drift is inconvenient. But by applying a change of measure (via Girsanov's theorem), we can find a different measure such that under it, $\widehat{W}$ is a Brownian motion. So that's our next step.*

5.4.2 Apply Girsanov's Theorem to transform $\widehat{W}$ into a Brownian motion

Observe that there exists a Brownian motion $\widetilde{W}(u), 0 \leq u \leq T-t$, namely $\widetilde{W}_u = W_{t+u} - W_t$, such that

$$\widehat{W}(u) = \widetilde{W}(u) + \alpha u, u \in [0, T-t].$$

Since $\widehat{W}$ has drift term αt, our change of measure kernel is

$$Z_{T-t} = \exp[-\alpha\widetilde{W}(T-t) - \frac{\alpha^2}{2}(T-t)].$$

Denoting our original measure as $\mathbb{Q}$ and define

$$d\widehat{\mathbb{P}} := Z_{T-t}d\mathbb{Q},$$

then note that

$$\begin{aligned} d\mathbb{Q} &= Z_{T-t}^{-1}d\widehat{\mathbb{P}} \\ &= \exp[\alpha\widehat{W}_{T-t} - \frac{\alpha^2}{2}(T-t)], \end{aligned}$$

So that

$$
\begin{aligned}
w(t,x) &= xe^{-r(T-t)}\widehat{\mathbb{E}}\Big[e^{\sigma\widehat{W}_{T-t}}\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}e^{\alpha\widehat{W}_{T-t}-\frac{\alpha^2}{2}(T-t)}\Big] \\
&\quad -Ke^{-r(T-t)}\widehat{\mathbb{E}}\Big[\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}e^{\alpha\widehat{W}_{T-t}-\frac{\alpha^2}{2}(T-t)}\Big] \\
&= xe^{-(r+\frac{\alpha^2}{2})(T-t)}\widehat{\mathbb{E}}\Big[e^{(\alpha+\sigma)\widehat{W}_{T-t}}\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}\Big] \\
&\quad -Ke^{-(r+\frac{\alpha^2}{2})(T-t)}\widehat{\mathbb{E}}\Big[e^{\alpha\widehat{W}_{T-t}}\mathbf{1}_{\{\widehat{W}_{T-t}\geq\frac{\log(K/x)}{\sigma}\}}\mathbf{1}_{\{\widehat{M}_{T-t}>\frac{\log(B/x)}{\sigma}\}}\Big],
\end{aligned}
\tag{5.9}
$$

where now *what we have gained is $\widehat{W}$ is a Brownian motion under $\widehat{\mathbb{P}}$.*

5.4.3 Writing $w(t,x)$ in terms of $H_s(\alpha,\beta,k,b)$

Let W_t be a Brownian motion and $M_t := \max_{[0,t]} W_s$ its running maximum. Recall that we defined

$$
H_s(\alpha,\beta,k,b) := E\Big[\mathbf{1}_{\{W_s\geq k\}}\mathbf{1}_{\{M_s>b\}}e^{\alpha W_s+\beta M_s}\Big].
$$

Then we have

$$
\begin{aligned}
w(t,x) &= e^{-(r+\frac{\alpha^2}{2})(T-t)}\Big[xH_{T-t}\Big(\alpha+\sigma,0,\frac{\log(K/x)}{\sigma},\frac{\log(B/x)}{\sigma}\Big) \\
&\quad -KH_{T-t}\Big(\alpha,0,\frac{\log(K/x)}{\sigma},\frac{\log(B/x)}{\sigma}\Big)\Big].
\end{aligned}
$$

and the original Knockout Barrier option price is:

$$
V_t = \mathbf{1}_{\{\max_{[0,t]} S_t\leq B\}}\big[c(t,S_t)-w(t,S_t)\big],
$$

where

$$
c(t,x) = xN(d_+(T-t,x)) - Ke^{-r(T-t)}N(d_-(T-t,x))
$$

is given by Black-Scholes formula.

5.5 Pricing Lookback Option via expectation

5.5.1 Preliminary discussion

Let S_t satisfies

$$
dS_t = rS_t dt + \sigma S_t dW_t.
$$

Consider the Knock-out Barrier option with barrier b and strike price K:

$$V_T = \max_{[0,T]} S_t - S_T.$$

The risk-neutral price V_t can be expressed as

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}V_T \middle| \mathcal{F}(t)\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)} \max_{[0,T]}\{S_t\} - S_T \middle| \mathcal{F}(t)\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)} \max_{[0,T]}\{S_t\} \middle| \mathcal{F}(t)\right] - S_t. \end{aligned}$$

Now to do further analysis (to reduce Conditional Expectation to an Expectation via the Independence Lemma) we may want to separate the term $\max_{[0,T]}\{S_t\}$ into some expression involving $S_u, u \in [0,t]$ and $S_u, u \in [t,T]$. One way to do this is

$$\max_{t\in[0,T]}\{S_t\} = \max_{u\in[0,t]}\{S_u\} \vee \max_{u\in[t,T]}\{S_u\}.$$

So far so good, but we need to do more work here, since the operator $\vee$ does not ”factor out" of the conditional expectation (we cannot factor $\max_{[0,t]}\{S_t\}$ out of $E(.\Big|\mathcal{F}(t))$). Looking at this in another way, the running max of S_t:

$$Y_t = \max_{u\in[0,t]} S_u$$

is not a Markov process.

However, there is a usual approach in studying Markov process like this: If $X(t)$ is not a Markov process, by increasing the components of $X(t)$, we may still yet obtain a Markov process.

In this case, we consider the two-component process (S_t, Y_t) instead of just Y_t. Then for $s > t$

$$Y(s) = \max\{Y_t, \max_{[t,s]} S_u\} = \max\{Y_t, S_t e^{\sigma \widehat{M}(s-t)}\},$$

where recall that we defined in Section 1

$$\begin{aligned} \alpha &:= \frac{r - \frac{1}{2}\sigma^2}{\sigma} \\ \widehat{W}(u) &:= W(t+u) - W_t + \alpha u, u \geq 0 \\ \widehat{M}(u) &:= \max_{s\in[0,u]} \widehat{W}_s. \end{aligned}$$

Then since $\widehat{M}$ and $\widehat{W}$ are independent of $\mathcal{F}(t)$ under the risk neutral measure, we get that (S_t, Y_t) is a Markov process under this measure as well (how to reach this conclusion is left as a homework exercise).

We then have

$$\begin{aligned}
V_t &= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)}V_T\Big|\mathcal{F}(t)\Big] \\
&= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)}\max_{[0,T]}\{S_t\}\Big|\mathcal{F}(t)\Big] - S_t \\
&= \mathbb{E}^{\mathbb{Q}}\Big[e^{-r(T-t)}\max\{Y_t, S_t e^{\sigma \widehat{M}_{T-t}}\}\Big|\mathcal{F}(t)\Big] - S_t \\
&= v(t, S_t, Y_t),
\end{aligned}$$

where

$$v(t,x,y) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\max(y, xe^{\sigma\widehat{M}_{T-t}})\Big] - x.$$

5.5.2 Apply Girsanov

Similar to the discussion in section 1, $v(t,x,y)$ involves the distribution of the running max of a Brownian motion with drift, so we want to apply Girsanov's theorem to transform it to a Brownian motion. The result is

$$v(t,x,y) = e^{-r(T-t)}\widehat{\mathbb{E}}\Big[e^{\alpha\widehat{W}_{T-t}-\frac{\alpha^2}{2}(T-t)}\max(y, xe^{\sigma\widehat{M}_{T-t}})\Big] - x,$$

where $\widehat{W}$ now is a Brownian motion under $\widehat{\mathbb{P}}$.

Note that the expression inside expectation is not (yet) of the form provided by the function $H_s(\alpha,\beta,k,b)$. Noting the fact that $x > 0$ since it is the stock price S_t, we have

$$\begin{aligned}
\max(y, xe^{\sigma\widehat{M}_{T-t}}) &= y \quad \text{if} \quad \widehat{M}_{T-t} < \frac{1}{\sigma}\log(y/x) \\
\max(y, xe^{\sigma\widehat{M}_{T-t}}) &= xe^{\sigma\widehat{M}_{T-t}} \quad \text{if} \quad \widehat{M}_{T-t} \geq \frac{1}{\sigma}\log(y/x).
\end{aligned}$$

Denoting

$$b := \frac{1}{\sigma}\log(y/x),$$

and note that the domain of interest for $v(t,x,y)$ is $y \geq x > 0$ thus $b \geq 0$. Then

$$\begin{aligned}
\max(y, xe^{\sigma\widehat{M}_{T-t}}) &= y\mathbf{1}_{\widehat{M}_{T-t}<b} + xe^{\sigma\widehat{M}_{T-t}}\mathbf{1}_{\widehat{M}_{T-t}\geq b} \\
&= y + \Big[xe^{\sigma\widehat{M}_{T-t}} - y\Big]\mathbf{1}_{\widehat{M}_{T-t}\geq b}.
\end{aligned}$$

Plug this back into the expectation, coupled with the fact that

$$\widehat{\mathbb{E}}\Big[ye^{\alpha\widehat{W}_{T-t}-\frac{\alpha^2}{2}(T-t)}\Big] = y,$$

after simplification we have

$$\begin{aligned} v(t,x,y) &= e^{-r(T-t)}y - x + xe^{-(r+\frac{\alpha^2}{2})(T-t)}\widehat{\mathbb{E}}\Big[\mathbf{1}_{\{\widehat{M}_{T-t}\geq b\}}e^{\alpha\widehat{W}_{T-t}+\sigma\widehat{M}_{T-t}}\Big] \\ &\quad -ye^{-(r+\frac{\alpha^2}{2})(T-t)}\widehat{\mathbb{E}}\Big[\mathbf{1}_{\{\widehat{M}_{T-t}\geq b\}}e^{\alpha\widehat{W}_{T-t}}\Big] \\ &= e^{-r(T-t)}y - x + xe^{-(r+\frac{\alpha^2}{2})(T-t)}H_{T-t}\big(\alpha,\sigma,-\infty,b\big) \\ &\quad -ye^{-(r+\frac{\alpha^2}{2})(T-t)}H_{T-t}\big(\alpha,0,-\infty,b\big). \end{aligned} \tag{5.10}$$

Remark 5.5.1. *You may question why we go into such length to derive the closed form expression for the value of the Barrier or Lookback Option via expectation. An alternative, as you may have already known, is ti simulate the paths of S_t and take the average over the simulated paths to obtain an approximation for the expectation. However, the work that we have done, for example, in expressing $v(t,x,y)$ in the form of (5.10) can be very helpful in increasing the efficiency of the computation. We have "simplified" the computation (not in the expression, of course, but in the actual computation time). The reason is the function $H_s(\alpha,\beta,k,b)$ is found explicitly via the cumulative distribution of the standard normal, which we have very efficient algorithms to compute. On the other hand, as you can already imagine, the efficiency of simulating the paths of S_t, also taking into account its running max or when it reaches the barrier and then take the average might not be as good as just computing the probability of a Normal distribution.*

CHAPTER 6 PDE for Lookback Option

6.1 Preliminary discussion

Let S_t satisfies

$$\begin{aligned} dS_t &= rS_t dt + \sigma(t, S_t) S_t dW_t \\ S(0) &= x > 0. \end{aligned}$$

Note that σ is a function of t, St here, in stead of being a constant. We call this the *local volatility* model and make the assumption that $\sigma(t, x) > 0$ for all t, x.

Consider the Lookback Option:

$$V_T = \max_{[0,T]} S_t - S_T.$$

Then by risk neutral pricing

$$V_t = E\big(e^{-r(T-t)} \max_{u \in [0,T]} S_u \Big| \mathcal{F}(t)\big) - S_t.$$

Similar to what we did in Lecture 5 notes, define

$$Y_t = \max_{[0,t]} S_u,$$

then for $s > t$

$$Y(s) = \max\{Y_t, \max_{u \in [t,s]} S_u\}$$

In Homework 5, we have discussed that when σ is constant, then $\{Y_t, S_t\}$ is a Markov process. The argument is by Independence Lemma. For the current local volatility model, the Independence Lemma no longer applies, since we cannot conclude that $\int_t^T \sigma(u, S_u) dWu$ is independent of $\mathcal{F}(t)$. However, it is still true that $\{Y_t, S_t\}$ is Markov. Indeed, we have the following principle:

Principle: If $S_t, t \geq 0$ is Markov with respect to $\mathcal{F}(t)$ and $Y_t = \max_{u \in [0,t]} S_u$ then $\{Y_t, S_t\}$ is also Markov with respect to $\mathcal{F}(t)$.

We call it a principle instead of a theorem because we will not give it a proof due to technical details. Thus we also have

$$\begin{aligned} V(t) &= \mathbb{E}\Big[e^{-r(T-t)}V_T\Big|\mathcal{F}(t)\Big] \\ &= \mathbb{E}\Big[e^{-r(T-t)}\max_{[0,T]}\{S_t\}\Big|\mathcal{F}(t)\Big] - S_t \\ &= \mathbb{E}\Big[e^{-r(T-t)}\max\{Y_t, \max_{u\in[t,T]} S_u\}\Big|\mathcal{F}(t)\Big] - S_t \\ &= v(t, S_t, Y_t), \end{aligned}$$

where

$$v(t,x,y) = \mathbb{E}\Big[e^{-r(T-t)}\max\{Y_t, \max_{u\in[t,T]} S_u\}\Big|S_t = x, Y_t = y\Big] - x.$$

Remark 6.1.1. *Note that here $V_T = G(Y_T, S_T)$ where $G(x,y) = y - x$. For this case, we call the option* **floating strike lookback option**. *Clearly one can consider other types of function G as well. The only difference this would affect on the PDE is the boundary conditions. See Section* (6.6) *for more details.*

Now assuming that v is $C^{1,2,2}$, that is once continuously differetiable in t and twice continuously differentiable in x, y, we would like to derive a PDE that v satisfies.

But note the following difference in our current case: *Y_t is not a $C^{1,2}$ function of S_t* so we cannot write down its dynamics using Ito's formula. In other words, we do not know what dY_t is explicitly.

However, observe that for $s < t$

$$Y(s) = \max_{u\in[0,s]} S_u \le \max_{u\in[0,t]} S_u = Y_t,$$

simply because the max over a bigger set is not smaller than the max over a (smaller) set contained in it. Therefore Y_t is *an increasing (meaning it is non-decreasing)* function.

From the discussion of the Lebesgue-Stieltjes integral of Chapter 11, we have learned how to integrate with respect to functions of bounded variation. Recall that increasing function is of bounded variation. Therefore, it makes sense to talk about dY_t (in the Lebesgue-Stieltjes integral sense, that is).

However, we did not discuss the Ito's formula for $v(t, S_t, Y_t)$ where S_t is an Ito process and Y_t is an increasing process. But suppose we just formally carry out the usual Ito's rule to $e^{-rt}v(t, S_t, Y_t)$, what we should get is

$$\begin{aligned} de^{-rt}v(t, S_t, Y_t) &= e^{-rt}\Big\{[-rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx \\ &\quad +\frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2]\Big|_{(x,y)=(S_t,Y_t)}\Big\}dt \\ &\quad +e^{-rt}\frac{\partial}{\partial x}v(t,S_t,Y_t)\sigma(t,S_t)S_t dW_t + e^{-rt}\frac{\partial}{\partial y}v(t,S_t,Y_t)dY_t \\ &\quad +e^{-rt}\frac{\partial^2}{\partial xy}v(t,S_t,Y_t)d\langle S,Y\rangle(t) + e^{-rt}\frac{\partial^2}{\partial y^2}v(t,S_t,Y_t)d\langle Y\rangle(t). \end{aligned}$$

Remark 6.1.2. *We will discuss what $d\langle S,Y\rangle(t)$ and $d\langle Y\rangle(t)$ means in the following section. For now, you can formally replace $d\langle S,Y\rangle(t)$ with $dS_t dY_t$ and $d\langle Y\rangle(t)$ with $[dY_t]^2$ to get an intuition.*

Since

$$e^{-rt}V_t = e^{-rt}v(t, S_t, Y_t),$$

$e^{-rt}v(t, S_t, Y_t)$ is a martingale. On the RHS of the above equation, the only martingale term we have is

$$\frac{\partial}{\partial x}v(t, S_t)\sigma(t, S_t)S_t dW_t.$$

The principle of deriving our PDE is that any other terms that do not contribute to the martingale property of the RHS should be set to 0. But before we can do that, we need to understand the following:

(i) Is the Ito's rule that we just formally applied correct? (If it is not correct there is no point in discussing the items below).

(ii) What are $d\langle S,Y\rangle(t)$ and $d\langle Y\rangle(t)$?

(iii) How to understand dY_t?

We will address these questions in the folowing order (ii), (i) and (iii) and then derive the PDE for $v(t, x, y)$ after that.

6.2 A summary of the main results

In what follows, we will discuss many technical details about the behavior of Y_t, the running max of S_t, and the extension of Ito's formula to functions depending on Y_t. Thus, it is easy to lose track of the main points of the discussion. To help the readers to follow, we list these points here.

a. Y_t is an increasing (non-decreasing) process in t.

b. Y_t has no dynamics. That is we cannot represent Y_t in the form

$$Y_t = Y_0 + \int_0^t \alpha_s ds + \int_0^t \sigma_s dW_s.$$

c. $\langle Y, S\rangle_t = \langle Y\rangle_t = 0$.
d. If $f(y)$ is a differentiable function then

$$f(Y_t) - f(Y_0) = \int_0^t f'(Y_s) dY_s.$$

e. Y_t can only increase on the set $\{t : Y_t = S_t\}$, which is closed and contains no interval.

f. If the term $\sigma(t)$ in the dynamics of S_t is positive for all t then

$$\int_0^t \alpha(s) ds + \int_0^t g(s) dY(s) = 0, \quad \forall t > 0$$

if and only if $\alpha(t) = 0$ and $g(t)\mathbf{1}_{\{Y_t = S_t\}} = 0$ for all $t > 0$.

(This allows us to set the terms other than the dW_t term equal to 0 when we derive the PDE).

6.3 The quadratic variation and covariation

Fix $T > 0$. Let $X(t), Y_t$ be functions defined on $[0, T]$. Recall the following definitions:

Definition 6.3.1. *The total variation of Y on $[0, T]$, denoted as $TV_Y(T)$ is defined as the smallest (finite) number such that for all partitions $0 = t_0 < t_1 < t_2 < ... < t_n = T$*

$$\sum_{i=0}^{n-1} |Y(t_{i+1}) - Y(t_i)| \leq TV_Y(T).$$

If there is no such number, we define $TV_Y(T) = \infty$.

We also say Y is a function of bounded variation (on $[0, T]$) if $TV_Y(T) < \infty$.

Definition 6.3.2. *The quadratic variation of Y on $[0, t]$, if it exists is defined as*

$$\langle Y\rangle(t) = \lim_{n\to\infty} \sum_{i=0}^{n-1} |Y(t_{i+1}^n) - Y(t_i^n)|^2,$$

where the limit is taken in probability, and for each fixed n, $0 = t_0^n < t_1^n < t_2^n < ... < t_n^n = t$ is a partition of $[0, t]$ such that its mesh size: $\max_i |t_{i+1}^n - t_i^n|$ goes to 0 as $n \to \infty$.

Definition 6.3.3. *The covariation between X and Y on $[0,t]$, if it exists is defined as*

$$\langle X,Y\rangle(t) = \lim_{n\to\infty}\sum_{i=0}^{n-1}\big(X(t_{i+1}^n)-X(t_i^n)\big)\big(Y(t_{i+1}^n)-Y(t_i^n)\big),$$

where the limit is taken in probability, and for each fixed n, $0 = t_0^n < t_1^n < t_2^n < ... < t_n^n = t$ is a partition of $[0,t]$ such that its mesh size: $\max_i |t_{i+1}^n - t_i^n|$ goes to 0 as $n \to \infty$.

Note: Some authors (including Shreve in our textbook, see Exercise 7.4) called the covariation the cross quadratic variation.

We will also state the following facts about quadratic variation and covariation. The proof is more or less contained in the extra credit problem in Homework 1.

(i) If Y is increasing then Y is of bounded variation.

(ii) If Y is continuous and of bounded variation, then $\langle Y\rangle(t)$ = 0.

(iii) If Y is of bounded variation and X is continuous, then $\langle X,Y\rangle(t)$ = 0.

(iv) The quadratic variation $\langle X\rangle(t)$ and covariation $\langle X,Y\rangle(t)$ of any two processes X,Y, if exist, are of bounded variation on $[0,T]$. Therefore, it makes sense to talk about $d\langle X\rangle(t)$ and $d\langle X,Y\rangle(t)$.

Applying these facts to our situation, we see that indeed

$$\langle Y\rangle(t) = 0$$
$$\langle S,Y\rangle(t) = 0$$

So question (ii) of Section 1 is answered.

6.4 An extension of Ito's formula

We now give answer to question (i) of Section 1. Let $W(t)$ be a Brownian Motion and $\mathcal{F}(t)$ a filtration for $W(t)$.

$$X^i(t) = X^i(0) + \int_0^t \alpha^i(s)ds + \int_0^t \sigma^i(s)dW(s) + A^i(t), i = 1,2$$

where α^i, σ^i, A^i are *stochastic processes* adapted to $\mathcal{F}(t)$, σ^i are chosen so that the stochastic integral is well-defined, and $A^i(t)$ are *continuous functions of bounded variations*. Let

$f(t, x_1)$ be a $C^{1,2}$ function. Then

$$\begin{aligned} df(t, X^1(t)) &= \Big[\frac{\partial}{\partial t} f + \left(\frac{\partial}{\partial x_1} f\right) \alpha_t^1 + \frac{1}{2}\left(\frac{\partial^2}{\partial (x_1)^2} f\right)(\sigma_t^1)^2\Big] dt \\ &\quad + \left(\frac{\partial}{\partial x_1} f\right) \sigma^1(t) dW(t) + \left(\frac{\partial}{\partial x_1} f\right) dA^1(t). \end{aligned}$$

Let $f(t, x_1, x_2)$ be a $C^{1,2,2}$ function. Then

$$\begin{aligned} df(t, X^1(t), X^2(t)) &= \Big[\frac{\partial}{\partial t} f + \sum_{i=1}^{2} \Big\{ \left(\frac{\partial}{\partial x_i} f\right) \alpha_t^i + \frac{1}{2}\left(\frac{\partial^2}{\partial (x_i)^2} f\right)(\sigma_t^i)^2 + \left(\frac{\partial^2}{\partial x_1 x_2} f\right) \sigma_t^1 \sigma_t^2 \Big\}\Big] dt \\ &\quad + \Big\{ \sum_{i=1}^{2} \left(\frac{\partial}{\partial x_i} f\right) \sigma^i(t) \Big\} dW(t) + \sum_{i=1}^{2} \left(\frac{\partial}{\partial x_i} f\right) dA^i(t). \end{aligned}$$

where by f we understand as $f(t, X^1(t), X^2(t))$.

Remark 6.4.1. *We do not give a proof of this extension. But you can see the formula is just the application of Ito's formula as we used to do, combined with the facts about quadratic variation and covariation of bounded variation process that we discussed in Section 2.*

Remark 6.4.2. *It is true that $\int_0^t \alpha^i(s) ds$ is also a continuous function of bounded variation. So what is the difference between $\int_0^t \alpha^i(s) ds$ and $A^i(t)$? Can we combine them into just 1 term? The answer is no, because these two terms have very different property. We say the term $\int_0^t \alpha^i(s) ds$ is* **absolutely continuous** *(with respect to the Lebesgue measure dt). Basically this means it can be represented as an integral with respect to dt (which it is already in that form). The term $A^i(t)$ in this formula is meant to be* **singularly continuous** *(with respect to the Lebesgue measure dt). For our purpose, what it means is that even though $A^i(t)$ is continuous we cannot represent $A^i(t)$ as an integral with respect to dt. Therefore, the two terms should be kept separate.*

Remark 6.4.3. *You should compare and contrast these Ito formulas with the ones we obtained in Chapter 11. There, the $A^i(t)$ are the pure jump processes. So while here we have the term $\left(\frac{\partial}{\partial x_i} f\right) dA^i(t)$; in Chapter 11 the corresponding term is $\sum_{0<s\leq t} f(X^i(s)) - f(X^i(s-))$. We mentioned that it's not always possible to get the differential form in the Ito's formula in Chapter 11. Here note that it is always in differential form.*

6.4.1 An intuition on the difference between the 2 Ito's formulae

For simplicity, let's just consider 2 cases:

(i) $X^1(t) = A(t)$

A is continuous and of bounded variation.
(ii) $X^2(t) = J(t)$,
J is pure jump.
Let f be C^1 function. Then (from Ito's formula) we have

$$f(X_t^1) = f(X_s^1) + \int_s^t f'(X_u^1)dX^1(u);$$
$$f(X_t^2) = f(X_s^2) + \sum_{s<u\leq t} f(X^2(u)) - f(X^2(u-)).$$

Note the derivative in the 1st case and the original function in the 2nd case. Why is this? We always have the following identity (assuming $X_t^1 - X_s^1 \neq 0$)

$$f(X_t^1) - f(X_s^1) = \frac{f(X_t^1) - f(X_s^1)}{X_t^1 - X_s^1}(X_t^1 - X_s^1).$$

And so intuitively, we have for s very close to t,

$$\begin{aligned} f(X_t^1) - f(X_s^1) &\approx f'(X^1(s))(X_t^1 - X_s^1) \\ &\approx f'(X^1(s))dX^1(s). \end{aligned}$$

This is correct, **if** $X_t^1 \to X_s^1$ **as** $t \to s$, which requires the continuity of $X^1(t)$ (so that the difference quotient approximates the derivative of f). But in the case of X^2, if X^2 is not continuous at s (it has a jump at s), then we cannot say the difference quotient approximates the derivative of f at $X^2(s)$ in any sense. Therefore, we cannot write it in the differential form, and can only write it as the form we always used in Chapter 11.

6.5 The integral dY_t

6.5.1 Some preliminary discussion

Recall that we define $Y_t := \max_{u\in[0,t]} S_u$ to be the running max of S_t. We have observed that Y_t is non-decreasing. But can we say more? For example, is there any interval where Y is strictly increasing, not just non-decreasing? To anwer that, we make the following observations (recall that by definition, $S_t \leq Y_t$):

(i) Suppose that $S_t < Y_t$ for some t then there must exist an interval $[a, b]$ around t (that is, $t \in (a, b)$)so that Y is constant on $[a, b]$.

Reason: Since S_t is continuous, if $S_t < Y_t$ there must exist an interval $[a, b]$ around t so that $S_u < Y_t$ for all $u \in [a, b]$. Since $Y_t = Y_a \vee \max_{u\in[a,t]} S_u$, it is clear that $Y_a = Y_t$.

Similar reasoning gives $Y_t = Y_b$. Because Y is increasing, it follows that Y is constant on $[a, b]$.

Remark: Observation (i) tells us that Y can only increase when $Y_t = S_t$. The next observation tells us how often this happens.

(ii) Suppose Y_t is strictly increasing on $[a, b]$ (that is for all $u < v$ in $[a, b], Y(u) < Y(v)$) then $Y_t = S_t$ for all $t \in [a, b]$. It also follows that S_t is also strictly increasing on $[a, b]$.

Reason: Suppose there is u in $[a, b]$ such that $S_u < Y(u)$ then by oservation (i) we can find an interval around u on which Y is constant, contradicting the assumption that Y is strictly increasing. Thus $S_t = Y_t$ for all t in $[a, b]$. The second conclusion is obvious.

Remark: Observation (ii) tells us that on any interval where Y_t is strictly increasing, S_t also has to be strictly increasing. Intuitively you can see that this will not happen on any interval $[a, b]$ if the volatility term σ_t of S_t is positive. The reason is if S_t is increasing, then its quadratic variation on $[a, b]$ must be 0 as we mentioned before. On the other hand, this should be $\int_a^b \sigma_t^2 dt$, which is a contradiction if $\sigma_t > 0$.

6.5.2 The set $C := \{t : Y_t = S_t\}$

From the above discussion, we see that Y_t can only increase on the set $C := \{t : Y_t = S_t\}$. We emphasize that C is a *random set*, that is we should write

$$C(\omega) := \{t : Y_t(\omega) = S_t(\omega)\}.$$

But our convention is that we just remember the fact that C is random and omit the writing of ω.

We also observe that C is a closed set (in the sense that its complement is an open set) from an elementary topological result that says the inverse image of a closed set is closed:

$$C = (Y_t - S_t)^{-1}(\{0\}).$$

From observation (ii) in the above section, C does not contain any interval. In this sense it is rather “small". Trivially then C cannot be the whole interval $[0, T]$. Equivalently, its nonempty complement C^c should be rather “large." But observe also that trivially, C^c cannot be too large (that is equal to $[0, T]$) because it would imply that

$$Y_T = Y_0,$$

which would force $S_t \leq S_0$ on $[0, T]$ and that is impossible.

Because C^c is open, for any point in C^c, we can find an open interval around it that is also in C^c. In other words, for any $t_0 \in C^c$, we can find $\varepsilon > 0$ so that

$$\int_{t_0-\varepsilon}^{t_0+\varepsilon} g(s)dY(s) = 0,$$

for any measurable function g. This is because Y is constant on the interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ by the fact that $(t_0 - \varepsilon, t_0 + \varepsilon) \subseteq C^c$.

From this observation, coupled with the fact that any open set in $\mathbb{R}$ can be written as a countable union of open intervals, it follows that for any measurable function g

$$\int_0^T \mathbf{1}_{C^c}(s) g_s dY_s = 0.$$

And

$$\int_0^T g_s dY_s = \int_0^T \mathbf{1}_C(s) g_s dY_s.$$

Finally, we discuss the fact that C is the set of increase of Y_t in the following sense: for any $s < t$, let E be the event that there is a point $u \in (s,t)$ such that $S_u = Y_u$. Then

$$P(Y_s < Y_t | E) = 1.$$

Indeed, let $\tau := \inf\{u \geq s : S_u = Y_u\}$. Then τ is a stopping time. Conditioned on E, $\tau < t$ with probability 1. Then $S_u - S_\tau, u \geq \tau$ is an Ito process starting at τ. By exactly the same argument as we discussed in lecture 4, with probability 1, there are infinitely many points u_n close to τ so that $S_{u_n} > S_\tau$. That implies $Y_{u_n} > Y_\tau$. Thus $P(Y_t > Y_s | E) = 1$.

Main result

Theorem 6.5.1. *Let S_t satisfies*

$$\begin{aligned} dS_t &= \alpha(t) S_t dt + \sigma(t) S_t dW_t \\ S(0) &= x > 0, \end{aligned}$$

where α, σ can be random processes **with** *$\sigma(t) > 0$. Let $\beta(s)$, $g(s)$ be continuous process. Then*

$$\int_0^t \beta(s)ds + \int_0^t g(s)dY(s) = 0, \quad \forall t > 0$$

if and only if $\beta(t) = 0$ and $g(t)\mathbf{1}_{\{Y_t = S_t\}} = 0$ for all $t > 0$.

Proof. Suppose $\beta(t) = 0$ and $g(t)\mathbf{1}_{\{Y_t = S_t\}} = 0$. We have

$$\int_0^t \beta(s)ds + \int_0^t g(s)dY(s) = \int_0^t g(s)\mathbf{1}_C dY(s) + \int_0^t g(s)\mathbf{1}_{C^c} dY(s) = 0.$$

We now show the other direction. Let $t_0 \in C^c$. Since C^c is open, we can find an interval (a, b) around t_0 so that $(a, b) \subseteq C^c$. Thus

$$\int_a^t g(s)dY_s = 0, a \leq t \leq b.$$

It follows that $\int_0^t g(s)dY(s)$ is a constant on (a, b) and hence differentiable at t_0 with derivative being 0.

On the other hand $\frac{\partial}{\partial t}\int_0^t \beta(s)ds = \beta(t)$ for all t. Thus by taking derivative of both sides of the equation

$$\int_0^t \beta(s)ds + \int_0^t g(s)dY(s) = 0$$

at t_0, we conclude that $\beta(t_0) = 0$ for all $t_0 \in C^c$.

Now let $t_0 \in C$. Since $\sigma_t > 0$, C contains no interval from our above discussion. Thus for any n, the interval $[t_0 - 1/n, t_0 + 1/n]$ has non empty intersection with C^c. That is we can find $t_n \in C^c$ in any interval $[t_0 - 1/n, t_0 + 1/n]$, $t_n \neq t_0$.

Since $t_n \in C^c, \beta(t_n) = 0$ from the previous paragraph. Obseve also that $t_n \to t_0$ and hence $\beta(t_0) = 0$ by continuity of β. Thus $\beta(t) = 0$ for all t.

We then have

$$\int_0^t g(u)dY(u) = 0, \forall t,$$

which implies

$$\int_s^t \mathbf{1}_C(u)g(u)dY(u) = 0, \forall s \leq t.$$

Note that since $\int_0^t \mathbf{1}_{C^c}(s)g(s)dY_s = 0$, we cannot conclude anything about the values of g on C^c.

Now since C is the set of increase of Y_t as discussed above, it follows that $g(s)\mathbf{1}_C(s) = 0$. Otherwise, suppose that there is a point $t_0 \in C$ so that $g(t_0) > 0$. Since g is continuous, we can find ε and an interval (a, b) around t_0 so that $g \geq \varepsilon$ on (a, b). But then

$$\int_a^b g(s)\mathbf{1}_C(s)dYs \geq \int_a^b \varepsilon\mathbf{1}_C(s)dYs \geq \varepsilon(Y_b - Y_a) > 0,$$

which is a contradiction.

6.6 Derivation of the PDE

Putting all the above information together, we have

$$\begin{aligned} e^{-rt}v(t,S_t,Y_t) &= v(0,S(0),Y(0)) + \int_0^t e^{-ru}[\mathcal{L}v(u,S_u,Y_u) - rv(u,S_u,Y_u)]du \\ &\quad + \int_0^t e^{-ru}\frac{\partial}{\partial y}v(u,S_u,Y_u)dY_u + \int_0^t e^{-ru}\frac{\partial}{\partial x}v(u,S_u,Y_u)\sigma(u,S_u)S_u dW_u. \end{aligned}$$

where

$$\mathcal{L}v(t,x,y) = \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2.$$

Since the LHS is a martingale, we set

$$\int_0^t e^{-ru}[\mathcal{L}v(u,S_u,Y_u) - rv(u,S_u,Y_u)]du + \int_0^t e^{-ru}\frac{\partial}{\partial y}v(u,S_u,Y_u)dY_u = 0, \forall t.$$

Apply Corollary (6.5.1) we conclude that

$$\begin{aligned} -rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2 &= 0; t < T, 0 < x \le y < \infty \\ \frac{\partial}{\partial y}v(t,y,y) &= 0; t \le T, y > 0 \qquad (6.1) \\ v(T,x,y) &= y - x; \qquad (6.2) \\ v(t,0,y) &= e^{-r(T-t)}y. \qquad (6.3) \end{aligned}$$

Condition (2),(3),(4) are boundary conditions. Condition (2) is called a Neumann condition, since it imposes the value of a derivative of v on the boundary . Condition (4) comes from the fact that once S_t hits 0 at time t it stays there so the running max is a constant on $[t,T]$: $Y(u) = Y_t, u \ge t$. Thus we get

$$\begin{aligned} v(t,0,Y_t) &= E(e^{-r(T-t)}Y(T)|\mathcal{F}(t)) \\ &= E(e^{-r(T-t)}Y_t|\mathcal{F}(t)) = e^{-r(T-t)}Y_t. \end{aligned}$$

More generally, suppose we consider the generalized lookback option:

$$V(T) = G(S(T),Y(T)),$$

then the same argument shows that $V(t) = v(t,S_t,Y_t)$ where $v(t,x,y)$ satisfies the PDE

$$\begin{aligned} -rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2 &= 0; t < T, 0 < x \le y < \infty \\ \frac{\partial}{\partial y}v(t,y,y) &= 0; t \le T, y > 0 \qquad (6.4) \\ v(T,x,y) &= G(x,y); 0 \le x \le y \quad (6.5) \\ v(t,0,y) &= e^{-r(T-t)}G(0,y). \qquad (6.6) \end{aligned}$$

6.7 Appendix: The dynamics of Y

The results in this section is not used in deriving the PDE for the Lookback Option. However, it is interesting in shedding some more light into the behavior of Y_t. In particular, we show here that Y is a singularly continuous process. That is it cannot be represented as a Ito process.

The question we'll address is: does there exist $\alpha(t), \beta(t)$ such that

$$dY_t = \alpha_t dt + \beta_t dW(t)? \tag{6.7}$$

This indeed cannot happen, as the following Lemma shows

Lemma 6.7.1. *Let Y_t be the running maximum of S_t where the volatility σ_t of S_t is positive. Then Y_t* **cannot be represented** *in the form of equation (6.7). In other words, Y_t is a singularly continuous process.*

Proof.

Suppose there exist $\alpha(t), \beta(t)$ such that

$$dY_t = \alpha_t dt + \beta_t dW(t).$$

Then since Y_t is increasing, $\langle Y \rangle_t = \int_0^t \beta_s^2 ds = 0$. But this means $\beta_s = 0, \forall s$. Then $Y_t = Y_0 + \int_0^t \alpha_s ds$. But that means Y is differentiable and $Y'(t) = \alpha_t$ for all t. From the above discussion, we learned that $Y'(t) = 0$ on the set $\{S_t < Y_t\}$. But that means $\alpha_t = 0$ on $\{S_t < Y_t\}$. Since the set $\{S_t = Y_t\}$ has Lebesgue measure 0 (see the following Lemma), it follows that

$$Y_t = Y_0 + \int_0^t \alpha_s ds = Y_0,$$

with probability 1. We have observed that this forces $S_u \leq S_0$ on $[0, t]$ with probability 1 and this is not possible.

The above proof relies on the following lemma that tells us precisely how small the set $\{Y_t = S_t\}$ is:

Lemma 6.7.2. *Let S_t satisfies*

$$\begin{aligned} dS_t &= \alpha(t) S_t dt + \sigma(t) S_t dW_t \\ S(0) &= x > 0, \end{aligned}$$

where α, σ can be random processes **with** *$\sigma(t) > 0$. Define the set $C := \{t : Y_t = S_t\}$. Then with probability 1,*

$$\int_0^T \mathbf{1}_{C(\omega)}(s) ds = 0.$$

In other words, the Lebesgue measure of the set $C(\omega)$, or the total length of $C(\omega)$, is 0 with probability 1.

Proof. By Fubini's theorem:

$$E\left(\int_0^T \mathbf{1}_{S_t=Y_t} dt\right) = \int_0^T E(\mathbf{1}_{S_t=Y_t}) dt.$$

Since $\sigma_t > 0$, (S_t, Y_t) has a joint p.d.f. that can be explicitly computed. Thus $E(\mathbf{1}_{S_t=Y_t}) = 0$ for all t. It follows that

$$E\left(\int_0^T \mathbf{1}_{S_t=Y_t} dt\right) = 0.$$

Since the term inside the expectation is non-negative, it must be 0 with probability 1.

CHAPTER 7 Asian option

7.1 Preliminary discussion

Let S_t satisfies

$$\begin{aligned} dS_t &= rS_t dt + \sigma(t, S_t) S_t dW_t \\ S(0) &= x > 0. \end{aligned}$$

Denote for $t \in [0, T]$

$$\begin{aligned} Y_t &= \int_0^t S(u) du; \\ S_{ave}(t) &= \frac{Y_t}{t}. \end{aligned}$$

Consider the Generalized Asian Option:

$$V_T = G\big(S(T), S_{ave}(T)\big).$$

Depends on the specific form G takes, we have the following types of Asian options:

(i) *Average price call*: $G(x, y) = (y - K)^+$;

(ii) *Average price put*: $G(x, y) = (K - y)^+$;

(iii) *Average strike call*: $G(x, y) = (x - y)^+$;

(iv) *Average strike put*: $G(x, y) = (y - x)^+$.

By risk neutral pricing

$$\begin{aligned} V_t &= E\Big\{e^{-r(T-t)} G\big(S_T, S_{ave}(T)\big)\Big|\mathcal{F}(t)\Big\} \\ &= E\Big\{e^{-r(T-t)} G\big(S_T, \frac{Y_T}{T}\big)\Big|\mathcal{F}(t)\Big\}. \end{aligned}$$

In the Lookback Option, we have discussed that the process (S_t, Y_t) (where $Y_t := \max_{[0,t]} S_u$) has Markov property. The same principle applies here

Principle: If $S_t, t \geq 0$ is Markov with respect to $\mathcal{F}(t)$ and $Y_t = \int_0^t S(u)du$ then $\{Y_t, S_t\}$ is also Markov with respect to $\mathcal{F}(t)$.

The intuitve reason why this principle is true is because we can write

$$Y(T) = Y_t + \int_t^T S(u)du.$$

Therefore, intuitively, to compute the conditional expectation of $Y(T)$ on $\mathcal{F}(t)$, we only need the value of Y_t plus the conditional expectation of $S(u)$ given $\mathcal{F}(t)$, which also only depends on S_t by the assumption on Markov property of S. In other words, the conditional expectation of $Y(T)$ on $\mathcal{F}(t)$ only depends on S_t, Y_t, thus the process S_t, Y_t is Markov.

Note here however that Y_t by itself is generally NOT a Markov process (same as the conclusion we draw for the running max of S_t in the Lookback option).

Thus there exists $v(t, x, y)$ such that

$$V(t) = v(t, S_t, Y_t)$$

where

$$\begin{aligned} v(t, x, y) &= E\Big\{e^{-r(T-t)}G\big(S_T, \frac{Y_T}{T}\big)\Big|S_t = x, Y_t = y\Big\} \\ &= E\Big\{e^{-r(T-t)}G\big(S_T, \frac{y + \int_t^T S_u du}{T}\big)\Big|S_t = x\Big\} \end{aligned}$$

To be able to proceed, one would need the knowledge of the joint distribution between $\int_t^T \sigma(u, S_u)dWu$ and $\int_t^T S_u du$. So without further assumption on S, this is the ultimate simplification that can be achieved to represent the option price as conditional expectation.

Remark 7.1.1. *If G is a linear function in x, y, however, then we can write down an explicit formula for $v(t, x, y)$. Do you see why? (Say for example, $G(x, y) = y - x$).*

7.2 PDE for Asian options

An explicit formula for the price in terms of expectation for an option of Asian type is not known, even if S follows the standard Black-Scholes (that is σ is a constant) model. To get an idea, observe that one will need to figure out the joint distribution of $(\int_0^t e^{W_u}du, e^{W_t})$ to start investigating how to express $v(t, x, y)$ above as an integral.

Therefore, it is important do drive a PDE with boundary conditions for $v(t, x, y)$ defined above.

Note that here, unlike the case of Lookback option, the process Y_t is an *absolutely continuous* process. Indeed, its dynamics is

$$\begin{aligned} dY_t &= S_t dt \\ Y(0) &= 0. \end{aligned}$$

Therefore, applying Ito's formula and setting the "dt" term to 0 is no problem:

$$de^{-rt}v(t,S_t,Y_t) = e^{-rt}\mathcal{L}v(t,S_t,Y_t)dt + e^{-rt}v_x(t,S_t,Y_t)S_t\sigma(t,S_t)dWt,$$

where

$$\mathcal{L}v(t,x,y) := -rv(t,x,y) + v_t(t,x,y) + v_x(t,x,y)rx + v_y(t,x,y)x + \frac{1}{2}v_{xx}(t,x,y)\sigma^2(t,x)x^2.$$

Recall that from the discussion on the quadratic variation and covariation in Lecture 6a, since Y_t is a function of bounded variation, $\langle Y\rangle_t = 0$ and $\langle S, Y\rangle_t = 0$.

Thus since $e^{-rt}v(t,S_t,Y_t)$ is a martingale, we set the dt term to 0 and get

$$-rv(t,x,y) + v_t(t,x,y) + v_x(t,x,y)rx + v_y(t,x,y)x + \frac{1}{2}v_{xx}(t,x,y)\sigma^2(t,x)x^2 = 0,$$
$$0 < x, y < \infty, 0 \le t < T.$$

But we also need to impose boundary conditions.

(i) At $t = T$ this is clear:

$$v(T,x,y) = G(x, \frac{y}{T}). \tag{7.1}$$

(ii) At $x = 0$: when the stock price hits 0, it stays there: $S(u) = 0, u \ge t$, so $Y(u)$ remains a constant on $[t,T]$ as well. Thus

$$\begin{aligned} v(t,0,Y_t) &= E(e^{-r(T-t)}G(S(T), \frac{Y_T}{T})|\mathcal{F}(t)) \\ &= E(e^{-r(T-t)}G(0, \frac{Y_t}{T})|\mathcal{F}(t)) = e^{-r(T-t)}G(0, \frac{Y_t}{T}). \end{aligned}$$

This implies that

$$v(t,0,y) = e^{-r(T-t)}G(0, \frac{y}{T}). \tag{7.2}$$

(iii) It's now natural to finish with the boundary condition at $y = 0$. However, note that

$$v(t,x,0) = E(e^{-r(T-t)}G(S(T), \frac{\int_t^T S_u du}{T})|S_t = x),$$

but in general we don't know what this is.

(iv) We instead then tries to seek the "boundary condition" for y at ∞. Suppose that

$$\lim_{y\to\infty} G(x,y) = 0.$$

Note that the average price put: $G(x,y) = (K-y)^+$ and the average strike call: $G(x,y) = (x-y)^+$ satisfy this condition. Then we have

$$\lim_{y\to\infty} v(t,x,y) \;=\; E(e^{-r(T-t)} \lim_{y\to\infty} G(S(T), \frac{y+\int_t^T S_u du}{T})|S_t = x) = 0.$$

Thus we can set the condition

$$\lim_{y\to\infty} v(t,x,y) = 0. \tag{7.3}$$

Then we have the following PDE for the Asian option, *assuming the condition* $\lim_{y\to\infty} G(x,y) = 0$

$$\begin{aligned}
&-rv(t,x,y) + v_t(t,x,y) + v_x(t,x,y)rx + v_y(t,x,y)x + \frac{1}{2}v_{xx}(t,x,y)\sigma^2(t,x)x^2 = 0,\\
&0 < x, y < \infty, 0 \le t < T;\\
&v(T,x,y) = G(x, \frac{y}{T});\\
&v(t,0,y) = e^{-r(T-t)}G(0, \frac{y}{T});\\
&\lim_{y\to\infty} v(t,x,y) = 0.
\end{aligned}$$

But then what about the average price call: $G(x,y) = (y-K)^+$ and the average strike put: $G(x,y) = (y-x)^+$? Intuitively we want to take $\lim_{y\to-\infty} G(x,y) = 0$. However, with our current definition of Y_t, this does not make sense, since $Y_t \ge 0$. So we need to extend our model by defining:

$$Y_t = Y(0) + \int_0^t S(u)du,$$

where $Y(0)$ *is a valued specified by the option contract, which can be negative or positive or zero.*

The payoff function G becomes

$$G(S(T), S_{ave}(T)) = G(S(T), \frac{1}{T}[Y(0) + \int_0^T S_u du]).$$

and the option value at time t is

$$v(t,x,y) = E\Big\{e^{-r(T-t)}G(S(T),Y(T))|S_t = x, Y_t = y\Big\},$$

as before. Adding a constant $Y(0)$ at time $t = 0$ clearly does not change the Markov property of V_t. Note that since $Y(0)$ can take any value (positive, negative, zero), y here also can take any value (positive, negative, zero).

In words, what we did here is just allow flexibility for dicussing our function $v(t,x,y)$ as $y \to -\infty$. But then arguing exactly as before, under the assumption that $\lim_{y\to-\infty} G(x,y) = 0$ we have

$$\lim_{y\to-\infty} v(t,x,y) \;=\; E\Big\{e^{-r(T-t)} \lim_{-y\to\infty} G(S(T), \frac{y + \int_t^T S_u du}{T})|S_t = x\Big\} = 0.$$

Then we have the following PDE for the Asian option, *assuming the condition* $\lim_{y\to-\infty} G(x,y) = 0$

$$\begin{aligned}
&-rv(t,x,y) + v_t(t,x,y) + v_x(t,x,y)rx + v_y(t,x,y)x + \frac{1}{2}v_{xx}(t,x,y)\sigma^2(t,x)x^2 = 0,\\
&0 < x < \infty, -\infty < y < \infty, 0 \le t < T;\\
&v(T,x,y) = G(x, \frac{y}{T});\\
&v(t,0,y) = e^{-r(T-t)}G(0, \frac{y}{T});\\
&\lim_{y\to\infty} v(t,x,y) = 0.
\end{aligned}$$

Note the change of domain for y on the first equation. Now y is defined on $(-\infty,\infty)$, not just $[0,\infty)$.

CHAPTER 8 American Option

8.1 Introduction

Let S_t have the Black-Scholes dynamics under a risk neutral measure

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dW_t \\ S_0 &= x. \end{aligned}$$

An American option purchased at time $t = 0$ with pay off function $g(x)$ and expiry T on an underlying stock S_t is a contract that gives the buyer the right, but not the obligation, to exercise the option at any time $t \in [0, T]$ and receive the pay off $g(S_t)$. If the option holder does not exercise at or before the expiry T then the option becomes worthless. We call $g(S_t)$ the *intrinsic value* of the option.

Example:
(i) American call option: $g(x) := (x - K)^+$;
(ii) American put option: $g(x) := (K - x)^+$;
(iii) Perpetual American put option: $g(x) := (K - x)^+$ and $T = \infty$.

Remark 8.1.1. *It is clear that if $g(S_t) < 0$ then the American option holder will not exercise the option at time t. Thus we can assume that $g(x) \geq 0$ for all x. Then $g(S_T) \geq 0$ and thus we can assume that the option is always exercised at or before time T.*

Remark 8.1.2. *We have the following simple, but important observation: Let V_t^A be the (no arbitrage) price of an American option and V_t^C the (no arbitrage) price of the corresponding European option purchased at time $t = 0$ with pay off function $g(x)$ and expiry T. That is $V_T^C = g(S_T)$ and the European option holder cannot exercise the option earlier than T. Then $V_t^A \geq V_t^C$. That is the price of an American option is always at least as expensive as its European counterpart. Note that this conclusion is model independent: we do not make any assumption on S_t.*

Reason: Suppose $V_t^C > V_t^A$. Then we take a long position (that is we buy 1 share) on the American option and a short position on the European option. Then we receive a positive amount equals $V_t^C - V_t^A$. At time T, we exercise the American option to close out our short position on the European option. This is an arbitrage opportunity. Therefore we must have $V_t^C \leq V_t^A$.

Remark 8.1.3. *Another important observation is that when $g(x) = (x-k)^+$, then $V_t^A = V_t^C$ for all t and thus the optimal exercise time for the American option is at the expiry T if the asset S_t* **does not pay dividend**. *This observation is also model-independent. You are asked to explore this in the extra-credit problem in homework 8. This conclusion does not extend to the case when S_t pays dividend.*

We now need a mathematical definition of the American option price.

8.2 Preliminary investigation

8.2.1 Exercise times

Exercise times on the time interval $[0, T]$

The first step in modeling American option is to formulate the notion of exercise time mathematically. Here we suppose that our current time is 0 and we are considering all the future possible exercise times for the option holder.

As usual, we assume the information available to investors as time progresses is encoded in a given filtration $\{\mathcal{F}(t);\, t \geq 0\}$. In general, the owner of an American option will decide when to exercise based on the current level of the price, its past history, and possibly other information is available about the economy and its past history (this depends on the construction of the filtration $\mathcal{F}_t$). It is assumed that investors cannot look into the future. So at the moment an investor decides to exercise, he or she can use only the information in the filtration up to that moment. Mathematically, this translates into the following assumption:

$$\textit{all exercise times are } \{\mathcal{F}(t);\, t \geq 0\}\textit{-stopping times.} \tag{8.1}$$

Most of the analysis of American options that is presented in this course will be made under the following additional assumptions:

$$\text{the price process } S \text{ is a Markov process;} \tag{8.2}$$

$$\{\mathcal{F}(t);\, t \geq 0\} \text{ is the filtration generated by } S \tag{8.3}$$

Assumption (8.3) says that our analysis of the American option in this lecture note will only deal with the case when the American option owner decides to exercise based on the current level of the price of the stock and its past history, but not on other additional information. The analysis we will do shows that, assuming (8.2) and (8.3), the best decision about whether to exercise or not at time t will use only the current value S_t of the price of the underlying. This is ultimately a consequence of the Markov property of S.

Exercise time on the time interval $[t, T]$

Consider now $t < T$. Here we want to look at all the possible exercise time beyond t for the option holder, *assuming the option has not been exercised before time* t. This is the same as entering into an American option type of contract at time t with payoff function g and expiry T.

Since S is Markov, the value of S_t contains all the information from the past relevant to the future when computing conditional expectations. Define

$$\mathcal{F}^{(t)}(u) \quad \text{be the } \sigma\text{-algebra generated by } S(v) \text{ for times } t \leq v \leq u.$$

This is the filtration of information generated by S for times after t. We will assume for any exercise time τ taking place at time t or later that

$$t \leq \tau \leq T; \text{ and,} \tag{8.4}$$

$$\tau \text{ is an } \{\mathcal{F}^{(t)}(u);\ u \geq t\}\text{-stopping time.} \tag{8.5}$$

The meaning of condition (8.5) is that for every $u \geq t$, the event $\{t \leq \tau \leq u\}$ must belong to $\mathcal{F}^{(t)}(u)$.

8.2.2 The mathematical definition of the value of an American option

Let $\{(\Omega, \mathcal{F}, P), \{S_t;\ t \geq 0\}$ be a risk-neutral price model. Let $\{\mathcal{F}(t);\ t \geq 0\}$ be the filtration generated by S and assume the risk free interest rate is the constant r.

Consider an American option that pays $g(S_t)$ if exercised at t and let T be its expiration date.

The following observations are important for us:

(i) $V_t \geq g(S_t)$.

Reason: If $V_t < g(S_t)$, one could buy the option and immediately exercise to realize a riskless positive profit $g(S_t) - V_t$.

(ii)Let τ be a given stopping time satisfying $P(\tau \leq T) = 1$. For example,

$$\tau = \inf\{t \geq 0 : S_t = L\} \wedge T,$$

where L is a positive constant. Consider a financial product that pays $g(S_\tau)$ at time τ (Note that this is not an American option - the payoff time (albeit being random) is specified in advance at time 0). According to risk-neutral pricing, the price V_0^τ of this product is

$$V_0^\tau = E\Big[e^{-r\tau} g(S_\tau)\Big].$$

Now consider an American option with payoff function $g(x)$ and expiry T. Let V_t denote the value of the option at time t. Then $V_0 \geq V_0^\tau$.

Reason: If $V_0 < V_0^\tau$, we can buy 1 share the American option and short 1 share of V^τ. We then receive a positive amount equals $V_0^\tau - V_0$ at time 0. At time τ, we simply excercise the American option to cover the payoff $g(S_\tau)$ from the financial product. This is an arbitrage opportunity.

(iii) It follows that V_0, the price of the American option, satisfies

$$V_0 \geq \sup \left\{ E\left[e^{-r\tau} g(S_\tau)\right];\ \tau \leq T, \tau \text{ is a stopping time} \right\}.$$

(iv) On the other hand, if the option holder plans to maximize her gain on the option, the only "strategy" available to her is to choose judiciously an optimal stopping time (since she cannot look into the future) to exercise the option. (For example, exercise the option when S_t has reached a low enough level $L < K$, if the option is American put). Therefore, one would not be willing to pay V_0 for the American option if she could not find an exercise time τ^* so that $V_0 = V_0^{\tau^*}$. That is

$$V_0 = E\left[e^{-r\tau^*} g(S(\tau^*))\right].$$

(v) It follows that

$$V_0 = \max \left\{ E\left[e^{-r\tau} g(S_\tau)\right];\ \tau \leq T, \tau \text{ is a stopping time} \right\}. \tag{8.6}$$

We shall take this formula as the *definition* of the value of the American option.

(vi) Actually, we have assumed in our argument of part (v) and (vi) that the maximum and the optimal exercise time τ^* exist. If they do not, we must use the supremum, and so the proper definition is

$$V(0) = \sup \left\{ E\left[e^{-r\tau} g(S_\tau)\right];\ \tau \leq T, \tau \text{ is a stopping time} \right\}. \tag{8.7}$$

(vii) The definition for the option value at time t is similar.

$$V_t = \sup \left\{ E\left[e^{-r(\tau-t)} g(S_\tau) \middle| \mathcal{F}_t\right];\ \tau \text{ satisfies (8.4) and (8.5).} \right\}. \tag{8.8}$$

(viii) Note that at a theoretical level, definition (8.8) tells us how to price and also when to exercise the American option. We say this is theoretical since at this point we do not know what V_t is. But assuming that we know V_t then the following principles apply.

At each time t, if $V_t > g(S_t)$ then continue to hold the option (since this implies there is a future exercise time $\tau > t$ that will give one a higher expected pay off than the immediate realization of $g(S_t)$).

If $V_t = g(S_t)$ then exercise the option since this is the best possible value one can realize at time t (among all other possible strategies that make one wait until a future time $\tau > t$ to exercise).

(ix) Define

$$v(t, x) = \sup \left\{ E\left[e^{-r(\tau - t)} g(S_\tau) \;\middle|\; S_t = x \right] ; \;\; \tau \text{ satisfies (8.4) and (8.5)} \right\} \tag{8.9}$$

From the perspective of time t this is the best, discounted, expected payoff the option can yield. Then by the Markov property of S_t:

$$V_t = v(t, S_t) \tag{8.10}$$

The function $v(t, x)$ is called the the *value function* of the option pricing problem. A stopping time τ^* for which

$$v(t, x) = E\left[e^{-r(\tau^* - t)} g(S(\tau^*)) \;\middle|\; S_t = x \right]$$

is called an *optimal exercise*, or *optimal stopping time*, for valuing the option starting at t.

8.2.3 Our goal

Item (ix) above characterizes the value of the American option ($v(t, S_t)$) and when to exercise (τ^*). Our goal for this note is to find an equation (a PDE) that $v(t, x)$ sastisfies and characterize τ^* (give a rule for τ^*). This will be accomplised in the sections on the perpetual put and American option with finite time of observation. Before that, we still need to discuss more about some properties of V_t.

8.3 The super-martingale property of V_t

8.3.1 Introduction

Definition 8.3.1. *Let X_t be a stochastic process and $\mathcal{F}(t)$ a filtration for X. We say that X_t is a* **super-martingale** *with respect to $\mathcal{F}(t)$ if for $s \leq t$*

$$E(X_t | \mathcal{F}_s) \leq X_s.$$

We say that X_t is a **sub-martingale** *with respect to $\mathcal{F}(t)$ if for $s \leq t$*

$$E(X_t|\mathcal{F}_s) \geq X_s.$$

Example: Let W_t be a Brownian motion and let

$$\begin{aligned} X_t^1 &= t + W_t; \\ X_t^2 &= -t + W_t. \end{aligned}$$

Then X_t^1 is a sub-martingale and X_t^2 is a super-martingale w.r.t. $\mathcal{F}_t^W$.

Remark 8.3.2. *Note that the definition of super and sub martingale allows for equality. Thus X_t is a martingale if and only if it is both a super-martingale and a sub-martingale.*

Remark 8.3.3. *Note that a probability measure is also implicitly involved in the definition of super-martingale and sub-martingale. In this note, we will always consider the context under the risk neutral measure (so that $e^{-rt}S_t$ is a martingale under our default measure). However, observe from the above example, that a super-martingale may no longer remain a super martingale under a change of measure (say we change the drift of the Brownian motion to 2t, for example, then X^2 will be come a sub-martingale under the new measure).*

8.3.2 Main results

We need to introduce some new notation. For a fixed $t \in [0, T]$, let $\mathcal{T}_{[t,T]}$ be the set of stopping times that satisfy (8.4) and (8.5). That is, $\mathcal{T}_{[t,T]}$ is a set of stopping times taking values in $[t, T]$ and adapted to the filtration generated by S_u for time $t \leq u \leq T$. Then observe that for $s \leq t$

$$\mathcal{T}_{[t,T]} \subseteq \mathcal{T}_{[s,T]}.$$

As time goes on (when t increases), the set of strategies available for the option holder (the stopping times) decreases. This intuitively suggests that on average (that is when taking conditional expectation), V_t should be decreasing (because if $A \subseteq B$ then $\sup A \leq \sup B$). That is V_t is a super-martingale.

The rigorous proof for this result is not easy. However, if we accept the fact that for any time t, there will exists $\tau^* \in \mathcal{T}_{[t,T]}$ such that

$$V_t = E\left[e^{-r(\tau^*-t)}g(S(\tau^*))\middle|\mathcal{F}_t\right],$$

then we can show that V_t is a super-martingale. (Later on we will show by an independent result that such a τ^* always exists. This discussion is a motivation for the rigorous results that will follow).

Theorem 8.3.4. *Assuming the optimal stopping time exists for each V_t, then the discounted American option value $e^{-rt}V_t$ is a super-martingale w.r.t $\mathcal{F}_t$.*

Proof. Let $s \le t$. Let $\tau^* \in \mathcal{T}_{[t,T]}$ such that

$$V_t = E\Big[e^{-r(\tau^*-t)}g(S(\tau^*))\Big|\mathcal{F}_t\Big],$$

Then

$$\begin{aligned} E(e^{-rt}V_t|\mathcal{F}_s) &= E\Big\{E(e^{-r\tau^*}g(S_{\tau^*})|\mathcal{F}_t)\Big|\mathcal{F}_s\Big\} \\ &\le \sup_{\tau\in\mathcal{T}_{[s,T]}} E\Big\{E(e^{-r\tau}g(S_\tau)|\mathcal{F}_t)\Big|\mathcal{F}_s\Big\} \\ &= \sup_{\tau\in\mathcal{T}_{[s,T]}} E(e^{-r\tau}g(S_\tau)\Big|\mathcal{F}_s) \\ &= e^{-rs}V_s, \end{aligned}$$

where the inequality follows from the fact that if $\tau^* \in \mathcal{T}_{[t,T]}$ then $\tau^* \in \mathcal{T}_{[s,T]}$.

Since $V_t \ge g(S_t)$, we say $e^{-rt}V_t$ is a super-martingale dominating $e^{-rt}g(S_t)$. The crucial fact about $e^{-rt}V_t$ is that it is also *the smallest super-martingale dominating* $e^{-rt}g(S_t)$ in the following sense:

Theorem 8.3.5. *Let X_t be a $\mathcal{F}_t$ super-martingale dominating $e^{-rt}g(S_t)$, that is $X_t \in \mathcal{F}_t$ a super-martingale and $X_t \ge e^{-rt}g(S_t)$ for all t. Then for all t, with probability 1:*

$$X_t \ge e^{-rt}V_t.$$

Proof. Let X_t be a $\mathcal{F}_t$ super-martingale dominating $e^{-rt}g(S_t)$. Then it also follows that $X_\tau \ge e^{-r\tau}g(S_\tau)$ for all stopping time τ. We have,

$$\begin{aligned} e^{-rt}V_t &= \sup_{\tau\in\mathcal{T}_{[t,T]}} E(e^{-r\tau}g(S_\tau)\Big|\mathcal{F}_t) \\ &\le \sup_{\tau\in\mathcal{T}_{[t,T]}} E(X_\tau\Big|\mathcal{F}_t) \\ &\le \sup_{\tau\in\mathcal{T}_{[t,T]}} X_t. \\ &= X_t \end{aligned}$$

We have used the optional sampling theorem in the inequality

$$E(X_\tau\Big|\mathcal{F}_t) \le X_t,$$

which basically says the super-martingale property of X_t also applies when we use a bounded stopping time $\tau \ge t$.

8.3.3 Implication

The result that $e^{-rt}V_t$ is the smallest super-martingale dominating $e^{-rt}g(S_t)$ is the key observation to solve the optimal stopping problem. To illustrate the use of the idea, we'll consider the simple senario where everything is deterministic. Then V_t being a super-martingale is just being a decreasing (i.e. non-increasing) function in t.

Example 1: Let $T = 1$ and $g(t) = t$ on $[0, 1]$. What is the smallest decreasing function dominating $g(t)$ on $[0, 1]$? You'll see quickly that it's the constant function $V_t = 1$ on $[0, 1]$.

Example 2: Let $T = 2$ and $g(t) = t$ on $[0, 1]$ and $2 - t$ on $[1, 2]$. What is the smallest decreasing function dominating $g(t)$ on $[0, 1]$? You'll see quickly that it's the function $V_t = 1$ on $[0, 1]$ and $V_t = 2 - t$ on $[1, 2]$.

So what's the observation here? In both cases, V_t is **always a constant** up to the first time it hits $g(t)$. Moreover, we can find out the value of V_0 from the value of $V(t^*)$, where t^* is the point that V_t meets $g(t)$.

Going back our stochastic setting. Then it is not hard to guess that being a constant in the deterministic setting "corresponds" to being a martingale in the stochastic setting. That is V_t should be a martingale up to the first time it hits $g(S_t)$. Moreover, the value of V_0 can be deduced by taking expectation of $E(V_{\tau^*})$, where τ^* is the first time V_t hits $g(S_t)$. All of this argument is intuitive of course, but it serves to illustrate the idea and will be made rigorous in the next section.

Example 3: A martingale intuition on why American call option price equals to European call option price:

The pay-off function in American call option is $g(S_t) = (S_t - K)^+$. We argue that it is "increasing" in the sense of taking the discounted price under the conditional risk neutral measure. In other words, $e^{-rt}(S_t - K)^+$ is a sub-martingale. That is for all $s < t$

$$E(e^{-rt}(S_t - K)^+|\mathcal{F}_s) \geq e^{-rs}(S_s - K)^+.$$

This is equivalent to showing

$$E(e^{-r(t-s)}(S_t - K)^+|\mathcal{F}_s) \geq (S_s - K)^+.$$

But we have, by Jensen's inequality

$$\begin{aligned} E(e^{-r(t-s)}(S_t - K)^+|\mathcal{F}_s) &\geq E\Big((e^{-r(t-s)}S_t - e^{-r(t-s)}K)^+|\mathcal{F}_s\Big) \\ &\geq \Big(E(e^{-r(t-s)}S_t|\mathcal{F}_s) - e^{-r(t-s)}K\Big)^+ \\ &\geq \Big(S_s - e^{-r(t-s)}K\Big)^+ \\ &\geq (S_s - K)^+. \end{aligned}$$

Thus, the smallest super-martingale $e^{-rt}V_t^A$ dominating $g(S_t)$, by the intuition of Example 1, must be a martingale that satisfies $V_T^A = g(S_T)$. But this is the same as saying the American call option price equals to the European call option price.

8.4 The American perpetual put

8.4.1 The value function

The perpetual put is defined by the payoff function $(K - x)^+$ and the expiration date $T = \infty$. In this case, exercise at time ∞ represents no exercise. To include no exercise in the simplest manner, we make the convention that

$$e^{-r\infty}(K - S(\infty))^+ = 0$$

This convention is consistent with taking a limit as $t \to \infty$; indeed

$$\lim_{t\to\infty} e^{-rt}(K - S_t)^+ = 0, \tag{8.11}$$

because $0 \le (K - x)^+ \le K$ for all $x \ge 0$.

For the rest of this discussion, S is the Black-Scholes price with risk free interest rate r and volatility σ. As usual P denotes the risk-neutral measure.

Define $v(t, x)$ as in (8.9), except now $T = \infty$ and τ is allowed to take the value ∞. Denote $v(0, x)$ by $v(x)$: thus,

$$v(x) = \sup\left\{ E\left[e^{-r\tau}(K - S_\tau)^+ \;\middle|\; S(0) = x \right];\, \tau \text{ satisfies (8.4) and (8.5) with } T = \infty. \right\}$$

Because the expiration date is infinite and the volatility and risk free rate are constant in time, all times should look the same for the perpetual put. That is, if today $S(0) = 10$ and the price of the perpetual put is 5, it should be 5 whenever the price is 10 in the future.

More precisely, observe that,

(i) $S_u = S(0)\exp\{\sigma W(u) + (r - \frac{\sigma^2}{2})u\}$ where $\{W(u); u \ge 0\}$ is a Brownian motion independent of $S(0)$, and

(ii) $S(t+u) = S_t \exp\{\sigma(W(t+u) - W(t)) + (r - \frac{\sigma^2}{2})u\}$, where, for fixed t, $\{W(t+u) - W(t);\ u \ge 0\}$ is a Brownian motion independent of S_t.

Therefore, the *process* $\{S(t+u);\ u \ge 0\}$ conditioned on $S_t = x$ is **identical in distribution** to the process $\{S_u;\ u \ge 0\}$ conditioned $S(0) = x$.

Since the value $v(t, x)$ *depends only on the distribution of the price process forward in time*, conditioned on $S_t = x$, $v(t, x)$ is independent of t. This proves:

Lemma 1. *For the perpetual put,* $v(t, x) = v(x)$ *for all* $t \ge 0$.

8.4.2 Martingale characterization

The next theorem is a rigorous statement of the intuitively derived martingale characterization of the price introduced in section 3 above.

Theorem 8.4.1. *(Martingale sufficient conditions for the value function.) Assume* $u(x)$*,* $x \geq 0$ *satisfies the following conditions:*

(a) $u(x) \geq (K - x)^+$ *for all* $x \geq 0$*;*

(b) u *is bounded—there is a constant* $M < \infty$ *such that* $u(x) \leq M$ *for all* $x \geq 0$*;*

(c) $e^{-rt}u(S_t)$ *is a* **supermartingale** *given any initial condition* $S(0) = x$*;*

(d) if τ^* *is the first time that* $u(S_t) = (K - S_t)^+$*, then* $e^{-r(\tau^* \wedge t)}u(S_{\tau^* \wedge t})$ *is a* **martingale** *given any initial condition* $S(0) = x$*;*

Then $u(x) = v(x)$*, the value function for the perpetual put, and* $\tau*$ *is the optimal exercise time.*

Note: We will be able to find such a u, so, in fact, conditions (a)—(d) uniquely characterize v.

Proof. To prove equality of u and v we will first prove $u(x) \geq v(x)$ for all x, and then prove $u(x) \leq v(x)$.

(i) $u(x) \geq v(x)$:

It suffices to show that

$$u(x) \geq E\left[e^{-r\tau}(K - S_\tau)^+ \;\middle|\; S(0) = x\right] \quad \text{for any stopping time } \tau. \tag{8.12}$$

Indeed, (8.12) would imply that

$$u(x) \geq \sup\left\{E\left[e^{-r\tau}(K - S_\tau)^+ \;\middle|\; S(0) = x\right];\ \tau \text{ satisfies (8.4) and (8.5) with } T = \infty\right\} = v(x).$$

So let τ be any stopping time and assume $S(0) = x$. We have,

$$\begin{aligned} u(x) &= e^{-r(\tau \wedge t)}u(S_{\tau \wedge t})\,\Big|_{t=0} \geq E\left[e^{-r(\tau \wedge t)}u(S_{\tau \wedge t}) \;\middle|\; S(0)=x\right] \\ &\geq E\left[e^{-r(\tau \wedge t)}\Big(K - S_{\tau \wedge t}\Big)^+ \;\middle|\; S(0)=x\right], \end{aligned}$$

where the first inequality follows from assumption (c), $e^{-rt}u(S_t)$ is a supermartingale, and by the optional stopping theorem for supermartingales, $e^{-r(\tau\wedge t)}u(S_{\tau\wedge t})$ is also a supermartingale and supermartingales decreasing in expectation. The second inequality follows from assumption (a): $u(S_{\tau\wedge t}) \geq (K - S_{\tau\wedge t})^+$.

Because $(K - S_{\tau\wedge t})^+$ is bounded above by K and $\lim_{t\to\infty} e^{-rt}(K - S_t)^+ = 0$, limits and expectation can be interchanged in the above equation; this is a consequence of the dominated convergence theorem and the convention that the value of the option at $t = \infty$ is 0. Therefore,

$$u(x) \geq E\left[\lim_{t\to\infty} e^{-r(\tau\wedge t)}\left(K - S_{\tau\wedge t}\right)^+ \;\middle|\; S(0)=x\right] = E\left[e^{-r\tau}\left(K - S_\tau\right)^+ \;\middle|\; S(0)=x\right].$$

This proves (8.12) and so finishes the proof that $u(x) \geq v(x)$.

(ii) $u(x) \leq v(x)$:

We will show

$$u(x) = E\left[e^{-r\tau^*}(K - S(\tau^*))^+ \;\middle|\; S(0) = x\right] \tag{8.13}$$

where $\tau*$ is defined as in part (d) of the Theorem statement. The value function $v(x)$ is the maximum discounted expected payoff and so it is certainly greater than or equal than the right-hand side of (8.13). Hence $u(x) \leq v(x)$ follows from (8.13).

To prove 8.13 we need only repeat the previous calculation with $\tau*$ replacing τ. But now since $e^{-r(\tau*\wedge t)}u(S_{\tau^*\wedge t})$ is a martingale, all inequalities are replaced by equalities and the result is (8.13). This completes the proof. ◇

8.4.3 PDE characterization

Theorem 1 assumed only that the price process was a path continuous Markov process. The following theorem provides analytic conditions on a function u in order that it satisfy the martingale conditions of Theorem 8.4.1, in the special case that S follows that Black-Scholes price model. Recall that under the risk-neutral measure,

$$dS_t = rS_t\, dt + \sigma S_t\, dW(t).$$

The theorem provides an effective way of finding the value function or at least obtaining differential equations for the value function. It can easily be extended to other price models.

Theorem 8.4.2. *(Hamilton-Jacobi-Bellman equations for the value function.) Assume* $u(x)$, $x \geq 0$, *satisfies the following conditions:*

(a') $u(x) \geq (K-x)^+$ *for all* $x \geq 0$*;*

(b') u *is bounded—there is a constant* $M < \infty$ *such that* $u(x) \leq M$ *for all* $x \geq 0$*;*

(c') u *and* u' *are continuous and* u'' *is continuous except possibly at a finite number of points, where it has jump discontinuities, and it satisfies*

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) \geq 0; \tag{8.14}$$

(d') *on the set where* $u(x) > (K-x)^+$ *(the continuation set),*

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) = 0. \tag{8.15}$$

Then $u(x) = v(x)$*, the value function for the perpetual put, and the optimal exercise time is*

$$\tau^* := \inf\Big\{t \geq 0 : u(S_t) = (K - S_t)^+\Big\}.$$

Condition (a') and the two equations in (c') and (d') are all contained in the following equation:

$$\min\{u(x) - (K-x)^+,\, ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x)\} = 0. \tag{8.16}$$

Remark 8.4.3. *The condition that* u'' *is continuous except possibly at a finite number of points may be strange at the beginning, because we're used to assuming that* u *has continuous second derivatives. This is because for this type of PDE, making the assumption that* $C^{1,2}$ *is no longer appropriate. To get a clue on why, note that this PDE is* **nonlinear** *in* u *(equation (8.16)) while the PDEs we considered so far (in chapter 7 and 11 of Shreve) is* **linear** *in* u*. This should lead you to expect that the type of regularity property we are expecting for this type of PDE is weaker than the type of regularity we expect for the PDE we encountered in chapter 7 of Shreve. Hence assumption* c'*.*

Remark 8.4.4. *The Ito's formula we have learned so far assumes that* $u \in C^{1,2}$*. We want to apply Ito's formula to the function* u *in the above theorem. So clearly the question is can we still do so? The answer is essentially yes, via an extension of Ito's rule. We will present it at the end of this section, to preserve the flow of discussion.*

Proof: We need to check that the conditions (a), (b), (c), and (d) of 8.4.1 are satisfied by u. Conditions (a) and (b) are automatic from (a') and (b'). For conditions (c) and (d), we use Itô's rule and (c') and (d').

(i) $e^{-rt}u(S_t)$ is a *supermartingale* given any initial condition $S(0) = x$:

Since u'' is assumed continuous except at a finite number of points where it is undefined and has a jump discontinuity, Itô's rule applies and

$$\begin{aligned} e^{-rt}u(S_t) &= u(S(0)) + \int_0^t e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \\ &\quad + \int_0^t \sigma e^{-rs}u'(S_s)\,dW(s). \end{aligned} \tag{8.17}$$

Let $t_1 < t_2$. The stochastic integral is a martingale so

$$E\left[\int_0^{t_2} e^{-rs}u'(S_s)\,dW(s) \;\Big|\; \mathcal{F}(t_1)\right] = \int_0^{t_1} e^{-rs}u'(S_s)\,dW(s)$$

On the other hand, condition (c') implies that the integrand of the ds integral is always non-positive. Hence,

$$\begin{aligned} &E\left[\int_0^{t_2} e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \;\Big|\; \mathcal{F}(t_1)\right] \\ &= \int_0^{t_1} e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \\ &\qquad + E\left[\int_{t_1}^{t_2} e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \;\Big|\; \mathcal{F}(t_1)\right] \\ &\le \int_0^{t_1} e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \end{aligned}$$

Putting these calculations together,

$$\begin{aligned} &E\left[e^{-rt_2}u(S(t_2)) \;\Big|\; \mathcal{F}(t_1)\right] \\ &\qquad \le u(0) + \int_0^{t_1} e^{-rs}\{-ru(S_s) + rS_s u'(S_s) + \frac{1}{2}\sigma^2 S^2(s)u''(S_s)\}\,ds \\ &\qquad\qquad + \int_0^{t_1} \sigma e^{-rs}u'(S_s)\,dW(s) \\ &\qquad = e^{-rt_1}u(S(t_1)) \end{aligned}$$

This proves that $e^{-rt}u(S_t)$ is a supermartingale.

(ii) $e^{-r(t\wedge\tau^*)}u(S_{t\wedge\tau^*})$ is a *martingale* given any initial condition $S(0) = x$:

Repeat the above calculation but only up to the stopping time τ^*, that is the first time S_t hits the set $\{x;\, v(x) = (K-x)^+\}$. Then for $t < \tau^*, u(S_t) > (K - S_t)^+$ by definition of τ^* and (a').

So by (d'), for $s < \tau^*$, $-ru(S_s) + rS_s u'(s) + (1/2)\sigma^2 S^2(s) u''(S_s) = 0$. Thus from (8.17), but carried out up to $\tau*$ only,

$$e^{-r(\tau^* \wedge t)} u(S_{\tau^* \wedge t}) = u(0) + \int_0^{\tau^* \wedge t} e^{-rs} u'(S_s) \sigma S_s \, dW(s),$$

and this is a martingale, thereby proving condition (d) of Theorem 8.4.1. $\diamond$

8.4.4 Finding the value function analytically

(i) Guess for the form of u:

To find the value function, we attempt to solve equation (16), or equivalently (14) and (15). The technique is to guess the form of the solution. For the perpetual put one expects that the exercise (stopping) region will have the form $0 \le x \le L^*$ where $0 < L^* < K$; equivalently the continuation region has the form $L^* < x < \infty$. It is intuitively clear that if it is optimal to continue when the price is at a level ℓ, it should also be optimal to continue at all higher prices, because higher the price, the lower is the payoff. Also, one will never exercise when the price is higher than K.

Therefore assume the optimal exercise region is $0 \le x \le L^* < K$. By definition of the optimal exercise region,

$$u(x) = (K - x)^+, \quad 0 \le x \le L^*.$$

According to (8.16) or (8.15), on $L^* < x < \infty$, $u(x)$ must solve

$$\frac{1}{2}\sigma^2 x^2 u''(x) + rxu'(x) - ru(x) = 0.$$

This is a homogeneous differential equation of Euler type. Observe the equation, since there is x^2 in front of the 2nd derivative and x in front of the 1st derivative, one can guess the general form of the solution if Ax^p, where A is a constant and p is to be determined. Plug this form in to the equation above, we found the equation for p is

$$\frac{1}{2}\sigma^2 p^2 + (r - \frac{1}{2}\sigma^2)p - r = 0.$$

It is easily checked that $p = -\frac{2r}{\sigma^2}$ and $p = 1$ are the solutions to the above quadratic equation. Thus the general solution of the ODE for $u(x)$ has the form,

$$u(x) = Ax^{-2r/\sigma^2} + Bx,$$

where A and B can be arbitrary constants. The condition that u be bounded requires $B = 0$. Thus,

$$u(x) = \begin{cases} K - x, & \text{if } 0 \leq x \leq L^*; \\ Ax^{-2r/\sigma^2}, & \text{if } x > L^*. \end{cases}$$

(ii) Impose conditions so that u is “smooth":

For this function u to be continuously differentiable, $(K - x)^+$ and Ax^{-2r/σ^2} and their first derivatives must be equal to each other at L^*. First, by matching the expression for u from the left and right of L^*, we get

$$K - L^* = u(L^*-) = u(L^*+) = A(L^*)^{-2r/\sigma^2}. \tag{8.18}$$

Second, noting that for $x < L^*$, $u'(x) = -1$ and for $x > L^*$, $u'(x) = -\frac{2r}{\sigma^2}Ax^{-2r/\sigma^2-1}$,

$$-1 = u'(L^*-) = u(L^*+) = -\frac{2r}{\sigma^2}A(L^*)^{-2r/\sigma^2-1}. \tag{8.19}$$

These are called the equations of *smooth fit.* They can be solved uniquely for A and L^*:

$$L^* = \frac{2r}{2r+\sigma^2}K \quad \text{and} \quad A = \frac{\sigma^2}{2r}(L^*)^{+2r/\sigma^2+1}.$$

Notice that $L^* < K$ as desired.

(iii) Verify that the function u we came up with is the actual solution:
To finish the proof, we must verify that with L^* and A so defined, equation (8.14) is true on $0 < x < L^*$, that is

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2x^2u''(x) \geq 0;$$

on $0 < x < L^*$ **and** $u(x) \geq (K - x)^+$ for $L^* < x < \infty$.

(iii - a) Verifying equation (8.14) is true on $0 < x < L^*$:

It is only necessary to verify (8.14) on $0 < x < L^*$, since it is true by construction on $x > L^*$, that is we found u on $L^* < x < \infty$ by solving

$$\frac{1}{2}\sigma^2x^2u''(x) + rxu'(x) - ru(x) = 0.$$

Since $u(x) = K - x$ on $0 < x < L^*$, it follows by direct calculation that that

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2x^2u''(x) = r(K - x) + rx = rK > 0,$$

which proves (14).

(iii - b) Verifying that $u \geq (K-x)^+$ on $L^* < x$:

By construction, on $0 < x \leq L^*$, $u(x) \geq (K-x)^+$ (since $L^* < K$). So we only need to verify $u \geq (K-x)^+$ on $L^* < x$.

Since $A > 0$, it follows that $u(x) > 0$ for all x. By taking two derivatives of Ax^{-2r/σ^2}, one see easily that it is strictly convex up. A and L^* are chosen so that $u'(L^*) = -1$ and $u(L^*) = K - L^*$. Hence, by convexity, $u'(x) > u'(L^*) = -1$ for all $x > L^*$. Thus if we set $g(x) = u(x) - (K-x)$, it follows that $g(L^*) = 0$ and that $g'(x) > 0$ for $x > L^*$. This means that g is increasing on $L^* < x < \infty$ and hence $g(x) = u(x) - (K-x) > 0$ on $L^* < x < \infty$. Thus we have shown both that $u(x) > 0$ **and** that $u(x) > K - x$ on $L^* < x < \infty$, which means that $u(x) > (K-x)^+$ on $L^* < x < \infty$.

In conclusion, we have identified the value function, and, more importantly, we have found that the optimal time to exercise the perpetual American put is when the price first hits the region $0 < x < L^*$, where $L^* = 2rK/(2r+\sigma^2)$.

8.4.5 An extension of Itô's rule

Previously in this course, Itô's rule was stated for $f(Y(t))$, where Y is an Itô process and f is twice continuously differentiable. Actually, Itô's rule will still work so long as $f(x)$ and $f'(x)$ are continuous and $f''(x)$ is defined and continuous everywhere except possibley at a finite number of points, where it has jump discontinuities. As an example, consider the function

$$f(x) = \begin{cases} x^2, & \text{if } x \geq 0; \\ -x^2, & \text{if } x < 0. \end{cases}$$

For this function, $f'(x) = 2|x|$, and $f''(x) = 2$ if $x > 0$ and $f''(x) = -2$ if $x < 0$. Thus f'' is not defined at $x = 0$, where it has a jump discontinuity. If W is a Brownian motion, then one can show directly that

$$\begin{aligned} f(W(t)) &= f(0) + \int_0^t f'(W(s))\,dW(s) + \frac{1}{2}\int_0^t f''(W(s))\mathbf{1}_{\{W(s)\neq 0\}}\,ds \\ &= f(0) + \int_0^t f'(W(s))\,dW(s) + \int_0^t \mathbf{1}_{\{W(s)>0\}}\,ds + \int_0^t -\mathbf{1}_{\{W(s)>0\}}\,ds. \end{aligned}$$

The reason this is true is that Brownian motion spends zero total time at $x = 0$, in the sense that

$$\int_0^t \mathbf{1}_{\{W(u)=0\}}\,du = 0, \quad \text{for all } t > 0, \text{ with probability one.}$$

Therefore, in the approximation arguments that lead to Itô's rule, the failure of f'' to exist at $x = 0$ is not "seen."

More generally, if $A = \{a_1, \dots, a_n\}$ is any finite collection of points and $X(t) = X(0) + \int_0^t \alpha(u)\,du + \int_0^t \beta(u)\,dW(u)$ and $\beta(t) > 0$ for all t, with probability one, then

$$\int_0^t \mathbf{1}_{\{X(u)\in A\}}\,du = 0, \quad \text{for all } t > 0\text{, with probability one.} \tag{8.20}$$

and Itô's rule will be valid for any f for which f and f' are continuous and f'' is piecewise continuous with jump discontinuities at $a_1, \dots, a_n$. That is,

$$f(X(t)) = f(0) + \int_0^t f'(X(s))\,dW(s) + \frac{1}{2}\int_0^t f''(X(s))\mathbf{1}_{\{X(s)\notin A\}}\,ds.$$

For simplicity, we will just write the last term as $\int_0^t f''(X(s))\,ds$; this is reasonable considering the fact (8.20). It is just necessary to understand that the second derivative can fail to exist at the points of A and we really mean $\int_0^t f''(X(s))\mathbf{1}_{\{X(s)\notin A\}}\,ds$.

In a similar way, Itô's rule extends to functions $f(t,x)$ such that $f(t,x)$, $f_t(t,x)$ and $f_x(t,x)$ are continuous and $f_{xx}(t,x)$ is continuous except possibly along a finite number of curves $x = a_i(t)$, where its value can jump.

8.4.6 Connection with optimal stopping problem

This section can be seen as giving the answer to the problem: Given that $S_0 = x$, finding the value function

$$V_0 = \max\left\{E\left[e^{-r\tau}(K - S_\tau)^+\right];\ \tau \text{ satisfies (8.4) and (8.5) with } T = \infty.\right\}$$

and characterize the optimal stopping time τ^* such that

$$V_0 = E\left[e^{-r\tau^*}(K - S_\tau^*)^+\right].$$

Then the answer is as followed. Solve for the value function $u(x)$ as given in Theorem 2. Define the continuation set as

$$\mathcal{C} = \{x; u(x) > (K - x)^+\}$$

and it is so-called because one should continue (not exercise) so long as $S_t \in \mathcal{C}$. The stopping set is the complement of $\mathcal{C}$:

$$\mathcal{S} = \{x; u(x) = (K - x)^+\}.$$

When $S_t \in \mathcal{S}$, it is optimal to exercise at t. We also call $\mathcal{S}$ the optimal exercise region.

Then $V_0 = u(S_0) = u(x)$ and the optimal time to exercise τ^* is

$$\tau^* = \min \{u \geq 0;\ (u, S_u) \in \mathcal{S}\} \wedge T \tag{8.21}$$

Note also that $e^{-rt}V_t = e^{-rt}v(S_t)$ is a martingale before time τ^* (that is $e^{-rt\wedge\tau^*}V_{t\wedge\tau^*}$ is a martingale) and a super-martingale in general (that is without stopping at τ^*). This follows from the following observation: if

$$dS_t = rS_t dt + \sigma S_t dW t$$

and u satisfies,

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) = 0,$$

then $u(S_t)$ is a martingale. Moreover, if

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) \geq 0,$$

then $u(S_t)$ is a super martingale. The proof is by Ito's formula.

Combined this with the result of Theorem 2, that is on $u(x) > (K - x)^+$

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) = 0,$$

and

$$ru(x) - rxu'(x) - \frac{1}{2}\sigma^2 x^2 u''(x) \geq 0,$$

it is easy to see that $u(S_t)$ is a martingale up to the first time $u(S_t)$ hits $(K - S_t)^+$ and generally a super-martingale.

8.5 American options with finite time of expiration

The analysis that we carried out can be generalized to American options with finite expiration.

As usual, assume the price process of the underlying is Black-Scholes with risk-free rate r and volatility σ^2. Suppose the option payoff is $g(x)$, and assume g is bounded. This assumption is made so that the results we state are rigorously true; extensions are possible that drop the boundedness condition, but then extra conditions are needed.

Let T be the time of expiration and let $v(t, x)$, $t \leq T$, be the value function as defined above in equation (8.9).

Theorems 1 and 2 have easy generalizations that require little more than inserting extra dependence on t.

8.5.1 Martingale characterization

Theorem 8.5.1. *(Martingale sufficient conditions for the value function.) Assume* $u(t, x)$*,* $x \geq 0$ *satisfies the following conditions:*

(a) $u(t, x) \geq g(x)$ *for all* $x \geq 0$*,* $0 \leq t \leq T$*;*

(b) u *is bounded—there is a constant* $M < \infty$ *such that* $u(t, x) \leq M$ *for all* $x \geq 0$*,* $0 \leq t \leq T$*;*

(c) $e^{-rt}u(t, S_t)$ *is a supermartingale given any initial condition* $S(0) = x$*;*

(d) *if* τ^* *is the first time that* $u(t, S_t) = g(S_t)$*, then* $e^{-r(\tau^* \wedge t)}u(\tau^* \wedge t, S_{\tau^* \wedge t})$ *is a martingale given any initial condition* $S(0) = x$*;*

Then $u(t, x) = v(t, x)$*, the value function for the value function for the American put with pay off* $g(S_t)$ *and expiry* T*, and* τ^* *is the optimal exercise time.*

8.5.2 Equation for the value function

Theorem 8.5.2. *(Hamilton-Jacobi-Bellman equations for the value function.) Assume* $u(t, x)$*,* $x \geq 0$ *satisfies the following conditions:*

(a') $u(t, x) \geq (K - x)^+$ *for all* $x \geq 0$*,* $0 \leq t \leq T$*;*

(b') $u(t, x)$ *is bounded on the set* $x \geq 0$*,* $0 \leq t \leq T$*;*

(c') $u(t, x)$ *and* $u_t(t, x)$ *and* $u_x(t, x)$ *are continuous and* $u_{xx}(t, x)$ *exists and is continuous except possibly along a finite number of curves* $x = a_i(t)$*, where it has jump discontinuities, and* u *satisfies*

$$ru(t, x) - u_t(t, x) - rxu_x(t, x) - \frac{1}{2}\sigma^2 x^2 u_{xx}(t, x) \geq 0;$$

(d') *on the set where* $u(t, x) > g(x)$ *(the continuation set)*

$$ru(t, x) - u_t(t, x) - rxu_x(t, x) - \frac{1}{2}\sigma^2 x^2 u_{xx}(t, x) = 0.$$

Then $u(t, x) = v(t, x)$*, the value function for the American put with pay off* $g(S_t)$ *and expiry* T*, and the optimal exercise time is the first time* τ^* *that* S_t *hits the set* $\{x;\, v(t, x) = g(x)\}$*.*

The proofs of these statements mimic closely the proofs written out above for the perpetual put.

8.5.3 Characterization of the optimal exercise region

For the American put, $g(x) = (K - x)^+$. In this case, explicit formulae for the value function and for the exercise region are not known when T is finite. However, the methodology of Theorems 3 and 4, which can be generalized greatly, is very important. The equations presented in parts (c') and (d') of Theorem 8.5.2 can be solved numerically to get prices.

Even though explicit solutions for the value function of the American put are not known, the following can be proved.

The optimal exercise region has the form

$$\{(t, x);\, 0 \le x \le L^*(T - t), \text{for } x \ge 0,\, 0 \le t \le T\},$$

where $L^*(u)$, $u \ge 0$, is a function defined independently of T satisfying,

(i) $L^*(0) = K$;

(ii) $L^*(u)$ is decreasing as u increases;

(iii) $L^*(u) > \frac{2r}{2r+\sigma^2}K$ for all $u \ge 0$.

All these conditions make good intuitive sense. The parameter u in $L^*(u)$ represents time until expiration. Thus, $L^*(0)$ is the boundary of the exercise region when $t = T$, and at this time one exercises only if the the price is less than K. As u increases, that is, as we go back further and further in time from expiration, the value of the option at all price levels should either increase or stay the same, because more time means a greater range of exercise opportunities. But if the price increases the optimal exercise region shrinks. Finally, as $u \to +\infty$, $L^*(u)$ should decrease to the optimal exercise boundary of the perpetual put, which was shown above to be $\frac{2r}{2r+\sigma^2}K$.

8.6 Miscellaneous - A related interview question

Can we explain that the American call option value V_t^{AC} is equal to the European call option value V_t^{EC} but NOT using the martingale theory that we developed in this chapter? (After all not everyone is familiar with optimal stopping theory). The answer is yes, via the Put-Call parity relation. The detail is as followed: by Put-Call parity

$$V_t^{EC} - V_t^{EP} = S_t - Ke^{-r(T-t)}.$$

Thus

$$V_t^{EC} - (S_t - K) = V_t^{EP} + (K - Ke^{-r(T-t)}).$$

The LHS is the difference between the European option value and the pay off we would have received from the American option if we exercised at time t (if S_t is at least K).

The RHS, on the other hand, is always positive if $t < T$. Thus the exercise value is always strictly less than V_t^{EC}, which in turn is less than or equal to V_t^{AC}, the value of the American call option. Thus it is not optimal to excerise for $t < T$. But if we exercise the American call option at time T then it is just like an European option.

The RHS of the above equation also can be interpreted as the “insurance" for the event $S_T < K$ through the put option V_t^{EC} and the time “premium" for early exercise $K - Ke^{-r(T-t)}$. This premium converges to 0 as $t \to T$. This also explains intuitively why the optimal exercise time is at T for the American call option.

CHAPTER 9 Change of numéraire

Main question:

Price a product that pays $V_T = Q_T$ at time T. Is there a domestic risk neutral measure? If so what is the dynamics of Q_t under that measure?

Price a risky asset quoted in foreign currency: $V_T = (\frac{S_T}{Q_T} - K)^+$, K in foreign currency. This is not convenient to multiply through with Q_T. More convenient to work with foreign risk neutral measure.

Put-call duality: If we hold currency in dollar and suspect Euro to go up we would want a call on Euro denominated in dollars. If we hold currency in Euro then we suspect the dollar to go down. Then we would want a put on dollar denominated in Euro. These two options should be equivalent.

9.1 Introduction

9.1.1 Another look at the Black-Scholes risk neutral model

Let $r > 0$ be the constant risk free rate. So far, we've considered the following Black-Scholes model of a stock:

$$dS_t = \alpha S_t dt + \sigma S_t dW_t,$$

where α is a constant and W_t a Brownian motion.

To price any financial derivative based on S, the first question we have to answer is: what is the risk neutral probability measure? In other words, we want to find a probability measure $\tilde{P}$ such that $e^{-rt}S_t$ is a martingale under $\tilde{P}$.

We're used to looking at $e^{-rt}S_t$ as the discounted stock price. And the risk neutral measure is interpreted as the probability such that the discounted stock price is a martingale.

There is yet a slightly different way of looking at this. If we denote

$$\begin{aligned} dNt &= rNtdt \\ N_0 &= 1, \end{aligned}$$

that is $N_t = e^{rt}$; then N_t is the price of one unit of the *money market account*. Then $e^{-rt}S_t$ is nothing but the price of the stock *expressed in the unit of the money market account*. The risk neutral measure above can be looked at as the probability such that the price of the stock, *expressed in the unit of the money market account*, is a martingale.

Note that there is another asset, which is also a martingale (albeit a trivial one), when expressed in the unit of the money market account: the money market price process itself. It is clear that the price of the money market is 1 when expressed under its own unit, thus it is a (trivial) martingale.

9.1.2 Main questions of this chapter

The process N_t in the above is a *numéraire*, and the risk neutral measure we've studied in the Black-Scholes model is the risk neutral measure associated to the (domestic) money market numéraire. To re-emphasize, it is the probability measure such that the price of all non-dividend paying assets are martingales when expressed in the unit of the domestic money market account.

It is clear that the domestic money market is not the only choice for a numéraire. In a world where there is a foreign currency, then the foreign money market is also a possible choice of numéraire. The obvious question is, *how do we determine the risk neutral probability associated with the foreign money market numéraire*? More generally, how do we decide a risk neutral probability associated with any numéraire, as long as we have a model for that particular numéraire?

Specifically, letting

$$D_t := e^{-\int_0^t R_u du},$$

be the discount process, and suppose we have two underlying assets S_t, N_t so that both D_tN_t and D_tS_t are martingales under the risk neutral measure $\widetilde{P}$. Denoting

$$S_t^{(N)} := \frac{S_t}{N_t}$$

as the "price" of S_t under the numéraire N_t, we will address the following questions:

a. Does there exist a measure $\widetilde{P}^{(N)}$ so that $S_t^{(N)}$ is a martingale under $\widetilde{P}^{(N)}$? If yes, we will call $\widetilde{P}^{(N)}$ the risk neutral measure associated with the numéraire N_t.

b. What is the dynamics of $S_t^{(N)}$ under $\widetilde{P}^{(N)}$?

c. How can we relate the pricing of a derivative V_t based on S_t and N_t under the risk neutral measure $\widetilde{P}$ with the pricing of

$$V_t^{(N)} := \frac{V_t}{N_t}$$

under the measure $\widetilde{P}^{(N)}$?

One important thing to note about the choice of numéraire: we shall take only *non-dividend* paying assets as numéraire. Another way to put it is that we will only use asset

N_t that satisfies D_tN_t is a martingale under the risk neutral measure as a numéraire. Using this criterion, the domestic currency itself **cannot** be used as a numéraire because its value stays constant for all time t.

9.1.3 New set up of this chapter

Since in this chapter, we will introduce the foreign exchange rate and foreign money market, it is natural that we are into a multiple risky assets setting. Moreover, the risk free rate *will no longer be a constant* r. We will consider a risk-free rate process $R(t)$ that can be stochastic. The associated discount process is

$$D_t = \exp\Big\{ - \int_0^t R(u)du \Big\},$$

and for the domestic money market risk neutral measure $\tilde{P}$, we will require that D_tS_t be a martingale under $\tilde{P}$. Also, as we use different numéraires, there will be different risk-neutral measures corresponding to these numéraires. It is important to clarify which risk-neutral probability we are discussing. For example, we will call the risk neutral measure associated with the domestic money market *the domestic risk neutral measure.* Similarly, we will call the risk neutral measure associated with the foreign money market *the foreign risk neutral measure.* In this note, by dollars we also mean the domestic currency and vice versa.

To prepare for these new set ups, we will review a few details on stochastic calculus and multi-asset model in the next couple sections.

9.1.4 Why study change of numéraire

(i) A risk neutral pricing formula when the financial product is quoted in foreign currency:

Suppose we have a Euro style derivative that pays V_T (in foreign currency) at time T. We want to find the no-arbitrage price V_0 of this derivative at time 0. Let $R^f(t)$ be the foreign interest rate, which is an adapted process. Intuitively, the pricing formula would be

$$V_0 = \widetilde{E}^f \Big[e^{-\int_0^T R^f(u)du} V_T \Big],$$

where $\widetilde{E}^f$ is a foreign risk neutral measure. How to define this $\widetilde{E}^f$ so that the above formula holds is a question that we will address in this chapter.

(ii) Modeling when the interest rate is random:

When the interest rate is random, the pricing formula for a Euro-style derivative on a stock S_t becomes complicated (See formula 9.4.6 in Shreve and the discussion after).

However, when we use a suitable numéraire, which is the zero-coupon bond in this case, the pricing formula becomes much simpler (formula 9.4.7 in Shreve). Thus this suggests one should model S_t under the risk-neutral measure associated with the zero-coupond bond (called the T-forward measure). Indeed, it turns out that the correct object to model is the forward price of the stock S_t (Section 9.4.3 in Shreve). The point is that our usual choice of numéraire (the domestic money market) may not be the best choice in all situations. Studying the change of numéraire suggests other choice of numéraire that would simplify the problem, both in terms of pricing and in terms of modeling.

9.2 Markets with multiple risky assets

Itô process models for markets with multiple risky assets are treated in Chapter 5 of Shreve. This is a brief review.

Consider a market with m risky assets. Prices are given in a domestic currency, which, for convenience, we will assume to be US dollars. A price model consists of a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a filtration $\{\mathcal{F}(t);\, t \geq 0\}$, and an m-vector-valued stochastic process $S(t) = (S_1(t), \ldots, S_m(t))$ that represents the asset prices and that is adapted to the filtration. The goal of modeling is to construct S so that its statistical behavior approximately matches what is actually observed in the market. The return of asset i over the small interval of time $[t, t+dt]$ is given by $\dfrac{dS_i(t)}{S_i(t)}$. From analysis of the historical data or from structural models for the market, the modeler can generate estimates for all assets of

(i) the mean rate, local of change of asset i: $\quad \mu_i(t)\,dt = E\left[\dfrac{dS_i(t)}{S_i(t)} \;\Big|\; \mathcal{F}(t)\right]$;

(ii) the local (square) volatility of asset i: $\sigma_i^2(t)\,dt = \text{Var}\left(\dfrac{dS_i(t)}{S_i(t)} \;\Big|\; \mathcal{F}(t)\right)$; and

(iii) the correlation between the returns of asset i and j, $i \neq j$;

$$\rho_{ij}(t), dt = \frac{\text{Cov}\left(\frac{dS_i(t)}{S_i(t)}, \frac{dS_j(t)}{S_j(t)} \;\Big|\; \mathcal{F}(t)\right)}{\sigma_i(t)\sigma_j(t)}.$$

A nice way to construct models that can fit these informally described parameters, and for which the price processes are continuous, is to use stochastic differential equations driven by a multi-dimensional Brownian motion $W(t) = (W_1(t), \ldots, W_d(t))$. Suppose we set

$$\frac{dS_i(t)}{S_i(t)} = \mu_i(t)\,dt + \sum_{k=1}^{d} \sigma_{ik}(t)\,dW_k(t), \qquad 1 \leq i \leq m. \tag{9.1}$$

Let $\{\mathcal{F}(t);\, t \geq 0\}$ be a filtration for W and assume $\mu_i(t)$ and $\sigma_{ij}(t)$, $t \geq 0$ are adapted to $\{\mathcal{F}(t);\, t \geq 0\}$. Recall that, by definition, $W_1, \ldots, W_d$ are independent Brownian motions.

Then, formally, $E[dW_i(t) \mid \mathcal{F}(t)] = 0$, $E[(dW_i(t))^2 \mid \mathcal{F}(t)] = dt$ and $E[dW_i(t))\,dW_j(t) \mid \mathcal{F}(t)] = 0$. Thus, for each asset,

$$E\left[\frac{dS_i(t)}{S_i(t)} \,\Big|\, \mathcal{F}(t)\right] = \mu_i(t)\,dt + \sum_{k=1}^{d} \sigma_{ik}(t) E[dW_k(t) \mid \mathcal{F}(t)] = \mu_i(t)\,dt,$$

in conformity with (i). On the other hand,

$$\begin{aligned} \text{Var}\left(\frac{dS_i(t)}{S_i(t)} \,\Big|\, \mathcal{F}(t)\right) &= E\left[\left(\sum_{k=1}^{d} \sigma_{ik}(t)\,dW_k(t)\right)^2 \,\Big|\, \mathcal{F}(t)\right] \\ &= \left[\sum_{k=1}^{d} \sigma_{ik}^2(t)\right] dt. \end{aligned} \tag{9.2}$$

By a similar calculation

$$\begin{aligned} \text{Cov}\left(\frac{dS_i(t)}{S_i(t)}, \frac{dS_j(t)}{S_j(t)} \,\Big|\, \mathcal{F}(t)\right) &= E\left[\sum_{k=1}^{d} \sigma_{ik}(t)\,dW_k(t) \cdot \sum_{l=1}^{d} \sigma_{jl}(t)\,dW_l(t) \,\Big|\, \mathcal{F}(t)\right] \\ &= \left[\sum_{k=1}^{d} \sigma_{ik}(t)\sigma_{jk}(t)\right] dt. \end{aligned} \tag{9.3}$$

Therefore, we can match the model (9.1) to the variances and correlations prescribed in (ii) and (iii) by choosing d and $\sigma_{ij}(t)$, $1 \leq i, j \leq m$, so that

$$\begin{aligned} \sum_{k=1}^{d} \sigma_{ik}^2(t) &= \sigma_i^2(t) \\ \sum_{k=1}^{d} \sigma_{ik}(t)\sigma_{jk}(t) &= \rho_{ij}(t)\sigma_i(t)\sigma_j(t) \end{aligned}$$

Because model (9.1) is flexible enough to capture price means, volatilities, and correlations in this manner, it is a standard model for multi-asset markets. Usually, it is expressed in the more familiar form

$$dS_i(t) = \mu_i(t)S_i(t)\,dt + S_i(t) \sum_{k=1}^{d} \sigma_{ik}(t)\,dW_k(t). \tag{9.4}$$

Example 1. Given $\sigma_1(t)$, $\sigma_2(t)$, and $\rho(t)$ satisfying $-1 \leq \rho(t) \leq 1$, we want to construct a model with two risky assets so that the volatility process of S_1 is $\sigma_1(t)$, that of S_2 is $\sigma_2(t)$

and

$$\rho(t) = \frac{\text{Cov}\left(\frac{dS_1(t)}{S_1(t)}, \frac{dS_2(t)}{S_2(t)}\right)}{\sigma_1(t)\sigma_2(t)}.$$

This can be achieved with the model

$$dS_1(t) = \mu_1(t)S_1(t)\,dt + \sigma_1(t)S_1(t)\,dW_1(t), \tag{9.5}$$

$$dS_2(t) = \mu_2(t)S_2(t)\,dt + \sigma_2(t)S_2(t)\left[\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right]. \tag{9.6}$$

Indeed, equation (9.5) is by itself just the usual model for an asset with volatility process $\sigma_1(t)$. As for S_2,

$$\begin{aligned}
\text{Var}\left(\frac{dS_2(t)}{S_2(t)} \;\Big|\; \mathcal{F}(t)\right) &= E\left[\left(\sigma_2(t)\left[\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right]\right)^2 \;\Big|\; \mathcal{F}(t)\right] \\
&= \sigma_2^2(t)\left[\rho^2(t)E[(dW_1(t))^2] + (1-\rho^2(t))E[(dW_2(t))^2] + 2\rho(t)\sqrt{1-\rho^2(t)}E[dW_1(t)\,dW_2(t)]\right] \\
&= \sigma_2^2(t)\left[\rho^2(t)\,dt + (1-\rho^2(t))\,dt\right] = \sigma_2^2(t)\,dt
\end{aligned}$$

Also,

$$\begin{aligned}
\text{Cov}\left(\frac{dS_1(t)}{S_1(t)}, \frac{dS_2(t)}{S_2(t)} \;\Big|\; \mathcal{F}(t)\right) &= E\left[\sigma_1(t)\,dW_1(t)\cdot\left(\sigma_2(t)\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right) \;\Big|\; \mathcal{F}(t)\right] \\
&= \sigma_1(t)\sigma_2(t)\left[\rho(t)E[(dW_1(t))^2] + \sqrt{1-\rho^2(t)}E[dW_1(t)dW_2(t)]\right] \\
&= \sigma_1(t)\sigma_2(t)\rho(t)\,dt
\end{aligned}$$

9.2.1 Generating multi-dimensional Brownian motion with a given covariance structure

Another way to phrase the discussion in the previous section is that we are interested in generating the following multi-dimensional model

$$dS_i(t) = \mu_i S_i(t)\,dt + \sigma^i S_i(t) dW_t^i, i = 1\cdots, n \tag{9.7}$$

where for simplicity we assume all parameters are constants. Here W_t^i are Brownian motions with

$$\text{Cov}(d\bar{W}_t) = S dt$$

where S is a given positive definite symmetric matrix . Here

$$\bar{W}_t = [W_t^1, W_t^2, \cdots, W_t^n]^T$$

represents the vector with length n whose ith component is W_t^i.

To achieve this we use the Cholesky decomposition of S. That is we find a lower triangular matrix L so that $S = LL^T$. This is always possible if S is a positive definite symmetric matrix. See e.g. this Wikipedia reference. To generate $\bar{W}_t$, we start out with an independent n-dimensional Brownian motion

$$\bar{Z}_t = [Z_t^1, Z_t^2, \cdots, Z_t^n]^T.$$

We claim that

$$\bar{W}_t = L\bar{Z}_t$$

has the covariance structure we want. That is

$$\text{Cov}(d\bar{W}_t) = Sdt.$$

To see this, recall the following elementary result of multi-dimensional random variables: for A a $m \times n$ constant matrix and X a n-dimensional random variable,

$$\text{Cov}(AX) = A\text{Cov}(X)A^T.$$

Reason: WLOG we can assume $E(X) = 0$ and hence

$$\text{Cov}(AX) = E((AX)(AX)^T) = E(AXX^TA^T) = A\text{Cov}(X)A^T.$$

Applying this to our case, with $\text{Cov}(d\bar{Z}_t) = I_n dt$ where I_n is the $n \times n$ identity matrix since $\bar{Z}_t$ are independent Brownian motions, we have

$$\text{Cov}(d\bar{W}_t) = \text{Cov}(Ld\bar{Z}_t) = LI_nL^T dt = Sdt,$$

as desired.

9.2.2 Market with foreign currency

Example 1 can be translated into a model for a market with one risky asset and a tradable foreign currency, which is the important setting we want to discuss in this chapter.

Let $S(t)$ be the price in dollars of the risky asset.

Let $Q(t)$ denote the price (in dollars) of one unit of the foreign currency at time t. Thus $Q(t)$ is the exchange rate. Q can be thought of as a second risky asset; it fluctuates randomly and these fluctuations may be correlated with those of S (if we have reason to believe there is no correlation between S and Q we just have to set $\rho(t) = 0$ in the model below). Hence the model of Example 1 is appropriate.

Following the notation in Shreve, (9.3.1)-(9.3.2), it shall be written:

$$dS(t)\,dt = \alpha(t)S(t)\,dt + \sigma_1(t)S(t)\,dW_1(t), \tag{9.8}$$

$$dQ(t)\,dt = \gamma(t)Q(t)\,dt + \sigma_2(t)Q(t)\left[\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right]. \tag{9.9}$$

Suppose that in this market one can purchase a foreign money market account earning the risk-free rate with one's foreign cash.

Let $R^f(t)$, $t \geq 0$, be the risk-free foreign rate. We will say that one unit of this money market is an account in which one unit of foreign currency is deposited at time $t = 0$ and never withdrawn.

Thus the price at time t in *foreign currency* of one foreign money market unit is

$$M^f(t) = \exp\{\int_0^t R^f(u)\,du\}.$$

The price at time t of one foreign money market unit *in dollars* is

$$N^f(t) = M^f(t)Q(t).$$

If we are investing in this market, we will certainly deposit any idle *foreign cash* in the *foreign money market*; otherwise, we forego the interest we could earn at rate R^f. Thus it is really more appropriate to write the model for the above market in terms of $S(t)$ and $N^f(t)$. An easy calculation shows that this model is:

$$dS(t)\,dt = \alpha(t)S(t)\,dt + \sigma_1(t)S(t)\,dW_1(t), \tag{9.10}$$

$$\begin{aligned} dN^f(t)\,dt &= \left[\gamma(t) + R^f(t)\right] N^f(t)\,dt \\ &\quad + \sigma_2(t)N^f(t)\left[\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right]. \end{aligned} \tag{9.11}$$

Remarks:

(i) Both equations (9.10), (9.11) are expressed in terms of dollars, not the foreign currency; and N^f is, again, the price of the foreign money market in dollars.

(ii) Both equations (9.10), (9.11) are not written in risk neutral measure setting.

(iii) Important: Even though $Q(t)$ and $N^f(t)$ are closely related, there is one crucial difference between them: under the domestic risk neutral measure $\widetilde{P}$, $D(t)Q(t)$ is **not** a martingale while $D(t)N^f(t)$ **is** a martingale. The reason is because a unit of foreign currency (without being invested into the foreign money market) loses its value over time at the rate $R^f(t)$. See also section (9.4.1) for more discussion.

9.3 A review of multi-dimensional stochastic calculus

This is more review material, collected for convenience of reference.

9.3.1 Multi-dimensional stochastic integration

Let $W(t) = (W_1(t), \ldots, W_d(t))$ be a d-dimensional Brownian motion, and let $\{\mathcal{F}(t);\, t \geq 0\}$ be a filtration for W. Let the vector-valued process $\Theta(t) = (\theta_1(t), \ldots, \theta(t))$ be adapted to $\{\mathcal{F}(t);\, t \geq 0\}$. For d-dimensional vectors, we use $x \cdot y = \sum_{k=1}^{d} x_i y_i$ to denote the inner product and $\|x\| = \sqrt{\sum_{k=1}^{d} x_k^2} = \sqrt{x \cdot x}$ to denote the norm of a vector. Accordingly, we use the following convenient notation:

$$\int_0^t \Theta(u) \cdot \, dW(u) = \sum_{k=1}^{d} \int_0^t \theta_k(u)\, dW_k(u).$$

For example, if we define $\underline{\sigma_i}(t) = (\sigma_{i1}(t), \ldots, \sigma_{id}(t))$, the equation for $S_i(t)$ in (9.4) can be written

$$dS_i(t) = \mu_i(t) S_i(t)\, dt + S_i(t) \left[\underline{\sigma_i}(t) \cdot dW(t) \right].$$

9.3.2 Linear stochastic differential equations

The following general fact is useful. The solution to the stochastic differential equation $dX(t) = \mu(t)X(t)\, dt + X(t)[\Theta(t) \cdot dW(t)]$ is

$$X(t) = X(0) \exp\{ \int_0^t \Theta(u) \cdot dW(u) - \frac{1}{2} \int_0^t \|\Theta(u)\|^2\, du + \int_0^t \mu(u)\, du \}. \tag{9.12}$$

To show this expression is a solution requires just an application of the multi-dimensional Itô rule. We will not show it is the unique solution; this is done in the theory of stochastic differential equations.

A simple calculation also shows that X solves

$$dX(t) = \mu(t)X(t)\, dt + X(t)[\Theta(t) \cdot dW(t)] \tag{9.13}$$

if and only if

$$d\left[e^{-\int_0^t \mu(u)\, du} X(t) \right] = \left[e^{-\int_0^t \mu(u)\, du} X(t) \right] [\Theta(t) \cdot dW(t)] \tag{9.14}$$

We will pass between these two equivalent equations frequently and without comment.

9.3.3 Girsanov's theorem

Let

$$Z(t)\exp\{-\int_0^t \Theta(u)\cdot dW(u) - \frac{1}{2}\int_0^t \|\Theta(u)\|^2\,du\}.$$

If it is assumed that $E[Z(T)] = 1$, then

$$\mathbf{P}^Z(A) = E[\mathbf{1}_A Z(T)], \quad A \in \mathcal{F},$$

defines a new probability measure. The multi-dimensional Girsanov theorem says that

$$W^Z(t) = W(t) + \int_0^t \Theta(u)\,du = \left(W_1(t) + \int_0^t \theta_1(u)\,du, \ldots, W_d(t) + \int_0^t \theta_d(u)\,du\right)$$

is a Brownian motion up to time T under $\mathbf{P}^Z$.

9.3.4 What happens when we use a general $\mathcal{F}_t^W$ martingale Z_t in the change of measure fomula

Suppose that $Z(t)$ is a $\mathcal{F}(t)$ martingale and $Z(0) = 1$. It follows that $E[Z(T)] = Z(0) = 1$. We can still define a new measure

$$\mathbf{P}^Z(A) = E[\mathbf{1}_A Z(T)], \quad A \in \mathcal{F},$$

as above (the measure $\mathbf{P}^Z$ is well-defined). However, this is a bit abstract. We did not impose any dynamics on Z_t. But we still want to learn, for example, the distribution of $W(t)$ under $\mathbf{P}^Z$. It turns out that when the filtration is generated by the Brownian motion, then the martingale representation will give us information about the dynamics of Z_t and the Girsanov's theorem will tell us about the behavior of W_t under P^Z.

(i) Martingale representation:

Assume now that the filtration $\{\mathcal{F}(t);\, t \geq 0\}$ is generated by W. Under this **important** assumption, if $Z(t)$ is a martingale with respect to $\{\mathcal{F}(t);\, t \geq 0\}$ under measure $\mathbf{P}$, then the martingale representation theorem says there exists an adapted, vector valued process $\Gamma(t) = (\gamma_1(t), \ldots, \gamma_d(t))$ such that

$$Z(t) = Z(0) + \int_0^t \Gamma(u)\cdot dW(t).$$

Suppose that $Z(0) = 1$ and that $Z(t) > 0$ for all $0 \leq t \leq T$ almost surely. By defining $\nu(t) = (\nu_1(t), \ldots, \nu_d(t)) = \frac{1}{Z(t)}\Gamma(t)$, one can write

$$Z(t) = 1 + \int_0^t Z(u)\frac{1}{Z(u)}\Gamma(u)\cdot dW(u) = 1 + \int_0^t Z(u)[\nu(u)\cdot dW(u)].$$

It then follows from equation (9.12) that

$$Z(t) = \exp\{\int_0^t \nu(u) \cdot dW(u) - \frac{1}{2}\int_0^t \|\nu(u)\|^2 \, du\}. \tag{9.15}$$

(ii) Girsanov's theorem:

By applying Girsanov's theorem to this expression we obtain:

Theorem 3. *Suppose that* $\{\mathcal{F}(t);\, t \geq 0\}$ *is generated by* W *and that* $Z(t)$ *is an* $\{\mathcal{F}(t);\, t \geq 0\}$*-martingale under* $\mathbf{P}$ *such that* $E[Z(t)] = 1$ *for all* t*. Define* $P^Z(A) = E[\mathbf{1}_A Z(T)]$, $A \in \mathcal{F}$.

Suppose in addition that $Z(T) > 0$ *almost surely. Then there is an* $\{\mathcal{F}(t);\, t \geq 0\}$*-adapted process* $\nu(t) = (\nu_1(t), \ldots, \nu_d(t))$ *so that equation (9.15) holds, and for this process,*

$$W^Z(t) = W(t) - \int_0^t \nu(u)\, du \quad \text{is a Brownian motion up to time } T \text{ under } P^Z.$$

The only point we have not justified (and will not) is that if $Z(T) > 0$ almost surely, then $Z(t) > 0$ for all $t \leq T$ almost surely.

This theorem is essentially Theorem 9.2.1 in Shreve; we have just stated it more generally. It is *one of the important theorems* in this Chapter. Later on, we will replace Z_t with $D_t N_t$, where D_t is the discounted process mentioned above and N_t the actual numéraire we want to study (for example, the domestic or foreign money market). Then P^Z is the risk neutral measure associated with that numéraire. And this theorem tells us how the distribution of the Brownian motion changes under this risk neutral measure. Note that at this level, the Theorem is a bit abstract: it only tells us that the process ν exists—it does not say how to find ν. In applications, one can often determine ν from other assumptions, as we shall see in studying numéraires.

9.4 The domestic risk-neutral measure

Consider the model for $S(t) = (S_1(t), \ldots, S_m(t))$ given in equation (9.4). Assume henceforth that $\{\mathcal{F}(t);\, t \geq 0\}$ is the filtration generated by W. This allows us to employ the martingale representation theorem.

Add also to the model a risk-free rate process $R(t)$, $t \geq 0$, which is assumed to be non-negative and adapted to $\{\mathcal{F}(t);\, t \geq 0\}$. The associated discount process is denoted by $D(t) = \exp\{-\int_0^t R(u)\, du\}$.

The price of S_i, in terms of the domestic money market, is $D_t S_i(t)$. We have the following important definition:

Definition 9.4.1. *The domestic risk-neutral measure for the model (9.4) is the probability measure $\tilde{P}$ such that $D_t S_i(t)$ is a martingale under $\tilde{P}$, for all i.*

The model (9.4) is equivalent to

$$d\,[D(t)S_i(t)] = (\mu_i(t) - R(t))D(t)S_i(t)\,dt + D(t)S_i(t)\sum_{k=1}^{d}\sigma_{ik}(t)\,dW_k(t), \qquad 1 \le i \le m, \tag{9.16}$$

as an easy calculation shows; (compare to equations (9.13) and (9.14)). We are interested in finding a domestic risk-neutral measure, assuming one exists. The essential ingredient is provided in the following theorem, which reviews material from Chapter 5 of Shreve. This review is useful because the procedure of finding the risk-neutral measure is a template for changing measure for numéraires.

Theorem 4. *Assume that there is a risk-neutral measure $\widetilde{\mathbf{P}}$ for model (9.16) given by $\widetilde{\mathbf{P}}(A) = E[\mathbf{1}_A Z]$, where Z is an $\mathcal{F}(T)$ measurable random variable for which $E[Z] = 1$ and $\mathbf{P}(Z > 0) = 1$. Then*

$$Z = \exp\left\{-\int_0^T \Theta(u)\cdot dW(u) - \frac{1}{2}\int_0^T \|\Theta(u)\|^2\,du\right\}, \tag{9.17}$$

where $\Theta(t) = (\theta_1(t), \ldots, \theta_d(t))$ is an $\{\mathcal{F}(t);\, t \ge 0\}$-adapted process that is a solution of the market price of risk equation

$$\begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) & \cdots & \sigma_{1d}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) & \cdots & \sigma_{2d}(t) \\ \vdots & \vdots & & \vdots \\ \sigma_{m1}(t) & \sigma_{m2}(t) & \cdots & \sigma_{md}(t) \end{pmatrix} \cdot \begin{pmatrix} \theta_1(t) \\ \theta_2(t) \\ \vdots \\ \theta_d(t) \end{pmatrix} = \begin{pmatrix} \mu_1(t) - R(t) \\ \mu_2(t) - R(t) \\ \vdots \\ \mu_m(t) - R(t) \end{pmatrix}, \qquad 0 \le t \le T. \tag{9.18}$$

If this equation has a unique solution, the risk-neutral measure is unique. Under $\widetilde{\mathbf{P}}$,

$$\widetilde{W}(t) = \left(W_1(t) + \int_0^t \theta_1(u)\,du, \ldots, W_d(t) + \int_0^t \theta_d(u)\,du\right) \tag{9.19}$$

is a Brownian motion up to time t.

Proof: The process $Z(t) = E[Z \mid \mathcal{F}(t)]$ is a martingale and since Z is $\mathcal{F}(T)$-measurable, $Z(T) = Z$. By Theorem 3, there is an adapted process ν such that $Z(t) = \exp\{\int_0^t \nu(u)\cdot dW(u) - \frac{1}{2}\int_0^t \|\nu(u)\|^2\,du\}$. Equation (9.17) then follows if we set $\Theta(t) = -\nu(t)$. By Girsanov, the process $\widetilde{W}$ defined in equation (9.19) is a Brownian motion up to time T

under $\widetilde{\mathbf{P}}$. From (9.19), $dW_i(t) = d\widetilde{W}_i(t) - \theta_i(t)\,dt$. By using this substitution in the equation (9.16) for $D(t)s_i(t)$,

$$d\left[D(t)S_i(t)\right] = \left(\mu_i(t) - R(t) - \sum_{k=1}^{d} \sigma_{ik}(t)\theta_k(t)\right) D(t)S_i(t)\,dt + D(t)S_i(t)\sum_{k=1}^{d} \sigma_{ik}(t)\,d\widetilde{W}_k(t). \tag{9.20}$$

This must be a martingale under the risk-neutral measure $\widetilde{\mathbf{P}}$ for all i; that is what it means for $\widetilde{\mathbf{P}}$ to be a risk-neutral measure. Thus the 'dt' term in (9.20) must be 0 for each i:

$$\sum_{k=1}^{d} \sigma_{ik}(t)\theta_k(t) = \mu_i(t) - R(t), \qquad 1 \leq i \leq m.$$

The matrix form of these equations is just equation (9.18) of the theorem statement. This completes the proof. $\diamond$

As a consequence of the proof, the stochastic differential equation model for the discounted prices under the risk-neutral measure is

$$d\left[D(t)S_i(t)\right] = D(t)S_i(t)\sum_{k=1}^{d} \sigma_{ik}(t)\,d\widetilde{W}_k(t), \qquad 1 \leq i \leq m.$$

Remarks:

1) The equations summarized by (9.18) are called the *market price of risk equations.* The difference $\mu_i(t) - R(t)$ can be regarded as a risk premium; it is the amount by which the expected rate of gain of the asset is larger than the risk-free rate. Investors typically demand $\mu_i(t) - R(t)$ to be positive before investing in i, to make up for the fact that the investment carries risk. The expression $\mu_i(t) - R(t) = \sum_{k=1}^{d} \sigma_{ik}(t)\theta_k(t)$ may be thought of as a decomposition of $\mu_i(t) - R(t)$ into a sum contributions from each source of random fluctations of $S_i(t)$; $\theta_k(t)$ is effectively a price per unit of volatility of the contribution $\sigma_{ik}(t)\theta_k(t)$.

2. Theorem 2 implies that a necessary condition for the existence of a risk-neutral measure is that (9.18) must have a solution $\Theta(t)$. However, having a solution to (9.18) is not by itself a sufficient condition for the existence of a risk neutral measure. If $\Theta(t)$ is a solution and $Z = \exp\left\{-\int_0^T \Theta(u)\cdot dW(u) - \frac{1}{2}\int_0^T \|\Theta(u)\|^2\,du\right\}$, one must check in addition that $E[Z] = 1$, in order that $\widetilde{\mathbf{P}}^Z$ define a probability measure.

9.4.1 Model with foreign money market under the domestic risk-neutral measure

Consider the model for a risky asset, a foreign currency and a foreign money market introduced above in Section (9.2.2). Now add a domestic money market, with risk-free rate $R(t), t \geq 0$.

Recall that the price at time t in *foreign currency* of one foreign money market unit is

$$M^f(t) = \exp\{\int_0^t R^f(u)\,du\}.$$

Given the exchange rate $Q(t)$, the price *in dollars* of a unit of the foreign money market is

$$N^f(t) = M^f(t)Q(t).$$

So there are 2 risky assets in this model:

$$\begin{aligned} dS(t)\,dt &= \alpha(t)S(t)\,dt + \sigma_1(t)S(t)\,dW_1(t), \\ dN^f(t)\,dt &= \left[\gamma(t) + R^f(t)\right] N^f(t)\,dt \\ &\quad + N^f(t)\sigma_2(t)\left[\rho(t)\,dW_1(t) + \sqrt{1-\rho^2(t)}\,dW_2(t)\right]. \end{aligned}$$

As definition (9.4.1) states, a domestic risk-neutral measure for this model must make $D(t)S(t)$ and $D(t)N^f(t) = D(t)M^f(t)Q(t)$ into martingales.

Equation (9.18) in this case is

$$\begin{pmatrix} \sigma_1(t) & 0 \\ \sigma_2(t)\rho(t) & \sigma_2(t)\sqrt{1-\rho^2(t)} \end{pmatrix} \begin{pmatrix} \theta_1(t) \\ \theta_2(t) \end{pmatrix} = \begin{pmatrix} \alpha(t) - R(t) \\ \gamma(t) + R^f(t) - R(t) \end{pmatrix} \tag{9.21}$$

Assume there is a unique, risk-neutral measure for (9.10)–(9.11). By Theorem 4 equation (9.21) must then have a unique solution. Indeed it will, if $\sigma_1(t) > 0$, $\sigma_2(t) > 0$, and $-1 < \rho(t) < 1$ for all t with probability one. This solution is

$$\theta_1(t) = \frac{\alpha(t) - R(t)}{\sigma_1(t)}, \quad \theta_2(t) = \frac{1}{\sigma_2(t)\sqrt{1-\rho^2(t)}}\left[\gamma(t) + R^f(t) - R(t) - \sigma_2(t)\rho(t)\theta_1(t)\right].$$

Let $\widetilde{W}(t) = (W_1(t) + \int_0^t \theta_1(u)\,du, W_2(t) + \int_0^t \theta_2(u)\,du)$. Then one easily derives

$$\begin{aligned} dS(t) &= R(t)S(t)\,dt + \sigma_1(t)S(t)\,d\widetilde{W}_1(t) \\ dN^f(t) &= R(t)N^f(t)\,dt + N^f(t)\sigma_2(t)\left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right]. \end{aligned}$$

Remark 9.4.2. *This is similar to the situation in the classical Black-Scholes model, in which we consider 2 assets: the (domestic) money market and the stock S_t. The only difference is in the Black-Scholes model, the dynamics of the domestic money market does not have a Brownian motion component:*

$$dN(t) = R(t)N(t)dt.$$

Also note that the Brownian motion component of the foreign money market comes from the dynamics of the exchange rate $Q(t)$, not from the dynamics of $M^f(t)$ itself.

Remark 9.4.3. *We* **cannot** *require the foreign exchange rate $Q(t)$ to satisfy the condition $D(t)Q(t)$ being a martingale under the domestic risk neutral measure $\widetilde{P}$. This is because $Q(t)$ is the price of the foreign currency, which loses value over time if not invested into the foreign money market. To see this consider a portfolio that invests Δ_t into the foreign money market N_t^f and $V_t - \Delta_t N_t^f$ into the domestic money market. Then because the portfolio is self-financing*

$$\begin{aligned} dV_t &= r_t V_t dt + (\gamma_t + R_t^f - r_t)\Delta_t N_t^f dt + \sigma_t \Delta_t N_t^f dW_t \\ &= r_t V_t dt + \sigma_t \Delta_t N_t^f (\alpha_t dt + dW_t), \end{aligned}$$

where α_t is such that $\sigma_t \alpha_t = \gamma_t + R_t^f - r_t$. The discounted value of this portfolio, namely $D_t V_t$ must be a martingale under the risk neutral measure $\widetilde{P}$. So under $\widetilde{P}$

$$d\tilde{W}_t = dW_t + \alpha_t dt$$

is a Brownian motion. On the other hand,

$$\begin{aligned} dQ_t &= \gamma_t Q_t dt + \sigma_t Q_t dW_t \\ &= (r_t - R^f)Q_t dt + \sigma_t Q_t d\widetilde{W}_t. \end{aligned}$$

It follows immediately that $D_t Q_t$ cannot be a martingale under $\widetilde{P}$.

In fact **if** $D(t)Q(t)$ were a martingale under the domestic risk neutral measure and the foreign interest rate is not identically 0 then we would have an arbitrage opportunity, as the following lemma shows.

Lemma 9.4.4. *Suppose that $D(t)Q(t)$ is a martingale under $\widetilde{P}$ and $R^f(t)$ is not identically 0. Then an arbitrage opportunity exists.*

Proof. Consider a contract that pays 1 unit of foreign currency at time T. The value of this contract at time 0 is

$$V_0 = \tilde{E}(D(T)Q(T)).$$

By the assumption that $D(t)Q(t)$ is a martingale under $\widetilde{P}$, we have $V_0 = Q_0$. But this implies an arbitrage opportunity since at time 0, we can sell such a contract for $Q(0)$, use $Q(0)$ to buy 1 unit of foreign currency and invest in the foreign money market. At time T we would have $e^{R^f(t)dt}$ in units of foreign currency which is larger than 1 since $R^f(t)$ is not identically 0. We can then use 1 foreign currency to close out the contract and make a riskless profit.

9.5 Numéraires

Up to now, we have always assumed that prices were given in units of a fixed, domestic currency, which for concreteness we take to be US dollars. One could choose other units to measure prices, and it is often convenient, even necessary, to do so.

Let the price in dollars of some given asset or financial instrument be denoted $N(t)$. Let $S(t)$ be the price in dollars of any other asset. Then the ratio

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

is the price of the asset corresponding to S in units of the asset corresponding to N. In this situation, N is referred as the numéraire. The asset used for the numéraire could in principle be almost anything— a risky asset in the market, a foreign currency, a money market account, an index based on the market, or the price of a derivative.

Example 9.5.1. *Let $R(t)$ denote the (domestic) risk-free rate. It is common to think of $R(t)$ as the rate available from a money market account which can be added to or withdrawn from at will. One unit of a money market account is defined to be the value of $\$1$ invested at time $t = 0$ and left in the account. This value in dollars at time t is $M(t) = \exp\{\int_0^t R(u)\,du\}$. If $S(t)$ is the price in dollars of an asset at time t, its price in units of the money market is*

$$\frac{S(t)}{\exp\{\int_0^t R(u)\,du\}} = e^{-\int_0^t R(u)\,du} S(t).$$

This is just the discounted price, or present value, of $S(t)$. So we can think of discounting as an example of pricing in money market account units.

Example 9.5.2. *(Non-example) Consider the market model studied in Examples 2 and 3. This consists of an asset with price (in dollars) $S(t)$, an exchange rate $Q(t)$ (dollars per* *unit of foreign currency), a domestic risk-free rate $R(t)$, and a foreign risk-free rate $R^f(t)$. Let $M(t) = \exp\{\int_0^t R(u)\,du\}$.*

There are many choices for denominating prices. A tempting example is to use the foreign currency as the numéraire. In this case, $S^{(Q)}(t) = S(t)/Q(t)$ is the price of the asset in units of the foreign currency, while a unit of the domestic money market in the foreign currency is $M^{(Q)}(t) = M(t)/Q(t)$. However, this should not be done, because under the domestic risk neutral measure, $Q(t)$ is NOT a martingale (See the discussion in section (9.4.1) and in section (9.5.1)). This is also consistent with our remark at the beginning of this note that we will only use non-dividend paying asset as numéraire. $Q(t)$, as denoting the price of the foreign currency, is a dividend paying asset with dividend rate R^f.

Example 9.5.3. *One could also use as numéraire the value in dollars $N^f(t) = M^f(t)Q(t)$ of a unit of the foreign money market. Then, in this unit*

$$S^{(N^f)}(t) = \frac{S(t)}{M^f(t)Q(t)} = e^{-\int_0^t R^f(u)\,du} S^{(Q)}(t).$$

is the price of the asset, and

$$\frac{Q(t)}{M^f(t)Q(t)} = e^{-\int_0^t R^f(u)\,du}$$

is the value of a unit of foreign currency.

9.5.1 Change of measure for change of numéraire

Risk-neutral pricing theory should not depend on the unit of price. If there is a risk-neutral measure when the price is in dollars, then there ought to be a risk-neutral measure $\widetilde{\mathbf{P}}^N$ for pricing with respect to N, for any numéraire N. This section addresses how to find $\widetilde{\mathbf{P}}^N$.

We will always start out with a risk-neutral model for $S(t) = (S_1(t), \ldots, S_d(t))$, given on a probability space $(\Omega, \mathcal{F}, \widetilde{\mathbf{P}})$, with filtration $\{\mathcal{F}(t);\, t \geq 0\}$, and a risk-free (domestic) rate $R(t)$, $t \geq 0$. As usual, $D(t) = \exp\{-\int_0^t R(u)\,, du\}$ denotes the discount factor. Thus we can also see this as starting out with *our default probability measure* as the domestic risk neutral measure.

Let $N(t)$, $t \geq 0$, be a strictly positive, numéraire process. Since $N(t)$ represents a price and the model is risk-neutral, $D(t)N(t)$, $t \geq 0$, is a martingale. In particular,

$$\tilde{E}\left[D(T)N(T)\right] = N(0), \quad \text{for any } T \geq 0. \tag{9.22}$$

Let T be the time horizon for which we want to study the market. It follows that

$$\widetilde{\mathbf{P}}^{(N)}(A) = \tilde{E}\left[\mathbf{1}_A \frac{D(T)N(T)}{N(0)}\right]$$

defines a new probability measure.

Theorem 5. *$\widetilde{\mathbf{P}}^{(N)}$ is a risk-neutral measure for pricing with respect to N in the following sense: for each i, $1 \le i \le m$, $S_i^{(N)}(t)$ is a martingale with respect to $\{\mathcal{F}(t);\, t \ge 0\}$ up to time T under $\widetilde{\mathbf{P}}^{(N)}$.*

Proof: The proof is an application of the formula for computing conditional expectations under a change of measure. Shreve states a special case in Lemma 5.2.2. The formula implies that for any sub-σ-algebra $\mathcal{G}$,

$$\tilde{E}^{(N)}\left[X \,\Big|\, \mathcal{G}\right] = \frac{\tilde{E}\left[X\frac{D(T)N(T)}{N(0)} \,\Big|\, \mathcal{G}\right]}{\tilde{E}\left[\frac{D(T)N(T)}{N(0)} \,\Big|\, \mathcal{G}\right]}. \tag{9.23}$$

In this formula $\tilde{E}^{(N)}$ represents expectation with respect to $\widetilde{\mathbf{P}}^{(N)}$. Apply equation (9.23) with $X = S_i^{(N)}(T)$ and $\mathcal{G} = \mathcal{F}(t)$. The result is

$$\tilde{E}^{(N)}\left[S_i^{(N)}(T) \,\Big|\, \mathcal{F}(t)\right] = \frac{\tilde{E}\left[\frac{S_i(T)}{N(T)}\frac{D(T)N(T)}{N(0)} \,\Big|\, \mathcal{F}(t)\right]}{\tilde{E}\left[\frac{D(T)N(T)}{N(0)} \,\Big|\, \mathcal{F}(t)\right]} = \frac{\tilde{E}\left[D(T)S_i(T) \,\Big|\, \mathcal{F}(t)\right]}{\tilde{E}\left[D(T)N(T) \,\Big|\, \mathcal{F}(t)\right]}.$$

But $D(u)S_i(u)$ and $D(u)N(u)$ are both martingales under $\widetilde{\mathbf{P}}$, and so $\tilde{E}\left[D(T)S_i(T) \,\Big|\, \mathcal{F}(t)\right] = D(t)S_i(t)$ and $\tilde{E}\left[D(T)N(T) \,\Big|\, \mathcal{F}(t)\right] = D(t)N(t)$. Hence

$$\tilde{E}^{(N)}\left[S_i^{(N)}(T) \,\Big|\, \mathcal{F}(t)\right] = \frac{D(t)S_i(t)}{D(t)N(t)} = \frac{S_i(t)}{N(t)} = S_i^{(N)}(t).$$

This shows that $S_i^{(N)}(u)$, $0 \le u \le T$, is a martingale under $\widetilde{\mathbf{P}}^{(N)}$. ◇

Remark 9.5.4. *Intuitively, the foreign risk neutral measure $\widetilde{P}^{Nf}$ should satisfy the condition that the discounted (under the foreign interest rate) risky asset price quoted in the foreign currency is a martingale under $\widetilde{P^{Nf}}$. This is indeed true in our framework: Let S_t be the dynamics of the risky asset quoted in dollars. Then $\frac{S_t}{Q_t}$ is the price of the risky asset quoted in foreign currency. By our construction,*

$$D^f(t)\frac{S_t}{Q_t} = \frac{S_t}{N^f(t)} = S^{(N^f)}(t)$$

is a martingale under $\widetilde{P^{Nf}}$.

9.5.2 Pricing under a change of numéraire

Suppose we have a financial product that pays $V(T)$ dollars at time T. Then the (domestic) risk neutral price of this product at time t is

$$V(t) = \tilde{E}\Big[\frac{D(T)}{D(t)}V(T)|\mathcal{F}(t)\Big]$$

(since $D(t)V(t)$ is a martingale under $\tilde{P}$).

What is the corresponding pricing formula when $V(t)$ is denoted under the unit of a numéraire $N(t)$? Arguing similar to the proof of Theorem (5) we will see that

$$\frac{V(t)}{N(t)} = \tilde{E}^{(N)}\Big[\frac{V(T)}{N(T)}|\mathcal{F}(t)\Big].$$

Thus denoting $V^{(N)}(t) := \frac{V(t)}{N(t)}$ we have

$$V^{(N)}(t) = \tilde{E}^{(N)}\Big[V^{(N)}(T)|\mathcal{F}(t)\Big].$$

This equation is meaningful by itself. It says that the price in the unit of numéraire $N(t)$ of the financial product is the conditional expectation under the corresponding risk neutral measure of the terminal value, also expressed in the same unit of numéraire. Note that the domestic risk neutral pricing formula is a special case of this when we use $N(t) = \frac{1}{D(t)}$, the domestic money market account.

It is also important to remember that here $V(t)$ is again in dollars, or the domestic currency, and $N(t)$ is the price in dollars of the numéraire of interest. To see a consequence of this, see the below section on pricing a financial product quoted in foreign currency.

9.5.3 Effect of change of numéraire

In section V, no assumptions were made on the nature of the price model. In this section, we specialize to the multi-asset model stated above and written under the risk-neutral measure as

$$d\left[D(t)S_i(t)\right] = D(t)S_i(t)\sum_{k=1}^{d}\sigma_{ik}(t)\,d\widetilde{W}_k(t), \qquad 1 \le i \le m. \tag{9.24}$$

In addition, we impose the assumption that $\{\mathcal{F}(t);\, t \ge 0\}$ is generated by $\widetilde{W}$.

Let N be a numéraire process and $\widetilde{\mathbf{P}}^{(N)}$ the risk-neutral measure for N. $\widetilde{W}$ is no longer a Brownian motion under $\widetilde{\mathbf{P}}^{(N)}$. The object of this section is use Theorem 3 and Girsanov's theorem to identify an appropriate Brownian motion $\widetilde{W}^{(N)}(t)$ under $\widetilde{\mathbf{P}}^{(N)}$ and to rewrite equation (9.24) using it.

Now $D(t)N(t)$ is a martingale under $\widetilde{\mathbf{P}}$, and $N(t)$, as a numéraire, is strictly positive for all t. So Theorem 1 implies there is a process ν so that

$$\frac{D(t)N(t)}{N(0)} = \exp\{\int_0^t \nu(u)\cdot d\widetilde{W}(u) - \frac{1}{2}\int_0^t \|\nu(u)\|^2\,du\} \tag{9.25}$$

Since $\widetilde{\mathbf{P}}^{(N)}(A) = \tilde{E}[\mathbf{1}_A \frac{D(T)N(T)}{N(0)}]$, it also follows from Theorem 1 that $\widetilde{W}^{(n)}(t) = \widetilde{W}(t) - \int_0^t \nu(u)\,du$ is a Brownian motion under $\widetilde{\mathbf{P}}^{(N)}$ up to time T.

Define $\sigma_i(t) = (\sigma_{i1}(t),\ldots,\sigma_{id}(t))$, so that (9.24) may be written compactly as

$$dD(t)S_i(t) = D(t)S_i(t)\left[\sigma_i(t)\cdot d\widetilde{W}(t)\right].$$

By (9.12), the solution to this equation is

$$D(t)S_i(t) = S_i(0)\exp\{\int_0^t \sigma_i(u)\cdot d\widetilde{W}(u) - \frac{1}{2}\int_0^t \|\sigma_i(u)\|^2\,du\}.$$

Thus, using the representation (9.25) for $D(t)N(t)$,

$$\begin{aligned} S_i^{(N)}(t) &= \frac{S_i(t)}{N(t)} = \frac{D(t)S_i(t)}{D(t)N(t)} \\ &= \frac{S_i(0)}{N(0)}\exp\{\int_0^t [\sigma_i(u)-\nu(u)]\cdot d\widetilde{W}(u) - \frac{1}{2}\int_0^t (\|\sigma_i(u)\|^2 - \|\nu(u)\|^2)\,du\} \end{aligned}$$

Replace $d\widetilde{W}$ in this expression by $d\widetilde{W}^{(N)}(t) + \nu(t)\,dt$. Note first that

$$\begin{aligned} &\exp\{\int_0^t [\sigma_i(u)-\nu(u)]\cdot[d\widetilde{W}^{(N)}(u) + \nu(u)\,du] \\ &= \int_0^t [\sigma_i(u)-\nu(u)]\cdot d\widetilde{W}^{(N)}(u) + \int_0^t \sigma_i(u)\cdot\nu(u)\,du - \int_0^t \nu(u)\cdot\nu(u)\,du \\ &= \int_0^t [\sigma_i(u)-\nu(u)]\cdot d\widetilde{W}^{(N)}(u) + \int_0^t \sigma_i(u)\cdot\nu(u)\,du - \int_0^t \|\nu(u)\|^2\,du \end{aligned}$$

It follows that

$$\begin{aligned} &S_i^{(N)}(t) \\ &\quad= \frac{S_i(0)}{N(0)}\exp\{\int_0^t [\sigma_i(u)-\nu(u)]\cdot d\widetilde{W}^{(N)}(u) - \frac{1}{2}\int_0^t (\|\sigma_i(u)\|^2 - 2\underline{\sigma}_i(u)\cdot\nu(u) + \|\nu(u)\|^2)\,du\}. \end{aligned}$$

But

$$\begin{aligned}\|\underline{\sigma}_i(u) - \nu(u)\|^2 &= [\sigma_i(u) - \nu(u)] \cdot [\sigma_i(u) - \nu(u)] \\ &= \|\sigma_i(u)\|^2 - 2\sigma_i(u) \cdot \nu(u) + \|\nu(u)\|^2)\end{aligned}$$

Thus

$$S_i^{(N)}(t) = \frac{S_i(0)}{N(0)} \exp\{\int_0^t [\sigma_i(u) - \nu(u)] \cdot d\widetilde{W}^{(N)}(u) - \frac{1}{2}\int_0^t \|\sigma_i(u) - \nu(u)\|^2 \, du\}$$

It follows from equation (9.12) that

$$dS_i^{(N)}(t) = S_i^{(N)}(t)[\sigma_i(t) - \nu(t)] \cdot d\widetilde{W}^{(N)}(t) = S_i^{(N)}(t) \sum_{k=1}^{d} (\sigma_{ik}(t) - \nu_k(t)) \, d\widetilde{W}_k^{(N)}(t) \tag{9.26}$$

This is an interesting equation because it shows exactly how the volatility of $S_i^{(N)}$ differs from that of S_i. Of course, we expect them to differ because N itself has volatility and $S_i^{(N)}(t) = S_i(t)/N(t)$. In fact, from the expression (9.25) and from (9.12) one finds that

$$dN(t) = R(t)N(t)\,dt + \sum_{k=1}^{d} \nu_k(t)\, d\widetilde{W}_k(t),$$

so $\nu_k(t)$ is the component of the volatility of N at time t due to $\widetilde{W}_k$.

9.6 Foreign risk-neutral measure

The discussion of the previous section established the existence of ν, but not a formula for it. In examples it can be found explicitly if the numéraire is defined explicitly.

Consider the example of a market with an asset and foreign currency formulated above. Its risk neutral version was derived in Example 3 and is

$$dS(t) = R(t)S(t)\,dt + \sigma_1(t)S(t)\,d\widetilde{W}_1(t) \tag{9.27}$$

$$dN^f(t) = R(t)N^f(t)\,dt + N^f(t)\sigma_2(t)\left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right] \tag{9.28}$$

Recall that $N^f(t) = \exp\{\int_0^t R^f(u)\,du\}Q(t)$ is the dollar value of one unit of **the foreign money market account**. *We shall use it as the numéraire in this section.* The domestic discount factor is $D(t) = \exp\{-\int_0^t R(u)\,du\}$.

From (9.28),

$$\begin{aligned} d[D(t)N^f(t)] &= D(t)N^f(t)\sigma_2(t)\left[\rho(t)\, d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\, d\widetilde{W}_2(t)\right] \\ &= D(t)N^f(t)\left(\sigma_2(t)\rho(t), \sigma_2(t)\sqrt{1-\rho^2(t)}\right)\cdot d\widetilde{W}(t). \end{aligned}$$

Note: $D(0)N^f(0) = Q(0)$. It follows from (9.12) that

$$D(t)N^f(t) = Q(0)\exp\{\int_0^t \nu(u)\cdot d\widetilde{W}(t) - \frac{1}{2}\int_0^t \|\nu(u)\|^2\, du\},$$

where $\nu(t) = \left(\sigma_2(t)\rho(t), \sigma_2(t)\sqrt{1-\rho^2(t)}\right)$.

The risk-neutral measure for denominating prices in units of the foreign money market up to time T, or *the foreign risk neutral measure* in short, is therefore defined by

$$\widetilde{\mathbf{P}}^{(N^f)}(A) = \tilde{E}\left[\mathbf{1}_A\frac{D(T)N^f(T)}{Q(0)}\right] = \tilde{E}\left[\mathbf{1}_A \exp\{\int_0^t \nu(u)\cdot d\widetilde{W}(t) - \frac{1}{2}\int_0^t \|\nu(u)\|^2\, du\}\right],$$

and

$$\widetilde{W}^{(N^f)}(t) = \left(\widetilde{W}_1(t) - \int_0^t \sigma_2(u)\rho(u)\, du,\ \widetilde{W}_2(t) - \int_0^t \sigma_2(u)\sqrt{1-\rho^2(u)}\, du\right)$$

is a Brownian motion up to time T under $\widetilde{\mathbf{P}}^{(N^f)}$.

The price of the asset with respect to numéraire $N^f(t)$ is $S^{(N^f)}(t) = S(t)/N^f(t)$. It is the *present value* of the asset price in units of the foreign currency (or just simply the value of the asset price at time t in units of the foreign money market). It is a martingale with respect to $\widetilde{\mathbf{P}}^{(N^f)}$. By applying (9.26),

$$dS^{(N^f)}(t) = S^{(N^f)}(t)\left[(\sigma_1(t) - \sigma_2(t)\rho(t))\, d\widetilde{W}_1^{(N^f)}(t) - \sigma_2(t)\sqrt{1-\rho^2(t)}\, d\widetilde{W}_2^{(N^f)}(t)\right], \quad t \le T.$$

Because $\widetilde{W}_2$ contributes to the volatility of $N^f(t)$, $d\widetilde{W}_2^{(N^f)}(t)$ contributes to the volatility of $S^{(N^f)}(t)$.

9.6.1 Pricing a financial product quoted in foreign currency

Suppose we have a financial product that pays $V^f(T) := \Phi(S\cdot)$ *units of foreign currency* at time T. Then we have the following lemma

Lemma 9.6.1. *The risk neutral price (in foreign currency) of the above product is*

$$V^f(t) = \tilde{E}^{(N^f)}\left(e^{-\int_t^T R^f(u)du}V(T)\middle|\mathcal{F}(t)\right).$$

Note that the result is very intuitive: to price a financial product quoted in foreign currency, we take conditional expectation under the foreign risk neutral measure, discounted under the foreign interest rate.

Proof. The proof of this Lemma relies on the result of Section (9.5.2). Using the foreign money market $N^f(t)$ as numéraire, the value of $V^f(T)$ denoted in the units of $N^f(T)$ is

$$\frac{V^f(T)Q(T)}{N^f(T)}.$$

To price $V^f(T)$ using $N^f(T)$ as numéraire we need to use $\widetilde{P}^{(N^f)}$. Thus we have

$$V^{(N^f)(t)} = \tilde{E}^{(N^f)}\Big[\frac{V^f(T)Q(T)}{N^f(T)}\Big|\mathcal{F}(t)\Big].$$

Note that $N^f(T) = M^f(T)Q(T)$, and $V^{(N^f)(t)} = \frac{V^f(t)}{M^f(t)}$. After simplifying, we get

$$\frac{V^f(t)}{M^f(t)} = \tilde{E}^{(N^f)}\Big[\frac{V^f(T)}{M^f(T)}\Big|\mathcal{F}(t)\Big].$$

Since $M^f(t) = e^{\int_0^t R^f(u)du}$ the conclusion follows.

Remark: Alternatively, the risk neutral price (in dollars) of this financial product is

$$V_t = \tilde{E}\Big[\frac{D(T)V^f(T)Q(T)}{D(t)}\Big|\mathcal{F}(t)\Big].$$

But we have

$$\begin{aligned}
\tilde{E}\Big[\frac{D(T)V^f(T)Q(T)}{D(t)}\Big|\mathcal{F}(t)\Big] &= \tilde{E}^{(N^f)}\Big[\frac{D(T)V^f(T)Q(T)}{D(t)}\frac{Q(t)M^f(t)D(t)}{Q(T)M^f(T)D(T)}\Big|\mathcal{F}(t)\Big] \\
&= \tilde{E}^{(N^f)}\Big[\frac{D^f(T)V^f(T)Q(t)}{D^f(t)}\Big|\mathcal{F}(t)\Big].
\end{aligned}$$

Thus dividing by $Q(t)$ on both sides gives

$$D^f(t)V^f(t) = \tilde{E}^{(N^f)}\Big[D^f(T)V^f(T)\Big|\mathcal{F}(t)\Big].$$

9.7 The exchange rate

Recall the exchange rate model. There is asset price $S(t)$, foreign exchange rate $Q(t)$, domestic money market rate $R(t)$ and foreign money market rate $R^f(t)$. Recall

$$N^f(t) = \exp\{\int_0^t R^f(u)\,du\}Q(t) \tag{9.29}$$

is the dollar value of one unit of the foreign money market account. The risk-neutral model when prices are in dollars is

$$\begin{aligned} dS(t) &= R(t)S(t)\,dt + \sigma_1(t)S(t)\,d\widetilde{W}_1(t) \\ dN^f(t) &= R(t)N^f(t)\,dt + N^f(t)\sigma_2(t)\left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right] \end{aligned} \quad (9.30)$$

The risk-neutral measure $\widetilde{\mathbf{P}}^{(N^f)}$ when prices are denominated using N^f as numéraire is given in Shreve, page 386, equation (9.3.17) and in the previous set of lecture notes. Here, we make some remarks concerning the exchange rate process $Q(t)$.

9.7.1 The exchange rate under the domestic risk-neutral measure

It follows from equations (9.29), (9.30) that

$$dQ(t) = [R(t) - R_f(t)]\,Q(t)\,dt + Q(t)\sigma_2(t)\left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right] \quad (9.31)$$

When dealing with Q alone it is convenient to write this in a simpler form. Define

$$\widetilde{W}_3(t) = \int_0^t \left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right].$$

Observe that

$$\begin{aligned} [d\widetilde{W}_3(t)]^2 &= \left[\rho(t)\,d\widetilde{W}_1(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2(t)\right]^2 \\ &= \rho^2(t)\,dt + (1-\rho^2(t))\,dt = dt. \end{aligned}$$

$\widetilde{W}_3(t)$ is a continuous martingale starting at 0 with quadratic variation t and so Lévy's theorem implies that $\widetilde{W}_3(t)$ is itself a Brownian motion. Using $\widetilde{W}_3$,

$$dQ(t) = [R(t) - R_f(t)]\,Q(t)\,dt + \sigma_2(t)Q(t)\,d\widetilde{W}_3(t). \quad (9.32)$$

Remark: The foreign exchange rate behaves exactly like a risky asset that pays dividends at rate $R^f(t)$. Equation (9.32) is the same as equation (5.5.6) in Shreve for a dividend-paying asset if $A(t)$ in that equation is replaced by $R^f(t)$.

9.7.2 Black-Scholes formula for a Call option on the exchange rate

Let $\sigma_2(t) = \sigma_2$ be constant, and also let $R(t) = r$ and $R^f(t) = r^f$ be constant. Then equation (9.32) becomes

$$dQ(t) = \left[r - r^f\right] Q(t)\, dt + \sigma_2 Q(t)\, d\widetilde{W}_3(t). \tag{9.33}$$

The solution to this equation is

$$Q(t) \quad = \quad Q(0) \exp\{\sigma_2 \widetilde{W}_3(t) + (r - r^f - \frac{1}{2}\sigma^2)t\}. \tag{9.34}$$

We can look at $Q(t)$ (*from a computational point of view*) as the Black-Scholes price of an asset following the geometric Brownian motion model, when the volatility is σ_2 and *the risk free rate is* $r^f - r$.

The fact that $Q(t)$ is a classical Black-Scholes price gives immediate formulas for options on the exchange rate in the constant coefficient case, which we will develop below.

Suppose that the risk free rate is r and under $\tilde{P}$, a stock S_t has dynamics:

$$dS_t = rS_t dt + \sigma S_t d\tilde{W}_t.$$

Let $C(T-t, x, K, r, \sigma)$ the price of at time t of a European call on S with strike K, conditioned on $S_t = x$. That is

$$C(T-t, x, K, r, \sigma) = \tilde{E}\Big(e^{-r(T-t)}(S_T - K)^+ \Big| S_t = x\Big).$$

Then the Black-Scholes formula for $C(T-t, x, K, r, \sigma)$ is

$$\begin{aligned} C(T-t, x, K, r, \sigma) \quad &= \quad e^{-r(T-t)}\tilde{E}\left[\left(xe^{\sigma\widetilde{W}(T-t))+(r-\sigma^2/2)(T-t)} - K\right)^+\right] \\ &= \quad xN\left(\frac{\ln(x/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\qquad - Ke^{-r(T-t)}N\left(\frac{\ln(x/K) + (r - \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

Consider now a European call option on $Q(T)$ at strike K, for the model of (9.34). This can also be looked at as a call option with strike K on *a unit of foreign currency, quoted in domestic currency.*

According to risk-neutral pricing, the value of this option at time t is

$$\begin{aligned} V(t) \quad &= \quad e^{-r(T-t)}\tilde{E}\left[(Q(T) - K)^+ \mid \mathcal{F}(t)\right] \\ &= \quad e^{-r^f(T-t)}e^{-(r-r^f)(T-t)}\tilde{E}\left[(Q(T) - K)^+ \mid Q(t)\right]. \end{aligned}$$

But evaluating $e^{-(r-r^f)(T-t)}\tilde{E}\left[(Q(T)-K)^+ \mid Q(t)\right]$ is exactly the same as evaluating the price of a European call when the risk free rate is $r-r^f$ and the volatility is σ_2. Therefore,

$$V(t) = e^{-r^f(T-t)}C(T-t, Q(t), K, r-r^f, \sigma_2).$$

This is called the Garman-Kohlhagen formula. You can also recover this formula from the formula (5.5.12) in Shreve for the price of a call on dividend-paying asset. Just replace a in this formula by r^f.

9.7.3 The exchange rate from the foreign currency viewpoint

Starting with the model (9.29)-(9.30), suppose we use the foreign currency money market $N^f(t)$ as the numéraire. In the previous lecture we found that

$$\widetilde{W}^{(N^f)}(t) = (\widetilde{W}_1(t) - \int_0^t \sigma_2(u)\rho(u)\,du, \widetilde{W}_2(t) - \int_0^t \sigma_2(u)\sqrt{1-\rho^2(u)}\,du)$$

is a Brownian motion under $\widetilde{\mathbf{P}}^{(N^f)}$ and we showed

$$dS^{(N^f)}(t) = S^{(N^f)}(t)\left[(\sigma_1(t) - \sigma_2(u)\rho(u))\,d\widetilde{W}_1^{(N^f)}(t) - \sigma_2(t)\sqrt{1-\rho^2(t)}\,d\widetilde{W}_2^{(N^f)}(t)\right]$$

To completely describe the model from the viewpoint of the foreign currency we should also look at the dollar to foreign currency exchange rate $1/Q(t)$, which is the value of one dollar in units of the foreign currency. The equation for this should have a form symmetrical to the equation (9.33) for $Q(t)$ when units are in dollars. Indeed,

$$d\left[\frac{1}{Q(t)}\right] = [R^f(t) - R(t)]\frac{1}{Q(t)}\,dt - \sigma_2(t)\frac{1}{Q(t)}\left[\rho(u)\,d\widetilde{W}_1^{(N^f)}(t) + \sqrt{1-\rho^2(t)}\,d\widetilde{W}_2^{(N^f)}(t)\right]. \tag{9.35}$$

This may be verified from Itô's rule, but one can see why it must be correct by the following reasoning. From the perspective of numéraire N^f, $R^f(t)$ is the domestic risk free rate and $R(t)$ is the domestic rate, so, where $R(t) - R^f(t)$ appears in (9.32), $R^f(t) - R(t)$ appears in (9.35). The volatility terms are essentially the same because the same stochastic fluctuation is obviously driving both $Q(t)$ and $1/Q(t)$. To explain why σ_2 appears in (9.32) but $-\sigma_2$ appears in (9.35) just note that $1/Q(t)$ goes down when Q goes up and vice-versa.

Concerning this topic, the student should read section 9.3.4 on Siegel's paradox (which is not really a paradox, but arises from a misunderstanding of the correct numéraire to use in interpreting a model.)

9.8 Zero coupon bonds as numéraire

In this section we assume given a risk-neutral model with a stochastic interest rate process $R(t), t \geq 0$.

9.8.1 Zero-coupon bonds

Bonds are financial instruments that promise fixed payoffs. Most bonds provide periodic payments called coupons and then a final payment consisting of a coupon and a lump sum called the *principal* or *face value*. A zero-coupon bond pays out only at the terminal time. We let $B(t,T)$ denote the price at time $t \leq T$ of a zero-coupon bond that pays \$1 at time T.

Given a risk-neutral model defined by a probability measure $\widetilde{\mathbf{P}}$, the no-arbitrage principle demands that $D(t)B(t,T)$ be a martingale in t up to time T. Since $B(T,T) = 1$, it follows that

$$B(t,T) = \frac{\tilde{E}[D(T)B(T,T) \mid \mathcal{F}(t)]}{D(t)} = \frac{\tilde{E}[D(T) \mid \mathcal{F}(t)]}{D(t)} = \frac{\tilde{E}\left[e^{-\int_0^T R(u)\,du} \mid \mathcal{F}(t)\right]}{e^{\int_0^t R(u)\,du}}. \tag{9.36}$$

Hence,

$$B(t,T) = \tilde{E}\left[e^{-\int_t^T R(u)\,du} \mid \mathcal{F}(t)\right] \tag{9.37}$$

This is an interesting formula. If $R(\cdot)$ is a random process, and we are at time t, we do do not know what R will be exactly after time t. But we do the market tells us what all zero-coupon bond prices are. Any model we create for R must be consistent with (in quant lingo, must be calibrated to) the zero-coupon bond prices via (9.37).

9.8.2 Forward prices

Suppose at time t, where $t < T$, Alice contracts to buy a unit of an asset from Bob at price F at time T. This is called a forward contract. No money changes hands at time t. Let $S(u)$ denote the price of the asset as a function of time u. From Alice's perspective she is getting an option that pays off $S(T) - F$, because she is purchasing something worth $S(T)$ dollars for F dollars at time T. The value of this option at t is $D^{-1}(t)\tilde{E}[D(T)(S(T) - F) \mid \mathcal{F}(t)] = S(t) - FB(t,T)$; remember, $D(t)S(t)$ is a martingale with respect to $\widetilde{\mathbf{P}}$! If she is paying or receiving no money for the contract at time t this value should be zero. Hence

$$F = \frac{S(t)}{B(t,T)}$$

is the fair price for this contract. It is called the T-forward price and denoted by $\mathrm{For}_S(t,T)$. Really, it is the price of $S(t)$ obtained using $B(t,T)$ as a numéraire.

A trivial but important observation is that the forward price and the market price concur at time T:

$$\mathrm{For}_S(T,T) = \frac{S(T)}{B(T,T)} = S(T). \tag{9.38}$$

9.8.3 The risk-neutral measure associated with the zero-coupon bond

Under the domestic risk neutral measure $\tilde{P}$, $D_t B(t,T)$ is a martingale. Therefore, $B(t,T)$ can be used as a numéraire. Indeed, the risk-neutral measure corresponding to numéraire $B(t,T)$, according to Theorem 3 of Lecture 9, is

$$\widetilde{\mathbf{P}}^T(A) = \tilde{E}[\mathbf{1}_A \frac{D(T)B(T,T)}{B(0,T)}] = \frac{1}{B(0,T)}\tilde{E}[\mathbf{1}_A D(T)]$$

We will call P^T, following Shreve (Definition 9.4.1), *the T-forward measure.*

Consider the special case in which the filtration in the risk neutral market is generated by a single Brownian motion $\widetilde{W}$. Then in this case we know from Theorem 9.1 of Shreve that there is a process $\nu_T(u)$ such that

$$\frac{D(T)}{B(0,T)} = e^{\int_0^T \nu_T(u)\, d\widetilde{W}(u) - \frac{1}{2}\int_0^T \nu_T^2(u)\, du}$$

and that $\widetilde{W}^T(t) = \widetilde{W}(t) - \int_0^t \nu_t(u)\, du$ is a Brownian motion under $\widetilde{\mathbf{P}}^T$. (In Shreve, 9.4.2, the notation $-\sigma^*(t,T)$ stands for our $\nu_T(t)$.

9.8.4 Pricing under the T-forward measure

Pricing under the domestic risk neutral measure with random interest rate

Suppose the interest rate is R_t, an adapted process. Then the risk neutral price V_t of a Euro style financial product that pays V_T at time T is

$$V_t = \tilde{E}\Big(e^{-\int_t^T R_u du} V_T \Big| \mathcal{F}_t\Big).$$

Since all we know about R_t is that it is an adpated process, we cannot go further with this pricing formula, unless we make some assumption on R_t (which is about modeling the interest rate, the topic of next Chapter). This is certainly a complex topic. Moreover, even if we have a model for R_t, it doesn't mean the pricing formula will be simple, if $\int_t^T R_u du$

has non zero correlation with V_T, for example. However, a nice observation here is that we do not have to compute this equation under $\tilde{E}$. Indeed, recall from the section 5.2 result, we have:

$$V_t^T = \tilde{E}^T\left(V_T^T \middle| \mathcal{F}_t\right).$$

where $V_t^T := \frac{V_t}{B(t,T)}$ is the price of the product denoted in the unit of zero-coupon bond. Note that since $B(T,T) = 1$, we have $V_T^T = V_T$.

The nice thing about the pricing formula under $\tilde{P}^T$ is that it is only a conditional expectation of the terminal value, not involving other quantities like the interest rate (this is not a pure gain, since the interest rate was absorbed into $\tilde{E}^T$). However, this suggests a new approach to the entire problem: we may directly model the assets under $\tilde{P}^T$, rather than under domestic measure $\tilde{P}$. Note that if we model the asset under $\tilde{P}^T$, then the unit of denomination (or the numéraire) is the price of zero coupon bond $B(t,T)$. In particular, if our objective is to model the stock price S_t (under $\tilde{P}$) then under $\tilde{P}^T$, we model

$$S_t^T := \frac{S_t}{B(t,T)} = For_S(t,T).$$

Pricing a call option on $S(t)$ under the domestic risk neutral measure is equivalent to pricing a call option on the forward price $For_S(t,T)$ under the T-forward measure. The advantage here is again about modeling. If we model under $\tilde{P}$ then necessarily we need to involve the model of R_t and need to know how to handle the expectation $\tilde{E}\left(e^{-\int_t^T R_u du} V_T \middle| \mathcal{F}_t\right)$. If we model under $\tilde{P}^T$ then we only need to model the forward price of S_t (which potentially maybe easier to calibrate to market parameters than modeling R_t) and then the expectation $\tilde{E}^T\left(V_T^T \middle| \mathcal{F}_t\right)$ is straight forward. The detailed computation is done in the following section.

Pricing a Call option under the T-forward measure

Here one assumes that the forward price of asset S, is given by the simple formula

$$d\text{For}_S(t,T) = \sigma \text{For}_S(t,T)\, d\widetilde{W}^T(t), \quad t \le T.$$

The point is that this is the Black-Scholes price model with $r = 0$, and if one looked at $S(t)$ under the original risk-neutral measure, it would not follow a Black-Scholes model with constant volatility. However it is possible to explicitly price a call. Indeed, let $C(T-t, x, K, r, \sigma)$ denote the Black-Scholes price of a European call when the price is x, the risk-free interest rate is r and the volatility is σ. Let $V(t)$ be the dollar price of the call. Then its forward price is $V^T(t) = V(t)/B(t,T)$. But, recalling from (9.38) that $\mathrm{For}_S(T,T) = S(T)$, we know from risk-neutral pricing that

$$\begin{aligned} V^T(t) &= \tilde{E}^T\left[\frac{(S(T)-K)^+}{B(T,T)} \,\Big|\, \mathcal{F}(t)\right] = \tilde{E}^T\left[(\mathrm{For}_S(T,T)-K)^+ \,\Big|\, \mathcal{F}(t)\right] \\ &= \tilde{E}^T\left[(\mathrm{For}_S(T,T)-K)^+ \,\Big|\, \mathrm{For}_S(t,T)\right]. \end{aligned}$$

But since $\mathrm{For}_S(t,T)$ follows the Black-Scholes price model with $r = 0$ and volatility σ,

$$V^T(t) = C(T-t, \mathrm{For}_S(t,T), K, 0, \sigma).$$

Hence,

$$V(t) = B(t,T)C(T-t, \mathrm{For}_S(t,T), K, 0, \sigma).$$

By substitution into the explicit formula for C (given above on page 2),

$$\begin{aligned} V(t) &= B(t,T)\mathrm{For}_S(t,T)N\left(\frac{\ln(\frac{\mathrm{For}_S(t,T)}{K}) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad - KB(t,T)N\left(\frac{\ln(\frac{\mathrm{For}_S(t,T)}{K}) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

This is essentially formula (9.4.9) in Shreve.

Clearly, this procedure could be applied to other cases where explicit pricing formulae are known for the Black-Scholes price model.

9.9 Miscellaneous - A related interview question

Consider the Black-Scholes formula:

$$V_0 = S_0 N(d+) - Ke^{-rT}N(d-).$$

What events are $N(d+), N(d-)$ probabilities of (and, maybe not obvious at first, under what measure)? This is an interview question that I was asked. The answer is as followed: by risk neutral pricing

$$\begin{aligned} V_0 &= \widetilde{E}(e^{-rT}(S_T-K)^+) \\ &= \widetilde{E}(e^{-rT}(S_T-K)\mathbf{1}_{\{S_T\geq K\}}) \\ &= \widetilde{E}(e^{-rT}S_T\mathbf{1}_{\{S_T\geq K\}} - Ke^{-rT}\mathbf{1}_{\{S_T\geq K\}}). \end{aligned}$$

Therefore it is clear that $N(d-)$ is the probability that $S_T \geq K$ under the risk neutral measure. What about $N(d+)$? The presence of the term $e^{-rT}S_T$ may make it seem like it is not a probability. However, if we note that $Z_t = \frac{e^{-rt}S_t}{S_0}$ is a martingale under the risk neutral measure with initial value $Z_0 = 1$, then we can write

$$\widetilde{E}(e^{-rT}S_T\mathbf{1}_{\{S_T \geq K\}}) = S_0\widetilde{E}(\frac{e^{-rT}S_T}{S_0}\mathbf{1}_{\{S_T \geq K\}}) = S_0P^{(S)}(S_T \geq K).$$

That is $N(d+)$ is the probability that $S_T \geq K$ under the measure using S itself as a numéraire.

CHAPTER 10 Interest rate models

10.1 Introduction

So far in this class, we have studied various financial derivatives connected with a stock model. The stock is a typical example of a risky asset. On the other hand, we also have the bond, whose price is directly related to the interest rate, which in turn influences the price of the risk free asset: the money market account. In this Chapter, we will study various models of the short rate and the forward rate, which leads us to the price of the bond.

10.1.1 Money market account versus zero-coupon bond

It is clear that both the money market account and the zero-coupon bond prices are determined by the interest rate. But just how the two are similar and how are they different?

When the interest rate is a deterministic constant r, then the price at time t of a zero-coupon bond is $B(t,T) = e^{-r(T-t)}$. This is the same as the price of a money market account that has initial deposit at time 0 equals $e^{-rT} = B(0,T)$. On the other hand, it's also clear that the price at time t of a money market account with initial deposit K is Ke^{rt}, which is also the price of Ke^{rT} shares of zero-coupon with maturity T.

The situation is not the same when the interest rate is stochastic. First, observe that if the interest rate is an adapted process $R(t)$, then the value of the money market with an initial deposit K at time T is

$$M(T) = K \exp(\int_0^T R_u du),$$

and is random. Thus one cannot determine an initial deposit amount so that $M(T) = B(T,T) = 1$: the money market account cannot replicate a zero-coupon bond.

Similarly, the price at time 0 of a zero-coupon bond is

$$B(0,T) = \tilde{E}\left(e^{-\int_0^T R_u du}\right),$$

where $\tilde{E}$ is the expectation under a risk neutral measure. So without a prior assumption (or a model) of $R(t)$ under the risk neutral measure $\tilde{P}$, one cannot compute what $B(0,T)$ is (We say the bond price $B(0,T)$ is determined by the market).

An interesting question is by observing $B(t,T)$ for all $0 \le t \le T$, can one determine what $R(t)$ is? I believe for a fixed T, the answer is no. However, if we know the price of $B(t,T)$ for various maturity T, then we can deduce what $R(t)$ is, see the next section.

10.1.2 Various rates connected with bond

Since the payment of the bond with maturity T at time T is fixed, one can use the bonds (with various maturities, if necessary) to “lock-in" certain interest rates. Thus, zero-coupon bond prices are used as the standard for calculating interest rates. Throughout, it is assumed we deal with the market for risk free bonds and loans. The price at time $t \le T$ of a zero-coupon bond that pays \$1 at time T shall always be denoted by $B(t,T)$. Notice that $B(T,T) = 1$. There are various interest rates associated with B.

(i) Continuous compounding:

In this discussion all interest rates are quoted assuming continuous compounding. Consider an account which at time S has \$$L_S$ and at time $T > S$ has \$$L_T$, where S and T are measured in years. Then the interest rate r, per annum, continuously compounded, earned over $[S,T]$ is determined by the equation $L_S e^{r(T-S)} = L_T$, or

$$r = \frac{1}{T-S} \ln\left(\frac{L_T}{L_S}\right) \tag{10.1}$$

(ii) The spot rate (yield to maturity)

The *zero rate* for the period $[t,T]$, also called the *spot rate*, or, more precisely, the *continuously compounded spot rate for the period* $[t,T]$ is the function which gives the interest rate of a zero coupon bond over the interval $[t,T]$. That is, if we denote this rate by $R(t,T)$ then

$$1 = B(t,T)\exp\Big(R(t,T)(T-t)\Big).$$

From which it follows that

$$R(t,T) = \frac{1}{T-t}\ln\left(\frac{1}{B(t,T)}\right) = -\frac{\ln B(t,T)}{T-t} \tag{10.2}$$

Thus

$$B(t,T) = e^{-(T-t)R(t,T)}. \tag{10.3}$$

For a fixed t, a plot of $R(t,T)$ as a function of T is called a spot rate curve. It gives the (continuously compounded) interest rates available for risk free zero coupon bonds for all maturities starting from t. The notable fact about the spot rate curve is that it is not constant—normally it tends to be upward sloping. This phenomenon is called the term structure of interest rates. $R(t, T)$ is also referred to as the yield to maturity of the zero coupon bond at time t for the *time to maturity* $\tau = T - t$.

(iii) The forward rate

Consider times $t < S < T$. Suppose at time t we would want to lock in certain spot rate for the time interval $[S, T]$. Let's call this rate $F(t, S, T)$. Then clearly this rate must be related with the bond price $B(t, S)$ and $B(t, T)$. So we should determine what $F(t, S, T)$ is and further inquire into whether we can indeed lock in this rate at time t by trading certain shares of the bonds with maturities at S and T.

To answer the first question, clearly what we want is if we invest 1 dollar at time S then we should receive $\exp\Big(F(t, S, T)(T - S\Big)$ at time T. Note that, 1 dollar at time S is equivalent to $B(t, S)$ at time t and $\exp\Big(F(t, S, T)(T - S)\Big)$ dollars at time T is equivalent to $B(t, T) \exp\Big(F(t, S, T)(T - S)\Big)$ at time t. These clearly should be equal if we want to lock in the rate $F(t, S, T)$ at time t. Therefore

$$B(t, S) = B(t, T) \exp\Big(F(t, S, T)(T - S)\Big),$$

or

$$F(t; S, T) = \frac{1}{T - S} \ln\left(\frac{B(t, S)}{B(t, T)}\right) = -\frac{\ln B(t, T) - \ln B(t, S)}{T - S}.$$

To answer the second question, first plug

$$F(t, S, T) = -\frac{\ln B(t, T) - \ln B(t, S)}{T - S}$$

into $B(t, T) \exp\Big(F(t, S, T)(T - S)\Big)$ and observe that

$$B(t, T) \exp\Big(F(t, S, T)(T - S)\Big) = B(t, T) \frac{B(t, S)}{B(t, T)}.$$

This suggests that we should hold $\frac{B(t,S)}{B(t,T)}$ shares of bond with maturity T at time t. To finance this position, we should sell 1 share of bond with maturity S at time t (since we expect to invest 1 dollar at time S). This turns out to be the right scheme because this costs nothing at time t: If at t we sell one zero-coupon bond maturing at S for $B(t, S)$

and with this money buy $B(t,S)/B(t,T)$ zero-coupon bonds maturing at T, the net value of this transaction for us is 0. At time S we pay out a dollar and at time T receive $B(t,S)/B(t,T)$. This is indeed equivalent to earning, at T, the amount $B(t,S)/B(t,T) = \exp\Big(F(t,S,T)(T-S)\Big)$ from a deposit of \$ 1 at S. The rate of interest earned by this transaction, $F(t,S,T)$ is called the *forward rate for* $[S,T]$ *contracted at* t.

Remark: $F(t,S,T)$ is known at time t by observing $B(t,S)$ and $B(t,T)$, that is $F(t,S,T) \in \mathcal{F}_t$, where $\mathcal{F}_t$ is the filtration generated by $B(t,S)$ and $B(t,T)$.

(iv) The instantaneous forward rate

The forward rate $F(t,S,T)$ has the formula

$$F(t;S,T) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}.$$

If we let T goes to S, then the right hand side should go to $-\frac{\partial}{\partial T}\ln[B(t,S)]$, if the derivative exists. Indeed, if we assume $B(t,T)$ is differentible in T, then this is the case. This is not an unreasonable assumption since for a fixed t, one can believe that the bond price is a smooth function of different maturities. (On the other hand, for a fix maturity T, the bond price should *not* be a smooth function of t. It should be very irregular, indeed, in t, similar to behavior of the graph of a Brownian motion in t).

So we define the *instantaneous forward rate at* t *for investing at time* T as

$$f(t,T) = -\frac{\partial}{\partial T}\ln[B(t,T)]. \tag{10.4}$$

By integrating in T, it follows that

$$B(t,T) = \exp\{-\int_t^T f(t,u)\,du\} \tag{10.5}$$

For brevity, we refer to $f(t,T)$ as the forward rate function.

Remark: Again, note that here $f(t,T)$ is known at time t, that is $f(t,T) \in \mathcal{F}_t$ where $\mathcal{F}_t$ is the filtration generated by $B(t,T)$.

(v) The short rate

The *short rate* is the rate available at time t for the shortest period loans. Formally it is defined as

$$R(t) = f(t,t) \tag{10.6}$$

Remark: It seems reasonable to define $R(t) = f(t,t)$ (just from the understanding of what the forward rate is). First, it is reasonable to believe the spot rate $R(t,T)$ should converge to the short rate $R(t)$ when $T \to t$. Recall

$$R(t,T) = -\frac{\ln B(t,T)}{T-t} = \frac{\int_t^T f(t,u)du}{T-t}.$$

The RHS converges to $f(t,t)$ (by Lesbegue differentation theorem) as $T \to t$. So if we expect $R(t,T)$ to converge to $R(t)$, then it is reasonable to set $R(t) = f(t,t)$.

Second, from comparing the risk neutral pricing formula

$$B(t,T) = \tilde{E}\big(e^{-\int_t^T R(u)du}\big)$$

with the definition of the forward rate:

$$B(t,T) = e^{-\int_t^T f(t,u)du},$$

we should expect $R(t) = f(t,t)$ as well. Indeed, suppose $R(t) > f(t,t)$. Then if we suppose $R(u)$ and $f(t,u)$ are continuous functions of u (which is reasonable) then there must exist some $T > t$ so that $R(u) > f(t,u)$ for $u \in [t,T]$. But then we have

$$B(t,T) = e^{-\int_t^T f(t,u)du} > \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big) = B(t,T),$$

which is a contradiction. So this cannot happen.

10.1.3 Remarks on modeling $B(t,T), R(t), f(t,T)$

At each t, the market presents us with the function $B(t,t+s)$, $s \geq 0$, or, equivalently, $f(t,t+s)$, $s \geq 0$, capturing, at each time t, the return on zero-coupon bonds of all maturities. As a function of s, this term structure of interest rates fluctuates as t changes, and, we regard these fluctuations as random, because we cannot predict them exactly for future values of t. Developing good stochastic process models for the term structure of interest rates is a major area of mathematical finance. These models are used to analyze and price derivatives that depend on credit markets.

A few, loosely stated principles guide the construction of the basic models covered in this course. First, the models should be simple enough for fairly explicit calculation or at least easy simulation. Second, they should be rich enough that they can be calibrated to the market; that is, it should be possible to choose the parameters of the model so that its statistical behavior mimics reasonably well actual market performance. Of course, these two criteria push in opposite directions—the richer the model, the harder it is to analyze and simulate—and one must strike a good balance between them. A third important principle is that the model should not admit arbitrage.

From the previous section, all three processes $B(t,T), R(t), f(t,T)$ are objects of interest in modeling and we would like to obtain models for all three of them (for a fixed maturity T, as a process in t). It is also clear that if we get a model for one then the other two can be deduced out of it, via the relations:

$$\begin{aligned} B(t,T) &= e^{-\int_t^T f(t,u)du} \\ B(t,T) &= \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big) \\ R(t) &= f(t,t). \end{aligned}$$

But there are subtle differences in which process we choose to model actually. First suppose we want to model $B(t,T)$ (which means we fix the maturity T and model $B(t,T)$ as a process in t). And let's say we go with the Geometric framework:

$$dB(t,T) = \alpha(t,T)B(t,T)dt + \sigma(t,T)B(t,T)dW_t,$$

under the physical measure P, where $B(0,T)$ is assumed known. The question is what should α and σ be? We recognize that they cannot be just any processes because we have the contraint:

$$B(T,T) = 1.$$

Indeed, unless for very trivial choices ($\sigma = 0$, α a constant) the terminal constraint cannot be easily satisfied. So modeling $B(t,T)$ as a function of t does not seem straightforward.

Note that the constraint $B(T,T) = 1$ is naturally built into the formula

$$B(t,T) = e^{-\int_t^T f(t,u)du}.$$

Thus, a good idea is to model the forward rate $f(t,T)$ and then derive $B(t,T)$ via the above formula. This is the HJM model and will be discussed in the next lecture.

The last approach is of course to model $R(t)$ and derive $B(t,T)$ via the relation

$$B(t,T) = \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big).$$

The subtle point to note here is if we take this approach, then *necessarily we need to model $R(t)$ in under the risk neutral measure $\tilde{P}$ to obtain a price for $B(t,T)$*. The reason is this: if we model $R(t)$ in the physical probability P, say

$$dR(t) = \alpha(t)dt + \sigma(t)dW_t,$$

under P; then we do not have a market price of risk equation: there is one random source and there is no asset here. (Remember that we do not have the bond price dynamics under P under this approach, whose goal is to induce the bond price from the risk neutral pricing formula). Starting to model $R(t)$ under the risk neutral measure doesn't seem unreasonable; all it takes is to declare

$$dR(t) = \alpha(t)dt + \sigma(t)d\tilde{W}_t,$$

under $\tilde{P}$. But this may pose a problem for model calibration when we need to determine α, σ. The reason is we live in the physical world, i.e. we observe distribution under P. Nevertheless, modeling $R(t)$ under the risk neutral measure is an approach that many have taken and it leads to a connection with the forward rate HJM model, which can be modeled under the physical measure P. So in this lecture, we will discuss modeling the short rate $R(t)$ under the risk neutral measure and leave the connection with the forward rate for the next lecture.

The last question you may ask is, if we do not have the market price of risk equation, then who determines the risk neutral measure $\tilde{P}$? The answer is: the market does. I.e., it decides the bond price $B(t,T)$, which in turns imply what the risk neutral measure $\tilde{P}$ is (This is the answer given by Björk in his book: Arbitrage theory in continuous time). I'll leave it to you to ponder more about the meaning of this answer.

In both approaches (modeling $R(t)$ and $f(t,T)$), the models we present have an *ad hoc* flavor. It would seem more reasonable to build models supported by some theory of how the economy works. Such models would try to incorporate in a quantitative manner the economic factors and indicators that influence term structure. But we shall proceed innocent of all economic theory. We just look for models based on stochastic differential equations that we hope are rich enough to capture actual market behavior.

10.2 Multi-factor short rate models

A multi-factor model will consist of a vector-valued process

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_m(t) \end{pmatrix},$$

that solves a stochastic differential equation and is a Markov process under the risk-neutral measure, and a function $\Phi(x_1, \dots, x_m)$. The short rate is then **defined** by

$$R(t) = \Phi(X_1(t), \dots, X_m(t)).$$

The factors $X_i, i = 1, \cdots, m$ are meant to model economic factors that might influence the interest rate, such as GDP, import-export rate, inflation etc. A particularly simple choice for $\Phi(X_1(t), \dots, X_m(t))$ would be $R(t) = \delta_0(t) + \sum_{i=1}^m \delta_i(t)X_i(t)$. That is $R(t)$ is just a linear combination of the factors. One can then use linear regression to determine $\delta_i, i = 0, \cdots, m$.

Recall that the relation between the bond price and the short rate is via the equation

$$B(t,T) \;=\; \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big) \qquad (10.7)$$

In the short-rate approach, being able to analyze the model comes down to calculating the conditional expectation in formula (10.7).

The key in evaluating this conditional expectation is an old idea that we have used over and over again in this course:

(i) Construct a model for the process $R(t)$ so that it is Markovian. Then (10.7) becomes

$$B(t,T) \;=\; \tilde{E}\left[\exp\{-\int_t^T \Phi(X_1(s),\ldots,X_m(s))\,ds\} \;\Big|\; X_1(t),\ldots,X_m(t)\right].$$

Thus we can find a function $c(t,x_1,x_2,...,x_m)$ so that $c\big(t,X_1(t),X_2(t),...,X_m(t)\big) = B(t,T)$.

(ii) Find a PDE that $c(t,x_1,x_2,...,x_n)$ satisfies. If we can solve the PDE (via numerical procedure, for example) then we can recover the bond price $B(t,T)$ as described in (i).

10.2.1 Affine-yield model

Definition 10.2.1. *The short rate model is called an affine-yield model if it turns out that the zero-coupon bond price can be written as*

$$B(t,T) = \exp\{-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T)\}, \qquad (10.8)$$

for some functions $C_1(t,T),\ldots,C_m(t,T)$, and $A(t,T)$.

Remark: Recall that the spot rate, or the yield to maturity $R(t,T)$ is defined such that

$$B(t,T) = e^{-(T-t)R(t,T)}.$$

Compare this with (10.8) we see that an affine yield model is such that the yield to maturity $R(t,T)$ is an affine combination of the factors $X_i, i = 1,\cdots,m$. The Vasicek, CIR and Hull-White models discussed below are affine-yield models.

The case in which $m = 1$ is called a *single-factor model*. In such a model one takes $R(t)$ itself to be a Markov process; no auxiliary process $\mathbf{X}(t)$ is defined.

10.2.2 Connection with the forward rate

Affine yield models are particularly nice because it is easy to read off of an affine model a model for the instantaneous forward rate:

$$\begin{aligned} B(t,T) &= \exp\{-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T)\} \\ &= \exp(-\int_t^T f(t,u)du). \end{aligned}$$

Thus

$$-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T) = -\int_t^T f(t,u)du.$$

By differentiating both sides of the equation with respect to T, (assuming $C_i(t,T)$ and $A(t,T)$ are differentiable w.r.t T) we have a model for $f(t,T)$.

The obvious question is then how can we come up with candidates for affine yield models? We'll give an idea of how this is done in the two factor model in the section below. The generalization of this procedure for multi-factor model is straight forward.

10.3 Affine yield model in general

We illustrate the idea for two factor models under the risk neutral measure $\tilde{P}$,

$$\begin{aligned} dX_1(t) &= a_1(t,X_1(t),X_2(t))\,dt + b_{11}(t,X_1(t),X_2(t))\,d\widetilde{W}_1(t) + b_{12}(t,X_1(t),X_2(t))\,d\widetilde{W}_2(t) \\ dX_2(t) &= a_2(t,X_1(t),X_2(t))\,dt + b_{21}(t,X_1(t),X_2(t))\,d\widetilde{W}_1(t) + b_{22}(t,X_1(t),X_2(t))\,d\widetilde{W}_2(t) \end{aligned}$$

There is a lot of freedom in this general set-up, and we will quickly be more specific. But an easy first observation is that to obtain an affine yield model, we need $R(t)$ to be an affine function of X_1, X_2.

Thus we assume that $R(t) = \delta_0(t) + \delta_1(t)X_1(t) + \delta_2(t)X_2(t)$, where $\delta_i(t), i = 0, ...2$ are parameters of the model that would be determined by model calibration.

Because (X_1, X_2) is a Markov process, we know that $B(t,T) = g(t, X_1(t), X_2(t))$ for some function g.

To find an equation for g we start with the observation that

$$D(t)g(t,X_1(t),X_2(t)) = D(t)B(t,T)$$

is a martingale under the risk-neutral measure and that

$$g(T,X_1(T),X_2(T)) = B(T,T) = 1.$$

Apply Ito's formula,

$$
\begin{aligned}
d[D(t)B(t,T)] &= D(t)\mathcal{L}g(t,X_1(t),X_2(t))\,dt \\
&\qquad + D(t)\mathcal{M}_1 g(t,X_1(t),X_2(t))\,d\widetilde{W}_1(t) + D(t)\mathcal{M}_2 g(t,X_1(t),X_2(t))\,d\widetilde{W}_2(t),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{L}g(t,x_1,x_2) &= -(\delta_0+\delta_1(t)x_1+\delta_2(t)x_2)g + g_t + a_1(t,x_1,x_2)g_{x_1} + a_2(t,x_1,x_2)g_{x_2} \\
&\quad +\frac{1}{2}[b_{11}^2+b_{12}^2](t,x_1,x_2)g_{x_1x_1} + [b_{11}b_{21}+b_{12}b_{22}](t,x_1,x_2)g_{x_1x_2} \\
&\quad + \frac{1}{2}[b_{21}^2+b_{22}^2](t,x_1,x_2)g_{x_2x_2}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{M}_1 g(t,X_1(t),X_2(t)) &= b_{11}(t,x_1,x_2)g_{x_1} + b_{21}(t,x_1,x_2)g_{x_2} \\
\mathcal{M}_2 g(t,X_1(t),X_2(t)) &= b_{12}(t,x_1,x_2)g_{x_1} + b_{22}(t,x_1,x_2)g_{x_2}.
\end{aligned}
$$

(In these expressions we have omitted writing the argument (t,x_1,x_2) of g and its partials.) In order that $D(t)g(t,X_1(t),X_2(t))$ be a martingale, g must be a solution of the parabolic pde,

$$
\mathcal{L}g(t,x_1,x_2)=0, \quad t\le T, \quad g(T,x_1,x_2)=1. \tag{10.9}
$$

The following is the key observation:

If all the coefficients of the operator $\mathcal{L}$—that is

$$
a_1(t,x_1,x_2), a_2(t,x_1,x_2), [b_{11}^2+b_{12}^2](t,x_1,x_2), \text{ etc.}
$$

are *affine functions*, i.e. functions of the form

$$
\eta_0(t,T)+\eta_1(t,T)x_1+\eta_2(t,T)x_2,
$$

then (10.9) has a solution of the affine-yield form. That is, we can find $\alpha(t,T), c_1(t,T), c_2(t,T)$ such that

$$
\begin{aligned}
g(t,x_1,x_2) &= \exp\{-c_1(t,T)x_1 - c_2(t,T)x_2 - \alpha(t,T)\}, \quad \text{with} \\
c_1(T,T) &= c_2(T,T) = \alpha(T,T) = 0.
\end{aligned}
$$

Remark:

a. We require the $b_{11}^2+b_{12}^2$ etc. to be affine, *not* b_{11} or b_{12} themselves. This explains the choice of the volatility in Vasicek model: constant in x_i and the choice of volatility in CIR model: $\sqrt{x_i}$ for b_{ii} and 0 for $b_{ij}, i\neq j$.

b. If $\eta_0(t,T), \eta_1(t,T), \eta_2(t,T)$ are constants, that is the coefficients in the affine form of $a_i(t, x_1, x_2$ etc. are constants, (which is the case for the standard Vasicek and CIR models) then we can check that $c_i(t,T)$ takes the form $c_i(T-t)$, similarly for $\alpha(t,T)$. That is their dependence on the two variables t, T is only on the difference $T-t$.

The conditions $c_1(T,T) = c_2(T,T) = \alpha(T,T) = 0$ imply that this function automatically satisfies the boundary condition $g(T, x_1, x_2) = B(T,T) = 1$.

Moreover, for this g, one can see by direct calculation that

$$\mathcal{L}g = D_1(t,T)x_1 g + D_2(t,T)x_2 g + D_3(t,T)g,$$

where $D_1(t,T), D_2(t,T), D_3(t,T)$ are defined **in terms of** $c_1(t,T), c_2(t,T), \alpha(t,T)$ and their first derivatives (in t).

Since $\mathcal{L}g(t, x_1, x_2) = 0$ we need the RHS of the above equation to be equal to 0 as well. This is accomplished by setting $D_i(t,T) = 0$, for all i.

So by setting $D_1(t,T) = 0$, $D_2(t,T) = 0$, $D_3(t,T) = 0$, *we obtained equations for determining* $c_1(t,T), c_2(t,T), \alpha(t,T)$ so that $\exp\{-c_1(t,T)x_1 - c_2(t,T)x_2 - \alpha(t,T)\}$ indeed solves (10.9).

If you examine the multi-factor CIR model or mixed models you will see that they are formulated exactly so that the coefficients of $\mathcal{L}g$ are affine. Following the derivation of the affine-yield expressions for these models in Shreve and doing Exercise 10.2 will help you understand this overall strategy.

10.4 One factor Hull-White model

10.4.1 Explicit formula for $R(t)$

The one-factor Hull-White model is defined by a short rate which solves a linear differential equation

$$dR(t) = k(\mu - R(t))\,dt + \sigma\,d\widetilde{W}(t),$$

where $\widetilde{W}$ is a Brownian motion under the risk-neutral measure. To find the explicit solution for $R(t)$, note that

$$d(e^{kt}R(t)) = e^{kt}(kR(t)dt + dR(t)).$$

Thus we get

$$e^{kt}R(t) = R(0) + \int_0^t k\mu e^{ks}ds + \int_0^t \sigma e^{ks}d\widetilde{W}_s.$$

And

$$R(t) = e^{-kt}R(0) + \mu(1 - e^{-kt}) + e^{-kt}\int_0^t \sigma e^{ks} d\widetilde{W}_s.$$

From this formula we can see that as t approaches infinity, $R(t)$ is centered around μ with variance $\frac{\sigma^2}{2k}$. The formula also explains why we refer to this model as mean-reverting and k as the rate of mean reversion.

10.4.2 Explicit formula for $B(t,T)$

From the risk neutral pricing formula

$$B(t,T) = \tilde{E}[e^{-\int_t^T R(u)du} | \mathcal{F}_t].$$

Note that for $u \geq t$

$$R(u) = e^{-k(u-t)}R(t) + \mu(1 - e^{-k(u-t)}) + e^{-k(u-t)}\int_t^u \sigma e^{k(s-t)} d\widetilde{W}_s.$$

We need to integrate

$$\begin{aligned}\int_t^T e^{-k(u-t)}\int_t^u \sigma e^{k(s-t)} d\widetilde{W}_s du &= \int_t^T \int_t^u e^{-k(u-t)}\sigma e^{k(s-t)} d\widetilde{W}_s du \\ &= \sigma\int_t^T \int_t^u e^{-k(u-s)} d\widetilde{W}_s du\end{aligned}$$

Accepting the fact that change of order of integration applies to this double integral of $d\widetilde{W}_s$ and du we have

$$\sigma\int_t^T \int_t^u e^{-k(u-s)} d\widetilde{W}_s du = \sigma\int_t^T \int_s^T e^{-k(u-s)} du d\widetilde{W}_s.$$

The above stochastic integral is independent of $\mathcal{F}_t$ and has normal distribution with mean 0 and variance

$$\begin{aligned}A(T-t) &= \sigma^2\int_t^T (\int_s^T e^{-k(u-s)}du)^2 ds \\ &= \frac{\sigma^2}{k^2}\Big[T - t + \frac{2}{k}(1 - e^{-k(T-t)}) + \frac{1}{2k}(1 - e^{-2k(T-t)})\Big].\end{aligned}$$

Thus the conditional distribution of the integral $\int_t^T R_u du$ on $\mathcal{F}_t$ has normal distribution with mean

$$B(T-t) = -\Big[R(t)\frac{1}{k}(1 - e^{-k(T-t)}) + \mu(T-t) - \frac{\mu}{k}(1 - e^{-k(T-t)})\Big]$$

and variance $A(T-t)$.

Thus by the Independence Lemma and the moment genrating function of a Normal random variable we have

$$B(t,T) = e^{B(T-t)+\frac{1}{2}A(T-t)}.$$

10.4.3 Model calibration

By model calibration we mean making the choices for μ, k, σ in the Hull-White model so that the resulting bond prices $B(t,T)$ most closely match the market data.

It should be remarked that we possibly do not observe the short rate $R(t)$ directly in the market; if by $R(t)$ we mean the process such that

$$B(t,T) = \tilde{E}(e^{-\int_t^T R(u)du}|\mathcal{F}_t).$$

There are some proxys for the risk free short rate, such as the OIS (overnight interest rate swap) but strictly speaking this is not the rate $R(t)$ we have in mind when plugging in the pricing formula for the treasury bond $B(t,T)$. This is why even though we have the dynamics of $R(t)$ directly dependent on μ, k, σ we cannot use that dynamics to calibrate these parameters. The yield to maturity $R(t,T)$ on the other hand, is readily observable as a function of the bond price $B(t,T)$.

Observe that the previous formula for $B(t,T)$ can be written as

$$B(t,T) = e^{-C(T-t)R(t)-D(T-t)}$$

where

$$\begin{aligned} C(T-t) &= \frac{1}{k}(1-e^{-k(T-t)}) \\ D(T-t) &= \mu(T-t) - \frac{\mu}{k}(1-e^{-k(T-t)}) - \frac{1}{2}A(T-t). \end{aligned}$$

Recall that in terms of the yield to maturity $R(t,T)$

$$B(t,T) = e^{-R(t,T)(T-t)}.$$

Thus we conclude that

$$R(t,T) = \frac{C(T-t)}{T-t}R(t) + \frac{D(T-t)}{T-t}.$$

Now we start to look at $R(t,T)$ as dependent instead on t and the time to maturity $\tau = T-t$ and we keep τ *fixed* for varying t. This is subtle conceptually because for $s \neq t, R(s,\tau)$ and $R(t,\tau)$ refers to the yields to maturity of *different* bonds with the same time to maturity τ. Nevertheless these quantities are readily observable in the market and we have

$$dR(t,\tau) = \frac{C(\tau)}{\tau} dR(t) + \frac{D(\tau)}{\tau}$$

Thus

$$dR(t,\tau) = k(\tau)(\mu(\tau) - R(t,\tau))dt + \sigma(\tau)d\widetilde{W}_t,$$

where the coefficients $k(\tau), \mu(\tau), \sigma(\tau)$ depend only on τ, k, μ, σ. Observe that $R(t,\tau)$ is again a mean-reverting process like $R(t)$. Since we can observe $R(t,\tau)$ at different t, there are readily availble methods to estimate the coefficients $k(\tau), \mu(\tau), \sigma(\tau)$ (see e.g. calibration for Ornstein-Uhlenbeck process using linear regression technique) from which we can get estimates of k, μ, σ.

Remark: The dynamics for $R(t,\tau)$ given above holds for any τ. In practice for each time to maturity τ we have a different set of data $R(t,\tau), 0 \le t \le T$ for example. Each set of data potentially gives rise to a different estimate of k, μ, σ. Thus we may have the issue of having too much data for too few parameters. Or put it in another way: the Hull-White model may not be flexible enough to fit the market term structure. A solution for this is of course to increase the number of parameters, by going to the multi-factor setting. However, it's good to keep in mind that since the bond market is huge, with many different time to maturity it probably is not easy to come up with a model that captures the term structure perfectly across the spectrum, from the short end to the long end.

10.5 Multi-factor Vasicek models

Multi-factor Vasicek models generalize the Hull-White model to the multi-factor case. In these models, $\mathbf{X}$ solves a linear stochastic differential equation with constant coefficients:

$$d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + B\,d\widetilde{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0 \tag{10.10}$$

where bX_0 is a given initial value, $\widetilde{W}$ is a multi-dimensional Brownian motion under the risk-neutral measure, and

$$R(t) = \delta_0 + \delta_1 X_1(t) + \cdots + \delta_m X_m(t).$$

Note that the choice of the liner coefficients $\delta_i, i = 0, \cdots, m$ in $R(t)$ are constants (not depending on t or ω). This is also a siginifcant simplifcation compared with the general multifactors model we proposed above.

Note also that no constant term $c\,dt$ enters (10.10). This causes no loss of generality if A is invertible, as we usually assume, because the constant term can be removed by an affine change of variables, as was shown in subsection (10.7.6) of this lecture.

Remark: The multi-factor Vasicek models have the advantage of having an explicit solution. More precisely, we can solve for an explicit formula for the factors $\mathbf{X}(t)$ - see the dicussion in section (10.7). This in turn leads to *explicit computation* for $B(t,T)$, which we will discuss below.

On the other hand, if one wants a model so that the short rate $R(t)$ is non-negative, then the Vasicek model may not satisfy this condition (since $X(t)$ may be negative). Thus if imposing this non-negativity condition takes piority over the explicit solution, then one should go with the CIR model, discussed below. The CIR model does not have an explicit solution, but the interest rate $R(t)$ under the CIR model is guaranteed to be non-negative.

Example *Two-factor Vasicek; Canonical form.* The canonical form of the two-factor Vasicek is obtained by a linear change of variables to get an equivalent system with as few free parameters as possible. It is derived in Shreve assuming that W is a 2-dimensional Brownian motion and that A and B are invertible. The factors solve the system of Example 2 in Section (10.7.4) with $\sigma_1 = \sigma_2 = 1$:

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix} \tag{10.11}$$

The short rate is

$$R(t) = \delta_0 + \delta_1 X_1(t) + \delta_2 X_2(t) = \delta_0 + (\delta_1, \delta_2) \cdot \mathbf{X}(t). \qquad \diamond$$

10.5.1 Explicit formula for $B(t,T)$ under the Vasicek model

In this subsection, we will obtain an explicit formula for $B(t,T)$ under the Vasicek model. This requires some details about stochastic calculus in multi-dimensional model. These details will be presented in section (10.7).

Recall that the factors under Vasicek model have the following dynamics

$$d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + B\,d\widetilde{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0 \tag{10.12}$$

Recall that $\{\mathbf{X}(u);\, u \geq 0\}$ is a Gaussian process. Because of this, it can be shown that the conditional distribution of $\int_t^T R(u)\,du$ given $(X_1(t), \ldots, X_m(t))$ is Gaussian with a mean of the form

$$\tilde{E}[\int_t^T R(u)\,du \,\Big|\, X_1(t), \ldots, X_m(t)] = C_1(T-t)X_1(t) + \cdots + C_m(T-t)X_m(t) + (T-t)\delta_0,$$

where

$$(C_1(\tau), \dots, C_m(\tau)) = \int_0^\tau (\delta_1, \cdots, \delta_m) e^{A \cdot u}\, du,$$

and with variance $\bar{A}(T-t)$, where

$$\bar{A}(\tau) = \int_0^\tau (C_1(u), \dots, C_m(u)) B B^* (C_1(u), \dots, C_m(u))^*\, du.$$

Here B^* denotes the transpose of B (the volatility in equation (10.12), don't confuse B with the bond price). $(C_1(u), \dots, C_m(u))^*$ is the column vector that is the transpose of the row vector $(C_1(u), \dots, C_m(u))$.)

Indeed, the explicit solution to equation (10.12) (given $\mathbf{X}(t)$ - see equation (10.24)) is

$$\mathbf{X}(u) = e^{A \cdot (u-t)} \mathbf{X}(t) + \int_t^u e^{A \cdot (u-s)} B\, d\tilde{W}(s).$$

Therefore

$$\begin{aligned} \int_t^T R(u) du &= \int_t^T \delta \cdot \mathbf{X}(u) du \\ &= \Big\{ \int_t^T \delta \cdot e^{A \cdot (u-t)} du \Big\} \mathbf{X}(t) + \int_t^T \int_t^u \delta \cdot e^{A \cdot (u-s)} B\, d\tilde{W}(s) du. \end{aligned}$$

By switching the order of integration between $d\tilde{W}(s)$ and du, we have

$$\begin{aligned} \int_t^T \int_t^u \delta \cdot e^{A \cdot (u-s)} B\, d\tilde{W}(s) du &= \int_t^T \int_s^T \delta \cdot e^{A \cdot (u-s)} B du\, d\tilde{W}(s) \\ &= \int_t^T \int_0^{T-s} \delta \cdot e^{A \cdot u} B du\, d\tilde{W}(s) \\ &= \int_t^T \mathbf{C}(T-s) B\, d\tilde{W}(s), \end{aligned}$$

where

$$\mathbf{C}(\tau) := (C_1(\tau), \dots, C_m(\tau)) = \int_0^\tau (\delta_1, \cdots, \delta_m) e^{A \cdot u}\, du,$$

is defined above.

We have

$$\int_t^T \mathbf{C}(T-s) B\, d\tilde{W}(s)$$

has Normal distribution with mean 0 and variance matrix

$$\int_t^T \mathbf{C}(T-s)BB^*\mathbf{C}^*(T-s)ds = \int_0^{T-t} \mathbf{C}(s)BB^*\mathbf{C}^*(s)ds.$$

Thus the distribution of $\int_t^T R(u)du = \int_t^T \delta \cdot \mathbf{X}(u)du$ is normal with mean

$$\Big\{ \int_t^T \delta \cdot e^{A\cdot(u-t)}du \Big\} \mathbf{X}(t) \tag{10.13}$$

and variance

$$\int_t^T \mathbf{C}(T-s)BB^*\mathbf{C}^*(T-s)ds = \int_0^{T-t} \mathbf{C}(s)BB^*\mathbf{C}^*(s)ds. \tag{10.14}$$

It is trivial but helpful to keep in mind that the mean and variance here are real numbers, not vectors.

It follows from the formula $E[e^{\lambda Y}] = e^{\lambda\mu+\sigma^2\lambda^2/2}$ for the moment generating function of a normal random variable with mean μ and variance σ^2 that

$$\begin{aligned} B(t,T) &= \tilde{E}\left[\exp\{-\int_t^T R(u)\,du\} \;\Big|\; X_1(t),\ldots,X_m(t) \right] \\ &= \exp\{-C_1(T-t)X_1(t) - \cdots - C_m(T-t)X_m(t) - (T-t)\delta_0 + \frac{1}{2}\bar{A}(T-t)\}, \end{aligned}$$

where

$$\mathbf{C}(\tau) := (C_1(\tau),\ldots,C_m(\tau)) = \int_0^\tau (\delta_1,\cdots,\delta_m)e^{A\cdot u}\,du$$

and

$$\bar{A}(\tau) = \int_0^\tau (C_1(u),\ldots,C_m(u))BB^*(C_1(u),\ldots,C_m(u))^*\,du.$$

are defined above. δ_0 is given from the model of $R(t)$.

Application of this formula to the canonical two-factor Vasicek model is carried out in this week's Assignment.

10.6 Cox-Ingersoll-Ross model

As mentioned above, the short rate $R(t)$ under Vasicek model can become negative. We consider instead the two-factor CIR model:

$$d\mathbf{X}(t) = \{\mu(t) - A(t)\mathbf{X}(t)\}\,dt + B(X_t)\,d\tilde{W}(t).$$

where

$$A(t) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \mu(t) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

$$B(X_t) = \begin{pmatrix} \sqrt{X_1(t)} & 0 \\ 0 & \sqrt{X_2(t)} \end{pmatrix}, \quad \tilde{W}(t) = \begin{pmatrix} \tilde{W}_1(t) \\ \tilde{W}_2(t) \end{pmatrix},$$

where $\mu_1, \mu_2, a_{11}, a_{22} > 0$ and $a_{12}, a_{21} \leq 0$.

The dynamics is constructed so that when $X_1(t) = 0$ then $\mu_1 - a_{12}X_2(t) \geq 0$ pushing X_1 above 0. Similarly for X_2. Thus one sees that $X_1(t), X_2(t)$ stays non-negative for all t if $X_1(0), X_2(0)$ are non-negative.

Thus by choosing

$$R(t) = \delta_0 + \delta_1 X_1(t) + \delta_2 X_2(t),$$

where we set $\delta_0 \geq 0, \delta_i > 0$ for all $i = 1, 2$ then $R(t)$ is non-negative as well.

There is no explicit solution for the CIR factor models (because of the $\sqrt{X_i(t)}$ term in the volatility). But also exactly because of this structure, and the independence of $\tilde{W}_1, \tilde{W}_2$ we also have an affine yield structure for the CIR model. In other words, the bond price $B(t, T)$ has the form

$$\begin{aligned} B(t,T) &= f(t, X_1(t), X_2(t)) \\ f(t, x_1, x_2) &= e^{-x_1 C_1(T-t) - x_2 C_2(T-t) - A(T-t)}. \end{aligned}$$

One can then set up a system of ODE equations for C_1, C_2, A in the fashion discussed in the Section (10.3) and solve for the bond price that way. See also Shreve's Section 10.2.2 for more details.

10.6.1 One factor CIR model

We can demonstrate the idea of the CIR model more clearly if we simplify it to a one dimensional case. Namely, suppose the short rate follows the dynamics

$$dR_t = k(\mu - R_t)dt + \sigma\sqrt{R_t}d\widetilde{W}_t.$$

This is a mean reveversion equation but it does not have an explicit solution. By Markov property, there exists a function $u(t, R_t)$ such that

$$u(t, R_t) = B(t,T) = \tilde{E}(e^{-\int_t^T R_u du} | \mathcal{F}_t).$$

The PDE that $u(t, x)$ satisfies is

$$\begin{aligned} -xu + u_t + u_x k(\mu - x) + \frac{1}{2}u_{xx}\sigma^2 x &= 0 \\ u(T, x) = 1. \end{aligned}$$

On the other hand, to be affine-yield, $u(t,x)$ has the representation

$$u(t,x) = e^{-C(t,T)x - A(t,T)}.$$

Plug this form into the PDE, factoring the terms that involves x gives an ODE system in t

$$\begin{aligned} A_t + k\mu C &= 0 \\ C_t - kC - \frac{1}{2}\sigma^2 C^2 + 1 = 0. \end{aligned}$$

The terminal conditions are $A(T,T) = C(T,T) = 0$. This is as far as we can go to describe the form of the explicit solution for $B(t,T)$. It is not our interest here to find the explicit solution to the above system of ODEs. The interested readers can find the solution in Shreve section 6.6.

We remark that similar to the one-factor Hull-White case, here we will also find that indeed $C(t,T) = C(T-t)$ and $A(t,T) = A(T-t)$. That is the coefficients only depend on the time to maturity $T-t$. Thus for calibration, we can fix $\tau = T-t$ and find the dynamics of $R(t,\tau)$ in t as described in the section on the one-factor Hull-White.

10.7 Linear Systems of Stochas tic Differential Equations

10.7.1 The setting

This is a purely mathematical section. Linear systems of stochastic differential equations appear frequently in applied modeling and it is useful for the mathematical finance practitioner to know the basics about them.

By a linear system of stochastic equations we mean a system of the form

$$\begin{aligned} dX_1(t) &= \{a_{11}(t)X_1(t) + a_{12}X_2(t) + \cdots + a_{1m}(t)X_m(t) + c_1(t)\}\, dt + \sum_{k=1}^{d} \sigma_{1k}(t)\, dW_k(t) \\ dX_2(t) &= \{a_{21}(t)X_1(t) + a_{22}X_2(t) + \cdots + a_{2m}(t)X_m(t) + c_2(t)\}\, dt + \sum_{k=1}^{d} \sigma_{2k}(t)\, dW_k(t) \\ \cdots &= \cdots\cdots\cdots\cdots\cdots \\ dX_m(t) &= \{a_{m1}(t)X_1(t) + a_{m2}X_2(t) + \cdots + a_{mm}(t)X_m(t) + c_m(t)\}\, dt + \sum_{k=1}^{d} \sigma_{mk}(t)\, dW_k(t) \end{aligned}$$

It is much more efficient to write this in vector notation. Define

$$\mathbf{X} = \begin{pmatrix} X_1(t) \\ \vdots \\ X_m(t) \end{pmatrix}, \quad A(t) = \begin{pmatrix} a_{11}(t) & \cdots & a_{1m}(t) \\ \vdots & \vdots & \vdots \\ a_{m1}(t) & \cdots & a_{mm}(t). \end{pmatrix}, \quad c(t) = \begin{pmatrix} c_1(t) \\ \vdots \\ c_m(t) \end{pmatrix},$$

$$B(t) = \begin{pmatrix} \sigma_{11}(t) & \cdots & \sigma_{1d}(t) \\ \vdots & \vdots & \vdots \\ \sigma_{m1}(t) & \cdots & \sigma_{md}(t). \end{pmatrix}, \quad W(t) = \begin{pmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{pmatrix}.$$

Then the system of equations becomes

$$d\mathbf{X}(t) = \{A(t)\mathbf{X}(t) + c(t)\}\, dt + B(t)\, dW(t). \tag{10.15}$$

Here $W(t)$ is a multi-dimensional Brownian motion, and $A(t)$, $c(t)$, and $B(t)$ are given functions of t. They could even be stochastic processes adapted to a filtration $\{\mathcal{F}(t);\ t \geq 0\}$ for W. In this lecture we shall present results only for the case in which $A(t) = A$, $B(t) = B$, and $c(t) = c$ are constant, deterministic matrices or vectors. This is the simplest, most often encountered case, and the theory for (10.15) is a fairly straightforward generalization from this case.

In equation (10.15) the components of $\mathbf{X}(t)$ do not appear in the '$dW(t)$' term. It is common to use the term "bilinear" for equations in which linear functions of the components of $\mathbf{X}(t)$ multiply $dW_i(t)$ terms. The Black-Scholes equation is bilinear in this sense.

Consider

$$d\mathbf{X}(t) = \{A\mathbf{X}(t) + c\}\, dt + B\, dW(t), \tag{10.16}$$

where A is an $m \times m$ matrix, B is an $m \times d$ matrix, c is an m-vector, and W is a d-dimensional Brownian motion. An explicit solution to this equation can be written down using the theory of ordinary linear systems of differential equations. This requires a bit of review.

10.7.2 The fundamental matrix for a linear system

Let I denote the $m \times m$ identity matrix. For a given $m \times m$ matrix A, define

$$e^{A\cdot t} = I + \sum_{k=1}^{\infty} A^k \frac{t^k}{k!}. \tag{10.17}$$

This is a matrix-valued, infinite series, and it can be proved that it converges for any A and any t, $-\infty < t < \infty$, and so $e^{A\cdot t}$ is well-defined.

Let $\frac{d}{dt} e^{A\cdot t}$ denote the matrix obtained by differentiating each entry of $e^{A\cdot t}$. Then it can be shown also that

$$\begin{aligned} \frac{d}{dt} e^{A\cdot t} &= \sum_{k=1}^{\infty} A^k \frac{d}{dt} \frac{t^k}{k!} = \sum_{k=1}^{\infty} A^k \frac{t^{k-1}}{(k-1)!} = A \cdot \left[I + \sum_{k=1}^{\infty} A^k \frac{t^k}{k!} \right] \\ &= A e^{A\cdot t}. \end{aligned} \tag{10.18}$$

Also, clearly, $e^{A\cdot 0} = I$. As a result, if $\mathbf{Z}_0$ is any m-vector, $Z(t) = e^{A\cdot t}\mathbf{Z}_0$ solves

$$\frac{d}{dt}\mathbf{Z}(t) = A\mathbf{Z}(t), \quad \mathbf{Z}(0) = \mathbf{Z}_0. \tag{10.19}$$

This is easily verified; $e^{A\cdot 0}\mathbf{Z}_0 = I \cdot \mathbf{Z}_0 = \mathbf{Z}_0$ and $(d/dt)eA \cdot t\mathbf{Z}_0 = [Ae^{A\cdot t}]\mathbf{Z}_0 = A[e^{A\cdot t}\mathbf{Z}_0]$. For this reason, $e^{A\cdot t}$ is called the fundamental matrix for the equation $\frac{d}{dt}\mathbf{Z}(t) = A\mathbf{Z}(t)$.

From the fact that solutions to (10.19) are unique, one can also deduce a converse statement:

$$\text{if } \Phi(t) \text{ is a matrix valued solution to} \frac{d}{dt}\Phi(t) = A\Phi(t), \ \ \Phi(0) = I, \quad \text{then } \Phi(t) = e^{A\cdot t}. \tag{10.20}$$

The following basic fact can be proved using either the definition (10.16) or the fact that $e^{A\cdot t}$ solves equation (10.17): for any $-\infty < s, t < \infty$, $e^{A\cdot t}e^{A\cdot s} = e^{A\cdot(t+s)}$; in particular, $e^{-A\cdot t}e^{A\cdot t} = e^{A\cdot 0} = I$, and hence e^{-At} is the inverse of e^{At}. However, if $C \neq A$, it is not in general true that $e^{A\cdot t}e^{C\cdot t} = e^{(A+C)\cdot t}$.

Another very useful fact when it come to computing matrix exponentials is the following. Suppose P is an invertible matrix. Observe that

$$[PAP^{-1}]^k = [PAP^{-1}][PAP^{-1}]\cdots[PAP^{-1}] = PA^kP^{-1}.$$

Thus

$$e^{[PAP^{-1}]\cdot t} = I + \sum_{k=1}^{\infty} PA^kP^{-1}\frac{t^k}{k!} = Pe^{A\cdot t}P^{-1}.$$

Example 1. Let

$$A = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2 \end{pmatrix}.$$

Then if $\lambda_1 \neq \lambda_2$,

$$e^{A\cdot t} = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1 - \lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix}. \tag{10.21}$$

If $\lambda_1 = \lambda_2$,

$$e^{A\cdot t} = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \lambda_{21}te^{-\lambda_1 t} & e^{-\lambda_1 t} \end{pmatrix}. \tag{10.22}$$

Shreve gives a derivation of these formulas in Lemma 10.2.3 on page 417. It is not necessary to study this derivation in detail. From the characterization of $e^{A\cdot t}$ in (10.20), it suffice to show that the given formula in each case solves $\frac{d}{dt}\Phi(t) = A\Phi(t)$ with $\Phi(0) = I$. Consider the case of (10.21). Obviously the given matrix is the identity matrix when $t = 0$. A simple calculation shows that

$$\frac{d}{dt}\begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t}-e^{-\lambda_2 t}\right) & e^{-\lambda_2 t}\end{pmatrix} = \begin{pmatrix} -\lambda_1 e^{-\lambda_1 t} & 0 \\ \frac{-\lambda_{21}}{\lambda_1-\lambda_2}\left(\lambda_1 e^{-\lambda_1 t}-\lambda_2 e^{-\lambda_2 t}\right) & -\lambda_2 e^{-\lambda_2 t}\end{pmatrix}.$$

It is left to the student to show that

$$A\cdot\begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t}-e^{-\lambda_2 t}\right) & e^{-\lambda_2 t}\end{pmatrix} = \begin{pmatrix} -\lambda_1 e^{-\lambda_1 t} & 0 \\ \frac{-\lambda_{21}}{\lambda_1-\lambda_2}\left(\lambda_1 e^{-\lambda_1 t}-\lambda_2 e^{-\lambda_2 t}\right) & -\lambda_2 e^{-\lambda_2 t}\end{pmatrix}.$$

This completes the verification of (10.21) and (10.22) may be checked in the same way.

10.7.3 Multi-dimensional Ito's formula

Let

$$\mu(t) = \begin{pmatrix} \mu_1(t) \\ \vdots \\ \mu_m(t)\end{pmatrix}, \qquad \sigma(t) = \begin{pmatrix} \sigma_{11}(t) & \cdots & \sigma_{1d}(t) \\ \vdots & \vdots & \vdots \\ \sigma_{m1}(t) & \cdots & \sigma_{md}(t). \end{pmatrix},$$

$$W(t) = \begin{pmatrix} W_1(t) \\ \vdots \\ W_d(t)\end{pmatrix}.$$

Let the m-dimensional process X have the dynamics given by

$$dX(t) = \mu(t)dt + \sigma(t)dW_t.$$

Let f be a smooth function that maps $(\mathbb{R}^+, \mathbb{R}^m) \to \mathbb{R}$. Then the process $f(t, X(t))$ has a stochastic differential given by

$$df = \Big\{\frac{\partial f}{\partial t} + \sum_{i=1}^{m} \mu_i(t)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m}(\sigma\sigma^*)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}\Big\}dt + \sum_{i=1}^{m}\frac{\partial f}{\partial x_i}\sigma_i(t)\cdot dW_t,$$

where σ_i denotes the ith row f the matrix σ:

$$\sigma_i = [\sigma_{i1}, ..., \sigma_{id}],$$

and σ^* denotes the transpose of σ.

All the expression of df and the partials of f in the above are evaluated at (t, X_t), which we suppressed in the formula for simplicity of notation.

Alternatively, if we denote

$$\begin{aligned}\nabla f &:= (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_m}) \\ \sigma(t) &= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_m} & \cdots & \frac{\partial^2 f}{\partial x_m^2}(t). \end{pmatrix},\end{aligned}$$

to be the gradient and the Hessian matrix of f respectively, then the Ito's formula can be written succinctly as

$$df = \Big\{\frac{\partial f}{\partial t} + \mu(t)\cdot\nabla f + \frac{1}{2}tr\big(\sigma\sigma^*(t)H_f\big)\Big\}dt + \nabla f\cdot\sigma dW_t,$$

where $tr(A) := \sum_i A_{ii}$ denotes the trace of a a square matrix A.

10.7.4 The solution to equation (10.16)

The matrix exponential function of A may be used to express the solution to (10.16). This solution is

$$\mathbf{X}(t) = e^{A\cdot t}\mathbf{X}(0) + \int_0^t e^{A\cdot(t-s)}B\,dW(s) + \int_0^t e^{A\cdot(t-s)}c\,ds \qquad (10.23)$$

In the expression $\int_0^t e^{A\cdot(t-s)}B\,dW(t)$, the term $e^{A\cdot(t-s)}B$ is an $m\times d$ matrix multiplying a d-dimensional vector $dW(t)$ of Brownian differentials; hence $\displaystyle\int_0^t e^{A\cdot(t-s)}B\,dW(t)$ is an m-dimensional, vector-valued process.

A variant of (10.23) is true for representing $\mathbf{X}(T)$ for $T > t$ in terms of T,

$$\mathbf{X}(T) = e^{A\cdot(T-t)}\mathbf{X}(t) + \int_t^T e^{A\cdot(T-s)}B\,dW(s) + \int_t^T e^{A\cdot(t-s)}c(s)\,ds \qquad (10.24)$$

The increments $dW(s)$ for times $s > t$ are independent of $X(t)$; hence (10.24) exhibits $\mathbf{X}(T)$ is the sum of a linear transformation of $\mathbf{X}(T)$ plus a random vector *independent of* $\mathbf{X}(t)$.

Example 2. Let $\lambda_1 \neq \lambda_2$. Consider

$$d\begin{pmatrix} X_1(t) \\ X_2(t)\end{pmatrix} = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2\end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t)\end{pmatrix}dt + \begin{pmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{pmatrix}\begin{pmatrix} dW_1(t) \\ dW_2(t)\end{pmatrix} \qquad (10.25)$$

Using the result of Example 1,

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix}$$
$$+ \int_0^t \begin{pmatrix} e^{-\lambda_1(t-s)} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1(t-s)} - e^{-\lambda_2(t-s)}\right) & e^{-\lambda_2(t-s)} \end{pmatrix} \begin{pmatrix} \sigma_1 dW_1(s) \\ \sigma_2 dW_2(s) \end{pmatrix}.$$

The student should verify that:

$$\begin{aligned} X_1(t) &= e^{-\lambda t} X_1(0) + \int_0^t e^{-\lambda_1(t-s)} \sigma_1 \, dW_1(t) \\ X_2(t) &= \frac{\lambda_{21}}{\lambda_1 - \lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) X_1(0) + e^{-\lambda_2 t} X_2(0) \\ &+ \int_0^t \frac{\lambda_{21}}{\lambda_1 - \lambda_2}\left(e^{-\lambda_1(t-s)} - e^{-\lambda_2(t-s)}\right) \sigma_1 \, dW_1(s) \\ &+ \int_0^t e^{-\lambda_2(t-s)} \sigma_2 \, dW_2(s). \qquad \diamond \end{aligned}$$

To show the validity of (10.23), write $e^{A\cdot(t-s)} = e^{A\cdot t} e^{-A\cdot s}$ and factor $e^{A\cdot t}$ out of the integrals to write,

$$\mathbf{X}(t) = e^{A\cdot t}\left[\mathbf{X}(0) + \int_0^t e^{-A\cdot s} B \, dW(t) + \int_0^t e^{-A\cdot s} c \, ds\right].$$

Let $Y(t) = \mathbf{X}(0) + \int_0^t e^{-A\cdot s} B \, dW(t) + \int_0^t e^{-A\cdot s} c \, ds$ be the vector-valued Itô process in this expression, and note that $dY(t) = e^{-A\cdot t}\left[c \, dt + B \, dW(t)\right]$. Then

$$\begin{aligned} dX(t) &= \left[\frac{d}{dt} e^{A\cdot t}\right] Y(t) \, dt + e^{A\cdot t} dY(t) \\ &= A e^{A\cdot t} Y(t) \, dt + c \, dt + B \, dW(t) = A\mathbf{X}(t) \, dt + c \, dt + B \, dW(t). \end{aligned}$$

10.7.5 Joint distribution of the solution to (10.16)

In Theorem 4.4.9, Shreve states and proves the important fact that the Itô integral of a deterministic integrand is a normal random variable. This fact generalizes. If $W(t)$ is a d-dimensional Brownian motion and $B(t)$ is a deterministic $m \times d$-matrix-valued function, then $\int_0^t B(s) \, dW(s)$ is a normally distributed random vector. Hence, its joint density is determined by its mean vector and covariance matrix. The proof is essentially the same as that of Theorem 4.4.9. This fact has the following consequence for the solution $\mathbf{X}(t)$ of (10.16): *for any* $0 \leq t \leq T$*, the conditional distribution of* $X(T)$ *given* $X(t)$ *is Gaussian (normal).* In particular, if $\mathbf{X}(0)$ is deterministic or is a normal random variable independent of W, then $\{\mathbf{X}(t);\, t \geq 0\}$ is a vector-valued, Gaussian process.

Exercise 10.1 in Shreve is about the mean vector and covariance matrix of the process defined in Example 2.

10.7.6 How (10.16) changes under affine change of variable

Let $\mathbf{X}$ solve equation (10.16), let P be an invertible $m \times m$ matrix, let a be an m-vector, and define

$$\mathbf{Y}(t) = P\mathbf{X}(t) + a.$$

Then $\mathbf{Y}(t)$ also satisfies a system of linear stochastic differential equations. Note that $\mathbf{X}(t) = P^{-1}(\mathbf{Y}(t) - a)$. Thus,

$$\begin{aligned} d\mathbf{Y}(t) = P\,d\mathbf{X}(t) &= P\left[(A\mathbf{X}(t) + c)\,dt + B\,dW(t)\right] \\ &= \left[PAP^{-1}\mathbf{Y}(t) + P(c - AP^{-1}a)\right]\,dt + PB\,dW(t). \end{aligned}$$

Linear transformations like this are extremely useful for simplifying linear systems. For example, if A is invertible, and we choose $a = PA^{-1}C$, then $(c - AP^{-1}a)$ equals the zero vector, and hence $d\mathbf{Y}(t) = PAP^{-1}\mathbf{Y}(t)\,dt + PB\,dW(t)$. More importantly, one can choose P so that the matrix PAP^{-1} has a canonical form that is simple to work with, in terms of calculating $e^{PAP^{-1}\cdot t} = Pe^{A\cdot t}P^{-1}$ and in terms of understanding how the different components of $Y_i(t)$ influence one another. For example, if A has a basis of eigenvectors with real eigenvalues $\lambda_1, \ldots, \lambda_m$, P can be chosen so that PAP^{-1} is the diagonal matrix

$$PAP^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ldots & \ldots & \ldots & \lambda_m \end{pmatrix}.$$

It is easily seen that

$$e^{PAP^{-1}\cdot t} = \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ldots & \ldots & \ldots & e^{\lambda_m t} \end{pmatrix}.$$

CHAPTER 11 The Heath-Jarrow-Morton model for forward rates

11.1 Set up of the HJM model

We start with a probability space, $(\Omega, \mathcal{F}, \mathbf{P})$, which supports a Brownian motion $W(t) = (W_1(t), \ldots, W_d(t))$. Let $\{\mathcal{F}(t);\, t \geq 0\}$ be the filtration generated by W.

The HJM model for the forward rate is

$$df(t,T) = \alpha(t,T)\,dt + \sum_{j=1}^{d} \sigma_j(t,T)\,dW_j(t), \qquad 0 \leq t \leq T \leq \bar{T}, \tag{11.1}$$

where for each T, $\{\alpha(t,T);\, 0 \leq t \leq T\}$ and $\{\gamma_1(t,T);;\, 0 \leq t \leq T\}, \ldots, \{\gamma_d(t,T);\, 0 \leq t \leq T\}$ are $\{\mathcal{F}(t);\, t \geq 0\}$-adapted process for each T, $0 < T \leq \bar{T}$.

The interpretation is that we fix a time horizon $\bar{T}$ and we investigate all bonds with expiry T, $T \leq \bar{T}$.

At time $t = 0$, we know the forward rate curve, $T \to f(0,T)$ from market price quotes. This provides an initial condition for equation (11.1) for each T, and by integrating (11.1) forward in time,

$$f(t,T) = f(0,T) + \int_0^t \alpha(u,T)\,du + \int_0^t \sum_{j=1}^{d} \sigma_j(u,T)\,dW_j(u). \tag{11.2}$$

In particular, the short rate is

$$R(t) = f(t,t) = f(0,t) + \int_0^t \alpha(u,t)\,du + \int_0^t \sum_{j=1}^{d} \sigma_j(u,t)\,dW_j(u).$$

From now on, for simplicity, we study the case in which $d = 1$:

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW(t), \qquad 0 \leq t \leq T \leq \bar{T}. \tag{11.3}$$

11.2 Equations for zero-coupon bond prices

What stochastic differential equation for $B(t,T)$ is implied by the HJM model? To answer this, set

$$Y(t,T) = -\int_t^T f(t,v)\,dv.$$

Since

$$B(t,T) = e^{Y(t,T)},$$

Itô's rule implies

$$dB(t,T) = B(t,T)\,dY(t,T) + \frac{1}{2}B(t,T)[dY(t,T)]^2,$$

so to derive an equation for $dB(t,T)$ we need only to compute $dY(t,T)$.

Remark: Note that by $dY(t,T)$ and $dB(t,T)$ here we mean the differentials of Y and B in t, not T. That is for δ small, we mean

$$\begin{aligned} dY(t,T) &\approx Y(t+\delta,T) - Y(t,T); \\ dB(t,T) &\approx B(t+\delta,T) - B(t,T). \end{aligned}$$

Now to calculate $dY(t,T)$, since this is the differential in t, note that t appears in 2 places in the formula

$$Y(t,T) = -\int_t^T f(t,v)dv.$$

If $f(t,v)$ is differentiable in t, then the computation falls under the Newton-Leibniz formula. Here we have a differential form of f, but a similar version of that formula still holds, and thus

$$dY(t,T) = f(t,t)dt - \int_t^T [df(t,v)]dv.$$

Define

$$\alpha^*(t,T) = \int_t^T \alpha(t,v)\,dv \quad \text{and} \quad \sigma^*(t,T) = \int_t^T \sigma(t,v)\,dv.$$

Substituting $df(t,v)$ by (11.3) and by switching the order of integration if follows that

$$\begin{aligned} dY(t,T) &= f(t,t)dt - \int_t^T \Big[\alpha(t,v)\,dt + \sigma(t,v)\,dW(t)\Big]dv \\ &= f(t,t)dt - \Big[\int_t^T \alpha(t,v)\,dv\Big]dt - \Big[\int_t^T \sigma(t,v)\,dv\Big]dW(t) \\ &= f(t,t)dt - \alpha^*(t,T)dt - \sigma^*(t,T)dW(t). \end{aligned}$$

Thus

$$\begin{aligned} Y(t,T) &= Y(0,T) + \int_0^t f(u,u)\,du - \int_0^t \alpha^*(u,T)\,du - \int_0^t \sigma^*(u,T)\,dW(u) \\ &= -\int_0^T f(0,u)\,du + \int_0^t f(u,u)\,du - \int_0^t \alpha^*(u,T)\,du \\ &- \int_0^t \sigma^*(u,T)\,dW(u). \end{aligned} \tag{11.4}$$

Since $f(t,t) = R(t)$, it follows that

$$dY(t,T) = \Big[R(t) - \alpha^*(t,T)\Big]\,dt - \sigma^*(t,T)\,dW(t), \tag{11.5}$$

and hence that

$$dB(t,T) = B(t,T)\Big[R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2\Big]\,dt - B(t,T)\sigma^*(t,T)\,dW(t). \tag{11.6}$$

It is interesting to note that $\sigma^*(T,T) = 0$. This makes sense: the volatility of $B(T,T)$ is zero because $B(T,T) = 1$.

Remark: Recall the dynamics for $f(t,T)$ is

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW(t). \tag{11.7}$$

It is important to observe that equations (11.5), (11.6), (11.7) are equivalent. That is if we have the dynamics of $B(t,T)$ following (11.6) then the dynamics of $Y(t,T)$ and $f(t,T)$ have to follow (11.5) and (11.7) respectively. The other cases are similar. The reason is because $B(t,T)$, $f(t,T)$ and $Y(t,T)$ are connected via the relations

$$\begin{aligned} B(t,T) &= e^{Y(t,T)} \\ \frac{\partial Y}{\partial T} &= -f(t,T). \end{aligned}$$

The implication of this will be seen in Section (11.4), where we conclude that the drift of $f(t,T)$ have to be $\sigma(t,T)\sigma^*(t,T)$ under the risk neutral measure, even when we derive f from a short rate model, for example the Hull-White model.

11.3 Existence of a risk-neutral model

In order that there be no arbitrage in the HJM model it is necessary that there exists an equivalent probability measure $\widetilde{\mathbf{P}}$ with respect to which $\{D(t)B(t,T);\, t \le T\}$ is a martingale for all $0 \le T \le \bar{T}$. Note that this is a subtle condition since for each T, we have a different financial product $B(t,T)$. So this statement requires that the discounted price of *infinitely many* financial products must be $\widetilde{\mathbf{P}}$ martingale. But in our model, we only have finitely many sources of noise. This has implication about the existence of solution to the market price of risk equation. (Recall from what we have learned that if the number of assets m is less than the number of random sources d then the risk neutral measure many not exist).

From the previous section, we have the dynamics of the bond $B(t,T)$ under the HJM model under the physical measure $\mathbf{P}$:

$$dB(t,T) = B(t,T)\left[R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2\right]dt - B(t,T)\sigma^*(t,T)\,dW(t).$$

This is a geometric Brownian motion model with drift term

$$\gamma(t,T) := R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2$$

and volatility $-\sigma^*(t,T)$.

Thus, the market price of risk equation for this *particular bond with expiry* T is: to find an adapted process $\theta(t)$ such that

$$\begin{aligned} -\sigma^*(t,T)\theta(t) &= \gamma(t,T) - R(t) \\ &= \left[R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2\right] - R(t) \\ &= -\alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2. \end{aligned}$$

Or equivalently

$$\sigma^*(t,T)\theta(t) = \alpha^*(t,T) - \frac{1}{2}(\sigma^*(t,T))^2.$$

Note that $\theta(t)$ does not depend on T (the risk neutral measure, which is defined from θ should not depend on an expiry). On the other hand, for each T, we have such an equation and thus this gives an infinite system of equations for $\theta(t)$.

By differentiating both sides of the above equation with respect to T, this reduces to the equivalent and simpler sufficient condition:

$$\sigma(t,T)\theta(t) = \alpha(t,T) - \sigma^*(t,T)\sigma(t,T), \quad \text{for all } 0 \le t \le T \le \bar{T}. \tag{11.8}$$

Indeed, if $\theta(t)$ satisfies (11.8) then by taking the anti-derivative (in T) we get

$$\sigma^*(t,T)\theta(t) = \alpha^*(t,T) - \frac{1}{2}(\sigma^*(t,T))^2 + C(T),$$

where $C(T)$ is a constant depending on T. Plug in $t = T$ we conclude $C(T) = 0$ and thus $\theta(t)$ solves the market price of risk equation.

Therefore, if equation (11.8) has a solution and if

$$E[\exp\{-\int_0^{\bar{T}} \theta(u)\, dW(u) - \int_0^{\bar{T}} \theta^2(u)\, du\}] = 1,$$

(this is a technical condition to make sure the change of measure is well-defined) then there is an equivalent risk-neutral measure.

Again, (11.8) places heavy restrictions on the relation of $\alpha(t,T)$ to $\sigma(t,T)$. If $\sigma(t,T)$ is non-zero always then $\theta(t) = \sigma^{-1}(t,T)\alpha(t,T) - \sigma^*(t,T)$ and the right-hand side *must not actually depend on* T. Thus $\alpha(t,T)$ and $\sigma(t,T)$ must be related in a very special way.

You are asked to derive a version of (11.8) when W is multidimensional in Shreve, Exercise 10.9.

11.4 The risk-neutral form of HJM

Assume that the market price of risk equations of (11.8) have a solution θ. Then under the risk-neutral measure, $\widetilde{W}(t) = W(t) + \int_0^t \theta(u)\, du$ is a Brownian motion, and

$$\begin{aligned} df(t,T) &= \alpha(t,T)\, dt + \sigma(t,T)\, [d\widetilde{W}(t) - \theta(t)\, dt] \\ &= [\alpha(t,T) - \sigma(t,T)\theta(t)]\, dt + \sigma(t,T)\, d\widetilde{W}(t) \\ &= \sigma^*(t,T)\sigma(t,T)\, dt + \sigma(t,T)\, d\widetilde{W}(t) \end{aligned} \tag{11.9}$$

Here is an important observation for the risk neutral form of the forward rate: *there is no freedom in choosing the drift of the forward rate under the risk-neutral model*. This, in some sense is similar to the situation we are used to with modeling the asset in Black-Scholes framework, where the drift term of the asset must be $R(t)S(t)dt$ under the risk neutral measure.

For the forward rate, the drift must be $\sigma^*(t,T)\sigma(t,T)$, which is completely determined by the volatility $\sigma(t,T)$. This fact limits the nature of the models that can be proposed for the forward rate.

For example, suppose we wanted to have a lognormal model for $f(t,T)$ uder the risk-neutral measure (So that we can have explicit computation for the distribution of $f(t,T)$

and possibly $B(t,T)$). A reasonable way to approach this would be to let $\sigma(t,T) = \sigma f(t,T)$, where σ is a constant. (This is possibly the simplest way to propose the volatility of a log-normal model).

However, if we do this the drift term will become

$$\sigma^*(t,T)\sigma(t,T) = \sigma^2(t)f(t,T)\int_t^T f(t,v)\,dv$$

which is *not a linear function* of $f(t,T)$. (So even if we want, a log-normal model under the risk neutral measure is not possible for the forward rate).

Even worse, with this drift (11.9) will not even have a solution defined for all sample paths (near the expiry T the forward rate may get very large - due to the drift being almost like quadratic in $f(t,T)$. See Shreve's Section 10.4.1 for the more details. This is also a motivating example for us to switch our attention to the forward LIBOR model in the 2nd part of this lecture.

On the other hand, we can use (11.9) to define *a risk-neutral model directly*. We suppose the Brownian motion $\widetilde{W}$ on $(\Omega, \mathcal{F}, \widetilde{\mathbf{P}})$ to be given, we assume we have a $\sigma(t,T)$ such that (11.9) *has a solution* (in the case when $\sigma(t,T)$ depends on $f(t,T)$), and this gives us a risk-neutral model with

$$R(t) = f(t,t) = f(0,t) + \int_0^t \sigma^*(u,t)\sigma(u,t)\,du + \int_0^t \sigma(u,t)\,d\widetilde{W}(u),$$

and with

$$B(t,T) = \exp\{-\int_t^T f(t,v)\,dv\}$$

satisfying

$$dB(t,T) = R(t)B(t,T)\,dt - \sigma^*(t,T)B(t,T)\,d\widetilde{W}(t) \tag{11.10}$$

(We know this has to be the right equation for $B(t,T)$ from equation (11.6) and from the fact that $D(t)B(t,T)$ is a martingale under $\widetilde{\mathbf{P}}$.)

11.5 Example of HJM models

The fact that the market price of risk equations

$$\sigma(t,T)\theta(t) = \alpha(t,T) - \sigma^*(t,T)\sigma(t,T), \quad \text{for all } 0 \le t \le T \le \bar{T}$$

is an infinite systems for $\theta(t)$ might give us the impression that it is not easy to have an example of HJM model, except perhaps for trivial cases when α and σ do not depend on

T. But this is not the case. All previous short rate models we discussed (Vasicek, CIR and also the one factor model Hull-White) are examples of HJM model as we will see. The key is again, if we *start directly under the risk neutral measure* (which is the case for all short rate models) then we do not have the market price of risk equation to deal with.

More generally, we can always obtain a risk-neutral HJM model as described in the previous section, if we choose an exogenously determined volatility $\sigma(t,T)$ (that is, a volatility that *is not a function of* $f(t,T)$ but is defined *a priori*.

Indeed this will be so for any affine yield model based on a model for the short rate, under the risk-neutral measure.

$$B(t,T) = \exp\{-\sum_{j=1}^{m} C_j(t,T)X_j(t) - A(t,T)\},$$

where $X_1(t),\ldots,X_m(t)$ solves a system of stochastic differential equations driven by $\widetilde{W}$ (possibly multi-dimensional), and $C_1(t,T),\ldots,C_m(t,T)$, and $A(t,T)$ are differentiable in T. Then

$$f(t,T) = \sum_{j=1}^{m} X_j(t)\frac{\partial C_j(t,T)}{\partial T} + \frac{\partial A(t,T)}{\partial T}$$

and

$$df(t,T) = \sum_{j=1}^{m} d[\frac{\partial C_j(t,T)}{\partial T} X_j(t)] + d[\frac{\partial A(t,T)}{\partial T}]$$

This enables one to recover an HJM type model for the forward rate.

11.5.1 Example with one-factor model

For simpliciy, suppose $m = 1$ (that is we have a one-factor model), from which $R(t)$ has dynamics under $\widetilde{\mathbf{P}}$

$$dR(t) = \beta(t)dt + \gamma(t)d\widetilde{W}(t),$$

where we do allow β, γ to depend on $R(t)$, we just do not explicitly write it out that way.

The form of the bond is

$$B(t,T) = \exp\{-C(t,T)R(t) - A(t,T)\},$$

where, $C(t,T)$ and $A(t,T)$ are functions of β, γ (that is obtained from the risk neutral pricing formula).

The forward rate $f(t,T)$ is derived from $B(t,T)$ via the relation

$$f(t,T) = -\frac{\partial}{\partial T}\log B(t,T) = R(t)\frac{\partial}{\partial T}C(t,T) + \frac{\partial}{\partial T}A(t,T).$$

Therefore, $f(t,T)$ has the dynamics

$$
\begin{aligned}
df(t,T) &= d[\frac{\partial C(t,T)}{\partial T}R(t)] + d[\frac{\partial A(t,T)}{\partial T}] \\
&= \Big[\frac{\partial C(t,T)}{\partial T}\beta(t) + R(t)\frac{\partial C'(t,T)}{\partial T} + \frac{\partial A'(t,T)}{\partial T}\Big]dt + \frac{\partial C(t,T)}{\partial T}\gamma(t)d\widetilde{W}(t),
\end{aligned}
$$

where $C'(t,T), A'(t,T)$ denotes the derivative of these functions with respect to t.

In order that this be a valid HJM model, the drift term must be related to the volatility term as in equation (11.9). That is denoting

$$\sigma(t,T) = \frac{\partial C(t,T)}{\partial T}\gamma(t),$$

then it must follow that

$$\frac{\partial C(t,T)}{\partial T}\beta(t) + R(t)\frac{\partial C'(t,T)}{\partial T} + \frac{\partial A'(t,T)}{\partial T} = \sigma(t,T)\sigma^*(t,T).$$

But you do not really need to check it to know that it must be true, because the Hull-White model was created under the risk-neutral measure so that $D(t)B(t,T)$ was automatically a martingale. And we know from the analysis of the previous section that if $D(t)B(t,T)$ is a martingale under the risk-neutral measure and if $df(t,T)$ has a stochastic differential, it must be of the form in (11.9).

Indeed, as discussed in the remark at the end of Section (11.2), once we know the dynamics of the bond $B(t,T)$, then the dynamics of $f(t,T)$ is forced. Recall the dynamics of $B(t,T)$ and $f(t,T)$ as discussed in Section (11.2) under the physical measure P:

$$
\begin{aligned}
dB(t,T) &= B(t,T)\Big[R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2\Big]dt - B(t,T)\sigma^*(t,T)\,dW(t); \\
df(t,T) &= \alpha(t,T)dt + \sigma(t,T)dW(t).
\end{aligned}
$$

Since the derivation is via Ito's formula, this relation also holds under the risk neutral measure $\widetilde{\mathbf{P}}$, we just need to change $W(t)$ to $\widetilde{W}(t)$ and modify the drift accordingly. Under risk neutral, the drift of $B(t,T)$ is $B(t,T)R(t)$. Thus one must have

$$\alpha^*(t,T) = \frac{1}{2}(\sigma^*(t,T))^2.$$

But this implies

$$\alpha(t,T) = \frac{\partial \alpha^*(t,T)}{\partial T} = \sigma(t,T)\sigma^*(t,T),$$

which is what the drift of $f(t,T)$ must be under risk neutral. So as long as we derive $B(t,T)$ and $f(t,T)$ from a consistent risk neutral model (which is any factor model for the short rate) then the dynamics of $f(t,T)$ and $B(t,T)$ would be consistent as discussed in the above.

Apply this to the Hull-White example above, we see that since

$$B(t,T) = \exp\{-C(t,T)R(t) - A(t,T)\},$$

the dynamics of $B(t,T)$ under $\widetilde{\mathbf{P}}$ is

$$dB(t,T) = \textit{ something } dt - B(t,T)C(t,T)\gamma(t)d\widetilde{W}(t).$$

(The fact that the *something* becomes $B(t,T)R(t)$ comes from the relation between $C(t,T), A(t,T), \beta(t), \gamma(t)$. The fact that the volatility has this form comes from the fact that only $R(t)$ has Brownian motion component in its dynamics and the Ito's formula applied to $B(t,T)$). This implies that

$$\sigma^*(t,T) = C(t,T)\gamma(t).$$

Note that this is consistent with the fact that we assign

$$\sigma(t,T) = \frac{\partial C(t,T)}{\partial T}\gamma(t)$$

for the dynamics of $f(t,T)$. Then the conclusion that

$$\frac{\partial C(t,T)}{\partial T}\beta(t) + R(t)\frac{\partial C'(t,T)}{\partial T} + \frac{\partial A'(t,T)}{\partial T} = \sigma(t,T)\sigma^*(t,T).$$

can be seen as a consequence of the dynamics of $f(t,T)$ being forced once we known the dynamics of $B(t,T)$ as discussed above.

Because the Hull-White, Vasicek, etc. models are examples of HJM models, all theory we develop later assuming the HJM model is valid for these earlier models.

CHAPTER 12 Forward LIBOR model

12.1 Forward LIBOR

12.1.1 Continuous vs simple compounding

Let $0 < t < S < T$. Suppose we want to price a product based on the forward rate, say a cap K on the forward rate for a loan taken on the interval from S to T, with the rate locked in at time t. The risk neutral pricing formula for this product would be

$$\tilde{E}\left(e^{-\int_S^T (f(t,u)-K)^+ du}\right).$$

(The meaning of such cap would be clearer when we discuss the Caplet below).

It is certainly desirable to be able to obtain a closed form solution to such formula, under the assumption that the volatility of $f(t,T)$ (as a process in t) is of the form $\sigma f(t,T)$ where σ is a constant. But we saw that from a purely mathematical point of view, this is not possible since it would force *the drift term* of $f(t,T)$ under the risk neutral measure $\tilde{P}$ to have certain form, which in turn makes the solution to $f(t,T)$ explode near T (Shreve's Section 10.4.1).

Furthermore, the problem of the drift term of $f(t,T)$ cannot be solved by a change of measure associated with a change of numéraire. You should try to derive the dynamics of $f(t,T)$ under the $\tilde{P}^{T'}$-forward measure, for some $T < T'$ for example, and convince yourself that the drift term of $f(t,T)$ cannot be eliminated.

The presence of the drift term, or equivalently the dynamics of $f(t,T)$, can be seen as coming from the *continuous compouding* used in its definition:

$$B(t,T) = e^{-\int_t^T f(t,u)du}.$$

It turns out that in order to "eliminate the drift term" in the "interest rate", so that we will be able to posit a log-normal distribution for it, we want to use *simple compounding* instead. That is, we denote $L_\delta(t,T)$ as the quantity that satisfies

$$B(t,T+\delta)(1+\delta L_\delta(t,T)) = B(t,T).$$

You should see that $L_\delta(t,T)$ is the interest rate one can lock in at time t for investing on the time interval $[T, T+\delta]$ with *simple compounding: repayment = investment* $\times$ *(1 + duration of investment* $\times$ *interest rate).*

$L_\delta(t,T)$ is called the simple forward LIBOR rate of tenor δ.

The reason why $L_\delta(t,T)$ should have a "better" dynamics than $f(t,T)$ (in terms of being able to posit a log normal distribution) is because of its definition:

$$1 + \delta L_\delta(t,T) = \frac{B(t,T)}{B(t,T+\delta)}.$$

Thus clearly the dynamics of $L_\delta(t,T)$ is related to the dynamics of $B(t,T)$ under the $\tilde{P}^{T+\delta}$ forward measure, using $B(t,T+\delta)$ as numéraire. Since $D(t)B(t,T)$ is a martingale under $\tilde{P}$ we expect the dynamics of $L_\delta(t,T)$ under $\tilde{P}^{T+\delta}$ is "nice" as well.

Thus in this note we will develop the dynamics of $L_\delta(t,T)$ under the $T+\delta$ forward measure and show how to price financial products based on it (cap and caplet), under *the assumption of determinsitic volatility*, using Black-Scholes type of calculation.

12.1.2 How to construct a portfolio that realize the simple interest rate $L_\delta(t,T)$

Suppose at time $t < T$, we go short one share of $B(t,T)$ and long $B(t,T)/B(t,T+\delta)$ shares of $B(t,T+\delta)$. The value of this portfolio is zero at time t; at time T it requires us to pay out one dollar and at time $T+\delta$ we receive $B(t,T)/B(t,T+\delta)$ dollars. Thus at time t we can lock in a deposit that multiplies to $B(t,T)/B(t,t+\delta)$ over $[T,T+\delta]$ and hence earns the simple interest rate $L_\delta(t,T)$ satisfying

$$1 + \delta L_\delta(t,T) = \frac{B(t,T)}{B(t,T+\delta)}$$

Thus

$$L_\delta(t,T) = \frac{1}{\delta}\left[\frac{B(t,T)}{B(t,T+\delta)} - 1\right] = \frac{1}{\delta}\frac{B(t,T) - B(t,T+\delta)}{B(t,T+\delta)}.$$

We have immediately that

$$1 + \delta L_\delta(T,T) = \frac{1}{B(T,T+\delta)}.$$

Thus $L_\delta(T,T)$ is the simple interest rate available at time T for a deposit over time period $[T,T+\delta]$. This is a financially important quantity, because it is often used for floating rate loans or as a benchmark for interest rate caps and floors.

12.1.3 Dynamics of $L_\delta(t,T)$

Here is an elementary, but very important observation:

$$\begin{aligned} L_\delta(t,T) &= \frac{1}{\delta}\frac{B(t,T) - B(t,T+\delta)}{B(t,T+\delta)} \\ &= \frac{\frac{1}{\delta}B(t,T) - \frac{1}{\delta}B(t,T+\delta)}{B(t,T+\delta)}. \end{aligned}$$

Thus $L_\delta(t,T)$, for $t \le T$ is the $T+\delta$ forward price of a portfolio that is long $1/\delta$ zero coupon bonds that mature at T and short $1/\delta$ zero coupon bonds that mature at $T+\delta$.

In this section, we will derive the model implied for the forward LIBOR rate by the risk-neutral HJM model. To start out, observe that since

$$\begin{aligned} L_\delta(t,T) &= \frac{1}{\delta}\frac{B(t,T)-B(t,T+\delta)}{B(t,T+\delta)} \\ &= \frac{1}{\delta}\frac{B(t,T)}{B(t,T+\delta)} - \frac{1}{\delta}, \end{aligned}$$

we have

$$dL_\delta(t,T) = \delta^{-1} d[B(t,T)/B(t,T+\delta)].$$

Following the notation of the change of numéraire section, we define

$$B^{T+\delta}(t,T) := B(t,T)/B(t,T+\delta)$$

as the $T+\delta$ forward price of $B(t,T)$.

Observe then, that it is most natural to express the model for $L_\delta(t,T)$ under the $T+\delta$ forward measure $\widetilde{\mathbf{P}}^{T+\delta}$. We know from Theorems 9.2.1 and 9.2.2 in Shreve that because

$$\begin{aligned} dD(t)B(t,T) &= -D(t)B(t,T)\sigma^*(t,T)\, d\widetilde{W}(t) \\ dD(t)B(t,T+\delta) &= -D(t)B(t,T+\delta)\sigma^*(t,T+\delta)\, d\widetilde{W}(t), \end{aligned}$$

we have

$$\begin{aligned} dL_\delta(t,T) &= \frac{1}{\delta} B^{T+\delta}(t,T)[\sigma^*(t,T+\delta) - \sigma^*(t,T)]\, d\widetilde{W}^{T+\delta}(t) \\ &= \frac{1}{\delta}\,[1+\delta L_\delta(t,T)]\,[\sigma^*(t,T+\delta) - \sigma^*(t,T)]\, d\widetilde{W}^{T+\delta}(t) \\ &= L_\delta(t,T)\left\{\frac{1+\delta L_\delta(t,T)}{\delta L_\delta(t,T)}[\sigma^*(t,T+\delta) - \sigma^*(t,T)]\right\} d\widetilde{W}^{T+\delta}(t), \quad (12.1) \end{aligned}$$

where $\widetilde{W}^{T+\delta}(t) = \widetilde{W}(t) + \int_0^t \sigma^*(u,T+\delta)\,du$ is a Brownian motion under $\widetilde{\mathbf{P}}^{T+\delta}$. From this equation we can easily derive the model for the forward LIBOR rate under the original risk-neutral measure $\widetilde{\mathbf{P}}$, but we will not have need for this.

Remark:

(i) If we denote

$$\gamma(t) := \frac{1+\delta L_\delta(t,T)}{\delta L_\delta(t,T)}[\sigma^*(t,T+\delta) - \sigma^*(t,T)],$$

then it follows that

$$dL_\delta(t,T) = L_\delta(t,T)\gamma(t)d\widetilde{W}^{T+\delta}(t).$$

If we **assume** $\gamma(t)$ is a constant then it is easy to see that $L_\delta(t,T)$ has log-normal distribution under $\widetilde{\mathbf{P}}^{T+\delta}$, which is a goal we have set out to achieve. This will help us to derive pricing equation in Black-Scholes style for financial products based on $L_\delta(t,T)$ as discussed in the Sections below.

(ii) Assuming $\gamma(t)$ is a constant is a **big** assumption if we start from the risk neutral model of $B(t,T)$ and $B(t,T+\delta)$. However, we can start modeling under the $T+\delta$-forward measure, where we are *free to assume* the fact that $\gamma(t)$ is a constant. The distribution of $B(t,T)$ and $B(t,T+\delta)$ under the risk neutral measure can then be derived from the $\tilde{P}^{T+\delta}$ model.

12.2 T-forward models

Previously, we defined a T-forward measure. This is a measure, $\widetilde{\mathbf{P}}^T$, if it exists, under which T-forward prices of all market assets are martingales. Recall that the T-forward price of an asset whose price in dollars is $S(t)$ is $S(t)/B(t,T)$. Now assume we have an HJM model driven by a single Brownian motion, and write it under the risk-neutral measure $\widetilde{\mathbf{P}}$. According to the theory developed in Chapter 9 of Shreve, the T-forward measure is defined by a change of measure from $\widetilde{\mathbf{P}}$ by the Radon-Nikodym derivative,

$$\frac{d\widetilde{\mathbf{P}}^T}{d\widetilde{\mathbf{P}}} = \frac{D(T)}{B(0,T)}. \tag{12.2}$$

That is, $\widetilde{\mathbf{P}}^T(A) = E[\mathbf{1}_A D(T)]/B(0,T)$, for $A \in \mathcal{F}$. But we know the solution to

$$dB(t,T) = R(t)B(t,T)dt - \sigma^*(t,T)B(t,T)dWt$$

is

$$D(t)B(t,T) = B(0,T)\exp\{-\int_0^t \sigma^*(u,T)\,dW(u) - \frac{1}{2}\int_0^t (\sigma^*)^2(u,T)\,du\}$$

and hence

$$\frac{d\widetilde{\mathbf{P}}^T}{d\widetilde{\mathbf{P}}} = \exp\{-\int_0^T \sigma^*(u,T)\,dW(u) - \frac{1}{2}\int_0^T (\sigma^*)^2(u,T)\,du\}. \tag{12.3}$$

It follows from Girsanov's theorem that

$$\widetilde{W}^T(t) = \widetilde{W}(t) + \int_0^t \sigma^*(u,T)\,du \tag{12.4}$$

is a Brownian motion under $\widetilde{\mathbf{P}}^T$, at least for times $t \leq T$.

All this is review of section 9.4 in Shreve.

12.3 Financial products based on forward LIBOR

12.3.1 Description

The forward LIBOR $L_\delta(t,T)$ is strictly not a financial asset by itself. However, if we think about investing a principal P at time T for the duration $[T, T+\delta]$ to realize the interest payment $P\delta L_\delta(T,T)$ at time $T+\delta$, then we have a product that is very much like a Euro style derivative, with expiry $T+\delta$.

One can also create another product that is in the spirit of the Euro Call option, in this case called an *interest rate cap*. For a constant K positive, we can consider a financial product that pays

$$V_{T+\delta} = \delta P\big(L_\delta(T,T) - K\big)^+$$

at time $T+\delta$. The interpretation is that if we borrow an amount P at time T, we may not want the interest rate $L_\delta(T,T)$ to go beyond K. Therefore to protect ourselves, we would want to get an interest rate cap that would pay us the difference should the interest rate go beyond K.

Moreover, since P and δ are deterministic (we think of them as determined at time 0), for simplicity we can take $P\delta = 1$. Thus, one can discuss the following products:

(i) A contract that pays $L_\delta(T,T)$ at time $T+\delta$. This is called a backset LIBOR on a notional amount of 1.

(ii) A contract that pays $(L_\delta(T,T) - K)^+$ at time $T+\delta$. This is called an *interest rate caplet.*

Clearly the question is what are the risk neutral prices of these products at time 0. We will give the formula for backset LIBOR in this section and give a detailed discussion of interest rate cap and caplet in the next section.

12.3.2 Risk neutral price of backset LIBOR

Theorem 12.3.1. *The no arbitrage price at time t of a contract that pays $L_\delta(T,T)$ at time $T+\delta$ is*

$$\begin{aligned} S(t) &= B(t, T+\delta)L_\delta(t,T),\ 0 \le t \le T \\ &= B(t, T+\delta)L_\delta(T,T),\ T \le t \le T+\delta. \end{aligned}$$

($S(t)$ is the notation Shreve used in the textbook. Don't confuse it with the stock price).

Proof:

By the risk neutral pricing formula

$$S(t) = \widetilde{E}\Big[e^{-\int_t^{T+\delta} R(u)du} L_\delta(T,T)\Big|\mathcal{F}(t)\Big].$$

If $T \leq t$ then $L_\delta(T,T)$ is $\mathcal{F}(t)$ measurable. Therefore

$$S(t) = L_\delta(T,T)\widetilde{E}\Big[e^{-\int_t^{T+\delta} R(u)du}\Big|\mathcal{F}(t)\Big] = B(t, T+\delta)L_\delta(T,T).$$

If $t < T$ then by the change of numéraire pricing formula under $\widetilde{\mathbf{P}}^{T+\delta}$ we have

$$\frac{S(t)}{B(t, T+\delta)} = \widetilde{E}^{T+\delta}\Big[L_\delta(T,T)\Big|\mathcal{F}(t)\Big].$$

But $L_\delta(t,T)$ is a martingale under $\widetilde{\mathbf{P}}^{T+\delta}$ (see equation 12.1 in Section 1). Therefore,

$$\frac{S(t)}{B(t, T+\delta)} = L_\delta(t,T)$$

and the conclusion follows.

12.4 Caps and caplets

12.4.1 Description

We will consider the following type of floating rate bond. It starts at $T_0 = 0$ and pays coupons $C_1, \ldots, C_{n+1}$ on principal P at dates $T_1 = \delta$, $T_2 = 2\delta, \ldots, T_j = j\delta, \ldots, T_{n+1} = (n+1)\delta$. The interest charged over $[T_{j-1}, T_j]$ is the LIBOR rate set at T_{j-1}. So coupon $C_j = \delta P L_\delta(T_{j-1}, T_{j-1})$.

Suppose now that Alice has issued such a bond. An equivalent interpretation is *she has taken out a floating rate loan*. For convenience, assume the principal is \$1. She can purchase an *interest rate cap* to protect herself against unacceptable increases in the floating rate.

A cap set at strike K and lasting until T_{n+1} will pay her $\delta(L_\delta(T_{j-1}, T_{j-1}) - K)^+$ at each time T_j, $1 \leq j \leq n+1$. This means that she will never pay more than rate K over any period; the cap will make up the difference between the $\delta L_\delta(T_{j-1}, T_{j-1})$ she owes the bond holder and the maximum δK she wishes to pay. We shall use $\mathrm{Cap}^{\mathrm{m}}(0, n+1)$ to denote the market price of this cap at time $T_0 = 0$.

Consider the derivative which pays the interest rate cap only at time T_j. So it consists of *a single payoff* $\delta(L_\delta(T_{j-1}, T_{j-1}) - K)^+$ at T_j. This is called a *caplet*. Caplets are not traded as such. However, we can imagine them for the purposes of pricing. Clearly, if $\text{Caplet}_j(0)$ denotes the price of this caplet at time $T_0 = 0$, the total price at $T_0 = 0$ of a cap of maturity T_{n+1} will be

$$\sum_{j=1}^{n+1} \text{Caplet}_j(0).$$

If caps of all maturities are available on the market, we can create a caplet with payoff at T_j by going long one cap maturing at T_j and short one cap maturing at T_{j-1}. Thus the market price of the caplet at T_j is

$$\text{Caplet}_j(0) = \text{Cap}^m(0, j) - \text{Cap}^m(0, j-1).$$

Just as there are interest rate caps, there are also interest floors. By going long a cap and short a floor, one can create also a *collar* that keeps the interest rate one pays between two levels.

Interest rate caps and floors are widely traded and their prices are readily available from the market.

12.4.2 A remark on the Black-Scholes formula

The pricing formula for the caplet follows the argument of the Black-Scholes formula. The derivation of the Black-Scholes formula is a direct consequence of the following result about normal random variables, which in turn is a consequence of Corollary 1 in the class lecture notes, *Review of Mathematical Finance I*.

Theorem 6. *If Y is a normal random variable with mean 0 and variance ν^2,*

$$E\left[\left(xe^{Y-\nu^2/2} - K\right)^+\right] = xN\left(\frac{\ln(x/K) + \nu^2/2}{\nu}\right) - KN\left(\frac{\ln(x/K) - \nu^2/2}{\nu}\right). \quad (12.5)$$

To see the connection to the Black-Scholes formula, note that the price at time 0 of a call with strike K is

$$e^{-rT}\tilde{E}\left[\left(xe^{\sigma\widetilde{W}(T)+rT-\frac{1}{2}\sigma^2T} - K\right)^+\right] = e^{-rT}\tilde{E}\left[\left(xe^{rT}e^{\sigma\widetilde{W}(T)-\frac{1}{2}\sigma^2T} - K\right)^+\right].$$

Since $\sigma\widetilde{W}(T)$ is a normal random variable with mean 0 and variance σ^2T, we are exactly in the situation of Theorem 6, and it is easy to derive the Black-Scholes formula from (12.5).

12.4.3 Black's caplet model and pricing formula

The idea behind Black's caplet model and price is to take advantage of Theorem 6 by positing lognormal models where possible. We already saw this strategy in section 9.4 of Shreve, where we assumed T-forward prices for a given T were lognormal. The idea for caplets is similar. Consider the caplet that pays $\delta(L_\delta(T_j,T_j)-K)^+$ at T_{j+1}. We posit that there is a risk-neutral model $\widetilde{\mathbf{P}}^{T_{j+1}}$ under which T_{j+1} forward prices are martingales, that there is a Brownian motion $\widetilde{W}^{T_{j+1}}$ under $\widetilde{\mathbf{P}}^{T_{j+1}}$ and that

$$dL_\delta(t,T_j)=\gamma(t,T_j)L_\delta(t,T_j)\,d\widetilde{W}^{T_{j+1}}, \tag{12.6}$$

where $\gamma(t,T_j)$ is deterministic. Equivalently,

$$L_\delta(t,T_j)=L_\delta(0,T_j)\exp\left\{\int_0^t\gamma(u,T_j)\,d\widetilde{W}^{T_{j+1}}(u)-\frac{1}{2}\int_0^t\gamma^2(u,T_j)\,du\right\}.$$

For convenience of notation, let

$$\bar{\gamma}^2(T_j)=\frac{1}{T_j}\int_0^{T_j}\gamma^2(u,T_j)\,du.$$

Let $\mathbf{Caplet}_{j+1}(0,\bar{\gamma}(T_j))$ denote the price at $T_0=0$ of the caplet maturing at T_{j+1}); (we will see that this price depends only on $\bar{\gamma}(T_j)$, if δ and K are fixed, so the notation is appropriate.) By the risk-neutral pricing formula, the T_{j+1}-forward price of the caplet is

$$\frac{\mathbf{Caplet}_{j+1}(0,\bar{\gamma}(T_j))}{B(0,T_{j+1})}=\delta\tilde{E}^{T_{j+1}}\left[\left(L_\delta(0,T_j)e^{\int_0^{T_j}\gamma(u,T_j)\,d\widetilde{W}^{T_{j+1}}(u)-\frac{1}{2}\int_0^{T_j}\gamma^2(u,T_j)\,du}-K\right)^+\right].$$

But, since $\gamma(t,T_j)$ is deterministic, $\int_0^{T_j}\gamma(u,T_j)\,d\widetilde{W}^{T_{j+1}}(u)$ is a normal random variable with mean 0 and variance $\int_0^{T_j}\gamma^2(u,T_j)\,du=T_j\bar{\gamma}(T_j)$. Thus from Theorem 6,

$$\begin{aligned}\frac{\mathbf{Caplet}_{j+1}(0,\bar{\gamma}(T_j))}{B(0,T_{j+1})} &= \delta L_\delta(0,T_j)N\left(\frac{\ln\frac{L_\delta(0,T_j)}{K}+\frac{1}{2}\bar{\gamma}^2(T_j)T_j}{\bar{\gamma}(T_j)\sqrt{T_j}}\right)\\ &\quad-\delta KN\left(\frac{\ln\frac{L_\delta(0,T_j)}{K}-\frac{1}{2}\bar{\gamma}^2(T_j)T_j}{\bar{\gamma}(T_j)\sqrt{T_j}}\right)\end{aligned}$$

In this way, we derive *Black's caplet formula*:

$$\begin{aligned}\mathbf{Caplet}_{j+1}(0,\bar{\gamma}(T_j)) &= B(0,T_{j+1})\Bigg[\delta L_\delta(0,T_j)N\left(\frac{\ln\frac{L_\delta(0,T_j)}{K}+\frac{1}{2}\bar{\gamma}^2(T_j)T_j}{\bar{\gamma}(T_j)\sqrt{T_j}}\right)\\ &\quad-\delta KN\left(\frac{\ln\frac{L_\delta(0,T_j)}{K}-\frac{1}{2}\bar{\gamma}^2(T_j)T_j}{\bar{\gamma}(T_j)\sqrt{T_j}}\right)\Bigg]\end{aligned} \tag{12.7}$$

The implied spot volatility is a number γ_j, which, when substituted into Black's caplet formula, gives the market value:

$$\mathbf{Caplet}_{j+1}(0, \gamma_j) = \text{Caplet}_{j+1}(0).$$

By finding the implied volatilities and then choosing $\gamma(t, T_j)$ for each j so that

$$\int_0^{T_j} \gamma^2(u, T_j)\, du = T_j \gamma_j^2,$$

we can fit Black's model to the market for all j.

We emphasize that this model is formulated directly for forward LIBOR and does not assume that one has formulated a prior model, such as an HJM model, for zero-coupon bond prices.

12.5 Calibration of forward LIBOR model

12.5.1 Motivation

From the above, we've seen that the forward LIBOR rates for different maturity $T_j, 1 \leq j \leq n$ have the dynamics:

$$dL_\delta(t, T_j) = \gamma(t, T_j) L_\delta(t, T_j)\, d\widetilde{W}^{T_{j+1}},$$

where $\widetilde{W}^{T_{j+1}}$ is a Brownian motion under the T_{j+1} forward measure.

The financial products associated with these LIBOR rates are the caplets that pay $\delta(L_\delta(T_{j-1}, T_{j-1}) - K)^+$ at T_j. The market price of these caplets can be derived from the price of the caps:

$$\text{Caplet}_j(0) = \text{Cap}^m(0, j) - \text{Cap}^m(0, j-1).$$

On the other hand, from the model of the LIBOR rates, we can also derive, under the assumption that $\gamma(t, T_j)$ are determinstic, via Black-Scholes formula, the theoretical price of these caplets. We denote these prices by $\mathbf{Caplet}_j(0, \bar{\gamma}(T_{j-1}))$.

The obvious question is: can we build a model of these forward LIBOR rates so that

$$\text{Caplet}_j(0) = \mathbf{Caplet}_j(0, \bar{\gamma}(T_{j-1}))?$$

The answer is of course yes. Since $\mathbf{Caplet}_j(0, \bar{\gamma}(T_{j-1}))$ is a function of $\bar{\gamma}(T_{j-1}))$ we can choose a number γ_{j-1} so that the above equation holds:

$$\text{Caplet}_j(0) = \mathbf{Caplet}_j(0, \gamma_{j-1}).$$

γ_{j-1} is called the implied volatility of the LIBOR rate with maturity T_j. In general, we do not have an explicit formula for γ_{j-1}. The way to find γ_{j-1} is via numerical procedure, but it can be done.

Next we can construct a determinsitic function $\gamma(t, T_{j-1})$ so that

$$\int_0^{T_{j-1}} \gamma^2(t, T_{j-1}) = T_{j-1}\gamma_{j-1}.$$

There is much freedom in choosing $\gamma(t, T_{j-1})$ of course.

We may think the next step is just to construct n Brownian motions: $\widetilde{W}^{T_{j+1}}, 1 \leq j \leq n$ (question: how are they related?) and from which we can derive n LiBOR rates $L_\delta(t, T_j), 1 \leq j \leq n$ from which the Caplet price will match the market data. But this is missing some details.

First, we want to build a consistent model for $L_\delta(t, T_j), 1 \leq j \leq n$, beyond just matching the market data at time 0. Recall the definition of the LIBOR rates:

$$\begin{aligned} L_\delta(t, T_j) &= \frac{1}{\delta}\frac{B(t, T_j) - B(t, T_j + \delta)}{B(t, T_j + \delta)} \\ &= \frac{1}{\delta}\frac{B(t, T_j) - B(t, T_{j+1})}{B(t, T_{j+1})} \end{aligned}$$

So even without knowing the details, we should suspect that $L_\delta(t, T_j)$ and $L_\delta(t, T_{j+1})$ are related at some level. If we simulate $\widetilde{W}^{T_{j+1}}$ and $\widetilde{W}^{T_{j+2}}$ without regards to this relation, we're missing certain things.

Second, suppose starting out from the risk neutral measure $\widetilde{P}$, we have the dynamics of the bond $B(t, T_j)$ as

$$dB(t, T_j) = R(t)B(t, T_j)dt + \sigma^*(t, T_j)B(t, T_j)\widetilde{W}(t).$$

Note that there is only one Brownian motion $\widetilde{W}$ here, which is *independent* of T_j. (The choice of how many Brownian motions we put in is up to us, of course, but the point is that we use the same Brownian motions to model the dynamics of $B(t, T_j)$ for different T_j). So from what we learned from the change of numéraire section, the Brownian motion $\widetilde{W}^{T_{j+1}}$ are all related to $\widetilde{W}$ via the equation:

$$d\widetilde{W}^{T_j}(t) = d\widetilde{W}(t) + \sigma^*(t, T_j)\, dt.$$

Thus all Brownian motions $\widetilde{W}^{T_j}$ are related actually. So to model $L_\delta(t, T_j)$ properly, beyond determining the $\gamma(t, T_j)$ to match the market data, we also need to learn about the relations of $L_\delta(t, T_j)$. We will do so in the next section.

Finally, as the bond price $B(t, T_j)$ and LIBOR rates $L_\delta(t, T_j)$ are clearly related, we will see that by modeling the $L_\delta(t, T_j)$ properly, this will also give us a handle on how to model the volatility $\sigma^*(t, T_j)$ of the bonds and the (discounted) value of the bond $B(t, T_j)$ themselves. The details will be given in the third section.

12.5.2 Consistent forward LIBOR models - Relation among the $L_\delta(t, T_j)$

Relation among $\widetilde{W}^{T_j}$

Recall from section (12.5.1) that for every j, $d\widetilde{W}^{T_j}(t) = d\widetilde{W}(t) + \sigma^*(t, T_j)\, dt$. In particular, it follows from this that

$$d\widetilde{W}^{T_j}(t) = d\widetilde{W}^{T_{j+1}}(t) + [\sigma^*(t, T_j) - \sigma^*(t, T_{j+1})]\, dt \tag{12.8}$$

Next, from the dynamics of $L_\delta(t, T_j)$ that we derived before:

$$dL_\delta(t, T_j) = L_\delta(t, T_j) \left\{ \frac{1 + \delta L_\delta(t, T_j)}{\delta L_\delta(t, T_j)} [\sigma^*(t, T_{j+1}) - \sigma^*(t, T_j)] \right\} d\widetilde{W}^{T_{j+1}}(t).$$

This will be the same as the Black model $dL_\delta(t, T_j) = \gamma(t, T_j) L_\delta(t, T_j)\, d\widetilde{W}^{T_{j+1}}(t)$ only if

$$\gamma(t, T_j) = \frac{1 + \delta L_\delta(t, T_j)}{\delta L_\delta(t, T_j)} [\sigma^*(t, T_{j+1}) - \sigma^*(t, T_j)], \quad t \le T_j$$

or equivalently,

$$\sigma^*(t, T_{j+1}) - \sigma^*(t, T_j) = \frac{\delta L_\delta(t, T_j)}{1 + \delta L_\delta(t, T_j)} \gamma(t, T_j), \quad t \le T_j. \tag{12.9}$$

Assume this is the case for all $j \le n$. By combining this result with equation (12.8),

$$d\widetilde{W}^{T_j}(t) = d\widetilde{W}^{T_{j+1}}(t) - \frac{\delta L_\delta(t, T_j)}{1 + \delta L_\delta(t, T_j)} \gamma(t, T_j)\, dt \tag{12.10}$$

The significance of this equation is that *the processes* $\sigma^*(t, T)$ *no longer explicitly appear—everything is expressed in terms of the LIBOR rates themselves and their volatility functions* $\gamma(t, T_j)$.

By working backward with (12.10), $d\widetilde{W}^{T_j}(t)$ can be expressed in terms of $d\widetilde{W}^{T_{n+1}}(t)$ for all j. Indeed,

$$d\widetilde{W}^{T_n}(t) = d\widetilde{W}^{T_{n+1}}(t) - \frac{\delta L_\delta(t, T_n)}{1 + \delta L_\delta(t, T_n)} \gamma(t, T_n)\, dt.$$

But then

$$\begin{aligned} d\widetilde{W}^{T_{n-1}}(t) &= d\widetilde{W}^{T_n}(t) - \frac{\delta L_\delta(t, T_{n-1})}{1 + \delta L_\delta(t, T_{n-1})} \gamma(t, T_{n-1})\, dt \\ &= d\widetilde{W}^{T_{n+1}}(t) - \left[\frac{\delta L_\delta(t, T_n)}{1 + \delta L_\delta(t, T_n)} \gamma(t, T_n) + \frac{\delta L_\delta(t, T_{n-1})}{1 + \delta L_\delta(t, T_{n-1})} \gamma(t, T_{n-1}) \right] dt. \end{aligned}$$

Continuing further, and using what has just been derived,

$$\begin{aligned} d\widetilde{W}^{T_{n-2}}(t) &= d\widetilde{W}^{T_{n-1}}(t) - \frac{\delta L_\delta(t,T_{n-2})}{1+\delta L_\delta(t,T_{n-2})}\gamma(t,T_{n-2})\,dt \\ &= d\widetilde{W}^{T_{n+1}}(t) - \left[\sum_{i=n-2}^{n} \frac{\delta L_\delta(t,T_i)}{1+\delta L_\delta(t,T_i)}\gamma(t,T_i)\right] dt. \end{aligned}$$

Clearly, this will yield for general $j \le n$ that

$$d\widetilde{W}^{T_j}(t) = d\widetilde{W}^{T_{n+1}}(t) - \left[\sum_{i=j}^{n} \frac{\delta L_\delta(t,T_i)}{1+\delta L_\delta(t,T_i)}\gamma(t,T_i)\right] dt. \tag{12.11}$$

The significance of this equation, compare with (12.10) is that *now all $\widetilde{W}^{T_j}$ is written in terms of $\widetilde{W}^{T_{n+1}}$. Thus, instead of generating n Brownian motions, we only need to generate one Brownian motion $\widetilde{W}^{T_{n+1}}$. This is consistent with what we mentioned before that we started out with only one Brownian Motion under risk neutral measure $\widetilde{W}$.*

The relation among the $L(t,T_j)$ - Their construction

Now we can write down a coherent system of equations for the LIBOR forward rates . First of all, Black's model for $j = n$ gives

$$dL_\delta(t,T_n) = L_\delta(t,T_n)\gamma(t,T_n)\,d\widetilde{W}^{T_{n+1}}(t), \quad t \le T_n. \tag{12.12}$$

Next, for arbitrary $j < n$, $dL_\delta(t,T_j) = L_\delta(t,T_j)\gamma(t,T_j)\,d\widetilde{W}^{T_{j+1}}(t)$, and so

$$dL_\delta(t,T_j) = L_\delta(t,T_j)\gamma(t,T_j)\left[-\sum_{i=j+1}^{n} \frac{\delta L_\delta(t,T_i)}{1+\delta L_\delta(t,T_i)}\gamma(t,T_i) + d\widetilde{W}^{T_{n+1}}(t)\right], \quad t \le T_j \tag{12.13}$$

This system of equations makes no reference to the original risk-neutral HJM model. In fact, it can stand alone as its own model. By working backwards on this set of equations using standard theorems, one can prove that it generates a consistent model for caplets of all maturities up to T_{n+1}, without assuming the prior existence of an HJM model for $B(t,T)$. We state this result and summarize the forward LIBOR model in the following theorem. The proof will be given in the next subsection where we discuss the relation between forward measures, see (8).

Theorem 7. *Let there be given a probability space with measure $\widetilde{\mathbf{P}}^{T_{n+1}}$ supporting a Brownian motion $\widetilde{W}^{T_{n+1}}$. Then there exists a unique solution $L_\delta(t,T_1),\dots,L_\delta(t,T_n)$ to the system of equations (12.12)–(12.13). If the measures $\widetilde{\mathbf{P}}^{T_j}$, $j = n, n-1, \dots, 1$ are defined* *recursively by*

$$\widetilde{\mathbf{P}}^{T_j}(A) = \tilde{E}^{T_{j+1}}\left[\mathbf{1}_A \frac{1+\delta L_\delta(T_j,T_j)}{1+\delta L_\delta(0,T_j)}\right],$$

and the processes $\widetilde{W}^{T_j}(t)$, $1 \le j \le n$, are defined recursively by

$$d\widetilde{W}^{T_j}(t) = d\widetilde{W}^{T_{j+1}}(t) - \frac{\delta L_\delta(t,T_j)}{1+\delta L_\delta(t,T_j)}\gamma(t,T_j)\,dt,$$

then $\widetilde{W}^{T_j}$ is a Brownian motion under $\widetilde{\mathbf{P}}^{T_j}$ for each $j \le n$ and

$$dL_\delta(t,T_j) = L_\delta(t,T_j)\gamma(t,T_j)\,d\widetilde{W}^{T_{j+1}}(t), \quad \textit{for each } j \le n.$$

Changing between T-forward measures

Let $0 < T < T'$. Suppose that we have a risk-neutral model for the T' forward prices of a market in which zero-coupon bonds are offered on all maturities. We are not assuming that this has necessarily been derived from an HJM model, just that we have a probability space with a measure $\widetilde{\mathbf{P}}^{T'}$ under which the T'-forward prices of all assets are martingales. Let us denote the T' forward price of an asset whose price in dollars is $S(t)$ by $S^{T'}(t) = S(t)/B(t,T')$. In particular, the T'-forward price of a zero-coupon bond maturing at T, which is

$$B^{T'}(t,T) = \frac{B(t,T)}{B(t,T')}, \quad t \le T,$$

is a martingale under $\widetilde{\mathbf{P}}^{T'}$. The T forward price of an asset whose T' forward price is $S^{T'}(t)$ is

$$S^T(t) = \frac{S(t)}{B(t,T)} = \frac{S(t)/B(t,T')}{B(t,T)/B(t,T')} = \frac{S^{T'}(t)}{B^{T'}(t,T)}.$$

We are interested in finding the $\widetilde{\mathbf{P}}^T$-forward measure that makes prices $S^T(t)$ into martingales. Since we are not starting from an HJM model as in the previous section, we want to derive this in terms of the T'-forward measure. Denote expectation with respect to $\widetilde{\mathbf{P}}^{T'}$ by $\tilde{E}^{T'}$.

Theorem 8. *Define, $\widetilde{\mathbf{P}}^T$ by*

$$\widetilde{\mathbf{P}}^T(A) = \frac{B(0,T')}{B(0,T)}\tilde{E}^{T'}[\mathbf{1}_A \frac{1}{B(T,T')}] \tag{12.14}$$

Then if an asset is such that its T'- forward price is a martingale under $\widetilde{\mathbf{P}}^{T'}$ then its T-forward price is also a martingale under $\widetilde{\mathbf{P}}^T$.

This theorem is a generalization of formula (9.2.7) in Shreve.

Heuristic idea:

The intuitive idea why formula (12.14) is true is as followed. We want to convert from $\widetilde{\mathbf{P}}^{T'}$ to $\widetilde{\mathbf{P}}^{T}$. The numéraire associated with $\widetilde{\mathbf{P}}^{T}$ is $B(t,T)$. The price process of this numéraire under $\widetilde{\mathbf{P}}^{T'}$ is

$$N(t) := \frac{B(t,T)}{B(t,T')}.$$

Thus the change of measure formula states that

$$\begin{aligned}\widetilde{\mathbf{P}}^{T}(A) &= \tilde{E}^{T'}[\mathbf{1}_A \frac{N(T)}{N(0)}] \\ &= \frac{B(0,T')}{B(0,T)}\tilde{E}^{T'}[\mathbf{1}_A \frac{1}{B(T,T')}].\end{aligned}$$

Compare this with what we did for change of measure from $\widetilde{\mathbf{P}}$ to $\widetilde{\mathbf{P}}^{(N)}$, for example. The numéraire under $\widetilde{\mathbf{P}}^{(N)}$ is clearly $N(t)$. Its "price" under $\widetilde{\mathbf{P}}$ is $D(t)N(t)$. Therefore the change of measure formula is

$$\widetilde{\mathbf{P}}^{(N)}(A) = \widetilde{E}[\mathbf{1}_A \frac{D(T)N(T)}{D(0)N(0)}]$$

Rigorous proof:

The proof is an application of Lemma 5.2.2 in Shreve: Suppose that $Z(t)$ is a positive martingale under a probability measure $\mathbf{P}$ and define

$$\mathbf{P}^Z(A) = E[\mathbf{1}_A Z(T)]/Z(0).$$

Then if $M(t)$ is a martingale under $\mathbf{P}$,

$$\{M(t)/Z(t);\, t \leq T\}$$

is a martingale under $\mathbf{P}^Z$. To prove the theorem, simply apply this principle with $\widetilde{\mathbf{P}}$ in place of $\mathbf{P}$ and $B^{T'}(t,T) = B(t,T)/B(t,T')$ in place of $Z(t)$. Note that the definition in (12.14) is the same as

$$\widetilde{\mathbf{P}}^{T}(A) = \tilde{E}^{T'}[\mathbf{1}_A B^{T'}(T,T)]/B^{T'}(0,T).$$

Since a T' forward price $S^{T'}(t)$ is a martingale under $\widetilde{\mathbf{P}}^{T'}$, it follows that the T forward price

$$S^T(t) = S^{T'}(t)/B^{T'}(t,T),$$

is a martingale under $\widetilde{\mathbf{P}}^{T}$ as defined in (12.14). This completes the proof.

12.5.3 Construction the T_j-Maturity Discounted Bonds

Construction of $\sigma^*(t,T_j)$

The above theorem does not give us a HJM model, which is defined in terms of functions $\sigma^*(t,T_j)$ on the risk-neutral probability for prices denominated in the domestic currency. This is done in Shreve on pages 444-447. We will only outline the main idea here.

With the deterministic functions $\gamma(t,T_j)$ in hand, we can construct the functions $\sigma^*(t,T_j)$ that are consistent with $\gamma(t,T_j)$

$$\sigma^*(t,T_{j+1}) - \sigma^*(t,T_j) = \frac{\delta L_\delta(t,T_j)}{1+\delta L_\delta(t,T_j)}\gamma(t,T_j) \quad t \le T_j.$$

By writing this as

$$\sigma^*(t,T_{j+1}) = \sigma^*(t,T_j) + \frac{\delta L_\delta(t,T_j)}{1+\delta L_\delta(t,T_j)}\gamma(t,T_j), \quad t \le T_j. \tag{12.15}$$

we see that $L_\delta(t,T_j)$, $\gamma(t,T_j)$, and $\sigma^*(t,T_j)$ *determine* $\sigma^*(t,T_{j+1})$ at least for $t \le T_j$. This leads to a recursive procedure for defining $\sigma^*(t,T_j)$. We outline the procedure of construction here:

1. Choose $\sigma^*(t,T_1)$ for $0 \le t \le T_1$. The only constraint is

$$\lim_{t\to T_1} \sigma^*(t,T_1) = 0.$$

2. Construct $\sigma^*(t,T_2)$ for $0 \le t \le T_1$ (note the time interval) using the relation

$$\sigma^*(t,T_{j+1}) = \sigma^*(t,T_j) + \frac{\delta L_\delta(t,T_j)}{1+\delta L_\delta(t,T_j)}\gamma(t,T_j), \quad t \le T_j.$$

3. Choose $\sigma^*(t,T_2)$ for $T_1 \le t \le T_2$ (again note the time interval). The only constraint is

$$\lim_{t\to T_2} \sigma^*(t,T_2) = 0.$$

4. Repeat this procedure to construct $\sigma^*(t,T_j)$ for $j \ge 3$.

Observe that in the above procedure, we had freedom to construct $\sigma^*(t,T_j)$ on the interval $T_{j-1} \le t \le T_j$ subject to the only constraint

$$\lim_{t\to T_j} \sigma^*(t,T_j) = 0.$$

Thus there is also much freedom in constructing $\sigma^*(t,T_j)$.

Construction the T_j-Maturity Discounted Bonds

Now that we have constructed $\sigma^*(t, T_j)$, the dynamics of the bond $B(t, T_j)$ under the risk neutral measure $\widetilde{P}$ is straightforward:

$$dB(t, T_j) = R(t)B(t, T_j)dt - \sigma^*(t, T_j)B(t, T_j)d\widetilde{W}(t).$$

Since we constructed the LIBOR rate under the forward measure $\widetilde{P}^{T_{n+1}}$ and the Brownian motion $\widetilde{W}^{T_{n+1}}$, it's also convenient to write the dynamics of $B(t, T_j)$ using these as well:

$$dB(t, T_j) = R(t)B(t, T_j)dt + \sigma^*(t, T_j)\sigma^*(t, T_{n+1})B(t, T_j)dt - \sigma^*(t, T_j)B(t, T_j)d\widetilde{W}^{T_{n+1}}(t).$$

Lastly, since we haven't constructed $R(t)$, it is better to write the dynamics of the discounted bond price instead:

$$d\Big(D(t)B(t, T_j)\Big) = \sigma^*(t, T_j)\sigma^*(t, T_{n+1})D(t)B(t, T_j)dt - \sigma^*(t, T_j)D(t)B(t, T_j)d\widetilde{W}^{T_{n+1}}(t).$$

We need the initial conditions to generate the bonds. They can be obtained from the LIBOR rates we have constructed as well:

$$D(0)B(0, T_j) = B(0, T_j) = \prod_{i=0}^{j-1} \frac{B(0, T_{i+1})}{B(0, T_i)} = \prod_{i=0}^{j-1} \Big(1 + \delta L_\delta(0, T_i)\Big)^{-1}.$$

The verification that our construction is consistent: $D(t)B(t, T_j)$ is a martingale under $\widetilde{P}$ is stated in Shreve's Theorem 10.4.4. (The only subtle point is we start out modeling under the forward measure $\widetilde{P}^{T_{n+1}}$. So we need to *define* the risk neutral measure $\widetilde{P}$ from $\widetilde{P}^{T_{n+1}}$. After that the verification is straightforward.)

References

[1] Kurtz, Thomas G. *Lectures on Stochastic Analysis.* Department of Mathematics and Statistics, University of Wisconsin, Madison, WI (2001): 53706-1388.

[2] Papapantoleon, Antonis. *An introduction to Lé vy processes with applications in finance.* arXiv preprint arXiv:0804.0482 (2008).

[3] Shreve, Steven E. *Stochastic calculus for finance II: Continuous-time models.* Vol. 11. Springer Science & Business Media, 2004.