

Lecture Notes for 5765/6895, Part II

The choices of material we cover after Chapter 5 have been more flexible, reflecting recent developments and our own interest. In the following, we discuss and summarize the materials under various topic listings.

1 Stochastic Volatility

1.1 Motivations

One obvious and major limitation in the classic Black-Scholes-Merton model is its assumption that the stock price follows a geometric Brownian motion with constant volatility. Even though there is no perfect way to determine the volatility of a stock, one thing we know for sure is that it varies in time in some *random* fashion. The implication of the constant volatility assumption leads to a log price that is supposed to have a normal distribution, a claim that is easily invalidated in reality. For actual pricing and trading of stock derivatives, this limitation gives rise to the phenomenon of so-called “volatility smile” or “volatility skew”, where the implied volatilities observed on the market vary according to the strike price and the expiration of individual contracts, which means that the most important input (underlying volatility) in pricing an option needs to be adjusted in practice. The overall purposes of developing stochastic volatility models are twofold: we want to model the actual volatility as realistic as possible, so that the stock price distribution used in the model is close to the observed data (for instance, the tails of the lognormal distribution are too thin for most observed stocks so other distributions are preferred); at the same time we want to develop a tool that fills in the gap in the implied volatility data set so that an appropriate volatility value can be used in derivative pricing and hedging.

First of all, let us revisit the concept of implied volatility σ_{imp} , which is a value that is associated with a call or put price. In the case of a call, this value is obtained by solving the following equation:

$$c_{BS}(t, S(t); K, T, \sigma_{imp}) = c_{market}, \quad (1)$$

where c_{market} is the observed call price, and c_{BS} is the Black-Scholes-Merton formula for European call. A similar relation defines the implied volatility for a put. The need for stochastic volatility becomes obvious when we notice that σ_{imp} would be different for different strike K , everything else being equal. This clearly contradicts the assumptions of the Black-Scholes model, where σ is just the volatility for the underlying stock, which has nothing to do with the strike price, and it introduces an extra source of ambiguity in option pricing. For example, if a 3-month call on stock X with strike 50 is being traded at implied volatility of 20%, while a similar call on the same stock with strike 45 is being quoted at 22%, what σ value would you use in the Black-Scholes-Merton formula when you need to sell another call at strike 55? As we know that the volatility is the single most crucial

parameter in pricing, this lack of preciseness will be pivotal in the eventual profit and loss of the trading.

When we develop stochastic volatility models, we need to keep in mind some key phenomena observed in stock price data: (1) volatility clustering, and (2) the common highly peaked, fat tail stock return distributions. One breakthrough in option pricing in the last 20 years is the realization of the fact that volatility as a process is mean reverting in time: abnormally high or low levels of volatility cannot be sustained for very long time, so mean-reverting processes (such as Ornstein-Uhlenbeck) are natural candidates for modeling stochastic volatility.

1.2 A General Form

A general form of stochastic volatility models can be expressed as a process for the stock price

$$\frac{dS(t)}{S(t)} = \mu dt + \sqrt{V(t)} dW_1(t), \quad (2)$$

where $V(t)$ is the instantaneous variance of stock return at time t , and a process for V itself

$$dV(t) = \alpha(S(t), V(t), t) dt + \eta\beta(S(t), V(t), t) \sqrt{V(t)} dW_2(t) \quad (3)$$

with a correlation between two Brownian motions W_1 and W_2 given by

$$\langle dW_1, dW_2 \rangle = \rho dt, \quad -1 \leq \rho \leq 1. \quad (4)$$

Notice that we use a general square root process for $V(t)$ here since we require V to remain positive given $V(0) > 0$. This will become important in the analytics later.

1.3 Pricing Equation

To derive the pricing equation for derivatives under the stochastic volatility model, we follow a similar no-arbitrage based approach to form a portfolio that will be hedged for risks in both stock price and volatility. More precisely, we seek a portfolio that is free of both W_1 and W_2 components. To do that, we need the Itô's formula in two dimensions and apply it to the price of a portfolio that consists of two derivatives and the underlying itself,

$$\Pi(t) = U(t) - \Delta(t)S(t) - \Delta_1(t)U_1(t), \quad (5)$$

where $U(t)$ is the price of the asset in question, and U_1 is the price of another asset that is dependent on the same underlying $S(t)$. The reason for two derivative assets in the portfolio is obvious: we have two risk factors, so more than one asset is needed in order to cancel the multi-factors in their prices. A key point in differentiating $\Pi(t)$ is to notice that we assume self-financing, therefore

$$d\Pi = dU - \Delta dS - \Delta_1 dU_1 \quad (6)$$

and the dependences are on S and V . After the application of Itô's formula and regrouping, we have two terms with W_1 and W_2 dependences, and we can set them both to zero by requiring

$$\frac{\partial U}{\partial S} - \Delta_1 \frac{\partial U_1}{\partial S} = \Delta, \quad (7)$$

$$\frac{\partial U}{\partial V} - \Delta_1 \frac{\partial U_1}{\partial V} = 0. \quad (8)$$

With risk factor dependences eliminated, we must have a riskless return for the portfolio:

$$d\Pi = r\Pi dt \quad (9)$$

where r is the riskless rate. However, this equation alone does not directly lead to a PDE for U or U_1 , as there are two unknown functions U and U_1 involved. Another important observation is to be made, which is motivated by a separation of variable technique. Here we need to separate U terms and U_1 terms into two sides of the equation, such as

$$\mathcal{L}U = \mathcal{L}_1U_1, \quad (10)$$

for some differential operators \mathcal{L} and \mathcal{L}_1 . Since U and U_1 are arbitrary derivative prices, both sides must be independent of either U or U_1 and we have

$$\mathcal{L}U = \mathcal{L}_1U_1 = -f(S, V, t). \quad (11)$$

Motivated by the definition of market price of risk, a specific form of f is chosen:

$$f = \alpha - \phi\beta\sqrt{V} \quad (12)$$

Notice that α and β notations are generic and they are not necessarily related to Eq.(3). The rationale behind this form is the following: if we form a delta hedged portfolio

$$\Pi_1 = U - U_S S \quad (13)$$

and the excess return over dt is

$$d\Pi_1 - r\Pi_1 dt = \beta\sqrt{V}U_V (\phi dt + \eta dW_2). \quad (14)$$

Here ϕ represents the excess return over the risk-less interest, scaled by the volatility of volatility η . It is natural to call it the market price of risk in regard to the volatility factor. As in the risk-neutral setting for constant volatility model, if we set $\phi = 0$, the PDE for U can be written as

$$\frac{\partial U}{\partial t} + \frac{1}{2}VS^2\frac{\partial^2 U}{\partial S^2} + \rho\eta\beta SV\frac{\partial^2 U}{\partial S\partial V} + \frac{1}{2}\eta^2\beta^2V\frac{\partial^2 U}{\partial V^2} + rS\frac{\partial U}{\partial S} + \alpha\frac{\partial U}{\partial V} = rU \quad (15)$$

1.4 Heston's Model (1993)

The most popular model in the above form is probably Heston's model (1993), where a mean-reversion is built into the volatility process:

$$\frac{dS(t)}{S(t)} = \mu dt + \sqrt{V(t)} dW_1(t), \quad (16)$$

$$dV(t) = \kappa(\theta - V)dt + \eta\sqrt{V(t)}dW_2(t). \quad (17)$$

We can see that we just need to set α to this mean reverting form, and $\beta = 1$ in Eq.(3). One advantage of this model is that we can easily interpret various parameters: κ represents the speed of mean reversion, and θ is the long time equilibrium of V . It should be pointed out that the process for V is just one example of the CIR process, where a crucial condition for the parameters is known ($\kappa\theta \geq \eta^2/2$) in order to make sure that V stays above zero if it starts out positive. The resulting PDE for Heston's model is

$$\frac{\partial U}{\partial t} + \frac{1}{2}VS^2\frac{\partial^2 U}{\partial S^2} + \rho\eta SV\frac{\partial^2 U}{\partial S\partial V} + \frac{1}{2}\eta^2\beta^2V\frac{\partial^2 U}{\partial V^2} + rS\frac{\partial U}{\partial S} + \kappa(\theta - V)\frac{\partial U}{\partial V} = rU \quad (18)$$

The success of this model is due to the fact that a closed-form solution to the above equation exists for the European call and put. There are several steps involved in solving this equation for the particular form of payoff functions (call or put).

- Step 1: Introduce changes of variables

$$x = \log(Se^{-r(T-t)}/K), \quad \tau = T - t, \quad c(x, V, \tau) = U(S, V, t),$$

so we can eliminate the variable dependence in the coefficients.

- Step 2: Motivated by the Black-Scholes formula, suggest a solution form

$$c(x, V, \tau) = K [e^x P_1(x, V, \tau) - P_0(x, V, \tau)],$$

so we have P_1 and P_0 satisfying two equations that are similar.

- Step 3: The terminal conditions for P_1 and P_0 have a nice form

$$\lim_{\tau \rightarrow 0} P_j = H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad j = 0, 1. \quad (19)$$

- Step 4: After the treatment regarding the terminal conditions, Fourier transform can be used to convert the PDE into several ODEs.
- Step 5: Once the ODEs are solved, take the inverse Fourier transform to express the solution to the original problem in a principal-value integral. The calculation is rather tedious but it is indeed a closed-form solution.

It should be noted that in many applications the condition $\kappa\theta \geq \eta^2/2$ is sometimes violated. One treatment to fix this artifact is to introduce a boundary condition (such as the reflection condition) for the V process, and impose a corresponding condition for the PDE at $V = 0$.

1.5 Volatility Smile

As a simple example that indeed stochastic volatility models can predict the smile observed on the market, we consider a toy model where the volatility is a Bernoulli random variable:

$$\bar{\sigma}^2 = \begin{cases} \sigma_1^2 & \text{with probability } p \\ \sigma_2^2 & \text{with probability } 1 - p \end{cases} \quad (20)$$

So the implied volatility is solved from

$$c_{BS}(K, \sigma_{imp}(p, K)) = p c_{BS}(K, \sigma_1) + (1 - p) c_{BS}(K, \sigma_2) = c_{market} \quad (21)$$

After some clever manipulations, it can be shown that

$$\frac{\partial^2 \sigma_{imp}}{\partial K^2} \geq 0.$$

So indeed there is a “smile” in the implied volatility curve.

2 Jump Diffusion

2.1 Introducing Poisson jump process

Stock prices are known to jump at unusual moments, such as at major political or economical events. It is natural to address the continuous path restriction in a diffusion process by introducing a jump component. We begin with adding Poisson jumps to a diffusion process:

$$X(t) = X(0) + \int_0^t \Delta(s) dW(s) + N(t), \quad (22)$$

where $\Delta(t)$ is an adapted process (depending on information up to t), and $N(t)$ is the number of jumps before t . The jump times $\tau_k, k = 0, 1, \dots$ with $\tau_0 = 0$ have the property that the time differences in between consecutive jumps $\tau_k - \tau_{k-1}$ are independent exponential rv's with parameter λ , which will be called the jump intensity that measures how often jumps appear in time. With the jump times defined, the number of jumps before t can be written as

$$N(t) = \max \left\{ n : \sum_{k=0}^n \tau_k \leq t \right\} \quad (23)$$

and we can show the following properties:

1. the distribution of $N(t)$ is given by

$$\mathbb{P} \{N(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots \quad (24)$$

2. Independent increment: $N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$ are stationary and independent, for any sequence $t_0 < t_1 < t_2 < \dots < t_n$.
3. $\mathbb{E}[N(t) - N(s)] = \text{Var}[N(t) - N(s)] = \lambda(t - s), \quad t > s$
4. $M(t) = N(t) - \lambda t$ is a martingale.

In the Poisson jump process $N(t)$, every time there is a jump, $N(t)$ gains an additional increment of size one. It is necessary to introduce processes with random jump sizes. For this reason, we consider the so-called compound Poisson process

$$Q(t) = \sum_{i=1}^{N(t)} Y_i \quad (25)$$

where rv's Y_i are described by specifying their distributions. If we assume $\mathbb{E}[Y_i] = \beta$, then

$$\mathbb{E}[Q] = \sum_{k=1}^{\infty} \mathbb{E} \left[\sum_{i=1}^k Y_i | N(t) = k \right] \cdot \mathbb{P}[N(t) = k] = \beta \lambda t \quad (26)$$

To conclude this subsection, we write the class of jump-diffusion process that we will consider here as

$$X(t) = X(0) + \int_0^t \Delta(s) dW(s) + Q(t). \quad (27)$$

2.2 Itô's formula for Compound Poisson Processes

Starting with a jump diffusion process $X(t)$, if we have a function evaluated at $X(t)$, we will need to study the differential of that function, in terms of the changes in X . In another word, we will need a version of Itô's formula for $X(t)$ that is described by Eq.(27). First we assume the simple Poisson process and notice

$$f(N(t)) = f(N(0)) + \int_0^t df$$

where

$$df = \begin{cases} f(N(t)) - f(N(t-)) & \text{jump at } t \\ 0 & \text{no jump at } t \end{cases}$$

Therefore,

$$f(N(t)) = f(N(0)) + \sum_{0 \leq s \leq t} (f(N(s)) - f(N(s-))),$$

where the sum is over all possible jump times s . It is straightforward to extend this to

$$f(Q(t)) = f(Q(0)) + \sum_{0 \leq s \leq t} (f(Q(s)) - f(Q(s-))) \quad (28)$$

for compound process Q . For more general jump diffusion processes (Brownian motion plus Poisson jumps), there is a version of Itô's formula that combines the above with the original formula for diffusion alone.

2.3 Merton's Model (1976) for Pricing European Options

Merton proposed the following model for stock prices that allow jumps

$$\frac{dS(t)}{S(t-)} = (\alpha - \lambda k) dt + \sigma dW(t) + dQ(t) \quad (29)$$

where $k = \mathbb{E}[Y - 1]$ and the term λt are there to be consistent with the notion that α is still the expected growth rate. Another way to see this equation is

$$\frac{dS(t)}{S(t-)} = \begin{cases} (\alpha - \lambda k) dt + \sigma dW(t) & \text{no jump at } t \\ (\alpha - \lambda k) dt + \sigma dW(t) + Y - 1 & \text{jump at } t \end{cases}$$

The reason to use $Y - 1$ instead of Y can be seen in the case

$$\frac{dS(t)}{S(t-)} = \frac{S(t) - S(t-)}{S(t-)} = Y - 1$$

which implies

$$S(t) = S(t-) + (Y - 1)S(t-) = YS(t-).$$

It is clear that the random variable Y models the ratio of the stock price when jumps happen, so it is necessary to use a distribution with positive values. In Merton's original model, the lognormal distribution is chosen for Y such that $Y = (1 + k) \exp(-\frac{1}{2}\nu^2 + \nu Z)$ and Z is the standard normal random variable. The solution to Eq.(29) is therefore

$$S(t) = S(0)Y(N(t)) \exp \left[\left(\alpha - \frac{\sigma^2}{2} - \lambda k \right) t + \sigma W(t) \right] \quad (30)$$

where

$$Y(n) = \begin{cases} 1 & n = 0, \\ \prod_{j=1}^n Y_j & n \geq 1, \end{cases} \quad (31)$$

and $N(t)$ is the number of jumps before time t .

One immediate calculation from this model is the expected payoff under a proper measure, such as the call price

$$\begin{aligned} c(0, S(0)) &= \mathbb{E} \left[e^{-rT} (S(T) - K)^+ \right] \\ &= \sum_{j=0}^{\infty} \mathbb{E} \left[e^{-rT} (S(T) - K)^+, N = j \right] \cdot \mathbb{P} [N = j] \\ &= \sum_{j=0}^{\infty} \frac{(\lambda T)^j e^{-\lambda T}}{j!} \mathbb{E} \left[e^{-rT} (S(T) - K)^+, N = j \right] \\ &= e^{-\lambda T} \sum_{j=0}^{\infty} \frac{(\lambda T)^j}{j!} \mathbb{E} \left[c_{BS} (S(0)Y(j)e^{-rT}, T, K, \sigma, r) \right] \\ &= e^{-\lambda(1+k)T} \sum_{j=0}^{\infty} \frac{(\lambda(1+k)T)^j}{j!} c_{BS} (S(0), T, K, \sigma_j, r_j) \end{aligned} \quad (32)$$

where

$$\sigma_j = \sqrt{\sigma^2 + \frac{j\nu^2}{T}}, \quad r_j = r - \lambda k + \frac{j \log(1+k)}{T}. \quad (33)$$

The last equation is derived by taking expectation with respect to the rv Y_j and it can be shown that this expectation can be expressed in terms of the Black-Scholes formula again, with modified parameters σ_j and r_j .

To justify this option price, we first try to follow the hedging argument in the original Black-Scholes derivation, but soon realize that it will not work. The difficulty can be explained by considering the simple case where

$$dS = S(t) - S(t-) = (Y - 1)S(t-),$$

and

$$df = f(S(t)) - f(S(t-)) = f(Y S(t-)) - f(S(t-)).$$

No matter what you choose to multiply in front of either dS or df , there will not be a combination that will eliminate the jump change in S , simply because Y is another random variable. In the textbook, it is shown that jump risk could be eliminated if the jump size Y is deterministic, which is an unrealistic assumption. A compromise is made if we want a price process to be a "martingale" under some loose conditions. If we look at the differential of $f(t, S(t))$ when $S(t)$ follows Eq.(29) and make sure that the drift term is zero, we end up with a partial integro-differential equation

$$f_t + (\alpha - \lambda k)Sf_s + \frac{1}{2}\sigma^2 S^2 f_{ss} + \lambda \mathbb{E}[f(Y S) - f(S)] = rf \quad (34)$$

Here the expectation is taken with respect to the random variable Y . This is the equation that leads to the Merton's call price formula when the appropriate terminal condition $f(T, S(T)) = (S(T) - K)^+$ is imposed. In fact, when other payoff functions are imposed, the solution to this equation should give the jump-diffusion model price for that particular derivative.

3 American Options

American options differ from their European counterparts in that they can be exercised *any time* before expiration. Therefore they are at least as valuable as, if not more than, the European counterparts. We have seen in the binomial model that the modification is quite straightforward in implementation, but the analysis, the proof that such policy is indeed optimal, are much more complicated. In fact, a beautiful and readily accessible mathematical presentation of the American option is only available in the form of a free boundary PDE problem, as presented below. We have argued that American calls on stocks that pay no dividends are best to be left to the expiration, so they are as valuable as the European calls. Because of this important observation, most American option discussions start with American

puts, in which case it's indeed optimal to exercise prior to the expiration of the option. The real question is what this criterion to exercise is.

First, we know that the optimal exercise time is not known ahead of time, therefore it is a random variable. Since we cannot rely on the future information to make a decision, this random variable is a **stopping time** with respect to the filtration generated by the process. An exercise policy is a procedure for the investors to follow, so we use the terms “exercise policy” and “stopping time” interchangeably. A natural way to search for a stopping time is to introduce a barrier, with the first time this barrier is crossed being the stopping time. In the American put, it has the form

$$\tau = \min \{t : S(t) \leq L(t)\}$$

where $L(t)$ is a pre-specified barrier and it must satisfy the following conditions

$$L(t) \leq K, \quad L(T) = K.$$

Once we have this function, it is easy for the investors to follow the strategy: at any time t before the expiration, if $S(t)$ is higher than $L(t)$, he/she should hold the put; on the other hand, if $S(t)$ is lower than $L(t)$, he/she should immediately exercise to obtain a payoff $K - S(t)$. The question for the optimal exercise policy is to find the barrier $L(t)$ such that the value of the option is maximized. As it turns out, the problem is solved via the following free boundary value problem, which can be summarized in the following. The region $\{(S, t), 0 \leq t \leq T, 0 < S < \infty\}$ will be divided into two regions by the barrier $L(t)$. In the region above $L(t)$, we have the option price V satisfying

$$V_t + rSV_S + \frac{1}{2}\sigma^2 S^2 V_{SS} = rV \tag{35}$$

$$V \geq (K - S)^+ \tag{36}$$

and in the region below $L(t)$,

$$V_t + rSV_S + \frac{1}{2}\sigma^2 S^2 V_{SS} < rV \tag{37}$$

$$V = K - S \tag{38}$$

This is a typical free boundary value problem where the boundary $L(t)$ itself is part of the solution, and we must be very careful about the boundary conditions at $L(t)$. For the American put, the following conditions are imposed:

$$V \text{ is continuous across } L, \quad \text{and} \quad \lim_{S \rightarrow L} V_S = -1. \tag{39}$$

How should we hedge an American option (pretending you sold the option), given that the price is obtained from the above formulation? The following are the guidelines:

1. As long as the option is not exercise, delta hedge;

2. If the option is exercised exactly when the boundary is crossed, you should be able to break even if hedged properly;
3. If the option holder chooses to exercise when the boundary is not crossed, so the payoff is actually lower than the value of the option, you should just close the option (sell the option on the market and pay the owner the payoff he/she is entitled) and walk away with the difference;
4. If the option holder does not exercise when the barrier is actually crossed, which means that the payoff is higher than the option value, you should cash out the difference and use the remaining balance to delta hedge the option as a new option.

4 Interest Rate Term Structures

The obvious challenge for modeling interest rates is that there are so many related rates on the market, and they must be modeled in a consistent way. To make things more manageable, all these rates can be compared in two aspects: the particular type and the maturity. The maturity is easier to understand: a two-year rate is a rate that is guaranteed for the next two years if you deposit the money with the particular financial institution that offers it. Here are some typical rates you find on the market:

1. one-month LIBOR rate, a measure of rates at which banks make loans to each other for a one-month period;
2. 2-year treasury bill (yield), the yield (to be described below) you received when you loan the money to US treasury for two years;
3. 15-year swap rate, a fixed rate that will be good for 15 years at which you exchange interest payments with some counterparty in exchange of the floating rate;
4. 30-year mortgage rate offered by a particular bank.

Suppose we restrict to one type of rate and just consider the differences in maturity, the goal would be to generate a curve that describes the rates for different maturities. This curve is called the yield curve, and it can be equivalently shown in three forms. Before we describe these three forms, it is important to understand the concept of an important class of financial securities - bonds. A bond is a certificate issued by an issuer to investors that promises to pay back the face value and subsequent coupons stated on the bond at some prearranged dates. The importance of bonds is made all significant in that they are securities and can be freely traded on the market. Imagine that a bond that promises \$100 in 5 years (for the purpose of illustration, assume no coupons) and it is selling at a price of \$50, which implies an annualized rate of about 14% (doubling your money in 5 years) and is all “guaranteed” if the issuer does not default in the next five years.

The tricky part is that the investor does not always hold it until maturity. He can buy the bond at \$50 today and sell it for \$50 again tomorrow. In that case he would lose the interest and end up with a zero return for the investment. On the other hand if the price moves up by one dollar in a week, the realized return would be $2\% \times 365/7 \approx 104\%$, annualized. The game in interest rate markets (so-called fixed-income markets) is to buy and sell these interest rate dependent securities in an environment where all kinds of rates are fluctuating with some form of dependences. With the trading practices in mind, we now introduce the following three descriptions of the yield curve.

1. yield $y(t, T)$: for each dollar deposited at time t , you will receive $e^{y(t, T)(T-t)}$ at the maturity T ;
2. zero-coupon bond price $B(t, T)$: for a zero-coupon bond maturing at T with face value \$1, the price at time $t < T$ is $B(t, T)$;
3. forward rate $f(t, T)$: observed at time t , the implied short rate for depositing at time T .

These three descriptions can be converted from one to another via the following relations.

$$B(t, T) = e^{-y(t, T) \cdot (T-t)} \quad (40)$$

$$f(t, T) = y(t, T) + (T - t) \frac{\partial y}{\partial T} \quad (41)$$

For a fixed time of observation t , the yield curve is a function of maturity T . However, when we focus on the dynamics of the yield curve, we are more concerned with the change in t , while T serves as a parameter. In terms of modeling, we want to describe $dy(t, T)$, $dB(t, T)$, or $df(t, T)$, where the differential is with respect to the time t . As we can see that the parameter T itself will also change, so potentially we will be dealing with a partial stochastic differential equation, which is far beyond the capability of our standard machineries. In the following, we briefly mention two classes of term structure models.

1. Short rate models: we choose to model the instantaneous change of the short rate $R(t)$, which applies to an instantaneous time period and a typical real life example is the overnight rate. Popular choices include
 - Vasicek model:

$$dR(t) = \kappa(\theta - R(t)) dt + \sigma dW(t)$$

- BDT (Black-Derman-Toy) model:

$$dy(t) = (\theta + \beta y(t)) dt + \sigma dW(t), \quad R(t) = \exp(y(t))$$

2. HJM model: a model that describes the dynamics of the forward rate $f(t, T)$. This is considered a complete model as the parameter T explicitly appears in the model and it extends to multidimensions.

We will limit ourselves to the discussion of short rate models, for which the PDE approach is a powerful tool. First we introduce the discount process

$$D(t) = \exp\left(-\int_0^t R(s) ds\right) \quad (42)$$

The usefulness of this process is that for no-arbitrage to hold, $D(t)B(t, T)$ must be a martingale under the risk-neutral measure. As a consequence,

$$B(t, T) = \mathbb{E}\left[\frac{D(T)}{D(t)}|\mathcal{F}(t)\right] \quad (43)$$

The martingale property allows us to derive an equation for the price of any interest rate derivative under this one-factor world assumption - we just need to differentiate and set the drift term to zero. Suppose the short rate model is given by

$$dR(t) = \beta dt + \gamma dW(t),$$

the PDE for the price of the derivative $f(t, R)$ is

$$f_t + \beta f_R + \frac{1}{2}\gamma^2 f_{RR} = Rf \quad (44)$$

with the terminal condition for $f(T, R(T))$ determined based on the derivative payoff. For a zero-coupon bond, the terminal condition is $f = 1$ when $t = T$.

How do we use a model to price interest-rate derivatives? First we choose a model and calibrate it (determine the parameters) to make sure that bond prices from the model are reasonably matched to the market. Next we should use the calibrated model to price other derivatives, such as the interest rate caps/floors and interest swap options. It is in these not-so-common, or structured instruments where the financial institutions can rake huge profits, if they can hedge the products they sell accurately with more common instruments. Equally possible they can also stand to lose a lot of money.

5 Credit Derivatives

One of the important motivations for credit derivatives is the need to address potential defaults of counterparties in various financial contracts, such as an interest rate swap, where a product called credit default swap (CDS) is often used to hedge the credit risk. But the innovators at that time never imagined the widespread use of credit products as a vehicle to speculate and make all kinds of bets.

The mathematical question is how to quantify the risk of a default of a particular entity. Here is a toy example to illustrate the principle: suppose there is a 5-year zero-coupon corporate bond issued by a company that matures in 5 years but the company may default at time τ within the next 5 years, in which case the company will not be able to meet the obligation to fully pay back the principal amount. We

assume the distribution of τ is an exponential distribution with parameter λ , that is

$$\mathbb{P}[\tau \leq 5] = 1 - e^{-5\lambda}.$$

Suppose the risk-free interest rate (annualized) for the 5-year investment is r , and there is no payment in case of a default, the discounted expected payoff at $T = 5$ is therefore

$$P = \mathbb{P}[\tau > 5] \cdot e^{-5r} + \mathbb{P}[\tau \leq 5] \cdot 0 = e^{-5(r+\lambda)}.$$

The yield of this bond is therefore $y = r + \lambda$, and we can see that the extra yield in addition to the risk-free rate is due to the intensity λ .

5.1 Reduced Form Models

Here the focus is on modeling the default intensity λ , for which $\lambda\Delta t$ can be viewed as the likelihood of default over a short period $(t, t + \Delta t)$. It is obvious that the intensity λ must be time-dependent to develop a particular default term structure. Not only λ should be time varying, but also it is most likely stochastic, thus opening the door for many models reflecting the forecast for the credit future of a particular entity, such as one versus multi factors, mean reversion, and so on. In developing a proper model, a balance should be maintained between analytic tractability and realistic and comprehensive description of the phenomena.

5.2 Structural Models

The other type of model, pioneered by Merton (1973), is called structural model, and they are based on the fundamental argument that a default is triggered when the company's total asset falls below the company's total liability. We need to understand the basic corporate financing principle that a company can choose between issuing debt or equity to raise the capital it needs to expand its business. The debt holders are capped at the gain, but they have the priority in collecting the remaining asset if the company goes bust.

Merton's original argument is the following. Let V be the total asset of the company and B the outstanding liability, and they are both time dependent. At maturity T ,

- If $V > B$, the bond holders receive their promised amount B , while the equity holders take away the rest $(V - B)$;
- If $V \leq B$, which is the case where the company would fail their obligations, the bond holders take everything that is left (V) and the stock holders get nothing.

The payment received by the stock holders is therefore exactly the same payoff of a call option

$$(V - B)^+ = \begin{cases} V - B, & V > B \\ 0, & V \leq B \end{cases}$$

So the value of the stock at $t < T$ can be expressed as the Black-Scholes formula (with a proper probability measure), and the value of the defaultable bond at t is

$$B(t) = V(t) - e^{-r(T-t)} \mathbb{E} [(V(T) - B)^+ | \mathcal{F}(t)]$$

This is one of the first clean expressions for modeling defaultable bonds, and the dependence on the volatility as required in the Black-Scholes formula highlights the risk in the company's credit conditions.

An obvious weakness of the model in this form is that defaults are allowed only at the maturity T . The extension of this model leads to an application of the famous "first exit" problem. Let $V(t)$ be the value of the asset at time t , the first time V falls below B

$$\tau = \min \{s : V(s) \leq B\}$$

gives the default time, and the probability of default before t

$$P(t) = \mathbb{P}[\tau < t]$$

as a function of time is the outcome of the model that can be calibrated to the market implied default probabilities. The problem becomes a calibration problem for the process $X(t) = V(t) - B$, that is the determination of the model parameters. In the case where X is a Brownian motion or Geometric Brownian motion, there are well-known results based on the reflection principle. We illustrate this approach by building a series of processes from the very basic.

1. $X(t) = W(t)$, with barrier $m > 0$. We define the exit time

$$\tau_m = \min \{t : X(t) \geq m\}$$

For any $w > 0$, the reflection principle gives

$$\mathbb{P}[\tau_m \leq t, W(t) \leq w] = \mathbb{P}[W(t) \geq 2m - w]$$

Therefore,

$$\begin{aligned} \mathbb{P}[\tau_m \leq t] &= \mathbb{P}[\tau_m \leq t, W(t) \leq m] + \mathbb{P}[\tau_m \leq t, W(t) > m] \\ &= 2\mathbb{P}[W(t) \geq m] \\ &= \frac{2}{\sqrt{2\pi t}} \int_m^\infty e^{-\frac{x^2}{2t}} dx = 2 \left(1 - N \left(\frac{m}{\sqrt{t}} \right) \right). \end{aligned}$$

2. $X(t) = W(t)$, $m \leq 0$. We define

$$\tau_m = \min \{t : X(t) \leq m\}$$

A similar calculation based on symmetry gives

$$\mathbb{P}[\tau_m \leq t] = 2 \left(1 - N \left(-\frac{m}{\sqrt{t}} \right) \right).$$

3. $X(t) = \sigma W(t)$, $m > 0$.

$$\tau_m = \min \{t : X(t) \geq m\}$$

We have

$$\mathbb{P}[\tau_m \leq t] = 2 \left(1 - N \left(\frac{m}{\sigma\sqrt{t}} \right) \right).$$

4. $X(t) = \alpha + \sigma W(t)$ where α is a constant, we note that $X(t) \geq m$ is equivalent to $W(t) \geq \frac{m-\alpha}{\sigma}$ so

$$\mathbb{P}[\tau_m \leq t] = 2 \left(1 - N \left(\frac{|m-\alpha|}{\sigma\sqrt{t}} \right) \right).$$

5. $X(t) = \alpha t + \sigma W(t)$ with constant $\alpha > 0$ and σ . This turns out to be a substantial problem that involves a change of measure:

$$\tilde{W}(t) = W(t) + \frac{\alpha t}{\sigma}$$

The calculation follows the change of measure formula

$$\mathbb{P}[\tau_m \leq t] = \mathbb{E} [I_{\{\tau_m \leq t\}}] = \tilde{\mathbb{E}} \left[\frac{1}{Z} I_{\{\tau_m \leq t\}} \right]$$

5.3 First-to-default (first-to-exit) models

In many CDS/CDO modeling, the first-to-default, or second-to-default issues would arise. Suppose $\tau_1, \tau_2, \dots, \tau_m$ are default times of companies $1, 2, \dots, m$, each with a survival probability

$$p_i = \mathbb{P}[\tau_i > t]$$

How do we determine the distribution of the first-to-default, or first-to-exit time $\tau = \min_i \{\tau_i\}$?

In the case all the company defaults are **independent** from each other,

$$\begin{aligned} \mathbb{P}[\tau > t] &= \mathbb{P}[\tau_1 > t, \tau_2 > t, \dots, \tau_m > t] \\ &= \mathbb{P}[\tau_1 > t] \cdot \mathbb{P}[\tau_2 > t] \cdots \mathbb{P}[\tau_m > t] \\ &= p_1 \cdot p_2 \cdots p_m \end{aligned}$$

This simplifies the problem quite a bit, and it can explain the popularity with such intensity-based models. However, this convenience in practice effectively encourages many practitioners to assume this all-too-important, but not necessary realistic, **independence** assumption. This can be devastating in many applications.

5.4 Copula Model

Now that we realized that the independence assumption is not a reasonable one, and the correlation factor is often a major issue in many structured products such as CDOs, it becomes clear to practitioners that the dominant issue is to model correlations. The copula model makes an often oversimplified attempt to address this issue. It is observed that the random variable τ can be transformed to another random variable with **uniform distribution**: Suppose $p(t) = \mathbb{P}[\tau_i > t]$ is the survival probability, the inverse CDF method suggests that if we take a uniformly distributed rv $U \sim \text{Unif}[0, 1]$, then $\tau = p^{-1}(U)$ has $1 - p(t)$ as its CDF. The idea of the copula model is that instead of working with rv's τ_i with individual distributions and a rather special correlation structure, it would be far simpler to work with transformed rv's U_i , for which the correlation structure may be much easier to specify. The copula model thus changes the problem of imposing a correlation structure for $\tau_1, \tau_2, \dots, \tau_m$ into a correlation structure for U_1, U_2, \dots, U_m . Namely, we call

$$C(u_1, u_2, \dots, u_m) = \mathbb{P}(U_1 \leq u_1, \dots, U_m \leq u_m),$$

a copula function for the transformed rv's U_1, U_2, \dots, U_m . A particular copula model specifies the form of the copula function C . Consider two rv's, for examples,

1. Independence: $C(u, v) = uv$;
2. Perfect correlation: $C(u, v) = \min(u, v)$;
3. Gaussian: $C(u, v) = P(N(X) \leq u, N(Y) \leq v)$ where X, Y are standard joint normal random variables with correlation coefficient ρ , and $N(x)$ is the cumulative normal distribution function.

Gaussian copula model is one of the most popular copula models in which a pair of joint Gaussian rv's with correlation coefficient ρ is simulated (which is easy to do), and they are turned to a pair of uniform distributed U, V , and then further turned to a pair τ_1, τ_2 . The problem, however, is that the correlation between X and Y is **not** the same as the correlation between τ_1 and τ_2 .

6 Exotic Options

Option payoffs can be divided into two categories according to its dependence on the underlying: those on the underlying value at expiration T , or those on all the underlying values over certain time period. The latter is often called path-dependent options as the payoff can depend on the whole path, rather than the ending value of the underlying as in the case of call/put options.

We use barrier options to illustrate some of the common features among exotic options. A barrier option usually has a barrier specified in advance, whether the barrier is crossed in the lifetime of the option will determine the payoff of the option. A knock-out option differs from a straightforward option in that the option can

be knocked out - that is, it can cease to exist, if some barrier $S = B$ is crossed before expiration. For options like that, we need the joint distribution for the random variables $S(t)$ and $M^S(t) = \max\{S(s), s \leq t\}$. For example, the payoff for a up-and-out call is

$$V(T) = (S(T) - K)^+ \cdot \mathcal{I}_{\{M^S(T) < B\}}$$

There is an elegant PDE approach to solve the valuation problem for the price $v(t, S)$ (assuming the standard Black-Scholes model for the underlying S):

$$v_t + rSv_S + \frac{1}{2}\sigma^2 S^2 v_{SS} = rv.$$

The heart of the approach is in its delicate prescription of the boundary conditions for the value function $v(t, S)$:

$$\begin{aligned} v(t, 0) &= 0, & 0 \leq t \leq T \\ v(t, B) &= 0, & 0 \leq t \leq T \\ v(T, S) &= (S - K)^+, & 0 \leq S \leq B \end{aligned}$$

It should be pointed out that this price $v(t, S)$ is the price of the call when S has not reached B before t .

7 Monte Carlo Methods

The power of Monte Carlo methods is that almost everything can be simulated, but the efficiency is the drag: the convergence to the true expectation can be slow. The problem can be illustrated by observing the central limit theorem: if $X_1, X_2, X_3, \dots, X_N$ are i.i.d's with mean μ and variance σ^2 , then $\frac{X_1 + X_2 + X_3 + \dots + X_N - N\mu}{\sqrt{N}}$ converges in distribution to $N(0, \sigma^2)$. We can rephrase this as

$$\begin{aligned} & \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} \\ = & \mu + \frac{X_1 + X_2 + X_3 + \dots + X_N - N\mu}{N} \\ = & \mu + \frac{1}{\sqrt{N}} \frac{X_1 + X_2 + X_3 + \dots + X_N - N\mu}{\sqrt{N}} \\ \rightarrow & \mu + N(0, \sigma^2/N) \end{aligned}$$

The error has a normal distribution of mean zero and variance σ^2/N . In another word, the standard deviation of the error is on the order of $1/\sqrt{N}$. Obviously this is not very satisfactory for reducing the error: if you double N , you are only going to **expect** the error to be reduced by a factor of $\sqrt{2}$, **on average**.

Most efforts made in improving the efficiency of Monte Carlo methods are therefore centered around reducing the variance, so-called the variance reduction techniques. The simplest one is called antithetic variate approach: suppose the distribution is

symmetric (X and $-X$ have the same probability being sampled), then for each sample value X , we will also include $-X$ in the simulations. This method leads to unbiased approximation at a minimum cost.

One of the main challenges in valuing derivatives using Monte Carlo methods is the American Option problem. The reason is obvious: for each point in a future time, you will need to simulate all the branches from that point to determine if you should exercise or not, and you will need to do this at every point in time. This made Monte Carlo methods almost impractical, until the Least-Square Monte Carlo approach came along (Longstaff-Schwartz, 2001). The approach is along the same line as the algorithm in binomial models, where a comparison is made at each node. Suppose at a time before expiration, the underlying has value X , the immediate exercise $f(X)$, and the value of continuation $E(Y|X)$ where Y is the value of the option if continued. The contribution of the LSM algorithm is in its approximation of the function $g(X) = E(Y|X)$ and the idea is the following. If we can approximate $g(X)$ by a quadratic function

$$g(X) = C_0 + C_1X + C_2X^2,$$

then we can argue (later prove) that the coefficients should be chosen so

$$\sum_k (C_0 + C_1X_k + C_2X_k^2 - Y_k)^2 = \min$$

where X_k and Y_k are the value of the underlying at early time t and the value of the option if continued along the path k . The determination based on a least-square problem gives rise to the term Least Square Monte Carlo (LSM). In actual implementations, more sophisticated function forms are used but the principle is the same. To build this tree-like structure, just like in a binomial tree, you start from the end, and move one step at a time till the beginning. The striking feature of the algorithm is that the same paths are generated and used at all different stages, thus avoiding the need to branch paths at later times. For implementation details, the numerical example in the original Longstaff/Schwartz paper is quite illuminating.

8 Statistical Arbitrage

Statistical arbitrage strategies refer to a class of strategies that often rely on some combined long-short positions, where you buy some stocks and sell some others to take advantage of some statistical features of the return differences. The key is the believe that certain features, such as mean-reversion, exist in the return difference, rather than the return of a stock by itself. The real question is how to find such pairs with a difference exhibiting such behaviors. Typically we will be looking at the pairing of relevant stocks, ETF relative to the index, etc. To find proper pairs, we will need to conduct substantial statistical studies on the differences/spreads between two or more correlated stocks/indices.

To give some idea, let us consider the following strategy based on a model for a stock and the index that includes the stock. Here S denotes a particular stock price, I is the relevant index, and we assume that the two are related via

$$\frac{dS(t)}{S(t)} = \beta \frac{dI(t)}{I(t)} + \epsilon(t)$$

$$\epsilon(t) = \alpha dt + dX(t)$$

$$dX = \kappa(m - X(t))dt + \sigma dW(t)$$

The reading is that the instantaneous returns of the stock and the index are proportional to each other, with some residual ϵ . The residual is supposed to have two components: a drift and a random portion, which is driven by a mean-reversion process. The idea is that when X is too high, it's about to go down, namely, the difference between the returns is getting smaller. Here is a quick example: suppose we know that S daily returns double the index daily returns on average. On a particular day, say the S return is 1.4% while the index return is 0.5%, we see that $\epsilon = 0.4\%$ so the X value is quite high, and we believe that it cannot be sustained so we can start selling the stock and buy the index. When the X value is reduced to certain level (say the S return is 1.6% and the index return is 0.8%), we can then close these two positions and rake the profit.

How should we generate signals to buy or sell? Here is one example that introduces this so-called s -score

$$s(t) = \frac{X(t) - m}{\sigma}$$

and we notice that this score is centered around the mean reversion level m and scaled properly by the volatility of the residual (not the stock itself). A typical strategy runs like the following:

- open long position (buy one share of S , sell $\beta S/I$ shares of the index), if $s < -a$;
- open short position (sell one share of S , buy $\beta S/I$ shares of the index), if $s > a$;
- close long position (sell one share of S , buy $\beta S/I$ shares of the index), if $s > -b$;
- close short position (buy one share of S , sell $\beta S/I$ shares of the index), if $s < b$;

a and b ($a > b > 0$) are parameters that should be determined from a data analysis. For real applications, if we believe the process above is a reasonable description of the relation between the returns, how do we estimate the model parameters α, κ, m , and β ? This remains the real test for a quant analyst.