Lecture 3

Continuous Time Finance

The Martingale Approach

I: Mathematics

(Ch 10-12)

Tomas Björk

a purely theoretical lecture

Introduction

which you probably know already

In order to understand and to apply the martingale approach to derivative pricing and hedging we will need to some basic concepts and results from measure theory. These will be introduced below in an informal manner - for full details see the textbook.

Many propositions below will be proved but we will also present a couple of central results without proofs, and these must then be considered as dogmatic truths. You are of course not expected to know the proofs of such results (this is outside the scope of this course) but you are supposed to be able to **use** the results in an operational manner.

Contents

1.

Events and sigma-algebras

Events and sigma-algebras

(book: Section A.2)

Consider a probability measure P on a sample space Ω. An **event** is simply a subset $A \subseteq \Omega$ and $P(A)$ is the probability that the event A occurs.

For technical reasons, a probability measure can only be defined for a certain "nice" class $\mathcal{F}$ of events, so for $A \in \mathcal{F}$ we are allowed to write $P(A)$ as the probability for the event A.

For technical reasons the class $\mathcal{F}$ must be a **sigma-algebra**, which means that $\mathcal{F}$ is closed under the usual set theoretic operations like complements, countable intersections and countable unions. !

Interpretation: We can view a σ-algebra $\mathcal{F}$ as formalizing the idea of information. More precisely: A σ-algebra $\mathcal{F}$ is a collection of events, and if we assume that we have access to the information contained in $\mathcal{F}$, this means that for every $A \in \mathcal{F}$ we know exactly if A has occured or not.

Borel sets

Definition: The **Borel algebra** $\mathcal{B}$ is the smallest sigma-algebra on R which contains all intervals. A set B in $\mathcal{B}$ is called a **Borel set**.

Remark: There is no constructive definition of $\mathcal{B}$, but almost all subsets of R that you will ever see will in fact be Borel sets, so the reader can without danger think about a Borel set as "an arbitrary subset of R".

alternatively, contains all open sets, or
contains all closed sets

Random variables

Section B.1

An $\mathcal{F}$-**measurable random variable** X is a a mapping

$$X : \Omega \to R$$

such that $\{X \in B\} = \{\omega \in \Omega : \ X(\omega) \in B\}$ belongs to $\mathcal{F}$ for all Borel sets B. This guarantees that we are allowed to write $P(X \in B)$. Instad of writing "X is $\mathcal{F}$-measurable" we will often write $X \in \mathcal{F}$.

This means that if $X \in \mathcal{F}$ then the value of X is completely determined by the information contained in $\mathcal{F}$.

If we have another σ-algebra $\mathcal{G}$ with $\mathcal{G} \subseteq \mathcal{F}$ then we interpret this as "$\mathcal{G}$ contains less information than $\mathcal{F}$".

2.

Conditional Expectation (Section B.5)

Formal definition:

Let $\mathcal{G} \subset \mathcal{F}$, $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$

Thm: "$\exists!$" $\hat{X} \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$ s.t. !

$$E\hat{X}1_G = EX1_G, \quad \forall G \in \mathcal{G}$$

uniqueness: another $\hat{X}'$ satisfies
$P(\hat{X} = \hat{X}') = 1$.

Such an $\hat{X}$ is a version of the conditional expectation of X given $\mathcal{G}$.

Conditional Expectation

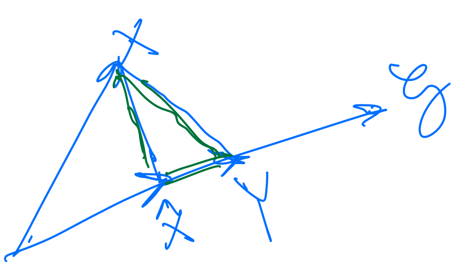
If $X \in \mathcal{F}$ and if $\mathcal{G} \subseteq \mathcal{F}$ then we write $E[X|\mathcal{G}]$ for the conditional expectation of X given the information contained in $\mathcal{G}$. Sometimes we use the notation $E_{\mathcal{G}}[X]$.

The following proposition contains everything that we will need to know about conditional expectations within this course. Also:

If EX^2, then for all $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ one has

$$E(X-Y)^2 = E(X-\hat{X})^2 + E(\hat{X}-Y)^2$$

[Think of Pythagoras and ordinary projections,



]

Main Results

Proposition 1: Assume that $X \in \mathcal{F}$, and that $\mathcal{G} \subseteq \mathcal{F}$. Then the following hold.

- The random variable $E[X|\mathcal{G}]$ is completely determined by the information in $\mathcal{G}$ so we have

$$E[X|\mathcal{G}] \in \mathcal{G}$$

(by definition)

- If we have $Y \in \mathcal{G}$ then Y is completely determined by $\mathcal{G}$ so we have

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

In particular we have

$$E[Y|\mathcal{G}] = Y$$

- If $\mathcal{H} \subseteq \mathcal{G}$ then we have the "law of iterated expectations"

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

(analogy with iterated projections)

- In particular we have

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

3.

Changing Measures

Changing Measures

Section B.6

Consider a probability measure P on $(\Omega, \mathcal{F})$, and assume that $L \in \mathcal{F}$ is a random variable with the properties that

$$L \geq 0$$

and

$$E^P[L] = 1.$$

For every event $A \in \mathcal{F}$ we now define the real number $Q(A)$ by the prescription

$$Q(A) = E^P[L \cdot I_A]$$

where the random variable I_A is the indicator for A, i.e.

$$I_A = \begin{cases} 1 & \text{if} \quad A \text{ occurs} \\ 0 & \text{if} \quad A^c \text{ occurs} \end{cases}$$

Recall that

$$Q(A) = E^P[L \cdot I_A]$$

We now see that $Q(A) \geq 0$ for all A, and that

$$Q(\Omega) = E^P[L \cdot I_\Omega] = E^P[L \cdot 1] = 1$$

We also see that if $A \cap B = \emptyset$ then

$$\begin{aligned} Q(A \cup B) &= E^P[L \cdot I_{A \cup B}] = E^P[L \cdot (I_A + I_B)] \\ &= E^P[L \cdot I_A] + E^P[L \cdot I_B] \\ &= Q(A) + Q(B) \end{aligned}$$

(extends to finite disjoint unions)

Furthermore we see that

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0$$

We have thus more or less proved the following

Proposition 2: If $L \in \mathcal{F}$ is a nonnegative random variable with $E^P[L] = 1$ and Q is defined by

$$Q(A) = E^P[L \cdot I_A]$$

then Q will be a probability measure on $\mathcal{F}$ with the property that

for which you need countable additivity (true!)

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0.$$

I turns out that the property above is a very important one, so we give it a name.

Absolute Continuity

Definition: Given two probability measures P and Q on $\mathcal{F}$ we say that Q is **absolutely continuous** w.r.t. P on $\mathcal{F}$ if, for all $A \in \mathcal{F}$, we have

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0$$

We write this as

$$Q << P.$$

If $Q << P$ and $P << Q$ then we say that P and Q are **equivalent** and write

$$Q \sim P$$

(this does NOT mean $Q = P$!)

Equivalent measures

It is easy to see that P and Q are equivalent if and only if

$$P(A) = 0 \quad \Leftrightarrow \quad Q(A) = 0$$

or, equivalently,

$$P(A) = 1 \quad \Leftrightarrow \quad Q(A) = 1$$

(look at complements)

Two equivalent measures thus agree on all certain events and on all impossible events, but can disagree on all other events.

Simple examples:

- All non degenerate Gaussian distributions on R are equivalent.

- If P is Gaussian on R and Q is exponential then $Q << P$ but not the other way around. (why?)

(?) end of lecture 3a

Absolute Continuity ct'd

We have seen that if we are given P and **define** Q by

$$Q(A) = E^P[L \cdot I_A] \quad (*)$$

for $L \geq 0$ with $E^P[L] = 1$, then Q is a probability measure and $Q << P$. .

A natural question is now if **all** measures $Q << P$ are obtained in this way. The answer is yes, and the precise (quite deep) result is as follows. The proof is difficult and therefore omitted.

that is, by
formula $(*)$
for some L.

The Radon Nikodym Theorem

Consider two probability measures P and Q on $(\Omega, \mathcal{F})$, and assume that $Q << P$ on $\mathcal{F}$. Then there exists a unique random variable L with the following properties

L is $\mathcal{F}$-measurable !

1. $Q(A) = E^P[L \cdot I_A], \quad \forall A \in \mathcal{F}$

2. $L \geq 0, \quad P - a.s.$

3. $E^P[L] = 1,$

4. $L \in \mathcal{F}$

The random variable L is denoted as

$$L = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}$$

and it is called the **Radon-Nikodym derivative** of Q w.r.t. P on $\mathcal{F}$, or the **likelihood ratio** between Q and P on $\mathcal{F}$.

A simple example

The Radon-Nikodym derivative L is intuitively the local scale factor between P and Q. If the sample space Ω is finite so $\Omega = \{\omega_1, \dots, \omega_n\}$ then P is determined by the probabilities $p_1, \dots, p_n$ where

$$p_i = P(\omega_i) \quad i = 1, \dots, n$$

Now consider a measure Q with probabilities

$$q_i = Q(\omega_i) \quad i = 1, \dots, n$$

If $Q << P$ this simply says that

$$p_i = 0 \quad \Rightarrow \quad q_i = 0$$

and it is easy to see that the Radon-Nikodym derivative $L = dQ/dP$ is given by

$$L(\omega_i) = \frac{q_i}{p_i} \quad i = 1, \dots, n \qquad (\text{if } p_i > 0),$$

If $p_i = 0$ then we also have $q_i = 0$ and we can define the ratio q_i/p_i arbitrarily.

If $p_1, \ldots, p_n$ as well as $q_1, \ldots, q_n$ are all positive, then we see that $Q \sim P$ and in fact

$$\frac{dP}{dQ} = \frac{1}{L} = \left(\frac{dQ}{dP}\right)^{-1}$$

as could be expected.

Computing expected values

A main use of Radon-Nikodym derivatives is for the computation of expected values.

Suppose therefore that $Q << P$ on $\mathcal{F}$ and that X is a random variable with $X \in \mathcal{F}$. With $L = dQ/dP$ on $\mathcal{F}$ then have the following result.

Proposition 3: With notation as above we have

$$E^Q[X] = E^P[L \cdot X]$$

Proof: We only give a proof for the simple example above where $\Omega = \{\omega_1, \ldots, \omega_n\}$. We then have

$$\begin{aligned} E^Q[X] &= \sum_{i=1}^{n} X(\omega_i) q_i = \sum_{i=1}^{n} X(\omega_i) \frac{q_i}{p_i} p_i \\ &= \sum_{i=1}^{n} X(\omega_i) L(\omega_i) p_i = E^P[X \cdot L] \end{aligned}$$

Recall notation:
$E^Q[X] = \int X dQ$ and $E^P[X . L] = \int X L \, dP$
and insert $L = \frac{dQ}{dP}$ to compare...
(careful!)

The Abstract Bayes' Formula

We can also use Radon-Nikodym derivatives in order to compute conditional expectations. The result, known as the abstract **Bayes' Formula**, is as follows.

Theorem 4: Consider two measures P and Q with $Q << P$ on $\mathcal{F}$ and with

$$L^{\mathcal{F}} = \frac{dQ}{dP} \quad \text{on } \mathcal{F}$$

Assume that $\mathcal{G} \subseteq \mathcal{F}$ and let X be a random variable with $X \in \mathcal{F}$. Then the following holds

$$E^{Q}\left[X \middle| \mathcal{G}\right] = \frac{E^{P}\left[L^{\mathcal{F}} X \middle| \mathcal{G}\right]}{E^{P}\left[L^{\mathcal{F}} \middle| \mathcal{G}\right]}$$

note the denominator,

different from $E^Q X = E^P[LX]$ - - - -

(see book Proposition B.41)

Dependence of the σ-algebra

Suppose that we have $Q << P$ on $\mathcal{F}$ with

$$L^{\mathcal{F}} = \frac{dQ}{dP} \quad \text{on } \mathcal{F}$$

Now consider smaller σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Our problem is to find the R-N derivative

$$L^{\mathcal{G}} = \frac{dQ}{dP} \quad \text{on } \mathcal{G}$$

Note that also $Q \ll P$ on $\mathcal{G}$!

We recall that $L^{\mathcal{G}}$ is characterized by the following properties

1. $Q(A) = E^P\left[L^{\mathcal{G}} \cdot I_A\right] \quad \forall A \in \mathcal{G}$

2. $L^{\mathcal{G}} \geq 0$

3. $E^P\left[L^{\mathcal{G}}\right] = 1$

4. $L^{\mathcal{G}} \in \mathcal{G}$ crucial!

A natural guess would perhaps be that $L^{\mathcal{G}} = L^{\mathcal{F}}$, so let us check if $L^{\mathcal{F}}$ satisfies points 1-4 above.

By assumption we have

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] \quad \forall A \in \mathcal{F}$$

Since $\mathcal{G} \subseteq \mathcal{F}$ we then have

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] \quad \forall A \in \mathcal{G}$$

so point 1 above is certainly satisfied by $L^{\mathcal{F}}$. It is also clear that $L^{\mathcal{F}}$ satisfies points 2 and 3. It thus seems that $L^{\mathcal{F}}$ is also a natural candidate for the R-N derivative $L^{\mathcal{G}}$, but the problem is that we do not in general have $L^{\mathcal{F}} \in \mathcal{G}$.

This problem can, however, be fixed. By iterated expectations we have, for all $A \in \mathcal{G}$,

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] = E^P\left[\underbrace{E^P\left[L^{\mathcal{F}} \cdot I_A \middle| \mathcal{G}\right]}_{\text{Consider this}}\right]$$

Then previous formula becomes

$Q(A) = E^P\left[E^P[L^{\mathcal{F}} | \mathcal{G}] I_A \right]$

Since $A \in \mathcal{G}$ we have

$$E^P\left[L^{\mathcal{F}} \cdot I_A \middle| \mathcal{G} \right] = E^P\left[L^{\mathcal{F}} \middle| \mathcal{G} \right] I_A$$

Let us now define $L^{\mathcal{G}}$ by

$$L^{\mathcal{G}} = E^P\left[L^{\mathcal{F}} \middle| \mathcal{G} \right]$$

We then obviously! have $L^{\mathcal{G}} \in \mathcal{G}$ and

$$Q(A) = E^P\left[L^{\mathcal{G}} \cdot I_A \right] \quad \forall A \in \mathcal{G}$$

It is easy to see that also points 2-3 are satisfied so we have proved the following result.

A formula for $L^{\mathcal{G}}$

Proposition 5: If $Q << P$ on $\mathcal{F}$ and $\mathcal{G} \subseteq \mathcal{F}$ then, with notation as above, we have

$$L^{\mathcal{G}} = E^P \left[L^{\mathcal{F}} \middle| \mathcal{G} \right] \qquad !$$

the point is that we want $L^{\mathcal{G}}$ to be $\mathcal{G}$-measurable

The likelihood process on a filtered space

We now consider the case when we have a probability measure P on some space Ω and that instead of just one σ-algebra $\mathcal{F}$ we have a **filtration**, i.e. an increasing family of σ-algebras $\{\mathcal{F}_t\}_{t \geq 0}$.

The interpretation is as usual that $\mathcal{F}_t$ is the information available to us at time t, and that we have $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$.

Now assume that we also have another measure Q, and that for some fixed T, we have $Q << P$ on $\mathcal{F}_T$. We define the random variable L_T by

$$L_T = \frac{dQ}{dP} \quad \text{on } \mathcal{F}_T$$

as on p.150

Since $Q << P$ on $\mathcal{F}_T$ we also have $Q << P$ on $\mathcal{F}_t$ for all $t \leq T$ and we define

$$L_t = \frac{dQ}{dP} \quad \text{on } \mathcal{F}_t \quad 0 \leq t \leq T$$

For every t we have $L_t \in \mathcal{F}_t$, so L is an adapted process, known as the **likelihood process**.

The L process is a P martingale

We recall that

$$L_t = \frac{dQ}{dP} \quad \text{on } \mathcal{F}_t \quad 0 \leq t \leq T$$

Since $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$ we can use Proposition 5 and deduce that

$$L_s = E^P \left[L_t | \mathcal{F}_s \right] \quad s \leq t \leq T$$

and we have thus proved the following result.

Proposition: Given the assumptions above, the likelihood process L is a P-martingale.

Where are we heading?

(and why do we have to know this?)

We are now going to perform measure transformations on Wiener spaces, where P will correspond to the objective measure and Q will be the risk neutral measure.

For this we need define the proper likelihood process L and, since L is a P-martingale, we have the following natural questions.

- What does a martingale look like in a Wiener driven framework? (like Black-Scholes setting)
- Suppose that we have a P-Wiener process W and then change measure from P to Q. What are the properties of W under the new measure Q?

These questions are handled by the Martingale Representation Theorem, and the Girsanov Theorem respectively.

Recall BS framework, with $dS = \mu S dt + \sigma S dW$ (under P)

and $dS = r S dt + \sigma S dW^Q$ (under Q)

end of lecture 36

4.

The Martingale Representation Theorem

Section 11.1

Intuition

from general Itô theory

Suppose that we have a Wiener process W under the measure P. We recall that if h is adapted (and integrable enough) and if the process X is defined by

$E\int_0^T h_s^2 ds < \infty$

$$X_t = x_0 + \int_0^t h_s dW_s$$

then X is a a martingale. We now have the following natural question:

Question: Assume that X is an arbitrary martingale. Does it then follow that X has the form

!

$$X_t = x_0 + \int_0^t h_s dW_s$$

for some adapted process h?

In other words: Are **all** martingales stochastic integrals w.r.t. W?

Answer

It is immediately clear that all martingales can **not** be written as stochastic integrals w.r.t. W. Consider for example the process X defined by

$$X_t = \begin{cases} 0 & \text{for} \quad 0 \leq t < 1 \\ Z & \text{for} \quad t \geq 1 \end{cases}$$

where Z is an random variable, independent of W, with $E[Z] = 0$.

compute $E[X_t \mid \mathcal{F}_s]$ for different $t>$

X is then a martingale (why?) but it is clear (how?) that it cannot be written as

with filtration generated by W (to which filtration is X adapted?)

$$X_t = x_0 + \int_0^t h_s dW_s$$

for any process h. See next page.

Intuition

The intuitive reason why we cannot write

$$X_t = x_0 + \int_0^t h_s dW_s$$

in the example above is of course that the random variable Z "has nothing to do with" the Wiener process W. In order to exclude examples like this, we thus need an assumption which guarantees that our probability space only contains the Wiener process W and nothing else.

This idea is formalized by assuming that the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ **is the one generated by the Wiener process** W,

$\mathcal{F}_t = \sigma(W_s, s \leq t)$

The Martingale Representation Theorem

Theorem. Let W be a P-Wiener process and assume that the filtation is the **internal** one i.e.

$$\mathcal{F}_t = \mathcal{F}_t^W = \sigma\left\{W_s;\ 0 \le s \le t\right\}$$

Then, for every $(P, \mathcal{F}_t)$-martingale X, there exists a real number x and an adapted process h such that

$$X_t = x + \int_0^t h_s dW_s,$$

i.e.

$$dX_t = h_t dW_t.$$

Proof: Hard. This is very deep result.

Crucial is that X is adapted to this special filtration

Note

For a given martingale X, the Representation Theorem above guarantees the existence of a process h such that

$$X_t = x + \int_0^t h_s dW_s,$$

The Theorem does **not**, however, tell us how to find or construct the process h.

5.

The Girsanov Theorem

Sections 11.2, 11.3

Setup

Let W be a P-Wiener process and fix a time horizon T. Suppose that we want to change measure from P to Q on $\mathcal{F}_T$. For this we need a P-martingale L with $L_0 = 1$ to use as a likelihood process, and a natural way of constructing this is to choose a process g and then define L by

$$\begin{cases} dL_t = g_t dW_t \\ L_0 = 1 \end{cases}$$

This definition does not guarantee that $L \geq 0$, so we make a small adjustment. We choose a process φ and define L by

take $g_t = L_t \varphi_t$ for some φ_t

$$\begin{cases} dL_t = L_t \varphi_t dW_t \\ L_0 = 1 \end{cases}$$

under integrability conditions!

The process L will again be a martingale and we easily obtain

$$L_t = e^{\int_0^t \varphi_s dW_s - \frac{1}{2} \int_0^t \varphi_s^2 ds}$$

(check by Itô formula)

Thus we are guaranteed that $L \geq 0$. We now change measure form P to Q by setting

$$dQ = L_t dP, \quad \text{on } \mathcal{F}_t,\ 0 \leq t \leq T$$

The main problem is to find out what the properties of W are, under the new measure Q. This problem is resolved by the **Girsanov Theorem**.

The Girsanov Theorem

Let W be a P-Wiener process. Fix a time horizon T.

Theorem: Choose an adapted process φ, and define the process L by

OK if L is a P_martingale

$$\begin{cases} dL_t &= L_t \varphi_t dW_t \\ L_0 &= 1 \end{cases}$$

Assume that $E^P[L_T] = 1$, and define a new mesure Q on $\mathcal{F}_T$ by

$$dQ = L_t dP, \quad \text{on } \mathcal{F}_t,\ 0 \le t \le T$$

Then $Q << P$ and the process W^Q, defined by

$$W_t^Q = W_t - \int_0^t \varphi_s ds$$

is Q-Wiener. We can also write this as

$$dW_t = \varphi_t dt + dW_t^Q$$

Changing the drift in an SDE

(Section 11.5)

The single most common use of the Girsanov Theorem is as follows. (has to with BS like models)

Suppose that we have a process X with P dynamics

$$dX_t = \mu_t dt + \sigma_t dW_t$$

where μ and σ are adapted and W is P-Wiener.

We now do a Girsanov Transformation as above, and the question is what the Q-dynamics look like.

From the Girsanov Theorem we have

$$dW_t = \varphi_t dt + dW_t^Q$$ (page 166)

and substituting this into the P-dynamics we obtain the Q dynamics as

$$dX_t = \{\mu_t + \sigma_t \varphi_t\} dt + \sigma_t dW_t^Q$$

Moral: The drift changes but the diffusion is unaffected, meaning that we keep on having the same σ_t in front of the new Brownian motion

The Converse Girsanov Theorem

Let W be a P-Wiener process. Fix a time horizon T.

Theorem. Assume that:

- $Q << P$ on $\mathcal{F}_T$, with likelihood process

(I think we even needs $Q \sim P$)

$$L_t = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}_t \; 0, \leq t \leq T$$

- The filtation is the **internal** one .i.e.

$$\mathcal{F}_t = \sigma \left\{ W_s; \; 0 \leq s \leq t \right\}$$

Then there exists a process φ such that

$$\begin{cases} dL_t &= L_t \varphi_t dW_t \\ L_0 &= 1 \end{cases}$$

note (p.155) that L is a P-martingale.

end of lecture 3C.