Summer Project 2024

Stochastic Calculus and Option Pricing

September 4, 2024

Abstract

This document was produced to aid learning about stochastic calculus in a bid to apply it to finance. It includes martingales, Ito integration, risk-neutral pricing and the Black-Scholes and Heston models.

Contents

1 Introduction

The creation of this document is in the pursuit of learning about stochastic calculus and its applications to finance. We begin by exploring martingales and then define Ito integration, before looking at some finance specific ideas like risk-neutral measures and option pricing.

2 Discrete Time Martingales

In this section we will introduce the definition of a martingale and we will establish some fundamental results.

Definition 2.1 (Discrete Time Martingale). A sequence of random variables $\{M_n\}_{n=0}^{\infty}$ is a martingale with respect to the random variables $\{X_n\}_{n=1}^{\infty}$ if for all $n \geq 1$ there exists $f_n : \mathbb{R}^n \to \mathbb{R}$ such that $M_n = f_n(X_1, ..., X_n)$ and:

$$E[M_n \mid X_1, ..., X_{n-1}] = M_{n-1} \tag{1}$$

Remark. There exists notions of sub and super martingales where we have an inequality in (1). For sub-martingale : $E[M_n \mid X_1, ..., X_{n-1}] \geq M_{n-1}$ and for super-martingale: $E[M_n \mid X_1, ..., X_{n-1}] \leq M_{n-1}$.

We will often use the notation: $E[M_n \mid \mathcal{F}_{n-1}] = E[M_n \mid X_1, ..., X_{n-1}]$, this will become more familiar as we explore continuous time martingales later and the idea of filtration.

Lemma 2.1. *If M_n is a martingale w.r.t. $\{\mathcal{F}\}$ and f is a convex function, then $f(M_n)$ is a sub-martingale.*

Proof. Immediate by Jensen's inequality. □

2.1 Examples

- Sum of Random Variables: define $M_n = X_1 + ... + X_n$ for $n \geq 1$ and $M_0 = 0$, then we have a martingale given X_i are a sequence of iid random variables and $E[X_i] = 0$ for all i.
- Product of Random Variables: define $M_n = X_1 \times ... \times X_n$ for $n \geq 1$ and $M_0 = 1$, then we have a martingale given X_i are a sequence of iid random variables and $E[X_i] = 1$ for all i.

2.2 Martingale Transforms and Stopping Times

Definition 2.2. Non-anticipating A sequence of random variables $\{A_n\}_{n=1}^{\infty}$ is called non-anticipating if for all $n \geq 1$ $A_n = f(X_1, ..., X_{n-1})$ or A_n is $\mathcal{F}_n$ measurable.

This definition is very intuitive, in a gambling setting, if we consider a random variable which decides how much we want to bet on the next round it does not make sense for it to require future information.

Theorem 2.2 (Martingale Transform). *If $\{M_n\}$ is a martingale wrt $\{\mathcal{F}\}$ and we have a sequence $\{A_n\}_{n=1}^{\infty}$ of non-anticipating, bounded random variables, then $\tilde{M}_n$ defined as follows is a martingale:*

$$\tilde{M}_n = A_1(M_1 - M_0) + A_2(M_2 - M_1) + ... + A_n(M_n - M_{n-1}) \tag{2}$$

Proof. [1] It is immediately clear $\tilde{M}_n$ is $\mathcal{F}_n$-measurable because M_n is a martingale and the A_n are non-anticipating.

$$E[\tilde{M}_n \mid \mathcal{F}_{n-1}] = \tilde{M}_{n-1} + E[A_n(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] \tag{3}$$
$$= \tilde{M}_{n-1} + A_n E[(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] \tag{4}$$
$$= \tilde{M}_{n-1} \tag{5}$$

where we use $E[(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] = 0$ as M_n is a martingale. $\square$

In a gambling setting, a gambler is likely to have thresholds where they will stop gambling. We must ensure that a stopping time/condition does not destroy the martingale structure.

Definition 2.3. A random variable τ taking values $\{0, 1, 2, ...\} \cup \{\infty\}$ is a called a stopping time for $\{\mathcal{F}_n\}$ if $\{\tau \leq n\} \in \mathcal{F}_n$ for all $0 \leq n < \infty$.

$\{\tau \leq n\} \in \mathcal{F}_n$ means that the most information the stopping time can have to decide whether to stop at n is what happens at time n.

Remark. We note that τ need not be finite and so we use $n \wedge \tau = min(n, \tau)$ to then define the stopped process $\{M_{n\wedge\tau}\}$ and we define M_τ as:

$$\sum_{k=0}^{\infty} \mathbb{1}_{\{\tau(\omega)=k\}} M_k \tag{6}$$

Theorem 2.3 (Stopping Time Theorem)**.** *If $\{M_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$, then the stopped process $\{M_{n\wedge\tau}\}$ is also a martingale with respect to $\{\mathcal{F}_n\}$.*

Proof. [1] WLOG we assume that $M_0 = 0$ - it clear that a martingale shifted by a constant is still a martingale. Now we look to apply the martingale transform and define $\{A_n\}_{n=1}^{\infty}$ as $A_n = \mathbb{1}_{\{\tau \geq n\}}$. A_n is clearly bounded and non-anticipating because τ is a stopping time.

$$\sum_{k=1}^{n} A_k(M_k - M_{k-1}) = \sum_{k=1}^{n} \mathbb{1}_{\{\tau \geq k\}}(M_k - M_{k-1}) \tag{7}$$
$$= \mathbb{1}_{\{\tau \geq 1\}}(M_1 - M_0) + \mathbb{1}_{\{\tau \geq 2\}}(M_2 - M_1) + ... + \mathbb{1}_{\{\tau \geq n\}}(M_n - M_{n-1}) \tag{8}$$
$$= M_\tau \mathbb{1}_{\{\tau \leq n-1\}} + M_n \mathbb{1}_{\{\tau \geq n\}} \tag{9}$$
$$= M_{n\wedge\tau} \tag{10}$$

Line (9) comes from fact that if $\tau \geq n$ then we get a telescoping sum leaving M_n as M_0 is zero and if $\tau \leq n-1$ then we know $M_{k+1} - M_k = 0$ for all $\tau \geq i$, meaning the telescoping sum abruptly stops and leaves M_τ. $\square$

Theorem 2.4 (Doob's Maximal Inequality)**.** *Let $\{M_n\}$ be a non-negative sub-martingale with respect to $\{\mathcal{F}_n\}$, and $\lambda > 0$, then:*

$$\lambda P(M_n^* \geq \lambda) \leq E[M_n \mathbb{1}_{\{M_n^* \geq \lambda\}}] \leq E[M_n] \tag{11}$$

where $M_n^* = \sup_{0 \leq m \leq n} M_m$ (called maximal sequence).

Proof. [1] The proof of this comes down to two observations. The first of which is that for $A \in \mathcal{F}_m$ and $m \leq n$:

$$E[M_m \mathbb{1}_A] \leq E[M_n \mathbb{1}_A] \tag{12}$$

We can prove this via the definition of a sub-martingale for m and $m+1$:

$$M_m \leq E[E[M_{m+2} \mid \mathcal{F}_{m+1}] \mid \mathcal{F}_m] \tag{13}$$
$$M_m \leq E[M_{m+2} \mid \mathcal{F}_m] \tag{14}$$

Where we use tower property of conditional expectation in line (14) (A.1). It is clear we can extend this for m and n, and so for $A \in \mathcal{F}_m$:

$$\begin{aligned} E[\mathbb{1}_A M_m] &\leq E[\mathbb{1}_A\, E[M_n \mid \mathcal{F}_m]] & (15)\\ &\leq E[E[\mathbb{1}_A M_n \mid \mathcal{F}_m]] & (16)\\ &= E[\mathbb{1}_A M_n] & (17)\end{aligned}$$

Now for the second observation, we consider a stopping time $\tau = \min\{m : M_m \geq \lambda\}$, which we can use to write $P(M_n^* \geq \lambda) = P(\tau \leq n)$ and importantly:

$$\lambda \mathbb{1}_{\{\tau \leq n\}} \leq M_\tau \mathbb{1}_{\{\tau \leq n\}} = \sum_{0 \leq m \leq n} M_m \mathbb{1}_{\{\tau = m\}} \tag{18}$$

Finally we can proceed with the proof - we let $A = \{\tau = m\}$, meaning $A \in \mathcal{F}_m$ giving $E[M_m \mathbb{1}_{\{\tau=m\}}] \leq E[M_n \mathbb{1}_{\{\tau=m\}}]$. Also taking expectations of line (18), we can yield the result:

$$\lambda P(\tau \leq n) \leq E[\sum_{0 \leq m \leq n} M_m \mathbb{1}_{\{\tau = m\}}] \leq E[M_n \mathbb{1}_{\{\tau \leq n\}}] \leq E[M_n] \tag{19}$$

$$\lambda P(M_n^* \geq \lambda) \leq E[M_n \mathbb{1}_{\{M_n^* \geq \lambda\}}] \leq E[M_n] \tag{20}$$

□

Corollary 2.4.1. *Let $\{M_n\}$ be a non-negative sub-martingale with respect to $\{\mathcal{F}_n\}$ and $\lambda > 0$, then for all $p \geq 1$:*

$$\lambda^p P(M_n^* \geq \lambda) \leq E[M_n^p] \tag{21}$$

Proof. [1] This is immediately clear as if M_n is a non-negative sub-martingale, then M_n^p is also by Jensen's inequality. Then we just have to realise that for random variable X $P(X \geq \lambda) \leq P(X^p \geq \lambda^p)$. □

Now we introduce a lemma for random variables X and Y in order to prove Doob's L^p inequality.

Lemma 2.5. *Let X and Y be non-negative random variables such that $Y \in L^p(dP)$ for some $p > 1$, if we have $\forall \lambda \geq 0$:*

$$\lambda P(X \geq \lambda) \leq E[Y \mathbb{1}_{\{X \geq \lambda\}}] \tag{22}$$

Then:

$$\|X\|_p \leq \frac{p}{p-1} \|Y\|_p \tag{23}$$

Proof. [1] We note in the assumptions for this theorem we do not have $X \in L^p(dP)$ and so to avoid issues we will work with $X_n = min(n, X)$. The first thing to note is that equation (22) still applies to X_n and Y. One can see this by considering the event $\{X_n \geq \lambda\}$ which is the same as $\{X \geq \lambda\} \cap \{n \geq \lambda\}$ and so for $n \geq \lambda$ nothing changes at all and otherwise we have $0 \leq 0$.

We use the following identity, for all $z \geq 0$:

$$z^p = p\int_0^z x^{p-1}dx = \int_0^\infty x^{p-1}\mathbb{1}_{\{z\geq x\}}dx \tag{24}$$

Then we can swap in X_n and proceed as follows:

$$\begin{align}
E[X_n^p] &= p\int_0^\infty x^{p-1}P(X_n \geq x)dx \tag{25}\\
&\leq p\int_0^\infty x^{p-2}E[Y\mathbb{1}_{\{X_n\geq x\}}]dx \tag{26}\\
&= pE[Y\int_0^\infty x^{p-2}\mathbb{1}_{\{X_n\geq x\}}dx] \tag{27}\\
&= \frac{p}{p-1}E[YX_n^{p-1}] \tag{28}
\end{align}$$

Now using Holder's inequality for p and q conjugate pair:

$$\begin{align}
\frac{p}{p-1}E[YX_n^{p-1}] = \|YX_n^{p-1}\|_1 &\leq \|Y\|_p\|X_n^{p-1}\|_q \tag{29}\\
&= \|Y\|_p(\int X_n^{(p-1)q})^{\frac{1}{q}} \tag{30}\\
&= \|Y\|_p\|X_n\|_p^{p-1} \tag{31}
\end{align}$$

Finally, dividing through and applying Fatou's lemma we yield the result. □

Corollary 2.5.1. *Let M_n be a non-negative sub-martingale with respect to $\{\mathcal{F}_n\}$, then for $p > 1$ and $n \geq 0$:*

$$\|M_n^*\|_p \leq \frac{p}{p-1}\|M_n\|_p \tag{32}$$

where $M_n^* = \sup_{0\leq m\leq n} M_m$.

Proof. Combine theorem 2.4 and lemma 2.5. □

3 Continuous Time Martingales and Brownian Motion

Definition 3.1 (Filtrations)**.** A filtration $\mathcal{F}(t)$ is a sequence of of sub sigma-algebras, meaning $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ for $s \leq t$.

Remark. This is the time continuous version of $\mathcal{F}_n$ that we introduced earlier. This contains the information we have so far.

Definition 3.2 (Adapted Processes)**.** A stochastic process $M(t)$ is adapted to filtration $\mathcal{F}(t)$ if $M(t)$ is $\mathcal{F}(t)$ measurable for all $t \geq 0$.

The obvious filtration for a stochastic process $M(t)$ is the sigma-algebra of everything up to that point i.e. $\mathcal{F}(t) = \sigma(M(s) : s \leq t)$. It can sometimes be advantageous to define a slightly bigger filtration: $\mathcal{F}(t) \subset \mathcal{F}^*(t)$, which of course does not change the process being adapted or not.

Definition 3.3 (Continuous Time Martingales)**.** $M(t)$ is a continuous martingale w.r.t. to $\mathcal{F}(t)$ if it is continuous with respect to time, adapted to $\mathcal{F}(t)$ and for $s \leq t$:

$$E[M(t) \mid \mathcal{F}(s)] = M(s) \tag{33}$$

3.1 Maximal Inequalities

Theorem 3.1. *Let M_t be a non-negative, time continuous sub- martingale and let $\lambda > 0$, then for all $p \geq 1$:*

$$P\left(\sup_{\{t:0\leq t\leq T\}} M_t > \lambda\right) \leq \frac{E[M_T^p]}{\lambda^p} \tag{34}$$

Also if $M_T \in L^p(dP)$ then for some $p > 1$:

$$\|\sup_{\{t:0\leq t\leq T\}} M_t\|_p \leq \frac{p}{p-1}\|M_T\|_p \tag{35}$$

Proof. [1] This is an easy extension from the discrete time version. In order to use the discrete version we define a set of times which split up $[0, T]$ and in the limit become a continuous interval:

$$S(n, T) = \{t_i : t_i = \frac{iT}{2^n}, 0 \leq i \leq 2^n\} \tag{36}$$

By the continuity of M_t we note that:

$$\lim_{n\to\infty} \sup_{t\in S(n,T)} M_t = \sup_{0\leq t\leq T} M_t \tag{37}$$

$$\lim_{n\to\infty} \mathbb{1}_{\{\sup_{t\in S(n,T)} M_t > \lambda\}} = \mathbb{1}_{\{\sup_{0\leq t\leq T} M_t > \lambda\}} \tag{38}$$

So by the discrete version in theorem 2.4 we have:

$$P\left(\sup_{t\in S(n,T)} M_t \geq \lambda\right) \leq \frac{E[M_T^p]}{\lambda^p} \tag{39}$$

Then we take the lim inf on both sides, acknowledging that the probability is nothing other than the integral of $\mathbb{1}_{\{\sup_{t\in S(n,T)} M_t > \lambda\}}$ with respect to the probability measure, we use Fatou's lemma which maintains the inequality we need.

For (35) simply apply Fatou's lemma to the discrete version (corollary 2.5.1) using the discretisation $S(n, T)$. □

Remark. As one might expect, there does exist a stopping time theorem for the continuous setting. We won't prove it here because it requires a bit of a diversion into uniform integrability.

3.2 Brownian Motion

In this section we will introduce the fundamental idea of Brownian motion and explore a construction of it with functions.

Definition 3.4. A time continuous stochastic process $\{W(t) : 0 \leq t < T\}$ is known as standard Brownian motion on the interval $[0, T)$ if the following hold:

- $W(0) = 0$
- for any finite set of times: $0 \leq t_1 < t_2 < ... < t_n < T$, the random variables defined as $W(t_2) - W(t_1), W(t_3) - W(t_2), ..., W(t_n) - W(t_{n-1})$ are independent

- random variable $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$
- $W(t)$ is a continuous a.s.

Brownian is indeed time continuous martingale with respect to its natural filtration $\mathcal{F}(t) = \sigma(W(s) : 0 \leq s \leq t)$. However, it is often helpful to expand this filtration to make it right-continuous and contain all subsets of measure zero sets. This makes it a complete filtration. We will not dwell on this going forward.

3.3 Quadratic Variation

Definition 3.5. For a partition $\pi = \{t_0, t_1, ..., t_n\}$ of interval $[0, T]$ and a function $f(t)$ defined on the interval, the quadratic variation [f,f](T) is defined as:

$$\lim_{\|\pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 \tag{40}$$

where $\|\pi\| = \max_j (t_{j+1} - t_j)$

Theorem 3.2. *Let $W(t)$ be Brownian motion, then [W,W](t) = t for all $t \in [0, T]$ a.s. (i.e. the set of $\omega \in \Omega$ such that the path of Brownian motion has quadratic variation t has full measure (probability one)).*

Proof. [2] Let $\pi = \{t_0, t_1, ..., t_n\}$ be a partition of $[0, T]$ then we define the random variable Q_π as follows:

$$Q_\pi = \sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2 \tag{41}$$

Now we want to show that this random variable converges in probability to T; we can do this by showing its expectation equals T and in the limit its variance tends to 0. This proves convergence in L^2 i.e $E[|Q_\pi - T|^2] \to 0$, and by Markov's inequality it is clear this implies convergence in probability as required.

$$\begin{aligned}
E[Q_\pi] &= \sum_{j=0}^{n-1} E[(W(t_{j+1}) - W(t_j))^2] && (42)\\
&= \sum_{j=0}^{n-1} Var[W(t_{j+1}) - W(t_j)] + \sum_{j=0}^{n-1} E[W(t_{j+1}) - W(t_j)] && (43)\\
&= \sum_{j=0}^{n-1} t_{j+1} - t_j && (44)\\
&= T && (45)\\
Var[Q_\pi] &= \sum_{j=0}^{n-1} Var[(W(t_{j+1}) - W(t_j))^2] && (46)\\
&= \sum_{j=0}^{n-1} E[(W(t_{j+1}) - W(t_j))^4] - \sum_{j=0}^{n-1} (t_{j+1} - t_j)^2 && (47)\\
&= 2 \sum_{j=0}^{n-1} (t_{j+1} - t_j)^2 \to 0 && (48)
\end{aligned}$$

□

Lemma 3.3. *Brownian motion has unbounded total variation*

Proof. Suppose it has bounded total variation:

$$\sum_{j=0}^{n-1}(W(t_{j+1})-W(t_j))^2 \leq \max_{0\leq k\leq n-1}|W(t_{k+1})-W(t_k)|\sum_{j=0}^{n-1}|W(t_{j+1})-W(t_j)| \xrightarrow{\|\pi\|\to 0} 0 \quad (49)$$

which is a contradiction. □

4 Ito Integration

In this section we explore Ito Integrals which enable us to integrate with respect to Brownian motion and other stochastic processes. This forms the backbone of stochastic calculus, for example, consider the following stochastic differential equation:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \quad (50)$$

This is known as geometric Brownian motion which we will explore later, but for now the thing to note is that so far $dW(t)$ is undefined, where $W(t)$ denotes Brownian motion. However, it does seem natural, as in normal calculus that you can say something changes with time dt, that one would be able to describe changes with respect to a random process, like Brownian motion.

We follow the standard presentation found in many sources, including [1],[2] and [3]. The proofs will be referenced individually.

4.1 Defining Ito Integrals

Before we begin we note that if Brownian motion had bounded total variation integrating with respect to it would not go beyond a Riemann-Stieltjes integral. However, its total variation is infinite and so we use the following construction for an integral of the form:

$$\int_0^t \Delta(\omega,t)dW(t) \quad (51)$$

where $W(t)$ is Brownian motion adapted to a filtration $\mathcal{F}(t)$, and we are integrating an adapted process Δ. Like one might do when defining the Lebesgue integral, we consider first what we want the integral of a simple process to be.

Consider a partition P of the time interval $[0,T]$ $P = \{0 = t_0 \leq t_1 \leq ... \leq t_n = T\}$, then $\Delta(\omega,t)$ is a simple process if it is constant on each $[t_i, t_{i+1})$.

$$\Delta(\omega,t) = \sum_{j=0}^{n-1} a_j(\omega)\mathbb{1}[t_j,t_{j+1}) \quad (52)$$

The integral of $\Delta(\omega,t)$ is defined as:

$$I(\Delta,t) = \int_0^t \Delta(\omega,s)dW(s) = \sum_{j=0}^{k-1} a_j(\omega)[W(t_{j+1})-W(t_j)] + a_k[W(t)-W(t_k)] \quad (53)$$

Unsurprisingly, we use this construction to define Ito integration for a larger set of processes. This involves approximating said processes with simple processes and how this is done is fundamental to definition of an Ito integral - let $\Delta(\omega, t)$ be an arbitrary process defined on $[0, T]$, then we could approximate it with:

$$\Delta_n(\omega, t) = \sum_{j=0}^{n-1} \Delta(\omega, t_j^*) \mathbb{1}[t_j, t_{j+1}) \tag{54}$$

where we have an even partition $P_n = \{0 = t_0, t_1, ..., t_n = T\}$. The choice is then where t_j^* lies within the interval $[t_j, t_{j+1})$. For Riemann integrals this does not matter, but here, as the following example shows, it does.

4.1.1 Example

We will consider $\int_0^T W(\omega, t) dW(\omega, t)$ and use the 2 following approximations [3]:

$$\phi_1(\omega, t) = \sum_{j=0}^{2^n-1} W(\omega, jT2^{-n}) \mathbb{1}[jT2^{-n}, (j+1)T2^{-n}) \tag{55}$$

$$\phi_2(\omega, t) = \sum_{j=0}^{2^n-1} W(\omega, (j+1)T2^{-n}) \mathbb{1}[jT2^{-n}, (j+1)T2^{-n}) \tag{56}$$

Clearly ϕ_1 uses the left endpoint and ϕ_2 uses the right endpoint, let's see how they integrate differently by taking the expectation:

$$E\left[\int_0^T \phi_1(\omega, t) dW(t)\right] = \sum_{j=0}^{2^n-1} E\left[W(\omega, jT2^{-n}) \cdot (W(\omega, (j+1)T2^{-n}) - W(\omega, jT2^{-n}))\right] = 0 \tag{57}$$

$$E\left[\int_0^T \phi_2(\omega, t) dW(t)\right] = \sum_{j=0}^{2^n-1} E\left[W(\omega, (j+1)T2^{-n}) \cdot (W(\omega, (j+1)T2^{-n}) - W(\omega, jT2^{-n}))\right] \tag{58}$$

$$= \sum_{j=0}^{2^n-1} \left[(W(\omega, (j+1)T2^{-n}) - W(\omega, jT2^{-n}))^2\right] = T \tag{59}$$

To define an Ito integral one must use the left endpoint (the Stratonovich integral comes from using the midpoint). The Ito integral is more useful for us and you can see that from a finance perspective in that one must make a decision at the start of the time period, for example, to buy a stock.

Now we will formally define the integral for the set of functions $\mathcal{V}$ defined below.

Definition 4.1. $\mathcal{V}$ is the set of functions $g(\omega, t) : \Omega \times [0, \infty) \to \mathbb{R}$ satisfying:

- $\mathcal{B} \times \mathcal{F}$ measurable
- $\mathcal{F}(t)$ adapted
- $E\left[\int_0^T g(\omega, s)^2 ds\right] < \infty$

Now we prove a result which will help us extend the integral to $\mathcal{V}$:

Lemma 4.1. *(Ito's Isometry for simple process) For simple processes f we have:* $\|I(f,T)\|_{L^2(dP)} = \|f\|_{L^2(dP\times dt)}$

Proof. [1] Given we can write f as $f(\omega,t) = \sum_{j=0}^{n-1} a_j(\omega)\mathbb{1}[t_j,t_{j+1})$ and given the subintervals are disjoint f^2 is simply $f^2(\omega,t) = \sum_{j=0}^{n-1} a_j^2(\omega)\mathbb{1}[t_j,t_{j+1})$. Hence we have:

$$\|f\|_{L^2(dP\times dt)} = \int\int\sum_{j=0}^{n-1} a_j^2(\omega)\mathbb{1}[t_j,t_{j+1})dtdP \tag{60}$$

$$= E\left[\sum_{j=0}^{n-1} a_j^2(\omega)(t_{j+1}-t_j)\right] \tag{61}$$

$$= \sum_{j=0}^{n-1} E\left[a_j(\omega)^2\right](\omega)(t_{j+1}-t_j) \tag{62}$$

Then using that the expectation of terms of the form $(W(t_{j+1}) - W(t_j))(W(t_{i+1}) - W(t_i))$ is zero for $i \neq j$, we have:

$$\|I(f,t)\|_{L^2(dP)} = \int(\sum_{j=0}^{n-1} a_j(\omega)[W(t_{j+1}) - W(t_j)])^2 dP \tag{63}$$

$$= \sum_{j=0}^{n-1} E\left[a_j(\omega)^2\right] E\left[(W(t_{j+1}) - W(t_j))^2\right] \tag{64}$$

$$= \sum_{j=0}^{n-1} E\left[a_j(\omega)^2\right](t_{j+1}-t_j) \tag{65}$$

□

One can prove that the simple processes defined before are dense in $\mathcal{V}$ (see [3] page 27 - this is a 3 step process.), so for every $g \in \mathcal{V}$, there exists ϕ_n such that:

$$E\left[\int_0^T (g-\phi_n)^2 dt\right] \to 0 \tag{66}$$

Hence we define the Ito integral for $g \in \mathcal{V}$:

$$\int_0^T g(\omega,t)dW(t) = \lim_{n\to\infty}\int_0^T \phi_n(\omega,t)dW(t) \tag{67}$$

Then the principal question becomes the existence of this limit. First because we know ϕ_n converges to g in $L^2(dt\times dP)$, meaning the ϕ_n form a Cauchy sequence in $L^2(dt\times dP)$. Ito's isometry then tells us that $I(\phi_n,t)$ forms a Cauchy sequence in $L^2(dP)$ and so we are done because $L^2(dP)$ is complete. One also ought to check it is well defined in the sense that the choice of approximating sequence does not matter - this again follows from the Ito isometry [1]

Now we can prove a few results about the process that results from Ito integration. The immediate one is that Ito's isometry extends to processes $\in \mathcal{V}$ - this is easy to see by limits.

Theorem 4.2. *Ito integral* $I(\Delta,t) = \int_0^t \Delta(\omega,s)dW(s)$ *has quadratic variation given by:*

$$[I,I](t) = \int_0^t \Delta(\omega,s)^2 ds \tag{68}$$

Proof. [2] Assume $\Delta(\omega, t)$ is constant on subintervals. So we have a partition of [0,t], where Δ is constant, $\pi = \{t_0, t_1, ..., t_n\}$ and then a sub partition for each $[t_i, t_{i+1}]$: $\tau = \{s_0, s_1, ..., s_k\}$. Now we can calculate the quadratic variation on $[t_i, t_{i+1}]$:

$$\sum_{j=0}^{k-1}[I(s_{j+1}) - I(s_i)]^2 = \sum_{j=0}^{k-1}[\Delta(\omega, t_j)(W(s_{j+1}) - W(s_j))]^2 \tag{69}$$

$$= \Delta(\omega, t_j)^2 \sum_{j=0}^{k-1}(W(s_{j+1}) - W(s_j))^2 \tag{70}$$

It is then clear that as $\|\tau\| \to 0$ we get $\Delta(\omega, t_j)^2$ multiplied by the quadratic variation of Brownian motion on $[t_i, t_{i+1}]$, which is $t_{i+1} - t_i$. Finally, letting $\|\pi\| \to 0$ and adding up the subintervals, we get the result. (Note that there are now no assumptions about Δ being constant as $\|\pi\| \to 0$). ☐

Theorem 4.3. *Let $f \in \mathcal{V}$ and consider $\int_0^t f(\omega, s)dW(s)$. Then there exists a time continuous stochastic process J_t such that for $t \in [0, T]$:*

$$P\left[J_t = \int_0^t f(\omega, s)dW(s)\right] = 1 \tag{71}$$

Proof. [3] We understand the stochastic integral via the simple process ϕ_n which defines it in the limit. We define the process $I_n(t) = \int_0^t \phi_n(\omega, s)dW(s)$. We will first prove this is a martingale.

Let $s < t$ and define $\phi_n(\omega, u) = \sum_{j=0}^n a_j(\omega)\mathbb{1}_{[t_j, t_{j+1})}(u)$, now we will assume that the time indexing falls perfectly on s and t. In other words, there exists k such that $t_k = s$ and $t_n = t$, if this is not the case then you can simply add the time and use the same random variable. This is an abuse of notation as ϕ_n indicates discretisation into n time steps but it is the same function and it makes it cleaner.

$$E[I_n(t) \mid \mathcal{F}(s)] = E[I_n(s) + \int_s^t \phi_n(\omega, u)dW(u) \mid \mathcal{F}(s)] \tag{72}$$

$$= I_n(s) + E[\sum_{j=k}^n a_j(\omega)(W(t_{j+1}) - W(t_j)) \mid \mathcal{F}(s)] \tag{73}$$

$$= I_n(s) + \sum_{j=k}^n E[E[a_j(\omega)(W(t_{j+1}) - W(t_j)) \mid \mathcal{F}(t_j)] \mid \mathcal{F}(s)] \tag{74}$$

$$= I_n(s) + \sum_{j=k}^n E[a_j E[(\omega)(W(t_{j+1}) - W(t_j)) \mid \mathcal{F}(t_j)] \mid \mathcal{F}(s)] \tag{75}$$

$$= I_n(s) \tag{76}$$

Using $I_n(t)$ is a martingale for all n, it follows $|I_n(t) - I_m(t)|$ is a non-negative sub-martingale by lemma 2.1, to which we can apply theorem 3.1 and Ito's isometry. Hence:

$$P(\sup_{0 \le t \le T} |I_n(t) - I_m(t)| \ge \lambda) \le \frac{E[|I_n(t) - I_m(t)|^2]}{\lambda^2} \tag{77}$$

$$= \frac{\|\phi_n - \phi_m\|_{L^2(dP \times dt)}}{\lambda^2} \xrightarrow{m,n \to \infty} 0 \tag{78}$$

So for all $\lambda > 0$ $P\left(\sup_{0\leq t\leq T}|I_n(t) - I_m(t)| \geq \lambda\right)$ tends to 0. It is then clear that one can construct a monotonically increasing sequence n_k such that:

$$P(\sup_{0\leq t\leq T}|I_{n_{k+1}}(t) - I_{n_k}(t)| \geq 2^{-k}) \leq 2^{-k} \tag{79}$$

Now considering the event for k of $\sup_{0\leq t\leq T}|I_{n_{k+1}}(t) - I_{n_k}(t)| \geq 2^{-k}$ it is clear by the geometric sum and the Borel-Cantelli lemma that it has probability 0 of happening for infinitely many k. This tells you there exists a $k^{'}(\omega)$ such that on a full measure set:

$$\sup_{0\leq t\leq T}|I_{n_{k+1}}(\omega,t) - I_{n_k}(\omega,t)| \leq 2^{-k} \text{ for } k \geq k^{'}(\omega) \tag{80}$$

Finally we can conclude that on a full measure set I_{n_k} is uniformly convergent - this means it has a continuous limiting process for each ω in the full measure set. We are done because $I_{n_k}(t)$ also converges to our Ito integral. □

Moving forward we will always assume we have the time continuous process when we use the Ito integral of a function $f \in \mathcal{V}$.

Corollary 4.3.1. *For $f \in \mathcal{V}$, $\int_0^t f(\omega,s)dW(s)$ is a martingale.*

Proof. This follows from taking the limit of the result in the first half of the above proof. We can use dominated convergence to handle the limits. □

4.2 Extending the Domain

In this section, we will extend the definition of the Ito integral to a wider range of functions such that we have a more natural integral definition. This follows the presentation by Steele in [1].

So far we have established Ito integration for adapted, measurable processes which satisfy the following constraint.

$$E\left[\int_0^t \Delta(\omega,s)^2 ds\right] < \infty \tag{81}$$

Whilst this is a wide range of functions, it does exclude some functions which one might think a good theory of integration can handle. For instance, if we have $f : \mathbb{R} \to \mathbb{R}$ continuous then one would hope $\int_0^t f(W(s))dW(s)$ would be defined, but it is not too difficult to find functions such that it isn't. Hence we consider a larger set of functions satsifying:

$$P\left(\int_0^T f(\omega,t)^2 dt < \infty\right) = 1 \tag{82}$$

Denote this set $\mathcal{L}^2_{LOC}$. It is clear that $\mathcal{V} \subset \mathcal{L}^2_{LOC}$ by Markov's inequality. Also by the continuity of Brownian motion we know then that $f(W(t))^2$ is a continuous function and so bounded on [0,T], meaning $\int_0^t f(W(s))dW(s)$ would be defined.

To define the Ito integral on this set we will now define some ideas that help us transition between $\mathcal{V}$ and $\mathcal{L}^2_{LOC}$

Definition 4.2. (Localising Sequences for $\mathcal{V}$) An increasing sequence of stopping times ν_n is called a $\mathcal{V}$ localising sequence for f provided that for all n $f_n(\omega, t) = f(\omega, t)\mathbb{1}_{\{t \leq \nu_n\}} \in \mathcal{V}$ and:

$$P\left(\bigcup_{n=1}^{\infty}\{\omega : \nu_n = T\}\right) = 1 \tag{83}$$

We can see how a localising sequence could be a very useful tool in order to extend to $\mathcal{L}^2_{LOC}$ and it turns out every function in $\mathcal{V}$ has one.

Lemma 4.4. *For any $f \in \mathcal{L}^2_{LOC}[0, T]$, the following sequence defines a localising sequence for $\mathcal{V}[0, T]$:*

$$\tau_n = \inf\{s : \int_0^s f(\omega, t)^2 dt \geq n \text{ or } s \geq T\} \tag{84}$$

Proof. [1] For any $f \in \mathcal{L}^2_{LOC}[0, T]$ it is immediately clear by the definition of the infimum that for $f_n(\omega, t) := f(\omega, t)\mathbb{1}_{\{t \leq \tau_n\}}$ $\|f_n\|^2_{L^2(dP \times dt} \leq n < \infty$. Suppose it was false and $\|f_n\|^2_{L^2(dP \times dt} > n$ then it contradicts τ_n being the infimum. Next we can recognise that:

$$\bigcup_{n=1}^{\infty}\{\omega : \tau_n = T\} = \{\omega : \int_0^T f(\omega, t)^2 dt < \infty\} \tag{85}$$

LHS $\subset$ RHS because if there exists an n such that $\tau_n = T$ by definition this implies $\int_0^T f(\omega, t)^2 dt < \infty$ because it means $\forall \epsilon > 0$ $\int_0^{T-\epsilon} f(\omega, t)^2 dt < n$ and so we have finiteness by continuity. RHS $\subset$ LHS is clear. Hence we are done because the probability of the RHS is 1 for $f \in \mathcal{L}^2_{LOC}[0, T]$. $\square$

We will now define the Ito integral for $f \in \mathcal{L}^2_{LOC}[0, T]$ by defining it for $f_n(\omega, t) := f(\omega, t)\mathbb{1}_{\{t \leq \tau_n\}}$ and then taking the limit as $n \to \infty$. We will prove a few results to make this concrete, but before this we will state without proof the following result.

Theorem 4.5. *Let $f, g \in \mathcal{V}$ and suppose ν is a stopping time such that $f(\omega, s) = g(\omega, s)$ for almost all $\omega \in \{\omega : s \leq \nu\}$. Then:*

$$X(t) = \int_0^t f(\omega, s)dW(s) \text{ and } Y(t) = \int_0^t g(\omega, s)dW(s) \tag{86}$$

are equal for almost all $\omega \in \{\omega : t \leq \nu\}$.

This is an intuitive result, but we omit the proof because it requires a lot of work for not a lot if insight. For the proof see [1] page 89.

Lemma 4.6. *(Sequential Consistency) Let $f \in \mathcal{V}$ and let $\{\nu_n\}$ be a localising sequence. Define $X_n(t)$ as $\int_0^t f\mathbb{1}_{\{t \leq \nu_n\}}$, then for all $t \in [0, T]$ and $n \geq m$:*

$$X_n(t) = X_m(t) \text{ for almost all } \omega \in \{\omega : t \leq \nu_m\} \tag{87}$$

Proof. This follows immediately from theorem 4.5 as $f_n(\omega, t) = f_m(\omega, t)$ for almost all $\omega \in \{\omega : t \leq \nu_m\}$. $\square$

Theorem 4.7. *There exists a continuous process $X(t)$ such that:*

$$P\left(X(t) = \lim_{n \to \infty} X_{t,n}\right) \quad \forall t \in [0, T] \tag{88}$$

Proof. [1] We leverage the previous result for processes in $\mathcal{V}$ namely theorem 4.3 and then using the localising sequence we consider when $\nu_n = T$.

Hence we define $N = \min\{n : \nu_n = T\}$ for ν_n a localising sequence and so by definition we know $P(N < \infty) = 1$. Denote by Ω_0 the probability 1 set where $X_n(t)$, as defined before, is continuous.

Now we work in $\Omega_1 = \Omega_0 \cap \{N < \infty\}$ and define $X(t)$ as $X_N(t)$. It follows immediately that this is continuous $\forall \omega \in \Omega_1$ and using lemma 4.6 we have $\forall t \in [0.T]$:

$$P\left(\lim_{n\to\infty} X_n(t) = X_N(t)\right) = 1 \tag{89}$$

which completes the proof after letting $X(t) = X_N(t)$. □

We now check 1 final thing: indepedence of the localising sequence.

Lemma 4.8. *Let $f \in \mathcal{L}^2_{LOC}[0,T]$ and let ν_n and τ_n be 2 localising sequences. Then, using previous notation, we have:*

$$\lim_{n\to\infty} X_n^{(\nu)}(t) = \lim_{n\to\infty} X_n^{(\tau)}(t) \tag{90}$$

Proof. [1] We define a new sequence of stopping times $\lambda_n = \min(\nu_n, \tau_n)$, then by theorem 4.5 we have for $n \geq m$:

$$X_n^{\nu}(t) = X_n^{\tau}(t) \text{ almost everywhere } \{t \leq \lambda_m\} \tag{91}$$

Then we have proven that the limit exists as $n \to \infty$ so:

$$\lim_{n\to\infty} X_n^{\nu}(t) = \lim_{n\to\infty} X_n^{\tau}(t) \text{ almost everywhere } \{t \leq \lambda_m\} \tag{92}$$

Now to conclude we must show that λ_n is actually a localiser sequence; this follows the following equivalence:

$$\bigcup_{m=1}^{\infty} \{\lambda_m = T\} = \bigcup_{m=1}^{\infty} \{\nu_m = T \text{ and } \tau_m = T\} \tag{93}$$

Hence with probability one there exists m such that $\lambda_m = T$ and so we have the result:

$$\lim_{n\to\infty} X_n^{\nu}(t) = \lim_{n\to\infty} X_n^{\tau}(t) \quad \forall t \in [0,T] \text{ with probability one} \tag{94}$$

□

To conclude, we remark that there exists an extension of theorem 4.5 to $\mathcal{L}^2_{LOC}$, for proof see [1] page 98.

4.3 Riemann and Gaussian Results

In this section we state without proof 2 Riemann results which help us derive a very important result.

Theorem 4.9. *(Riemann Representation) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and let $t_i = \frac{iT}{n}$ such that $\{t_0, t_1, ..., t_m\}$ partitions $[0, T]$. Then:*

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(W(t_i))(W(t_{i+1}) - W(t_i)) = \int_0^T f(W(s))dW(s) \tag{95}$$

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i^*)(W(t_{i+1}) - W(t_i)) = \int_0^T f(s)dW(s) \tag{96}$$

where both limits are in probability and $t_i^* \in [t_i, t_{i+1}]$

The proofs of these results are analysis and so are omitted for brevity, see [1] page 99.

Theorem 4.10. *(Integral of Deterministic Functions) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and let $t_i = \frac{iT}{n}$ such that $\{t_0, t_1, ..., t_m\}$ partitions $[0, T]$. Then:*

$$\int_0^t f(s)dW(s) \sim N(0, \int_0^t f(s)^2 ds) \tag{97}$$

Proof. [1] Using the Riemann Representation, let $X_n(t)$ denote $\sum_{i=0}^{n-1} f(t_i)(W(t_{i+1}) - W(t_i))$. It is immediate that $E[X_n(t)] = 0$ by definition of Brownian motion. Using the independent intervals of BM we can calculate variance as follows:

$$Var[X_n(t)] = \sum_{i=0}^{n-1} f(t_i)^2 Var[(W(t_{i+1}) - W(t_i))] = \sum_{i=0}^{n-1} f(t_i)^2 (t_{i+1} - t_i) \tag{98}$$

Thus we have $X_n(t) \sim N(0, \sum_{i=0}^{n-1} f(t_i)^2 (t_{i+1} - t_i))$ which converges in distribution to (121) as $n \to \infty$. ☐

4.4 Local Martingales

Definition 4.3. A process $M(t)$ which is adapted to $\mathcal{F}(t)$ is called a local martingale if there exists a non-decreasing sequence $\{\tau_n\}$ which $\to \infty$ with probability one such that $M_k(t)$, defined below, is a martingale with respect to $\mathcal{F}(t)$

$$M_k(t) = M_{t\wedge\tau_k} - M_0 \tag{99}$$

One can see the immediate connection to the localiser sequences defined previously and how they allow us to control a more general object. Next is a very important result and is the more general version of Corollary 5.3.1.

Theorem 4.11. *Let $f \in \mathcal{L}^2_{LOC}$, then there is a continuous local martingale $X(t)$ such that:*

$$P\left(X(t) = \int_0^t f(\omega, s)dW(s)\right) = 1 \tag{100}$$

Proof. We take the localising sequence for f as before as $\tau_n = \{t : \inf \int_0^t f(\omega, s)^2 ds \geq n \text{ or } t \geq T\}$, then $f_n(\omega, s) = f(\omega, s)\mathbb{1}_{\{s\leq\tau_n\}} \in \mathcal{V}$.

Now we know $I_n(t) = \int_0^t f_n(\omega, s)dW(s)$ is a martingale for all n and we take $n \to \infty$ to define the Ito integral $I(t)$ of f. It is then immediately obvious that using the same τ_n one can make $I(t \wedge \tau_n)$ a martingale. Therefore, $I(t)$ is a local martingale on $[0, T]$. ☐

Lemma 4.12. *Let $X(t)$ be a local martingale and τ a stopping time, then $X(t \wedge \tau)$ is a local martingale.*

Proof. WLOG assume $X(0) = 0$. $X(t)$ being a local martingale means there exists ν_k such that $X(t \wedge \nu_k)$ is a martingale. Then we know $X(t \wedge \nu_k \wedge \tau)$ is a martingale by Doob's stopping time theorem. Hence ν_k also work to make $X(t \wedge \tau)$ a martingale so it is a local martingale. □

Lemma 4.13. *Let $X(t)$ be a time continuous local martingale and B a constant such that $|X(t)| \leq B$ for all $t \geq 0$, then $X(t)$ is a martingale.*

Proof. [1] WLOG assume $X(0) = 0$. By definition there exists non-decreasing sequence τ_k which tends to infinity with probability 1 such that $X(t \wedge \tau_k)$ is a time continuous martingale for each k. Hence we have the martingale identity:

$$E\left[X(t \wedge \tau_k) \mid \mathcal{F}(s)\right] = X(s \wedge \tau_k) \tag{101}$$

Now we know $X(t \wedge \tau_k) \xrightarrow{k \to \infty} X(t)$ and $X(s \wedge \tau_k) \xrightarrow{k \to \infty} X(s)$, we can use the dominated convergence theorem to conclude:

$$E[X(t) \mid \mathcal{F}(s)] = X(s) \tag{102}$$

□

5 Ito's Formula

In this section we want to work more easily with the Ito integral and subsequent stochastic differential equations. Before stating and proving Ito's formula we review some ideas which will be useful.

Firstly, we have seen $\lim_{\|\pi\| \to 0} \sum_{j=0}^{n-1} (W(t_{j+1} - W(t_j))^2 \to T$ a.s. and so we get the idea that the quadratic variation increases linearly in time with Brownian motion and so we could write $dW(t)dW(t) = dt$. Next we consider cross variation between $W(t)$ and t and between t and itself:

$$|\sum_{j=0}^{n-1} (W(t_{j+1}) - W(t_j))(t_{j+1} - t_j)| \leq \max_{0 \leq k \leq n-1} |W(t_{k+1}) - W(t_k)| \cdot T \xrightarrow{\|\pi\| \to 0} 0 \tag{103}$$

$$\sum_{j=0}^{n-1} (t_{j+1} - t_j)^2 \leq \sum_{j=0}^{n-1} (t_{j+1} - t_j) \cdot \max_{0 \leq k \leq n-1} (t_{k+1} - t_k) \leq T \cdot \|\pi\| \xrightarrow{\|\pi\| \to 0} 0 \tag{104}$$

Hence we can conclude $dW(t)dt = 0$ and $dtdt = 0$.

Theorem 5.1. *Let $f(t, x)$ be a function such that f_x, f_t and f_{xx} all exist and are continuous. Then if $W(t)$ is Brownian for all $T \geq 0$:*

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))dt + \int_0^T f_x(t, W(t))dW(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))dt \tag{105}$$

Proof. [2] We use the Taylor expansion of $f(t, x)$:

$$\begin{aligned} f(t_{j+1}, x_{j+1}) - f(t_j, x_j) &= f_t(t_j, x_j)(t_{j+1} - t_j) + f_x(t_j, x_j)(x_{j+1} - x_j) && (106)\\ &+ \frac{1}{2} f_{xx}(t_j, x_j)(x_{j+1} - x_j)^2 + f_{tx}(t_j, x_j)(t_{j+1} - t_j)(x_{j+1} - x_j) && (107)\\ &+ \frac{1}{2} f_{tt}(t_j, x_j)(t_{j+1} - t_j)^2 + \text{higher-order} && (108) \end{aligned}$$

Now subbing in and using $f(T, W(T)) - f(0, W(0)) = \sum_{j=0}^{n-1} f(t_{j+1}, x_{j+1}) - f(t_j, x_j)$:

$$f(T, W(T)) = f(0, W(0)) + \sum_{j=0}^{n-1} f_t(t_j, W(t_j))(t_{j+1} - t_j) + \sum_{j=0}^{n-1} f_x(t_j, W(t_j))(W(t_{j+1}) - W(t_j)) \tag{109}$$

$$+ \frac{1}{2} \sum_{j=0}^{n-1} f_{xx}(t_j, W(t_j))(W(t_{j+1}) - W(t_j))^2 + \frac{1}{2} \sum_{j=0}^{n-1} f_{tt}(t_j, W(t_j))(t_{j+1} - t_j)^2 \tag{110}$$

$$+ \sum_{j=0}^{n-1} f_{tx}(t_j, W(t_j))(t_{j+1} - t_j)(W(t_{j+1}) - W(t_j)) + \text{higher-order} \tag{111}$$

Now looking at each term in turn we have: First term converges to the Lebesgue integral: $\int_0^T f_t(t, W(t))dt$.

Second term converges to the Ito integral: $\int_0^T f_x(t, W(t))dW(t)$.

Third term converges to the Lebesgue integral: $\int_0^T f_{xx}(t, W(t))dt$.

Finally we show the last 2 terms converge to 0:

$$|\sum_{j=0}^{n-1} f_{tt}(t_j, W(t_j))(t_{j+1} - t_j)^2| \leq \sum_{j=0}^{n-1} |f_{tt}(t_j, W(t_j))|(t_{j+1} - t_j)^2 \tag{112}$$

$$\leq \max_{0 \leq k \leq n-1} (t_{k+1} - t_k) \sum_{j=0}^{n-1} |f_{tt}(t_j, W(t_j))|(t_{j+1} - t_j) \xrightarrow{\|\pi\| \to 0} 0 \cdot \int_0^T |f_{tt}(t, W(t))|dt = 0 \tag{113}$$

$$|\sum_{j=0}^{n-1} f_{tx}(t_j, W(t_j))(t_{j+1} - t_j)(W(t_{j+1}) - W(t_j))| \leq \sum_{j=0}^{n-1} |f_{tx}(t_j, W(t_j)|(t_{j+1} - t_j)|(W(t_{j+1}) - W(t_j))| \tag{114}$$

$$\leq \max_{0 \leq k \leq n-1} |(W(t_{k+1}) - W(t_k))| \sum_{j=0}^{n-1} |f_{tx}(t_j, W(t_j)|(t_{j+1} - t_j) \xrightarrow{\|\pi\| \to 0} 0 \cdot \int_0^T |f_{tx}(t, W(t))|dt = 0 \tag{115}$$

It is clear that the higher order terms have the same fate. To rigorously complete this proof, one would have to consider types of convergence in order to get a.s. and also whether the function f has a compact support. To see this refer to [1] page 114. □

5.1 Example

Using what we have proven so far we can shown the well-known result that $W(t)^2 - t$ is a martingale. Firstly, using Ito's formula with $f(t, x) = x^2 - t$ we have:

$$W(t)^2 - t = \int_0^t -1ds + \int_0^t 2W(s)dW(s) + \frac{1}{2} \int_0^t 2ds \tag{116}$$

$$= \int_0^t 2W(s)dW(s) \tag{117}$$

One must check $W(t) \in \mathcal{V}$:

$$E\left[\int_0^T W(s)^2 ds\right] = \int_0^T E\left[W(s)^2\right] ds = \frac{t^2}{2} \tag{118}$$

Hence we can conclude by Corollary 5.3.1 that $W(t)^2 - t$ is a martingale.

5.2 Ito Processes

Definition 5.1. Let $W(t)$ for $t \geq 0$ be Brownian motion and let $\mathcal{F}(t)$ be its filtration. Then for adapted stochastic processes Δ and Θ, belonging to $\mathcal{L}^2_{LOC}$ and L^1 respectively, we define an Ito process $X(t)$ as:

$$X(t) = X(0) + \int_0^t \Delta(\omega, s) dW(s) + \int_0^t \Theta(\omega, s) ds \tag{119}$$

5.2.1 Quadratic Variation

Lemma 5.2. *For the Ito process $X(t)$ defined above $[X, X](t) = \int_0^t \Delta(\omega, s)^2 ds$*

Proof. [2] Consider partition π of the interval $[0, t]$: $\pi = \{t_0, t_1, ..., t_n\}$ and we define $I(t) = \int_0^t \Delta(\omega, s) dW(s)$ and $R(t) = \int_0^t \Theta(\omega, s) ds$. We note that both $I(t)$ and $R(t)$ are continuous.

$$\begin{aligned}
Q_\pi &= \sum_{i=0}^{n-1} [X(t_{i+1}) - X(t_i)]^2 &(120)\\
&= \sum_{i=0}^{n-1} [I(t_{j+1}) - I(t_j) + R(t_{j+1}) - R(t_j)]^2 &(121)\\
&= \sum_{i=0}^{n-1} (I(t_{j+1}) - I(t_j))^2 + 2(I(t_{j+1}) - I(t_j))(R(t_{j+1}) - R(t_j)) + (R(t_{j+1}) - R(t_j))^2 &(122)
\end{aligned}$$

From left to right, as $\|\pi\| \to 0$, the first term converges to $\int_0^t \Delta(\omega, s)^2 ds$ as in theorem 4.2

We can bound the second term as follows:

$$\begin{aligned}
\sum_{i=0}^{n-1} (I(t_{j+1}) - I(t_j))(R(t_{j+1}) - R(t_j)) &\leq \max_{0 \leq k \leq n-1} |I(t_{k+1}) - I(t_k)| \sum_{i=0}^{n-1} |R(t_{j+1}) - R(t_j)| &(123)\\
&\leq \max_{0 \leq k \leq n-1} |I(t_{k+1}) - I(t_k)| \sum_{i=0}^{n-1} \int_{t_j}^{t_{j+1}} |\Theta(\omega, s)| ds &(124)\\
&\leq \max_{0 \leq k \leq n-1} |I(t_{k+1}) - I(t_k)| \int_0^t |\Theta(\omega, s)| ds &(125)
\end{aligned}$$

Then by the continuity of $I(t)$ as $\|\pi\| \to 0$ this tends to 0. For the third term it is clear we have the bound:

$$\max_{0 \leq k \leq n-1} |R(t_{k+1}) - R(t_k)| \int_0^t |\Theta(\omega, s)| ds \tag{126}$$

This tends to zero with the same reasoning. □

5.2.2 Integrating with respect to an Ito process

The following is Shreve

Definition 5.2. Let $X(t)$ be an Ito process as defined above, and let $\beta(t)$ be an adapted process. We define the integral of this process with respect to the Ito process as:

$$\int_0^t \beta(s)dX(s) = \int_0^t \beta(s)\Delta(s)dW(s) + \int_0^t \beta(s)\theta(s)ds \tag{127}$$

Theorem 5.3. *(Ito Formula for an Ito Process) Let $f(t,x)$ be a function such that f_x, f_x and f_{xx} all exist and are continuous, then for $X(t)$ an Ito process:*

$$f(T,X(T)) = f(0,X(0)) + \int_0^T f_t(t,X(t))dt + \int_0^T f_x(t,X(t))dX(t) + \frac{1}{2}\int_0^T f_{xx}(t,X(t))d[X,X](t) \tag{128}$$

Using the above definition and Lemma 5.2 we can rewrite this as:

$$f(T,X(T)) = f(0,X(0)) + \int_0^T f_t(t,X(t))dt + \int_0^T f_x(t,X(t))\Delta(t)dW(t) \tag{129}$$

$$+ \int_0^T f_x(t,X(t))\theta(t)dt + \frac{1}{2}\int_0^T f_{xx}(t,X(t))\Delta(t)^2 dt \tag{130}$$

Proof. Proceeds very similar to theorem 5.1, using Taylor's theorem. □

Theorem 5.4. *Let $X(t)$ and $Y(t)$ be Ito processes and let $f(t,x,y)$ have all the necessary derivatives, then:*

$$df(t,X(t),Y(t)) = f_t(t,X(t),Y(t))dt + f_x(t,X(t),Y(t))dX(t) + f_y(t,X(t),Y(t))dY(t) \tag{131}$$

$$+ \frac{1}{2}f_{xx}(t,X(t),Y(t))dX(t)dX(t) + f_{xy}(t,X(t),Y(t))dX(t)dY(t) \tag{132}$$

$$+ \frac{1}{2}f_{yy}(t,X(t),Y(t))dY(t)dY(t) \tag{133}$$

Proof. Proceeds very similar to theorem 5.1, using Taylor's theorem. □

Corollary 5.4.1. *For $X(t)$ and $Y(t)$ Ito processes:*

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + dX(t)dY(t) \tag{134}$$

Proof. [2] Let $f(t,x,y) = xy$ in theorem 5.4. □

5.3 Applications

5.3.1 Geometric Brownian Motion

Geometric Brownian motion is a very useful stochastic process, commonly used to simulate stock prices. It is defined by the following differential equation:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{135}$$

μ and σ are referred to as the drift and volatility respectively.

We consider a solution $S(t) = f(t, W(t))$ and match coefficients[1]:

$$dS(t) = (f_t(t, W(t)) + \frac{1}{2}f_{xx}(t, W(t)))dt + f_x(t, W(t))dW(t) \tag{136}$$

Hence we require:

$$\mu f(t, W(t)) = f_t(t, W(t)) + \frac{1}{2}f_{xx}(t, W(t)) \tag{137}$$

$$\sigma f(t, W(t)) = f_x(t, W(t)) \tag{138}$$

The second equation immediately tells us $f(t, x) = exp(\sigma x + g(t))$ for some $g(t)$. Working with this we find:

$$f(t, x) = A\exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma x\} \tag{139}$$

Now we consider the expectation and variance of GBM. This requires $E[e^{\alpha X}]$, where $X \sim N(\mu, \sigma^2)$. This is calculated as $\exp\{\mu\alpha + \frac{1}{2}\sigma^2\alpha^2\}$. Hence:

$$E[S(t)] = S(0)\exp{(\mu - \frac{\sigma^2}{2})t}E[\exp{\sigma W(t)}] = e^{\mu t} \tag{140}$$

$$Var[S(t)] = E[S(t)^2] - E[S(t)]^2 = S(0)^2 e^{2\mu t}(e^{\sigma^2 t} - 1) \tag{141}$$

An important property is that GMB is log-normal distributed - this is easy to see given that $W(t)$ is normally distributed.

$$log(S(t)) \sim N(log(S(0)) + (\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t) \tag{142}$$

5.3.2 Levy's Characterisation Theorem

The following theorem is very useful when checking whether a stochastic process is Brownian motion.

Theorem 5.5. *Let $M(t)$ be a martingale w.r.t $\mathcal{F}(t)$ satisfying the following:*

- *$M(t)$ has continuous paths*
- $M(0) = 0$
- $[M, M](t) = t$ *a.s.*

Then $M(t)$ is Brownian motion.

Proof. [4] We consider the function $f(t, x) = e^{i\lambda x + \frac{1}{2}\lambda^2 t}$ and apply Ito's formula to $f(t, M(t))$ (we note that we can do this as in the proof of theorem 5.1 we only used the assumptions of this theorem):

$$df(t, M(t)) = \frac{1}{2}\lambda^2 e^{i\lambda x + \frac{1}{2}\lambda^2 t}dt + i\lambda e^{i\lambda x + \frac{1}{2}\lambda^2 t}dM(t) - \frac{1}{2}\lambda^2 e^{i\lambda x + \frac{1}{2}\lambda^2 t}d[M(t), M(t)] \tag{143}$$

Using the quadratic variation of $M(t)$ this simplifies to:

$$df(t, M(t)) = i\lambda e^{i\lambda x + \frac{1}{2}\lambda^2 t} dM(t) \tag{144}$$

Hence one can prove (similarly to how we did for standard Ito integrals) that $f(t, M(t))$ is martingale.

$$E[e^{i\lambda M(t) + \frac{1}{2}\lambda^2 t} \mid \mathcal{F}(s)] = e^{i\lambda M(s) + \frac{1}{2}\lambda^2 s} \tag{145}$$

$$E[e^{i\lambda (M(t) - M(s))} \mid \mathcal{F}(s)] = e^{-\frac{1}{2}\lambda^2 (t-s)} \tag{146}$$

Hence by moment generating functions we conclude that $M(t) - M(s) \sim N(0, t-s)$.

To conclude that $M(t)$ is Brownian motion we have to show that for $0 \leq t_1 < t_2 < ... < t_n \leq T$ the random variables $M(t_1) - M(0)$, $M(t_2) - M(t_1)$, ..., $M(Mt_n) - M(t_{n-1})$ are independent.

$$E[e^{i\lambda_1 (M(t_1) - M(0))} e^{i\lambda_2 (M(t_2) - M(t_1))} ... e^{i\lambda_n (M(t_n) - M(t_{n-1}))}] \tag{147}$$

$$= E[e^{iM(t_1)(\lambda_1 - \lambda_2)} e^{iM(t_2)(\lambda_2 - \lambda_3)} ... e^{iM(t_{n-1})\lambda_{n-1} - \lambda_n)} e^{iM(t_n)\lambda_n}] \tag{148}$$

$$= E[E[e^{iM(t_1)(\lambda_1 - \lambda_2)} e^{iM(t_2)(\lambda_2 - \lambda_3)} ... e^{iM(t_{n-1})(\lambda_{n-1} - \lambda_n)} e^{iM(t_n)\lambda_n} \mid \mathcal{F}(t_{n-1})]] \tag{149}$$

$$= E[e^{iM(t_1)(\lambda_1 - \lambda_2)} e^{iM(t_2)(\lambda_2 - \lambda_3)} ... e^{iM(t_{n-1})(\lambda_{n-1} - \lambda_n)} E[e^{iM(t_n)\lambda_n} \mid \mathcal{F}(t_{n-1})]] \tag{150}$$

$$= E[e^{iM(t_1)(\lambda_1 - \lambda_2)} e^{iM(t_2)(\lambda_2 - \lambda_3)} ... e^{iM(t_{n-1})(\lambda_{n-1} - \lambda_n)} e^{i\lambda M(t_{n-1})} e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})}] \tag{151}$$

$$= E[e^{iM(t_1)(\lambda_1 - \lambda_2)} e^{iM(t_2)(\lambda_2 - \lambda_3)} ... e^{iM(t_{n-1})\lambda_{n-1}}] e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} \tag{152}$$

$$\vdots \tag{153}$$

$$= e^{-\frac{1}{2}\lambda_1^2 (t_1)} e^{-\frac{1}{2}\lambda_2^2 (t_2 - t_1)} ... e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} \tag{154}$$

Hence $M(t)$ is Brownian motion. □

5.3.3 Black-Scholes PDE

Definition 5.3. (Law of One Price) If two assets have identical future cash flows then they should be priced the same.

In this section we present the beginning of a derivation of the Black-Scholes formula that would be most similar to the way Fischer Black and Myron Scholes derived their formula in their 1973 paper. We follow the presentation in [1]. This method does not make use of any risk neutral measures/pricing to avoid arbitrage and instead uses a brute force equation such that arbitrage is not possible. The main idea is that once constructs a continuously adjusting portfolio such that its movements match those of the option payoff and then uses the above definition to glean a price; we have 2 investment options in our portfolio, the first being going long/short in the stock and the second lending/borrowing from the money market at the risk free rate r. Let $V(t)$ denote the value of the portfolio at time t

$$V(t) = a(t)S(t) + b(t)\beta(t) \tag{155}$$

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{156}$$

$$d\beta(t) = r\beta(t)dt \tag{157}$$

where $a(t)$ is the number of shares (positive means a long position and negative a short position) and $b(t)$ is the amount we have lent(+ve) or borrowed (-ve) from the money market. If the derivative pays $h(S(T))$ at time T, then we require $V(T) = h(S(T))$. Another condition is that the portfolio must be self-financing (any change in $a(t)$ must be accompanied by a change in $b(t)$, no external flows of money). This is enforced by this equation:

$$dV(t) = a(t)dS(t) + b(t)d\beta(t) \tag{158}$$

Now we want to find $a(t)$ and $b(t)$; we assume $V(t) = f(t, S(t))$ where f is sufficiently differentiable. Then using Ito's formula:

$$dV(t) = f_t(t, S(t))dt + f_x(t, S(t))dS(t) + \frac{1}{2}f_{xx}(t, S(t))dS(t)dS(t) \tag{159}$$

$$dV(t) = \left[f_t(t, S(t)) + \mu f_x(t, S(t))S(t) + \frac{1}{2}\sigma^2 f_{xx}(t, S(t))S(t)^2\right]dt + \sigma f_x(t, S(t))S(t)dW(t) \tag{160}$$

From (261) we have:

$$dV(t) = [\mu a(t)S(t) + r\beta(t)b(t)]\,dt + \sigma a(t)S(t)dW(t) \tag{161}$$

So comparing coefficients we have:

$$a(t) = f_x(t, S(t)) \tag{162}$$

$$b(t) = \frac{1}{r\beta(t)}\left[f_t(t, S(t)) + \frac{\sigma^2}{2}(t, S(t))S(t)^2\right] \tag{163}$$

Now subbing into (259) and replacing $S(t)$ with x, we arrive at the famous Black-Scholes PDE:

$$f_t(t, x) + rxf_x(t, x) + \frac{\sigma^2}{2}f_{xx}(t, x)x^2 - rf(t, x) = 0 \tag{164}$$

$$f(T, x) = h(x) \tag{165}$$

One can solve this to find the Black-Scholes formula; we will derive this later using a different method.

6 Existence and Uniqueness of Solutions of SDEs

In this section we will explore just one theorem about the existence and uniqueness of the solution to an SDE of the form:

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t) \quad X(0) = x_0 \tag{166}$$

We follow the presentation in [1] and remark about the similarities to the Picard Lindelof theorem for ODEs.

Theorem 6.1. *If the coefficients of equation (166) satisfy the following two conditions:*

- $|\mu(t, x) - \mu(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \leq K|x - y|^2$ *(space-variable Lipschitz condition)*
- $|\mu(t, x)|^2 + |\sigma(t, x)|^2 \leq K(1 + |x|^2))$ *(spatial growth condition)*

Then there exists a continuous, adapted solution $X(t)$ which is uniformly bounded in $L^2(dP)$:

$$\sup_{0 \leq t \leq T} E[X(t)^2] < \infty \tag{167}$$

Additionally, if $X(t)$ and $Y(t)$ are continuous and $L^2(dP)$ bounded solutions then for all $t \in [0, T]$ $X(t) = Y(t)$ a.s.

The proof is presented in two halves: uniqueness and existence.

6.1 Uniqueness

We begin by supposing we have 2 solutions $X(t)$ and $Y(t)$:

$$X(t) - Y(t) = \int_0^t \mu(s, X(s)) - \mu(s, Y(s))ds + \int_0^t \sigma(s, X(s)) - \sigma(s, X(s))dW(s) \tag{168}$$

Now we can use $(u+v)^2 \leq 2u^2 + 2v^2$ - this is a very easy inequality to show using $(u-v)^2 \geq 0$.

$$E[|X(t) - Y(t)|^2] \leq 2E\left[\left(\int_0^t \mu(s, X(s)) - \mu(s, Y(s))ds\right)^2\right] + 2E\left[\left(\int_0^t \sigma(s, X(s)) - \sigma(s, Y(s))dW(s)\right)^2\right] \tag{169}$$

Now bounding the first term we can use Cauchy-Schwarz inequality $(\int_0^t f(t)g(t)dt)^2 \leq \int_0^t f^2(t)dt \int_0^t g^2(t)dt$ letting $g(t) = 1$, arriving at the bound:

$$2tE[\int_0^t |\mu(s, X(s)) - \mu(s, Y(s))|^2 dt] \tag{170}$$

For the second term we note it can be written $\|I(\sigma(s, X(s)) - \sigma(s, Y(s)), t)\|_{L^2(dP)}$ and so we would like to use Ito's isometry. We can easily verify $\sigma(s, X(s)) - \sigma(s, Y(s)) \in \mathcal{V}$ using the Lipschitz condition and that $X(t)$ and $Y(t)$ were bounded in $L^2(dP)$. Hence, the second term equals:

$$2E[\int_0^t |\sigma(s, X(s)) - \sigma(s, Y(s))|^2 dt] \tag{171}$$

Now using the Lipschitz condition again and defining $C = 2K\max(T, 1)$ we have the overall bound:

$$E[|X(t) - Y(t)|^2] \leq CE[\int_0^t |X(s) - Y(s)|^2 ds] = C\int_0^t E[|X(s) - Y(s)|^2]ds < \infty \tag{172}$$

We are almost done as if we define $g(t) = E[|X(t) - Y(t)|^2]$ then we have $g(t) \leq C\int_0^t g(s)ds$ for all $t \in [0, T]$. Hence we let $M = \sup\{g(t) : 0 \leq t \leq T\}$ and buy sucessive iterations arrive at:

$$g(t) \leq \frac{MC^n t^n}{n!} \tag{173}$$

This implies $g(t) = 0$ for all $t \in [0, T]$ as $n!$ grows very fast and so if we fix t we can use Markov's inequality which gives:

$$P(X(t) = Y(t)) = 1 \tag{174}$$

Through countability of the rationals and the fact that we have this for all $t \in [0, T]$:

$$P(X(t) = Y(t)t \in [0, T] \cap \mathbb{Q}) = 1 \tag{175}$$

We then conclude the result via the density of $\mathbb{Q}$ in $\mathbb{R}$ and the continuity of $X(t)$ and $Y(t)$[1].

6.2 Existence

As we alluded to earlier we use an ODE style technique here as as we formulate a Picard-like iterate:

$$X_{n+1}(t) = x_0 + \int_0^t \mu(s, X_n(s))ds + \int_0^t \sigma(s, X_n(s))dW(s) \tag{176}$$

Before we go steaming into the proof we have to check that the RHS remains defined throughout this iteration.

Lemma 6.2. *If $X_n(t) \in L^2(dP)$ for all $t \in [0,T]$, then:*

- $\mu(s, X_n(s)) \in L^2([0,T] \times \Omega)$
- $\sigma(s, X_n(s)) \in \mathcal{V}$
- $X_{n+1} \in L^2(dP)$ *for all* $t \in [0,T]$

Proof. [1] Using the spatial growth condition it is easy to show the first two points. We we know by the L^2 boundedness of $X_n(t)$ that there exists B such that:

$$\sup_{t \in [0.T]} E[|X_n(t)|^2] = B < \infty \tag{177}$$

Then:

$$E\left[\int_0^T \mu(s, X_n(s))^2 ds\right] \leq TK(1+B) < \infty \tag{178}$$

$$E\left[\int_0^T \sigma(s, X_n(s))^2 ds\right] \leq TK(1+B) < \infty \tag{179}$$

For the last point we enlist Cauchy-Schwarz and Ito's isometry which we are now clear to use:

$$E\left[\left(\int_0^T \mu(s, X_n(s) ds\right)^2\right] \leq T^2 K(1+B) < \infty \tag{180}$$

$$E\left[\left(\int_0^T \sigma(s, X_n(s)) dW(s)\right)^2\right] = E\left[\int_0^T \sigma(s, X_n(s))^2 ds\right] \leq TK(1+B) < \infty \tag{181}$$

□

So we have proven that the iteration makes sense, next we prove convergence. This is a multiple step process: first we will prove convergence of $X_n(t)$ to a continuous $X(t)$ on a full probability set; second we will show this convergence in $L^2(dP)$ and show the limit is uniformly bounded; and last we will look at both sides of equation (176) and ensure that for our limit we have:

$$X(t) = x_0 + \int_0^t \mu(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW(s) \quad \text{a.s. } \forall t \in [0,T] \tag{182}$$

The following lemma will help us on our way.

Lemma 6.3. *Under the conditions of theorem 6.1 and using the iteration (176), there exists C such that:*

$$E\left[\sup_{0 \leq s \leq t} |X_{n+1}(s) - X_n(s)|^2\right] \leq C \int_0^t E(|X_n(s) - X_{n-1}(s)|^2) ds \tag{183}$$

Proof. [1] We can write $X_{n+1}(s) - X_n(s)$ as:

$$\int_0^s \mu(\alpha, X_n(\alpha)) - \mu(\alpha, X_{n-1}(\alpha)) d\alpha + \int_0^s \sigma(\alpha, X_n(\alpha)) - \sigma(\alpha, X_{n-1}(\alpha)) dW(\alpha) \tag{184}$$

We denote the LHS term $D(s)$ and the RHS $M(s)$. We note from lemma 6.2 it follows that $M(s)$ is a martingale. As we did for uniqueness we use the inequality $(u+v)^2 \leq 2u^2 + 2v^2$:

$$\sup_{0 \leq s \leq t} |X_{n+1}(s) - X_n(s)|^2 = \sup_{0 \leq s \leq t} |M(s) + D(s)|^2 \leq 2 \sup_{0 \leq s \leq t} M(s)^2 + 2 \sup_{0 \leq s \leq t} D(s)^2 \tag{185}$$

We now want to handle these terms to eventually use the Lipschitz condition to conclude. First using theorem 3.1 and Ito isometry we have:

$$E[\sup_{0\leq s\leq t} M(s)^2] \leq 4E[M(t)^2] = 4E\left[\int_0^t |\sigma(s, X_n(s)) - \sigma(s, X_{n-1}(s))|^2 ds\right] \tag{186}$$

Then using Cauchy-Schwarz:

$$\sup_{0\leq s\leq t} D(s)^2 \leq \sup_{0\leq s\leq t} s\int_0^s (\mu(\alpha, X_n(\alpha)) - \mu(\alpha, X_{n-1}(\alpha)))^2 d\alpha = t\int_0^t |\mu(s, X_n(s)) - \mu(s, X_{n-1}(s))|^2 ds \tag{187}$$

Putting everything together we define $C := 8K\max(1, T)$:

$$\begin{aligned} E[\sup_{0\leq s\leq t} |X_{n+1}(s) - X_n(s)|^2] &\leq 8E\left[\int_0^t |\sigma(s, X_n(s)) - \sigma(s, X_{n-1}(s))|^2 ds\right] &(188)\\ &+ 2tE\left[\int_0^t |\mu(s, X_n(s)) - \mu(s, X_{n-1}(s))|^2 ds\right] &(189)\\ &\leq CE\left[\int_0^t |X_n(s) - X_{n-1}(s)|^2 ds\right] &(190)\\ &= C\int_0^t E(|X_n(s) - X_{n-1}(s)|^2) ds &(191)\end{aligned}$$

□

Now very similarly to how we proved uniqueness we define $g_n(t) = E\left[\sup_{0\leq s\leq t}|X_{n+1}(s) - X_n(s)|^2\right]$ and so the by the above lemma:

$$g_n(t) \leq C\int_0^t g_{n-1}(s) ds \tag{192}$$

Now if we look at $g_0(t) = E\left[\sup_{0\leq s\leq t}|X_1(s) - X_0(s)|^2\right]$ we know this is bounded above for all $t \in [0, T]$ and so as before $g_n(t) \leq \frac{Mc^n t^n}{n!}$. Then by Markov's inequality we have:

$$P\left(\sup_{0\leq s\leq t} |X_{n+1}(s) - X_n(s)| \geq 2^{-n}\right) \leq P\left(\sup_{0\leq s\leq t} |X_{n+1}(s) - X_n(s)|^2 \geq 2^{-2n}\right) \leq \frac{MC^n T^n 2^{2n}}{n!} \tag{193}$$

Now if we consider the event $A_n = \{\sup_{0\leq s\leq t}|X_{n+1}(s) - X_n(s)| \geq 2^{-n}\}$ then by the above probability and the Borel-Cantelli lemma we know the event A_n cannot happen infinitely often, meaning with probability 1 $\exists N$ s.t. $\forall n \geq N$ A_n does not occur. In other words, there exists a full probability set Ω_0 where we have a Cauchy sequence under the supremum norm on the space of continuous functions on $[0, T]$. Note that the summability required for Borel-Cantelli is easily shown via the ratio test and demonstrating Cauchy on Ω_0 is trivial via the triangle inequality.

We can now conclude as the supremum norm on $C[0, T]$ is a complete space and so Cauchy implies convergent and so there exists a continuous $X(t)$ that $X_n(t)$ converges uniformly to.

So far we have shown that the 'Picard' iterates make sense and that there exists a continuous limit for them. Now we need to check that this limit actually solves the SDE and that is is uniformly $L^2(dP)$ bounded.

The second of these is actually very easy as we can show convergence in $L^2(dP)$ for all $t \in [0, T]$ using $g_n(t)$. By definition we have:

$$\|X_{n+1}(t) - X_n(t)\|_{L^2(dP)} \leq g_n(T)^{\frac{1}{2}} \text{ for all } t \in [0, T] \tag{194}$$

The summability of $g_n(T)^{\frac{1}{2}}$ (easily shown by ratio test) means we have a Cauchy sequence and by completeness we know we have convergence to a limit. By uniqueness of limits, we know this must be the continuous $X(t)$ process we found earlier.

Showing uniformly bounded in $L^2(dP)$ is done by the triangle inequality and the summability from before:

$$\|X(t)\|_{L^2(dP)} \leq g_0(T)^{\frac{1}{2}} + \sum_{n=0}^{\infty} g_n(T)^{\frac{1}{2}} \text{ for all } t \in [0,T] \tag{195}$$

Now we will look particularly at the RHS of equation (176). Looking at equations (188-191) it is immediately clear we have:

$$\|\mu(s, X_n(s)) - \mu(s, X_{n-1}(s))\|_{L^2(dP\times dt)} \leq C^{\frac{1}{2}} T^{\frac{1}{2}} g_{n-1}(T)^{\frac{1}{2}} \tag{196}$$
$$\|\sigma(s, X_n(s)) - \sigma(s, X_{n-1}(s))\|_{L^2(dP\times dt)} \leq C^{\frac{1}{2}} T^{\frac{1}{2}} g_{n-1}(T)^{\frac{1}{2}} \tag{197}$$

where we used that $g_{n-1}(t)$ is monotonically increasing.

The summability of $g_n(T)^{\frac{1}{2}}$ immediately gives $\forall t \in [0,T]$:

$$\int_0^t \mu(s, X_n(s))ds \xrightarrow{L^2(dP)} \int_0^t \mu(s, X(s))ds \tag{198}$$

Using Ito's isometry, we have the following also:

$$\int_0^t \sigma(s, X_n(s))dW(s) \xrightarrow{L^2(dP)} \int_0^t \sigma(s, X(s))dW(s) \tag{199}$$

Finally we can conclude the proof by bringing everything together. We consider the LHS and RHS of the equation:

$$X_{n+1}(t) = x_0 + \int_0^t \mu(s, X_n(s))ds + \int_0^t \sigma(s, X_n(s))dW(s) \tag{200}$$

For the LHS we have that with probability 1 it converges uniformly on $[0,T]$. The RHS is more difficult - we use this flow:

$$\xrightarrow{L^2(dP)} \implies \xrightarrow{L^1(dP)} \implies \xrightarrow{\text{probability}} \implies \exists \text{ subsequence } n_j \xrightarrow{\text{a.s.}} \tag{201}$$

So for both integrals on RHS if we fix t there exists a subsequence such that they convergence almost-surely. Then via the countability of the rationals we have the same result for $t \in [0,T] \cap \mathbb{Q}$. Taking the limit we find there exists Ω_*, $P(\Omega_*) = 1$ such that:

$$X(t) = x_0 + \int_0^t \mu(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s) \ \forall t \in [0,T] \cap \mathbb{Q} \ \forall \omega \in \Omega_* \tag{202}$$

Continuity completes the proof extending to $t \in [0,T]$[1].

7 Risk Neutral

Definition 7.1. (Arbitrage) A trading strategy is an arbitrage opportunity if its value process satisfies:

- $V(0) \leq 0$
- $P(V(t) \geq 0) = 1$
- $P(V(t) > 0) > 0$

In this section we introduce the fundamental notion of risk neutral pricing and the risk neutral measure. We will present a mathematical argument for it and also a more intuitive one about the risk aversion of financial markets in general. We begin the discrete setting.

7.1 Discrete

We first consider an example which comes in the form of a one period binomial model; an asset has value x at time 0 and with probability p at time 1 it has value ux, or with probability $q = 1 - p$ value dx, where $d < 1 < u$. Now the question is how do we price this asset, given a risk free rate of r?

An intuitive answer might be:

$$\frac{1}{1+r}(p \times ux + q \times dx) \tag{203}$$

where we have take the expectation of the value at time 1 and discounted it at the risk free rate. Unfortunately, this price would certainly be above the market price because the market and it participants are risk averse, meaning they feel the downside more than the upside and so would not pay this price. However, this motivates the idea of a risk neutral probability as we know there must exist $\tilde{p}$ and $\tilde{q}$ such that this price is correct. One could explain this idea of risk aversion by the fact that there exists a risk free rate r and this will be very important going forward.

In the following lemma we will establish conditions to avoid arbitrage.

Lemma 7.1. *No arbitrage iff* $d \leq (1+r) \leq u$

Proof. [5] This is very easy to see with the following 2 portfolios: First, we can short the asset and save x at the risk free rate. This means at time 1 the portfolio is worth one of the following:

$$\begin{aligned} x(1+r) &- xd \quad (204) \\ &- xu \quad (205) \end{aligned}$$

It is then clear given $u > d$ that there exists arbitrage if $(1+r) > u$

Second, we can borrow x and by the asset. Then at time 1 we have:

$$\begin{aligned} -x(1+r) &+ xd \quad (206) \\ &+ xu \quad (207) \end{aligned}$$

Now it is clear that there exists arbitrage if $d > (1+r)$

To show that $d < (1+r) < u$ implies no arbitrage we can consider an arbitrary portfolio $f(a,b) = ax + b$, where a can be positive or negative and is the number the asset we buy/short and b can be positive or negative and is the amount we save/borrow. We begin with $ax = -b$ and consider what the portfolio is worth at time 1:

$$\begin{aligned} aux + b(1+r) &= ax(u - (1+r)) \quad \text{with probability p} \quad (208) \\ adx + b(1+r) &= ax(d - (1+r)) \quad \text{with probability q} \quad (209) \\ & \quad (210) \end{aligned}$$

Then if we consider the two cases $a > 0$ and $a < 0$ we can see that the inequality guarantees that the outcomes cannot both be positive, hence no arbitrage. $\square$

Using this inequality we continue the idea from earlier that there exists $\tilde{p}$ and $\tilde{q}$ such that $(1+r) = \tilde{p}u + \tilde{q}d$. Now using these new probabilities we have the following fundamental result:

$$\frac{1}{1+r}(\tilde{p} \times xu + \tilde{q} \times xd) = E^Q[S_1] = x \tag{211}$$

We denote with a Q above the new probability measure which we explore now for the continuous setting.

7.2 Continous

Using a non negative random variable with expectation 1 it very easy to define a new, equivalent probability measure. Given the triplet $(\Omega, \mathcal{F}(t), P)$ and random variable Z such that $E[Z] = 1$, we can define Q:

$$Q(A) = \int_A Z dP \quad \text{for } A \in \mathcal{F} \tag{212}$$

Using Radon-Nikodym notation we have $Z = \frac{dQ}{dP}$ and using this theorem it follows Q is absolutely continuous with respect to P. Proving P is absolutely continuous wrt. Q, and so their equivalence, is easily given by the positive expectation of Z and that it is non negative.

It is useful prove a few results for a process which can be defined from our random variable Z. For an arbitrary filtration $\mathcal{F}(t)$, a Radon-Nikodym process is defined as:

$$Z(t) = E[Z \mid \mathcal{F}(t)] \tag{213}$$

Lemma 7.2. *For a random variable Y which is $\mathcal{F}(t)$ measurable and where $\tilde{E}$ denotes the expectation under Q we have:*

- $\tilde{E}[Y] = E[YZ(t)]$
- $\tilde{E}[Y \mid \mathcal{F}(s)] = \frac{1}{Z(s)} E[YZ(t) \mid \mathcal{F}(s)]$

Proof. [2]

$$\tilde{E}[Y] = E[YZ] = E[E[YZ \mid \mathcal{F}(t)]] = E[YZ(t)] \tag{214}$$

For the second point we use the definition of conditional expectation from A.1, meaning we have to prove:

$$\int_A Y dQ = \int_A \frac{1}{Z(s)} E[YZ(t) \mid \mathcal{F}(s)] dQ \quad \text{for } A \in \mathcal{F}(s) \tag{215}$$

We work with the RHS and use what we just proved:

$$\tilde{E}[\mathbb{1}_A \frac{1}{Z(s)} E[YZ(t) \mid \mathcal{F}(s)]] = E[\mathbb{1}_A E[YZ(t) \mid \mathcal{F}(s)]] \tag{216}$$

$$= E[\mathbb{1}_A YZ(t)] \tag{217}$$

$$= \tilde{E}[\mathbb{1}_A Y] \tag{218}$$

$$= \int_A Y dQ \tag{219}$$

$\square$

The following theorem helps us construct a random variable Z to define an equivalent probability measure and also understand Brownian motion in this space.

Theorem 7.3. *(Girsanov Theorem) In the probability space $(\Omega, \mathcal{F}, P)$, let $W(t)$ on $[0, T]$ be Brownian motion with standard filtration $\mathcal{F}(t)$. Let $\Theta(t) \in \mathcal{V}$ be an adapted process, then we can define:*

$$Z(t) = \exp\{-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta^2(u)du\} \tag{220}$$

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du \tag{221}$$

Now if we let $Z(T) = Z$ then $E[Z] = 1$ and we can define a new measure Q as described above and with respect to Q, $\tilde{W}(t)$ is Brownian motion, given we assume:

$$E\left[\int_0^T \Theta^2(u)Z^2(u)du\right] < \infty \tag{222}$$

Proof. [2] We can use Levy's theorem to prove $\tilde{W}(t)$ is Brownian motion. Firstly, it is clear that $[\tilde{W}, \tilde{W}](t) = t$ as this follows from very similar ideas as lemma 5.2. We clearly have continuity and $\tilde{W}(0) = 0$, meaning the last thing we must check is that it is a martingale wrt. Q. Define X(t):

$$X(t) = -\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta^2(u)du \tag{223}$$

Using Ito's formula:

$$\begin{aligned} dZ(t) &= Z(t)dX(t) + \frac{1}{2}Z(t)dX(t)dX(t) && (224)\\ &= Z(t)[-\Theta(t)dW(t) - \frac{1}{2}\Theta^2(t)dt] + \frac{1}{2}Z(t)\Theta^2(t)dt && (225)\\ &= -Z(t)\Theta(t)dW(t) && (226) \end{aligned}$$

Hence given the assumption in the theorem we have $Z(t)$ is a martingale, which immediately gives $E[Z] = 1$ given that $Z(0) = 1$.

Given $Z(t)$ is martingale, we have $E[Z \mid \mathcal{F}(t)] = Z(t)$ meaning it is a Radon-Nikodym process. Looking at the second point in lemma 7.2, it is clear if we show $\tilde{W}(t)Z(t)$ is martingale w.r.t P then we are done.

$$\begin{aligned} d(\tilde{W}(t)Z(t)) &= \tilde{W}(t)dZ(t) + Z(t)d\tilde{W}(t) + dZ(t)d\tilde{W}(t) && (227)\\ &= \tilde{W}(t)(-Z(t)\Theta(t)dW(t)) + Z(t)(dW(t) + \Theta(t)dt) && (228)\\ &+ (-Z(t)\Theta(t)dW(t))(dW(t) + \Theta(t)dt) && (229)\\ &= (-\tilde{W}(t)Z(t)\Theta(t) + Z(t))dW(t) && (230) \end{aligned}$$

Hence, under mild integrability conditions, we have a martingale under P which completes the proof by lemma 7.2. □

In the continuous setting we have the following theorem regarding avoiding arbitrage:

Theorem 7.4. *(Fundamental Theorem of Asset Pricing) Market is arbitrage free if and only if there exists an equivalent local martingale measure.*

We can consider a risk-free rate given by $r(t)$ and an asset $S(t)$:

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t) \tag{231}$$
$$dB(t) = -r(t)B(t)dt \tag{232}$$

An equivalent local martingale measure means there exists an measure which is equivalent to P such that $S(t)B(t)$ is a local martingale.

The proof of theorem 7.4 is challenging but we will look at how to formulate the local martingale measure and how using it avoids arbitrage.

One can show that the following satisfy the above SDEs:

$$S(t) = S(0)\exp\left(\int_0^t \sigma(s)dW(s) + \int_0^t \mu(s) - \frac{1}{2}\sigma^2(s)ds\right) \tag{233}$$
$$B(t) = \exp\left(-\int_0^t r(s)ds\right) \tag{234}$$

We can consider $d(S(t)B(t))$ and use the previous theory about stochastic integrals to see how to make $S(t)B(t)$ a local martingale.

$$\begin{aligned} d(S(t)B(t)) &= S(t)dB(t) + B(t)dS(t) + dS(t)dB(t) && (235)\\ &= -S(t)r(t)B(t)dt + B(t)(\mu(t)S(t)dt + \sigma(t)S(t)dW(t)) && (236)\\ &= S(t)B(t)dt(\mu(t) - r(t)) + S(t)B(t)\sigma(t)dW(t) && (237)\\ & && (238)\end{aligned}$$

If we look back at Girsanov's theorem we can see how the new measure Q can help us change the drift term as $\tilde{W}(t)$ can absorb it. Hence if we define $\Theta(t)$ from this theorem as $\frac{\mu(t)-r(t)}{\sigma(t)}$ then we can remove the drift term.

$$\begin{aligned} d(S(t)B(t)) &= S(t)B(t)\sigma(t)(dW(t) + \Theta(t)) && (239)\\ &= S(t)B(t)\sigma(t)d\tilde{W}(t) && (240)\end{aligned}$$

Therefore, under Q $S(t)B(t)$ has no drift term and so given $S(t)B(t)\sigma(t) \in \mathcal{L}^2_{LOC}$ we can conclude $S(t)B(t)$ is a local martingale.

Now let $S_i(t)B(t)$ represent the present value of an asset i so we can build a portfolio $V(t,h)$ which contains these assets weighted by $h(t) = [h_1(t), h_2(t), ..., h_n(t)]$, where each $h_i(t)$ is an adapted process. Now we consider an arbitrage opportunity where we have $P(V(T,h) \geq 0) = 1$ and $P(V(T,h) > 0) > 0$, meaning using the risk-neutral measure we must show $V(0,h) > 0$ to prove it avoids arbitrage.

Firstly P and Q are equivalent probability measures meaning $Q(V(T,h) \geq 0) = 1$ and $Q(V(T,h) > 0) > 0$ and under Q we have:

$$dV(t,h) = \sum_{i=1}^{n} h_i(t)S_i(t)B(t)\sigma_i(t)d\tilde{W}(t) \tag{241}$$

Hence under mild integrability constraints we conclude $V(t,h)$ is a martingale and so $V(0,h) = E^Q[V(T,h)] > 0$, which implies no arbitrage[5].

7.3 Derivative Pricing via Risk-Neutral Measure

In this section we will assume that the risk free rate r, the rate of return μ and volatility σ are all constants. Hence we have the SDE for the asset $S(t)$:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{242}$$

Utilising the equivalent measure Q we can make the substitution from (154) to get the SDE under Q:

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t) \tag{243}$$

We remark how the risk-neutral measure has changed expected rate of return from μ to r and this motivates its name in that it brings everything down to the same level.

For a derivative which pays out $f(S(T))$ at time T we have the following formula for the price at time 0:

$$P(0) = e^{-rT}E^Q[f(S(T))] \tag{244}$$

$$= e^{-rT}\int_{-\infty}^{\infty} f\left(S(0)\exp\left((r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}x\right)\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \tag{245}$$

Now for a European call we know $f(S(T)) = (S(T)-K)^+$ so subbing this in we immediately note a large portion where the integral is zero. Let x^* be such that:

$$S(0)\exp\left((r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}x^*\right) = K \tag{246}$$

$$\tag{247}$$

Hence:

$$E^Q[(S(T)-K)^+] = \int_{x^*}^{\infty}\left(S(0)\exp\left((r-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}x\right) - K\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \tag{248}$$

$$= -K(1-\Phi(x^*)) + \frac{S(0)e^{rT}}{\sqrt{2\pi}}\int_{x^*}^{\infty}\exp(-\frac{x^2}{2}+\sigma\sqrt{T}x-\frac{\sigma^2}{2}T)dx \tag{249}$$

$$= -K(1-\Phi(x^*)) + \frac{S(0)e^{rT}}{\sqrt{2\pi}}\int_{x^*}^{\infty}\exp\left(-\frac{(x-\sigma\sqrt{T})^2}{2}\right)dx \tag{250}$$

$$= -K(1-\Phi(x^*)) + \frac{S(0)e^{rT}}{\sqrt{2\pi}}\int_{x^*-\sigma\sqrt{T}}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx \tag{251}$$

$$= -K(1-\Phi(x^*)) + S(0)e^{rT}(1-\Phi(x^*-\sigma\sqrt{T})) \tag{252}$$

After some rearrangements and using the symmetry of the normal distribution we land upon the familiar Black-Scholes formula [6]:

$$S(0)\Phi(d_1) - e^{-rT}K\Phi(d_2) \tag{253}$$

$$\tag{254}$$

where

$$d_1 = \frac{log(\frac{S(0)}{K}) + (r+\frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \tag{255}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{256}$$

Remark. We note that we have implicitly used some Markov assumptions by assuming that one has a formula for the price of an option at a certain time, which is only a function of the information at that time. This is above board we just haven't covered it here.

8 Stochastic Volatility

In this section we explore how one can address one of the big assumptions in the Black Scholes model: constant volatility. We will first introduce a general framework and then hone in on a specific example known as the Heston model.

8.1 General pricing model

Before we begin we specify some notation for when we are working with multiple Brownian motions:

$$\mathbf{W}(t) = (W_1(t), W_2(t), ..., W_d(t)) \quad \text{d-dimensional Brownian motion} \tag{257}$$
$$\mathbf{\Theta}(t) = (\Theta_1(t), \Theta_2(t), ..., \Theta_d(t)) \quad \text{d-dimensional adapted process} \tag{258}$$

We note that in this setting an adapted process is measurable with respect to $\mathcal{F}(t)$, which is the filtration such that each $W_i(t)$ is $\mathcal{F}(t)$ measurable. We have the following calculations:

$$\int_0^t \mathbf{\Theta}(\mathbf{u})d\mathbf{W}(\mathbf{u}) = \sum_{i=1}^{d} \int_0^t \Theta_i(u)dW_i(u) \tag{259}$$
$$\int_0^t \mathbf{\Theta}(\mathbf{u})du = (\int_0^t \Theta_1(u)du, ..., \int_0^t \Theta_d(u)du) \tag{260}$$
$$\int_0^t \|\mathbf{\Theta}(\mathbf{u})\|^2 du = \int_0^t \sum_{i=1}^{d} \Theta_i^2(u)du \tag{261}$$

In this section we follow [7] closely. In the following setup, we will have 2 independent Brownian motions $W(t)$ and $Z(t)$ and formulate a third $\hat{Z}(t)$:

$$\hat{Z}(t) = \rho W(t) + \sqrt{1-\rho^2}Z(t) \quad \rho \in [-1, 1] \tag{262}$$

This gives Brownian motion such that $W(t)$ and $\hat{Z}(t)$ are correlated.

$$Cov[W(t), \hat{Z}(t)] = E[W(t)\hat{Z}(t)] = E[\rho W(t)^2 + \sqrt{1-\rho^2}Z(t)W(t)] \tag{263}$$
$$= \rho Var[W(t)] + \sqrt{1-\rho^2}Cov[W(t), Z(t)] \tag{264}$$
$$= \rho t \tag{265}$$

Hence they have correlation ρ. As we formulate the pricing model we will see how this correlation can connect the asset price and its volatility.

So we begin with the SDE for the stock price $S(t)$:

$$dS(t) = \mu S(t)dt + \sigma(t)S(t)dW(t) \tag{266}$$

Then $\sigma(t) = f(Y(t))$ where $Y(t)$ is an Ornstein-Uhlenbeck process:

$$dY(t) = \alpha(m - Y(t))dt + \beta(t)d\hat{Z}(t) \tag{267}$$

As before, we want to use risk neutral measures to ensure no-arbitrage. To do this we employ a multi-dimensional Girsanov theorem:

Theorem 8.1. *Let $\mathbf{\Theta}(t)$ be a d-dimensional adapted process and $\mathbf{W}(t)$ a d-dimensional Brownian motion on $[0, T]$. Then define:*

$$Z(t) = \exp\{-\int_0^t \mathbf{\Theta}(u) d\mathbf{W}(u) - \frac{1}{2}\int_0^t \|\mathbf{\Theta}(\mathbf{u})\|^2\} \tag{268}$$

$$\tilde{\mathbf{W}}(t) = \mathbf{W}(t) + \int_0^t \mathbf{\Theta}(\mathbf{u}) du \tag{269}$$

Assume that:

$$E[\int_0^T \|\mathbf{\Theta}(u)\|^2 Z(u)^2] \infty \tag{270}$$

Then if we let $Z = Z(T)$ we have $E[Z] = 1$ and defining the measure Q in the usual way $\tilde{W}(t)$ is d-dimensional Brownian motion under Q.

We won't prove this here, but one can do it very similarly to the one-dimensional version via a multidimensional Levy Characterisation.

Applying this theorem to our setup we let $d = 2$, $(W_1(t), W_2(t)) = (W(t), Z(t))$,$(\tilde{W_1}(t), \tilde{W_2}(t)) = (\tilde{W}(t), \tilde{Z}(t))$ and define $\mathbf{\Theta}(t) = (\frac{\mu - r}{\sigma(t)}, \gamma(t))$, where $\gamma(t)$ is arbitrary for now. We know from earlier that under Q this makes $D(t)S(t)$ a martingale and that:

$$dS(t) = rS(t)dt + \sigma(t)S(t)d\tilde{W}(t) \tag{271}$$

Next we explore how $Y(t)$ works under Q, before doing this we will look back at what exactly Q does to the SDE describing $S(t)$. We know that the drift term changes and we can understand that as a risk premium which equals $\mu - r$, which the extra an investor demands for taking on the risk in the first place. Equally, now we have stochastic volatility we can expect a volatility risk premium which for now we will say equals $\lambda(t, S(t), Y(t))$ [8].

We can sub in to $dY(t)$ as follows:

$$\begin{aligned} dY(t) &= \alpha(m - Y(t))dt + \beta(t)d\hat{Z}(t) &(272)\\ &= \alpha(m - Y(t))dt + \beta(t)(\rho dW(t) + \sqrt{1-\rho^2}dZ(t)) &(273)\\ &= \alpha(m - Y(t))dt + \beta(t)(\rho[d\tilde{W}(t) - \frac{\mu - r}{\sigma(t)}dt] + \sqrt{1-\rho^2}[d\tilde{Z}(t) - \gamma(t)dt]) &(274)\\ &= \alpha(m - Y(t))dt + \beta(t)\hat{\tilde{Z}}(t) + (-\rho\frac{\mu - r}{\sigma(t)} - \sqrt{1-\rho^2}\gamma(t))\beta(t)dt &(275) \end{aligned}$$

where $\hat{\tilde{Z}}(t) = \rho d\tilde{W}(t) + \sqrt{1-\rho^2}d\tilde{Z}(t)$.

Hence to have $\lambda(t, S(t), Y(t))$ as the volatility risk premium we define it to equal:

$$(\rho\frac{\mu - r}{\sigma(t)} + \sqrt{1-\rho^2}\gamma(t))\beta(t) \tag{276}$$

The exact formulation of the volatility risk premium has to be theorised via other means in order to describe how the market responds to volatility. We will see a more concrete example next as we explore the Heston model.

8.2 Heston Model

In this section, we will cover a bit of Heston's 1993 paper [9], particularly the equations under the risk-neutral measure and Heston's choice for the volatility risk premium.

The Heston model has the following governing equations:

$$dS(t) = \mu S(t)dt + \sqrt{\nu(t)}S(t)dW(t) \tag{277}$$
$$d\sqrt{\nu(t)} = -\beta\sqrt{\nu(t)}dt + \delta d\hat{Z}(t) \tag{278}$$

where $W(t)$ and $\hat{Z}(t)$ are as in the general model above.

It is a very easy exercise using Ito's Formula to get:

$$d\nu(t) = (\delta^2 - 2\beta\nu(t)) + 2\sqrt{\nu(t)}d\hat{Z}(t) \tag{279}$$
$$= \kappa(\theta - \nu(t))dt + \xi\sqrt{\nu(t)}d\hat{Z}(t) \tag{280}$$

where we have adjusted constant to simplify things.

Comparing to the general model, we have $f(x) = \sqrt{x}$ and $\beta(t) = \xi\sqrt{\nu(t)}$.

Using theorem 8.1, as we did for the general model, we find the risk neutral dynamics under probability measure defined as before. In Heston's 1993 paper, he decides that the volatility risk premium should be $\lambda\nu(t)$, where λ is then a parameter. Hence under Q we have:

$$d\nu(t) = (\kappa^P(\theta^P - \nu(t)) - \lambda\nu(t))dt + \xi\sqrt{\nu(t)}d\hat{\hat{Z}}(t) \tag{281}$$
$$= (\kappa^P + \lambda)(\frac{\kappa^P\theta^P}{\kappa^P + \lambda} - \nu(t)) + \xi\sqrt{\nu(t)}d\hat{\hat{Z}}(t) \tag{282}$$
$$= \kappa^Q(\theta^Q - \nu(t)) + \xi\sqrt{\nu(t)}d\hat{\hat{Z}}(t) \tag{283}$$

where κ^Q and θ^Q are the new parameter values under Q, note that ξ does not change. Hence to summarise, under Q, we have:

$$dS(t) = rS(t)dt + \sqrt{\nu(t)}S(t)d\tilde{W}(t) \tag{284}$$
$$d\nu(t) = \kappa^Q(\theta^Q - \nu(t)) + \xi\sqrt{\nu(t)}d\hat{\hat{Z}}(t) \tag{285}$$

where $\tilde{W}(t)$ and $\hat{\hat{Z}}(t)$ are Brownian motions with correlation ρ and r is the risk free rate.

8.3 Simulation

One of the main reasons that the Heston model is so popular is that there does exist a solution for pricing European options using it - this is what Heston presented in his paper [9]. We will not cover this here as the derivation is complicated. Instead we will consider simulation ideas for the model.

We follow the presentation of the Milstein scheme as produced in [10]. Consider:

$$dX(t) = \alpha(X(t))dt + \beta(X(t))dW(t) \tag{286}$$

This gives rise to the integral equation:

$$X(t + \Delta t) = X(t) + \int_t^{t+\Delta t} \alpha(X(s))ds + \int_t^{t+\Delta t} \beta(X(s))dW(s) \tag{287}$$

Using left-point rule, you immediately get what's known as Euler discretisation:

$$X(t + \Delta t) = X(t) + \alpha(X(t))\Delta t + \beta(X(t))\sqrt{\Delta t}Z \tag{288}$$

where $Z \sim N(0,1)$. The Milstein scheme is an extension of this as instead of using the left-point rule immediately we use Ito's formula to 'get more information' about the movement of α and β.

$$d\alpha(X(t)) = [\alpha'(X(t))\alpha(X(t)) + \frac{1}{2}\alpha''(X(t))\beta(X(t))^2]dt + \alpha'(X(t))\beta(X(t))dW(t)) \tag{289}$$

$$d\beta(X(t)) = [\beta'(X(t))\alpha(X(t)) + \frac{1}{2}\beta''(X(t))\beta(X(t))^2]dt + \beta'(X(t))\beta(X(t))dW(t) \tag{290}$$

Taking the integral equations for the above we can sub into equation (287) and if we discount higher order terms, we get for $\tau \in [t, t+\Delta t]$:

$$X(t+\Delta t) = X(t) + \alpha(X(t))\Delta t + \beta(X(t))\Delta t + \int_t^{t+\Delta t}\int_t^{\tau} \beta'(X(s))\beta(X(s))dW(s)dW(\tau) \tag{291}$$

We can then apply left-point rule to this extra term:

$$= \beta'(X(t))\beta(X(t))\int_t^{t+\Delta t}\int_t^{\tau} dW(s)dW(\tau) \tag{292}$$

$$= \beta'(X(t))\beta(X(t))\int_t^{t+\Delta t} W(\tau) - W(t)dW(\tau) \tag{293}$$

$$= \beta'(X(t))\beta(X(t))(\frac{1}{2}W(t+\Delta t)^2 + \frac{1}{2}W(t)^2 - W(t+\Delta t)W(t) - \frac{1}{2}t + \Delta t) \tag{294}$$

$$= \frac{1}{2}\beta'(X(t))\beta(X(t))[(W(t+\Delta t) - W(t))^2 - \Delta t] \tag{295}$$

Hence we finally conclude:

$$X(t+\Delta t) = X(t) + \alpha(X(t))\Delta t + \beta(X(t))\Delta t + \frac{1}{2}\beta'(X(t))\beta(X(t))(Z^2-1)\Delta t \tag{296}$$

Now applying this to the Heston model, first for $\nu(t)$:

$$\nu(t+\Delta t) = \nu(t) + \kappa(\theta - \nu(t))\Delta t + \xi\sqrt{\nu(t)}\sqrt{\Delta t}Z + \frac{1}{4}\xi^2(Z^2-1)\Delta t \tag{297}$$

$$= (\sqrt{\nu(t)} + \frac{1}{2}\xi\sqrt{\Delta t}Z)^2 + \kappa(\theta - \nu(t))\Delta t - \frac{1}{4}\xi^2\Delta t \tag{298}$$

We can now explain why this is advantageous - given we use $\sqrt{\nu(t)}$ in the SDE describing the stock price - it is important that $\nu(t)$ remains non-negative. However, this is not always the case with the Milstein scheme, but the frequency at which it becomes negative is much lower than the standard Euler discretisation [11]. To address the rare occurence when $\nu(t)$ does become negative we will use $\max(\nu(t), 0)$.

We note that for modelling purposes $ln(S(t))$ is more useful in many cases and Milstein and Euler are the same as $\beta' = 0$.

A First appendix

Here we clarify some ideas about conditional expectation, working the probability space $(\Omega, \mathcal{H}, P)$.

Definition A.1 (Conditional Expectation). Let X be a random variable and $\mathcal{F}$ a sub sigma algebra. Then $E[X \mid \mathcal{F}]$ is the $\mathcal{F}$-measurable random variable satisfying for all $A \in \mathcal{F}$:

$$\int E[X \mid Y]\mathbb{1}_A dP = \int X\mathbb{1}_A dP \tag{299}$$

Lemma A.1. *(Tower Property) Let $\mathcal{F}$ and $\mathcal{G}$ be 2 sigma-algebras such that $\mathcal{F} \subset \mathcal{G}$, then:*

$$E[E[X \mid \mathcal{G}] \mid \mathcal{F}] = E[X \mid \mathcal{F}] \tag{300}$$

Proof. Let $A = E[X \mid \mathcal{G}]$ and $B = E[A \mid \mathcal{F}]$. Now applying the definition we have:

$$E[X \mathbb{1}_G] = E[A \mathbb{1}_G] \quad G \in \mathcal{G} \tag{301}$$
$$E[A \mathbb{1}_F] = E[B \mathbb{1}_F] \quad F \in \mathcal{F} \tag{302}$$

Then using $\mathcal{F} \subset \mathcal{G}$:

$$E[X \mathbb{1}_F] = E[A \mathbb{1}_F] = E[B \mathbb{1}_F] \quad F \in \mathcal{F} \tag{303}$$

Now applying the definition:

$$E[X \mid \mathcal{F}] = B \tag{304}$$
$$= E[E[X \mid \mathcal{G}] \mid \mathcal{F}] \tag{305}$$

□

References

[1] J. M. Steele. *Stochastic Calculus and Financial Applications*. Springer, 2012.

[2] S. E. Shreve. Stochastic calculus for finance 2, 2004.

[3] B. Oksendal. *Stochastic Differential Equations*. Springer-Verlag, 2000.

[4] H.-H. Kuo. *Introduction to Stochastic Integration*. Springer, 2006.

[5] T. Bjork. *Arbitrage Theory in Continuous Time*. Oxford University Press, 1998.

[6] F. Black and M. Scholes. *The Pricing of Options and Corporate Liabilities*. The Journal of Political Economy, 81, 637-654, 1973.

[7] M. Puig. An introduction to stochastic volatility models, 2017.

[8] K. Schneider. Quant stackexchange. https://quant.stackexchange.com/questions/61927/heston-stochastic-volatility-girsanov-theorem, 2021. [Online; accessed 3-September-2024].

[9] S. L. Heston. *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*. The Review of Financial Studies 1993, 6, 327-343, 1993.

[10] F. D. Rouah. Euler and milstein discretization. https://www.frouah.com/finance%20notes/Euler%20and%20Milstein%20Discretization.pdf, 2013. [Online; accessed 3-September-2024].

[11] J. Gatheral. *The Volatility Surface: A Practitioner's Guide*. John Wiley & Sons, 2006.