Notes on Math Finance

Haosheng Zhou

Sept, 2022

Contents

Black-Scholes Model

Basic Idea in Option Pricing

The pricing of financial derivatives starts with the BS model where one assumes the stock price to follow the geometric Brownian motion (GBM)

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \tag{1}$$

we assume the risk-free interest rate to be r which is deterministic and time-independent (so are μ, σ). The option is assumed to be of European style with payoff function $h(S_T)$, only depending on the stock price on the maturity date.

The first idea comes from **PDE** that one can build up a self-financing replicating portfolio using stock and bond which has exactly the same payoff as the option. To ensure that there's no arbitrage, any two financial derivatives sharing the same payoff must have the same price, so the price of this replicating portfolio is just the price of the option. To be specific, assume the self-financing replicating portfolio consists of a_t units of stock and b_t units of bond at time t (bond has initial price 1 at time 0) so the value of this portfolio at time t is denoted

$$V_t = a_t S_t + b_t e^{rt} \tag{2}$$

Ito formula tells us that

$$dV_t = a_t \, dS_t + S_t \, da_t + \, d\left\langle S, a\right\rangle_t + r b_t e^{rt} \, dt + e^{rt} \, db_t \tag{3}$$

what does self-financing property mean? It means that the change in the value of the portfolio is completely due to the market. Among all those terms in dV_t, it's clear that the change in the value of the portfolio only has the part $a_t \, dS_t + r b_t e^{rt} \, dt$ completely due to the market. To be specific, $S_t \, da_t + e^{rt} \, db_t$ denotes the change in the value of the portfolio that is due to the change of the position in stock and bond and $d\left\langle S, a\right\rangle_t$ is the cross variation term. In other words, **self-financing portfolio** is defined to satisfy

$$S_t \, da_t + \, d\left\langle S, a\right\rangle_t + e^{rt} \, db_t = 0 \tag{4}$$

so the change in the value of the portfolio can be represented in two terms

$$dV_t = a_t \, dS_t + r b_t e^{rt} \, dt \tag{5}$$

with the **assumption that** $V_t = u(t, S_t)$ **has the Markovian form** with deterministic time-independent feedback function u, we can proceed with Ito formula

$$
\begin{aligned}
dV_t &= \partial_t u\, dt + \partial_x u\, dS_t + \frac{1}{2}\partial_{xx} u\, d\left\langle S, S\right\rangle_t && (6)\\
&= \left(\partial_t u + \mu S_t \partial_x u + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} u\right) dt + \sigma S_t \partial_x u\, dW_t && (7)\\
&\overset{match}{=} a_t\, dS_t + r b_t e^{rt}\, dt && (8)\\
&= (\mu a_t S_t + r b_t e^{rt})\, dt + \sigma a_t S_t\, dW_t && (9)
\end{aligned}
$$

it's clear that the coefficients match

$$
\begin{cases}
\sigma a_t S_t = \sigma S_t \partial_x u\\
\mu a_t S_t + r b_t e^{rt} = \partial_t u + \mu S_t \partial_x u + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} u
\end{cases} \tag{10}
$$

so the replicating portfolio is given by (a_t, b_t) satisfying

$$
\begin{cases}
a_t = \partial_x u(t, S_t)\\
b_t = \frac{\partial_t u(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} u(t,S_t)}{re^{rt}}
\end{cases} \tag{11}
$$

the only problem is to find a PDE characterization for $u = u(t, x)$ to solve for u. Let's recall that the self-financing property requires

$$
a_t S_t + b_t e^{rt} = V_t = u(t, S_t) \tag{12}
$$

plug in a_t, b_t calculated above to get the **BS-PDE** as a characterization for u

$$
\partial_t u + rx\partial_x u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \tag{13}
$$

paired with the natural **terminal condition** that

$$
u(T, x) = h(x) \tag{14}
$$

to match the payoff of the option. As a result, the option price at time 0 is given by $V_0 = u(0, S_0)$.

Remark. *The pros of PDE approach is that we simultaneously get the self-financing replicating portfolio (also called the Delta-hedging strategy) when solving for option price.*

One of the main cons is that in order to solve for one value $u(0, S_0)$, one has to start from the terminal condition and solve $u(t, x)$ for any pair (t, x). In other words, one solve the PDE at all the points just to figure out a single value, which is very inefficient.

The other approach is **risk-neutral pricing**. As is known to all, the arbitrage-free price of a financial derivative is NOT equal to its discounted expected payoff since one has to compensate the buyers for the risk, i.e. extra risk premium is needed such that the buyer is willing to admit risk. However, if an investor is risk-neutral, he has no positive or negative preference on taking any risk thus the price of the derivative is just the expected payoff. How do we create another world that is risk-neutral? We have to make sure that in that world the discounted stock price $\{e^{-rt}S_t\}$ is a martingale! As a result, the drift term in the dynamics of $\{S_t\}$ shall be rS_t and all other extra drift terms shall be absorbed into the diffusion term, leading to the change of measure, an application of Girsanov theorem. Following this motivation, we call the original probability measure (physical measure, i.e. the probability measure of the real world) as $\mathbb{P}$ and the risk-neutral measure as $\mathbb{Q}$, then

$$\begin{aligned} dS_t &= rS_t\,dt + ((\mu - r)S_t\,dt + \sigma S_t\,dW_t) && (15)\\ &= rS_t\,dt + \sigma S_t\left(\frac{\mu - r}{\sigma}\,dt + \,dW_t\right) && (16)\\ &= rS_t\,dt + \sigma S_t\,dW_t^{\mathbb{Q}} && (17)\end{aligned}$$

where $dW_t^{\mathbb{Q}} = \frac{\mu-r}{\sigma}\,dt + dW_t$ is a BM under $\mathbb{Q}$. Clearly, Girsanov theorem enables us to find the connection between $\mathbb{P}$ and $\mathbb{Q}$. Write $L_t = -\frac{\mu-r}{\sigma}W_t$ such that $\langle W, L\rangle_t = -\frac{\mu-r}{\sigma}t$, then

$$W_t^{\mathbb{Q}} = W_t - \langle W, L\rangle_t \tag{18}$$

according to Girsanov,

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{L_T - \frac{1}{2}\langle L,L\rangle_T} = e^{-\frac{\mu-r}{\sigma}W_T - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 T} \tag{19}$$

provides the Radon-Nikodym derivative. It's easy to check that this RN derivative has expectation 1 so it's not only a local MG but also a MG, which satisfies the condition of Girsanov theorem. In this risk-neutral world, the price of the option at time 0 is just the discounted expected payoff

$$e^{-rT}\mathbb{E}_{\mathbb{Q}}[h(S_T)] \tag{20}$$

which concludes the pricing procedure.

Remark. *The pros of risk-neutral pricing is that it simplifies the calculation to a single expectation, which also enables us to use Monte Carlo methods numerically.*

The cons of risk-neutral pricing is that the Delta hedging strategy is hidden behind and requires extra efforts to find.

Remark. *Actually, those two approaches are two sides of the same coin, connected through **Feynman-Kac formula** that*

$$V_t = e^{-r(T-t)}\mathbb{E}_Q[h(S_T)|\mathscr{F}_t] \tag{21}$$

*It's quite obvious that $\{e^{-rt}V_t\}$ is also a MG under $\mathbb{Q}$ so an application of **MG representation theorem** tells* *us that there exists some L^2 adapted process $\{\xi_t\}$ such that*

$$e^{-rt}V_t = V_0 + \int_0^t \xi_s \, dW_s^{\mathbb{Q}} \tag{22}$$

interestingly, let's take derivative w.r.t. t on both sides

$$d(e^{-rt}V_t) = \xi_t \, dW_t^{\mathbb{Q}} = \xi_t dW_t + \frac{\mu - r}{\sigma} \xi_t \, dt \tag{23}$$

with

$$d(e^{-rt}V_t) = -re^{-rt}V_t \, dt + e^{-rt} \, dV_t = -re^{-rt}V_t \, dt + e^{-rt}(a_t \, dS_t + rb_t e^{rt} \, dt) \tag{24}$$

$$= -re^{-rt}V_t \, dt + re^{-rt}a_t S_t \, dt + \sigma e^{-rt} a_t S_t \, dW_t + rb_t \, dt \tag{25}$$

comparing the diffusion coefficient, we know that

$$\sigma e^{-rt} a_t S_t = \xi_t \tag{26}$$

so $\{\xi_t\}$ has a close connection with the option Delta $a_t = \partial_x u(t, S_t)$.

B-L Formula

For European call option with payoff function $h(S_T) = (S_T - K)_+$, it turns out that we can recover the density of the stock price using the information on the option price $C(T, K)$, which is defined as the price of the European call option at time 0 with time to maturity T and strike K.

Assume that the density of stock price S_t under risk-neutral measure $\mathbb{Q}$ is denoted as $p(t, \cdot)$, then from risk-neutral pricing,

$$C(T, K) = e^{-rT}\mathbb{E}_{\mathbb{Q}}(S_T - K)_+ = e^{-rT}\int_K^{\infty}(x - K)\cdot p(T, x)\,dx \tag{27}$$

take derivative w.r.t. K (assume it's admissible),

$$\begin{aligned}\frac{\partial C}{\partial K} &= e^{-rT}\left[-Kp(T, K) - \int_K^{\infty} p(T, x)\,dx + Kp(T, K)\right] &(28)\\ &= -e^{-rT}\int_K^{\infty} p(T, x)\,dx &(29)\end{aligned}$$

take again derivative w.r.t. K,

$$\frac{\partial^2 C}{\partial K^2} = e^{-rT}p(T, K) \tag{30}$$

rename variable K as x, then we have the **B-L formula**

$$p(T, x) = e^{rT}\frac{\partial^2 C}{\partial K^2}(T, x) \tag{31}$$

showing us how to derive the density function of S_T under $\mathbb{Q}$. Of course, this formula is restrictive in practice since it requires us to know a continuum of European call price w.r.t. different strike K. However, it is not model specific and will help us build up more useful results in a later context.

Carr-Madan Formula

Carr-Madan formula provides an equation for the price of a European-style call option with any payoff function $h(S_T)$. Let's denote $C_h(T,K)$ as the price of this option at time 0 with payoff h, time to maturity T and strike K. Denote $C(T,K)$ as the price of European call option at time 0 with payoff $h(S_T) = (S_T - K)_+$, time to maturity T and strike K.

It turns out that if the density of S_T under $\mathbb{Q}$ is assumed to exist,

$$C_h(T,K) = e^{-rT}\mathbb{E}_Q[h(S_T)] \tag{32}$$

$$= e^{-rT}\int_0^\infty h(x)\cdot p(T,x)\,dx \tag{33}$$

plug in the B-L formula for $p(T,x)$ to get

$$C_h(T,K) = \int_0^\infty h(x)\cdot \frac{\partial^2 C}{\partial K^2}(T,x)\,dx \tag{34}$$

$$= h(x)\frac{\partial C}{\partial K}(T,x)\Big|_{x=0}^\infty - \int_0^\infty h'(x)\cdot \frac{\partial C}{\partial K}(T,x)\,dx \tag{35}$$

notice that $\frac{\partial C}{\partial K}(T,K) = -e^{-rT}\int_K^\infty p(T,x)\,dx$ from the derivation of B-L formula,

$$C_h(T,K) = h(0)e^{-rT} - \int_0^\infty h'(x)\cdot \frac{\partial C}{\partial K}(T,x)\,dx \tag{36}$$

$$= h(0)e^{-rT} - h'(x)C(T,x)\Big|_{x=0}^\infty + \int_0^\infty h''(x)\cdot C(T,x)\,dx \tag{37}$$

notice that $C(T,\infty) = 0$ since one never exercise the call option if the strike price is very high, and $C(T,0) = S_0$ since call option with zero strike has payoff S_T, the same as a forward contract,

$$C_h(T,K) = h(0)e^{-rT} + h'(0)S_0 + \int_0^\infty h''(x)\cdot C(T,x)\,dx \tag{38}$$

gives the **Carr-Madan formula**. Be careful here that C_h and C are the prices of possibly different European style options, depending on the payoff function h.

Remark. *Carr-Madan formula implies put-call parity. To see this, take $h(x) = (K-x)_+$ so that $C_h(T,K) = P(T,K)$ is the price of the European put. Plug in Carr-Madan formula, $h(0) = K, h'(x) = -\mathbb{I}_{x<K}, h''(x) = \delta_{\{K\}}$ where h', h'' are in the sense of weak derivative and h'' is a Dirac point mass.*

$$P(T,K) = Ke^{-rT} - S_0 + C(T,K) \tag{39}$$

is exactly put-call parity.

Obviously, Carr-Madan requires the existence of $h(0), h'(0)$, which is sometimes a too strong condition for some of the options we are interested in, e.g. options with log payoff $h(x) = \log x$. In this case, we have to use the other form of Carr-Madan formula as an extension.

Lemma 1. *For any $x, x_0 > 0$, the equation holds*

$$G(x) = G(x_0) + G'(x_0)(x - x_0) + \int_{x_0}^{\infty} G''(K)(x-K)_+ \, dK + \int_0^{x_0} G''(K)(K-x)_+ \, dK \tag{40}$$

Proof. Split into case $x > x_0, x = x_0, x < x_0$, we only prove for $x > x_0$ since the proof is similar for other cases.

$$RHS = G(x_0) + G'(x_0)(x - x_0) + \int_{x_0}^{x} (x-K)G''(K) \, dK \tag{41}$$

$$= G(x_0) + G'(x_0)(x - x_0) + x[G'(x) - G'(x_0)] - \int_{x_0}^{x} KG''(K) \, dK \tag{42}$$

$$= G(x_0) - x_0G'(x_0) + xG'(x) - xG'(x) + x_0G'(x_0) + \int_{x_0}^{x} G'(K) \, dK \tag{43}$$

$$= G(x_0) + \int_{x_0}^{x} G'(K) \, dK \tag{44}$$

$$= G(x) \tag{45}$$

concludes the proof. □

The smart point here is to not only introduce the European call but also introduce the European put. Apply the lemma with $x = S_T, G(x) = \log x, x_0 = S_0$ specified,

$$\log S_T = \log S_0 + \frac{1}{S_0}(S_T - S_0) + \int_{S_0}^{\infty} \left(-\frac{1}{K^2}\right)(S_T - K)_+ \, dK + \int_0^{S_0} \left(-\frac{1}{K^2}\right)(K - S_T)_+ \, dK \tag{46}$$

take expectation under $\mathbb{Q}$ and discount on both sides

$$e^{-rT}\mathbb{E}_{\mathbb{Q}} \log S_T = e^{-rT} \log S_0 + e^{-rT} \frac{\mathbb{E}_Q S_T - S_0}{S_0} + \int_{S_0}^{\infty} \left(-\frac{1}{K^2}\right) C(T,K) \, dK + \int_0^{S_0} \left(-\frac{1}{K^2}\right) P(T,K) \, dK \tag{47}$$

is the **extension of Carr-Madan** that connects the BS formula with the price of the European option with log payoff. At this point, we see the importance of Carr-Madan that it's telling us **any European style option with twice differentiable payoff function can be replicated using European call and put**.

Remark. *Under BS model, the* ***log contract*** *has payoff $h(S_T) = \log \frac{S_T}{S_0}$ so it has price*

$$e^{-rT}\mathbb{E}_{\mathbb{Q}} \log \frac{S_T}{S_0} = e^{-rT}\mathbb{E}_{\mathbb{Q}} \left[\left(r - \frac{\sigma^2}{2}\right) T + \sigma W_T^{\mathbb{Q}} \right] \tag{48}$$

the RHS can be written as

$$e^{-rT}\mathbb{E}_{\mathbb{Q}} \left(\int_0^T \frac{1}{S_t} \, dS_t - \frac{\sigma^2}{2} T \right) \tag{49}$$

in more general cases where the volatility is deterministic but time-dependent, i.e. $\sigma = \sigma_t$, the RHS can be written as

$$e^{-rT}\mathbb{E}_{\mathbb{Q}}\left(\int_0^T \frac{1}{S_t}\,dS_t - \frac{1}{2}\int_0^T \sigma_t^2\,dt\right) \tag{50}$$

the first term on RHS $\int_0^T \frac{1}{S_t}\,dS_t$ is the payoff of a continuous-time hedging portfolio. The portfolio always maintains exactly one dollar invested in the stock (buy stock when its price decreases and sell stock when its price increases). However, Carr-Madan formula tells us that log contract can be replicated using European call and put so ***it's actually always possible to replicate $\int_0^T \sigma_t^2\,dt$ as a portfolio of European call and put in a model-independent way****. That's why this hedging portfolio is always called a* ***variance swap****. The variance swap is closely related to VIX, the implied volatility index.*

Path-dependent Option

The **barrier option** (down-and-out call) has payoff $h(S_{[0,T]}) = (S_T - K)_+ \mathbb{I}_{\inf_{t\in[0,T]} S_t > L}$ to be path-dependent. Under risk-neutral pricing, it's only necessary to calculate

$$\mathbb{E}_{\mathbb{Q}}\left((S_T - K)_+ \mathbb{I}_{\inf_{t\in[0,T]} S_t > L}\right) \tag{51}$$

explicit calculation can be done using the joint distribution of $(\inf_{s\le t} W_s, W_t)$ coming from reflection principle. Under the PDE approach, a smart move is taken to organize the existence of the barrier as an extra boundary condition. BS-PDE still holds

$$\partial_t u + rx\partial_x u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \tag{52}$$

with boundary conditions

$$u(T,x) = K, u(t,L) = 0 \tag{53}$$

once u hits zero, BS-PDE ensures that u permanently stays at zero, which is consistent with the behavior of the barrier. The method of images (reflection principle) can then be applied to solve this modified PDE. First use the log transform to change the BS-PDE into a heat equation and then do reflection in the variable x to extend the boundary conditions and connect with the normal BS-PDE.

The **asian option** has payoff $h(S_{[0,T]}) = \left(S_T - \frac{1}{T}\int_0^T S_t\,dt\right)_+$. Under risk-neutral pricing, Monte-Carlo helps us get an estimated price simulating SDE. Under PDE approach, keep in mind the Markovian assumption that we have to keep track of not only t, S_t, but also the running integral of the stock price $I_t = \int_0^t S_u\,du$. As a result, a PDE can still be established at the cost of increasing the space dimension. Let $V_t = u(t, S_t, I_t)$ be the value of the replicating portfolio,

$$dV_t = \partial_t u\,dt + \partial_x u\,dS_t + \partial_i u\,dI_t + \frac{1}{2}\partial_{xx} u\,d\langle S,S\rangle_t + \frac{1}{2}\partial_{ii} u\,d\langle I,I\rangle_t + d\langle S,I\rangle_t \tag{54}$$

$$= \partial_t u\,dt + rS_t\partial_x u\,dt + S_t\partial_i u\,dt + \frac{\sigma^2 S_t^2}{2}\partial_{xx} u\,dt + \sigma S_t \partial_x u\,dW_t^{\mathbb{Q}} \tag{55}$$

since $dI_t = S_t\,dt$, the terms $d\langle I,I\rangle_t$, $d\langle S,I\rangle_t$ are zero. Here we combine the risk-neutral pricing and PDE approach, plugging in the dynamics of S_t under $\mathbb{Q}$ and recalling the fact that $\{e^{-rt}V_t\}$ is a MG under $\mathbb{Q}$. As a result, the drift of dV_t under $\mathbb{Q}$ must match rV_t, which results in the PDE

$$\partial_t u + rx\partial_x u + x\partial_i u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \tag{56}$$

with terminal condition

$$u(T,x,i) = \left(x - \frac{1}{T}i\right)_+ \tag{57}$$

Similarly, **lookback option** has payoff $h(S_{[0,T]}) = \sup_{t\in[0,T]} S_t - S_T$ then under risk-neutral pricing the price can be calculated similar to barrier option. Under PDE approach, we still have to make it Markovian by adding another space variable j with $J_t = \sup_{s\in[0,t]} S_s$ as the running sup. The value of replicating portfolio $V_t = u(t, S_t, J_t)$ induces BS-PDE

$$\partial_t u + rx\partial_x u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \ (\forall x < j) \tag{58}$$

so what's the boundary conditions? The terminal condition is natural that

$$u(T, x, j) = j - x \tag{59}$$

but we still need another boundary condition. When stock price is equal to the running sup, i.e. the current stock price is the highest in the history, a small change in variable j actually makes no difference to the option value (since x has the dominating effect in this case). With this intuition, the other boundary condition can be formed as

$$\partial_j u(t, j, j) = 0 \tag{60}$$

however, the choice of boundary condition is never unique, another boundary condition

$$\partial_x u(t, x, x) = 1 \tag{61}$$

also has a nice interpretation that when stock price is the highest in history, the sensitivity of option price w.r.t. stock price is 1, due to the term $\sup_{t\in[0,T]} S_t$ in the payoff. This Neumann boundary condition has a deep connection with the local time of stochastic process.

Stock with Dividend

Continuous-time dividend yield is always denoted as δ, meaning that in an infinitesimal time interval $[t, t+dt]$, one gets $\delta S_t\, dt$ dividend payment from the stock. Since the dividend payment is from the stock, stock value decreases as dividend payment is made. The rule of dividend payment is actually equivalent to saying that the number of shares investors hold increases at rate δ continuous in time (increase in the number of share results in less stock price per share which matches the effect of dividend payment). In short, if one is holding 1 share of stock at time 0, and the stock has dividend yield δ, then one is actually holding $e^{\delta t}$ share of stock at time t.

With this interpretation in mind, it's pretty clear that dividend payment is discounting the price of the stock as time goes in the forward direction, i.e. it has exactly opposite effect compared to that of the interest rate. When it comes to risk-neutral pricing, in the risk-neutral world we would expect to see $\left\{e^{-(r-\delta)t}S_t\right\}$ as a MG, so the dynamics under risk-neutral measure $\mathbb{Q}$ is

$$dS_t = (r-\delta)S_t\, dt + \sigma S_t\, dW_t^{\mathbb{Q}} \tag{62}$$

just replacing r with $r-\delta$ in the dynamics.

Under PDE approach, recall the BS-PDE

$$\partial_t u + rx\partial_x u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \tag{63}$$

shall we replace both r with $r-\delta$? The answer is NO and we have to be clear with the different meanings of those two r in BS-PDE. The term ru comes from the fact that $\{e^{-rt}V_t\}$ is a MG under $\mathbb{Q}$, which still holds in the dividend-paying situation. On the other hand, $rx\partial_x u$ comes from the Ito formula expansion of dV_t which is directly related to dS_t under $\mathbb{Q}$ and this is exactly the place where we change r to $r-\delta$. As a result, the BS-PDE in dividend-paying situation is

$$\partial_t u + (r-\delta)x\partial_x u + \frac{1}{2}\sigma^2 x^2 \partial_{xx} u - ru = 0 \tag{64}$$

Volatility Models

Motivation and History

The implied volatility curve is symmetric before 1987, i.e. the slope of the curve $K \mapsto I(T,K)$ is close to zero at $K = S_0$ (at-the-money) but it became asymmetric after 1987, i.e. the slope of the curve $K \mapsto I(T,K)$ is negative at $K = S_0$. Empirical observation from the stock market implies that the implied volatility based on options with time to maturity T and strike K denoted $I(T,K)$ is not flat in both T and K. Naturally, the GBM is not a good model for stock price since it assumes constant volatility σ. As a result, changes are made to the BS model in mainly two directions, to introduce jumps in the stock price or to change the volatility setting.

The modelling of volatility has its history as the following

- constant volatility σ
- time-dependent volatility $\sigma(t)$
- local volatility $\sigma(t, S_t)$
- stochastic volatility $\{\sigma_t\}$
- uncertain volatility (volatility can be any process taking value between two given bounds, results in non-linear PDE and the connection with optimal transport)
- fractional BM models for volatility (stock price must be semi-MG to avoid arbitrage but volatility can be something other than a semi-MG)

Time-dependent Volatility Model

With σ replaced with σ_t, a deterministic function in time, the volatility now becomes time-dependent. The BS formula for European call and put still holds if σ^2 is replaced with $\frac{1}{T}\int_0^T \sigma_s^2\, ds$. To see this easy fact, consider the stock price dynamics under $\mathbb{Q}$ as

$$dS_t = S_t(r\, dt + \sigma_t\, dW_t^{\mathbb{Q}}) \tag{65}$$

apply log transformation and Ito formula to see the solution

$$S_t = S_0 e^{\int_0^t (r - \frac{\sigma_s^2}{2})\, ds + \int_0^t \sigma_s\, dW_s^{\mathbb{Q}}} \tag{66}$$

as a result, we just replace $\sigma^2 t$ with $\int_0^t \sigma_s^2\, ds$ and replace $\sigma W_t^{\mathbb{Q}}$ with $\int_0^t \sigma_s\, dW_s^{\mathbb{Q}}$, all the calculation still holds. The implied volatility surface $I(T,K)$ is not flat in T but still flat in K, still far from being an ideal model.

Local Volatility Model

The local volatility model is assuming the volatility to have the form $\sigma = \sigma(t, S_t)$. A simple derivation of BS-PDE tells us that if $V_t = u(t, S_t)$ denotes the value of the replicating portfolio at time t, then

$$\partial_t u + rx\partial_x u + \frac{1}{2}\sigma^2(t,x)x^2\partial_{xx}u - ru = 0 \tag{67}$$

as mentioned above, BS-PDE is quite inefficient in solving option price since it requires the knowledge at all (t, x) just to solve out a single number.

In the local volatility model, however, there is a model-dependent result called **Dupire's PDE** that efficiently solves out option price. We consider the European call option under local volatility model, it has time to maturity T and strike K, with the option price denoted $C(T, K)$. It's clear that

$$C(T,K) = e^{-rT}\mathbb{E}_Q(S_T - K)_+ = e^{-rT}\int_0^\infty (x-K)_+ \cdot p(T,x)\,dx \tag{68}$$

where $p(T, \cdot)$ is the density of S_T under $\mathbb{Q}$. Take derivative w.r.t. T to get

$$\frac{\partial C}{\partial T} = -re^{-rT}\int_K^\infty (x-K)\cdot p(T,x)\,dx + e^{-rT}\int_K^\infty (x-K)\cdot\frac{\partial p}{\partial T}(T,x)\,dx \tag{69}$$

notice that $\frac{\partial p}{\partial T}(T, x)$ is the time derivative of the density of S_T under $\mathbb{Q}$, and $\{S_t\}$ has dynamics under $\mathbb{Q}$ as

$$dS_t = rS_t\,dt + \sigma(t,S_t)S_t\,dW_t^{\mathbb{Q}} \tag{70}$$

the time evolution of density is described by Fokker-Planck equation that

$$L_t^* p(t,x) = \frac{\partial p}{\partial T}(t,x) \tag{71}$$

where L_t^* is the adjoint of the infinitesimal generator L_t of the diffusion and $L_t = rx\partial_x + \frac{1}{2}\sigma^2(t, x)x^2\partial_{xx}$. As a result,

$$\begin{aligned} e^{-rT}\int_K^\infty (x-K)\cdot\frac{\partial p}{\partial T}(T,x)\,dx &= \int_0^\infty e^{-rT}(x-K)_+ \cdot L_T^* p(T,x)\,dx \quad (72)\\ &= \int_0^\infty L_T[e^{-rT}(x-K)_+]\cdot p(T,x)\,dx \quad (73)\end{aligned}$$

from the definition of adjoint. Calculate the action of infinitesimal generator

$$L_T[e^{-rT}(x-K)_+] = e^{-rT}\left(rx\mathbb{I}_{x>K} + \frac{1}{2}\sigma^2(T,x)x^2\delta_{\{K\}}\right) \tag{74}$$

so the integral has value

$$\int_0^\infty L_T[e^{-rT}(x-K)_+]\cdot p(T,x)\,dx = e^{-rT}\left(\int_K^\infty rxp(T,x)\,dx + \frac{1}{2}\sigma^2(T,K)K^2\cdot p(T,K)\right) \tag{75}$$

combine those equations

$$\begin{aligned}\frac{\partial C}{\partial T} &= -re^{-rT}\int_K^\infty (x-K)\cdot p(T,x)\,dx + e^{-rT}\left(\int_K^\infty rxp(T,x)\,dx + \frac{1}{2}\sigma^2(T,K)K^2\cdot p(T,K)\right) &(76)\\ &= rKe^{-rT}\int_K^\infty p(T,x)\,dx + \frac{1}{2}\sigma^2(T,K)K^2e^{-rT}p(T,K) &(77)\end{aligned}$$

the last step is to use B-L formula

$$\frac{\partial C}{\partial T}(T,K) = -rK\frac{\partial C}{\partial K}(T,K) + \frac{1}{2}\sigma^2(T,K)K^2\frac{\partial^2 C}{\partial K^2}(T,K) \tag{78}$$

this is the **Dupire's PDE with natural initial condition**

$$C(0,K) = (S_0-K)_+ \tag{79}$$

since when $T=0$ option price is just the immediate payoff.

Dupire's PDE is a forward equation and it's much more efficient to use than BS-PDE since on knowing one point $(0,K)$ for some given strike K, one can figure out the call price for all (T,K) at all time to maturity. Besides, Dupire's PDE also provides **explicit formula for implied volatility** in the local volatility model. Keep in mind that those results are model-dependent and only hold in the local volatility model.

The local volatility model is expected to produce implied volatility surface $I(T,K)$ not flat in T and not flat in K. So what's the problem with local volatility model? It suffers from the so-called **"(t,T,K) problem"**. In the setting of local volatility, $\sigma=\sigma(t,x)$ so the volatility is always the same given that the time and stock price is the same. As a result, at time 0 if we do calibration and get the surface $I(0,T,K)$ using options that is sold at time 0 and matures at time T, we expect to see the same surface $I(t,T+t,K)$ based on options that is sold at time t and matures at time $T+t$. However, empirical data implies that those two implied volatility surfaces are not the same under time translation, which contradicts local volatility model.

Stochastic Volatility Model

We introduce the general framework of the stochastic volatility model where $\sigma_t = f(Y_t)$ is the volatility process and $\{Y_t\}$ has dynamics

$$dY_t = \alpha(Y_t)\,dt + \beta(Y_t)\,dW_t^1 \tag{80}$$

the stock price has dynamics

$$dS_t = \mu(Y_t)S_t\,dt + \sigma_t S_t\,dW_t^0 \tag{81}$$

and W^0, W^1 has a given correlation coefficient ρ. Equivalently speaking,

$$W^1 = \rho W^0 + \sqrt{1-\rho^2}W^\perp \tag{82}$$

for independent BM $W^0, W^\perp$.

Under risk-neutral pricing, we still want to find $\mathbb{Q}$ under which $\{e^{-rt}S_t\}$ is a MG. A similar application of Girsanov theorem tells us

$$\begin{aligned} dS_t &= rS_t\,dt + (\mu(Y_t)-r)S_t\,dt + f(Y_t)S_t\,dW_t^0 &(83)\\ &= rS_t\,dt + f(Y_t)S_t\,dW_t^{0,\mathbb{Q}} &(84)\end{aligned}$$

so $\mathbb{Q}$ shall be chosen such that

$$W_t^{0,\mathbb{Q}} = W_t^0 + \int_0^t \frac{\mu(Y_s)-r}{f(Y_s)}\,ds \tag{85}$$

is a BM under $\mathbb{Q}$. Different from BS model, in stochastic volatility model we also have the dynamics for $\{Y_t\}$ which we shall care about under the change of measure. Luckily,

$$W_t^{\perp,\mathbb{Q}} = W_t^\perp + \int_0^t \gamma_s\,ds \tag{86}$$

as a BM under $\mathbb{Q}$ does not change the drift term in the stock price dynamics. As a result, different $\{\gamma_t\}$ can be chosen to do the change of measure and they result in $W^{0,\mathbb{Q}}, W^{\perp,\mathbb{Q}}$ being independent BM under $\mathbb{Q}$. It's thus necessary to denote the risk-neutral measure as $\mathbb{Q}^{(\gamma)}$, meaning that it's induced by a choice of $\{\gamma_t\}$.

The RN derivative is then

$$\frac{d\mathbb{Q}^{(\gamma)}}{d\mathbb{P}} = e^{-\int_0^T \frac{\mu(Y_s)-r}{f(Y_s)}\,dW_s^0 - \frac{1}{2}\int_0^T\left(\frac{\mu(Y_s)-r}{f(Y_s)}\right)^2 ds - \int_0^T \gamma_s\,dW_s^\perp - \frac{1}{2}\int_0^T \gamma_s^2\,ds} \tag{87}$$

from Girsanov theorem and it remain as an issue whether the condition of Girsanov holds. Here we assume that it's always legitimate by forcing e.g. the stochastic Sharpe ratio $\frac{\mu(Y_t)-r}{f(Y_t)}$ is always almost surely bounded (which is not necessarily true in practice).

Remark. *In BS model, the risk-neutral measure $\mathbb{Q}$ always exists and is unique, implying that the market is complete. However, in stochastic volatility model, the risk-neutral measure is not unique since any choice of $\{\gamma_t\}$ results in a risk-neutral measure. In other words,* ***the market is incomplete in stochastic volatility model****, meaning that the replicating portfolio cannot always be carried out. Intuitively, this is due to the fact that there is only one stock in the market but there are two independent BM, i.e. two sources of randomness. The existence of excess randomness causes the failure of exact replication. Actually, the second fundamental theorem of asset pricing tells us that the market is complete iff the number of independent stocks equal the number of independent sources of randomness.*

Under risk-neutral measure $\mathbb{Q}^{(\gamma)}$, the dynamics of $\{S_t\}$ now becomes

$$dS_t = rS_t\,dt + f(Y_t)S_t\,dW_t^{0,\mathbb{Q}^{(\gamma)}} \tag{88}$$

and the dynamics of $\{Y_t\}$ now becomes

$$\begin{aligned} dY_t &= \alpha(Y_t)\,dt + \rho\beta(Y_t)\,dW_t^{0,\mathbb{Q}^{(\gamma)}} - \rho\beta(Y_t)\frac{\mu(Y_t)-r}{f(Y_t)}\,dt + \sqrt{1-\rho^2}\beta(Y_t)\,dW_t^{\perp,\mathbb{Q}^{(\gamma)}} - \sqrt{1-\rho^2}\beta(Y_t)\gamma_t\,dt & (89)\\ &= \left[\alpha(Y_t) - \rho\beta(Y_t)\frac{\mu(Y_t)-r}{f(Y_t)} - \sqrt{1-\rho^2}\beta(Y_t)\gamma_t\right]\,dt + \beta(Y_t)\,dW_t^{1,\mathbb{Q}^{(\gamma)}} & (90)\end{aligned}$$

where

$$W_t^{1,\mathbb{Q}^{(\gamma)}} = \rho W_t^{0,\mathbb{Q}^{(\gamma)}} + \sqrt{1-\rho^2}W_t^{\perp,\mathbb{Q}^{(\gamma)}} \tag{91}$$

let's denote

$$\Lambda_t = \rho\frac{\mu(Y_t)-r}{f(Y_t)} + \sqrt{1-\rho^2}\gamma_t \tag{92}$$

as the **total risk premium**. The dynamics of $\{Y_t\}$ under risk-neutral measure is

$$dY_t = \left(\alpha(Y_t) - \beta(Y_t)\Lambda_t\right)\,dt + \beta(Y_t)\,dW_t^{1,\mathbb{Q}^{(\gamma)}} \tag{93}$$

the change of measure is done for now and it's time to price the option under the stochastic volatility model.

Remark. $\{\gamma_t\}$ *is called volatility risk premium or market price of volatility risk and it parametrizes the space of all risk-neutral measures. The total risk premium is just an average of the stochastic Sharpe ratio and the volatility risk premium.*

A fixed $\{\gamma_t\}$ induces the risk-neutral measure $\mathbb{Q}^{(\gamma)}$ and if the payoff function is $h(S_{[0,T]})$ then the arbitrage-free value of the option at time t is just

$$e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}^{(\gamma)}}\left[h(S_{[0,T]})|\mathscr{F}_t\right] \tag{94}$$

under the logic of risk-neutral pricing. Obviously, the selection of $\{\gamma_t\}$ is not unique so this leads to a family of arbitrage-free prices of the option, not a single value.

Different choice of $f, \alpha, \beta, \Lambda$ leads to different stochastic volatility models. Empirically, volatility is observed to have mean-reverting pattern so it's natural to set up an OU dynamics for $\{Y_t\}$ together with $\Lambda \equiv 0, f(y) = e^y$, i.e. under this specific risk-neutral measure $\mathbb{Q}$ (induced by such Λ), we have

$$dS_t = rS_t\,dt + e^{Y_t}S_t\,dW_t^{0,\mathbb{Q}} \tag{95}$$

$$dY_t = \alpha(m - Y_t)\,dt + \beta\,W_t^{1,\mathbb{Q}} \tag{96}$$

where $W^{0,\mathbb{Q}}, W^{1,\mathbb{Q}}$ has correlation coefficient ρ. This model is called the **exp(OU) model**.

A different approach is not to use OU but to use CIR as the dynamics for $\{Y_t\}$ together with $\Lambda \equiv 0, f(y) = \sqrt{y}$, under this risk-neutral measure $\mathbb{Q}$, we have

$$dS_t = rS_t\,dt + \sqrt{Y_t}S_t\,dW_t^{0,\mathbb{Q}} \tag{97}$$

$$dY_t = \alpha(m - Y_t)\,dt + \beta\sqrt{Y_t}\,W_t^{1,\mathbb{Q}} \tag{98}$$

where $W^{0,\mathbb{Q}}, W^{1,\mathbb{Q}}$ has correlation coefficient ρ. This model is called the **Heston model**. There are some constraints on the parameters so that $\{Y_t\}$ stays positive. Heston model is popular since under the PDE approach after log transform, one gets a PDE with constant coefficient, solvable through Fourier transform.

Fixed Income Market

Definition of Rates

Let's assume that a zero-coupon bond that pays 1 at time T has price $p(t,T)$ contracted at time t. Naturally, $\forall t, p(t,t) = 1$ and $p(t,T)$ is well-defined for $\forall t \in \mathbb{R}$. Assume that $p(t,T)$ is also well-defined for $\forall T \in \mathbb{R}$ and is differentiable w.r.t. T. If the accumulation of interest can be equivalently described by a constant continuous-time interest rate on $[t,T]$ denoted $y(t,T)$. Then

$$p(t,T)e^{y(t,T)(T-t)} = 1 \tag{99}$$

solves

$$y(t,T) = -\frac{1}{T-t}\log p(t,T) \tag{100}$$

called the **yield curve** viewed as a function in T after fixing t. Some other rates of people's interest can also be defined in terms of the bond price.

Consider time t with two maturity dates $S < T$. It's interesting to check the forward interest rate on $[S,T]$ contracted at time t. A riskless portfolio can be constructed with no initial endowment that at time t one can short 1 unit of bond that matures at S. This brings immediate cashflow $p(t,S)$ and a debt of returning 1 at time S. However, one can long $\frac{p(t,S)}{p(t,T)}$ unit of bond that matures at T using the $p(t,S)$ immediate cashflow, this brings with a cashflow of $\frac{p(t,S)}{p(t,T)}$ at time T. In all, this portfolio has no immediate cashflow, but we have to pay 1 at time S and receive $\frac{p(t,S)}{p(t,T)}$ at time T. If there exists a constant simple interest rate on $[S,T]$ denoted $L(t;S,T)$, then

$$1 \cdot [1 + L(t;S,T)(T-S)] = \frac{p(t,S)}{p(t,T)} \tag{101}$$

solves

$$L(t;S,T) = \frac{1}{T-S}\left(\frac{p(t,S)}{p(t,T)} - 1\right) \tag{102}$$

called the simple forward rate on $[S,T]$ contracted at t, also called **LIBOR forward** rate. If one use continuous-time compounded interest rate model instead and denote the rate as $R(t;S,T)$, then

$$1 \cdot e^{R(t;S,T)(T-S)} = \frac{p(t,S)}{p(t,T)} \tag{103}$$

solves

$$R(t;S,T) = \frac{1}{T-S}\log\frac{p(t,S)}{p(t,T)} \tag{104}$$

called the continuously compounded forward rate on $[S,T]$ contracted at t.

The spot rates on $[S,T]$ is just the forward rate contracted at time S, i.e. by setting $t = S$, we get the spot rates

$$L(S;S,T) = \frac{1}{T-S}\left(\frac{1}{p(S,T)} - 1\right) \tag{105}$$

$$R(S;S,T) = \frac{1}{T-S}\log\frac{1}{p(S,T)} \tag{106}$$

$L(S;S,T)$ is also called the **LIBOR spot rate**.

Remark. *Future spot rate is different from the current forward rate. Current forward rate reflects the estimation of asset performance on a future time interval based on available current information while future spot rate reflects the estimation of asset performance based on available future information. When constructing riskless arbitrage portfolio, only current forward rate can be used since future spot rate contains risk.*

The **instantaneous forward rate** $f(t,T)$ describes the forward rate in $[T,T+dt]$ contracted at t when dt is small enough. Obviously, all instantaneous rates are built upon continuously compounded interest model. From the definition, it's clear that

$$f(t,T) = \lim_{dt\to 0} R(t;T,T+dt) = \lim_{dt\to 0}\frac{1}{dt}\log\frac{p(t,T)}{p(t,T+dt)} = -\frac{\partial \log p(t,T)}{\partial T} \tag{107}$$

the **instantaneous short rate** $r(t)$ describes the spot rate in $[t,t+dt]$ contracted at t when dt is small enough. It's clear that

$$r(t) = f(t,t) \tag{108}$$

is just plugging $T = t$ into the instantaneous forward rate $f(t,T)$. The short rate has the interpretation as the spot interest rate in a future infinitesimal time interval, so it's exactly the **risk-free interest rate** r we referred to in the context of option pricing. When one opens a savings account, denote the amount of money he has in the account at time t as S_t^0, then it's clear that

$$dS_t^0 = r_t S_t^0\,dt \tag{109}$$

where $r_t = r(t)$ is the short rate. As a result, S_t^0 can be represented in terms of the short rate that

$$S_t^0 = e^{\int_0^t r_s\,ds} \tag{110}$$

tells us how to discount using the short rate.

Modelling the Bond Market

How do we model the bond market? There are several choices to model $p(t,T)$ or $f(t,T)$ or $r(t)$ in terms of t. Typically speaking, directly modelling bond price $p(t,T)$ is hard. If we have the model

$$dp(t,T) = p(t,T)[m(t,T)\,dt + \Sigma(t,T)\,dW_t] \tag{111}$$

finding $m(t,T)$ is easy since under MG measure $m(t,T) = r(t)$ must hold. However, it remains a problem how we interpret and model the diffusion coefficient $\sigma(t,T)$. Another problem here is the selection of BM W, currently all $p(t,T)$ are sharing the same BM, causing the modelling to be too rigid. One may want to change the BM on $\mathbb{R}$ to a BM on an infinite dimensional space $W^{(T)}$ such that each $p(t,T)$ occupies its own BM $W^{(T)}$ for a fixed value of T.

As a result, we don't often model bond price but turn to modelling forward rate $f(t,T)$ and short rate $r(t)$ instead. The forward rate modelling gives the infinite-dimensional Heath-Jarrow-Morton (HJM) model we introduce in the next section (the infinite dimensional structure comes from the dependence on T).

In this section, we talk about modelling short rate $r(t)$ which is a much easier work to do since short rate does not contain T. Typically, a dynamics of $\{r_t\}$ is built as

$$dr_t = \mu(t,r_t)\,dt + \sigma(t,r_t)\,dW_t \tag{112}$$

and the price of the bond can be recovered through

$$p(t,T) = \mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_t^T r_s\,ds}\middle|\mathscr{F}_t\right) \tag{113}$$

the discounted expected payoff. Here $\mathbb{P}^*$ is a pricing measure / MG measure.

Vasicek Model

Vasicek model picks OU dynamics to model the short rate under $\mathbb{P}$ as

$$dr_t = \alpha(r_\infty - r_t)\, dt + \sigma\, dW_t \tag{114}$$

let's try to solve out the bond price $p(0,T)$. This requires setting up the pricing measure $\mathbb{P}^*$, but wait, this pricing measure is not unique! This is due to the fact that there's no risky asset in this market but there is one BM introducing randomness, i.e. the market is not complete. Obviously, infinitely many pricing measure exists and we have to fix **the market price of risk**

$$\lambda_t = \frac{\alpha(r_\infty - r_t)}{\sigma} \tag{115}$$

in order to proceed. As a result, **on assuming that the market price of risk is constantly a given λ for any** t, T, the dynamics under pricing measure $\mathbb{P}^*$ is

$$dr_t = \alpha\left(r_\infty - \frac{\lambda\sigma}{\alpha} - r_t\right) dt + \sigma\, dW_t^* \tag{116}$$

and we denote $m = r_\infty - \frac{\lambda\sigma}{\alpha}$ so the dynamics of $\{r_t\}$ under pricing measure $\mathbb{P}^*$ is given by

$$dr_t = \alpha\left(m - r_t\right)\, dt + \sigma\, dW_t^* \tag{117}$$

still an OU with a new mean-reverting level m.

Remark. *It's important to keep in mind that* ***a short rate model of*** $\{r_t\}$ ***cannot uniquely determine the bond price*** $p(t,T)$***!*** *This is intuitive since the definition* $r(t) = f(t,t)$ *loses information.*

Probabilistic calculation gives

$$p(0,T) = \mathbb{E}_{\mathbb{P}^*} e^{-\int_0^T r_s\, ds} \tag{118}$$

notice that under $\mathbb{P}^*$, $\{r_t\}$ is a Gaussian process with mean μ and covariance kernel $C(s,t) = \frac{\sigma^2}{2\alpha}e^{-\alpha|t-s|}$ so $\int_0^T r_s\, ds$ must still be Gaussian under $\mathbb{P}^*$. One can thus calculate its mean and variance, so $p(0,T)$ is just the MGF of such Gaussian and can be calculated easily.

Here we do not present all the calculations but use PDE method to solve for bond price. Let $u^T(t,x)$ denote the value of the T-maturity bond at time t seeing $r_t = x$, keep in mind that $\left\{e^{-\int_0^t r_s\, ds} u^T(t, r_t)\right\}$ must be a MG under $\mathbb{P}^*$ so we apply Ito formula under $\mathbb{P}^*$

$$du^T(t, r_t) = \partial_t u^T\, dt + \partial_x u^T\, dr_t + \frac{1}{2}\partial_{xx} u^T\, d\langle r, r\rangle_t \tag{119}$$

$$= \partial_t u^T\, dt + \alpha(m - r_t)\partial_x u^T\, dt + \sigma\partial_x u^T\, dW_t^* + \frac{1}{2}\sigma^2\partial_{xx} u^T\, dt \tag{120}$$

match the drift coefficient with $r_t u^T$ to get the **PDE (term structure equation)**

$$\partial_t u^T + \alpha(m-x)\partial_x u^T + \frac{1}{2}\sigma^2 \partial_{xx} u^T - xu^T = 0, u^T(T,x) = 1 \tag{121}$$

this equation is similar to BS-PDE, but the difference is in the xu^T term. This affine structure tells us to use the ansatz

$$u^T(t,x) = e^{b^T(t)x + a^T(t)} \tag{122}$$

the PDE turns into

$$\dot{a}^T(t) + \dot{b}^T(t)x + \alpha(m-x)b^T(t) + \frac{1}{2}\sigma^2[b^T(t)]^2 - x = 0 \tag{123}$$

collect coefficients to turn it into ODEs with terminal conditions

$$\begin{cases} \dot{a}^T(t) + \alpha m b^T(t) + \frac{1}{2}\sigma^2[b^T(t)]^2 = 0 \\ \dot{b}^T(t) - \alpha b^T(t) - 1 = 0 \\ a^T(T) = 0, b^T(T) = 0 \end{cases} \tag{124}$$

we first solve b^T that

$$b^T(t) = \frac{1}{\alpha}e^{\alpha(t-T)} - \frac{1}{\alpha} \tag{125}$$

then plug in the other ODE to solve out $a^T(t)$

$$a^T(t) = \alpha m \int_t^T b^T(s)\, ds + \frac{1}{2}\sigma^2 \int_t^T [b^T(s)]^2\, ds \tag{126}$$

some easy calculation tells us that

$$a^T(t) = \frac{m}{\alpha}(1-e^{\alpha(t-T)}) - m(T-t) + \frac{\sigma^2}{2\alpha^2}\left(\frac{1}{2\alpha}(1-e^{2\alpha(t-T)}) - \frac{2}{\alpha}(1-e^{\alpha(t-T)}) + (T-t)\right) \tag{127}$$

to conclude,

$$p(t,T) = u^T(t,r_t) = \exp\left\{-\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)r_t + \left(m - \frac{\sigma^2}{2\alpha^2}\right)\left[\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right) - (T-t)\right] - \frac{\sigma^2}{4\alpha}\left[\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)\right]^2\right\} \tag{128}$$

recovers the **bond price**.

When it comes to the yield curve,

$$y(t,T) = -\frac{1}{T-t}\log p(t,T) \tag{129}$$

$$= -\frac{1}{T-t}\left\{-\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)r_t + \left(m - \frac{\sigma^2}{2\alpha^2}\right)\left[\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right) - (T-t)\right] - \frac{\sigma^2}{4\alpha}\left[\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)\right]^2\right\} \tag{130}$$

has a certain form depending on the three parameters α, m, σ.

It's interesting that after figuring out $p(t,T)$, we are able to explicitly model the bond price a posteriori. Assume that under $\mathbb{P}^*$

$$dp(t,T) = p(t,T)[m(t,T)\,dt + \Sigma(t,T)\,dW_t^*] \tag{131}$$

then

$$d\log p(t,T) = \left(r_t - \frac{1}{2}\Sigma^2(t,T)\right)\,dt + \Sigma(t,T)\,dW_t^* \tag{132}$$

as argued before, m has to be identified as short rate r_t. Apply Ito formula for $\log p(t,T)$ we have derived above to get

$$d\log p(t,T) = \left(r_t - \frac{\sigma^2}{2}\left[\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)\right]^2\right)\,dt - \sigma\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right)\,dW_t^* \tag{133}$$

compare those two expressions, if we absorb the minus sign in the diffusion coefficient into the BM W^*, then it identifies

$$\Sigma(t,T) = \sigma\frac{1}{\alpha}\left(1-e^{\alpha(t-T)}\right) \tag{134}$$

and the drift of $d\log p(t,T)$ also matches! As a result, the bond price admits the representation

$$dp(t,T) = p(t,T)[m(t,T)\,dt - \Sigma(t,T)\,dW_t^*] \tag{135}$$

under $\mathbb{P}^*$. It turns out that in the Vasicek model the volatility is time-dependent and converges to zero as $t \to T$.

Remark. *Interestingly, the dynamics of $p(t,T)$ is a forward SDE with deterministic terminal condition $p(T,T) = 1$. This might seem very weird at first glance since it actually looks more similar to a BSDE due to its terminal condition. However, Vasicek model has the volatility shrink to zero as the time of maturity comes near, which eliminates the randomness in $p(t,T)$ and results in the deterministic terminal condition.*

It's worth noting that modelling $p(t,T)$ a priori would be hard since one has to guarantee that the deterministic terminal condition $p(T,T) = 1$ of the forward SDE is satisfied. However, by first modelling short rate and then come back to the dynamics of $p(t,T)$, we successfully circumvent this problem.

One can also investigate the dynamics of the forward rate in Vasicek model recalling

$$f(t,T) = -\frac{\partial \log p(t,T)}{\partial T} \tag{136}$$

through some calculation, we get

$$f(t,T) = r_t e^{\alpha(t-T)} + \left(m - \frac{\sigma^2}{2\alpha^2}\right)\left(1 - e^{\alpha(t-T)}\right) + \frac{\sigma^2}{2\alpha^2}\left(1 - e^{\alpha(t-T)}\right)e^{\alpha(t-T)} \tag{137}$$

differentiate w.r.t. t on both sides

$$df(t,T) = \frac{\sigma^2}{\alpha}\left(1 - e^{\alpha(t-T)}\right)e^{\alpha(t-T)}\,dt + \sigma e^{\alpha(t-T)}\,dW_t^* \tag{138}$$

specify

$$\sigma(t,T) = \sigma e^{\alpha(t-T)} \tag{139}$$

as the volatility for forward rate dynamics, then the forward rate dynamics can be represented as

$$df(t,T) = \sigma(t,T)\Sigma(t,T)\,dt + \sigma(t,T)\,dW_t^* \tag{140}$$

Remark. *The relationship*

$$\Sigma(t,T) = \int_t^T \sigma(t,u)\,du \tag{141}$$

holds, i.e. the volatility of bond price is the integral of the volatility of forward rate in terms of T. This is actually the ***no-arbitrage condition for HJM models for forward rates!***

Comments and Improvements on Vasicek Model

Vasicek model is the simplest model with everything to have closed-form solution. Main criticism of Vasicek model focuses on the following points: (i): r_t has positive probability to be negative since it's Gaussian distributed (ii): the calibration of yield curve cannot be well done at each time t using only three parameters α, m, σ.

Naturally, improvements are made on Vasicek model. To solve the problem that r_t might take negative values, one can use CIR model instead of OU model, i.e.

$$dr_t = \alpha(m - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t^* \tag{142}$$

as the dynamics of the short rate under pricing measure $\mathbb{P}^*$. CIR dynamics can be restricted to be positive if α, m, σ

satisfies certain condition. Let's redo the PDE approach that recovers the bond price

$$du^T(t,r_t) = \partial_t u^T\,dt + \partial_x u^T\,dr_t + \frac{1}{2}\partial_{xx}u^T\,d\left\langle r,r\right\rangle_t \tag{143}$$

$$= \partial_t u^T\,dt + \alpha(m-r_t)\partial_x u^T\,dt + \sigma\sqrt{r_t}\partial_x u^T\,dW_t^* + \frac{1}{2}\sigma^2 r_t\partial_{xx}u^T\,dt \tag{144}$$

which provides the PDE

$$\partial_t u^T + \alpha(m-x)\partial_x u^T + \frac{1}{2}\sigma^2 x\partial_{xx}u^T - xu = 0, u^T(T,x) = 1 \tag{145}$$

use the ansatz

$$u^T(t,x) = e^{b^T(t)x+a^T(t)} \tag{146}$$

the PDE turns into

$$\dot{a}^T(t) + \dot{b}^T(t)x + \alpha(m-x)b^T(t) + \frac{1}{2}\sigma^2 x[b^T(t)]^2 - x = 0 \tag{147}$$

collect coefficients to turn it into ODEs with terminal conditions

$$\begin{cases} \dot{a}^T(t) + \alpha m b^T(t) = 0 \\ \dot{b}^T(t) - \alpha b^T(t) + \frac{1}{2}\sigma^2[b^T(t)]^2 - 1 = 0 \\ a^T(T) = 0, b^T(T) = 0 \end{cases} \tag{148}$$

one can still first solve b^T and then solve a^T, notice that the ODE for b^T is now a Ricatti equation but not a linear one.

To solve the calibration problem, $m = m(t)$ is allowed to depend on time in the **Hull-White model**. Now the calibration of yield curve can be perfectly done at each time t yet another problem arises. The calibration of yield curve is not stable in time, i.e. it only calibrates the current yield curves but the same set of parameters does not work in the future so recalibration is required. This problem is the similar to that of the local volatility model for stock price. In order to solve this problem, the natural idea is to either add jumps or to introduce stochastic volatility, similar to what we have done before.

Ho-Lee Model

Ho-Lee model models the short rate through

$$dr_t = \Theta(t)\,dt + \sigma\,dW_t^* \tag{149}$$

under $\mathbb{P}^*$. Interestingly, Θ is unknown when the model is built but can be found through empirical observations. Similar to before, we write down the PDE for bond price w.r.t. $u^T(t,x)$ that

$$\partial_t u^T + \Theta(t)\partial_x u^T + \frac{1}{2}\sigma^2\partial_{xx}u^T - xu^T = 0, u^T(T,x) = 1 \tag{150}$$

put up the ansatz under affine term structure $u^T(t,x) = e^{b^T(t)x + a^T(t)}$ to get ODEs

$$\begin{cases} \dot{b}^T(t) - 1 = 0 \\ \dot{a}^T(t) + \Theta(t)b^T(t) + \frac{\sigma^2}{2}[b^T(t)]^2 = 0 \\ a^T(T) = 0, b^T(T) = 0 \end{cases} \tag{151}$$

solve the ODEs

$$a^T(t) = \int_t^T \Theta(s)(s-T)\,ds + \frac{\sigma^2}{6}(T-t)^3 \tag{152}$$

$$b^T(t) = t - T \tag{153}$$

let's determine the function Θ for now.

Assume that we have observed all the initial term structure $\{p^*(0,T)\}_T$ currently (where $*$ means empirical observation), our goal is to **find Θ such that $\forall T, p^*(0,T) = p(0,T)$, i.e. the model matches exactly with the empirical observation on the initial term structure**. It turns out that

$$\Theta(t) = \frac{\partial f^*(0,t)}{\partial T} + \sigma^2 t \tag{154}$$

we first calculate the empirical forward rates $f^*(0,t)$ based on $\{p^*(0,T)\}_T$ and then use it to build up Θ. Let's prove that this Θ satisfies the requirement. Now that

$$a^T(0) = \int_0^T \frac{\partial f^*}{\partial T}(0,s)(s-T)\,ds \tag{155}$$

$$= (s-T)f^*(0,s)\Big|_{s=0}^T - \int_0^T f^*(0,s)\,ds \tag{156}$$

$$= Tf^*(0,0) - \int_0^T f^*(0,s)\,ds \tag{157}$$

from the affine term structure of bond price,

$$p(0,T) = e^{a^T(0)+b^T(0)r_0} \tag{158}$$
$$= e^{Tf^*(0,0)-\int_0^T f^*(0,s)\,ds - Tr_0} \tag{159}$$
$$= e^{-\int_0^T f^*(0,s)\,ds} = p^*(0,T) \tag{160}$$

notice that $r_0^* = f^*(0,0) = f(0,0) = r_0$ from the definition of short rate and the fact that the short rate at time 0 is observed. The last equation follows from the definition of forward rate.

Given initial term structure $\{p^*(0,T)\}_T$ to solve for Θ, we can proceed and calculate the bond price explicitly as

$$p(t,T) = \frac{p^*(0,T)}{p^*(0,t)} e^{(T-t)f^*(0,t)-\frac{\sigma^2}{2}t(T-t)^2-(T-t)r_t} \tag{161}$$

one of the criticism of this model is that **it is not consistent with the empirical observation that forward rate curve typically has a horizontal asymptote in terms of** T. Using the bond price in Ho-Lee model and the definition of forward rate,

$$f(t,T) = -\frac{\partial \log p^*(0,T)}{\partial T} - f^*(0,t) + \sigma^2 t(T-t) + r_t \tag{162}$$
$$= f^*(0,T) - f^*(0,t) + \sigma^2 t(T-t) + r_t \tag{163}$$

empirically, horizontal asymptote exists so the observed $f^*(0,T)$ has some constant upper bound. As a result, $\sigma^2 t(T-t)$ determines the asymptotic behavior of $f(t,T)$ when $T \to \infty$. Since the term is linear in T, no horizontal asymptote exists for $f(t,T)$. To sum up, if we assume that $\lim_{T\to\infty} f^*(0,T) \le C$, the forward rate curve in Ho-Lee model still has $\forall t, \lim_{T\to\infty} f(t,T) = +\infty$ proves the conclusion.

Derivative on Bonds

Consider the European call bond option with strike K, time of maturity T, the bond option is built upon underlying bond with time of maturity $T_0 > T$. The payoff function of the bond option is $h(p(T, r_T, T_0)) = (p(T, r_T, T_0) - K)_+$ and the bond price is modelled through the Vasicek short rate model. Here we write the bond price as $p(T, r_T, T_0)$ since in Vasicek model the bond price is determined through short rate so $p(T, T_0)$ has dependence on r_T.

The probabilistic approach tells us that the value of this bond option at time t seeing $r_t = x$ denoted $Q(t, x; T)$ has the representation

$$Q(t, x; T) = \mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_t^T r_s\,ds} h(p(T, r_T, T_0))\middle| r_t = x\right) \tag{164}$$

as the discounted expected payoff under the pricing measure. As a result, the joint distribution of $\left(\int_t^T r_s\,ds, r_T\right)$ is required to do the calculation.

PDE approach tells us that if we use the fact that the discounted value process $\left\{e^{-\int_0^t r_s\,ds}Q(t, r_t; T)\right\}$ is a MG under $\mathbb{P}^*$,

$$\begin{aligned} dQ(t, r_t; T) &= \partial_t Q\,dt + \partial_x Q\,dr_t + \frac{1}{2}\partial_{xx}Q\,d\langle r, r\rangle_t &(165)\\ &= \partial_t Q\,dt + \alpha(m - r_t)\partial_x Q\,dt + \sigma\partial_x Q\,dW_t^* + \frac{1}{2}\sigma^2\partial_{xx}Q\,dt &(166)\end{aligned}$$

so we get the same term structure equation with a different terminal condition

$$\partial_t Q + \alpha(m - x)\partial_x Q + \frac{1}{2}\sigma^2\partial_{xx}Q - xQ = 0, Q(T, x; T) = (p(T, x, T_0) - K)_+ \tag{167}$$

notice that exponential affine ansatz no longer works due to a different terminal condition so it's still hard to solve using PDE approach.

The **pricing via forward measure**, on the other hand, helps simplify the calculation through a change of measure. It suffices to compute $Q(0, r_0; T)$, by probabilistic representation,

$$\begin{aligned} Q(0, r_0; T) &= \mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_0^T r_s\,ds}(p(T, r_T, T_0) - K)_+\right) &(168)\\ &= \mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_0^T r_s\,ds}(p(T, r_T, T_0) - K)\mathbb{I}_{p(T, r_T, T_0) > K}\right) &(169)\\ &= \mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_0^T r_s\,ds}p(T, r_T, T_0)\mathbb{I}_{p(T, r_T, T_0) > K}\right) - K\mathbb{E}_{\mathbb{P}^*}\left(e^{-\int_0^T r_s\,ds}\mathbb{I}_{p(T, r_T, T_0) > K}\right) &(170)\end{aligned}$$

notice that $p(0, r_0, T) = \mathbb{E}_{\mathbb{P}^*}e^{-\int_0^T r_s\,ds}$, a smart move is to scale the discount factor with $p(0, r_0, T)$ to get

$$p(0, r_0, T)\mathbb{E}_{\mathbb{P}^*}\left(\frac{e^{-\int_0^T r_s\,ds}}{p(0, r_0, T)}p(T, r_T, T_0)\mathbb{I}_{p(T, r_T, T_0) > K}\right) - Kp(0, r_0, T)\mathbb{E}_{\mathbb{P}^*}\left(\frac{e^{-\int_0^T r_s\,ds}}{p(0, r_0, T)}\mathbb{I}_{p(T, r_T, T_0) > K}\right) \tag{171}$$

since $\frac{e^{-\int_0^T r_s\,ds}}{p(0,r_0,T)}$ has expectation one under $\mathbb{P}^*$, it can be seen as a Radon-Nikodym derivative that admits a change of measure, a frequently used trick (typically with MGF)

$$\frac{d\mathbb{P}^T}{d\mathbb{P}^*} = \xi_T = \frac{e^{-\int_0^T r_s\,ds}}{p(0,r_0,T)} \tag{172}$$

with $\mathbb{P}^T$ to be a legal probability measure (forward measure). As a result, the expression is simplified

$$Q(0,r_0;T) = p(0,r_0,T)\mathbb{E}_{\mathbb{P}^T}\left(p(T,r_T,T_0)\mathbb{I}_{p(T,r_T,T_0)>K}\right) - Kp(0,r_0,T)\mathbb{E}_{\mathbb{P}^T}\left(\mathbb{I}_{p(T,r_T,T_0)>K}\right) \tag{173}$$

it suffices to find the distribution of $p(T,r_T,T_0)$ under $\mathbb{P}^T$. Notice that we no longer have to find the joint distribution of $\left(\int_t^T r_s\,ds, r_T\right)$, which greatly simplifies the calculation.

Finally, let's figure out the distribution of $p(T,r_T,T_0)$ under $\mathbb{P}^T$. Why don't we check the Radon-Nikodym of this change of measure restricted on $\mathscr{F}_t$? Clearly, it should be

$$\xi_t = \mathbb{E}_{\mathbb{P}^*}(\xi_T|\mathscr{F}_t) = e^{-\int_0^t r_s\,ds}\frac{\mathbb{E}(e^{-\int_t^T r_s\,ds}|\mathscr{F}_t)}{p(0,r_0,T)} = e^{-\int_0^t r_s\,ds}\frac{p(t,r_t,T)}{p(0,r_0,T)} \tag{174}$$

recall that under Vasicek model the bond price admits the representation under $\mathbb{P}^*$ that

$$dp(t,T) = p(t,T)(r_t\,dt - \Sigma(t,T)\,dW_t^*) \tag{175}$$

which is a GBM with coefficients to be time-dependent. However, it's still easy to derive the solution

$$p(t,T) = p(0,T)e^{\int_0^t (r_s - \frac{1}{2}\Sigma^2(s,T))\,ds - \int_0^t \Sigma(s,T)\,dW_s^*} \tag{176}$$

so we can plug it into ξ_t and check that

$$\xi_t = e^{-\frac{1}{2}\int_0^t \Sigma^2(s,T)\,ds - \int_0^t \Sigma(s,T)\,dW_s^*} \tag{177}$$

it has the form of exponential local MG and actually it is an exponential MG! We immediately think of Girsanov theorem, it's telling us that

$$W_t^T = W_t^* + \int_0^t \Sigma(s,T)\,ds \tag{178}$$

is a BM under forward measure $\mathbb{P}^T$. So the short rate dynamics under $\mathbb{P}^T$ is

$$dr_t = [\alpha(m - r_t) - \sigma\Sigma(t,T)]\,dt + \sigma\,dW_t^T \tag{179}$$

solve again bond price p under $\mathbb{P}^T$ through PDE approach (as what we have done above) to get the distribution of $p(T,r_T,T_0)$ under $\mathbb{P}^T$ and the price of this bond option can be explicitly solved.

BS Style Formula for Bond Option

Let's consider the short rate Vasicek model, under which the price of a European call bond option is

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}^*} e^{-\int_0^T r_s\,ds}(p(T,r_T,T_0)-K)_+ && (180)\\
&= \mathbb{E}_{\mathbb{P}^*} e^{-\int_0^T r_s\,ds}(p(T,r_T,T_0)\mathbb{I}_{p(T,r_T,T_0)>K}) - K\cdot \mathbb{E}_{\mathbb{P}^*} e^{-\int_0^T r_s\,ds}\mathbb{I}_{p(T,r_T,T_0)>K} && (181)\\
&= p(0,T_0)\cdot \mathbb{E}_{\mathbb{P}^*}\frac{e^{-\int_0^T r_s\,ds}}{p(0,T_0)} p(T,r_T,T_0)\mathbb{I}_{p(T,r_T,T_0)>K} - K\cdot p(0,T)\cdot \mathbb{E}_{\mathbb{P}^*}\frac{e^{-\int_0^T r_s\,ds}}{p(0,T)}\mathbb{I}_{p(T,r_T,T_0)>K} && (182)\\
&= p(0,T_0)\cdot \mathbb{E}_{\mathbb{P}^*}\frac{e^{-\int_0^T r_s\,ds}}{p(0,T_0)} p(T,r_T,T_0)\mathbb{I}_{p(T,r_T,T_0)>K} - K\cdot p(0,T)\cdot \mathbb{E}_{\mathbb{P}^T}\mathbb{I}_{p(T,r_T,T_0)>K} && (183)
\end{aligned}
$$

let's rewrite the expectation in the first term on the RHS

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}^*}\frac{e^{-\int_0^T r_s\,ds}}{p(0,T_0)} p(T,r_T,T_0)\mathbb{I}_{p(T,r_T,T_0)>K} &= \mathbb{E}_{\mathbb{P}^*}\left(\frac{e^{-\int_0^T r_s\,ds}}{p(0,T_0)}\cdot \mathbb{E}_{\mathbb{P}^*}\left[e^{-\int_T^{T_0} r_s\,ds}\middle|\mathscr{F}_T\right]\cdot \mathbb{I}_{p(T,r_T,T_0)>K}\right) && (184)\\
&= \frac{\mathbb{E}_{\mathbb{P}^*}\left(\mathbb{E}_{\mathbb{P}^*}\left[e^{-\int_0^{T_0} r_s\,ds}\mathbb{I}_{p(T,r_T,T_0)>K}\middle|\mathscr{F}_T\right]\right)}{p(0,T_0)} && (185)\\
&= \mathbb{E}_{\mathbb{P}^*}\frac{e^{-\int_0^{T_0} r_s\,ds}}{p(0,T_0)}\mathbb{I}_{p(T,r_T,T_0)>K} && (186)\\
&= \mathbb{E}_{\mathbb{P}^{T_0}}\mathbb{I}_{p(T,r_T,T_0)>K} && (187)
\end{aligned}
$$

where

$$
\frac{d\mathbb{P}^T}{d\mathbb{P}^*} = \frac{e^{-\int_0^T r_s\,ds}}{p(0,T)}, \frac{d\mathbb{P}^{T_0}}{d\mathbb{P}^*} = \frac{e^{-\int_0^{T_0} r_s\,ds}}{p(0,T_0)} \tag{188}
$$

at this point the call price has a BS style formula

$$
C = p(0,T_0)\cdot p_1 - K\cdot p(0,T)\cdot p_2 \tag{189}
$$

where

$$
\begin{cases}
p_1 = \mathbb{E}_{\mathbb{P}^{T_0}}\mathbb{I}_{p(T,r_T,T_0)>K}\\
p_2 = \mathbb{E}_{\mathbb{P}^T}\mathbb{I}_{p(T,r_T,T_0)>K}
\end{cases} \tag{190}
$$

are probabilities under measure $\mathbb{P}^{T_0}$ and $\mathbb{P}^T$.

Recall that the bond price under Vasicek model has affine term structure

$$
p(T,r_T,T_0) = \exp\left\{-\frac{1}{\alpha}\left(1-e^{\alpha(T-T_0)}\right)r_T + \left(m-\frac{\sigma^2}{2\alpha^2}\right)\left[\frac{1}{\alpha}\left(1-e^{\alpha(T-T_0)}\right)-(T_0-T)\right] - \frac{\sigma^2}{4\alpha}\left[\frac{1}{\alpha}\left(1-e^{\alpha(T-T_0)}\right)\right]^2\right\} \tag{191}
$$

so $\{p(T, r_T, T_0) > K\}$ if and only if $\{r_T < R\}$ for some value R. Therefore, p_1, p_2 has the following representation

$$\begin{cases} p_1 = \mathbb{P}^{T_0}(r_T < R) \\ p_2 = \mathbb{P}^T(r_T < R) \end{cases} \tag{192}$$

and we only need to figure out the distribution of r_T under $\mathbb{P}^T, \mathbb{P}^{T_0}$ if possible. The following lemma helps us figure out the distribution.

Lemma 2. *Let (X, Y) be Gaussian random vector under $\mathbb{P}^*$ and $\mathbb{P} << \mathbb{P}^*$ with*

$$\frac{d\mathbb{P}}{d\mathbb{P}^*} = \frac{e^{-\lambda X}}{\mathbb{E}_{\mathbb{P}^*} e^{-\lambda X}} \tag{193}$$

then Y is still Gaussian under $\mathbb{P}$ and its distribution under $\mathbb{P}$ is determined by

$$\mathbb{E}_{\mathbb{P}} Y = \mathbb{E}_{\mathbb{P}^*} Y - cov_{\mathbb{P}^*}(X, Y) \cdot \lambda, Var_{\mathbb{P}}(Y) = Var_{\mathbb{P}^*}(Y) \tag{194}$$

Proof. Assume (X, Y) has mean vector $\mu = (\mu_1, \mu_2)^T$ and covariance matrix Σ under $\mathbb{P}^*$. Then it has MGF under $\mathbb{P}^*$ as

$$M_{\mathbb{P}^*}(t) = e^{\mu^T t + \frac{1}{2} t^T \Sigma t} \tag{195}$$

where $t = (t_1, t_2)^T$. Denote the random vector as $Z = (X, Y)^T \sim N(\mu, \Sigma)$ under $\mathbb{P}^*$. Conduct the change of measure

$$\begin{aligned} M_{\mathbb{P}}(t) &= \mathbb{E}_{\mathbb{P}} e^{t^T Z} = \mathbb{E}_{\mathbb{P}^*} \frac{d\mathbb{P}}{d\mathbb{P}^*} e^{t^T Z} && (196) \\ &= \frac{1}{\mathbb{E}_{\mathbb{P}^*} e^{-\lambda X}} \mathbb{E}_{\mathbb{P}^*} e^{t^T Z - \lambda X} = \frac{1}{\mathbb{E}_{\mathbb{P}^*} e^{-\lambda X}} \mathbb{E}_{\mathbb{P}^*} e^{(t_1 - \lambda) X + t_2 Y} && (197) \\ &= \frac{M_{\mathbb{P}^*}((t_1 - \lambda, t_2)^T)}{M_{\mathbb{P}^*}((-\lambda, 0)^T)} && (198) \\ &= \frac{e^{\mu^T t - \lambda \mu_1 + \frac{1}{2}[\Sigma_{11}(t_1 - \lambda)^2 + 2\Sigma_{12}(t_1 - \lambda) t_2 + \Sigma_{22} t_2^2]}}{e^{-\lambda \mu_1 + \frac{1}{2} \lambda^2 \Sigma_{11}}} && (199) \\ &= e^{\mu^T t + \frac{1}{2}[\Sigma_{11}(t_1 - \lambda)^2 + 2\Sigma_{12}(t_1 - \lambda) t_2 + \Sigma_{22} t_2^2] - \frac{1}{2} \lambda^2 \Sigma_{11}} && (200) \end{aligned}$$

to derive the MGF of random vector Z under $\mathbb{P}$. Since we only care about the marginal of Y, set $t_1 = 0$ to get the MGF of Y under $\mathbb{P}$ that

$$M_{\mathbb{P}}(t_2) = e^{(\mu_2 - \Sigma_{12} \lambda) t_2 + \frac{1}{2} \Sigma_{22} t_2^2} \tag{201}$$

so $Y \sim N(\mu_2 - \Sigma_{12} \lambda, \Sigma_{22})$ under $\mathbb{P}$. □

Specify $X = -\int_0^T r_s \, ds, Y = r_T$ in the lemma above, it's clear that (X, Y) is a Gaussian random vector under $\mathbb{P}^*$ so $Y = r_T$ is still Gaussian under $\mathbb{P}^T$ with its distribution known. Similarly, $Y = r_T$ is also Gaussian under $\mathbb{P}^{T_0}$

with its distribution known. It's immediately clear that

$$\begin{cases} p_1 = \mathbb{P}^{T_0}(r_T < R) \\ p_2 = \mathbb{P}^{T}(r_T < R) \end{cases} \tag{202}$$

p_1, p_2 are actually Gaussian CDF values and can both be written in terms of Φ. As a result, the European call price of bond option under Vasicek model has the form

$$C = p(0,T_0) \cdot \Phi(d_1) - K \cdot p(0,T) \cdot \Phi(d_2) \tag{203}$$

of BS style.

Forward Rate Model

The easiest forward rate model assumes that bond with different time of maturity T shares the same BM, i.e.

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t^* \tag{204}$$

under pricing measure $\mathbb{P}^*$. As we have mentioned earlier, this model is too rigid to use in practice, and a better model is to assign a separate BM for each time of maturity T, which results in an infinite dimensional model described below as the HJM framework. Nevertheless, playing with this trivial forward rate model provides some intuition on modelling bond price. Let's try to recover the bond price dynamics from this forward rate model using the definition

$$p(t,T) = e^{-\int_t^T f(t,s)\,ds} \tag{205}$$

specify $X(t,T) = \int_t^T f(t,s)\,ds$ so $p(t,T) = e^{-X(t,T)}$. First derive the dynamics of X, differentiate w.r.t. t to get (notice that both the integration domain and the integrand has dependence on t so we have to use chain rule!)

$$\frac{d}{dt}X(t,T) = \left(\frac{d}{dt}\int_t^T f(u,s)\,ds\right)\Big|_{u=t} + \left(\frac{d}{dt}\int_u^T f(t,s)\,ds\right)\Big|_{u=t} \tag{206}$$

$$= -f(t,t) + \left(\frac{d}{dt}\int_u^T f(t,s)\,ds\right)\Big|_{u=t} \tag{207}$$

the second term on RHS requires more calculation, but we assume the differentiation interchanges with integration arbitrarily in this section for simplicity. It remains to calculate

$$\left(\int_u^T \frac{d}{dt}f(t,s)\,ds\right)\Big|_{u=t} = \int_t^T \frac{d}{dt}f(t,s)\,ds \tag{208}$$

$$= \int_t^T \alpha(t,s)\,ds + \int_t^T \sigma(t,s)\,ds\frac{dW_t^*}{dt} \tag{209}$$

where $\frac{dW_t^*}{dt}$ is just for notation. Take differential w.r.t. t for $X(t,T)$ to see

$$dX(t,T) = -f(t,t)\,dt + \left(\int_t^T \alpha(t,s)\,ds\right)dt + \left(\int_t^T \sigma(t,s)\,ds\right)dW_t^* \tag{210}$$

provides the dynamics, apply Ito formula for $p(t,T)$, differentiate w.r.t. t to get

$$dp(t,T) = -e^{-X(t,T)}\,dX(t,T) + \frac{1}{2}e^{-X(t,T)}\,d\left\langle X(\cdot,T), X(\cdot,T)\right\rangle_t \tag{211}$$

$$= p(t,T)\left[r_t\,dt - \left(\int_t^T \alpha(t,s)\,ds\right)dt - \left(\int_t^T \sigma(t,s)\,ds\right)dW_t^* + \frac{1}{2}\left(\int_t^T \sigma(t,s)\,ds\right)^2 dt\right] \tag{212}$$

recovers the dynamics of bond price.

From our reasoning above, under $\mathbb{P}^*$ the drift term must match $r_t\,dt$ so **the consistency condition** must hold

$$\frac{1}{2}\left(\int_t^T \sigma(t,s)\,ds\right)^2 = \int_t^T \alpha(t,s)\,ds \tag{213}$$

must be satisfied to provide a reasonable dynamics for forward rate under $\mathbb{P}^*$. Compare the diffusion coefficient, we see that

$$\Sigma(t,T) = \int_t^T \sigma(t,s)\,ds \tag{214}$$

by recalling that $-\Sigma(t,T)p(t,T)$ is the diffusion coefficient in the dynamics of the bond price under $\mathbb{P}^*$. This is exactly the **no-arbitrage condition for forward rate model** and we have checked that this condition holds for Vasicek model.

Let's check for Vasicek model the first consistency condition mentioned above. Recall that under Vasicek,

$$\sigma(t,T) = \sigma e^{\alpha(t-T)}, \alpha(t,T) = \frac{\sigma^2}{\alpha}\left(1 - e^{\alpha(t-T)}\right)e^{\alpha(t-T)} \tag{215}$$

compute LHS

$$\frac{1}{2}\left(\int_t^T \sigma(t,s)\,ds\right)^2 = \frac{1}{2}\Sigma^2(t,T) = \frac{\sigma^2}{2\alpha^2}\left(1 - e^{\alpha(t-T)}\right)^2 \tag{216}$$

compute RHS

$$\int_t^T \alpha(t,s)\,ds = \frac{\sigma^2}{\alpha}\int_t^T e^{\alpha(t-s)} - e^{2\alpha(t-s)}\,ds = \frac{\sigma^2}{2\alpha^2}\left(e^{2\alpha(t-T)} - 2e^{\alpha(t-T)} + 1\right) \tag{217}$$

so they match each other, the consistency condition holds.

Remark. *As we can see, when modelling forward rate, $\alpha(t,T), \sigma(t,T)$ cannot be arbitrarily picked and has to satisfy the consistency condition. After figuring out the dynamics of the forward rate, the dynamics of the bond price is automatically given. This is the evidence towards the same intuition that directly modelling bond price is almost impossible (to match the deterministic terminal condition of an FSDE) and one shall model short rate (less restriction) or forward rate (more restriction) instead. The short rate and forward rate model then recovers the model of bond price in a very natural way.*

Credit Risk

In reality, bond, especially company-issued bond defaults. To model the exposure to credit risk, we shall consider the **defaultable bond** whose price at time t is denoted $p^D(t,T)$ for a defaultable bond that matures at time T. Clearly, the payoff of this bond is 1 if default does not happen and is 0 if default happens. Since we care about modelling credit risk, the interest rate is assumed to be fixed for simplicity. There are two main approaches to modelling credit risk, the structural models and the intensity-based models.

Structural Models

The structural models considers the possibility of default to be closely related to the value of the company. It's assumed that the value of the company is completely reflected through its stock price. Whenever stock price hits level below D, default happens and we denote the hitting time as τ.

Clearly, the price of the (defaultable) bond is given by

$$p^D(t,T) = \mathbb{E}_{\mathbb{P}^*} e^{-r(T-t)} \mathbb{I}_{\tau > T} = e^{-r(T-t)} \cdot \mathbb{P}^*(\tau > T) \tag{218}$$

under pricing measure $\mathbb{P}^*$. The price only has something to do with the distribution of τ under $\mathbb{P}^*$ and the formula of the price reminds us of the price of the down-and-out barrier binary option, for which we have an explicit formula from the method of images. Let's assume for now that the stock price follows BS model, then $p^D(t,T)$ has a formula and we can calculate the yield curve $y^D(t,T)$ in closed form. However, empirical data does not match the pattern of $y^D(t,T)$. The shortfall is very obvious in the case of shorter maturity, the reason is that the stock price model is typically a continuous adapted process, so the hitting time is a predictable stopping time, i.e. we can see the default approaching gradually.

The improvements on this vanilla Black-Cox model is made in several ways, e.g. adding stochastic volatility to the stock price dynamics, using a Bayesian way to make default level D stochastic etc.

Intensity-based Models

Intensity-based models introduce jumps to model "surprise". In reality, defaults always happen as surprise instead of approaching gradually. Since only the distribution of τ is needed, the simplest model sets up τ to follow exponential distribution $\mathcal{E}(\lambda)$ due to its memoryless property, i.e. defaults happen as memoryless arrivals. Simple calculation shows

$$p^D(0,T) = e^{-rT-\lambda T}, y^D(0,T) = r + \lambda \tag{219}$$

the yield curve is constant at time 0, which obviously does not match empirical observations.

Naturally, improvements are made by picking deterministic time-dependent intensity $\lambda(t)$ so now $\mathbb{P}(\tau > T) = e^{-\int_0^T \lambda(t)\,dt}$, this allows us to fit today's yield curve $y^D(0,T)$ perfectly but it has the issue of instability. After a period of time, the original $\lambda(t)$ does not work any more so we have to recalibrate the model to match the data.

Further improvements are made by considering the doubly stochastic process (Cox process) by setting $\{\lambda_t\}$ as a stochastic process that has its own dynamics. At this point, the model for $\{\lambda_t\}$ is very alike the short rate model for $\{r_t\}$, but there is a slight difference that $\{r_t\}$ taking negative values does not cause an essential issue but $\{\lambda\}$ has to stay strictly positive. As a result, the CIR process is used to model $\{\lambda_t\}$ and stochastic volatility can be added to the volatility in CIR etc.

Multi-name Setting

We briefly talk about the difficulties in the multi-name setting, i.e. there exists multiple companies in the economy. The difficulty lies in **modelling the correlation**.

For structural models, one has to deal with the correlations between BMs. If each company owns a separate default level and has separate hitting time as the default time, calculating the joint distribution of all hitting times is hard. This requires using reflection principle for infinitely many times and would result in an infinite series. On the other hand, if we solve the joint distribution numerically we have to do Monte Carlo. Monte Carlo typically has problem estimating small probability so importance sampling is needed to make sure that the stock price process hits the default level. Moreover, if the model is too complicated to design an importance sampling scheme, use evolution selection algorithm combining with the idea in particle system.

For intensity-based models, one has to think about correlating two exponentially distributed r.v. in a reasonable way. This is hard typically but we can use **Copula** to put up the model. Since a pair of exponentially distributed r.v. (τ_1, τ_2) is typically sampled through inverse CDF method based on uniform random variables (U_1, U_2), an easy idea is to model the correlation between two uniformly distributed r.v. and then transit the correlation to exponentially distributed r.v. In other words, Copula means **the transition of correlation**. However, it's also hard to model the correlation between two uniformly distributed r.v. but we notice that the correlation between Gaussian r.v. is easy to model. As a result, we shall sample Gaussian distributed r.v. with correlation ρ which can be done easily, transit the correlation to (U_1, U_2) by applying the Gaussian CDF and then transit the correlation to (τ_1, τ_2) by applying the inverse of the exponential distribution CDF. This method is called Gaussian Copula but it's not the unique Copula to use.

One can also build Copula based on Bernoulli experiments using Bernoulli experiments to approximate Poisson process (shrink time lag) and use first hitting time of Poisson process to produce exponential distribution. In this sense, one can build correlation between X, Y, two $B(1, p)$ r.v., which is very simple since p depends on λ so we only need to know $\mathbb{P}(X = 1, Y = 1)$ to fix the joint distribution. Transit this correlation to exponential r.v. provides the Marshall-Olkin Copula which makes use of the fact that the joint distribution of Bernoulli is restrictive.

Another way to deal with multi-name intensity-based model is to correlate the λ instead of directly the exponential r.v. This is more natural in the setting of SDE since we just need to correlate the BM in the dynamics of λ. Probabilistic and PDE approach can be used to solve for $\mathbb{P}(\tau_1 > T, ..., \tau_N > T)$.

To build credit derivative, we introduce the **loss distribution**

$$L_t = \sum_{i=1}^{N} a_i \mathbb{I}_{\tau_i \leq t} \tag{220}$$

as the default amount at time t with company i defaulting amount a_i. Assume the market capitalization of company i is $a_i = 1$ in the case of **symmetric-name case** so the credit derivative paying 1 at time T if at least one company defaults has payoff function

$$\mathbb{I}_{L_T \geq 1} \tag{221}$$

it's just a European digital call built upon $\{L_t\}$. Typically, one trades in the market credit derivatives that pay when e.g. $3\%, 7\%, 15\%$ of the names in the basket default. If a credit derivative does not pay if less than 3% of the names in the basket default, pays 1 if more than 7% of the names in the basket default and pays linearly within 3% to 7%, then it's just a straddle on $\{L_t\}$ and can be decomposed into two European calls with payoff $\mathbb{I}_{L_T \geq 0.03N}$ and $\mathbb{I}_{L_T \geq 0.07N}$.

Generally computing the loss distribution is hard but if one knows the family of the probability of survival in the symmetric-name case

$$S_m = \mathbb{P}\left(\tau_1 > T, ..., \tau_m > T\right),\ 1 \leq m \leq N \tag{222}$$

the loss distribution can be easily calculated. Notice that the probability above only depends on m since all names are identical.

Lemma 3. *Let $F_k^{(N)} = \mathbb{P}\left(L_T = k\right)$ denote the probability that exactly k names are defaulting at time T in the symmetric-name case, then*

$$F_k^{(N)} = \binom{N}{k} \sum_{j=0}^{k} \binom{k}{j} (-1)^j S_{N+j-k} \tag{223}$$

Proof. $F_k^{(N)} = \mathbb{P}\left(\sum_{i=1}^N \mathbb{I}_{\tau_i \leq T} = k\right)$ and we consider

$$\phi(z) = \mathbb{E} z^{\sum_{i=1}^N \mathbb{I}_{\tau_i > T}} = \mathbb{E} z^{N - \sum_{i=1}^N \mathbb{I}_{\tau_i \leq T}} = \sum_{k=0}^{N} z^{N-k} F_k^{(N)} = \sum_{k=0}^{N} z^k F_{N-k}^{(N)} \tag{224}$$

on the other hand, since the indicator takes value only zero or one,

$$\begin{aligned}
\phi(z) &= \mathbb{E}\prod_{i=1}^{N} z^{\mathbb{I}_{\tau_i>T}} = \mathbb{E}\prod_{i=1}^{N}[1+(z-1)\mathbb{I}_{\tau_i>T}] && (225)\\
&= \mathbb{E}\sum_{\{i_1,\ldots,i_n\}\subset\{1,\ldots,N\}} (z-1)^n \mathbb{I}_{\tau_{i_1}>T,\ldots,\tau_{i_n}>T} && (226)\\
&= \sum_{\{i_1,\ldots,i_n\}\subset\{1,\ldots,N\}} (z-1)^n S_n = \sum_{n=0}^{N}\binom{N}{n}(z-1)^n S_n && (227)\\
&= \sum_{n=0}^{N}\binom{N}{n}\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}z^k S_n && (228)
\end{aligned}$$

change the order of summation to get

$$\phi(z) = \sum_{k=0}^{N}\left(\sum_{n=k}^{N}\binom{N}{n}\binom{n}{k}(-1)^{n-k}S_n\right)z^k \tag{229}$$

compare coefficients to get

$$F_{N-k}^{(N)} = \sum_{n=k}^{N}\binom{N}{n}\binom{n}{k}(-1)^{n-k}S_n \tag{230}$$

so

$$F_k^{(N)} = \sum_{n=N-k}^{N}\binom{N}{n}\binom{n}{N-k}(-1)^{n-N+k}S_n \tag{231}$$

set $j = N-n$ to get

$$\begin{aligned}
F_k^{(N)} &= \sum_{j=0}^{k}\binom{N}{j}\binom{N-j}{N-k}(-1)^{k-j}S_{N-j} && (232)\\
&= \binom{N}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^j S_{N-k+j} && (233)
\end{aligned}$$

concludes the proof. □

Remark. *This proof mainly uses the zero-one nature of the indicator and the generation function in combinatorics which deserves our attention.*

Heath-Jarrow-Morton (HJM) Framework

Here we model the forward rate dynamics for fixed T

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t \tag{234}$$

Musiela Parametrization

We now focus on the forward rate dynamics and want to change the parametrization of forward curve to $r_t(\cdot) = f(t, \cdot + t)$ so $r_t(x) = f(t, x+t)$. It can be understood as $T = x + t$ so we only care about **time to maturity** x **instead of time of maturity** T.

Why would we like to use this different parametrization? It has the advantage that it allows non-varying state space (for the set of variables (t,T), the range $T > t$ depends on t but now for (t,x), $x > 0$ does not depend on t) and allows non-local state dependence, not only state dependent coefficients of the form $\sigma(t,T,f(t,T))$ but also those having dependence on the whole curve instead.

Let $\{S(t)\}_{t\geq 0}$ be the set of right shift operators of unit t defined by $[S(t)f](x) = f(x+t)$ for any function $f : \mathbb{R}_+ \to \mathbb{R}$. Recall the forward rate model, $f(t,T) = f(0,T) + \int_0^t \alpha(s,T)\,ds + \int_0^t \sigma(s,T)\,dW_s$ and reformulate it in terms of (t,x) with $T = t + x$ to get

$$f(t,x+t) = [S(t)f](0,x) + \int_0^t [S(t-s)\alpha](s,x+s)\,ds + \int_0^t [S(t-s)\sigma](s,x+s)\,dW_s \tag{235}$$

where the shift operator is acting on the second argument of f, α, σ. One might recall at this step that the equation above is similar to the OU process. This inspires us to first investigate the derivation of OU process and match each part correspondingly in the equation above.

Recall that OU process is the solution to the following SDE

$$dX_t = (AX_t + \alpha_t)\,dt + \sigma_t\,dW_t \tag{236}$$

where $A \in \mathbb{R}^{n\times n}$ is a matrix and the SDE is an n-dimensional one. Consider

$$Y_t = e^{-At}X_t \tag{237}$$

and apply Ito formula to derive an SDE of Y_t that

$$dY_t = e^{-At}\,dX_t - Ae^{-At}X_t\,dt \tag{238}$$

plug in the original SDE to see

$$dY_t = e^{-At}(\alpha_t\,dt + \sigma_t\,dW_t) \tag{239}$$

solve it to get

$$Y_t = Y_0 + \int_0^t e^{-As}\alpha_s\, ds + \int_0^t e^{-As}\sigma_s\, dW_s \tag{240}$$

so the OU process is given by

$$X_t = X_0 e^{At} + \int_0^t e^{A(t-s)}\alpha_s\, ds + \int_0^t e^{A(t-s)}\sigma_s\, dW_s \tag{241}$$

Compare those two expressions to find that the $S(t-s)$ part corresponds to $e^{A(t-s)}$ and the $S(t)$ part corresponds to e^{tA} so in order to find an SDE to characterize the forward dynamics under Musiela parametrization, one has to find what A corresponds to in the SDE of OU process. A direct observation is that

$$\frac{d}{dt}e^{tA}\Big|_{t=0} = A \tag{242}$$

so A is actually the **infinitesimal generator** of e^{tA} which forms a semi-group $\left\{e^{tA}\right\}_{t\geq 0}$ under matrix multiplication. As a result, one naturally think about the infinitesimal generator of the shift operator semi-group $\{S(t)\}_{t\geq 0}$ under function composition.

C_0 Semi-group

A C_0 **semigroup on a Banach space** $\mathscr{H}$ is a mapping $\mathbb{R}_+ \to \mathscr{L}(\mathscr{H})$ from non-negative real numbers to the set of all linear bounded operators $\mathscr{H} \to \mathscr{H}$ (here for fixed time t, $S(t)$ maps the function $f \in \mathscr{H}$ to another function $S(t)f \in \mathscr{H}$) with the property that

$$S(0) = id,\ \forall s, t \geq 0, S(t+s) = S(t) \circ S(s),\ \forall f \in \mathscr{H}, ||S(t)f - f|| \to 0\ (t \to 0) \tag{243}$$

notice the third condition requires the existence of the norm and the continuity of $S(t)$ when t is close to 0, that's why it's called a C_0 semi-group.

The infinitesimal generator $\mathscr{A} \in \mathscr{L}(\mathscr{H})$ on the C_0 semi-group ($\mathscr{A}$ maps function $f \in \mathscr{H}$ to function $\mathscr{A}f \in \mathscr{H}$) is defined as

$$\mathscr{A}f = \lim_{t \searrow 0} \frac{[S(t) - id]f}{t} \tag{244}$$

for $f \in \mathscr{H}$ if it exists, its domain is denoted as $D(\mathscr{A})$ and the domain is a linear dense subspace of $\mathscr{H}$. One often writes $S(t) = e^{t\mathscr{A}}$ to express its similarity to the exponential.

Remark. *In our Musiela parametrization, S is the element in the shift semi-group and it maps time $t \geq 0$ to $S(t) \in \mathscr{L}(\mathscr{H})$ as a linear bounded operator. Here $\mathscr{H}$ is the normed function space containing the forward curve f and the coefficients in the dynamics α, σ.*

Naturally, we would expect to see that the infinitesimal generator of the shift semi-group is the **derivative operator** O such that $Of = f'$ since

$$\lim_{t \searrow 0} \frac{[(S(t) - id)f](x)}{t} = \lim_{t \searrow 0} \frac{f(x+t) - f(x)}{t} = f'(x) \tag{245}$$

if f has enough regularity.

At this point, we would expect to see a characterization for the forward rate dynamics under Musiela parametrization

$$r_t(x) = f(t, x+t) = [S(t)f](0, x) + \int_0^t [S(t-s)\alpha](s, x+s)\, ds + \int_0^t [S(t-s)\sigma](s, x+s)\, dW_s \tag{246}$$

that

$$dr_t(x) = Or_t(x) + \alpha(t, t+x)\, dt + \sigma(t, t+x)\, dW_t \tag{247}$$

where $Or_t(x) = \frac{\partial}{\partial x} r_t(x)$ is the image of $r_t(x)$ under derivative operator and W is $\mathscr{H}$-BM. This is the SPDE characterization that naturally arises. Notice that here we **find the functional relationship of $r_t(x)$ depending on time to maturity x and view t as the family index**.

Remark. *Generally SPDE refers to the type of equation above where one can view it as a PDE in both t and x with the stochastic noise. The main difference between SPDE and SDE is that in SDE there's only differential w.r.t. the time variable t and we often require drift and diffusion coefficient to be regular enough (Lipschitz and growth condition) so that it has existence and uniqueness of the solution. On the other hand, for SPDE, there is derivative* *operator O standing for the derivative w.r.t. the space variable x and it is not a bounded operator so there's not such good regularity condition.*

Wiener Process in $\mathscr{H}$

Let's consider a general form of SPDE that

$$dX_t = (\mathscr{A} X_t + \alpha_t)\, dt + \sigma_t\, dW_t \tag{248}$$

where $X_t = \{X(t)(x)\}_{t \geq 0}$ is view as a family of function in space variable x indexed by time t. So X_t takes values in the function space $\mathscr{H}$ and so does α_t. In the following context, we omit the space variable x if possible. For the simple setting, we may have W_t taking value in $\mathscr{H}$ as a Wiener process on $\mathscr{H}$ and σ_t taking values in $\mathscr{L}(\mathscr{H})$ so that $\sigma_t\, dW_t$ takes values in $\mathscr{H}$. Of course, for more general setting, we can have W_t as a Wiener process on some space V and σ_t taking values in $\mathscr{L}(V, \mathscr{H})$. Here $\mathscr{A}$ is the infinitesimal generator of the C_0 semi-group. In order to make such SPDE well-defined, we have to find a way to define the Wiener process on a general Banach function space $\mathscr{H}$.

Process W_t is called $\mathscr{H}$ **Wiener process** if W_t takes values in $\mathscr{H}$, W has continuous trajectory, $W_0 = 0$, W has independent increments and $\forall t \geq s \geq 0, W_t - W_s \overset{d}{=} W_{t-s}$ and $W_t \overset{d}{=} -W_t$. Notice that the Wiener process in general space does not have anything to do with Gaussian any longer, but we require symmetricity w.r.t. 0.

It then follows that $\forall h \in \mathscr{H}$, we can define the quadratic variation process $\{\langle W_t, h\rangle\}_{t \geq 0}$. Notice that here $W_t \in \mathscr{H}$ is a function and $h \in \mathscr{H}$ is also a function so their quadratic variation can be defined. Different from the setting in $\mathbb{R}^n$, the bracket is not defined for the whole BM trajectory W, but for the value of the trajectory at only one time point W_t.

The good thing for such process is that it takes values in $\mathbb{R}$ and is still bilinear with the good properties of the quadratic variation process preserved. We see that

$$\mathbb{E} \langle W_t, u\rangle \langle W_s, u\rangle = t \wedge s \mathbb{E} \langle W_1, u\rangle^2 \tag{249}$$

calculated by

$$\begin{aligned}
\forall 0 \leq s \leq t, \mathbb{E} \langle W_t, u\rangle \langle W_s, u\rangle &= \mathbb{E} \langle W_t - W_s, u\rangle \langle W_s, u\rangle + \mathbb{E} \langle W_s, u\rangle^2 &(250)\\
&= \mathbb{E} \langle W_t - W_s, u\rangle \mathbb{E} \langle W_s, u\rangle + \mathbb{E} \langle W_s, u\rangle^2 &(251)\\
&= \mathbb{E} \langle W_{t-s}, u\rangle \mathbb{E} \langle W_s, u\rangle + \mathbb{E} \langle W_s, u\rangle^2 &(252)
\end{aligned}$$

where $W_s = \sum_{i=1}^{s}(W_i - W_{i-1})$ is the *i.i.d.* sum of s terms so

$$\mathbb{E}\left\langle W_s, u\right\rangle^2 = \mathbb{E}\left\langle \sum_{i=1}^{s}(W_i - W_{i-1}), u\right\rangle^2 \tag{253}$$

$$= \mathbb{E}\left[\sum_{i=1}^{s}\left\langle W_i - W_{i-1}, u\right\rangle\right]^2 \tag{254}$$

$$= \sum_{i,j=1}^{s}\mathbb{E}\left\langle W_i - W_{i-1}, u\right\rangle\left\langle W_j - W_{j-1}, u\right\rangle \tag{255}$$

$$= s\mathbb{E}\left\langle W_1, u\right\rangle^2 + s(s-1)\mathbb{E}^2\left\langle W_1, u\right\rangle \tag{256}$$

on the other hand

$$\mathbb{E}\left\langle W_s, u\right\rangle = \sum_{i=1}^{s}\mathbb{E}\left\langle (W_i - W_{i-1}), u\right\rangle \tag{257}$$

$$= \sum_{i=1}^{s}\mathbb{E}\left\langle W_1, u\right\rangle = s\mathbb{E}\left\langle W_1, u\right\rangle = 0 \tag{258}$$

since $\mathbb{E}\left\langle W_1, u\right\rangle = \mathbb{E}\left\langle -W_1, u\right\rangle$, this infers $2\mathbb{E}\left\langle W_1, u\right\rangle = 0$, so

$$\forall 0 \leq s \leq t, \mathbb{E}\left\langle W_t, u\right\rangle\left\langle W_s, u\right\rangle = s(t-s)\mathbb{E}^2\left\langle W_1, u\right\rangle + s\mathbb{E}\left\langle W_1, u\right\rangle^2 + s(s-1)\mathbb{E}^2\left\langle W_1, u\right\rangle \tag{259}$$

$$= s\mathbb{E}\left\langle W_1, u\right\rangle^2 \tag{260}$$

proves the equation. Similarly we can prove that for $u, v \in \mathscr{H}$,

$$\mathbb{E}\left\langle W_t, u\right\rangle\left\langle W_s, v\right\rangle = t \wedge s\mathbb{E}\left\langle W_1, u\right\rangle\left\langle W_1, v\right\rangle \tag{261}$$

From the perspective of Gaussian process for BM, we know that the covariance kernel determines the finite-dimensional distribution of the process. Now denote

$$\mathbb{E}\left\langle W_1, u\right\rangle\left\langle W_1, v\right\rangle = \left\langle Qu, v\right\rangle \tag{262}$$

and such $Q \in \mathscr{L}(\mathscr{H})$ is called **the covariance operator of the Gaussian measure** and we also call such W_t a Q-Wiener process.

Remark. *For more generalized Wiener process, one might first need to find the quadratic variation process as a linear bounded functional in $h \in \mathscr{H}$ and then apply Riesz representation theorem to define W_t as the unique element in $\mathscr{H}$ such that at each fixed time t the quadratic variation process is equal to $\langle W_t, h\rangle$.*

Mild Solution

For general SPDE

$$dX_t = (\mathscr{A} X_t + \alpha_t)\, dt + \sigma_t\, dW_t \tag{263}$$

we call the process

$$X_t = S(t)x_0 + \int_0^t S(t-s)\alpha_s\, ds + \int_0^t S(t-s)\sigma_s\, dW_s \tag{264}$$

a **mild solution to the SPDE**. The existence and uniqueness of mild solution requires Lipschitz condition and growth condition in α_s, σ_s along with some bound that $||S(t)||_{op} \le C\rho^{t\lambda}$. Here one can consider a more general form of SPDE that α_t, σ_t are changed into $\alpha(t, X_t), \sigma(t, X_t)$ so in this case $\alpha : [0, T] \times \mathscr{H} \to \mathscr{H}, \sigma : [0, T] \times \mathscr{H} \to \mathscr{L}(\mathscr{H})$ so there are similar results.

Remark. *Notice that here $\alpha(t, X_t)$ is actually $\alpha(t, X_t)(x)$ with space variable x always hidden in the notation. Recall that the C_0 semi-group element S (taken as the shift semi-group mentioned in the Musiela parametrization above) acts only on the space variable x such that*

$$[S(t-s)\alpha_s](x) = \alpha_s(x + t - s) \tag{265}$$

but won't change the time variable.

To make this mild solution well-defined, one has to notice that $\int_0^t S(t-s)\alpha_s\, ds$ is a Bochner integral as an extension of Lebesgue integral and $\int_0^t S(t-s)\sigma_s\, dW_s$ is an extension of Ito integral on a more general space. The definition of mild solution, as mentioned above, comes from the analogue of OU process.

Spectrum Decomposition of W_t by Covariance Operator

Since Q is self-adjoint, in Hilbert space $\mathscr{H}$ there exists $\{e_h\}$ as an orthonormal system in $\mathscr{H}$ and λ_h as real numbers such that

$$Qe_h = \lambda_h e_h \tag{266}$$

(e_h is eigenvector and λ_h is eigenvalue) and this naturally leads to

$$W_t = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\beta_j(t)e_j \tag{267}$$

where $\beta_j(t) = \frac{1}{\sqrt{\lambda_j}} \langle W_t, e_j\rangle$ are real-valued mutually independent (due to orthogonality between $\{e_h\}$) Wiener process. Such representation is well-defined if the infinite sum converges, i.e. in Hilbert space $\mathscr{H}$,

$$\sum_{j=1}^{\infty} (\sqrt{\lambda_j})^2 = \sum_{j=1}^{\infty} \lambda_j < \infty \tag{268}$$

This way is much easier to deal with in practice to build up an $\mathscr{H}$-Wiener process to do simulations. In order to find those eigenvalues from the data, PCA will always be helpful and can also provide some ideas on how to choose the orthonormal system.

Paper Summary: Pricing Options on Flow Forwards by Neural Networks in Hilbert Space

Consider forward contracts in electricity market over time period $[T_1, T_2]$ signed at time $t \leq T$ with price $\hat{F}(t, T_1, T_2)$. Turn to artificial fixed delivery contracts $F(t, T)$ and recover the forward price with $\hat{F}(t, T_1, T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} F(t, T)\, dt$ so now we only have to provide a model for the forward curve $F(t, T)$. Notice that using finite dimensional models for $\{F(t, \cdot)\}_{t \geq 0}$ is not a good choice since there are always more forwards added into the market so one always has to reformulate the model when there are new derivatives coming in. Instead, we use the infinite dimensional model given by the Musiela parametrization that assumes all derivatives are internally consistent

$$X(t, \xi) = F(t, t + \xi) \tag{269}$$

where ξ denotes the time to maturity and we model the forward curve $\{X(t, \cdot)\}_{t \geq 0}$, i.e. finding the functional relationship between X and ξ for each fixed time t.

Form the **state space** $\mathscr{X}$ as the space of weakly differentiable functions $x : \mathbb{R}_+ \to \mathbb{R}$ such that

$$||x||_W^2 = x(0)^2 + \int_0^\infty w(\xi) x'(\xi)^2\, d\xi < \infty \tag{270}$$

where the weight function w is non-decreasing and $w(0) = 1$. The intuition here is that we want the derivative of x to be fast decaying since we expect to see a forward of 30 and 31 years have almost the same price. The reason we are not using the classical L^2 space is that the points in L^2 space are equivalent classes under almost everywhere equal relationship. However, we want to have a model that can be evaluated at each time point in practice so being almost everywhere equal is not acceptable and we want the function to be in a good space with pointwise evaluation. Those two conditions are satisfied through the construction of $\mathscr{X}$ and such $\mathscr{X}$ has the structure of Banach algebra.

Now consider the forward curve $X(t) \in \mathscr{X}$ as a function in ξ, put up the **dynamics** with SPDE shown above

$$dX(t) = \partial_\xi X(t)\, dt + \mu(t, X(t))\, dt + \sigma(t, X(t))\, dW(t) \tag{271}$$

where $S_t = e^{t\partial_\xi}$ are elements in the C_0 semi-group. This SPDE has mild solution

$$X(t) = S_t X_0 + \int_0^t S_{t-s} \mu(s, X(s))\, ds + \int_0^t S_{t-s} \sigma(s, X(s))\, dW(s) \tag{272}$$

One step further, for **options written on such forward contracts with time of maturity** τ, the **option value function** $V(t, x)$ (value of the option at time t on seeing the current forward curve $x \in \mathscr{X}$ at time t) is given by

$$V(t, x) = \mathbb{E} q(X^{t,x}(\tau)) \tag{273}$$

for some Lipschitz function $q : \mathscr{X} \to \mathbb{R}$ mapping a function to real number. Here $X^{t,x}(\tau)$ is the solution to the SPDE model stated above with initial value condition on the forward curve at time t that $X(t) = x \in \mathscr{X}$. The option pricing problem now transforms into numerically solving for V. With a measure μ put up on the space $\mathscr{X}$ such that

$\mathbb{E}\left[\int_{\mathscr{X}} q^2(X^{t,x}(\tau))\mu(dx)\right] < \infty$, the variance decomposition gives

$$\mathbb{E}\left[\int_{\mathscr{X}} \left|q(X^{t,x}(\tau)) - g(x)\right|^2 \mu(dx)\right] = \int_{\mathscr{X}} Var(q^2(X^{t,x}(\tau)))\,\mu(dx) + \int_{\mathscr{X}} \left|V(t,x) - g(x)\right|^2 \mu(dx) \tag{274}$$

for $g : \mathscr{X} \to \mathbb{R}, g \in L^2(\mu)$. This provides a variational characterization for the option value function that

$$V(t,\cdot) = \underset{g\in L^2(\mu)}{\arg\min}\, \mathbb{E}\left[\int_{\mathscr{X}} \left|q(X^{t,x}(\tau)) - g(x)\right|^2 \mu(dx)\right] \tag{275}$$

it means that $V(t,x) = g(x)$ for fixed t if the argmin is achieved by $g(x)$. This minimization is numerically done by the Frechet space neural network that approximates the mapping from a function space to the set of real numbers with the Schauder basis of $\mathscr{X}$ specified.

For the following stochastic portfolio theory part, refer to *Portfolio Theory and Arbitrage: A Course in Mathematical Finance* by Karatzas and Kardaras for full details. Only key points, solution to exercises and motivations will be provided in the notes.

Basic Setting of Portfolio Theory

Probabilistic Setting

Throughout the context, assume $\mathscr{F}_0$ to be trivial in the filtration and the filtration to be right-continuous, i.e.

$$\forall t \in \mathbb{R}_+, \mathscr{F}_t = \mathscr{F}_{t+} \overset{def}{=} \bigcap_{s>t} \mathscr{F}_s \tag{276}$$

for a stochastic process $\{X_t\}$ (for asset price) adapted to the filtration $\{\mathscr{F}_t\}$, investment decision process shall always be predictable only based on past knowledge. A stochastic process is called **simple predictable** if it has the form $P_t = \sum_{j=1}^m h_j \mathbb{I}_{(t_{j-1},t_j]}(t)$ where $0 = t_0 < t_1 < ... < t_m$ with random variable $h_j \in \mathscr{F}_{t_{j-1}}$. Notice that here the time intervals are open on the left endpoint but closed on the right endpoint with h_j already realized at time t_{j-1}, the start of the time interval. General predictable process is defined through the **predictable sigma field** $\mathscr{P}$ as the smallest sigma field on $\Omega \times \mathbb{R}_+$ such that all simple predictable processes are measurable w.r.t. $\mathscr{P}$. As a result, a general stochastic process is said **predictable** if it's measurable w.r.t. $\mathscr{P}$.

Lemma 4 (Exercise 1.1)**.** *All left-continuous, adapted processes are predictable.*

Proof. X_t has left-continuous sample path and is adapted. Consider fixing $t \in \mathbb{R}_+$ and setting up a partition $0 = t_0^n < t_1^n < ... < t_{p(n)}^n = t$ where $t_i^n = \frac{i}{2^n}t, p(n) = 2^n$. Now construct the simple predictable process

$$P_t^n = \sum_{i=1}^{p(n)} X_{t_{i-1}^n} \mathbb{I}_{(t_{i-1}^n, t_i^n]}(t) \in \mathscr{P} \tag{277}$$

with

$$|X_t - P_t^n| \leq \sum_{i=1}^{p(n)} |X_t - X_{t_{i-1}^n}| \mathbb{I}_{(t_{i-1}^n, t_i^n]}(t) \tag{278}$$

since X_t is left-continuous at t, $\forall \varepsilon > 0, \exists \delta > 0, \forall t - \delta < s < t, |X_s - X_t| < \varepsilon$. Take $\delta < 2^{-n}$ to find $\sum_{i=1}^{p(n)} |X_t - X_{t_{i-1}^n}| \mathbb{I}_{(t_{i-1}^n, t_i^n]}(t) < \varepsilon$ so $|X_t - P_t^n| \leq \varepsilon, \lim_{n\to\infty} P_t^n = X_t \in \mathscr{P}$. $\square$

Remark. *Intuitively, a process P_t is predictable iff at each time t its value P_t only has something to do with $\{P_s\}_{s<t}$, which is obviously the case when a process is adapted and left-continuous since $P_t = \lim_{s\to t^-} P_s$.*

The notion of optionality is very similar to predictability. A stochastic process is called **simple optional** if it has the form $O_t = \sum_{j=1}^m h_j \mathbb{I}_{[t_{j-1},t_j)}(t)$ where $0 = t_0 < t_1 < ... < t_m$ with random variable $h_j \in \mathscr{F}_{t_{j-1}}$. Notice that here the time intervals are closed on the left endpoint but open on the right endpoint with h_j already realized at time t_{j-1}, the start of the time interval. General optional process is defined through the **optional sigma field** $\mathscr{O}$ as the smallest sigma field on $\Omega \times \mathbb{R}_+$ such that all simple optional processes are measurable w.r.t. $\mathscr{O}$. As a result, a general stochastic process is said **optional** if it's measurable w.r.t. $\mathscr{O}$.

Lemma 5 (Exercise 1.2)**.** *All right-continuous, adapted processes are optional. All simple predictable processes are optional so $\mathscr{P} \subset \mathscr{O}$.*

Proof. Similar to what is done above, $X_t = \lim_{s\to t^+} X_s$, notice that this does not cause measurability issues since $\lim_{s\to t^+} X_s \in \mathscr{F}_{t+} = \mathscr{F}_t$ under the right continuity assumption of filtration.

For simple predictable process

$$P_t = \sum_{j=1}^{m} h_j \mathbb{I}_{(t_{j-1},t_j]}(t) \tag{279}$$

to prove that it's optional, we just need to show that $h\mathbb{I}_{(a,b]}(t)$ is optional for $h \in \mathscr{F}_a, 0 \le a < b$. Construct a sequence $\{a_n\}$ such that $a_n \in (a,b), a_n \searrow a\ (n \to \infty)$, consider

$$O_t = \sum_{j=0}^{N} h\mathbb{I}_{[a_{j+1},a_j)}(t) \in \mathscr{O} \tag{280}$$

simple optional such that $\forall a < t < a_0, |O_t - h\mathbb{I}_{(a,b]}(t)| = |h\mathbb{I}_{(a,a_{N+1})}(t)| \to 0\ (N \to \infty)$. It shows that $h\mathbb{I}_{(a,b]}(t) \in \mathscr{O}$ since one can similarly deal with the right endpoint b and thus all simple predictable processes are optional.

□

Remark. *Predictable process must be optional and optional process must be progressive. In some sense, the set of all optional processes is the smallest set that contains all right-continuous adapted process which is closed under the limit of a sequence of process.*

Consider R as a $\mathbb{R}^n$-valued continuous semi-MG with components $R_1, ..., R_n$ starting from 0 with the decomposition $R_i(t) = M_i(t) + A_i(t)$ where M_i is a continuous local MG and A_i is a finite variation process. The **operational clock** for R is defined as

$$O_t = \sum_{i=1}^{n} \left(\int_0^t |\, dA_i(s)| + \langle M_i, M_i \rangle_t \right) \tag{281}$$

this process is scalar, continuous and increasing and most importantly, both processes $A_i, \langle M_i, M_j \rangle$ are absolute continuous w.r.t. this clock. Applying Radon-Nikodym theorem gives **predictable rate processes** $\alpha_1, ..., \alpha_n, c_{11}, ..., c_{1n}, ..., c_{nn}$ such that

$$A_t = \int_0^t \alpha_s \, dO_s \tag{282}$$

where $A = (A_1, ..., A_n)^T$ and $\alpha = (\alpha_1, ..., \alpha_n)^T$ are vector-valued processes.

$$C_t = \langle M, M \rangle_t = \int_0^t c_s \, dO_s \tag{283}$$

where C, c are matrix-valued processes with c_{ij} as the entries of c and $C_{ij} = \langle M_i, M_j \rangle$ as the covariation. Here we can understand R as the cumulative return of n stocks with the noise part (continuous local MG) and the drift part (finite variation process). As a result, α represent the local mean rate of return of the stocks and c represent the local covariation rates of the stocks. The introduction of the operational clock is just to turn price processes M, A into rate processes α, c, of course c with the interpretation of covariance matrix shall take value as symmetric SPD matrix.

The **stochastic exponential** $\mathcal{E}(Z)$ of continuous semi-MG Z with $Z_0 = 0$ is defined as the unique solution to the SDE

$$d\mathcal{E}(Z)_t = \mathcal{E}(Z)_t \, dZ_t \tag{284}$$

set $Y_t = \log \mathcal{E}(Z)_t$ and apply Ito formula to get

$$dY_t = dZ_t - \frac{1}{2} d\langle Z, Z\rangle_t \tag{285}$$

and solve out $\mathcal{E}(Z)_t = e^{Z_t - \frac{1}{2}\langle Z, Z\rangle_t}$. As the inverse of stochastic exponential, there is **stochastic logarithm** defined as

$$\mathcal{L}(Y)_t = \int_0^t \frac{1}{Y_s} \, dY_s \tag{286}$$

for strictly positive continuous semi-MG Y such that $Y_0 = 1$.

Lemma 6 (Exercise 1.5)**.** *For continuous semi-MG Z with $Z_0 = 0$, $\mathcal{L}(\mathcal{E}(Z)) = Z$. For continuous semi-MG Y with $Y_0 = 1$, $\mathcal{E}(\mathcal{L}(Y)) = Y$.*

Proof. Let $Y_t = \mathcal{E}(Z)_t$ then

$$\begin{aligned}
\mathcal{L}(Y)_t &= \int_0^t \frac{1}{Y_s} \, dY_s &(287)\\
&= \int_0^t \frac{1}{Y_s} Y_s \, dZ_s &(288)\\
&= Z_t &(289)
\end{aligned}$$

let $P_t = \mathcal{L}(Y)_t$ then

$$\begin{aligned}
\mathcal{E}(P)_t &= e^{P_t - \frac{1}{2}\langle P, P\rangle_t} = e^{\int_0^t \frac{1}{Y_s} \, dY_s - \frac{1}{2}\int_0^t \frac{1}{Y_s^2} \, d\langle Y, Y\rangle_s} &(290)\\
&= e^{\log Y_t} = Y_t &(291)
\end{aligned}$$

by the Ito formula expansion of $d \log Y_t$.

□

Lemma 7 (Exercise 1.7)**.** *For any two continuous semi-MG X, Z with $X_0 = Z_0 = 0$, there is* ***Yor formula***

$$\mathcal{E}(X)\mathcal{E}(Z) = \mathcal{E}(X + Z + \langle X, Z\rangle) \tag{292}$$

Proof.

$$\mathcal{E}(X)\,\mathcal{E}(Z) = e^{X+Z-\frac{1}{2}(\langle X,X\rangle+\langle Z,Z\rangle)} \tag{293}$$

$$= e^{X+Z+\langle X,Z\rangle-\frac{1}{2}(\langle X,X\rangle+2\langle X,Z\rangle+\langle Z,Z\rangle)} \tag{294}$$

$$= e^{X+Z+\langle X,Z\rangle-\frac{1}{2}\langle X+Z,X+Z\rangle} \tag{295}$$

$$= e^{X+Z+\langle X,Z\rangle-\frac{1}{2}\langle X+Z+\langle X,Z\rangle,X+Z+\langle X,Z\rangle\rangle} \tag{296}$$

$$= \mathcal{E}(X+Z+\langle X,Z\rangle) \tag{297}$$

□

To consider asset price with jumps, one might consider general semi-MG without continuous sample path with Cadlag modification. Similarly, such semi-MG can be decomposed into two parts, one as continuous local MG and the other as a finite variation process with jumps. One define the jump process $\Delta X_t = X_t - X_{t-}$ and the covariation between two semi-MG is now defined as $\langle X,Z\rangle_t = \langle X^c, Z^c\rangle_t + \sum_{s\le t}\Delta X_s \Delta Z_s$ where X^c denotes the continuous local MG part of X.

Assets

Now we understand continuous semi-MG R as the **cumulative return of n stocks** stated above with $R_i(0) = 0$. Now when there's both components and time appearing, we write the component index as subscript and the time as a variable. The **stock price process** S_i then has dynamics

$$S_i = S_i(0)\mathcal{E}\left(R_i\right) \tag{298}$$

by the definition of stochastic exponential, it's equivalent to saying

$$dS_i(t) = S_i(0)\mathcal{E}\left(R_i\right)(t)\, dR_i(t) = S_i\, dR_i(t) \tag{299}$$

which looks similar to the Black Scholes model we are familiar with (by setting $dR_t = \mu\, dt + \sigma\, dW_t$ one gets the BS model). As long as $S_i(0) > 0$, we have $S_i(t) > 0$. Take log on both sides

$$\log S_i = \log S_i(0) + \log \mathcal{E}\left(R_i\right) = \log S_i(0) + R_i - \frac{1}{2}\left\langle R_i, R_i\right\rangle \tag{300}$$

to see that

$$\log S_i = \log S_i(0) + A_i - \frac{1}{2}C_{ii} + M_i \tag{301}$$

since $\langle R_i, R_i\rangle = \langle M_i, M_i\rangle = C_{ii}$ by definition. Notice that $C_{ii} = \langle \log S_i, \log S_i\rangle$ since M_i is the only continuous local MG part on the RHS of the equation. We define

$$\Gamma_i \overset{def}{=} A_i - \frac{1}{2}C_{ii} \tag{302}$$

as the **cumulative growth process** for stock i.

Remark. *The reason why we call Γ_i as cumulative growth process comes from the fact that it's the drift part in the representation of $\log S_i$, i.e. the asymptotic behavior of Γ_i determines the asymptotic behavior of S_i. When $\Gamma_i(t) \to +\infty$ $(t \to \infty)$, we see that $S_i(t) \to +\infty$ $(t \to \infty)$ is always growing. On the other hand, when $\Gamma_i(t) \to -\infty$ $(t \to \infty)$, we see that $S_i(t) \to 0$ $(t \to \infty)$ shrinks to 0.*

Naturally, take the Radon-Nikodym derivative of Γ_i w.r.t. the operational clock O to get the **local growth rate** γ_i such that

$$\Gamma_i(t) = \int_0^t \gamma_i(s)\, dO_s \tag{303}$$

it immediately follows that

$$\gamma_i = \alpha_i - \frac{1}{2}c_{ii} \tag{304}$$

where α_i is the **local mean rate of return** and c_{ij} is the **local covariation rate** defined in the previous context.

Example: Black Scholes Model with Single Stock

Till here, we have finished stating the setup for modelling the asset price. Let's then look at an example how the conventional **BS model for a single stock** is represented under those settings.

Now $n = 1$ so we use subscript for time, and start from building the continuous semi-MG as cumulative return of stock that

$$dR_t = \mu \, dt + \sigma \, dW_t \tag{305}$$

solving this SDE gives us

$$R_t = \mu t + \sigma W_t \tag{306}$$

interpreted as that the cumulative return of the stock till time t is equal to the sum of the drift part μt linear in time with μ as the mean return rate of the stock and the diffusion part σW_t linear in BM W_t with σ as the volatility of the stock. This R is natural as the simplest model since it assumes that the drift and risk in stock price is uniform in time. Now the decomposition of semi-MG gives $M_t = \sigma W_t, A_t = \mu t, C_t = \sigma^2 t$. Let's check the operational clock defined as

$$O_t = \int_0^t |\, dA_s| + \langle M, M \rangle_t = \mu t + \sigma^2 t \tag{307}$$

so it's a linear function in t. For simplicity, let's choose $O_t = t$ (the most often used **Lebesgue clock**) instead since it's still positive increasing such that $A_t, \langle M, M \rangle_t$ are absolute continuous w.r.t. O_t. The Lebesgue clock is just the time in our real life so we don't have to change our time scale for this model and all processes above can be turned into local rate processes in a well-defined manner. (If the $A_t, \langle M, M \rangle_t$ are no longer linear in time t, one might have to use another operational clock under which time runs in different speed from that in the reality such that the rates are well-defined)

Under such operational clock, the rate functions are just derivatives of A_t, C_t w.r.t. t so

$$\alpha_t = \mu, c_t = \sigma^2 \tag{308}$$

the rate processes are constant. The local growth rate is thus

$$\gamma_t = \alpha_t - \frac{1}{2} c_t^2 = \mu - \frac{\sigma^2}{2} \tag{309}$$

and the stock price dynamics becomes

$$\frac{dS_t}{S_t} = dR_t = \mu \, dt + \sigma \, dW_t \tag{310}$$

whose solution is given by

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} \tag{311}$$

now it's clear that when the local growth rate is positive, $S_t \to +\infty$ $(t \to \infty)$ and when it's negative, $S_t \to 0$ $(t \to \infty)$. This example should provide all intuition for the concepts defined above.

Investment and Admissibility

An investor has initial capital $x \in \mathbb{R}_+$ and his control process is predictable and vector valued denoted $\theta = (\theta_1, ..., \theta_n)^T$ with $\theta_i(t)$ to denote **the number of shares of stock i in the investor's portfolio at time t** so it stands for the investment strategy. The investor cares about his **total wealth process** defined as

$$X(t; x, \theta) \stackrel{def}{=} x + \int_0^t \theta(s)\, dS(s) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) \tag{312}$$

one might be confused since this seems different with the one in Merton's problem. Actually they are the same except that here we **assume a zero interest rate** for simplicity so the part of wealth that is invested into the money market will not accumulate interest.

For given initial wealth x and strategy θ, $X(t; x, \theta)$ is **admissible** if solvency $X(t; x, \theta) \geq 0$. In other words, we stick to the practice that **there is a credit constraint**. Luckily, this assumption actually makes our life easier by ensuring that the solvency exists. Other common possible constraints include a long-only investment strategy $\forall i, \theta_i \geq 0$ or prohibiting borrowing from money market $\forall t, X(t; x, \theta) - \sum_{i=1}^n \theta_i(t) S_i(t) \geq 0$ that fits in with the reality in some situations.

An example of **doubling strategy** in the book tells us what might happen if infinite credit is allowed. Consider a single stock $n = 1$ with the cumulative return $R_t = W_t$ as the standard BM, restrict to the time interval $[0, 1)$ and define

$$\eta_t = \frac{1}{S_t\sqrt{1-t}}, N_t = \int_0^t \eta_s\, dS_s = \int_0^t \frac{1}{\sqrt{1-s}}\, dR_s \tag{313}$$

with η to be predictable and N to be a continuous local MG (since it's a stochastic integral w.r.t. BM). Now consider the process (any continuous local MG is a time-changed BM)

$$B_t = N_{1-e^{-t}} \tag{314}$$

since $\langle B, B\rangle_t = \int_0^{1-e^{-t}} \frac{1}{1-s}\, d\langle R, R\rangle_s = \int_0^{1-e^{-t}} \frac{1}{1-s}\, ds = t$,by Levy's characterization, it is a BM in its natural filtration $\left\{\mathscr{F}_t^B\right\}$. Define the stopping time τ^m as the first time N is above m

$$\tau^m = \inf\left\{t \in [0, 1) : N_t \geq m\right\} \tag{315}$$

and consider the investment strategy as a predictable process

$$\theta_t^m = \eta_t \mathbb{I}_{(0,\tau^m]}(t) \tag{316}$$

that holds η_t share of stock when N_t is below m and does not hold any stock when N_t is above m. Simple calculations give

$$X_1 = x + \int_0^1 \theta_t^m\, dS_t = x + \int_0^{\tau^m} \eta_t\, dS_t = x + N_{\tau^m} = x + m \tag{317}$$

so starting from initial wealth x, we can always reach wealth $x + m$ at time 1 for some positive m.

The point here is that $X_t = x + N_{t\wedge\tau^m}$ is a local MG but is **not bounded from below**(if bounded from below, local MG is super-MG so $x = \mathbb{E}X_0 \geq \mathbb{E}X_1 = x + m$ for some positive m), resulting in the situation where infinitely deep credit line exists, i.e. one can borrow as much as they want from the money market even if they have already been bankrupt. The lack of credit constraint enables one to reach any level of wealth at any deterministic time they want. On the other hand, if admissibility is ensured, it makes X_t a non-negative local MG, thus a super-MG, by optional stopping theorem, $x = \mathbb{E}X_0 \geq \mathbb{E}X_\tau$ for any almost surely finite stopping time τ, the doubling strategy above no longer works.

Remark. *Actually, admissibility can be weakened a little bit to just ensure that the total wealth process is bounded from below. What the doubling strategy is doing is actually buying the stock when it's cheap and selling it when it's more valuable. Typically when one has finite credit constraint, this does not bring with profit because of the optional stopping theorem but when one has infinite credit one can always borrow "enough amount of money" to make it work even if the stock has no cumulative return on average (the R_t above has no finite variation part, only pure noise).*

However, notice that here η_t violates the integrability condition at $t = 1$. That's why doubling strategy works on finite time horizon.

Capital Withdrawal

Cumulative capital withdrawal till time t is denoted by an increasing adapted right-continuous process $K_t \in \mathscr{K}$ so the total wealth at time t considering capital withdrawal has the form

$$X(t; x, \theta, K) \overset{def}{=} x + \int_0^t \theta(s)\, dS(s) - K(t) \tag{318}$$

a simple subtraction of cumulative amount $K(t)$. The total wealth process is said to **finance** $K \in \mathscr{K}$ if $X \geq K$ holds (one can afford the capital withdrawal under the admissible condition). Naturally, we define a subset $\mathscr{K}(x)$ of the set of all possible cumulative capital withdrawal process $\mathscr{K}$ such that $\mathscr{K}(x)$ consists of all $K \in \mathscr{K}$ **financeable from initial capital** x **under some strategy** θ, i.e. $\exists \theta, \exists F \in \mathscr{K}, \forall t \in \mathbb{R}_+, X(t; x, \theta, F) \geq K(t)$. Equivalently, we can write

$$\mathscr{K}(x) \overset{def}{=} \{K \in \mathscr{K} : \exists \theta, \forall t \in \mathbb{R}_+, X(t; x, \theta, K) \geq 0\} \tag{319}$$

some properties for the capital withdrawal processes $K \in \mathscr{K}(x), x \in \mathbb{R}_+$ from the definition above are that

- The trivial capital withdrawal process $0 \in \mathscr{K}(0)$
- $\mathscr{K}(x)$ is convex in x since $X(\cdot; x, \theta, K)$ is linear in K
- $\forall x_1 \leq x_2, \mathscr{K}(x_1) \subset \mathscr{K}(x_2)$, the capital withdrawal financeable from less initial capital must also be financeable from more initial capital
- $\forall x > 0, \mathscr{K}(x) = x\mathscr{K}(1)$ is homogeneous in x
- $\forall x, K \in \mathscr{K}(x)$ if $\exists \overline{K} \in \mathscr{K}, \overline{K} \leq K$, it's always true that $\overline{K} \in \mathscr{K}(x)$, if K is financeable then capital withdrawal less than K must be financeable

Money Market and Zero Interest Rate Assumption

We have previously made the assumption that the money market always has zero interest rate. Here let's argue why this assumption makes sense, we will see that the interest rate does not make a difference in our setting above.

Assume $\check{R}_0$ is an adapted continuous process of finite variation with $\check{R}_0(0) = 0$ that denotes the cumulative return of the riskless asset in the market (let's say, bond, or bank account). Now under this non-trivial money market setting let $\check{R}_i$ denote the cumulative return of stock i to be continuous semi-MG with $\check{R}_i(0) = 0$ as mentioned above. The asset price dynamics is given by

$$\frac{d\check{S}_i(t)}{\check{S}_i(t)} = d\check{R}_i(t), i \in \{0, 1, ..., n\} \tag{320}$$

where $\check{S}_0(0) = 1$ is required since 1 dollar at time 0 obviously has the value of 1 dollar. Now consider $R_i = \check{R}_i - \check{R}_0, S_i = \frac{\check{S}_i}{\check{S}_0}$ so by Ito formula

$$dS_i(t) = \frac{1}{\check{S}_0(t)}\, \check{S}_i(t) - \frac{\check{S}_i(t)}{[\check{S}_0(t)]^2}\, d\check{S}_0(t) \tag{321}$$

$$= S_i(t)\, d\check{R}_i(t) - S_i(t)\, d\check{R}_0(t) \tag{322}$$

$$= S_i(t)\, dR_i(t) \tag{323}$$

since $\check{R}_0$ has finite variation. This recovers the model in the previous context and such R_i stands for the cumulative return of stock i **in excess of the interest paid by the money market** (subtracted the opportunity cost).

Consider the total wealth process $\check{X}$ with capital withdrawal $\check{K}$, we find

$$\check{X}(t; x, \theta, \check{K}) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, d\check{S}_i(s) + \int_0^t \left(\check{X}(s) - \sum_{i=1}^n \theta_i(s)\check{S}_i(s) \right) d\check{R}_0(s) - \check{K}(t) \tag{324}$$

$$= x + \int_0^t \sum_{i=1}^n \theta_i(s)\, d\check{S}_i(s) + \int_0^t \left(\check{X}(s) - \check{S}_0(s) \sum_{i=1}^n \theta_i(s) S_i(s) \right) d\check{R}_0(s) - \check{K}(t) \tag{325}$$

where the second term in the first equality on the RHS comes from investing all remaining wealth into the money market. Let's define $X(\cdot; x, \theta, K) = \frac{\check{X}(\cdot; x, \theta, \check{K})}{\check{S}_0}$ and $K(t) = \int_0^t \frac{1}{\check{S}_0(s)}\, d\check{K}(s)$.

Lemma 8 (Exercise 1.14)**.** *Under such setting, we have*

$$X(t; x, \theta, K) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) - K(t) \tag{326}$$

consistent with the total wealth process defined in the zero interest rate case.

Proof. Differentiate the equation for $\check{X}$ above w.r.t. t

$$d\check{X}(t) = \sum_{i=1}^{n} \theta_i(t)\, d\check{S}_i(t) + \left(\check{X}(t) - \check{S}_0(t) \sum_{i=1}^{n} \theta_i(t) S_i(t) \right) d\check{R}_0(t) - \, d\check{K}(t) \tag{327}$$

since $\frac{\check{X}(t)}{\check{S}_0(t)} = X(t)$, we know that by Ito formula

$$\begin{aligned}
dX(t) &= \frac{1}{\check{S}_0(t)}\, d\check{X}(t) - \frac{\check{X}(t)}{[\check{S}_0(t)]^2}\, d\check{S}_0(t) && (328)\\
&= \frac{1}{\check{S}_0(t)} \sum_{i=1}^{n} \theta_i(t)\, d\check{S}_i(t) + \left(X(t) - \sum_{i=1}^{n} \theta_i(t) S_i(t) \right) d\check{R}_0(t) - \frac{1}{\check{S}_0(t)}\, d\check{K}(t) - X(t)\, d\check{R}_0(t) && (329)\\
&= \frac{1}{\check{S}_0(t)} \sum_{i=1}^{n} \theta_i(t)\, d\check{S}_i(t) - \sum_{i=1}^{n} \theta_i(t) S_i(t)\, d\check{R}_0(t) - \, dK(t) && (330)\\
&= \sum_{i=1}^{n} \theta_i(t) \left(\frac{d\check{S}_i(t)}{\check{S}_0(t)} - S_i(t)\, d\check{R}_0(t) \right) - \, dK(t) && (331)\\
&= \sum_{i=1}^{n} \theta_i(t)\, dS_i(t) - \, dK(t) && (332)
\end{aligned}$$

where the second last equation follows from the fact we have applied previously that

$$dS_i(t) = S_i(t)\, dR_i(t) = S_i(t)\, d\check{R}_i(t) - S_i(t)\, d\check{R}_0(t) = \frac{d\check{S}_i(t)}{\check{S}_0(t)} - S_i(t)\, d\check{R}_0(t) \tag{333}$$

since $\check{R}_0$ has finite variation. Notice that $X(0) = \frac{\check{X}(0)}{\check{S}(0)} = x$ since $\check{S}(0) = 1$ concludes the proof.

$\square$

It's not hard to see that our previous zero interest rate model makes sense and is different only by a simple transformation from the general situation. That's why it's reasonable to assume there's no interest rate in the money market and we stick to this assumption which can also be denoted as $R_0 \equiv 0, S_0 \equiv 1$.

Numeraire and Portfolio

The definition of numeraire and portfolio has the motivation coming from **proportional investment**, referring to investors not caring about the number of shares they invest but only care about the proportion of wealth they invest into each asset. As a result, θ_i as the number of shares invested in stock i does not meet our requirement and we define

$$\pi_i \stackrel{def}{=} \frac{S_i\theta_i}{X} \tag{334}$$

as a predictable, vector-valued process with $\pi_i(t)$ to be **the proportion of total wealth invested in stock i at time t**. Under the proportional investment setting, π_i shall be the true control of the investor that affects θ_i.

Naturally, to ensure such π_i is well-defined, we have to make sure that $\forall t \in \mathbb{R}_+, X_i(t) > 0$. This gives rise to the definition of numeraire that a wealth process is called **numeraire** if it starts with initial wealth 1 and remains strictly positive all the time (different concept with admissibility where we just require the wealth process to be non-negative).

Remark. *Notice that $\sum_{i=1}^n \pi_i(t)$ is the total proportion of wealth in all the stocks at time t but it does not have to be 1. If this sum is less than 1, the remaining wealth will be invested in the money market and results in zero interest. If this sum is larger than 1, the investor is actually leveraging.*

The total wealth process X can now be represented using π instead of θ

$$X(t) = 1 + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) = 1 + \int_0^t \sum_{i=1}^n \frac{\pi_i(s)X(s)}{S_i(s)}\, dS_i(s) = 1 + \int_0^t X(s) \sum_{i=1}^n \pi_i(s)\, dR_i(s) \tag{335}$$

recall the fact that $dR_i(t) = \frac{dS_i(t)}{S_i(t)}$. Let's denote

$$R_\pi(t) \stackrel{def}{=} \int_0^t \pi^T(s)\, dR(s) = \int_0^t \sum_{i=1}^n \pi_i(s)\, dR_i(s) \tag{336}$$

as the **cumulative return weighted by** π (a scalar process), then it's clear that $dX(t) = X(t)\, dR_\pi(t)$ can be written in terms of stochastic exponential $X_\pi = \mathcal{E}(R_\pi)$ as the total wealth process following proportion π. A predictable vector-valued process π is called a **portfolio** if the total wealth process it generates, i.e.

$$X_\pi = \mathcal{E}(R_\pi) \tag{337}$$

is a numeraire.

Remark. *Examples of portfolio includes the constant portfolio $\pi \equiv p \in \mathbb{R}^n$ keeping constant proportion p_i of wealth in stock i throughout the time horizon, the equal-weighted portfolio $\pi \equiv \frac{1}{n}\vec{1}$ always putting equal proportion of wealth in all stocks.*

Nevertheless, portfolios can also have constraints just as θ_i does. Long-only portfolios requires π to take values in the simplex

$$\Delta_n = \left\{ x \in \mathbb{R}^n : x \geq 0, x^T \vec{1} \leq 1 \right\} \tag{338}$$

where $\geq$ between vectors is component-wise. Stock portfolio requires π to take values in the simplex

$$H_{n-1} = \left\{ x \in \mathbb{R}^n : x^T \vec{1} = 1 \right\} \tag{339}$$

so all wealth shall be invested in stock market with no leverage allowed. If one combine those two constraints, the long-only stock portfolio can only take values in

$$\Delta_{(n-1)} = \left\{ x \in \mathbb{R}^n : x \geq 0, x^T \vec{1} = 1 \right\} \tag{340}$$

the probability simplex.

Since we have defined the weighted version of cumulative return R_π, it's possible to generalize all the previously defined rates into their weighted versions. For two portfolios π, ρ, R_π, R_ρ are continuous semi-MG with decomposition $R_\pi = M_\pi + A_\pi, R_\rho = M_\rho + A_\rho$

$$C_{\pi\rho} \overset{def}{=} \langle R_\pi, R_\rho \rangle, \Gamma_\pi \overset{def}{=} A_\pi - \frac{1}{2} C_{\pi\pi} \tag{341}$$

and take Radon-Nikodym derivative w.r.t. the operational clock to get rates $\alpha_\pi, c_{\pi\rho}, \gamma_\pi$ as the **mean rate of return, covariation rate, growth rate of portfolio** respectively. It's quite obvious that $\alpha_\pi = \pi^T \alpha, c_{\pi,\rho} = \pi^T c \rho$. In the following context, $c_{i\rho}$ denotes the covariation rate between portfolio e_i and portfolio ρ, note that this notation is consistent with the previously defined notation since the covariation rate between e_i, e_j is just c_{ij}, the (i, j) entry of matrix c.

Excess Growth

For the rates of the portfolio defined above, it's clear that

$$\gamma_\pi = \alpha_\pi - \frac{1}{2}c_{\pi\pi} = \pi^T\alpha - \frac{1}{2}\pi^T c\pi \tag{342}$$

however, it we try to directly calculate the weighted average of the growth rate

$$\pi^T\gamma = \sum_{i=1}^n \pi_i\gamma_i = \sum_{i=1}^n \pi_i\left(\alpha_i - \frac{1}{2}c_{ii}\right) = \pi^T\alpha - \frac{1}{2}\pi^T\mathrm{diag}(c) \neq \gamma_\pi \tag{343}$$

the weighted average of growth rate is not equal to the growth rate of the portfolio and we call the difference **excess growth rate** γ_π^*.

$$\gamma_\pi^* \overset{def}{=} \gamma_\pi - \pi^T\gamma = \frac{1}{2}(\pi^T\mathrm{diag}(c) - \pi^T c\pi) \tag{344}$$

integrate this excess growth rate w.r.t. the operational clock gives the **cumulative excess growth** $\Gamma_\pi^*(t) = \int_0^t \gamma_\pi^*(s)\,dO(s)$.

Lemma 9 (Exercise 1.19)**.** *Long-only portfolio π has non-negative excess growth rate.*

Proof. Now $\forall t\in\mathbb{R}_+, \forall i, \pi_i(t)\geq 0, \sum_{i=1}^n \pi_i(t)\leq 1$. Notice that $c_{ij}\leq\sqrt{c_{ii}c_{jj}}$ by Cauchy-Schwarz, and $\sum_{i=1}^n \pi_i c_{ii}\geq 0$

$$2\gamma_\pi^* = \sum_{i=1}^n \pi_i c_{ii} - \sum_{i,j=1}^n \pi_i c_{ij}\pi_j \tag{345}$$

$$\geq \sum_{i=1}^n \pi_i c_{ii} - \sum_{i,j=1}^n \pi_i\sqrt{c_{ii}c_{jj}}\pi_j \tag{346}$$

$$= \sum_{i=1}^n \pi_i c_{ii} - \left(\sum_{i=1}^n \pi_i\sqrt{c_{ii}}\right)^2 \tag{347}$$

$$\geq \sum_{i=1}^n \pi_i c_{ii} - \sum_{i=1}^n \pi_i c_{ii}\sum_{i=1}^n \pi_i \tag{348}$$

$$\geq 0 \tag{349}$$

where the second inequality still follows from Cauchy-Schwarz by viewing $\pi\sqrt{c_{ii}} = (\sqrt{\pi_i}\sqrt{c_{ii}})(\sqrt{\pi_i})$. □

Lemma 10 (Exercise 1.20)**.** *$\pi^1, ..., \pi^l$ are l portfolios in a market with n assets available, and π^k generates the wealth process $V^k \equiv X_{\pi^k}$ with $V^k(0) = 1$. Now $V^1, ..., V^l$ are all continuous semi-MG so they form an aggregated market with ρ as a portfolio on such aggregated market. Naturally, ρ ($\mathbb{R}^l$-valued process) induces a portfolio Π ($\mathbb{R}^n$-valued process) in the original market that $\Pi_i = \sum_{k=1}^l \rho^k\pi_i^k$. Then the excess growth rate of Π in the original market is*

$$\gamma_\Pi^* = \sum_{k=1}^l \rho^k\gamma_{\pi^k}^* + \frac{1}{2}\left(\sum_{k=1}^l \rho^k c_{\pi^k,\pi^k} - \sum_{k=1}^l\sum_{j=1}^l \rho^k c_{\pi^k\pi^j}\rho^j\right) \tag{350}$$

Proof.

$$
\begin{aligned}
\gamma_\Pi^* &= \frac{1}{2}\sum_{i=1}^n \Pi_i c_{ii} - \frac{1}{2}\sum_{i,j=1}^n \Pi_i c_{ij}\Pi_j && (351)\\
&= \frac{1}{2}\sum_{k=1}^l\sum_{i=1}^n \rho^k\pi_i^k c_{ii} - \frac{1}{2}\sum_{k=1}^l\sum_{p=1}^l\sum_{i,j=1}^n \rho^k\pi_i^k c_{ij}\rho^p\pi_j^p && (352)\\
&= \sum_{k=1}^l \rho^k\left(\gamma_{\pi^k}^* + \frac{1}{2}\sum_{i,j=1}^n \pi_i^k c_{ij}\pi_j^k\right) - \frac{1}{2}\sum_{k=1}^l\sum_{j=1}^l \rho^k c_{\pi^k\pi^j}\rho^j && (353)\\
&= \sum_{k=1}^l \rho^k\gamma_{\pi^k}^* + \frac{1}{2}\left(\sum_{k=1}^l \rho^k c_{\pi^k,\pi^k} - \sum_{k=1}^l\sum_{j=1}^l \rho^k c_{\pi^k\pi^j}\rho^j\right) && (354)
\end{aligned}
$$

□

Remark. *In the original market there are n stocks, so each portfolio can be understood as a fund. In the example above there are l different funds and the investor is constructing a portfolio in the fund market, that's called **fund-of-funds investing** as investing in a higher-level aggregate market.*

Market Portfolio

The market portfolio is an important concept in finance used in deriving asset pricing models like CAPM etc. The idea is to construct a portfolio such that the weight of each stock in the portfolio is proportional to its market capitalization. The market portfolio can be understood as a concept of equilibrium meaning on the financial market, i.e. everybody shall hold the market portfolio when the stock market reaches an economic equilibrium.

In the framework built above, we first build up the **relative market capitalizations** and the **total capitalization**

$$\mu_i \overset{def}{=} \frac{S_i}{\Sigma}, \Sigma \overset{def}{=} \sum_{i=1}^{n} S_i \tag{355}$$

where μ_i is the market weight process of stock i. As a result, μ is defined as the **market portfolio**. It's predictable, bounded and actually a long-only stock portfolio $\mu \in \Delta_{(n-1)}$.

We can verify that the cumulative return generated by the market portfolio is

$$dR_\mu(t) = \sum_{i=1}^{n} \mu_i(t) \frac{1}{S_i(t)}\, dS_i(t) = \sum_{i=1}^{n} \frac{1}{\Sigma(t)}\, dS_i(t) = \frac{d\Sigma(t)}{\Sigma(t)} \tag{356}$$

solve the SDE to get $R_\mu(t) = \int_0^t \frac{1}{\Sigma(s)}\, d\Sigma(s) = \mathcal{L}\left(\frac{\Sigma}{\Sigma(0)}\right)(t)$, so the total wealth process generated is

$$X_\mu = \mathcal{E}\left(R_\mu\right) = \frac{\Sigma}{\Sigma(0)} = \frac{S_1 + ... + S_n}{S_1(0) + ... + S_n(0)} \tag{357}$$

since stochastic exponential and stochastic logarithm eliminates iff $\frac{\Sigma}{\Sigma(0)}(0) = 1$ the process has initial value 1. The total wealth process generated by market portfolio is just **the sum of market capitalization of all stocks divided by the initial sum of market capitalization of all stocks**. In other words, one just buy the same share of all stocks at time 0 and hold them without doing any modifications. This can also be seen by

$$\theta_i(t) = \frac{\mu_i(t)X(t)}{S_i(t)} = \frac{1}{\Sigma(0)} \tag{358}$$

is constant in all time and stocks and deterministic that's why it's a very nice characterization of the **buy-and-hold** strategy.

Relative Performance

Now we want to compare a portfolio π with a **baseline portfolio** ρ, so we introduce the **relative wealth process**

$$X_\pi^\rho \overset{def}{=} \frac{X_\pi}{X_\rho} \tag{359}$$

the bigger value it takes, the better portfolio π is performing compared to the baseline. This relative wealth process induces the relative version of all processes defined above compared to the baseline.

Imagine an **auxiliary market** where the stock prices are $S_i^\rho \overset{def}{=} \frac{S_i}{X_\rho}$ for $0 \leq i \leq n$. Define the **relative cumulative returns** of each stock to be $R_i^\rho \overset{def}{=} R_0^\rho + (R_i - C_{i\rho})$ $(1 \leq i \leq n)$ where $R_0^\rho \overset{def}{=} C_{\rho\rho} - R_\rho$. Let's verify that all processes in this auxiliary market sticks still have the connections shown above.

Firstly, let's verify $S_i^\rho = S_i(0)\mathcal{E}\left(R_i^\rho\right)$, the relationship between cumulative return and stock price. When $i = 0$,

$$\mathcal{E}\left(R_0^\rho\right) = e^{C_{\rho\rho} - R_\rho - \frac{1}{2}\langle R_\rho, R_\rho\rangle} = \frac{1}{\mathcal{E}\left(R_\rho\right)} = \frac{S_0}{X_\rho} = S_0^\rho \tag{360}$$

since $S_0 \equiv 1, X_\rho = \mathcal{E}\left(R_\rho\right)$. When $1 \leq i \leq n$,

$$\mathcal{E}\left(R_i^\rho\right) = e^{C_{\rho\rho} - R_\rho + R_i - C_{i\rho} - \frac{1}{2}\langle R_i - R_\rho, R_i - R_\rho\rangle} = e^{R_i - R_\rho + \frac{1}{2}(C_{\rho\rho} - C_{ii})} = \frac{\mathcal{E}\left(R_i\right)}{\mathcal{E}\left(R_\rho\right)} = \frac{\frac{S_i}{S_i(0)}}{X_\rho} = \frac{S_i^\rho}{S_i(0)} \tag{361}$$

concludes the verification.

Secondly, let's calculate $R_\pi^\rho = \pi^T R^\rho$

$$R_\pi^\rho = \sum_{i=0}^n \pi_i R_i^\rho = \pi_0 R_0^\rho + \sum_{i=1}^n \pi_i (R_0^\rho + R_i - C_{i\rho}) = \left(\sum_{i=0}^n \pi_i\right) R_0^\rho + R_\pi - C_{\pi\rho} \tag{362}$$

$$= C_{\rho\rho} - R_\rho + R_\pi - C_{\pi\rho} = R_{\pi-\rho} - C_{\pi-\rho,\rho} \tag{363}$$

notice that in the previous context we are considering $\pi_1, ..., \pi_n$ and allow the sum $\sum_{i=1}^n \pi_i$ to take any values without special constraint. Essentially, this is due to the fact that we are always setting $\pi_0 = 1 - \sum_{i=1}^n \pi_i$ to clear our position in the money market. As a result, $\sum_{i=0}^n \pi_i = 1$ must hold in general.

Remark. *In such auxiliary market we can never ignore all the processes with subscript 0 any longer since it's no longer the trivial money market after comparing to the baseline ρ since*

$$S_0^\rho = \frac{S_0}{X_\rho} = \frac{1}{X_\rho}, R_0^\rho = C_{\rho\rho} - R_\rho \tag{364}$$

that's why we always have to consider the subscript running from 0 to n in the auxiliary market.

One last thing to verify is that $X_\pi^\rho = \mathcal{E}\left(R_\pi^\rho\right)$.

$$\mathcal{E}\left(R_\pi^\rho\right) = e^{R_{\pi-\rho} - C_{\pi-\rho,\rho} - \frac{1}{2}\langle R_{\pi-\rho}, R_{\pi-\rho}\rangle} = e^{R_\pi - R_\rho - C_{\pi-\rho,\rho} - \frac{1}{2}C_{\pi-\rho,\pi-\rho}} \tag{365}$$

$$= \frac{\mathcal{E}\left(R_\pi\right)}{\mathcal{E}\left(R_\rho\right)} e^{\frac{1}{2}C_{\pi\pi} - \frac{1}{2}C_{\rho\rho} - C_{\pi-\rho,\rho} - \frac{1}{2}C_{\pi-\rho,\pi-\rho}} = \frac{X_\pi}{X_\rho} \tag{366}$$

so everything is still consistent with our setting without the baseline! In other words, we can **view the baseline** ρ **as a condition on the market**. Under such condition, only cumulative returns are changed from R into R^ρ but all the generation rules stay the same (notice that $X^0 \equiv X, R^0 \equiv R$). It's also worth noting that under the baseline ρ, the money market cumulative return changes from 0 into $C_{\rho\rho} - R_\rho$ and the difference between the cumulative return for stock i and the cumulative return for money market changes from R_i to $R_i - C_{i\rho}$, which has something to do with the correlation of return between stock i and baseline ρ. Intuitively, if a stock and baseline portfolio ρ are highly correlated, then in the auxiliary market it shall behave like the money market since it's close to baseline, consistent with the observation that $R_i - C_{i\rho}$ is close to 0. Typically we call the following formulas the **change of numeraire formula** since it tells us how to go from numeraire in the original market to the numeraire in the auxiliary market with baseline ρ.

$$\begin{cases} R_\pi^\rho = R_{\pi-\rho} - C_{\pi-\rho,\rho} \\ X_\pi^\rho = \mathcal{E}\left(R_\pi^\rho\right) \end{cases} \tag{367}$$

Now let's proceed to defining the relative covariations

$$C_{ij}^\rho \overset{def}{=} \left\langle R_i^\rho, R_j^\rho \right\rangle \tag{368}$$

a little bit of simple calculations show

$$C_{ij}^\rho = \langle R_0^\rho + R_i - C_{i\rho}, R_0^\rho + R_j - C_{j\rho}\rangle = \langle R_0^\rho + R_i, R_0^\rho + R_j\rangle \tag{369}$$

$$= \langle R_i - R_\rho, R_j - R_\rho\rangle = C_{ij} - C_{i\rho} - C_{\rho j} + C_{\rho\rho} \tag{370}$$

leading to the definition of **relative rate of covariation** as the Radon-Nikodym derivative of C_{ij}^ρ w.r.t. the operational clock

$$c_{ij}^\rho \overset{def}{=} c_{ij} - c_{i\rho} - c_{\rho j} + c_{\rho\rho} \tag{371}$$

notice that here $i, j \in \{0, 1, ..., n\}$ and we just have to set $c_{ij} = 0$ if either i or j is 0 to be consistent with previous notations. Now it's even possible to talk about the relative performance of three portfolios, ρ as the baseline and π, κ as portfolios to compare. In such setting, $c_{\pi\kappa}^\rho$ denotes the rate of covariation of portfolio π, κ under the baseline ρ and it's obvious that $c_{\pi\kappa}^\rho = \pi^T c^\rho \kappa$ is the quadratic form with

$$c_{\pi\kappa}^\rho = \sum_{i,j=0}^n \pi_i c_{ij}^\rho \kappa_j = \sum_{i,j=0}^n \pi_i (c_{ij} - c_{i\rho} - c_{\rho j} + c_{\rho\rho})\kappa_j \tag{372}$$

$$= (\pi - \rho)^T c(\kappa - \rho) \tag{373}$$

shows the very natural connection between c and $c^\rho_{\pi\kappa}$ (one just have to subtract the baseline from the portfolio in the quadratic form).

Lemma 11 (Exercise 1.24)**.** *For any portfolios ρ,π, the excess growth rate for π is*

$$\gamma_\pi^* = \frac{1}{2}\sum_{i=0}^n \pi_i c_{ii}^\rho - \frac{1}{2}\sum_{i,j=0}^n \pi_i c_{ij}^\rho \pi_j \tag{374}$$

Proof. By definition, we know

$$\gamma_\pi^* = \frac{1}{2}\sum_{i=1}^n \pi_i c_{ii} - \frac{1}{2}\sum_{i,j=1}^n \pi_i c_{ij}\pi_j \tag{375}$$

$$= \frac{1}{2}\sum_{i=1}^n \pi_i(c_{ii}^\rho + 2c_{i\rho} - c_{\rho\rho}) - \frac{1}{2}\sum_{i,j=1}^n \pi_i(c_{ij}^\rho + c_{i\rho} + c_{\rho j} - c_{\rho\rho})\pi_j \tag{376}$$

$$= \frac{1}{2}\sum_{i=0}^n \pi_i c_{ii}^\rho - \frac{1}{2}\sum_{i,j=0}^n \pi_i c_{ij}^\rho \pi_j - \frac{1}{2}\pi_0 c_{00}^\rho + \frac{1}{2}\sum_{i=1}^n \pi_i c_{i0}^\rho \pi_0 + \frac{1}{2}\sum_{j=1}^n \pi_0 c_{0j}^\rho \pi_j + \frac{1}{2}\pi_0 c_{00}^\rho \pi_0 \tag{377}$$

$$+ \frac{1}{2}\sum_{i=1}^n \pi_i(2c_{i\rho} - c_{\rho\rho}) - \frac{1}{2}\sum_{i,j=1}^n \pi_i(c_{i\rho} + c_{\rho j} - c_{\rho\rho})\pi_j \tag{378}$$

simplify the sum of last four terms on the first line

$$-\frac{1}{2}\pi_0 c_{00}^\rho + \frac{1}{2}\sum_{i=1}^n \pi_i c_{i0}^\rho \pi_0 + \frac{1}{2}\sum_{j=1}^n \pi_0 c_{0j}^\rho \pi_j + \frac{1}{2}\pi_0 c_{00}^\rho \pi_0 = \frac{1}{2}c_{\rho\rho}\pi_0(1-\pi_0) - \pi_0\sum_{i=1}^n \pi_i c_{i\rho} \tag{379}$$

also the sum of two terms on the second line

$$\frac{1}{2}\sum_{i=1}^n \pi_i(2c_{i\rho} - c_{\rho\rho}) - \frac{1}{2}\sum_{i,j=1}^n \pi_i(c_{i\rho} + c_{\rho j} - c_{\rho\rho})\pi_j = \pi_0\sum_{i=1}^n \pi_i c_{i\rho} - \frac{1}{2}c_{\rho\rho}\pi_0(1-\pi_0) \tag{380}$$

so they add up to 0 and we conclude the proof (just use the fact that $\sum_{i=1}^n \pi_i = 1-\pi_0$ and plug in the definition of relative rate of covariation).

□

Remark. *Set $\rho = \pi$ to get*

$$\gamma_\pi^* = \frac{1}{2}\sum_{i=0}^n \pi_i c_{ii}^\pi - \frac{1}{2}\sum_{i,j=0}^n \pi_i c_{ij}^\pi \pi_j = \frac{1}{2}\sum_{i=0}^n \pi_i c_{ii}^\pi \tag{381}$$

due to the fact that $\pi^T c^\pi \pi = c^\pi_{\pi\pi} = (\pi-\pi)^T c(\pi-\pi) = 0$ since $c^\rho_{\pi\kappa} = (\pi-\rho)^T c(\kappa-\rho)$. As a result, the excess growth rate of portfolio π has a nice representation under c^π. The lemma above also shows us the ***numeraire invariance of excess growth*** *since γ_π^* always has the same representation no matter what baseline ρ one takes.*

Bounds on Excess Growth

Since the excess growth rate γ_π^* is a function of π and c, it's natural to expect its bound to depend on the spectrum of the matrix-valued process c. Define predictable process $\lambda(t)$ as the **maximum eigenvalue** of $c(t)$ and define predictable process $l(t)$ such that $l(t)\lambda(t)$ is the **minimum eigenvalue** of $c(t)$. Since c takes values as symmetric SPD matrices, we immediately know that λ is non-negative and l takes values in $[0, 1]$. Whenever $\lambda(t) = 0$, we define $l(t) = 1$.

Lemma 12. *For any long-only portfolio π,*

$$\frac{l}{2}\lambda\left(1 - \max_{0\leq i\leq n}\pi_i\right) \leq \gamma_\pi^* \leq \lambda\left(1 - \max_{0\leq i\leq n}\pi_i\right) \tag{382}$$

Proof. Use the nice representation of excess growth rate in the lemma above,

$$\gamma_\pi^* = \frac{1}{2}\sum_{i=0}^{n}\pi_i c_{ii}^\pi \tag{383}$$

the inequality

$$\forall 0 \leq i \leq n, l\lambda||e_i - \pi||_2^2 \leq c_{ii}^\pi \leq \lambda||e_i - \pi||_2^2 \tag{384}$$

follows from the fact that $c_{ii}^\pi = (e_i - \pi)^T c(e_i - \pi)$ concludes the proof. □

Note that this bound can also be built for the cumulative excess growth process Γ_π^*, just integrate λ, l w.r.t. the operational clock and a similar inequality still holds.

Market Portfolio as Baseline

A natural baseline to select is the market portfolio μ. Recall that $\forall 1 \leq i \leq n, \mu_i = \frac{S_i}{\Sigma}$ is always the relative market capitalization and it's a long-only stock portfolio so $\mu_0 = 0$. Taking market portfolio as baseline is equivalent to comparing with the performance of the market.

Recall that **for stock portfolio** π, it's true that $\pi_0 = 0$ and due to $X_\pi^\mu = \mathcal{E}\left(R_\pi^\mu\right)$ where

$$R_\pi^\mu = R_{\pi-\mu} - C_{\pi-\mu,\mu} = \sum_{i=1}^n (\pi_i - \mu_i) R_i - \sum_{i,j=1}^n (\pi_i - \mu_i) C_{ij} \mu_j \tag{385}$$

recall that $X_\mu = \frac{\Sigma}{\Sigma(0)}$, we also have

$$S_i^\mu = \frac{S_i}{X_\mu} = \Sigma(0)\mu_i, R_i^\mu = \mathcal{E}\left(S_i^\mu\right) \tag{386}$$

combine those two facts to see

$$\frac{1}{X_\pi^\mu(t)}\, dX_\pi^\mu(t) = dR_\pi^\mu(t) = \sum_{i=1}^n \pi_i(t)\, dR_i^\mu(t) = \sum_{i=1}^n \pi_i(t) \frac{1}{\mu_i(t)}\, d\mu_i(t) \tag{387}$$

gives the **dynamics of the relative wealth process**.

Now we can calculate the excess growth rate of the market portfolio with the nice representation that

$$\gamma_\mu^* = \frac{1}{2} \sum_{i=0}^n \mu_i c_{ii}^\mu \tag{388}$$

integrating both sides w.r.t. operational clock gives the cumulative excess growth process

$$\Gamma_\mu^*(t) = \frac{1}{2} \int_0^t \sum_{i=0}^n \mu_i(s)\, dC_{ii}^\mu(s) \tag{389}$$

notice that $\mu_i = \frac{S_i^\mu}{\Sigma(0)} = \mu_i(0)\mathcal{E}\left(R_i^\mu\right)$ so $\langle \log \mu_i, \log \mu_i \rangle = \langle R_i^\mu, R_i^\mu \rangle = C_{ii}^\mu$ and

$$\Gamma_\mu^*(t) = \frac{1}{2} \sum_{i=1}^n \int_0^t \mu_i(s)\, d\left\langle \log \mu_i, \log \mu_i \right\rangle(s) \tag{390}$$

this characterization of the cumulative excess growth of market portfolio defines $2\Gamma_\mu^*$ as the **cumulative intrinsic variation** since it's the capitalization-weighted cumulative average stock variation relative to the market.

In order to **estimate the cumulative intrinsic variation historically**, notice that

$$\int_0^t \mu_i(s)\, d\left\langle \log \mu_i, \log \mu_i \right\rangle(s) = \int_0^t \frac{1}{\mu_i(s)}\, d\left\langle \mu_i, \mu_i \right\rangle(s) = \left\langle \mu_i, \log \mu_i \right\rangle(t) \tag{391}$$

since

$$d\left<\mu_i, \mu_i\right> = \frac{1}{\Sigma^2(0)}\, d\left<S_i^\mu, S_i^\mu\right> = \frac{(S_i^\mu)^2}{\Sigma^2(0)}\, d\left<\log\mu_i, \log\mu_i\right> = \mu_i^2\, d\left<\log\mu_i, \log\mu_i\right> \tag{392}$$

$$d\left<\mu_i, \log\mu_i\right> = \frac{1}{\Sigma(0)} d\left<S_i^\mu, \log\mu_i\right> = \frac{S_i^\mu}{\Sigma(0)}\, d\left<\log\mu_i, \log\mu_i\right> = \mu_i\, d\left<\log\mu_i, \log\mu_i\right> \tag{393}$$

this gives us an easy characterization of the cumulative intrinsic variation to be estimated historically

$$2\Gamma_\mu^*(t) = \sum_{i=1}^n \left<\mu_i, \log\mu_i\right>(t) \tag{394}$$

it's just the sum of the quadratic variation between market portfolio μ_i and its logarithm among all stocks.

Last but not least, let's check the bound for cumulative excess growth applied for market portfolio μ, set $\Lambda(t) = \int_0^t \lambda(t)\, dO(t)$ as the cumulative process, the previous bound extends to

$$\frac{1}{2}\int_0^t l(s)\left(1 - \max_{0\le i\le n}\mu_i(s)\right) d\Lambda(s) \le \Gamma_\mu^*(t) = \int_0^t \gamma_\mu^*(s)\, dO(s) \le \int_0^t \left(1 - \max_{0\le i\le n}\mu_i(s)\right) d\Lambda(s) \tag{395}$$

and the process

$$1 - \max_{0\le i\le n}\mu_i(s) \tag{396}$$

has the interpretation as **market diversity**, i.e. when such value is low, there exists a stock with dominant market capitalization, leading to a low cumulative excess growth. Conversely, when market diversity is high, market capitalization is distributed among different stocks so the cumulative excess growth will be high. We see the very important fact that **intrinsic market variation and market diversity are two sides of the same coin**.

Functional Generation of Stock Portfolio

With all our setting, now we are able to analyze the behavior of the portfolio with a market baseline. In previous context, we have already provided the dynamics that

$$\frac{1}{X_\pi^\mu(t)}\,dX_\pi^\mu(t)=\sum_{i=1}^n \pi_i(t)\frac{1}{\mu_i(t)}\,d\mu_i(t) \tag{397}$$

if π is a stock portfolio. For simplicity, here we **only focus on stock portfolio** π (where the dynamics has simple and compact representation). Our goal now is to figure out the behavior of X_π^μ for given π but it does not seem easy solving this SDE directly.

However, we may put up **a special ansatz**

$$\log X_\pi^\mu = J_\pi^\mu + H_\pi^\mu \tag{398}$$

where J_π^μ is continuous nondecreasing and H_π^μ is a continuous semi-MG dominated by J_π^μ, i.e. when J_π^μ is large, $\frac{H_\pi^\mu}{J_\pi^\mu}$ is always small. Under such ansatz the asymptotic behavior of $\log X_\pi^\mu$ is dominated by the finite variation process J_π^μ. Although this idea seems to work, we still need more restrictions on the form of H_π^μ. One simple way is **to assume**

$$H_\pi^\mu=\log\frac{F(\mu)}{F(\mu(0))} \tag{399}$$

where the function $F:\mathbf{ri}\left(\Delta_{(n-1)}\right)\to(0,+\infty), F\in C^2$ is a mapping from the relative interior of the space of all long-only stock portfolios to the set of strictly positive real numbers. Simple calculations from Ito formula gives

$$\frac{dX_\pi^\mu(t)}{X_\pi^\mu(t)}=dJ_\pi^\mu(t)+\sum_{i=1}^n\frac{F_i(\mu(t))}{F(\mu(t))}\,d\mu_i(t)+\frac{1}{2}\sum_{i,j=1}^n\frac{F_{ij}(\mu(t))}{F(\mu(t))}\,d\left\langle\mu_i,\mu_j\right\rangle(t) \tag{400}$$

where F_i, F_{ij} denotes the partial derivatives of F w.r.t. its components. Compare with the original dynamics to find

$$\forall 1\le i\le n, \frac{\pi_i}{\mu_i}=\frac{F_i(\mu)}{F(\mu)} \tag{401}$$

but notice that this does not guarantee $\sum_{i=1}^n \pi_i=1$ is a stock portfolio, so we have to use a different π that

$$\pi_i^F\overset{def}{=}\mu_i\left(\frac{F_i(\mu)}{F(\mu)}+1-\sum_{j=1}^n\mu_j\frac{F_j(\mu)}{F(\mu)}\right) \tag{402}$$

let's first verify that $\pi^F\in H_{n-1}$ is always a stock portfolio

$$\sum_{i=1}^n\pi_i^F=\sum_{i=1}^n\mu_i\frac{F_i(\mu)}{F(\mu)}+\sum_{i=1}^n\mu_i-\sum_{i,j=1}^n\mu_i\mu_j\frac{F_j(\mu)}{F(\mu)}=\sum_{i=1}^n\mu_i\frac{F_i(\mu)}{F(\mu)}+1-\sum_{i=1}^n\mu_i\frac{F_i(\mu)}{F(\mu)}=1 \tag{403}$$

on the other hand, recall the global structure of market portfolio as stock portfolio that $\sum_{i=1}^n \mu_i = 1$,

$$\int_0^t \left(\sum_{i=1}^n \frac{\pi_i^F(s)}{\mu_i(s)} - \sum_{i=1}^n \frac{F_i(\mu(s))}{F(\mu(s))}\right) d\mu_i(s) = \int_0^t \sum_{i=1}^n \left(\frac{\pi_i^F(s)}{\mu_i(s)} - \frac{F_i(\mu(s))}{F(\mu(s))}\right) d\mu_i(s) \tag{404}$$

$$= \int_0^t \sum_{i=1}^n \left(1 - \sum_{j=1}^n \mu_j(s)\frac{F_j(\mu(s))}{F(\mu(s))}\right) d\mu_i(s) \tag{405}$$

$$= \int_0^t \left(1 - \sum_{j=1}^n \mu_j(s)\frac{F_j(\mu(s))}{F(\mu(s))}\right) d\sum_{i=1}^n \mu_i(s) \tag{406}$$

$$= 0 \tag{407}$$

we see that **the definition of π^F is very subtle that it makes sure π^F is a stock portfolio and that $\int_0^t \sum_{i=1}^n \frac{\pi_i^F(s)}{\mu_i(s)}\, d\mu_i(s) = \int_0^t \sum_{i=1}^n \frac{F_i(\mu(s))}{F(\mu(s))}\, d\mu_i(s)$ holds**. Now let's look back at the dynamics so

$$\frac{dX_\pi^\mu(t)}{X_\pi^\mu(t)} = dJ_\pi^\mu(t) + \sum_{i=1}^n \frac{\pi_i^F(t)}{\mu_i(t)}\, d\mu_i(t) + \frac{1}{2}\sum_{i,j=1}^n \frac{F_{ij}(\mu(t))}{F(\mu(t))}\, d\left\langle \mu_i, \mu_j\right\rangle(t) \tag{408}$$

must hold. If we further define

$$J^F(t) \overset{def}{=} -\frac{1}{2}\int_0^t \sum_{i,j=1}^n \frac{F_{ij}(\mu(s))}{F(\mu(s))}\, d\left\langle \mu_i, \mu_j\right\rangle(s) \tag{409}$$

then the dynamics becomes

$$\frac{dX_\pi^\mu(t)}{X_\pi^\mu(t)} = dJ_\pi^\mu(t) + \sum_{i=1}^n \frac{\pi_i^F(t)}{\mu_i(t)}\, d\mu_i(t) - dJ^F(t) \tag{410}$$

so setting $\pi = \pi^F, J_\pi^\mu = J^F$ would now recover the original dynamics of X_π^μ w.r.t. π_i, μ_i.

In short, **we have proved above that**

$$X_{\pi^F}^\mu = e^{J^F}\frac{F(\mu)}{F(\mu(0))} \tag{411}$$

the ansatz is still true replacing π with π^F and J_π^μ with J^F where

$$\begin{cases} \pi_i^F \overset{def}{=} \mu_i\left(\frac{F_i(\mu)}{F(\mu)} + 1 - \sum_{j=1}^n \mu_j \frac{F_j(\mu)}{F(\mu)}\right) \\ J^F(t) \overset{def}{=} -\frac{1}{2}\int_0^t \sum_{i,j=1}^n \frac{F_{ij}(\mu(s))}{F(\mu(s))}\, d\left\langle \mu_i, \mu_j\right\rangle(s) \end{cases} \tag{412}$$

this is a clever trick since now everything depends on the selection of F. Whenever one specifies an F, one can figure out π^F, J^F and one gets $X_{\pi^F}^\mu$ immediately! The π^F is called **the stock portfolio generated by function F**. Of course, the price to pay is that we can only investigate the relative performance of all stock portfolios with the market baseline that can be represented as such π^F. Despite the restriction, it still contains a very wide class of stock portfolios including a lot of important examples.

Most frequently, such $F : \text{ri}\left(\Delta_{(n-1)}\right) \to (0,\infty), F \in C^2$ can be taken as the **linear function** $F(x) = \sum_{i=1}^n x_i w_i$ with $w \in \Delta_{(n-1)}$ given, in this case $\frac{\partial F(x)}{\partial x} = w, F_i(x) = w_i$, we have

$$\pi_i^F = \mu_i \left(\frac{w_i}{w^T\mu} + 1 - \sum_{j=1}^n \mu_j \frac{w_j}{w^T\mu} \right) = \frac{\mu_i w_i}{w^T\mu} = \frac{w_i S_i}{\sum_{j=1}^n w_j S_j} \tag{413}$$

so the stock portfolio π^F buys $w_i \geq 0$ shares of stock i at time 0 and holds it without doing any modifications. This is the **buy-and-hold strategy**.

On the other hand, we might also consider the **geometric average** $F(x) = \prod_{i=1}^n (x_i)^{p_i}$ with $p \in \Delta_{(n-1)}$ given, in this case $\frac{\partial F(x)}{\partial x_i} = F_i(x) = F(x)\frac{p_i}{x_i}$, we have

$$\pi_i^F = \mu_i \left(\frac{p_i}{\mu_i} + 1 - \sum_{j=1}^n \mu_j \frac{p_j}{\mu_j} \right) = p_i \tag{414}$$

so the stock portfolio π_i^F always invest a constant proportion p_i of the total wealth into stock i, it's a **constant-proportion portfolio**. A special case is when $p_i = \frac{1}{n}$ so it's an **equal-weighted portfolio**. If one always invests equal proportional of wealth in each stock, when one stock has a higher price, the share one holds is smaller and when a stock has a lower price, the share one holds is larger. So this naturally corresponds to the **buy-low sell-high strategy**.

If one already has stock portfolios $\pi_1, ..., \pi_m$ generated by functions $F^1, ..., F^m$, then setting $F(x) = \prod_{i=1}^m [F_i(x)]^{p_i}$ for given $p \in \Delta_{(n-1)}$ gives one the **mixture portfolio** $\pi = \sum_{i=1}^m p_i \pi_i$.

Lemma 13 (Exercise 1.30)**.** *Let the generating function F be concave, then π^F is long-only stock portfolio and J^F is nondecreasing.*

Proof. It's obvious that π^F is always a stock portfolio. To prove it's long-only, just need to prove

$$\forall 1 \leq i \leq n, \frac{F_i(\mu)}{F(\mu)} + 1 - \sum_{j=1}^n \mu_j \frac{F_j(\mu)}{F(\mu)} \geq 0 \tag{415}$$

just need to show

$$F(\mu) \geq (\mu - e_i)^T \nabla F(\mu) \tag{416}$$

by the first-order characterization of concavity,

$$0 < F(e_i) \leq F(\mu) + (e_i - \mu)^T \nabla F(\mu) \tag{417}$$

concludes the proof.

For J^F,

$$\frac{dJ^F(t)}{dt} = -\frac{1}{2} \sum_{i,j=1}^n \frac{F_{ij}(\mu(t))}{F(\mu(t))} \frac{d\left\langle \mu_i, \mu_j \right\rangle(t)}{dt} \tag{418}$$

is a actually a quadratic form w.r.t. $\nabla^2 F(\mu(t))$, the Hessian of F. F is concave implies $\nabla^2 F \leq 0$ so $\frac{dJ^F(t)}{dt} \geq 0$, it's nondecreasing in time.

□

Lemma 14 (Exercise 1.31)**.** *Let the generating function F be concave and*

$$\lim_{m\to\infty} \sup_{t\geq 0} \mathbb{P}\left(F(\mu(t)) \leq \frac{1}{m}\right) = 0 \tag{419}$$

then on the event $\left\{J^F(\infty) = \infty\right\}$, we have

$$\lim_{T\to\infty} \frac{1}{J^F(T)} \log \frac{X_{\pi^F}(T)}{X_\mu(T)} = 1 \tag{420}$$

the limit is in the sense of convergence in probability.

Proof. By the functional generation of stock portfolio, we have

$$\frac{1}{J^F(T)} \log \frac{X_{\pi^F}(T)}{X_\mu(T)} = 1 + \frac{\log F(\mu(T))}{J^F(T)} - \frac{\log F(\mu(0))}{J^F(T)} \tag{421}$$

now we just need to prove that

$$\frac{\log F(\mu(T))}{J^F(T)} \xrightarrow{p} 0 \ (T \to \infty) \tag{422}$$

since $J^F(T) \to \infty$ $(T \to \infty)$, the third term in the equation above has limit 0 in probability. For $\forall m \in \mathbb{N}$,

$$\frac{\log F(\mu(T))}{J^F(T)} = \frac{\log F(\mu(T))\mathbb{I}_{F(\mu(T))\leq \frac{1}{m}}}{J^F(T)} + \frac{\log F(\mu(T))\mathbb{I}_{F(\mu(T))> \frac{1}{m}}}{J^F(T)} \tag{423}$$

with $\frac{\log F(\mu(T))\mathbb{I}_{F(\mu(T))> \frac{1}{m}}}{J^F(T)} \xrightarrow{p} 0$ $(T \to \infty)$ due to the concavity of F that under event $\left\{F(\mu(T)) > \frac{1}{m}\right\}$, for given $p \in \Delta_{(n-1)}$

$$\frac{1}{m} < F(\mu(T)) \leq F(p) + (\mu(T) - p)^T \nabla F(p) \leq F(p) + ||\mu(T) - p||_2 \cdot ||\nabla F(p)||_2 \leq F(p) + \sqrt{n}||\nabla F(p)||_2 \tag{424}$$

since $\forall i, |\mu_i(T)-p_i| \leq 1$. As a result, $\log F(\mu(T))$ is bounded both from above and below under event $\left\{F(\mu(T)) > \frac{1}{m}\right\}$. So now the only problem is to deal with the first term where $F(\mu(T))$ is not bounded from below. Let's apply concavity again to see

$$\log F(\mu(T)) \geq \log(F(p) - \sqrt{n}||\nabla F(\mu(T))||_2) \geq \log\left(F(p) - \sqrt{n}\sup_{t\geq 0} ||\nabla F(\mu(t))||_2\right) > -\infty \tag{425}$$

since ∇F is a continuous function on the compact set $\Delta_{(n-1)}$. Now by the condition given, we see that

$$\frac{\log F(\mu(T))\mathbb{I}_{F(\mu(T))\leq \frac{1}{m}}}{J^F(T)} \geq \log\left(F(p) - \sqrt{n}\sup_{t\geq 0}||\nabla F(\mu(t))||_2\right)\frac{\mathbb{I}_{F(\mu(T))\leq \frac{1}{m}}}{J^F(T)} \xrightarrow{p} 0 \ (T\to\infty) \tag{426}$$

so we see that

$$\frac{\log F(\mu(T))\mathbb{I}_{F(\mu(T))\leq \frac{1}{m}}}{J^F(T)} \xrightarrow{p} 0 \ (T\to\infty) \tag{427}$$

concludes the proof.

□

Remark. *Those two examples show us that if* ***the generating function*** F ***is concave, the stock portfolio must be long-only, the finite variation process*** J^F ***must be increasing****.*

In addition, if we know ***the probability of*** $F(\mu)$ ***taking values close to 0 to be uniformly small (which is satisfied if the range of*** F ***is bounded away from 0), then if the finite variation process increases to infinity,*** $\log X^\mu_{\pi^F}(T) \sim J^F(T)$ $(T\to\infty)$, ***the log-relative wealth process with market baseline has the same asymptotic behavior as*** J^F*, which provides a rigorous statement for our motivation that when* J^μ_π *is large, it dominates the asymptotic behavior of* $\log X^\mu_{\pi^F}$.

Example: Diversity Weighting

Consider taking $F(x) = (\sum_{i=1}^n (x_i)^p)^{\frac{1}{p}}$ for fixed $p \in (0,1)$. This is a concave function and the minimum is achieved iff $x_1 = ... = x_n = \frac{1}{n}$ so $F(x) \geq n^{\frac{1-p}{p}}$ is bounded away from 0. Now calculate the **diversity-weighted portfolio**

$$\pi_i^F = \frac{\mu_i^p}{\sum_{j=1}^n \mu_j^p} \tag{428}$$

is the average of the relative capitalization to the power p. It's clear that when $p = 1$, this gives the market portfolio and when $p = 0$, this gives the equal-weighted portfolio. It's an intermediary between these two portfolios.

Let's calculate the process J^F for this example.

$$J^F(t) = -\frac{1}{2}\int_0^t \sum_{i,j=1}^n \frac{F_{ij}(\mu(s))}{F(\mu(s))}\, d\left\langle \mu_i, \mu_j \right\rangle(s) = (1-p)\Gamma^*_{\pi^F}(t) \tag{429}$$

where $\Gamma^*_{\pi^F}$ is the cumulative excess growth for portfolio π^F. Since the F is bounded away from 0, we use the result in the lemma proved above to conclude

$$\lim_{T\to\infty} \frac{1}{\Gamma^*_{\pi^F}(T)} \log X^\mu_{\pi^F}(T) = 1 - p \tag{430}$$

when $\Gamma^*_{\pi^F}(\infty) = \infty$. Intuitively, this means that the log relative wealth with market baseline is always $1-p$ multiple of the cumulative excess growth.

Example: Entropy Weighting

Consider $F(x) = -\sum_{i=1}^n x_i \log x_i$ as the entropy of the probability mass sequence $x_1, ..., x_n$. Calculate the portfolio to see

$$\pi_i^F = \frac{\mu_i \log \mu_i}{\sum_{j=1}^n \mu_j \log \mu_j} \tag{431}$$

and

$$J^F(t) = \int_0^t \frac{1}{F(\mu(s))}\, d\Gamma^*_\mu(s) \tag{432}$$

since the entropy goes to 0 if x goes close to e_i, this F is no longer bounded away from 0 and we need some other conditions to ensure that $\lim_{T\to\infty} \frac{1}{J^F(T)} \log \frac{X_{\pi^F}(T)}{X_\mu(T)} = 1$.

Super Martingale Numeraire

For portfolio ν, X_ν is called **super-MG (local-MG) numeraire** if for every π, X_π^ν is super-MG (local-MG). Of course, super-MG numeraire does not necessarily exist in the market and needs some effort to find. However, if the super-MG numeraire exists in the market, it means that X_π^ν shall go down for any portfolio π. In other words, given such a baseline ν, any other portfolio does not behave better than ν as time goes by. Such ν is called a **super-MG numeraire portfolio** and is typically a nice investment strategy. Notice that since numeraire is strictly positive, local-MG and super-MG are the same things under such context.

Some simple properties of super-MG numeraire are listed below.

Lemma 15 (Exercise 2.2)**.** *Let Z be strictly positive continuous semi-MG such that $Z, \frac{1}{Z}$ are both super-MG, then Z is constant.*

Proof. We have by Jensen's inequality

$$\forall s \le t, \log Z_s \ge \log \mathbb{E}(Z_t|\mathscr{F}_s) \ge \mathbb{E}(\log Z_t|\mathscr{F}_s) = -\mathbb{E}\left(\log \frac{1}{Z_t}|\mathscr{F}_s\right) \ge -\log \mathbb{E}\left(\frac{1}{Z_t}|\mathscr{F}_s\right) \ge -\log \frac{1}{Z_s} = \log Z_s \tag{433}$$

so we see that $Z_s = \mathbb{E}(Z_t|\mathscr{F}_s)$ meaning that Z is a MG and we have Jensen's inequality to be equality. Since $\log x$ is not linear, it must be the case that $Z_t \in \mathscr{F}_s$ so we have $\forall s \le t, Z_t = \mathbb{E}(Z_t|\mathscr{F}_s) = Z_s$ concludes that it's constant in time. □

Lemma 16 (Exercise 2.3)**.** *For ν, X_ν is local MG numeraire iff all processes S_i^ν are local MG for $0 \le i \le n$.*

Proof. By definition, for any portfolio π, $X_\pi^\nu = \mathcal{E}\left(R_\pi^\nu\right), R_\pi^\nu(t) = \int_0^t \sum_{i=0}^n \pi_i(s)\, dR_i^\nu(s)$ and $S_i^\nu = S_i(0)\mathcal{E}\left(R_i^\nu\right)$.

All S_i^ν are local MG iff all R_i^ν as the stochastic logarithm of S_i^ν are local MG iff R_π^ν is local MG iff X_π^ν as the stochastic exponential of R_π^ν is local MG.

□

Remark. *Super-MG numeraire is **unique** if it exists. TO see this, assume that ν, η are both super-MG numeraire portfolio, then $\frac{X_\nu}{X_\eta}, \frac{X_\eta}{X_\nu}$ are both super-MG so $\frac{X_\nu}{X_\eta}$ is constant in time, they are actually the same super-MG numeraire.*

Theorem 1. *The Following statements are equivalent:*

1. *X_ν is a super-MG numeraire*
2. *X_ν is local MG numeraire*
3. *$\forall 1 \le i \le n, A_i = C_{i\nu}$*

Proof. It's obvious that 1 and 2 are equivalent.

If 3 holds, then $dR_i(t) - dC_{i\nu}(t) = dR_i(t) - dC_{i\nu}(t) = dM_i(t)$ so $R_i - C_{i\nu}$ is a local MG. Naturally, $R_0^\nu = C_{\nu\nu} - R_\nu$ is a local MG and R_i^ν is local MG so $X_\pi^\nu = \mathcal{E}\left(R_\pi^\nu\right)$ is local MG for any portfolio π. One can prove conversely to show the equivalence.

□

Remark. *This theorem gives an equivalent condition for the existence of super-MG numeraire. We see that if super-MG numeraire exists, then*

$$\forall 1 \leq i \leq n, \int_0^t \alpha_i(s)\, dO(s) = \int_0^t c_{i\nu}(s)\, dO(s) \tag{434}$$

with $c_{i\nu} = e_i^T c\nu$. This means that vector α is always in the range of matrix c and $c\nu = \alpha$ so $\nu = c^{-1}\alpha$ ***gives the formula for the super-MG numeraire portfolio****. This formula works if c is non-singular and one may replace the matrix inverse with pseudoinverse if c is singular.*

Let's look at an example where **super-MG numeraire does not exist**. Consider $n = 1$ with only one stock in the market and $R_t = |W_t|$ where W_t is 1-dimensional BM. By Ito-Tanaka formula, we have

$$R_t = \int_0^t sgn(W_s)\, dW_s + L_t \tag{435}$$

where

$$L_t = \lim_{\varepsilon \to 0} \frac{|\{0 \leq s \leq t : |W_s| < \varepsilon\}|}{2\varepsilon} \tag{436}$$

is the BM local time at 0. It's clear that L_t has finite variation but is singular w.r.t. Lebesgue measure so under the Lebesgue clock, $A \neq C_{1\nu}$ since $C_{1\nu}$ is absolute continuous w.r.t. the Lebesgue clock. This shows that no super-MG numeraire exists in this market.

In order to show another non-trivial example, we prepare ourselves with the lemma below

Lemma 17 (Exercise 2.9)**.** *In filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbb{Q})$, Z is 1-dimensional BM such that $Z_0 = z > 0$. Consider the first hitting time to 0 that $\tau_0 = \inf\{t \geq 0 : Z_t = 0\}$, fix $T > 0$ and define a new probability measure $\mathbb{P} = \mathbb{P}_T$ absolute continuous w.r.t. $\mathbb{Q}$ via*

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{Z_{T\wedge\tau_0}}{z} \tag{437}$$

then show that

$$Z_t = z + \int_0^t \frac{1}{Z_s}\, ds + W_t^T \tag{438}$$

where W^T is standard BM on time interval $[0, T]$ under $\mathbb{P}$. Show that

$$m(T) \overset{def}{=} \mathbb{E}_{\mathbb{P}} \frac{1}{Z_T} = \frac{1}{z}\mathbb{Q}(\tau_0 > T) = \frac{1}{z}\left(2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1\right) \tag{439}$$

and m is strictly decreasing with $m(0) = \frac{1}{z}, m(\infty) = 0$.

Proof. $Z_t - z$ is standard BM under $\mathbb{Q}$ and now we want to prove that $W_t^T = Z_t - z - \int_0^t \frac{1}{Z_s}\, ds$ is standard BM under $\mathbb{P}$. Since

$$\mathbb{E}_{\mathbb{Q}} \frac{Z_{T\wedge\tau_0}}{z} = \frac{1}{z}\mathbb{E}_{\mathbb{Q}} Z_{T\wedge\tau_0} = 1 \tag{440}$$

by optional stopping theorem, Girsanov theorem holds and we only need to verify that

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = e^{\int_0^T \frac{1}{Z_s}\,dZ_s - \frac{1}{2}\int_0^T \frac{1}{Z_s^2}\,ds} = \mathcal{E}\left(\mathcal{L}\left(Z\right)\right)(T) \tag{441}$$

it's quite obvious that if $\tau_0 \le T$, i.e. Z_t has hit 0 before time T, then this process takes value 0. Otherwise we can eliminate the stochastic logarithm and stochastic exponential with a constant multiple. So when $\tau_0 > T$, we have

$$\mathcal{E}\left(\mathcal{L}\left(Z\right)\right)(T) = \mathcal{E}\left(\mathcal{L}\left(\frac{Z}{z}\right)\right)(T) = \frac{Z_T}{z} \tag{442}$$

as a result, $\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{Z_{T\wedge\tau_0}}{z}$ proves the equation for Z_t under change of measure.

To calculate such m, we notice that

$$m(T) = \mathbb{E}_{\mathbb{P}}\frac{1}{Z_T} = \frac{1}{z}\mathbb{E}_{\mathbb{Q}}\frac{Z_{T\wedge\tau_0}}{Z_T} \tag{443}$$

$$= \frac{1}{z}\mathbb{Q}(\tau_0 > T) \tag{444}$$

$\mathbb{Q}(\tau_0 > T)$ is the probability BM Z_t has not hit 0 before time T. Let $B_t = Z_t - z$ be standard BM under $\mathbb{Q}$, then by reflection principle, $\inf_{t\le T} B_t \overset{d}{=} -|B_T|$ and

$$\mathbb{Q}(\tau_0 > T) = \mathbb{Q}\left(\inf_{t\le T} B_t > -z\right) = \mathbb{Q}(|B_T| < z) = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1 \tag{445}$$

concludes the proof. □

Now we can look at the example $n = 1$, $S_t = Z_t$ where $Z_t = z + \int_0^t \frac{1}{Z_s}\,ds + W_t$ so Z is the solution to such SDE. By Ito formula

$$d\frac{1}{Z_t} = -\frac{1}{Z_t^2}\,dZ_t + \frac{1}{Z_t^3}\,d\left\langle Z, Z\right\rangle_t = -\frac{1}{Z_t^2}\,dW_t \tag{446}$$

so $\frac{1}{Z}$ is local-MG. Recall from stochastic calculus that the 3-dimensional Bessel process R_t is defined as $R_t = ||B_t||_2$ where $B_t = (B_t^1, B_t^2, B_t^3)$ is a 3-dimensional standard BM. By Ito formula,

$$dR_t = \sum_{j=1}^3 \frac{B_t^j}{R_t}\,dB_t^j + \frac{1}{R_t}\,dt \tag{447}$$

however, it's an Ito diffusion since $\sum_{j=1}^3 \frac{(B_t^j)^2}{(R_t)^2} = 1$ so we may denote $\sum_{j=1}^3 \frac{B_t^j}{R_t}\,dB_t^j = d\tilde{B}_t$ for a new 1-dimensional BM $\tilde{B}_t$. As a result, Bessel process is the solution (weak sense) to

$$dR_t = d\tilde{B}_t + \frac{1}{R_t}\,dt \tag{448}$$

so here in our example $Z \overset{d}{=} R$, the stock price process has the same distribution as the **3D Bessel process**.

We see that

$$\frac{1}{S_t}\,dS_t = \frac{1}{Z_t^2}\,dt + \frac{1}{Z_t}\,dW_t \tag{449}$$

so the semi-MG decomposition of R_t is given by

$$A_t = \int_0^t \frac{1}{Z_s^2}\,ds, M_t = \int_0^t \frac{1}{Z_s}\,dW_s, C_t = \langle M, M\rangle_t = \int_0^t \frac{1}{Z_s^2}\,ds \tag{450}$$

so $A_t = C_t$, which implies that super-MG numeraire exists. From our remark above, **the super-MG numeraire portfolio shall be $\nu = \frac{\alpha}{c} = 1$, always investing all wealth into the stock**. The super-MG numeraire is given by

$$X_\nu = \mathcal{E}\left(R_\nu\right) = \mathcal{E}\left(R\right) = \frac{S}{S_0} = \frac{Z}{z} \tag{451}$$

if we investigate the expectation of the relative wealth of the trivial portfolio π with baseline ν where $X_\pi \equiv 1$,

$$\mathbb{E}X_\pi^\nu(T) = \mathbb{E}\frac{z}{Z_T} = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1 \tag{452}$$

from the lemma above. One may find that this expectation is strictly decreasing in T with $\mathbb{E}X_\pi^\nu(0) = 1, \mathbb{E}X_\pi^\nu(T) \to 0$ $(T \to \infty)$. This is consistent with our observation that for any portfolio π, X_π^ν is always a super-MG.

Remark. *One might realize that since $\mathbb{E}\frac{z}{Z_T}$ is strictly decreasing in T, $\frac{1}{Z_t}$ cannot be a MG. However, we do prove that for 3D Bessel process Z_t, its reciprocal $\frac{1}{Z_t}$ is a local MG. This gives the classical example of a process being **local MG but not MG**.*

Market Viability

Consider capital withdrawal $K \in \mathscr{K}$ such that $K(0) = 0$, define

$$x(K) \stackrel{def}{=} \inf \{x > 0 : K \in \mathscr{K}(x)\} \tag{453}$$

as **the approximately minimum initial wealth** x **required such that a given capital withdrawal** K **is financeable from** x **under some investment strategy** θ.

Remark. *Notice that when one has initial wealth as exactly $x(K)$, one cannot necessarily finance the capital withdrawal stream K. This is due to the inf in the definition of $x(K)$ so there might be the case that $x(K) = 0$ so any positive amount is enough to finance K but K is not financeable with exactly zero initial wealth.*

If we think about the capital withdrawal as consumption, then $x(K)$ is the lowest amount of money an investor shall set aside initially such that he can stick to his consumption plan K being financed successfully under some appropriate investment strategy. An example for K might be the **European contingent claim** $K(t) = h\mathbb{I}_{[\tau,\infty)}(t)$ where $h \in \mathscr{F}_\tau$ is non-negative for some stopping time τ. It means a single capital withdrawal of size $h \in \mathscr{F}_\tau$ happens at time τ.

The following are properties of $x(K)$, we do not write the proof here since it's obvious.

Lemma 18 (Exercise 2.17)**.** *Prove the properties of $x : \mathscr{K} \to [0, \infty]$ that it's monotone in $K \in \mathscr{K}$, has positive homogeneity and subadditivity.*

The questions that arise in this context are: whether the inf in the definition of $x(K)$ is attained and how can we find the corresponding investment strategy θ and wealth process X if it can be attained. To provide answers for these questions, we introduce the definition of market viability. **A market is viable if** $K \equiv 0$ **whenever** $\exists K \in \mathscr{K}, x(K) = 0$. In other words, we call a market viable if **one cannot afford to finance any capital withdrawal / consumption with arbitrarily small amount of initial wealth** (notice that it's not zero initial wealth but arbitrarily small initial wealth since there is an inf in the definition of $x(K)$). It's easy to see that market viability is a kind of **non-arbitrage criterion** for the market.

Notice that market viability is equivalent to saying $\mathscr{K}(0+) = \{0\}$ (due to the inf in the definition of $x(K)$) where

$$\mathscr{K}(0+) \stackrel{def}{=} \cap_{x>0} \mathscr{K}(x) \tag{454}$$

since $\mathscr{K}(0+)$ contains all the capital withdrawal processes such that they can be financed with any positively small amount of initial wealth. It's clear that $\{0\} \subset \mathscr{K}(0) \subset \mathscr{K}(0+)$ and generally $\mathscr{K}(0) \neq \mathscr{K}(0+)$.

The following theorem provides characterization for market viability. Basically, market is viable iff the probability that any numeraire at a fixed time takes large values is uniformly small.

Theorem 2. *Market is viable iff*

$$\forall T \geq 0, \lim_{m\to\infty} \sup_{X\in\mathscr{X}} \mathbb{P}(X(T) > m) = 0 \tag{455}$$

where $\mathscr{X}$ is the set of all numeraires. This condition is called ***boundedness in probability****.*

Local MG Deflator

If $Y > 0, Y(0) = 1$ and if YX is a local MG for any numeraire $X \in \mathscr{X}$, then we call Y to be a **local MG deflator**. Denote $\mathscr{Y}$ as the set of all local MG deflators. It's obvious that local MG deflator Y itself must also be a local MG. If X_ν is super-MG numeraire, then for any portfolio π, $X_\pi^\nu = \frac{X_\pi}{X_\nu}$ is super-MG so $\frac{1}{X_\nu}$ is a local MG deflator. As a result, the existence of super-MG numeraire implies $\mathscr{Y} \neq \emptyset$.

Actually, local MG deflator and super-MG numeraire are two sides of the same coin. The name deflator comes from the fact that if portfolio ν is used as a baseline, then it's as if one is in an auxiliary market with asset prices $\forall 0 \leq i \leq n, S_i^\nu = \frac{S_i}{X_\nu}$, deflated by the baseline wealth process $\frac{1}{X_\nu}$. Notice that if the asset prices S_i are local MG, and it happens that $\frac{1}{X_\nu}$ is a local MG deflator (e.g. ν is the super-MG numeraire portfolio), the local MG structure of S_i^ν is preserved.

Let's consider again the 3D Bessel Process example mentioned before where

$$X_\nu(t) = \frac{Z_t}{z} \tag{456}$$

is a super-MG numeraire. Then it's immediate that the local MG deflator is

$$\frac{1}{X_\nu(t)} = \frac{z}{Z_t} \tag{457}$$

now consider the capital withdrawal

$$K(t) = \mathbb{I}_{[T,\infty)}(t) \tag{458}$$

with a withdrawal of size 1 happening at time T. It's obvious that $x(K) \leq 1$ since setting aside 1 initial wealth must be enough to finance such K (always more than enough).

Actually, the previous calculations tell us that

$$\mathbb{E}\frac{1}{X_\nu(T)} = \mathbb{E}\frac{z}{Z_T} = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1 \tag{459}$$

and the initial wealth has a lower bound (Exercise 2.28)

$$x(K) \geq \sup_{Y \in \mathscr{Y}} \mathbb{E}\int_0^\infty Y(t)\,dK(t) \geq \int_0^\infty \frac{1}{X_\nu(t)}\,dK(t) = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1 \tag{460}$$

now let's show that actually $x(K) = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1$ can be attained. Consider the following investment strategy

$$\theta(t) = \frac{2}{\sqrt{T-t}}\frac{1}{\sqrt{2\pi}}e^{-\frac{S^2(t)}{2(T-t)}}\mathbb{I}_{[0,T)}(t) \tag{461}$$

so according to $dX(t) = \theta(t)\,dS(t)$, when $t \geq T$ there's no investment so $X(t) = 1$. When $t < T$, we have

$$dX(t) = \frac{2}{\sqrt{T-t}}\frac{1}{\sqrt{2\pi}}e^{-\frac{S^2(t)}{2(T-t)}}\,dS(t) \tag{462}$$

with the solution given as $X(t) = 2\Phi\left(\frac{S(t)}{\sqrt{T-t}}\right) - 1$. To verify this, apply Ito formula and recall that $S = Z$ is the Bessel process

$$dX(t) = 2\varphi\left(\frac{S(t)}{\sqrt{T-t}}\right)\frac{1}{\sqrt{T-t}}\,dS(t) + \varphi'\left(\frac{S(t)}{\sqrt{T-t}}\right)\frac{1}{T-t}\,d\left\langle S,S\right\rangle(t) + \varphi\left(\frac{S(t)}{\sqrt{T-t}}\right)S(t)(T-t)^{-\frac{3}{2}}\,dt \tag{463}$$

$$= \theta(t)\,dS(t) + \theta(t)S(t)\frac{1}{2(T-t)}\,dt - \frac{S(t)\theta(t)}{2(T-t)}\,dt = \theta(t)\,dS(t) \tag{464}$$

so we have verified that under such $\theta(t)$, the wealth process is

$$X(t) = \begin{cases} 2\Phi\left(\frac{S(t)}{\sqrt{T-t}}\right) - 1 & t < T \\ 1 & t \geq T \end{cases} \tag{465}$$

notice that this wealth process just has 1 wealth at time T with initial wealth $X(0) = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1$, so $x(K) \geq 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1$ and

$$x(K) = 2\Phi\left(\frac{z}{\sqrt{T}}\right) - 1 \tag{466}$$

in this specific example, we can derive the closed-form solution for $x(K)$ and the θ, X when such $x(K)$ is attained.

The Fundamental Theorem

Theorem 3. *The following statements are equivalent:*

1. *Market is viable*
2. $\mathscr{Y} \neq \emptyset$ *so there exists local MG deflator*
3. *Super-MG numeraire exists*
4. *Market has locally finite growth, i.e.* $G(t) = \int_0^t \frac{1}{2}\alpha^T(s)c^{-1}(s)\alpha(s)\, dO(s) < \infty$ *for* $\forall t \geq 0$
5. *Boundedness in probability condition holds*
6. *The growth optimal portfolio exists (to be defined)*
7. *The relative log optimal portfolio exists (to be defined)*

Remark. *The* ***maximal growth rate process*** *is defined as*

$$g \overset{def}{=} \sup_{p\in\mathbb{R}^n} \left\{ p^T\alpha - \frac{1}{2}p^T cp \right\} \tag{467}$$

as the maximal possible growth rate achievable in the market and G is the cumulative version by integrating g w.r.t. the operational clock. It's quite easy to see that g is the pointwise maximum of the growth rate when one holds portfolio p, i.e. the best possible action one can take in the market.

Simple matrix calculus tells us that the sup in the definition is achieved when $\alpha = cp$. If α is not even in the range of c, then this sup is ∞, otherwise the sup is achieved at $p = c^{-1}\alpha$ so

$$g = \frac{1}{2}\alpha^T c^{-1}\alpha \tag{468}$$

if α is in the range of c and it's clear that one can always replace the inverse with the pseudoinverse if c is singular.

Due to the importance of this theorem, let's raise some examples to understand market viability. Consider the single asset case where $n = 1$ in the examples.

If $R_t = \mu t + \sigma W_t$, the Black-Scholes model, we have $\alpha_t = \mu, c_t = \sigma^2$ so under the Lebesgue clock

$$G(t) = \frac{\mu^2}{2\sigma^2}t < \infty \tag{469}$$

the market with a single stock following Black-Scholes model is viable.

If $R_t = \mu\sqrt{t} + \sigma W_t$ with a drift term increasing slower, $\alpha_t = \frac{\mu}{2\sqrt{t}}, c_t = \sigma^2$ so under the Lebesgue clock

$$G(t) = \frac{\mu^2}{8\sigma^2}\int_0^t \frac{1}{s}\, ds = \infty \tag{470}$$

so **the market with a single stock following such dynamics is not viable**. Generally, if one has constant diffusion coefficient for R_t as σ, the market is viable iff

$$\int_0^1 \alpha^2(s)\, ds < \infty \tag{471}$$

to explain why this is happening, notice that by the law of iterated logarithm, BM W_t has order $\sqrt{2t\log\log t}$ for large enough t. Whether the market is viable or not has something to do with the comparison of the order of drift and diffusion coefficient. Intuitively, if diffusion coefficient is of higher order, the local MG part takes the lead so the stock price shows some randomness and noise in its asymptotic behavior, making the market not viable.

Remark. *Actually when the market is viable, one can define $v \overset{def}{=} \sqrt{\alpha^T c^{-1}\alpha}$ as the* ***signal-to-noise ratio*** *of the market, a larger v results in a better investing environment since the maximal growth rate will be larger. In this context, $G(t) = \frac{1}{2}\int_0^t v^2(s)\, dO(s)$ is called the* ***market opportunity clock****. It's absolute continuous w.r.t. the operational clock O but not vice versa and it has the nice interpretation as the kinetic energy.*

Lemma 19 (Exercise 2.32)**.** *Assume the market is viable, now there's initial wealth $x \geq 0$, with investment strategy θ and cumulative capital withdrawal $F \in \mathscr{K}$ such that $X(\cdot; x, \theta, F) \geq 0$ (not required to be a numeraire). Define the bankruptcy time*

$$\tau_X = \inf\{t \geq 0 : X(t) = 0\} \tag{472}$$

then $\forall t \geq \tau_X, X(t) = 0$.

Proof. On the event $\{\tau_X < \infty\}$, consider a new wealth process Y such that $Y(t) = X(t + \tau_X)$. Since X has continuous sample path, $Y(0) = 0$. Now consider $F'(t) = F(t + \tau_X) \in \mathscr{K}, F'(0) = 0$ and we have $y(F') = 0$ since $\forall t \geq 0, X(t; x, \theta, F) \geq 0$ there's no need adding any extra initial wealth to Y. Since the market is viable, $F' \equiv 0$. This tells us that for any capital withdrawal F that is financeable for X, $\forall t \geq \tau_X, F(t) = 0$ so $\forall t \geq \tau_X, X(t) = 0$ otherwise one can always put all the wealth into the money market as soon as one reaches positive wealth and then a positive capital withdrawal shall be financeable. ☐

The lemma above gives another interpretation of market viability that **in viable markets bankruptcy is permanent**.

Growth Optimal Portfolio

Start with defining the **cumulative relative growth** process of portfolio π compared to baseline ρ as

$$\Gamma_\pi^\rho \stackrel{def}{=} \Gamma_\pi - \Gamma_\rho \tag{473}$$

and a portfolio ρ is called **growth optimal** if for any portfolio π, Γ_π^ρ is non-increasing. In short, as time goes by the relative growth is always shrinking so the cumulative growth of ρ has the highest increasing rate. The definition of cumulative relative growth comes from the log relative wealth that

$$\log X_\pi^\rho = \log X_\pi - \log X_\rho = \log \mathcal{E}\left(R_\pi\right) - \log \mathcal{E}\left(R_\rho\right) \tag{474}$$

$$= R_\pi - \frac{1}{2}\left\langle R_\pi, R_\pi\right\rangle - R_\rho + \frac{1}{2}\left\langle R_\rho, R_\rho\right\rangle \tag{475}$$

$$= A_\pi - \frac{1}{2}C_{\pi\pi} - A_\rho + \frac{1}{2}C_{\rho\rho} + M_\pi - M_\rho \tag{476}$$

$$= \Gamma_\pi - \Gamma_\rho + M_\pi - M_\rho \tag{477}$$

so naturally the cumulative relative growth is just the finite variation part in the log relative wealth.

Lemma 20 (Exercise 2.39)**.** *Suppose ν is local MG numeraire portfolio, then for any portfolio π, we have*

$$\langle \log X_\pi^\nu, \log X_\pi^\nu\rangle = C_{\pi,\pi}^\nu = 2\Gamma_\nu^\pi = -2\Gamma_\pi^\nu \tag{478}$$

Proof. From direct calculations,

$$\langle \log X_\pi^\nu, \log X_\pi^\nu\rangle = \langle M_\pi - M_\nu, M_\pi - M_\nu\rangle \tag{479}$$

$$= C_{\pi\pi} - C_{\nu\pi} - C_{\pi\nu} + C_{\nu\nu} \tag{480}$$

$$= C_{\pi\pi}^\nu \tag{481}$$

now since X_π^ν is super-MG numeraire, we have $\forall 1 \le i \le n, A_i = C_{i\nu}$ so $A_\pi = C_{\pi\nu}, A_\nu = C_{\nu\nu}$ and $\Gamma_\pi = C_{\pi\nu} - \frac{1}{2}C_{\pi\pi}, \Gamma_\nu = C_{\nu\nu} - \frac{1}{2}C_{\nu\nu}$ and

$$-\Gamma_\pi^\nu = \Gamma_\nu^\pi = \Gamma_\nu - \Gamma_\pi = \frac{1}{2}C_{\nu\nu} - C_{\pi\nu} + \frac{1}{2}C_{\pi\pi} = \frac{1}{2}C_{\pi\pi}^\nu \tag{482}$$

concludes the proof. □

The connection between super-MG numeraire portfolio and the growth optimal portfolio is given by the following theorem. Notice that **market viability is not required for this theorem to hold**!

Theorem 4. *A portfolio is growth optimal iff it's the super-MG numeraire portfolio.*

Proof. Let's just prove one direction here that super-MG numeraire portfolio ν must be growth optimal.

For any portfolio π, X_π^ν is a super-MG and from the lemma above we see $\Gamma_\pi^\nu = -\frac{1}{2}\langle \log X_\pi^\nu, \log X_\pi^\nu \rangle$. It's obvious that the quadratic variation of any process must be non-decreasing so Γ_π^ν must be non-increasing, ν is growth-optimal.

$\square$

Now let's discuss on a viable market with super-MG numeraire portfolio ν, such that X_π^ν is an non-negative super-MG and thus converges almost surely as $t \to \infty$. As a result, for any adapted non-decreasing process F such that $F(0) = 0, F(\infty) = \infty$, it's always true that

$$\limsup_{T\to\infty} \frac{1}{F(T)} \log \frac{X_\pi(T)}{X_\nu(T)} \le 0 \tag{483}$$

however, this conclusion is too coarse and not useful. Instead, if we choose $F = \Gamma_\nu^\pi$ as the cumulative relative growth, under some condition we would get a nicer conclusion for the **long-term growth of the relative wealth**.

Theorem 5. *Suppose markt is viable and X_ν is super-MG numeraire, for any portfolio π,*

$$\left\{ \lim_{T\to\infty} \frac{X_\pi(T)}{X_\nu(T)} \in (0,\infty) \right\} = \{\Gamma_\nu^\pi(\infty) < \infty\} \tag{484}$$

for non-decreasing Γ_ν^π and

$$\lim_{T\to\infty} \frac{1}{\Gamma_\nu^\pi(T)} \log \frac{X_\pi(T)}{X_\nu(T)} = -1 \tag{485}$$

on the event $\{\Gamma_\nu^\pi(\infty) = \infty\}$.

Proof. Although the statement seems hard to prove, there is an incredibly simple way to achieve that.

By the lemma above, $\langle \log X_\pi^\nu, \log X_\pi^\nu \rangle = 2\Gamma_\nu^\pi$ when ν is the super-MG numeraire portfolio. As a result, by the Dambis-Dubins-Schwarz representation (random time change), there exists BM $W(t)$ under a different enlarged filtration such that

$$\log X_\pi^\nu(t) = -\Gamma_\nu^\pi(t) + \sqrt{2}W(\Gamma_\nu^\pi(t)) \tag{486}$$

since $\left\langle \sqrt{2}W(\Gamma_\nu^\pi), \sqrt{2}W(\Gamma_\nu^\pi) \right\rangle(t) = 2\Gamma_\nu^\pi(t)$. It's clear that then $X_\pi^\nu(\infty) < \infty$ iff $\Gamma_\nu^\pi(\infty) < \infty$.

On the other hand,

$$\lim_{T\to\infty} \frac{1}{\Gamma_\nu^\pi(T)} \log \frac{X_\pi(T)}{X_\nu(T)} = \lim_{T\to\infty} \frac{-\Gamma_\nu^\pi(T) + \sqrt{2}W(\Gamma_\nu^\pi(T))}{\Gamma_\nu^\pi(T)} = -1 \tag{487}$$

since $\frac{W_t}{t} \overset{a.s.}{\to} 0$ $(t \to \infty)$, that's why we require the event $\{\Gamma_\nu^\pi(\infty) = \infty\}$ to be happening. $\square$

In natural language, **the log relative wealth behaves asymptotically same as cumulative relative growth with super-MG numeriare portfolio (optimal growth portfolio) as baseline** in a viable market when one of them blows up in terms of time.

Relative Log Optimal Portfolio

ρ is called **relative log-optimal portfolio** if for any portfolio π and any stopping time τ we always have

$$\mathbb{E}(\log X_\pi^\rho(\tau))_+ < \infty, \mathbb{E}\log X_\pi^\rho(\tau) \le 0 \tag{488}$$

for every π and stopping time τ. In other words, the log relative wealth of any portfolio w.r.t. ρ stopped at any time τ cannot have positive expectation.

Theorem 6. *ν is super-MG numaraire portfolio iff it's relative log optimal.*

Proof. We only provide the easy part of the proof.

If ν is super-MG numeraire portfolio, then for any portfolio π, X_π^ν is non-negative super-MG. So

$$0 = \log 1 = \log \mathbb{E}X_\pi^\nu(0) \ge \log \mathbb{E}X_\pi^\nu(\tau) \ge \mathbb{E}\log X_\pi^\nu(\tau) \tag{489}$$

by optional stopping theorem and Jensen's inequality. $\mathbb{E}(\log X_\pi^\rho(\tau))_+ < \infty$ is very easy to verify. $\square$

Expected Log Wealth

As we have discussed the long-term behavior of log wealth in previous sections, under some conditions the log relative wealth shall behave asymptotically the same as cumulative relative growth with the super-MG numeraire portfolio as baseline in a viable market. Naturally, one might expect to see that those two processes also have same expectation, which makes sense since $\log X_\pi^\nu$ and Γ_π^ν are just different by a local MG.

The following theorem makes the statement rigorous. It uses a standard technique in MG theory.

Theorem 7. *In a viable market with ν as super-MG numeraire portfolio, π as any portfolio and τ as any stopping time,*

$$\mathbb{E}\log X_\pi^\nu(\tau) = \mathbb{E}\Gamma_\pi^\nu(\tau) = -\mathbb{E}\Gamma_\nu^\pi(\tau) \tag{490}$$

Proof. Denote $N = X_\pi^\nu$, a continuous strictly positive super-MG with $N(0) = 1$, by Ito formula

$$\log N(t) = \int_0^t \frac{1}{N(s)}\, dN(s) - \frac{1}{2}\left\langle \log N, \log N\right\rangle(t) \tag{491}$$

denote the stochastic integral as L and the finite variation part as B so $\log N = L - B$.

Discuss whether $\mathbb{E}B(\tau) < \infty$ or not since $2B$ is the quadratic variation of L. If $\mathbb{E}B(\tau) < \infty$, $L(t \wedge \tau)$ is L^2 MG so it's done. Otherwise, use a sequence of stopping time to truncate B that

$$\tau_k = \inf\{t : B(t) \geq k\} \wedge \tau \tag{492}$$

and the monotone convergence still proves the result.

$\square$

Remark. *With π taken as the trivial portfolio, it immediately follows that*

$$\mathbb{E}\log X_\nu(\tau) = \mathbb{E}\log X_\nu^\pi(\tau) = -\mathbb{E}\log X_\pi^\nu(\tau) = \mathbb{E}\Gamma_\nu^\pi(\tau) = \mathbb{E}\Gamma_\nu(\tau) = \mathbb{E}G(\tau) \tag{493}$$

for any stopping time τ and super-MG numeraire portfolio ν. Recall that $G(t) = \int_0^t \frac{1}{2}\alpha^T(s)c^{-1}(s)\alpha(s)\, dO(s)$ is the maximal growth rate one can get. The last equality holds since ν is always growth optimal.

More generally, for any portfolio π,

$$\mathbb{E}\log X_\pi(\tau) = \mathbb{E}\Gamma_\pi(\tau) \tag{494}$$

if $\mathbb{E}\Gamma_\pi(\tau) < \infty$ follows from subtracting the two equations above but this is no longer equal to $\mathbb{E}G(\tau)$.

Minimum Expected Opportunity Time for Given Level of Wealth

Under the framework of portfolio theory, one may put up a lot of interesting optimization problems. The first one to consider is to **minimize the expected opportunity time to reach a given level of wealth**. Let's say we have initial log wealth $\log X_0 = 0$ and we want to achieve log wealth $l > 0$ by picking up some portfolio π that helps us minimize the expected time we have to wait until the log wealth reaches l. Naturally, set up the stopping time

$$\tau_{\pi,l} = \inf\{t : \log X_\pi(t) = l\} \tag{495}$$

however, this $\tau_{\pi,l}$ corresponds to the time in the real life but under some conditions the market may behave very bad such that it's impossible to reach the objective level. As a result, we want the expected waiting time to be as short as possible **as market opportunities permit**. This leads us choosing the market opportunity clock G instead of the operational clock O. The optimization problem is formed as

$$\inf_\pi G(\tau_{\pi,l}) \tag{496}$$

and the solution is given by $\pi = \nu$ as the super-MG portfolio if it exists.

Theorem 8. *In a viable market with $G(\infty) = \infty$, and ν to be super-MG numeraire portfolio, for any portfolio π,*

$$\forall l > 0, \mathbb{E}G(\tau_{\nu,l}) = \inf_\pi G(\tau_{\pi,l}) = l \tag{497}$$

Proof. $G(\infty) = \infty$ is just to make sure $\forall l > 0, \tau_{\nu,l} < \infty$ *a.s.*. Now by the expected log wealth formula

$$l = \mathbb{E}\log X_\nu(\tau_{\nu,l}) = \mathbb{E}G(\tau_{\nu,l}) \tag{498}$$

and for any π, since ν has optimal growth so $\Gamma_\pi^\nu = \Gamma_\pi - \Gamma_\nu = \Gamma_\pi - G$ is non-increasing so $\Gamma_\pi - G \leq 0$,

$$\mathbb{E}G(\tau_{\pi,l}) \geq \mathbb{E}\Gamma_\pi(\tau_{\pi,l}) = \mathbb{E}\log X_\pi(\tau_{\pi,l}) = l \tag{499}$$

follows from expected log wealth formula for any portfolio π. It concludes the proof. □

Robust Optimization on Long Term Growth

The **robust optimization on long term growth** aims to pick a portfolio that has the best long term growth in the worst scenario. To describe the problem setting, let's introduce some notations. Let $\Omega = C(\mathbb{R};\mathbb{R}^n)$ be the Wiener space, i.e. the space of all continuous functions from $\mathbb{R}$ to $\mathbb{R}^n$. Equip $(\Omega, \mathscr{F})$ with probability measure $\mathbb{Q}$ as a Wiener measure so any sample points $\omega \in \Omega$ is a BM trajectory. The filtration $\mathscr{F}_t$ is set as the right-continuous augmentation of the natural filtration $\sigma(W_s, s \leq t)$ with $\mathscr{F} = \mathscr{F}_\infty$. Define the stochastic process W on this probability space as $W(\omega, t) \overset{def}{=} \omega(t)$ so it's an n-dimensional BM under $\mathbb{Q}$.

Now consider $\Pi = \{\mathbb{P} \in \mathscr{P}(\Omega) : \mathbb{P} << \mathbb{Q}\}$ as the set of all probability measures $\mathbb{P}$ that is absolute continuous w.r.t. $\mathbb{Q}$ on $(\Omega, \mathscr{F})$. By Girsanov theorem, there exists a predictable process $\lambda^\mathbb{P}$ such that $\int_0^T ||\lambda^\mathbb{P}(t)||^2\,dt < \infty$ $\mathbb{P}-a.e.$ and that

$$W^\mathbb{P}(t) = W(t) - \int_0^t \lambda^\mathbb{P}(s)\,ds \tag{500}$$

is an n-dimensional BM under $\mathbb{P}$. Assume σ is matrix-valued non-singular predictable process, and define $c = \sigma\sigma^T$ such that $\forall T > 0, \int_0^T \sum_{i=1}^n c_{ii}(t)\,dt < \infty$ $\mathbb{Q} - a.e.$.

The stock market is built up with the dynamics

$$S_i(t) = S_i(0)e^{\int_0^t \sum_{j=1}^n \sigma_{ij}(s)\,dW_j(s) - \frac{1}{2}\int_0^t c_{ii}(s)\,ds} = S_i(0)\mathcal{E}\left(\int_0^t \sum_{j=1}^n \sigma_{ij}(s)(\lambda_j^\mathbb{P}(s)\,ds + dW_j^\mathbb{P}(s))\right) \tag{501}$$

for $1 \leq i \leq n$ so there are n different stocks under any given probability measure $\mathbb{P}$. The reason we are rebuilding everything here is that we hope the $\mathbb{P}$ to be changing in our context of optimization. Notice that $\lambda^\mathbb{P}$ is interpreted as the **market price of risk under probability measure** $\mathbb{P}$.

Remark. *Consistent with the notation in portfolio theory, the cumulative return can be formed as*

$$R_i(t) = \int_0^t \sum_{j=1}^n \sigma_{ij}(s)\,dW_j(s) = \int_0^t \sum_{j=1}^n \sigma_{ij}(s)(\lambda_j^\mathbb{P}(s)\,ds + dW_j^\mathbb{P}(s)) \tag{502}$$

so that $S_i = S_i(0)\mathcal{E}(R_i)$ still holds. So for given probability measure $\mathbb{P}$, under the Lebesgue clock, the covariation rate matrix is $c = \sigma\sigma^T$ and the mean return rate is $\alpha^\mathbb{P} = \sigma\lambda^\mathbb{P}$. This explains why $\lambda^\mathbb{P}$ is called the market price of risk.

Now we define **the long term growth (asymptotic growth) under portfolio** π **and probability measure (environment)** $\mathbb{P}$ as

$$\mathrm{ag}(\pi;\mathbb{P}) = \sup\left\{g \in \mathbb{R} : \liminf_{T\to\infty} \frac{1}{T}\log X_\pi(T) \geq g,\ \mathbb{P} - a.e.\right\} \tag{503}$$

the largest g asymptotically dominated by the log wealth per time unit under measure $\mathbb{P}$.

Our optimization shall not take place for any $\mathbb{P} \in \Pi$ so we define an non-empty convex closed set $\Lambda \subset \mathbb{R}^n$ and Π_Λ as the collection of probability measure $\mathbb{P}$ for which $\lambda^\mathbb{P} \in \Lambda$ *a.e.*. In our optimization problem, we restrict ourselves to the probability measure $\mathbb{P}$ in Π_Λ by considering only the $\lambda^\mathbb{P}$ in Λ.

The optimization problem is given by

$$\mathrm{ag}_\Lambda^* = \sup_\pi \inf_{\mathbb{P}\in\Pi_\Lambda} \mathrm{ag}(\pi;\mathbb{P}) \tag{504}$$

the interpretation of the problem is crucial here. **The market chooses $\mathbb{P}$ from Π_Λ as the environment probability measure that any single investor has to accept, but we want to make sure that no matter how the market environment changes, our worst scenario long term growth is always maximized by taking portfolio π.** This is also called the **robust** optimization problem since one has the ability to stay in a rather good state even when the market environment is bad.

Remark. *This can be understood as a two-agent zero-sum game and we are finding the Stackelberg equilibrium to this game. In the more general setting when the conditions listed below such that the simplification of the problem does not work, one can naturally think of using reinforcement learning algorithm to solve this problem. Besides, thinking about different notions of equilibrium in the financial background and multi-agent settings also makes sense.*

Theorem 9. *The solution to the optimization problem above is given by the following characterization*

$$ag_\Lambda^* = \frac{1}{2}\inf_{l\in\Lambda} ||l||^2 \tag{505}$$

the optimal l and portfolio π is given as

$$l_\Lambda^* = \arg\min_{l\in\Lambda} ||l|| \in \Lambda, \pi_\Lambda^* = (\sigma^T)^{-1} l_\Lambda^* \tag{506}$$

such that

$$\forall\mathbb{P}\in\Pi_\Lambda, ag(\pi_\Lambda^*;\mathbb{P}) \geq ag_\Lambda^* \tag{507}$$

Proof. The proof shall make use of the super-MG numeraire but the problem is how shall we find the super-MG numeraire under this framework. We have proved above that if ν is the super-MG numeraire portfolio then it shall satisfy $\nu = c^{-1}\alpha^{\mathbb{P}}$ under some measure $\mathbb{P}$. Notice that $l_\Lambda^* \in \Lambda$ and that $W^{\mathbb{P}}(t) + \int_0^t \lambda^{\mathbb{P}}(s)\,ds = W(t)$, so why don't we **consider the measure $\mathbb{P}_\Lambda^* \in \Pi_\Lambda$ under which W becomes the BM with drift l_Λ^***, i.e. $\int_0^t \lambda^{\mathbb{P}_\Lambda^*}(s)\,ds = l_\Lambda^* t, \lambda^{\mathbb{P}_\Lambda^*}(t) = l_\Lambda^*$, it's clear that

$$c^{-1}\alpha^{\mathbb{P}_\Lambda^*} = (\sigma^T)^{-1}\sigma^{-1}\sigma\lambda^{\mathbb{P}_\Lambda^*} = (\sigma^T)^{-1}\lambda^{\mathbb{P}_\Lambda^*} = (\sigma^T)^{-1} l_\Lambda^* = \pi_\Lambda^* \tag{508}$$

so this clever trick leads to the fact that **under $\mathbb{P}_\Lambda^*$, π_Λ^* is the super-MG numeraire portfolio**. It's then obvious that

$$\mathrm{ag}_\Lambda^* \leq \sup_\pi \mathrm{ag}(\pi;\mathbb{P}_\Lambda^*) \leq \mathrm{ag}(\pi_\Lambda^*;\mathbb{P}_\Lambda^*) = \frac{1}{2}||l_\Lambda^*||^2 = \frac{1}{2}\inf_{l\in\Lambda} ||l||^2 \tag{509}$$

where we are using the fact that super-MG numeraire portfolio achieves the largest long term growth (formed as a lemma below). The first equality comes from the lemma below that

$$\frac{1}{T}\log X_{\pi_\Lambda^*}(T) = \frac{1}{T}\int_0^T \left[(l_\Lambda^*)^T\lambda^{\mathbb{P}_\Lambda^*}(t) - \frac{1}{2}||l_\Lambda^*||^2\right]\,dt + \frac{1}{T}(l_\Lambda^*)^T W^{\mathbb{P}_\Lambda^*}(T) \tag{510}$$

$$= \frac{1}{2}||l_\Lambda^*||^2 + \frac{1}{T}(l_\Lambda^*)^T W^{\mathbb{P}_\Lambda^*}(T) \overset{\mathbb{P}_\Lambda^*-a.s.}{\to} \frac{1}{2}||l_\Lambda^*||^2\ (T\to\infty) \tag{511}$$

For the second half of the proof, take $\forall\mathbb{P}\in\Pi_\Lambda$, by the lemma below,

$$\frac{1}{T}\log X_{\pi_\Lambda^*}(T) = \frac{1}{T}\int_0^T \left[(l_\Lambda^*)^T\lambda^{\mathbb{P}}(t) - \frac{1}{2}||l_\Lambda^*||^2\right]\,dt + \frac{1}{T}(l_\Lambda^*)^T W^{\mathbb{P}}(T) \tag{512}$$

with the definition of l_Λ^* and the fact that Λ is convex, we know $||l_\Lambda^*||^2 \le (l_\Lambda^*)^T\lambda^{\mathbb{P}}(t)$ by the projection of points onto convex sets so the first term on the RHS is larger than $\frac{1}{T}\int_0^T \frac{1}{2}||l_\Lambda^*||^2\,dt$.

However, the second term on the RHS goes to 0 as $T\to\infty$ since $\frac{W^{\mathbb{P}}(T)}{T} \overset{\mathbb{P}-a.s.}{\to} 0\ (T\to\infty)$. We have thus proved that $\liminf_{T\to\infty}\frac{1}{T}\log X_{\pi_\Lambda^*}(T) \ge \frac{1}{2}||l_\Lambda^*||^2 = ag_\Lambda^*\ \mathbb{P}-a.s.$ for $\forall\mathbb{P}\in\Pi_\Lambda$.

Combine both parts to see that $\text{ag}_\Lambda^* = \frac{1}{2}\inf_{l\in\Lambda}||l||^2$ concludes the proof.

□

The two extra properties we use in the proof are stated as lemmas below.

Lemma 21. *$\forall\mathbb{P}\in\Pi_\Lambda$, we always have the representation for log wealth under portfolio π_Λ^* defined above in the theorem that*

$$\frac{1}{T}\log X_{\pi_\Lambda^*}(T) = \frac{1}{T}\int_0^T \left[(l_\Lambda^*)^T\lambda^{\mathbb{P}}(t) - \frac{1}{2}||l_\Lambda^*||^2\right]\,dt + \frac{1}{T}(l_\Lambda^*)^T W^{\mathbb{P}}(T) \tag{513}$$

Proof. $\forall\mathbb{P}\in\Pi_\Lambda$, since $X_{\pi_\Lambda^*} = \mathcal{E}\left(R_{\pi_\Lambda^*}\right)$,

$$\log X_{\pi_\Lambda^*}(T) = \left[R_{\pi_\Lambda^*}(T) - \frac{1}{2}\left\langle R_{\pi_\Lambda^*}, R_{\pi_\Lambda^*}\right\rangle(T)\right] \tag{514}$$

notice that $\left\langle R_{\pi_\Lambda^*}, R_{\pi_\Lambda^*}\right\rangle = (\pi_\Lambda^*)^T c\pi_\Lambda^* = (l_\Lambda^*)^T\sigma^{-1}\sigma\sigma^T(\sigma^T)^{-1}l_\Lambda^* = ||l_\Lambda^*||^2$, combined with the basic setting of portfolio theory,

$$\log X_{\pi_\Lambda^*}(T) = R_{\pi_\Lambda^*}(T) - \frac{1}{2}||l_\Lambda^*||^2 \tag{515}$$

$$= \int_0^T \sum_{i=1}^n (\pi_\Lambda^*)_i(t)\sum_{j=1}^n \sigma_{ij}(t)(\lambda_j^{\mathbb{P}}(t)\,dt + dW_j^{\mathbb{P}}(t)) - \frac{1}{2}||l_\Lambda^*||^2 \tag{516}$$

$$= \int_0^T [\pi_\Lambda^*(t)]^T\sigma(t)\lambda^{\mathbb{P}}(t)\,dt + \int_0^T d[\pi_\Lambda^*(t)]^T\sigma(t)W^{\mathbb{P}}(t) - \frac{1}{2}||l_\Lambda^*||^2 \tag{517}$$

$$= \int_0^T (l_\Lambda^*)^T\lambda^{\mathbb{P}}(t)\,dt + \int_0^T d(l_\Lambda^*)^T W^{\mathbb{P}}(t) - \frac{1}{2}||l_\Lambda^*||^2 \tag{518}$$

$$= \int_0^T (l_\Lambda^*)^T\lambda^{\mathbb{P}}(t)\,dt + (l_\Lambda^*)^T W^{\mathbb{P}}(T) - \frac{1}{2}||l_\Lambda^*||^2 \tag{519}$$

concludes the proof. □

Lemma 22 (Exercise 2.44)**.** *In a viable market, if there is $G(\infty) = \infty$, then for super-MG numeraire portfolio ν and any portfolio π,*

$$\limsup_{T\to\infty} \frac{\log X_\pi(T)}{G(T)} \le 1 = \lim_{T\to\infty} \frac{\log X_\nu(T)}{G(T)} \tag{520}$$

so ***the super-MG numeraire portfolio achieves the optimal long term growth****.*

Proof. It's clear that $G(T) = \Gamma_\nu(T)$ so by taking π as the trivial portfolio, and notice that now $\{\Gamma_\nu^\pi(\infty) = \infty\}$ happens, we have proved in previous context the long term growth of the log relative wealth so

$$-\lim_{T\to\infty} \frac{1}{G(T)} \log X_\nu(T) = \lim_{T\to\infty} \frac{1}{\Gamma_\nu^\pi(T)} \log X_\pi^\nu(T) = -1 \tag{521}$$

this proves that $1 = \lim_{T\to\infty} \frac{\log X_\nu(T)}{G(T)}$. The other part follows from the fact that

$$\limsup_{T\to\infty} \left(\frac{\log X_\pi(T)}{G(T)} - \frac{\log X_\nu(T)}{G(T)} \right) = \limsup_{T\to\infty} \frac{\log X_\pi^\nu(T)}{\Gamma_\nu(T)} \le 0 \tag{522}$$

since $\Gamma_\nu(\infty) = \infty$ and X_π^ν is non-negative super-MG thus converges. □

Remark. *The robust optimization on long term growth is a variational (max-min) problem so it's generally hard to solve. However, by restricting ourselves to Π_Λ and by the help of super-MG numeraire portfolio, this problem is very easy to solve, equivalent to finding the point l_Λ^* with the least norm in a convex set Λ.*

The conditions that make this problem easy is the ***convexity of*** *Λ and the* ***existence of super-MG numeraire****. Combined with the optimal long term growth property, super-MG numeraire structure helps us prove the upper bound of the variational problem in a very easy way. Actually, whenever we are doing some optimization problems in the framework of portfolio theory, as long as* ***the optimality is shared by the super-MG numeraire portfolio****, we would always expect it to help us reduce the problem into a simpler one.*

Local-MG Numeraire for Stock Portfolio

Restrict to the stock portfolios, we want to investigate what conclusions remain the same and what will change. Consider all portfolios below to be stock portfolios, $\rho \in H_{n-1}$ is called the **local-MG numeraire stock portfolio** if for $\forall \pi \in H_{n-1}$, $\frac{X_\pi}{X_\rho}$ is always a local-MG. As previously shown, this can be characterized by the fact that $\forall 1 \leq i \leq n$, all stock prices in the auxiliary market $S_i^\rho = \frac{S_i}{X_\rho}$ as if discounted by the baseline ρ are local MG.

Theorem 10. *ρ is local MG numeraire stock portfolio iff there exists continuous adapted process F of finite variation on compact time interval such that*

$$\forall 1 \leq i \leq n, A_i = F + C_{i\rho} \tag{523}$$

Proof. ρ is local MG numeraire stock portfolio iff S_i^ρ are all local MG iff R_i^ρ are all local MG for $1 \leq i \leq n$. Since $R_i^\rho = C_{\rho\rho} - R_\rho + R_i - C_{i\rho}$, the finite variation part in R_i^ρ which is $C_{\rho\rho} - A_\rho + A_i - C_{i\rho}$ is equal to 0 iff R_i^ρ is local MG by Doob-Meyer decomposition. As a result, one can set $F = A_\rho - C_{\rho\rho}$ to see that ρ is local MG numeraire stock portfolio iff $F = A_i - C_{i\rho}$.

Lastly, easy to verify that F is continuous, adapted and of finite variation on compact time interval. □

Remark. *If one compares the theorem above with the theorem that characterizes super-MG numeraire portfolio, one will find the difference that in a general market with money market one has to also make sure that $R_0^\rho = C_{\rho\rho} - R_\rho$ is a local MG, resulting in $F \equiv 0$ in the theorem above.*

However, for local-MG numeraire stock portfolio, we never invest in the money market so we don't care if R_0^ρ is local MG, that's why the finite variation part of R_0^ρ is set as $-F$. From another perspective, one can understand the theorem above by absorbing F into A_i to create a new market where $\forall 1 \leq i \leq n, R_i^* = R_i - F$ and it's clear that $\langle R_i^*, R_i^* \rangle = \langle R_i, R_i \rangle = C_{ii}$ does not change all the covariations. In this new market, $A_i^* = A_i - F, C_{i\rho}^* = C_{i\rho}$ so $\forall 1 \leq i \leq n, A_i^* = C_{i\rho}^*$. By the characterization of super-MG numeraire portfolio, ρ is now a super-MG numeraire portfolio in this new market with no investment in the money market.

*In the theorem above F does not depend on i so it's not a very weak formulation. Other than that, notice that saying ρ is a local MG numeraire stock portfolio is weaker than saying ρ is a super-MG numeraire portfolio which is also a stock portfolio. The previous statement means that ρ beats all other stock portfolios but the latter one means that ρ beats all other portfolios (not only stock portfolios) with a coincidence that itself is a stock portfolio. Naturally, **the existence of local MG numeraire stock portfolio cannot imply the existence of of super-MG numeraire portfolio and market viability**.*

An explicit formula of the local MG numeraire stock portfolio can be provided under market viability condition by simple linear algebra.

Theorem 11. *Assume market is viable, then the local MG numeraire stock portfolio is given by*

$$\rho = c^\dagger \alpha + (1 - \vec{1}^T c^\dagger \alpha) \frac{c^\dagger \vec{1}}{\vec{1}^T c^\dagger \vec{1}} \tag{524}$$

if $\vec{1}$ is in the range of c and

$$\rho = c^\dagger \alpha + (1 - \vec{1}^T c^\dagger \alpha) \frac{\vec{1} - cc^\dagger \vec{1}}{||\vec{1} - cc^\dagger \vec{1}||^2} \tag{525}$$

if $\vec{1}$ is not in the range of c. The pseudoinverse here is given by $c^\dagger = \lim_{m\to\infty}\left(c+\frac{1}{m}I\right)^{-2}c$.

Proof. ρ is local MG numeraire stock portfolio iff there exists adapted continuous finite variation process F such that $\forall 1\le i\le n, A_i = F + C_{i\rho}$. Written f as the rate process of F under operational clock, we see $\alpha_i = f + c_{i\rho}$ so $\alpha = f\vec{1} + c\rho, c\rho = \alpha - f\vec{1}$. We want to take pseudoinverse on both sides but since $cc^\dagger\alpha = \alpha$ iff α is in the range of c, we need to discuss whether $\alpha - f\vec{1}$ is in the range of c, i.e. if $\vec{1}$ is in the range of c (by the market viability, super-MG numeraire portfolio ν exists and thus α is in the range of c since $c\nu = \alpha$).

The first case is where $\vec{1}$ is in the range of c, then it's easy to see that $\rho = c^\dagger(\alpha - f\vec{1})$. To derive f, consider $\vec{1}^T\rho = 1$ so

$$\vec{1}^T c^\dagger(\alpha - f\vec{1}) = 1 \tag{526}$$

solve to get $f = \frac{\vec{1}^T c^\dagger \alpha - 1}{\vec{1}^T c^\dagger \vec{1}}$ gives the first formula.

The second case is where $\vec{1}$ is not in the range of c, so $cc^\dagger\vec{1} \neq \vec{1}$ and we have to make use of the property that $cc^\dagger$ is the projection operator on the range of c so $\vec{1} - cc^\dagger\vec{1}$ is orthogonal to the range of c and $c(\vec{1} - cc^\dagger\vec{1}) = 0$ so we verify that

$$\vec{1}^T\rho = \vec{1}^T\left(c^\dagger\alpha + (1 - \vec{1}^T c^\dagger\alpha)\frac{\vec{1} - cc^\dagger\vec{1}}{||\vec{1} - cc^\dagger\vec{1}||^2}\right) = 1 \tag{527}$$

by noticing that fact that $||\vec{1} - cc^\dagger\vec{1}||^2 = \vec{1}^T(\vec{1} - cc^\dagger\vec{1})$ since $c\vec{1} = cc c^\dagger\vec{1}$. Other than that, $c\rho = \alpha$ for such ρ so this corresponds to the case where $f = 0$.

After figuring out the motivation of those two formulas, we give a proof that these formulas do give ρ as a local MG numeraire stock portfolio. For any stock portfolio π, it's always true that

$$(\pi - \rho)^T c(\rho - \nu) = 0 \tag{528}$$

since $c\nu = \alpha$ and $c\rho = \alpha - f\vec{1}$ with $f = \frac{\vec{1}^T c^\dagger\alpha - 1}{\vec{1}^T c^\dagger \vec{1}}\mathbb{I}_{\vec{1}\in\text{range}(c)}$ for given ρ. When $\vec{1}$ is not in the range of c, $c\rho - \alpha = 0$ and when $\vec{1}$ is in the range of c, $(\pi-\rho)^T(c\rho - \alpha) = -\frac{\vec{1}^T c^\dagger\alpha - 1}{\vec{1}^T c^\dagger\vec{1}}(\pi - \rho)^T\vec{1} = 0$ since both π, ρ are stock portfolios.

As a result, $\alpha_{\pi-\rho} = c_{\pi-\rho,\rho}$ so $R_{\pi-\rho} - C_{\pi-\rho,\rho}$ is local MG thus $\frac{X_\pi}{X_\rho}$ is local MG. □

Remark. *The construction of local MG numeraire stock portfolio takes advantage of the existence of the super-MG numeraire portfolio ν since such existence ensures that $\alpha \in$ range(c).*

Perfectly Balanced Market

The market portfolio μ is the most natural example of a stock portfolio so a question to ask is what happens if the market portfolio itself is the local MG numeraire stock portfolio beating all other stock portfolios. If this happens, the market is called **perfectly balanced**, i.e.

$$\forall 1 \leq i \leq n, S_i^\mu = \frac{S_i}{X_\mu} = \Sigma(0)\frac{S_i}{\Sigma} = \Sigma(0)\mu_i \tag{529}$$

are local MG (recall that $X_\mu = \frac{\Sigma}{\Sigma(0)}$). In a perfectly balanced market, the market portfolio weights μ_i are themselves local MG, thus

$$\frac{d\mu_i(t)}{\mu_i(t)} = dR_i^\mu(t) = dM_i(t) - dM_\mu(t) = (e_i - \mu(t))^T dM(t) \tag{530}$$

where the first equation follows from the definition of market portfolio and the second equation follows from μ being the local MG numeraire stock portfolio. As a result, in a perfectly balanced market the market portfolio has the **direct SDE characterization**

$$\forall 1 \leq i \leq n, \mu_i(t) = \mu_i(0)\mathcal{E}\left(\int_0^\cdot (e_i - \mu(s))^T dM(s)\right)(t) \tag{531}$$

fully determined by the $M_1, ..., M_n$, the noise part of the cumulative return.

Lemma 23 (Exercise 2.66)**.** *Let $M_1, ..., M_n$ be continuous local MG starting from 0, then for any initial condition $\mu(0) \in ri\left(\Delta_{(n-1)}\right)$, it can be uniquely extended to a process μ that solves*

$$\forall 1 \leq i \leq n, \mu_i(t) = \mu_i(0)\mathcal{E}\left(\int_0^\cdot (e_i - \mu(s))^T dM(s)\right)(t) \tag{532}$$

Moreover, for any initial condition such that $\mu(0)$ is a long-only stock portfolio, such SDE always have a unique solution and the solution must be a stock portfolio (existence and uniqueness argument of SDE).

At this point, we can make sure that given any M as continuous local MG and the initial value $\mu(0)$, one can always construct a perfectly balanced market. Firstly, since perfectly balanced market indicates the existence of the local MG numeraire stock portfolio, we pick some continuous adapted finite variation process F, and define

$$\forall 1 \leq i \leq n, R_i = F + C_{i\mu} + M_i \tag{533}$$

(such that $A_i = F + C_{i\mu}$ holds), following the definition of market portfolio, set

$$S_i(0) = \Sigma(0)\mu_i(0) \tag{534}$$

where μ_i is given in the lemma above and $S_i = S_i(0)\mathcal{E}(R_i)$ provides the stock price dynamics. It's easy to see that such market is perfectly balanced and μ_i as bounded local MG must be MG. By the MG convergence theorem,

$\mu_i(t) \stackrel{a.s.}{\to} \mu_i(\infty)$ $(t \to \infty)$ with the limit $\mu_i(\infty)$ as a random variable taking values in $[0,1]$. At this point, a natural question to ask is that whether $\mu_i(\infty)$ can reach 1, i.e. in long-term all capitalization aggregates to a single company, and what's the probability that it reaches 1. It turns out that in perfectly balanced market, the answer is very intuitive: **the probability of all capitalization aggregating to company i is proportional to the initial capitalization of company i.**

Theorem 12. *Assume the market is perfectly balanced and the stock prices are sufficiently distinct, i.e.*

$$\forall i \neq j, \mathbb{P}\left(\langle M_i - M_j, M_i - M_j \rangle (\infty) < \infty\right) = 0 \tag{535}$$

then $\mathbb{P}\left(\mu_i(\infty) = 1\right) = \mu_i(0)$.

Proof. On the event $\{\mu_i(\infty) > 0\}$, we have $C^\mu_{ii}(\infty) = \langle \log \mu_i, \log \mu_i \rangle (\infty) < \infty$. Since μ is local MG numeraire stock portfolio, $C^\mu_{ii} = 2\Gamma^i_\mu$ by the previously proved lemma for relative growth (we proved for super-MG numeraire portfolio but here a slight generalization still holds).

Now consider on the event $\{\mu_i(\infty) > 0, \mu_j(\infty) > 0\}$ where $i \neq j$, we have $\int_0^\infty |d\Gamma^j_i(t)| \leq \Gamma^j_\mu(\infty) + \Gamma^i_\mu(\infty) = \frac{1}{2}(C^\mu_{ii}(\infty) + C^\mu_{jj}(\infty)) < \infty$ so the total variation of Γ^j_i on $(0,\infty)$ is finite. Recall that $M_i - M_j = \log \frac{\mu_i}{\mu_j} - \log \frac{\mu_j(0)}{\mu_i(0)} - \Gamma^j_i$ is continuous local MG so on the event $\{\mu_i(\infty) > 0, \mu_j(\infty) > 0\}$ we must have

$$(M_i - M_j)(\infty) = \log \frac{\mu_i(\infty)}{\mu_j(\infty)} - \log \frac{\mu_j(0)}{\mu_i(0)} - \Gamma^j_i(\infty) \tag{536}$$

exists and is finite almost surely so $\langle M_i - M_j, M_i - M_j \rangle (\infty) < \infty$ holds, which has probability 0.

As a result, we have proved that $\forall i \neq j$, almost surely the two events $\{\mu_i(\infty) > 0\}, \{\mu_j(\infty) > 0\}$ do not happen simultaneously. This tells us that $\mu_i(\infty)$ takes values in $\{e_1, ..., e_n\}$ almost surely and by optional stopping theorem, since μ_i is U.I. MG,

$$\mathbb{P}\left(\mu_i(\infty) = 1\right) = \mathbb{P}\left(\mu_i(\infty) = e_i\right) = \mathbb{E}\mu_i(\infty) = \mu_i(0) \tag{537}$$

concludes the proof.

$\square$

Remark. *The condition of this theorem above means that for any two stocks i, j, the aggregate relative covariation blows up at ∞, i.e. the relative noise keep varying at all time, meaning that* ***any two stocks are sufficiently distinct****. Notice that $\langle M_i - M_j, M_i - M_j \rangle = \langle \log \mu_i - \log \mu_j, \log \mu_i - \log \mu_j \rangle$ is easier to calculate in practice.*

We prove the theorem above by first arguing that ***in long term all capitalization must aggregate in one of the companies****, then optional stopping can be applied to conclude the proof. The conclusion is intuitively correct since the market is perfectly balanced, meaning that whoever has the largest capitalization at the beginning has the largest probability of winning at last. Now we can also see the motivation of defining the perfectly balanced market as a market where market portfolio is the local MG numeraire stock portfolio. In other words, in such a market, everyone will eventually hold the market portfolio and trading is then discouraged so the* ***market equilibrium*** *will be achieved (supply and demand of each stock is cleared).*

Lemma 24 (Exercise 2.68)**.** *Consider process N defined by*

$$N(t) = \frac{1}{2} + \int_0^t N(s)(1 - N(s))\, dW(s) \tag{538}$$

let Σ be any strictly positive continuous semi-MG such that $S_1 = N\Sigma, S_2 = (1 - N)\Sigma$, figure out the long-term behavior of the market portfolio μ.

Proof. Obviously $\mu_1 = N, \mu_2 = 1 - N$, from $dN(t) = N(t)(1 - N(t))\, dW(t), N(0) = \frac{1}{2}$ we know that N always stays in $(0, 1)$ (leave the proof for later) so everything is well-defined, and N is a bounded continuous local MG, thus a continuous MG. This tells us that μ_1, μ_2 are both continuous MG, the market is perfectly balanced.

Notice that $\langle M_2 - M_1, M_2 - M_1\rangle = \langle \log \mu_2 - \log \mu_1, \log \mu_2 - \log \mu_1\rangle$ where

$$d \log N(t) = \frac{1}{N(t)}\, dN(t) - \frac{1}{2N^2(t)}\, d\langle N, N\rangle(t) \tag{539}$$

$$d \log(1 - N(t)) = -\frac{1}{1 - N(t)}\, dN(t) - \frac{1}{2(1 - N(t))^2}\, d\langle N, N\rangle(t) \tag{540}$$

so $\langle M_2 - M_1, M_2 - M_1\rangle(t) = \int_0^t \frac{1}{N^2(s)(1-N(s))^2}\, d\langle N, N\rangle(s) = t$ (actually $\log \mu_2 - \log \mu_1$ is a BM). As a result, $\langle M_2 - M_1, M_2 - M_1\rangle(\infty) = \infty$, the condition of the theorem holds and we get $\mathbb{P}(\mu_1(\infty) = 1) = \mu_1(0) = \frac{1}{2} = \mathbb{P}(\mu_2(\infty) = 1)$ Intuitively, N is completely symmetric w.r.t. $\frac{1}{2}$ so two companies have the same probability of winning at last. $\square$

Here in order to prove the behavior of the solution to the SDE for N, we use **Feller's test for explosion**. It's good to know such technique since it helps figure out if the solution to a SDE exits an area without solving out the SDE.

Theorem 13. *Consider SDE*

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t \tag{541}$$

with deterministic initial value condition $X_0 = x \in (l, r)$ for some $l < r$ and X is a weak solution to the SDE. The SDE satisfies the condition that $\forall x \in \mathbb{R}, \sigma^2(x) > 0$ and $\forall x \in \mathbb{R}, \exists \varepsilon > 0, \int_{x-\varepsilon}^{x+\varepsilon} \frac{|b(y)|}{\sigma^2(y)}\, dy < \infty$. Define

$$S = \inf\{t \geq 0 : X_t \notin (l, r)\} \tag{542}$$

and for some fixed constant $c \in (l, r)$,

$$p(x) = \int_c^x e^{-2\int_c^\xi \frac{b(t)}{\sigma^2(t)}\, dt}\, d\xi \tag{543}$$

which can also be written as $p''(x) = -2\frac{b(x)}{\sigma^2(x)}p'(x)$, there are four cases below (in (b),(c),(d), we do not ensure $S < \infty$ a.s.)

(a): If $p(l+) = -\infty, p(r-) = +\infty$ then $\mathbb{P}(S = \infty) = 1$ and such X is recurrent.

(b): If $p(l+) > -\infty, p(r-) = +\infty$ then $\mathbb{P}(\lim_{t\nearrow S} X_t = l) = \mathbb{P}\left(\sup_{0\leq t<S} X_t < r\right) = 1$.

(c): If $p(l+) = -\infty, p(r-) < +\infty$ then $\mathbb{P}(\lim_{t\nearrow S} X_t = r) = \mathbb{P}(\inf_{0\leq t<S} X_t > l) = 1$.

(d): If $p(l+) > -\infty, p(r-) < +\infty$ then $\mathbb{P}\left(\lim_{t \nearrow S} X_t = l\right) = 1 - \mathbb{P}\left(\lim_{t \nearrow S} X_t = r\right) = \frac{p(r-)-p(x)}{p(r-)-p(l+)}$.

Theorem 14. ***(Feller's Test for Explosions)*** *With the same setting as the theorem above, $\mathbb{P}(S = \infty) = 1$ or $\mathbb{P}(S = \infty) < 1$, according to whether $v(l+) = v(r-) = \infty$ or not where*

$$v(x) = \int_c^x p'(y) \int_c^y \frac{2}{p'(z)\sigma^2(z)}\,dz\,dy \tag{544}$$

for some fixed $c \in (l, r)$.

Now we use the example above for N to illustrate the usage of these theorems.

Lemma 25. *Consider process N defined by*

$$N(t) = \frac{1}{2} + \int_0^t N(s)(1 - N(s))\,dW(s) \tag{545}$$

then N always lives in $(0, 1)$.

Proof. Here $N_0 = \frac{1}{2} \in (0, 1)$ is deterministic and $dN_t = N_t(1 - N_t)\,dW_t$ with $b(x) = 0, \sigma(x) = x(1 - x) > 0$ on $(0, 1)$. Take $l = 0, r = 1$, compute $p(x)$ to find

$$p(x) = \int_c^x d\xi = x - c \tag{546}$$

then compute $v(x)$ to find

$$v(x) = \int_c^x \int_c^y \frac{2}{z^2(1-z)^2}\,dz\,dy = 2\int_c^x 2\log\frac{y}{1-y} - \frac{2y-1}{y(y-1)} - C\,dy \tag{547}$$

so obviously $v(0+) = v(1-) = \infty$, by Feller's test for explosion, $\mathbb{P}(S = \infty) = 1$ so the solution will always live in $(0, 1)$. $\square$

Remark. *For the SDE*

$$\forall 1 \le i \le n, \mu_i(t) = \mu_i(0)\mathcal{E}\left(\int_0^\cdot (e_i - \mu(s))^T\,dM(s)\right)(t) \tag{548}$$

that works as the characterization of market portfolio in a perfectly balanced market determined by M, a more general version of Feller's test for explosion can be applied to prove that the solution always live in $\Delta_{(n-1)}$ if one starts from $\mu(0) \in \Delta_{(n-1)}$, the space of all long-only stock portfolios. As a result, the solution to this SDE is always a legal market portfolio.

CAPM (Capital Asset Pricing Model)

A market is called **CAPM** if

$$\forall 1 \leq i \leq n, R_i(t) = \int_0^t \beta_i(s)\, dR_\mu(s) + N_i(t) \tag{549}$$

for predictable β_i and continuous local MG N_i starting from 0 such that the orthogonality property $\langle N_i, R_\mu \rangle = 0$ holds. In other words, **in a CAPM market the cumulative return of each asset can be decomposed into the part that depends on the market portfolio cumulative return R_μ through β_i and the noise N_i orthogonal to the market cumulative return R_μ**. It's not hard to see that **in a CAPM market individual stocks cannot systematically outperform or underperform the market**, we will explain more about this point afterwards.

Theorem 15. *A market is CAPM iff the following two conditions hold:*

(A): There exists predictable leverage process b such that

$$\forall T \geq 0, \sum_{i=1}^n \int_0^T |b(s)| \mathbb{I}_{c_{\mu\mu}(s) > 0} |dC_{i\mu}(s)| < \infty \tag{550}$$

and $\alpha_i = bc_{i\mu}$ for $1 \leq i \leq n$ on $\{c_{\mu\mu} > 0\}$.

(B): On $\{c_{\mu\mu} = 0\}$, $\alpha_\mu = 0$ iff $\forall 1 \leq i \leq n, \alpha_i = 0$.

*So **condition (A), (B) is a characterization for a market being CAPM**.*

Proof. First assume that market is CAPM.

$$C_{i\mu}(t) = \langle R_i, R_\mu \rangle (t) = \int_0^t \beta_i(s)\, dC_{\mu\mu}(s) \tag{551}$$

by the decomposition of R_i w.r.t. R_μ. This implies that $\forall 1 \leq i \leq n, c_{i\mu} = \beta_i c_{\mu\mu}$. Since $R_i(t) - \int_0^t \beta_i(s)\, dR_\mu(s) = N_i(t)$ is local MG, take finite variation part on both sides to see

$$A_i(t) = \int_0^t \beta_i(s)\, dA_\mu(s) \tag{552}$$

so $\forall 1 \leq i \leq n, \alpha_i = \beta_i \alpha_\mu$.

On the event $\{c_{\mu\mu} = 0\}$, if all $\alpha_i = 0$ then obviously $\alpha_\mu = \mu^T \alpha = 0$. Conversely, if $\alpha_\mu = 0$ then $\alpha_i = \beta_i \alpha_\mu = 0$ so condition (B) holds.

On the event $\{c_{\mu\mu} = 0\}$, we would have the representation $\alpha_i = \frac{c_{i\mu}}{c_{\mu\mu}} \alpha_\mu$ so it's fine to set $b = \frac{\alpha_\mu}{c_{\mu\mu}}$ such that

$\alpha_i = bc_{i\mu}$. For the other half of condition (A),

$$\forall T \geq 0, \sum_{i=1}^{n} \int_0^T |b(s)|\mathbb{I}_{c_{\mu\mu}(s)>0}|dC_{i\mu}(s)| = \sum_{i=1}^{n} \int_0^T |b(s)|\mathbb{I}_{c_{\mu\mu}(s)>0}|c_{i\mu}(s)|dO(s) \tag{553}$$

$$= \sum_{i=1}^{n} \int_0^T |\alpha_i(s)|dO(s) = \sum_{i=1}^{n} \int_0^T |dA_i(s)| < \infty \tag{554}$$

since A_i has finite variation.

Conversely, assume that condition (A) and (B) holds, let's define $\beta_i = \frac{c_{i\mu}}{c_{\mu\mu}}\mathbb{I}_{c_{\mu\mu}>0} + \frac{\alpha_i}{\alpha_\mu}\mathbb{I}_{c_{\mu\mu}=0,\alpha_\mu\neq 0}$ and let's verify that such β_i provides the decomposition under CAPM. Let's split the integral below according to if the event $\{c_{\mu\mu} > 0\}$ happens

$$\int_0^t \beta_i(s)\, dR_\mu(s) = \int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)>0}\, dA(s) + \int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)=0}\, dA(s) + \int_0^t \beta_i(s)\mu^T(s)\, dM(s) \tag{555}$$

$$= \int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)>0}b(s)c(s)\mu(s)\, dO(s) + \int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)=0}\, dA(s) + \int_0^t \beta_i(s)\mu^T(s)\, dM(s) \tag{556}$$

$$= \int_0^t \beta_i(s)\mathbb{I}_{c_{\mu\mu}(s)>0}b(s)\, dC_{\mu\mu}(s) + \int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)=0}\, dA(s) + \int_0^t \beta_i(s)\mu^T(s)\, dM(s) \tag{557}$$

the first integral

$$\int_0^t \beta_i(s)\mathbb{I}_{c_{\mu\mu}(s)>0}b(s)\, dC_{\mu\mu}(s) = \int_0^t \frac{c_{i\mu}(s)}{c_{\mu\mu}(s)}\mathbb{I}_{c_{\mu\mu}(s)>0}b(s)\, dC_{\mu\mu}(s) \tag{558}$$

$$= \int_0^t \mathbb{I}_{c_{\mu\mu}(s)>0}b(s)\, dC_{i\mu}(s) \tag{559}$$

$$= \int_0^t \mathbb{I}_{c_{\mu\mu}(s)>0}\, dA_i(s) \tag{560}$$

and the second integral

$$\int_0^t \beta_i(s)\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)=0}\, dA(s) = \int_0^t \frac{\alpha_i(s)}{\alpha_\mu(s)}\mu^T(s)\mathbb{I}_{c_{\mu\mu}(s)=0,\alpha_\mu(s)\neq 0}\, dA(s) \tag{561}$$

$$= \int_0^t \mathbb{I}_{c_{\mu\mu}(s)=0}\, dA_i(s) \tag{562}$$

so the first two integrals sum up to $\int_0^t dA_i(s) = A_i(t)$, the finite variation part of $R_i(t)$. So we get

$$\int_0^t \beta_i(s)\, dR_\mu(s) = A_i(t) + \int_0^t \beta_i(s)\mu^T(s)\, dM(s) = R_i(t) - \int_0^t (e_i^T - \beta_i(s)\mu^T(s))\, dM(s) \tag{563}$$

naturally we set $N_i(t) = \int_0^t (e_i^T - \beta_i(s)\mu^T(s))\, dM(s)$ and we only need to prove that it's a continuous local MG orthogonal to R_μ. This can be seen by

$$\langle N_i, R_\mu \rangle (t) = \int_0^t \sum_{j,k=1}^n (e_i - \beta_i(s)\mu(s))_j c_{jk}(s)\mu_k(s)\, dO(s) = 0 \tag{564}$$

since $(e_i - \beta_i\mu)^T c\mu = c_{i\mu} - \beta_i c_{\mu\mu} = 0$ is always true. This concludes the proof. □

Remark. *In the proof one gets some "side product" that* ***if a market is CAPM then the leverage process has formula***

$$b = \frac{\alpha_\mu}{c_{\mu\mu}} \mathbb{I}_{c_{\mu\mu}>0} \tag{565}$$

and ***the beta of each asset has formula***

$$\beta_i = \frac{c_{i\mu}}{c_{\mu\mu}} \mathbb{I}_{c_{\mu\mu}>0} + \frac{\alpha_i}{\alpha_\mu} \mathbb{I}_{c_{\mu\mu}=0,\alpha_\mu\neq 0}, \ (\forall 1 \leq i \leq n) \tag{566}$$

under CAPM assumption and consider the non-degenerate case where $\{c_{\mu\mu} > 0\}$, we have $\beta_i = \frac{c_{i\mu}}{c_{\mu\mu}}$ as the ratio of covariation. However, if one consider the CAPM without time horizon (fix a time t), we see that

$$\beta_i = \frac{cov(R_i, R_\mu)}{Var(R_\mu)} \tag{567}$$

is a deterministic real number consistent with ***the beta of asset*** *i and it stands for the quantity of* ***systematic risk*** *of asset i. From its special structure, one immediately recognizes that it is the slope of the security market line (SML) regressing the excess return of asset i w.r.t. the excess return of market portfolio.*

For a market to really be CAPM, we are assuming that each asset cannot outperform or underperform the market systematically, equivalent to saying the SML shall pass the origin. The intercept of the SML is typicaally called ***the alpha of the asset*** *and a market is CAPM if the alpha of all assets are 0, which is often checked in practice by doing hypothesis testing on the intercept term. It's worth noting that CAPM has the meaning of* ***market equilibrium****, i.e. the supply and demand of all assets shall clear up on the market so one would expect to see that all investors are holding the market portfolio. That's why it starts from the assumption that market can never be systematically outperformed or underperformed (in the CAPM without time horizon, there is a nice derivation based on the perturbation of the portfolio adopting the similar idea).*

Notice that the equation in condition (A) can be written as $\alpha = bc\mu$ and recall that the super-MG numeraire portfolio ν satisfies $c\nu = \alpha$ if exists, so if super-MG numeraire portfolio exists in a CAPM market **(market is CAPM and viable)**, we have

$$\alpha = c\nu, \nu = b\mu \tag{568}$$

showing the **connection** between those two concepts. However notice that **there exists CAPM market that is not viable** since such ν may not satisfy the integrability condition (which we neglect in this notes but is important to verify). So far we see that b is called the leverage process because investing b portions of market portfolio (with leverage) beats all other portfolios on the market.

Lemma 26 (Exercise 2.72)**.** *Consider an original viable market with n stocks and l funds built upon those assets $V^1, ..., V^l$ such that $V^k = X_{\pi^k}$. The investors can only form portfolio ρ upon those funds $V^1, ..., V^l$.* ***The investor can achieve the maximal growth that is attainable in the original market iff the super-MG numeraire portfolio ν in the original market can be induced by some portfolio λ on the fund market****, i.e.*

$$\exists\lambda, \forall 1 \leq i \leq n, \nu_i = \sum_{k=1}^{l} \lambda^k \pi_i^k \tag{569}$$

such that holding portfolio λ in the fund market is equivalent to holding ν in the original market. Prove that this condition is equivalent to saying there exists β_i^k, N_i such that

$$\forall 1 \leq i \leq n, R_i(t) = \int_0^t \sum_{k=1}^{l} \beta_i^k(s)\, dR_{\pi^k}(s) + N_i(t) \tag{570}$$

where N_i is some continuous local MG starting from 0 and orthogonal to R_{π^k} for $\forall 1 \leq k \leq l$.

The idea and the procedure of the proof is basically the same as that for CAPM, just verify the characterization $\alpha_i = c_{i\nu}$ for super-MG numeraire portfolio.

Financing, Optimization, Maximality

Assume a viable market so $\mathscr{Y} \neq \emptyset$ where $\mathscr{Y}$ denotes the set of local MG deflators. Now we consider the problem of financing, i.e. **how much initial wealth one has to begin with in order to be able to finance a given capital withdrawal, or conversely, given certain amount of initial wealth, what kinds of capital withdrawal is financeable**. As a result, we allow the total wealth process $X(\cdot; x, \theta, K)$ to be non-negative (not necessarily a numeraire) and the investment strategy θ is considered instead of the portfolio π since π may not exists.

Optional Decomposition

By the definition of local MG deflator, $\forall Y \in \mathscr{Y}, \forall X \in \mathscr{X}, YX$ is non-negative local MG, thus a super-MG. Naturally, if we relax the constraint on such wealth process X from being a numeraire (strictly positive) to being non-negative, we would still expect to see that YX is an non-negative super-MG. The following lemma provides the rigorous argument.

Lemma 27 (Exercise 2.28). *Assume a viable market ($\mathscr{Y} \neq \emptyset$) and fix a cumulative capital withdrawal $K \in \mathscr{K}$, if K can be financed with initial wealth $x \geq 0$, there exists $\theta, F \in \mathscr{K}$ such that $\forall t \geq 0, X(t) \equiv X(t; x, \theta, F) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) - F(t) \geq K(t)$. Then $\forall Y \in \mathscr{Y}, Y(X-K) + \int_0^\cdot Y(t)(dK(t) + dF(t))$ is non-negative local MG. In particular, $Y(X-K), Y(X-K) + \int_0^\cdot Y(t)dK(t)$ are non-negative super-MG.*

Proof. Consider

$$Z(t) = Y(t)(X(t) - K(t)) + \int_0^t Y(s)(dK(s) + dF(s)) \tag{571}$$

it's obvious to be non-negative and by Ito formula

$$\begin{aligned} dZ(t) &= X(t)\, dY(t) - K(t)\, dY(t) + Y(t)\, dX(t) + Y(t)\, dF(t) + d\langle Y, X-K\rangle(t) && (572)\\ &= \sum_{i=1}^n \theta_i(t)\, d(S_iY)(t) + \left(X(t) - K(t) - \sum_{i=1}^n \theta_i(t)S_i(t)\right) dY(t) && (573) \end{aligned}$$

by plugging in the dynamics of X. Since the local MG deflator Y itself must be a local MG and that S_i is numeraire so S_iY is local MG, we have proved that Z is local MG. One step further, Z is also a super-MG.

Set $F \equiv 0$ to see that $Y(t)(X(t) - K(t)) + \int_0^t Y(s)dK(s)$ is a non-negative super-MG. Since $\int_0^t Y(s)dK(s)$ is increasing, $Y(X-K)$ must also be an non-negative super-MG. □

Remark. *Obviously, $K \equiv 0$ is always financeable for $\forall x \geq 0$. In this case, $F \equiv 0$, plugging into the lemma above gives that **for any non-negative wealth process X and any local MG deflator Y, YX is always an non-negative super-MG**.*

Optional decomposition tells us that the converse is also true. Whenever X is an non-negative process such that YX is a super-MG for $\forall Y \in \mathscr{Y}$, such X corresponds to a wealth process in the market with asset prices S_i and some capital withdrawal K.

Theorem 16. ***(Optional Decomposition)*** *In a viable market for non-negative process X with $X(0) = x \geq 0$, the following statements are equivalent:*

(1): $\forall Y \in \mathscr{Y}$, YX is super-MG.

(2): There exists $\theta, K \in \mathscr{K}$ such that

$$X(t) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) - K(t) \tag{574}$$

Remark. *The theorem has its name from the fact that generally K is adapted with right-continuous sample path so it's **optional** but not necessarily predictable. Notice that we are putting no assumptions on the path regularity of X so the path can be discontinuous, bringing with much more complexity.*

The proof of this theorem is too technical so we skip it here. However, it's easy to see the uniqueness of the optional decomposition.

If any wealth process X (even does not have to be non-negative) can be decomposed in two ways

$$X(t) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) - K(t) \tag{575}$$

$$X(t) = \tilde{x} + \int_0^t \sum_{i=1}^n \tilde{\theta}_i(s)\, dS_i(s) - \tilde{K}(t) \tag{576}$$

then $x = \tilde{x}$ obviously, with

$$\tilde{K}(t) - K(t) = \int_0^t \sum_{i=1}^n [\tilde{\theta}_i(s) - \theta_i(s)]\, dS_i(s) \tag{577}$$

the LHS being finite variation and the RHS having continuous sample path. As a result, set $D \overset{def}{=} \tilde{K} - K$. From market viability, super-MG numeraire X_ν exists so $Y = \frac{1}{X_\nu} \in \mathscr{Y}$ is a local MG deflator. The integration by parts gives $\int_0^t Y(s)\, D(s) = Y(t)D(t) - Y(0)D(0) - \int_0^t D(s)\, dY(s) = Y(t)D(t) - \int_0^t D(s)\, dY(s)$ since $D(0) = 0$ has finite variation.

Now that $\int_0^t D(s)\, dY(s)$ is a continuous local MG. On the other hand, YD is also a continuous local MG. This is due to the lemma below that since

$$D(t) = \int_0^t \sum_{i=1}^n [\tilde{\theta}_i(s) - \theta_i(s)]\, dS_i(s) \tag{578}$$

has such representation as a wealth process in the market with no capital withdrawal, $\forall Y \in \mathscr{Y}$, YD is a local MG.

Now that $\int_0^t Y(s)\, D(s)$ is a continuous local MG with finite variation, so $\forall t \geq 0, \int_0^t Y(s)\, D(s) = 0$. Since $Y > 0$ we know D must be constant. Since $D(0) = 0$, $D \equiv 0$ proves $K \equiv \tilde{K}$ and this results in $\theta \equiv \tilde{\theta}$. This is a classical application of the fact that a continuous local MG with finite variation must be trivial in the proof of uniqueness argument.

In the context above we investigated the case where $X(0) = x \geq 0$ as an non-negative process such that $\forall Y \in \mathscr{Y}$,

YX is a super-MG. One would naturally ask if this ensures that such YX is also a local MG. The equivalent conditions of such YX being a local MG are given by the following lemma.

Lemma 28 (Exercise 3.4)**.** *In a viable market, X is non-negative process with $X(0) = x \geq 0$, the following statements are equivalent:*

(1): $\forall Y \in \mathscr{Y}$, YX is local MG.

(2): $\forall Y \in \mathscr{Y}$, YX is super-MG and $\exists Y_0 \in \mathscr{Y}$, $Y_0 X$ is local MG.

(3): There exists θ such that

$$X(t) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) \tag{579}$$

Proof. For simplicity, we only prove for the case where all processes have continuous sample path.

(1) implies (2) is trivial.

(2) implies (3): by optional decomposition theorem, there exists $\theta, K \in \mathscr{K}$ such that

$$X(t) = x + \int_0^t \sum_{i=1}^n \theta_i(s)\, dS_i(s) - K(t) \tag{580}$$

apply Ito formula to see

$$d(Y_0 X)(t) = Y_0(t)\, dX(t) + X(t)\, dY_0(t) + d\left\langle X, Y_0 \right\rangle(t) \tag{581}$$

$$= \sum_{i=1}^n \theta_i(t)\, d(S_i Y_0)(t) + \left(X(t) - \sum_{i=1}^n \theta_i(t) S_i(t)\right) dY_0(t) - Y_0(t)\, dK(t) \tag{582}$$

since $Y_0 X$ is a local MG, $S_i Y_0$ and Y_0 are also local MG, we must have $\int_0^t Y_0(s)\, dK(s)$ as a continuous local MG with finite variation so it's constant. Since $K(0) = 0$, we must have $K \equiv 0$ gives the representation in (3).

(3) implies (1): By Ito formula,

$$d(YX)(t) = Y(t)\, dX(t) + X(t)\, dY(t) + d\left\langle X, Y \right\rangle(t) \tag{583}$$

$$= \sum_{i=1}^n \theta_i(t)\, d(S_i Y)(t) + \left(X(t) - \sum_{i=1}^n \theta_i(t) S_i(t)\right) dY(t) \tag{584}$$

with $S_i Y$ and Y to be local MG, so YX is local MG. □

Remark. *If $\forall Y \in \mathscr{Y}$, YX is local MG, then in the optional decomposition of X no capital withdrawal is allowed. This is intuitive since the non-negative increasing K contributes the finite variation part and it's the reason why YX is a super-MG. An important fact this tells us it that* ***for any wealth process X (may be negative) generated by investment strategy θ and initial wealth $x \in \mathbb{R}$ (but no capital withdrawal), YX is always local MG for any local MG deflator Y.***

Financing Duality

The following theorem provides answer to the financing question that for a given capital withdrawal how much initial wealth does one need in order to make it financeable.

Theorem 17. ***(Financing Duality)*** *In a viable market, for any given $K \in \mathscr{K}$, the financing capital has the equation that*

$$x(K) = \sup_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}} \int_0^\infty Y(t)\, dK(t) \tag{585}$$

so when $x(K) < \infty$, K is financeable starting from initial wealth $x(K)$. Moreover, the inf in the definition of $x(K)$ is ***attained*** *by some investment strategy θ_K and capital withdrawal $F_K \in \mathscr{K}$ that depends on K such that $\forall t \geq 0, X(t; x(K), \theta_K, F_K) \geq K(t)$.*

If the inf in $x(K)$ can be attained, $x(K)$ corresponds to the exactly (not approximate) minimum initial wealth required to finance K. Financing duality tells us that this happens when the market is viable and $x(K) < \infty$.

Remark. *Recall the notation $\mathscr{K}(x)$ as the set of all capital withdrawal processes that is financeable with initial wealth x. By the definition of $\mathscr{K}(x)$ and the representation of $x(K)$ above,*

$$\mathscr{K}(x) = \left\{ K \in \mathscr{K} : \sup_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}} \int_0^\infty Y(t)\, dK(t) \leq x \right\} \tag{586}$$

gives ***an important characterization of*** $\mathscr{K}(x)$.

It's natural to think about the **minimal financing problem**, i.e. a wealth process X minimally finances K if it finances K and for any wealth process $\overline{X}$ that finances K, $X \leq \overline{X}$ always holds. In other words, X denotes the lowest possible wealth such that K is financeable. If such minimal financing wealth process X exists, it is unique and it must start from initial wealth $x(K) < \infty$. By the theorem above, when $x(K) < \infty$ there exists $\theta_K, F_K \in \mathscr{K}$ such that $X(\cdot; x(K), \theta_K, F_K)$ is the minimal financing wealth process. As a result, **the minimal financing wealth process for $K \in \mathscr{K}$ exists iff** $x(K) < \infty$.

However, such F_K is not necessarily constantly zero. In such cases, the minimal financing wealth process X for capital withdrawal K can also finance capital withdrawal $K + F_K$, but the minimal financing wealth process for those two different capital withdrawal streams are the same so $x(K) = x(K + F_K)$. This shows that the existence of such non-zero F_K destroys the maximality of K in the sense that its minimal financing wealth process is actually able to finance a larger capital withdrawal stream. Oppositely, $F_K \equiv 0$ is equivalent to **the maximality of K in** $\mathscr{K}(x(K))$, i.e. K is the largest capital withdrawal such that it's financeable with the initial capital as that in the minimal financing of K.

Actually a more general representation for the minimal financing wealth process is given as the following which we will turn back to later.

Theorem 18. *In a viable market, for any finite stopping time σ and minimal financing wealth process X w.r.t. $K \in \mathscr{K}$,*

$$X(\sigma) = K(\sigma) + \operatorname*{ess\,sup}_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}} \left[\int_\sigma^\infty \frac{Y(t)}{Y(\sigma)}\, dK(t) \middle| \mathscr{F}(\sigma) \right] \tag{587}$$

Although we are skipping the technical proof of the financing duality theorem and leaving the proof of the theorem above to later, we may see that one side of the proof is not hard to complete.

Lemma 29 (Exercise 2.28)**.** *Assume market is viable and fix $K \in \mathscr{K}$, if K can be financed with initial capital $x \geq 0$ such that $X \equiv X(\cdot; x, \theta, F) \geq K$ for some investment strategy θ and some $F \in \mathscr{K}$, then*

$$X(\sigma) \geq K(\sigma) + \mathbb{E}_{\mathbb{P}}\left[\int_{\sigma}^{\infty} \frac{Y(t)}{Y(\sigma)}\, dK(t) \Big| \mathscr{F}(\sigma)\right] \tag{588}$$

holds for any $\mathbb{P}-a.e.$ finite stopping time σ w.r.t. filtration $\mathscr{F}(\cdot)$ and $\forall Y \in \mathscr{Y}$. We further deduce that

$$x(K) \geq \sup_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}} \int_0^{\infty} Y(t)\, dK(t) \tag{589}$$

Proof. Recall from the lemma in the last section that we have proved $\forall Y \in \mathscr{Y}$,

$$Z(t) = Y(t)(X(t) - K(t)) + \int_0^t Y(s)(\,dK(s) + \,dF(s)) \tag{590}$$

is non-negative local MG so it's a super-MG. By optional stopping theorem, $Z(\sigma) \geq \mathbb{E}[Z(\infty)|\mathscr{F}(\sigma)]$ with $Z(\infty)$ well-defined by MG convergence theorem. This gives

$$Y(\sigma)(X(\sigma) - K(\sigma)) \geq \mathbb{E}[Y(\infty)(X(\infty) - K(\infty))|\mathscr{F}(\sigma)] + \mathbb{E}\left[\int_{\sigma}^{\infty} Y(s)(\,dK(s) + \,dF(s)) \Big| \mathscr{F}(\sigma)\right] \tag{591}$$

take $F \equiv 0$ to see

$$X(\sigma) - K(\sigma) \geq \frac{1}{Y(\sigma)} \mathbb{E}\left[Y(\infty)X(\infty) - Y(\infty)K(\infty) \Big| \mathscr{F}(\sigma)\right] + \mathbb{E}\left[\int_{\sigma}^{\infty} \frac{Y(s)}{Y(\sigma)}\, dK(s) \Big| \mathscr{F}(\sigma)\right] \tag{592}$$

noticing $X - K \geq 0, Y > 0$ concludes the proof. Set $\sigma = 0$ and take initial wealth $X(0) = x > x(K)$ gives

$$\forall Y \in \mathscr{Y}, \forall x > x(K), x \geq \mathbb{E} \int_0^{\infty} Y(t)\, dK(t) \tag{593}$$

take inf on both sides w.r.t. such x and sup on both sides w.r.t. $Y \in \mathscr{Y}$ proves the lower bound of $x(K)$. $\square$

Remark. *This is exactly the one-sided version of the two theorems above and it's much easier to prove.*

*Intuitively, one may think of the **local MG deflator as a "reasonable" discounting** process such that the auxiliary market with asset price $S_i^Y = YS_i$ only has the noise part (local MG). When the market is viable, $\frac{1}{X_\nu}$ works as local MG deflator for super-MG numeraire X_ν.*

To explain what it means for a discounting process to be reasonable, we use a simple example. If the continuous-time risk-free rate is r, and one has cash flow 1 at time t, it's immediate that the present value of this cash flow is e^{-rt}. To look into this example, we can think about the money market asset price as $S_0(t)$ with $S_0(0) = 1$ whose *dynamics is given by*

$$\frac{1}{S_0(t)}\,dS_0(t) = r\,dt \tag{594}$$

so $S_0(t) = e^{rt}$ *by taking* $Y(t) = e^{-rt}$*, we make sure that* $YS_0 \equiv 1$ *is a local MG so in the auxiliary market with asset price* $S_0^Y = YS_0 \equiv 1$*, no finite variation "tendency" is exhibited in the asset price. As a result, a "reasonable" discounting process shall make sure that it* ***eliminates the finite variation "tendency" part in all asset price dynamics****, which is just the motivation of defining local MG deflators.*

Remark. *To interpret the integral, notice that* $Y(t_i)[K(t_{i+1}) - K(t_i)]$ *contributes to* $\int_0^T Y(t)\,dK(t)$ *when there is new capital withdrawal happening in the time period and the contribution is larger if* Y *takes a larger value at that time. As a result,* $\mathbb{E}\int_0^\infty Y(t)\,dK(t)$ *stands for* ***the expected total amount of capital withdrawal in the whole time horizon discounted w.r.t.*** Y*.*

The theorems and lemmas above can thus be interpreted as the fact that ***the minimum initial wealth has to be no less than the expected present value of capital withdrawal under all possible "reasonable" discounting****, making it natural to understand.*

Model-consistent Probability

$\mathbb{Q}$ is called **model-consistent probability measure** on the same filtered measurable space as $\mathbb{P}$ if $\mathbb{P} \sim \mathbb{Q}$ and M_i are still local MG under $\mathbb{Q}$. Let Π denote the set of all model-consistent probability measures. It's called model-consistent because $\mathbb{P}$ and $\mathbb{Q}$ are not too different (equivalent) and the semi-MG decomposition of R_i remains the same so asset prices still follow the same dynamics. It's easy to see that if X_ν is super-MG numeraire under $\mathbb{P}$, it remains the super-MG numeraire under $\mathbb{Q}$ since the equivalent condition $\forall 1 \leq i \leq n, A_i = C_{i\nu}$ is still satisfied. When $\Pi = \{\mathbb{P}\}$, i.e. $\mathbb{P}$ is the only model-consistent probability measure, it's the case of **market completeness**.

The following lemma shows the connection between local MG deflators under $\mathbb{P}$ and $\mathbb{Q}$.

Lemma 30 (Exercise 3.13)**.** *$\forall \mathbb{Q} \in \Pi$, consider super-MG numeraire X_ν under $\mathbb{P}$ and define*

$$Y^{\mathbb{Q}} \overset{def}{=} \frac{1}{X_\nu} Z^{\mathbb{Q}}, Z^{\mathbb{Q}}(\cdot) \overset{def}{=} \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}(\cdot)} \tag{595}$$

so $Z^{\mathbb{Q}}$ is the density process of $\mathbb{Q}$ w.r.t. $\mathbb{P}$. Show that $Y^{\mathbb{Q}}$ is a local MG deflator under $\mathbb{Q}$.

Proof. We just need to show that $Z^{\mathbb{Q}} = \mathcal{E}(L)$ for some $L \in \mathscr{M}_{\text{loc}}^{\perp}(M)$ where $\mathscr{M}_{\text{loc}}^{\perp}(M)$ is the set of all local MG L starting from 0 such that $\forall 1 \leq i \leq n, \langle L, M_i \rangle = 0$ (all local MG starting from 0 in the orthogonal complement of M). Notice that since $\mathbb{Q}$ is a model-consistent probability measure, M_i does not change, that's why the space $\mathscr{M}_{\text{loc}}^{\perp}(M)$ stays the same. Refer to the lemma below for why this argument suffices.

Let's set $L = \mathcal{L}\left(Z^{\mathbb{Q}}\right)$ so it's a local MG starting from 0 since $Z^{\mathbb{Q}}$ is a MG with $\mathbb{E}_{\mathbb{P}} Z^{\mathbb{Q}}(t) = 1$. By Girsanov theorem (it holds on the infinite time horizon since $\mathbb{E} Z^{\mathbb{Q}}(\infty) = 1$), since M_i is continuous local MG under $\mathbb{P}$, $M_i - \langle M_i, L \rangle$ is local MG under $\mathbb{Q}$. However, M_i is still continuous local MG under $\mathbb{Q}$ (model-consistent) so $\langle M_i, L \rangle$ is continuous local MG under $\mathbb{Q}$ with finite variation. So $\langle L, M_i \rangle = 0$, $L = \mathcal{L}\left(Z^{\mathbb{Q}}\right) \in \mathscr{M}_{\text{loc}}^{\perp}(M)$ and $Y^{\mathbb{Q}}$ is a local MG deflator under $\mathbb{Q}$. $\square$

Remark. *We denote*

$$\mathscr{Y}_\Pi \overset{def}{=} \left\{Y^{\mathbb{Q}} : \mathbb{Q} \in \Pi\right\} \tag{596}$$

*as the set of all local MG deflators under $\mathbb{Q}$ of the form above. It's obvious that $Y^{\mathbb{Q}}(t) X_\nu(t)$ is non-negative local MG thus a super-MG but $Y^{\mathbb{Q}}(\infty) X_\nu(\infty)$ **is strictly positive (since $\mathbb{P} \sim \mathbb{Q}$) with expectation 1** (since $Y^{\mathbb{Q}} X_\nu = Z^{\mathbb{Q}}$ is MG), that's why this family of local MG deflators is interesting.*

Lemma 31 (Exercise 2.27)**.** *M_i are continuous local MG parts of R_i as previously defined. If super-MG numeraire X_ν exists, then*

$$Y = \frac{1}{X_\nu} \mathcal{E}(L) = \mathcal{E}(L - M_\nu) \in \mathscr{Y} \tag{597}$$

for $\forall L \in \mathscr{M}_{loc}^{\perp}(M)$ so such Y is a local MG deflator (we simplify the setting such that all processes have continuous sample path). In fact, this gives the representation of all local MG deflators in $\mathscr{Y}$ so it characterizes $\mathscr{Y}$.

Proof. For any portfolio π, we have

$$Y X_\pi = \frac{X_\pi}{X_\nu} \mathcal{E}(L) = \mathcal{E}(R_\pi^\nu) \mathcal{E}(L) = \mathcal{E}(R_\pi^\nu + L + \langle R_\pi^\nu, L \rangle) \tag{598}$$

by Yor formula. To prove that YX_π is local MG, it suffices to prove $R_\pi^\nu + L + \langle R_\pi^\nu, L\rangle$ is local MG where $R_\pi^\nu = R_{\pi-\nu} - C_{\pi-\nu,\nu}$. Some simple calculation tells us

$$\langle R_\pi^\nu, L\rangle (t) = \langle R_{\pi-\nu}, L\rangle (t) = \int_0^t \sum_{i=1}^n [\pi_i(s) - \nu_i(s)]\, d\langle R_i, L\rangle (s) = 0 \tag{599}$$

by the orthogonality of L with M_i while on the other hand

$$R_\pi^\nu + L = R_{\pi-\nu} - C_{\pi-\nu,\nu} + L = (A_{\pi-\nu} - C_{\pi-\nu,\nu}) + (L + M_{\pi-\nu}) \tag{600}$$

is local MG. This is because $L + M_{\pi-\nu}$ is local MG and for super-MG numeraire portfolio ν we have $\forall 1 \le i \le n, A_i = C_{i\nu}$ so $A_{\pi-\nu} = C_{\pi-\nu,\nu}$. This concludes the proof that such Y is a local MG deflator.

Let's then verify the other representation of Y that

$$\begin{aligned} X_\nu \mathcal{E}(L - M_\nu) &= \mathcal{E}(R_\nu)\mathcal{E}(L - M_\nu) && (601)\\ &= \mathcal{E}(R_\nu + L - M_\nu + \langle R_\nu, L - M_\nu\rangle) && (602)\\ &= \mathcal{E}(A_\nu + L + \langle M_\nu, L - M_\nu\rangle) && (603)\\ &= \mathcal{E}(A_\nu + L - C_{\nu\nu}) && (604)\\ &= \mathcal{E}(L) && (605) \end{aligned}$$

since $A_\nu = C_{\nu\nu}$ by super-MG numeraire portfolio property.

At last, we prove that all local MG deflators has this representation. $\forall Y \in \mathscr{Y}$, YX_ν must be a local MG so let's define $L = \mathcal{L}(YX_\nu)$ such that L is a local MG starting from zero. Let's verify that $L \in \mathscr{M}_{\text{loc}}^\perp(M)$. Apply the Yor formula for stochastic logarithm to get

$$\begin{aligned} L &= \mathcal{L}(Y) + \mathcal{L}(X_\nu) + \langle \mathcal{L}(Y), \mathcal{L}(X_\nu)\rangle && (606)\\ &= \mathcal{L}(Y) + R_\nu + \langle \mathcal{L}(Y), M_\nu\rangle && (607) \end{aligned}$$

then it's clear that

$$\begin{aligned} \forall 1 \le i \le n, \langle L, M_i\rangle &= \langle \mathcal{L}(Y) + M_\nu, M_i\rangle && (608)\\ &= C_{i\nu} + \langle \mathcal{L}(Y), R_i\rangle && (609)\\ &= C_{i\nu} + \mathcal{L}(YS_i) - \mathcal{L}(Y) - \mathcal{L}(S_i) && (610)\\ &= (C_{i\nu} - A_i) + \mathcal{L}(YS_i) - \mathcal{L}(Y) - M_i && (611)\\ &= \mathcal{L}(YS_i) - \mathcal{L}(Y) - M_i && (612) \end{aligned}$$

with Yor formula applied once more for $R_i = \mathcal{L}(S_i)$ and use $A_i = C_{i\nu}$ for super-MG numeraire portfolio ν. This shows us that $\langle L, M_i\rangle$ a continuous local MG with finite variation, so it's constantly zero, the orthogonality holds. $\square$

Remark. *We have verified previously that $\frac{1}{X_\nu}$ is a local MG deflator if super-MG numeraire X_ν exists. Actually,* ***the characterization of*** $\mathscr{Y}$ *above shows us that all local MG deflators are not too far away from $\frac{1}{X_\nu}$ induced by the super-MG numeraire. To be specific, they are only different up to $\mathcal{E}(L)$ for any $L \in \mathscr{M}_{loc}^{\perp}(M)$.*

Let's now come back to the model-consistent probability measure and introduce another way to prove financing duality. Let's introduce the notations

$$K^\nu(t) \overset{def}{=} \int_0^t \frac{1}{X_\nu(s)}\, dK(s) \tag{613}$$

as the cumulative capital withdrawal discounted w.r.t. $Y = \frac{1}{X_\nu} \in \mathscr{Y}$. As we have proved in the last section, in a viable market a lower bound of $x(K)$ is easy to derive as

$$x(K) \geq \sup_{Y \in \mathscr{Y}(\mathbb{P})} \mathbb{E}_\mathbb{P} \int_0^\infty Y(t)\, dK(t) \geq \mathbb{E}_\mathbb{P} \int_0^\infty \frac{1}{X_\nu(t)}\, dK(t) = \mathbb{E}_\mathbb{P} K^\nu(\infty) \tag{614}$$

now consider model-consistent probability measure $\mathbb{Q} \in \Pi$, since it does not change the model, it does not change $x(K)$ and the super-MG numeraire portfolio is also retained. As a result, we think about the lower bound under $\mathbb{Q}$

$$x(K) \geq \sup_{Y \in \mathscr{Y}(\mathbb{Q})} \mathbb{E}_\mathbb{Q} \int_0^\infty Y(t)\, dK(t) \geq \mathbb{E}_\mathbb{Q} \int_0^\infty \frac{1}{X_\nu(t)}\, dK(t) = \mathbb{E}_\mathbb{Q} K^\nu(\infty) \tag{615}$$

the main point to understand here is that **model-consistent probability measure keeps $x(K)$ and ν but results in different lower bounds**. Naturally, one shall let $\mathbb{Q}$ run through all probability measures in Π to see

$$x(K) \geq \sup_{\mathbb{Q} \in \Pi} \mathbb{E}_\mathbb{Q} K^\nu(\infty) \tag{616}$$

a very similar bound to what we have seen from the perspective of local MG deflator! Since the lower bound of local MG deflator is actually tight under market viability, we would also expect this bound to be tight, giving another version of financing duality.

Theorem 19. ***(Financing Duality from model-consistent probability)*** *In a viable market, for $\forall K \in \mathscr{K}$,*

$$x(K) = \sup_{\mathbb{Q} \in \Pi} \mathbb{E}_\mathbb{Q} K^\nu(\infty) \tag{617}$$

Proof. We have already proved the RHS as the lower bound of $x(K)$. Conversely, $\forall Y \in \mathscr{Y}$, by the characterization of $\mathscr{Y}$, the representation holds that

$$\exists L \in \mathscr{M}_{\text{loc}}^{\perp}(M), Y = \frac{1}{X_\nu} \mathcal{E}(L) \tag{618}$$

denote $Z = \mathcal{E}(L)$ as a local MG starting from 1, there exists a sequence of stopping times $\tau_m \leq m$ as the localizing sequence of Z with probability measure $\mathbb{Q}_m$ induced by

$$\frac{d\mathbb{Q}_m}{d\mathbb{P}} = Z(\tau_m) \tag{619}$$

verify that $\mathbb{Q}_m \in \Pi$. Since Z is strictly positive, two probability measures are equivalent. To see the model consistency, by Girsanov theorem ($Z(\tau_m)$ is U.I. MG), $M_i - \langle L, M_i\rangle$ is still a local MG under $\mathbb{Q}_m$. Combined with the orthogonality $\langle L, M_i\rangle = 0$ we see that M_i is local MG under $\mathbb{Q}_m$.

Now denote $Z^{\mathbb{Q}_m}(t) = Z(t\wedge\tau_m) = \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}(t)}$ and apply the lemma below to see

$$\mathbb{E}_{\mathbb{Q}_m} K^\nu(\infty) = \mathbb{E}_{\mathbb{P}} \int_0^\infty Z^{\mathbb{Q}_m}(t)\, dK^\nu(t) \tag{620}$$

where the RHS is actually

$$\begin{aligned}\mathbb{E}_{\mathbb{P}} \int_0^\infty Z^{\mathbb{Q}_m}(t)\, dK^\nu(t) &= \mathbb{E}_{\mathbb{P}} \int_0^\infty X_\nu(t\wedge\tau_m) Y(t\wedge\tau_m)\, dK^\nu(t) &(621)\\ &= \mathbb{E}_{\mathbb{P}} \int_0^{\tau_m} X_\nu(t) Y(t)\, dK^\nu(t) &(622)\\ &= \mathbb{E}_{\mathbb{P}} \int_0^{\tau_m} Y(t)\, dK(t) &(623)\end{aligned}$$

so we conclude

$$\forall Y\in\mathscr{Y}, \sup_{\mathbb{Q}\in\Pi} \mathbb{E}_{\mathbb{Q}} K^\nu(\infty) \ge \sup_m \mathbb{E}_{\mathbb{Q}_m} K^\nu(\infty) = \sup_m \mathbb{E}_{\mathbb{P}} \int_0^{\tau_m} Y(t)\, dK(t) = \mathbb{E}_{\mathbb{P}} \int_0^\infty Y(t)\, dK(t) \tag{624}$$

by monotone convergence theorem for $\tau_m \nearrow \infty$ $(m\to\infty)$. Derive $\sup_{\mathbb{Q}\in\Pi} \mathbb{E}_{\mathbb{Q}} K^\nu(\infty) \ge \sup_{Y\in\mathscr{Y}} \mathbb{E}_{\mathbb{P}} \int_0^\infty Y(t)\, dK(t) = x(K)$ by the financing duality theorem for local MG deflator. □

Lemma 32. *$\forall \mathbb{Q} << \mathbb{P}$ as probability measures on the same filtered measurable space with $Z^{\mathbb{Q}}(\cdot) = \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}(\cdot)}$, then*

$$\forall K\in\mathscr{K}, \mathbb{E}_{\mathbb{Q}} K(\infty) = \mathbb{E}_{\mathbb{P}} \int_0^\infty Z^{\mathbb{Q}}(t)\, dK(t) \tag{625}$$

Proof. Integration by parts gives (K is of finite variation)

$$\begin{aligned}\int_0^\infty Z^{\mathbb{Q}}(t)\, dK(t) &= Z^{\mathbb{Q}} K\Big|_0^\infty - \int_0^\infty K(t)\, dZ^{\mathbb{Q}}(t) &(626)\\ &= Z^{\mathbb{Q}}(\infty)K(\infty) - \int_0^\infty K(t)\, dZ^{\mathbb{Q}}(t) &(627)\end{aligned}$$

taking expectation under $\mathbb{P}$ gives

$$\mathbb{E}_{\mathbb{P}} \int_0^\infty Z^{\mathbb{Q}}(t)\, dK(t) = \mathbb{E}_{\mathbb{Q}} K(\infty) - \mathbb{E}_{\mathbb{P}} \int_0^\infty K(t)\, dZ^{\mathbb{Q}}(t) \tag{628}$$

where the last term has value zero by a localization argument and we omit the details here. □

Remark. *We see in the proof of the financing duality theorem that* ***local MG deflators and model-consistent***

probability are two sides of the same coin. *They are closely connected through the characterization of* $\mathscr{Y}$. *Actually,* ***the difference between*** $Y \in \mathscr{Y}$ ***and*** $\frac{1}{X_\nu}$ ***which is*** $\mathcal{E}(L), L \in \mathscr{M}_{loc}^{\perp}(M)$***, corresponds to the difference between*** $\mathbb{P}$ ***and*** $\mathbb{Q} \in \Pi$***, with the Radon-Nikodym derivative given by the localized version of*** $\mathcal{E}(L)$.

As a result, such $\mathcal{E}(L)$ *can be interpreted as the* ***nonhedgeable source of uncertainty*** *since* L *is a local MG orthogonal with all* M_i *and* $\mathcal{E}(L)$ *explains the difference between model-consistent probability measures* $\mathbb{P}$ *and* $\mathbb{Q} \in \Pi$.

Example: Running Sup and Inf of the Super-MG Numeraire

Consider

$$X_\nu^*(t) \overset{def}{=} \sup_{0\le s\le t} X_\nu(s) \tag{629}$$

as the running sup with F nondecreasing and right-continuous with $F(1) = 0$ so that the capital withdrawal is formed as $K = F(X_\nu^*) \in \mathscr{K}$. In other words, the cumulative capital withdrawal until time t now only depends on the sup of the super-MG numeraire until time t. In this section we always assume $X_\nu(\infty) = \infty$ $\mathbb{P} - a.s.$ and $x(K) = x(F(X_\nu^*))$ is investigated. From financing duality,

$$x(K) = \sup_{Y\in\mathscr{Y}} \mathbb{E}_\mathbb{P} \int_0^\infty Y(t)\, dK(t) \tag{630}$$

$$\ge \mathbb{E}_\mathbb{P} \int_0^\infty \frac{1}{X_\nu(t)}\, dF(X_\nu^*(t)) \tag{631}$$

notice that within a time interval (s,t) there is contribution to this integral only when $X_\nu^*(u)$ is not constant for $u\in(s,t)$, i.e. $\forall u\in(s,t), X_\nu(u) = X_\nu^*(u)$ since running maximum stays the same iff X_ν is not breaking the past maximum. This clever trick results in

$$x(K) \ge \mathbb{E}_\mathbb{P} \int_0^\infty \frac{1}{X_\nu^*(t)}\, dF(X_\nu^*(t)) \tag{632}$$

$$= \mathbb{E}_\mathbb{P} \int_1^\infty \frac{1}{z}\, dF(z) \tag{633}$$

$$= \int_1^\infty \frac{1}{z}\, dF(z) \tag{634}$$

where change of variable $z = X_\nu^*(t)$ (notice that X_ν^* is increasing) and $X_\nu(\infty) = \infty$ is applied.

However, the equality actually holds for $x(K)$. This can be seen by constructing an investment strategy

$$\theta_i = \frac{f(X_\nu^*)X_\nu \nu_i}{S_i} \tag{635}$$

where $f(w) = \int_w^\infty \frac{1}{z}\, dF(z)$. Denote $X \equiv X(\cdot; f(1), \theta)$ as the total wealth process following this investment strategy with initial wealth $f(1)$, then

$$dX(t) = \sum_{i=1}^n \theta_i(t)\, dS_i(t) = \sum_{i=1}^n \frac{f(X_\nu^*(t))X_\nu(t)\nu_i(t)}{S_i(t)}\, dS_i(t) \tag{636}$$

$$= f(X_\nu^*(t))X_\nu(t) \sum_{i=1}^n \frac{\nu_i(t)}{S_i(t)}\, dS_i(t) \tag{637}$$

$$= f(X_\nu^*(t))X_\nu(t)\, dR_\nu(t) \tag{638}$$

$$= f(X_\nu^*(t))\, dX_\nu(t) \tag{639}$$

since $X_\nu = \mathcal{E}(R_\nu)$. Now that $X(t) = f(1) + \int_0^t f(X_\nu^*(s))\, dX_\nu(s)$, integration by parts combined with the same trick as above gives

$$X(t) = f(1) + \int_0^t f(X_\nu^*(s))\, dX_\nu(s) \tag{640}$$
$$= f(X_\nu^*(t))X_\nu(t) - \int_0^t X_\nu(s)\, df(X_\nu^*(s)) \tag{641}$$
$$= f(X_\nu^*(t))X_\nu(t) - \int_0^t X_\nu^*(s)\, df(X_\nu^*(s)) \tag{642}$$
$$= f(X_\nu^*(t))X_\nu(t) - \int_1^{X_\nu^*(t)} w\, df(w) \tag{643}$$

Now we want to compare $X(t)$ with the capital withdrawal $K = F(X_\nu^*)$ to see if K is financeable under such condition, consider

$$X - K = f(X_\nu^*)X_\nu - \int_1^{X_\nu^*} w\, df(w) - F(X_\nu^*) \tag{644}$$
$$= f(X_\nu^*)X_\nu + \int_1^{X_\nu^*} dF(w) - F(X_\nu^*) \tag{645}$$
$$= f(X_\nu^*)X_\nu \geq 0 \tag{646}$$

K is financeable, so

$$x(K) \leq f(1) \tag{647}$$

combined with the previous conclusion, we see

$$x(F(X_\nu^*)) = f(1) = \int_1^\infty \frac{1}{z}\, dF(z) \tag{648}$$

gives the **exact formula for $x(K)$ attainable**. Change the notation of investment strategy into portfolio to see that

$$\pi_i = \frac{\theta_i S_i}{X - K} = \nu_i \tag{649}$$

so **in order to finance this capital withdrawal $F(X_\nu^*)$ starting from minimal possible initial wealth $x(F(X_\nu^*))$, one always has to stick to the super-MG numeraire portfolio at all times**.

When it comes to the running inf of super-MG numeraire

$$I_\nu \overset{def}{=} \inf_{0\leq s\leq t} X_\nu(s) \tag{650}$$

we similarly fix H to be non-increasing continuous on $(0,1]$, $H(1)=0$ so $K=H(I_\nu)\in\mathscr{K}$ is considered. Notice that I_ν is bounded and decreasing and $H(0)$ is defined as the right limit of H at 0. The lower bound of $x(K)$ is similar to what we have done above

$$x(K)\geq \mathbb{E}_\mathbb{P}\int_0^\infty \frac{1}{I_\nu(t)}\,dH(I_\nu(t)) = \mathbb{E}_\mathbb{P}\int_1^{I_\nu(\infty)} \frac{1}{z}\,dH(z) \tag{651}$$

to get rid of $I_\nu(\infty)$, we can define a stopping time

$$\tau_z \overset{def}{=} \inf\{t\geq 0: I_\nu(t)=z\} \tag{652}$$

to rewrite the formula above as

$$x(K)\geq -\mathbb{E}_\mathbb{P}\int_0^1 \frac{\mathbb{I}_{\tau_z<\infty}}{z}\,dH(z) = -\int_0^1 \frac{\mathbb{P}(\tau_z<\infty)}{z}\,dH(z) \tag{653}$$

to calculate the probability involved, $\mathbb{P}(\tau_z<\infty)=\mathbb{P}(I_\nu(\infty)\leq z)$ but the distribution of $I_\nu(\infty)$ is actually known to be $U(0,1)$ (see the lemma below) so $\mathbb{P}(\tau_z<\infty)=z$ gives

$$x(K)\geq H(0) \tag{654}$$

the equality actually holds when $H(0)<\infty$ and it's easy to see since $H(0)\geq H(I_\nu)$, one shall be able to finance $K=H(I_\nu)$ with initial wealth $H(0)$ (put all wealth into money market) so $x(K)\leq H(0)$ resulting in

$$x(H(I_\nu)) = H(0) \text{ if } H(0)<\infty \tag{655}$$

gives the **exact formula for $x(K)$ attainable**. **In order to finance this capital withdrawal $H(I_\nu)$ starting from minimal possible initial wealth $H(0)$, one always has to stick to put all the wealth into the money market**.

Remark. *It's interesting to see that when the capital withdrawal is a function of the running sup of super-MG numeraire, one has to always trade in order to maintain the super-MG numeraire portfolio at all times. On the other hand, when capital withdrawal is a function of the running inf, one always sets aside the minimum amount of money needed to finance capital withdrawal in the money market in order to finance.*

Lemma 33 (Exercise 2.56)**.** *Let M be positive continuous MG such that $M(\infty)=0$ with $M^*(t)=\sup_{s\leq t}M(s)$, then*

$$\forall x>0, \mathbb{P}(M^*(\infty)\geq x|\mathscr{F}_0) = 1\wedge\frac{M_0}{x} \tag{656}$$

Proof. Consider stopping time $\tau=\inf\{t\geq 0: M(t)\geq x\}$ so $\mathbb{P}(M^*(\infty)\geq x|\mathscr{F}_0)=\mathbb{P}(\tau<\infty|\mathscr{F}_0)$ so $M(t\wedge\tau)$ is bounded thus U.I. MG. Optional stopping theorem tells us $\mathbb{E}(M(\tau)|\mathscr{F}_0) = M_0$. It's clear that $M(\tau)$ takes value either x or $M(\infty) = 0$ so

$$x\mathbb{P}\left(\tau < \infty|\mathscr{F}_0\right) = \mathbb{E}(M(\tau)|\mathscr{F}_0) = M_0 \tag{657}$$

concludes the proof. $\square$

Remark. *This property also holds for continuous local MG M from a localization argument. To see its implications, $\frac{1}{X_\nu}$ is positive continuous local MG starting from one with limit zero under the condition $X_\nu(\infty) = \infty$. As a result, its running sup process is $\frac{1}{I_\nu}$ with*

$$\mathbb{P}\left(\frac{1}{I_\nu(\infty)} \geq x\right) = \frac{1}{x} \tag{658}$$

that's why under the condition $X_\nu(\infty) = \infty$, $I_\nu(\infty) \sim U(0,1)$ has uniform distribution.

European Contingent Claim

A **European contingent claim** is a pair $(T, P(T))$ where T is the time of maturity and $P(T)$ is the payoff. Here $T \in \mathscr{T}$ is a stopping time and $P(T) \in \mathscr{F}_T$, denote Eu as the set of all European contingent claims. Naturally, one thinks of the European options where T is formed as a deterministic real number and $P(T)$ is taken as $(S_i(T)-k)_+$ for call option and $(k-S_i(T))_+$ for put option with strike price k. A wealth process $X \equiv X(\cdot; x, \theta, F) \geq 0$ is said to **hedge** $(T, P(T)) \in \text{Eu}$ if

$$X(T) \geq P(T) \text{ on } \{T < \infty\} \tag{659}$$

similarly we can define the **minimal hedge** $X \geq 0$ as the lowest wealth process that can hedge $(T, P(T)) \in \text{Eu}$ and the **hedging capital** $x_{\text{Eu}}(T, P(T))$ is the inf of all levels of initial wealth for which it's possible to choose an investment strategy to hedge $(T, P(T))$ (same as $x(K)$ defined previously).

It's immediate that the setting can be translated into the language of capital withdrawal where

$$K^{(T,P(T))} = P(T)\mathbb{I}_{[T,\infty)} \in \mathscr{K} \tag{660}$$

the financing duality tells us that the hedging capital has the characterization

$$x_{\text{Eu}}(T, P(T)) = \sup_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}}[Y(T)P(T)\mathbb{I}_{T<\infty}] = \sup_{\mathbb{Q} \in \Pi} \mathbb{E}_{\mathbb{Q}} \frac{P(T)}{X_\nu(T)} \mathbb{I}_{T<\infty} \tag{661}$$

under local MG deflators and model-consistent probability measure. When those two quantities are finite, the hedging capital is attainable (thus the **price of the option**). From this perspective, financing duality is very useful and provides us with a tool for asset pricing.

Lemma 34 (Exercise 3.26)**.** *Consider $(T, P(T)) \in Eu$ in a viable market with super-MG numeraire X_ν and assume there exists $b > 0$ such that $P(T) \leq bX_\nu(T)$ is true on the event $\{T < \infty\}$. Show that $x_{Eu}(T, P(T)) \leq b$ and the minimal hedge X for this contingent claim satisfies $X \leq bX_\nu$ on $[0, T]$.*

Proof. This is just an easy application of the conclusions for capital withdrawal

$$x_{\text{Eu}}(T, P(T)) = \sup_{\mathbb{Q} \in \Pi} \mathbb{E}_{\mathbb{Q}} \frac{P(T)}{X_\nu(T)} \mathbb{I}_{T<\infty} \leq b \sup_{\mathbb{Q} \in \Pi} \mathbb{Q}(T < \infty) \leq b \tag{662}$$

recall the characterization of minimal hedge that

$$\forall t \in [0, T], X(t) = \operatorname*{ess\,sup}_{\mathbb{Q} \in \Pi} \mathbb{E}_{\mathbb{Q}} \left[\frac{X_\nu(t)}{X_\nu(T)} P(T)\mathbb{I}_{T<\infty} \middle| \mathscr{F}_t \right] \leq b \operatorname*{ess\,sup}_{\mathbb{Q} \in \Pi} \mathbb{E}_{\mathbb{Q}} \left[X_\nu(t)\mathbb{I}_{T<\infty} \middle| \mathscr{F}_t \right] \leq bX_\nu(t) \tag{663}$$

concludes the proof. □

Lemma 35 (Exercise 3.27 (1)-(3))**.** ***(Samuelson-Black-Scholes-Merton Model)*** *Consider a market with single risky asset $S \equiv S_1$ such that $S(t) = S(0)e^{\gamma t + \sigma W(t)}$ where γ is the growth rate and $\sigma > 0$ as volatility.*

(1): Show that the market is viable and identify the super-MG numeraire X_ν.

(2): Show $\frac{1}{X_\nu}$ and $\frac{S}{X_\nu}$ are both MG under any $\mathbb{Q} \in \Pi$.

(3): If filtration $\mathscr{F}$ is generated by BM W, show that $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$, equivalently $\Pi = \{\mathbb{P}\}$.

Proof. (1): We can see that

$$dR(t) = \frac{dS(t)}{S(t)} = \left(\gamma + \frac{\sigma^2}{2}\right) dt + \sigma\, dW(t) \tag{664}$$

gives the semi-MG decomposition of $R(t)$ that $A(t) = \left(\gamma + \frac{\sigma^2}{2}\right) t, M(t) = \sigma W(t), C(t) = \sigma^2 t$ so $\alpha(t) = \gamma + \frac{\sigma^2}{2}, c(t) = \sigma^2$. For convenience, we denote $\gamma = \mu - \frac{\sigma^2}{2}$.

By the equivalent conditions for market viability, the maximal growth

$$\forall t \geq 0, G(t) = \frac{1}{2}\int_0^t \alpha^T(s)c^{-1}(s)\alpha(s)\, dO(s) = \frac{1}{2}\int_0^t \left(\gamma + \frac{\sigma^2}{2}\right)^2 \frac{1}{\sigma^2}\, ds = \frac{\left(\gamma + \frac{\sigma^2}{2}\right)^2}{2\sigma^2} t < \infty \tag{665}$$

so market has locally finite growth, is viable.

To identify super-MG numeraire portfolio ν, notice that

$$\nu = c^{-1}\alpha = \frac{\gamma + \frac{\sigma^2}{2}}{\sigma^2} = \frac{\mu}{\sigma^2} \tag{666}$$

so it's the best to always invest $\frac{\mu}{\sigma^2}$ (constant proportion) of the total wealth into this stock. Calculate R_ν to get $R_\nu(t) = \int_0^t \nu(s)\, dR(s) = \frac{\mu}{\sigma^2} R(t)$, from $X_\nu = \mathcal{E}\left(R_\nu\right)$ we see

$$\frac{1}{X_\nu(t)}\, dX_\nu(t) = \frac{\mu^2}{\sigma^2}\, dt + \frac{\mu}{\sigma}\, dW(t) \tag{667}$$

solve this SDE to get the super-MG numeraire

$$X_\nu(t) = e^{\frac{\mu^2}{2\sigma^2} t + \frac{\mu}{\sigma} W(t)} \tag{668}$$

(2): Here we have

$$\frac{S(t)}{X_\nu(t)} = S(0)e^{\left(\gamma - \frac{\mu^2}{2\sigma^2}\right)t + \left(\sigma - \frac{\mu}{\sigma}\right)W(t)} \tag{669}$$

where $-\frac{1}{2}\left\langle \left(\sigma - \frac{\mu}{\sigma}\right) W, \left(\sigma - \frac{\mu}{\sigma}\right) W \right\rangle (t) = -\frac{1}{2}\left(\sigma - \frac{\mu}{\sigma}\right)^2 t = \left(\gamma - \frac{\mu^2}{2\sigma^2}\right) t$ so its the stochastic exponential of $\left(\sigma - \frac{\mu}{\sigma}\right) W$, a MG. On the other hand,

$$\frac{1}{X_\nu(t)} = e^{-\frac{\mu^2}{2\sigma^2} t - \frac{\mu}{\sigma} W(t)} \tag{670}$$

with $-\frac{1}{2}\left\langle -\frac{\mu}{\sigma} W, -\frac{\mu}{\sigma} W \right\rangle (t) = -\frac{\mu^2}{2\sigma^2} t$ so its the stochastic exponential of $-\frac{\mu}{\sigma} W$, a MG. As a result, they are both MG under $\mathbb{P}$.

$\forall \mathbb{Q} \in \Pi$, $M(t) = \sigma W(t)$ is still a continuous local MG under $\mathbb{Q}$ but W still has quadratic variation t on time interval $[0, t]$ (pathwise property, does not change with measure). By Levy's characterization of BM, W is still BM under $\mathbb{Q}$ so $\frac{S}{X_\nu}, \frac{1}{X_\nu}$ are still MG under $\mathbb{Q}$.

(3):

Take $\forall \mathbb{Q} \in \Pi$ and consider $L(\cdot) = \mathcal{L}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathscr{F}(\cdot)}\right)$ is well-defined since $0 < \frac{d\mathbb{Q}}{d\mathbb{P}} < \infty$ $\mathbb{P}-a.s.$, apply Girsanov theorem to see that

$$\tilde{W} = W - \langle L, W \rangle \tag{671}$$

is a BM under $\mathbb{Q}$. However, W is BM under both $\mathbb{P}$ and $\mathbb{Q}$ so $L \equiv 0$ by the uniqueness of RN-derivative and $\frac{d\mathbb{Q}}{d\mathbb{P}} \equiv 1$ proves that $\mathbb{P} = \mathbb{Q}$ so the model-consistent probability measure is unique.

Let's then argue that $\mathscr{Y}$ only contains one element. The characterization of the elements in $\mathscr{Y}$ is given by

$$Y = \frac{1}{X_\nu}\mathcal{E}(L), L \in \mathscr{M}_{\text{loc}}^{\perp}(M) \tag{672}$$

since L is continuous local MG starting from zero such that $\langle L, W \rangle = 0$, by the same argument as above, $L \equiv 0$ and $\mathcal{E}(L) \equiv 1$, so $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$. $\square$

Remark. *In such a market with a single stock following Black-Scholes model, the* ***market is viable and complete****. As we have previously mentioned, local MG deflator and model-consistent probability measure are two sides of the same coin so it's natural to expect that* $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$ *iff* $\Pi = \{\mathbb{P}\}$.

Lemma 36 (Exercise 3.27 (4))**.** *For fixed* $\mathbb{Q} \in \Pi$ *and* $T > 0$*, define probability measure* $\mathbb{Q}_T^* \sim \mathbb{Q} \sim \mathbb{P}$ *such that*

$$\frac{d\mathbb{Q}_T^*}{d\mathbb{Q}} = \frac{1}{X_\nu(T)} \tag{673}$$

show that

$$\forall t \in [0, T], S(t) = S(0)e^{\sigma W^{\mathbb{Q}_T^*}(t) - \frac{\sigma^2}{2}t} \tag{674}$$

where $W^{\mathbb{Q}_T^*}$ *is BM under* $\mathbb{Q}_T^*$*. In particular,* S *is a MG under* $\mathbb{Q}_T^*$*.*

Proof. Obviously if the equation holds, S is MG under $\mathbb{Q}_T^*$ since it's the stochastic exponential of $\sigma W^{\mathbb{Q}_T^*}$. Notice that

$$S(t) = S(0)e^{\gamma t + \sigma W(t)} = S(0)e^{-\frac{\sigma^2}{2}t + [\sigma W(t) + \mu t]} \tag{675}$$

since $\frac{1}{X_\nu} = \mathcal{E}\left(-\frac{\mu}{\sigma}W\right)$ is MG under $\mathbb{Q}$, apply Girsanov theorem to see that

$$\tilde{W}(t) = W(t) - \left\langle -\frac{\mu}{\sigma}W, W \right\rangle(t) = W(t) + \frac{\mu}{\sigma}t \tag{676}$$

is a BM under $\mathbb{Q}_T^*$ with $S(t) = S(0)e^{\sigma \tilde{W}(t) - \frac{\sigma^2}{2}t}$ proves the result. Here we don not distinguish between $\mathbb{P}$ and $\mathbb{Q}$ since market is complete. $\square$

Remark. *This shows us the way to build up the* ***martingale measure (risk-neutral measure)*** *crucial for asset pricing. The RN-derivative is just set as the only local MG deflator* $\frac{1}{X_\nu}$*. Under* $\mathbb{Q}_T^*$ *the stock price dynamics does not exhibit any "tendency".*

Lemma 37 (Exercise 3.27 (5)-(6))**.** *For fixed $T > 0, k > 0$, calculate the minimal hedge associated with the European put $(T, (k - S(T))_+) \in Eu$. Argue that put-call parity*

$$x_{Eu}(T, (S(T) - k)_+) = S(0) - k + x_{Eu}(T, (k - S(T))_+) \tag{677}$$

holds.

Proof. Minimal hedging is

$$\begin{aligned}
\forall t \in [0, T], X(t) &= \operatorname*{ess\,sup}_{\mathbb{Q}\in\Pi} \mathbb{E}_{\mathbb{Q}}\left[\frac{X_\nu(t)}{X_\nu(T)} P(T)\mathbb{I}_{T<\infty}\Big|\mathscr{F}_t\right] &(678)\\
&= \mathbb{E}_{\mathbb{P}}\left[\frac{X_\nu(t)}{X_\nu(T)}(k - S(T))_+\Big|\mathscr{F}_t\right] &(679)\\
&= X_\nu(t)\mathbb{E}_{\mathbb{Q}_T^*}[(k - S(T))_+|\mathscr{F}_t] &(680)\\
&= e^{\frac{\mu^2}{2\sigma^2}t + \frac{\mu}{\sigma}W(t)}\mathbb{E}_{\mathbb{Q}_T^*}\left[\left(k - S(0)e^{\sigma W^{\mathbb{Q}_T^*}(T) - \frac{\sigma^2}{2}T}\right)_+\Big|\mathscr{F}_t\right] &(681)
\end{aligned}$$

setting $t = 0$ gives $x_{\mathrm{Eu}}(T, (k - S(T))_+)$

$$X(0) = \mathbb{E}_{\mathbb{Q}_T^*}\left(k - S(0)e^{\sigma W^{\mathbb{Q}_T^*}(T) - \frac{\sigma^2}{2}T}\right)_+ = \left[k\Phi\left(\frac{\log\frac{k}{S(0)} + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) - S(0)\Phi\left(\frac{\log\frac{k}{S(0)} - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right)\right] \tag{682}$$

recovers the **Black-Scholes formula with** $r = 0$. Notice that the minimal hedging is just the **value function** of the option.

Similar calculations give

$$x_{\mathrm{Eu}}(T, (S(T) - k)_+) = \left[-k\Phi\left(\frac{\log\frac{k}{S(0)} - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) + S(0)\Phi\left(\frac{\log\frac{k}{S(0)} + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right)\right] \tag{683}$$

so

$$x_{\mathrm{Eu}}(T, (S(T) - k)_+) - x_{\mathrm{Eu}}(T, (k - S(T))_+) = S(0) - k \tag{684}$$

put-call parity holds. $\square$

Now consider a viable complete market with n stock dynamics $S_1, ..., S_n$, fix $T > 0, p_j : (0, \infty)^n \to \mathbb{R}_+$ as payoff functions, then it's clear that

$$x_{\mathrm{Eu}}(T, (p_2(S(T)) - p_1(S(T)))_+) = x_{\mathrm{Eu}}(T, p_2(S(T))) - x_{\mathrm{Eu}}(T, p_1(S(T))) + x_{\mathrm{Eu}}(T, (p_1(S(T)) - p_2(S(T)))_+) \tag{685}$$

as a **general parity relationship**. To be more specific, set $p_1(z) = k, p_2(z) = z_i$ to see the **European contingent**

claim put-call parity

$$x_{\mathrm{Eu}}(T,(S_i(T)-k)_+) = \mathbb{E}_{\mathbb{P}}\frac{S_i(T)-k}{X_\nu(T)} + x_{\mathrm{Eu}}(T,(k-S_i(T))_+) \tag{686}$$

notice that $\frac{1}{X_\nu}, \frac{S_i}{X_\nu}$ are both non-negative local MG thus super-MG, we have the estimates from optional stopping

$$\mathbb{E}_{\mathbb{P}}\frac{S_i(T)}{X_\nu(T)} \le \frac{S_i(0)}{X_\nu(0)} = S_i(0), \mathbb{E}_{\mathbb{P}}\frac{k}{X_\nu(T)} \le k \tag{687}$$

in Black-Scholes model, however, $\frac{1}{X_\nu}, \frac{S_i}{X_\nu}$ are both MG, that's why the inequalities become equalities and gives

$$\mathbb{E}_{\mathbb{P}}\frac{S_i(T)-k}{X_\nu(T)} = S_i(0)-k \tag{688}$$

so we get the **"pure" version of European put-call parity**

$$x_{\mathrm{Eu}}(T,(S_i(T)-k)_+) = S_i(0)-k+x_{\mathrm{Eu}}(T,(k-S_i(T))_+) \tag{689}$$

under the extra condition that $\frac{1}{X_\nu}, \frac{S_i}{X_\nu}$ are both MG.

Remark. *It might be somewhat surprising (since put-call parity is derived based on no-arbitrage criterion in elementary finance courses) to see that* ***the pure version of European put-call parity as a model-free result might not hold even in a viable complete market (i.e. enough different assets with no-arbitrage) where one of $\frac{1}{X_\nu}, \frac{S_i}{X_\nu}$ fails to be a MG***.

Recall the example shown previously where there's a single stock in the market with $S_t = Z_t$ as a 3D Bessel process starting from z. The market is then viable and complete with $X_\nu = \frac{Z}{z}$ such that

$$\mathbb{E}_{\mathbb{P}}\frac{1}{X_\nu(T)} = \mathbb{E}\frac{z}{Z(T)} = 2\Phi\left(\frac{z}{\sqrt{T}}\right)-1 \tag{690}$$

now we see that

$$\mathbb{E}_{\mathbb{P}}\frac{S(T)-k}{X_\nu(T)} = z-k\left[2\Phi\left(\frac{z}{\sqrt{T}}\right)-1\right] \tag{691}$$

while $S(0)-k = z-k$, so the pure version of European put call parity fails and we get

$$x_{Eu}(T,(S_i(T)-k)_+) > S_i(0)-k+x_{Eu}(T,(k-S_i(T))_+) \tag{692}$$

this is due to the fact that $\frac{1}{X_\nu}$ as the reciprocal of 3D Bessel process is a strict local MG.

Market Completeness, Replicability

In a viable market, a European contingent claim $(T, P(T)) \in \text{Eu}$ is called **replicable** if the minimal hedge associated does not involve any capital withdrawal, i.e. $X \equiv X(\cdot; x, \theta)$ is the minimal hedge for some x, θ. A viable market is **complete** if every European contingent claim with $x_{\text{Eu}}(T, P(T)) < \infty$ is replicable. Actually, a European contingent claim is replicable if one can build up the same payoff using the assets in the market. Naturally, the European contingent claim and the replicating portfolio shall have the same price since they have exactly the same payoff (option pricing from the perspective of Delta hedging). A market is complete if there are enough different types of assets available so one can replicate any reasonable European contingent claim, which turns out to be equivalent to saying $\Pi = \{\mathbb{P}\}$ (two notions of market completeness are equivalent). On the other hand, we know that in BSDE

$$dY_t = -f(t, Y_t, Z_t)\, dt + Z_t\, dB_t, Y_T = \xi \tag{693}$$

Y_t is often interpreted as option price and Z_t as the hedging strategy. As a result, market completeness naturally connects with the hedging strategy and the MG representation theorem (to derive Z_t).

Theorem 20. ***(Second Fundamental Theorem of Asset Pricing)*** *In a viable market, the following are equivalent:*

(1): Market is complete.
(2): $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$
(3): $\Pi = \{\mathbb{P}\}$

Proof. (1) implies (2):

Assume (2) fails so $\mathbb{P}\left(Y(T) = \frac{1}{X_\nu(T)}\right) < 1$ for some $T > 0$, due to characterization of $\mathscr{Y}$,

$$\forall Y \in \mathscr{Y}, \exists L \in \mathscr{M}_{\text{loc}}^{\perp}(M), Y = \frac{1}{X_\nu}\mathcal{E}(L), \mathbb{P}(L(T) = 0) < 1 \tag{694}$$

WLOG we assume $|L| \leq \frac{1}{2}$ (otherwise we can set τ as the stopping time $|L|$ first go beyond $\frac{1}{2}$ and consider $L(t \wedge \tau) = L^\tau(t)$, then $\langle L^\tau, M_i\rangle(t) = \langle L, M_i\rangle(t \wedge \tau) = 0$ still orthogonal with all M_i). Consider

$$P(T) = \left(\frac{1}{2} + L(T)\right) X_\nu(T) \in \mathscr{F}_T \tag{695}$$

as the payoff of European contingent claim $(T, P(T)) \in \text{Eu}$ such that $0 \leq P(T) \leq X_\nu(T)$. The previous lemma tells us that $0 \leq x_{\text{Eu}}(T, P(T)) \leq 1$ and the minimal hedge $0 \leq X(\cdot; x, \theta) \leq X_\nu(\cdot)$ for some investment strategy θ such that $X(T; x, \theta) = P(T)$.

Consider the minimal hedge discounted by super-MG numeraire $X^\nu = \frac{X}{X_\nu} \in [0, 1]$ as a bounded local MG, thus a MG with $X^\nu(T; x, \theta) = \frac{1}{2} + L(T)$. Now that L is also a bounded local MG, $X^\nu - L$ is a bounded local MG, thus a MG that takes value $\frac{1}{2}$ at time T. By the definition of a MG, $X^\nu - L \equiv \frac{1}{2}$. Notice that X^ν can be represented as a stochastic integral w.r.t. M, we know

$$\langle L, L\rangle = \langle L, X^\nu\rangle . = 0 \tag{696}$$

so $L \equiv 0$ proves the uniqueness of local MG deflator.

(2) implies (1):

In order to prove that for any European contingent claim, there always exists x, θ such that the minimal hedge is $X(\cdot; x, \theta)$, recall the optional decomposition theorem (the corollary with no capital withdrawal) that it suffices to prove that $\forall Y \in \mathscr{Y}$, YX is local MG. Since $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$ and we have the representation

$$X(t) = \operatorname*{ess\,sup}_{Y \in \mathscr{Y}} \mathbb{E}_{\mathbb{P}}\left[\frac{Y(T)}{Y(t)} P(T) \Big| \mathscr{F}_t\right] = X_\nu(t) \mathbb{E}_{\mathbb{P}}\left[\frac{1}{X_\nu(T)} P(T) \Big| \mathscr{F}_t\right] \tag{697}$$

with $\mathbb{E}_{\mathbb{P}}\left[\frac{1}{X_\nu(T)} P(T) \Big| \mathscr{F}_t\right]$ as a MG in t. It's clear that

$$YX = \frac{X}{X_\nu} = \mathbb{E}_{\mathbb{P}}\left[\frac{1}{X_\nu(T)} P(T) \Big| \mathscr{F}_t\right] \tag{698}$$

is local MG and the proof is done.

(2) implies (3) and (3) implies (2):

Notice the connection previously shown between local MG deflator and model-consistent probability measure that $\forall Y \in \mathscr{Y}, Y = \frac{1}{X_\nu} \mathcal{E}(L), L \in \mathscr{M}_{\text{loc}}^{\perp}(M)$ with localization sequence $\tau_m \leq m$ for $\mathcal{E}(L)$ that induces $\mathbb{Q}_m \in \Pi$ through

$$\frac{d\mathbb{Q}_m}{d\mathbb{P}} = \mathcal{E}(L)(\tau_m) \tag{699}$$

proves the uniqueness of local MG deflator when model-consistent probability measure is unique.

Conversely, $\mathbb{Q} \in \Pi$ induces the special local MG deflator $Y^{\mathbb{Q}}(\cdot) = \frac{1}{X_\nu} \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathscr{F}(\cdot)}$, proves the uniqueness of model-consistent probability measure when the local MG deflator is unique. $\square$

Lemma 38 (Exercise 3.43). *$(W_1, ..., W_m)$ is m-dimensional BM and assume $n \leq m$ (more source of uncertainty than the number of assets), there are n stock price dynamics*

$$S_i(t) = S_i(0) + \int_0^t S_i(s) \alpha_i(s)\, ds + \int_0^t S_i(s) \sum_{j=1}^m \sigma_{ij}(s)\, dW_j(s) \tag{700}$$

where $\alpha, c = \sigma\sigma^T$ satisfies integrability conditions and the operational clock is taken as Lebesgue clock. Notice that here the filtration is taken as the one generated by $W_1, ..., W_m$ (important condition). Show that market is viable and complete iff the following conditions hold:

(1): $m = n$

(2): c is non-singular.

(3): $\forall T \geq 0, \int_0^T \alpha^T(s) c^{-1}(s) \alpha(s)\, ds < \infty$ $\mathbb{P} - a.s.$

Remark. *We do not prove this lemma right here but remind the reader of the interpretation of the lemma that a complete market requires enough quantity of different assets in order to replicate any European contingent claim. As* *a result, the number of asset shall be not too small and σ shall have full row rank (assets are different enough from each other).*

A rigorous characterization of this intuition is based on the **martingale representation property**. We say $L = (L_1, ..., L_m)^T$ has martingale representation property if for any scalar local MG on the same filtered probability space, there exists predictable $\eta = (\eta_1, ..., \eta_m)^T$ such that

$$N(t) = N(0) + \int_0^t \sum_{j=1}^m \eta_j(s)\, dL_j(s) \tag{701}$$

in other words, L spans a large enough space to contain all local MGs. It's natural to expect that **market is complete iff M has the martingale representation property**.

Theorem 21. *L is a continuous local MG, the following statements are equivalent:*

(1): L has martingale representation property.

(2): If $\mathbb{Q} \sim \mathbb{P}$ is a probability measure such that L is still local MG under $\mathbb{Q}$, then $\mathbb{P} = \mathbb{Q}$.

Proof. If L has martingale representation property and L is local MG under $\mathbb{Q}$, Girsanov theorem tells us that there exists local MG Λ starting from 0 such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}(\cdot)} = \mathcal{E}\left(\Lambda\right)(\cdot) \tag{702}$$

and $\tilde{L}_j = L_j - \langle L_j, \Lambda \rangle$ is local MG under $\mathbb{Q}$ so $\forall j, \langle L_j, \Lambda \rangle = 0$. Now Λ can be represented under L that

$$\Lambda(t) = \int_0^t \sum_{j=1}^m \eta_j(s)\, dL_j(s) \tag{703}$$

calculating the quadratic variation shows

$$\langle \Lambda, \Lambda \rangle (t) = \int_0^t \sum_{j=1}^m \eta_j(s)\, d\left\langle L_j, \Lambda \right\rangle (s) = 0 \tag{704}$$

proves $\Lambda \equiv 0$ so $\mathbb{P} = \mathbb{Q}$.

Conversely, if N is a continuous local MG that cannot be represented under L, by **Kunita-Watanabe decomposition** (unique),

$$N(t) = \int_0^t H(s)\, dL(s) + \Lambda(t) \tag{705}$$

for some H predictable and $\Lambda \not\equiv 0$ starting from 0 that is strongly orthogonal to L, i.e. $\forall j, \langle \Lambda, L_j \rangle = 0$. it's natural to think about applying Girsanov once again such that $\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}(\cdot)} = 1 + \Lambda(\cdot)$. If this can be done, then $\tilde{L}_j = L_j - \langle \Lambda, L_j \rangle = L_j$ is local MG under $\mathbb{Q}$ and $\Lambda \not\equiv 0$ so $\mathbb{P} \neq \mathbb{Q}$, a contradiction! The only problem is that such argument requires Λ to be a MG. Notice that in the proof of second fundamental theorem of asset pricing that

WLOG one can assume $|\Lambda| \leq \frac{1}{2}$ when $\mathbb{P}(\Lambda(T) = 0) < 1$ for some time $T \geq 0$ and this makes Λ a bounded local MG, thus a MG. We conclude the proof by Girsanov. $\square$

Remark. *From the perspective of martingale representation property, the lemma above can be easily proved since $n < m$ or a singular c causes the dimension of the range of $M = \sigma W$ to be strictly less than m, violating the martingale representation property that is equivalent to market completeness.*

Utility Maximization

The capital withdrawal can be understood as consumption from which one gets utility. The **utility random field** is formed as

$$U\left(t, \frac{dK(t)}{dQ(t)}\right),\ U : \Omega \times \mathbb{R}_+ \times (0,\infty) \to \mathbb{R} \tag{706}$$

mapping the time and the rate of consumption to a utility level. Notice that here we are not taking the operational clock $O(t)$ to get the rate but an **internal consumption clock** $Q(t)$. This idea of changing clock has appeared for many times previously and the intuition will be mentioned later. **With initial wealth x, one shall choose the best financeable consumption stream $K \in \mathscr{K}(x)$ such that it maximizes the expected utility**

$$\mathbb{U}(K) = \mathbb{E}_{\mathbb{P}} \int_0^\infty U\left(t, \frac{dK(t)}{dQ(t)}\right) dQ(t) \tag{707}$$

written in a formal way, we want to identity $K_x \in \mathscr{K}(x)$ such that

$$\mathbb{U}(K_x) = \sup_{K \in \mathscr{K}(x)} \mathbb{U}(K) \tag{708}$$

we first formalize the internal consumption clock Q and utility random field U and then give the solution to this optimization problem.

The **saturation time** of consumption stream K is defined as the time when one expects not to see any future consumption based on the current information, i.e.

$$\tau_K \stackrel{def}{=} \inf\left\{t \geq 0 : \mathbb{P}\left(K(t) < K(\infty) | \mathscr{F}(t)\right) = 0\right\} \tag{709}$$

the space of all admissible internal consumption clock is formed as

$$\mathscr{Q} \stackrel{def}{=} \left\{Q \in \mathscr{K} : \mathbb{P}\left(Q(\infty) = 0\right) = 0, \mathbb{P}\left(\tau_Q < \infty, Q(\tau_Q) = Q(\tau_Q -)\right) = 0\right\} \tag{710}$$

whenever we specify an internal consumption clock Q in the following context, it's always assumed that $Q \in \mathscr{Q}$.

Remark. *For the intuition of such $\mathscr{Q}$, notice that $\mathbb{P}\left(Q(\infty) = 0\right) = 0$ makes sure that Q is not trivial and the second condition $\mathbb{P}\left(\tau_Q < \infty, Q(\tau_Q) = Q(\tau_Q -)\right) = 0$ makes sure that whenever the saturation time of Q is finite, the agent can always consume the remaining wealth at the saturation time and derive utility.*

In different problem settings, the internal consumption clock can be specified in different ways. If the agent cares about infinite time horizon, Q can be formed as strictly increasing so $\tau_Q = \infty$ $\mathbb{P}-a.s.$. On the other hand, if one cares about only the consumption at some finite stopping time T, $Q = \mathbb{I}_{[T,\infty)}$ can be taken such that $\tau_Q = T$ $\mathbb{P}-a.s.$. In brief, set up Q such that the clock under Q moves iff one cares about the consumption in such time period.

An example for Q in the infinite horizon setting can be stated as

$$Q(t) = 1 - e^{-\int_0^t q(s)\, ds} \tag{711}$$

for q non-negative locally integrable and optional such that $\int_0^\infty q(t)\,dt = \infty$. It's clear that Q is strictly increasing in t with $Q(\infty) = 1$ so $\tau_Q = \infty, Q \in \mathscr{Q}$. The interpretation of this model is provided by $\frac{dQ(t)}{1-Q(t)} = q(t)\,dt$ with q as **local impatience rate**. The higher q it takes, one is more impatient so today's consumption is more important than the consumption in the future.

The random utility field $U(t,\cdot)$ takes values in $\mathscr{U}$, the collection of all increasing, strictly concave $\phi \in C^1$ that satisfies the Inada condition $\phi'(0+) = +\infty, \phi'(\infty) = 0$. The formulation of concave utility function shows the risk averse of investors and is a frequently used practice so we do not comment much on this. Notice that by saying $U \in \mathscr{U}$, we actually mean that on viewing $U = U(t,x)$ as a deterministic function (with sample point ω fixed) in x, it satisfies the restrictions in $\mathscr{U}$.

To introduce the conclusion, let's introduce the concave conjugate of U as U_*

$$V(t,y) = U_*(t,y) = \inf_{x>0} \{xy - U(t,x)\} \tag{712}$$

with the **primal value function and the dual value function** given by

$$u(x) \overset{def}{=} \sup_{K\in\mathscr{K}(x)} \mathbb{E}_{\mathbb{P}} \int_0^\infty U\left(t, \frac{dK(t)}{dQ(t)}\right) dQ(t) \tag{713}$$

$$v(y) \overset{def}{=} \sup_{Y\in\mathscr{Y}} \mathbb{E}_{\mathbb{P}} \int_0^\infty V\left(t, yY(t)\right) dQ(t) \tag{714}$$

under the assumption that $\forall x,y > 0, u(x) > -\infty, v(y) > -\infty$. Under this setting, the following theorem is provided as a conclusion without a proof.

Theorem 22. ***(Utility Maximization)*** *Fix $Q \in \mathscr{Q}$ and an admissible U, assume market is viable and $\forall x,y > 0, u(x) > -\infty, v(y) > -\infty$, then*

(1): $\forall x,y > 0, u(x) < \infty, v(y) < \infty$ with $u,v \in C^1$ to be strictly increasing, strictly concave with the Inada conditions and u,v are concave conjugate of each other

$$v(y) = \inf_{x>0} \{xy - u(x)\}, u(x) = \inf_{y>0} \{xy - v(y)\} \tag{715}$$

(2): $\forall x > 0$, $K_x \in \mathscr{K}(x)$ as the maximizer of utility maximization problem exists and is unique. $\forall y > 0$, $Y_y \in \mathscr{Y}$ as the maximizer that attains $v(y)$ exists and is unique on $[0,\tau_Q]$. Whenever $y = u'(x)$ holds for (x,y), there is duality relationship

$$Y_y(t) = \frac{1}{y} U'\left(t, \frac{dK_x}{dQ}\right) \Longleftrightarrow \frac{dK_x}{dQ}\Big|_{\mathscr{F}(t)} = V'(t, yY_y) \tag{716}$$

with the equality $u(x) + v(y) = xy$ and the complementary slackness

$$\mathbb{E}_{\mathbb{P}} \int_0^\infty Y_y(t)\,dK_x(t) = x \tag{717}$$

Remark. *It's not hard to see from the structure of the theorem that y as the dual variable actually works as the Lagrange multiplier in the optimization problem of maximizing utility with the constraint that K is financeable with* *initial wealth x. As a result, the conclusion is a standard argument in optimization theory stating the bi-conjugacy between u (primal problem) and v (dual problem).*

Let's look at some examples in order to understand the implications of this theorem.

Lemma 39. *Assume the market is viable and complete, define function w as*

$$w(y) = \mathbb{E}_{\mathbb{P}} \int_0^\infty V'\left(t, \frac{y}{X_\nu(t)}\right) \frac{1}{X_\nu(t)}\, dQ(t) \tag{718}$$

with all the conditions of the theorem above to hold, show that

(1): w is strictly decreasing, $w(0+) = \infty, w(\infty) = 0$ with $w = v', w^{-1} = u'$.

(2): For $\forall x > 0, y_x = w^{-1}(x) = u'(x)$, the $K_x \in \mathscr{K}(x)$ that attains the maximal utility is given by

$$K_x(t) = \int_0^t V'\left(s, \frac{y_x}{X_\nu(s)}\right) dQ(s) \tag{719}$$

Proof. Since now the market is complete, $\mathscr{Y} = \left\{\frac{1}{X_\nu}\right\}$, the dual problem is trivial with

$$v(y) = \mathbb{E}_{\mathbb{P}} \int_0^\infty V\left(t, \frac{y}{X_\nu(t)}\right) dQ(t) \tag{720}$$

it's easy to see that $w = v'$, apply utility maximization theorem to see that $w(0+) = \infty, w(\infty) = 0$ follows from the Inada condition of v. Since v is strictly concave, $w = v'$ is strictly decreasing. Now the derivative of $xy - v(y)$ w.r.t. y gives $x - w(y)$ so $xw^{-1}(x) - v(w^{-1}(x)) = u(x)$. Taking derivative on both sides gives

$$u'(x) = w^{-1}(x) + x\frac{d}{dx}w^{-1}(x) - w(w^{-1}(x))\frac{d}{dx}w^{-1}(x) = w^{-1}(x) \tag{721}$$

By the second part of utility maximization, now that $y_x = u'(x)$ holds, check the complementary slackness

$$\mathbb{E}_{\mathbb{P}} \int_0^\infty Y_{y_x}(t)\, dK_x(t) = x \tag{722}$$

by plugging in the given K_x that

$$\mathbb{E}_{\mathbb{P}} \int_0^\infty Y_{y_x}(t) V'\left(t, \frac{y_x}{X_\nu(t)}\right) dQ(t) = \mathbb{E}_{\mathbb{P}} \int_0^\infty \frac{1}{X_\nu(t)} V'\left(t, \frac{y_x}{X_\nu(t)}\right) dQ(t) = w(y_x) = x \tag{723}$$

the complementary slackness condition holds. Since the optimizer K_x is unique, this proves the conclusion. ∎

Ramifications and Extensions

Fractional Brownian Motion (FBM)

Please refer to the paper *Stochastic Integration with respect to fractional Brownian motion and applications* by David Nualart for a detailed introduction to FBM. Here I would like to list some of the important properties of FBM.

The definition of FBM $\{B_t\}_{t\geq 0}$ is from the Gaussian process (GP) perspective that it's a **centered GP with covariance kernel**

$$R_H(t,s) \overset{def}{=} \mathbb{E}B_tB_s = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \tag{724}$$

where $H \in (0,1)$ is the **Hurst parameter**. From the covariance kernel, it's immediate that FBM has **self-similar property**, i.e.

$$\forall a > 0, \left\{a^{-H}B_{at}\right\}_{t\geq 0} \overset{d}{=} \{B_t\}_{t\geq 0} \tag{725}$$

it's easy to see that FBM is BM when $H = \frac{1}{2}$ so we eliminate the case where $H = \frac{1}{2}$ in our discussion below and refer to FBM as a fractional Brownian motion that is not a true BM. Calculation for the increments gives

$$\mathbb{E}|B_t - B_s|^2 = |t-s|^{2H}, \rho_H(n) \overset{def}{=} cov(B_{t+h} - B_t, B_{s+h} - B_s) = \frac{1}{2}h^{2H}\left((n+1)^{2H} + (n-1)^{2H} - n^{2H}\right) \tag{726}$$

where $t - s = nh$ with $s + h \leq t$. It's immediate that **FBM has stationary but not independent increments**. Taylor series expansion tells us

$$\rho_H(n) \sim H(2H-1)h^{2H}n^{2H-2} \to 0 \ (n \to \infty) \tag{727}$$

so when $H > \frac{1}{2}$, $\rho_H(n) > 0, \sum_{n=1}^{\infty} |\rho_H(n)| = \infty$. **When Hurst parameter is larger than $\frac{1}{2}$, there is aggregation behavior, auto-correlation is positive and is not decaying fast to zero**. On the other hand, when $H < \frac{1}{2}$, $\rho_H(n) < 0, \sum_{n=1}^{\infty} |\rho_H(n)| < \infty$. **When Hurst parameter is less than $\frac{1}{2}$, there is intermittency behavior, auto-correlation is negative and is decaying fast enough**. It's interesting to see that the behavior of FBM varies according to the Hurst parameter. Notice that no matter what value Hurst parameter takes, FBM always exhibits **asymptotic independence** and by Kolmogorov's continuity lemma, it has **a modification with α-Holder continuous sample path for $\forall \alpha \in (0, H)$**.

The relationship between FBM and BM can be given by the moving average representation, whose proof is a simple calculation and comparison of covariance kernel since FBM is GP.

Theorem 23. ***(Moving Average Representation)*** *Let $\{W_t\}_{t\in\mathbb{R}}$ be a BM on the real line, then*

$$B_t = \frac{1}{C_1(H)}\int_{\mathbb{R}} \left[(t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}\right] dW_s \tag{728}$$

where

$$C_1(H) = \sqrt{\frac{1}{2H} + \int_0^{\infty} \left((1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}}\right)^2 ds} \tag{729}$$

is a constant as a function of H*.*

Remark. *Here* $(t-s)_+^{H-\frac{1}{2}}$ *actually means* $(\max\{t-s,0\})^{H-\frac{1}{2}}$*, notice that here* $0^{H-\frac{1}{2}}$ *is always defined to be* 0 *no matter what value* H *takes. In other words,*

$$B_t = \frac{1}{C_1(H)}\left(\int_{-\infty}^0 \left[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right] dW_s + \int_0^t (t-s)^{H-\frac{1}{2}}\, dW_s\right) \tag{730}$$

The BM on the whole real line is defined through taking two independent BM W^1, W^2 *on* $\mathbb{R}_+$ *and stick them together through*

$$W_t = \begin{cases} W^1_{-t} & t<0 \\ W^2_t & t\geq 0 \end{cases} \tag{731}$$

one may verify that Ito's isometry still holds for the integral w.r.t. such W *in the calculation of the covariance kernel.*

The motivation of FBM comes from some statistical phenomenon that some time series data $\{X_n\}$ *satisfies* $X_n \overset{d}{=} n^H X_1$ *for* $H \neq \frac{1}{2}$*. By the self-similar property stated above, we immediately know that such* X *can be modelled as a FBM with Hurst parameter* H*.*

Although the moving average representation provides connections between FBM and BM, it's not very useful since it's an integral on an unbounded interval. Actually, there is another important representation of FBM as a Volterra process. Before entering into that, we first argue that FBM is not a semi-MG, raising a different class of stochastic process for which Ito formula does not work any longer. A natural question to ask is that if FBM could be a good model for asset price, it turns out that if the asset price follows FBM there always exists arbitrage in the market.

Theorem 24. *For FBM* B *with Hurst parameter* $H \neq \frac{1}{2}$*, it's not a semi-MG.*

Proof. The proof is smart and inspirational so we list the details here.

It's easy to see that $B_1 - B_0, B_2 - B_1, ...$ is a strongly stationary sequence (FBM has stationary increments and is GP) with asymptotic independence (mixing) so it must be ergodic. By Birkhoff's ergodic theorem,

$$\forall p>0, \frac{1}{n}\sum_{j=1}^n |B_j - B_{j-1}|^p \overset{a.s.\ L^1}{\to} \mathbb{E}|B_1|^p \ (n\to\infty) \tag{732}$$

use the self-similar property of FBM that

$$\frac{1}{n}\sum_{j=1}^n |B_j - B_{j-1}|^p \overset{d}{=} \frac{1}{n^{1-pH}}\sum_{j=1}^n |B_{\frac{j}{n}} - B_{\frac{j-1}{n}}|^p \tag{733}$$

to conclude $\frac{1}{n^{1-pH}}\sum_{j=1}^n |B_{\frac{j}{n}} - B_{\frac{j-1}{n}}|^p \overset{d}{\to} \mathbb{E}|B_1|^p > 0\ (n\to\infty)$ and the convergence can be lifted to convergence in probability since the limit is deterministic. As a result, the asymptotic behavior of

$$\sum_{j=1}^n |B_{\frac{j}{n}} - B_{\frac{j-1}{n}}|^p \tag{734}$$

is known (whose limit is the p-variation of B on time interval $[0,1]$). When $pH<1$, it has limit ∞ in probability and when $pH>1$ it has limit 0 in probability.

When $H<\frac{1}{2}$, $\exists p>2, pH<1$ so FBM has infinite quadratic variation on $[0,1]$ so it can't be semi-MG.

When $H>\frac{1}{2}$, for $\forall 1<p<\frac{1}{H}$ the p-variation is infinite so FBM has infinite total variation. For $\forall \frac{1}{H}<p<2$, the p-variation is zero so the quadratic variation must be zero. This contradicts FBM being semi-MG since infinite total variation indicates that there exists nontrivial local MG part but zero quadratic variation indicates the opposite. $\square$

Remark. *The proof above does not work for BM when $H=\frac{1}{2}$ since $\frac{1}{H}=2$ which corresponds to the quadratic variation. An interesting fact is that if one has FBM B with Hurst parameter H and an independent BM W, consider their sum $M_t=B_t+W_t$. Then M is not semi-MG when $H\le\frac{3}{4}, H\ne\frac{1}{2}$ but M is semi-MG with same distribution as BM when $H>\frac{3}{4}$ (FBM is covered by BM).*

Theorem 25. ***(Representation of FBM on an Interval)*** *Fix time interval $[0,T]$ for $\forall T>0$, let B be FBM with Hurst parameter $H\ne\frac{1}{2}$, then*

$$\forall t\in[0,T], B_t=\int_0^t K_H(t,s)\,dW_s \tag{735}$$

where W is a BM and K_H is a square-integrable kernel given by

$$\forall t\ge s, K_H(t,s)=\begin{cases} C_1(H)s^{\frac{1}{2}-H}\int_s^t (u-s)^{H-\frac{3}{2}}u^{H-\frac{1}{2}}\,du & H>\frac{1}{2}\\ C_2(H)\left[\left(\frac{t}{s}\right)^{H-\frac{1}{2}}(t-s)^{H-\frac{1}{2}}-\left(H-\frac{1}{2}\right)s^{\frac{1}{2}-H}\int_s^t u^{H-\frac{3}{2}}(u-s)^{H-\frac{1}{2}}\,du\right] & H<\frac{1}{2}\end{cases} \tag{736}$$

with the constants given by

$$\begin{cases} C_1(H)=\sqrt{\frac{H(2H-1)}{B(2-2H,H-\frac{1}{2})}}\\ C_2(H)=\sqrt{\frac{2H}{(1-2H)B(1-2H,H+\frac{1}{2})}}\end{cases} \tag{737}$$

(B denotes the Beta function) and the property holds for both cases that

$$\int_0^{t\wedge s} K_H(t,u)K_H(s,u)\,du=R_H(t,s) \tag{738}$$

intuitively, K_H is the square root of R_H as an operator.

Remark. *The representation is crucial since the integral is now on finite interval $[0,t]$, but the proof includes a lot of tricks so we do not present it here. To mention one of the points, fractional Gaussian white noise G can be constructed as an isometry from a Hilbert space $L^2(E,\mathscr{E},\mu)$ to the centered Gaussian space generated by FBM.*

In detail, here intensity μ is a sigma-finite measure on $(E,\mathscr{E})$ and the Gaussian white noise G satisfies

$$\langle f,g\rangle_{L^2}\overset{def}{=}\mathbb{E}G(f)G(g) \tag{739}$$

taking $f=\mathbb{I}_A, \mu(A)<\infty$ gives $G(\mathbb{I}_A)\sim N(0,\mu(A))$ and $A_1,...,A_n\in\mathscr{E}$ as disjoint sets implies the independence between $G(A_1),...,G(A_n)$. In particular, $W_t=G(\mathbb{I}_{[0,t]})$ has the same distribution as the BM by taking $\mu=\lambda$ as the

Lebesgue measure and $E = \mathbb{R}_+$. Such construction is essentially based on the fact that the covariance kernel of BM is $C(s,t) = s \wedge t$ so

$$\left\langle \mathbb{I}_{[0,t]}, \mathbb{I}_{[0,s]} \right\rangle_{L^2(\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+}, \lambda)} = \int_0^{t\wedge s} \lambda(du) = t \wedge s = C(s,t) \tag{740}$$

In the context of FBM, however, the covariance kernel changes but the idea of using covariance kernel to build up a Hilbert space still works. Shortly speaking, we just want to define an inner product such that

$$\left\langle \mathbb{I}_{[0,t]}, \mathbb{I}_{[0,s]} \right\rangle_{\mathscr{H}} = R_H(s,t) \tag{741}$$

on Hilbert space $\mathscr{H}$. This space $\mathscr{H}$ is actually called the ***reproducing kernel Hilbert space (RKHS) of*** *R_H and we will turn back to this idea in a later context. The G is an isometry between such Hilbert space and the Gaussian space, and it is just the* ***fractional Gaussian white noise****.*

Remark. *Such integral gives B as a FBM but not a semi-MG since $K_H(t,s)$ has dependence on t which cannot necessarily go out of the integral (recall that when $K_H(t,s) = e^{-(t-s)}$ is the exponential kernel such integral provides the OU process). This type of integral equation is called a* ***Volterra integral equation****. Nevertheless, the Volterra integral is much easier to approximate numerically so such characterization of FBM is always applied to numerically simulate its trajectories.*

One can refer to the paper *Arbitrage for Fractional Brownian Motion by L.C.G. Rogers* for interesting results of FBM in finance. The paper starts from the moving average representation of FBM (on the whole real line $\mathbb{R}$) and constructs an arbitrage strategy based on the self-similar property and the stationarity of increments of FBM. The basic idea is to argue that the so-called promising investment period exists and appears infinitely often due to ergodicity. Here we do not talk about the details of the paper but we mention something about the Volterra type integral and its connection with Gaussian process and arbitrage in financial market.

From the moving average representation of FBM, one naturally thinks about Gaussian processes that are in the form

$$\forall t \in \mathbb{R}, X_t = \int_{-\infty}^t \varphi(t-s)\, dW_s - \int_{-\infty}^0 \varphi(-s)\, dW_s \tag{742}$$

where W is now a BM on the whole real line and φ is some convolution kernel. Those processes can be seen as a convolution between $\varphi(s)$ and dW_s as a Volterra-type integral. Rogers argued in his paper that the behavior of φ near 0 determines whether arbitrage exists in a market if there is only one asset whose price at time t is X_t while the behavior of φ at ∞ determines whether long-range dependence of X is allowed. The conditions he comes up with are that

$$\varphi \in C^2(\mathbb{R}), \varphi(0) = 1, \varphi'(0) = 0, \lim_{t\to\infty} \varphi''(t) t^{\frac{5}{2}-H} = L \in (0,\infty) \tag{743}$$

ensures no arbitrage on market together with the long-range dependence of X same as that of FBM with Hurst parameter H. The no arbitrage condition comes from $\varphi'(0) = 0$ which restricts such X generated to be a semi-MG while the long-range dependence obviously comes from $\lim_{t\to\infty} \varphi''(t) t^{\frac{5}{2}-H} = L \in (0,\infty)$.

We omit the details of the proof here (through integration by parts and changing the order of integration) but present some examples of the convolution kernel such that the reader can understand what we are talking about.

For FBM with Hurst parameter $H < \frac{1}{2}$, the convolution kernel is

$$\varphi(s) = s^{H-\frac{1}{2}}\mathbb{I}_{s>0} \tag{744}$$

so $\varphi'(0)$ is undefined (arbitrage exists) and $\lim_{t\to\infty}\varphi''(t)t^{\frac{5}{2}-H} > 0$ (long-range dependence exists). For BM, the convolution kernel is

$$\varphi \equiv 1 \tag{745}$$

so $\varphi'(0) = 0$ (no arbitrage) and $\lim_{t\to\infty}\varphi''(t)t^{\frac{5}{2}-H} = 0$ (no long-range dependence). Consider a subtly constructed convolution kernel

$$\varphi(s) = (\varepsilon + s^2)^{\frac{2H-1}{4}} \tag{746}$$

for some $\varepsilon > 0$, $\varphi'(0) = 0$ (no arbitrage) and $\lim_{t\to\infty}\varphi''(t)t^{\frac{5}{2}-H} > 0$ (long-range dependence exists) provides a **reasonable model of asset price with long-range dependence**. In short, FBM provides possibility for us to generalize the existing setting of portfolio theory to a more general one with preferable properties added to the original model. Although sometimes FBM itself may not be a perfect fit, careful investigation and modification based on FBM is often rewarding, as shown above.

Reproducing Kernel Hilbert Space (RKHS)

As shown in the last section, the definition of Gaussian white noise and fractional Gaussian white noise both require the existence of a Hilbert (function) space whose inner product of indicators is defined as the covariance kernel. As a result, how to build a Hilbert space correspondent to a given covariance kernel is the problem we want to solve.

Let I denote the index set (often uncountable in our purpose, taken as $\mathbb{R}_+$, the time index set of stochastic process), $c \in \mathbb{R}^{I\times I}$ denotes the **positive definite kernel** on I, i.e. it's symmetric $\forall (i,j) \in I\times I, c_{ij} = c_{ji}$ and positive definite $\forall J \in Fin(I), \theta \in \mathbb{R}^J, \sum_{(i,j)\in J\times J}\theta_i c_{ij}\theta_j \geq 0$. Note that here $Fin(I)$ denotes the set of all finite non-empty subsets of I and $Cou(I)$ denotes the set of all countable non-empty subsets of I. The positive definiteness defined above is actually the semi-positive definiteness we refer to in linear algebra and it's defined through **finite truncation**, i.e. for any finite subset $J = \{j_1, ..., j_n\} \subset I$, the matrix $C \in \mathbb{R}^{n\times n}$ whose (p,q)-th entry is $C_{pq} = c_{j_p,j_q}$ is a semi-positive definite matrix. Intuitively, one can understand c as an infinite-dimensional matrix so $c_{Ij} \overset{def}{=} \{c_{ij} : i \in I\} \in \mathbb{R}^I$ can be understood as the j-th column of this matrix.

Let's start with the **finite-dimensional index set** $|I| < \infty$ and later generalize it to infinite dimension. It's natural that

$$\mathscr{R}(c) \overset{def}{=} span\{c_{Ij} : j \in I\} \tag{747}$$

is the **range of kernel** c with c to be seen as a mapping from $\theta \in \mathbb{R}^I$ to $c\theta \overset{def}{=} \sum_{j\in I}\theta_j c_{Ij} \in \mathbb{R}^I$. The **inner product on** $\mathscr{R}(c)$ is defined as

$$\forall f = c\theta \in \mathscr{R}(c), h = c\eta \in \mathscr{R}(c), \langle f, h\rangle_c \overset{def}{=} \theta^T c\eta \tag{748}$$

which does not depend on the representation of f, h in $\mathscr{R}(c)$. The **reproducing kernel property** is defined as

$$\forall i \in I, \forall f \in \mathscr{R}(c), \langle c_{Ii}, f\rangle_c = f_i \tag{749}$$

with a simple verification for $f = c\theta$ that

$$\langle c_{Ii}, f\rangle_c = e_i^T c\theta = e_i^T f = f_i \tag{750}$$

so $(\mathscr{R}(c), \langle\cdot,\cdot\rangle_c)$ is called **RKHS associated with** c with the norm $||\cdot||_c$ induced by the inner product and $\forall f \notin \mathscr{R}(c), ||f||_c \overset{def}{=} \infty$.

Let id denote the identity operator, then regardless of if c is invertible, the **generalized inverse** is defined through the following steps

$$\theta^{f;n} \overset{def}{=} \left(c + \frac{1}{n}id\right)^{-1} f \tag{751}$$

take the limit

$$\theta^f \overset{def}{=} \lim_{n\to\infty}\theta^{f;n} \tag{752}$$

such that the following theorem holds.

Theorem 26.

$$\forall f \in \mathscr{R}(c), f = c\theta^f \tag{753}$$

in particular, $\forall f \in \mathbb{R}^I, \lim_{n\to\infty} \left\langle \theta^{f;n}, f \right\rangle_{\mathbb{R}^I} = ||f||_c^2$.

Proof. The proof is based on the decomposition $f = c\theta + \eta$ where $c\eta = 0$, i.e. the part in $\mathscr{R}(c)$ and the part in the null space of c and the diagonalization of c. □

Remark. *The definition of generalized inverse is subtle. If the definition of* $\theta^{f;n}$ *is replaced with* $\left(c^2 + \frac{1}{n}id\right)^{-1} cf$, *then taking limit gives the Moore-Penrose inverse. However, this is not what we want since then* $\lim_{n\to\infty} \left\langle \theta^{f;n}, f \right\rangle_{\mathbb{R}^I}$ *is always finite so one cannot distinguish if* $f \in \mathscr{R}(c)$ *or not through the limit.*

For the **general case** where $|I| = \infty$, the idea of construction still comes from finite truncation. Take $\forall J \in Fin(I)$ where $\mathscr{R}(c;J)$ is defined as the linear span of finitely many columns $\{c_{Ij} : j \in J\}$. An inner product $\langle \cdot, \cdot \rangle_{c,J}$ can be defined on $\mathscr{R}(c;J)$ same as above, making $(\mathscr{R}(c;J), \langle \cdot, \cdot \rangle_{c,J})$ an inner product space. Naturally, consider the union of all those finite truncation

$$\mathscr{R}(c;Fin) \overset{def}{=} \bigcup_{J\in Fin(I)} \mathscr{R}(c;J) \tag{754}$$

equipped with the inner product $\langle \cdot, \cdot \rangle_c$ such that $\langle \cdot, \cdot \rangle_{c,J}$ is just the restriction of $\langle \cdot, \cdot \rangle_c$ on $J \in Fin(I)$. Why does such $\langle \cdot, \cdot \rangle_c$ necessarily exist? Because of the consistency of $\langle \cdot, \cdot \rangle_{c,J}$ that if $J \subset Q$ are both elements in $Fin(I)$, then $\mathscr{R}(c;J) \subset \mathscr{R}(c;Q)$ and $\langle \cdot, \cdot \rangle_{c,J}$ is a restriction of $\langle \cdot, \cdot \rangle_{c,Q}$ on $\mathscr{R}(c;J)$. The **reproducing kernel property** still holds on $\mathscr{R}(c;Fin)$ the the Cauchy-Schwarz inequality implies the continuity of linear evaluation functional $h_i : \mathscr{R}(c;Fin) \to \mathbb{R}$ that maps f to f_i, i.e. $h_i \in \mathscr{R}(c;Fin)^*$ is in the dual space.

One last problem for the space $\mathscr{R}(c;Fin)$ is that it's not necessarily complete. However, this is not a big problem since a completion can be made by manually adding the limits of Cauchy sequences into $\mathscr{R}(c;Fin)$ to make it complete. The resulting complete Hilbert space is denoted $(\mathscr{R}(c), \langle \cdot, \cdot \rangle_c)$, called the **RKHS associated with kernel** c. Those are the details with the construction of RKHS. Actually much more theorems can be stated for such RKHS but we do not mention them here. Instead, we provide some examples to see why RKHS is useful.

Consider BM on $[0,1]$ that $\{W_t\}_{t\in[0,1]}$, the information is given as

$$I = [0,1], c_{ts} = t \wedge s \ (\forall t, s \in [0,1]) \tag{755}$$

with c as a positive definite kernel. As a result, its RKHS $(\mathscr{R}(c), \langle \cdot, \cdot \rangle_c)$ exists. If we take $J = \{0\} \in Fin(I)$ as an example,

$$\mathscr{R}(c;J) = span \{f : [0,1] \to \mathbb{R} | f(s) = s \wedge 0 = 0\} = \{f | f \equiv 0\} \tag{756}$$

is a function space whose elements are functions from $[0,1]$ to $\mathbb{R}$. Then it's clear that $\mathscr{R}(c)$ must also be a function Hilbert space as a subset of $\{f | f : [0,1] \to \mathbb{R}\}$. Taking $J = \{j\} \in Fin(I)$ gives

$$\mathscr{R}(c;J) = span \{f_j : [0,1] \to \mathbb{R} | f_j(s) = s \wedge j, j \in [0,1]\} \tag{757}$$

whose elements are functions $f_j = c_{Ij}$ with bounded derivative $f_j'(s) = \mathbb{I}_{[0,j]}(s)$. From those examples, we can already conclude that the functions in $\mathscr{R}(c)$ must pass through he origin. On the other hand, notice that by the reproducing kernel property

$$\forall \varphi \in \mathscr{R}(c), \varphi(0) = 0, \langle \varphi, f_j \rangle_c = \varphi(j) = \int_0^1 \mathbb{I}_{[0,j]}(t)\varphi'(t)\, dt = \int_0^1 f_j'(t)\varphi'(t)\, dt \tag{758}$$

makes use of the indicator structure of the derivative of f_j. Here $\langle \varphi, f_j \rangle_c = \varphi(j)$ is from the reproducing kernel property since $f_j = c_{Ij}$. So far, we can see that $\langle \varphi, f_j \rangle_c$ is just the $L^2([0,1])$ inner product between their derivatives. To make it rigorous, only need to notice that the set of all step functions is dense in $L^2([0,1])$ so

$$\forall \varphi, \psi \in \mathscr{R}(c), \langle \varphi, \psi \rangle_c = \int_0^1 \varphi'(t)\psi'(t)\, dt \tag{759}$$

as a result, **the RKHS of BM** is given by

$$\mathscr{R}(c) = \{f : [0,1] \to \mathbb{R} | f(0) = 0, ||f||_c < \infty\} \tag{760}$$

$$= \left\{f : [0,1] \to \mathbb{R} | f(0) = 0, f' \in L^2([0,1])\right\} \tag{761}$$

it's called the **Cameron-Martin Hilbert space**, a very crucial correspondence!

In portfolio theory, RKHS is typically used to extend the theory from finitely many assets to infinitely many assets. One example is the HJM framework we have mentioned above where there are infinitely many different bonds on the market and one would like to find an infinite-dimensional model for the interest rate. The difficulty lies in building up BM on a infinite-dimensional space and defining stochastic integration on such space. The theory heavily depends on RKHS but we are not presenting it here due to limited scope.

Drawdown Constraint

Drawdown refers to the difference between total wealth process and its running maximum. Considering the similar portfolio optimization and financing problems under the drawdown constraint has significant importance. In real life, the numeraire constraint (total wealth to be strictly positive) is too weak and the investors would withdraw when their total wealth drops to a large extent before their wealth reaches zero. On the other hand, drawdown has something to do with high-watermark provisions that affects fund managers' incentives (typically fund manager receive extra bonus depending on the drawdown). The running maximum of total wealth process X is denoted as X^* as mentioned above and $X^* - X$ is the **absolute drawdown** of X while $\frac{X^*-X}{X^*}$ is the **relative drawdown** of X.

The readers might recall the example we have raised in the section of financing duality regarding X^* where the capital withdrawal $K = F(X_\nu^*)$ is a function of the running sup of the super-MG numeraire. We have used some tricks in that example saying that the integral w.r.t. the path of X_ν^* does not contribute unless X_ν is breaking the running sup, i.e. $X_\nu = X_\nu^*$. If we apply the same trick and go a little bit further, we can see the following Azema-Yor transform.

Theorem 27. ***(Azema-Yor Transform)*** *X is strictly positive continuous semi-MG with $X(0) = 1$ and $F : [1,\infty) \to [1,\infty)$ is increasing and C^1 with $F(1) = 1$. Define*

$$Z \overset{def}{=} F(X^*) - F'(X^*)(X^* - X) \tag{762}$$

as the Azema-Yor transform of X by F, then we have integral representation

$$Z(t) = 1 + \int_0^t F'(X^*(s))\,dX(s) \tag{763}$$

and $Z^* = F(X^*)$, $\frac{Z^*-Z}{Z^*} = \frac{F'(X^*)X^*}{F(X^*)}\frac{X^*-X}{X^*}$.

Proof. We want to apply integration by parts for $\int_0^t F'(X^*(s))\,dX(s)$ since it gives us an integral w.r.t. the path of $F'(X^*(s))$ and the same trick mentioned above applies. However, we hope to see that $F'(X^*(s))$ is of finite variation since then there would be no Ito correction term in the integration by parts formula. Notice that since $F \in C^1$, we can only say that F' is continuous, but not necessarily of finite variation. Naturally, we first prove the result for $F \in C^2$.

When $F \in C^2$, $F' \in C^1$, $F'(X^*(s))$ must have finite variation on any compact time interval. Integration by parts tells us

$$\int_0^t F'(X^*(s))\,dX(s) = X(t)F'(X^*(t)) - F'(1) - \int_0^t X(s)\,dF'(X^*(s)) \tag{764}$$

since $dX^*(s)$ has contribution only on the event $\{X^* = X\}$ (otherwise X^* is flat), $\forall t \geq 0, \int_0^t [X^*(s) - X(s)]\,dF'(X^*(s)) =$

0, so

$$\int_0^t F'(X^*(s))\,dX(s) = X(t)F'(X^*(t)) - F'(1) - \int_0^t X(s)\,dF'(X^*(s)) \tag{765}$$

$$= X(t)F'(X^*(t)) - F'(1) - \int_0^t X^*(s)\,dF'(X^*(s)) + \int_0^t [X^*(s) - X(s)]\,dF'(X^*(s)) \tag{766}$$

$$= Z(t) - F(X^*(t)) + X^*(t)F'(X^*(t)) - F'(1) - \int_0^t X^*(s)\,dF'(X^*(s)) \tag{767}$$

notice that the last three terms comes from the integration by parts of $\int_0^t F'(X^*(s))\,dX^*(s)$, but this integral can be calculate easily as $F(X^*(t)) - F(X^*(0)) = F(X^*(t)) - 1$ so

$$\int_0^t F'(X^*(s))\,dX(s) = Z(t) - F(X^*(t)) + F(X^*(t)) - 1 = Z(t) - 1 \tag{768}$$

proves the representation. In particular, since $X \leq X^*$,

$$Z - F(X^*) = -F'(X^*)(X^* - X) \leq 0 \tag{769}$$

tells us $Z \leq F(X^*)$ so $Z^* \leq F(X^*)$. On the other hand, set

$$\psi(t) \overset{def}{=} \sup\{s \in [0,t] : X(s) = X^*(s)\} \tag{770}$$

as the last time before time t when X is still breaking the running sup. $Z^*(t) \geq Z^*(\psi(t))$ since running sup is increasing but $Z(\psi(t)) = F(X^*(\psi(t))) = F(X^*(t))$ since within time interval $[\psi(t), t]$, X^* must stay flat. As a result, $Z^*(t) \geq F(X^*(t))$ proves $Z^* = F(X^*)$. The relative drawdown of Z is also proved by plugging in the proved conclusions.

For general $F \in C^1$, an approximation argument holds that there exists F_m as a sequence of C^2 increasing functions with $F_m(1) = 1$ such that $F_m' \to F', F_m \to F$ $(m \to \infty)$ uniformly on compact subset of $[1, \infty)$ (e.g. use mollifier). □

Remark. *Azema-Yor transform provides us with a simple construction of Z such that the running sup of Z is just the image of the running sup of X under some incresing C^1 function F. However, in terms of financial interest, such Z is not necessarily a numeraire (strictly positive), so one typically adds a sufficient condition that*

$$\forall x \geq 1, \frac{xF'(x)}{F(x)} < 1 \tag{771}$$

to make sure that Z is also strictly positive. To see this, notice that

$$Z = F(X^*)\left(1 - \frac{F'(X^*)}{F(X^*)}(X^* - X)\right) = F(X^*)\left(1 - \frac{F'(X^*)X^*}{F(X^*)}\frac{X^* - X}{X^*}\right) \tag{772}$$

with the relative drawdown to be always between 0 and 1.

One interesting application of the Azema-Yor transform can be seen in the following exercise.

Lemma 40 (Exercise 4.3)**.** *X is continuous semi-MG with $X_0 = x_0 > 0$ and suppose that $\varphi : [x_0, \infty) \to (0, \infty)$ is continuous with $\int_{x_0}^{\infty} \frac{1}{\varphi(y)}\, dy = \infty$, then derive the strong solution to the SDE (actually exist and is unique)*

$$\begin{cases} dZ(t) = \varphi(Z^*(t))\, dX(t) \\ Z(0) = x_0 \end{cases} \tag{773}$$

Proof. Consider $Y = \frac{X}{x_0}, W = \frac{Z}{x_0}$ so $Y(0) = W(0) = 1$ with $dW(t) = \varphi(x_0 W^*(t))\, dY(t)$. Compare with the Azema-Yor transform, we naturally assume $W^* = F(Y^*)$ for some increasing function $F \in C^1$ such that $F(1) = 1, F \geq 1$, then $dW(t) = F'(Y^*(t))\, dY(t)$. Plug into the SDE to see

$$F'(Y^*(t))\, dY(t) = dW(t) = \varphi(x_0 W^*(t))\, dY(t) = \varphi(x_0 F(Y^*(t)))\, dY(t) \tag{774}$$

set $F'(y) = \varphi(x_0 F(y))$. So we just need to solve this ODE

$$\begin{cases} F'(y) = \varphi(x_0 F(y)) \\ F(1) = 1 \end{cases} \tag{775}$$

for F and then the strong solution is given by

$$Z(t) = x_0 W(t) = x_0 \left[F(Y^*) - F'(Y^*)(Y^* - Y)\right] = x_0 F\left(\frac{X^*}{x_0}\right) - F'\left(\frac{X^*}{x_0}\right)(X^* - X) \tag{776}$$

□

Now we proceed to define the set of **drawdown constrainted numeriares**

$$^{\delta}\mathscr{X} \overset{def}{=} \left\{ X \in \mathscr{X} : \frac{X^* - X}{X^*} < \delta \right\} \tag{777}$$

whose relative drawdown is less than some fixed $\delta \in (0, 1]$. Obviously the set $^{\delta}\mathscr{X}$ is increasing w.r.t. δ and $^{1}\mathscr{X} = \mathscr{X}$ is just the set of all numeraires. The Azema-Yor transform provides us with an important 1-to-1 correspondence between $\mathscr{X}$ and $^{\delta}\mathscr{X}$.

Theorem 28. ***(Characterization of*** $^{\delta}\mathscr{X}$***)*** *$\forall \delta \in (0, 1]$, we have the characterization*

$$^{\delta}\mathscr{X} = \left\{ ^{\delta}X : X \in \mathscr{X} \right\} \tag{778}$$

where

$$^{\delta}X \overset{def}{=} (1 - \delta)(X^*)^{\delta} + \delta X (X^*)^{\delta - 1} \tag{779}$$

Proof. Consider applying Azema-Yor transform for $X \in \mathscr{X}$ by $F(x) = x^\delta \geq 1$ $(x \geq 1)$ increasing, C^1 such that $F(1) = 1$. We get

$$Z = (X^*)^\delta - \delta(X^*)^{\delta-1}(X^* - X), Z^* = (X^*)^\delta \tag{780}$$

to find that $Z = {}^\delta X$. As a result,

$$({}^\delta X)^* = (X^*)^\delta, d\,{}^\delta X(t) = \delta(X^*(t))^{\delta-1}\,dX(t) \tag{781}$$

it's obvious that $\frac{({}^\delta X)^* - {}^\delta X}{({}^\delta X)^*} = \delta\frac{X^* - X}{X^*} < \delta$ so ${}^\delta X \in {}^\delta\mathscr{X}$.

Conversely, apply Azema-Yor transform for $Z \in {}^\delta\mathscr{X}$ by $F(x) = x^{\frac{1}{\delta}} \geq 1$ $(x \geq 1)$ increasing, C^1 such that $F(1) = 1$. We get

$$X = (Z^*)^{\frac{1}{\delta}} - \frac{1}{\delta}(Z^*)^{\frac{1}{\delta}-1}(Z^* - Z) \tag{782}$$

with the property of such transform that

$$X^* = (Z^*)^{\frac{1}{\delta}}, dX(t) = \frac{1}{\delta}(Z^*(t))^{\frac{1}{\delta}-1}\,dZ(t) \tag{783}$$

now we want to verify that $Z = {}^\delta X$. Plug in the definition of ${}^\delta X$ to get

$$\begin{aligned} {}^\delta X &= (1-\delta)(X^*)^\delta + \delta X(X^*)^{\delta-1} && (784)\\ &= (1-\delta)Z^* + \delta\left[Z^* - \frac{1}{\delta}(Z^* - Z)\right] = Z && (785) \end{aligned}$$

the last step is to verify $X \in \mathscr{X}$ but this is obvious since

$$\frac{X^* - X}{X^*} = \frac{1}{\delta}\frac{Z^* - Z}{Z^*} < 1 \tag{786}$$

□

Remark. *It's a useful conclusion that $({}^\delta X)^* = (X^*)^\delta$ and we call ${}^\delta X$ the drawdown constrainted numeraire* ***induced by*** *$X \in \mathscr{X}$, which is actually a convex combination of $(X^*)^\delta$ and $X(X^*)^{\delta-1}$.*

After introducing the basic settings, we introduce the portfolio and consider ${}^\delta X_\pi \in {}^\delta\mathscr{X}$, the drawdown constrainted numeriare induced by X_π. A natural question to ask is that if there exists a portfolio δ_π such that it directly satisfies the drawdown constraint ${}^\delta X_\pi = X_{\delta_\pi}$. The answer to this question is provided in the theorem below.

Theorem 29. $\exists \delta_\pi \overset{def}{=} \left(\frac{1-D_\pi}{\frac{1}{\delta}-D_\pi}\right)\pi$ *such that ${}^\delta X_\pi = X_{\delta_\pi}$ where*

$$D_\pi \overset{def}{=} \frac{X_\pi^* - X_\pi}{X_\pi^*} \tag{787}$$

is the relative drawdown of the numaraire generated by portfolio π.

Proof. By the definition of ${}^{\delta}X_\pi$, ${}^{\delta}X_\pi = (1-\delta D_\pi)(X_\pi^*)^\delta$, by Azema-Yor,

$$d\ {}^{\delta}X_\pi(t) = \delta(X_\pi^*(t))^{\delta-1}\,dX_\pi(t) \tag{788}$$

so

$$\frac{1}{{}^{\delta}X_\pi(t)}d\ {}^{\delta}X_\pi(t) = \frac{\delta(X_\pi^*(t))^{\delta-1}\,dX_\pi(t)}{{}^{\delta}X_\pi(t)} \tag{789}$$

but since $X_\pi = \mathcal{E}\left(R_\pi\right)$, $\frac{1}{X_\pi(t)}\,dX_\pi(t) = \,dR_\pi(t)$ so

$$\frac{1}{{}^{\delta}X_\pi(t)}d\ {}^{\delta}X_\pi(t) = \frac{\delta(X_\pi^*(t))^{\delta-1}X_\pi(t)}{{}^{\delta}X_\pi(t)}\,dR_\pi(t) \tag{790}$$

similarly, $\frac{1}{X_{\delta_\pi}(t)}\,dX_{\delta_\pi}(t) = \,dR_{\delta_\pi}(t)$ so if ${}^{\delta}X_\pi = X_{\delta_\pi}$ holds then

$$\frac{\delta(X_\pi^*(t))^{\delta-1}X_\pi(t)}{{}^{\delta}X_\pi(t)}\,dR_\pi(t) = \,dR_{\delta_\pi}(t) \tag{791}$$

simplify further to see

$$\frac{\delta(X_\pi^*(t))^{\delta-1}X_\pi(t)}{{}^{\delta}X_\pi(t)}\pi^T\,dR(t) = \delta_\pi^T\,dR(t) \tag{792}$$

gives $\delta_\pi = \left(\frac{1-D_\pi}{\frac{1}{\delta}-D_\pi}\right)\pi$ since

$$\frac{\delta(X_\pi^*)^{\delta-1}X_\pi}{{}^{\delta}X_\pi} = \frac{\delta(X_\pi^*)^{\delta-1}(1-D_\pi)X_\pi^*}{(1-\delta D_\pi)(X_\pi^*)^\delta} = \frac{\delta(1-D_\pi)}{1-\delta D_\pi} \tag{793}$$

an easy verification proves that ${}^{\delta}X_\pi = X_{\delta_\pi}$ does hold. □

Remark. *It's obvious that*

$$D_{\delta_\pi} = \frac{X_{\delta_\pi}^* - X_{\delta_\pi}}{X_{\delta_\pi}^*} = \frac{({}^{\delta}X_\pi)^* - {}^{\delta}X_\pi}{({}^{\delta}X_\pi)^*} = \delta D_\pi \tag{794}$$

as presented in the proof of the characterization of ${}^{\delta}\mathscr{X}$*. As a result,* δ_π *is always putting* $\frac{\delta-D_{\delta_\pi}}{1-D_{\delta_\pi}} \in (0,\delta]$ *proportion of wealth in the fund* X_π *and* $\frac{1-\delta}{1-D_{\delta_\pi}} \in [1-\delta,1)$ *proportion of wealth in the money market. Such* δ_π *is called the* ***drawdown constrained portfolio associated with*** π*.*

In other words, if there is a fund on the market generated by the portfolio π *and a* ***mutual fund manager wants to make sure that he invests in such fund but the investors won't face a relative drawdown larger than*** δ*, he/she shall always* ***just invest*** $\frac{1-D_\pi}{\frac{1}{\delta}-D_\pi} = \frac{\delta-D_{\delta_\pi}}{1-D_{\delta_\pi}} \in (0,\delta]$ ***proportion of the total wealth of the mutual fund into this fund*** *and the rest into the money market. Interestingly, the proportion invested in the fund shall never exceed* δ *(the risk of drawdown totally comes from the fund) and it's decreasing in* D_π *(higher relative drawdown, worse current performance compared to the best in history, less investment in the fund). In addition,* D_π *is* ***observable*** *(not depending on all the assets in the market) so this provides a guideline of how mutual fund manager shall deal with a pre-planned objective for drawdown.*

Local-MG Numeraire Under Drawdown Constraint

Since we have already considered the drawdown constrainted numeraire induced by X_π, it's natural to see what will happen if π is taken as ν, the local-MG numeraire portfolio. Of course, we assume that the market is viable so ν exists, besides we also assume $X_\nu(\infty)=\infty$. The first thing we can say about X_ν is that its relative drawdown D_ν has a fluctuating asymptotic behavior.

Theorem 30. $\liminf_{t\to\infty} D_\nu(t)=0, \limsup_{t\to\infty} D_\nu(t)=1.$

Proof. It's obvious that since $X_\nu(\infty)=\infty$, D_ν has liminf zero. The proof of limsup requires us to play with two stopping times

$$\tau_m=\inf\left\{t:\log X_\nu(t)=m\right\}, \sigma_m=\inf\left\{t>\tau_{m-1}:D_\nu(t)=1-\frac{1}{m}\right\} \tag{795}$$

both sequences of stopping times are increasing and the conclusion holds from

$$\forall m, \mathbb{P}\left(\sigma_m\le\tau_m|\mathscr{F}(\tau_{m-1})\right)\ge\frac{1}{m} \tag{796}$$

(derived from the special structure of $\frac{X_\nu^*}{X_\nu}$) and the conditional Borel-Cantelli lemma (it implies $\mathbb{P}\left(\tau_{m-1}\le\sigma_m\le\tau_m\ i.o.\right)=1$). As a result, since $\tau_m<\infty\ a.s.$ and when $\sigma_m<\infty$ then $D_\nu(\sigma_m)=1-\frac{1}{m}$, there exists infinitely many times when D_ν is close enough to 1, so limsup is 1.

□

Remark. *This fact tells us that ν **does not behave very nice in terms of relative drawdown**. Although ν has a lot of optimality properties and can be considered the best portfolio on the market, the relative drawdown may get very high. This is because bounding the drawdown is a portfolio insurance but the super-MG numeraire portfolio has optimality properties because it's closely related to the market. It's impossible to get a high return and a portfolio insurance at the same time.*

*From this point of view, it's natural to expect that ${}^\delta X_\nu$ **loses its super-MG numeraire property within** ${}^\delta\mathscr{X}$. A counterexample is provided in example 4.8.*

The following theorem provides result for the long-term behavior of ${}^\delta X_\nu$.

Theorem 31.

$$\forall\delta\in(0,1], \lim_{t\to\infty}\frac{{}^\delta X_\pi(t)}{{}^\delta X_\nu(t)}=\lim_{t\to\infty}\frac{{}^\delta X_\pi^*(t)}{{}^\delta X_\nu^*(t)}=\left(\lim_{t\to\infty}\frac{X_\pi(t)}{X_\nu(t)}\right)^\delta \tag{797}$$

where all limits are well-defined under almost surely convergence.

Proof. We provide the sketch of the proof. By the Azema-Yor transform, ${}^\delta X_\pi=(1-\delta D_\pi)(X_\pi^*)^\delta$ so the ratio is

$$\frac{{}^\delta X_\pi(t)}{{}^\delta X_\nu(t)}=\left(\frac{X_\pi^*(t)}{X_\nu^*(t)}\right)^\delta\frac{1-\delta D_\pi}{1-\delta D_\nu} \tag{798}$$

again by Azema-Yor transform, we have proved $(X_\pi^*)^\delta = {}^\delta X_\pi^*$ so the ratio can also be written as

$$\frac{{}^\delta X_\pi(t)}{{}^\delta X_\nu(t)} = \frac{{}^\delta X_\pi^*(t)}{{}^\delta X_\nu^*(t)} \frac{1-\delta D_\pi}{1-\delta D_\nu} \tag{799}$$

proves the second equality. For the first equality, just prove that $\frac{1-\delta D_\pi}{1-\delta D_\nu} \to 1$ $(t \to \infty)$. In order to prove this limit, notice that

$$\left| \frac{1-\delta D_\pi}{1-\delta D_\nu} - 1 \right| \leq \delta \frac{|D_\nu - D_\pi|}{1-D_\nu} \tag{800}$$

it suffices to prove $\frac{|D_\nu - D_\pi|}{1-D_\nu} \to 0$ $(t \to \infty)$, implied by $\frac{1-D_\pi}{1-D_\nu} \to 1$ $(t \to \infty)$. As a result, one just need to prove the equality for $\delta = 1$.

When $\delta = 1$, the proof is standard with the trick of two stopping time presented above

$$\psi_\nu(t) = \sup\{s \in [0,t] : X_\nu(s) = X_\nu^*(s)\}, \psi_\pi(t) = \sup\{s \in [0,t] : X_\pi(s) = X_\pi^*(s)\} \tag{801}$$

with $X_\nu^*(t) = X_\nu(\psi_\nu(t)) = X_\nu^*(\psi_\nu(t))$ and $\psi_\nu(t) \overset{a.s.}{\to} \infty$ $(t \to \infty)$. □

Lemma 41 (Exercise 4.10)**.** *Under those assumptions, for any portfolio π,*

$$\limsup_{T\to\infty} \frac{\log {}^\delta X_\pi(T)}{G(T)} \leq \delta = \limsup_{T\to\infty} \frac{\log {}^\delta X_\nu(T)}{G(T)} \tag{802}$$

Proof. We have proved similar conclusion for super-MG numeraire without drawdown constraint that

$$\limsup_{T\to\infty} \frac{\log X_\pi(T)}{G(T)} \leq 1 = \limsup_{T\to\infty} \frac{\log X_\nu(T)}{G(T)} \tag{803}$$

recalling that super-MG numeraire achieves optimal long-term growth. The connection between X_π and ${}^\delta X_\pi$ is provided by the Azema-Yor transform

$$\log {}^\delta X_\pi = \log\left[(1-\delta)(X_\pi^*)^\delta + \delta X_\pi (X_\pi^*)^{\delta-1}\right] \tag{804}$$

$$= \delta \log X_\pi^* + \log\left(1 - \delta + \delta \frac{X_\pi}{X_\pi^*}\right) \tag{805}$$

it's then clear that

$$\limsup_{T\to\infty} \frac{\log {}^\delta X_\pi(T)}{G(T)} = \delta \limsup_{T\to\infty} \frac{\log X_\pi^*(T)}{G(T)} \leq \delta \tag{806}$$

we only have to argue the last inequality here.

When $\pi = \nu$, it's clear that such inequality holds since $X_\nu(\infty) = \infty$ and a same trick with stopping time $\psi_\nu(t) = \sup\{s \in [0,t] : X_\nu(s) = X_\nu^*(s)\}$ applies. The point is that since X_ν goes to ∞, when time is large enough we expect to see that $X_\nu = X_\nu^*$ always holds. As a result, $\limsup_{T\to\infty} \frac{\log X_\nu^*(T)}{G(T)} = \limsup_{T\to\infty} \frac{\log X_\nu(T)}{G(T)} = 1$. For general π, on the event $\{X_\pi(\infty) < \infty\}$, since $G(\infty) = \infty$ we have $\limsup_{T\to\infty} \frac{\log X_\pi^*(T)}{G(T)} = 0$, otherwise a same argument concludes the proof. □

Remark. *When one adds drawdown constraint such that relative drawdown does not exceed* δ*,* ${}^{\delta}X_{\nu}$ ***has the optimal long-term growth property within*** ${}^{\delta}\mathscr{X}$*, which is not surprising since super-MG numeraire is long-term growth optimal. However, the best one can achieve in terms of maximal long-term growth from investment is just* δ *of the unconstrained optimal long-term growth. This is the* ***price to pay for the portfolio insurance****, the better portfolio insurance one gets, the less maximal long-term growth one gets.*

Interestingly, since $\frac{{}^{\delta}X_{\pi}}{{}^{\delta}X_{\nu}}(\infty) = \left(\frac{X_{\pi}}{X_{\nu}}(\infty)\right)^{\delta}$*, we see that*

$$\forall 0 < \delta \leq \varepsilon \leq 1, \left|\frac{{}^{\delta}X_{\pi}}{{}^{\delta}X_{\nu}}(\infty) - 1\right| = \left|\left(\frac{X_{\pi}}{X_{\nu}}(\infty)\right)^{\delta} - 1\right| \tag{807}$$

$$\leq \left|\left(\frac{X_{\pi}}{X_{\nu}}(\infty)\right)^{\varepsilon} - 1\right| \tag{808}$$

$$= \left|\frac{{}^{\varepsilon}X_{\pi}}{{}^{\varepsilon}X_{\nu}}(\infty) - 1\right| \tag{809}$$

since the long-term growth behavior tells us

$$\lim_{T\to\infty} \frac{1}{\Gamma_{\nu}^{\pi}(T)} \log \frac{X_{\pi}(T)}{X_{\nu}(T)} = -1 \tag{810}$$

resulting in $\frac{X_{\pi}}{X_{\nu}}(\infty) \leq 1$*. An easy interpretation is that* ***enforcing harsher drawdown constraints results in a reduction of the long-term difference in the performance of the drawdown constrained process against the long-term growth optimum****.*

Time of Maximum

For the last part of drawdown constraints, let's turn to **the failure of the super-MG property of $^{\delta}X_\nu$ within $^{\delta}\mathscr{X}$** and talk about how to recover such super-MG property. Firstly, let's look at a counterexample saying that $\exists \eta, X_\eta \in {}^{\delta}\mathscr{X}$ with $\frac{X_\eta}{^{\delta}X_\nu}$ fails to be a super-MG.

For fixed $\delta \in (0,1)$ consider the stopping time when relative drawdown of X_ν first reaches $\frac{\delta}{2}$

$$T \overset{def}{=} \inf\left\{t > 0 : D_\nu(t) = \frac{\delta}{2}\right\} \tag{811}$$

since the relative drawdown of super-MG numeraire fluctuates, i.e. $\liminf_{t\to\infty} D_\nu(t) = 0, \limsup_{t\to\infty} D_\nu(t) = 1$, we know $T < \infty$ *a.s.*. The portfolio $\eta(t) = \nu(t)\mathbb{I}_{[0,T]}(t)$ sticks to ν before time T but stops investing after time T (intuitively, when the relative drawdown D_ν is large enough, X_ν is not performing as well as it used to be, so one stops investing). It's obvious that $X_\eta \in {}^{\delta}\mathscr{X}$. Now if $\frac{X_\eta}{^{\delta}X_\nu}$ is a super-MG, it's non-negative so optional stopping theorem tells us

$$\mathbb{E}\frac{X_\eta(T)}{^{\delta}X_\nu(T)} \le 1 \tag{812}$$

on the other hand, when the time $t \le T$, η is the same as ν so $\frac{X(t\wedge T)}{X_\eta(t\wedge T)}$ is a super-MG for $\forall X \in {}^{\delta}\mathscr{X}$. Take X as $^{\delta}X_\nu \in {}^{\delta}\mathscr{X}$ and apply optional stopping theorem again to get

$$\mathbb{E}\frac{^{\delta}X_\nu(T)}{X_\eta(T)} \le 1 \tag{813}$$

however, by Jensen's inequality for any strictly positive r.v. Y, $\mathbb{E}\frac{1}{Y} \ge \frac{1}{\mathbb{E}Y}$ the equality is attained iff Y is almost surely constant. Combine two equations above to see $\mathbb{E}\frac{X_\eta(T)}{^{\delta}X_\nu(T)} = 1, X_\eta(T) = {}^{\delta}X_\nu(T)$ *a.s.* leading to $X_\nu(T) = {}^{\delta}X_\nu(T)$ *a.s.*. Since $\frac{^{\delta}X_\nu(t\wedge T)}{X_\nu(t\wedge T)}$ is non-negative local MG, thus a super-MG with $\forall t > 0, \mathbb{E}\frac{^{\delta}X_\nu(t\wedge T)}{X_\nu(t\wedge T)} = 1$, we know $\{X_\nu(t)\}_{t\in[0,T]} \equiv \left\{^{\delta}X_\nu(t)\right\}_{t\in[0,T]}$ (super-MG with constant expectation function is a MG and now this MG takes value 1 at ∞). However, this holds only when $\delta = 1$, a contradiction!

The counterexample is constructed based on the intuition that if one **invest following the super-MG numeraire portfolio ν but stops investing when the relative drawdown is higher than some value**, $^{\delta}X_\nu$ cannot be ensured to be superior to the numeraire generated by this portfolio. The key point here is that before stopping η still shares the optimality of ν.

Despite the general failure of super-MG property, the definition of time of maximum allows such property to hold along a sequence of stopping times. Stopping time τ is called **time of maximum for local-MG numeraire** X_ν if $X_\nu(\tau) = X_\nu^*(\tau)$ holds on the event $\{\tau < \infty\}$. In other words, whenever τ is finite, X_ν is always breaking its running sup at time τ. Since $^{\delta}X_\nu^* = (X_\nu^*)^\delta$, whenever τ whenever X_ν is breaking its running sup at time τ, $^{\delta}X_\nu$ is also breaking its running sup at time τ. As a result, **time of maximum for X_ν is the same as time of maximum for $^{\delta}X_\nu$.**

Theorem 32. *Fix $\delta \in (0,1]$ and a portfolio π, then for any two times of maximum $\sigma \le \tau$, we have*

$$\mathbb{E}\left[\frac{{}^\delta X_\pi}{{}^\delta X_\nu}(\tau)\Big|\mathscr{F}(\sigma)\right] \le \frac{{}^\delta X_\pi}{{}^\delta X_\nu}(\sigma)\ a.s. \tag{814}$$

Proof. First assume $\sigma = 0, \tau < \infty$ *a.s.*, by Azema-Yor transform,

$${}^\delta X_\pi(\tau) = (1-\delta)(X_\pi^*(\tau))^\delta + \delta X_\pi(\tau)(X_\pi^*(\tau))^{\delta-1} \tag{815}$$

now τ as time of maximum of both X_ν and ${}^\delta X_\nu$ tells us (the following equation is the key reason why super-MG property holds at times of maximum)

$${}^\delta X_\nu(\tau) = {}^\delta X_\nu^*(\tau) = (X_\nu^*(\tau))^\delta = (X_\nu(\tau))^\delta \tag{816}$$

consider the relative wealth process evaluated at time τ

$$\begin{aligned}\frac{{}^\delta X_\pi(\tau)}{{}^\delta X_\nu(\tau)} &= (1-\delta)\frac{(X_\pi^*(\tau))^\delta}{{}^\delta X_\nu(\tau)} + \delta\frac{X_\pi(\tau)(X_\pi^*(\tau))^{\delta-1}}{{}^\delta X_\nu(\tau)} &(817)\\ &= (1-\delta)\frac{(X_\pi^*(\tau))^\delta}{(X_\nu^*(\tau))^\delta} + \delta\frac{X_\pi(\tau)(X_\pi^*(\tau))^{\delta-1}}{X_\nu(\tau)(X_\nu^*(\tau))^{\delta-1}} &(818)\\ &= (1-\delta)\left(\frac{X_\pi^*(\tau)}{X_\nu^*(\tau)}\right)^\delta + \delta\frac{X_\pi(\tau)}{X_\nu(\tau)}\left(\frac{X_\pi^*(\tau)}{X_\nu^*(\tau)}\right)^{\delta-1} &(819)\end{aligned}$$

noticing that the RHS can be connected to the Azema-Yor transform of $\frac{X_\pi}{X_\nu}$ by $F(x) = x^\delta$ that

$$(1-\delta)\left(\frac{X_\pi^*(\tau)}{X_\nu^*(\tau)}\right)^\delta + \delta\frac{X_\pi(\tau)}{X_\nu(\tau)}\left(\frac{X_\pi^*(\tau)}{X_\nu^*(\tau)}\right)^{\delta-1} \le (1-\delta)\left[\left(\frac{X_\pi}{X_\nu}\right)^*(\tau)\right]^\delta + \delta\frac{X_\pi(\tau)}{X_\nu(\tau)}\left[\left(\frac{X_\pi}{X_\nu}\right)^*(\tau)\right]^{\delta-1} \tag{820}$$

since $\frac{X_\pi^*(\tau)}{X_\nu^*(\tau)} \le \left(\frac{X_\pi(\tau)}{X_\nu(\tau)}\right)^*$ so we conclude

$$\frac{{}^\delta X_\pi(\tau)}{{}^\delta X_\nu(\tau)} \le {}^\delta\left(\frac{X_\pi}{X_\nu}\right)(\tau) \tag{821}$$

since ${}^\delta\left(\frac{X_\pi}{X_\nu}\right)$ still bears the structure as a local MG (by Azema-Yor transform it's a stochastic integral w.r.t. the trajectory of $\frac{X_\pi}{X_\nu}$, a local MG so itself is also local-MG) and it's non-negative thus a super-MG. By optional stopping theorem,

$$\mathbb{E}\frac{{}^\delta X_\pi(\tau)}{{}^\delta X_\nu(\tau)} \le \mathbb{E}^\delta\left(\frac{X_\pi}{X_\nu}\right)(\tau) \le 1 \tag{822}$$

When τ has positive probability of taking ∞, truncating with stopping time τ_m as the first time $\log X_\nu$ reaches m, i.e. considering $\tau \wedge \tau_m$ proves the conclusion (Fatou and notice that $\tau_m \nearrow \infty (m \to \infty)$ under the assumption $X_\nu(\infty) = \infty$).

When σ is no longer constantly 0, construct $V \in {}^\delta\mathscr{X}$ with ${}^\delta X_\nu, {}^\delta X_\pi \in {}^\delta\mathscr{X}$ as in the lemma below and consider $\frac{V}{{}^\delta X_\nu}(\tau)$ to get $\mathbb{E}\frac{V}{{}^\delta X_\nu}(\tau) \le 1$ (use the conclusion proved above when $\sigma \equiv 0$) proves the conclusion. ☐

The lemma (switching lemma) used here in the proof enables us to build $V \in {}^{\delta}\mathscr{X}$ out of $X, Z \in {}^{\delta}\mathscr{X}$ such that V is identical to X before the time of maximum for X and V switches to Z after such a time.

Lemma 42 (Exercise 4.14)**.** *σ is the time of maximum for X while $X, Z \in {}^{\delta}\mathscr{X}, \Lambda \in \mathscr{F}(\sigma)$, then*

$$V(t) \overset{def}{=} X(t)\mathbb{I}_{[0,\sigma)}(t) + \left(X(t)\mathbb{I}_{\Omega-\Lambda}(\omega) + \frac{X(\sigma)}{Z(\sigma)}Z(t)\mathbb{I}_{\Lambda}(\omega)\right)\mathbb{I}_{[\sigma,\infty)}(t) \tag{823}$$

satisfies $V \in {}^{\delta}\mathscr{X}$ where ω is the sample point and Ω is the sample space.

Proof. Let's fix sample point ω and prove the result. When $\omega \in \Omega - \Lambda$, $V \equiv X$ so it's proved. When $\omega \in \Lambda$, V is the same as X before time σ with $V(\sigma) = X(\sigma)$ and $V = \frac{X(\sigma)}{Z(\sigma)}Z$ after time σ (WLOG assume $\sigma < \infty$ otherwise it's trivial). It's obvious that $\forall t \le \sigma, D_V(t) < \delta$ so the only problem lies in checking D_V when $t > \sigma$.

Now for $t > \sigma$, $V^*(t) = \max\left\{X^*(\sigma), \frac{X(\sigma)}{Z(\sigma)}\sup_{\sigma\le s\le t} Z(s)\right\} = X(\sigma)\max\left\{1, \frac{\sup_{\sigma\le s\le t} Z(s)}{Z(\sigma)}\right\} = X(\sigma)\frac{\sup_{\sigma\le s\le t} Z(s)}{Z(\sigma)}$ since $X(\sigma) = X^*(\sigma)$. It's direct that

$$D_V(t) = 1 - \frac{V(t)}{V^*(t)} = 1 - \frac{Z(t)}{\sup_{\sigma\le s\le t} Z(s)} \tag{824}$$

since $Z \in {}^{\delta}\mathscr{X}$, $\frac{Z(t)}{\sup_{\sigma\le s\le t} Z(s)} \ge \frac{Z(t)}{Z^*(t)} > 1 - \delta$ so $D_V(t) < \delta$ concludes the proof. □

Remark. *This theorem tells us that* ***super-MG property of ${}^{\delta}X_\nu$ holds within ${}^{\delta}\mathscr{X}$ along an increasing sequence of time of maximum for*** *X_ν. There are actually more facts to state w.r.t. the structure of ${}^{\delta}X_\nu$. In the counterexample at the beginning of this section, we have shown that the super-MG property of ${}^{\delta}X_\nu$ fails since $\frac{X_\eta}{{}^{\delta}X_\nu}$ cannot be a super-MG for $\eta = \nu\mathbb{I}_{[0,T]}$. Actually* ***there does not exists any $\hat{X} \in {}^{\delta}\mathscr{X}$ such that $\hat{X}$ has super-MG property within ${}^{\delta}\mathscr{X}$ on the whole time horizon****. This is because of the theorem above telling us that ${}^{\delta}X_\nu$ is the only hopeful numeraire in ${}^{\delta}\mathscr{X}$ for which super-MG property holds.*

With $\eta = \nu\mathbb{I}_{[0,T]}$ and T as the first time D_ν reaches $\frac{\delta}{2}$, it's obvious that within time interval $[0, T]$, X_η is the same as X_ν so super-MG property holds. Actually ***such $X_\eta \in {}^{\delta}\mathscr{X}$ is the unique process for which super-MG property within ${}^{\delta}\mathscr{X}$ holds over*** *$[0, T]$. Otherwise if $\hat{X} \in {}^{\delta}\mathscr{X}$ is another process with such property, $\frac{\hat{X}}{X_\eta}$ is super-MG and $\frac{X_\eta}{\hat{X}}$ is super-MG over $[0, T]$, resulting in $\{X_\eta(t)\}_{t\in[0,T]} = \left\{\hat{X}(t)\right\}_{t\in[0,T]}$ (refer to the uniqueness of super-MG numeraire proved in the previous sections).*

However, if one relaxes a little bit to the time interval $[0, \tau]$ where τ is the first time X_ν reaches $X_\nu^*(T)$ after time T

$$\tau \overset{def}{=} \inf\{t > T : X_\nu(t) = X_\nu^*(T)\} \tag{825}$$

then ***super-MG property within ${}^{\delta}\mathscr{X}$ never holds over*** *$[0, \tau]$* ***for any*** *$\hat{X} \in {}^{\delta}\mathscr{X}$. One can consider portfolio η and δ_ν (drawdown constrainted portfolio associated with ν). We have proved $\mathbb{E}\frac{X_\eta(T)}{X_{\delta_\nu}(T)} > 1$ (notice ${}^{\delta}X_\nu = X_{\delta_\nu}$) but from the theorem above*

$$\mathbb{E}\left[\frac{X_\eta}{X_{\delta_\nu}}(\tau)\right] = \mathbb{E}\left[\frac{X_\eta}{{}^{\delta}X_\nu}(\tau)\right] \le 1 \tag{826}$$

since τ is a time of maximum for X_ν (draw a picture to see this fact), a contradiction! Even if T and τ are generally *not that far away from each other, the existence super-MG property changes drastically. As a result, the super-MG property on different time intervals is subtle and requires careful calculations.*

Asset Price as Semi-MG

As shown in the section of FBM, arbitrage exists if there is only one asset on the market whose price process is an FBM, not a semi-MG. We naturally ask if it's necessary to restrict the price process to be a semi-MG and it turns out that the loss of semi-MG property results in the failure of weak viability.

Assume S is the price process of n assets on the market but not necessarily a semi-MG with the investment strategy still formed as θ. However, stochastic integral w.r.t. S is not well-defined so θ must have a special form

$$\forall i \in \{1,2,...,n\}, \theta_i(t) = \sum_{j=1}^{m} \eta_{ji} \mathbb{I}_{(\tau_{j-1},\tau_j]}(t) \tag{827}$$

with $\tau_1, ..., \tau_m$ as a strictly increasing sequence of stopping times and $\eta_j \in \mathscr{F}(\tau_{j-1})$ taking values in $\mathbb{R}^n$. With this special form as a **simple trading strategy**, the wealth process generated by such strategy is

$$X(t) = 1 + \int_0^t \sum_{i=1}^{n} \theta_i(s)\, dS_i(s) \overset{def}{=} 1 + \sum_{j=1}^{m}\sum_{i=1}^{n} \eta_{ji}[S_i(t\wedge\tau_j) - S_i(t\wedge\tau_{j-1})] \tag{828}$$

with given initial wealth $x > 0$ and capital withdrawal $K \in \mathscr{K}$, the wealth process is

$$X(t;x,\theta,K) = x + \sum_{j=1}^{m}\sum_{i=1}^{n} \eta_{ji}[S_i(t\wedge\tau_j) - S_i(t\wedge\tau_{j-1})] - K(t) \tag{829}$$

does not actually rely on the stochastic integral (just a notation). Here we not only requires X to be strictly positive, but also **prohibit short-selling**, i.e. $\forall i \in \{1,2,...,n\}, \theta_i \geq 0$ and **prohibit borrowing**, i.e. $X \geq \sum_{i=1}^n \theta_i S_i$. Since viability is defined through financeable capital withdrawal, we naturally define

$$\mathscr{K}_\Delta(x) \overset{def}{=} \left\{ K \in \mathscr{K} : \exists\theta, \forall i \in \{1,2,...,n\}, \theta_i \geq 0, s.t.\ X(\cdot;x,\theta,K) \geq \sum_{i=1}^{n} \theta_i S_i \right\} \tag{830}$$

as **the set of all capital withdrawal that is financeable with initial wealth x under some long-only non-borrowing investment strategy**. It's obvious that $\mathscr{K}_\Delta(x) \subset \mathscr{K}(x)$ and if we introduce $\mathscr{X}_\Delta$ as the set of wealth process starting from initial wealth 1 under simple long-only non-borrowing investment strategy, then $\mathscr{X}_\Delta \subset \mathscr{X}$.

Weak viability is defined similarly through the minimal initial wealth needed to finance some given $K \in \mathscr{K}_\Delta(x)$

$$x_\Delta(K) \overset{def}{=} \inf\{x > 0 : K \in \mathscr{K}_\Delta(x)\} \tag{831}$$

a market is **weakly viable** if $x_\Delta(K) = 0$ implies $K \equiv 0$. If the cumulative withdrawal stream can be financed by a simple long-only non-borrowing investment strategy starting from arbitrarily small initial wealth, then this stream has to be identically equal to zero. We list the important theorems below without providing the proof.

Theorem 33. *Market is weakly viable iff*

$$\forall T > 0, \lim_{m\to\infty} \sup_{X\in\mathscr{X}_\Delta} \mathbb{P}\left(X(T) > m\right) = 0 \tag{832}$$

this is an extension of the "boundedness in probability" characterization of market viability.

Theorem 34. ***(Bichteler-Dellacherie)*** *With strictly positive adapted continuous asset price S, the following statements are equivalent:*

(1): Market is weakly viable.

(2): There exists strictly positive super-MG Y such that $\forall i \in \{1, 2, ..., n\}, YS_i$ is super-MG.

(3): $\forall i \in \{1, 2, ..., n\}, S_i$ is semi-MG.

The theorem justifies that **under very weak non-arbitrage assumptions, asset prices must be semi-MG**. Such Y is called the super-MG deflator and it turns out that $Y = \frac{1}{X_\nu}$ if the super-MG numeraire X_ν exists.

Decentralized Finance (DeFi) and Blockchain Technology

Overview

Blockchain is a sequence of blocks maintained by an achieving consensus in a peer-to-peer network. Most importantly, it's decentralized and everyone can join so blockchain can be very large. Once a transaction is written in a block, it's immutable since blockchain is designed such that the reverse transaction cannot be done. It's incentive compatible (people believe in the value of cryptocurrency). Consensus protocol is the mechanism to make all the nodes agree on a common data history. Three dimensions: efficiency (throughput), decentralization (fair distribution of the accounting right), security (resistance to attacks).

Application of blockchain includes cryptocurrency (transaction anonymity, don't need trusted third party like banks). However, as long as bitcoin is traded with dollars, it's related to currency in reality so that the bank keeps record of this. People are trying to trade bitcoins without going through the bank so that the trading is exactly anonymous (maybe through the collapse of quantum states?).

Consensus Protocol

Consensus protocol is an algorithm that allows all the nodes in a distributed blockchain network to agree on a common data history. Simply speaking, it's like the rules of a game all participants shall agree upon. The protocol is typically constructed based on scarce resources within the network, e.g., bandwidth, computational power and storage.

There are two main types of consensus protocols: voting-based and leader-based. The voting-based protocol is democratic since the decision is based on the outcomes of several rounds of voting. However, this protocol is not efficient if the network is large in size. The leader-based protocol selects a group of leaders based on some certain criterion and only the leaders make a decision. This is the state-of-the-art consensus protocol and the leader selection criterion is based on computational power (proof-of-work), storage (proof-of-spacetime), bandwidth (proof-of-interaction) and tokens (proof-of-stake). The most often used proof-of-work protocol select the leaders based on computational effort to append the next block. However, leader-based protocol might suffer from DDoS attacks (to crash the current leader that results in a new leader) and there might be ethical issues with the leader (e.g. the leader might not put his rival's transactions into the block or put inaccurate timestamps).

Proof-of-work

In blockchain architecture, a **block** consists of a header and a list of transactions that represents the information recorded through the blockchain. The header includes the timestamp of creation of the block, the block height (index inside the blockchain), the hash of the block, and the hash of the previous block (to indicate the parent node). The hash is produced through a hash function $h : \{0,1\}^* \to \{0,1\}^d$, mapping a bit sequence of arbitrary size to a bit sequence of a fixed size. We hope to have a hash function that is deterministic, efficient to compute, one-way (cannot recover the original information from the hash value), collision-resistant (hard to find two messages sharing the same hash value), and chaotic (two similar messages do not share similar hash values). One of the hash functions

```
Block Hash: 1fc23a429aa5aaf04d17e9057e03371f59ac8823b1441798940837fa2e318aaa
Block Height: 0
Time:2022-02-25 12:42:04.560217
Nonce:0
Block data: [{'sender': 'Coinbase', 'recipient': 'Satoshi', 'amount': 100, 'fee': 0}, {'sen
der': 'Satoshi', 'recipient': 'Pierre-O', 'amount': 5, 'fee': 2}]
Previous block hash: 0
Mined: False
--------------
```

Figure 1: An unmined block

```
Block Hash: 0869032ad6b3e5b86a53f9dded5f7b09ab93b24cd5a79c1d8c81b0b3e748d226
Block Height: 0
Time:2022-02-25 13:41:48.039980
Nonce:2931734429
Block data: [{'sender': 'Coinbase', 'recipient': 'Satoshi', 'amount': 100, 'fee': 0}, {'sen
der': 'Satoshi', 'recipient': 'Pierre-O', 'amount': 5, 'fee': 2}]
Previous block hash: 0
Mined: True
--------------
```

Figure 2: Caption

satisfying those properties is the SHA-256 function that converts any message into a hash value of 256 bits (written as 64 hexadecimal digits).

For example, an unmined block looks like that in Fig 1. The block hash is generated by applying the SHA-256 function to the block data. The timestamp indicates the specific time this block is created. The block data contains two records in this case: coinbase sending Satoshi 100 amount and Satoshi sending Pierre-O 5 amount with fee 2. The fee is a small amount of cryptocurrency that users pay to complete a transaction on the blockchain network. The fee is paid to the network's miners who validate and add it to the blockchain network. As a result, a higher fee makes the block preferred by the miners and it's more likely to be added to the blockchain network. Since this block has not been connected to the blockchain, there is no information for the block height and the previous block hash, with "Mined" set as false.

The nounce is the answer to a crypto problem. Denote $T_{max} = 2^{256} - 1$ as the largest possible value SHA-256 output can take. Given T as a target, the difficulty of the problem is $D = \frac{T_{max}}{T}$ as a given number. One's goal is to find the nounce, a number uniformly random from $\left\{0, 1, ..., 2^{32} - 1\right\}$, whose concatenation with the block information provides a number less than target T after the application of the hash function h. In other words, one tries uniformly random nounce until $h(\text{nounce}|\text{block info}) < T$. When a nounce satisfying this property is found, the crypto problem is said to be solved. As a result, the number of trial to solve this problem is geometrically distributed, there is exponential inter-block time and the length of the blockchain can be modelled as a Poisson process. See Fig. 2 for an example of a mined block where the nounce is found prior to all other opponents.

The BTC protocol in reality: one block every 10 minutes on average depending on the hashrate of the network, difficulty adjustment every 2016 blocks (baseline is that 2016 blocks are mined in 2 weeks, make sure the problem is not too easy or too hard). The BTC protocol is designed such that every 2 weeks $6 \times 24 \times 14 = 2016$ new blocks are built, and this exactly matches the baseline of the difficulty! Reward halving every 210000 blocks (the intension is to avoid inflation). When one uses mining software, there is a pool manager and several users are organized as a pool to mine a single block. After a block is mined, the pool manager distributes payoff to the pool members so there is a relative steady payoff. However, if one mines a block on his own, it's very risky and the payoff is not guaranteed at all.

Remark. *Proof-of-stake is another possible consensus protocol to use in the blockchain framework. Proof-of-work is slow and resource consuming (requires calculating the artificial cryptography problem). In proof-of-stake, each node is selected with probability proportional to the share of cryptocoins it has.*

Overview of how Blockchain works

We briefly conclude how the blockchain works. It's actually an iteration of the following steps:

1. Mining a block: miners compete to solve a cryptography problem (proof-of-work).

2. Adding the block: validate the block and append it to the blockchain once a miner finds the nounce.

3. Propagation: announce the newly mined block throughout the blockchain such that all nodes recognize and update their own copy of the blockchain.

4. Create the next block: transactions are extracted from the pending pool and put into a new block, the new block references the hash of the block that was just mined.

Since block appending time is sometimes smaller than propagation delay (the time for information to get to all blocks in the blockchain), there might be disagreement between nodes (e.g., a newly mined block refers to the previous block which is not the true latest appended block in the blockchain because two blocks are mined kind of simultaneously), which results in **forks** (several branches in the blockchain). In this case, the rule is to stick to the LCR (longest chain rule) and only trust the branch with the longest chain. The blocks in the other branches are called orphaned and are discarded (but they are not totally ignorable, refer to the double spending attack). Propagation delay is a problem that cannot be neglected when the blockchain network is very large, but typically forks do not often happen since block appending time is often longer.

Motivation of DeFi (Decentralized Finance)

TradFi has the following problems: access barrier (bank account), banks keep the record (cyber risk), there are transaction fees and transfer fees etc., slow transaction settlement, and difficulty to verify the accuracy of transactions and asset holdings. TradFi has censorship and restrictions (government), global accessibility (24/7), fractional ownership (tokenized assets) and innovation (interoperability).

Although one still has to add fees to make the transaction being processed faster, the fee depends on the information appended in the block (small number of transactions results in lower fees). If the transaction amount is large but the number of transaction is not large, then the fee could be very low (totally different from that in reality, which depends on the amount).

Decentralized Exchange

Sometimes one sees centralized exchange where one can buy bitcoins using currency. Those companies has the business to receive a lot of buy and sell, ask for a fee from the buyer or seller, but reorganize and group the transactions (optimization) so that the transactions actually have a lower fee. The centralized exchange then earns the difference in the fee as the profit. Those exchanges include Binance, Coinbase, etc., but they keep the order book to match buyer and seller to trade. They act like market makers.

On the other hand, decentralized exchanges use algorithm automated market maker (AMM). One can exchange one token against another with liquidity providers. The stable coin is a bridge from fiat to cryptocurrency

constant product marmet maker (CPMM): $xy = k$, x, y are amounts of tokens and k is constant.

Example: price of one ETH is $P = 500$, first LP provides 20 ETH, 10000 DAI (this number has to be 20×500 to match the price, here one DAI is some ETH locked to back up for one dollar), sets $k = 200000$. The liquidity provided is measured by $L = \sqrt{xy} = \sqrt{200000}$ as the geometric mean (here $x = 20, y = 10000$). One can only trade on the curve $xy = k$ so the pool never runs out of some tokens (price becomes infinity).

Swap X for Y: acquire dy of token Y, then deposit dx such that $(x + dx)(y - dy) = k$ and pay a fee adx to LP (liquidity providers). This causes the price of Y to rise in the pool.

Example: takes 2 ETH then deposits $dx = 1111$, give out $0.3 \times 1111 = 333$ worth of fees. The price of ETH then rises to $\frac{x+dx}{y-dy} = 617$.

Add liquidity: another LP provides dx of token X, to maintain the same price, $dy = y/xdx$. New level $k' = (x + dx)(y + dy)$ then liquidity rises $L' = \sqrt{xy} + \sqrt{dxdy}$. LPs are weighted according to the liquidity they provide, fees distributed according to weights (rewards for LPs depend on contribution).

Example: new LP deposits 5000 worth of tokens to pool, $dx = 2500, dy = 5$, weights computed based on liquidity $\frac{L}{L'}$. This results in the shift of the curve.

flaws: price slippage, divergent

sandwich attack for CPMM

Security

Double Spending Attack

Goal: replace a transaction by another one inside the blockchain

Scenario: Marie sends BTC 10 to John, this trans recorded in a block. John waits for integer α confirmation (have $\alpha - 1$ blocks added after the block with trans). Once we reach α confirmation then John ships the good. Marie issue a conflicting trans: M to M with BTC 10 transferred (the point is that Marie has to make it conflicting so that the fork happens, e.g. she only has BTC 10 but not BTC 20). Marie has malicious nodes working on a private branch in which M to J is replaced by M to M. At shipment date, the private branch is behind the public branch by x blocks. Marie wants to make the private chain longer than the public one such that M to M gets admitted so she does not have to pay BTC 10.

Random walk model: $\{X_n\}$ be the difference in length between two branches, then $X_n = x + \xi_1 + ... + \xi_n$ where ξ_i are *i.i.d.* such that $\mathbb{P}(\xi = 1) = p, \mathbb{P}(\xi = -1) = 1 - p$. Whenever each block mined found by public branch w.p. p, private branch w.p. $q = 1 - p$. Assume $p > q$ (public branch has the majority of computing power), double spending time τ_0 first hitting time to zero.

Obviously $\mathbb{P}(\tau_0 < \infty) = (\frac{q}{p})^x$ and the distribution of τ_0 is known.

Counting Process Model

Levy process X independent stationary increment with Cadlag sample paths. Model $x + N_t$ as length of public branch, and M_t as length of private branch. N intensity λ while M intensity $\mu < \lambda$. Consider first time they meet τ, then $\mathbb{P}(\tau < \infty) = (\frac{\mu}{\lambda})^x$. ($Y_t = M_t - N_t$ is Levy, consider its exp MG $e^{\theta Y_t - tk(\theta)}$, OST). Want to find $\gamma > 0$ such that $e^{\gamma Y_t}$ is MG, so $k(\gamma) = 0$, see that $\gamma = \log \frac{\lambda}{\mu}$. Since $Y_t \to -\infty$, OST proves.

Model-free Finance and OT

The main idea of model-free finance in derivative pricing is to avoid using any pre-assumed dynamics of the stock price, e.g., the Black-Scholes model, the stochastic volatility model. This approach avoids model calibration and problems with model specification.

Super-Replication: Duality

We take the perspective of the European option seller's side. In order to hedge the risk of a short position on the option, the seller buys H stock at time 0. We denote the portfolio value (of selling one option and buying H stocks) at time t by π_t, the option payoff by $F_T := F(S_T)$ and the option price by C. It is clear that

$$\pi_0 = -HS_0 + C. \tag{833}$$

At time T (the time of maturity of the option),

$$\pi_T = (-HS_0 + C)e^{rT} + HS_T - F_T, \tag{834}$$

with the risk coming essentially from S_T. Here we denote $\mathbb{P}^{\text{hist}}$ as a probability measure that describes the historical fluctuations of S_T and assume that S_T is well-modelled by $\mathbb{P}^{\text{hist}}$. In the general setting, the market is incomplete so we cannot necessarily perform a perfect replication of the option payoff F_T, i.e. $\pi_T = 0,\ \mathbb{P}^{\text{hist}} - a.s.$. Instead, we can perform super-replication which ensures that $\pi_T \geq 0,\ \mathbb{P}^{\text{hist}} - a.s.$. This provides a natural motivation for the definition of the **super-replication price**:

$$C_{sel} := \inf\left\{C : \exists H, \pi_T \geq 0,\ \mathbb{P}^{\text{hist}} - a.s.\right\}. \tag{835}$$

The duality theorem, as the fundamental result in model-free finance, states that the super-replication price shall be equal to the maximum martingale price.

Theorem 35 (Super-Replication Duality)**.** *If $F_T \in L^\infty(\mathbb{P}^{hist})$ and $\mathscr{M}_1$ is not empty (which holds when the market is viable), then*

$$C_{sel} = \sup_{\mathbb{Q}\in\mathscr{M}_1} \mathbb{E}_{\mathbb{Q}} e^{-rT} F_T, \tag{836}$$

where

$$\mathscr{M}_1 := \left\{\mathbb{Q} \sim \mathbb{P}^{hist} : \mathbb{E}_{\mathbb{Q}} e^{-rT} S_T = S_0\right\}. \tag{837}$$

Remark. *$\mathscr{M}_1$ contains all martingale (risk-neutral) measures that are equivalent to $\mathbb{P}^{hist}$, under which $e^{-rt}S_t$ is a martingale. As a result, the RHS in the theorem above can be interpreted as the maximum martingale price in the sense of martingale pricing. The duality tells us that super-replication and martingale measures are two sides of the* *same coin.*

The proof of the theorem relies on the Fenchel-Rockafeller duality, which provides sufficient conditions for the strong duality of a general optimization problem.

Lemma 43 (Fenchel-Rockafeller Duality)**.** *E is a normed vector space, with $\Theta, \Sigma : E \to \mathbb{R} \cup \{\infty\}$ to be convex functions. Assume that $\exists z_0 \in E, \Theta(z_0) < \infty, \Sigma(z_0) < \infty$ and Θ is continuous at z_0, then*

$$\inf_{z\in E} \{\Theta(z) + \Sigma(z)\} = \max_{z^*\in E^*} \{-\Theta^*(-z^*) - \Sigma^*(z^*)\}, \tag{838}$$

where $\Theta^*, \Sigma^* : E^* \to \mathbb{R} \cup \{\infty\}$ are Fenchel conjugates. Moreover, the maximum on the RHS can be attained.

Proof. It suffices to prove the LHS is larger than the RHS (the other direction is implied by the weak duality).

Step 1: Apply Hahn-Banach.

Let $A := \text{epi}(\Theta), B := \text{hypo}(M - \Sigma) \subset E \times \mathbb{R}$ where $M := \inf_{z\in E} \{\Theta(z) + \Sigma(z)\}$. Both A, B are convex and non-empty so by Hahn-Banach, there exists a hyperplane $H = \{(x, t) \in E \times \mathbb{R} : f(x) + kt = \alpha, f \in E^*\}$ that separates the disjoint convex open sets $C = A^\circ$ and B.

Step 2: Prove $k \neq 0$ so that the hyperplane is not parallel to the axis of t.

Denote $\Phi(x, t) := f(x) + kt$ and WLOG, assume

$$\forall (x, t) \in C, \Phi(x, t) \geq \alpha, \quad \forall (x, t) \in B, \Phi(x, t) \leq \alpha. \tag{839}$$

Due to the continuity of Φ, $\forall (x, t) \in A, \Phi(x, t) \geq \alpha$. In this case, $(z_0, t) \in A$ whenever t is large enough, which implies that $k \geq 0$. Hereby we prove that actually $k > 0$.

We prove by contradiction. If $k = 0$, then $H = \{x : f(x) = \alpha\}$, which implies that $z_0 \in H$ (consider $(z_0, \Theta(z_0)) \in A$ and $(z_0, M - \Sigma(z_0)) \in B$). Since Θ is continuous at z_0, there exists $r > 0$ such that $\Theta < \infty$ on $B(z_0, r)$. This means that for any z such that $\|z\| < r$ and any $|\delta| < 1$, $\Theta(z_0 + \delta z) < \infty$. As a result, $f(z_0) + \delta f(z) = f(z_0 + \delta z) \geq \alpha$ since $(z_0 + \delta z, \Theta(z_0 + \delta z)) \in A$ and that f is linear, which implies $\delta f(z) \geq 0$. Due to δ and z taken under arbitrary sense, f is always non-negative in a neighborhood of the origin, indicating that the linear functional $f \equiv 0$ on E. In this situation, $H = \{x : 0 = \alpha\}$ is trivial, a contradiction!

Step 3: Prove the inequality and identify the maximizer.

We check that $\frac{f}{k}$ is the maximizer. Simple calculations based on the definition of Fenchel conjugate and the separation hyperplane show that

$$\Theta^*\left(\frac{f}{k}\right) = -\frac{1}{k} \inf_{z\in E} \{f(z) + k\Theta(z)\} \leq -\frac{\alpha}{k}, \tag{840}$$

$$\Sigma^*\left(\frac{f}{k}\right) = \frac{1}{k} \sup_{z\in E} \{f(z) - k\Sigma(z)\} \leq -M + \frac{\alpha}{k}. \tag{841}$$

Combining both yields

$$M \leq -\Theta^*\left(-\frac{f}{k}\right) - \Sigma^*\left(\frac{f}{k}\right) \leq M, \tag{842}$$

which concludes the proof. □

At this point, we could provide a formal proof of the duality theorem.

Proof of super-replication duality. The primal optimization problem is:

$$\begin{cases} \inf_{C,H} C \\ s.t. \ -HS_0 + C + (HS_T - F_T)e^{-rT} \geq 0, \ \mathbb{P}^{\text{hist}} - a.s. \end{cases}. \tag{843}$$

Write down the Lagrangian:

$$L(C,H,q) = C - \int_0^\infty [-HS_0 + C + (Hs - F(s))e^{-rT}]\, dq(s), \tag{844}$$

where q is a positive measure equivalent to $\mathbb{P}^{\text{hist}}$. We discuss in more details the motivation of identifying the Langrange multiplier q:

1. The constraint in the primal problem is an inequality of a random variable, whose dual variable shall take value as a measure, i.e., $\langle X, \mathbb{P}\rangle = \int X(\omega)\, d\mathbb{P}(\omega)$ (Riesz-Markov representation theorem).

2. A well-known result from functional analysis states that generally $(L^\infty(\mathbb{P}^{\text{hist}}))^* \neq L^1(\mathbb{P}^{\text{hist}})$. The Yosida-Hewitt decomposition provides the detailed structure of $(L^\infty(\mathbb{P}))^*$. However, for the purpose of financial applications, we always restrict attention to the regular part of $(L^\infty(\mathbb{P}))^*$, i.e., formally accepting that

$$(L^\infty(\mathbb{P}^{\text{hist}}))^* \overset{formally}{=} L^1(\mathbb{P}^{\text{hist}}) \cong \mathscr{Q} := \left\{\mathbb{Q} : \mathbb{Q} \text{ is a signed measure, and } \mathbb{Q} << \mathbb{P}^{\text{hist}}\right\}. \tag{845}$$

One can also understand this part as if $L^\infty(\mathbb{P}), L^1(\mathbb{P})$ are assumed to be reflexive spaces, which holds true under strong restrictions, e.g., the probability measures only concentrates on a finite support. ***This part is ambiguious in the references, and this is the best I can do...*** After accepting those facts, the dual variable $q \in \mathscr{Q}$ could be naturally identified.

3. The introduction of the Lagrangian enables us to write the primal problem in a minimax form (minimizing w.r.t. the primal variable and maximizing w.r.t. the dual variable). In this sense, we hope that

$$\sup_q \left\{C - \int_0^\infty [-HS_0 + C + (Hs - F(s))e^{-rT}]\, dq(s)\right\} \tag{846}$$

$$= \begin{cases} C & \text{if} - HS_0 + C + (HS_T - F_T)e^{-rT} \geq 0, \ \mathbb{P}^{\text{hist}} - a.s. \\ +\infty & \text{else} \end{cases}. \tag{847}$$

This explains why we add the positivity of q and the equivalence between q and $\mathbb{P}^{\text{hist}}$.

The next step is to exploit strong duality and interchange inf and sup.

$$\sup_q \inf_{C,H} \left\{ C - \int_0^\infty [-HS_0 + C + (Hs - F(s))e^{-rT}]\, dq(s) \right\} \tag{848}$$

$$= \sup_q \inf_{C,H} \left\{ C\left(1 - \int_0^\infty dq(s)\right) + H\left(S_0 \int_0^\infty dq(s) - e^{-rT}\int_0^\infty s\, dq(s)\right) + e^{-rT}\int_0^\infty F(s)\, dq(s) \right\}. \tag{849}$$

Since the function is linear in C and H, the inf w.r.t. C, H is finite iff $\int_0^\infty dq(s) = 1$, and $S_0 \int_0^\infty dq(s) = e^{-rT}\int_0^\infty s\, dq(s)$. We get the dual problem:

$$\begin{cases} \sup_q e^{-rT}\mathbb{E}_q F_T \\ s.t.\ q \sim \mathbb{P}^{\text{hist}} \text{ is a probability measure}, S_0 = e^{-rT}\mathbb{E}_q S_T \end{cases}, \tag{850}$$

which concludes the proof.

Finally, we apply the Fenchel-Rockafeller duality to justify the interchange of inf and sup. We split the justification into several steps for the sake of clarity.

Step 1: Rewrite the primal optimization problem. Denote $\mathscr{X} := \left\{HS_0 - (HS_T - F_T)e^{-rT} : H \in \mathbb{R}\right\} \subset L^\infty(\mathbb{P}^{\text{hist}})$ so that the primal problem is equivalent to

$$\inf_C C \tag{851}$$

$$s.t.\ \exists \xi \in \mathscr{X}, C \geq \xi,\ \mathbb{P}^{\text{hist}} - a.s. \tag{852}$$

Rewrite the objective as a function in ξ:

$$\inf_\xi \operatorname{esssup}(\xi) \tag{853}$$

$$s.t.\ \xi \in \mathscr{X}. \tag{854}$$

Step 2: Identify Θ and Σ.
Consider $E = L^\infty(\mathbb{P}^{\text{hist}})$ so that $E^* \overset{formally}{=} \mathscr{Q}$ as remarked above. Set

$$\Theta(\xi) = \begin{cases} 0 & \text{if } \xi \in \mathscr{X} \\ +\infty & \text{else} \end{cases}, \Sigma(\xi) = \operatorname{esssup}(\xi). \tag{855}$$

Simple calculations show that

$$\Theta^*(-q) = -e^{-rT}\int F(s)\,dq(s) + \sup_H \left\{ H\left(e^{-rT}\int s\,dq(s) - S_0\right)\right\} \tag{856}$$

$$= \begin{cases} -e^{-rT}\int F(s)\,dq(s) & \text{if } e^{-rT}\int s\,dq(s) = S_0 \\ +\infty & \text{else} \end{cases}. \tag{857}$$

Similarly,

$$\Sigma^*(q) = \begin{cases} 0 & \text{if } q \text{ is a probability measure equivalent to } \mathbb{P}^{\text{hist}} \\ +\infty & \text{else} \end{cases}. \tag{858}$$

Applying the Fenchel-Rockafeller duality concludes the proof. □

Remark. *From the proof, we see an extra side product that the super-replication price can be attained by some martingale measure $\mathbb{Q}_{sel}$, but cannot necessarily be attained by some hedging strategy H.*

The similar definition holds on the buyer's side. The buyer's price is defined as

$$C_{buy} := \sup\left\{C : \exists H, -\pi_T \geq 0,\ \mathbb{P}^{\text{hist}} - a.s.\right\}, \tag{859}$$

and it is clear that $C_{buy} \leq C_{sell}$. Consider a simple example where $r = 0$, then $\pi_T = C - HS_0 + HS_T - F_T$. If we assume that the support of the r.v. $-HS_0 + HS_T - F_T$ is $[-2, 3]$, then $C_{sel} = 2$ since moving the support upward by 2 units makes it always non-negative. On the other hand, $C_{buy} = -3$ since moving the support downward by 3 units makes it always non-positive.

The super-replication theorem tells us that

$$C_{buy} = \inf_{\mathbb{Q}\in\mathscr{M}_1} \mathbb{E}_{\mathbb{Q}} e^{-rT} F_T, \tag{860}$$

with the inf to be attained by some MG measure $\mathbb{Q}_{buy}$.

As a result, any arbitrage-free price $C \in [C_{buy}, C_{sel}]$, and there always exists a MG measure $\mathbb{Q} \in \overline{\mathscr{M}_1}$ that attains this price, which is a convex combination of $\mathbb{Q}_{buy}$ and $\mathbb{Q}_{sel}$ (since the expectation is linear in the measure).

Remark. *$\mathbb{Q}_{buy}$ and $\mathbb{Q}_{sel}$ both **depend on the payoff** F_T, which implies that $\mathbb{E}_{\mathbb{Q}_{sel}} e^{-rT} F_T$ is generally not linear in F_T. Given two option payoffs F_T^1, F_T^2, we can only guarantee $\mathbb{E}_{\mathbb{Q}_{sel}} e^{-rT}(F_T^1 + F_T^2) \leq \mathbb{E}_{\mathbb{Q}_{sel}} e^{-rT} F_T^1 + \mathbb{E}_{\mathbb{Q}_{sel}} e^{-rT} F_T^2$ but not the equation.*

Pricing Option with Joint Payoff

Let's assume that $r = 0$, and the price of all European call options with strike price K and maturity T is observable on the market for a continuum of K. By B-L formula, this provides the distribution of the stock price at time T under the market measure $\mathbb{P}_{mkt}$ (which is a MG measure, $\mathbb{E}_{\mathbb{P}^1} S_T^1 = S_0^1$). In practice, a continuum of K is impossible to directly observe. One uses the BS formula (without assuming the BS model!) to turn the observed call prices on the market into points on the implied volatility surface, then interpolates the implied volatility surface, which corresponds to the European call prices for a continuum of K. Under this assumption, we can always assume that we have full knowledge of the distribution of two stock prices $S_1 := S_T^1 \sim \mathbb{P}^1, S_2 := S_T^2 \sim \mathbb{P}^2$ at maturity.

In this setting, we hope to calculate the super-replication price of an European option with payoff $c(S_1, S_2)$. However, the joint distribution of (S_1, S_2) is unknown while the marginals are known. In this case, the super-replication price is defined as

$$\mathrm{MK}_2 := \inf_{(\lambda_1,\lambda_2)\in\mathscr{P}^*(\mathbb{P}^1,\mathbb{P}^2)} \mathbb{E}_{\mathbb{P}^1}\lambda_1(S_1) + \mathbb{E}_{\mathbb{P}^2}\lambda_2(S_2), \tag{861}$$

where $\mathscr{P}^*(\mathbb{P}^1, \mathbb{P}^2)$ is the collection of (λ_1, λ_2) satisfying $\lambda_1(s_1) + \lambda_2(s_2) \geq c(s_1, s_2)$.

Intuitively, from the seller's side, the portfolio consists of selling one option with joint payoff, buying $\lambda_1(s_1)$-value of stock S^1 and buying $\lambda_2(s_2)$-value of stock S^2. The constraint is equivalent to saying that the terminal payoff of this portfolio is non-negative almost surely, and the seller hopes to figure out the minimum cost of super-replication.

Notice that if inf is changed to a sup, this is the Kantorovich dual problem. By Kantorovich duality, the following theorem is proved.

Theorem 36. *If $\mathscr{P}(\mathbb{P}^1, \mathbb{P}^2)$ is not empty (which is always true since there exists a trivial coupling),*

$$\mathrm{MK}_2 = \sup_{\mathbb{P}\in\mathscr{P}(\mathbb{P}^1,\mathbb{P}^2)} \mathbb{E}_{\mathbb{P}}[c(S_1, S_2)], \tag{862}$$

where $\mathbb{P} \in \mathscr{P}(\mathbb{P}^1, \mathbb{P}^2)$ is the collection of probability measures, under which $S_1 \sim \mathbb{P}^1$ and $S_2 \sim \mathbb{P}^2$.

Remark. *The wisdom here is to borrow the coupling interpretation of the Kantorovich formulation, which optimizes w.r.t. the joint distribution based only on knowledge of marginals. In other words, we can say that this problem has no interpretation as a transport problem. Instead, we only borrow tools from OT.*

Correspondence to OT framework:

- *cost functional – joint option payoff*
- *Kantorovich potentials – super-hedging strategy*

Remark. *It's well-known that the inf in the definition of* MK_2 *is attainable. OT theory proves the existence of the solution to the Kantorovich primal and dual problems, but the minimizers λ_1, λ_2 are obviously not unique.*

By the Kantorovich-Rubinstein theorem, the inf is attained by $\lambda_1 = \lambda_2^c$, where λ_2^c is the c-transform of λ_2, defined as:

$$\lambda_2^c(s_1) := \sup_{s_2} c(s_1, s_2) - \lambda_2(s_2). \tag{863}$$

As a result, the MK price can also be written as

$$\mathrm{MK}_2 = \inf_{\lambda} \mathbb{E}_{\mathbb{P}^1} \lambda^c(S_1) + \mathbb{E}_{\mathbb{P}^2} \lambda(S_2). \tag{864}$$

Remark. *In the multi-dimensional case, S_1 takes values in $\mathbb{R}^{d_1}$ while S_2 takes values in $\mathbb{R}^{d_2}$, with each of them denoting the stock price of a sector of stocks at time T. In the setting of OT, the distribution of S_1 and S_2 are known but the joint distribution of S_1, S_2 is unknown, while the option has joint payoff $c(S_1, S_2)$.*

Different from the previous case, $S_1 = (S_1^1, ..., S_1^{d_1})$ itself is also a joint distribution, and the distribution actually comes from empirical option prices on the market after an application of the B-L formula, which requires observing European call option prices with payoff $(S_1 \cdot w - K)_+$ for a continuum of w and K (not practically observable). As a result, one typically assumes the correlation structure within components of S_1 through the copula. For example, the uniform copula

$$\mathbb{E}_{\mathbb{P}^1} \lambda(S_1) = \mathbb{E}[\lambda_1(F_1^{-1}(U_1), ..., F_{d_1}^{-1}(U_{d_1})) \cdot co(U_1, ..., U_{d_1})], \tag{865}$$

where $U_i \sim U(0,1)$ are i.i.d. Copula delivers the pre-assumed correlation structure on U_i onto components of S_1.

The next step is to explicitly solve the Kantorovich problem. Notice that unlike the Brenier's theorem that requires a certain convex structure on the cost functional, in 1-dimensional cases ($\mathbb{R}$), the Kantorovich problem can be solved for general cost functionals as long as $c_{12} := \partial_{s_1, s_2} c > 0$. The restriction to 1-dimension is crucial, since this argument does not hold in high-dimensional spaces.

Theorem 37 (Solution to MK Price)**.** *For Kantorovich problem on $\mathbb{R}$, if $c \in C^2$ and $c_{12} > 0$, $\mathbb{P}^1$ has no point mass, the unique optimal coupling is*

$$\mathbb{P}^* = (id, T)_\# \mathbb{P}^1, \tag{866}$$

where $T := F_2^{-1} \circ F_1$. In this case,

$$\mathrm{MK}_2 = \int_0^1 c(F_1^{-1}(u), F_2^{-1}(u))\, du. \tag{867}$$

This bound is attained by a (not unique generally) static hedging strategy

$$\lambda_2(x) := \int_0^x c_2(T^{-1}(y), y)\, dy, \tag{868}$$

$$\lambda_1(x) := c(x, T(x)) - \lambda_2(T(x)). \tag{869}$$

Proof. The T is exactly the solution to 1-dim OT on $\mathbb{R}$ under quadratic cost functionals. The same T is expected to work here due to the monotone arrangement principle.

By Kantorovich duality, there exists λ_1, λ_2 attaining the bound. To prove that $\mathbb{P}^*$ and λ_i are primal and dual optimizers, it suffices to check the constraints, the primal and dual objectives, and the complementary slackness, i.e., $\mathbb{P}^*(ds_1, ds_2) > 0$ iff $\lambda_1(s_1) + \lambda(s_2) = c(s_1, s_2)$.

Firstly, $\mathbb{P}^*$ is a coupling of $\mathbb{P}^1$ and $\mathbb{P}^2$. We only check the second marginal as an example. For any f continuous bounded,

$$\int f(s_2)\, d\mathbb{P}^*(s_1, s_2) = \int f(T(s_1))\, d\mathbb{P}^1(s_1) \overset{u=F_1(s_1)}{=} \int_0^1 f \circ F_2^{-1}(u)\, du = \int f(s_2)\, d\mathbb{P}^2(s_2). \tag{870}$$

This results in the primal objective

$$\mathbb{E}_{\mathbb{P}^*} c(S_1, S_2) = \int c(s, T(s))\, d\mathbb{P}^1(s) \overset{u=F_1(s)}{=} \int_0^1 c(F_1^{-1}(u), F_2^{-1}(u))\, du. \tag{871}$$

Secondly, check dual constraint

$$\lambda_1(s_1) + \lambda_2(s_2) = c(s_1, T(s_1)) - \lambda_2(T(s_1)) + \int_0^{s_2} c_2(T^{-1}(y), y)\, dy \tag{872}$$

$$= c(s_1, s_2) + \int_{s_2}^{T(s_1)} c_2(s_1, y) - \lambda_2'(y)\, dy \geq c(s_1, s_2), \tag{873}$$

if $\forall y \in [s_2, T(s_1)], c_2(s_1, y) - \lambda_2'(y) \geq 0$. When $c_{12} > 0$,

$$c_2(s_1, y) - \lambda_2'(y) = c_2(s_1, y) - c_2(T^{-1}(y), y) = \int_{T^{-1}(y)}^{s_1} c_{12}(x, y)\, dx \geq 0. \tag{874}$$

The dual objective is

$$\mathbb{E}_{\mathbb{P}^1} \lambda_1(S_1) + \mathbb{E}_{\mathbb{P}^2} \lambda_2(S_2) = \mathbb{E}_{\mathbb{P}^1} c(S_1, T(S_1)) = \mathbb{E}_{\mathbb{P}^*} c(S_1, S_2). \tag{875}$$

Lastly, it suffices to check the complementary slackness. The support of $\mathbb{P}^*$ is exactly the graph of T. On the other hand, $\lambda_1(s_1) + \lambda_2(s_2) = c(s_1, s_2)$ iff $\forall y \in [s_2, T(s_1)], c_2(s_1, y) - \lambda_2'(y) = 0$ iff $\forall y \in [s_2, T(s_1)], \int_{T^{-1}(y)}^{s_1} c_{12}(x, y)\, dx = 0$. Since $c_{12} > 0$, it's equivalent to saying $T^{-1}(s_2) = s_1$, i.e. (s_1, s_2) is in the support of $\mathbb{P}^*$. □

Remark. *The Spence–Mirrlees condition $c_{12} > 0$ indicates that the cost has a positive interaction between s_1 and s_2. For example, this holds for $c(x, y) = -(x - y)^2$ (quadratic cost in OT), $c(x, y) = h(x - y)$ where h is strictly concave (the case in Brenier's theorem), $c(x, y) = \sqrt{xy}$, etc. The MK price is attained when $T(S_1) \overset{d}{=} S_2$, i.e., two stocks are perfectly correlated.*

From the proof above, ***when*** $c_{12} \geq 0$ ***but not necessarily strictly positive****, the complementary slackness fails. In this case, the theorem still holds but the uniqueness of the optimizers is not guaranteed. The proof can be modified in terms of weak duality. We plug $\mathbb{P}^*$ in the primal objective to get a lower bound of the MK price denoted D. We* *plug λ_1, λ_2 in the dual objective to get an upper bound of the MK price denoted P and actually $D = P$ as verified in the proof above.*

Remark. *The construction of $\mathbb{P}^*$ is natural from the Brenier's theorem. The construction of optimizers λ_i, on the other hand, follows from the idea that $\lambda_1(s_1) = \lambda_2^c(s_1)$, meaning that*

$$\lambda_1(s_1) = \sup_{s_2} \{c(s_1, s_2) - \lambda_2(s_2)\}. \tag{876}$$

We expect the sup in the c-transform to be attained when $s_2 = T(s_1)$. Taking derivative w.r.t. s_2 and evaluated at $s_2 = T(s_1)$ yields

$$c_2(s_1, T(s_1)) = \lambda_2'(T(s_1)). \tag{877}$$

On the other hand, we expect that the optimizer shall lie on the boundary of the constraint, i.e.,

$$\lambda_1(s_1) + \lambda_2(T(s_1)) = c(s_1, T(s_1)). \tag{878}$$

Combining both equations yields

$$\lambda_2(s_2) = \int_0^{s_2} c_2(T^{-1}(y), y)\, dy, \quad \lambda_1(s_1) = c(s_1, T(s_1)) - \lambda_2(T(s_1)), \tag{879}$$

which provides the construction of the dual optimizer.

Remark. *If $S_2 = \varphi(S_1)$, then it's the same as adding an extra constraint on the optimization problem, which results in c as a function purely in s_1. In this case, $\lambda_1(s_1) = c(s_1, \varphi(s_1))$ is the trivial hedging strategy and $T = \varphi$. Notice that we have been considering the seller's side so the MK price we calculate is the highest possible arbitrage-free price on the market.*

An example is given by the payoff

$$c(s_1, s_2) = (s_1 - K_1)_+ \mathbb{I}_{s_2 > K_2}, \tag{880}$$

which is a call in S_1 jointly with a digital in S_2. Calculation shows (notice $c_2(s_1, s_2) = (s_1 - K_1)_+ \delta_{K_2}(s_2)$) that

$$T = F_2^{-1} \circ F_1, \tag{881}$$
$$\lambda_1(s_1) = \left[(s_1 - K_1)_+ - (T^{-1}(K_2) - K_1)_+\right] \mathbb{I}_{T(s_1) > K_2}, \tag{882}$$
$$\lambda_2(s_2) = (T(K_2) - K_1)_+ \mathbb{I}_{s_2 > K_2}, \tag{883}$$
$$\mathrm{MK}_2 = \int_{\max\{F_1(K_1), F_2(K_2)\}}^1 [F_1^{-1}(u) - K_1]\, du. \tag{884}$$

This tells the seller the highest arbitrage-free price, together with the super-hedging strategy. When $F_1 = F_2$, i.e., two stocks have the terminal price to be of the same distribution,

$$\lambda_1(s_1) = [(s_1 - K_1)_+ - (K_2 - K_1)_+]\, \mathbb{I}_{s_1 > K_2}, \tag{885}$$

$$\lambda_2(s_2) = (K_2 - K_1)_+ \mathbb{I}_{s_2 > K_2}, \tag{886}$$

$$\mathrm{MK}_2 = \int_{\max\{F_1(K_1), F_2(K_2)\}}^1 [F_1^{-1}(u) - K_1]\, du. \tag{887}$$

Now λ_2 is the pure digital payoff for S_2, and λ_1 consists of buying a digital call in S_1 while selling a pure digital in S_1. As a result, the super-hedging portfolio consists of buying λ_1, λ_2 and selling the joint option, whose payoff equals

$$[(s_1 - K_1)_+ - (K_2 - K_1)_+](\mathbb{I}_{s_1 > K_2} - \mathbb{I}_{s_2 > K_2}) \geq 0, \tag{888}$$

and can be checked to be always non-negative.

Martingale OT and Path-Dependent Option

Consider the option payoff $c(S_1, S_2)$, where S_1, S_2 are **different** from the previous setting. We consider a market with one stock, S_1 being the stock price at time t_1 and S_2 being the stock price at time $t_2 > t_1$. The option payoff is no longer European, but actually path-dependent (similar to an Asian option payoff). In this situation, we assume the knowledge of the marginal distribution of S_1, S_2 (denoted $\mathbb{P}^1, \mathbb{P}^2$) by observing a continuum of European call price with maturity t_1, t_2 on the market, while the joint distribution of (S_1, S_2) is still unknown.

As the seller of the option, one hopes to super-replicate the option payoff in order to hedge the risk. The hedging portfolio consists of selling one option, buying European payoffs with different maturity $\lambda_1(S_1), \lambda_2(S_2)$, and buying the difference of stock price $H(S_1)(S_2 - S_1)$.

Remark. *The quantity $H(S_1)$ is a function of S_1 since it is determined at time t_1. The difference term corresponds to selling $H(S_1)$ unit of stock at time t_1 and buying back $H(S_1)$ unit of stock at time t_2. Assuming the current time to be zero, one can also trade $H(S_0)(S_1 - S_0)$ and $H(S_0)(S_2 - S_0)$. However, since the current stock price S_0 is deterministic, those two terms have already been absorbed into $\lambda_1(S_1)$ and $\lambda_2(S_2)$.*

One might be wondering: why are we only trading at time t_1 and t_2 but not at some time t_i and t_{i+1} in between, e.g., trading $H_i(S_0, S_1, S_{t_i})(S_{t_{i+1}} - S_{t_i})$ for $t_1 < t_i < t_{i+1} < t_2$. As shown later, the optimal solution is attained by $H_i \equiv 0$ in this case.

At this point, we put up the optimization problem

$$\inf_{(\lambda_1,\lambda_2,H)\in\mathscr{M}^*(\mathbb{P}^1,\mathbb{P}^2)} \left\{\mathbb{E}_{\mathbb{P}^1}\lambda_1(S_1) + \mathbb{E}_{\mathbb{P}^2}\lambda_2(S_2)\right\}, \tag{889}$$

$$s.t.\ \lambda_1(s_1) + \lambda_2(s_2) + H(s_1)(s_2 - s_1) - c(s_1, s_2) \geq 0. \tag{890}$$

$\tilde{\text{MK}}_2$ is defined as the infimum of the optimization problem, which is the best seller's price on the market. Here $\mathscr{M}^*(\mathbb{P}^1, \mathbb{P}^2)$ is requiring H to be bounded continuous, and λ_i to be integrable w.r.t. the measures $\mathbb{P}^i$.

Remark. *Notice that $\mathbb{E}_{\mathbb{P}_{mkt}} H(S_1)(S_2 - S_1) = \mathbb{E}_{\mathbb{P}_{mkt}}[H(S_1)\mathbb{E}_{\mathbb{P}_{mkt}}(S_2 - S_1|S_1)] = 0$ since $\mathbb{P}_{mkt}$ is a MG measure (notice $r = 0$). In other words, the trading of the difference produces no cost.*

Theorem 38 (MOT Duality)**.** *If $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ is not empty (which requires extra conditions),*

$$\tilde{\text{MK}}_2 = \sup_{\mathbb{P}\in\mathscr{M}(\mathbb{P}^1,\mathbb{P}^2)} \mathbb{E}_{\mathbb{P}}[c(S_1, S_2)], \tag{891}$$

where $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2) := \left\{\mathbb{P} : S_1 \overset{\mathbb{P}}{\sim} \mathbb{P}^1, S_2 \overset{\mathbb{P}}{\sim} \mathbb{P}^2, \mathbb{E}_{\mathbb{P}}(S_2|S_1) = S_1\right\}$ is the set of MG measures as couplings of given marginals.

Proof. Write down the Lagrangian of the optimization problem with sup:

$$L(\mathbb{P}, \lambda_1, \lambda_2, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}}[c(S_1, S_2)] + \int \lambda_1(s_1)\, d(\mathbb{P}^1 - \mathbb{P}^{S_1}) + \int \lambda_2(s_2)\, d(\mathbb{P}^2 - \mathbb{P}^{S_2}) - \int [\mathbb{E}_{\mathbb{P}}(S_2|S_1 = s_1) - s_1]\, d\mathbb{Q}(s_1), \tag{892}$$

where $\mathbb{P}^{S_1}, \mathbb{P}^{S_2}$ denote the law of S_1, S_2 under $\mathbb{P}$, and λ_1, λ_2 are functions, while $\mathbb{Q}$ is a signed measure on the space where S_1 lives and $\mathbb{Q} << \mathbb{P}^1$. It's easy to verify that

$$\inf_{\lambda_1,\lambda_2,\mathbb{Q}} L(\mathbb{P}, \lambda_1, \lambda_2, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}}[c(S_1, S_2)] \tag{893}$$

holds when $\mathbb{P} \in \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$, otherwise the inf is $-\infty$. Exchanging sup and inf through Fenchel-Rockafeller duality and denoting $H = \frac{d\mathbb{Q}}{d\mathbb{P}^1}$ (H uniquely identifies $\mathbb{Q}$) yields

$$\int [\mathbb{E}_{\mathbb{P}}(S_2|S_1 = s_1) - s_1]\, d\mathbb{Q}(s_1) = \int [\mathbb{E}_{\mathbb{P}}(S_2|S_1 = s_1) - s_1] H(s_1)\, d\mathbb{P}^1(s_1) \tag{894}$$

$$= \mathbb{E}_{\mathbb{P}^1}[(\mathbb{E}_{\mathbb{P}}(S_2|S_1) - S_1)H(S_1)] = \mathbb{E}_{\mathbb{P}} H(S_1)(S_2 - S_1). \tag{895}$$

Hence we have the dual objective

$$\sup_{\mathbb{P}} L(\mathbb{P}, \lambda_1, \lambda_2, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}^1}\lambda_1(S_1) + \mathbb{E}_{\mathbb{P}^2}\lambda_2(S_2), \tag{896}$$

if $\lambda_1(s_1) + \lambda_2(s_2) + H(s_1)(s_2 - s_1) - c(s_1, s_2) \geq 0$, otherwise the sup gives $+\infty$. This concludes the proof. Moreover, we proved that the sup can be attained while the inf is not necessarily attained. $\square$

Remark. *The sup-optimization problem is called martingale OT. An extra constraint that the transport trajectory must lie in the space of MG measures is added to the traditional OT problem. This problem does not have an interpretation in the sense of transportation problems, but is originated from finance. The martingale condition comes from the financial setting that the time evolution of the stock price must be a MG under the risk-neutral measure.*

It is worth discussing the interpretation of the condition: "$\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ is not empty". This condition is saying that there exists some MG measure, under which $\mathbb{P}^1$, $\mathbb{P}^2$ are the laws of stock prices subsequently observed in time. Due to the introduction of the MG condition, the trivial coupling is no longer guaranteed to lie within $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ (most often it is not in this space). Instead, we take the intuition from the European call options working as the "basis" in the financial context (Carr-Madan).

Two measures $\mathbb{P}^1, \mathbb{P}^2$ are defined to be in **convex order** (denoted $\mathbb{P}^1 \leq \mathbb{P}^2$) iff

$$\forall K, \mathbb{E}_{\mathbb{P}^1}(S_1 - K)_+ \leq \mathbb{E}_{\mathbb{P}^2}(S_2 - K)_+. \tag{897}$$

Remark. *When $r = 0$, both sides are European call prices. In this sense, the convex order implies that for any strike price, the European call with the shorter time-to-maturity (terminal stock price distribution $\mathbb{P}^1$) has a lower price than the European call with the longer time-to-maturity (terminal stock price distribution $\mathbb{P}^2$).*

We emphasize that this is true under the Black-Scholes model. When $r = 0$, BS formula tells us that the European call with time-to-maturity T has price

$$C(T) = S_0\Phi(d_1) - K\Phi(d_2), \tag{898}$$

where

$$d_1 = \frac{\log\frac{S_0}{K} + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}. \tag{899}$$

Through simple calculations,

$$\frac{\partial C(T)}{\partial T} = [K\varphi(d_2) - S_0\varphi(d_1)]\log\frac{S_0}{K}\cdot\frac{1}{2\sigma T^{\frac{3}{2}}} + [S_0\varphi(d_1) + K\varphi(d_2)]\frac{\sigma}{4\sqrt{T}}. \tag{900}$$

It's not immediately clear that this partial derivative is always non-negative. Further calculations show that

$$K\varphi(d_2) - S_0\varphi(d_1) = \frac{K}{\sqrt{2\pi}}e^{-\frac{d_2^2}{2}}\left(1 - e^{\log\frac{S_0}{K} - \frac{d_1^2}{2} + \frac{d_2^2}{2}}\right) = 0, \tag{901}$$

since $\log\frac{S_0}{K} - \frac{d_1^2}{2} + \frac{d_2^2}{2} \equiv 0$*. At this point, we checked that if the stock price* $\{S_t\}$ *is generated by the Black-Scholes model, then* $\mathscr{L}(S_{t_1}) \le \mathscr{L}(S_{t_2})$ *if* $t_1 < t_2$*.*

One way is clear that if $\mathbb{P} \in \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$, then $\mathbb{P}^1$ and $\mathbb{P}^2$ share the same mean S_0, and by Jensen's inequality,

$$\forall K, \mathbb{E}_{\mathbb{P}}((S_2 - K)_+|S_1) \ge [\mathbb{E}_{\mathbb{P}}(S_2|S_1) - K]_+ = (S_1 - K)_+. \tag{902}$$

Taking expectation on both sides yields $\mathbb{E}_{\mathbb{P}}(S_2 - K)_+ \ge \mathbb{E}_{\mathbb{P}}(S_1 - K)_+$, so $\mathbb{P}^1 \le \mathbb{P}^2$. Actually, the converse is also true by construction, and the following theorem holds as a consistency guarantee for the market observations to be following some martingale measure.

Theorem 39. $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ *is not empty iff* $\mathbb{P}^1$, $\mathbb{P}^2$ *share the same mean* S_0 *and* $\mathbb{P}^1 \le \mathbb{P}^2$*.*

So far, after clarifying the duality argument, we finally come back to prove that any intermediate trading of the stock price differences is not necessary in the problem formulation.

Theorem 40. *Given* $j \in (1,2)$ *(e.g.,* $j = \frac{3}{2}$*) such that* S_j *is the stock price at time* $t_j \in (t_1, t_2)$*. Define*

$$\tilde{\mathrm{MK}}_2^j := \inf_{(\lambda_1,\lambda_2,H,H_j)\in\mathscr{M}_j^*(\mathbb{P}^1,\mathbb{P}^2)} \left\{\mathbb{E}_{\mathbb{P}^1}\lambda_1(S_1) + \mathbb{E}_{\mathbb{P}^2}\lambda_2(S_2)\right\}, \tag{903}$$

$$s.t.\ \lambda_1(s_1) + \lambda_2(s_2) + H(s_1)(s_j - s_1) + H_j(s_1, s_j)(s_2 - s_j) - c(s_1, s_2) \ge 0. \tag{904}$$

Then $\tilde{\mathrm{MK}}_2^j = \tilde{\mathrm{MK}}_2$*.*

Proof. A similar duality argument provides

$$\tilde{MK}_2^j = \sup_{\mathbb{P}\in\mathscr{M}_j(\mathbb{P}^1,\mathbb{P}^2)} \mathbb{E}_{\mathbb{P}}[c(S_1, S_2)], \tag{905}$$

where $\mathscr{M}_j(\mathbb{P}^1, \mathbb{P}^2) := \left\{\mathbb{P} : S_1 \overset{\mathbb{P}}{\sim} \mathbb{P}^1, S_2 \overset{\mathbb{P}}{\sim} \mathbb{P}^2, \mathbb{E}_{\mathbb{P}}(S_j|S_1) = S_1, \mathbb{E}_{\mathbb{P}}(S_2|S_1, S_j) = S_j\right\}$. This proves $\tilde{MK}_2^j \le \tilde{MK}_2$ since $\mathscr{M}_j(\mathbb{P}^1, \mathbb{P}^2) \subset \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$.

On the other hand, $\tilde{MK}_2$ can be attained by some measure $\mathbb{P}^* \in \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$. Setting

$$\mathbb{P}^{*,j} := \mathscr{L}(X_1, X_j, X_2),\ (X_j, X_2) \sim \mathbb{P}^*,\ X_1 = X_j \tag{906}$$

constructs $\mathbb{P}^{*,j}$ based on $\mathbb{P}^*$. It's easy to check that under $\mathbb{P}^{*,j}$, $S_1 \sim \mathbb{P}^1, S_2 \sim \mathbb{P}^2$. We check two extra MG conditions:

$$\mathbb{E}_{\mathbb{P}^{*,j}}(S_j|S_1) = \mathbb{E}_{\mathbb{P}^{*,j}}(S_1|S_1) = S_1, \tag{907}$$

$$\mathbb{E}_{\mathbb{P}^{*,j}}(S_2|S_1, S_j) = \mathbb{E}_{\mathbb{P}^{*,j}}(S_2|S_j) = \mathbb{E}_{\mathbb{P}^*}(S_2|S_j) = S_j. \tag{908}$$

This proves $\mathbb{P}^{*,j} \in \mathscr{M}_j(\mathbb{P}^1, \mathbb{P}^2)$. However,

$$\tilde{MK}_2^j \geq \mathbb{E}_{\mathbb{P}^{*,j}}[c(S_1, S_2)] = \mathbb{E}_{\mathbb{P}^{*,j}}[c(S_j, S_2)] = \mathbb{E}_{\mathbb{P}^*}[c(S_j, S_2)] = \mathbb{E}_{\mathbb{P}^*}[c(S_1, S_2)] = \tilde{MK}_2. \tag{909}$$

This concludes the proof. □

Remark. *From the inf-optimization problem, it's not immediately clear why all intermediate tradings of stock price differences are unnecessary. However, under its dual, the structure becomes clear since we can always extend the two-marginal coupling $\mathbb{P}^*$ to the three-marginal coupling $\mathbb{P}^{*,j}$ by setting $S_j = S_1$ (act as if the stock price remains constant between time t_1 and t_j while keeping the MG property). Of course, such an extension does not change the objective. That is exactly the wisdom of duality.*

Example. *Finally, we compute an example for illustration. Take $\mathbb{P}^1 = U(-1,1), \mathbb{P}^2 = U(-2,2)$ and option payoff $c(S_1, S_2) = |S_1 - S_2|$, with $S_0 = 0$. Firstly, we check that $\mathbb{P}^1, \mathbb{P}^2$ share mean zero, which is S_0, and $\mathbb{P}^1 \leq \mathbb{P}^2$ so that $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ is not empty.*

Remark. *For any convex function ϕ, we can prove*

$$\frac{1}{2}\int_{-1}^{1} \phi(x)\, dx \leq \frac{1}{4}\int_{-2}^{2} \phi(x)\, dx = \frac{1}{2}\int_{-1}^{1} \phi(2x)\, dx. \tag{910}$$

This is because convexity implies $\phi(x) - \phi(2x) \leq \phi(0) - \phi(x)$, and Jensen's inequality implies $\phi(0) \leq \frac{1}{2}\int_{-1}^{1}\phi(x)\, dx$. This proves the convex order condition.

Then, one first finds an upper bound for $\tilde{MK}_2$. Consider $\lambda_1(s_1) = \frac{1-s_1^2}{2}, \lambda_2(s_2) = \frac{s_2^2}{2}, H(s_1) = -s_1$, then

$$\lambda_1(s_1) + \lambda_2(s_2) + H(s_1)(s_2 - s_1) - c(s_1, s_2) \geq 0. \tag{911}$$

As a result, $\tilde{MK}_2 \leq 1$ from the primal.

On the other hand, consider $S_1 + Z$ where Z is independent of S_1 and takes ± 1 w.p. $\frac{1}{2}$. This provides a MG coupling $\mathscr{L}(S_1, S_1 + Z)$ since $S_1 + Z \sim U(-2,2)$ and $\mathbb{E}(S_1 + Z|S_1) = S_1 + \mathbb{E}Z = S_1$. As a result, $\tilde{MK}_2 \geq 1$ from the dual.

*Combining both parts yields $\tilde{MK}_2 = 1$, which tells us: (i). there **exists a MG measure** under which the stock price evolves from $U(-1,1)$ to $U(-2,2)$ as time goes by. (ii). the **seller's price** (highest arbitrage-free price) of* *an option with path-dependent payoff* $|S_1 - S_2|$ *where the marginals* $S_1 \sim U(-1,1), S_2 \sim U(-2,2)$ *is* 1*. (iii). the* ***seller's super-replication strategy*** *is to buy European payoffs* $\lambda_1(S_1) = \frac{1-S_1^2}{2}$ *and* $\lambda_2(S_2) = \frac{S_2^2}{2}$ *at time* 0*, buy* S_1 *unit of stock at time* t_1 *and sell those stocks at time* t_2*.*

The Solution to Martingale OT

Normal OT problems on $\mathbb{R}$ has its solution $\mathbb{P} = (id, T)_{\#}\mathbb{P}^1$ given by the Brenier map $T : \mathbb{R} \to \mathbb{R}$ under the condition $c_{12} \geq 0$. Such a map, however, violates the martingale constraint, in the sense that $\mathbb{E}_{\mathbb{P}}(S_2|S_1) = T(S_1) \neq S_1$ generally. The problem lies in the contradiction between the deterministic attribute of the transport plan T and the probabilistic attribute of the martingale constraint. Naturally, we introduce T_u, T_d as the upper and lower bounds of the transport plan. The mass at x is probabilistically transported to $\{T_d(x), T_u(x)\}$ (a distribution supported on those two points).

The increasing property of the optimal transport plan in normal OT problems (Brenier's theorem) is generalized to the **left-monotonicity**, stating that $[T_d(x), T_u(x)]$ is increasing under the set inclusion relationship as x gets larger. A measure $\mathbb{P} \in \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$ is left-monotone if there exists Borel $\Gamma \subset \mathbb{R}^2$, such that (X, Y) is supported on Γ under $\mathbb{P}$, and $\forall (x, y_1), (x, y_2), (x', y') \in \Gamma$ with $x < x'$, it holds that $y' \notin (y_1, y_2)$. A typical example to bear in mind is when T_d, T_u share the same plot before some threshold, after which a splitting happens. T_u goes up after the splitting while T_d goes down after the splitting. As an extension of Brenier's theorem, the optimal coupling for martingale OT is the unique left-monotone coupling as a transport map between $\mathbb{P}^1$ and $\mathbb{P}^2$ (under certain condition).

Recall that the solution to normal OT requires a condition $c_{12} \geq 0$. Similarly, the solution to martingale OT requires $c_{122} > 0$ and that $F_2 - F_1$ has a unique maximum m. In the proof of the theorem below, we assume that there exists hedging strategy $\lambda_1^*, \lambda_2^*, H^*$ that attain the inf in martingale OT (which is not generally true).

Theorem 41. *Assume $\mathbb{P}^1 \leq \mathbb{P}^2$ with the same mean S_0 (so that $\mathscr{M}(\mathbb{P}^1, \mathbb{P}^2) \neq \emptyset$). Assume $c_{122} > 0$ and that $F_2 - F_1$ has a unique maximum m. Assume that there exists hedging strategy $\lambda_1^*, \lambda_2^*, H^*$ that attain the inf in martingale OT. The optimizers of martingale OT are given by:*

$$\mathbb{P}^*(ds_1, ds_2) = \mathbb{P}^1(ds_1)[q(s_1)\delta_{T_u(s_1)}(ds_2) + (1 - q(s_1))\delta_{T_d(s_1)}(ds_2)], \tag{912}$$

$$q(x) = \frac{x - T_d(x)}{T_u(x) - T_d(x)}, \tag{913}$$

$$T_u(x) = T_d(x) = x \text{ when } x \leq m, \tag{914}$$

$$T_u(x) = F_2^{-1}(F_1(x) + (F_2 - F_1)(T_d(x))) \text{ when } x > m, \tag{915}$$

$$T_d'(x) = -\frac{T_u(x) - x}{T_u(x) - T_d(x)} \frac{F_1'(x)}{F_2'(T_d(x)) - F_1'(T_d(x))},\ T_d(m) = m,\ \text{when } x > m \tag{916}$$

$$H^{*\prime}(s_1) = \frac{c_1(s_1, T_u(s_1)) - c_1(s_1, T_d(s_1))}{T_u(s_1) - T_d(s_1)},\ \text{when } s_1 \geq m, \tag{917}$$

$$\lambda_2^{*\prime}(s_2) = \begin{cases} c_2(T_u^{-1}(s_2), s_2) - H^* \circ T_u^{-1}(s_2) & \text{when } s_2 \geq m \\ c_2(T_d^{-1}(s_2), s_2) - H^* \circ T_d^{-1}(s_2) & \text{when } s_2 < m \end{cases}, \tag{918}$$

$$\lambda_1^*(s_1) = \mathbb{E}_{\mathbb{P}^*}[c(S_1, S_2) - \lambda_2^*(S_2)|S_1 = s_1]. \tag{919}$$

If only a derivative is given, the optimizer can be different up to an arbitrary constant.

Proof. To prove the primal and dual optimality, we split the proof into three steps. We check that the primal and dual optimizers satisfy the constraints, provide the same objective, and satisfy the complementary slackness. Before going into the proof, we provide interpretation for the optimizers.

1. Under the optimal coupling $\mathbb{P}^*$, (S_1, S_2) has the same law as (X, Y), where $X \sim \mathbb{P}^1$ and

$$Y|_X = \begin{cases} T_u(X) & \text{w.p. } q(X) \\ T_d(X) & \text{w.p. } 1 - q(X) \end{cases} . \tag{920}$$

   Obviously, (X, Y) has its support on the graph of (T_u, T_d), with $q(x)$ interpreted as the probability of the mass located at x being transported to the T_u branch.

2. Actually $T_d(x) \le x \le T_u(x)$, and when $x > m$, T_u is increasing while T_d is decreasing, as shown in the proof below. The solution pair (T_u, T_d) is actually left-monotone.

3. Compare to the binomial tree model for option pricing. At location x, the stock price in the next time step either rises by $T_u(x)$ or drops by $T_d(x)$. From this perspective, q is the risk-neutral probability of seeing the stock price rise in the next time step and $\mathbb{P}^*$ is nothing but a risk-neutral measure (identify $x = e^{(r-\delta)h}$). The martingale constraint is satisfied by constructing a risk-neutral measure.

4. The construction of the optimizers is inspired by the structure of the optimizer that

$$\lambda_1^*(s_1) = \sup_{s_2} \{c(s_1, s_2) - \lambda_2^*(s_2) - H^*(s_1)(s_2 - s_1)\}, \ \forall s_1. \tag{921}$$

   This has a similar motivation to the c-transform in normal OT. Such sup is attained when $s_2 = T_u(s_1)$ and $s_2 = T_d(s_1)$, resulting in first-order conditions

$$c_2(s_1, T_u(s_1)) - \lambda_2^{*\prime}(T_u(s_1)) - H^*(s_1) = 0, \ \forall s_1, \tag{922}$$

$$c_2(s_1, T_d(s_1)) - \lambda_2^{*\prime}(T_d(s_1)) - H^*(s_1) = 0, \ \forall s_1, \tag{923}$$

   that determines λ_2^* and H^* through ODEs simultaneously.

Step 1: Dual Feasibility

Firstly, check the dual constraint $\mathbb{P}^* \in \mathscr{M}(\mathbb{P}^1, \mathbb{P}^2)$. Check the marginals by taking f to be any continuous bounded function, and denoting the law of (S_1, S_2) under $\mathbb{P}^*$ to be the same as the law of (X, Y).

$$\int f(s_1)\, d\mathbb{P}^*(s_1, s_2) = \mathbb{E}f(X) = \mathbb{E}_{\mathbb{P}^1} f(S_1), \tag{924}$$

$$\int f(s_2)\, d\mathbb{P}^*(s_1, s_2) = \mathbb{E}f(Y) = \mathbb{E}[\mathbb{E}(f(Y)|X)] = \mathbb{E}\left(f(T_u(X))q(X) + f(T_d(X))[1 - q(X)]\right). \tag{925}$$

We hope to prove $\mathbb{E}\left(f(T_u(X))q(X) + f(T_d(X))[1 - q(X)]\right) = \mathbb{E}_{\mathbb{P}^2} f(S_2)$. From the definition of optimizers T_u, T_d, when $x > m$,

$$\frac{d}{dx}[(F_2 - F_1)(T_d(x))] = [F_2'(T_d(x)) - F_1'(T_d(x))]T_d'(x) = -[1-q(x)]F_1'(x), \tag{926}$$
$$\frac{d}{dx}[F_2(T_u(x))] = F_1'(x) + [(F_2' - F_1')(T_d(x))]T_d'(x) = q(x)F_1'(x). \tag{927}$$

Replacing $q(x), 1-q(x)$ with F_1, F_2 yields

$$\mathbb{E}\left(f(T_u(X))q(X) + f(T_d(X))[1-q(X)]\right) \tag{928}$$
$$= \int \left(f(T_u(x))q(x) + f(T_d(x))[1-q(x)]\right)\, dF_1(x) \tag{929}$$
$$= \int_{-\infty}^{m} f(x)\, dF_1(x) + \int_m^{+\infty} f(T_u(x))\, dF_2(T_u(x)) - \int_m^{+\infty} f(T_d(x))\, d(F_2 - F_1)(T_d(x)). \tag{930}$$

When $x > m$, $F_2'(T_u(x))T_u'(x) = q(x)F_1'(x) \geq 0$ so T_u is increasing, while $T_d' \leq 0$ so T_d is decreasing (left-monotone). Applying change of variables while noticing $T_u(m) = T_d(m) = m$ yields

$$\mathbb{E}\left(f(T_u(X))q(X) + f(T_d(X))[1-q(X)]\right) \tag{931}$$
$$= \int_{-\infty}^{m} f(x)\, dF_1(x) + \int_m^{+\infty} f(y)\, dF_2(y) + \int_{-\infty}^{m} f(y)\, d(F_2 - F_1)(y) = \int f(y)\, dF_2(y). \tag{932}$$

So far, we have proved that $\mathbb{P}^*$ is indeed a coupling. Then check the martingale constraint:

$$\mathbb{E}_{\mathbb{P}^*}(S_2|S_1 = s_1) = \mathbb{E}(Y|X = s_1) = q(s_1)T_u(s_1) + [1-q(s_1)]T_d(s_1) = s_1. \tag{933}$$

So far, we have checked the dual feasibility.

Step 2: Primal Feasibility

Secondly, we check that the primal optimizers satisfy the primal constraint:

$$\lambda_1^*(s_1) + \lambda_2^*(s_2) + H^*(s_1)(s_2 - s_1) - c(s_1, s_2) \geq 0,\ \forall (s_1, s_2). \tag{934}$$

It is hard to directly prove this fact. Instead, we check the special cases on the support of $\mathbb{P}^*$: when $s_2 = T_u(s_1)$ or $s_2 = T_d(s_1)$. We only perform computations for the case $s_2 = T_u(s_1)$ and the other counterpart can be computed similarly. Use the definition of λ_1^* to get:

$$\lambda_1^*(s_1) + \lambda_2^*(T_u(s_1)) + H^*(s_1)(T_u(s_1) - s_1) - c(s_1, T_u(s_1)) \tag{935}$$
$$= \left[\lambda_2^*(T_u(s_1)) - \lambda_2^*(T_d(s_1)) + c(s_1, T_d(s_1)) - c(s_1, T_u(s_1))\right][1-q(s_1)] + H^*(s_1)(T_u(s_1) - s_1). \tag{936}$$

This expression takes value zero when $s_1 \leq m$. To investigate the case where $s_1 > m$, it suffices to investigate the derivative. Differentiating both sides w.r.t. s_1 yields

$$\begin{aligned} D_{s_1} := \Big[&\lambda_2^{*\prime}(T_u(s_1))T_u'(s_1) - \lambda_2^{*\prime}(T_d(s_1))T_d'(s_1) + c_1(s_1, T_d(s_1)) + c_2(s_1, T_d(s_1))T_d'(s_1) \\ &- c_1(s_1, T_u(s_1)) - c_2(s_1, T_u(s_1))T_u'(s_1)\Big][1 - q(s_1)] \\ &- [\lambda_2^*(T_u(s_1)) - \lambda_2^*(T_d(s_1)) + c(s_1, T_d(s_1)) - c(s_1, T_u(s_1))]\, q'(s_1) \\ &+ H^{*\prime}(s_1)(T_u(s_1) - s_1) + H^*(s_1)(T_u'(s_1) - 1). \end{aligned} \tag{937}$$

Plug in the definition of $\lambda_2{}', H^{*\prime}, q$. Recall that when $s_1 > m$, $T_u(s_1) > m$, $T_d(s_1) < m$, hence $\lambda_2^{*\prime}(T_u(s_1)) = c_2(s_1, T_u(s_1)) - H^*(s_1)$ and $\lambda_2^{*\prime}(T_d(s_1)) = c_2(s_1, T_d(s_1)) - H^*(s_1)$. We get

$$\begin{aligned} D_{s_1} = \Big[&\cancel{c_1(s_1, T_d(s_1)) - c_1(s_1, T_u(s_1))} + H^*(s_1)(T_d'(s_1) - T_u'(s_1))\Big] \frac{T_u(s_1) - s_1}{T_u(s_1) - T_d(s_1)} \\ &- [\lambda_2^*(T_u(s_1)) - \lambda_2^*(T_d(s_1)) + c(s_1, T_d(s_1)) - c(s_1, T_u(s_1))]\, q'(s_1) \\ &+ \cancel{\frac{c_1(s_1, T_u(s_1)) - c_1(s_1, T_d(s_1))}{T_u(s_1) - T_d(s_1)}}(T_u(s_1) - s_1) + H^*(s_1)(T_u'(s_1) - 1). \end{aligned} \tag{938}$$

Check the coefficient of $H^*(s_1)$, which is $[1-q(s_1)]T_d'(s_1) + q(s_1)T_u'(s_1) - 1 = -(T_u(s_1) - T_d(s_1))q'(s_1)$. The derivative admits the simplification

$$\begin{aligned} D_{s_1} &= [-(T_u(s_1) - T_d(s_1))q'(s_1)]H^*(s_1) - \left[\int_{T_d(s_1)}^{T_u(s_1)} \lambda_2^{*\prime}(y) - c_2(s_1, y)\, dy\right] q'(s_1) \\ &= [-(T_u(s_1) - T_d(s_1))q'(s_1)]H^*(s_1) \\ &\quad - \left[\int_{T_d(s_1)}^{m} c_2(T_d^{-1}(y), y) - H^*(T_d^{-1}(y)) - c_2(s_1, y)\, dy + \int_m^{T_u(s_1)} c_2(T_u^{-1}(y), y) - H^*(T_u^{-1}(y)) - c_2(s_1, y)\, dy\right] q'(s_1) \\ &= [-(T_u(s_1) - T_d(s_1))q'(s_1)]H^*(s_1) \\ &\quad - \Big[-\int_m^{s_1} \Big[c_2(z, T_d(z)) - H^*(z) - c_2(s_1, T_d(z))\Big]T_d'(z)\, dz + \int_m^{s_1} \Big[c_2(z, T_u(z)) - H^*(z) - c_2(s_1, T_u(z))\Big]T_u'(z)\, dz\Big] q'(s_1) \\ &= [-(T_u(s_1) - T_d(s_1))q'(s_1)]H^*(s_1) \\ &\quad - \left[\int_m^{s_1}\int_z^{s_1} c_{12}(x, T_d(z))T_d'(z) - c_{12}(x, T_u(z))T_u'(z)\, dx\, dz + \int_m^{s_1} H^*(z)(T_d'(z) - T_u'(z))\, dz\right] q'(s_1), \end{aligned} \tag{939}$$

where we plug in $\lambda_2^{*\prime}$, conduct a change of variables, and rewrite the differences in terms of integrals. Doing integration by parts for the red part and plugging in $H^{*\prime}$ yields

$$\int_m^{s_1} H^*(z)(T_d'(z) - T_u'(z))\, dz = H^*(s_1)(T_d(s_1) - T_u(s_1)) + \int_m^{s_1} c_1(z, T_u(z)) - c_1(z, T_d(z))\, dz. \tag{940}$$

Combining both equations allows us to cancel the green terms and yield

$$D_{s_1} = -q'(s_1)\left[\int_m^{s_1}\int_z^{s_1} c_{12}(x,T_d(z))T_d'(z) - c_{12}(x,T_u(z))T_u'(z)\,dx\,dz + \int_m^{s_1} c_1(z,T_u(z)) - c_1(z,T_d(z))\,dz\right]. \quad (941)$$

Applying Fubini's theorem for the blue term yields

$$\int_m^{s_1}\int_z^{s_1} c_{12}(x,T_d(z))T_d'(z) - c_{12}(x,T_u(z))T_u'(z)\,dx\,dz \quad (942)$$
$$= \int_m^{s_1}\int_m^{x} c_{12}(x,T_d(z))T_d'(z) - c_{12}(x,T_u(z))T_u'(z)\,dz\,dx \quad (943)$$
$$= \int_m^{s_1} c_1(x,T_d(x)) - c_1(x,m) - c_1(x,T_u(x)) + c_1(x,m)\,dx \quad (944)$$
$$= \int_m^{s_1} c_1(x,T_d(x)) - c_1(x,T_u(x))\,dx, \quad (945)$$

which exactly cancels out the second integral! At this point, we have proved that when $s_2 = T_u(s_1)$, $D_{s_1}\equiv 0$ is constantly zero. As a result, the primal constraint is always tight on the support of $\mathbb{P}^*$, i.e.,

$$\lambda_1^*(s_1) + \lambda_2^*(T_u(s_1)) + H^*(s_1)(T_u(s_1) - s_1) - c(s_1,T_u(s_1)) = 0, \quad (946)$$
$$\lambda_1^*(s_1) + \lambda_2^*(T_d(s_1)) + H^*(s_1)(T_d(s_1) - s_1) - c(s_1,T_d(s_1)) = 0. \quad (947)$$

On the other hand, by the definition of ${\lambda_2}^*$,

$$c_2(s_1,T_u(s_1)) - \lambda_2^{*\prime}(T_u(s_1)) - H^*(s_1) = 0,\ \forall s_1, \quad (948)$$
$$c_2(s_1,T_d(s_1)) - \lambda_2^{*\prime}(T_d(s_1)) - H^*(s_1) = 0,\ \forall s_1. \quad (949)$$

With those tools in hands, it turns out that we can prove the primal feasibility. For any (s_1,s_2), WLOG assume $s_1\ge m, s_2 > T_u(s_1)\ge m$ (the T_d counterpart can be similarly proved). The key equations that are used below are colored in red above. By the first red equation,

$$P := \lambda_1^*(s_1) + \lambda_2^*(s_2) + H^*(s_1)(s_2 - s_1) - c(s_1,s_2) \quad (950)$$
$$= -\lambda_2^*(T_u(s_1)) - H^*(s_1)(T_u(s_1) - s_1) + c(s_1,T_u(s_1)) + \lambda_2^*(s_2) + H^*(s_1)(s_2 - s_1) - c(s_1,s_2). \quad (951)$$

Representing the differences as integrals and replacing H^* with the second red equation yields

$$P = \int_{T_u(s_1)}^{s_2} \lambda_2^{*\prime}(y) - c_2(s_1,y)\,dy - \int_{T_u(s_1)}^{s_2} \lambda_2^{*\prime}(T_u(s_1)) - c_2(s_1,T_u(s_1))\,dy. \quad (952)$$

Plugging in the definition of $\lambda_2^{*\prime}$ yields

$$P = \int_{T_u(s_1)}^{s_2} c_2(T_u^{-1}(y), y) - H^*(T_u^{-1}(y)) - c_2(s_1, y) - \cancel{c_2(s_1, T_u(s_1))} + H^*(s_1) + \cancel{c_2(s_1, T_u(s_1))}\, dy. \tag{953}$$

Rewrite it as a double integral:

$$P = \int_{T_u(s_1)}^{s_2} \int_{s_1}^{T_u^{-1}(y)} c_{12}(x, y) - H^{*\prime}(x)\, dx\, dy. \tag{954}$$

By definition, $H^{*\prime}(x) = \frac{c_1(x, T_u(x)) - c_1(x, T_d(x))}{T_u(x) - T_d(x)}$, using the intermediate value theorem,

$$\exists z(x) \in [T_d(x), T_u(x)],\ H^{*\prime}(x) = c_{12}(x, z(x)). \tag{955}$$

This results in the intermediate value theorem once more

$$P = \int_{T_u(s_1)}^{s_2} \int_{s_1}^{T_u^{-1}(y)} c_{12}(x, y) - c_{12}(x, z(x))\, dx\, dy \tag{956}$$

$$= \int_{T_u(s_1)}^{s_2} \int_{s_1}^{T_u^{-1}(y)} c_{122}(x, \xi(x))(y - z(x))\, dx\, dy \geq 0, \tag{957}$$

for some $\xi(x)$. Since $y \geq T_u(x) \geq z(x)$, the primal constraint is satisfied under the condition $c_{122} > 0$.

Compute the primal objective:

$$\mathbb{E}_{\mathbb{P}^1} \lambda_1^*(S_1) + \mathbb{E}_{\mathbb{P}^2} \lambda_2^*(S_2) = \mathbb{E}_{\mathbb{P}^*}[c(S_1, S_2) - \lambda_2^*(S_2)] + \mathbb{E}_{\mathbb{P}^2} \lambda_2^*(S_2) = \mathbb{E}_{\mathbb{P}^*}[c(S_1, S_2)]. \tag{958}$$

The primal and dual objective align with each other.

Step 3: Complementary Slackness

The consistency conditions hold for those optimizers. By the KKT conditions, the optimality can be established if the complementary slackness condition holds, i.e., (s_1, s_2) is in the support of $\mathbb{P}^*$ iff $\lambda_1^*(s_1) + \lambda_2^*(s_2) + H^*(s_1)(s_2 - s_1) - c(s_1, s_2) = 0$. From the expression of P in step 2, since $c_{122} > 0$, the primal constraint is tight iff $s_2 = T_u(s_1)$, which is equivalent to saying (s_1, s_2) lies on the graph of T_u. The counterpart for T_d can be similarly proved.

Final remark: The primal constraint still holds when $c_{122} \geq 0$, but the complementary slackness condition fails (sufficient but not necessary). In this case, we can still use the weak duality to prove this theorem (obvious). □