F-Notes

Alessandro Cesarini

Jan 31, 2014

Contents

Chapter 1

Math

1 The Binomial Model

1.1 The One Period Binomial Model

We follow [2]. Let us consider a model with a bond B and a stock S and with only two points in time $t = 0$, $t = 1$. We assume

$$B_0 = 1 \tag{1.1}$$
$$B_1 = 1 + R, \text{being } R \text{ the interest rate} \tag{1.2}$$
$$S_0 = s \tag{1.3}$$
$$S_1 = s \cdot u, \text{ with probability } p_u \tag{1.4}$$
$$S_1 = s \cdot d, \text{ with probability } p_d \tag{1.5}$$
$$p_u + p_d = 1 \tag{1.6}$$
$$Z = u, d \text{ with probability } p_u, p_d \text{ respectively} \tag{1.7}$$
$$S_1 = s \cdot Z \tag{1.8}$$
$$u > d \tag{1.9}$$

Consider a portfolio $h = (x, y)$, where x is the number of bonds held at $t = 0$, y is the number of stocks held at $t = 0$. The value of h at $t = 0, 1$ will be

$$V_t^h = xB_t + yS_t \tag{1.10}$$

Definition: An arbitrage portfolio is a portfolio h such that

$$V_0^h = 0 \tag{1.11}$$
$$V_1^h > 0, \text{ with probability } 1 \tag{1.12}$$

Proposition: The model above is free of arbitrage if and only if

$$d \leq (1 + R) \leq u \tag{1.13}$$

Proof:

Suppose

$$(1 + R) < d \leq u$$

then $h = (-s, 1)$ is such that $V_0^h = 0$, $V_1^h \geq -s(1 + R) + s * d > 0$. The case $d \leq u < (1 + R)$ is treated similarly using $h = (s, -1)$. We hence proved that if (1.13) is violated, arbitrage exists. Hence absence of arbitrage implies (1.13) is respected. On the other way, we now prove that if

(1.13) is respected, then absence of arbitrage follows. Suppose *per assurdo* that $h = (-ys, y)$ is an arbitrage opportunity, i.e. that

$$V_0^h = -ys + ys = 0$$

and that

$$V_1^h = -ys(1+R) + ysd > 0 \tag{1.14}$$
$$V_1^h = -ys(1+R) + ysu > 0 \tag{1.15}$$

This clearly violates the hypothesis (1.13), hence (1.13) implies absence of arbitrage. Now notice that (1.13) implies that there exists $\{q_u, q_d\}$ such that

$$1 + R = q_u \cdot u + q_d \cdot d \tag{1.16}$$
$$q_u + q_d = 1 \tag{1.17}$$

We interpret $\{q_u, q_d\}$ as a new probability measure (the risk neutral or risk adjusted or martingale measure, to be contrasted with the objective measure we used above $\{p_u, p_d\}$) and also compute what follows

$$\frac{1}{1+R} E^Q [S_1] = \frac{1}{1+R} (q_u \cdot su + q_d \cdot sd) = \frac{s(1+R)}{1+R} = s = S_0 \tag{1.18}$$

Definition: A probability measure Q is called a martingale measure if the following condition holds

$$S_0 = \frac{1}{1+R} E^Q [S_1] \tag{1.19}$$

From what we have seen up to now, we can formulate the following proposition.

Proposition: The one time period binomial model is arbitrage free if and only if there exists a martingale measure. Notice also that

$$q_u = \frac{(1+R) - d}{u - d} \tag{1.20}$$
$$q_d = \frac{u - (1+R)}{u - d} \tag{1.21}$$

Definition: A contingent claim (or financial derivative) is any stochastic variable $X_t = \Phi(S_t)$, where Φ is any arbitrary function.

Definition: A claim is reachable if there exists an hedging (or replicating) portfolio h such that

$$X_1 = V_1^h, \text{(for any outcome of } S_1). \tag{1.22}$$

Definition: A market such that all claims are reachable is said to be complete.

Proposition: Let X be a claim reachable with he replicating portfolio $h = (x, y)$. Then the only price for X_0 (i.e. at time $t = 0$) that would not imply an arbitrage is

$$X_0 = V_0^h = x + ys \tag{1.23}$$

Proof:

Obvious.

Proposition: If the binomial model is free of arbitrage, then it is complete.
Proof:

Suppose to consider any claim X such that

$$X_1 = \Phi(u), \text{ if } Z = u \tag{1.24}$$
$$X_1 = \Phi(d), \text{ if } Z = d \tag{1.25}$$

then we have to find $h = (x, y)$ such that

$$x(1+R) + ysu = \Phi(u) \tag{1.26}$$
$$x(1+R) + ysd = \Phi(d) \tag{1.27}$$

Since the model is free of arbitrage by hypothesis, (1.13) holds and hence the above system has a unique solution (remind $u > d$)

$$x = \frac{1}{1+R}\left(\Phi(u) - su\frac{\Phi(u) - \Phi(d)}{s(u-d)}\right) = \frac{1}{1+R}\frac{u\Phi(d) - d\Phi(u)}{(u-d)} \tag{1.28}$$
$$y = \frac{\Phi(u) - \Phi(d)}{s(u-d)} \tag{1.29}$$

The no arbitrage price at $t = 0$ of the X claim will then be

$$X_0 = V_0^h = x + ys = \frac{1}{1+R}\frac{u\Phi(d) - d\Phi(u)}{(u-d)} + \frac{\Phi(u) - \Phi(d)}{s(u-d)}s = \tag{1.30}$$
$$= \frac{1}{1+R}\left[\Phi(d)\left(\frac{u}{(u-d)} - \frac{1+R}{(u-d)}\right) + \Phi(u)\left(\frac{-d}{(u-d)} + \frac{1+R}{(u-d)}\right)\right] = \tag{1.31}$$
$$= \frac{1}{1+R}\left[q_d\Phi(d) + q_u\Phi(u)\right] = \frac{1}{1+R}E^Q[X_1] \tag{1.32}$$

Summarizing:

- The binomial model is arbitrage free if and only if a martingale measure exists
- If the model is arbitrage free, any claim is reachable (that is the market is complete)
- The arbitrage free price of the claim at $t = 0$ is the discounted price of the claim at $t = 1$, using the martingale measure.

$$X_0 = \frac{1}{1+R}E^Q[X_1] \tag{1.33}$$

- The ratio between the claim and the bank account is a martingale.
- Any price for X_0 different from (1.33) would lead to an arbitrage opportunity, since the price comes from a perfect replication strategy.

1.2 The Multiperiod Binomial Model

Let $t = 0, 1, 2, \cdots, T$ be the time step indexes and write the bank account evolution as

$$B_0 = 1 \tag{1.34}$$
$$B_{n+1} = B_n \cdot (1+R) \tag{1.35}$$

Similarly for the stock

$$S_0 = s \tag{1.36}$$
$$S_{n+1} = S_n \cdot Z_n \tag{1.37}$$
$$Z_n = u_n, d_n, \text{with } u_n > d_n \text{ and with objective probabilities } p_{n,u}, p_{n,d} \tag{1.38}$$

Definition: A portfolio strategy is a stochastic process

$$\{h_t = (x_t, y_t); t = 1, \cdots, T\} \tag{1.39}$$

such that h_t is a function of 'past' values of the stock $S_0, S_1, \cdots, S_{t-1}$, meaning that (x_t, y_t) are the quantities of bond and stock respectively, that one decides to hold from time $t-1$ to time t (the decision of the exact amount of the quantities is taken at $t-1$, observing the whole past trajectory of the stock until $t-1$ included (but not observing the future with respect to $t-1$: $t, t+1, \cdots$)). Notice that we conventionally impose

$$h_0 = h_1 \tag{1.40}$$

The corresponding value process will be

$$V_t^h = x_t(1+R) + y_t S_t \tag{1.41}$$

Definition: The portfolio strategy is self-financing if it satisfies the following condition (budget equation, no cash injection or withdrawal):

$$x_t(1+R) + y_t S_t = x_{t+1} + y_{t+1} S_t \tag{1.42}$$

This means that the portfolio we decided to hold at $t-1$ has produced the amount of money $x_t(1+R) + y_t S_t$ at time t. At time t, we then rebalance the assets (stock versus cash bond) according to (x_{t+1}, y_{t+1}) without injection or withdrawal of 'exogenous money'.

Definition: An arbitrage portfolio strategy h will be such that:

$$V_0^h = 0 \tag{1.43}$$
$$P\left(V_T^h \geq 0\right) = 1 \tag{1.44}$$
$$P\left(V_T^h > 0\right) > 0 \tag{1.45}$$

Notice that we will easily have that the following condition is necessary to avoid the existence of arbitrage strategies:

$$d_n \leq (1+R) \leq u_n, \forall n \geq 0 \tag{1.46}$$

We will see that (1.46) is also sufficient to exclude the existence of arbitrage strategies in the model.

Definition: For simplicity, but the following arguments could be extended, we will consider a claim X depending only on the values of the stock at the end of the path: S_T:

$$X_T = \Phi(S_T), \text{ i.e. a non path-dependent pay off} \tag{1.47}$$

Definition: As in the one period case, we will say that the X claim is reachable if there exists a self-financing (hedging, replicating) strategy h such that

$$X_T = V_T^h, \text{ with probability 1} \tag{1.48}$$

A market is complete if all claims are reachable.

Proposition: If the X claim can be reached with the self-financing strategy h, then the only ('fair') price of the claim, at any time $t \geq T$, that allows to avoid arbitrage opportunities trading X against h is:

$$X_t = V_t^h, \text{ for any state of the world at time } t \tag{1.49}$$

Let us now suppose to perform a tree style roll back (hedging) procedure:

- Compute $X_T = \Phi(S_T)$ at final time T for all possible states of S_T.
- Consider any state $S_{T-1}(i)$ at time $T-1$ (notice that the i runs on all possible values that S_{T-1} could realize at time $T-1$, according to the chosen tree geometry). For such (i) state, it is possible to form a replicating self financing strategy by buying/selling suitable amounts of bonds and stock, as explained in the previous sub section (one period binomial model, see 1.1). This can be done considering that from the (i) state, the payoff could move, at time T, only towards two states, whose values are known: $\{S_{T-1}(i) \cdot u_{T-1}, S_{T-1}(i) \cdot d_{T-1}\}$. The value of the replicating strategy at (i) state of time $T-1$ will hence be the discounted value of the claim at such two states at time T, as we saw in the previous sub section (see 1.1).
- Roll back the previous step down to time 0 and notice that each step of the tree replication involves a self financing strategy by construction.

The above discussion has hence shown that if eq. (1.46) is satisfied, it is possible to replicate any claim through a self financing strategy, i.e. the model is complete. Let us now consider any self financing strategy h such that:

- $V_0^h = 0$
- $P\left(V_T^h \geq 0\right) = 1$
- $P\left(V_T^h > 0\right) > 0$

Considering the roll back procedure described above, we can observe that (remind sub section 1.1):

$$V_0^h = \frac{1}{(1+R)^T} E^Q\left[V_T^h\right] > 0 \tag{1.50}$$

that contradicts $V_0^h = 0$ (remind E^Q is the expectation according to the risk neutral probabilities introduced in sub section 1.1). We can then conclude that (1.46) is a necessary and (also) sufficient condition for absence of arbitrage in the model, that is:

Proposition: The model is arbitrage free if and only if the martingale measure exists for all n:

$$1 + R = q_{n,u} \cdot u_n + q_{n,d} \cdot d_n \tag{1.51}$$
$$q_{n,u} + q_{n,d} = 1 \tag{1.52}$$
$$u_n > q_n \tag{1.53}$$

Furthermore, if the martingale measure exists the model is (also) complete.

2 A More General One Period Model

We go back to a one period model, but we slightly generalize the structure of the assets w.r.t. 1.1. Let $t = 0, 1$ be the time schedule and consider the column vector of asset values at time t:

$$S_t = \begin{pmatrix} S_t^1 \\ . \\ . \\ . \\ S_t^N \end{pmatrix} \tag{1.54}$$

The randomness of the system is modeled by assuming that we have a finite sample space

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_M\} \tag{1.55}$$
$$P(\omega_i) > 0, \text{ i.e. probability of } \omega = \omega_i \text{ is strictly positive} \tag{1.56}$$

The state of the system at time $t = 1$ is represented w.r.t. Ω by the following matrix:

$$D = \begin{pmatrix} S_1^1(\omega_1) & S_1^1(\omega_2) & \cdots & S_1^1(\omega_M) \\ S_1^2(\omega_1) & S_1^2(\omega_2) & \cdots & S_1^2(\omega_M) \\ \cdots & \cdots & \cdots & \cdots \\ S_1^N(\omega_1) & S_1^N(\omega_2) & \cdots & S_1^N(\omega_M) \end{pmatrix} = \begin{pmatrix} | & | & \cdots & | \\ | & | & \cdots & | \\ d_1 & d_2 & \cdots & d_M \\ | & | & \cdots & | \\ | & | & \cdots & | \end{pmatrix} \tag{1.57}$$

Let us now define

$$h = [h^1, \cdots, h^N] \tag{1.58}$$

as a portfolio, i.e. h^i is the number of units of S^i that we buy at $t = 0$ and hold up to time $t = 1$. Using scalar product $\cdot$, we will have

$$V_t^h = h \cdot S_t \tag{1.59}$$

Definition: An arbitrage portfolio is defined as an h such that

$$V_0^h < 0 \tag{1.60}$$

$$V_1^h \geq 0, \text{with probability } 1 \tag{1.61}$$

Proposition: (Farkas' Lemma) Suppose $d_0, d_1, \cdots, d_M$ are column vectors in $\Re^N$, then exactly one of the following two problems possesses a solution:

- (P1): find non negative numbers $z_1, \cdots, z_M$ such that

$$d_0 = \sum_{i=1}^{M} z_j d_j \tag{1.62}$$

- (P2): find a row vector $h \in \Re^N$ such that

$$hd_0 < 0 \tag{1.63}$$

$$hd_j \geq 0, \forall j = 1, \cdots, M \tag{1.64}$$

Proof:

Let K be the set of all non negative linear combinations of vectors $d_1, \cdots, d_M$. K is easily seen to be a convex cone containing the origin. Then only one of the following two cases holds:

- $d_0 \in K$, i.e. (P1) has a solution
- $d_0 \notin K$. In this case, it means that it exists an hyperplane H such that d_0 is strictly on one side of H whereas K is on the other side. Let us then consider an h that is normal to H and points to the direction where K lies. h will satisfy the conditions in (P2).

Using Farkas' lemma and the definition of arbitrage portfolio, we can now directly state what follows:
Proposition: The model is arbitrage free if and only if there exist a set of non negative numbers $z_1, \cdots, z_M$ such that

$$S_0 = \sum_{j=1}^{M} z_j S_1(\omega_j) \tag{1.65}$$

We can now define

$$q_j = \frac{z_j}{\beta}, j = 1, \cdots, M \tag{1.66}$$

$$\beta = \sum_{j=1}^{M} z_j \tag{1.67}$$

$$S_0 = \beta \sum_{j=1}^{M} q_j S_1(\omega_j) = \beta E^Q [S_1] \tag{1.68}$$

where we have interpreed the q_j as probabilities.

Proposition: We can then restate the above proposition by affirming that the model is arbitrage free if and only if there exists a probability distribution Q (named: risk adjusted measure, martingale measure, risk neutral measure) such that

$$S_0 = \beta \sum_{j=1}^{M} q_j S_1(\omega_j) = \beta E^Q [S_1] \tag{1.69}$$

We now assume that S_t^1 is a risk free asset and in particular that

$$S_1^1(\omega_j) = 1, \forall j = 1, \cdots, M \tag{1.70}$$

which, using the first component of (1.69) yields

$$S_0^1 = \beta = \frac{1}{1+R} \tag{1.71}$$

where we introduced an interest rate R to parametrize β.

Proposition: (First fundamental theorem) Eq. (1.69) now gets

$$S_0 = \frac{1}{1+R} E^Q [S_1] \tag{1.72}$$

Assuming that there exists a risk free asset and denoting by R the corresponding interest rate, then the market is arbitrage free if and only if there exists a probability measure Q such that (1.72) holds. Notice that $\forall i = 0, 1, \cdots, N$ it holds that

$$\frac{S_0^i}{S_0^1} = E^Q \left[\frac{S_1^i}{S_1^1} \right] = E^Q \left[\frac{S_1^i}{1} \right] \tag{1.73}$$

that explains why Q is called named 'martingale measure'. Now consider a claim X, that is any real valued random variable defined on the Ω space. It is possible to consider the extended market ($t = 0, 1$):

$$\{S_t^1, S_t^2, \cdots, S_t^N, \Pi(t, X)\} \tag{1.74}$$

where we treat the new asset claim X as an original a priori given asset $S_t^i, i = 1, \cdots, N$. By extending what we got up to now in this section, it is then clear that the extended model will be arbitrage free if and only if there exists a martingale measure such that

$$\Pi(0, X) = \frac{1}{1+R} E^Q [\Pi(1, X)] = \frac{1}{1+R} E^Q [X] \tag{1.75}$$

$$S_0 = \frac{1}{1+R} E^Q [S_1] \tag{1.76}$$

This reasoning has produced a unique way of pricing the contingent claim X at time $t = 0$ once the Q measure is chosen and fixed. The problem now is to investigate the conditions for which the martingale measure Q is unique and hence for which the claim price $\Pi(0, X)$ at $t = 0$ is unique and well determined.

2.1 A More General One Period Model: Completeness

In this subsection, we assume that $S^1, \cdots, s^N$ is an arbitrage free market and that there exists a risk free asset (as defined in the previous section).

Definition: A contingent claim X (that is a real valued random variable depending on the future state at time $t = 1$: $\omega_i \in \Omega$) is replicated by the strategy h if

$$V_1^h = X, \text{ with probability } 1 \tag{1.77}$$

The market is complete if any claim can be replicated.

Proposition: The market is complete if and only if the matrix D defined in (1.57) has rank equal to M.

Proof:

It must hold that the following equation has a solution for any X:

$$V_1^h = h \cdot D = X \tag{1.78}$$

Being V_1^h a linear combination of the rows of D, the equation has a solution for any X if and only if

$$rank(D) = M \tag{1.79}$$

We hence see that if the market is complete the number of assets (N) must be greater or equal to the number of possible outcome states at time $t = 1$, i.e. M. We can now also propose a new pricing formula for $\Pi(0, X)$ based on the replicating portfolio h:

$$\Pi(0, X) = V_0^h = hS_0 = h\frac{1}{1+R}E^Q\left[S_1\right] = \frac{1}{1+R}E^Q\left[hS_1\right] = \frac{1}{1+R}E^Q\left[X\right] \tag{1.80}$$

where we used that by definition $\Pi(1, X) \equiv X = hS_1$. Notice that (1.80) agrees with (1.75). Now remind that we assumed that the model is arbitrage free. This means that if two strategies h_a and h_b both replicate X, it must hold that (otherwise it an arbitrage would exists contradicting the hypothesis):

$$h_a S_0 = h_b S_0 \tag{1.81}$$

and hence

$$\Pi(0, X) = h_a S_0 = h_b S_0 \tag{1.82}$$

is well defined. We now give another characterization of complete markets:

Proposition: (Second Fundamental Theorem) Assuming that the market is arbitrage free, it is true that the market is complete if and only if the martingale measure is unique (remind that at least one martingale measure exists since the market is by hypothesis arbitrage free).

Proof:

We already know from a previous characterization that the market is complete if and only $rank(D) = M$ (see (1.79)), i.e. if and only if

$$Im\left(D^*\right) = \Re^M \tag{1.83}$$

where we are now regarding D^* (i.e. the transpose of D) as a linear mapping from $\Re^N$ to $\Re^M$. After this consideration, we also remind that, due to absence of arbitrage, there exists a solution z (in particular $z_i \geq 0, \forall i = 1, \cdots, M$) such that

$$S_0 = Dz \tag{1.84}$$

Such solution z is unique if and only if the kernel (null space) of D is trivial, i.e. if and only if

$$Ker\left(D\right) = 0 \tag{1.85}$$

Using a known duality result[1]

$$(Im\,(D^*))^{\perp} = Ker\,(D) \tag{1.86}$$

we conclude that z is unique, and hence the martingale measure is unique, if and only if $(Im\,(D^*))^{\perp} = 0$, that is if and only if $Im\,(D^*) = \Re^M$, that is if and only if the market is complete. This concludes the proof. We now summarize the results:

- The market is arbitrage free if and only if (at least one) martingale measure exists.
- Assuming the market is arbitrage free, the market is (also) complete if and only if the martingale measure is unique.
- For any claim X, the only prices that are consistent with absence of arbitrage are of the form

$$\Pi(0, X) = \frac{1}{1+R} E^Q[X] \tag{1.87}$$

- If the market is not complete (but arbitrage free), it can happen that different choices of the martingale measure Q in (1.87) can give rise to different prices at time $t = 0$. Notice that in case of different martingale prices for the same claim X, if absence of arbitrage should actually hold, all market players should nonetheless agree to use the same Q.
- Even in an incomplete (but arbitrage free) market, if X can be replicated by a strategy h it holds that

$$V_0^h = \Pi(0, X) = \frac{1}{1+R} E^Q[X] \tag{1.88}$$

  for any possible choice of the martingale measure Q.

Definition: (Stochastic discount factor, Radon-Nikodym derivative)

$$\Pi(0, X) = \frac{1}{1+R} E^Q[X] = \frac{1}{1+R} \sum_{i=1}^{M} q_i X(\omega_i) = \tag{1.89}$$

$$= \frac{1}{1+R} \sum_{i=1}^{M} p_i \frac{q_i}{p_i} X(\omega_i) = \frac{1}{1+R} E^P[LX] = E^P[\Lambda X] \tag{1.90}$$

where $L(\omega_i) = \frac{q_i}{p_i}$ is the Radon-Nikodym derivative, p_i is the probability in the objective measure (remind that by hypothesis $p_i > 0, \forall i = 1, \cdots, M$),

$$\Lambda(\omega_i) = \frac{1}{1+R} \frac{q_i}{p_i} = \frac{1}{1+R} L(\omega_i) \tag{1.91}$$

is the stochastic discount factor. Notice the one to one correspondence between stochastic discount factors and martingale measures.

2.2 A More General One Period Model: Different Numeraire Measures

Let us assume that the k-th market asset S^k pays no dividends and is not generally risk free (i.e. it can possibly hold that $S_1^k(\omega_i) \neq S_1^k(\omega_j)$ for some $i \neq j$). We then write

$$\Pi(0, X) = \frac{1}{1+R} E^Q[X] = \frac{1}{1+R} \sum_{i=1}^{M} q_i X(\omega_i) = \frac{1}{1+R} \sum_{i=1}^{M} q_i \frac{X(\omega_i)}{S_1^k(\omega_i)} \frac{S_1^k(\omega_i) S_0^k}{S_0^k} \tag{1.92}$$

[1] Let D be a matrix with a rows and b cols. $Im(D^*) = rank(D)$, $(Im(D^*))^{\perp} = a - rank(D)$ since $Im(D^*)$ is a sub space of $\Re^a$. Furthermore $Ker(D) = a - rank(D)$ since $Ker(D)$ is the sub space of $\Re^a$ that is mapped to 0. Hence (1.86) follows. Notice that in this short explanation of (1.86) we abused notation and denoted with $Im(D^*)$ and $Ker(D)$ both the spaces and their dimensions, but everything should be clear from the context.

that reads out as

$$\frac{\Pi(0,X)}{S_0^k} = \frac{1}{1+R}\sum_{i=1}^{M} q_i \frac{X(\omega_i)}{S_1^k(\omega_i)}\frac{S_1^k(\omega_i)}{S_0^k} = \sum_{i=1}^{M} q_i^{(k)} \frac{X(\omega_i)}{S_1^k(\omega_i)} \tag{1.93}$$

where

$$q_i^{(k)} = \frac{1}{1+R} q_i \frac{S_1^k(\omega_i)}{S_0^k} = \frac{1}{1+R} q_i \frac{S_1^k(\omega_i)}{\frac{1}{1+R}\sum_{j=1}^{M} q_l S_1^k(\omega_l)} = \frac{q_i S_1^k(\omega_i)}{\sum_{j=1}^{M} q_l S_1^k(\omega_l)} \tag{1.94}$$

$$\sum_{i=1}^{M} q_i^{(k)} = 1 \tag{1.95}$$

that is clearly the right identity for $k = 1$, considering that we assumed that $S_1^1(\omega_i) = 1, \forall i = 1, \cdots, M$, being S^1 the risk free asset. We then write

$$\frac{\Pi(0,X)}{S_0^k} = E^{(k)}\left[\frac{X}{S_1^k}\right] \tag{1.96}$$

We now also compute for a different numeraire S^j

$$L^{j/k}(\omega_i) = \frac{q_i^{(j)}}{q_i^{(k)}} = \frac{S_1^j(\omega_i)}{S_0^j}\frac{S_0^k}{S_1^k(\omega_i)} \tag{1.97}$$

and

$$\frac{\Pi(0,X)}{S_0^j} = E^{(j)}\left[\frac{X}{S_1^j}\right] = E^{(k)}\left[\frac{X(\omega)L^{j/k}(\omega)}{S_1^j(\omega)}\right] = E^{(k)}\left[\frac{X(\omega)}{S_1^j(\omega)}\frac{S_1^j(\omega)}{S_0^j}\frac{S_0^k}{S_1^k(\omega)}\right] \tag{1.98}$$

that correctly gives back (1.96). Finally

$$E^{(k)}\left[L^{j/k}(\omega)\right] = \sum_{i=1}^{M} \frac{S_1^j(\omega_i)}{S_0^j}\frac{S_0^k}{S_1^k(\omega_i)}\frac{q_i S_1^k(\omega_i)}{\sum_{j=1}^{M} q_l S_1^k(\omega_l)} = \frac{S_0^k(1+R)S_0^j}{S_0^j(1+R)S_0^k} = 1 \tag{1.99}$$

where we used (1.72), that it was in any case a priori obvious since

$$E^{(k)}\left[L^{j/k}(\omega)\right] = \sum_{i=1}^{M} \frac{q_i^{(j)}}{q_i^{(k)}} q_i^{(k)} = \sum_{i=1}^{M} q_i^{(j)} = 1 \tag{1.100}$$

3 Measure and Integration

3.1 σ-algebra

Definition: A family $\mathcal{F}$ of subsets is a sigma-algebra if

- $\emptyset \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- countable unions and intersections of sets in the σ-algebra are in the σ-algebra

Proposition: let $\{\mathcal{F}_\alpha; \alpha \in A\}$ be a family of σ-algebras on the set X. Then $\mathcal{F} = \bigcap_{\alpha \in A} \mathcal{F}_\alpha$ is also a σ-algebra.

Definition: Given $\mathcal{S}$ a family of subsets of X, there's a unique minimal extension of $\mathcal{S}$ as a σ-algebra , which is the intersecion of all σ-algebras that contain $\mathcal{S}$. Such σ-algebra is called the

σ-algebra generated by $\mathcal{S}$ and denoted by $\sigma\{\mathcal{S}\}$.

Definition: Let $\{f_\gamma; \gamma \in \Gamma\}$ be an indexed family of functions from X to $\Re$. $\sigma\{f_\gamma; \gamma \in \Gamma\}$ is the smallest σ-algebra such that f_γ is measurable for all $\gamma \in \Gamma$.

Definition: $\mathcal{B}(\Re^n) = \sigma\{\text{open sets in } \Re^n\}$ is the Borel σ-algebra (i.e. the smallest σ-algebra generated by open sets in $\Re^n$. Notice that there are a number of other generators for this σ-algebra beyond open sets.)

Definition: given a measure space $(X, \mathcal{F}, \mu)$, $f : X \to \Re$ is measurable w.r.t. the sigma-algebra $\mathcal{F}$ if for every interval $I \subset \Re$, it holds that $f^{-1}(I) \in \mathcal{F}$. It can also be seen that f is measurable w.r.t. $\mathcal{F} \Leftrightarrow f^{-1}(B) \in \mathcal{F}$ for every Borel set $B \subset \Re$.

3.2 L^p Spaces

Definition: Let $p \in [1, \infty)$, $L^p(X, \mathcal{F}, \mu)$ is the class of measureable functions $f : X \to \Re$ such that

$$\|f\|_p = \left(\int_X |f(x)|^p d\mu(x)\right)^{1/p} < \infty \tag{1.101}$$

Definition: A sequence $\{f_n\}_n$ is Cauchy in L^p if for every $\epsilon > 0$ there exists a $N > 0$ such that

$$\|f_n - f_m\|_p \leq \epsilon, \forall n, m \geq N$$

Proposition: For $1 \leq p \leq +\infty$, every L^P space is complete, in the sense that every sequence $\{f_n\}_n$, that is Cauchy in L^p, converges (L^p-convergence) to a unique element $f \in L^p$, i.e.

$$\lim_{n \to +\infty} \|f_n - f\|_p = 0 \tag{1.102}$$

3.3 Radon-Nikodyn Theorem

Definition: Consider a measurable space $(X, \mathcal{F})$, on which two different measures are defined μ and ν. Then:

- if $\forall A \in \mathcal{F}$ it holds that $\mu(A) = 0 \Rightarrow \nu(A) = 0$, then one says that $\nu << \mu$ (in words: ν is absolutely continous w.r.t. μ on $\mathcal{F}$)
- if at the same time $\nu << \mu$ and $\mu << \nu$, then $\nu \sim \mu$, that is ν and μ are equivalent measures on $\mathcal{F}$
- if there exist $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$ and $\mu(A) = \nu(B) = 0$, μ and ν are said mutally singular, or in short $\mu \perp \nu$.

Definition: Take $(X, \mathcal{F}, \mu)$ and a non-negative $f : X \to \Re_0^+$ which is also measurable in $L^1(X, \mathcal{F}, \mu)$. Define for $A \in \mathcal{F}$:

$$\nu(A) = \int_A f(x) d\mu(x) \tag{1.103}$$

It can be proved that $\nu(A)$ is indeed a measure, that is

- $\nu(\emptyset) = 0$
- $\nu(A) \geq 0, \forall A \in \mathcal{F}$
- if $A_n \in \mathcal{F}$ for $n \in \mathcal{N}$ and $A_i \cap A_j = \emptyset, \forall i \neq j$, $\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$

and furthermore that $\nu << \mu$.

Proposition: The Radon-Nikodyn Theorem. Consider a measure space $(X, \mathcal{F}, \mu)$, with finite μ, i.e. $\mu(X) < \infty$. Assume exists a ν measure on $(X, \mathcal{F})$ such that $\nu << \mu$ on $\mathcal{F}$. Then there's $f : X \to \Re_0^+$ such that:

- f is $\mathcal{F}$-measurable
- $\int_X f(x)d\mu(x) < \infty$
- $\nu(A) = \int_A f(x)d\mu(x), \forall A \in \mathcal{F}$

f is called *the Radon-Nikodyn derivative* of ν w.r.t. μ. Furthermore f is uniquely determined μ almost surely and one usually writes:

- $f(x) = \frac{d\nu(x)}{d\mu(x)}$
- $d\nu(x) = f(x)d\mu(x)$

(Sketch of)*Proof:*

Define a new measure: $\lambda(A) = \mu(A) + \nu(A)$, $\forall A \in \mathcal{F}$ and choose any $g \in L^2(\lambda)$. It is now possible to define a linear mapping $\Phi : L^2(\lambda) \to \Re$ by

$$\Phi(g) = \int_X g(x)d\nu(x)$$

such that

$$\begin{aligned} |\Phi(g)| &= \left|\int_X g(x)d\nu(x)\right| \leq \int_X |g(x)|d\nu(x) \leq \int_X |g(x)|d(\nu(x) + \mu(x)) = && (1.104) \\ & && (1.105) \\ &= \int_X |g(x)|d\lambda(x) = (|g|, 1) \leq \|g\|_2 \cdot \|1\|_2 = \left(\int_X |g(x)|^2 d\lambda\right)^{\frac{1}{2}} \cdot \left(\int_X |1|^2 d\lambda\right)^{\frac{1}{2}} = && (1.106) \\ &= \sqrt{\lambda(X)} \cdot \|g\|_{L^2(\lambda)} && (1.107) \end{aligned}$$

Having prooved that $\phi(g)$ is bounded in $L^2(\lambda)$, we can use a known result (Riesz representation theorem) that states that there exists a unique $f \in L^2(\lambda)$ such that $\phi(g) = (g, f)$, $\forall g \in L^2(\lambda)$. In practice we then have

$$\Phi(g) = \int_X g(x)d\nu(x) = (g, f)_\lambda = \int_X g(x)f(x)d\lambda(x), \forall g \in L^2(\lambda)$$

Now choose any $A \in \mathcal{F}$ and set $g = I_A$ and remind that $0 \leq \nu(A) \leq \lambda(A)$:

$$\Phi(I_A) = \int_A 1d\nu(x) = \nu(A) = (1, f)_\lambda = \int_X f(x)d\lambda(x)$$

From the arbitrary of A it then follows that $0 \leq f(x) \leq 1$ a.e. in X. We now write:

$$\int_X g(x)d\nu(x) = \int_X g(x)f(x)d\lambda(x) = \int_X g(x)f(x)d\nu(x) + \int_X g(x)f(x)d\mu(x)$$

that is

$$\int_X g(x)(1 - f(x))d\nu(x) = \int_X g(x)f(x)d\mu(x)$$

Consider now the set $A = \{x \in X : f(x) = 1\} \in \mathcal{F}$ and set $g = I_A$ to get

$$0 = \int_A 1(1 - 1))d\nu(x) = \int_A f(x)d\mu(x)$$

from which (being $f \geq 0$) one gets $\mu(A) = 0$. Hence outside A one can write for any $B \in \mathcal{F}$ and taking $g = I_B$:

$$\int_B (1 - f(x)) d\nu(x) = \int_B f(x) d\mu(x)$$

from which

$$\int_B (1 - f(x)) \frac{d\nu(x)}{d\mu(x)} d\mu(x) = \int_B f(x) d\mu(x)$$

which gives

$$(1 - f(x)) \frac{d\nu(x)}{d\mu(x)} = f(x)$$

and finally ($A = \{x \in X : f(x) = 1\} \in \mathcal{F}$, with $\mu(A) = 0$):

$$\frac{d\nu(x)}{d\mu(x)} = \frac{f(x)}{1 - f(x)}$$

(Of course the last steps were not formal at all, unless B can be any singleton $\{x\}, \forall x \in X$).

4 Probability

A measure space $(\Omega, \mathcal{F}, P)$ is a probability space if the measure P is such that $P(\Omega) = 1$ (Ω is called the sample space, elements $F \in \mathcal{F}$ are called events).

Definition: A random variable X is a mapping $X : \Omega \to \Re$ that is $\mathcal{F}$-measurable, i.e. for any Borel set $B \in \mathcal{B}(\Re)$ it is true that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$.

Definition: The **distribution measure** μ_X for a random variable X is a measure on $(\Re, \mathcal{B}(\Re))$, that is defined by (notice that $X^{-1}(B) \in \mathcal{F}$ since X is measurable on $\mathcal{F}$):

$$\mu_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X^{-1}(B)) = P(X \in B), \forall B \in \mathcal{B}(\Re)$$

Similarly the **(cumulative) distribution function** of X is defined as

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x)$$

Definition: Taking $X \in L^1(\Omega, \mathcal{F}, P)$ define the **expected value** as

$$E[X] = \int_\Omega X(\omega) dP(\omega)$$

and, if $X \in L^2(\Omega, \mathcal{F}, P)$, the **variance**

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

Proposition: Let $g : \Re \to \Re$ be a Borel function such that $g(X(\omega))$ is integrable on $(\Omega, \mathcal{F}, P)$. Then is holds that

$$E[g(X)] = \int_\Omega g(X(\omega)) dP(\omega) = \int_\Re g(x) d\mu_X(x) \tag{1.108}$$

Proof:

First notice that $g(X)$ is $\mathcal{F}$-measurable since g is Borel. Now start with $g = I_A$ for any $A \in \mathcal{B}(\Re)$ to get that

$$\begin{aligned} E[I_A(X)] &= \int_\Omega I_A(X(\omega)) dP(\omega) = \int_{\{\omega \in \Omega : X(\omega) \in A\}} dP(\omega) = P(\{\omega \in \Omega : X(\omega) \in A\}) = \mu_X(A) = \\ &= \int_\Re I_A(x) d\mu_X(x) \end{aligned}$$

Now if g is a simple function for a certain family of pairs (c_i, A_i), $c_i \in \Re$, $A_i \in \mathcal{B}(\Re)$, $i = 1, ..., n$:

$$g(x) = \sum_{i=1}^{n} c_i \cdot I_{A_i}(x)$$

the same result holds by linearity. At this point one can approximate g with simple functions and get the result in the limit.

Proposition: Given X non-negative random variable, it holds that

$$E[X] = \int_0^\infty P(X > t)dt$$

Proof:

$$\begin{aligned} E[X] &= \int_\Omega X(\omega)dP(\omega) = \int_\Omega \left[\int_0^{X(\omega)} 1 \cdot dt\right] dP(\omega) = \int_\Omega \left[\int_0^\infty I\{X(\omega) > t\}dt\right] dP(\omega) && (1.109) \\ &= \int_0^\infty \left[\int_\Omega I\{X(\omega) > t\}dP(\omega)\right] dt = \int_0^\infty P(X > t)dt && (1.110) \end{aligned}$$

Definition: A random process $X(t, \omega)$ on $(\Omega, \mathcal{F}, P)$ is a mapping

$$X : \Re_0^+ \times \Omega \to \Re$$

such that for any $t \in \Re_0^+$

$$X(t, \cdot) : \Omega \to \Re$$

is a $\mathcal{F}$-measurable random variable.

Definition: A partition $\mathcal{P}$ on a set Ω is a finite collection os sets $\{A_1, ..., A_n\}$, such that

- the sets cover Ω, that is $\bigcup_{i=1}^n A_i = \Omega$
- $A_i \cap A_j = \emptyset, \forall i \neq j$

Take a mapping $f : \Omega \to \Re$ and suppose f takes values only in the finite set $\{x_1, ..., x_K\}$. Then defining

$$A_i = \{\omega \in \Omega : f(\omega) = x_i\} = f^{-1}(x_i), i = 1, ..., K$$

we see that

$$\mathcal{P}(f) = \{A_1, ..., A_K\}$$

is a partition of Ω (the partition generated by f).

Definition: A mapping is defined to be measurable w.r.t. a partition $\mathcal{P}$ if and only if it is constant on the components of the partition.

Definition: Consider X random variable on Ω. $\sigma\{X\}$ is defined to be the smallest σ-algebra $\mathcal{F}$ such that X is $\mathcal{F}$-measurable. More precisely:

$$\sigma\{X\} = \bigcap_{\mathcal{G}} \{\mathcal{G}\}$$

where $\mathcal{G}$ is any σ-algebra w.r.t. X is measurable.

Proposition: $\sigma\{X\} = \{X^{-1}(B); B \in \mathcal{B}(\Re)\}$

Proof:

First it can be proved that $\sigma\{X\}$ defined as above is a σ-algebra , since

- $X^{-1}(\emptyset) = \emptyset \Rightarrow \emptyset \in \sigma\{X\}$
- if $A \in \sigma\{X\}$ then exists $B \in \mathcal{B}(\Re)$ such that $A = X^{-1}(B)$. Since $\mathcal{B}(\Re)$ is a σ-algebra , we know that $B^c \in \mathcal{B}(\Re)$ and hence that $A^c = X^{-1}(B^c) \in \sigma\{X\}$.
- similarly, considering any $A_1, A_2, ... \in \sigma\{X\}$ and the corresponding $B_1, B_2, ... \in \mathcal{B}(\Re)$ such that $A_i = X^{-1}(B_i)$, one has

$$\bigcap_i A_i = \bigcap_i X^{-1}(B_i) = X^{-1}\left(\bigcap_i B_i\right) \in \sigma\{X\}$$

  since $\bigcap_i B_i \in \mathcal{B}(\Re)$
- similarly for countable unions
- $\sigma\{X\}$ is the smallest σ-algebra for which X is measurable, since any σ-algebra $\mathcal{G}$ on which X is measurable must contain $\sigma\{X\}$ due to the definition of measurability of X on $\mathcal{G}$.

Definition: Let $\mathcal{K}$ be an arbitrary family of mappings from Ω to $\Re$. Then $\sigma\{\mathcal{K}\}$ is defined to be the smallest σ-algebra $\mathcal{G}$ such that X is $\mathcal{G}$-measurable for any $X \in \mathcal{K}$.

Proposition: (No proof). Let $X_1, ..., X_N$ be mappings $X_n : \Omega \to \Re$ and assume that $X : \Omega \to \Re$ is measurable w.r.t. $\sigma\{X_1, ..., X_N\}$. Then there exists a Borel function $f : \Re^n \to \Re$ such that $\forall \omega \in \Omega$ it holds that

$$X(\omega) = f(X_1(\omega), ..., X_N(\omega)) \tag{1.111}$$

Definition: Let $\{X_t; t \geq 0\}$ be a random process defined on the probability space $(\Omega, \mathcal{F}, P)$. The σ-algebra generated by X over the interval $[0, t]$ is denoted by

$$\mathcal{F}_t^X = \sigma\{X_s; s \leq t\}$$

It can be proved that $\mathcal{F}_t^X$ is generated by all the subsets of Ω of the form

$$\{X_s \in B\} = \{\omega \in \Omega; X_s(\omega) \in B\}$$

for any $B \in \mathcal{B}(\Re)$ and for any $s \leq t$. Notice that $\mathcal{F}_s^X \subseteq \mathcal{F}_t^X$ for $s \leq t$.

Definition: A filtration $\underline{\mathcal{F}} = \{F_t\}_{t \geq 0}$ on the probability space $(\Omega, \mathcal{F}, P)$ is an indexed family of σ-algebra on Ω such that

- $\mathcal{F}_t \subseteq \mathcal{F}, \forall t \geq 0$
- $s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$

Furthermore, given a $\underline{\mathcal{F}}$, one can define the smallest σ-algebra that contains all the $\mathcal{F}_t, \forall t \geq 0$:

$$\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t$$

Definition: Considering a filtration $\underline{\mathcal{F}}$ on the probability space $(\Omega, \mathcal{F}, P)$ and a random process X_t on the same space, we say that X_t is adapted to $\underline{\mathcal{F}}$ if X_t is measurable w.r.t. $\mathcal{F}_t$ for every $t \geq 0$, that is if

$$X_t \in \mathcal{F}_t, \forall t \geq 0$$

We could say that if for a certain time $t \geq 0$ we take a measure of the random variable X_t and gather the information that its outcome is contained in a certain Borel set $B \in \mathcal{B}(\Re)$, then, being $X_t \in \mathcal{F}_t$, we gather the information that $\omega \in X_t^{-1}(B)$, where $X_t^{-1}(B) \in \mathcal{F}_t$. In other words the information we gatherd on ω does not go beyond $\mathcal{F}_t$: if for example it was the case that $X_t^{-1}(B) \notin \mathcal{F}_t$ but

$X_t^{-1}(B) \in \mathcal{F}_s$ with $s > t$, then X would not have been $\mathcal{F}_t$-measurable and was instead 'looking into the future' w.r.t. $\mathcal{F}_t$.

Definition: Two events $A, B \in \mathcal{F}$ and within $(\Omega, \mathcal{F}, P)$ are said to be independent if

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

Definition: Similarly:

- Two σ-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ (within $(\Omega, \mathcal{F}, P)$) are defined independent if

  $$P(G \cap H) = P(G) \cdot P(H)$$

  for every $G \in \mathcal{G}$ and every $H \in \mathcal{H}$.

- Two random variables X, Y are independent if $\sigma\{X\}$ and $\sigma\{Y\}$ are independent.

- Two stochastic processes $\{X_t; t \geq 0\}, \{Y_t; t \geq 0\}$ are independent if $\sigma\{X_t; t \geq 0\}$ and $\sigma\{Y_t; t \geq 0\}$ are independent.

- An indexed family of σ-algebras $\{\mathcal{G}_\gamma; \gamma \in \Gamma\}$, with $\{\mathcal{G}_\gamma \in \mathcal{F}; \forall \gamma \in \Gamma\}$, has mutually independent elements if

  $$P\left(\bigcap_{i=1}^{n} G_i\right) = \prod_{i=1}^{n} P(G_i)$$

  for every finite collection $\{G_1, ..., G_n\}$, with $\gamma_i \neq \gamma_j$ for $i \neq j$ and $G_i \in \mathcal{G}_{\gamma_i}$. This definition straightly extends to random variables and processes.

Proposition: Suppose that X and Y are independent random variables and that they both are $L^1(\Omega, \mathcal{F}, P)$, together with their product $X \cdot Y$. Then we have: $E[X \cdot Y] = E[X] \cdot E[Y]$.

Proof:

The proposition is straightforward for indicator functions: $X = I_A$, $A \in \sigma\{X\}$, $Y = I_B$, $B \in \sigma\{Y\}$, since

$$E[X \cdot Y] = E[I_A \cdot I_B] = E[I_{A \cap B}] = \int_{A \cap B} dP(\omega) = P(A \cap B) = P(A) \cdot P(B) = E[X] \cdot E[Y]$$

and the extensions to aribitrary simple functions and then to arbitrary random variables in L^1 are as usual.

4.1 Conditional Expectations

Definition: Given $(\Omega, \mathcal{F}, P)$, the probability of A conditional on B is defined as

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

Definition: Given $(\Omega, \mathcal{F}, P)$, suppose $B \in \mathcal{F}$ with $P(B) > 0$ and $X \in L^1(\Omega, \mathcal{F}, P)$. Then the conditional expectation of X given B is defined by

$$E[X|B] = \frac{1}{P(B)} \int_B X(\omega) dP(\omega)$$

Let us consider a partition $\mathcal{P} = \{A_1, ..., A_K; A_i \in \mathcal{F}\}$ of Ω, always within $(\Omega, \mathcal{F}, P)$. Given a ω, one can determine a unique n such that $\omega \in A_n$ end hence compute $E[X|A_n]$. In other words one can define the following $\Omega \to \Re$ mapping:

$$\omega \to E[X|A_{n(\omega)}]$$

Definition: More sinthetically one can define the following random variable (mapping $\Omega \to \Re$):

$$E[X|\mathcal{P}](\omega) = \sum_{n=1}^{K} I_{A_n}(\omega) \cdot E[X|A_n]$$

under the hypotheis $P(A_n) > 0, \forall n = 1, .., K$.

Proposition: Consider $(\Omega, \mathcal{F}, P)$ and X, $\mathcal{P}$ as above. Consider also $\mathcal{G} = \sigma\{\mathcal{P}\} \subseteq \mathcal{F}$. Then $Z(\omega) = E[X|\mathcal{P}](\omega)$ is such that

- Z is $\mathcal{G}$-measurable
- for evey $G \in \mathcal{G}$

$$\int_G Z(\omega)dP(\omega) = \int_G X(\omega)dP(\omega)$$

The reverse is also true: if the two properties are satisfied by any $Z(\omega)$ then $Z(\omega) = E[X|\mathcal{P}](\omega)$.

Proof:

- Noticing that $Z(\omega) = E[X|\mathcal{P}](\omega)$ is constant on A_i, we can choose any $z \in \Re$ and compute $\{\omega \in \Omega; Z(\omega) \le z\} = \bigcup_i \{A_i : E[X|A_i] \le z\} \in \mathcal{G}$. This proves that $Z(\omega)$ is $\mathcal{G}$-measusrable.
-

$$\begin{aligned} \int_G Z(\omega)dP(\omega) &= \int_G \sum_{n=1}^{K} I_{A_n}(\omega) \cdot E[X|A_n]dP(\omega) = \sum_{n=1}^{K} E[X|A_n] \int_{G \cap A_n} dP(\omega) = \\ &= \sum_{n=1}^{K} \frac{P(G \cap A_n)}{P(A_n)} \int_{A_n} X(\omega)dP(\omega) \end{aligned}$$

At this point we have to consider that since $G \in \mathcal{G}$, it can only be the case that:

$$G = \bigcup_{i=1}^{J} A_{n_i}$$

for some J and some set of $n_i \in 1, ..., K$. This is so since being $\mathcal{P}$ a partition the A_i are disjoint and hence $\mathcal{G} = \sigma\{\mathcal{P}\} \subseteq \mathcal{F}$ can only have sets of the above type. We hence get:

$$\int_G Z(\omega)dP(\omega) = \sum_{n=1}^{K} \frac{P(G \cap A_n)}{P(A_n)} \int_{A_n} X(\omega)dP(\omega) = \sum_{n=1}^{K} \frac{P(\bigcup_{i=1}^{J} A_{n_i} \cap A_n)}{P(A_n)} \int_{A_n} X(\omega)dP(\omega) \quad (1.112)$$

$$= \sum_{i=1}^{J} \int_{A_{n_i}} X(\omega)dP(\omega) = \int_G X(\omega)dP(\omega) \quad (1.113)$$

Up to now, we have proved that $Z(\omega) = E[X|\mathcal{P}](\omega)$ satisfies the above two properties (i.e. that it is $\mathcal{G}$-measurable and it has the same expecations as $X(\omega)$ on all sets $G \in \mathcal{G}$). Now, on the other hand we prove the reverse. First we notice that if a $Z^{'}$ random variable is $\mathcal{G}$-measurable, it is constant on elements of the partition $A_i, i = 1, ..., K$ (as our target $Z(\omega)$). It then only remains to determine this constant values. Taking any $i = 1, ..., K$ and any $\omega \in A_i$ we have by hypothesis:

$$\int_{A_i} Z^{'}(\omega)dP(\omega) = Z(A_i) \cdot P(A_i) = \int_{A_i} X(\omega)dP(\omega)$$

from which we get (as desired) for all $\omega \in A_i$:

$$Z(\omega) = Z(A_i) = \frac{1}{P(A_i)} \int_{A_i} X(\omega) dP(\omega)$$

Definition: Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable il $L^1(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a σ-algebra such that $\mathcal{G} \subset \mathcal{F}$. If Z is a random variable such that

- Z is $\mathcal{G}$-measurable
- For every $G \in \mathcal{G}$ it holds that

$$\int_G Z(\omega) dP(\omega) = \int_G X(\omega) dP(\omega)$$

Then we say that Z is the conditional expectation of X given the σ-algebra $\mathcal{G}$ and we introduce the notation

$$Z(\omega) = E[X|\mathcal{G}](\omega)$$

Proposition: Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable il $L^1(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a σ-algebra such that $\mathcal{G} \subset \mathcal{F}$. Then

- there always exists a random variable Z such that
  - Z is $\mathcal{G}$-measurable
  - For every $G \in \mathcal{G}$ it holds that

$$\int_G Z(\omega) dP(\omega) = \int_G X(\omega) dP(\omega)$$

- Z is unique P almost surely, i.e. if both Y and Z satisfy the two properties then $Y(\omega) = Z(\omega)$ $\forall \omega \in A^c$ with $P(A) = 0$.

Proof:

Define the new measure ν on $(\Omega, \mathcal{G})$ by

$$\nu(G) = \int_G X(\omega) dP(\omega)$$

We have $\nu << P$ and using the Radon-Nikodyn theorem we see that exists (unique P-a.s.)

$$\frac{d\nu}{dP}$$

such that

$$\nu(G) = \int_G \frac{d\nu}{dP}(\omega) dP(\omega) = \int_G X(\omega) dP(\omega)$$

(the last equality by definition of $\nu(G)$). This means that $\frac{d\nu}{dP}(\omega)$ is the $Z(\omega)$ we are looking for (notice that it is also $\mathcal{G}$-measurable by definition). Notice in particular that if $\mathcal{G} = \{\emptyset, \Omega\}$ then it is sure that $E[X|\mathcal{G}]$ is constant on the entire Ω and that such constant is such that

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

hence $E[X] = E[X|\sigma\{\emptyset, \Omega\}]$.

Proposition: Assume $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then the following is true:

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}] \tag{1.114}$$

$$E[X] = E[E[X|\mathcal{G}]] \tag{1.115}$$

Proof:

First of all, being $E[X] = E[X|\sigma\{\emptyset,\Omega\}]$, if (1.114) is true we can use it with $\mathcal{H} = \sigma\{\emptyset,\Omega\}$ and get

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

which has proved (1.115). Then we denote $Z(\omega) = E[E[X|\mathcal{G}]|\mathcal{H}](\omega)$ and we observe that by definition of conditional expectation, Z is $\mathcal{H}$ measurable. Furthermore we compute for any $H \in \mathcal{H}$

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

where we have again used (twice) the definition of conditional expectation together with the fact that being $\mathcal{H} \subseteq \mathcal{G}$ we also have $H \in \mathcal{G}$. This last equaion, using also the previous theorem, proves that

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

at least P-a.s.

Proposition: (No proof) Consider f, g integrable in $(\Omega, \mathcal{F}, P)$ then the following is true:

$$f(\omega) = g(\omega), P - a.s. \Leftrightarrow \forall A \in \mathcal{F}, \int_A f(\omega)dP(\omega) = \int_A g(\omega)dP(\omega) \tag{1.116}$$

We know that if X is $\mathcal{G}$-measurable then if we can observe the space Ω with degree of detail given by the sets of $\mathcal{G}$, we can fully "determine/observe" the outcomes of X. This will hence intuitively mean that conditioning X, that is $\mathcal{G}$-measurable, on $\mathcal{G}$ itself will mean that we can treat it as deterministic (conditional on $\mathcal{G}$). Formally:

Proposition: If X is $\mathcal{G}$-measurable ($\mathcal{G} \subseteq \mathcal{F}$) and if X, Y, XY are in $L^1(\Omega, \mathcal{F}, P)$ then

$$E[X|\mathcal{G}](\omega) = X(\omega), P - a.s. \tag{1.117}$$

$$E[XY|\mathcal{G}](\omega) = X(\omega) \cdot E[Y|\mathcal{G}](\omega), P - a.s. \tag{1.118}$$

Proof:

For every $G \in \mathcal{G}$ we have

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

and being both integrands $\mathcal{G}$-measurable, by (1.116), we directly get (1.117). Regarding (1.118), for every $G \in \mathcal{G}$ and starting by $X = I_A$ for any $A \in \sigma\{X\}$:

$$\begin{aligned}\int_G E[I_A Y|\mathcal{G}](\omega)dP(\omega) &= \int_G I_A(\omega)Y(\omega)dP(\omega) = \int_{A\cap G} Y(\omega)dP(\omega) = \\ &= \int_G I_A(\omega) \cdot E[Y|\mathcal{G}](\omega)dP(\omega) = \int_{A\cap G} E[Y|\mathcal{G}](\omega)dP(\omega) = \\ &= \int_{A\cap G} Y(\omega)dP(\omega)\end{aligned}$$

where we used that $A \cap G \in \mathcal{G}$ being $\sigma\{X\} \subseteq \mathcal{G}$ and hence the last equality holds by definition of $E[Y|\mathcal{G}]$. We hence proved for any $G \in \mathcal{G}$ it holds that

$$\int_G E[I_A Y|\mathcal{G}](\omega)dP(\omega) = \int_G I_A(\omega) \cdot E[Y|\mathcal{G}](\omega)dP(\omega)$$

which together with the measurability on $\mathcal{G}$ of both $E[I_A Y|\mathcal{G}](\omega)$ and $I_A(\omega) \cdot E[Y|\mathcal{G}](\omega)$ and with (1.116), proves that

$$E[I_A Y|\mathcal{G}](\omega) = I_A(\omega) \cdot E[Y|\mathcal{G}](\omega), P - a.s.$$

Notice that we have used the key assumption that X is $\mathcal{G}$-measurable, i.e. $\sigma\{X\} \subseteq \mathcal{G}$. Having proved (1.118) for a general indicator functions on $\sigma\{X\}$, one proceeds with simple functions and then, taking the limit, for X itself to end the proof.

Definition: For any integrable Y in $(\Omega, \mathcal{F}, P)$ and for any $X \in \mathcal{F}$ we define

$$E[Y|X] = E[Y|\sigma\{X\}]$$

Since $E[Y|X]$ is by definition $\sigma\{X\}$ measurable, we know by proposition (1.111) that there exists a Borel function $g : \Re \to \Re$ such that

$$E[Y|X](\omega) = g(X(\omega)), P - a.s. \tag{1.119}$$

Definition: Now, using this Borel function g, we also define

$$E[Y|X = x] = g(x), x \in \Re \tag{1.120}$$

Proposition: Consider $(\Omega, \mathcal{F}, P)$ and let μ_X the distribution measure for X and Y any random variable. Then

$$E[Y] = \int_{\Re} E[Y|X = x] d\mu_X(x) \tag{1.121}$$

Proof:

By using

$$E[Y|X](\omega) = g_Y(X(\omega)), P - a.s.$$

we can write

$$\begin{aligned} E[Y] &= E[E[Y|X](\omega)] = E[g_Y(X(\omega))] = \int_{\Omega} g_Y(X(\omega)) dP(\omega) = \int_{\Re} g_Y(x) d\mu_X(x) = \\ &= \int_{\Re} E[Y|X = x] d\mu_X(x) \end{aligned}$$

where we also used (1.108).

Proposition: Consider $(\Omega, \mathcal{F}, P)$ and let μ_X the distribution measure for X and Y any random variable. Then for any $\bar{x} \in \Re$:

$$E\left[Y 1_{(X=\bar{x})}\right] = E\left[Y|X = \bar{x}\right] P\left(X = \bar{x}\right) \tag{1.122}$$

$$E\left[Y \delta\left(X - \bar{x}\right)\right] = E\left[Y|X = \bar{x}\right] f_\mu\left(\bar{x}\right) \tag{1.123}$$

$$\int_{\Re} d\mu_X(x) = \int_{\Re} f_\mu\left(x\right) dx = 1 \tag{1.124}$$

Proof:

Let us take any $\bar{x} \in \Re$ and consider the Borel set $\{\bar{x}\}$. Being X measurable w.r.t. $\sigma\{X\}$, it holds that $B = X^{-1}(\{\bar{x}\}) \in \sigma\{X\}$. Using the definition of $E\left[Y|X\right](\omega)$ and using (1.119), (1.120), we compute what follows:

$$E\left[Y 1_{(X=\bar{x})}\right] = E\left[Y 1_B\right] = \int_{\Omega} Y(\omega) 1_B(\omega) dP(\omega) = \tag{1.125}$$

$$= \int_B Y(\omega) dP(\omega) = \int_B E\left[Y|X\right](\omega) dP(\omega) = \int_B E\left[Y|X = X(\omega)\right] dP(\omega) = \tag{1.126}$$

$$= \int_B E\left[Y|X = \bar{x}\right] dP(\omega) = E\left[Y|X = \bar{x}\right] P(B) = E\left[Y|X = \bar{x}\right] P\left(X = \bar{x}\right) \tag{1.127}$$

where we also used that, by construction, $\forall \omega \in B, X(\omega) = \bar{x}$. Now we further compute:

$$E[Y] = E\left[\int_{\Re} Y\delta(X-\bar{x})d\bar{x}\right] = \int_{\Omega}\left[\int_{\Re} Y(\omega)\delta(X(\omega)-\bar{x})d\bar{x}\right]dP(\omega) = \tag{1.128}$$

$$= \int_{\Re}\left[\int_{\Omega} Y(\omega)\delta\left(X(\omega)-\bar{x}\right)\right]dP(\omega)d\bar{x} = \int_{\Re} E\left[Y\delta\left(X-\bar{x}\right)\right]d\bar{x} = \tag{1.129}$$

$$= \int_{\Re} E\left[Y|X=\bar{x}\right] f_{\mu}\left(\bar{x}\right)d\bar{x} \tag{1.130}$$

In the last equation we used (1.121) and hence we found the proper normalization when using Dirac delta functions:

$$E\left[Y\delta\left(X-\bar{x}\right)\right] = E\left[Y|X=\bar{x}\right] f_{\mu}\left(\bar{x}\right) \tag{1.131}$$

Finally, using the same hypothesis of the calculations above, for any Borel function $\phi : \Re \to \Re$, we get:

$$E\left[Y\phi\left(X\right)1_{(X=\bar{x})}\right] = \int_{\Omega} Y(\omega)\phi\left(X(\omega)\right)1_B(\omega)dP(\omega) = \tag{1.132}$$

$$= \phi(\bar{x})\int_{\Omega} Y(\omega)1_B(\omega)dP(\omega) = \phi(\bar{x})E\left[Y|X=\bar{x}\right]P\left(X=\bar{x}\right) \tag{1.133}$$

and for Dirac delta functions:

$$E\left[Y\phi(X)\delta\left(X-\bar{x}\right)\right] = \phi(\bar{x})E\left[Y|X=\bar{x}\right] f_{\mu}\left(\bar{x}\right) \tag{1.134}$$

Proposition: Consider $(\Omega, \mathcal{F}, P)$ and consider $Y \in \mathcal{G} \subseteq \mathcal{F}$ and also $X \in \mathcal{F}$ such that $\sigma\{X\} \perp \mathcal{G}$. Then:

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

Proof:

We start by $Y = I_A$ with $A \in \mathcal{G}$ and for any $B \in \sigma\{X\}$

$$\int_B E[Y|X](\omega)dP(\omega) = \int_B Y(\omega)dP(\omega) = \int_B I_A(\omega)dP(\omega) = \int_{A\cap B} dP(\omega) = \tag{1.135}$$

$$= P(A\cap B) = P(A)\cdot P(B) = \left(\int_{\Omega} I_A dP(\omega)\right)\cdot P(B) = \tag{1.136}$$

$$= E[Y]\cdot P(B) = \int_B E[Y]dP(\omega) \tag{1.137}$$

which shows, being $B \in \mathcal{G}$ arbitrary, that $E[Y|X](\omega) = E[Y], P-a.s.$. To go on, the usual generalization to simple functions $\in \mathcal{G}$ and then to generic functions $\in \mathcal{G}$ is applied.

Proposition: Let $(\Omega, \mathcal{F}, P)$ be a given probability space and $\mathcal{G}$ a sub-sigma-algebra of $\mathcal{F}$. Let $X \in L^2(\Omega, \mathcal{F}, P)$. Now consider the problem of minimizing

$$E\left[(X-Z)^2\right]$$

where Z is allowed to vary in $L^2(\Omega, \mathcal{G}, P)$. The optimal solution is

$$\hat{Z}(\omega) = E[X|\mathcal{G}](\omega)$$

Proof:

We start by choosing any $A \in \mathcal{G}$ and considering $Z = I_A$, then, using the definition of $E[X|\mathcal{G}]$ and the fact that $A \in \mathcal{G}$

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

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

from which we get

$$E[Z \cdot (X - E[X|\mathcal{G}])] = 0 \tag{1.138}$$

Having proved (1.138) for indicators, we extend it in the usual way to simple and then aribitrary $\mathcal{G}$ measurable functions. Notice that (1.138) can be summurazied in words by saying that $(X - E[X|\mathcal{G}])$ is orthogonal to $\mathcal{G}$, or in again other words, by saying that $E[X|\mathcal{G}]$ is the orthogonal projection of $X \in L^2(\Omega, \mathcal{F}, P)$ onto $L^2(\Omega, \mathcal{G}, P)$. Now we write

$$\begin{aligned}
E\left[(X-Z)^2\right] &= E\left[\left[(X - E[X|\mathcal{G}]) + (E[X|\mathcal{G}] - Z)\right]^2\right] = &(1.139)\\
&= E\left[(X - E[X|\mathcal{G}])^2\right] + E\left[(E[X|\mathcal{G}] - Z)^2\right] + &(1.140)\\
&+ 2E\left[(X - E[X|\mathcal{G}]) \cdot E[X|\mathcal{G}]\right] - 2E\left[(X - E[X|\mathcal{G}])Z\right] &(1.141)
\end{aligned}$$

By applying (1.138) twice (once with Z and once with $E[X|\mathcal{G}] \in \mathcal{G}$ in the role of Z), we get

$$E\left[(X-Z)^2\right] = E\left[(X - E[X|\mathcal{G}])^2\right] + E\left[(E[X|\mathcal{G}] - Z)^2\right]$$

and then the desired result.

4.2 Equivalent Probability Measures

Proposition: Consider $(\Omega, \mathcal{F})$ equipped with two probability measures P, Q. It is true that the relation $P \sim Q$ holds if and only if

$$P(A) = 1 \Leftrightarrow Q(A) = 1, \forall A \in \mathcal{F} \tag{1.142}$$

Proof:

Suppose $P \sim Q$. Then take any $A \in \mathcal{F}$ such that $P(A) = 1$. We can compute

$$P(A) + P(A^c) = 1 \Rightarrow P(A^c) = 0 \Rightarrow Q(A^c) = 0 \Rightarrow Q(A) = 1 - Q(A^c) = 1$$

where we used the property that each measure is such that $P(A \cup B) = P(A) + P(B), \forall A \cap B = \emptyset$ and $P(\Omega) = 1$. We then shown that $P(A) = 1 \Rightarrow Q(A) = 1$. $Q(A) = 1 \Rightarrow P(A) = 1$ can be shown in the same way. Now suppose (1.142) holds. Then take any A such that $P(A) = 0$. Then as before

$$P(A^c) = 1 \Rightarrow Q(A^c) = 1 \Rightarrow Q(A) = 0$$

showing that $Q << P$. In the same way one shows $P << Q$ and hence we have $P \sim Q$. As a further observation, suppose that $P \sim Q$ and that $P(A) > 0$. Now *per assurdo* suppose that $Q(A) = 0$. We will directly get $P(A) = 0$ violating the hypothesis. Then:

$$P \sim Q \Leftrightarrow \{P(A) > 0 \Leftrightarrow Q(A) > 0\}$$

Proposition: Let us know consider a probability space $(\Omega, \mathcal{F}, P)$ and reminding the Radon-Nikodym theorem (section 3.3), we can say that for any other measure Q in the same probability space it holds that

$$Q << P \Leftrightarrow \exists L \in L^1(\Omega, \mathcal{F}, P) : \forall A \in \mathcal{F}, \int_A dQ(\omega) = \int_A L(\omega)dP(\omega)$$

Proof:

The $\Leftarrow$ is obvious, while $\Rightarrow$ is indeed the Radon-Nikodym theorem.

Being Q a probability measure, it must hold that

$$\int_\Omega L(\omega)dP(\omega) = E^P[L] = 1$$

that is to say that the Radon-Nikodym derivative is a certain non-negative random variable, measurable in $\mathcal{F}$, with the property $E^P[L] = 1$ (the likelihood ratio of Q w.r.t. P). Furthermore:

Proposition: For any random variable $X \in L^1(\Omega, \mathcal{F}, Q)$ it holds that

$$E^Q[X] = E^P[L \cdot X] \tag{1.143}$$

Proof:

Starting with any $B \in \mathcal{F}$ and with $X(\omega) = I_B(\omega)$, one has

$$E^Q[X] = \int_B dQ(\Omega) = \int_B L(\omega)dP(\omega) = \int_\Omega I_B(\omega)L(\omega)dP(\omega) = \int_\Omega X(\omega)L(\omega)dP(\omega) = E^P[L \cdot X]$$

As usual, one extends to simple and then arbitrary functions $\in \mathcal{F}$.

Proposition: Let us take again $(\Omega, \mathcal{F}, P)$ and a sub-sigma-algebra $\mathcal{G} \subseteq \mathcal{F}$. Consider then a measure $Q << P$ on $\mathcal{F}$. The we will have the possibility to see Q both on $(\Omega, \mathcal{F}, Q)$ and on the restricted measurable space $(\Omega, \mathcal{G}, Q)$. The same holds of course for the measure P. We will then have two different Radon-Nikodym derivatives: $L^{\mathcal{F}}$ on $\mathcal{F}$ and $L^{\mathcal{G}}$ on $\mathcal{G}$. Notice that they will in general be different, since $L^{\mathcal{F}}$ is $\mathcal{F}$ measurable but not $\mathcal{G} \subseteq \mathcal{F}$ measurable. Furthermore it holds that

$$L^{\mathcal{G}}(\omega) = E^P[L^{\mathcal{F}}|\mathcal{G}](\omega) \tag{1.144}$$

Proof:

First of all we notice that $E^P[L^{\mathcal{F}}|\mathcal{G}]$ is $\mathcal{G}$-measurable. Then, for any $G \in \mathcal{G}$, we notice that it also holds that $G \in \mathcal{F}$ and hence

$$\int_G dQ(\omega) = \int_G L^{\mathcal{F}}(\omega)dP(\omega) = \int_G E^P[L^{\mathcal{F}}|\mathcal{G}](\omega)dP(\omega)$$

where we used the definition of conditional expectation. This is enough to prove the result. Let us show an example:

- $\Omega = \{1, 2, 3\}$
- $\mathcal{F} = 2^\Omega$
- $\mathcal{G} = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$
- $P(1) = 1/4, P(2) = 1/2, P(3) = 1/4$
- $Q(1) = 1/3, Q(2) = 1/3, Q(3) = 1/3$
- $L^{\mathcal{F}}(1) = 4/3, L^{\mathcal{F}}(2) = 2/3, L^{\mathcal{F}}(3) = 4/3$ which is not $\mathcal{G}$ measurable
- Notice that $P(\{2, 3\}) = 3/4, Q(\{2, 3\}) = 2/3$ and $Q(\{2, 3\})/P(\{2, 3\}) = 8/9$
- We hence pose $L^{\mathcal{G}}(1) = 4/3, L^{\mathcal{G}}(2) = 8/9, L^{\mathcal{G}}(3) = 8/9$ which is $\mathcal{G}$ measurable

- Furthermore $E^P[L^{\mathcal{F}}|\{2,3\}] = \frac{P(2)L^{\mathcal{F}}(2)+P(3)L^{\mathcal{F}}(3)}{P(\{2,3\})} = 8/9 = L^{\mathcal{G}}(2) = L^{\mathcal{G}}(3)$ which agrees with (1.144).

Proposition: (Bayes' Theorem) Assume X is any random variable $\in L^1(\Omega, \mathcal{F}, P)$. Let Q be another probability measure on $(\Omega, \mathcal{F})$ with Radon-Nikodym derivative on $\mathcal{F}$ given by

$$L(\omega) = \frac{dQ}{dP}(\omega)$$

Let $\mathcal{G} \subseteq \mathcal{F}$ be a sigma-sub-algebra. Then it holds that

$$E^Q[X|\mathcal{G}] = \frac{E^P[L \cdot X|\mathcal{G}]}{E^P[L|\mathcal{G}]}, Q - a.s. \tag{1.145}$$

Proof:

First of all we notice that $Q << P$, being Q defined through a Radon-Nikodym derivative. Then we prove:

$$E^Q[X|\mathcal{G}] \cdot E^P[L|\mathcal{G}] = E^P[L \cdot X|\mathcal{G}], P - a.s. \tag{1.146}$$

We start by choosing any $G \in \mathcal{G}$ and by computing the integral on G of left-hand side of (1.146) as follows:

$$\int_G E^Q[X|\mathcal{G}] \cdot E^P[L|\mathcal{G}] dP(\omega) = \int_G E^P[L \cdot E^Q[X|\mathcal{G}]|\mathcal{G}] dP(\omega) = \tag{1.147}$$

$$= \int_G L \cdot E^Q[X|\mathcal{G}] dP(\omega) = \int_G E^Q[X|\mathcal{G}] dQ(\omega) = \int_G X dQ(\omega) \tag{1.148}$$

where we used

- in the first passage: the fact that $E^Q[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and hence it can pass inside $E^P[L|\mathcal{G}]$
- in the second passage: the definition of $E^P[\cdot|\mathcal{G}]$
- in the third passage: the property in (1.143)
- in the fourth passage: the definition of $E^Q[\cdot|\mathcal{G}]$

Now we compute the same integral on G but on the right-hand side of (1.146):

$$\int_G E^P[L \cdot X|\mathcal{G}] dP(\omega) = \int_G L \cdot X dP(\omega) = \int_G X dQ(\omega)$$

Being G arbitrary we can conlude that (1.146) holds. Since $Q << P$ as observed before, we can then directly conclude that (1.145) also holds (i.e. holds $Q - a.s.$). To complete the proof, we have to prove that (1.145) is well defined, i.e. that $E^P[L|\mathcal{G}] \neq 0, Q - a.s.$. We can calculate:

$$Q\left(\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega) = 0\}\right) = \int_{\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega)=0\}} dQ(\omega) = \tag{1.149}$$

$$= \int_{\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega)=0\}} L(\omega) dP(\omega) = \int_{\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega)=0\}} E^P[L|\mathcal{G}](\omega) dP(\omega) = \tag{1.150}$$

$$= \int_{\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega)=0\}} 0 \cdot dP(\omega) = 0 \tag{1.151}$$

which completes the proof. Notice that in deriving (1.151), we used that

- $\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega) = 0\} \in \mathcal{G}$ since it is the counter-image of the Borel set $\{0\}$ through the $\mathcal{G}$-measurable random variable $E^P[L|\mathcal{G}]$
- equation (1.143)
- the definition of $E^P[\cdot|\mathcal{G}]$ taking into account that $\{\omega \in \Omega : E^P[L|\mathcal{G}](\omega) = 0\} \in \mathcal{G}$.

5 Martingales and Stopping Times

Definition: Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and $\{X_t\}_t$ a random process in continuous or discrete time. The process X is and $\underline{\mathcal{F}}$-martingale if

- X is $\underline{\mathcal{F}}$-adapted
- $X_t \in L^1(\Omega, \mathcal{F}_t, P), \forall t$
- for every $0 \leq s \leq t$ it holds that

$$X_s(\omega) = E[X_t(\omega)|\mathcal{F}_s](\omega), P - a.s. \tag{1.152}$$

- if the $=$ sign is replaced by $\leq$ $(\geq)$, X is said to be a submartingale (supermartingale).

Proposition: Since $E[X_s|\mathcal{F}_s] = X_s, P - a.s.$, if X is a $\underline{\mathcal{F}}$-martingale, we have $E[X_t - X_s|\mathcal{F}_s] = 0, \forall s \leq t$ For processes in discrete time the martingale property must hold just for adjacent times (and will extend to all times automatically). We can indeed state what follows.

Proposition: An adapted integrable discrete time process $\{X_n; n = 0, 1, \cdots\}$ is a martingale w.r.t. the filtration $\{\mathcal{F}_n; n = 0, 1, \cdots\}$ if and only if

$$E[X_{n+1}|F_n] = X_n, \forall n = 0, 1, \cdots \tag{1.153}$$

Proof:

$\Rightarrow$ follows from the martingale definition. Take instead any integer $d > 0$ and compute

$$E[X_{n+d}|\mathcal{F}_n] = E[E[\cdots E[X_{n+d}|\mathcal{F}_{n+d-1}]\cdots|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \tag{1.154}$$

$$= E[E[\cdots E[X_{n+d-1}|\mathcal{F}_{n+d-2}]\cdots|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \cdots = E[X_{n+1}|\mathcal{F}_n] = X_n \tag{1.155}$$

which is what we were looking for.

Proposition: Given Y any integrable random variable on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$, define $X_t(\omega) = E[Y|\mathcal{F}_t](\omega)$. Then X_t ia an $\underline{\mathcal{F}}$-martingale since for $s \leq t$ (being $\mathcal{F}_s \subseteq \mathcal{F}_t$):

$$E[X_t|\mathcal{F}_s] = E[E[Y|\mathcal{F}_t]|\mathcal{F}_s] = E[Y|\mathcal{F}_s] = X_s$$

Proposition: Now consider a compact interval $[0, T]$ and take any martingale process M_t. We notice that for any $t \in [0, T]$ it holds that $M_t = E[M_T|\mathcal{F}_t], P - a.s.$. This means that any martingale M_t on a compact interval $[0, T]$ can always be represented $(P - a.s.)$ as expectation of its final time random variable ($M_T(\omega)$, on T). Notice that for non-compact intervals the situation will be more complicated.

Proposition: If X is a process with independent increments on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ and also if $E[X_t - X_s] = 0, \forall s, t$, then X is a martingale.

Proposition: Let X be a process on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$. Then

- If X is a martingale and if $f : \Re \to \Re$ is a convex (concave) function such that $f(X_t)$ is integrable $\forall t$, then the process
$$Y_t = f(X_t)$$
is a submartingale (supermartingale).
- If X is a submartingale and $f : \Re \to \Re$ is a convex non-decreasing function such that $f(X_t)$ is integrable $\forall t$, then the process
$$Y = f(X_t)$$
is also submartingale.

Proof:

For the first statement, if f is convex (concave) we have by Jensen inequality and for $s \leq t$:

$$E[Y_t|\mathcal{F}_s] = E[f(X_t)|\mathcal{F}_s] \geq (\leq) f\left(E[X_t|\mathcal{F}_s]\right) = f(X_s) = Y_s, P - a.s.$$

For the second statement, we have by Jensen inequality and for $s \leq t$:

$$E[Y_t|\mathcal{F}_s] = E[f(X_t)|\mathcal{F}_s] \geq f\left(E[X_t|\mathcal{F}_s]\right) \geq f(X_s) = Y_s, P - a.s.$$

where we have used that by hypothesis $X_s \leq E[X_t|\mathcal{F}_s]$ and that f is also non-decreasing.

We saw that on a finite interval $[0, T]$, every martingale can be represented by

$$X_t = E[X_T|\mathcal{F}_t]$$

but, in general on an infinite interval $[0, \infty]$ this does not hold, i.e. it is not always true that there exists an integrable random variable X_∞ such that

$$X_t = E[X_\infty|\mathcal{F}_t] \tag{1.156}$$

Proposition: (No proof, see below for a proof in a less general case.) Let X be a submartingale on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ satisfying the condition

$$\sup_{t\geq 0} E[X_t^+] < \infty \tag{1.157}$$

Then there exists a random variable Y such that $X_t \to Y, P - a.s.$, i.e. such that

$$P\left(\omega \in \Omega : \lim_{t\to+\infty} X_t(\omega) = Y(\omega)\right) = 1 \tag{1.158}$$

Definition: A martingale X on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ is defined to be **squared integrable** if there exists a constant M such that

$$E[X_t^2] \leq M, \forall t \geq 0 \tag{1.159}$$

Proposition: (**Martingale Convergence**) Assume X is a square integrable martingale. Then there exists a random variable, which we denote X_∞, such that, for $t \to +\infty$, one has convergence of $X_t \to X_\infty$ both in $L^2(\Omega, \mathcal{F}, P)$ and also $P - a.s.$, i.e. both the following two relations hold:

$$\lim_{t\to+\infty} \|X_t - X_\infty\|_2 = \lim_{t\to+\infty} \left(E[(X_t - X_\infty)^2]\right)^{\frac{1}{2}} = 0 \tag{1.160}$$

$$P\left(\omega \in \Omega : \lim_{t\to+\infty} X_t(\omega) = X_\infty(\omega)\right) = 1 \tag{1.161}$$

Furthermore the following representation holds:

$$X_t = E[X_\infty|\mathcal{F}_t], \forall t \geq 0 \tag{1.162}$$

Proof:

The function $x \to x^2$ is convex and hence the process X_t^2 is a submartingale (see above). This means that the mapping

$$m_t = E[X_t^2] \tag{1.163}$$

is such that for $s \leq t$

$$m_s = E[X_s^2] \leq E[E[X_t^2|\mathcal{F}_s]] = E[X_t^2] = m_t$$

i.e. m_t is non-decreasing. The assumption that X_t is square integrable is then equivalent to the existence of a real number $c < +\infty$ such that

$$\lim_{t\to+\infty} m_t = \lim_{t\to+\infty} E[X_t^2] = c$$

We will now prove the L^2 convergence, i.e. (1.160), by showing that X_t is Cauchy in L^2 (remind (1.102)). Reminding that X_t is a martingale and taking any $0 \le s \le t$, we have

$$E[(X_t - X_s)^2] = E[X_t^2 - 2X_sX_t + X_s^2] = E[E[X_t^2 - 2X_sX_t + X_s^2|\mathcal{F}_s]] = \tag{1.164}$$

$$= E[X_t^2] + E[X_s^2] - 2E[X_sE[X_t|\mathcal{F}_s]] = E[X_t^2] - E[X_s^2] = m_t - m_s \tag{1.165}$$

Since we know that $\lim_{t\to+\infty} m_t = c$, we conclude that m_s is Cauchy and hence that also X_t is Cauchy in L^2. Having proved the L^2 convergence, we have proved the existence of $X_\infty \in L^2$. Now we should also prove the $P - a.s.$ convergence to such X_∞, i.e. (1.161). We can do this by directly using proposition (1.158), since we know for sure that $\sup_{t\ge 0} E[X_t^+] < \infty$ being X_t squared integrable by hypothesis. Now it only remains to prove the representation (1.162). Let us fix $s \ge 0$ and take any $A \in \mathcal{F}_s$ and try to prove that it holds that

$$\int_A X_s(\omega)dP(\omega) = \int_A X_\infty(\omega)dP(\omega) = \int_A E[X_\infty|\mathcal{F}_s](\omega)dP(\omega) \tag{1.166}$$

Being A arbitrary and both X_s and $E[X_\infty|\mathcal{F}_s]$ $\mathcal{F}_s$-measurale, (1.166) would be enough to prove (1.162), i.e. that

$$X_s = E[X_\infty|\mathcal{F}_s], P - a.s.$$

Being X_t a martingale, we know that for every $s \le t$:

$$X_s(\omega) = E[X_t|\mathcal{F}_s](\omega), P - a.s. \tag{1.167}$$

and hence

$$\int_A X_s(\omega)dP(\omega) = \int_A E[X_t|\mathcal{F}_s](\omega)dP(\omega)$$

Being $A \in \mathcal{F}_s$ and using the definition of $E[\cdot|\mathcal{F}_s]$

$$\int_A X_s(\omega)dP(\omega) = \int_A X_t(\omega)dP(\omega) \tag{1.168}$$

At this point we remind that, taken $r > s \ge 1$, we have that L^r convergences implies L^s convergence:

$$\lim_{n\to\infty} \|f_n - f\|_r = 0 \Rightarrow \lim_{n\to\infty} \|f_n - f\|_s = 0 \tag{1.169}$$

We use this fact together with the fact that we know that $\|X_t - X_\infty\|_2 \to 0$ for $t \to \infty$ and hence we get:

$$0 = \lim_{t\to\infty} \|X_t - X_\infty\|_1 = \lim_{t\to\infty} E[|X_t - X_\infty|] = \lim_{t\to\infty} \int_\Omega |X_t - X_\infty|dP(\omega) \tag{1.170}$$

Finally

$$\lim_{t\to\infty} \left| \int_A X_t(\omega)dP(\omega) - \int_A X_\infty(\omega)dP(\omega) \right| = \lim_{t\to\infty} \left| \int_A (X_t(\omega) - X_\infty(\omega))\, dP(\omega) \right| \le \tag{1.171}$$

$$\le \lim_{t\to\infty} \int_A |X_t(\omega) - X_\infty(\omega)|\, dP(\omega) \le \lim_{t\to\infty} \int_\Omega |X_t(\omega) - X_\infty(\omega)|\, dP(\omega) = 0 \tag{1.172}$$

that, together with (1.168), proves (1.166) that we supposed to hold in the discussion above.

5.1 Discrete Stochastic Integrals

Definition: Consider a filtered space $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ in discrete time, i.e. $n = 0, 1, 2, \cdots$. Then

- A random process X is $\underline{\mathcal{F}}$-predictable if, $\forall n$, X_n is $\mathcal{F}_{n-1}$-measurable, with the convention $\mathcal{F}_{-1} = \mathcal{F}_0$. Notice that a predictable process is known one step ahead in time.
- For any random process X, the increment process ΔX is defined by

$$(\Delta X)_n = X_n - X_{n-1} \tag{1.173}$$

with the convention $X_{-1} = 0$. Notice that ΔX is adapted to $\underline{\mathcal{F}}$ since X is adapted and the icrement is taken backward.

- For any two processes X and Y, the discrete stochastic integral process $X * Y$ is defined by

$$(X * Y)_n = \sum_{k=0}^{n} X_k (\Delta Y)_k \tag{1.174}$$

that can also be denoted by $\int_0^n X_s dY_s$.

Proposition: Consider a filtered space $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ in discrete time, i.e. $n = 0, 1, 2, \cdots$. Let X be a predictable process, M a martingale and suppose that the process $X_n(\Delta M)_n$ is integrable for each n. Then

$$X * M \tag{1.175}$$

is also a martingale.

Proof:

Being X_{n+1} measurable w.r.t. $\mathcal{F}_n$:

$$E[(X * Y)_{n+1}|\mathcal{F}_n] = E\left[\sum_{k=0}^{n+1} X_k (\Delta Y)_k | \mathcal{F}_n\right] = (X * Y)_n + E[X_{n+1}(\Delta M)_{n+1}|\mathcal{F}_n] =$$
$$= (X * Y)_n + E[X_{n+1}(M_{n+1} - M_n)|\mathcal{F}_n] = (X * Y)_n + X_{n+1}\left(E[(M_{n+1})|\mathcal{F}_n] - M_n\right) = (X * Y)_n$$

5.2 Likelihood Processes

Let us consider a filtered space $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ and a compact interval $[0, T]$. Suppose that L_T is some non negative integrable random variable $\in \mathcal{F}_T$. We can define a new measure Q on $\mathcal{F}_T$ by setting

$$dQ = L_T \cdot dP, \text{on } \mathcal{F}_T$$

and is $E^P[L_T] = 1$, then Q will also be a probability (measure). L_T will hence be the Radon-Nikodym derivative of Q w.r.t. P on $\mathcal{F}_T$ and it will hold that $Q << P$ on $\mathcal{F}_T$. Since $\mathcal{F}_t \subseteq \mathcal{F}_T$, it will also hold that $Q << P$ on $\mathcal{F}_t$, $\forall t \in [0, T]$. By the Radon-Nikodym theorem, for any t, we will hence imply the existence of an L_t of Q w.r.t. P on $\mathcal{F}_t$:

$$L_t = \frac{dQ}{dP}, \text{on } \mathcal{F}_t$$

In other words, we have defined a process: the likelihood process for the measure transformation from P to Q on $[0, T]$. By proposition (1.144), it holds that

$$L_t = E[L_T|\mathcal{F}_t] \tag{1.176}$$

and hence we see that L_t is a $(P, \underline{\mathcal{F}})$-martingale. Furthermore, let us take a process M and, for any $0 \leq s \leq t$, let us use Bayes' theorem (1.145):

$$E^Q[M_t|\mathcal{F}_s] = \frac{E^P[M_t \cdot L_t|\mathcal{F}_s]}{E^P[L_t|\mathcal{F}_s]} = \frac{E^P[M_t \cdot L_t|\mathcal{F}_s]}{E^P[E^P[L_T|\mathcal{F}_t]|\mathcal{F}_s]} = \frac{E^P[M_t \cdot L_t|\mathcal{F}_s]}{L_s} \tag{1.177}$$

from which we conclude that

$$E^Q[M_t|\mathcal{F}_s] = M_s \Leftrightarrow E^P[M_t \cdot L_t|\mathcal{F}_s] = M_s L_s$$

which relates the martingale condition in two different measures.

5.3 Stopping Times

Definition: Let us consider a filtered probability space $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$. A stopping time w.r.t. the filtration $\underline{\mathcal{F}}$, is a non negative random variable T, such that

$$\{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t, \forall t \geq 0 \tag{1.178}$$

Suppose X is a discrete time adapted process and define

$$T \equiv \inf\{n \geq 0; X_n \in A\} \tag{1.179}$$

where A is some Borel set. Then

$$\{\omega \in \Omega : T(\omega) \leq n\} = \{\omega \in \Omega : X_t(\omega) \in A, t \leq n\} = \bigcup_{t=0}^{n} \{\omega \in \Omega : X_t(\omega) \in A\}$$

Since $X_t \in \mathcal{F}_t \subseteq \mathcal{F}_n$ (X_t is adapted), we have $\{\omega \in \Omega : X_t(\omega) \in A\} \in \mathcal{F}_t$ and hence $\{\omega \in \Omega : T(\omega) \leq n\} \in \mathcal{F}_n$, which formally shows that (1.179) defines a stopping time.

Definition: Let T be an $\underline{\mathcal{F}}$ stopping time on a filtered probability space $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$. The sigma algebra $\mathcal{F}_T$, i.e. the sigma algebra generated by the stopping time T, is defined as the class of events satisfying

$$\left\{ \begin{matrix} A \in \mathcal{F}_\infty \\ A \cap \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t, \forall t \geq 0 \end{matrix} \right\} \tag{1.180}$$

Proposition: $\mathcal{F}_T$ is indeed a sigma algebra.

Proof:

If $A \in \mathcal{F}_T$ then

$$\mathcal{F}_t \ni A \cap \{T \leq t\} = (A^c \cup \{T > t\})^c$$

Since $\mathcal{F}_t$ is a sigma algebra:

$$\mathcal{F}_t \ni (A \cap \{T \leq t\})^c = A^c \cup \{T > t\}$$

Now it holds that

$$\mathcal{F}_t \ni (A^c \cup \{T > t\}) \cap \{T \leq t\} = A^c \cap \{T \leq t\}$$

where the $\ni$ holds since both $(A^c \cup \{T > t\})$ and $\{T \leq t\}$ are $\in \mathcal{F}_t$ (which is a sigma algebra). So we proved that if $A \in \mathcal{F}_T$ then A^c is also $\in \mathcal{F}_T$. If instead we have $A_1, A_2, \cdots \in \mathcal{F}_T$ then

$$\left(\bigcup_i A_i\right) \cap \{T \leq t\} = \bigcup_i (A_i \cap \{T \leq t\}) \in \mathcal{F}_t$$

since each $A_i \cap \{T \leq t\} \in \mathcal{F}_t$ which is a sigma algebra. Hence $(\bigcup_i A_i) \in \mathcal{F}_T$. This completes the prove that $\mathcal{F}_T$ is a sigma algebra ($\emptyset \in \mathcal{F}_T$ is obvious).

Proposition: Let S and T be two stopping times on a filtered probability space $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$. Let X be an adapted process, which, in continuous time, is assumed to have either left or right continuous trajectories. For any two real numbers x, y, define $x \vee y = \max(x, y)$, $x \wedge y = \min(x, y)$. Define then

$$(S \vee T)(\omega) = S(\omega) \vee T(\omega) \tag{1.181}$$
$$(S \wedge T)(\omega) = S(\omega) \wedge T(\omega) \tag{1.182}$$

Then the following hold:

- if $S \leq T, P - a.s.$ then $\mathcal{F}_S \subseteq \mathcal{F}_T$
- $(S \vee T)$ and $(S \wedge T)$ are stopping times
- if T is $P - a.s.$ finite or if X_∞ is well defined in $\mathcal{F}_\infty$, then X_T is $\mathcal{F}_T$ measurable.

Proof:

- (First item). We restrict to the case $S \leq T$ always (i.e. $\forall \omega \in \Omega$) and not just $P - a.s.$ Let us fix any $t \geq 0$ and define

$$A_S^t = \{\omega \in \Omega : S(\omega) \leq t\} \in \mathcal{F}_t \tag{1.183}$$
$$A_T^t = \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t \tag{1.184}$$

  If $\omega \in A_T^t$ then $S(\omega) \leq T(\omega) \leq t$ and hence $\omega \in A_S^t$, which means that $A_T^t \subseteq A_S^t$. Now take any $A \in \mathcal{F}_\infty$ and suppose that

$$A \cap \{\omega \in \Omega : S(\omega) \leq t\} = A \cap A_S^t \in \mathcal{F}_t, \forall t \geq 0$$

  and compute

$$A \cap A_T^t = A \cap (A_S^t \cap A_T^t) = (A \cap A_S^t) \cap A_T^t \in \mathcal{F}_t$$

  where the last conclusion "$\in \mathcal{F}_t$" follows from the fact that $(A \cap A_S^t) \cap A_T^t$ is the intersection between two sets in $\mathcal{F}_t$, i.e. $(A \cap A_S^t)$ and A_T^t. We have then proved that an event A that generates $\mathcal{F}_S$ also generates $\mathcal{F}_T$, i.e. $\mathcal{F}_S \subseteq \mathcal{F}_T$.

- (Second item). Let us compute

$$\{\omega \in \Omega : (S \vee T)(\omega) \leq t\} = A_S^t \cap A_T^t \in \mathcal{F}_t$$

  being $A_S^t, A_T^t \in \mathcal{F}_t$, which proves that $(S \vee T)$ is a stopping time. Similarly we can prove that that $(S \wedge T)$ is a stopping time:

$$\{\omega \in \Omega : (S \wedge T)(\omega) \leq t\} = A_S^t \cup A_T^t \in \mathcal{F}_t$$

- (Third item). We restrict to the discrete time case. In order to prove that X_T is $\mathcal{F}_T$ measurable, for any Borel set $B \in \mathcal{B}(\Re)$, we have to prove that

$$\{\omega \in \Omega : X_T(\omega) \in B\} \in \mathcal{F}_T$$

  This means that we have to show that, for every time n, it holds that

$$\{\omega \in \Omega : X_T(\omega) \in B\} \cap \{\omega \in \Omega : T(\omega) \leq n\} \in \mathcal{F}_n \tag{1.185}$$

We can then compute:

$$\{X_T \in B\} \cap \{T \leq n\} = \{X_T \in B\} \cap \bigcup_{k=0}^{n} \{T = k\} = \tag{1.186}$$

$$= \bigcup_{k=0}^{n} \{X_T \in B\} \cap \{T = k\} = \bigcup_{k=0}^{n} \{X_k \in B\} \cap \{T = k\} \tag{1.187}$$

We know that $\{X_k \in B\} \in \mathcal{F}_k$, since X is adapted by hypothesis, and furthermore that $\{T = k\} \in \mathcal{F}_k$, since T is a stopping time. Since $\mathcal{F}_k \subseteq \mathcal{F}_n$, we can conclude (1.185), i.e. $X_T \in \mathcal{F}_T$.

Proposition: Let X be a martingale and T a stopping time. Then the stopped process X^T defined by

$$X_t^T = X_{T \wedge t} \tag{1.188}$$

is a martingale.

Proof:

We restrict to the discrete time case. Let us define the process h:

$$h_n(\omega) = I\{n \leq T(\omega)\}, n = 0, 1, 2, \cdots$$

For a fixed n, we can state that

$$\{\omega \in \Omega : n \leq T(\omega)\} = \{\omega \in \Omega : n > T(\omega)\}^c = \{\omega \in \Omega : n - 1 \geq T(\omega)\}^c \in \mathcal{F}_{n-1}$$

We thus see that $h_n in \mathcal{F}_{n-1}$, i.e. h is predictable. Furthermore we can write

$$X_n^T = X_{T \wedge n} = \sum_{k=0}^{n} I\{k \leq T\}(X_k - X_{k-1}) = \sum_{k=0}^{n} h_k \cdot (X_k - X_{k-1})$$

which, by proposition (1.175), we deduce that the stopped process X_n^T is also a martingale.

Proposition: (The Optional Sampling Theorem). Consider as usual a filtered probability space $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$. Let X be a martingale such that

$$\sup_{t \geq 0} E[X_t^2] < \infty \tag{1.189}$$

Let then S and T be two stopping times such that $S \leq T$. The we can conclude that

$$E[X_T | \mathcal{F}_S] = X_S, P - a.s. \tag{1.190}$$

Furthermore, if X is a submartingale, always satisfying (1.189), then (1.190) transforms to

$$X_S \leq E[X_T | \mathcal{F}_S], P - a.s.$$

Proof:

We restrict to the case in which X is a martingale and time is discrete. Due to the martingale representation theorem (1.162), we know that there exists an integrable random variable, say Y, such that

$$X_n = E[Y | \mathcal{F}_n], n = 0, 1, 2, \cdots \tag{1.191}$$

At this point let us try to prove that

$$X_T = E[Y|\mathcal{F}_T] \tag{1.192}$$

If we manage to prove (1.192), the thesis will be proved since

$$E[X_T|\mathcal{F}_S] = E[E[Y|\mathcal{F}_T]|\mathcal{F}_S] = E[Y|\mathcal{F}_S] = X_S$$

where we used that $\mathcal{F}_S \subseteq \mathcal{F}_T$ since $S \leq T$ (see proposition above). We will then prove that for any $A \in \mathcal{F}_T$ we have

$$\int_A Y(\omega)dP(\omega) = \int_A X_T(\omega)dP(\omega)$$

For this purpose, let us write

$$A = \bigcup_n (A \cap \{T = n\})$$

and notice that $A \cap \{T = n\} \in \mathcal{F}_n$, to get that

$$\int_A Y(\omega)dP(\omega) = \sum_{n=0}^{\infty} \int_{A\cap\{T=n\}} Y(\omega)dP(\omega) = \sum_{n=0}^{\infty} \int_{A\cap\{T=n\}} X_n(\omega)dP(\omega) = \tag{1.193}$$

$$= \sum_{n=0}^{\infty} \int_{A\cap\{T=n\}} X_T(\omega)dP(\omega) = \int_A X_T(\omega)dP(\omega) \tag{1.194}$$

A Wiener process on $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$ is a process such that

- $W_0 = 0$
- $(W_t - W_s)$ is distributed as a Gaussian $\mathcal{N}(0, t - s)$
- increments are independent, i.e. $(W_t - W_s) \perp \mathcal{F}_s$
- trajectories are continuous

It is possible to prove that $W_t, (W_t^2 - t), \exp(\lambda W_t - \frac{1}{2}\lambda^2 t)$ (for any $\lambda \in \Re$) are all martingales. Now we define the stopping time T as the first time the process W_t hits the barriers $b < 0 < a$, i.e.

$$T = \inf\{t \geq 0 : X_t = a \text{ or } X_t = b\} \tag{1.195}$$

We then define

$$p_a = P(W_t \text{ hits a before hitting b}) = P(W_T = a) \tag{1.196}$$

$$p_b = P(W_t \text{ hits b before hitting a}) = P(W_T = b) \tag{1.197}$$

Then we prove the following facts.

- We know that $W_t^T = W_{T\wedge t}$ remains a martingale (see (1.188)) and hence

  $$0 = \lim_{t\to\infty} W_0 = \lim_{t\to\infty} E[W_{T\wedge t}|\mathcal{F}_0] = E[W_T|\mathcal{F}_0] = p_a a + p_b b \text{ and } p_a + p_b = 1$$

  from which

  $$p_a = \frac{-b}{a-b}, p_b = \frac{a}{a-b}$$

  where we used (without proof) that $P(T < \infty) = 1$.

- In the same way $W_{T\wedge t}^2 - T \wedge t$ must remain a martingale and then

  $$0 = \lim_{t\to\infty} W_0^2 - 0^2 = \lim_{t\to\infty} E[W_{T\wedge t}^2 - T \wedge t|\mathcal{F}_0] = E[W_T^2 - T|\mathcal{F}_0] = p_a a^2 + p_b b^2 - E[T]$$

  from which

  $$E[T] = |ab| \tag{1.198}$$

- Using that $\exp(\lambda W_{T\wedge t} - \frac{1}{2}\lambda^2 T \wedge t)$ is still a martingale, and assuming $a = -b$:

$$1 = \lim_{t\to\infty} \exp\left(\lambda W_0 - \frac{1}{2}\lambda^2 0\right) = \lim_{t\to\infty} E\left[\exp\left(\lambda W_{T\wedge t} - \frac{1}{2}\lambda^2 T \wedge t\right) |\mathcal{F}_0\right] = \tag{1.199}$$

$$= E\left[\exp\left(\lambda W_T - \frac{1}{2}\lambda^2 T\right) |\mathcal{F}_0\right] = E\left[\exp\left(\lambda W_T - \frac{1}{2}\lambda^2 T\right)\right] = \tag{1.200}$$

$$= \int_{\{W_T=a\}} e^{\lambda a} e^{-\frac{1}{2}\lambda^2 T} + \int_{\{W_T=-a\}} e^{-\lambda a} e^{-\frac{1}{2}\lambda^2 T} \tag{1.201}$$

Notice that we used that $\Omega = \{W_T = -a\} \cup \{W_T = a\}$ and, being $a > 0$, $\{W_T = -a\} \cap \{W_T = a\} = \emptyset$. Now, by symmetry, we know that

$$\int_{\{W_T=a\}} e^{-\frac{1}{2}\lambda^2 T} = \int_{\{W_T=-a\}} e^{-\frac{1}{2}\lambda^2 T} = \frac{E[e^{-\frac{1}{2}\lambda^2 T}]}{2}$$

and hence

$$1 = E[e^{-\frac{1}{2}\lambda^2 T}] \frac{e^{\lambda a} + e^{-\lambda a}}{2}$$

Defining $\frac{\lambda^2}{2} = \rho$ with $\rho \geq 0$, we get

$$E[e^{-\rho T}] = \frac{2}{e^{a\sqrt{2\rho}} + e^{-a\sqrt{2\rho}}} \tag{1.202}$$

For small ρ, notice that

$$E[e^{-\rho T}] \approx 1 - \rho E[T] = \frac{2}{e^{a\sqrt{2\rho}} + e^{-a\sqrt{2\rho}}} \approx \frac{1}{1 + \rho a^2} \approx 1 - \rho a^2$$

which correctly agrees with (1.198) (notice that (1.202) is different from the result claimed in [2] at page 452, point (vi) of exercise C.8 and in particular such claim does not agree with its own point (v), hence [2] is probably wrong on the computation of (1.202)).

6 Girsanov's Theorem

Proposition: (From [2]) Let W^P be a d-dimensional standard (i.e. independent components) P-Wiener process on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ and let φ be any d-dimensional adapted column vector process. Choose a fixed $T > 0$ and define the (scalar) process L on $[0, T]$ by

$$dL_t = \varphi_t^* L_t dW_t^P ; L_0 = 1$$

that is

$$L_t = \exp\left[\int_0^t \varphi_s^* dW_s^P - \frac{1}{2}\int_0^t \|\varphi_s\|^2 ds\right]$$

Assume that

$$E^P[L_T] = 1$$

and define the new probability measure Q on $\mathcal{F}_T$ by

$$L_T = \frac{dQ}{dP}$$

Then

$$dW_t^P = \varphi_t dt + dW_t^Q \tag{1.203}$$

$$W_t^Q = W_t^P - \int_0^t \varphi_s ds \tag{1.204}$$

where W^Q is a Q-Wiener process.

Proof:

TODO:

7 Changing measures between equivalent martingale measures

For every non dividend paying tradable asset H, measurable on $\mathcal{F}_\tau$, and considering two numerairs $N(\tau)$, $M(\tau)$, we know that the following holds ($t \leq \tau$):

$$H(t) = N(t)E^N\left[\frac{H(\tau)}{N(\tau)}|\mathcal{F}_t\right] = M(t)E^M\left[\frac{H(\tau)}{M(\tau)}|\mathcal{F}_t\right]$$

which, defining $G(\tau) = \frac{H(\tau)}{N(\tau)}$, yields

$$E^N\left[G(\tau)|\mathcal{F}_t\right] = E^M\left[G(\tau)\frac{M(t)}{M(\tau)}\frac{N(\tau)}{N(t)}|\mathcal{F}_t\right] = E^M\left[G(\tau)\frac{dQ^N}{dQ^M}(\tau)|\mathcal{F}_t\right]$$

which gives

$$\frac{dQ^N}{dQ^M}(\tau) = \frac{M(t)}{M(\tau)}\frac{N(\tau)}{N(t)} \tag{1.205}$$

8 Changing measures between T-fwd martingale measures

Let $t_0 < t_1 < \cdots < t_N$ be a time schedule and referring to (1.205) we assume

$$M(t) = P(t, t_k) \tag{1.206}$$
$$N(t) = P(t, t_h) \tag{1.207}$$

with $t \leq \tau \leq \min(t_k, t_h)$ and $h, k \geq 0$. By $P(\tau, T)$ we denote the risk free zero coupon bond price as observed from τ for maturity T, delivering 1 unit of currency at T. We then get

$$\frac{dQ^N}{dQ^M}(\tau) = \frac{dQ^h}{dQ^k}(\tau) = \frac{P(t, t_k)}{P(\tau, t_k)}\frac{P(\tau, t_h)}{P(t, t_h)} = L_{t,h/k}(\tau) \tag{1.208}$$

where $L_{t,h/k}(t) = 1$ and

$$E^h\left[G(\tau)|\mathcal{F}_t\right] = E^k\left[G(\tau)L_{t,h/k}(\tau)|\mathcal{F}_t\right] \tag{1.209}$$

We now write

$$P(\tau, h/k) \equiv \frac{P(\tau, t_h)}{P(\tau, t_k)} = \left\{\Pi_{i=m_{h,k}+1}^{M_{h,k}}\left[1 + F_i(\tau)\tau_i\right]\right\}^{s(h,k)} \tag{1.210}$$
$$m_{h,k} = \min(h, k) \tag{1.211}$$
$$M_{h,k} = \max(h, k) \tag{1.212}$$
$$s(h, k) = 1 \text{ if } h \leq k, -1 \text{ if } h > k \tag{1.213}$$

and

$$F_i(\tau) \equiv \left(\frac{P(\tau, t_{i-1})}{P(\tau, t_i)} - 1\right)\frac{1}{\tau_i} \tag{1.214}$$

We now compute the following Ito differential in the t_k-forward measure

$$dP(\tau, h/k) = \sum_{l=m_{h,k}+1}^{M_{h,k}} dF_l(\tau) \frac{\partial}{\partial F_l(\tau)} P(\tau, h/k) = \tag{1.215}$$

$$= \sum_{l=m_{h,k}+1}^{M_{h,k}} \tau_l s(h,k) dF_l(\tau) \left[1 + F_l(\tau)\tau_l\right]^{s(h,k)-1} \left\{\Pi_{i=m_{h,k}+1, i\neq l}^{M_{h,k}} \left[1 + F_i(\tau)\tau_i\right]\right\}^{s(h,k)} = \tag{1.216}$$

$$= s(h,k) P(\tau, h/k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l dF_l(\tau)}{1 + F_l(\tau)\tau_l} \tag{1.217}$$

From the last equation and (1.208) we get (always in the t_k-forward measure)

$$dL_{t,h/k}(\tau) = L_{t,h/k}(\tau) s(h,k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l dF_l(\tau)}{1 + F_l(\tau)\tau_l} \tag{1.218}$$

We now use this last equation, together with the Girsanov's theorem (1.203) to get the following result. Suppose to consider a 1-dimensional Wiener process $Z^{(k)}(\tau)$ where the (k)-apex is used to stress that it is a Wiener process in the t_k-fwd measure. We can then conclude that $Z^{(h)}(\tau)$ as defined in the following is a Wiener process in the t_h-fwd measure:

$$dZ^{(h)}(\tau) = dZ^{(k)}(\tau) - s(h,k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l \left\langle dF_l(\tau) \cdot dZ^{(k)}(\tau) \right\rangle}{1 + F_l(\tau)\tau_l} \tag{1.219}$$

8.1 Shifted log-normal fwd rate model and measure change between T-fwd martingale measures

We now assume that

$$F_l(\tau) = \lambda_l + f_l(\tau) \tag{1.220}$$

where λ_l is a constant and $f_l(\tau)$ is the following (martingale) log-normal process in the t_l-fwd measure:

$$\frac{df_l(\tau)}{f_l(\tau)} = -\frac{1}{2}\sigma_l^2 d\tau + \sigma_l dW_l^{(l)}(\tau) \tag{1.221}$$

where we used a notation that stresses that $W_l^{(l)}$ is the Wiener process driving $f_l(\tau)$ in the t_l-fwd measure and we introduced an annual volatility process σ_l (not a constant in general). We now further observe the obvious relation

$$dF_l(\tau) = df_l(\tau) \tag{1.222}$$

Equation (1.246) then becomes

$$dZ^{(h)}(\tau) = dZ^{(k)}(\tau) - s(h,k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l \sigma_l f_l(\tau) \left\langle dW_l^{(l)}(\tau) \cdot dZ^{(k)}(\tau) \right\rangle}{1 + F_l(\tau)\tau_l} = \tag{1.223}$$

$$= dZ^{(k)}(\tau) - s(h,k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l \sigma_l (F_l(\tau) - \lambda_l) \left\langle dW_l^{(l)}(\tau) \cdot dZ^{(k)}(\tau) \right\rangle}{1 + F_l(\tau)\tau_l} \tag{1.224}$$

We further notice that one could apply the 'freezing the drift approximation' in the last equation to get

$$dZ^{(h)}(\tau) = dZ^{(k)}(\tau) - s(h,k) \sum_{l=m_{h,k}+1}^{M_{h,k}} \frac{\tau_l \sigma_l (F_l(t) - \lambda_l) \left\langle dW_l^{(l)}(\tau) \cdot dZ^{(k)}(\tau) \right\rangle}{1 + F_l(t)\tau_l} \tag{1.225}$$

where t has been substituted in place of τ in some of the factors.

8.2 Changing measures between T-fwd martingale measures of different currencies

Let us consider a tradable asset A quoted in currency α and suppose to know its market-agreed forward value as seen from time t for maturity t_k. For example, A could be a stock with t_k any future maturity or a fwd-rate ibor index with t_k equal to its 'natural pay date' (typically its calendar adjusted end accrual date). By definition of fwd contract we can write

$$F_A^\alpha(t,t_k)P^\alpha(t,t_k) = E^\alpha\left[A(f_k)D^\alpha(t,t_k)|F_t\right] = P^\alpha(t,t_k)E^{k,\alpha}\left[A(f_k)|F_t\right] \tag{1.226}$$

where D^α denotes the stochastic risk neutral discount factor in α currency and E^α the expectation operator in the corresponding measure. Furthermore $f_k \geq t_k$ deones the fixing date of A associated to the forward contract delivering at t_k (typically $f_k = t_k$ for stocks or $f_k = t_{k-1}$ for ibor indexes). We can then write

$$F_A^\alpha(t,t_k) = E^{k,\alpha}\left[A(f_k)|F_t\right] \tag{1.227}$$

We now consider the problem of quanto-ing the forward contract in a β-currency and at the same time of changing its payment date from t_k to t_h, with $t_h \geq f_k$. We define X as the value of 1 unit of β currency expressed in α currency. The quanto fwd price will be

$$E^\beta\left[A(f_k)D^\beta(t,t_h)|F_t\right] = P^\beta(t,t_h)E^{h,\beta}\left[A(f_k)|F_t\right] = F_A^\beta(t,t_h)P^\beta(t,t_h) = \phi_A(t,t_h,\beta) \tag{1.228}$$

where the t_h-fwd quantoed in β-currency is then defined as

$$F_A^\beta(t,t_h) = E^{h,\beta}\left[A(f_k)|F_t\right] \tag{1.229}$$

By no arbitrage, it must also hold that

$$\phi_A(t,t_h,\beta) = E^\alpha\left[A(f_k)D^\alpha(t,t_h)X(t_h)|F_t\right]\frac{1}{X(t)} = \tag{1.230}$$

$$= E^\alpha\left[E^\alpha\left[A(f_k)D^\alpha(t,t_h)\frac{X(t_h)}{X(t)}|F_{f_k}\right]|F_t\right] = \tag{1.231}$$

$$= E^\alpha\left[D^\alpha(t,f_k)A(f_k)E^\alpha\left[D^\alpha(f_k,t_h)\frac{X(t_h)}{X(t)}|F_{f_k}\right]|F_t\right] \tag{1.232}$$

Again by no arbitrage it must hold that

$$E^\alpha\left[D^\alpha(f_k,t_h)\frac{X(t_h)}{X(f_k)}|F_{f_k}\right] = E^\beta\left[1\cdot D^\beta(f_k,t_h)|F_{f_k}\right] = P^\beta(f_k,t_h) \tag{1.233}$$

meaning that the contract delivering 1 unit of β currency at t_h musthave the same price as seen form f_k irrespective of the measure we use to compute the price. Then we get

$$\phi_A(t,t_h,\beta) = E^\alpha\left[D^\alpha(t,f_k)A(f_k)\frac{X(f_k)}{X(t)}P^\beta(f_k,t_h)|F_t\right] \tag{1.234}$$

Equation (1.228) to (1.234) we obtain

$$P^\beta(t,t_h)E^{h,\beta}\left[A(f_k)|F_t\right] = E^\alpha\left[D^\alpha(t,f_k)A(f_k)\frac{X(f_k)}{X(t)}P^\beta(f_k,t_h)|F_t\right] = \tag{1.235}$$

$$= E^\alpha\left[\frac{D^\alpha(t,t_k)}{P^\alpha(f_k,t_k)}A(f_k)\frac{X(f_k)}{X(t)}P^\beta(f_k,t_h)|F_t\right] = E^{k,\alpha}\left[\frac{P^\alpha(t,t_k)}{P^\alpha(f_k,t_k)}A(f_k)\frac{X(f_k)}{X(t)}P^\beta(f_k,t_h)|F_t\right] \tag{1.236}$$

where we also applied a payoff deferring formula and change from the risk neutral expectation E^{α} to the t_k-fwd measure expectation of the α currency. Summarizing, we proved that for any asset A it holds that

$$E^{h,\beta}\left[A(f_k)|F_t\right] = E^{k,\alpha}\left[A(f_k)\frac{L_X(f_k;\alpha,\beta,t_h,t_k)}{L_X(t;\alpha,\beta,t_h,t_k)}|F_t\right] \tag{1.237}$$

where

$$L_X(\tau;\alpha,\beta,t_h,t_k) = X(\tau)\frac{P^{\beta}(\tau,t_h)}{P^{\alpha}(\tau,t_k)} = \frac{dQ^{h,\beta}}{dQ^{k,\alpha}}(\tau) \tag{1.238}$$

is the Radon-Nikodyn derivative. We also write L in another illuminating form

$$L_X(\tau;\alpha,\beta,t_h,t_k) = \frac{dQ^{h,\beta}}{dQ^{k,\alpha}}(\tau) = X(\tau)\frac{P^{\beta}(\tau,t_h)}{P^{\alpha}(\tau,t_h)}\frac{P^{\alpha}(\tau,t_h)}{P^{\alpha}(\tau,t_k)} \tag{1.239}$$

We notice that $L_X(\tau;\alpha,\beta,t_h,t_k)$ is a martingale in the t_k-fwd measure of the α currency (being a $X(\tau)P^{\beta}(\tau,t_h)$ a tradable asset). We now define the process of the forward of the X fx rate for date t_h as

$$F_X(\tau,t_h) = X(\tau)\frac{P^{\beta}(\tau,t_h)}{P^{\alpha}(\tau,t_h)} \tag{1.240}$$

and finally obtain (see (1.210))

$$L_X(\tau;\alpha,\beta,t_h,t_k) = \frac{dQ^{h,\beta}}{dQ^{k,\alpha}}(\tau) = F_X(\tau,t_h)\cdot\frac{P^{\alpha}(\tau,t_h)}{P^{\alpha}(\tau,t_k)} = F_X(\tau,t_h)\cdot P^{\alpha}(\tau,h/k) \tag{1.241}$$

We now write (in the t_k-fwd measure of the α currency)

$$dF_X(\tau,t_h) = \cdots dt + \sigma_{F_X(t_h)}F_X(\tau,t_h)dW_X(\tau) \tag{1.242}$$

We now consider the following Ito differential (see (1.217)):

$$\frac{dL_X(\tau;\alpha,\beta,t_h,t_k)}{L_X(\tau;\alpha,\beta,t_h,t_k)} = \cdots dt + \sigma_{F_X(t_h)}dW_X(\tau) + \frac{1}{P^{\alpha}(\tau,h/k)}dP^{\alpha}(\tau,h/k) = \tag{1.243}$$

$$= \cdots dt + \sigma_{F_X(t_h)}dW_X(\tau) + s(h,k)\sum_{l=m_{h,k}+1}^{M_{h,k}}\frac{\tau_l dF_l^{\alpha}(\tau)}{1+F_l^{\alpha}(\tau)\tau_l} \tag{1.244}$$

We now use this last equation, together with the Girsanov's theorem (1.203) to get the following result. Suppose to consider a 1-dimensional Wiener process $Z^{(k,\alpha)}(\tau)$ where the (k,α)-apex is used to stress that it is a Wiener process in the t_k-fwd measure of the α currency. We can then conclude that $Z^{(h,\beta)}(\tau)$ as defined in the following is a Wiener process in the t_h-fwd measure of the β currency:

$$dZ^{(h,\beta)}(\tau) = dZ^{(k,\alpha)}(\tau) - s(h,k)\sum_{l=m_{h,k}+1}^{M_{h,k}}\frac{\tau_l\left\langle dF_l^{\alpha}(\tau)\cdot dZ^{(k,\alpha)}(\tau)\right\rangle}{1+F_l^{\alpha}(\tau)\tau_l} - \tag{1.245}$$

$$-\sigma_{F_X(t_h)}\left\langle dW_X(\tau)\cdot dZ^{(k,\alpha)}(\tau)\right\rangle \tag{1.246}$$

9 Differential Equations

Definition: Let us consider a filtered probability space $(\Omega, \mathcal{F}, \underline{\mathcal{F}}, P)$ and a vector process

$$X = (X_1, \cdots, X_n)^* \tag{1.247}$$

where $*$ stands for transpose and for each $i = 1, \ldots, n$ it holds that

$$dX_i(t) = \mu_i(t,x)dt + \sum_{r=1}^{d} \sigma_{ir}(t,x)dW_r(t) \tag{1.248}$$

where $W_1, \cdots, W_d$ are d independent Wiener processes. Consider also a $C^{1,2}$ mapping $f : \Re^+ \times \Re^n \to \Re$ defining the process

$$Z(t) = f(t, X(t)) \tag{1.249}$$

It holds that

$$df(t, X(t)) = \frac{\partial f}{\partial t} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dX_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_i} dX_i dX_j \tag{1.250}$$

with formally

$$dX_i dX_j = \sum_{r,s=1}^{d} \sigma_{ir}\sigma_{js}\delta_{rs}dt = (\sigma\sigma^*)_{ij}dt \equiv C_{ij}(t,x)dt \tag{1.251}$$

and so

$$df(t, X(t)) = \left[\frac{\partial f}{\partial t} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \mu_i(t,x) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_i} C_{ij}(t,x) \right] dt + \tag{1.252}$$

$$+ \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \sum_{r=1}^{d} \sigma_{ir}(t,x)dW_r(t) \tag{1.253}$$

In other terms

$$df(t, X(t)) = \left[\frac{\partial f}{\partial t} + \mathcal{A}f \right] dt + \nabla_x f \cdot \sigma \cdot dW \tag{1.254}$$

where $\mathcal{A}$ is the infinitesimal (or Dynkin or Ito or Kolmogorov backward) operator defined by

$$\mathcal{A}h(t,x) = \sum_{i=1}^{n} \mu_i(t,x) \frac{\partial h}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} C_{ij}(t,x) \frac{\partial^2 h}{\partial x_i \partial x_i} \tag{1.255}$$

Definition: A process g belongs to $\ell^2[a,b]$ if the following conditions are met

- $\int_a^b E[g^2(s)]ds < \infty$
- g is adapted to $\mathcal{F}_t^W$-filtration

Furthermore we say that the process g belongs to ℓ^2, if $\forall t > 0$ it holds that $g \in \ell^2[0,t]$.

Proposition: (Feynman-Kac) Suppose that the function $f : \Re^+ \times \Re^n \to \Re$ is such that

$$\frac{\partial f}{\partial t} + \mathcal{A}f - r_t f = 0 \tag{1.256}$$

$$f(T,x) = \Phi(x) \tag{1.257}$$

with

$$\frac{\partial f}{\partial x_i} \sigma_{ir}(t,x) \in \ell^2, \forall i, r \tag{1.258}$$

Then it holds that

$$f(t, X_t) = E_{t,X_t} \left[e^{-\int_t^T r_u du} \Phi(X_T) \right] \tag{1.259}$$

Proof. For any $t \leq s \leq T$, define

$$z(s, X_s) = e^{-\int_t^s r_u du} f(s, X_s)$$

and compute

$$dz(s, X_s) = e^{-\int_t^s r_u du} \left[-r_s f + \frac{\partial f}{\partial t} + \mathcal{A} f + \nabla_x f \cdot \sigma \cdot dW \right] \tag{1.260}$$

Using (1.256,)by integrating in $[t, T]$ we get

$$z(T, X_T) = z(t, X_t) + \int_t^T e^{-\int_t^s r_u du} \nabla_x f \cdot \sigma \cdot dW_s$$

Considering that

$$E_{t,X_t} \left[\int_t^T e^{-\int_t^s r_u du} \nabla_x f \cdot \sigma \cdot dW_s \right] = 0$$

by taking expectations of both sides and using (1.257) we get what needed:

$$E_{t,X_t} \left[e^{-\int_t^T r_u du} \phi(X_T) \right] = z(t, X_t) = f(t, X_t)$$

9.1 Kolmogorov Equations

Let us go back to (1.256) in the hypothesis $r = 0$ and for $\Phi(y) = I_B(y)$ (the indicator function of B). In other words, let us suppose $u(s, y)$ to be a solution of the following boundary value problem:

$$\frac{\partial u(s,y)}{\partial s} + \mathcal{A} u(s, y) = 0 \tag{1.261}$$

$$u(T, y) = I_B(y) \tag{1.262}$$

We know that the solution is

$$u(s, y) = E_{s,y} \left[I_B(X_T) \right] = P \left(X_T \in B | X_s = y \right) = P(s, y, T, B) \tag{1.263}$$

where $P(s, y, T, B)$ is the transition probability from (s, y) to (T, B).

Proposition: (**Kolmogorov backward equation**) It is possible to get a similar formula also for the transition density between (s, y) and any (t, x) as solution of the following boundary value problem:

$$\frac{\partial p(s,y;t,x)}{\partial s} + \mathcal{A} p(s, y; t, x) = 0 \tag{1.264}$$

$$p(t, y; t, x) = \delta(y - x) \tag{1.265}$$

with $(s, y) \in (0, t) \times \Re^n$. Notice that (1.264) is called backward since $\frac{\partial}{\partial s}$ and $\mathcal{A}$ are both acting on backward variables (s, y). We now move on deriving the corresponding forward equations where the differential operators act on the forward variables (t, x). Let us consider an infinite differentiable arbitrary test function $h(t, x)$ on the compact support $(s, T) \times \Re$ (so we are in the scalar case) and compute by the Ito formula:

$$h(T, X_T) = h(s, X_s) + \int_s^T \left[\frac{\partial h}{\partial t} + \mathcal{A} h \right] (t, X_t) dt + \int_s^T \frac{\partial h}{\partial t}(t, X_t) \cdot \sigma(t, X_t) \cdot dW_t \tag{1.266}$$

Let us now suppose that $h(T, x) = h(s, x) = 0 \; \forall x$, $h(t, \pm\infty) \to 0 \; \forall t$ and take expectations of both sides. Considering that the expectation of the dW_t integral is zero, we get

$$E_{s,y} \left[\int_s^T \left[\frac{\partial h}{\partial t} + \mathcal{A} h \right] (t, X_t) dt \right] = 0 \tag{1.267}$$

where $X_s \equiv y$. This last equation can be written also in terms of the transition density $p(s, y; t, x)$:

$$\int_{-\infty}^{\infty} \int_{s}^{T} p(s, y; t, x) \left[\frac{\partial h}{\partial t} + \mathcal{A}h \right] (t, x) dx dt = 0 \tag{1.268}$$

We now perform partial integration in $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial x}$:

$$\int_{-\infty}^{\infty} \int_{s}^{T} \left[-h(t, x) \frac{\partial p(s, y; t, x)}{\partial t} - h(t, x) \frac{\partial}{\partial x} \left(\mu(t, x) p(s, y; t, x) \right) \right] (t, x) dx dt - \tag{1.269}$$

$$- \int_{-\infty}^{\infty} \int_{s}^{T} \left[\frac{\partial}{\partial x} h(t, x) \frac{1}{2} \frac{\partial}{\partial x} \left(\sigma^2(t, x) p(s, y; t, x) \right) \right] (t, x) dx dt = 0 \tag{1.270}$$

Partial-integrating again in $\frac{\partial}{\partial x}$ we get

$$\int_{-\infty}^{\infty} \int_{s}^{T} h(t, x) \left[- \frac{\partial p(s, y; t, x)}{\partial t} - \frac{\partial}{\partial x} \left(\mu(t, x) p(s, y; t, x) \right) \right] (t, x) dx dt + \tag{1.271}$$

$$+ \int_{-\infty}^{\infty} \int_{s}^{T} \left[h(t, x) \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\sigma^2(t, x) p(s, y; t, x) \right) \right] (t, x) dx dt = 0 \tag{1.272}$$

which must be valid for any h satisfying the boundary condition described above. We finally got the **Kolmogorov forward equation** also known as **Fokker-Planck equation**:

$$\frac{\partial}{\partial t} p(s, y; t, x) = - \frac{\partial}{\partial x} \left(\mu(t, x) p(s, y; t, x) \right) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\sigma^2(t, x) p(s, y; t, x) \right) \tag{1.273}$$

$$p(s, y; t, x) \to \delta(y - x), t \to s^+ \tag{1.274}$$

In the multidimensional case, one can similarly get

$$\frac{\partial}{\partial t} p(s, y; t, x) = \mathcal{A}^* p(s, y; t, x) \tag{1.275}$$

where $\mathcal{A}^*$ is the adjoint Fokker-Planck forward operator:

$$\left(\mathcal{A}^* f \right) (t, x) = - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[\mu_i(t, x) f(t, x) \right] + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_i} \left[C_{ij}(t, x) f(t, x) \right] \tag{1.276}$$

Chapter 2

Models

1 Dupire Local Volatility

We follow [4]. Let us start consider an asset with the following dynamics in the risk neutral measure

$$dS_t = \mu_t S_t dt + \sigma(S_t, t) S_t dW_t \tag{2.1}$$

$$\mu_t = r_t - q_t \tag{2.2}$$

where q_t is the continuous dividend yield. We assume deterministic interest rates and write down the following notation for the zero coupon bond price of maturity $T > t$:

$$P(t, T) = e^{-\int_t^T r_s ds} \tag{2.3}$$

1.1 Derivation of the Dupire local volatility from the Fokker-Planck equation

The Fokker-Planck equation (1.275) for the particular case of (2.1) is

$$\frac{\partial}{\partial T} p(t, S_t; T, S_T) = -\frac{\partial}{\partial S_T} \left(\mu_T S_T p(t, S_t; T, S_T) \right) + \frac{1}{2} \frac{\partial^2}{\partial S_T^2} \left(\sigma^2(T, S_T) S_T^2 p(t, S_t; T, S_T) \right) \tag{2.4}$$

where one should remind that μ and σ are the ones of (2.1). The call price will be computed as

$$C(t, S_t; K) = P(t, T) \int_K^\infty (S_T - K) p(t, S_t; T, S_T) dS_T \tag{2.5}$$

We also compute call derivatives w.r.t. the strike variable:

$$\frac{\partial C(t, S_t; K)}{\partial K} = -P(t, T) \int_K^\infty p(t, S_t; T, S_T) dS_T \tag{2.6}$$

$$\frac{\partial^2 C(t, S_t; K)}{\partial^2 K} = P(t, T) p(t, S_t; T, K) \tag{2.7}$$

$$\frac{\partial C(t, S_t; K)}{\partial T} = -r_T P(t, T) \int_K^\infty (S_T - K) p(t, S_t; T, S_T) dS_T + \tag{2.8}$$

$$+ P(t, T) \int_K^\infty (S_T - K) \frac{\partial p(t, S_t; T, S_T)}{\partial T} dS_T = \tag{2.9}$$

$$= -r_T C(t, S_t; K) + P(t, T) \int_K^\infty (S_T - K) \frac{\partial p(t, S_t; T, S_T)}{\partial T} dS_T \tag{2.10}$$

where we used that

$$\frac{\partial P(t, T)}{\partial T} = -r_T P(t, T) \tag{2.11}$$

Using (2.4) in (2.10) we get

$$\frac{\partial C(t,S_t;K)}{\partial T}+r_T C(t,S_t;K)=P(t,T)\cdot \tag{2.12}$$

$$\cdot\int_K^\infty (S_T-K)\left\{-\frac{\partial}{\partial S_T}\left(\mu S_T p(t,S_t;T,S_T)\right)+\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\left(\sigma^2 S_T^2 p(t,S_t;T,S_T)\right)\right\}dS_T= \tag{2.13}$$

$$=\left(-I_1+\frac{1}{2}I_2\right)P(t,T) \tag{2.14}$$

with

$$I_1=\mu_T\int_K^\infty (S_T-K)\frac{\partial}{\partial S_T}\left(S_T p(t,S_t;T,S_T)\right)dS_T \tag{2.15}$$

$$I_2=\int_K^\infty (S_T-K)\frac{\partial^2}{\partial S_T^2}\left(\sigma(T,S_T)^2 S_T^2 p(t,S_t;T,S_T)\right)dS_T \tag{2.16}$$

1.1.1 Useful identities

The first identity we want to state come directly from (2.7):

$$p(t,S_t;T,K)=\frac{1}{P(t,T)}\frac{\partial^2 C(t,S_t;K)}{\partial^2 K} \tag{2.17}$$

The second identity starts from

$$\frac{C(t,S_t;K)}{P(t,T)}=\int_K^\infty S_T p(t,S_t;T,S_T)dS_T-K\int_K^\infty p(t,S_t;T,S_T)dS_T \tag{2.18}$$

which using (2.6) yields to

$$\int_K^\infty S_T p(t,S_t;T,S_T)dS_T=\frac{C(t,S_t;K)}{P(t,T)}-\frac{K}{P(t,T)}\frac{\partial C(t,S_t;K)}{\partial K} \tag{2.19}$$

1.1.2 Evaluate I_1

Integrating I_1 by parts:

$$I_1=\mu_T\int_K^\infty (S_T-K)\frac{\partial}{\partial S_T}\left(S_T p(t,S_t;T,S_T)\right)dS_T= \tag{2.20}$$

$$=\left[\mu_T(S_T-K)\left(S_T p(t,S_t;T,S_T)\right)\right]_K^\infty-\mu_T\int_K^\infty\left(S_T p(t,S_t;T,S_T)\right)dS_T \tag{2.21}$$

Assuming that $\left[\mu_T(S_T-K)\left(S_T p(t,S_t;T,S_T)\right)\right]_\infty=0$ and using (2.19) we get

$$I_1=-\mu_T\left[\frac{C(t,S_t;K)}{P(t,T)}-\frac{K}{P(t,T)}\frac{\partial C(t,S_t;K)}{\partial K}\right] \tag{2.22}$$

1.1.3 Evaluate I_2

Integrating I_2 by parts:

$$I_2=\int_K^\infty (S_T-K)\frac{\partial^2}{\partial S_T^2}\left(\sigma(T,S_T)^2 S_T^2 p(t,S_t;T,S_T)\right)dS_T= \tag{2.23}$$

$$=\left[(S_T-K)\frac{\partial}{\partial S_T}\left(\sigma(T,S_T)^2 S_T^2 p(t,S_t;T,S_T)\right)\right]_K^\infty- \tag{2.24}$$

$$-\int_K^\infty\frac{\partial}{\partial S_T}\left(\sigma(T,S_T)^2 S_T^2 p(t,S_t;T,S_T)\right)dS_T \tag{2.25}$$

Assuming $\left[(S_T - K)\frac{\partial}{\partial S_T}\left(\sigma(T,S_T)^2 S_T^2 p(t,S_t;T,S_T)\right)\right]_{\infty} = 0$, we get

$$I_2 = \sigma(T,K)^2 K^2 p(t,S_t;T,K) = \frac{\sigma(T,K)^2 K^2}{P(t,T)} \frac{\partial^2 C(t,S_t;K)}{\partial^2 K} \tag{2.26}$$

1.1.4 Dupire equation

Using (2.22) and (2.26) in (2.12):

$$\frac{\partial C(t,S_t;K)}{\partial T} + r_T C(t,S_t;K) = \left(-I_1 + \frac{1}{2} I_2\right) P(t,T) = \tag{2.27}$$

$$= \left\{\mu_T \left[\frac{C(t,S_t;K)}{P(t,T)} - \frac{K}{P(t,T)} \frac{\partial C(t,S_t;K)}{\partial K}\right] + \frac{\sigma(T,K)^2 K^2}{2P(t,T)} \frac{\partial^2 C(t,S_t;K)}{\partial^2 K}\right\} P(t,T) \tag{2.28}$$

from which

$$\sigma(T,K)^2 = \frac{\frac{\partial C(t,S_t;K)}{\partial T} + q_T C(t,S_t;K) + (r_T - q_T) K \frac{\partial C(t,S_t;K)}{\partial K}}{\frac{K^2}{2} \frac{\partial^2 C(t,S_t;K)}{\partial^2 K}} \tag{2.29}$$

where we also used (2.2).

1.2 Derivation of the Dupire local volatility as expected volatility

We write the following preliminaries:

$$\frac{\partial}{\partial S}(S-K)^+ = 1_{(S>K)}, \frac{\partial}{\partial K}(S-K)^+ = -1_{(S>K)} \tag{2.30}$$

$$\frac{\partial}{\partial S} 1_{(S>K)} = \delta(S-K), \frac{\partial}{\partial K} 1_{(S>K)} = -\delta(S-K) \tag{2.31}$$

$$\frac{\partial C}{\partial K} = -P(t,T) E[1_{(S>K)}], \frac{\partial^2 C}{\partial^2 K} = P(t,T) E[\delta(S-K)] \tag{2.32}$$

Let us now define the following function

$$f(S_T, T) = P(t,T)(S_T - K)^+ \tag{2.33}$$

where the process for S_t is always given by (2.1). By Ito's lemma we get

$$df = \left[\frac{\partial f}{\partial T} + \mu_T S_T \frac{\partial f}{\partial S_T} + \frac{1}{2}\sigma^2(T,S_T,\omega) S_T^2 \frac{\partial^2 f}{\partial^2 S_T}\right] dT + \left[\sigma(T,S_T,\omega) S_T \frac{\partial f}{\partial S_T}\right] dW_T(\omega) \tag{2.34}$$

where we outlined the dependence on the sample space variable ω in $\sigma(T,S_T,\omega)$ (and in $dW_T(\omega)$) in order to underline that, in the present discussion, the volatility σ is not restricted to be just a local volatility, i.e. a function of the form $\sigma(T,S_T)$, but can depend on any other general random factor through ω. Using (2.30), (2.31), (2.32) and (2.11) in (2.34), we get:

$$df = P(t,T)\left[-r_T(S_T-K)^+ + \mu_T S_T 1_{(S_T>K)} + \frac{1}{2}\sigma^2(T,S_T,\omega) S_T^2 \delta(S_T-K)\right] dT + \tag{2.35}$$

$$+ P(t,T)\left[\sigma(T,S_T,\omega) S_T 1_{(S_T>K)}\right] dW_T(\omega) \tag{2.36}$$

The following holds

$$-r_T(S_T-K)^+ + \mu_T S_T 1_{(S_T>K)} = \left[-r_T(S_T-K) + \mu_T S_T\right] 1_{(S_T>K)} = \tag{2.37}$$

$$= \left[r_T K - q_T S_T\right] 1_{(S_T>K)} \tag{2.38}$$

and hence

$$df(T,S_T) = P(t,T)\left[\left[r_T K - q_T S_T\right] 1_{(S_T>K)} + \frac{1}{2}\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] dT + \tag{2.39}$$

$$+P(t,T)\left[\sigma(T,S_T,\omega)S_T 1_{(S_T>K)}\right] dW_T(\omega) \tag{2.40}$$

Integrating this last equation between t and τ we get

$$\int_t^\tau df(T,S_T) = f(\tau,S_\tau) - f(t,S_t) = P(t,\tau)(S_\tau-K)^+ - P(t,t)(S_t-K)^+ = \tag{2.41}$$

$$= \int_t^\tau P(t,T)\left[\left[r_T K - q_T S_T\right] 1_{(S_T>K)} + \frac{1}{2}\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] dT + \tag{2.42}$$

$$+\int_t^\tau P(t,T)\left[\sigma(T,S_T,\omega)S_T 1_{(S_T>K)}\right] dW_T(\omega) \tag{2.43}$$

Taking expectation of both sides and considering that the diffusive term has vanishing expectation, we get

$$C(t,S_t;\tau,K) - C(t,S_t;t,K) = \tag{2.44}$$

$$= \int_t^\tau P(t,T)E\left[\left[r_T K - q_T S_T\right] 1_{(S_T>K)} + \frac{1}{2}\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] dT \tag{2.45}$$

By deriving both sides w.r.t. τ (and by calling $\tau = T$ just to keep the same notation as above), we get

$$\frac{\partial C(t,S_t;T,K)}{\partial T} = P(t,T)E\left[\left[r_T K - q_T S_T\right] 1_{(S_T>K)} + \frac{1}{2}\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] \tag{2.46}$$

We now use (2.32)

$$\frac{\partial C(t,S_t;T,K)}{\partial T} = -r_T K\frac{\partial C(t,S_t;T,K)}{\partial K} - q_T\left(C(t,S_t;T,K) - K\frac{\partial C(t,S_t;T,K)}{\partial K}\right) + \tag{2.47}$$

$$+\frac{1}{2}P(t,T)E\left[\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] \tag{2.48}$$

Using (1.123) and (1.134) with:

- $S_T(\omega)$ acting as $X(\omega)$,
- $\sigma^2(T,S_T,\omega)$ acting as $Y(\omega)$,
- $\phi(X) = X^2$,

we get:

$$E\left[\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] = E\left[\sigma^2(T,S_T,\omega)S_T^2\delta(S_T-K)\right] = \tag{2.49}$$

$$= K^2E\left[\sigma^2(T,S_T,\omega)|S_T=K\right]p(t,S_t;T,K) = \tag{2.50}$$

$$= K^2E\left[\sigma^2(T,S_T,\omega)|S_T=K\right]\frac{1}{P(t,T)}\frac{\partial^2 C(t,S_t;K)}{\partial^2 K} \tag{2.51}$$

where we also used (2.17) in the last passage. Using (2.49) into (2.47), we finally arrive at

$$E\left[\sigma^2(T,S_T,\omega)|S_T=K\right] = \frac{\frac{\partial C(t,S_t;K)}{\partial T} + q_T C(t,S_t;K) + (r_T-q_T)K\frac{\partial C(t,S_t;K)}{\partial K}}{\frac{K^2}{2}\frac{\partial^2 C(t,S_t;K)}{\partial^2 K}} \tag{2.52}$$

Comparing (2.29) and (2.52) and recognizing that the right hand sides are equivalent, we then observe that the Dupire local volatility at $(T, S_T = K)$ can be interpreted as the expectation of the asset volatility, conditional on the asset hitting K at T.

1.3 Local Volatility in Terms of Implied Volatility

Let us define the standard Black-Scholes formula for a call option with strike x, forward f, log-asset cumulated standard deviation w, interest rates r_s, time to maturity τ:

$$BS(f, x, \tau, w, \{r\}) = e^{-\int_0^\tau r_s ds} \left[f \cdot \mathcal{N}\left(\frac{\log\left(\frac{f}{x}\right)}{w} + \frac{w}{2} \right) - x \cdot \mathcal{N}\left(\frac{\log\left(\frac{f}{x}\right)}{w} - \frac{w}{2} \right) \right] \tag{2.53}$$

Considering the function $C(t, S_t; K)$ appearing in both (2.29) and (2.52), we then define the following parametrization of the call price in terms of the market *implied* Black-Scholes volatility (we assume $t = 0$ without loss of generality):

$$C(0, S_0; K) \equiv BS\left(S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, K, T, \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right) \tag{2.54}$$

where we performed the following substitutions:

$$F_T = S_0 \cdot e^{\int_0^T (r_s - q_s) ds} \to f \tag{2.55}$$
$$K \to x \tag{2.56}$$
$$\sigma_i(K, T)\sqrt{T} \to w \tag{2.57}$$
$$T \to \tau \tag{2.58}$$

Notice that we introduce the implied volatility surface $\sigma_i(K, T)$. It is now possible to use (2.54) in (2.29). Let us for example examine the term of (2.29) involving $\frac{\partial C(t, S_t; K)}{\partial K}$. Using (2.54) one has to make the following computation:

$$\frac{\partial C(t, S_t; K)}{\partial K} = \frac{d}{dK} BS\left(S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, K, T, \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right) =$$
$$= \frac{\partial}{\partial x} BS\left(f = S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, x = K, \tau = T, w = \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right) +$$
$$+ \left[\frac{\partial}{\partial w} BS\left(f = S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, x = K, \tau = T, w = \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right)\right] \cdot \frac{\partial\left[\sigma_i(K, T) \cdot \sqrt{T}\right]}{\partial K}$$

where we used the total derivative symbol in order to underline that the K derivative must act also on $\sigma_i(K, T)$. Similarly

$$\frac{\partial C(t, S_t; K)}{\partial T} = \frac{d}{dT} BS\left(S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, K, T, \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right) =$$
$$= \frac{\partial}{\partial \tau} BS\left(f = S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, x = K, \tau = T, w = \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right) +$$
$$+ \left[\frac{\partial}{\partial w} BS\left(f = S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, x = K, \tau = T, w = \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right)\right] \cdot \frac{\partial\left[\sigma_i(K, T) \cdot \sqrt{T}\right]}{\partial T} +$$
$$+ \left[\frac{\partial}{\partial f} BS\left(f = S_0 \cdot e^{\int_0^T (r_s - q_s) ds}, x = K, \tau = T, w = \sigma_i(K, T) \cdot \sqrt{T}, \{r\}\right)\right] \cdot \frac{\partial\left[S_0 \cdot e^{\int_0^T (r_s - q_s) ds}\right]}{\partial T}$$

Using Mathematica software we have then verified that the local volatility can also be expressed in terms of the Black-Scholes implied volatility as follows

$$\sigma(T, K)^2 = \frac{\frac{\partial C(t, S_t; K)}{\partial T} + q_T C(t, S_t; K) + (r_T - q_T) K \frac{\partial C(t, S_t; K)}{\partial K}}{\frac{K^2}{2} \frac{\partial^2 C(t, S_t; K)}{\partial^2 K}} = \tag{2.59}$$

$$= \frac{\sigma_i^2(K, T) + 2\sigma_i(K, T) T \left[\frac{\partial \sigma_i(K, T)}{\partial T} + (r_T - q_T) K \frac{\partial \sigma_i(K, T)}{\partial K}\right]}{\left(1 - \frac{K y(K, T)}{\sigma_i(K, T)} \frac{\partial \sigma_i(K, T)}{\partial K}\right)^2 + K \sigma_i(K, T) T \left[\frac{\partial \sigma_i(K, T)}{\partial K} - \frac{K \sigma_i(K, T) T}{4} \left(\frac{\partial \sigma_i(K, T)}{\partial K}\right)^2 + K \frac{\partial^2 \sigma_i(K, T)}{\partial^2 K}\right]} \tag{2.60}$$

where we defined the log-moneyness

$$y(K,T) = \log\left[\frac{K}{S_0 e^{\int_0^T (r_s - q_s)ds}}\right] = \log\left[\frac{K}{F_T}\right] \tag{2.61}$$

Eq. (2.60) is coherent with eq. (2.25) of [5], but there's a different sign in the y term in denominator with respet to [4]. Another way to write the local volatility in terms of the implied one is the following[1], that can be found also in eq. (22.20) of [7] and in eq. (2.12) of [6]:

$$\sigma(T,K)^2 = \frac{\frac{\sigma_i(K,T)}{T} + 2\frac{\partial\sigma_i(K,T)}{\partial T} + 2(r_T - q_T)K\frac{\partial\sigma_i(K,T)}{\partial K}}{K^2\left[\frac{\partial^2\sigma_i(K,T)}{\partial^2 K} - d_1\sqrt{T}\left(\frac{\partial\sigma_i(K,T)}{\partial K}\right)^2 + \frac{1}{\sigma_i(K,T)}\left(\frac{1}{K\sqrt{T}} + d_1\frac{\partial\sigma_i(K,T)}{\partial K}\right)^2\right]} \tag{2.62}$$

with

$$d_{1,2} = d_\pm = \frac{\log\left[\frac{S_0}{K}\right] + \int_0^T (r_s - q_s)ds}{\sigma_i(K,T)\sqrt{T}} \pm \sigma_i(K,T)\sqrt{T} \tag{2.63}$$

If we choose

$$\sigma_i(K,T) = \sigma_0\left(\frac{K}{F_T}\right)^\beta \tag{2.64}$$

we can compute

$$\sigma(T,K)^2 = \frac{\sigma_0^2\left(\frac{K}{F_T}\right)^{2\beta}}{1 - \frac{\beta^2}{4}\left(\frac{K}{F_T}\right)^{2\beta}\sigma_0^2 T\left(\sigma_0^2 T\left(\frac{K}{F_T}\right)^{2\beta} - 4\right) + 2\beta\log\frac{F_T}{K} + \beta^2\left(\log\frac{F_T}{K}\right)^2} \tag{2.65}$$

We hence see that

$$\sigma(T,F_T)^2 = \frac{\sigma_0^2}{1 - \frac{\beta^2}{4}\sigma_0^2 T\left(\sigma_0^2 T - 4\right)} \tag{2.66}$$

that shows that the at the money forward local volatility is of the order of the at the money forward implied volatility (the denominator is of order 1).

2 CSA Derivative Pricing

The following section is taken from [3]. Let $S(t)$ be an asset that, in the real worlds measure, follows the following dynamics:

$$\frac{dS(t)}{S(t)} = \mu_S(t)dt + \sigma_S(t)dW(t) \tag{2.67}$$

and let $V(t,S)$ be a derivative security on S. By Ito's lemma we get

$$dV(t,S) = \left(\mathcal{L}V(t,S)\right)dt + \Delta(t,S)dS(t) \tag{2.68}$$

where

$$\mathcal{L} = \frac{\partial}{\partial t} + \frac{\sigma_S^2 S^2}{2}\frac{\partial^2}{\partial^2 S} \tag{2.69}$$

$$\Delta(t,S) = \frac{\partial V(t,S)}{\partial S} \tag{2.70}$$

[1] Using the Mathematica software, we checked that (2.60 and (2.62) are coherent.

Let us now write

$$V(t) = V(t) - C(t) + C(t) \tag{2.71}$$

where $C(t)$ is the collateral (cash) amount held against $V(t)$[2]. In order to replicate the derivative we hold $\Delta(t)$ of stock and $\gamma(t)$ unit of cash, in such a way that

$$V(t) = \Delta(t)S(t) + \gamma(t) \tag{2.72}$$

We hence can write

$$\gamma(t) = V(t) - \Delta(t)S(t) = [C(t)] + [V(t) - C(t)] - [\Delta(t)S(t)] \tag{2.73}$$

This means that

$$d\gamma(t) = r_c(t)C(t)dt + r_f(t)\left[V(t) - C(t)\right]dt - r_r(t)\left[\Delta(t)S(t)\right]dt + r_d(t)\left[\Delta(t)S(t)\right]dt \tag{2.74}$$

where we introduced

- r_c, the collateral rate (the risk free rate), i.e. the rate of funding when securing the loan with $C(t)$ (i.e. with cash)
- r_f, the rate at which the derivative desk, that is pricing V, can lend and borrow money (we assume equal rates for lending and borrowing are available). This is the rate taking into account the credit spread for the bank the desk belongs to, i.e. the rate of its treasury function.
- r_r, the repo rate for $S(t)$, i.e. the rate of funding when securing the loan with $S(t)$
- r_d, the dividend rate paid by the stock.

By imposing the self-financing condition, it must hold that

$$dV(t) - \Delta(t)dS(t) = d\gamma(t) \tag{2.75}$$

that yields

$$\left(\frac{\partial}{\partial t} + \frac{\sigma_S^2 S^2}{2}\frac{\partial^2}{\partial^2 S}\right)V(t,S) = \tag{2.76}$$

$$= r_c(t)C(t) + r_f(t)\left[V(t) - C(t)\right] + \left(r_d(t) - r_r(t)\right)\left[S(t)\frac{\partial V(t,S)}{\partial S}\right] \tag{2.77}$$

that is

$$\left(\frac{\partial}{\partial t} + \left(r_r(t) - r_d(t)\right)S\frac{\partial}{\partial S} + \frac{\sigma_S^2 S^2}{2}\frac{\partial^2}{\partial^2 S}\right)V(t,S) = r_f(t)V(t,S) - \left(r_f(t) - r_c(t)\right)C(t) \tag{2.78}$$

Let us know define

$$\tilde{V}(t) = e^{\int_t^T r_f(u)du}V(t) \tag{2.79}$$

which, in the measure in which the S-drift is

$$\frac{dS(t)}{S(t)} = \left(r_r(t) - r_d(t)\right)dt + \sigma_S(t)dW(t) \tag{2.80}$$

[2] For example, V(t) could be a negative amount, meaning that a derivative desk owe some money (option sold) to bank clients and for this reason the desk keeps a cash account to cover and guarantee its debt, eliminating the risk of not getting paid for the clients (credit risk).

yields to

$$d\tilde{V}(t,S) = e^{\int_t^T r_f(u)du}\left(\frac{\partial}{\partial t} + (r_r(t) - r_d(t))\, S\frac{\partial}{\partial S} + \frac{\sigma_S^2 S^2}{2}\frac{\partial^2}{\partial^2 S} - r_f(t)\right) V(t,S)dt + \tag{2.81}$$

$$+ (r_r(t) - r_d(t))\, S\frac{\partial \tilde{V}(t,S)}{\partial S} dW(t) \tag{2.82}$$

$$= -e^{\int_t^T r_f(u)du}\left(r_f(t) - r_c(t)\right) C(t)dt + (r_r(t) - r_d(t))\, S\frac{\partial \tilde{V}(t,S)}{\partial S} dW(t) \tag{2.83}$$

Let us now integrate both sides between t and T

$$\int_t^T d\tilde{V}(t,S) = \tilde{V}(T,S_T) - \tilde{V}(t,S_t) = V(T,S_T) - e^{\int_t^T r_f(u)du} V(t,S_t) = \tag{2.84}$$

$$= -\int_t^T e^{\int_\nu^T r_f(u)du}\left(r_f(\nu) - r_c(\nu)\right) C(\nu)\nu + \int_t^T (r_r(\nu) - r_d(\nu))\, S_\nu \frac{\partial \tilde{V}(\nu,S_\nu)}{\partial S_\nu} dW(\nu) \tag{2.85}$$

which becomes

$$V(t,S_t) = e^{-\int_t^T r_f(u)du} V(T,S_T) + \int_t^T e^{-\int_t^\nu r_f(u)du}\left(r_f(\nu) - r_c(\nu)\right) C(\nu)\nu - \tag{2.86}$$

$$-e^{-\int_t^T r_f(u)du}\int_t^T (r_r(\nu) - r_d(\nu))\, S_\nu \frac{\partial \tilde{V}(\nu,S_\nu)}{\partial S_\nu} dW(\nu) \tag{2.87}$$

By taking expectations of both sides, and using that the expectation of (2.87) vanishes, we get

$$V(t,S_t) = E_{t,S_t}\left[e^{-\int_t^T r_f(u)du} V(T,S_T) + \int_t^T e^{-\int_t^\nu r_f(u)du}\left(r_f(\nu) - r_c(\nu)\right) C(\nu)d\nu\right] \tag{2.88}$$

always valid in the measure such that (2.80) holds. With a completely similar computation, but starting from $\bar{V}(t) = e^{\int_t^T r_c(u)du} V(t)$ instead of $\tilde{V}(t) = e^{\int_t^T r_f(u)du} V(t)$ and using that

$$\left(\frac{\partial}{\partial t} + (r_r(t) - r_d(t))\, S\frac{\partial}{\partial S} + \frac{\sigma_S^2 S^2}{2}\frac{\partial^2}{\partial^2 S} - r_c(t)\right) V(t,S) = -\left(r_f(t) - r_c(t)\right)\left(C(t) - V(t)\right)$$

it is possible to arrive at

$$V(t,S_t) = E_{t,S_t}\left[e^{-\int_t^T r_c(u)du} V(T,S_T) - \int_t^T e^{-\int_t^\nu r_c(u)du}\left(r_f(\nu) - r_c(\nu)\right)\left(V(\nu) - C(\nu)\right)d\nu\right] \tag{2.89}$$

From (2.78) we can observe that

$$E_{t,S_t}\left[dV(t)\right] = \left[r_f(t)V(t,S) - \left(r_f(t) - r_c(t)\right)C(t)\right]dt = \left[r_f(t)V(t,S) - s_f(t)C(t)\right]dt \tag{2.90}$$

with $s_f(t) = (r_f(t) - r_c(t))$ equal to the spread between the bank funding rate $r_f(t)$ and the collateral funding rate $r_c(t)$.

2.1 Zero strike call option under no CSA

Assuming no dividends for simplicity, we can find:

$$V_{zsc}(t) = E_{t,S_t}\left[e^{-\int_t^T r_f(u)du} S(T)\right] \tag{2.91}$$

$$S_t = E_{t,S_t}\left[e^{-\int_t^T r_r(u)du} S(T)\right] \tag{2.92}$$

showing that the price of the zero strike call is different from the stock price S_t. The reason for this is that the call is subject to the credit risk of the bank, while the stock can be used as collateral in a transaction of type

- At time t, A lends B be 1 unit of currency
- At time t, B lends $1/S_t$ stocks to A
- At time T, B pays A an amount of $e^{+\int_t^T r_r(u)du}$ units of currency
- At time T, A returns $1/S_t$ stocks to B

In other words, the zero strike call is not a collateral equivalent (in theory as good as) to the stock due to the bank credit spread.

2.2 Forward without CSA

By definition, the forward price seen by the bank at time t for expiry T is the value that solves the following equation

$$0 = E_{t,S_t}\left[e^{-\int_t^T r_f(u)du}\left(S(T) - F_{noCSA}(t,T)\right)\right] \tag{2.93}$$

We then have

$$F_{noCSA}(t,T) = \frac{E_{t,S_t}\left[e^{-\int_t^T r_f(u)du}S(T)\right]}{E_{t,S_t}\left[e^{-\int_t^T r_f(u)du}\right]} \tag{2.94}$$

We then define the risky zero coupon issued by the bank as

$$P_f(t,T) = E_t\left[e^{-\int_t^T r_f(u)du}\right] = E_t\left[\frac{1}{B_f(t,T)}\right] \tag{2.95}$$

where

$$B_f(t,T) = e^{\int_t^T r_f(u)du} \tag{2.96}$$

2.3 Forward with CSA

Similarly the condition is

$$0 = E_{t,S_t}\left[e^{-\int_t^T r_c(u)du}\left(S(T) - F_{CSA}(t,T)\right)\right] \tag{2.97}$$

which yields

$$F_{CSA}(t,T) = \frac{E_{t,S_t}\left[e^{-\int_t^T r_c(u)du}S(T)\right]}{E_{t,S_t}\left[e^{-\int_t^T r_c(u)du}\right]} \tag{2.98}$$

2.4 Convexity adjustment

$$F_{noCSA}(t,T) = \frac{E_{t,S_t}\left[e^{-\int_t^T r_f(u)du}S(T)\right]}{P_f(t,T)} = \frac{E_{t,S_t}\left[e^{-\int_t^T r_c(u)du}e^{-\int_t^T s_f(u)du}S(T)\right]}{P_f(t,T)} = \tag{2.99}$$

$$= P_c(t,T)\frac{E^T_{t,S_t}\left[e^{-\int_t^T s_f(u)du}S(T)\right]}{P_f(t,T)} \tag{2.100}$$

where we changed measure from the one having the risk free bank account as numeraire[3] to the one having the risk free zero coupon bond $P_c(t,T) = E_t\left[e^{-\int_t^T r_c(u)du}\right]$ as numeraire. Now we define

$$M(t,T) = \frac{P_f(t,T)}{P_c(t,T)} e^{-\int_0^t s_f(u)du} \tag{2.101}$$

$$M(T,T) = e^{-\int_0^T s_f(u)du} \tag{2.102}$$

and compute

$$F_{noCSA}(t,T) = E^T_{t,S_t}\left[\frac{M(T,T)}{M(t,T)} S(T)\right] \tag{2.103}$$

It holds that

$$M(t,T) = \frac{P_f(t,T)}{P_c(t,T)} e^{-\int_0^t s_f(u)du} = E_t\left[e^{-\int_t^T r_f(u)du}\right] \frac{e^{-\int_0^t s_f(u)du}}{P_c(t,T)} = \tag{2.104}$$

$$= \frac{E_t\left[e^{-\int_0^t s_f(u)du} e^{-\int_t^T r_f(u)du}\right]}{P_c(t,T)} = \frac{E_t\left[e^{-\int_0^t s_f(u)du} e^{-\int_t^T (r_f(u)-r_c(u))du} e^{-\int_t^T r_c(u)du}\right]}{P_c(t,T)} = \tag{2.105}$$

$$= \frac{E_t\left[e^{-\int_0^T s_f(u)du} e^{-\int_t^T r_c(u)du}\right]}{P_c(t,T)} = \frac{P_c(t,T) E_t^T\left[e^{-\int_0^T s_f(u)du}\right]}{P_c(t,T)} = E_t^T\left[e^{-\int_0^T s_f(u)du}\right] = \tag{2.106}$$

$$= E_t^T\left[M(T,T)\right] \tag{2.107}$$

which clearly shows that $M(t,T)$ is a martingale in the risk free T forward measure. Furthermore it is at this point trivial that

$$E_t^T\left[\frac{M(T,T)}{M(t,T)}\right] = 1 \tag{2.108}$$

We hence compute the convexity adjustment as

$$F_{noCSA}(t,T) - F_{CSA}(t,T) = E_t^T\left[\left(\frac{M(T,T)}{M(t,T)}\right) S(T) - S(T)\right] = \tag{2.109}$$

$$= E_t^T\left[\left(\frac{M(T,T)}{M(t,T)}\right) S(T) - E_t^T\left[\frac{M(T,T)}{M(t,T)}\right] S(T)\right] = \tag{2.110}$$

$$= E_t^T\left[\left(\frac{M(T,T)}{M(t,T)} - E_t^T\left[\frac{M(T,T)}{M(t,T)}\right]\right) (S(T) - F_{CSA}(t,T))\right] = \tag{2.111}$$

$$= \frac{1}{M(t,T)} \mathrm{Cov}_t^T\left[M(T,T), F_{CSA}(T,T)\right] \tag{2.112}$$

2.5 Vanilla options

We have (the put case is similar):

$$\Pi_{noCSA}(t) = E_{t,S_t}\left[e^{-\int_t^T r_f(u)du} (S(T) - K)^+\right] \tag{2.113}$$

$$\Pi_{CSA}(t) = E_{t,S_t}\left[e^{-\int_t^T r_c(u)du} (S(T) - K)^+\right] \tag{2.114}$$

which can also be written in the non risky T forward measure:

$$\Pi_{noCSA}(t) = P_c(t,T) E^T_{t,S_t}\left[e^{-\int_t^T s_f(u)du} (S(T) - K)^+\right] \tag{2.115}$$

$$\Pi_{CSA}(t) = P_c(t,T) E^T_{t,S_t}\left[(S(T) - K)^+\right] \tag{2.116}$$

[3] always the same measure in which (2.80) holds.

Using (2.202) we can recast $\Pi_{noCSA}(t)$ as follows

$$\Pi_{noCSA}(t) = P_f(t,T) E_{t,S_t}^T \left[\frac{M(T,T)}{M(t,T)} (S(T)-K)^+ \right] = \tag{2.117}$$
$$= P_f(t,T) E_{t,S_t}^T \left[E_{t,S_T}^T \left[\frac{M(T,T)}{M(t,T)} (S(T)-K)^+ \right] \right] = \tag{2.118}$$
$$= P_f(t,T) E_{t,S_t}^T \left[(S(T)-K)^+ E_{t,S_T}^T \left[\frac{M(T,T)}{M(t,T)} \right] \right] \tag{2.119}$$

where it should be noticed the difference between the small t and capital T in E_{t,S_T}^T. Defining

$$\alpha(t,T,x) = E_t^T \left[\frac{M(T,T)}{M(t,T)} | S_T = x \right] \tag{2.120}$$

we get

$$\Pi_{noCSA}(t) = P_f(t,T) E_{t,S_t}^T \left[(S(T)-K)^+ \alpha(t,T,S_T) \right] \tag{2.121}$$

Now we approximate (this is an hypothesis):

$$\alpha(t,T,x) = \alpha_0(t,T) + \alpha_1(t,T)x \tag{2.122}$$

and notice that

$$E_{t,S_t}^T [\alpha(t,T,S_T)] = E_{t,S_t}^T \left[E_t^T \left[\frac{M(T,T)}{M(t,T)} | S_T \right] \right] = E_{t,S_t}^T \left[\frac{M(T,T)}{M(t,T)} \right] = 1 = \tag{2.123}$$
$$= \alpha_0(t,T) + \alpha_1(t,T) E_{t,S_t}^T [S_T] = \alpha_0(t,T) + \alpha_1(t,T) F_{CSA}(t,T) \tag{2.124}$$

where we used

$$F_{CSA}(t,T) = E_{t,S_t}^T [S_T] \tag{2.125}$$

Now we go on calibrating the two parameters $\alpha_0(t,T)$, $\alpha_1(t,T)$ (we already found one equation, we need another one). We saw above (see (2.103)) that

$$F_{noCSA}(t,T) = E_{t,S_t}^T \left[\frac{M(T,T)}{M(t,T)} S(T) \right] = E_{t,S_t}^T \left[\frac{M(T,T)}{M(t,T)} S(T) \right] = \tag{2.126}$$
$$= E_{t,S_t}^T \left[E_{t,S_T}^T \left[\frac{M(T,T)}{M(t,T)} S(T) \right] \right] = E_{t,S_t}^T [\alpha(t,T,S_T) S_T] = \tag{2.127}$$
$$= E_{t,S_t}^T [(\alpha_0(t,T) + \alpha_1(t,T) S_T) S_T] = \tag{2.128}$$
$$= \alpha_0(t,T) F_{CSA}(t,T) + \alpha_1(t,T) \left(\mathrm{Var}_t^T (S_T) + F_{CSA}^2(t,T) \right) \tag{2.129}$$

Using (2.124) and (2.129) we get:

$$\alpha_0(t,T) = 1 - \alpha_1(t,T) F_{CSA}(t,T) \tag{2.130}$$
$$\alpha_1(t,T) = \frac{F_{noCSA}(t,T) - F_{CSA}(t,T)}{\mathrm{Var}_t^T (S_T)} \tag{2.131}$$

Let us now go back to (2.121):

$$\Pi_{noCSA}(t) = P_f(t,T) \int_K^\infty dS_T f_t^T(S_T) (S_T - K) \alpha(t,T,S_T) = \tag{2.132}$$
$$= P_f(t,T) \int_K^\infty dS_T f_t^{\tilde{T}}(S_T) (S_T - K) \tag{2.133}$$

where $f_t^T(S_T)$ is the density in the non risky T forward measure, while $f_t^{\tilde{T}}(S_T)$ is the density in the measure such that

$$\Pi_{noCSA}(t) = P_f(t,T) E_{t,S_t}^{\tilde{T}} \left[(S(T) - K)^+ \right] \tag{2.134}$$

Differentiating twice in K we get

$$f_t^{\tilde{T}}(S_T) = f_t^T(S_T) \alpha(t,T,S_T) \tag{2.135}$$

which shows that the probability density under no CSA, that is $f_t^{\tilde{T}}(S_T)$, is obtained by multiplying the probability density under CSA, by a linear function of the stock level, which amounts in a slope distortion of the volatility smile.

2.6 Numeric Example

Suppose the stock follows a log-normal process with volatility σ_s

$$dS(t) = \cdots dt + \sigma_s S(t) dW_s(t) \tag{2.136}$$

and let the funding spread of the bank be governed by

$$ds_f(t) = -a_f(\theta - s_f(t))dt + \sigma_f dW_f(t) \tag{2.137}$$

Suppose also that $r_c(t), r_r(t)$ are deterministic and that $r_d(t) = 0$. Under the T forward measure $F_{CSA}(t,T)$ is a martingale and we can write

$$F_{CSA}(t,T) = F_{CSA}(0,T) e^{-\frac{1}{2}\int_0^t \sigma_s^2(u)du + \int_0^t \sigma_s(u)dW_s(u)} \tag{2.138}$$

It also holds that

$$\frac{M(t,T)}{M(0,T)} = e^{-\int_0^t s_f(u)du} \frac{P_c(0,T)}{P_f(0,T)} \frac{P_f(t,T)}{P_c(t,T)} \tag{2.139}$$

We then observe that (as usual $b_x(t) = (1 - e^{-xt})/x$):

$$\int_0^t s_f(u)du = \cdots + \int_0^t \sigma_f b_a(t-u) dW_f(u) \tag{2.140}$$

or better, reminding that $M(t,T)$ must be a martingale (see before), we get

$$\frac{M(T,T)}{M(0,T)} = \exp\left[-\frac{1}{2}\int_0^T \sigma_f^2 b_a^2(T-u)du - \int_0^T \sigma_f b_a(T-u) dW_f(u) \right] \tag{2.141}$$

We now remind (2.103)

$$F_{noCSA}(0,T) = E_{0,S_0}^T \left[\frac{M(T,T)}{M(0,T)} S(T) \right] = E_{0,S_0}^T \left[\frac{M(T,T)}{M(0,T)} F_{CSA}(T,T) \right] = \tag{2.142}$$

$$\tag{2.143}$$

and using both (2.141) and (2.138) we finally arrive at

$$F_{noCSA}(0,T) = F_{CSA}(0,T) \exp\left[-\int_0^T \sigma_f \sigma_s b_a(T-u) \rho_{s,f} du \right] \tag{2.144}$$

where $\rho_{s,f}$ is the instantaneous correlation between $dW_f(u)$ and $dW_s(u)$. Equation (2.144) gives a convexity adjustment estimation and it in particular shows that the adjustment grows for increasing T, as expected.

3 Options on defaultable bonds

3.1 Lognormal credit model

Let $t_0 = 0$ be calculation date and consider the following time schedule

$$0 = t_0 < t_1 < t_2 < \cdots < t_N \tag{2.145}$$

Let us define the probability to default after time t_i to be Q_i, i.e.

$$Q_i = Prob\,(\tau \geq t_i) \tag{2.146}$$

where τ is the default stopping time. It holds that

$$1 = Q_0 \geq Q_1 \geq Q_2 \geq \cdots \geq Q_N \tag{2.147}$$

where we supposed that default hasn't happened up to time t_0 included. We introduce the following further notation $\forall i = 1, 2, \cdots, N$:

$$Q_i(t_0) = Q_{i-1}(t_0)\frac{1}{1 + \delta_i q_i(t_0, t_{i-1}, t_i)} \leq Q_{i-1}(t_0) \tag{2.148}$$

$$\delta_i = t_i - t_{i-1} \tag{2.149}$$

where we outlined the fact that the probabilities are observed at time t_0 and we introduced the positive quantity $q_i(t_0, t_{i-1}, t_i)$. Equation (2.148) tries to resemble the way one usually writes interest rate zero coupon bond values in a libor market model framework, with the following identifications:

$$Q_i(t) \leftrightarrow P_i(t, t_i) \tag{2.150}$$

$$q_i(t, t_{i-1}, t_i) \leftrightarrow F_i(t, t_{i-1}, t_i), t \leq t_{i-1} \tag{2.151}$$

Let us now notice that if all $q_i(t, t_{i-1}, t_i)$ are taken to be constant, i.e. if

$$q_i(t, t_{i-1}, t_i) = q_i(t_0, t_{i-1}, t_i), t_0 \leq t \leq t_{i-1}$$

then the model reduces to a deterministic default intensity model, where the integral of the default intensity between any two times (t_{i-1}, t_i) is specified by $(Q_{i-1}(t_0), Q_i(t_0))$, or equivalently by $(Q_{i-1}(t_0), q_i(t_0, t_{i-1}, t_i))$. Notice that if the condition $q_i(t, t_{i-1}, t_i) \geq 0$ is satisfied, the default intensity is by construction non negative. Let us now introduce some processes $\phi_i(t)$, $i = 1, \cdots, d$, such that

$$d\ln\phi_i(t) = -\frac{1}{2}\sigma_i^2(t)dt + \sigma_i(t)dW_i(t) \tag{2.152}$$

for some deterministic non negative functions $\sigma_i(t), t \geq 0$. We then write

$$q_i(t, t_{i-1}, t_i) = q_i(t_0, t_{i-1}, t_i) \cdot \psi_i(t) \cdot \prod_{k=1}^{d} \phi_k(t)^{\gamma_{i,k}} \tag{2.153}$$

In words, $q_i(t, t_{i-1}, t_i)$ is proportional to its value at present date $q_i(t_0, t_{i-1}, t_i)$ through a log normal stochastic factor $\prod_{k=1}^{d} \phi_k(t)^{\gamma_{i,k}}$ and a deterministic normalization function $\psi_i(t)$. Let us now imagine to perform a Monte Carlo evolution from time t_0 to time t_1. As seen from t_0, the unconditional probability to default between any two times (t_{i-1}, t_i), $i > 0$, is given by

$$D_i(t_0) \equiv Q_{i-1}(t_0) - Q_i(t_0) \geq 0 \tag{2.154}$$

We suppose that the strip of probabilities $D_1(t_0), D_2(t_0), \cdots, D_N(t_0)$ is calibrated to the market[4] and hence the Monte Carlo evolution we are about to consider should mandatory produce the exact

[4]Notice that t_N could formally be ∞ and $Q_N(t_0) = 0$.

strip of $D_i(t_0), i \geq 0$ in order to be calibrated to the market. Simulating from t_0 to t_1 there is a probability $D_1(t_0)$ to default in (t_0, t_1) and we assume that as seen from t_1 and conditional to not having defaulted in (t_0, t_1), the new strip of survival probabilities will be given by

$$1 = Q_1(t_1) \geq Q_2(t_1) \geq \cdots \geq Q_N(t_1) \tag{2.155}$$

with

$$Q_i(t_1) = Q_{i-1}(t_1)\frac{1}{1+\delta_i q_i(t_1, t_{i-1}, t_i)} \leq Q_{i-1}(t_1), i \geq 2 \tag{2.156}$$

$$\tag{2.157}$$

Assuming independence between the stochastic event that decides whether the default in (t_0, t_1) happens or not - given the probability $D_1(t_0)$ - and the stochastic process factor that drives $q_i(t_0, t_{i-1}, t_i)$ to $q_i(t_1, t_{i-1}, t_i)$, in order to fit the market, we should impose that $D_1(t_0), D_2(t_0), \cdots, D_N(t_0)$ will be correcly reproduced. This implies that

$$Q_1(t_0) \cdot E\left[D_i(t_1)|\mathcal{F}_0\right] = D_i(t_0), i \geq 2 \tag{2.158}$$

that yields to

$$E\left[1 - \frac{1}{1+\delta_2 q_2(t_1, t_1, t_2)}|\mathcal{F}_0\right] = \frac{D_2(t_0)}{Q_1(t_0)} \tag{2.159}$$

$$E\left[\frac{1}{1+\delta_2 q_2(t_1, t_1, t_2)}\left(1 - \frac{1}{1+\delta_3 q_3(t_1, t_2, t_3)}\right)|\mathcal{F}_0\right] = \frac{D_3(t_0)}{Q_1(t_0)} \tag{2.160}$$

$$\cdots \tag{2.161}$$

that, using a numerical integration, should allow to calibrate the parameters of (2.152), (2.153)

$$\psi_2(t_1), \sigma_2(t_1) \tag{2.162}$$

$$\psi_3(t_1), \sigma_3(t_1) \tag{2.163}$$

$$\cdots \tag{2.164}$$

in a bootstrap-like procedure. Notice that going on with the calibration of time intervals, i.e. considering the further simulation step (t_1, t_2), the analogous of (2.158) will be given by

$$E\left[D_i(t_2)1_{\tau>t_2}|\mathcal{F}_0\right] = D_i(t_0), i \geq 3 \tag{2.165}$$

and there is dependence between $D_i(t_2)$ and $1_{\tau>t_2}$ which, even if the default probability of the model is calibrated up to t_2, implies that

$$E\left[D_i(t_2)1_{\tau>t_2}|\mathcal{F}_0\right] \neq Q_2(t_0) \cdot E\left[D_i(t_2)|\mathcal{F}_0\right] \tag{2.166}$$

This difficulty implies that the calibration of interval (t_1, t_2) and subsequent ones will be possible only through direct computations (for example: Monte Carlo calibration).

3.2 Stochastic credit spread model

Let $t_0 = 0$ be calculation date and consider the following time schedule

$$0 = t_0 < t_1 < t_2 < \cdots < t_N \tag{2.167}$$

Consider the fixed rate bond paying the cashflow c_i at time t_i, $i > 0$, whose value at present date is[5]

$$B(t_0) = E\left[\sum_{i=1}^{N}\left(c_i 1_{\tau>t_i} + Rc_i 1_{\tau\leq t_i}\right)D(t_0, t_i)|\mathcal{G}_0\right] = B_{mkt}(t_0) \tag{2.168}$$

[5] $\mathcal{G}_t = \mathcal{F}_t \vee \sigma\left(\{\tau \leq u\}, u \leq t\right)$ is the usual market filtration completed with the observation of default events.

where R is the recovery rate, $D(t_0, t_i)$ is the stochastic discount factor and we also introduced the price $B_{mkt}(t_0)$ quoted by the market for the bond at time t_0. We can now introduce the following additional way to write the price of the bond

$$B(t_0) = 1_{\tau > t_0} \sum_{i=1}^{N} c_i P(t_0, t_i) e^{-s(t_0)\cdot(t_i - t_0)} + 1_{\tau \leq t_0} \sum_{i=1}^{N} R(t_0) c_i P(t_0, t_i) = B_{mkt}(t_0) \tag{2.169}$$

which is just a way to implicitly define an effective credit spread w.r.t. the risk free rates, that has to be chosen with the only prescription to match $B_{mkt}(t_0)$. In words, if the issuer is already defaulted in t_0 the bond is worth the (estimated) recovery rate times the risk free price of the remaining cash flows, otherwise the bond values is the discounted value of the future cash flows using discount factors that have a spread w.r.t. the risk free rates calibrated in order to take into account that the issuer can default after t_0 and also possible liquidity issues. Notice that we added a dependence on the time t_0 in the recovery rate $R(t_0)$ in order to outline that it can in principle be stochastic. Let us now compute the price as seen from t_0 of a vanilla call option[6] written on the bond with maturity t_m, $m \leq N$, pay date t_p:

$$\Pi\left(t_0, t_m, K\right) = E\left[\left(B(t_m) - K\right)^+ D(t_0, t_p) | \mathcal{G}_0\right] = \tag{2.170}$$

$$= P(t_0, t_p) E^p \left[\left(B(t_m) - K\right)^+ | \mathcal{G}_0\right] = P(t_0, t_p) \cdot \tag{2.171}$$

$$\cdot E^p \left[\left(1_{\tau > t_m} \sum_{i \geq m}^{N} c_i P(t_m, t_i) e^{-s(t_m)\cdot(t_i - t_m)} + 1_{\tau \leq t_m} \sum_{i \geq m}^{N} R(t_m) c_i P(t_m, t_i) - K\right)^+ | \mathcal{G}_0\right] = \tag{2.172}$$

$$= P(t_0, t_p) E^p \left[1_{\tau > t_m} \left(\sum_{i \geq m}^{N} c_i P(t_m, t_i) e^{-s(t_m)\cdot(t_i - t_m)} - K\right)^+ | \mathcal{G}_0\right] + \tag{2.173}$$

$$+ P(t_0, t_p) E^p \left[1_{\tau \leq t_m} \left(\sum_{i \geq m}^{N} R(t_m) c_i P(t_m, t_i) - K\right)^+ | \mathcal{G}_0\right] \tag{2.174}$$

Assuming that the stopping time τ is independent from all other random variables we get

$$\Pi\left(t_0, t_m, K\right) = \tag{2.175}$$

$$= P(t_0, t_p) Q\left(\tau > t_m | \mathcal{G}_0\right) E^p \left[\left(\sum_{i \geq m}^{N} c_i P(t_m, t_i) e^{-s(t_m)\cdot(t_i - t_m)} - K\right)^+ | \mathcal{G}_0\right] + \tag{2.176}$$

$$+ P(t_0, t_p) \left[1 - Q\left(\tau > t_m | \mathcal{G}_0\right)\right] E^p \left[\left(\sum_{i \geq m}^{N} R(t_m) c_i P(t_m, t_i) - K\right)^+ | \mathcal{G}_0\right] \tag{2.177}$$

where $Q\left(\tau > t_m | \mathcal{G}_0\right)$ is the probability of the default event occurring after t_m as observed from t_0 knowing $\mathcal{G}_0$. At this point we assume a normal distribution for $s(t_m)$

$$s(t_m) \sim \mathcal{N}\left[\mu_s(t_m), \Sigma_s^2\left(t_m, t_n - t_m, s_m^*(K)\right) | \mathcal{F}_0\right] \tag{2.178}$$

where Σ_s^2 is the variance that is supposed to depend on the maturity of the option, the remaining time to maturity of the underlying bond $(t_n - t_m)$ and the par discounting spread s_m^* that would satisfy the following equation:

$$\sum_{i > m}^{N} c_i \bar{P}(t_m, t_i) e^{-s_m^*(K)\cdot(t_i - t_m)} = K \tag{2.179}$$

[6] put would be analogous.

where $\bar{P}$ is the expected forward zero coupon bond value for delivering 1 unit of cash in t_i as seen from t_0 and assuming deterministic interest rates for simplicity. In words, (2.179) finds the credit discounting spread that would make the bond worth as the strike K assuming default is not possible. Notice that when $t_m \to t_n$, it could happen that $s^*_m(K)$ diverges, but in this case the actual strike $s^*_m(K)$ to extract Σ^2_s is no more important due to the fact that the variance of the stochastic term $P(t_m, t_i)e^{-s(t_m)\cdot(t_i - t_m)}$ in (2.176) approaches in any case zero in reason of the vanishing time to maturity. The expect value $\mu_s(t_m)$ appearing in (2.178) will be implied by the following equation that ensures the calibration to market price of the bond as seen from t_0[7]

$$\Pi\left(t_0, t_m, K = 0\right) = B_{mkt}(t_0) - \sum_{t_i < t_m} \Pi_{mkt}\left(c_i\right) \tag{2.180}$$

computed with the condition

$$\Sigma^2_s = \Sigma^2_s\left(t_m, t_n - t_m, s^*_m(K)\right) \tag{2.181}$$

and assuming that the distribution of $R(t_m)$ is known to be a log normal with exogenously given mean and volatility parameters[8]:

$$\ln R(t_m) = \mathcal{N}\left[\mu_R(t_m), \Sigma^2_R\left(t_m, t_n - t_m, K\right)|\mathcal{F}_0\right] \tag{2.182}$$

It is now worth noticing from (2.175) that for the case $t_p = t_m = t_N$, $K = 0$, $C_N = 1$, one has

$$\Pi\left(t_0, t_N, K = 0\right) = P(t_0, t_N)Q\left(\tau > t_N|\mathcal{G}_0\right) + P(t_0, t_N)\left[1 - Q\left(\tau > t_N|\mathcal{G}_0\right)\right]E^p\left[R(t_N)|\mathcal{G}_0\right] \tag{2.183}$$

This last equation correctly ensures that

$$\Pi\left(t_0, t_N, K = 0\right) = B_{mkt}(t_0) - \sum_{t_i < t_N} \Pi_{mkt}\left(c_i\right) \tag{2.184}$$

by construction, if one assumes that $Q\left(\tau > t_N|\mathcal{G}_0\right)$ has been correctly calibrated to the market price of the bond, taking into account the exogenously given mean of the recovery rate $E^p\left[R(t_m)|\mathcal{G}_0\right]$.

3.3 Interest rate model and detailed computation

We assume that interest rates are stochastic, driven by a gaussian short rate model. The basic dynamics of the short rate is given by

$$r\left(t\right) = \sum_i x_i\left(t\right) + \varphi\left(t\right) \tag{2.185}$$

with $\varphi\left(t\right)$ deterministic, $x_i\left(0\right) = 0$ and with

$$dx_i\left(t\right) = -a_i\, x_i\left(t\right) dt + \eta_i\left(t\right) dW_i\left(t\right) \tag{2.186}$$

We consider a_i as constants, $\eta_i\left(t\right)$ as deterministic functions of time and we set the correlation of the Brownian motions to be

$$dW_i\left(t\right) dW_j\left(t'\right) = \rho_{ij}\left(t\right)\, \delta\left(t - t'\right)\, dt\, dt' \tag{2.187}$$

again with $\rho_{ij}\left(t\right)$ deterministic and with $\rho_{ii}\left(t\right) = 1$. For $t_0 < t < s$ one then has the following dynamics for the short rate component x_i in the T-forward measure

$$x_i\left(s\right) = x_i\left(t\right)\, b_i\left(t, s\right) + \int_t^s \eta_i\left(u\right)\, b_i\left(u, s\right)\, dW_i\left(u\right) + \mu^T_i\left(t, s\right) \tag{2.188}$$

[7] $\sum_{t_i < t_m} \Pi_{mkt}\left(c_i\right)$ in (2.180) is the sum of the market present values of the cashflows of the bond occurring before the maturity date of the option t_m.

[8] The recovery random variable is assumed to be independent from all other stochastic quantities.

with

$$b_i(t,T) = e^{-a_i(T-t)} \tag{2.189}$$

$$B_i(t,T) = \frac{1}{a_i}\left(1-e^{-a_i(T-t)}\right) \tag{2.190}$$

and

$$\mu_i^T(t,s) = \sum_j \mu_{ij}^T(t,s) \tag{2.191}$$

$$\mu_{ij}^T(t,s) = -\int_t^s b_i(u,s)\, B_j(u,T)\, \eta_{ij}(u)\, du \tag{2.192}$$

The zero coupon bond valueas seen from $\mathcal{F}_T$ is

$$P(T,S) = \frac{P(0,S)}{P(0,T)} e^{-M(T,S)+\frac{1}{2}[V(T,S)-V(0,S)+V(0,T)]} \tag{2.193}$$

where

$$M(t,T) = \sum_i x_i(t)\, B_i(t,T) \tag{2.194}$$

and for the V terms expressions refer to standard computations (see ....). If we now go back to (2.175) that we rewrite here for clarity

$$\Pi(t_0, t_m, K) = \tag{2.195}$$

$$= P(t_0,t_p)Q(\tau > t_m|\mathcal{G}_0)\, E^p\left[\left(\sum_{i\geq m}^{N} c_i P(t_m,t_i) e^{-s(t_m)\cdot(t_i-t_m)} - K\right)^+ |\mathcal{G}_0\right] + \tag{2.196}$$

$$+P(t_0,t_p)\left[1-Q(\tau > t_m|\mathcal{G}_0)\right] E^p\left[\left(\sum_{i\geq m}^{N} R(t_m) c_i P(t_m,t_i) - K\right)^+ |\mathcal{G}_0\right] \tag{2.197}$$

we realize that the calculation of the two expectations reduce to multivariate Gaussian integrations[9] w.r.t. either the following vector of correlated Gaussian variables

$$(s(t_m), x_i(t_m)) \tag{2.198}$$

or

$$(\ln R(t_m), x_i(t_m)) \tag{2.199}$$

This ends the description of the model and its implementation.

4 Stochastic volatility's orderly smiles

The following is taken from [8]. Let us consider a stock S_t, $x_t = \ln S_t$ and the dynamics

$$dx_t = \left(-\frac{1}{2}\xi_t^t + m_t\right)dt + \sqrt{\xi_t^t} dZ_t^1, x_0 = x \tag{2.200}$$

$$d\xi_t^u = \lambda(t,u,\xi_t)\cdot dZ_t, \xi_0^u = \xi^u \tag{2.201}$$

[9] to be performed numerically

where $\xi_t \equiv (\xi_t^u, u \geq t)$ is the instantaneous forward variance curve as seen from time t for time u, dZ_t is a d-dimensional standard Brownian motion with orthogonal components, m_t is the drift component:

$$m_t = r_t - d_t \tag{2.202}$$

with r_t the (deterministic) short interest rate, d_t the continuous dividend yield. Furthermore we choose

$$\lambda(t, u, \xi_t) = (\lambda_1, \lambda_2, \cdots, \lambda_d) \tag{2.203}$$

with λ_1 driving the covariance between the spot and the variance processes. One could infer the instantaneous forward variance curve at time 0 from the variance swap market[10] as

$$\xi_0^u = \frac{d}{du}\left(\hat{\sigma}_u^2 u\right) \tag{2.204}$$

where $\hat{\sigma}_u$ is the variance swap implied volatility for maturity u. Befor going on, we outline the following:

- the model has no local volatility component
- the model is a second generation stochastic volatility model, meaning that it models the evolution of the all curve $(\xi_t^u, u \geq t)$ and not only ξ_t^t (like for example heston models does)

We will proceed with a (second order) perturbative expansion for small ϵ, making the following substitution in the dynamics:

$$\lambda \to \epsilon\lambda \tag{2.205}$$

Let us consider any derivative (undiscounted) price within the model specified by (2.200):

$$\Pi\left(t, x, \xi^u\right) \tag{2.206}$$

By Ito formula, it holds that

$$d\Pi\left(t, x, \xi^u\right) = \Pi_t dt + \Pi_x\left(-\frac{1}{2}\xi_t^t dt + m_t dt + \sqrt{\xi_t^t} dZ_t^1\right) + \frac{1}{2}\Pi_{xx}\left(\xi_t^t dt\right) + \tag{2.207}$$

$$+\int_t^T \left(\Pi_u \lambda(t, u, \xi_t)\cdot dZ_t\right) du + \int_t^T\left(\int_t^T\left(\frac{\Pi_{uu'}}{2}\sum_{i=1}^d \lambda_i(t, u, \xi_t)\lambda_i(t, u', \xi_t)dt\right) du\right) du' + \tag{2.208}$$

$$+\int_t^T\left(\Pi_{xu}\lambda_1(t, u, \xi_t)\sqrt{\xi_t^t}dt\right) du \tag{2.209}$$

where we used the short notation

$$\frac{\partial}{\partial z}\Pi\left(t, x, \xi^u\right) = \Pi_z \text{ for } z = t, x \tag{2.210}$$

and

$$\frac{\partial}{\partial \xi_t^u}\Pi\left(t, x, \xi^u\right) = \Pi_u, u \geq t \tag{2.211}$$

We now define

$$\mu\left(t, u, \xi_t\right)\lambda_1(t, u, \xi_t)\sqrt{\xi_t^t} \tag{2.212}$$

$$\nu\left(t, u, u', \xi_t\right) = \sum_{i=1}^d \lambda_i(t, u, \xi_t)\lambda_i(t, u', \xi_t) \tag{2.213}$$

[10]or from ATM vanilla options at time 0, or using other similar sensible choices.

and get

$$d\Pi\left(t,x,\xi^{u}\right)=\Pi_{t}dt+\Pi_{x}\left(-\frac{1}{2}\xi_{t}^{t}dt+m_{t}dt+\sqrt{\xi_{t}^{t}}dZ_{t}^{1}\right)+\frac{1}{2}\Pi_{xx}\left(\xi_{t}^{t}dt\right)+ \tag{2.214}$$

$$+dZ_{t}\cdot\int_{t}^{T}\left(\Pi_{u}\lambda(t,u,\xi_{t})\right)du+dt\int_{t}^{T}\left(\int_{t}^{T}\left(\frac{\Pi_{uu'}}{2}\nu\left(t,u,u',\xi_{t}\right)\right)du\right)du'+ \tag{2.215}$$

$$+dt\int_{t}^{T}\left(\Pi_{xu}\mu\left(t,u,\xi_{t}\right)\right)du \tag{2.216}$$

We now integrate and take expectations of both sides between t and any τ such that $t\le\tau\le T$:

$$E_{t}\left[\int_{t}^{\tau}d\Pi\left(t,x,\xi^{u}\right)\right]=E_{t}\left[\Pi\left(\tau,x_{\tau},\xi_{\tau}^{u}\right)-\Pi\left(t,x,\xi^{u}\right)\right]=E_{t}\left[(\cdots)_{1}dt\right]+E_{t}\left[(\cdots)_{2}\cdot dZ_{t}\right] \tag{2.217}$$

where we used dots to shorty group the dt terms and the dZ_t terms, that one obtains when integrating the right hand side of (2.214). We now know that $E_t\left[(\cdots)_2\cdot dZ_t\right]=0$[11] and observe that, being $\Pi\left(t,x,\xi^u\right)$ the undiscounted derivative price, it must be a martingale. This brings us to the pricing equation $(\cdots)_1(s)=0$ for all $s\ge t$[12], that explicitly amounts to

$$\Pi_{t}+\Pi_{x}\left(-\frac{1}{2}\xi_{t}^{t}+m_{t}\right)+\frac{1}{2}\xi_{t}^{t}\Pi_{xx}+ \tag{2.218}$$

$$+\int_{t}^{T}\left(\int_{t}^{T}\left(\frac{\Pi_{uu'}}{2}\nu\left(t,u,u',\xi_{t}\right)\right)du\right)du'+\int_{t}^{T}\left(\Pi_{xu}\mu\left(t,u,\xi_{t}\right)\right)du=0 \tag{2.219}$$

$$\Pi\left(T,x_{T},\xi_{T}^{u}\right)=g(x_{T}) \tag{2.220}$$

where $g(x_T)$ is the non path dependent derivative payoff at maturity time T. Without loss of generality, we now assume $m_t=0$[13]. We can define

$$H_{t}^{0}=\frac{\xi_{t}^{t}}{2}\left(\frac{\partial^{2}}{\partial x^{2}}-\frac{\partial}{\partial x}\right) \tag{2.221}$$

$$W_{t}^{1}=\int_{t}^{T}du\mu\left(t,u,\xi_{t}\right)\frac{\partial^{2}}{\partial x\partial\xi_{t}^{u}} \tag{2.222}$$

$$W_{t}^{2}=\frac{1}{2}\int_{t}^{T}du\int_{t}^{T}du'\nu\left(t,u,u',\xi_{t}\right)du\frac{\partial^{2}}{\partial\xi_{t}^{u'}\partial\xi_{t}^{u}} \tag{2.223}$$

$$H_{t}=H_{t}^{0}+W_{t}^{1}+W_{t}^{2} \tag{2.224}$$

that permits to rewrite (2.218) as

$$\left(\frac{\partial}{\partial t}+H_{t}\right)\Pi\left(t,x,\xi^{u}\right)=0 \tag{2.225}$$

$$\Pi\left(T,x_{T},\xi_{T}^{u}\right)=g(x_{T}) \tag{2.226}$$

[11] notice that $(\cdots)_2$ is adapted to dZ_t

[12] remind that τ is generic and hence the martingale condition implies that $(\cdots)_1(s)=0$ for all $s\ge t$

[13] if this is not the case, one should work out all the subsequent calculations with a shifted x_t without drift and plug m_t in just in the final results.

Bibliography

[1] Antoon Pelsser, 'Efficient Methods for Valuing Interest Rate Derivatives', Springer Finance

[2] Bjork, 'Arbitrage Theory in Continuos Time', 2nd ed, Oxford

[3] Vladimir Piterbarg, 'Funding beyond discounting: collateral agreements and derivatives pricing, Risk Magazine'

[4] 'Derivation of Local Volatility', by Fabrice Douglas Rouah (seems not published, search the internet..)

[5] 'LOCAL VOLATILITY MODELLING',Roel van der Kamp, July 13, 2009, 'A DISSERTATION SUBMITTED FOR THE DEGREE OF Master of Science in Applied Mathematics (Financial Engineering)'

[6] 'Pricing Barrier Options under Local Volatility', Artur Sepp, Mail: artursepp@hotmail.com, Web: www.hot.ee/seppar, 16 November 2002

[7] 'Foreign Exchange Risk: Models, Instruments and Stategies', Risk Books, 2002 (ed. Hakala, Wystup)

[8] 'Stochastic volatility's orderly smiles', Lorenzo Bergomi, Julien Guyon, Risk Magazine, May 2012, page 60