# An Introduction to Asset Pricing Theory

Junhui Qian

# Preface

This note introduces asset pricing theory to Ph.D. students in finance. The emphasis is put on dynamic asset pricing models that are built on continuous-time stochastic processes.

It is very preliminary. Please let me know if you discover any mistake.

Shanghai, China, *Junhui Qian*
February 2019 jhqian@sjtu.edu.cn

# Contents

# Chapter 1

# Introduction to Asset Pricing Theory

The theory of asset pricing is concerned with explaining and determining prices of financial assets in a uncertain world.

The asset prices we discuss would include prices of bonds and stocks, interest rates, exchange rates, and derivatives of all these underlying financial assets. Asset pricing is crucial for the allocation of financial resources. Mispricing of financial assets would lead to inefficiency in investment and consumption in the real economy.

The "uncertainty" in this book is, rather simplistically, described by probability distributions. A more sophisticated treatment would differentiate uncertainty from risk as in Knight (1921). Here we treat uncertainty and risk as the same thing: future variation that can be characterized by some distribution without ambiguity. In this book, uncertainty is assumed in both how an asset would pay in the future and how agents would discount the payoff.

In this book we also take the simplistic view that the uncertainty is given and that it is not influenced by the evolution of prices. It is generally believed in the investment community, however, that prices may affect future payoffs. For example, a surge in stock price would lower financing cost for the company and boost future earnings. We do not go into this direction.

In this first chapter, we get familiarized with some basic theoretical abstractions. Then we study no-arbitrage pricing in a simple context. Key concepts such as state prices, risk-neutral probability, and stochastic discount factor, are introduced. Finally, we connect the no-arbitrage pricing to a representative consumer problem and endow the stochastic discount factor with economic meaning.

Classical asset pricing models, such as CAPM and APT (Arbitrage Pricing

Theory), are discussed as special cases of modern asset pricing theory using stochastic discount factor. A classical derivation of CAPM is offered in the Appendix.

## 1.1 Basic Abstractions

### Commodity

A commodity is a "good" at a particular time and a particular place when a particular "state" happens. For a commodity to be interesting to economists, it must cost something. Besides physical characteristics, the key characteristic of a commodity is its availability in space and time, and conditionality. A cup of water in the desert and another one in Shanghai, although physically the same, are different commodities. And an umbrella when it rains is also different from that when it does not. It is in the sense of conditionality that we call a commodity a "contingent claim".

In the real world, forwards and futures on oil, ores, gold, and other metals can be understood as commodities. The price of these commodities are in US dollar. In principle, a market of commodities can operate without any money. However, money facilitates trading and more importantly, money is essential for the introduction of a financial market. A commodity economy with a financial market, with much less markets open, can achieve the same allocational efficiency with the Arrow-Debreu (1954) competitive equilibrium.

### Security

Security is financial commodity. Stocks, bonds, and their derivatives are all securities. The payoffs of physical goods are physical goods, but the payoffs of securities may be simply money, as well as other securities. A financial market is a market where securities are exchanged.

Suppose we live at time $t$, the next period is $t+1$. A security has

- price: $p_t$
- payoff: $x_{t+1}$ – a r.v.

For examples, we understand

- stock: $x_{t+1} = p_{t+1} + d_{t+1}$, where $d_{t+1}$ denotes dividend payment.
- bond (zero-coupon, riskless): $x_{t+1} = 1$.

- forwards/futures (on a stock with strike price $K$, long position):

$$x_{t+1} = p_{t+1} - K.$$

- option (European, on a stock with strike price $K$)

$$\begin{aligned} \text{long call:} \quad & x_{t+1} = \max\{0, p_{t+1} - K\} \\ \text{long put:} \quad & x_{t+1} = \max\{0, K - p_{t+1}\}. \end{aligned}$$

For now, we assume $x_{t+1}$ is a finite-dimensional discrete random variable, taking value in $\mathbb{R}^S$. In other words, there are $S$ possible values for $x_{t+1}$, $(x^1, x^2, \cdots, x^S)'$, corresponding to states $s = 1, 2, \cdots, S$ and probability $\pi = (\pi^1, \pi^2, \cdots, \pi^S)'$. We assume $\pi^s > 0$ for all $s$.

## Portfolio

Suppose there are $J$ securities in the market, with prices described by a vector $p = (p_1, p_2, \ldots, p_J)'$, where $p_j$ is the price of security $j$. Individuals may build portfolios of these securities. Mathematically, a portfolio is characterized by a $J$-dimensional vector $h \in \mathbb{R}^J$. The total price of a portfolio $h$ is thus $p'h$.

We use a matrix to describe payoffs of all securities at time $t+1$:

$$X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_J' \end{pmatrix} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^S \\ x_2^1 & x_2^2 & \cdots & x_2^S \\ \vdots & & & \\ x_J^1 & x_J^2 & \cdots & x_J^S \end{pmatrix}.$$

Note that rows correspond to securities and columns correspond to states. The payoff of a portfolio $h$ at time $t+1$ is thus, $X'h$.

## Complete Market

If $\text{rank}(X) = S$, the financial market is complete, meaning that every payoff vector in $\mathbb{R}^S$ can be realized by trading these $J$ securities. More formally, for all $x \in \mathbb{R}^S$, there exists $h \in \mathbb{R}^J$ such that $x = X'h$.

**Definition 1.1.1 (Asset Span)** *For any financial market, the asset span is the space spanned by columns of $X'$:*

$$\mathcal{M} = \{X'h, h \in \mathbb{R}^J\} = span(X').$$

If $\mathcal{M} = \mathbb{R}^S$, the market is complete.

## 1.2 No-Arbitrage Pricing

### 1.2.1 The Law of One Price

The law of one price (LOP) states that portfolios with the same payoff must have the same price:

$$X'h = X'\tilde{h} \quad \Rightarrow \quad p'h = p'\tilde{h},$$

where $p \in \mathbb{R}^J$ is the price vector.

**Theorem 1.2.1** *A necessary and sufficient condition for LOP is: zero payoff has zero price.*

**Proof** : (1) If LOP holds, $X'(h - \tilde{h}) = 0 \quad \Rightarrow \quad p'(h - \tilde{h}) = 0$. (2) If LOP does not hold, then $X'h = X'\tilde{h}$ but $p'h \neq p'\tilde{h}$, this is, there exists a $h^*$ such that $X'h^* = 0$ but $p'h^* \neq 0$.

**Theorem 1.2.2** *For any $z \in \mathcal{M}$, there is a linear pricing functional $q(z)$ iff LOP holds.*

**Proof** (1) Linear functional $\Rightarrow$ LOP holds. If LOP does not hold, then $q(0) = \epsilon \neq 0$, then $q(z + 0) = q(z) + \epsilon$, contradicting the definition of linear functional. (2) LOP holds $\Rightarrow$ linearity. For any $z, \tilde{z} \in \mathcal{M}$, we can find $h$ and $\tilde{h}$ such that $z = X'h$, $\tilde{z} = X'\tilde{h}$. So $\alpha z + \beta \tilde{z} = \alpha X'h + \beta X'\tilde{h}$. So

$$q(\alpha z + \beta \tilde{z}) = \alpha p'h + \beta p'\tilde{h} = \alpha q(z) + \beta q(\tilde{z}).$$

### 1.2.2 No Arbitrage

For a vector $x = (x_1, \ldots, x_n)'$, we define

$x \geq 0$ if $x_i \geq 0$ for all $i$,

$x > 0$ if $x_i \geq 0$ for all $i$ and $x_i > 0$ for some $i$,

**Definition 1.2.3 (Arbitrage)** *An arbitrage is a portfolio $h$ that satisfies $X'h \geq 0$ and $p'h < 0$.*

This definition of arbitrage is sometimes called strong arbitrage (LeRoy and Werner, 2001). An arbitrage portfolio generates nonnegative payoff but has a negative price. If there are arbitrage opportunities, the market is obviously not stable or efficient. The no-arbitrage assumption is thus a weak form of equilibrium or efficiency.

**Theorem 1.2.4** *The payoff functional is linear and positive iff there is no arbitrage.*

**Proof** (1) linear & positivity $\Rightarrow$ no arbitrage: For any $h$ that satisfies $X'h \geq 0$, $p'h = q(X'h) \geq 0$. (2) no arbitrage $\Rightarrow$ linear & positivity: no arbitrage $\Rightarrow$ LOP $\Rightarrow$ linearity, and positivity follows from $z = X'h \geq 0 \Rightarrow q(z) = p'h \geq 0$, for any $h$.

Note that a functional is positive if it assigns nonnegative value to every positive element of its domain. It is strictly positive if it assigns strictly positive value to positive elements.

## An Example

Let's have some flavor of no-arbitrage pricing. Consider a financial market with a money account, a stock, and an European call option on the stock with strike price 98. Suppose there are two future states. If state 1 realizes, the stock price declines to 84 from the current price 100. If state 2 happens, the stock price rises to 112. Suppose the interest rate on the money account is 5%, we want to obtain a no-arbitrage price $c_0$ for the call option. The following table lists the payoff structure of our simple financial market.

| | Time 0 | Time 1 | |
|---|---|---|---|
| | | State 1 | State 2 |
| Money | 1 | 1.05 | 1.05 |
| Stock | 100 | 84 | 112 |
| Call | $c_0$ | 0 | (112-98)=14 |

The idea of no-arbitrage pricing is to form a portfolio of Money and Stock, $h = (\alpha, \beta)'$, that replicates the payoff of the call option, and to deduce the option price from the current price of the replication portfolio. Obviously, we solve the following set of equations for $\alpha$ and $\beta$,

$$\begin{aligned} 1.05\alpha + 84\beta &= 0 \\ 1.05\alpha + 112\beta &= 14 \end{aligned}$$

We obtain,

$$\alpha = -40 \quad \beta = 1/2.$$

This means that we borrow 40 from bank and buy a "half" stock. At time 0, this portfolio has a value of

$$100 \cdot \frac{1}{2} - 40 = 10.$$

And this should be the price of the call option. If the price is 5, then we can form the following portfolio

$$(\alpha, \beta, c) = (45, -1/2, 1)$$

The current price of the portfolio is zero, but the payoff will be 5.25 next period whichever state realizes. This is an arbitrage. If the price is 15, verify that the following portfolio achieves an arbitrage

$$(\alpha, \beta, c) = (-35, 1/2, -1).$$

Obviously, short sell must be allowed to make the above analysis valid. Borrowing from banks can be considered as shorting the money.

## 1.3 State Prices

Consider Arrow-Debreu securities,

$$e^s = (\underbrace{\overbrace{0, 0, \cdots, 0}^{s-1}, 1, 0, \cdots, 0}_{S})'$$

When and only when state $s$ happens, we obtain one unit of payoff. In a complete market, all Arrow-Debreu securities are available. Using Arrow-Debreu securities, we can represent any asset payoff $x$ by a portfolio:

$$x = (x^1, \cdots, x^S)' = \sum_{s=1}^{S} x^s e^s.$$

Let $\varphi^s = q(e^s)$, and

$$\varphi = (\varphi^1, \varphi^2, \cdots, \varphi^S)'$$

$\varphi$ is called state price vector. The no-arbitrage price of $x$ would be

$$q(x) = \sum_{s=1}^{S} x^s q(e^s) = \sum_{s=1}^{S} x^s \varphi^s = x'\varphi.$$

**Theorem 1.3.1** *There is no arbitrage iff there is a state price vector.*

To prove this theorem, we need the Stiemke's Lemma, which is a theorem of alternatives.

**Lemma 1.3.2 (The Stiemke's Lemma)** *For an $m$-by-$n$ matrix $A$, one and only one of the following statements is true:*

*(a) There exists an $x \in \mathbb{R}^n$ and $x > 0$ such that $Ax = 0$.*

*(b) There exists a $y \in \mathbb{R}^m$ with $y'A > 0$.*

Recall that an arbitrage-portfolio is one that satisfies $X'h \geq 0$ and $p'h < 0$. This can be stated mathematically,

$$h'(-p, X) > 0.$$

According to the Stiemke's Lemma, there is no such $h$ iff there exists a vector $\varphi \in \mathbb{R}^S$ and $\varphi > 0$ such that

$$(-p, X)\begin{pmatrix} 1 \\ \varphi \end{pmatrix} = 0.$$

Put differently, we have

$$p = X\varphi.$$

The vector $\varphi$ is the desired state price vector. To see this, let $X$ be an $S$-by-$S$ identity matrix, which characterizes a market for $S$ Arrow-Debreu securities. We then have $p = \varphi$, which means that $q(e^s) = \varphi^s$ for all $s$.

## Example (continued)

In the previous example (the market with a money account and a stock), the payoff matrix and the price vector are as follows,

$$X = \begin{bmatrix} 1.05 & 1.05 \\ 84 & 112 \end{bmatrix}, \quad p = \begin{bmatrix} 1 \\ 100 \end{bmatrix}.$$

The state price $\varphi$ can be determined by solving

$$X\varphi = p,$$

which yields

$$\varphi = \begin{bmatrix} 0.2381 \\ 0.7143 \end{bmatrix}.$$

Using the state price vector, we can price the European call option:

$$c_0 = 0.2381 \times 0 + 0.7143 \times 14 \approx 10.$$

## 1.4 Risk-Neutral Probability

Let

$$\varphi^0 = \sum_{s=1}^{S} \varphi^s = \sum_{s=1}^{S} q(e^s) = q(I),$$

where $I$ is the vector of 1's. $\varphi^0$ is the price of risk-free zero-coupon bond, and

$$R_f = \frac{1}{\varphi^0}$$

is then the yield.

The price of $x$ can be written as

$$p = q(x) = \sum_{s=1}^{S} \varphi^s x^s = \varphi^0 \sum_{s=1}^{S} \frac{\varphi^s}{\varphi_0} x^s.$$

Let

$$\tilde{\pi}^s = \frac{\varphi^s}{\varphi^0},$$

Obviously,

$$\tilde{\pi}^s > 0 \forall s \quad \text{and} \quad \sum_{s=1}^{S} \tilde{\pi}^s = 1.$$

$(\tilde{\pi}^s)$ is called "Risk-Neutral Probability". Using this probability,

$$p = \varphi^0 \sum_{s=1}^{S} \tilde{\pi}^s x^s = R_f^{-1} \tilde{\mathbb{E}} x,$$

where the expectation $\tilde{\mathbb{E}}$ is taken with respect to risk-neutral probability.

Following the example in Section 6, let the risk-neutral probability of state 1 be $\tilde{p}$, then the stock price should satisfy

$$100 = \frac{1}{1.05}(84\tilde{p} + 112(1 - \tilde{p})).$$

So $\tilde{p} = 1/4$. So the price of call option is

$$c_0 = \frac{1}{1.05}(0 \cdot \tilde{p} + 14 \cdot (1 - \tilde{p})) = \frac{1}{1.05} \cdot 14 \cdot \frac{3}{4} = 10.$$

## 1.5 Stochastic Discount Factor

Suppose the objective probability of state $s$ is $\pi^s$,

$$p = R_f^{-1} \sum_{s=1}^{S} \pi^s \frac{\tilde{\pi}^s}{\pi^s} x^s.$$

Let $m^s = R_f^{-1} \frac{\tilde{\pi}^s}{\pi^s}$, $m^s > 0$,

$$\begin{aligned} p &= \sum_{s=1}^{S} \pi^s (m^s x^s) \\ &= \mathbb{E}(mx). \end{aligned} \tag{1.1}$$

$m$ is called "Stochastic Discount Factor". It is also called "change of measure" for obvious reasons.

The equation (1.1) is called the "Central Asset Pricing Formula" (Cochrane, 2005). It holds in more general settings than is spelled so far in this course. In fact, it exists for continuous-valued payoff.

### 1.5.1 Continuous-State World

Now, we use continuous-state random variables to describe future payoffs. Recall that a continuous-state random variable is defined as a mapping from the sample space to the real line,

$$x \equiv x(\omega) : \ \Omega \to \mathbb{R}.$$

In this case, $m$ would also be a $\mathbb{R}$-valued random variable.

We define $\mathcal{M} = \{x \in \mathbb{R} : \mathbb{E}x^2 < \infty\}$. This set contains all "reasonable" payoffs. And we define inner product on $\mathcal{M}$ as,

$$\langle x_1, x_2 \rangle = \mathbb{E}(x_1 x_2).$$

It is well known that $\mathcal{M}$ is a Hilbert space with the above inner product. If there is no arbitrage, then $q$ is a linear positive functional on $\mathcal{M}$. According to Riesz's Representation Theorem, every bounded linear pricing functional $q$ on $\mathcal{M}$ can be represented in terms of the inner product,

$$q(x) = \langle x, m \rangle = \mathbb{E}(mx)$$

for some $m \in \mathcal{M}$. Since $q$ must be positive to rule out arbitrage, $m > 0$ a.s. (almost sure). The reverse is also true. Hence we may conclude that there is no arbitrage iff $m > 0$ a.s. For more details, see Hansen & Richard (1987).

### 1.5.2 Some Ramifications

#### Gross Return

Let $R = x/p$, we have

$$1 = \mathbb{E}mR.$$

#### Risk-Free Rate

If $x = I$, $p = \mathbb{E}m$, So

$$R_f = \frac{1}{\mathbb{E}m}.$$

#### Risk Premium

Covariance decomposition.

$$p = \mathbb{E}(mx) = \text{cov}(m, x) + (\mathbb{E}m)(\mathbb{E}x).$$

So

$$\begin{aligned}\frac{\mathbb{E}x}{p} &= \frac{1}{\mathbb{E}m} - \frac{\text{cov}(m, R)}{\mathbb{E}m} \\ &= R_f - R_f \text{cov}(m, R)\end{aligned}$$

So

$$\mathbb{E}R - R_f = -R_f \text{cov}(m, R).$$

Risk premium is proportional to $\text{cov}(m, R)$.

Sometimes we may decompose $x$ into

$$x = x_m + \varepsilon,$$

where $x_m = \text{proj}(x|m)$ is the systematic component of $x$ and $\varepsilon = x - \text{proj}(x|m)$ is called the idiosyncratic component. It is easy to see that in this case

$$\mathbb{E}R - R_f = -R_f \text{cov}(m, R_m),$$

where $R_m = x_m/p$. In other words, idiosyncratic risk is not priced.

### $\beta$-Pricing

For asset $i$, the gross return can be written as

$$\mathbb{E}R_i = R_f + \left(\frac{\mathrm{cov}(R_i, m)}{\mathrm{var}(m)}\right)\left(-\frac{\mathrm{var}(m)}{\mathbb{E}m}\right)$$

Or,

$$\mathbb{E}R_i = R_f + \beta_{i,m}\lambda_m$$

$\beta_{i,m}$ measures the systematic risk contained in asset $i$.

$\lambda_m$ "price of risk".

### Factor Models

When

$$m = a + b'f, \tag{1.2}$$

where $f$ is a vector of "factors", $b$ factor loadings, and $a$ a constant, we call it "factor models". Without loss of generality, we assume $\mathbb{E}f = 0$, so $\mathbb{E}m = a = 1/R_f$.

Since $1 = \mathbb{E}(mR_i)$, we have

$$\mathbb{E}(R_i) = \frac{1}{\mathbb{E}m} - \frac{\mathrm{cov}(m, R_i)}{\mathbb{E}m} = \frac{1}{a} - \frac{\mathbb{E}(R_i f')b}{a}.$$

Let $\beta_i$ be the regression coefficient of $R_i$ on $f$, thus, $\beta_i \equiv \mathbb{E}(ff')^{-1}\mathbb{E}(fR_i)$. So

$$\mathbb{E}(R_i) = \frac{1}{a} - \frac{\mathbb{E}(R_i f')\mathbb{E}(ff')^{-1}\mathbb{E}(ff')b}{a} = \frac{1}{a} - \beta'\frac{\mathbb{E}(ff')b}{a}.$$

Note that $\mathbb{E}(ff')b = \mathbb{E}mf$. If we define

$$\lambda \equiv -R_f\mathbb{E}(mf),$$

we have

$$\mathbb{E}R_i = R_f + \lambda'\beta_i. \tag{1.3}$$

This generalizes the $\beta$-pricing model.

Reversely, we can also obtain (1.2) from (1.3). Hence factor models are equivalent to beta pricing models. The factor models includes CAPM and APT as special cases.

**Mean-Variance Frontier**

We have

$$\begin{aligned} 1 = \mathbb{E}(mR_i) &= \mathbb{E}m\mathbb{E}R_i + \mathrm{cov}(m, R_i) \\ &= \mathbb{E}m\mathbb{E}R_i + \rho_{m,R_i}\sigma(m)\sigma(R_i) \end{aligned}$$

So

$$\begin{aligned} \mathbb{E}R_i &= \frac{1}{\mathbb{E}m} - \frac{1}{\mathbb{E}m}\rho_{m,R_i}\sigma(m)\sigma(R_i) \\ &= R_f - \rho_{m,R_i}\frac{\sigma(m)}{\mathbb{E}m}\sigma(R_i). \end{aligned}$$

Then, since $|\rho| \leq 1$,

$$|\mathbb{E}R_i - R_f| \leq \frac{\sigma(m)}{\mathbb{E}m}\sigma(R_i).$$

Notice $\frac{\mathbb{E}R_i - R_f}{\sigma(R_i)}$ is the Sharpe ratio, and

$$\frac{\mathbb{E}R_i - R_f}{\sigma(R_i)} \leq R_f\sigma(m).$$

**Rational Bubble and Discounted Cash Flow Valuation**

Consider a stock with price process $p_t$ and dividend process $d_t$. Using the formula $p_t = \mathbb{E}_t m_{t+1} x_{t+1}$, we have

$$\begin{aligned} p_t &= \mathbb{E}_t[m_{t+1}(p_{t+1} + d_{t+1})] \\ &= \mathbb{E}_t[m_{t+1}(\mathbb{E}_{t+1}[m_{t+2}(p_{t+2} + d_{t+2})] + d_{t+1})] \\ &\vdots \\ &= \mathbb{E}_t\left(\prod_{j=1}^{n} m_{t+j}\right)p_{t+n} + \sum_{i=1}^{n}\mathbb{E}_t\left(\prod_{j=1}^{i} m_{t+j}\right)d_{t+i} \\ &= \mathbb{E}_t\left(\prod_{j=1}^{\infty} m_{t+j}\right)p_{\infty} + \sum_{i=1}^{\infty}\mathbb{E}_t\left(\prod_{j=1}^{i} m_{t+j}\right)d_{t+i} \end{aligned}$$

The second term on the last line is the DCF value of the stock. If we assume $(d_{t+i}, i \geq 1)$ are known at time $t$, and $m_t = 1/(1+r)$ with $r > 0$, then the second term reduces to $\sum_{i=1}^{\infty}\frac{1}{(1+r)^i}d_{t+i}$.

The first term is called the rational bubble. For it to exist as $n \to \infty$, $p_{t+n}$ must also go to infinity.

## 1.6 Consumption-based Asset Pricing

We have shown that a discount factor $m$ exists (and positive) under LOP (and no-arbitrage) condition. Now we give this discount factor economic meaning.

Consider a representative agent with a utility function $u$. He lives for two periods, $t$ and $t+1$. In period $t$, he makes investment/consumption decision, trying to maximize his expected life-time utility:

$$\max_{\xi} u(c_t) + \beta \mathbb{E}_t u(c_{t+1})$$

subject to

$$\begin{aligned} c_t &= y_t - p_t \xi \\ c_{t+1} &= y_{t+1} + x_{t+1} \xi \end{aligned}$$

– $\mathbb{E}_t$ is conditional expectation on the information available on time $t$.

– $\beta$ is called "subjective discount factor", or "impatience factor"

– Implicitly assumes separability of utility

– $(y_t)$: endowment (labor income or/and bequeathment)

– $\xi$: amount of asset.

The first order condition gives,

$$p_t u'(c_t) = \mathbb{E}_t \left[ \beta u'(c_{t+1}) x_{t+1} \right] \tag{1.4}$$

The left hand side is the marginal loss of utility of holding asset. And the right hand side is the marginal increase of expected utility of holding asset. Rearrange (1.4), we have

$$\begin{aligned} p_t &= \mathbb{E}_t \left[ \beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right] \\ &= \mathbb{E}_t m_{t+1} x_{t+1}, \end{aligned} \tag{1.5}$$

where

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}. \tag{1.6}$$

In light of (1.6), $m$ is often called "marginal rate of substitution", the rate at which the investor is willing to substitute consumption at time $t+1$ for consumption at time $t$.

For example, if the representative agent has power utility, $u(c) = \frac{1}{1-\gamma}c^{1-\gamma}$, then

$$m_{t+1} = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}.$$

In particular, when $\gamma \to 1$, $u(c) = \log c$, then $m_{t+1}$ would inversely proportional to $c_{t+1}/c_t$,

$$m_{t+1} = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-1}$$

If for some reason $c_{t+1}$ is expected to be tiny (e.g., expect a disaster to happen), the price for asset would be huge.

For another example, if the utility function is quadratic, $u(c) = u_0 - \gamma(c_* - c)^2$ with $\gamma > 0$, $c_* > c$ and $u_0 = \gamma c_*^2$, then

$$m_{t+1} = \beta\frac{c_* - c_{t+1}}{c_* - c_t}.$$

### 1.6.1 CAPM

The Capital Asset Pricing Model (CAPM) model is most frequently stated as:

$$\mathbb{E}R_i = R_f + \beta_{i,R_W}(\mathbb{E}R_W - R_f), \tag{1.7}$$

where $R_W$ denotes the return on the "wealth portfolio". We usually proxy $R_W$ by the return on a stock market index such as S&P 500.

The CAPM model is equivalent to

$$m = a + bR_W. \tag{1.8}$$

The CAPM model can be understood as a consumption-based asset pricing model. To see this, assume quadratic utility, $u(c) = -\frac{1}{2}c_*^2 - \frac{1}{2}(c_* - c)^2$. Thus

$$m_{t+1} = \beta\frac{c^* - c_{t+1}}{c^* - c_t}.$$

And suppose there is no endowment in the second period., $y_{t+1} = 0$. (We may think of $y_t$ as bequeathed wealth instead of labor income.) Then we have

$$c_{t+1} = R_W(y_t - c_t).$$

Hence
$$m_{t+1} = \beta \frac{c_* - R_W(y_t - c_t)}{c_* - c_t} = \frac{\beta c_*}{c_* - c_t} - \frac{\beta(y_t - c_t)}{c_* - c_t} R_W.$$
This is exactly (1.8). We may, of course, derive (1.8) under other assumptions. See Section 9.1 in Cochrane (2005).

### 1.6.2 Arbitrage Pricing Theory

The Arbitrage Pricing Theory (APT) is a linear factor model. It is assumed that we may characterize the asset payoff by a factor structure:
$$R_i = \mathbb{E}R_i + \beta_i' \tilde{f} + \varepsilon_i, \tag{1.9}$$
where $\tilde{f}$ a $K$-by-1 vector of demeaned factors, $\mathbb{E}\varepsilon_i = 0$, $\mathbb{E}f\varepsilon_i = 0$, and $\mathbb{E}\varepsilon_i\varepsilon_j = 0$ for $i \neq j$. The price errors ($\varepsilon_i$) represent idiosyncratic risks that can be diversified away and hence not priced. If we assume $\sigma(m) < \infty$ and LOP holds, we have
$$q(R_i) = \mathbb{E}R_i q(I) + \beta_i' q(\tilde{f}).$$
Since $q(R_i) = 1$, $q(I) = 1/R_f$, we solve the above equation for $\mathbb{E}R_i$:
$$\mathbb{E}R_i = R_f + \beta_i' \lambda,$$
where $\lambda = -R_f q(\tilde{f})$. Recall that the $\beta$-pricing model above is equivalent to linear factor model.

Note that the above derivation starts from statistics and factors are selected to best predict asset payoffs. The general linear factor model can also start from economic intuition. Recall that the stochastic discount factor $m_{t+1}$ is a nonlinear function of consumption $c_{t+1}$. Suppose we find factors $z_t$ that forecast future consumption, say, $c_{t+1} = g(z_t)$. Then $m_{t+1} = \beta u'(g(z_{t+1}))/u'(g(z_t))$, which we can linearize to obtain some forms of linear factor models.

## 1.7 Summary

In this first part of the course, we introduce fundamental concepts and laws of asset pricing. In particular, we show that if the law of one price (LOP) holds, there exists a stochastic discount factor (SDF). If there is no arbitrage, this SDF is strictly positive. Furthermore, if the market is complete, the SDF is unique. We show that the SDF framework incorporates many theories of classical finance. And we show that the SDF has a natural economic interpretation derived from consumer decision. From this we further connect SDF framework to CAPM and APT.

# References

Arrow, K.J. and Debreu, G. (1954), "Existence of an equilibrium for a competitive economy", Econometrica, 22, 265-290.

John H. Cochrane, Asset Pricing, Princeton University Press, 2005.

Chen, Nai-Fu; Roll, Richard; Ross, Stephen (1986). "Economic Forces and the Stock Market". Journal of Business 59 (3): 383-403.

Darrell Duffie, Dynamic Asset Pricing Theory, Princeton University Press, 2001.

Knight, F. H. (1921), "Risk, Uncertainty, and Profit". Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Hansen, L.P. and Richard S.F. (1987), "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587-613.

Lucas R.E.J. (1978), "Asset prices in an exchange economy", Econometrica, 46(6), 1429-1445.

Ross, Stephen (1976). "The arbitrage theory of capital asset pricing". Journal of Economic Theory 13 (3): 341C360.

Stephen F. LeRoy and Jan Werner, Principles of Financial Economics, Cambridge University Press, 2001.

# Chapter 2

# Mathematical Background for Continuous-time Finance

## 2.1 Probability Background

### 2.1.1 Random Variable

A random variable $X$ is defined as a mapping from sample space $\Omega$ to $\mathbb{R}$ with a probability measure $\mathbb{P}$ defined on an $\sigma$-field of $\Omega$, $\mathcal{F}$[1]. $(\Omega, \mathcal{F}, \mathbb{P})$ is called a "probability triple", or "probability measure space".

A $\sigma$-field (or $\sigma$-algebra) of $\Omega$ is a collection of subsets of $\Omega$ containing $\Omega$ itself and the empty set $\phi$, and closed under complements, countable unions. In mathematical language, a random variable is a $\mathcal{F}$-measurable function $X$ from $\Omega$ to $\mathbb{R}$. Being $\mathcal{F}$-measurable means,

$$\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}, \quad x \in \mathbb{R}.$$

Or, $X^{-1}(B) \in \mathcal{F}$ for every Borel set $B \in \mathcal{B}(\mathbb{R})$. $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-field containing all open sets of $\mathbb{R}$.

In intuitive terms, the $\sigma$-field $\mathcal{F}$ is a collection of all events. To see this, consider throwing dimes for three times, the sample space $\Omega$ is

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

[1]We may generalize the definition of random variable. So far what we have defined is "real-valued" random variable. We can also define "vector-valued" random variable, when $X : \Omega \to \mathbb{R}^d$. Or even "function-valued" random variable, when $X : \Omega \to H$, where $H$ is an appropriate function space.

One example of $\sigma$-fields on $\Omega$ is $\mathcal{F} = 2^{\Omega}$, the power set of $\Omega$ that contains all subsets of $\Omega$ including $\emptyset$ and $\Omega$. One element of $\mathcal{F}$ is

$$\{HHH, HHT, HTH, HTT\}.$$

This is an event that may be called "Head appear in the first throw". For another example, the event $\{HTH, HTT\}$ may be called "First Head, second Tail". So $X$ being $\mathcal{F}$-measurable means that $X^{-1}(B)$ is indeed an event. So it makes sense to talk about $\mathbb{P}(X \in B)$.

### 2.1.2 Stochastic Process

A stochastic process is a collection of random variables. We denote a stochastic process by $X = (X_t)$, $t \in \mathcal{T}$, where $\mathcal{T}$ is an index set. The index set $\mathcal{T}$ can be a discrete set such as $\{1, 2, ...\}$, or a continuous set, say $[0, 1]$.

More rigorously, $X$ is a mapping from the product space of $\Omega \times \mathcal{T}$ to $\mathbb{R}$. So we may write $X_t = X_t(\omega) = X(\omega, t)$. $X(t, \cdot)$ is a random variable, and $X(\cdot, \omega)$ is a sample path. We say that $X$ is measurable if for all $A \in \mathcal{B}(\mathbb{R})$,

$$\{(t, \omega)|X_t(\omega) \in A\} \in \mathcal{B}([0, 1]) \otimes \mathcal{F},$$

where $\otimes$ denotes product of sigma fields.

Continuous-time process is stochastic process with a continuous index set $\mathcal{T}$, e.g., $[0, 1]$. Continuous process refers to $X$ that takes value in continuous subsets of $\mathbb{R}$, rather than discrete subsets such as $\mathbb{N}$.

### 2.1.3 Filtration

A filtration is a non-decreasing indexed sequence of $\sigma$-fields $(\mathcal{F}_t)$. $(\mathcal{F}_s) \subset (\mathcal{F}_t)$ if $s < t$. As previously argued, a $\sigma$-field is a collection of events. The more inclusive a $\sigma$-field is, the more we may possibly know about the state of the nature $\Omega$. So we can roughly think of $\sigma$-fields as information sets. A filtration is thus an ever-finer sequence of info sets.

The natural (or standard) filtration of $X = (X_t)$ is defined by $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$, that is, the $\sigma$-filed generated by $((X_s)_{s \leq t})$, or intuitively speaking, the information contained in the stochastic process up to time $t$. Recall that $\sigma$-field generated by a random variable $X$, $\sigma(X)$, is defined as

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

where $\mathbb{B}(\mathbb{R})$ denotes the Borel $\sigma$-field of $\mathbb{R}$. Roughly speaking, $\sigma(X)$ is the set of all information we may know through observation of $X$.

A stochastic process $X = (X_t)$ is said to be adapted to a filtration $(\mathcal{F}_t)$ if, for every $t \geq 0$, $X_t$ is a $\mathcal{F}_t$-measurable random variable.

### 2.1.4 Brownian Motion

A one-dimensional Brownian motion is defined as a $\mathbb{R}$-valued process $W = (W_t)$ such that

- Continuous sample path a.s.
- Independent Gaussian increment, $W_t - W_s$ independent of $\mathcal{F}_s$ and $W_t - W_s | \mathcal{F}_s \sim N(0, t-s)$ for $t \geq s$.
- $W_0 = 0$ a.s.

**Results**

(1) $W_t \sim N(0, t)$

(2) $(W_{t_1}, W_{t_2}, ..., W_{t_n})' \sim$ Multivariate normal. To see this,

$$\begin{pmatrix} W_s \\ W_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} W_s \\ W_t - W_s \end{pmatrix}.$$

(3) $W_t | W_s = x \sim N(x, t-s)$.

**Properties**

(a) Time-homogeneity
$V_t = W_{t+s} - W_s$ for any fixed $s$ is a BM ind of $\mathcal{F}_s$.

(b) Symmetry
$V_t = -W_t$ is a BM.

(c) Scaling
$V_t = cW_{t/c^2}$ is a BM

(d) Time inversion
$V_0 = 0$, $V_t = tW_{1/t}$, $t > 0$ is a BM.

(e) Law of iterated logarithm.
$W_t \leq \sqrt{2t \log\log t}$.

(f) Sample path: continuous, but no where differentiable. In fact, it is Höolder continuous:

$$W_t - W_s \leq k|t-s|^{1/2-\delta}, \forall \delta > 0.$$

(g) $V_t = W_t^2$ is self-similar (of order 1):

$$(V_{ct}, t \geq 0) =^d (cV_t, t \geq 0).$$

### 2.1.5 Martingales

$(M_t, \mathcal{F}_t) \sim$ a martingale if $M$ is adapted to $\mathcal{F}$ (or, $M_t$ is $\mathcal{F}_t$-measurable) and

$$\mathbb{E}(M_t|\mathcal{F}_s) = M_s \quad \text{for} \quad s < t.$$

When $\mathbb{E}(M_t|\mathcal{F}_s) \geq M_s$, we call $M_t$ a sub-martingale. When $\mathbb{E}(M_t|\mathcal{F}_s) \leq M_s$, we call it a sup-martingale.

**Remarks:**

(1) BM is a martingale. "prototype" martingale

(2) $W_t^2 - t$ is a martingale.
Proof: use the "incrementalize" method.

(3) $\exp(\lambda W_t - \lambda^2 t/2)$ is a positive martingale.

(4) For any r.v. $X$, let $\xi_t = \mathbb{E}(X|\mathcal{F}_t)$, then $\xi = (\xi_t)$ is a martingale wrt $\mathcal{F}$.

(5) $|M_t|^p$ is a submartingale if $p \geq 1$ and $\mathbb{E}|M_t|^p < \infty$.

### 2.1.6 Markov Processes

$(X_t, \mathcal{F}_t) \sim$ a Markov process if the distribution of $X_t$ conditional on $\mathcal{F}_s$ is identical to the distribution of $X_t$ conditional on $\sigma(X_s)$ for all $s < t$.

The likelihood of the discrete samples of a Markov process has a nice iterative representation. Choose $t_1, t_2, ..., t_n$ arbitrarily. The likelihood of $(X_{t_1}, X_{t_1}, ..., X_{t_n})$, in general, is given by

$$f(X_{t_1}, X_{t_1}, ..., X_{t_n}) = f(X_{t_1}) \cdot f(X_{t_2}|X_{t_1}) \cdot f(X_{t_3}|X_{t_1}, X_{t_2}) \cdots f(X_{t_n}|X_{t_1}, ..., X_{t_{n-1}}).$$

For Markov processes, we have

$$f(X_{t_1}, X_{t_1}, ..., X_{t_n}) = f(X_{t_1}) \cdot f(X_{t_2}|X_{t_1}) \cdot f(X_{t_3}|X_{t_2}) \cdots f(X_{t_n}|X_{t_{n-1}}).$$

So to determine the distribution of a continuous Markov process, we just need to determine the distribution of $X_t|X_s = x$ for any $t$ and $s$.

**Definition 2.1.1 (Transition Probability)** *The transition probability of a Markov process $(X_t)$ is given by*

$$P_{s,t}(x, A) = \mathbb{P}\{X_t \in A | X_s = x\}.$$

Transition probability completely determines distribution of continuous process.

**Definition 2.1.2 (Homogenous Markov process)** *If the transition probability of a Markov process $(X_t)$ satisfies*

$$P_t(x, A) = \mathbb{P}\{X_t \in A | X_0 = x\} = \mathbb{P}\{X_{s+t} \in A | X_s = x\},$$

*we call it a homogenous Markov process.*

**A different notation.** $P_t(x, A)$ describes a distribution by assigning probability values to subsets $A$. We can also describe distribution by its generalized moments $\mathbb{E}f(Z)$, where $Z = X_t|X_s = x$ in our case. If we know $\mathbb{E}f(Z)$ for enough number of $f$, then we know the distribution of $Z$. In fact, we can choose $f$ to be

$$f(\cdot) = I\{\cdot \in A\}.$$

Then

$$\mathbb{E}f(Z) = \mathbb{E}I\{Z \in A\} = \mathbb{P}\{Z \in A\}.$$

In Markov literature, the following notation is often used,

$$P_t f(x) \equiv P_t(x, f) = \mathbb{E}(f(X_t)|X_0 = x).$$

$P_t$ may be regarded as a functional operator:

$$P_t : \quad f \longmapsto P_t f.$$

Transition density may be defined with respect to transition probability,

$$P_t(x, A) = \int_A p_t(x, y)dy.$$

Since $P_t f(x)$ is a conditional expectation, we have

$$P_t f(x) = \int_{-\infty}^{\infty} f(y)p_t(x, y)dy.$$

**Example:** Let $X$ be a Brownian Motion. We have $X_{t+s}|X_s = x \sim N(x, t)$. Hence the transition density,

$$p_t(x, y) = \frac{1}{2\pi\sqrt{t}} \exp(-\frac{(y-x)^2}{2(t)}).$$

## 2.2 Ito Calculus

### 2.2.1 Stochastic Integral

In this section we study the following integral:

$$\int_0^t K_s dM_s,$$

where $(M_t, \mathcal{F}_t)$ is a continuous martingale and $K$ is adapted to $\mathcal{F}$.

First we study the ordinary Lebesgue-Stieltjes integral:

$$\int_0^t f(s)dg(s).$$

For any partition $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$, define

$$S = \sum_i f(s_i)[g(t_i) - g(t_{i-1})],$$

where $t_{i-1} \leq s_i \leq t_i$. Let $\pi_t = \max_i |t_i - t_{i-1}|$. We say that the Lebesgue-Stieltjes integral exists, if $\lim_{\pi_t \to 0} S$ exists for any $s_i \in [t_{i-1}, t_i]$.

Suppose that $f$ is continuous and $g$ is of bounded variation, ie,

$$\sum_i |g(t_i) - g(t_{i-1})| < \infty.$$

For example, if $g$ is monotonely increasing, then it is of bounded variation. We look at

$$\begin{aligned} S1 &= \sum f(s_i)[g(t_i) - g(t_{i-1})] \\ S2 &= \sum f(t_{i-1})[g(t_i) - g(t_{i-1})]. \end{aligned}$$

And find that

$$\begin{aligned} |S_1 - S_2| &\leq \left(\max_i |f(s_i) - f(t_{i-1})|\right) \left(\sum_i |g(t_i) - g(t_{i-1})|\right) \\ &\to 0. \end{aligned}$$

So the Lebesgue-Stieltjes integral exists.

However, it is well known that a martingale $M$ is of bounded variation iff $M$ is constant. In other words, $\int_0^t K_s dM_s$ cannot be defined "path-by-path" as a Lebesgue-Stieltjes integral. Instead, we define

$$\int_0^t K_s dM_s = p \lim_{\pi_t \to 0} \sum_i K_{t_{i-1}}(M_{t_i} - M_{t_{i-1}}). \tag{2.1}$$

The choice of $K_{t_{i-1}}$ in (2.1) makes (2.1) an Ito integral. If we choose $K_{(t_{i-1}+t_i)/2}$, then we have Stratonovich integral. If $K$ is bounded, measurable, and $\mathcal{F}_t$-adapted, then the Ito integral in (2.1) is always well defined.

### 2.2.2 Quadratic Variation

For a continuous martingale $M$, the *quadratic variation* of $M$, denoted by $[M]$, is defined as

$$[M]_t = \text{plim}_{\pi_t \to 0} \sum_{i=1}^n (M_{t_i} - M_{t_{i-1}})^2. \tag{2.2}$$

It is clear that $[M]_t$ is non-decreasing, thus of bounded variation. Thus it is integrable in the Stieltjes sense.

For any continuous process $X$, the first-order variation on $[0,t]$ is captured by $\sum |X_{t_i} - X_{t_{i-1}}|$, and $\sum |X_{t_i} - X_{t_{i-1}}|^2$ captures the second order. Intuitively, for locally smooth stochastic processes, the first-order variation dominates. For locally volatile processes, the first-order variation explodes, but the second-order variation may be well defined.

**Example:** For a Brownian motion $W$, we have $[W]_t = t$. To show this, partition $[0,t]$ into $n$ intervals of equal length $\Delta = t/n$. we have by the law of large number,

$$n\frac{1}{n}\sum_i (W_{i\Delta} - W_{(i-1)\Delta})^2 \to_p n\mathbb{E}(W_{i\Delta} - W_{(i-1)\Delta})^2 = n\Delta = t.$$

## Quadratic Covariation

Given two continuous martingales, $M$ and $N$, their quadratic covariation is defined by

$$[M,N]_t = \text{plim}_{\pi_t \to 0} \sum_{i=1}^{n} (M_{t_i} - M_{t_{i-1}})(N_{t_i} - N_{t_{i-1}}).$$

It is straightforward to show that

$$[X+Y]_t = [X]_t + [Y]_t + 2[X,Y]_t.$$

It is also easy to show that, given a continuous martingale $M$ and bounded-variation process $A$, we have

$$[M,A]_t = 0.$$

### 2.2.3 Semimartingale

If $X$ can be written as $X_t = A_t + M_t$, where $(M_t)$ is a continuous martingale and $(A_t)$ a continuous adapted process of finite variation, then $X$ is called a continuous semimartingale. $A$ constitutes trend, while $M$ determines local variation. A continuous semimartingale $X = A + M$ has a finite quadratic variation and $[X]_t = [M]_t$.

It is clear that Ito integral with respect to semimartingale, $\int_0^t K_s dX_s$, is well defined. We have

$$\int_0^t K_s dX_s = \int_0^t K_s dA_s + \int_0^t K_s dM_s.$$

The second item is Ito integral, and the first item is essentially a Stieltjes integral.

### 2.2.4 Properties of Ito Integral

Consider $P_t = \int_0^t K_s dM_s = \text{plim}_{|\pi_t|\to 0} \sum_i K_{t_{i-1}}(M_{t_i} - M_{t_{i-1}})$, where $(M_t, \mathcal{F}_t) \sim$ is continuous martingale, $K_t$ is adapted, and $\int_0^t K_s^2 ds < \infty$ for all $t$. $P_t$ has the following properties,

(a) $P_t$ is a Martingale.

(b) $[P]_t = \int_0^t K_s^2 d[M]_s$.

(c) If $P_t = \int_0^t K_s dM_s$ and $Q_t = \int_0^t H_s dN_s$, then

$$[P, Q]_t = \int_0^t K_s H_s d[M, N]_s.$$

To understand (a) intuitively, note that

$$P_{t_i} - P_{t_{i-1}} \approx K_{t_{i-1}}(M_{t_i} - M_{t_{i-1}})$$

is a martingale difference sequence. To understand (b), we write

$$\begin{aligned} [P]_t &= \text{plim} \sum (P_{t_i} - P_{t_{i-1}})^2 \\ &= \text{plim} \sum K_{t_{i-1}}^2 (M_{t_i} - M_{t_{i-1}})^2 \\ &\approx \text{plim} \sum K_{t_{i-1}}^2 ([M]_{t_i} - [M]_{t_{i-1}}). \end{aligned}$$

Recall that $[M]_{t_i} = \text{plim} \sum_{k=1}^{i} (M_{t_i} - M_{t_{i-1}})^2$.

**Example: (a)** If $M_t = \int_0^t W_s dW_s$, then $[M]_t = \int_0^t W_s^2 ds$.

**(b)** If $M_t = \int_0^t W_s dW_s$, then $[M, W]_t = \int_0^t W_s d[W, W]_s = \int_0^t W_s ds$.

### 2.2.5 Ito's Formula

**Lemma 2.2.1 (Integration by Parts)** *If $X$ and $Y$ are two continuous semimartingales, then*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t. \tag{2.3}$$

*In particular,*

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + [X]_t.$$

**Proof** It suffices to prove the second statement. We have

$$X_{t_i}^2 - X_{t_{i-1}}^2 = 2X_{t_{i-1}}(X_{t_i} - X_{t_{i-1}}) + (X_{t_i} - X_{t_{i-1}})^2.$$

Taking sum and limit, we obtain the desired result. To prove the first statement, note that $X_tY_t = [(X_t + Y_t)^2 - X_t^2 - Y_t^2]/2$.

In differential form, we may rewrite (2.3) as

$$d(X_tY_t) = X_t dY_t + Y_t dX_t + d[X, Y]_t.$$

Recall that for ordinary functions $f(t)$ and $g(t)$, we have

$$\int_a^b f(t)dg(t) = f(t)g(t)|_a^b - \int_a^b g(t)df(t).$$

Rearranging terms, we have

$$f(b)g(b) = f(a)g(a) + \int_a^b f(t)dg(t) + \int_a^b g(t)df(t).$$

Now we introduce the celebrated Ito's formula.

**Theorem 2.2.2 (Ito's Formula)** *Let $X$ be a continuous semimartingale, and $f \in C^2(\mathbb{R})$, then $f(X)$ is a continuous semimartingale and*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d[X]_s. \tag{2.4}$$

In differential form, we may write the Ito's formula as

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d[X]_t. \tag{2.5}$$

**Proof** We prove by induction. Suppose $dX_t^n = nX_t^{n-1}dX_t + \frac{n(n-1)}{2}X_t^{n-2}d[X]_t$, we prove $dX_t^{n+1} = (n+1)X_t^n dX_t + \frac{n(n+1)}{2}X_t^{n-1}d[X]_t$. It is obviously true for $n = 1$. For arbitrary $n$,

$$\begin{aligned} d(X_t \cdot X_t^n) &= X_t^n dX_t + X_t dX_t^n + d[X, X^n]_t \\ &= X_t^n dX_t + X_t(nX_t^{n-1}dX_t + \frac{n(n-1)}{2}X_t^{n-2}d[X]_t) + nX_t^{n-1}d[X]_t \\ &= (n+1)X_t^n dX_t + \frac{n(n+1)}{2}X_t^{n-1}d[X]_t. \end{aligned}$$

So (2.5) is valid for polynomial functions. We can infer it remains true for all $f \in C^2$.

**Example** Since $dW_t^2 = 2W_t dW_t + dt$, $W_1^2 = \int_0^1 W_t dW_t + 1$, so we have

$$\int_0^1 W_t dW_t = (W_1^2 - 1)/2.$$

Next we introduce the multivariate Ito's formula. Let $X = (X^1, ..., X^d)$ be a vector of continuous semimartingales and $f \in C^2(\mathbb{R}^d, \mathbb{R})$; then $f(X)$ is a continuous semimartingale and

$$f(X_t) = f(X_0) + \sum_i \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s. \quad (2.6)$$

In differential form, we have

$$df(X_t) = \sum_i \frac{\partial f}{\partial x_i}(X_t) dX_t^i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) d[X^i, X^j]_t. \quad (2.7)$$

In particular, for the bivariate case,

$$\begin{aligned} df(X_t, Y_t) &= f_1(X_t, Y_t) dX_t + f_2(X_t, Y_t) dY_t \\ &+ \frac{1}{2} f_{11}(X_t, Y_t) d[X]_t + f_{12}(X_t, Y_t) d[X, Y]_t + \frac{1}{2} f_{22}(X_t, Y_t) d[Y]_t \end{aligned}$$

Furthermore, if $A$ is of bounded variation, we have

$$df(X_t, A_t) = f_1(X_t, A_t) dX_t + f_2(X_t, A_t) dA_t + \frac{1}{2} f_{11}(X_t, A_t) d[X]_t.$$

In particular, if $dX_t = \mu_t dt + \sigma_t dW_t$, then

$$\begin{aligned} df(X_t, t) &= f_1(X_t, t) dX_t + f_2(X_t, t) dt + \frac{1}{2} f_{11}(X_t, t) d[X]_t \\ &= (\mu_t f_1 + f_2 + \frac{1}{2} \sigma_t^2 f_{11}) dt + \sigma_t f_1 dW_t. \end{aligned}$$

This special case often appears as “Ito's formula”.

**Example:** Consider $\xi_t = \exp(\lambda M_t - \frac{\lambda^2}{2}[M]_t) \equiv f(M_t, [M]_t)$, where $M$ is a continuous martingale. $\xi_t$ is called an exponential martingale. We have

$$\begin{aligned} f_1 &= \lambda f \\ f_2 &= -\frac{\lambda^2}{2} f \\ f_{11} &= \lambda^2 f, \end{aligned}$$

So

$$d\exp(\lambda M_t - \frac{\lambda^2}{2}[M]_t) = \lambda\exp(\lambda M_t - \frac{\lambda^2}{2}[M]_t)dM_t,$$

or

$$\exp(\lambda M_t - \frac{\lambda^2}{2}[M]_t) = 1 + \lambda\int_0^t \exp(\lambda M_s - \frac{\lambda^2}{2}[M]_s)dM_s.$$

Note that the exponential martingale is positive.

In general, if $M_t$ is a martingale, $f(M_t)$ is not necessarily a martingale, but it is always a semimartingale. If $X_t$ is semimartingale, the $f(X_t)$ is still semimartingale. So we say that the class of semimartingales is “invariant” under composition with $C^2$-functions.

## 2.3 Diffusions

### 2.3.1 Definition

A diffusion is a continuous-time semimartingale that is characterized by the following stochastic differential equation,

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

where $\mu(\cdot)$ is called the drift function and $\sigma(\cdot)$ is called the diffusion function. E-quivalently, we may represent $X_t$ in integral form,

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

In physics, diffusions describe the movement of a particle suspended in moving liquid.

Let $\Delta$ be a short time interval. On $[t, t+\Delta]$, we have

$$X_{t+\Delta} - X_t = \int_t^{t+\Delta} \mu(X_s)ds + \int_t^{t+\Delta} \sigma(X_s)dW_s.$$

It is clear that

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}(X_{t+\Delta} - X_t|X_t = x) = \mu(x).$$

So $\mu$ measures the rate of instantaneous changes in conditional mean. We also have,

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\text{var}(X_{t+\Delta} - X_t|X_t = x) = \sigma^2(x).$$

So $\sigma^2$ measures the rate of instantaneous changes in conditional volatility.

It is easy to see that if $\mu$ is bounded,

$$\int_t^{t+\Delta} \mu(X_s)ds = O(\Delta),$$

and that if $\sigma$ is bounded,

$$\int_t^{t+\Delta} \sigma(X_s)dW_s = O(\Delta^{1/2}).$$

So if we look at small intervals, diffusions dominates. In fact, drift term is not identifiable in small intervals. At long intervals, the drift part dominates, since

$$\int_0^{T} \mu(X_s)ds = O(T),$$

while

$$\int_0^{T} \sigma(X_s)dWs = O(T^{1/2}).$$

## Linear Drift

The linear (or affine) drift function is widely used in modeling processes with mean reversion. Specifically, we may have

$$\mu(x) = \kappa(u - x),$$

where $\kappa$ and $u$ are parameters. Since $\mathbb{E}(X_{t+\Delta} - X_t) \approx \Delta\kappa(u - X_t)$. So when $\kappa > 0$, the linear drift tend to be "mean reverting", producing downward correction when $X_t > u$. However, $u$ may or may not be the mean of the process. When $\kappa = 0$, the process is a martingale. When $\kappa < 0$, the process is unstable.

## Constant-Elasticity Diffusion

Many diffusion processes are endowed with the following form of diffusion function,

$$\sigma(x) = c|x|^\rho,$$

where $c$ and $\rho$ are constants. From

$$\log \sigma^2(x) = \log c^2 + 2\rho \log |x|,$$

we have

$$\frac{d \log \sigma(x)}{d \log |x|} = \rho.$$

This form of diffusion function is hence called constant-elasticity diffusion.

### 2.3.2 Useful Diffusion Models

In this section we introduce a number of useful parametric diffusion models.

**Brownian Motion with Drift**

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

$$X_t = X_0 + \mu t + \sigma W_t$$

Transition distribution: $X_{t+\Delta}|X_t = x \sim N(x + \mu\Delta, \sigma^2\Delta)$.

**Geometric Brownian Motion**

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

By Ito's formula,

$$d \log X_t = \frac{1}{X_t} dX_t - \frac{1}{2X_t^2} d[X]_t$$

Since $d[X]_t = \sigma^2 X_t^2 dt$,

$$d \log X_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t.$$

**Ornstein-Uhlenbeck Process** The Ornstein-Uhlenbeck process has the following SDE representation,

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

To derive the transition distribution, we define $Y_t = X_t - \mu$. Then

$$dY_t = -\kappa Y_t dt + \sigma dW_t.$$

$$\begin{aligned} d(\exp(\kappa t)Y_t) &= \kappa \exp(\kappa t)Y_t dt + \exp(\kappa t)dY_t \\ &= \kappa \exp(\kappa t)Y_t dt + \exp(\kappa t)(-\kappa Y_t dt + \sigma dW_t) \\ &= \sigma \exp(\kappa t)dW_t \end{aligned}$$

So

$$\exp(\kappa t)Y_t = Y_0 + \sigma \int_0^t \exp(\kappa s)dW_s,$$

ie,

$$Y_t = \exp(-\kappa t)Y_0 + \sigma \int_0^t \exp(-\kappa(t-s))dW_s.$$

So

$$X_t = \mu + \exp(-\kappa t)(X_0 - \mu) + \sigma \int_0^t \exp(-\kappa(t-s))dW_s.$$

Given $Y_0 = y$, what is the distribution of $Y_t$?

$$\begin{aligned} Y_t &= \exp(-\kappa t)Y_0 + \sigma \int_0^t \exp(-\kappa(t-s))dW_s \\ &\sim N(\exp(-\kappa t)y, \sigma^2 \frac{1-\exp(-2\kappa t)}{2\kappa}). \end{aligned}$$

Let $t \to \infty$,

$$Y_t \sim N(0, \frac{\sigma^2}{2\kappa}).$$

So if $Y_0 \sim N(0, \frac{\sigma^2}{2\kappa})$, $Y_t$ is stationary and $Y_t \sim N(0, \frac{\sigma^2}{2\kappa})$.

**Feller's Squared-Root Process** The Feller's squared-root process has the following representation,

$$dX_t = \kappa(\mu - X_t)dt + \sigma\sqrt{X_t}dW_t.$$

If $\frac{2\kappa\mu}{\sigma^2} \geq 1$, then $X_t \in [0, \infty)$. Like Ornstein-Uhlenbeck Process, Feller's SR Process is also a stationary process. And it's transition distribution is non-central Chi-square, and marginal distribution gamma. It is used by Cox, Ingersol, and Ross (CIR) to model interest rates.

### 2.3.3 Discrete-Time Approximations

For many diffusions, the transition distributions are very complicated. Often they do not have closed-form density functions. It is thus desirable to have approximations of transition distributions. The approximation error shall go to zero as intervals of discrete-time observations go to zero.

**Euler Approximation**

Suppose that we observe a discrete-time sequence, $X_\Delta$, $X_{2\Delta}$, ..., $X_{n\Delta}$. The interval between observations is $\Delta$. We seek an approximation of the conditional distribution of $X_{n\Delta}|X_{(n-1)\Delta}$. Let $\Delta$ be small. We have

$$\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} &= \int_{(i-1)\Delta}^{i\Delta} \mu(X_t)dt + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t)dW_t \\
&= \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
&\quad + \int_{(i-1)\Delta}^{i\Delta} [\mu(X_t) - \mu(X_{(i-1)\Delta})]dt + \int_{(i-1)\Delta}^{i\Delta} [\sigma(X_t) - \sigma(X_{(i-1)\Delta})]dW_t \\
&\approx \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})
\end{aligned}$$

This is called Euler Approximation. Under this approximation, $X_{i\Delta}|X_{(i-1)\Delta} = x \sim N(\Delta\mu(x), \Delta\sigma^2(x))$.

**Milstein Approximation**

We have something better. Consider

$$\begin{aligned}
&\mu(X_t) - \mu(X_{(i-1)\Delta}) \\
= &\int_{(i-1)\Delta}^{t} \mu'(X_s)dX_s + \frac{1}{2}\int_{(i-1)\Delta}^{t} \mu''(X_s)d[X]_s \\
= &\int_{(i-1)\Delta}^{t} (\mu'(X_s)\mu(X_s) + \frac{1}{2}\mu''(X_s)\sigma^2(X_s))ds + \int_{(i-1)\Delta}^{t} \mu'(X_s)\sigma(X_s)dW_s,
\end{aligned}$$

and

$$\begin{aligned}
&\sigma(X_t) - \sigma(X_{(i-1)\Delta}) \\
= &\int_{(i-1)\Delta}^{t} \sigma'(X_s)dX_s + \frac{1}{2}\int_{(i-1)\Delta}^{t} \sigma''(X_s)d[X]_s \\
= &\int_{(i-1)\Delta}^{t} (\sigma'(X_s)\mu(X_s) + \frac{1}{2}\sigma''(X_s)\sigma^2(X_s))ds + \int_{(i-1)\Delta}^{t} \sigma'(X_s)\sigma(X_s)dW_s.
\end{aligned}$$

And

$$\begin{aligned}
\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} (\mu\mu' + \frac{\sigma^2\mu''}{2}) &= O(\Delta^2) \\
\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} (\sigma\mu')dW_s dt &= O(\Delta^{3/2}) \\
\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} (\sigma\sigma')(X_s)dW_s dW_t &= O(\Delta) \qquad (*)
\end{aligned}$$

So if we want to have an accuracy of $O(\Delta)$, the last term cannot be ignored. To have a better approximation,

$$\begin{aligned}
(*) &= \sigma\sigma'(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} dW_s dW_t \\
&\quad + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} [(\sigma\sigma')(X_s) - (\sigma\sigma')(X_{(i-1)\Delta})] dW_s dW_t \\
&= \sigma\sigma'(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (W_t - W_{(i-1)\Delta}) dW_t + o(\Delta) \\
&= \frac{1}{2}[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta]\sigma\sigma'(X_{(i-1)\Delta}) + o(\Delta).
\end{aligned}$$

The last equality is obtained by applying Ito's formula,

$$d(\frac{1}{2}(W_t - W_{(i-1)\Delta})^2) = (W_t - W_{(i-1)\Delta})dW_t + \frac{1}{2}dt.$$

So here's Milstein Approximation,

$$\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} &= \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
&\quad + \frac{1}{2}[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta]\sigma\sigma'(X_{(i-1)\Delta}) + o(\Delta).
\end{aligned}$$

# Chapter 3

# Arbitrage Pricing in Continuous Time

## 3.1 Basic Setup

Consider a financial market with an interest-paying *money account* and a *stock*.

**Money Account** The interest rate may be fixed, time-varying, or even state-contingent. Let $M_0$ be the initial deposit and $M_t$ be the cash value of the account at time $t$. We may represent $M_t$ in stochastic differential equation form as follows,

- $r$ fixed
$$dM_t = rM_0e^{rt}dt = rM_tdt, \quad M_t = M_0e^{rt}.$$
- $r$ time-varying, $r_t$
$$dM_t = r_tM_tdt, \quad M_t = M_0e^{\int_0^t r_s ds}.$$
- $r$ state-contingent, $r(X_t)$
$$dM_t = r(X_t)M_tdt, \quad M_t = M_0e^{\int_0^t r(X_s)ds}.$$

**Stock** Let $S_t$ be stock price at time $t$ that follows an Ito process,

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t.$$

The use of continuous diffusion process implicitly asserts that there is no "surprise" in the market.

One special case is when $\mu(S_t) = 0$ and $\sigma(S_t) = 1$. $S_t$ then becomes a Brownian motion. Louis Bachelier, the pioneer of mathematical finance, used Brownian motion in describing the fluctuation in financial markets. Another well-known special case is the geometric Brownian motion, which is used by Black and Scholes (1973) to price European options.

If we assume $S_t > 0$ a.s., we may also represent $S_t$ in the geometric form,

$$dS_t = \mu(S_t)S_t dt + \sigma(S_t)S_t dW_t.$$

Note that, throughout the text, we use homogeneous diffusions to model stock prices. More generally, at the cost of technical complication, we may also use heterogenous diffusions.

**Multivariate Case** $S_t$ can be a $N \times 1$ price vector describing the prices of $N$ securities. Accordingly, $W$ may be a $d \times 1$ vector of independent Brownian motions, each of which represents a source of new information or innovation. In such case, $\mu \in \mathbb{R}^N$, $\sigma \in \mathbb{R}^{N \times d}$.

**Portfolio/trading strategy** Portfolio or trading strategy $(h_t)$ is an adapted vector process:

$$h_t = \begin{pmatrix} a_t \\ b_t \end{pmatrix},$$

where $a_t$ is the holding of money account, and $b_t$ the holding of stocks.

**Trading Gain**

$$G_t = \int_0^t b_s dS_s.$$

We usually impose the following integrability condition:

$$\int_0^t b_s^2 ds < \infty \text{a.s. } \forall t$$

If $S_t$ is a martingale (e.g., BM), then we know $G_t$ is also a martingale.

$G_t = \int_0^t b_s dS_s$ is often called "gains process". To see this, imagine an investor who makes decisions in discrete time: $0 = t_0 < t_1 < \cdots < t_n = T$. Let $b_{t_i}$ be the number of stocks the investor holds over the period $[t_i, t_{i+1})$. Then the gains process is described by the following stochastic difference equation:

$$G_0 = 0, \; G(t_{i+1}) - G(t_i) = b_{t_i}(S(t_{i+1}) - S(t_i)).$$

Or in summation form,

$$G(t_{i+1}) = G(0) + \sum_{j=0}^{i} b_{t_j}(S(t_{j+1}) - S(t_j)).$$

Note that as in the definition of Ito integral, $b_{t_i}$ must be $\mathcal{F}_{t_i}$-measurable, meaning that the investor cannot anticipate the future (exclusion of inside trading).

Let $X_t = (M_t, S_t)'$, and let $H_t$ be the value of the portfolio $(h_t)$. Then

$$H_t = h_t \cdot X_t = a_t M_t + b_t \cdot S_t.$$

**Definition 3.1.1 (Self-financing)** *$(h_t)$ is self-financing iff*

$$dH_t = h_t \cdot dX_t = a_t dM_t + b_t \cdot dS_t.$$

**Definition 3.1.2 (Arbitrage)** *Let $(h_t)$ be a self-financing portfolio and $(H_t)$ be its value, an arbitrage portfolio is one such that*

$$H_0 = 0, \quad \textit{and} \quad H_T > 0 \ \textit{a.s.}$$

**Lemma 3.1.3** *If there is no arbitrage opportunities, and if $(h_t)$ is self-financing and $dH_t = v_t H_t dt$, then $v_t = r_t$, the risk-free short rate.*

In other words, there is only one risk-free short rate.

**Numeraire** A numeraire is a strictly positive Ito process used for the "units" of pricing. If there is a riskfree rate $r_t$, the typical numeraire is the reciprocal of the price of riskfree zero-coupon bond, $Y_t = M_t^{-1} = \exp(-\int_0^t r_s ds)$. We denote the numeraire-deflated price process of $X_t$ by $Y_t$ as $X^Y$, $X_t^Y = X_t Y_t$. For example, if $X_t = (M_t, S_t)'$ and $Y_t = M_t^{-1}$, then $X_t^Y = (1, S_t/M_t)$.

**Theorem 3.1.4 (Numeraire Invariance Theorem)** *Suppose $Y$ is a numeraire. Then a trading strategy $(h_t)$ is self-financing w.r.t. $X$ iff $(h_t)$ is self-financing w.r.t. $X^Y$.*

**Proof** Let $H_t = h_t X_t$, and $H_t^Y = H_t Y_t$. If $dH_t = h_t dX_t$, then

$$\begin{aligned} dH_t^Y &= Y_t dH_t + H_t dY_t + d[H,Y]_t \\ &= Y_t h_t dX_t + (h_t X_t) dY_t + h_t d[X,Y]_t \\ &= h_t (Y_t dX_t + X_t dY_t + d[X,Y]_t) \\ &= h_t dX_t^Y. \end{aligned}$$

So $(h_t)$ is self-financing w.r.t. $X^Y$. The reverse is also true.

It follows that $h$ is an arbitrage w.r.t. $X$ iff it is an arbitrage w.r.t. $X^Y$. All this says that renormalization of security prices by a numeraire does not have economic effects.

## 3.2 The Black-Scholes Model

In a market of $(M_t, S_t)$ we price an European call option using no-arbitrage argument. We specify

$$\begin{cases} dM_t = rM_t dt \\ dS_t = \mu S_t dt + \sigma S_t dW_t \end{cases}$$

Recall that $C_T = \max(S_T - K, 0)$. In general, we may use the same argument to price any European option with the final payoff $g(S_T)$, where $g$ is a known function.

Let $C_t = F(S_t, t)$. We assume $F \in C^{2,1}(\mathbb{R} \times [0,T))$, ie, $F_1$, $F_2$, and $F_{11}$ exist and are continuous. We have

$$\begin{aligned} dC_t &= F_2(S_t,t)dt + F_1(S_t,t)dS_t + \frac{1}{2}F_{11}(S_t,t)d[S]_t \\ &= (F_2(S_t,t) + \mu S_t F_1(S_t,t) + \frac{1}{2}F_{11}(S_t,t)\sigma^2 S_t^2)dt + \sigma S_t F_1(S_t,t)dW_t \end{aligned}$$

The pricing strategy is to replicate $C_t$ using a self-financing portfolio of $M_t$ and $S_t$. The price of the replication portfolio must then be the price of the option, if no arbitrage is allowed.

We have $H_t = a_t M_t + b_t S_t$. Since $h_t$ is self-financing,

$$\begin{aligned} dH_t &= a_t dM_t + b_t dS_t \\ &= a_t r M_t dt + b_t(\mu S_t dt + \sigma S_t dW_t) \\ &= (a_t r M_t + b_t \mu S_t)dt + \sigma b_t S_t dW_t \end{aligned}$$

By the unique decomposition property of diffusion processes, since $C_t = H_t$ a.s. for all $t$, we have

$$\begin{aligned} b_t &= F_1(S_t, t) \\ a_t &= \frac{F(S_t, t) - F_1(S_t, t) S_t}{M_t}, \end{aligned}$$

and by the equality of drift terms between $C_t$ and $H_t$,

$$F_2(S_t, t) + r S_t F_1(S_t, t) + \frac{1}{2}\sigma^2 S_t^2 F_{11}(S_t, t) - r F(S_t, t) = 0. \tag{3.1}$$

For (3.1) to hold, $F$ must be the solution to the following partial differential equation (PDE):

$$F_2(x, t) + r x F_1(x, t) + \frac{1}{2}\sigma^2 x^2 F_{11}(x, t) - r F(x, t) = 0 \tag{3.2}$$

with the boundary condition

$$F(x, T) = \max(x - K, 0). \tag{3.3}$$

We can check that the Black-Scholes Option Pricing Formula solves the PDE (3.2) and (3.3). The formula is as follows,

$$F(x, t) = x\Phi(z) - \exp(-r(T - t)) K \Phi(z - \sigma\sqrt{T - t}), \tag{3.4}$$

with

$$z = \frac{\log(x/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}, \tag{3.5}$$

where $\Phi$ is the cdf of standard normal distribution.

A byproduct of this derivation is a popular dynamic hedging strategy called "delta hedging". Consider a bank that has sold an European call option and now it wants to hedge its position. All it has to do is to maintain opposite positions of $h_t = (a_t, b_t)$. In fact $b_t = F_1(S_t, t)$ is called the "delta" of the option in practice.

## The General Case

Now assume that $(M_t, S_t)$ are such that

$$\begin{aligned} dM_t &= r_t M_t dt \\ dS_t &= \mu(S_t) dt + \sigma(S_t) dW_t \end{aligned}$$

We may mimic the argument in previous section and show that under the no-arbitrage condition, the price process for an European call option $F(S_t, t)$ must satisfy:

$$F_2(x,t) + r_t x F_1(x,t) + \frac{1}{2}\sigma(x)^2 F_{11}(x,t) - r_t F(x,t) = 0 \tag{3.6}$$

with the boundary condition

$$F(x,T) = \max(x-K, 0).$$

Note that for general European option with payoff $g(S_T)$, the price process still satisfy (3.6) with following boundary condition

$$F(x,T) = g(x). \tag{3.7}$$

## 3.3 The Feynman-Kac Solution

### Constant Riskfree Rate

Consider the following boundary value problem:

$$F_2(x,t) + rxF_1(x,t) + \frac{1}{2}\sigma^2(x)F_{11}(x,t) - rF(x,t) = 0, \tag{3.8}$$

with

$$F(x,T) = g(x).$$

This problem differs from (3.6) only in the form of $r$, which is a constant here.

Construct an Ito process $Z$ such that $Z_t = x$

$$dZ_s = rZ_s ds + \sigma(Z_s)dW_s, \ \ s > t.$$

By Ito's formula,

$$\begin{aligned}
& d(e^{-rs}F(Z_s,s)) \\
= \ & [-re^{-rs}F(Z_s,s) + e^{-rs}F_2(Z_s,s)]ds + e^{-rs}F_1(Z_s,s)dZ_s + \frac{1}{2}F_{11}(Z_s,s)d[Z]_s \\
= \ & e^{-rs}[-rF + F_2 + rZ_sF_1 + \frac{1}{2}\sigma^2(Z_s)F_{11}]ds + e^{-rs}\sigma(Z_s)F_1(Z_s,s)dW_s
\end{aligned}$$

If $F(x,t)$ satisfies (3.8), then the term in bracket is zero. Hence $e^{-rs}F(Z_s,s)$ is martingale. So

$$e^{-rT}F(Z_T,T) = e^{-rt}F(Z_t,t) + \int_t^T e^{-rs}\sigma(Z_s)F_1(Z_s,s)dW_s.$$

Taking conditional expectation given $Z_t = x$ gives:

$$\mathbb{E}(e^{-rT}F(Z_T, T)|Z_t = x) \equiv \mathbb{E}(e^{-rT}g(Z_T)|Z_t = x) = e^{-rt}F(x, t).$$

Hence

$$F(x, t) = \mathbb{E}(e^{-r(T-t)}g(Z_T)|Z_t = x).$$

The message is that we can solve certain PDE's by calculating a conditional expectation of an imagined random variable $g(Z_T)$.

## Stochastic Riskfree Rate

Now we work to solve (3.6) which is reproduced below,

$$F_2(x, t) + r_t x F_1(x, t) + \frac{1}{2}\sigma(x)^2 F_{11}(x, t) - r(x)F(x, t) = 0$$

with

$$F(x, T) = g(x).$$

It can be shown that if we construct $Z$ such that $Z_t = x$ and

$$dZ_s = r_s Z_s ds + \sigma(Z_s)dW_s, \ \ s > t, \tag{3.9}$$

then

$$F(x, t) = \mathbb{E}\left\{\exp\left[-\int_t^T r_s ds\right] g(Z_T)|Z_t = x\right\}. \tag{3.10}$$

In particular, the Black-Scholes option price is given by

$$F(x, t) = \mathbb{E}\left\{e^{-r(T-t)}g(Z_T)|Z_t = x\right\}, \tag{3.11}$$

where $Z$ is such that $Z_t = x$ and

$$dZ_s = rZ_s ds + \sigma Z_s dW_s, \ \ s > t. \tag{3.12}$$

The expectation in the Feynman-Kac solution (3.10) is taken with respect to objective probability on an imagined r.v. $g(Z_T)$. We can also represent the solution as an expectation taken with respect to an imagined probability (risk-neutral probability) on a "real" r.v., for example, $g(S_T)$. This will be explored in the next section.

## Calculation of Black-Scholes Formula

Now we derive Black-Scholes formula from (3.11) and (3.12). We have

$$d\log Z_s = (r - \sigma^2/2)dt + \sigma dW_t,$$

which implies

$$\log Z_T - \log Z_t = (r - \sigma^2/2)(T - t) + \sigma(W_T - W_t).$$

This is,

$$Z_T = Z_t e^{(r-\sigma^2/2)(T-t)+\sigma(W_T - W_t)}.$$

Hence

$$Z_T|_{Z_t=x} =^d x e^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}Z}, \; Z \sim N(0,1).$$

So

$$\begin{aligned} F(x,t) &= \mathbb{E}\left\{e^{-r(T-t)}g(Z_T)|Z_t\right\} \\ &= e^{-r(T-t)}\int_{-\infty}^{\infty} \max(xe^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z} - K, 0)\phi(z)dz \end{aligned}$$

Some calculations yield the Black-Scholes formula in (3.4).

## Feynman-Kac in Multivariate Case

Now we consider the market of a money account and multiple stocks containing $d$ dimensions of risk. We have

$$\begin{aligned} dM_t &= r_t M_t dt \\ dS_t &= \mu(S_t)dt + \sigma(S_t)dW_t, \end{aligned}$$

where $W_t \in \mathbb{R}^d$, $S_t \in \mathbb{R}^N$, $\mu : \mathbb{R}^N \times [0,\infty) \to \mathbb{R}^N$, $\sigma : \mathbb{R}^N \times [0,\infty) \to \mathbb{R}^{N\times d}$.

The option price $F(x,t) : \mathbb{R}^N \times [0,\infty) \to \mathbb{R}$ solves the following boundary value pde:

$$F_1(x,t)r_t x + F_2(x,t) + \frac{1}{2}\mathrm{tr}\left[(\sigma\sigma')(x)F_{11}(x,t)\right] - r_t F(x,t) = 0 \tag{3.13}$$

with $F(x,T) = g(x)$. The solution of (3.13) is of the same form as (3.10).

## 3.4 Risk-Neutral Pricing

We have shown that in the market of $M_t$ and $S_t$ such that

$$\begin{aligned} dM_t &= r_t M_t dt \\ dS_t &= \mu(S_t)dt + \sigma(S_t)dW_t \end{aligned}$$

The price of a general derivative can be represented as

$$F(x,t) = \mathbb{E}\left\{\exp\left[-\int_t^T r_s ds\right] g(Z_T)|Z_t = x\right\},$$

where $Z$ is an imagined Ito process such that $Z_s = x$ for $s \le t$ and

$$dZ_s = r_s Z_s ds + \sigma(Z_s)dW_s, \ \ s > t.$$

In this section we show that there exists a probability measure $\tilde{\mathbb{P}}$ and a $\tilde{\mathbb{P}}$-BM $\tilde{W}$ such that

$$dS_t = r_t S_t dt + \sigma(S_t)d\tilde{W}_t.$$

So the price function of a general European option ($C_T = g(S_T)$) can be written as

$$F(x,t) = \tilde{\mathbb{E}}\left\{\exp\left[-\int_t^T r_s ds\right] g(S_T)|S_t = x\right\}.$$

Put it differently,

$$\frac{C_t}{M_t} = \tilde{\mathbb{E}}_t\left(\frac{C_T}{M_T}\right),$$

where $M_t$ is the money account and acts as a numeraire. In other words, $(C_t/M_t)$ is a martingale under $\tilde{\mathbb{P}}$. So $\tilde{\mathbb{P}}$ is sometimes called "martingale probability measure". And since the price of an asset is the expectation of its payoff $C_T/M_T$ taken with respect to a probability measure, we call this measure "risk-neutral" measure or probability.

### Change of Measure

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a nonnegative r.v. $\xi$ that satisfies $\mathbb{E}\xi = 1$, then we may define a new probability measure as follows,

$$\tilde{\mathbb{P}}(A) = \int_A \xi(\omega)dP(\omega) \ \text{ for all } A \in \mathcal{F}.$$

We may check that $\tilde{\mathbb{P}}(\omega) > 0$ for all $\omega$ and $\tilde{\mathbb{P}}(\Omega) = 1$.

If $\xi > 0$ a.s., then $\tilde{\mathbb{P}}$ is called an "equivalent probability measure" of $\mathbb{P}$, ie, for any set $A$, $\tilde{\mathbb{P}}(A) = 0$ iff $\mathbb{P}(A) = 0$. $\xi$ is called the Radon-Nikodym density of $\tilde{\mathbb{P}}$ w.r.t. $\mathbb{P}$. In differential form, we may write

$$d\tilde{\mathbb{P}} = \xi d\mathbb{P}.$$

For any r.v. $X$, it is easy to show that

$$\tilde{\mathbb{E}}X = \mathbb{E}\xi X,$$

and that

$$\mathbb{E}X = \tilde{\mathbb{E}}\frac{X}{\xi}.$$

**Example** Let $X$ be standard normal, $\xi = \exp(\lambda X - \lambda^2/2)$, and $\tilde{\mathbb{P}}$ be defined as above. For any function $f$, we have

$$\begin{aligned}
\tilde{\mathbb{E}}f(X) &= \mathbb{E}\xi f(X) \\
&= \int \exp(\lambda x - \lambda^2/2)\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)f(x)dx \\
&= \frac{1}{\sqrt{2\pi}}\int \exp(-(x-\lambda)^2/2)f(x)dx.
\end{aligned}$$

So under $\tilde{\mathbb{P}}$, $X \sim N(\lambda, 1)$. In other words, this particular change of measure shifts $X$ by a constant, without changing the variance.

## Density Process

**Definition 3.4.1 (Martingale Equivalent Measure)** *$\tilde{\mathbb{P}}$ is a martingale equivalent measure of $\mathbb{P}$ for a process $(X_t)$ if $(X_t)$ is a martingale under $\tilde{\mathbb{P}}$.*

If the price process $X$ admits an equivalent martingale measure, then there is no arbitrage. To see this, note that for any admissible trading strategy, $\tilde{\mathbb{E}}(\int_0^t h_s dX_s) = 0$. Hence, the self-financing condition $h_t X_t = h_0 X_0 + \int_0^t h_s dX_s$ implies

$$h_0 X_0 = \tilde{\mathbb{E}}(h_t X_t).$$

Thus, if $h_t X_t > 0$, then $h_0 X_0 > 0$.

**Density Process.** Let $\xi > 0$ a.s. and $\mathbb{E}\xi = 1$ and define $\tilde{\mathbb{P}}$ as above. We define $\xi_t = \mathbb{E}(\xi|\mathcal{F}_t)$. We call $(\xi_t)$ the density process for $\tilde{\mathbb{P}}$ w.r.t. $\mathbb{P}$. Obviously $(\xi_t, \mathcal{F}_t)$ is a positive martingale. If $X_t$ is $\mathcal{F}_t$-measurable, then

$$\begin{aligned}
\mathbb{E}(\xi_t X_t) &= \mathbb{E}(\xi X_t) \\
\tilde{\mathbb{E}} X_t &= \mathbb{E}\xi_t X_t \\
\tilde{\mathbb{E}}\xi_t^{-1} X_t &= \mathbb{E} X_t.
\end{aligned}$$

The first statement is an immediate consequence of the iterative expectation, and the second statement is due to

$$\tilde{\mathbb{E}} X_t = \mathbb{E}\xi X_t = \mathbb{E}[\mathbb{E}\xi X_t|\mathcal{F}_t] = \mathbb{E}[\mathbb{E}[\xi|\mathcal{F}_t]X_t] = \mathbb{E}\xi_t X_t.$$

The third statement is similarly established. We also have

**Bayes Rule.**

$$\tilde{\mathbb{E}}_s W = \xi_s^{-1}\mathbb{E}_s(\xi_t W), \ s < t, \tag{3.14}$$

where $W$ is $\mathcal{F}_t$-measurable and $\tilde{\mathbb{E}}|W| < \infty$.

**Proof** For any $A \in \mathcal{F}_s$, since $d\mathbb{P} = \xi^{-1} d\tilde{\mathbb{P}}$,

$$\begin{aligned}
\tilde{\mathbb{E}}\left[I_A \xi_s^{-1}\mathbb{E}_s(\xi_t W)\right] &= \mathbb{E}\left[I_A \mathbb{E}_s(\xi_t W)\right] \\
&= \mathbb{E}\left[I_A \xi_t W\right] \\
&= \tilde{\mathbb{E}}\left[I_A W\right] \\
&= \tilde{\mathbb{E}}\left[I_A \tilde{\mathbb{E}}_s W\right].
\end{aligned}$$

Since it is true for every $A$, the proof is complete.

## Girsanov Theorem

Since

$$\tilde{\mathbb{E}}(\tilde{M}_t|\mathcal{F}_s) = \xi_s^{-1}\mathbb{E}(\xi_t \tilde{M}_t|\mathcal{F}_s) = \tilde{M}_s,$$

we have the following lemma,

**Lemma:** If $(\tilde{M}_t \xi_t)$ is $\mathbb{P}$-martingale, then $(\tilde{M}_t)$ is $\tilde{\mathbb{P}}$-martingale.

**Girsanov Theorem.** If $(M_t)$ is $\mathbb{P}$-martingale, and $(\tilde{M}_t)$ is defined as $d\tilde{M}_t = dM_t - \xi_t^{-1} d[M, \xi]_t$, then $(\tilde{M}_t)$ is $\tilde{\mathbb{P}}$-martingale. In other words, $\tilde{\mathbb{P}}$ is the martingale equivalent measure for $\tilde{M}$.

**Proof** It suffices to show that $(\tilde{M}_t \xi_t)$ is $\mathbb{P}$-martingale.

$$\begin{aligned} d(\tilde{M}_t \xi_t) &= \tilde{M}_t d\xi_t + \xi_t d\tilde{M}_t + d[\tilde{M}, \xi]_t \\ &= \tilde{M}_t d\xi_t + \xi_t dM_t. \end{aligned}$$

Note that $[\tilde{M}]_t = [M]_t$, since $\tilde{M} = M + A$, where $A$ is of bounded variation.

If $M_t = W_t$, then $d\tilde{W}_t = dW_t - \xi_t^{-1} d[W, \xi]_t$ is $\tilde{\mathbb{P}}$-BM.

**A Light Version of Girsanov Theorem.** Let $\eta_t$ be adapted to $\mathcal{F}_t$ and let $\xi_t = \exp\left( - \int_0^t \eta_s dW_s - \frac{1}{2} \int_0^t \eta_s^2 ds \right)$. If $W_t$ is $\mathbb{P}$-BM, then $\tilde{W}_t = W_t + \int_0^t \eta_s ds$ is BM under $\tilde{\mathbb{P}}$.

**Proof** Let $L_t = - \int_0^t \eta_s dW_s$, then $\xi_t = \exp\left( L_t - \frac{1}{2}[L]_t \right)$. Then we have

$$d \log \xi_t = d(L_t - \frac{1}{2}[L]_t) = dL_t - \frac{1}{2} d[L]_t,$$

and

$$d \log \xi_t = \xi_t^{-1} d\xi_t - \frac{1}{2} \frac{1}{\xi_t^2} d[\xi]_t.$$

By unique decomposition of semimartingale, we have

$$dL_t = \xi_t^{-1} d\xi_t.$$

Hence

$$\xi_t^{-1} d[W, \xi]_t = d[W, L]_t = -\eta_t dt.$$

To show that $\tilde{W}_t$ is $\tilde{\mathbb{P}}$-BM, we note that $[\tilde{W}]_t = t$.

Note that in the above notations, $\xi_t$ is called the stochastic exponential of $L_t$.

## Black-Scholes Again

**Black-Scholes Using Girsanov** We may assume a riskfree rate $r_t$ and the stock price $S_t$ satisfies:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t,$$

where $\mu_t$ and $\sigma_t$ are adapted to $\mathcal{F}_t$. We may show that

$$d\left(\frac{S_t}{M_t}\right) = \sigma_t \frac{S_t}{M_t}\left(\frac{\mu_t - r_t}{\sigma_t}dt + dW_t\right).$$

If we define

$$\eta_t = \frac{\mu_t - r_t}{\sigma_t}$$

$$L_t = -\int_0^t \eta_s dW_s,$$

and similarly

$$\xi_t = \exp\left(L_t - \frac{1}{2}[L]_t\right)$$

and $\tilde{\mathbb{P}}$ such that $d\tilde{\mathbb{P}} = \xi d\mathbb{P}$.

Then we have

$$d\left(\frac{S_t}{M_t}\right) = \sigma_t \frac{S_t}{M_t} d\tilde{W}_t.$$

Or equivalently,

$$dS_t = r_t S_t dt + \sigma_t S_t d\tilde{W}_t.$$

Then the price of an European call option would be

$$C_t = \tilde{\mathbb{E}}\left\{\exp\left(-\int_t^T r_s ds\right)\max(S_T - K, 0)|\mathcal{F}_t\right\} = M_t \tilde{\mathbb{E}}\left\{\frac{C_T}{M_T}|\mathcal{F}_t\right\},$$

where $\tilde{\mathbb{E}}$ is taken w.r.t. $\tilde{\mathbb{P}}$.

Note that the process $\eta_t$ satisfies

$$\mu_t - r_t = \eta_t \sigma_t.$$

$\eta_t$ measures the excess return per unit of risk the market offers. For this reason $\eta_t$ is called the "market price of risk" process.

## Forwards and Futures

A forward contract is an agreement to pay a specified delivery price $K$ for an asset at a delivery date $T$. Suppose the asset price process is $S_t$. At time $T$, the value of the contract is $S_T - K$. At the time of reaching an agreement, say $t$, the value of the contract must be zero,

$$\tilde{\mathbb{E}} M_t / M_T (S_T - K) = \tilde{\mathbb{E}} \exp\left(-\int_t^T r_s ds\right)(S_T - K) = S_t - P(t,T)K = 0,$$

where $P(t,T) = \tilde{\mathbb{E}}\left(\exp\left(-\int_t^T r_s ds\right)|\mathcal{F}_t\right)$. Thus $K$ must be $S_t/P(t,T)$. $K$ is called the forward price of the asset. As it is a function of both $t$ and $T$, we denote it as $\mathrm{Fo}(t,T)$.

After a forward contract is signed on $t$, the value of this agreement will most likely diverge from zero, often substantially. Let $u$ be such that $t < u < T$. For the party on the long position, who receives $S_T$ and pays $Fo(t,T)$ at time $T$, the value of the agreement at time $u$ is

$$\begin{aligned} V_{t,u} &= M_u \tilde{\mathbb{E}}\left(M_T^{-1}\left(S_T - \frac{S_t}{P(t,T)}\right)|\mathcal{F}_u\right) \\ &= S_u - S_t\frac{P(u,T)}{P(t,T)}. \end{aligned}$$

If the riskfree rate is a constant $r$, then

$$V_{t,u} = S_u - \exp(r(u-t))S_t.$$

If the asset price rises more rapidly than the money account, then the long (short) position has a positive (negative) value. If the growth rate is less than the riskfree rate, then the long (short) position has a negative (positive) value. Whichever happens, one party will have an incentive to default.

A futures contract alleviates the risk of default by margin requirement (initial margin, marking to margin) and by trading in an exchange market. The futures price of an underlying asset whose spot price process is $S_t$ must be

$$\mathrm{Fu}(t,T) = \tilde{\mathbb{E}}(S_T|\mathcal{F}_t).$$

In other words, $\mathrm{Fu}(t,T)$ must be a martingale process under $\tilde{\mathbb{P}}$ with terminal price $\mathrm{Fu}(T,T) = S_T$. It is obvious that $\mathrm{Fu}(t,T)$ is uniquely determined given an $S_T$.

To see that any futures price is a $\tilde{\mathbb{P}}$-martingale, consider a partition of the life span of a futures contract, $0 = t_0 < t_1 < \cdots < t_n = T$. Each interval $[t_k, t_{k+1})$ represents a “day”. A long position in the futures contract is an agreement to receive as a cash flow the changes in the futures price, $\mathrm{Fu}(t_{k+1},T) - \mathrm{Fu}(t_k,T)$, during the time the position is held. A short position receives the opposite. Suppose the riskfree rate $r_t$ is constant within each day. $M_{t_{k+1}}$ is $\mathcal{F}_{t_k}$-measurable, since

$$M_{t_{k+1}} = \exp\left(-\int_0^{t_{k+1}} r_s ds\right) = \exp\left(-\sum_{k=0}^{k} r_{t_k}(t_{k+1}-t_k)\right).$$

For a futures agreement to be reached, any future cash flow must have a current value of zero. That is, for all $k$,

$$M_{t_k}\tilde{\mathbb{E}}\left(M_{t_{k+1}}^{-1}\left(\mathrm{Fu}(t_{k+1},T) - \mathrm{Fu}(t_k,T)\right)|\mathcal{F}_{t_k}\right) = 0.$$

Since $M_{t_{k+1}}$ is $\mathcal{F}_{t_k}$-measurable, we have

$$\frac{M_{t_k}}{M_{t_{k+1}}}\tilde{\mathbb{E}}\left(\mathrm{Fu}(t_{k+1},T)-\mathrm{Fu}(t_k,T)|\mathcal{F}_{t_k}\right)=0.$$

Hence $\mathrm{Fu}(t_k,T)$ must be a martingale sequence.

**Forwards-Futures Spread** The difference between forward and futures prices,

$$D_t = S_t/P(t,T) - \tilde{\mathbb{E}}(S_T|\mathcal{F}_t),$$

is called the forward-futures spread. It is obvious that $D_t \to 0$ as $t \to T$.

Since $S_t = \tilde{\mathbb{E}}_t\left(\frac{M_t}{M_T}S_T\right)$ and $P(t,T) = \tilde{\mathbb{E}}_t\frac{M_t}{M_T}$, we have

$$D_t = \frac{1}{P(t,T)}\left(\tilde{\mathbb{E}}_t\left(\frac{M_t}{M_T}S_T\right) - \tilde{\mathbb{E}}_t\frac{M_t}{M_T}\tilde{\mathbb{E}}_tS_T\right) = \frac{1}{P(t,T)}\widetilde{\mathrm{cov}}\left(\frac{M_t}{M_T}, S_T|\mathcal{F}_t\right).$$

If $r_t$ is a constant, $D_t = 0$. If $P(t,T)$ is positively correlated with $S_T$, which means that a higher $S_T$ goes together with a lower interest rate, then the forward price is higher than the futures price.

## Pricing of Cash Flow

Suppose an asset pays $D_t$ between time 0 and $t$. Then a long position of the asset gives us a gain process that satisfies

$$dG_t = dD_t + r_tG_td_t,$$

or

$$d(G_t/M_t) = 1/M_tdD_t.$$

The risk-neutral price at time $t$ of the cash flow between $t$ and $T$ is thus,

$$M_t\tilde{\mathbb{E}}(G_T/M_T|\mathcal{F}_t) = M_t\tilde{\mathbb{E}}\left(\int_t^T 1/M_sdD_s|\mathcal{F}_t\right).$$

The cash flow may be negative, in which case $D_t$ is decreasing. The cash flow is most likely discrete, ie,

$$D_t = \sum_{i=1}^n d_iI_{[0,t]}(t_i),$$

where $0 < t_1 < t_2 < \cdots < t_n \le T$ and $d_i$ is random payment at time $t_i$. In this case, the risk neutral price at time $t$ is given by

$$\sum_{i=1}^{n} I_{[t,T]}(t_i) \left( M_t \tilde{\mathbb{E}}(M_{t_i} d_i | \mathcal{F}_t) \right).$$

If $d_i$ is deterministic, then the above formula reduces to the pricing formula for bond with fixed coupons.

## The Multivariate Case

We first state the multivariate Girsanov theorem. It follows easily from the general Girsanov theorem.

**Multivariate Girsanov** Let $\eta_t \in \mathbb{R}^d$ be adapted to $\mathcal{F}_t$ and let

$$\xi_t = \exp\left( -\int_0^t \eta_s \cdot dW_s - \frac{1}{2}\int_0^t \|\eta_s\|^2 ds \right). \tag{3.15}$$

If $W_t$ is $\mathbb{P}$-BM, then $\tilde{W}_t = W_t + \int_0^t \eta_s ds$ is BM under $\tilde{\mathbb{P}}$.

We consider a stock market of $N$ stocks. Let $S_t = (S_t^1, ..., S_t^N)$ be the stock prices and let $W_t = (W_t^1, ..., W_t^d)$ be a $d$-dimensional independent Brownian Motions. Assume that for each stock,

$$\begin{aligned} dS_t^i &= \mu_t^i S_t^i dt + S_t^i \sum_{j=1}^{d} \sigma_t^{ij} dW_t^j, \\ &= \mu_t^i S_t^i dt + S_t^i \sigma_t^i \cdot dW_t, \quad i = 1, ..., N, \end{aligned}$$

where $\sigma_t^i = (\sigma_t^{i1}, ..., \sigma_t^{Nd})$.

If we find an adapted process $(\eta_t)$ such that

$$\mu_t^i - r_t = \eta_t \cdot \sigma_t^i, \quad i = 1, ..., N, \tag{3.16}$$

we may define $\xi$ as in (3.15) and define accordingly $\tilde{\mathbb{P}}$ such that $\tilde{W}_t = W_t + \int_0^t \eta_s ds$ is BM under $\tilde{\mathbb{P}}$. Hence

$$dS_t^i = r_t S_t^i dt + S_t^i \sigma_t^i \cdot d\tilde{W}_t, \quad i = 1, ..., N.$$

Or

$$d\left(\frac{S_t^i}{M_t}\right) = \left(\frac{S_t^i}{M_t}\right) \sigma_t^i \cdot d\tilde{W}_t.$$

In other words, each numeraire-denominated stock price is a martingale under $\tilde{\mathbb{P}}$.

We restate the crucial condition in (3.16) in matrix form. Let $\mu_t = (\mu_t^1, ..., \mu_t^N)'$, we have

$$\sigma_t \eta_t = \mu_t - r_t. \tag{3.17}$$

Recall that $(\eta_t)$ is called the "market-price-of-risk" (MPR) process and measures the drift in price the investor get compensated for taking each unit of risk.

When the equation (3.17) has no solution, martingale equivalent measure for this market does not exist. In this case, it is always possible to find arbitrage strategies. This says that no arbitrage implies the existence of the "market-price-of-risk" process, hence the existence of martingale equivalent measure.

When $N > d$, some of the securities are "redundant" (derivatives, for example) and can be replicated by a linear combination of other stocks. Thus we may assume $N = d$. If $\text{rank}(\sigma) = d$, there is at most one MPR process, and accordingly, an equivalent martingale measure. The market is said to be complete.

**Market Completeness** Let $\mathcal{M} = \{h_T \cdot X_T | (h_t) \text{is self-financing}\}$. If $\mathcal{M} = \mathcal{L}^2$, the space of all finite-variance random variables, we say the market is *complete*.

A necessary and sufficient condition for complete market is that there is an MPR process and $\text{rank}(\sigma_t) = d$. In other words, there exists a unique MPR process.

## Hedging a General Contingent Claim

We first state an important theorem.

**Theorem 3.4.2 (Martingale Representation Theorem)** *Let $W = (W^1, W^2, ..., W^d)$, and $\mathcal{F}_t$ be the natural filtration of $W$. If $(M_t)$ is a martingale w.r.t. $\mathcal{F}_t$, then there exists $K = (K^1, K^2, ..., K^d)$ such that $\int_0^t (K_s^j)^2 ds < \infty$ for each $j$ and*

$$M_t = M_0 + \int_0^t K_s \cdot dW_s.$$

Given a $\mathcal{F}_T$-measurable contingent claim $C_T$, the no arbitrage price at time $t$ satisfies,

$$\frac{C_t}{M_t} = \tilde{\mathbb{E}} \left( \frac{C_T}{M_T} | \mathcal{F}_t \right).$$

The process of $(C_t/M_t, \mathcal{F}_t)$ is a $\tilde{\mathbb{P}}$ martingale. By martingale representation theorem, we can find an adapted process $(\gamma_t)$ such that

$$\frac{C_t}{M_t} = C(0) + \int_0^t \gamma_s d\tilde{W}_s.$$

Let the value of the hedging portfolio be $H_t$. Suppose we hold $\Delta_t$ amount of stock at time $t$, we have

$$\begin{aligned} dH_t &= \Delta_t dS_t + r_t(H_t - \Delta_t S_t)dt \\ &= r_t H_t dt + \Delta_t \sigma_t S_t(\eta_t dt + dW_t), \end{aligned}$$

where $\eta_t$ is the market price of risk process. Hence

$$d\left(\frac{H_t}{M_t}\right) = \Delta_t \sigma_t S_t(\eta_t dt + dW_t) = \Delta_t \sigma_t S_t M_t^{-1} d\tilde{W}_t.$$

So we have

$$\left(\frac{H_t}{M_t}\right) = H_0 + \int_0^t \Delta_v \sigma_v S_v M_v^{-1} d\tilde{W}_v.$$

To hedge $C_t$, we must have $H_0 = C_0$ and

$$\Delta_t = \frac{M_t}{\sigma_t S_t}\gamma_t.$$

## From Risk Neutral Pricing to PDE

Consider the Black-Scholes Model,

$$\begin{aligned} dM_t &= r_t M_t dt \\ dS_t &= \mu_t S_t dt + \sigma_t S_t dW_t \end{aligned}$$

The MPR $\eta$ exists,

$$\eta_t = \frac{\mu_t - r_t}{\sigma_t}.$$

And $\sigma$ is trivially full rank. Hence the market is complete.

Since $S_T \in \mathcal{L}^2$ , $C_T = \max(S_T - K, 0) \in \mathcal{L}^2$. We know that $M_t^{-1}C_t = \tilde{\mathbb{E}}_t M_T^{-1} C_T$. Let $F(S_t, t) = C_t$, we have

$$\begin{aligned} d\left(\frac{F(S_t,t)}{M_t}\right) &= \frac{1}{M_t}(-r_t F(S_t,t)dt + dF(S_t,t)) \\ &= \frac{1}{M_t}(F_2 + r_t S_t F_1 + \frac{1}{2}F_{11}\sigma_t^2 S_t^2 - r_t F)dt + \frac{S_t}{M_t}\sigma_t F_1 d\tilde{W}_t. \end{aligned}$$

The bounded variation part gives the PDE. And the martingale part gives the hedging strategy, which is

$$\Delta_t = F_1(S_t, t).$$

A complete market admits only one martingale equivalent measure, in which case $C_t$ is unique.

## 3.5 State Prices

### Definition

A state-price deflator is a deflator $m$ for a price process $X$ such that $X^m$ is a martingale w.r.t. the natural filtration.

Other names for $m$ are stochastic discount factor, state-price density, marginal rates of substitution, and pricing kernel.

Given a numeraire $M_t$ and an equivalent martingale measure $\xi$, the state-price deflator is

$$m_t = \frac{\xi_t}{M_t}$$

To see this,

$$\mathbb{E}_s m_t X_t = \mathbb{E}_s \xi_t \frac{X_t}{M_t} = \xi_s \tilde{\mathbb{E}}_s \frac{X_t}{M_t} = \xi_s \frac{X_s}{M_s} = m_s X_s.$$

Conversely, given $m_t$, we can construct $\xi_t$ by

$$\xi_t = M_t m_t.$$

### Risk Premium

$m_t$ is an Ito process, hence it can be characterized by

$$dm_t = m_t \mu_{m,t} dt + m_t(\sigma_{m,t} \cdot dW_t).$$

Since

$$dm_t = d\left(\frac{\xi_t}{M_t}\right) = -r_t m_t dt + M_t^{-1} d\xi_t,$$

So

$$\mu_{m,t} = -r_t.$$

Let a price process be $S_t^i$,

$$dS_t^i = S_t^i \mu_t^i dt + S_t^i(\sigma_t^i \cdot dW_t).$$

We have

$$\begin{aligned} d(m_t S_t^i) &= m_t dS_t^i + S_t^i dm_t + d[m, S^i]_t \\ &= m_t S_t^i(\mu_t^i + \mu_{m,t} + \sigma_{m,t} \cdot \sigma_t^i)dt + m_t S_t^i(\sigma m, t + \sigma_t^i) \cdot dW_t. \end{aligned}$$

$(m_t S_t)$ is a martingale, so

$$\mu_t^i + \mu_{m,t} + \sigma_{m,t} \cdot \sigma_t^i = 0.$$

Hence,

$$\mu_t^i - r_t = -\sigma_t^i \cdot \sigma_{m,t}, \tag{3.18}$$

where both $\sigma_t$ and $\sigma_{m,t}$ can be negative. (3.18) characterizes the excess expected return or risk premium of the stock $S^i$.

Furthermore, if we define

$$\beta_t^i = -\frac{\sigma_t^i \cdot \sigma_{m,t}}{\sigma_{m,t} \cdot \sigma_{m,t}} \quad \text{and } \lambda_{m,t} = \sigma_{m,t} \cdot \sigma_{m,t} \equiv \|\sigma_{m,t}\|^2,$$

then we have

$$\mu_t^i = r_t + \beta_t^i \lambda_{m,t}.$$

$\beta_t^i$ measures the systematic risk in $S_t^i$, and $\lambda_{m,t}$ measures the price of the systematic risk. Note that $\eta_t = -\sigma_{m,t}$, since

$$\eta_t \cdot \sigma_t^i = \mu_t^i - r_t = -\sigma_{m,t} \cdot \sigma_t^i \text{ for all } t.$$

So

$$\lambda_{m,t} = \|\eta_t\|^2.$$

## 3.6 Treatment of Dividends

We discuss how to treat dividend payment in risk-neutral pricing framework.

### 3.6.1 Continuous Payment

We assume that if a stock withholds dividends, the stock price follows a diffusion process,

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t.$$

Now, we may assume that the stock pays dividends continuously at a rate of $d_t$. Then

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t - d_t S_t dt.$$

To replicate the price process of a contingent claim, $(C_t, t \in [0, T])$, we construct a self-financing portfolio $h_t$ which holds $\Delta_t$ stocks at time $t$. Let $H_t$ be the value of the portfolio and let $D_t$ denote the dividends paid cumulatively up to time $t$ and satisfies

$$dD_t = d_t S_t dt.$$

Then we have

$$\begin{aligned} dH_t &= \Delta_t dS_t + \Delta_t dD_t + r_t(H_t - \Delta_t S_t)dt \\ &= r_t H_t dt + \Delta_t S_t(\mu_t - r_t)dt + \Delta_t S_t \sigma_t dW_t \\ &= r_t H_t dt + \sigma_t \Delta_t S_t(\eta_t dt + dW_t), \end{aligned}$$

where

$$\eta_t = \frac{\mu_t - r_t}{\sigma_t}$$

is the MPR process. Define $d\tilde{W}_t = dW_t + \eta_t dt$. Under $\tilde{\mathbb{P}}$, density process of which is defined as the exponential martingale of $L_t = -\int_0^t \eta_s dW_s$, we have

$$dH_t = r_t H_t dt + \sigma_t \Delta_t S_t d\tilde{W}_t.$$

In other words, under $\tilde{\mathbb{P}}$, the numeraire-deflated process $H_t/M_t$ is a martingale,

$$d\left(\frac{H_t}{M_t}\right) = \sigma_t \Delta_t S_t / M_t d\tilde{W}_t.$$

Hence the price of the contingent claim would be given by

$$C_t = M_t \tilde{\mathbb{E}}_t \frac{C_T}{M_T}.$$

It can be shown that under $\tilde{\mathbb{P}}$, the stock price follows

$$dS_t = (r_t - d_t)S_t dt + \sigma_t S_t d\tilde{W}_t.$$

Now let $d_t = d$, $r_t = r$, and $\sigma_t = \sigma$, we have

$$S_T = S_0 \exp\left[(r - d - \frac{1}{2}\sigma^2)T + \sigma \tilde{W}_T\right].$$

From this we can easily calculate Black-Scholes Formula with continuous dividend yield $d$.

### 3.6.2 A Different Perspective

Obviously, if we reinvest the dividends, the "gain process" of the stock follows

$$dG_t = \mu_t G_t dt + \sigma_t G_t dW_t.$$

Consider the market that consists of the stock with dividend reinvestment and a money account, $X_t = (G_t, M_t)$. This market admits no arbitrage if and only if $X_t/M_t$ admits an equivalent martingale measure, say, $\tilde{\mathbb{P}}$. Then the price of a contingent claim $(C_t, t \in [0,T])$ in the unit of a numeraire should be a martingale under $\tilde{\mathbb{P}}$.

To find such a probability measure, we obtain

$$d\left(\frac{G_t}{M_t}\right) = (\mu_t - r_t)G_t/M_t dt + \sigma_t G_t/M_t dW_t = \sigma_t G_t/M_t \left(\frac{\mu_t - r_t}{\sigma_t} + dW_t\right).$$

Define $\tilde{W}_t = W_t + \int_0^t \eta_s ds$, $\eta_t = (\mu_t - r_t)/\sigma_t$, $L_t = -\int_0^t \eta_s dW_s$, and $\xi_t = \exp(L_t - [L]_t/2)$, which corresponds to an equivalent probability measure $\tilde{\mathbb{P}}$. According to Girsanov theorem, $\tilde{W}_t \sim \tilde{\mathbb{P}}$-Brownian Motion. Then $G_t/M_t$ is $\tilde{\mathbb{P}}$-martingale.

### 3.6.3 Discrete Payment

Suppose the dividends are paid at $n$ time points on $[0,T]$, $0 < t_1 < t_2 < ... < t_n < T$. At each time point $t_i$, the dividend payment is $d_i S(t_{i-})$, where $d_i$ is $\mathcal{F}_{t_i}$-measurable and $S(t_{i-})$ denotes the stock price just prior to the payment. The stock price after the payment is

$$S_{t_i} = S_{t_i-} - d_i S_{t_i-} = (1-d_i)S_{t_i-}.$$

We assume that between dividend payment dates the stock price follows,

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t, \quad t_i \leq t < t_{i+1}.$$

Hence, for time $t \in [t_i, t_{i+1})$, the value of hedging portfolio $H_t$ follows

$$\begin{aligned} dH_t &= \Delta_t dS_t + (H_t - \Delta_t S_t) dM_t \\ &= r_t H_t dt + \Delta_t S_t(\mu_t - r_t)dt + \Delta_t S_t \sigma_t dW_t \\ &= r_t H_t dt + \sigma_t \Delta_t S_t(\eta_t dt + dW_t), \end{aligned}$$

where

$$\eta_t = \frac{\mu_t - r_t}{\sigma_t}.$$

Since the portfolio collects the dividend payment, the portfolio value does not jump at payment dates. Hence the above SDE describes the portfolio value for all $t$.

We may define $d\tilde{W}_t = dW_t + \eta_t dt$ and define $\tilde{\mathbb{P}}$ as usual. Then $H_t/M_t$ would be martingale under $\tilde{\mathbb{P}}$ for all $t$. And

$$dS_t = r_t S_t dt + \sigma_t S_t \tilde{d}W_t, \quad t \in [t_i, t_{i+1}), \quad i = 0, ..., n.$$

Now let $r_t = r$ and $\sigma_t = \sigma$, we have

$$S_{t_{i+1}-} = S_i \exp\left[(r - \frac{1}{2}\sigma^2)(t_{i+1} - t_i) + \sigma(\tilde{W}_{t_{i+1}} - \tilde{W}_{t_i})\right],$$

and

$$S_{t_{i+1}} = (1 - d_{i+1})S_i \exp\left[(r - \frac{1}{2}\sigma^2)(t_{i+1} - t_i) + \sigma(\tilde{W}_{t_{i+1}} - \tilde{W}_{t_i})\right].$$

Then we have

$$S_T = \left(S_0 \prod_{i=0}^{n-1}(1 - d_{i+1})\right) \cdot \exp\left[(r - \frac{1}{2}\sigma^2)T + \sigma\tilde{W}_T\right].$$

It is easy to see that we can use the Black-Scholes formula to calculate the price of European call options on the stock $S_t$, with initial value being replaced by $\left(S_0 \prod_{i=0}^{n-1}(1 - d_{i+1})\right)$.

# Chapter 4

# Term Structure Modeling

## 4.1 Basics

We study the price of money, i.e., the default-free interest rate. (Think of the interest rate on bills/notes/bonds issued by the US treasury.) This differs from the price of a particular bond in that the latter depends on factors other than the time value of money, such as the credit history of the borrower, the liquidity condition of the market, and so on.

### Term Structure

We assume that a continuum of default-free discount bonds trade continuously at time $t$ with differing maturities $T$ and prices $P(t,T)$. Assume $P(T,T) = 1$. $P(t,T)$ is called the term structure.

$P(t,T)$ can be read along two dimensions:

1. Fix $t$ and let $T$ vary: prices for different maturities.

2. Fix $T$ and let $t$ vary: historical price series of a particular maturity.

### Yield Curve

The interest rate implied by the zero-coupon bond is called spot rate, which is given by

$$R(t,T) = -\frac{\log P(t,T)}{T-t}.$$

Fixing a $t$ and varying $T$, we call $R(t,T)$ the yield curve. Yield curves can be increasing or decreasing functions of $T$. In practice, yields to maturity on coupon bonds are often calculated. The yield to maturity is the internal rate of return for a coupon bond, or, the constant interest rate that makes the present value of future cash flow (coupon payments and the face value) equal to the market price of the bond. Suppose the coupon bond in question pays a series of coupons in the remaining period ($c_i$ at $t < t_i \leq T$, including the principal), then the yield to maturity $y(r,T)$ solves the following equation,

$$P_c(t,T) = \sum_{t_i > t} c_i \exp(-(t_i - t)y(t,T)), \tag{4.1}$$

where $P_c(t,T)$ is the market price of the coupon bond. Obviously the yield to maturity is not only a function of $T$, but also how coupons are paid (annual, biannual, or quarterly). For coupon bonds, the duration of a coupon bond is defined as:

$$D_t = \frac{\sum_{t_i>t}(t_i - t)c_i \exp(-(t_i - t)y(t,T))}{\sum_{t_i>t} c_i \exp(-(t_i - t)y(t,T))}.$$

The duration $D_t$ thus defined is in fact the derivative of $-\log P_c(t,T)$ in (4.1) with respect to $y(t,T)$. For discount bonds, duration is exactly the term to maturity. Duration may be understood as a measure of risk for coupon bonds.

## Short Rate

The short rate, or the instantaneous rate, measures the *current* cost of short-term borrowing.

$$r_t = \lim_{\Delta \to 0} R(t, t+\Delta) = -\lim_{\Delta \to 0} \frac{\log(P(t,t+\Delta))}{\Delta}.$$

Or,

$$\begin{aligned} r_t &= R(t,t), \text{ and} \\ r_t &= -\frac{\partial}{\partial T} \log P(t,t) \end{aligned}$$

## Forward Rate

Let $t < T_1 < T_2$. Consider a forward contract on a bond that matures at $T_2$: an agreement at time $t$ to make a payment at $T_1$ and receive a payment in return at $T_2$.

We can replicate the contract, at time $t$, by

- buying a $T_2$ bond
- selling $k$ units of $T_1$ bond.

The cash flow of this portfolio is

- At $t$, $-P(t,T_2)+kP(t,T_1)$
- At $T_1$, -k
- At $T_2$, 1

Since the value of any forward contract should be zero at the time of agreement $t$, $k$ must satisfies

$$k=\frac{P(t,T_2)}{P(t,T_1)}.$$

Obviously, $k$ should be called the forward price of the $T_2$-bond. The corresponding yield of holding the $T_2$-bond in the interval of $[T_1,T_2]$, denoted as $F(t,T_1,T_2)$, is

$$F(t,T_1,T_2)=-\frac{\log k}{T_2-T_1}=-\frac{\log P(t,T_2)-\log P(t,T_1)}{T_2-T_1}.$$

Now it is ready to define forward rate, the forward price for instantaneous borrowing at time $T$,

$$f(t,T)=\lim_{T_2\to T}F(t,T,T_2)=\lim_{\Delta\to 0}-\frac{\log P(t,T+\Delta)-\log P(t,T)}{\Delta}=-\frac{\partial}{\partial T}\log P(t,T).$$

The forward rate $f(t,T)$ contains all information about $P(t,T)$ and $R(t,T)$. Specifically, we have

$$P(t,T)=\exp\left(-\int_t^T f(t,u)du\right),$$

and

$$R(t,T)=\frac{\int_t^T f(t,u)du}{T-t}.$$

And the short rate $r_t$ can be recovered using

$$r_t=f(t,t).$$

We also have

$$\begin{aligned}\frac{\partial R(t,T)}{\partial T} &= -\frac{\partial \log P(t,T)}{\partial T}\frac{1}{T-t}+\frac{\log P(t,T)}{(T-t)^2}\\ &= \frac{f(t,T)}{T-t}-\frac{1}{T-t}R(t,T).\end{aligned}$$

Hence

$$f(t,T) = R(t,T) + (T-t)\frac{\partial R(t,T)}{\partial T}.$$

When $T = t$, $f(t,t) = R(t,t) = r_t$. Otherwise, $f(t,T)$ is greater (less) than $R(t,T)$ when $R(t,T)$ is increasing (decreasing). Finally, we have

$$F(t,T_1,T_2) = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} f(t,s)ds.$$

## 4.2 The Single-Factor Heath-Jarrow-Morton Model

### 4.2.1 The Risk-Neutral Pricing

**The Model** $P(t,T)$, $R(t,T)$, and $f(t,T)$ contain the same information. The HJM model (Heath, Jarrow, and Morton, 1992) is a model on $f(t,T)$:

$$f(t,T) = f(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T)dW_s. \tag{4.2}$$

Or, in its differential form,

$$d_t f(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t.$$

- $\alpha(t,T)$ and $\sigma(t,T)$ may depend on $(W_s, s \le t)$ and $f(t,T)$ itself.
- $f(0,T)$ is deterministic.
- $\int_0^T \int_0^u |\alpha(t,u)|dtdu < \infty$ and $\mathbb{E}\left(\int_0^T \left|\int_0^u \sigma(t,u)dW_t\right| du\right) < \infty$.

**The Numeraire** We use $M_t$, which satisfies

$$M_0 = 1, \quad \text{and} \quad dM_t = r_t M_t dt.$$

Or

$$M_t = \exp\left(\int_0^t r_s ds\right).$$

Recall that $r_t = f(t,t)$. So

$$r_t = f(0,t) + \int_0^t \alpha(s,t)ds + \int_0^t \sigma(s,t)dW_s.$$

So

$$
\begin{aligned}
M_t &= \exp\left(\int_0^t f(0,s)ds + \int_0^t \int_0^u \alpha(s,u)dsdu + \int_0^t \int_0^u \sigma(s,u)dW_s du\right) \\
&= \exp\left(\int_0^t f(0,s)ds + \int_0^t \int_s^t \alpha(s,u)duds + \int_0^t \int_s^t \sigma(s,u)dudW_s\right).
\end{aligned}
$$

**The Bond** We can choose any bond price to construct martingale equivalent measure under no arbitrage condition. Consider $P(t,T)$,

$$
\begin{aligned}
P(t,T) &= \exp\left(-\int_t^T f(t,u)du\right) \\
&= \exp\left(-\left[\int_t^T f(0,u)du + \int_0^t \int_t^T \alpha(s,u)duds + \int_0^t \int_t^T \sigma(s,u)dudW_s\right]\right)
\end{aligned}
$$

We can check that $P(0,T) = \exp\left(-\int_0^T f(0,u)du\right)$, and $P(T,T) = 1$.

**Deflation** Define

$$
\begin{aligned}
Z(t,T) &= M_t^{-1} P(t,T) \\
&= \exp\left[-\int_0^T f(0,u)du - \int_0^t \int_s^T \alpha(s,u)duds - \int_0^t \left(\int_s^T \sigma(s,u)du\right) dW_s\right] \\
&= \exp\left[-\int_0^T f(0,u)du - \int_0^t \int_s^T \alpha(s,u)duds + \int_0^t \Sigma(s,T)dW_s\right],
\end{aligned}
$$

where $\Sigma(t,T) = -\int_t^T \sigma(t,u)du$. Let $X_t$ be the term in the bracket,

$$
\begin{aligned}
d_t Z(t,T) &= d(\exp(X_t)) \\
&= \exp(X_t)dX_t + \frac{1}{2}\exp(X_t)d[X]_t \\
&= Z(t,T)\left(\left(\frac{1}{2}\Sigma(t,T)^2 - \int_t^T \alpha(t,u)du\right) dt + \Sigma(t,T)dW_t\right).
\end{aligned}
$$

**Change of Measure** Then the market-price-of-risk process would be

$$
\begin{aligned}
\eta_t &= \frac{\frac{1}{2}\Sigma^2(t,T) - \int_t^T \alpha(t,u)du}{\Sigma(t,T)} \\
&= \frac{1}{2}\Sigma(t,T) - \Sigma^{-1}\int_t^T \alpha(t,u)du.
\end{aligned}
$$

Then we can define $\tilde{\mathbb{P}}$ and a $\tilde{W}_t$ such that

$$d\tilde{W}_t = dW_t + \eta_t dt.$$

Then

$$d_t Z(t,T) = Z(t,T)\Sigma(t,T)d\tilde{W}_t.$$

Hence $Z(t,T)$ is a $\tilde{\mathbb{P}}$-martingale. And under $\tilde{\mathbb{P}}$, the bond price $P(t,T)$ has a drift term $r_t$:

$$d_t P(t,T) = P(t,T)(r_t dt + \Sigma(t,T)d\tilde{W}_t).$$

**Other Bonds** We use $P(t,T)$ to construct a martingale equivalent measure $\tilde{\mathbb{P}}$. What about other bonds, such as $P(t,S)$, $S < T$?

Let $X = 1$ be a claim that pays off at time $S$. Then $P(t,S)$ is the price of $X$ at time $t$,

$$P(t,S) = M_t \tilde{\mathbb{E}}_t(M_S^{-1}) = \tilde{\mathbb{E}}_t \left( \exp\left( -\int_t^S r_s ds \right) \right).$$

And the deflated price process is

$$Z(t,S) = M_t^{-1} P(t,S) = \tilde{\mathbb{E}}_t(M_S^{-1}).$$

So the deflated prices of all other bonds are $\tilde{\mathbb{P}}$-martingale. This means that their $\tilde{\mathbb{P}}$-drifts are restricted such that $\eta_t$ is the same market-price-of-risk process for all bonds. In particular, for all $S \in [0,T]$,

$$\int_t^S \alpha(t,s)ds = \frac{1}{2}\Sigma^2(t,S) - \Sigma(t,S)\eta_t.$$

Taking $\partial/\partial S$ on both sides,

$$\begin{aligned} \alpha(t,S) &= -\Sigma(t,S)\sigma(t,S) + \sigma(t,S)\eta_t \\ &= \sigma(t,S)(\eta_t - \Sigma(t,S)). \end{aligned}$$

If there exists an $(\eta_t)$ such that the above holds, among other regularity conditions, then the market is complete. As in Section **??**, we may find a self-financing portfolio of $(M_t, P(t,T))$, that replicates any contingent claim $U_S$ which pays off at time $S < T$.

### 4.2.2 A Direct Approach

We have

$$\begin{aligned} f(t,T) &= f(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T)(d\tilde{W}_s - \eta_s ds) \\ &= f(0,T) + \int_0^t (\alpha(s,T) - \sigma(s,T)\eta_s)dt + \int_0^t \sigma(s,T)d\tilde{W}_s \\ &= f(0,T) + \int_0^t (-\Sigma(s,T)\sigma(s,T))dt + \int_0^t \sigma(s,T)d\tilde{W}_s. \end{aligned}$$

So as in Duffie (2001), we may directly assume no arbitrage opportunities and there exists a martingale equivalent measure $\tilde{\mathbb{P}}$ such that we may specify $f(t,T)$ as

$$f(t,T) = f(0,T) + \int_0^t \mu(s,T)ds + \int_0^t \sigma(s,T)d\tilde{W}_s,$$

where

$$\begin{aligned} \mu(t,T) &= -\sigma(t,T)\Sigma(t,T) \\ &= \sigma(t,T)\int_t^T \sigma(t,s)ds. \end{aligned}$$

Then

$$r_t = f(0,t) + \int_0^t \sigma(s,t)\int_s^t \sigma(s,u)duds + \int_0^t \sigma(s,T)d\tilde{W}_s.$$

## 4.3 Short-Rate Models

The short-rate model is a model on $r_t$. Assume there exists a martingale equivalent measure $\tilde{\mathbb{P}}$. $r_t$ is usually specified as a Markov diffusion:

$$dr_t = \nu(r_t)dt + \rho(r_t)d\tilde{W}_t. \tag{4.3}$$

Then the term structure $P(t,T)$ is given as

$$P(t,T) = \tilde{\mathbb{E}}_t\left(\exp(-\int_t^T r_s ds)\right).$$

Note that the short rate process alone does not recover the term structure, which is determined by the risk appetite of the market as well as the future short-term borrowing cost.

### 4.3.1 Connection with Forward-Rate Models

Given $f(t,T)$, we can easily recover $r_t$. And we may recover the forward rate $f(t,T)$ from the short rate $r_t$ as follows. We define $g(x,t,T)$ as

$$g(x,t,T) = -\log\left[\tilde{\mathbb{E}}\left(\exp\left(-\int_t^T r_s ds\right) \mid r_t = x\right)\right]$$

Then we have

$$g(r_t,t,T) = -\log P(t,T) = \int_t^T f(t,u)du.$$

In other words,

$$f(t,T) = \frac{\partial}{\partial T} g(r_t,t,T).$$

Hence,

$$\begin{aligned} d_t f(t,T) &= \frac{\partial^2 g}{\partial T \partial t} dt + \frac{\partial^2 g}{\partial T \partial x} dr_t + \frac{1}{2}\frac{\partial^3 g}{\partial T \partial x^2} d[r]_t \\ &= \left(\frac{\partial^2 g}{\partial T \partial t}\nu(r_t) + \frac{\partial^2 g}{\partial T \partial t} + \frac{1}{2}\frac{\partial^3 g}{\partial T \partial x^2}\rho^2(r_t)\right) dt + \frac{\partial^2 g}{\partial x \partial T}\rho(r_t) d\tilde{W}_t. \end{aligned}$$

So

$$\begin{aligned} \sigma(t,T) &= \frac{\partial^2 g(r_t,t,T)}{\partial T \partial x}\rho(r_t) \\ \Sigma(t,T) &= -\frac{\partial g(r_t,t,T)}{\partial x}\rho(r_t). \end{aligned}$$

Note also that

$$f(0,T) = \frac{\partial g(r_0,0,T)}{\partial T}.$$

$f(0,T)$ together with $\sigma(t,T)$ determines $f(t,T)$ under $\tilde{\mathbb{P}}$.

### 4.3.2 Examples

#### Ho and Lee Model

The short rate process $r_t$ satisfies

$$dr_t = \nu_t dt + \rho d\tilde{W}_t, \tag{4.4}$$

where $\nu_t$ is deterministic and bounded and $\rho$ is constant.

For $s \geq t$,
$$r_s = r_t + \int_t^s \nu_u du + \int_t^s \rho d\tilde{W}_u.$$
Hence
$$\begin{aligned}
\int_t^T r_s ds &= r_t(T-t) + \int_t^T \int_t^s \nu_u du ds + \int_t^T \int_t^s \rho d\tilde{W}_u ds \\
&= r_t(T-t) + \int_t^T \int_u^T \nu_u ds du + \int_t^T \int_u^T \rho ds d\tilde{W}_u \\
&= r_t(T-t) + \int_t^T \nu_u(T-t) du + \int_t^T \rho(T-u) d\tilde{W}_u
\end{aligned}$$
Let $M_T = -\int_t^T \rho(T-u)d\tilde{W}_u$. $M_T$ is a $\tilde{\mathbb{P}}$-martingale and $M_t = 0$. We have
$$[M]_T = \rho^2 \int_t^T (T-u)^2 du.$$
Then $\exp(M_T - \frac{1}{2}[M]_T)$ is a (positive) $\tilde{\mathbb{P}}$-martingale with $\exp(M_t - \frac{1}{2}[M]_t) = 1$. Hence
$$\begin{aligned}
\tilde{\mathbb{E}}_t \exp(M_T) &= \exp\left(\frac{1}{2}[M]_T\right) \tilde{\mathbb{E}}_t \exp\left(M_T - \frac{1}{2}[M]_T\right) \\
&= \exp\left(\frac{1}{2}[M]_T\right) \exp(M_t - \frac{1}{2}[M]_t) \\
&= \exp\left(\frac{1}{2}\rho^2 \int_t^T (T-u)^2 du\right).
\end{aligned}$$
We have
$$\exp\left(-\int_t^T r_s ds\right) = \exp\left(-r_t(T-t)\right) \exp\left(-\int_t^T \nu_u(T-u)du\right) \exp\left(M_T\right)$$
Hence
$$\tilde{\mathbb{E}}\left[e^{-\int_t^T r_s ds} | r_t = x\right] = e^{-x(T-t)} \cdot e^{-\int_t^T \nu_u(T-u)du} \cdot e^{\frac{1}{2}\rho^2 \int_t^T (T-u)^2 du}.$$
Hence
$$\begin{aligned}
g(x,t,T) &= -\log\left(\tilde{\mathbb{E}}\left(e^{-\int_t^T r_s ds} | r_t = x\right)\right) \\
&= x(T-t) + \int_t^T \nu_u(T-u)du - \frac{1}{2}\rho^2 \int_t^T (T-u)^2 du \\
&= x(T-t) + \int_t^T \nu_u(T-u)du - \frac{1}{6}\rho^2 (T-t)^3.
\end{aligned}$$

The HJM volatility $\sigma(t,T)$ is then

$$\sigma(t,T) = \rho \frac{\partial^2 g(r_t,t,T)}{\partial x \partial T} = \rho,$$

which does not depend on $t$ or $T$.

And $\Sigma(t,T)$ is

$$\Sigma(t,T) = -\rho(T-t).$$

So Ho and Lee model is equivalent to the following HJM model:

$$d_t f(t,T) = \rho^2 (T-t)dt + \rho d\tilde{W}_t,$$

with

$$f(0,T) = \frac{\partial g(r_0,0,T)}{\partial T} = r_0 - \frac{1}{2}\rho^2 T^2 + \int_0^T \nu_s ds$$

.

We may easily generalize (4.4) as follows,

$$dr_t = \nu_t dt + \rho_t d\tilde{W}_t.$$

The HJM counterpart would be

$$d_t f(t,T) = \rho_t^2 (T-t)dt + \rho_t d\tilde{W}_t,$$

with

$$f(0,T) = r_0 - \int_t^T \rho_s^2 (T-s)ds + \int_0^T \nu_s ds.$$

Now the HJM volatility depends on time $t$.

### Vasicek Model

Now we allow the drift depend on $r_t$ itself,

$$dr_t = (\theta - \alpha r_t)dt + \rho d\tilde{W}_t, \tag{4.5}$$

where $\theta$, $\alpha$, and $\rho$ are constants. The process described by (4.5) is the well-known Orstein-Uhlenbeck process. This model translates into a HJM model with volatility that depends on maturity $T$ as well as time $t$.

Exercise: Show that the HJM representation of the Vasicek model is

$$\sigma(t,T) = \rho \exp(-\alpha(T-t)),$$

with

$$f(0,T) = \theta/\alpha + e^{-\alpha T}(r_0 - \theta/\alpha) - \frac{\rho^2}{2\alpha^2}(1 - e^{-\alpha T})^2.$$

We may generalize (4.5) into the following form,

$$dr_t = (\theta_t - \alpha_t r_t)dt + \rho_t d\tilde{W}_t,$$

where $\theta_t$, $\alpha_t$, and $\rho_t$ are deterministic processes.

**Cox-Ingersoll-Ross Model**

Both Ho and Lee model and Vasicek model may display negative short rates. The Cox-Ingersoll-Ross model avoids this problem.

$$dr_t = (\theta - \alpha r_t)dt + \rho\sqrt{r_t}d\tilde{W}_t, \tag{4.6}$$

where $\theta$, $\alpha$, and $\rho$ are constants. If $\theta \geq \rho^2/2$, $r_t$ is positive a.s. The process described by (4.6) is the Feller's Square Root Process. We may easily generalize the CIR model to allow deterministic processes $\theta_t$, $\alpha_t$ and $\rho_t$ in place of the corresponding constants.

The HJM equivalent model needs a special function $B(t,T)$ which is the solution to the Riccati differential equation

$$\frac{\partial B(t,T)}{\partial t} - \alpha B(t,T) - \frac{1}{2}\rho^2 B^2(t,T) + 1 = 0, \text{ with } B(T,T) = 0.$$

Then we have

$$g(x,t,T) = xB(t,T) + \theta \int_t^T B(s,T)ds.$$

Define $D(t,T) = \partial B/\partial T$. Then the HJM volatility can be written as

$$\begin{aligned} \sigma(t,T) &= \rho\sqrt{r_t}D(t,T) \\ \Sigma(t,T) &= -\rho\sqrt{r_t}B(t,T). \end{aligned}$$

The initial value can also be easily calculated,

$$f(0,T) = r_0 D(0,T) + \theta \int_0^T D(s,T)ds.$$

#### Black-Karasinski Model

The Black-Karasinski model forces the short rate to be positive by taking exponential of an Orstein-Uhlenbeck process:

$$r_t = \exp(X_t), \tag{4.7}$$

where

$$dX_t = (\theta_t - \alpha_t X_t)dt + \rho_t d\tilde{W}_t.$$

Using Ito's formula, we may write the Black-Karasinski Model as

$$dr_t = \left((\theta_t - \alpha_t \log r_t)r_t + \frac{1}{2}\rho_t^2 r_t\right) dt + \rho_t r_t d\tilde{W}_t.$$

#### The General Parametric Model

In general, we may write the short rate model in the following form,

$$dr_t = [c_0(t) + c_1(t)r_t + c_2(t)r_t \log r_t]\, dt + [d_0(t) + d_1(t)r_t]^v\, d\tilde{W}_t \tag{4.8}$$

Here are some special cases.

- $c_2 = 0$, $d_0 = 0$, $v = 0.5$, CIR
- $c_1 = 0$, $c_2 = 0$, $d_1 = 0$, $v = 1$, Ho and Lee
- $c_2 = 0$, $d_1 = 0$, $v = 1$, Vasicek
- $d_0 = 0$, $v = 1$, Black-Karasinski

In particular, $c_1$ is usually called "mean-reversion" parameter.

If $c_0 = d_1 = 0$, $r_t$ is Gaussian. It can be shown that $g(x, t, T)$ satisfies

$$g(x, t, T) = A(t, T) + B(t, T)x.$$

### 4.3.3 Affine Models

When $g(x, t, T)$ is affine in $x$, ie,

$$g(x, t, T) = A(t, T) + B(t, T)x,$$

then we call the associated term structure model an affine term structure model.

The term structure is affine if and only if $\nu$ and $\rho^2$ in the definition of short-rate models (4.3) is affine, ie,

$$\begin{aligned}\nu(r_t) &= c_0(t) + c_1(t) r_t \\ \rho^2(r_t) &= d_0(t) + d_1(t) r_t.\end{aligned}$$

Given $\nu$ and $\rho$, we may recover $A$ and $B$. First, $B$ satisfies the following Riccati equation,

$$\frac{\partial B(t,T)}{\partial t} + c_1(t)B(t,T) - \frac{1}{2}d_1(t)B^2(t,T) + 1 = 0, \quad \text{with} \quad B(T,T) = 0.$$

And

$$A = \int_t^T \left( c_0(s)B(s,T) - \frac{1}{2}d_0(s)B^2(s,T) \right) ds.$$

In particular, Gaussian models and the CIR model have explicit solution. Others may be solved numerically.

### 4.3.4 The Feynman-Kac Formulation

In a single-factor model, the evolution of short rate depends on one factor only. To emphasize this point, we may write the short rate model in (4.3) as

$$r_t = X_t, \quad \text{and} \quad dX_t = \nu(X_t)dt + \rho(X_t)d\tilde{W}_t. \tag{4.9}$$

We will see that this formulation extends easily to multi-factor models.

By the Markovian nature of $r_t$, $P(t,T)$ can be represented as $P(t,T) = F(r_t, t)$. Recall that

$$\begin{aligned}F(x,t) &= \tilde{\mathbb{E}}_t \left[ \exp\left( -\int_t^T r_s ds \right) | X_t = x \right] \\ &= \tilde{\mathbb{E}}_t \left[ \exp\left( -\int_t^T r_s ds \right) | r_t = x \right].\end{aligned}$$

It is clear that $F$ solves the following partial differential equation,

$$F_2(x,t) + \nu(x)F_1(x,t) + \frac{1}{2}\rho^2(x)F_{11}(x,t) - xF(x,t) = 0 \tag{4.10}$$

with

$$F(x,T) = 1.$$

We can thus solve the above pde numerically for $F(x,t)$, and thus $P(t,T)$.

## 4.4 Multi-factor Models

Now we assume the economy is subject to more than one "shocks". Let $W = (W^1, ..., W^d)'$ be a $d$-dimensional Brownian Motion.

### 4.4.1 Multi-factor Heath-Jarrow-Morton Model

Let $\sigma = (\sigma_1, ..., \sigma_d)'$, we can specify the forward rate as

$$\begin{aligned} f(t,T) &= f(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T) \cdot dW_s \\ &= f(0,T) + \int_0^t \alpha(s,T)ds + \sum_{i=1}^d \int_0^t \sigma_i(s,T)dW_s^i. \end{aligned}$$

In differential form,

$$d_t f(t,T) = \alpha(t,T)dt + \sigma(t,T) \cdot dW_t.$$

For example, we may have

$$d_t f(t,T) = \alpha(t,T)dt + \sigma_1 dW_t^1 + \sigma_2 e^{-\kappa(T-t)} dW_t^2,$$

where $\sigma_1$, $\sigma_2$, and $\kappa$ are constants. In this model, $W^1$ provides "shocks" that are felt equally by all points on the yield curve and $W^2$ "shocks" that are felt only in the short term.

For a small interval $\Delta$,

$$f(t+\Delta,T) - f(t,T) \approx \Delta\alpha(t,T) + \sum_{i=1}^d \sigma_i(t,T)(W_{t+\Delta}^i - W_t^i).$$

Hence

$$\lim_{\Delta\to 0} \frac{1}{\Delta} \text{var}\left(f(t+\Delta,T) - f(t,T)\right) = \sum_{i=1}^d \sigma_i^2(t,T).$$

And

$$\lim_{\Delta\to 0} \frac{1}{\Delta} \text{cov}\left(f(t+\Delta,T) - f(t,T), f(t+\Delta,S) - f(t,S)\right) = \sum_{i=1}^d \sigma_i(t,T)\sigma_i(t,S).$$

We may define an instantaneous correlation coefficient for the increments of the forward rate,

$$\frac{\sum_{i=1}^d \sigma_i(t,T)\sigma_i(t,S)}{\sqrt{\sum_{i=1}^d \sigma_i^2(t,T) \cdot \sum_{i=1}^d \sigma_i^2(t,S)}}.$$

If $d = 1$, the increments of the forward rates are perfectly correlated everywhere on the yield curve.

The results of single-factor HJM may be easily generalized. Again we use $M_t$ as numeraire and denote the discounted bond price as $Z(t,T) = M_t^{-1}P(t,T)$. We have

$$d_t Z(t,T) = Z(t,T)\left[\left(\frac{1}{2}(\Sigma\cdot\Sigma)(t,T) - \int_t^T \alpha(t,u)du\right)dt + \Sigma(t,T)\cdot dW_t\right]$$

Then we seek an $\eta = (\eta_t) \in \mathbb{R}^d$ such that

$$\Sigma(t,T)\cdot\eta_t = \frac{1}{2}(\Sigma\cdot\Sigma)(t,T) - \int_t^T \alpha(t,u)du.$$

For the above to have solution, it is necessary that the matrix $\Sigma \equiv (\Sigma^i(t,T_j))$ be full rank for all $t$ and $T_j$. The we define a $\tilde{\mathbb{P}}$ such that $\tilde{W}_t$ defined below is $\tilde{\mathbb{P}}$-BM,

$$d\tilde{W}_t = dW_t + \eta_t dt.$$

Under $\tilde{\mathbb{P}}$, $Z_t(t,T)$ is then martingale,

$$d_t Z(t,T) = Z(t,T)\Sigma(t,T)\cdot d\tilde{W}_t.$$

The bond price satisfies

$$d_t P(t,T) = P(t,T)\left(r_t dt + \Sigma(t,T)\cdot d\tilde{W}_t\right).$$

And the forward rate,

$$d_t f(t,T) = -(\sigma\cdot\Sigma)(t,T)dt + \sigma\cdot d\tilde{W}_t.$$

### 4.4.2 Multi-factor Short Rate Models

Let $\tilde{W} = (\tilde{W}^1, \tilde{W}^2, ..., \tilde{W}^d)$ be a $d$-dimensional Brownian Motion under $\tilde{\mathbb{P}}$. And Let $X = (X^1, X^2, ..., X^N)$ be the $N$ factors that determines the short rate $r_t$. $X$ generally includes a factor that is the short rate itself.

We write

$$r_t = r(X_t), \tag{4.11}$$

where $X_t$ satisfies

$$dX_t = \nu(X_t)dt + \rho(X_t)\cdot d\tilde{W}_t. \tag{4.12}$$

The term structure $P(t,T)$ can then be represented as

$$P(t,T) = \tilde{\mathbb{E}}_t \left[ \exp\left( -\int_t^T r(X_s)ds \right) \right].$$

Of course, any derivative that has a terminal payment of $g(X_T)$ may be priced as

$$\tilde{\mathbb{E}}_t \left[ \exp\left( -\int_t^T r(X_s)ds \right) g(X_T) \right].$$

### 4.4.3 Feynman-Kac Formulation

It is easy to extend the Feynman-Kac formulation of single-factor short rate model in (4.10) to the multi-factor case. Let $F(x,t) = P(t,T)$, we have

$$F_2(x,t) + \nu(x) \cdot F_1(x,t) + \frac{1}{2}\mathrm{tr}\left[\rho(x)\rho(x)' F_{11}(x,t)\right] - r(x)F(x,t) = 0, \tag{4.13}$$

with

$$F(x,T) = 1. \tag{4.14}$$

Obviously, if we change the boundary condition in (4.14) to $F(x,T) = g(x)$, $F(x,t)$ prices any general derivative with terminal payment $g(X_T)$.

## 4.5 Pricing Interest Rate Products

In this section we briefly review the pricing of some popular interest rate products, given the term structure $P(t,T)$.

### 4.5.1 Bond with Fixed Coupons

Suppose the coupon rate (uncompounded) is $k$ and the payment is made at a sequence of dates $T_i = T_0 + i\Delta$. The cash flow is shown in the diagram

This is equivalent to owning a $T_n$-bond and $k\Delta$ units of $T_i$-bond for each $i = 1, ..., n$:

$$\begin{cases} P(T_0, T_n) \\ k\Delta P(T_0, T_i), \; i = 1, ..., n. \end{cases}$$

From

$$k\Delta \sum_{i=1}^{n} P(T_0, T_i) + P(T_0, T_n) = 1,$$

we can determine the appropriate coupon rate,

$$k = \frac{1 - P(T_0, T_n)}{\Delta \sum_{i=1}^{n} P(T_0, T_i)}.$$

### 4.5.2 Floating-Rate Bond

Now the coupon rate paid at time $T_i$ is the floating rate at previous payment date $T_{i-1}$, which is defined as

$$L(T_{i-1}) = \frac{1}{\Delta}\left(\frac{1}{P(T_{i-1}, T_i)} - 1\right).$$

The cash flow is illustrated in the diagram.

The value of $\Delta L(T_{i-1})$ at $T_0$ is

$$\begin{aligned}
M_{T_0}\tilde{\mathbb{E}}_{T_0}\left(M_{T_i}^{-1}\Delta L(T_{i-1})\right) &= M_{T_0}\tilde{\mathbb{E}}_{T_0}\left[M_{T_i}^{-1}\left(\frac{1}{P(T_{i-1}, T_i)} - 1\right)\right] \\
&= M_{T_0}\tilde{\mathbb{E}}_{T_0}\left[P^{-1}(T_{i-1}, T_i)\tilde{\mathbb{E}}_{T_{i-1}}\left(M_{T_i}^{-1}\right) - M_{T_i}^{-1}\right] \\
&= M_{T_0}\tilde{\mathbb{E}}_{T_0}\left[\tilde{\mathbb{E}}_{T_{i-1}} M_{T_{i-1}}^{-1} - M_{T_i}^{-1}\right] \\
&= M_{T_0}\tilde{\mathbb{E}}_{T_0}\left[M_{T_{i-1}}^{-1} - M_{T_i}^{-1}\right] \\
&= P(T_0, T_{i-1}) - P(T_0, T_i).
\end{aligned}$$

Note that $P^{-1}(T_{i-1}, T_i)$ is $\mathcal{F}_{T_{i-1}}$-measurable and $P(T_{i-1}, T_i) = M_{T_{i-1}}\tilde{\mathbb{E}}_{T_{i-1}} M_{T_i}^{-1}$. The contingent claim $\Delta L(T_{i-1})$ can be replicated by buying a $T_{i-1}$-bond and sell a $T_i$-bond. At time $T_{i-1}$, buy $P^{-1}(T_{i-1}, T_i)$ units of $T_i$-bond. The value of floating rate bond is then

$$P(T_0, T_n) + \sum_{i=1}^{n} \left(P(T_0, T_{i-1}) - P(T_0, T_i)\right) = P(T_0, T_0) = 1.$$

### 4.5.3 Swaps

A swap contract exchanges a sequence of floating rate payments for a sequence of fixed-rate payments or vice sersa. The cash flow is shown in the diagram.

Entering a swap contract is equivalent to buying a fixed-coupon bond and selling a floating-rate bond. The former is worth

$$P(T_0, T_n) + k\Delta \sum_{i=1}^{n} P(T_0, T_i),$$

and the latter is worth 1. So the fixed coupon rate $k$ must satisfy

$$k = \frac{1 - P(T_0, T_n)}{\Delta \sum_{i=1}^{n} P(T_0, T_i)}.$$

### 4.5.4 Forward Swaps

The value of the swap at time $T_0$ is

$$X = P(T_0, T_n) + k\Delta \sum_{i=1}^{n} P(T_0, T_i) - 1.$$

The value of $X$ at time $t$ must be

$$\begin{aligned}
M_t \tilde{\mathbb{E}}_t \left( M_{T_0}^{-1} X \right) &= M_t \tilde{\mathbb{E}}_t \left( M_{T_0}^{-1} P(T_0, T_n) + k\Delta \sum_{i=1}^{n} M_{T_0}^{-1} P(T_0, T_i) - M_{T_0}^{-1} \right) \\
&= M_t \tilde{\mathbb{E}}_t \left( \tilde{\mathbb{E}}_{T_0} M_{T_n}^{-1} + k\Delta \sum_{i=1}^{n} \tilde{\mathbb{E}}_{T_0} M_{T_i}^{-1} - M_{T_0}^{-1} \right) \\
&= M_t \tilde{\mathbb{E}}_t M_{T_n}^{-1} + k\Delta \sum_{i=1}^{n} M_t \tilde{\mathbb{E}}_t M_{T_i}^{-1} - M_t \tilde{\mathbb{E}}_t M_{T_0}^{-1} \\
&= P(t, T_n) + k\Delta \sum_{i=1}^{n} P(t, T_i) - P(t, T_0) = 0.
\end{aligned}$$

So the forward swap rate must be

$$k = \frac{P(t, T_0) - P(t, T_n)}{\Delta \sum_{i=1}^{n} P(t, T_i)}.$$

When $t = T_0$, $k$ is equal to the swap rate.

### 4.5.5 Swaptions

A swaption is a contract to enter a swap at time $T_0$ with swap rate $k$. The value of the swaption at time $T_0$ is

$$X = \max\left(P(T_0, T_n) + k\Delta \sum_{i=1}^{n} P(T_0, T_i) - 1, 0\right).$$

We can price this like any other contingent claims.

### 4.5.6 Caps and Floors

A caps contract is an agreement that never pays more than a fixed rate $k$. So a cap contract pays at time $T_i$

$$X = \Delta \max\left(L(T_{i-1}) - k, 0\right).$$

$X$ is called a "caplet". Note that

$$\begin{aligned}
\Delta L(T_{i-1}) - \Delta k &= \frac{1}{P(T_{i-1}, T_i)} - 1 - \Delta k \\
&= (1 + \Delta k)\left(\frac{1}{P(T_{i-1}, T_i)(1 + \Delta k)} - 1\right) \\
&= (1 + \Delta k)P^{-1}(T_{i-1}, T_i)\left(\frac{1}{1 + \Delta k} - P(T_{i-1}, T_i)\right).
\end{aligned}$$

So

$$\begin{aligned}
X &= \Delta \max\left(L(T_{i-1}) - k, 0\right) \\
&= (1 + \Delta k)P^{-1}(T_{i-1}, T_i) \max\left(\frac{1}{1 + \Delta k} - P(T_{i-1}, T_i), 0\right).
\end{aligned}$$

So the value of $X$ at time $t$ is

$$\begin{aligned}
& M_t \tilde{\mathbb{E}}_t\left(M_{T_i}^{-1} X\right) \\
= \; & (1 + \Delta k) M_t \tilde{\mathbb{E}}_t\left(M_{T_i}^{-1} P^{-1}(T_{i-1}, T_i) \max\left(\frac{1}{1 + \Delta k} - P(T_{i-1}, T_i), 0\right)\right) \\
= \; & (1 + \Delta k) M_t \tilde{\mathbb{E}}_t\left(\left(\tilde{\mathbb{E}}_{T_{i-1}} M_{T_i}^{-1}\right) P^{-1}(T_{i-1}, T_i) \max\left(\frac{1}{1 + \Delta k} - P(T_{i-1}, T_i), 0\right)\right) \\
= \; & (1 + \Delta k) M_t \tilde{\mathbb{E}}_t\left(M_{T_{i-1}}^{-1} \max\left(\frac{1}{1 + \Delta k} - P(T_{i-1}, T_i), 0\right)\right).
\end{aligned}$$

This is $(1 + \Delta k)$ units of put option on the $T_i$-bond with strike price $(1 + \Delta k)$ and maturity date $T_{i-1}$.

A floor contract is an agreement to never pay less than $k$ at each $T_i$.

# Chapter 5

# Econometric Issues

## 5.1 Basics on Markov Processes

**Notation.** We consider a time-homogeneous Markov process $(X_t)$. In this case,

$$P_t(x, A) = \mathbb{P}\left(X_t \in A | X_0 = x\right) = \mathbb{P}\left(X_{t+s} \in A | X_s = x\right).$$

$P_t(x, A)$ is called transition probability and may be characterized by transition density $p_t(x, y, t)$,

$$P_t(x, dy) = p(t, x, y)dy.$$

$P_t(x, A)$ can also be characterized by conditional moments:

$$P_t f(x) \equiv \mathbb{E}\left(f(X_t)|X_0 = x\right) = \int P_t(x, dy) f(y) = \int p(t, x, y) f(y) dy.$$

It is clear that $P_t$ is a functional operator,

$$P_t: \; f \to P_t f.$$

**Chapman-Kolmogorov Equation.** For any homogeneous Markov process, we have

$$P_{s+t}(x, A) = \int P_s(x, dy) P_t(y, A). \tag{5.1}$$

Proof:

$$\begin{aligned} P_{s+t}(x, A) &= \mathbb{P}\left(X_{t+s} \in A | X_0 = x\right) \\ &= \mathbb{E}\left[\mathbb{P}\left(X_{t+s} \in A | X_s\right) | X_0 = x\right] \\ &= \mathbb{E}\left[f(X_s) | X_0 = x\right] \quad \hookleftarrow \; f(y) = \mathbb{P}\left(X_{t+s} \in A | X_s = y\right) = P_t(y, A) \end{aligned}$$

$$
\begin{aligned}
&= P_s f \\
&= \int P_s(x,dy)f(y) \\
&= \int P_s(x,dy)P_t(y,A).
\end{aligned}
$$

In terms of conditional moments, we have for any positive measurable function $f$,

$$P_{s+t}f = P_s P_t f. \tag{5.2}$$

Proof:

$$
\begin{aligned}
P_{t+s}f(x) &= \int P_{s+t}(x,dz)f(z) \\
&= \int\int P_s(x,dy)P_t(y,dz)f(z) \\
&= \int P_s(x,dy)\int P_t(y,dz)f(z) \\
&= (P_s P_t f)(x).
\end{aligned}
$$

Set $f = 1_A$, this becomes the Chapman-Kolmogorov Equation (5.1).

**Infinitestimal Generator.** We may denote $P_0 \equiv$ identity operator. Define

$$Af(x) = \lim_{t\to 0}\frac{P_t f(x) - P_0 f(x)}{t} = \lim_{t\to 0}\frac{P_t f(x) - f(x)}{t}. \tag{5.3}$$

If $X_t$ follows the following diffusion,

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

We have

$$f(X_t) = f(X_0) + \int_0^t (\mu f' + \frac{1}{2}\sigma^2 f'')(X_s)ds + \int_0^t (\sigma f')(X_s)dW_s.$$

Hence

$$P_t f(x) = f(x) + (\mu f' + \frac{1}{2}\sigma^2 f'')(x)t + O(t^2).$$

So

$$Af = \lim_{t\to 0}\frac{P_t f - f}{t} = \mu f' + \frac{1}{2}\sigma^2 f''.$$

Using this infinitestimal generator, we can write $P_t f(x) = \mathbb{E}(f(X_t)|X_0 = x)$ in the form of a Taylor series expansion,

$$P_t f(x) = f(x) + Af(x)t + \frac{1}{2}A^2 f(x)t^2 + \cdots + \frac{1}{j!}A^j f(x)t^j + O(t^{j+1}). \tag{5.4}$$

**Kolmogrov Forward and Backward Equations.** We have

$$\begin{aligned}\frac{d}{dt}P_t f = \lim_{s\to 0}\frac{P_{t+s}f - P_t f}{s} &= \lim_{s\to 0}\frac{P_t(P_s f - f)}{s} &= P_t A f \qquad (5.5)\\ &= \lim_{s\to 0}\frac{P_s(P_t f) - (P_t f)}{s} &= A P_t f \qquad (5.6)\end{aligned}$$

(5.5) and (5.6) are Kolmogrov forward and backward equations, respectively. The above also proves that $P_t$ and $A$ commutes.

Note that

$$\begin{aligned}P_t A f(x) &= \int (Af)(y)p(t,x,y)dy = \int (\mu f' + \frac{1}{2}\sigma^2 f'')(y)p(t,x,y)dy\\ &= -\int f(y)\frac{\partial}{\partial y}(\mu(y)p(t,x,y))dy + \int f(y)\frac{\partial^2}{\partial y^2}(\frac{1}{2}\sigma^2(y)p(t,x,y))dy.\end{aligned}$$

Since $P_t f(x) = \int f(y)p(t,x,y)dy$,

$$\frac{d}{dt}P_t f(x) = \int f(y)\frac{\partial}{\partial t}p(t,x,y).$$

So (5.5) results in

$$\frac{\partial}{\partial t}p(t,x,y) = -\frac{\partial}{\partial y}(\mu(y)p(t,x,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(t,x,y)), \qquad (5.7)$$

which is the more common form of the Kolmogrov forward equation. Similarly,

$$\begin{aligned}AP_t f(x) &= A\int f(y)p(t,x,y)dy\\ &= (\mu(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2})\int f(y)p(t,x,y)dy\\ &= \int f(y)\left(\mu(x)\frac{\partial}{\partial x}p(t,x,y) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}p(t,x,y)\right)dy.\end{aligned}$$

Hence

$$\frac{\partial}{\partial t}p(t,x,y) = \mu(x)\frac{\partial}{\partial x}p(t,x,y) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}p(t,x,y), \qquad (5.8)$$

which is the backward Kolmogrov equation.

## 5.2 MLE of Parametric Diffusion Models

Suppose the data generating process is parametric,

$$dX_t = \mu(X_t,\theta_0) + \sigma(X_t,\theta_0)dW_t,$$

where $\theta_0$ is a parameter vector.

We observe $X_t$ at evenly spaced time points, $\Delta, 2\Delta, ..., n\Delta \equiv T$. From these observations we want to estimate $\theta_0$.

### 5.2.1 Simple Cases

When $p(t, x, y)$ has closed-form expression (GBM, Ornstein-Uhlenbeck, CIR), we can easily form the log likelihood function as

$$L = \sum_{i=1}^{n} l(\Delta, X_{(i-1)\Delta}, X_{i\Delta}),$$

where

$$l(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) = \log p(\Delta, X_{(i-1)\Delta}, X_{i\Delta}).$$

Here we may safely ignore the log likelihood function of $X_0$.

### 5.2.2 Naive MLE

When $p(t, x, y)$ does not have a closed-form expression, we may apply MLE to the Euler approximation of the original diffusion,

$$X_{i\Delta} = X_{(i-1)\Delta} + \mu(X_{(i-1)\Delta}, \theta_0)\Delta + \sigma(X_{(i-1)\Delta}, \theta_0)Z_i,$$

where $(Z_i)$ are a sequence of independent $N(0, \Delta)$ random variables.

### 5.2.3 Exact MLE

We may also obtain $p(\Delta, X_{(i-1)\Delta}, X_{i\Delta})$ by solving Kolmogrov's forward or backward equation numerically. A boundary problem for the forward equation can be specified as

$$\frac{\partial}{\partial t} p(t, x, y) = -\frac{\partial}{\partial y}(\mu(y)p(t, x, y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(t, x, y)),$$

with

$$\begin{aligned} p(0, x, y) &= \delta(x - y) \\ p(t, x, \infty) &= p(t, x, -\infty) = 0. \end{aligned}$$

For each $i$, let $x = X_{(i-1)\Delta}$, we solve for $p(\Delta, X_{(i-1)\Delta}, X_{i\Delta})$. For more details, see Lo (1988).

### 5.2.4 Approximate MLE

The idea is to construct a closed-form sequence of approximations to $p(\Delta, X_{(i-1)\Delta}, z)$: $p^{(J)}(\Delta, X_{(i-1)\Delta}, z)$, $J = 1, 2, 3, ...$ As $J \to \infty$, $p^{(J)} \to p$.

Ait-Sahalia (2002) first transform $X_t$ into a process $Z_t$ whose transition density $p_Z$ is close to $N(0,1)$, making possible an expansion of $p_Z$ around $N(0,1)$. This involves two steps.

(1) Transform $X_t$ into $Y_t$ by

$$Y_t = \int^{X_t} \frac{ds}{\sigma(X_s)} =: \gamma(X_t).$$

$\gamma$ is obviously increasing and hence invertible. Using Ito's formula,

$$d\gamma(X_t) = \left( \frac{\mu(X_t)}{\sigma(X_t)} - \frac{1}{2}\sigma'(X_t) \right) dt + dW_t.$$

Hence

$$dY_t = \mu_Y(Y_t)dt + dW_t,$$

where

$$\mu_Y(Y_t) = \frac{\mu(\gamma^{-1}(Y_t))}{\sigma(\gamma^{-1}(Y_t))} - \frac{1}{2}\sigma'(\gamma^{-1}(Y_t)).$$

(2) Transform $Y_t$ into $Z_t$ by

$$Z_t = \Delta^{-1/2}(Y_t - y_0).$$

Now define the approximation to $p_Z(\Delta, z_0, z)$ as

$$p_Z^{(J)}(\Delta, z_0, z) = \phi(z) \sum_{j=0}^{J} \xi^{(j)} H_j(z),$$

where $\phi$ is the density function for standard normal distribution and $(H_j(z))$ are Hermite polynomials:

$$H_j(z) = e^{z^2/2} \frac{d^j}{dz^j} \left( e^{-z^2/2} \right), \ j \geq 0.$$

$\xi^{(j)}$ satisfies

$$\xi^{(j)} = \frac{1}{j!} \int H_j(z) p_Z(\Delta, z_0, z) dz$$

$$\begin{aligned}
&= \frac{1}{j!}\int H_j(z)\Delta^{1/2}p_Y(\Delta, y_0, \Delta^{1/2}z + y_0)dz \\
&= \frac{1}{j!}\int H_j(\Delta^{-1/2}(y - y_0))p_Y(\Delta, y_0, y)dy \\
&= \frac{1}{j!}\mathbb{E}\left(H_j(\Delta^{-1/2}(Y_\Delta - y_0))|Y_0 = y_0\right). \qquad (5.9)
\end{aligned}$$

Note that $p_Y(\Delta, y_0, y) = \Delta^{-1/2}p_Z(\Delta, z_0, \Delta^{-1/2}(y-y_0))$. Now let $f(y) = H_j(\Delta^{-1/2}(y-y_0))$. (5.9) reduces to $P_\Delta f(y_0)$, which allows Taylor-type expansion,

$$P_\Delta f(y_0) = f(y_0) + \sum_{k=1}^{K}\frac{1}{k!}(A^k f)(y_0)\Delta^k + O(\Delta^{K+1}).$$

We choose the orders of approximation $J$ and $K$. Then $\xi$ and thus $p_Z^{(J)}(\Delta, z_0, z)$ can be explicitly calculated. We then transform $p_Z^{(J)}(\Delta, z_0, z)$ back to $p_X^{(J)}(\Delta, x_0, x)$, which is an approximation of $p_X(\Delta, x_0, x)$.

## 5.3 GMM of Parametric Diffusions

### 5.3.1 Naive GMM

We have

$$dX_t = \mu(X_t, \theta_0) + \sigma(X_t, \theta_0)dW_t.$$

By Euler approximation,

$$X_{t+\Delta} \approx X_t + \mu(X_t, \theta_0)\Delta + \sigma(X_t, \theta_0)(W_{t+\Delta} - W_t).$$

Let $\varepsilon_{t+\Delta} = X_{t+\Delta} - X_t + \mu(X_t, \theta_0)\Delta$, we have

$$\begin{aligned}
\mathbb{E}(\varepsilon_{t+\Delta}|X_t) &= 0 \\
\mathbb{E}(\varepsilon_{t+\Delta}^2|X_t) &= \sigma^2(X_t, \theta_0)\Delta.
\end{aligned}$$

So we at least have four moment conditions:

$$\begin{aligned}
\mathbb{E}(\varepsilon_{t+\Delta}) &= 0 \\
\mathbb{E}(\varepsilon_{t+\Delta}X_t) &= 0 \\
\mathbb{E}(\varepsilon_{t+\Delta}^2 - \sigma^2(X_t, \theta_0)\Delta) &= 0 \\
\mathbb{E}[(\varepsilon_{t+\Delta}^2 - \sigma^2(X_t, \theta_0)\Delta)X_t] &= 0
\end{aligned}$$

For more details, see Chan et. al. (1992).

### 5.3.2 Simulated Moment Estimation

The idea is to use simulation to generate simulated moments, which are matched with sample moments.

The sample moment is simply

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_{i\Delta}).$$

For each choice of parameter vector $\theta$, we simulate a sequence of $X_{b\Delta}^{\theta}$, $b = 1, 2, ..., B$, where $B$ is a large number. The simulated moments are thus

$$\tilde{G}(\theta) = \frac{1}{B}\sum_{b=1}^{B} f(X_{b\Delta}^{\theta}).$$

Let

$$G_n(\theta) = \tilde{G}(\theta) - \hat{G}_n.$$

The GMM estimator is given as

$$\hat{\theta}_n = \text{argmin}_{\theta} G_n'(\theta) W_n G_n'(\theta),$$

where $W_n$ is an appropriate distance matrix. See Gallant and Tauchen (1996) for more details.

The key assumption for the above strategy to work is that $X_t$ is geometrically ergodic. Geometrical Ergodicity means that for some $\rho \in (0, 1)$, there is a probability measure $P$ such that for any initial point $x$,

$$\rho^{-t}\|P_t(x, \cdot) - P\|_v \to 0 \quad \text{as } t \to \infty,$$

where $\| \cdot \|_v$ is the total variation norm defined as

$$\|u\|_v = \sup_A |u(A)|.$$

### 5.3.3 Exact GMM

If we assume that $X_t$ is stationary, then $\mathbb{E}f(X_t)$ does not depend on $t$. This leads to

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}f(X_t) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\left(\mathbb{E}f(X_{t+\Delta}) - \mathbb{E}f(X_t)\right) \\
&= \mathbb{E}\left[\lim_{\Delta\to 0}\frac{1}{\Delta}\left(\mathbb{E}f(X_{t+\Delta}) - \mathbb{E}f(X_t)\right)|X_t\right] \\
&= \mathbb{E}\left[\lim_{\Delta\to 0}\frac{1}{\Delta}\left(\mathbb{E}f(X_\Delta) - \mathbb{E}f(X_0)|X_0 = X_t\right)\right] \\
&= \mathbb{E}Af(X_t) = 0 \qquad (5.10)
\end{aligned}$$

(5.10) holds for any measurable function and may serve enough number of moment conditions for GMM.

We can find more moment conditions. Define

$$P_t^* f^*(y) = \mathbb{E}\left(f^*(X_0)|X_t = y.\right),$$

where $f^*$ is any measurable function. Obviously $P_0^* f^* = f^*$. And we define the backward infinitestimal generator

$$A^* f^* = \lim_{t\to 0} \frac{P_t^* f^* - f^*}{t}.$$

$P_t^*$ is the adjoint of $P_t$. To see this,

$$\begin{aligned}
\langle f^*(X_0), P_t f(X_0)\rangle &\equiv \mathbb{E}\left[f^*(X_0)\mathbb{E}(f(X_t)|X_0)\right] \\
&= \mathbb{E}\left[f^*(X_0) f(X_t)\right] \\
&= \mathbb{E}\left[\mathbb{E}(f(X_0)|X_t) f(X_t)\right] \\
&= \langle P_t^* f^*(X_t), f(X_t)\rangle \\
&= \langle P_t^* f^*(X_0), f(X_0)\rangle
\end{aligned}$$

The last equality uses the stationarity of $X_t$. We can also show that $A^*$ is the adjoint of $A$. Then we have

$$\langle P_t A f(X_0), f^*(X_0)\rangle = \langle A P_t f(X_0), f^*(X_0)\rangle = \langle f(X_0), P_t^* A^* f^*(X_0)\rangle.$$

The inner product on the left,

$$\langle P_t A f(X_0), f^*(X_0)\rangle = \mathbb{E}[\mathbb{E}(Af(X_t)|X_0) f^*(X_0)] = \mathbb{E}[Af(X_t) f^*(X_0)].$$

The inner product on the right,

$$\begin{aligned}
\langle f(X_0), P_t^* A^* f^*(X_0)\rangle &= \langle f(X_t), P_t^* A^* f^*(X_t)\rangle \\
&= \mathbb{E}[f(X_t)\mathbb{E}(A^* f^*(X_0)|X_t)] \\
&= \mathbb{E}[f(X_t) A^* f^*(X_0)].
\end{aligned}$$

Hence

$$\mathbb{E}[Af(X_t) f^*(X_{t-\Delta}) - f(X_t) A^* f^*(X_{t-\Delta})] = 0. \tag{5.11}$$

(5.11) offer more choices of moment conditions for GMM. In particular, if $f^*$ is a constant function, (5.11) reduces to (5.10). For more details, Hansen and Scheinkman (1995).

### 5.3.4 Eigen GMM

Consider the infinitestimal generator $A$ of $X_t$. It is well known from the spectral theory of diffusion processes that for many diffusions, the set of eigenvalues (spectrum) $\Lambda_\theta$ for $A$ are positive and discrete. So $\Lambda_\theta$ can be written as $(\lambda_1, \lambda_2, ..., \lambda_n, ...)$, where $0 \leq \lambda_1 < \lambda_2 < ... < \lambda_n < ....$

Let $(\lambda, \phi)$ be any eigen-pair of $A$. We have

$$A\phi = -\lambda\phi.$$

Then

$$\begin{aligned}
\frac{dP_t\phi}{dt} &= \lim_{\Delta\to 0} \frac{P_{t+\Delta}\phi - P_t\phi}{\Delta} \\
&= P_t A\phi = -\lambda P_t\phi.
\end{aligned}$$

This is ordinary differential equation on $P_t\phi$. It is well known that

$$P_t\phi = e^{-\lambda t}\phi.$$

Now apply Ito's formula to $e^{\lambda t}\phi(X_t)$,

$$\begin{aligned}
de^{\lambda t}\phi(X_t) &= \lambda e^{\lambda t}\phi(X_t)dt + e^{\lambda t}\phi'(X_t)dX_t + \frac{1}{2}e^{\lambda t}\phi''(X_t)d[X]_t \\
&= \left(\lambda e^{\lambda t}\phi(X_t) + e^{\lambda t}\phi'(X_t)\mu(X_t) + \frac{1}{2}e^{\lambda t}\phi''(X_t)\sigma^2(X_t)\right) dt \\
&\quad + e^{\lambda t}\phi'(X_t)\sigma(X_t)dW_t \\
&= e^{\lambda t}(\lambda\phi(X_t) + A\phi(X_t))dt + e^{\lambda t}\phi'(X_t)\sigma(X_t)dW_t \\
&= e^{\lambda t}\phi'(X_t)\sigma(X_t)dW_t.
\end{aligned}$$

Hence,

$$\phi(X_t) = e^{-\lambda t}\phi(X_0) + \int_0^t e^{-\lambda(t-s)}\phi'(X_s)\sigma(X_s)dW_s.$$

So

$$\mathbb{E}(\phi(X_{t+\Delta})|X_t) = e^{-\lambda\Delta}\phi(X_t),$$

which leads to desired moment condition,

$$E\left[\left(\phi(X_{t+\Delta}) - e^{\lambda\Delta}\phi(X_t)\right) g(X_t)\right] = 0, \tag{5.12}$$

where $g$ can be any measurable function.

## 5.4 Estimation of Nonparametric Models

The methodology of MLE and GMM presupposes correct parameterization of the diffusion models (or equivalently the infinitestimal generator). Any misspecification leads to inconsistency of those estimators. The problem of parameterization can be avoided by the use of nonparametric diffusion models, the time-homogeneous version of which is given as,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t.$$

Ignoring terms in (5.4) that are of smaller order than $O(\Delta)$ , we obtain

$$\mathbb{E}[f(X_{t+\Delta})|X_t = x] = f(x) + Af(x)\Delta + O(\Delta^2),$$

where $Af(x) = \mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$. Then we have

$$Af(x) = \frac{1}{\Delta}\mathbb{E}[(f(X_{t+\Delta}) - f(X_t))|X_t = x] + O(\Delta).$$

Let $f$ be such that $f(x) = x$. Then we have

$$Af(x) = \mu(x) = \frac{1}{\Delta}\mathbb{E}[(X_{t+\Delta} - X_t)|X_t = x] + O(\Delta).$$

Let $f(x) = (x - X_t)^2$. We have $Af(x) = 2\mu(x)(x - X_t) + \sigma^2(x)$, and

$$\sigma^2(x) = \frac{1}{\Delta}\mathbb{E}[(X_{t+\Delta} - X_t)^2|X_t = x] + O(\Delta).$$

We can estimate $\mathbb{E}[(X_{t+\Delta} - X_t)|X_t = x]$ and $\mathbb{E}[(X_{t+\Delta} - X_t)^2|X_t = x]$ using Nadaraya-Watson kernel estimator.

$$\begin{aligned}
\hat{\mu}(x) &= \frac{\sum_{i=1}^n K(\frac{X_{(i-1)\Delta}-x}{h})(X_{i\Delta} - X_{(i-1)\Delta})}{\Delta\sum_{i=1}^n K(\frac{X_{(i-1)\Delta}-x}{h})} \\
\hat{\sigma}^2(x) &= \frac{\sum_{i=1}^n K(\frac{X_{(i-1)\Delta}-x}{h})(X_{i\Delta} - X_{(i-1)\Delta})^2}{\Delta\sum_{i=1}^n K(\frac{X_{(i-1)\Delta}-x}{h})}.
\end{aligned}$$

The Nadaraya-Watson estimators are consistent. To see this, first note that

$$\begin{aligned}
\frac{1}{nh}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta}}{h}\right) &\approx \frac{1}{Th}\int_0^T K\left(\frac{X_s - x}{h}\right) ds \\
&= \frac{1}{Th}\int_{-\infty}^{\infty} K\left(\frac{s - x}{h}\right) L(T, s)ds
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\int_{-\infty}^{\infty} K\left(s\right) L(T, x+hs)ds \\
&\to \frac{1}{T}\int_{-\infty}^{\infty} K\left(s\right) L(T, x)ds \\
&= \frac{1}{T}L(T,x),
\end{aligned}
$$

where $L(T, x)$ denotes the local time of $X_t$, and the derivation uses the Occupation Time Formula,

$$\int_0^t f(X_s)ds = \int_{-\infty}^{\infty} f(x)L(T,x)dx.$$

And

$$
\begin{aligned}
& \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)\left(\frac{X_{i\Delta}-X_{(i-1)\Delta}}{\Delta}\right) \\
\approx\ & \frac{1}{Th}\sum K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)\mu(X_{(i-1)\Delta})\Delta \\
\approx\ & \frac{1}{Th}\int_0^t K\left(\frac{X_s-x}{h}\right)\mu(X_s)ds \\
=\ & \frac{1}{Th}\int_{-\infty}^{\infty} K\left(\frac{s-x}{h}\right)\mu(s)L(T,s)ds \\
\to\ & \frac{1}{T}\mu(x)L(T,x).
\end{aligned}
$$

See Jiang and Knight (1997) and Bandi and Phillips (2003).

We may obtain better precision by keeping more terms in (5.4). For example, we have

$$\mathbb{E}[f(X_{t+\Delta})|X_t = x] = f(x) + Af(x)\Delta + \frac{1}{2}A^2f(x)\Delta^2 + O(\Delta^3), \tag{5.13}$$

and

$$\mathbb{E}[f(X_{t+2\Delta})|X_t = x] = f(x) + Af(x)2\Delta + \frac{1}{2}A^2f(x)4\Delta^2 + O(\Delta^3). \tag{5.14}$$

4(5.13)-(5.14) would give us

$$Af(x) = \frac{1}{2\Delta}\{4\mathbb{E}[f(X_{t+\Delta}) - f(X_t)|X_t = x] - \mathbb{E}[f(X_{t+2\Delta}) - f(X_t)|X_t = x]\} + O(\Delta^2).$$

More precise estimators of $\mu$ and $\sigma^2$ (in the order of $\Delta^2$) then follow.

## 5.5 Semiparametric Models

We consider stationary diffusions satisfying either

$$dX_t = \mu(X_t)dt + \sigma(X_t, \theta)dW_t, \tag{5.15}$$

or

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t)dW_t. \tag{5.16}$$

From Kolmogrov Forward Equation, we have

$$\begin{aligned}\frac{\partial}{\partial \Delta} p(\Delta, x, y) &= -\frac{\partial}{\partial y}(\mu(y)p(\Delta, x, y)) \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(\Delta, x, y)).\end{aligned}$$

By stationarity, the density of marginal distribution is time-invariant, ie,

$$\begin{aligned}\frac{\partial}{\partial \Delta}\pi(y) &= \frac{\partial}{\partial \Delta}\int_0^\infty p(\Delta, x, y)\pi(x)dx \\ &= \int_0^\infty \frac{\partial}{\partial \Delta} p(\Delta, x, y)\pi(x)dx \\ &= \int_0^\infty \left(-\frac{\partial}{\partial y}(\mu(y)p(\Delta, x, y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(\Delta, x, y))\right)\pi(x)dx \\ &= -\frac{\partial}{\partial y}\left(\mu(y)\left(\int_0^\infty p(\Delta, x, y)\pi(x)dx\right)\right) \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(X_{t+\Delta})\left(\int_0^\infty p(\Delta, x, y)\pi(x)dx\right)\right) \\ &= -\frac{\partial}{\partial y}\left(\mu(y)\pi(y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)\pi(y)\right) \\ &= 0.\end{aligned}$$

Hence

$$\frac{d^2}{x^2}(\sigma^2(x)\pi(x)) = 2\frac{d}{dx}(\mu(x)\pi(x)). \tag{5.17}$$

Suppose $\pi(0) = 0$, we have

$$\mu(x) = \frac{1}{2\pi(x)}\frac{d}{dx}(\sigma^2(x)\pi(x)), \tag{5.18}$$

and

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_0^x \mu(u)\pi(u)du. \tag{5.19}$$

For model (5.15), in which we have prior knowledge on the structure of $\sigma^2$, we may parametrically estimate $\sigma^2$ and nonparametrically estimate $\mu$ using (5.18). For model (5.16), we may similarly estimate $\sigma^2$ nonparametrically using (5.19), with prior knowledge of $\mu$. See Ait-Sahalia (1996) for more details.

# References

Ait-Sahalia, Y., 1996, Nonparametric Pricing of Interest Rate Derivative Securities, Econometrica, 64, 527-560

Ait-Sahalia, Y., 2002, Maximum likelihood estimation of discretely sampled diffusions: A closed-form approach, Econometrica, 70, 223-262

Bandi, F. M. and Phillips, P. C., 2003, Fully Nonparametric Estimation of Scalar Diffusion Models Econometrica, 71, 241-283

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 48, 1209-1227.

Gallant, A. and Tauchen, G., 1996, Which Moments to Match? Econometric Theory, 12, 657-681

Hansen, L. P. and Scheinkman, J. A., 1995, Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes, Econometrica, 63, 767-804

Jiang, G. J. and Knight, J. L., 1997, A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model Econometric Theory, 13, 615-645

Kessler, M. and Sorensen, M., 1999, Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process Bernoulli, 5, 299-314

Lo, A. W., 1988, Maximum Likelihood Estimation of Generalized It? Processes with Discretely Sampled Data, Econometric Theory, 4, 231-247

# Appendix A

# Appendix to Chapter 1

## A.1 Classical Derivation of CAPM

### A.1.1 Efficiency Frontier without Riskfree Asset

Suppose there are $N$ risky assets, and the return vector of these assets has a mean of $\mu$ and a covariance matrix $\mathbf{\Sigma}$. Each agent selects a portfolio of these assets $h$, where $h'\iota = 1$ and $\iota$ is a vector of ones. Thus the portfolio return has mean $\mu_h = h'\mu$ and variance $\sigma_h^2 = h'\Sigma h$. We assume that all agents in the economy are identical with utility function $u(\mu_h, \sigma_h^2)$. It is understood that $u(\cdot,\cdot)$ is increasing in $\mu_h$ and decreasing in $\sigma_h^2$.

Given an objective mean return of portfolio, agents try to find a portfolio that minimizes the variance. Mathematically, the following problem is to be solved,

$$\min_h \frac{1}{2} h'\Sigma h,$$

subject to

$$h'\mu = \mu_h \quad \text{and} \quad h'\iota = 1.$$

The Lagrangian function is given by

$$L = \frac{1}{2} h'\Sigma h + \lambda_1 (h'\mu - \mu_h) + \lambda_2 (h'\iota - 1).$$

The first order conditions are

$$\Sigma h = \lambda_1 \mu + \lambda_2 \iota, \tag{A.1}$$
$$\mu' h = \mu_h, \tag{A.2}$$
$$\iota' h = 1. \tag{A.3}$$

(A.1) yields

$$h = \Sigma^{-1}(\lambda_1 \mu + \lambda_2 \iota). \tag{A.4}$$

Pre-multiply (A.4) with $\mu'$ and $\iota'$, respectively, we obtain

$$\begin{aligned} A\lambda_1 + B\lambda_2 &= \mu_h \\ B\lambda_1 + C\lambda_2 &= 1, \end{aligned}$$

where

$$A = \mu'\Sigma^{-1}\mu, \quad B = \mu'\Sigma^{-1}\iota, \quad C = \iota'\Sigma^{-1}\iota.$$

Define $D = AC - B^2$. We obtain

$$\lambda_1 = \frac{1}{D}(C\mu_h - B), \quad \lambda_2 = \frac{1}{D}(-B\mu_h + A).$$

Plug in (A.4), we obtain

$$h = g_0 + g_1 \mu_h, \tag{A.5}$$

where

$$g_0 = \frac{1}{D}(A\Sigma^{-1}\iota - B\Sigma^{-1}\mu), \quad g_1 = \frac{1}{D}(C\Sigma^{-1}\mu - B\Sigma^{-1}\iota).$$

The variance of the return on $h$ is given by

$$\sigma_h^2 = h'\Sigma h = g_0'\Sigma g_0 + 2\mu_h g_0'\Sigma g_1 + \mu_h^2 g_1'\Sigma g_1.$$

Hence the pairs $(\mu_h, \sigma_h)$ trace a hyperbola boundary, the upper boundary of which is called the *efficiency frontier*.

Note that the minimum-variance portfolio $h$ (A.5) is linear in $\mu_h$. If we know two minimum-variance portfolios $h_1$ and $h_2$ with mean returns $\mu_1$ and $\mu_2$, respectively, then we know all minimum-variance portfolios. Indeed, for all expected return $\mu_a$, the corresponding minimum-variance portfolio can be constructed by $h_a = \alpha h_1 + (1-\alpha)h_2$, where $\alpha$ is obtained by solving $\mu_a = \alpha\mu_1 + (1-\alpha)\mu_2$. This observation is often called the *two mutual fund theorem*.

### A.1.2 CAPM

Suppose there is a money account with a risk-free return of $r$. Now the agents' problem becomes

$$\min_h \frac{1}{2} h'\Sigma h,$$

subject to

$$\mu' h + (1 - \iota' h) r = \mu_h.$$

The Lagrangian is

$$L = \frac{1}{2}h'\Sigma h + \lambda(\mu_h - \mu' h - (1 - \iota' h)r).$$

The first-order conditions are

$$\begin{aligned} \Sigma h &= \lambda(\mu - r\iota) \\ \mu' h + (1 - \iota' h)r &= \mu_h. \end{aligned}$$

Solving this set of equations, we obtain

$$h = \frac{\mu_h - r}{(\mu - r\iota)'\Sigma^{-1}(\mu - r\iota)} \cdot \Sigma^{-1}(\mu - r\iota). \tag{A.6}$$

Note that this portfolio equals a scalar that depends on $\mu_h$ times a vector that does *not* depend on $\mu_h$. In other words, for all expected return, the exactly same proportion of each risky assets are chosen. We normalize $\Sigma^{-1}(\mu - r\iota)$ to obtain the so-called tangency portfolio,

$$h_m = \frac{\Sigma^{-1}(\mu - r\iota)}{\iota'\Sigma^{-1}(\mu - r\iota)}.$$

The normalization makes the elements in $h_m$ adding up to one. In the idealized world of CAPM, therefore, everyone will choose the tangency portfolio. Individuals differ only in the percentage of cash holding or leverage. In equilibrium, the *market portfolio* must be the tangency portfolio.

Let $R_m$ be the market return. The variance of the market return is given by

$$\mathrm{var}(R_m) = h_m'\Sigma h_m = \frac{(\mu - r\iota)'\Sigma^{-1}(\mu - r\iota)}{\left(\iota'\Sigma^{-1}(\mu - r\iota)\right)^2}.$$

The expected market premium over the risk-free return,

$$\mathbb{E}R_m - r = h_m'\mu - r = \frac{(\mu - r\iota)'\Sigma^{-1}(\mu - r\iota)}{\iota'\Sigma^{-1}(\mu - r\iota)}.$$

Let $R_i$ be the $i$-th asset. We have $\mathbb{E}R_i = e_i'\mu$ and

$$\mathrm{cov}(R_i, R_m) = e_i'\Sigma h_m,$$

where $e_i$ is a vector that has 1 on the $i$-th element and 0 on others. Now we have

$$(\mathbb{E}R_m - r)\frac{\mathrm{cov}(R_i, R_m)}{\mathrm{var}(R_m)} = e_i'(\mu - r\iota) = \mathbb{E}R_i - r.$$

Rearranging terms, we obtain the celebrated CAPM model,

$$\mathbb{E}R_i = r + \beta_i(\mathbb{E}R_m - r), \tag{A.7}$$

where

$$\beta_i = \frac{\text{cov}(R_i, R_m)}{\text{var}(R_m)}. \tag{A.8}$$

In the CAPM model, the expected payoff of a security is a linear function of the security's beta, which characterizes the systematic risk contained in the security. The linear function is called the *security market line* (SML). Obviously, the intercept of the SML is the risk-free rate ($r$) and the slope of the SML is the market risk premium ($\mathbb{E}R_m - r$).