UiO : Department of Mathematics
University of Oslo

Polynomial Processes and Applications

Ville Jalasto

Master's Thesis, Spring 2020

This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Finance, Insurance and Risk*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

ABSTRACT. The goal of this thesis is to introduce polynomial processes and to present some of the most important findings and applications in a clear and pedagogical way. A polynomial process is a particular type of Markov process. Polynomial processes are defined by being in a sense polynomial preserving in expectation. This enables us to show that the calculation of moments can be done using matrix exponentials. Furthermore, the matrix is easily obtained from the extend generator of the process. As pricing of options and hedging often depends on the computation of moments, polynomial processes are attractable in financial modeling. The class of polynomial processes contains many of the processes already used in application, for example Ornstein-Uhlenbeck processes and Jacobi processes. Electricity markets provide an interesting application for polynomial processes. A case of hedging long-term electricity commitments with a risk-minimizing rolling hedge is introduced. Polynomial processes can also be applied in many other areas from interest rate models to computing life insurance liabilities.

Acknowledgments

I would like thank my patient and knowledgeable supervisor Salvador Ortiz-Latorre for providing me with a challenging but interesting topic for the thesis and for guiding me throughout the project. I appreciate our discussions and the time you have put into this.

I am eternally grateful for the loving support of my wonderful girlfriend Cathrine Bell. This would not have been possible without you. In the end I also want thank our newborn son Casper for giving me the extra motivation needed to finish the thesis.

Contents

Acknowledgments 3

Introduction 6
Outline of the Thesis 6
Polynomial Processes in Actuarial Sciences 7

Chapter 1. Stochastic Analysis 8
1.1. Stochastic Processes 8
1.2. Brownian Motion 12
1.3. Stochastic Calculus 14
1.4. Lévy Processes 20
1.5. Semimartingales and Stochastic Calculus 24
1.6. Markov Processes 31

Chapter 2. Polynomial Processes 34
2.1. Some Preliminaries 34
2.2. Polynomial Processes 35
2.3. Semimartingales and Polynomial Processes 44
2.4. Examples of Polynomial Processes 56
2.5. Computation of Moments 57
2.6. Applications 60

Chapter 3. Application on Electricity Markets 62
3.1. Electricity Markets 62
3.2. The Model 64
3.3. Market Price of Risk 66
3.4. Risk Minimizing Rolling Hedge 68
3.5. Simulation 71
3.6. Stress Test 73
3.7. Conclusion 74

Appendix A. Some Technical Results 80
A.1. Stochastic Analysis 80
A.2. Multi-notation and Itô's Formula 81

A.3. Instantaneous Covariation Matrix and Hedging Strategy 83
Appendix B. R-code 85
Appendix. Bibliography 91

Introduction

Outline of the Thesis

Although polynomial processes were first time mentioned already in the 1960's by Wong [24], the use of polynomial processes in financial applications is a quite new invention dating back to early 2000's, studied by among others Zhou [25]. First systematical accounts treating polynomial processes as Markov processes (with jumps) were done by Cuchiero et al. [9, 8], which we will follow closely in the definition of polynomial processes in Chapter 2. However, first we give a short introduction to stochastic processes and stochastic calculus in Chapter 1. Basic concepts in stochastic analysis, such as filtration, adapted processes and martingale are introduced. Further we will give definitions of several stochastic processes, in an increasing order of complexity, starting from Brownian motion and ending up with general semimartingales. Traditional calculus is not enough when dealing with a stochastic process, which is why we also give some basics on stochastic calculus. This gives the tools to say something about the dynamics and properties of stochastic processes. In the end of Chapter 1 we will also take a brief look at Markov processes, as they are the building block of polynomial processes. In Chapter 2 we define polynomial processes as a type of Markov process, which is in a sense polynomial preserving in expectation. The key point being that we can state the expectation of a polynomial function of a polynomial process explicitly with the help of a matrix exponential. The form of the matrix exponential is acquired through a connection between the (extended) generator of a Markov process and characteristics of a special semimartingale. Chapter 2 relies heavily on the work of Cuchiero et al. [9]. At the end of the chapter we briefly review some examples of applications of polynomial processes. In the last chapter we will give an example of the use of polynomial process in electricity markets by studying a case of hedging long-term commitments with a risk-minimizing rolling hedge, based on the work of Kleisinger-Yu et al. [18]. A model for spot price of electricity is given by a quadratic function on the underlying d-dimensional process that is going to be a polynomial process. Using the matrix exponential we can then explicitly calculate both forward prices and a particular hedging strategy for a long-term commitment to sell electricity. A simulation in the end of Chapter 3 shows the efficiency of the hedging strategy, by comparing hedged and unhedged exposures with each other. We will also test the robustness of our hedging strategy by doing a stress test to see how changing some of the key parameters affects the hedge.

Polynomial Processes in Actuarial Sciences

This thesis is written under the master's program Stochastic Modeling, Statistics and Risk with specialization in actuarial sciences. That is the reason we want to explain here why polynomial processes can be used in actuarial sciences and by insurance companies in general. The methods for calculating moments and conditional expectations of polynomial processes can be used to pricing financial derivatives. Companies providing life insurance have to deal with liabilities that can extend far in to the future. Hedging these liabilities is of interest to life insurance companies. Polynomial processes can be used in calculations of risk-minimizing hedging strategies, see Example 2.22. Furthermore, whenever working with financial products that stretch over time one encounters interest rates, either deterministic or stochastic. These are used to calculate current values of future payments. Modeling stochastic interest rates can be done by using polynomial processes, for instance by using the state price density approach, see Example 2.21.

CHAPTER 1

Stochastic Analysis

In this chapter we will give a brief introduction to stochastic processes and stochastic calculus, in particular by introducing semimartingales and Markov processes, which are going to be important when we consider polynomial processes in Chapter 2. As most of the results and definitions in this chapter are well known from most stochastic analysis books, we will not be giving rigorous proof in most of the cases. Most of this chapter is based on the work of Klebaner [17] and Eberlein and Kallsen [10], while for more rigorous approach we recommend for example Jacod and Shiryaev [14].

1.1. Stochastic Processes

A stochastic process $X = (X_t)$ is a collection of random variables defined on the same *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ and takes values in Euclidean space $\mathbb{R}^d$ for some $d \in \mathbb{N}$. We will be writing X for both real- and vector-valued processes. It should be clear from the context which one is meant. The index $t \in [0, \infty)$ (or $[0, T]$ for some $T > 0$) is usually interpreted as time and we are interested in how the process evolves through time. The fact that the random variables X_t are defined on the same probability space means that for any outcome $\omega \in \Omega$, we can consider the *path* $t \mapsto X_t(\omega)$ of the process. For example when modeling financial data, X_t could be the price of a stock at time t and we would like to know how the price changes through time. One usually knows what has happened up to time t, i.e. one observes the path $s \mapsto X_s(\omega)$ for all $s \leq t$, and would like to know how the process is going to evolve from there. We call the "history" of the process the *filtration* and define it as follows:

Definition 1.1. (Filtration) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and for every $t \geq 0$ let $\mathcal{F}_t$ be a sub-σ-algebra of $\mathcal{F}$, then $(\mathcal{F}_t)_{t \geq 0}$ is called a filtration if

$$\mathcal{F}_s \subseteq \mathcal{F}_u \subseteq \mathcal{F}$$

for all $s \leq u$. Filtration is thus a non-decreasing family of sub-σ-fields. We call $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$ a filtered probability space.

A closely related concept to the filtration is called adaptedness. An adapted process is a process that cannot "see in the future". In the case of stock price it means that the price X_t is known at time t, but we do not know what X_u is going to be at a future time $u > t$. We define it in the following way:

Definition 1.2. (Adaptedness) Let X_t be a stochastic process on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. It is adapted, to the filtration, if it is $\mathcal{F}_t$-measurable for all $t \geq 0$.

Remark 1.3. There are two common assumptions made about filtration, called the usual conditions. The first one is that the filtration is assumed complete in a sense that each $\mathcal{F}_t$ contains every null set and the second one is the right continuity of the filtration, i.e.

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s.$$

These assumptions are rather technical, but allow us to simplify the results, as they for example enable us to choose càdlàg versions of processes.

Basically one can think of a filtration as the flow of information through time. A natural assumption when considering for example stock prices is that they are based on the information available at each time. This gives rise to a definition of *natural filtration.* Starting with a process X, we can define a filtration with respect to which X is adapted, namely the filtration generated by the process X defined as $\mathcal{F}_t^{00} = \sigma\left(\{X_s, 0 \leq s \leq t\}\right)$. We use a slightly larger filtration called *minimal augmented filtration* $\mathcal{F}_t^X$, defined to be the smallest filtration that satisfied the usual conditions and with respect to which X is adapted. If nothing else is mentioned, we assume that the filtration in question is the minimal augmented filtration of the process.

When considering the paths of the process X, we assume them to be *càdlàg*, meaning that they are right continuous with finite left limits. We have the following definition for jumps and left limits of a càdlàg process:

Definition 1.4. If X is a càdlàg process, its left limits are given

$$X_{t-} = \begin{cases} \lim_{s\uparrow t} X_s & \text{if } t > 0 \\ X_0 & \text{if } t = 0 \end{cases}$$

and its jumps

$$\Delta X_t = X_t - X_{t-}$$

are well defined.

Much of this paper is based on *Markov processes*, which are stochastic processes that have the *Markov property.* Roughly speaking, this means that the future of a Markov process depends only on its current state i.e. the process has “no memory”. Again we can think of a stock price being a Markov process, meaning that all the relevant information that affects the future of the stock price is included in the current price, which is not an unreasonable assumption [1]. We will give a definition of Markov property here and leave rest of the theory to the end of this chapter, in Section 1.6.

Definition 1.5. (Markov property) A stochastic process X satisfies the Markov property if for any $0 \leq s < t$ we have

$$\mathbb{E}\left[f\left(X_t\right) \mid \mathcal{F}_s\right] = \mathbb{E}\left[f\left(X_t\right) \mid X_s\right]$$

where $\mathcal{F}_s$ denotes the σ-field generated by the process up to time s and f is a Borel-measurable function satisfying $\mathbb{E}\left[|f(X_t)|\right] < \infty$ for all $t \geq 0$.

When working with stochastic processes one often needs to use the concept of *stopping time*. It is a random variable, whose value is interpreted as the time that a given stochastic process exhibits a certain behavior, often defined by a stopping rule. One way to think of stopping time is through a closely related concept of hitting time, which is the first time a given process "hits" a given subset of the state space. Stopping time plays important role in the branch of mathematics called decision theory, but it is also used in mathematical proofs in order to control the time variable of stochastic processes. The formal definition is as follows:

Definition 1.6. (Stopping time) A $[0, \infty]$-valued random variable τ is called a stopping time (w.r.t $\mathcal{F}_t$) if

$$\{\tau \leq t\} \in \mathcal{F}_t$$

for all $t \in [0, \infty]$.

Basically this means that the "decision" to stop the process at time t has to be based on the information available at that time ($\mathcal{F}_t$). Furthermore we define the stopped process X^τ as

$$X_t^\tau := X_{\tau \wedge t},$$

which evolves normally until it is stopped and is constant after the stopping time is reached.

The concepts of *(true) martingale* and *local martingale* are essential in the study of stochastic processes. The martingale property means that if we know the values of the process up to time t, and let $X_t = x$, then expected future value of the process, given the information up to time t, is $\mathbb{E}\left[X_s | \mathcal{F}_t\right] = x$. Martingales also play a major role in pricing of assets, in what is called the martingale pricing theorem. This involves finding a risk-neutral measure under which the asset in question is a martingale.

Definition 1.7. (Martingale) A real valued process $M = (M_t)_{t \geq 0}$ on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$ is called a martingale if

- $\mathbb{E}\left[M_t \mid \mathcal{F}_s\right] = M_s$ for all $s \leq t$.

[1] If we accept the efficient market hypothesis to hold. This is a somewhat controversial assumption and not entirely supported by data from finance markets.

- $\mathbb{E}\left[|M_t|\right] < \infty$ for all t.
- M_t is $\mathcal{F}_t$-measurable for all $t \geq 0$.

A local martingale is defined similarly except that it is restricted to the stopped processes:

DEFINITION 1.8. (Local martingale) A stochastic process M_t on $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ is called a local martingale if there exists an increasing sequence of stopping times τ_n such that

$$\tau_n \to \infty \text{ a.s as } n \to \infty$$

and the stopped process $M_{t\wedge\tau_n}$ is a martingale for all n.

REMARK 1.9. Every true martingale is clearly also a local martingale, but the converse is not necessarily true. Processes that are strictly local martingales exists and we will encounter one in Chapter 2. This concept of localization extends to many other processes as well. Basically when we say that a process X belongs to a localized class, we mean that the stopped process $X^\tau = X_{t\wedge\tau}$ belongs to that class.

Finally we are giving a version of the very useful and well known Fubini's theorem, that will be used regularly in this paper. We will omit the proof here, as it is quite theoretical.

THEOREM 1.10. *(Fubini's theorem) Let X_t be an adapted càdlàg stochastic process. Then we can change expectation and integral*

$$\int_0^T \mathbb{E}\left[|X_t|\right] dt = \mathbb{E}\left[\int_0^T |X_t|\, dt\right].$$

Furthermore if this quantity is finite we have

$$\mathbb{E}\left[\int_0^T X_t dt\right] = \int_0^T \mathbb{E}\left[X_t\right] dt.$$

Another useful result we want to present here are the Doob's inequalities:

THEOREM 1.11. *Let M be a martingale with càdlàg paths, then*

$$\mathbb{P}\left(\sup_{s\leq t} |M_s| \geq \lambda\right) \leq \frac{\mathbb{E}\left[|M_t|\right]}{\lambda}$$

and for $1 < p < \infty$

$$\mathbb{E}\left[\sup_{s\leq t} |M_s|\right]^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|M_t|^p\right]. \tag{1.1.1}$$

The first inequality is called Doob's inequality, while the second one is referred to as Doob's maximal L^p-inequality.

PROOF. Proof can be found for example from [2, Theorem 3.6]. $\square$

In the next sections we will introduce a specific stochastic process called Brownian motion and a way of dealing with integrals of Brownian motion called Itô calculus.

1.2. Brownian Motion

A Scottish botanist Robert Brown observed in 1828 that pollen particles suspended in liquid moved in an irregular way. It turns out that this movement is best described by a stochastic process and was later mathematically formalized by N. Wiener and P. Levy, and is known as Brownian motion. This is the main process used in the stochastic calculus of continuous processes and is defined as follows:

DEFINITION 1.12. (Brownian motion) The standard Brownian motion, sometimes called Wiener process, is a stochastic process $(B_t)_{t\geq 0}$ with the following properties:

(1) $B_0 = 0$.
(2) B_t is a continuous function of t, for all t.
(3) B_t has independent increments, that is
$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$$
are independent for all $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$.
(4) Every increment $B_t - B_s$ is normally distributed with mean 0 and variance $t - s$, that
$$B_t - B_s \sim N(0, t-s).$$

A d-dimensional Brownian motion is a random vector $B_t = \left(B_t^1, B_t^2, \ldots, B_t^d\right)$, with all coordinates B_t^i being independent one-dimensional Brownian motions.

REMARK 1.13. One can show the existence of a Brownian motion using results by Kolmogorov (extension theorem and continuity criterion).

REMARK 1.14. It follows from the property (3) of the definition above and substitution property of conditional expectation (A.1) that B_t is a Markov process. Indeed for $h > 0$, all $t \geq 0$ and f a Borel-measurable function satisfying $\mathbb{E}\left[|f(B_t)|\right] < \infty$ we define a function $\hat{f}(x) = \mathbb{E}\left[f(B_{t+h} - B_t + x)\right]$. Then
$$\begin{aligned}\mathbb{E}\left[f(B_{t+h}) \mid \mathcal{F}_t\right] &= \mathbb{E}\left[f(B_{t+h} - B_t + B_t) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}\left[f(B_{t+h} - B_t + x)\right]|_{x=B_t} = \hat{f}(B_t).\end{aligned}$$

On the other hand
$$\mathbb{E}\left[f(B_{t+h} - B_t + B_t) \,|B_t\right] = \mathbb{E}\left[f(B_{t+h} - B_t + x)\right]|_{x=B_t} = \hat{f}(B_t),$$
which leads to
$$\mathbb{E}\left[f(B_{t+h}) \mid \mathcal{F}_t\right] = \mathbb{E}\left[f(B_{t+h}) \,|B_t\right]$$

as wanted. We have only used the independent increments property of Brownian motion here, which means that every stochastic process with independent increments is a Markov process.

We present three examples of martingales constructed from Brownian motion.

THEOREM 1.15. *The following processes are martingales for Brownian motion B:*
(1) B_t,
(2) $B_t^2 - t$,
(3) $e^{uB_t - t\frac{u^2}{2}}$.

PROOF. Indeed $\mathbb{E}[B_t] = 0 < \infty$ and

$$\mathbb{E}[B_t \mid \mathcal{F}_s] = \mathbb{E}[(B_t - B_s) + B_s \mid \mathcal{F}_s] = B_s,$$

since $B_{t-s} \sim N(0, t-s)$ and is independent of $\mathcal{F}_s$. For the second martingale we have $\mathbb{E}[B_t^2] = t < \infty$. We use the following identity

$$B_t^2 = B_s^2 + 2B_s(B_t - B_s) + (B_t - B_s)^2$$

together with

$$\mathbb{E}[B_s(B_t - B_s) | \mathcal{F}_s] = B_s \mathbb{E}[B_t - B_s] = 0,$$

which follows from the $\mathcal{F}_s$-measurability of B_s and the fact that $B_{t-s} \sim N(0, t-s)$ and is independent of $\mathcal{F}_s$. Combining these results gives

$$\mathbb{E}\left[B_t^2 - t \mid \mathcal{F}_s\right] = B_s^2 - t + t - s = B_s^2 - s,$$

as wanted. Regarding the last one we notice that

$$\begin{aligned}\mathbb{E}\left[e^{uB_t} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[e^{uB_s + u(B_t - B_s)} \mid \mathcal{F}_s\right] \\ &= e^{uB_s}\mathbb{E}\left[e^{u(B_t - B_s)}\right] = e^{uB_s} e^{(t-s)\frac{u^2}{2}}.\end{aligned}$$

The martingale property follows when we multiply both sides with $e^{-t\frac{u^2}{2}}$. In addition to the same techniques used in the first two proofs, we have also used the moment generating function of Brownian motion $\mathbb{E}\left[e^{uB_t}\right] = e^{t\frac{u^2}{2}} < \infty$. □

All of these martingales are used in the theory of stochastic integration. The second one, $B_t^2 - t$, provides an alternative way of describing Brownian motion, called Lèvy characterization, and the exponential martingale is useful for changing distributional properties of a process, for example Girsanov's theorem for changing probability measure.

What we are often interested in, is the paths of the Brownian motion and their properties:

PROPOSITION 1.16. *Brownian paths have the following properties, almost surely:*

(1) Is not monotone at any interval, no matter how small.
(2) Is nowhere differentiable, i.e. not differentiable at any point.
(3) Has infinite variation on any interval.
(4) Has quadratic variation on $[0,t]$ equal to t, for any t.

PROOF. We will not give a rigorous proof here, rather a few comments are presented. Property (1) is straight from the definition of Brownian motion. Property (4) follows from property (5) and continuity, as the assumption of finite variation leads to contradiction, since all continuous finite variation processes have zero quadratic variation. Property (2) then follows from (4), as a monotone function has finite variation. □

When considering a stock price or some other random process, we would like to know the dynamics of that process, that is how it evolves through time. For a non-random process this is done by considering a differential equation. In order to incorporate randomness in to this we introduce a stochastic differential equation (SDE), where one or more of the terms are stochastic. A typical SDE is of the form:

$$dX_t = b\left(t\right)dt + \sigma\left(t\right)dW_t, \tag{1.2.1}$$

where b and σ are some given functions and W_t denotes a Brownian motion. The fact that the paths of Brownian motion are nowhere differentiable makes the last term undefined. Stochastic calculus is introduced in the next section to deal with this problem.

1.3. Stochastic Calculus

We are going to define the stochastic integral with respect to Brownian motion and give some of the main properties of these integrals. The integral we are interested in is of the form:

$$\int_0^T \sigma\left(s\right)dB_s,$$

where $\sigma\left(t, X_t\right)$ is a suitable function and B is Brownian motion. These integrals are called Itô integrals, named after Kiyoshi Itô. One starts with simple processes and by using limiting procedures they are defined for more general processes. We are not going into all the details of this construction, instead we state the existence and some fundamental properties with a short sketch of the proof. First some restrictions for the integrands is needed for the Itô integral to make sense and to satisfy desirable properties.

DEFINITION 1.17. Let $\mathcal{V} = \mathcal{V}\left(S,T\right)$ be the class of functions

$$X_t\left(\omega\right) : [0,\infty) \times \Omega \to \mathbb{R}$$

such that

- X_t is $\mathcal{F}_t$-adapted.
- X_t is jointly measurable, i.e. $(t,\omega) \to X_t(\omega)$ is $\mathcal{B} \times \mathcal{F}$-measurable.
- $\mathbb{E}\left[\int_S^T X_t^2 dt\right] < \infty$.

Next theorem provides the existence and some useful properties of Itô integrals.

THEOREM 1.18. *Let $X_t, Y_t \in \mathcal{V}(S,T)$ and let $0 \leq S < U < T$, then the Itô integral $\int_S^T X_t dB_t$ is well defined and it has following properties.*

(1) Itô isometry

$$\mathbb{E}\left[\left(\int_S^T X_t dB_t\right)^2\right] = \mathbb{E}\left[\int_S^T X_t^2 dt\right].$$

(2) $\int_S^T X_t dB_t = \int_S^U X_t dB_t + \int_U^T X_t dB_t$.

(3) Linearity

$$\int_S^T (cX_t + Y_t)\, dB_t = c\int_S^T X_t dB_t + \int_S^T Y_t dB_t.$$

(4) Zero mean property

$$\mathbb{E}\left[\int_S^T X_t dB_t\right] = 0.$$

(5) $\int_S^T X_t dB_t$ is $\mathcal{F}_T$-measurable.

PROOF. First we show that the integral $\int_S^T X_t dB_t$ is well defined for $X_t \in \mathcal{V}(S,T)$. The idea is to define the integral first for simple functions $\phi \in \mathcal{V}$ and then by a limit procedure to extend it to more general processes. We start with a simple process of the form

$$\phi(t) = \sum_{i=0}^{n-1} \xi_i 1_{[t_i, t_{i+1})}(t),$$

for partition $S = t_0 < t_1 < \cdots < t_n = T$. We also require ξ_i to be $\mathcal{F}_{t_i}$-measurable (since $\phi \in \mathcal{V}$). For simple functions it is natural to define the Itô integral as

$$\int_S^T \phi(t)\, dB_t = \sum_{i=0}^{n-1} \xi_i \left(B_{t_{i+1}} - B_{t_i}\right).$$

The martingale property of B and the measurability of ξ_i's can be used to get the Itô isometry for simple functions:

$$\begin{aligned}\mathbb{E}\left[\left(\int_S^T \phi(t)\,dB_t\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n-1}\xi_i\left(B_{t_{i+1}}-B_{t_i}\right)\right)^2\right] = \sum_{i=1}^{n-1}\mathbb{E}\left[\xi_i^2\left(B_{t_{i+1}}-B_{t_i}\right)^2\right] \\ &+ 2\sum_{i<j}\mathbb{E}\left[\xi_i\xi_j\left(B_{t_{i+1}}-B_{t_i}\right)\left(B_{t_{j+1}}-B_{t_j}\right)\right] \\ &= \mathbb{E}\left[\sum_{i=0}^{n-1}|\xi_i|^2\left(t_{i+1}-t_i\right)\right] = \mathbb{E}\left[\int_S^T |\phi(t)|^2\,dt\right],\end{aligned} \tag{1.3.1}$$

where second to last equality follows when we condition on $\mathcal{F}_{t_i}$ and use the martingale property. It is possible to show that there exists a sequence of simple functions $\{\phi_n\}$ such that

$$\mathbb{E}\left[\int_S^T \left(X_t-\phi_n(t)\right)^2 dt\right] \to 0 \qquad \textit{as } n\to\infty. \tag{1.3.2}$$

Then the Itô integral can be defined as the limit

$$\int_S^T X_t dB_t = \lim_{n\to\infty}\int_S^T \phi_n(t)dB_t, \tag{1.3.3}$$

which by the Itô isometry exists in L^2. Furthermore 1.3.3 does not depend on the actual choice of $\{\phi_n\}$ as long as (1.3.2) holds, see Oksendal [21] for more details. For the properties (1)-(5) it is enough to see that they hold for simple functions and thus by taking limits we obtain them for all $X, Y \in \mathcal{V}(S,T)$. □

The Itô integral with respect to the Brownian motion is itself a stochastic process and also a martingale:

THEOREM 1.19. *Given $X_t \in \mathcal{V}(0,T)$ the Itô integral*

$$Y(t) = \int_0^t X_s dB_s$$

is a martingale. Furthermore it is square integrable on $[0,T]$, i.e. its second moments are bounded

$$\sup_{t\le T}\mathbb{E}\left[Y_t^2\right] < \infty.$$

REMARK 1.20. The Itô integral can be defined for a larger class of integrands than $\mathcal{V}$. One can relax the condition that X_t is $\mathcal{F}_t$-adapted to X_t being $\mathcal{H}_t$-adapted for an increasing family of σ-algebras such that B_t is a martingale w.r.t $\mathcal{H}_t$ (implies that

$\mathcal{F}_t \subset \mathcal{H}_t$). This allows the introduction of the multi-dimensional Itô integral. Moreover we can weaken the square integrability in the following sense

$$\mathbb{P}\left(\int_S^T X_t^2 dt < \infty\right) = 1.$$

The construct of the more general Itô integral relies to convergence in probability instead of L^2 as in the proof of Theorem 1.18. For these extended Itô integrals Theorem 1.19 is no longer true, instead we have that the extended Itô integral is in general only a *local* martingale.

Itô's formula is an important tool in stochastic calculus and is also known as the change of variables formula. Here is a version for Brownian motion:

THEOREM 1.21. *(Itô's formula for Brownian motion) Let B_t be a Brownian motion on $[0,T]$ and $g(x)$ a twice continuously differentiable function ($g \in C^2$) on $\mathbb{R}$, then for any $t \leq T$*

$$g(B_t) = g(0) + \int_0^t g'(B_s)\, dB_s + \frac{1}{2}\int_0^t g''(B_s)\, ds. \tag{1.3.4}$$

PROOF. Both integrals in (1.3.4) are well defined, since both $g'(B_s)$ and $g''(B_s)$ are continuous and adapted. Let $\{t_i^n\}$ be a partition of $[0,t]$, we have

$$g(B_t) = g(0) + \sum_{i=0}^{n-1}\left(g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)\right).$$

Applying Taylor's formula to $g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)$ gives

$$\begin{aligned} g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right) = {} & g'\left(B_{t_i^n}\right)\left(g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)\right) \\ & + \frac{1}{2} g''\left(\theta_i^n\right)\left(g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)\right)^2, \end{aligned}$$

where $\theta_i^n \in \left(B_{t_i^n}, B_{t_{i+1}^n}\right)$, resulting in

$$\begin{aligned} g(B_t) = {} & g(0) + \sum_{i=0}^{n-1} g'\left(B_{t_i^n}\right)\left(g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)\right) \\ & + \frac{1}{2}\sum_{i=0}^{n-1} g''\left(\theta_i^n\right)\left(g\left(B_{t_{i+1}^n}\right) - g\left(B_{t_i^n}\right)\right)^2. \end{aligned}$$

Taking limits as $\pi_n = \max_i \left\{\left(t_{i+1}^n - t_i^n\right)\right\} \to 0$, the first sum converges to the extended Itô integral $\int_0^t g'(B_s)\, dB_s$ while the second sum converges to the Lebesgue integral $\int_0^t g''(B_s)\, ds$, see [17, Theorem 4.14] for the proof. This completes the proof. $\square$

Remember the equation 1.2.1 in the end section on Brownian motion. We can now give a more precise definition of this type of equations called Itô processes.

DEFINITION 1.22. (Itô process) An Itô process has the form

$$Y_t = Y_0 + \int_0^t b(s)\,ds + \int_0^t \sigma(s)\,dB_s, \quad 0 \le t \le T, \tag{1.3.5}$$

where Y_0 is $\mathcal{F}_0$-measurable, processes $b(t)$ and $\sigma(t)$ are jointly measurable, such that

$$\int_0^T |b(t)|\,dt < \infty \text{ and } \int_0^T \sigma(t)^2\,dt < \infty, P - a.s.$$

REMARK 1.23. Note that the processes $\mu(t)$ and $\sigma(t)$ can and often do depend on B_t and/or Y_t as well, or even in the whole past of those processes. Itô processes are thus not in general Markov processes. Often the stochastic differential version is given (see equation 1.2.1)

$$dY_t = b(t)\,dt + \sigma(t)\,dB_t,$$

but this representation has only meaning in the sense of equation 1.3.5.

Next we will give the Itô's formula for Itô processes.

THEOREM 1.24. *Let X_t have the stochastic differential for* $0 \le t \le T$

$$dX_t = b(t)\,dt + \sigma(t)\,dB_t.$$

Let $f(x)$ be a twice continuously differentiable function, then the stochastic differential of the process $Y_t = f(X_t)$ exists and is given by

$$df(X_t) = f'(X)_t)\,dX_t + \frac{1}{2}f''(X_t)\,d[X,X]_t\,.$$

where $d[X,X]_t = (dX_t)^2$(quadratic variation) and dX_t are calculated according to the rules

$$dB_t dt = 0,\ (dt)^2 = 0$$

and

$$(dB_t)^2 = d[B,B]_t = dt.$$

PROOF. The proof is omitted here, but it is based on the same kind of ideas as the proof of Theorem 1.21. □

We will only consider Markov processes in the next chapter and thus we restrict our interest to Itô diffusions, which are Itô processes with some extra restrictions on the functions b and σ.

DEFINITION 1.25. (Itô diffusion) A time-homogeneous process X is called an Itô diffusion if it satisfies a stochastic differential equation

$$dX_t = b\left(X_t\right) dt + \sigma\left(X_t\right) dB_t, \tag{1.3.6}$$

where $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ do not depend on time, unlike in Definition 1.22.

These processes have many useful properties, among others they satisfy the Markov property. The uniqueness and existence of a solution to 1.3.6 is guaranteed when b and σ are uniformly Lipschitz, meaning that there exists a constant $C < \infty$ such that for all $x, y \in \mathbb{R}$,

$$\left|b\left(x\right) - b\left(y\right)\right| + \left|\sigma\left(x\right) - \sigma\left(y\right)\right| \leq C\left(x - y\right).$$

This result is a simplified form of the existence and uniqueness theorem, see [21, Definition 7.1.1].

Most of the results in this section can be expanded to vector-valued processes. As previously described a d-dimensional Brownian motion $B_t = \left(B_t^{(1)}, B_t^{(2)}, \ldots, B_t^{(d)}\right)$ is a vector where all the coordinates are independent 1-dimensional Brownian motions. Let $\mathcal{H}_t$ be a σ-field generated by B_s, $s \leq t$ and let $H\left(t\right)$ be a jointly measurable, adapted d-dimensional vector process. If for each j,

$$\int_0^T H_j^2\left(t\right) dt < \infty, \quad \mathbb{P} - a.s.,$$

or equivalently

$$\int_0^T \left|H\left(t\right)\right|^2 dt < \infty, \quad \mathbb{P} - a.s.,$$

where $\left|H\left(t\right)\right|^2 = \sum_{j=1}^d H_j\left(t\right)^2$, then the (extended) Itô integrals $\int_0^T H_j\left(t\right) dB_t^{(j)}$ are well defined and we write

$$\int_0^T H\left(t\right) dB_t = \sum_{j=1}^d \int_0^T H_j\left(t\right) dB_t^{(j)}.$$

This allows us to have an n-dimensional Itô process X_t driven by a d-dimensional Brownian motion

$$dX_t = b\left(t\right) dt + \sigma\left(t\right) dB_t,$$

where σ is an $n \times d$ matrix valued function, b_t and X_t are n-dimensional vector-valued functions and B_t is a d-dimensional Brownian motion. Again, the dependence of $b\left(t\right)$ and $\sigma\left(t\right)$ on time can be via the whole path of the process or the Brownian motion, with the only restriction being that they are adapted and integrable. Furthermore Itô diffusion can similarly be defined for vector-valued processes. Also, the Itô's formula extends to a function of several variables. Let $f(x_1, x_2, \ldots, x_n)$ be a C^2 function of n variables, then the stochastic differential is given by

$$df\left(X_t^{(1)}, \ldots, X_t^{(n)}\right) = \sum_{i=1}^{n} D_i f\left(X_t^{(1)}, \ldots, X_t^{(n)}\right) dX_i^t + \frac{1}{2}\sum_{i,j=1}^{n} D_{ij} f\left(X_t^{(1)}, \ldots, X_t^{(n)}\right) d\left[X^{(i)}, X^{(j)}\right]_t,$$

where D_i and D_{ij} stand for the partial derivatives and $\left[X^{(i)}, X^{(j)}\right]_t$ is the covariation. From the multidimensional Itô's formula we can deduce the useful integration by parts formula.

LEMMA 1.26. *(Integration by parts) Let X and Y be two Itô processes, then the following holds*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X_t, Y_t].$$

PROOF. An application of two dimensional Itô's formula with $f((x, y)) = xy$ leads directly to the result. □

In the next section we are giving a larger class of stochastic processes, which will allow us to include jumps in to our models.

1.4. Lévy Processes

We have previously defined Brownian motion, which is a continuous stochastic process. In many real life situations it is not enough with a continuous path process. Instead one needs processes that will allow jumps. A class of such processes is called Lévy processes, named after the french mathematician Paul Lévy. We start with the definition:

DEFINITION 1.27. (Lévy process) A stochastic process X_t is called a Lévy process on $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ if it satisfies the following properties:

(1) $X_0 = 0$ almost surely.
(2) X_t has independent increments, that is for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n < \infty$,
$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \cdots X_{t_n} - X_{t_{n-1}}$$
are independent.
(3) X_t has stationary increments, meaning that for any $s < t$, $X_t - X_s$ is equal in distribution to X_{t-s}.
(4) Continuity in probability. For any $\epsilon > 0$ and $t \geq 0$ it holds that
$$\lim_{h \to 0} \mathbb{P}\left(|X_{t+h} - X_t| > \epsilon\right) = 0.$$

If X_t is a Lévy process then it is possible to construct a version of X that is almost surely right continuous with left limits (càdlàg). From now on we assume such a construction is chosen.

REMARK 1.28. We notice straight away that a Brownian motion is a Lévy process. It is also true that any *continuous* Lévy process is of the form $a + \sigma B$, where B is a Brownian motion. Since Lévy process has independent increments it is a Markov process, see Remark 1.14.

The characteristics of an arbitrary Lévy process are uniquely given by its *Lévy-Khintchine triplet.*

THEOREM 1.29. *Any $\mathbb{R}^d$-valued Lévy process X is characterized by its characteristic function, given by Lévy-Khintchine formula. The characteristics function is of the form*

$$\varphi_{X_t}(\theta) = exp\left(\psi(\theta)\, t\right),\ \theta \in \mathbb{R}^d$$

where the characteristic exponent has a unique representation (for a fixed truncation function satisfying $h(y) = y$ for small y) given by

$$\psi(\theta) = i\theta^\top b - \frac{1}{2}\theta^\top c\theta + \int_{\mathbb{R}^d}\left\{e^{i\theta y} - 1 - i\theta^\top h(y)\right\} v(dy), \tag{1.4.1}$$

where $b \in \mathbb{R}^d$, $c \in \mathbb{R}^{d\times d}$ a non-negative definite matrix and v a Borel measure on $\mathbb{R}^d$ satisfying

$$\int_{\mathbb{R}^d} 1 \wedge |y|^2\, v(dy) < \infty$$

and $v(\{0\}) = 0$ (and thus called a Lévy measure).

PROOF. For the proof see [14, Corollary II.4.19 and Theorem II.5.2]. □

REMARK 1.30. The parameter a depends on the truncation function. Changing the truncation function to $\tilde{h}$ in (1.4.1) leads to

$$\tilde{b} = b + \int\left(\tilde{h}(y) - h(y)\right) v(dy).$$

The truncation function makes the theory of Lévy processes, and more general semimartingales, more complicated, as they can be confusing and lead to complicated expressions. The point of truncation is to insure that the integral in equation (1.4.1) is well defined. Fortunately many processes in mathematical finance have finite expectation and in that case truncation function $h(y) = y$ will work.

Because characteristic function completely determine the law of underlying probability distributions, every Lévy process is uniquely determined by the triplet (b, c, v): b being the drift, c a covariance matrix of the Brownian motion component and v a Lévy measure. This triplet suggests that Lévy process can be seen as having three independent components: a linear drift, a Brownian motion and a Lévy jump process respectively. The notion in the Remark 1.28, that every continuous Lévy process is a scaled Brownian motion, follows from the characteristics function when we set $v = 0$ (no jumps). The moments of a Lévy process can be stated with the Lévy-Khintchine triplet as long as the moment condition is satisfied.

THEOREM 1.31. *(Moments condition) Let X be Lévy process with Lévy-Khintchine triplet given in Theorem 1.29. Then X has moment of order $p \geq 1$ if and only if*

$$\int_{\{|y|>1\}} |y|^p \, v\,(dy) < \infty. \tag{1.4.2}$$

Furthermore the moments are obtained from the characteristic function by differentiation.

PROOF. We sketch the idea of the proof for real-valued Lévy processes while leaving the details out, see [10, Theorem 2.19] for more details. The proof is based on the fact that moments of random variable can be obtained by differentiating the characteristics function. So for a Lévy process X we get

$$\mathbb{E}\left[X_t^p\right] = (-i)^p \frac{d^p}{du^p} \exp\left(t\psi\left(\theta\right)\right)\bigg|_{\theta=0},$$

which by simple derivation leads to the first moment

$$\mathbb{E}\left[X_t\right] = -i\psi'\left(0\right)t = \left(b + \int y - h\left(y\right)v(dy)\right)t.$$

The integral in the equation above only makes sense when $\int |y - h\left(y\right)| \, v\left(dy\right) < \infty$, which is equivalent to $\int_{\{|y|>1\}} |y| \, v\left(dy\right) < \infty$. If we continue with higher order derivation, we notice that the p'th derivative of ψ involves integrals of the form $\int y^p v\left(dy\right)$. Every Lévy measure satisfy $\int_{\{|y|\leq 1\}} |y|^p \, v\left(dy\right)$, thus (1.4.2) is a necessary condition. It is also possible to show that its also sufficient condition for both real-valued and vector-valued Lévy processes. □

Since a general Lévy process is uniquely determined by the triplet (a, c, v), the question arises whether we can decompose an arbitrary Lévy process X in to a sum of three different parts - a deterministic drift, a Brownian motion with a certain covariance structure and a compound Poisson process. For this decomposition to exist we need to add the limit of a compensated compound Poisson process to deal with the possibility of having infinitely many jumps, such a decomposition is called Lévy-Itô decomposition:

THEOREM 1.32. *(Lévy-Itô decomposition) Let X be a Lévy process with Lévy-Khintchine triplet (b, c, v) with truncation function $h(y) = y1_{\{|y|\leq 1\}}$. Then we can write X as*

$$X_t = bt + \sqrt{c}B_t + \sum_{s\leq t} \Delta X_s 1_{\{|\Delta X_s|>1\}}$$
$$+ \lim_{\varepsilon\to 0}\left(\sum_{s\leq t} \Delta X_s 1_{\{\varepsilon\leq|\Delta X_s|\leq 1\}} - \int y1_{\{\varepsilon\leq|y|\leq 1\}} v(dx)\, t\right),$$

with B being a standard Brownian motion. In the vector valued case $\sqrt{c}$ is a square root of c, that is, a matrix A such that $AA^\top = c$.

PROOF. For an idea of the proof see [10, Theorem 2.33]. □

REMARK 1.33. We shall see a more general version of this decomposition in the next section, when we discuss general semimartingales.

One of the best known Lévy process, in addition to Brownian motion, is the Poisson process. It is a counting process representing the number of "events" that have occurred by a certain time. In other words it is a piecewise constant process that jumps by one at random times, where the number of events at some interval follows the Poisson distribution.

DEFINITION 1.34. (Poisson process) A counting process N_t is called a Poisson process with rate $\lambda > 0$ if it satisfies the following:

(1) $N_0 = 0$
(2) N_t has independent increments
(3) N_t has stationery increments and for any $s < t$,

$$N_{t-s} = N_t - N_s \sim Poisson\left(\lambda\left(t-s\right)\right).$$

REMARK 1.35. Unlike Brownian motion, the Poisson process is of positive and finite variation.

Poisson process is increasing and thus not martingale itself, however the compensated process $N_t - \lambda t$ is a martingale.

PROPOSITION 1.36. *We have the following martingales for Poisson process N_t:*

(1) $N_t - \lambda t$
(2) $(N_t - \lambda t)^2 - \lambda t$
(3) $e^{\ln(1-u)N_t + u\lambda t}$, *for any* $0 < u < 1$.

PROOF. For the first one we have for any $s < t$

$$\begin{aligned}\mathbb{E}\left[N_t - \lambda t \,|\, \mathcal{F}_s\right] &= \mathbb{E}\left[N_t - N_s \,|\, \mathcal{F}_s\right] + \mathbb{E}\left[N_s \,|\, \mathcal{F}_s\right] - \lambda t \\ &= \lambda\left(t-s\right) + N_s - \lambda t = N_s - \lambda s,\end{aligned}$$

where we have used independence of increments and the Poisson distribution of increments. Same idea works for the other two as well, so we skip the proofs. □

Other examples of Lévy processes include compound Poisson processes, the Cauchy process, gamma processes and the variance gamma process. Interested reader is referred to [10, section 2.4] for more details on these process and other Lèvy processes.

Stochastic calculus defined in the previous section is only defined for Brownian motion. In the next section we introduce a more general process called semimartingales. In fact it is the largest class of stochastic processes for which the stochastic integral is defined. We will also give a generalized version of Itô's formula, that will work for processes with jumps as well, including Lévy processes.

1.5. Semimartingales and Stochastic Calculus

Semimartingales are the most general processes for which stochastic calculus is developed. A semimartingale process is a sum of a local martingale and a finite variation process. Due to this representation the stochastic integral with respect to a semimartingale is a sum of two integrals, one w.r.t a local martingale and the other w.r.t a finite variation process. The latter can be dealt as Stieltjes integral and is thus already familiar. The integral w.r.t local martingale on the other hand is not familiar yet, as we have so far only shown the existence of stochastic integral with respect to Brownian motion. We are going to see that stochastic integral can be generalized to semimartingales. We are also introducing some new concepts such as compensator, quadratic variation and sharp bracket, which will help us to give the Itô's formula in its more general form. First we define semimartingales:

Definition 1.37. A real valued càdlàg process X on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a *semimartingale* if it can be written as

$$X_t = X_0 + A_t + M_t, \tag{1.5.1}$$

where M is a local martingale and A is an adapted càdlàg process of finite variation. A d-dimensional semimartingale is a process $X = (X^i)_{1\leq i\leq d}$ where all the components X^i are real valued semimartingales. Furthermore X is called a *special* semimartingale if the finite variation process A is in addition *predictable* (see Definition 1.38 below).

Example. Some examples of semimartingales.

(1) $X_t = B_t^2$. Here $M_t = B_t^2 - t$ is a local martingale and $A_t = t$ is a finite variation process.
(2) $X_t = B_t$ as B_t is a martingale, and thus local martingale.
(3) $X_t = N_t$ where N is a Poisson process with rate λ, as the Poisson process is of finite variation.

(4) A Lévy process is a semimartingale (special if bounded or locally bounded jumps, see [10, Proposition 2.14])
(5) Am Itô diffusion is a semimartingale. Indeed the integral with respect dB_t is a local martingale and integral with respect to dt is a predictable finite variation process.
(6) The application of Itô's formula. Meaning that for a function $f \in C^2$, $f(X_t)$ is a semimartingale whenever X is a semimartingale.

Before we can move to integrals with respect to semimartingales and specifically the general version of Itô's formula, we need to define some important concepts. We start with predictability which is a sort of stronger version of adaptedness and crucial to stochastic calculus, since only processes that are predictable are integrable with respect to general semimartingales. The definition of predictability is not very informative, but we will give some examples of subclasses of predictable processes after the definition. Intuitively one can think that a predictable process is being known a nanosecond in advance. For example an investment strategy is usually assumed predictable (w.r.t filtration created by the assets), as the amount of assets owned at time t has to be decided before the prices of those assets at time t is known.

DEFINITION 1.38. (Predictability) Let X be a stochastic process on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. We call X predictable if it is $\mathcal{P}$-measurable, $\mathcal{P}$ being the predictable σ-algebra on $\mathbb{R}_+ \times \Omega$ which is generated by all left-continuous adapted processes.

REMARK 1.39. A process H is predictable if it is one of the following:
(1) a left-continuous adapted process, thus also a continuous adapted process.
(2) a limit of left-continuous adapted process.
(3) a càdlàg process such that for any stopping time τ, H_τ is $\mathcal{F}_{\tau-}$-measurable.
(4) a Borel-measurable function of a predictable process.

As said earlier the stochastic integral with respect to semimartingales is a sum of two integrals, one with respect to a local martingale $\int_0^t H_s dM_s$, which is the one that is undefined so far. When M is a Brownian motion, the integral is the Itô integral from Section 1.3. But now martingales are allowed to jump and this makes the theory more complicated. It turns out that it can be shown, that for a locally square integrable martingale M, the integral

$$\int_0^t H_s dM_s$$

exists for the class of locally bounded predictable processes H, for the idea of the proof see [10, Section 3.2.3].

Next we define quadratic variation and covariation for semimartingales and give some important properties of quadratic (co)variation.

DEFINITION 1.40. Let X and Y be semimartingales on a common space, then the covariation process, also called the square bracket process, denoted by $[X, Y]$, is defined as

$$[X,Y]_t = \lim_{\pi_n \to 0} \sum_{i=0}^{n-1} \left(X_{t_{i+1}} - X_{t_i}\right)\left(Y_{t_{i+1}} - Y_{t_i}\right),$$

where partition of the interval $[0, t]$ is $0 = t_0 < t_1 < \cdots < t_n = t$ and the mesh $\pi_n = \max\{t_{i+1} - t_i : i = 0, \ldots, n-1\}$. Limit here is understood in probability. Taking $Y = X$ we get the quadratic variation process $[X, X]$.

This way of defining quadratic covariation is seldom used in actual calculation. In fact an equivalent definition can be given by the integration by parts formula, already stated for Brownian motion earlier (Lemma 1.26). We give it here as a theorem that follows from the Definition 1.40.

THEOREM 1.41. *If X, Y are semimartingales, then the quadratic covariation is given by*

$$[X,Y] = XY - X_0 Y_0 - \int X_- dY - \int Y_- dX.$$

PROOF. For proof we refer to [14, Def I.4.46 and Theorem I.4.47], Jacod's proof defines first quadratic covariation with integration by parts and then shows that the sum in Definition 1.40 converges in probability to $[X, Y]$. □

We state some of the properties of (co)variation processes:

PROPOSITION 1.42. *Let X and Y be semimartingales, then the quadratic (co)variation process has the following fundamental properties:*

(1) *$[X, Y]$ is bi-linear and symmetric. i.e. $[X, Y] = [Y, X]$ and*

$$[aX + Y, bU + V] = ab[X,U] + a[X,V] + b[U,Y] + [V,Y].$$

(2) *Polarization identity*

$$[X,Y] = \frac{1}{2}\left([X+Y, X+Y] - [X,X] - [Y,Y]\right).$$

(3) *$[X, X]$ is a non-decreasing function in t.*
(4) *$[X, Y]$ is a process of finite variation.*
(5) *If one of the processes Y or X is of finite variation, then*

$$[X,Y]_t = \sum_{s \le t} \Delta X_s \Delta Y_s.$$

(6) *If X is a continuous and X or Y is of finite variation, then $[X, Y] = 0$.*

PROOF. Statement (1) can be derived straight form the definition of quadratic co-variation. The polarization identity (2) is a well known formula and follows from the bi-linearity property. Property (3) is also obvious from the definition, since for $t > s$ and $s = t_m$ for some m

$$[X,X]_t - [X,X]_s = \lim_{\pi_n \to 0} \sum_{i=m+1}^{n-1} \left(X_{t_{i+1}} - X_{t_i}\right)^2 \geq 0.$$

Finite variation is a consequence of property (2) and (3), as $[X,Y]_t$ is a difference of non-decreasing processes and as such it is of finite variation. Property (6) follows from property (5), as a continuous process has no jumps, i.e $\Delta X_s = 0$ for all s if X is continuous. □

EXAMPLE 1.43. As seen earlier the quadratic variation of Brownian motion $[B,B]_t = t$, while the quadratic variation of a Poisson process is

$$[N,N]_t = \sum_{s \leq t} (\Delta N_s)^2 = \sum_{s \leq t} \Delta N_s = N_t,$$

where we have used that Poisson process is of finite variation with jump size equal to one.

In addition to quadratic variation we are going to need the concept of a *compensator.*

DEFINITION 1.44. Let N be an adapted process of integrable or locally integrable variation. It's compensator A is the unique predictable process such that

$$M = N - A$$

is a local martingale. Existence of compensators is assured by Doob-Meyer decomposition, see [17, Theorem 8.21].

Combining the idea of a compensator with the fact that quadratic variation of a semimartingale X exists and is of finite variation, see Proposition 1.42. We have the following definition for *predictable quadratic variation.*

DEFINITION 1.45. For a semimartingale X, the predictable quadratic variation (or the sharp bracket) $\langle X,X \rangle$ process is the compensator of $[X,X]$, i.e. it is the unique predictable process such that

$$[X,X] - \langle X,X \rangle$$

is a local martingale. Likewise the compensator of $[X,Y]_t$ is called the predictable covariation and denoted by $\langle X,Y \rangle_t$.

REMARK 1.46. Let X, Y be semimartingales, then we have $[X,Y] = \langle X,Y \rangle$ if either X or Y is continuous and from this it naturally follows that $[X,X] = \langle X,X \rangle$ for continuous semimartingale X. We will later see that the concept of compensator also extends to jump measures and is used in the canonical decomposition of a semimartingale.

EXAMPLE 1.47. We recall that for Poisson process N_t the quadratic variation is the process itself and we also know that $N_t - t$ is a martingale, thus the compensator of Poisson process is $\langle N, N\rangle_t = t$.

We call a martingale M uniformly integrable if it converges, when $t \to \infty$, to an integrable limit random variable M_∞ such that

$$\mathbb{E}\left[M_t|\mathcal{F}_s\right] = M_s$$

holds for $t = \infty$. The whole martingale is generated from its limit M_∞ by conditional expectation, thus we can identify the set of uniformly integrable martingales with the set $L^1\left(\Omega, \mathcal{F}_\infty, \mathbb{P}\right)$ of integrable random variables. Furthermore we denote the space of square-integrable martingales with $\mathscr{H}^2$, where we also require that the limit variable has a finite second moment, $\mathbb{E}\left[M_\infty^2\right] < \infty$. Correspondingly we identify the set $L^2\left(\Omega, \mathcal{F}_\infty, \mathbb{P}\right)$, which is a Hilbert space of square-integrable martingales, endowed with a scalar product defined as $\mathbb{E}\left[UV\right]$ for $U, V \in L^2\left(\Omega, \mathcal{F}_\infty, \mathbb{P}\right)$. This naturally induces $\mathscr{H}^2$ with a scalar product $\mathbb{E}\left[M_\infty N_\infty\right]$ for $N, M \in \mathscr{H}^2$, turning it into a Hilbert space. Using the scalar product we can define the concept of *purely discontinuous (square-integrable) martingale* $M \in \mathscr{H}^2$ that is orthogonal to all *continuous* square-integrable martingales, i.e.

$$\mathbb{E}\left[M_\infty N_\infty\right] = 0,$$

for any continuous $N \in \mathscr{H}^2$. This can be used to decompose any square-integrable martingale uniquely as an orthogonal sum

$$M = M_0 + M^c + M^d$$

of continuous martingale M^c (with $M_0 = 0$) and purely discontinuous martingale M^d. Moreover this decomposition can be extended to local martingales. One calls local martingales M, with $M_0 = 0$, purely discontinuous if they are strongly orthogonal (i.e. to all continuous local martingales N, i.e. $\langle M, N\rangle = 0$. If we now have a semimartingale $X = X_0 + A + M$, we denote the continuous part M^c of the local martingale with X^c and call it the continuous martingale part of X. It can be shown that this does not depend on the choice of M. This allows us to decompose the quadratic covariation of two semimartingales into a continuous martingale part and a pure jump part.

PROPOSITION 1.48. *For any two semimartingales X, Y we have*

$$[X, Y]_t = \langle X^c, Y^c\rangle_t + \sum_{s\le t} \Delta X_s \Delta Y_s.$$

PROOF. For the idea of of the proof, see [10, Proposition 3.9]. □

Now we can state the key result in this section, namely the general form of the famous Itô's formula.

THEOREM 1.49. *Itô's formula for semimartingales. Let X be a semimartingale and $f \in C^2$, then $f(X_t)$ is a semimartingale and Itô's formula is given by*

$$f(X_t) - f(X_0) = \int_0^t f'(X_{s-})\, dX_s + \frac{1}{2}\int_0^t f''(X_{s-})\, d\langle X^c, X^c\rangle_s + \sum_{s\le t}\left(f(X_s) - f(X_{s-}) - f'(X_{s-})\,\Delta X_s\right),$$

where we have used Proposition 1.48, to divide the quadratic variation into two parts, a continuous martingale and a purely discontinuous part. Similarly we have the multi-dimensional Itô's formula. Let X_t be an n-dimensional semimartingale and let f be a function of n-variables, then we have

$$\begin{aligned} f(X_t) - f(X_0) = &\int_0^t \sum_{i=1}^n D_i f(X_{s-})\, dX_s^i \\ &+ \frac{1}{2}\int_0^t \sum_{i,j=1}^n D_{ij} f(X_{s-})\, d\left\langle X^{i,c}, X^{j,c}\right\rangle_s \\ &+ \sum_{s\le t}\left(f(X_s) - f(X_{s-}) - \sum_{i=1}^n D_i f(X_{s-})\,\Delta X_s\right). \end{aligned}$$

PROOF. For an idea of the proof see [10, Theorem 3.16]. □

REMARK 1.50. Notice the left limits used in the formula, this ensures that X_{s-} is left-continuous and thus predictable, which ensures that the two integrals are well defined, while the jumps $\Delta X_s = X_s - X_{s-}$ are dealt by the last term.

In the end we are going to quickly go through the generalization of Lévy-Itô decomposition for semimartingales. In order to do that we need the concept of random measure and stochastic integral w.r.t a random measures. We will do this rather informally and refer to [10, Section 3.3] for more details. We let X be a semimartingale, we call μ the *integer-valued random measure* of the jumps of the semimartingale X when the following is satisfied:

(1) $\mu(\omega, \cdot)$ is a measure on $\mathbb{R}_+ \times \mathbb{R}^d$ for fixed $\omega \in \Omega$,
(2) $\mu\left(\{0\} \times \mathbb{R}^d\right) = 0$ for any $\omega \in \Omega$,
(3) $\mu(\omega, \cdot)$ has values in $\mathbb{N} \cup \{\infty\}$ for any $\omega \in \Omega$,
(4) $\mu\left(\omega, \{t\} \times \mathbb{R}^d\right) \le 1$ for any $\omega \in \Omega$, $t \ge 0$,
(5) μ is adapted, meaning that $\mu(\cdot, [0,t] \times B)$ if $\mathcal{F}_t$-measurable for fixed $t \ge 0$ and any Borel-measurable $B \in \mathcal{B}^d$.
(6) μ is predictably σ-finite.

From now on we will mostly omit the argument ω, as we have done for processes. Essentially the measure $\mu([0,t] \times B)$ counts the number of events belonging to B, these events being the jumps of the process X. If (3) and (4) are not satisfied we call the μ more generally a random measure. It is possible to integrate functions with respect to these measures. To achieve this we denote $\xi(t,x)$ for a predictable function on $\Omega \times \mathbb{R}_+ \times \mathbb{R}^d$. It is predictable in the sense that ξ is measurable w.r.t product-σ-field $\mathcal{P} \otimes \mathcal{B}^d$ on $\mathbb{R}_+ \times \mathbb{R}$. The integral process $\xi * \mu$ is defined by pathwise integration of ξ relative to μ, that is

$$\xi * \mu(t) := \int_0^t \int_{\mathbb{R}} \xi(s,x)\, \mu(ds,dx), \tag{1.5.2}$$

whenever the right-hand side makes sense, that is if

$$\int_0^t \int_{\mathbb{R}} |\xi(s,x)|\, \mu(ds,dx) < \infty, \quad t \geq 0.$$

The integral 1.5.2 is an adapted càdlàg process of finite variation and is in a semimartingale. For an integer-valued random measure it reduces to the sum

$$\int_0^t \int_{\mathbb{R}} \xi(x)\, \mu(ds,dx) = \sum_{s \leq t} \xi(s, \Delta X_s). \tag{1.5.3}$$

This is sometimes used to simplify the general Itô's formula and we use it in Chapter 2 (see also Appendix A.2). We define the compensator of μ is the unique predictable random measure v such that

$$\int_0^t \int_{\mathbb{R}} \xi(x)\, (\mu(ds,dx) - v(ds,dx)) \tag{1.5.4}$$

is a local martingale. Moreover, there exists a predictable process A and a kernel $K(t;dx)$ from $(\Omega \times \mathbb{R}_+, \mathcal{P})$ to $\mathbb{R}^d$ such that

$$v(\omega; dt, dx) = dA_t(\omega)\, K(\omega, t; dx). \tag{1.5.5}$$

The kernel stands for a local jump intensity and it is called the intensity measure of μ.

We can now state the generalized version of the Lévy-Itô decomposition, that exists for all semimartingales. At the same time we extend the three quantities given by the Lévy-Khintchine triplet (b,c,v) to more general processes (B,C,v), that does not need to be linear in t, contrasting to the Lévy case.

Definition 1.51. For a fixed truncation function h, we call the characteristics of X the triplet (B,C,v) consisting of:

- $B = (B^i)_{i \leq d}$, a predictable finite variation process (depends on the choice of h).

- $C = (C^{ij})_{i,j \le d}$, a continuous finite variation process, given by

$$C^{ij} = \left\langle X^{i,c}, X^{j,c} \right\rangle .$$

- v, a predictable random measure, namely the compensator of the jump measure μ of jumps of X.

THEOREM 1.52. *The canonical decomposition of a semimartingale X with characteristics triplet (B, C, v) is given by*

$$X_t = X_0 + B_t + X_t^c + \int_0^t \int_{\mathbb{R}} \bar{h}(x)\, \mu(ds, dx) + \int_0^t \int_{\mathbb{R}} h(x)\, (\mu(ds, dx) - v(ds, dx)), \tag{1.5.6}$$

where h is some fixed truncation function, for example $h(x) = 1_{\|x\|<1}(x)$, and $\bar{h}(x) := x - h(x)$. For special semimartingales we can use $h(x) = x$, so that 1.5.6 simplifies to

$$X_t = X_0 + B_t + X_t^c + \int_0^t \int_{\mathbb{R}} x\, (\mu(ds, dx) - v(ds, dx)) .$$

PROOF. For rigorous proof we refer to [14, Theorem II.2.34 and Corollary II.2.38]. □

1.6. Markov Processes

Many process in application are Markov process, including all the processes we are considering in this paper. As mentioned earlier Markov process is a process that has a lack of memory property. The theory of Markov processes gives us an alternative way of expressing local behavior of a stochastic process. In previous section we introduced semimartingale characteristics which can be used to characterize local dynamics of a semimartingale process. In this section we introduce generators of Markov processes and see that these can also be used for figuring out the local dynamics. We are going to be combining these two methods in the next chapter when defining polynomial processes, which turn out to be both Markov processes and special semimartingales with particular characteristics. We will here give a brief introduction to the theory of Markov processes, concentrating on the results that are needed in next chapters. We give a slightly different definition of Markov processes than in the beginning of this chapter. Here we give it in terms of a family of probability measures and a single stochastic process X. We start with a measurable space $(\Omega, \mathcal{F})$ and a filtration $\{\mathcal{F}_t\}_{t\ge 0}$. Furthermore we denote a Borel set with B and Borel-σ-field with $\mathcal{B}$.

DEFINITION 1.53. Let E be a closed subset of $\mathbb{R}^d$, $E \subseteq \mathbb{R}^d$. We call an adapted E-valued process X a time-homogeneous Markov process relative to $\mathcal{F}_t$ with state space

E if

$$\mathbb{P}_x\left(X_{s+t}\in B\,|\,\mathcal{F}_s\right)=\mathbb{P}_{X_s}\left(X_t\in B\right),\qquad s,t\geq 0, B\in\mathcal{B}\left(E\right)$$

holds for some family $(\mathbb{P}_x : x\in E)$ of probability measures on E such that $x\mapsto\mathbb{P}_x(X_t\in B)$ is measurable for a fixed B.

The function $(t,x,B)\mapsto\mathbb{P}_x\left(X_t\in B\right)$ is called the transition function of X and is denoted by $(p_t)_{t\geq 0}$ from now on. The definition 1.53 says basically the same as the Definition 1.5. Markov process is without memory, meaning that the future evolution of X_{t+s}, given the past up to time s, depends only on the present value X_s. Furthermore we only consider time homogeneous Markov processes, so that $X_{t+s}-X_s$ may depend on X_s, but not on s itself. The transition function satisfies the Chapman-Kolmogorov equation (up to a set of points almost never visited by the process),

$$p_{t+s}\left(x,B\right)=\int p_t\left(\xi,B\right)p_s\left(x,d\xi\right)$$

for $s,t\geq 0$ and $B\in\mathcal{B}\left(E\right)$.

The transition function leads to a family of operators $(P_t)_{t\geq 0}$ on Borel-measurable functions $f:E\to\mathbb{R}$, called Markov semigroup. These operators are defined by

$$P_tf\left(x\right):=\int f\left(\xi\right)p_t\left(x,d\xi\right),$$

for $t\geq 0$, $x\in E$ and all f for which the integral is well defined. The equation above can be interpreted as the expectation of $f(X_t)$ i.e.

$$P_tf\left(x\right)=\mathbb{E}_x\left[f\left(X_t\right)\right],$$

the expectation being with respect to $\mathbb{P}_x$.

Remark 1.54. We say that X is Markovian relative to $\mathcal{F}_t$, if for each $x\in E$, there exist a probability measure $\mathbb{P}_x$ such that the following holds

$$\mathbb{E}_x\left[f\left(X_{t+s}\right)|\mathcal{F}_s\right]=\mathbb{E}_{X_s}\left[f\left(X_t\right)\right]=P_tf\left(X_s\right)$$

for all $s,t\in[0,\infty)$ and all Borel-measurable functions f satisfying $\mathbb{E}[|f(X_t)|]<\infty$.

The family of operators $(P_t)_{t\geq 0}$ is a semigroup, because it satisfies

$$P_{s+t}=P_sP_t \text{ and } P_0=I,$$

where I is the identity operator. The first identity can be derived from the Chapman-Kolmogorov equation and second one is quite obvious as the expected value of X_0 is clearly the value of the process at time zero.

Sometimes it is useful to consider a Markov process up until a stopping time T_Δ, called the lifetime of the process. In order for this make sense we add a cemetery state Δ to our state space E and consider the state space $E_\Delta=E\cup\{\Delta\}$ instead of E. Further it is often assumed that $f\left(\Delta\right)=0$ for every Borel measurable function f, this to ensure that results like backward equation still hold. One reason to include a cemetery point is to deal with a possible explosion of the process. Instead of letting the process explode to infinity we kill it and set $X = \Delta$ after the explosion. In many applications this killing is not necessary as there often exist growth conditions such that processes won't reach the cemetery state.

In the case of semimartingales we described the dynamics of a process with the help of semimartingale characteristics. For Markov process similar kind of tool is called the extended (infinitesimal) generator $\mathcal{G}$.

DEFINITION 1.55. (Extended generator) An operator $\mathcal{G}$, with domain $D_{\mathcal{G}}$, is called an extended (infinitesimal) generator for a Markov process if for any function in $D_{\mathcal{G}}$

$$M_t^f := f(X_t) - f(x) - \int_0^t \mathcal{G}f(X_s)\, ds \tag{1.6.1}$$

is a $(\mathcal{F}_t, \mathbb{P}_x)$-local martingale for every $x \in E$. When M_t^f is a *true* martingale we write $\mathcal{G} = \mathcal{A}$ and call it the generator of X.

REMARK 1.56. The extended generator is a generalization of the infinitesimal generator $\mathcal{A}$, for which (1.6.1) is a *true* martingale and the generator is equivalently given by

$$\mathcal{A}f(x) = \lim_{t \geq 0} \frac{P_t f(x) - f(x)}{t},$$

for the set of all functions $f \in C_0(E)$ for which the limit exists in $C_0(E)$, this being the set of functions that vanish at infinity. Clearly when both generators exist, we have $D_{\mathcal{A}} \subset D_{\mathcal{G}}$ and $\mathcal{A}f = \mathcal{G}f$ for $f \in D_{\mathcal{A}}$.

In the next chapter we will use the results given so far to define polynomial process as a kind of Markov process.

CHAPTER 2

Polynomial Processes

Pricing and hedging different financial derivatives often involves calculating the expected value of the price processes under some martingale measure. Polynomial processes allow us to calculate the expectation by computing matrix exponentials, which is a big computational advantage in contrast to general Markov processes, where one often needs to solve higher dimensional partial differential equations. The class of polynomial processes involves many of the processes currently used in financial modeling. We will show later that for example Ornstein-Uhlenbeck processes, Jacobi diffusions and Lèvy processes are indeed polynomial processes. After giving some preliminaries, we move on to defining polynomial processes. The characterizing attribute of polynomial processes is that the expected value of any polynomial of the process is again a polynomial in the initial value of the process. This property is then used to show that the expected value of a polynomial process can be computed by matrix exponential. Furthermore the particular matrix can be deduced from the extended generator of the process. Next a connection between characteristics of special semimartingales and Markov processes is shown. This section is rather technical with somewhat complicated proofs, but in return we will achieve an easy way to verify that a given process is in fact a polynomial process. Having established the theory, a special case of Itô diffusion as polynomial processes is presented and some examples of polynomial processes are given. Moreover, we will give a formula for calculation of moments of a polynomial process. The theory presented in this chapter is mainly due Cuchiero et al. [9], with contributions from the work of Filipović and Larsson [13, 12]..

2.1. Some Preliminaries

We start by going through some general assumptions and notations used in this chapter. We write X for a càdlàg Markov process $\{X_t\}_{t\geq 0}$ defined on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ with state space E, a closed subset of $\mathbb{R}^d$. We let p_t denote the transition function defined in the Section 1.6 and we consider a time-homogeneous Markov semigroup $(P_t)_{t\geq 0}$ acting on all Borel-measurable functions $f : E \to \mathbb{R}$ for which the following expectation is well defined

$$P_t f(x) = \mathbb{E}_x \left[f\left(X_t\right)\right] := \int_E f\left(\xi\right) p_t\left(x, d\xi\right).$$

We add a cemetery state Δ to the state space, with the convention $f(\Delta) = 0$ for any function f on $E_\Delta = E \cup \{\Delta\}$. The transition function p_t satisfies the following additional properties:

(1) for all $x \in E_\Delta$, $p_0\left(x, \cdot\right) = \delta_x$;
(2) for all $t \geq 0$ and $x \in E$,$p_t\left(x, \{\Delta\}\right) = 1 - p_t\left(x, E\right)$ and $p_t\left(\Delta, \{\Delta\}\right) = 1$.

Property (1) says that the process starts at x. Property (2) gives the probability for the process to enter the cemetery state and states that when it enters there it does not leave. We let $(\mathcal{F}_t)_{t\geq 0}$ be the minimal augmented filtration mentioned in the start of Chapter 1. We assume that a probability measure $\mathbb{P}^x$ exists for each $x \in E_\Delta$ such that X is Markovian relative to $(\mathcal{F}_t)$ with semigroup (P_t).

We define $\mathrm{Pol}(\mathbb{R}^d)$ as the set of polynomials up to degree $m \geq 0$ for $m \in \mathbb{N}$ and use the following multi-index notation

$$\mathrm{Pol}_m\left(\mathbb{R}^d\right) := \left\{ x \mapsto \sum_{|\mathbf{k}|=0}^{m} \alpha_{\mathbf{k}} x^{\mathbf{k}} \mid \alpha_{\mathbf{k}} \in \mathbb{R} \right\},$$

with $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$, $|\mathbf{k}| = k + \cdots + k_d$ and $x^{\mathbf{k}} = x_1^{k_1} + \cdots + x_d^{k_d}$. The dimension of Pol_m is given by $N < \infty$ and it depends on the state space E . For example if the state space is one dimensional, i.e. $E \subseteq \mathbb{R}$, the dimension N of $\mathrm{Pol}_m\left(E\right)$ is m. We also define functions $f_{\mathbf{k}}$ by setting

$$f_{\mathbf{k}}\left(x\right) = \begin{cases} x^{\mathbf{k}} & \text{if } x \in E, \\ 0 & \text{if } x = \Delta. \end{cases}$$

Finally we set the restriction of polynomials to E as

$$\mathrm{Pol}_m(E) = \left\{ p|_E : p \in \mathrm{Pol}_m\left(\mathbb{R}^d\right) \right\}.$$

2.2. Polynomial Processes

We will start by defining polynomial process and then state some properties of the process. Our definition will follow along the lines of Cuchiero (2012) [9].

Definition 2.1. We say that a E_Δ-valued time-homogeneous Markov process X is *m-polynomial* if we for all $k \in \{0, \ldots, m\}$, all $f \in \mathrm{Pol}_m$, $x \in E$ and $t \geq 0$ have that

$$x \mapsto P_t f\left(x\right) \in \mathrm{Pol}_m.$$

Further, we assume that $t \mapsto P_t f(x)$ is continuous at $t = 0$ for all $f \in \mathrm{Pol}_m$. If X is m-polynomial for all $m \geq 0$, then it is called *polynomial.*

It is worth noticing that in the definition above it is implicitly assumed that

$$P_t|f|(x) = E_x\left[|f(X_t)|\right] < \infty, \tag{2.2.1}$$

for all $f \in \mathrm{Pol}_m$, $x \in E$ and $t \geq 0$. This ensures that the expectation is well defined. Further it is important to assume $P_t(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$ for all $k \in \{0, ..., m\}$, not just for m. Otherwise the proofs of some theorems, that are important for applications, fail. The Definition 2.1 says that if we are given a polynomial process X, we know that the expected value of any polynomial of the process at any time t is again a polynomial. There is an equivalent definition given in terms of $\mathcal{G}$, the extended generator of X. We will come back to that later, as we are going to need some results before we are able to prove this equivalence. First we recall the definition of the extended generator.

Definition 2.2. An operator $\mathcal{G}$, with domain $D_{\mathcal{G}}$, is called an extended (infinitesimal) generator for a Markov process if for any function in $D_{\mathcal{G}}$

$$M_t^f := f\left(X_t\right) - f\left(x\right) - \int_0^t \mathcal{G}f\left(X_s\right)ds \tag{2.2.2}$$

is a $(\mathcal{F}_t, \mathbb{P}_x)$-local martingale for every $x \in E_\Delta$.

We define the lifetime of the process X_t as

$$T_\Delta(\omega) = \inf\left\{t \mid X_t\left(\omega\right) = \Delta\right\},$$

where $\inf \emptyset = \infty$. We notice that

$$\{T_\Delta < t\} = \bigcup_{s<t, s\in\mathbb{Q}} \{X_s = \Delta\} \in \mathcal{F}_t.$$

Since $(\mathcal{F}_t)$ is assumed right-continuous, we have $\{T_\Delta < t\} = \{T_\Delta \leq t\}$ and thus $T_\Delta$is an $\mathcal{F}_t$-stopping time. This means, due to the convention $f(\Delta) = 0$, that the local martingale property reduces to

$$f(X_t)1_{\{t<T_\Delta\}} - f(x) - \int_0^{t\wedge T_\Delta} \mathcal{G}f(X_s)ds$$

being a local martingale.

REMARK 2.3. It is worth noticing that, if f lies in the domain of $D_{\mathcal{G}}$ and also satisfies $P_t|f|(x) < \infty$ for all $t \geq 0$ and $x \in E_\Delta$. Then, M^f is a true martingale if and only if all the increments of

$$f(X_t) - f(x) - \int_0^t \mathcal{G}f(X_u)du,$$

have vanishing expectation. That is for all $s \leq t$,

$$\mathbb{E}_x\left[f(X_t) - f(X_s) - \int_s^t \mathcal{G}f(X_u)\,du\right] = P_tf(x) - P_sf(x) - \int_s^t P_u\mathcal{G}f(x)du = 0,$$

This implies, by Fubini, that $\int_0^t P_s\mathcal{G}f(x)\,ds$ exists on finite intervals, since

$$\int_0^t P_u\mathcal{G}f(x)du = P_tf(x) - f(x) < \infty,$$

and thus $P_s|\mathcal{G}f|(x)$ exists for almost all s.

Our first theorem will establish a link between m-polynomial and the extended generator of a Markov process. We will also get an important result that provides us an effective way to calculate moments by computing matrix exponentials. In order to prove the theorem we need a lemma connecting the Kolmogorov backward equation with the extended generator.

LEMMA 2.4. *Let X be a time-homogeneous Markov process with semigroup (P_t) and $f : E_\Delta \to \mathbb{R}$ some function satisfying $P_t|f|(x) < \infty$ for all $t \geq 0$ and $x \in E_\Delta$. If $f \in D_{\mathcal{G}}$ and M^f is a true martingale, then:*

1. *For any $s \geq 0$, M^{P_sf} is a true martingale, $P_sf \in D_{\mathcal{G}}$ and $\mathcal{G}P_sf = P_s\mathcal{G}f$.*
2. *If $P_t\mathcal{G}f(x)$ is continuous at $t = 0$, then P_tf solves the Kolmogorov backward equation*

$$\frac{\partial u(t,x)}{\partial t} = \mathcal{G}u(t,x), \qquad u(0,x) = f(x)$$

PROOF. First we notice that as M^f is a true martingale, by Remark 1.56 the infinitesimal generator is given by

$$\mathcal{G}f(x) = \lim_{t\to 0}\frac{\mathbb{E}_x[f(X_t)] - f(x)}{t} = \lim_{t\to 0}\frac{P_tf(x) - f(x)}{t}.$$

Suppose that M^{P_sf} is a true martingale, then

$$\begin{aligned} P_s\mathcal{G}f(x) &= P_s\lim_{t\to 0}\frac{P_tf(x) - f(x)}{t} \\ &= \lim_{t\to 0}\frac{P_tP_sf(x) - P_sf(x)}{t} \\ &= \mathcal{G}P_sf(x), \end{aligned} \tag{2.2.3}$$

and trivially $P_s f \in D_{\mathcal{G}}$. Hence we need to show that $M^{P_s f}$ is indeed a true martingale to finish the proof of (1). This is done by showing that

$$P_s f(X_t) - P_s f(x) - \int_0^t P_s \mathcal{G} f(X_u)\, du$$

is a true martingale and this is shown in the rest of the proof. From the assumption $P_t |f| < \infty$ and Remark 2.3 we can deduce that $f(X_t)$ and $\mathcal{G} f(X_t)$ are integrable for all $t \geq 0$, which in turn means that $P_s f(X_t)$ and $P_s \mathcal{G} f(X_t)$ are integrable as well. So the expectation below is well defined for $u \leq t$,

$$\begin{aligned}
&\mathbb{E}_x \left[P_s f(X_t) - P_s f(x) - \int_0^t P_s \mathcal{G} f(X_r)\, dr | \mathcal{F}_u \right] \\
&= P_s f(X_u) - P_s f(x) - \int_0^u P_s \mathcal{G} f(X_r)\, dr \\
&+ \mathbb{E}_x \left[P_s f(X_t) - P_s f(X_u) - \int_u^t P_s \mathcal{G} f(X_r)\, dr | \mathcal{F}_u \right],
\end{aligned}$$

which is a martingale if the conditional expectation on the right hand side of the equality above is zero. Using the Markov property of X, the expectation is equal to

$$\mathbb{E}_{X_u} \left[P_s f(X_{t-u}) - P_s f(X_0) - \int_0^{t-u} P_s \mathcal{G} f(X_r)\, dr \right].$$

And we know that for any $y \in E_\Delta$, we have

$$\begin{aligned}
&\mathbb{E}_y \left[P_s f(X_{t-u}) - P_s f(X_0) - \int_0^{t-u} P_s \mathcal{G} f(X_r) dr \right] \\
&= P_{s+t-u} f(y) - P_s f(y) - \int_s^{s+t-u} P_r \mathcal{G} f(y) dr \\
&= 0,
\end{aligned}$$

where the last equality follows from Remark 2.3. For (2) we use the definition of the derivative

$$g'(x) = \lim_{h \to 0} \frac{g(x+h) - g(x)}{h},$$

the assumption of continuity of $P_t \mathcal{G} f(x)$ at $t = 0$, Remark 2.3 and the results from (1) to get the following

$$\begin{aligned}
\frac{\partial P_t f(x)}{\partial t} &= \lim_{h \to 0} \frac{P_{t+h} f(x) - P_t f(x)}{h} = P_t \lim_{h \to 0} \frac{P_h f(x) - f(x)}{h} \\
&= P_t \lim_{h \to 0} \frac{1}{h} \int_0^h P_s \mathcal{G} f(x) ds = P_t P_0 \mathcal{G} f(x) = \mathcal{G} P_t f(x),
\end{aligned}$$

which shows that $P_t f$ solves the Kolmogorov backward equation. □

We can now move on the first theorem of the chapter, which will give the connection between m-polynomials and the extended generator as promised. Some elementary results from semigroup theory are needed to prove this theorem.

THEOREM 2.5. *For a time-homogeneous process X, with state space E_Δ and semigroup (P_t), the following are equal:*

(1) X is a m-polynomial for $m \geq 0$.
(2) There exists a linear map A on Pol_k such that for all $t \geq 0$, (P_t) restricted to Pol_k can be written as

$$P_t|_{\mathrm{Pol}_k} = e^{At}.$$

(3) For all $f \in \mathrm{Pol}_m$, $x \in E_\Delta$ and $t \geq 0$

$$M_t^f = f(X_t) - f(x) - \int_0^t \mathcal{G}f(X_s)ds$$

is a true martingale and the extended generator satisfies $\mathcal{G}(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$ for all $k \in \{0, 1, ..., m\}$.

PROOF. First we fix a $k \in \{0, ..., m\}$ for the duration of the proof. We start by showing that (1) $\Longrightarrow$ (2) follows from the basic properties of semigroups. We know from the definition that the Markov semigroup (P_t) satisfies $P_{t+s} = P_t P_s$ and $P_0 = I$. Furthermore, Pol_m is finite dimensional and $t \to P_t f(x)$ is continuous at $t = 0$. Standard results from semigroup theory then implies that some linear map A exists such that $P_t|_{\mathrm{Pol}_k} = e^{At}$, see for example [11, Theorem 2.9].

In order to show (2) $\Longrightarrow$ (3), we have by (2) that for every $f \in \mathrm{Pol}_k$, $Af \in \mathrm{Pol}_k$ and

$$P_t f - f - \int_0^t P_s A f ds = e^{At} f - f - \int_0^t e^{As} f ds = 0$$

This shows that $f(X_t) - f(x) - \int_0^t Af(X_s)ds$ is a true martingale, thus we have $f \in D_{\mathcal{G}}$ which in turn implies $\mathcal{G}f = Af$. Finally the fact that A is a linear map on Pol_k together with $\mathcal{G}f = Af$ imply that $\mathcal{G}(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$. That completes the proof of the implication (2) $\Longrightarrow$ (3). For the last part, (3) $\Longrightarrow$ (1), we want to show that Definition 2.1 is satisfied. We do this by using the Kolmogorov backward equation for an initial value $u(0, \cdot) = f \in \mathrm{Pol}_k$, i.e.

$$\frac{\partial u(t,x)}{\partial t} = \mathcal{G}u(t,x).$$

From Lemma 2.4 we know that $P_t f$ solves the Kolmogorov equation if $t \mapsto P_t \mathcal{G} f(x)$ is continuous at $t = 0$ for any f, which follows from the fact that $\mathcal{G}$ maps Pol_k to itself and the martingale property of M^f. Since M^f is a martingale Remark 2.3 gives

$$P_t f(x) = f(x) + \int_0^t P_s \mathcal{G} f(x) ds,$$

which implies the continuity of $P_t f(x)$ for any $f \in \mathrm{Pol}_k$. We can define a linear map A such that $\mathcal{G}|_{\mathrm{Pol}_k} = A$, by choosing a basis $\{e_1, ..., e_N\}$ of Pol_k and setting

$$\mathcal{G}e_i =: \sum_{j=1}^{N} A_{ij} e_j.$$

The Kolmogorov backward equation reduces now to the linear ordinary differential equation

$$\frac{\partial u(t)}{\partial t} = Au(t), \quad u(0) = f,$$

with unique solution $e^{tA}f$, see [11, Proposition 2.8]. Thus on Pol_k, $P_t f$ is equal to $e^{tA}f$ which in turn is a polynomial of same or smaller degree than k by definition of matrix exponentials. This holds for all $k \in \{0, ..., m\}$ and thus X is an *m-polynomial.* □

REMARK 2.6. Part (2) of the previous proposition plays an important role in the applications of polynomial processes in the fields of mathematical finance and insurance mathematics. It namely allows us to calculate the moments of all orders of any process $f(X)$ with matrix exponentials, as long as f is a polynomial function and X is a polynomial process. The matrix exponential is given as the power series

$$e^A := \sum_{k=1}^{\infty} \frac{A^k}{k!}, \tag{2.2.4}$$

which is always convergent. Finding efficient and reliable methods for computing matrix exponential is an ongoing research topic in mathematics and numerical analysis. We will come back to this at the end of this chapter.

Theorem 2.5 characterizes *m-polynomial* processes in terms of the extended generator, but only when we assume that M^f is a *true* martingale. It turns out that M^f is a *true* martingale whenever $m \geq 2$ is an even number. We state this in the following theorem.

THEOREM 2.7. *Let X be a Markov process with state space E_Δ and let $m \geq 2$ be an even number. Then X is an m-polynomial process if and only if the following conditions are satisfied:*

- *For all $f \in \mathrm{Pol}_m$,$x \in E_\Delta$, and $t \geq 0$, M^f is a local martingale, i.e Pol_m lies in the domain of the extended generator.*
- $\mathcal{G}(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$ *for all $k \in \{1, \ldots, m\}$.*

PROOF. It is straightforward to see from Theorem 2.5 (1) $\Longrightarrow$ (3) that when X is an m-polynomial process, the conditions above are satisfied. In fact they are satisfied for all m, not just even numbers. In order to show the opposite way, we are going to prove for every even number $m \geq 2$, $f \in \mathrm{Pol}_m$, $x \in E_\Delta$ and $t \geq 0$, that

$$P_t|f|(x) = E_x[|f(X)|] < \infty$$

and that M_t^f is a *true* martingale. After this, Theorem 2.5 (3) $\Longrightarrow$ (1) gives us the wanted result. We start by showing that $P_t|f|(x) = E_x[|f(X_t)|] < \infty$. First we fix a $T > 0$ and an increasing sequence of stopping times $\{T_j\}_{j\in\mathbb{N}}$ with $\lim_{j\to\infty} T_j = \infty$ such that $(M^f_{t\wedge T_j})_{t\geq 0}$ are martingales for all $f \in \mathrm{Pol}_m$. We have $f_i(x) = x_i$ and set

$$F(x) \coloneqq 1 + \sum_{i=1}^{n} f_i(x)^m.$$

Notice that $F(x)$ is an always positive polynomial with degree m, meaning that $F(x) \in Pol_m$ and hence $(M^F_{t\wedge T_j})_{t\geq 0}$ are martingales. By assumption $\mathcal{G}(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$, so $\mathcal{G}f(x) \subset \mathrm{Pol}_m$ when $f \in \mathrm{Pol}_m$ and it is not hard to see that for some finite constant K

$$|\mathcal{G}f(x)| \leq KF(x)$$

for all $x \in E_\Delta$. Using the above inequality and the martingale property we get the following estimate

$$\begin{aligned}\mathbb{E}_x\left[F(X_{t\wedge T_j})\right] &= F(x) + \mathbb{E}_x\left[\int_0^{t\wedge T_j} \mathcal{G}F(X_u)du\right]\\ &\leq F(x) + K\mathbb{E}_x\left[\int_0^{t\wedge T_j} F(X_u)du\right]\\ &\leq F(x) + K\mathbb{E}_x\left[\int_0^{t} F(X_{u\wedge T_j})du\right].\end{aligned}$$

Using Fubini we can change the order of expectation and integral, then Gronwall's lemma (Appendix A, A.6) gives us the following inequality

$$\mathbb{E}_x\left[F(X_{t\wedge T_j})\right] \leq F(x)\, e^{Kt} \tag{2.2.5}$$

for all $t \leq T$, $j \in \mathbb{N}$ and $x \in E_\Delta$. Since $\lim_{j\to\infty} T_j = \infty$, we have $F(X_t) = F(X_{t\wedge T_j})$ as $j \to \infty$. Thus we can use Fatou's lemma (since F is always positive) to conclude that

$$\begin{aligned}\mathbb{E}_x[F(X_t)] &= \mathbb{E}_x\left[\lim_{j\to\infty} F\left(X_{t\wedge T_j}\right)\right]\\ &\leq \liminf_{j\to\infty} \mathbb{E}_x\left[F\left(X_{t\wedge T_j}\right)\right] \leq F(x)\, e^{Kt} < \infty,\end{aligned} \tag{2.2.6}$$

which shows that

$$P_t|f(x)| = E_x[|f(X_t)|] < \infty$$

for $f \in \mathrm{Pol}_m$.

We move on to showing that for any $f \in \mathrm{Pol}_m$ and $x \in E_\Delta$, M_t^f is a true martingale. We do this by showing that for each $f \in \mathrm{Pol}_m$ and $x \in E_\Delta$

$$\mathbb{E}_x \left[\sup_{t \le T} \left| M_t^f \right| \right] < \infty. \tag{2.2.7}$$

This allows us to use the dominated convergence theorem to show that M_s^f is a true martingale,

$$M_s^f = \lim_{j\to\infty} M_{s\wedge T_j}^f = \lim_{j\to\infty} \mathbb{E}\left[M_{t\wedge T_j}^f | \mathcal{F}_s \right] = \mathbb{E}\left[M_t^f | \mathcal{F}_s \right].$$

In order to show (2.2.7), we let $f \in \mathrm{Pol}_k$ for $k < m$ be fixed and set $p = m/k$. Also in this case there exists a constant K such that following inequalities hold

$$|f(x)|^p \le KF(x) \text{ and } |\mathcal{G}f(x)|^p \le KF(x)$$

for all $x \in E_\Delta$. Using inequality

$$|a+b|^p \le 2^{p-1}\left(|a|^p + |b|^p\right)$$

for $p \ge 0$, we arrive to the following estimate

$$\begin{aligned} |M_{t\wedge T_j}^f|^p &= \left| f\left(X_{t\wedge T_j}\right) - f(x) - \int_0^{t\wedge T_j} \mathcal{G}f\left(X_u\right) du \right|^p \\ &\le C\left(F\left(X_{t\wedge T_j}\right) + F(x) + \int_0^t F\left(X_u\right) du \right) \\ &\le C\left(F\left(X_{t\wedge T_j}\right) + F(x) + \int_0^t F\left(X_{u\wedge T_j}\right) du \right), \end{aligned} \tag{2.2.8}$$

for a positive constant C (depending on p) and $t \le T$. Next we take expectation and use (2.2.5) to get for each fixed x

$$\mathbb{E}_x \left[\left| M_{t\wedge T_j}^f \right|^p \right] \le C_x$$

for all $j \in \mathbb{N}$ and $t \le T$. Further, by Doob's maximal L^p-inequality (1.1.1), for $p > 1$, we have for all j that

$$\mathbb{E}_x \left[\sup_{t \le T} \left| M_{t\wedge T_j}^f \right|^p \right] \le C\mathbb{E}_x \left[\left| M_{T\wedge T_j}^f \right|^p \right] \le C_x.$$

We can see that the left hand side is increasing in j and $M_{t\wedge T_j}^f \to M_t^f$ as $j \to \infty$, thus monotone convergence gives us (2.2.7) for $k < m$, and in particular

$$\sup_{t\le T} \left| M_t^f \right| \in L^p. \tag{2.2.9}$$

Finally we take a look at the case $k = m$. We use the fact that $q = m/2$ is by assumption an integer. We consider the polynomial $f(x) = f_i(x)^q = x_i^q$ for $i \in \{1, ..., n\}$. Also, to make notation clearer, we write $N = M^f$. Estimating again gives

$$\begin{aligned} f(X_t)^2 &= \left(N_t + f(x) + \int_0^t \mathcal{G} f(X_u)\, du \right)^2 \\ &\leq C \left(N_t^2 + f(x)^2 + \int_0^t |\mathcal{G} f(X_u)|^2 du \right) \\ &\leq C \left(N_t^2 + f(x)^2 + \int_0^t F(X_u)\, du \right). \end{aligned}$$

We take the supremum on both sides

$$\sup_{t \leq T} f(X_t)^2 \leq C \left(\sup_{t \leq T} N_t^2 + f(x)^2 + \int_0^T F(X_u)\, du \right).$$

It follows from (2.2.9) that $\sup_{t \leq T} |N_t| \in L^2$ and from (2.2.6) that $\mathbb{E}_x \left[\int_0^T F(X_u)\, du \right] < \infty$. So the right hand side is integrable. Since $f(x) = f_i(x)^q$ we just sum over all i to get the integrability of $\sup_{t \leq T} F(X_t)$,which implies (2.2.7) for all $f \in \mathrm{Pol}_m$ and we are done. $\square$

We end this section by giving a couple of remarks about the conditions in the previous theorem.

Remark 2.8. It is worth noticing that the condition $m \geq 2$ is necessary for M^f to be a true martingale in the proof of Theorem 2.7. It turns out that for example the inverse 3-dimensional Bessel process defined as $\frac{1}{\|B\|}$, where B denotes the 3-dimensional Brownian motion starting at $B_0 \neq 0$, has an extended generator that maps Pol_1 to Pol_0, while

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G} f(X_s)\, ds$$

is a strictly local martingale (not a true martingale) for $f \in \mathrm{Pol}_1$. Indeed, the 3-dimensional Bessel process is the solution to the stochastic differential equation

$$dX_t = -X_t^2 dW_t, \quad X_0 = \frac{1}{\|B_0\|},$$

where W is a 1-dimensional standard Brownian motion. The process has extended generator given by

$$\mathcal{G} f(x) = \frac{1}{2} x^4 \frac{d^2 f(x)}{dx^2}$$

and it is clear that for a first degree polynomial f, we have $\mathcal{G}f(x) = 0$. On the other hand as

$$X_t - x - \int_0^t \mathcal{G}X_s ds = X_t - x,$$

is a strict local martingale, see [17, Example 7.10], X is not a 1-polynomial process (Theorem 2.5 (1) $\to$ (3)).

Remark 2.9. We also need m-polynomial processes to be k-polynomial for all $k \in \{0, ..., m\}$, as required in the Definition 2.1. Meaning that we implicitly exclude process that have an extended generator that maps polynomials of degree $k < m$ to polynomials of degree greater than $k \leq m$, while $\mathcal{G}(\mathrm{Pol}_m) \subset \mathrm{Pol}_m$ holds true. Consider for example the process

$$dX_t = \left(\frac{1}{2} - bX_t + \frac{1}{2}X_t^2\right) dt + \sqrt{X_t^2 (1 - X_t)} dW_t, \qquad X_0 = x \in [0, 1]$$

where $b \geq 1$, W is a standard Brownian motion and the state space is the interval $E = [0, 1]$. The generator of X is

$$\mathcal{G}f(x) = \left(\frac{1}{2} - bx + \frac{1}{2}x^2\right) \frac{df(x)}{dx} + \frac{1}{2}x^2 (1 - x) \frac{d^2 f(x)}{dx^2}.$$

We see that $\mathcal{G}(\mathrm{Pol}_1) \subset \mathrm{Pol}_2$, while $\mathcal{G}(\mathrm{Pol}_2) \subset \mathrm{Pol}_2$. It can be shown (due to compactness of E) that M^f is a true martingale for $f \in \mathrm{Pol}_2$ and thus $P_t(\mathrm{Pol}_2) \subset \mathrm{Pol}_2$, but $P_t(\mathrm{Pol}_1) \not\subseteq \mathrm{Pol}_1$ by the same arguments as in the proof of Theorem 2.5. This means that X_t is not a 1-polynomial process and therefore not a 2-polynomial process either.

In this section we have given the definition of polynomial processes and shown the useful connection to matrix exponentials. We still need to establish conditions which guarantee that X is an m-polynomial process for *all* $m \geq 2$, not just for even numbers. This is done in the next section, where we characterize polynomial processes as special semimartingales with particular characteristics.

2.3. Semimartingales and Polynomial Processes

We now turn to special semimartingales and their connection to polynomial processes. Semimartingales play a huge role in financial mathematics, as a process can model an asset price in a fair market with no arbitrage opportunity only if it is a semimartingale. Characterization of polynomial processes as special semimartingales allows us to define conditions that make it possible to check whether a Markov process X is an m-polynomial process for all $m \geq 2$, not just for even numbers. First we recall the definition of a special semimartingale. A real valued process Y is called a special semimartingale if it can be decomposed as

$$Y = Y_0 + M + A,$$

where M is a local martingale and A is a predictable process of finite variation. Furthermore a d-dimensional process $Y = (Y^i)_{1\leq i\leq d}$ is a special semimartingale iff all of its components Y^i are special semimartingales. We denote the characteristics triplet of a special semimartingale with (A, C, v) where:

- $A = (A_{ij})_{1\leq i\leq d}$ is the predictable process with finite variation.
- $C = (C_{ij})_{1\leq i,j\leq d}$ is the quadratic variation of the continuous local martingale part X^c of X, i.e. $C_{ij} = \langle X^{i,c}, X^{j,c}\rangle$.
- v is the (predictable) compensating measure of the jump measure $\mu = \mu^X$of X, as defined in Chapter 1.

In the next proposition we will see that the process $X1_{\{t<T_\Delta\}}$ is a special semimartingale whose characteristics are polynomials in X, if the generator of X satisfies the conditions in the Theorem 2.7. Furthermore, by an application of Itô's formula we can figure out the explicit form of the extended generator. These results enable us to later state the conditions which guarantee that X is an m-polynomial for all $m \geq 2$, not just for even m as in Theorem 2.7.

To help formulate the proposition we set

$$Y_t := X_t 1_{\{t<T_\Delta\}} = \left(f_1\left(X_t\right), ..., f_n\left(X_t\right)\right)^\top,$$

where $f_i(x)$ is the i'th component of x and $f_i(\Delta) = 0$. By the notation introduced in Section 2.1 this corresponds to $f_i(x) = f_{\mathbf{k}}(x)$, when $\mathbf{k} = e_i$ is the standard basis vector in $\mathbb{R}^n$. We write C^2_m for twice continuously differentiable functions $g : E \to \mathbb{R}$ for which there exists a constant $\tilde{C}$ such that

$$|g(x)| + \sum_{i=1}^{n} |D_i g(x)| + \sum_{i,j=1}^{n} |D_{ij} g(x)| \leq \tilde{C}\left(1 + \|x\|^m\right). \tag{2.3.1}$$

The inequality above is called the polynomial growth condition.

Proposition 2.10. *Let X be a Markov process with state space E_Δ and let $m \geq 2$. Then the following assertions are equivalent:*

(1) Pol_k lies in the domain of $\mathcal{G}$ and $\mathcal{G}\mathrm{Pol}_k \subseteq \mathrm{Pol}_k$ for all $k \in \{0, ..., m\}$.
(2) $X_t 1_{\{t\leq T_\Delta\}}$ is a semimartingale with respect to the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_x)$. Furthermore

$$\mathbb{P}_x\left[t \leq T_\Delta\right] = e^{-\gamma t}$$

for some constant $\gamma \geq 0$ called the killing rate and the semimartingale characteristics (B, C, v) associated with the truncation function $h(\xi) = \xi$ satisfy the following

$$B_{t,i} = \int_0^t b_i\left(X_s\right) ds, \tag{2.3.2}$$

$$C_{t,ij} + \int_0^t \int_{\mathbb{R}^n} \xi_i \xi_j v(ds, d\xi) = \int_0^t a_{ij}(X_s) ds, \tag{2.3.3}$$

where $b_i \in \mathrm{Pol}_1$ and $a_{ij} \in \mathrm{Pol}_2$. Moreover, the characteristics C and v can be written as

$$C_{t,ij} = \int_0^t c_{ij}(X_s) ds, \quad v(\omega; dt, d\xi) = K(X_t(\omega), d\xi) dt, \tag{2.3.4}$$

where c is a predictable process and K is a predictable random measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ satisfying $K(x, \{0\}) = 0$ and $\int_{\{\|\xi\| \leq 1\}} \|\xi\|^2 K(x, d\xi) < \infty$. Finally, for all $|\mathbf{k}| \in \{3, ..., m\}$ we have

$$\int_{\mathbb{R}^n} \xi^{\mathbf{k}} K(x, d\xi) = \sum_{|\mathbf{j}|=0}^{|\mathbf{k}|} \alpha_{\mathbf{j}} f^{\mathbf{j}}(x), \tag{2.3.5}$$

where $\alpha_{\mathbf{j}}$ denote some finite coefficients.

(3) C_m^2 lies in the domain of the extended generator of X, and for all $g \in C_m^2$, the generator is given by

$$\begin{aligned} \mathcal{G}g(x) = &\sum_{i=1}^n D_i g(x) \left(b_i(x) + \gamma f_i(x)\right) + \frac{1}{2} \sum_{i,j=1}^n D_{ij} g(x) \left(c_{ij}(x) - \gamma\right) \\ &+ \int_{\mathbb{R}^n} \left(g(x+\xi) - g(x) - \sum_{i=1}^n D_i g(x) \xi_i \right) \\ &\times \left(K(x, d\xi) - \gamma f_{\mathbf{0}}(x) \delta_{-x}(d\xi)\right), \end{aligned} \tag{2.3.6}$$

where γ and (b, c, K) satisfy the conditions of (2) and $f_{\mathbf{0}}(x) = 1 - 1_{\Delta}(x)$.

All the conditions above imply that Y is a special semimartingale.

Proof. We start by showing (1) $\Longrightarrow$ (2). Since Pol_k lies in the domain of $\mathcal{G}$ for all $k \in \{1, ..., m\}$, it follows that for all $f \in \mathrm{Pol}_m$

$$M_t^f = f(X_t) - f(x) - \int_0^t \mathcal{G} f(X_s)\, ds \tag{2.3.7}$$

is a *local* martingale. By rearranging the terms we can see that $f(X_t)$ is a sum of a local martingale M_t^f and a predictable process $\int_0^t \mathcal{G} f(X_s) ds$, so $f(X_t)$ is a special semimartingale. The convention $f(\Delta) = 0$ leads to $f(X_t) \equiv f(X_t) 1_{t<T_\Delta}$ and we note that $f(X_t)$ has cádlág paths, implying $\lim_{t \uparrow T_\Delta} |f(X_t)| < \infty$, so $f(X)$ cannot explode. By setting $f_i(x) = x_i$ for $i \in \{0, ..., n\}$, we get that $Y_t = X_t 1_{t<T_\Delta}$ is an (n-dimensional) special semimartingale. Next we take a look at (2.3.7) with $f = f_{\mathbf{0}}$, where $f_{\mathbf{0}}(X_t)$ is equal to one for $t < T_\Delta$ and zero elsewhere i.e. $f_{\mathbf{0}}(X_t) = 1_{\{t<T_\Delta\}}$. As $\mathcal{G} f_{\mathbf{0}} \in \mathrm{Pol}_0$, there

exists a constant γ such that $\gamma 1_{\{t<T_\Delta\}} := -\mathcal{G} f_{\mathbf{0}}(X_t)$ and $M^{f_{\mathbf{0}}}$ is a true martingale, we get the following,

$$\begin{aligned}\mathbb{P}_x\left(t<T_\Delta\right) =&\mathbb{E}_x\left[1_{t<T_j}\right] = \mathbb{E}_x\left[f_{\mathbf{0}}(X_t)\right] = 1+\int_0^t \mathbb{E}_x\left[\mathcal{G} f_{\mathbf{0}}(X_s)\right]ds\\ =&1-\gamma\int_0^t \mathbb{E}_x\left[1_{s\wedge T_j}\right]ds.\end{aligned}$$

Which implies that $\mathbb{P}_x(t<T_\Delta)=e^{-\gamma t}$.

We now turn to the characteristics (B, C, v) of Y with respect to the truncation function $\chi(\xi)=\xi$. We apply (generalized) Itô's formula (Theorem 1.49) to $f_{\mathbf{k}}(X_t)$ for $k=|\mathbf{k}|\in\{1,...,m\}$ and $X_{t,i}=x_i+M_t^{f_i}+B_{t,i}$ for $i\in\{1,...,n\}$ to get the following (see appendix A for more details):

$$\begin{aligned}f_{\mathbf{k}}(X_t) = f_{\mathbf{k}}(x) &+ \int_0^t\sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-})dM_s^{f_i} + \int_0^t\sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-})dB_{s,i}\\ &+\frac{1}{2}\int_0^t\sum_{i,j=1}^n D_{ij} f_{\mathbf{k}}(X_{s-})dC_{s,ij} + \int_0^t\int_{\mathbb{R}^n} W(s,\xi)\mu^Y(ds,d\xi),\end{aligned} \tag{2.3.8}$$

where μ^Ydenotes the random measure of the jumps of Y and

$$W(s,\xi) := \sum_{|\mathbf{j}|=2}^k \binom{\mathbf{k}}{\mathbf{j}} f_{\mathbf{k}-\mathbf{j}}\left(X_s\right)\xi^{\mathbf{j}}.$$

As $f_{\mathbf{k}}\left(X\right)$ is a special semimartingale it can thus be decomposed uniquely into a local martingale and a predictable finite variation process. We take a look at each term on the right side of Equation (2.3.8) in order to figure out which of them are local martingales. The first term is just a function of the initial value of the process and thus not a local martingale. The second term is an integral of a càdlàg adapted process with respect to a local martingale and thus a local martingale itself . The third and fourth terms are predictable processes of finite variation, in particular they are of locally integrable variation by [14, Lemma I.3.10]. For the last term it is enough to notice that it is of finite variation and for special semimartingales the finite variation part is of locally integrable variation (see [14, Proposition I.4.23]. We know from the section on semimartingales, see equation (1.5.4), that the jump measure μ has a compensator $\int_0^t\int_{\mathbb{R}^n} W\left(s,\xi\right)v\left(ds,d\xi\right)$ such that

$$\int_0^t\int_{\mathbb{R}^n} W\left(s,\xi\right)\mu^Y\left(ds,d\xi\right) - \int_0^t\int_{\mathbb{R}^n} W\left(s,\xi\right)v\left(ds,d\xi\right)$$

is a local martingale. Rearranging Equation (2.3.8) we get

$$\begin{aligned}
&\int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_s) dM_s^{f_i} + \int_0^t \int_{\mathbb{R}^n} W(s,\xi)\left(\mu^Y(ds,d\xi) - v(ds,d\xi)\right) \\
&= f_{\mathbf{k}}(X_t) - f_{\mathbf{k}}(x) - \int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) dB_{s,i} \\
&- \frac{1}{2}\int_0^t \sum_{i,j=1}^n D_{i,j} f_{\mathbf{k}}(X_{s-}) dC_{s,ij} - \int_0^t \int_{\mathbb{R}^n} W(s,\xi) v(ds,d\xi),
\end{aligned}$$

where the left hand side is a local martingale. Now we can combine (2.3.7) with (2.3.8) and use the fact that local martingale parts have to be equal due to the uniqueness of the decomposition, so we get

$$\begin{aligned}
M_t^{f_{\mathbf{k}}} = &\int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_s) dM_s^{f_i} + \int_0^t \int_{\mathbb{R}^n} W(s,\xi)\left(\mu^Y(ds,d\xi) - v(ds,d\xi)\right) \\
=& f_{\mathbf{k}}(X_t) - f_{\mathbf{k}}(x) - \int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) dB_{s,i} \\
&- \frac{1}{2}\int_0^t \sum_{i,j=1}^n D_{i,j} f_{\mathbf{k}}(X_{s-}) dC_{s,ij} - \int_0^t \int_{\mathbb{R}^n} W(s,\xi) v(ds,d\xi).
\end{aligned}$$

It then follows that the extended generator has the form

$$\begin{aligned}
\int_0^t \mathcal{G} f_{\mathbf{k}}(X_s)\, ds = &\int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) dB_{s,i} + \frac{1}{2}\int_0^t \sum_{i,j=1}^n D_{ij} f_{\mathbf{k}}(X_{s-}) dC_{s,ij} \\
&+ \int_0^t \int_{\mathbb{R}^n} W(s,\xi) v(ds,d\xi).
\end{aligned} \tag{2.3.9}$$

In order to find the characteristics of $B_{t,i}$ we take a look at (2.3.9) with $|\mathbf{k}| = 1$, that is the polynomials $f_i(x) = x_i$ for $i \in \{1, ..., n\}$. We find that the last to terms on the right hand side vanish and the first integral on the right hand side is just $B_{t,i}$,so we have

$$\int_0^t \mathcal{G} f_i(X_s)\, ds = B_{t,i}. \tag{2.3.10}$$

Defining $b_i(x) \coloneqq \mathcal{G} f_i(x)$ implies then that $b_i \in Pol_1$, as $\mathcal{G}(\mathrm{Pol}_1) \subset \mathrm{Pol}_1$ and we have thus shown (2.3.2).

Next we take a look at (2.3.9) for quadratic polynomials $f_{ij}(x) = x_i x_j$ for $i, j \in \{1, ..., n\}$. We set

$$\int_0^t a_{ij}(X_s)\, ds := \int_0^t \mathcal{G} f_{ij}(X_s)\, ds - \int_0^t \sum_{k=1}^n D_k f_{ij}(X_{s-}) dB_{s,i}$$

and notice that both $\mathcal{G} f_{ij}(x)$ and $D_i f_{ij}(x) b_i(x) = f_j(x) b_i(x)$ are in Pol_2, which implies that $a_{ij} \in \mathrm{Pol}_2$ as well. For the second term on the right hand side of (2.3.9) we see that

$$\frac{1}{2} \int_0^t \sum_{k,l=1}^n D_{kl} f_{ij}(X_{s-}) dC_{s,ij} = C_{t,ij}$$

and since $W(s,\xi) := \sum_{|\mathbf{j}|=2}^k \binom{\mathbf{k}}{\mathbf{j}} f_{\mathbf{k}-\mathbf{j}}(X_s)\, \xi^{\mathbf{j}}$, it is easy to see that it simplifies to $W(s,\xi) = \xi^{\mathbf{j}} = \xi_i \xi_j$ when $k = |\mathbf{k}| = 2$. Thus we end up with

$$\int_0^t a_{ij}(X_s)\, ds = C_{t,ij} + \int_0^t \int_{\mathbb{R}^n} \xi_i \xi_j v(ds, d\xi), \tag{2.3.11}$$

which is (2.3.3) as wanted.

We move on to proving the equalities (2.3.4). We define

$$A'_t(\omega) = \int_0^t \int_{\mathbb{R}^n} \|\xi\|^2\, v(\omega; ds, d\xi)$$

and using arguments from [14, Proposition II.2.9.b], we know there exists a random measure $K'(\omega, t; d\xi)$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that

$$v(\omega; ds, d\xi) = K'(\omega, t; d\xi)\, dA_t(\omega).$$

Furthermore by (2.3.3) with $i = j$ we have

$$\sum_{i=1}^n C_{t,ii}(\omega) + A'_t(\omega) = \sum_{i=1}^n \int_0^t a_{ii}(X_s(\omega))\, ds =: \int_0^t a_s(\omega)\, ds$$

and since C_{ii} and A are non-negative increasing processes, they are also absolutely continuous with respect to the Lebesque measure. Now [14, Proposition I.3.13] implies the existence of predictable processes $\tilde{c}_{ii}$ and H such that $C_{t,ii} = \int_0^t c_{s,ii} ds$ and $A'_t = \int_0^t H_s ds$. We thus have a predictable random measure

$$\tilde{K}_{\omega,t}(d\xi) = H_t(\omega)\, K'(\omega, t; d\xi)$$

satisfying $v(\omega; dt, d\xi) = \tilde{K}(d\xi)\, dt$ almost surely. The equation (2.3.11) now becomes

$$C_{t,ij} = \int_0^t \left(a_{ij}(X_s) - \int_{\mathbb{R}^n} \xi_i \xi_j \tilde{K}_{\omega,s}(d\xi) \right) ds,$$

implying that C_{ij} for $i \neq j$ is absolutely continuous with respect to the Lebesque measure and can therefore be written as $C_{t,ij} = \int_0^t \tilde{c}_{s,ij} ds$. By [6, Theorem 6.27] we can choose homogeneous versions for the processes $\tilde{c}$ and $\tilde{K}$ such that $\tilde{c}_t = c\left(X_t\left(\omega\right)\right)$ and $\tilde{K}_{\omega,t} = K\left(X_t\left(\omega\right), d\xi\right)$.

Finally we are left with (2.3.5). We take a look at (2.3.9) again, this time for $\|\mathbf{k}\| = k \geq 3$, and notice that we can rewrite it using previous results getting

$$\begin{aligned}\mathcal{G} f_{\mathbf{k}}\left(x\right) &= \sum_{i=1}^{n} D_i f_{\mathbf{k}}(x) b_i\left(x\right) + \frac{1}{2}\sum_{i,j=1}^{n} D_{ij} f_{\mathbf{k}}(x)\left(c_{ij}\left(x\right) + \int_{\mathbb{R}^n} \xi_i \xi_j K\left(x, d\xi\right)\right) \\ &\quad + \int_{\mathbb{R}^n}\left(\sum_{|\mathbf{j}|=3}^{k}\binom{\mathbf{k}}{\mathbf{j}}\right) f_{\mathbf{k}-\mathbf{j}}\left(x\right)\xi^{\mathbf{j}} K\left(x, d\xi\right).\end{aligned}$$

As $\mathcal{G} f_{\mathbf{k}}\left(x\right)$, $D_i f_{\mathbf{k}}\left(x\right) b\left(x\right)$ and $D_{ij} f_{\mathbf{k}}\left(x\right)\left(c_{ij}\left(x\right) + \int_{\mathbb{R}^n} \xi_i \xi_j K\left(x, d\xi\right)\right)$ all lie in Pol_k for all $k = |\mathbf{k}| \leq m$, then

$$\int_{\mathbb{R}^n} \xi^{\mathbf{k}} K\left(x, d\xi\right) = \sum_{|\mathbf{j}|=0}^{|\mathbf{k}|} \alpha_{\mathbf{j}} f^{\mathbf{j}}\left(x\right),$$

for some finite coefficients $\alpha_{\mathbf{j}}$, simply follows by induction, which is (2.3.5) as wanted.

Next we prove the implication (2) $\Longrightarrow$ (3), that is the characteristics of the extended generator of X. In order to simplify the notation we set

$$V\left(Y_s, \xi\right) \coloneqq g\left(Y_s + \xi\right) - g\left(Y_s\right) - \sum_{i=1}^{n} D_i g\left(Y_s\right)\xi_i$$

for $g \in C_m^2$. We know from (2.3.3) and (2.3.5) that $\int \|\xi\|^k K\left(X_s, d\xi\right) < \infty$ for all $k \in \{1, ..., m\}$, it follows that

$$\int_{\mathbb{R}^n} \left|V\left(Y_s, \xi\right)\right| K\left(X_s, d\xi\right) \leq h\left(Y_s\right) + H\left(Y_s\right)\int_{\mathbb{R}^n}\left(\|\xi\|^2 \wedge \|\xi\|^m\right) K\left(X_s, d\xi\right) < \infty,$$

where h and H denote some positive finite- valued functions. So we have that the process $\int_0^{\cdot} \int V\left(Y_s, \xi\right) K\left(X_s, d\xi\right)$ is of locally integrable variation and Itô's formula then implies that

$$\begin{aligned}M_t^g &\coloneqq g\left(Y_t\right) - g\left(x\right) - \int_0^t \sum_{i=1}^{n} D_i g\left(Y_s\right) b_i\left(X_s\right) ds - \int_0^t \frac{1}{2}\sum_{i,j=1}^{n} D_{ij} g\left(Y_s\right) c_{ij}\left(X_s\right) ds \\ &\quad - \int_0^t \int_{\mathbb{R}^n} V\left(Y_s, \xi\right) K\left(X_s, d\xi\right) ds\end{aligned}$$

is a local martingale. We also see that

$$-1_{\{T_\Delta \le t\}} + \gamma\left(T_\Delta \wedge t\right) = 1_{\{t<T_\Delta\}} - 1 + \int_0^t \gamma 1_{\{s<T\}} ds$$

is a true martingale. Remembering that $\mathbb{E}_x\left[1_{\{t<T_\Delta\}}\right] = \mathbb{P}_x\left(t < T_\Delta\right) = e^{-\gamma t}$, we indeed have

$$\mathbb{E}_x\left[1_{\{t<T_\Delta\}} - 1 + \int_0^t \gamma 1_{\{s<T\}} ds\right] = e^{-\gamma t} - 1 + \int_0^t \gamma e^{-\gamma s} ds = 0.$$

Denoting the right-hand side of (2.3.6) with $\mathcal{G}^{\#}$, we now need to prove that

$$M_t^{\#g} := g\left(X_t\right) - g\left(x\right) - \int_0^t \mathcal{G}^{\#} g\left(X_s\right) ds$$

is a local martingale. By the definition of $\mathcal{G}^{\#}$ we have

$$M_t^{\#g} = M_t^g - g\left(0\right)\left(1_{\{T_\Delta \le t\}} - \gamma\left(t \wedge T_\Delta\right)\right)$$

and since both terms on the right-hand side are local martingales, $M_t^{\#g}$ is a local martingale.

Finally we prove the implication (3) $\Longrightarrow$ (1). We obviously have $\mathrm{Pol}_m \subset C_m^2$ and thus Pol_k lies in the domain of $\mathcal{G}$ for all $k \in \{0, ..., m\}$. Also due to the assumptions on the characteristics of the process, we have $\mathcal{G}\left(\mathrm{Pol}_k\right) \subseteq \mathrm{Pol}_k$ for all $k \in \{0, ...m\}$. $\square$

Since every m-polynomial process satisfies the conditions in the Theorem 2.7, we have the following corollary as a direct result of Proposition 2.10.

COROLLARY 2.11. *Let X be an m-polynomial process with $m \geq 2$. Then the process $Y_t = \left(X_t 1_{\{t<T_\Delta\}}\right)$ is a special semimartingale satisfying the conditions (2.2.5)-(2.3.5) and its extended generator is of the form* (2.3.6).

Theorem 2.7 together with Proposition 2.10 provide the opposite direction, whenever m is an even number. The general case will be proven in the subsequent theorem, where necessary conditions are given for the compensator of the jump measure, so that the converse direction holds for all $m \geq 2$. In order to prove it, we will need a maximal inequality for semimartingales satisfying conditions (2.3.2)-(2.3.5). The lemma dealing with this inequality is given afterwards.

THEOREM 2.12. *Let X be a time-homogeneous Markov process with state space E_Δ and let $m \geq 2$. Suppose that $Y_t = \left(X_t 1_{\{t<T_\Delta\}}\right)$ is a semimartingale satisfying the conditions (2.3.2)-(2.3.5) or equivalently that $C_m^2 \subset \mathcal{D}_{\mathcal{G}}$ and that its extended generator $\mathcal{G}$ is given by (2.3.6). If one of the following conditions is satisfied, then X is an m-polynomial process.*

$$\mathbb{E}_x\left[\int_{\mathbb{R}^n} \left\|\xi\right\|^m K\left(X_t, d\xi\right)\right] < \infty \tag{2.3.12}$$

or

$$\int_{\mathbb{R}^n} \|\xi\|^m K(X_t, d\xi) \le \tilde{C}(1 + \|Y_t\|^m), \tag{2.3.13}$$

for almost all $t \ge 0$ and some constant $\tilde{C}$.

Proof. As in the proof of Theorem 2.7, it is enough to show that for each $f \in \mathrm{Pol}_m, x \in E_\Delta$ and every fixed $t \ge 0$,

$$P_t |f|(x) = \mathbb{E}_x[|f(X_t)|] < \infty$$

and

$$\mathbb{E}_x\left[\sup_{s\le t} \left|M_s^f\right|\right] < \infty,$$

which shows that M_s^f is a true martingale and by Theorem 2.5 X is an m-polynomial. The inequalities follow from the Lemma 2.14 below together with one of the assumptions (2.3.12) or (2.3.13). □

Remark 2.13. The theorem above introduces a tool to verify that a stochastic process is m-polynomial. One needs to have either the characteristics of the process or its extended generator available and check that the conditions in Proposition 2.10 are satisfied. Checking the conditions for Itô diffusions is simple, since the assumptions in the Theorem 2.12 are trivially satisfied for processes without jumps.

Lemma 2.14. *Fix $t > 0$ and let $m \ge 2$. Let Y be a semimartingale with respect to $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P}_x\right)$ whose characteristics (B, C, v) associated with the truncation function $\chi(\xi) = \xi$ satisfy the conditions (2.3.2)-(2.3.4) in Proposition 2.10. Then there exists a constant $\tilde{C}$ such that*

$$\mathbb{E}_x\left[\sup_{s\le t}\|Y_s\|^m\right] \le \tilde{C}\left(\|x\|^m + 1 + \int_0^t \mathbb{E}_x\left[\int_{\mathbb{R}^n}\|\xi\|^m K(X_s, d\xi)\right]ds + \int_0^t \mathbb{E}_x\left[\|Y_s\|^m\right]ds\right). \tag{2.3.14}$$

In particular, if either of the conditions (2.3.12) or (2.3.13) from Theorem 2.12 is satisfied, then there exists finite constants K and $\tilde{C}$ such that

$$\mathbb{E}_x\left[\sup_{s\le t}\|Y_s\|^m\right] \le K e^{\tilde{C}t}. \tag{2.3.15}$$

Proof. In order to make the notation simpler we only look at the case where Y is a one-dimensional. We know from Proposition 2.10 and from the assumptions on the characteristics of Y, that Y is a special semimartingale. Its unique canonical decomposition is given by $Y = x + M + \int_0^{\cdot} b(X_u)\,du$ where $M = M^{f_1}$. We denote the quadratic variation of the purely discontinuous martingale part of Y with Z,

$$Z_t = \sum_{s \le t} (\Delta Y_s)^2 = \int_0^t \int_{\mathbb{R}^n} \xi^2 \mu^Y(ds, d\xi)\,.$$

We also define two stopping times as follows

$$\begin{aligned} T_j^Y &= \inf\{t \ge 0 \mid |Y_t| \ge j \; or \; |Y_{t-}| \ge j\}\,, \\ T_j^Z &= \inf\{t \ge 0 \mid |Z_t| \ge j \; or \; |Z_{t-}| \ge j\} \end{aligned}$$

and put $T_j = T_j^Y \wedge T_j^Z$. Now we estimate

$$\sup_{s \le t} \left|Y_s^{T_j}\right|^m \le \tilde{C}\left(|x|^m + \sup_{s \le t}\left|M_s^{T_j}\right|^m + \sup_{s \le t}\left|\int_0^{s \wedge T_j} b(X_u)\,du\right|^m\right),$$

where $\tilde{\mathrm{C}}$ denotes some constant that may vary from line to line during the proof. First we deal with with the integral on the right hand side of the inequality above. Since, by assumption, $b \in \mathrm{Pol}_1$ we have

$$\sup_{s \le t}\left|\int_0^{s \wedge T_j} b(X_u)\,du\right|^m \le \sup_{s \le t} \tilde{C}\left(1 + \int_0^{s \wedge T_j} |X_u|^{,m} du\right) \tag{2.3.16}$$

$$\le \tilde{C}\left(1 + \int_0^t \left|Y_u^{T_j}\right|^m du\right), \tag{2.3.17}$$

where the last inequality follows as $Y_u = X_u$ for $u < T_\Delta$. Next we consider $\sup_{s \le t}\left|M_s^{T_j}\right|$. We can use Burkholder-Davis-Gundy inequality (see appendix) to get

$$\mathbb{E}_x\left[\sup_{s \le t}\left|M_s^{T_j}\right|^m\right] \le \tilde{C}\mathbb{E}_x\left[[M, M]_{t \wedge T_j}^{\frac{m}{2}}\right] \le \tilde{C}\mathbb{E}_x\left[C_{t \wedge T_j}^{\frac{m}{2}} + Z_{t \wedge T_j}^{\frac{m}{2}}\right]. \tag{2.3.18}$$

We know that C satisfies (2.3.3), thus we can estimate it by $C_t \le \int_0^t a(X_s)\,ds$,where $a \in \mathrm{Pol}_2$ is non-negative. We then get the following estimate

$$\mathbb{E}_x\left[C_{t \wedge T_j}^{\frac{m}{2}}\right] \le \tilde{C}\left(1 + \int_0^t \mathbb{E}_x\left[\left|Y_s^{T_j}\right|^m\right] ds\right).$$

We are left with $Z_{t \wedge T_j}^{\frac{m}{2}}$. Here the same approach as [15, Lemma 5.1] is used. Since Z is purely discontinuous, non-decreasing and $\Delta Z_s = |\Delta Y_s|^2$, we get

$$\begin{aligned}Z_{t\wedge T_j} &= \sum_{s\le t\wedge T_j} \left(Z_{s-}+\Delta Z_s\right)^{\frac{m}{2}} - \left(Z_{s-}\right)^{\frac{m}{2}}\\ &= \int_0^{t\wedge T_j}\int_{\mathbb{R}^n}\left(\left(Z_{s-}+\xi^2\right)^{\frac{m}{2}} - \left(Z_{s-}\right)^{\frac{m}{2}}\right)\mu^Y\left(ds,d\xi\right).\end{aligned}$$

Taking expectation and remembering that v is the predictable compensator of μ^Y, we have

$$\mathbb{E}_x\left[Z_{t\wedge T_j}^{\frac{m}{2}}\right] = \mathbb{E}_x\left[\int_0^{t\wedge T_j}\int_{\mathbb{R}^n}\left(\left(Z_{s-}+\xi^2\right)^{\frac{m}{2}} - \left(Z_{s-}\right)^{\frac{m}{2}}\right) v\left(ds,d\xi\right)\right]. \tag{2.3.19}$$

In the next part of the proof we will use the inequalities (see [15, Lemma 5.1, proof])

$$(z+x)^p - z^p \le 2^{p-1}\left(z^{p-1}x + x^p\right), \tag{2.3.20}$$

$$z^{p-1}x \le \epsilon z^p + \frac{x^p}{\epsilon^{p-1}}, \tag{2.3.21}$$

for $p\ge 1$, $x,z\ge 0$ and $\epsilon>0$. We apply (2.3.20) to the equation (2.3.19) to get the following

$$\mathbb{E}_x\left[Z_{t\wedge T_j}^{\frac{m}{2}}\right] \le \mathbb{E}_x\left[\int_0^{t\wedge T_j}\int_{\mathbb{R}} 2^{\frac{m}{2}-1}\left(Z_{s-}^{\frac{m}{2}-1}\xi^2 + |\xi|^m\right) v\left(ds,d\xi\right)\right].$$

First we check the part with ξ^2. Due to the assumption on $v\left(ds,d\xi\right)$, we can use (2.3.4) and (2.3.5) from Proposition 2.10 to get

$$\begin{aligned}&\mathbb{E}_x\left[\int_0^{t\wedge T_j}\int_{\mathbb{R}} 2^{\frac{m}{2}-1}\left(Z_{s-}^{\frac{m}{2}-1}\xi^2\right) v\left(ds,d\xi\right)\right]\\ &= \mathbb{E}_x\left[\int_0^{t\wedge T_j} 2^{\frac{m}{2}-1} Z_{s-}^{\frac{m}{2}-1}\left(\int_{\mathbb{R}^n}\xi^2 K\left(X_s,d\xi\right)\right) ds\right]\\ &\le \mathbb{E}_x\left[\int_0^{t\wedge T_j} 2^{\frac{m}{2}-1} Z_{s-}^{\frac{m}{2}-1} a\left(X_s\right) ds\right]\\ &\le \mathbb{E}_x\left[\int_0^{t\wedge T_j} \tilde{C}\left(\epsilon Z_s^{\frac{m}{2}} + \frac{1+|Y_s|^m}{\epsilon^{\frac{m}{2}-1}}\right) ds\right],\end{aligned}$$

where we have used that Z is non-negative, $a\in \mathrm{Pol}_2$ and the second inequality (2.3.21). We estimate again

$$\int_0^{t\wedge T_j} Z_s^{\frac{m}{2}}\, ds \le j^{\frac{m}{2}} \wedge Z_{t\wedge T_j}^{\frac{m}{2}},$$

which follows since Z_s is non-decreasing and $Z_{s-} \leq j$ for $s \leq T_j$. Combining the inequalities, we get

$$\mathbb{E}_x\left[Z_{t\wedge T_j}^{\frac{m}{2}}\right] \leq \tilde{C}\epsilon\mathbb{E}_x\left[j^{\frac{m}{2}} \wedge Z_{t\wedge T_j}^{\frac{m}{2}}\right] + \mathbb{E}_x\left[\int_0^{t\wedge T_j} \frac{\tilde{C}}{\epsilon^{\frac{m}{2}-1}}\left(1+|Y_s|^m\right) ds\right]$$
$$+ \mathbb{E}_x\left[\int_0^{t\wedge T_j}\int_{\mathbb{R}} 2^{\frac{m}{2}-1}|\xi|^m K\left(X_s, d\xi\right) ds\right].$$

Choosing $\epsilon = \frac{1}{2\tilde{C}}$ gives

$$\begin{aligned}\frac{1}{2}\mathbb{E}_x\left[Z_{t\wedge T_j}^{\frac{m}{2}}\right] \leq& \mathbb{E}_x\left[Z_{t\wedge T_j}^{\frac{m}{2}}\right] - \frac{1}{2}\mathbb{E}_x\left[j^{\frac{1}{2}} \wedge Z_{t\wedge T_j}^{\frac{m}{2}}\right]\\ \leq& \tilde{C}\mathbb{E}_x\left[\int_0^t\left(1+\left|Y_s^{T_j}\right|^m + \int_{\mathbb{R}}|\xi|^m K\left(X_{s\wedge T_j}, d\xi\right)\right) ds\right].\end{aligned}$$

Putting this together with inequalities (2.3.16) and (2.3.18), we arrive at

$$\mathbb{E}_x\left[\sup_{s\leq t}\left|Y_s^{T_j}\right|^m\right] \leq \tilde{C}\left(|x|^m + 1 + \int_0^t \mathbb{E}_x\left[\int_{\mathbb{R}}|\xi|^m K\left(X_{s\wedge T_j}, d\xi\right)\right] ds + \int_0^t \mathbb{E}_x\left[\left|Y_s^{T_j}\right|^m\right] ds\right) \tag{2.3.22}$$

and by monotone convergence theorem (2.3.14) follows.

For (2.3.15) we notice that under assumptions (2.3.12) or (2.3.13), it follows from (2.3.22) that

$$\mathbb{E}_x\left[\sup_{s\leq t}\left|Y_s^{T_j}\right|^m\right] \leq K + \tilde{C}\int_0^t \mathbb{E}_x\left[\sup_{i\leq s}\left|Y_u^{T_j}\right|^m\right] ds$$

for some finite constants K and $\tilde{C}$. Now the right hand side of above inequality is finite (due to the assumptions) and we have the following estimate

$$\begin{aligned}\sup_{u\leq s}\left|Y_u^{T_j}\right|^m &\leq |x|^m + j^m + \mathbb{E}_x\left[\left|\Delta Y_{s\wedge T_j}\right|^m\right]\\ &\leq |x|^m + j^m + \mathbb{E}_x\left[\int_0^{s\wedge T_j}\int_{\mathbb{R}}|\xi|^m K\left(X_s, d\xi\right) ds\right] < \infty,\end{aligned}$$

thus Gronwall's lemma yields

$$\mathbb{E}_x\left[\sup_{s\leq t}\left|Y_s^{T_j}\right|^m\right] \leq K e^{\tilde{C}t}$$

for all $j \in \mathbb{N}$ and monotone convergence gives the wanted result. $\square$

The last results have been quite technical and also general in nature, as it considers the class of Markov processes that are special semimartingales with triplets satisfying certain conditions. In the next subsection we are going to look at some examples of processes that are polynomial. We will see that the conditions in Proposition 2.10 and in Theorem 2.12 are not always so complicated especially when we consider processes with no jumps or even in the case of Lévy processes that allow for jumps.

2.4. Examples of Polynomial Processes

Let us consider Itô diffusions defined in Definition 1.25. For them the conditions for being polynomial processes are simple to check.

Theorem 2.15. *(Polynomial diffusion) Let X_t be a d-dimensional Itô diffusion satisfying*

$$dX_t = b\left(X_t\right)dt + \sigma\left(X_t\right)dW_t,$$

where W is an m-dimensional Brownian motion, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$. If μ and σ satisfy the following,

$$b\left(x\right) \in \mathrm{Pol}_1$$
$$\sigma\left(x\right)\sigma\left(x\right)^\top \in \mathrm{Pol}_2,$$

then X_t is a polynomial process. We call these processes polynomial diffusions.

Proof. As Itô diffusions are continuous processes, there are no jumps. So by Theorem 2.12 it is enough that the characteristic triplet (B, C, v) satisfy the conditions in Proposition 2.10. The triplet of Itô diffusion is given by $B_t = \int_0^t b\left(X_s\right)ds$, $C_t = \int_0^t \sigma\left(X_s\right)\sigma\left(X_s\right)^\top ds$ and $v = 0$. Clearly X_t is a polynomial process whenever the conditions on b and σ are satisfied. $\square$

This also means that the generator of a polynomial diffusion is given by (3) in Proposition 2.10 (with $\gamma = v = 0$) as

$$\mathcal{A}g\left(x\right) = \sum_{i=1}^{d} b_i\left(x\right)\frac{\partial f}{\partial x_i}\left(x\right) + \frac{1}{2}\sum_{i,j}^{d}\sigma\left(x\right)\sigma\left(x\right)^\top\frac{\partial^2 f}{\partial x_i \partial x_j}\left(x\right).$$

Example 2.16. (Ornstein-Uhlenbeck process) Ornstein-Uhlenbeck process (OU) is mean-reverting stochastic process with applications in both mathematical finance and physical sciences. The Ornstein-Uhlenbeck process X on $\mathbb{R}$ is defined by the following SDE

$$dX_t = \theta\left(\mu - X_t\right)dt + \sigma dW_t, \qquad X_0 = x,$$

where $\theta, \sigma > 0$, $\mu \in \mathbb{R}$ and W is a standard Brownian motion. We have $b\left(X_t\right) = \theta\left(\mu - X_t\right) \in \mathrm{Pol}_1$ and $\sigma\left(X_t\right)^2 = \sigma^2 \in \mathrm{Pol}_0 \subset \mathrm{Pol}_2$, so OU process is a polynomial diffusion. This is also true for vector-valued OU process $X \in \mathbb{R}^d$, with $\mu \in \mathbb{R}^d$, $\theta \in \mathbb{R}^{d\times d}$

and an m-dimensional Brownian motion B_t, with σ an $m \times d$ matrix. In the next section we are going to show how the moments of this process may be calculated by using matrix exponentials.

EXAMPLE 2.17. (Jacobi process) A solution of the SDE

$$dX_t = -\beta (X_t - \theta) \, dt + \sigma \left(\sqrt{X_t (1 - X_t)} \right) dW_t, \quad X_0 = x \in [0, 1],$$

where $\beta, \sigma > 0$, $\theta \in [0, 1]$ and the state space $E = [0, 1]$. Again it is easy to see that $b (X_t) = -\beta (X_t - \theta) \in \mathrm{Pol}_1$ and $\sigma(X_t)^2 = \sigma^2 (X_t - X_t^2) \in \mathrm{Pol}_2$.

EXAMPLE 2.18. (Lévy processes) Let L be a Lévy process on $\mathbb{R}^n$ with characterizing triplet (b, c, v) satisfying the moment condition $\int_{\|\xi\|>1} \|\xi\|^m v (d\xi)$, then the Markov process $X = X_0 + L$ is m-polynomial. From Theorem 1.29 it follows that the characteristics b and c are constant for Lévy processes, which means that they satisfy the conditions in Proposition 2.10 (2) and when the moment condition is satisfied, so does the Lévy measure v. Again by Theorem 2.12 X is an m-polynomial.

2.5. Computation of Moments

One of the important properties of m-polynomial processes is that their moments

$$\mathbb{E}_x \left[X_t^{\mathbf{k}} \right] = P_t x^{\mathbf{k}},$$

can be calculated in a fairly easy and efficient way. We know from Theorem 2.5 that, if X is an m-polynomial process then there exists a linear map A such that the moments of X can simply be calculated by computing e^{tA}. We can define $H : \mathbb{R}^d \to \mathbb{R}^N$ to be a function whose components form a basis of polynomials for Pol_m, i.e. $H (x) = (h_1 (x), ..., h_N (x))$, then for each $f \in \mathrm{Pol}_m$ there exists a unique coordinate vector $\overrightarrow{p}_f = (\alpha_1, ..., \alpha_N)$, such that

$$f (x) = H (x)^\top \overrightarrow{p}_f$$

and

$$\mathcal{G} f (x) = H (x)^\top A \overrightarrow{p}_f,$$

where A is the matrix obtained from the generator $\mathcal{G}$.

THEOREM 2.19. *(Moment formula) Let f be a polynomial of order m with coordinate representation given above and let X be an m-polynomial process. Then for $m \geq 2$ and $0 \leq t \leq T$*

$$\mathbb{E}_x [f (X_T) \mid \mathcal{F}_t] = H (X_t)^\top e^{(T-t)A} \overrightarrow{p}_f.$$

By setting $t = 0$, we get

$$\mathbb{E}_x [f (X_T)] = H (X_0)^\top e^{TA} \overrightarrow{p}_f.$$

PROOF. From Theorem 2.5 we know that

$$M_T^f = f(X_T) - f(x) - \int_0^T \mathcal{G}f(X_s)\,ds$$

is a true martingale. Taking expectation conditioned on $\mathcal{F}_t$ on both sides and rearranging we get

$$\begin{aligned}\mathbb{E}_x\left[f(X_T)\,|\mathcal{F}_t\right] &= f(X_t) - \int_0^t \mathcal{G}f(X_s)\,ds + \mathbb{E}_x\left[\int_0^T \mathcal{G}f(X_s)\,ds|\mathcal{F}_t\right]\\ &= f(X_t) + \mathbb{E}_x\left[\int_t^T \mathcal{G}f(X_s)\,ds|\mathcal{F}_t\right].\end{aligned}$$

Changing to coordinate representation and writing $F(T) = \mathbb{E}_x\left[H(X_T)\,|\mathcal{F}_t\right]$, gives

$$\begin{aligned}F(T)\,\overrightarrow{p}_f &= H(X_t)^\top\,\overrightarrow{p}_f + \mathbb{E}_x\left[\int_t^T H(X_s)^\top A\overrightarrow{p}_f ds|\mathcal{F}_t\right]\\ &= H(X_t)^\top\,\overrightarrow{p}_f + \mathbb{E}_{X_t}\left[\int_0^{T-t} H(X_s)^\top A\right]\overrightarrow{p}_f\\ &= H(X_t)^\top\,\overrightarrow{p}_f + \int_0^{T-t} H(X_t)^\top P_s A ds\,\overrightarrow{p}_f\\ &= H(X_t)^\top\,\overrightarrow{p}_f + H(X_t)^\top\int_0^{T-t} Ae^{As} ds\,\overrightarrow{p}_f\\ &= H(X_t)^\top e^{(T-t)A}\overrightarrow{p}_f.\end{aligned}$$

where we have used (2) from Theorem 2.5 and Markov property of X. $\square$

EXAMPLE 2.20. (OU process) We take again look at the one-dimensional OU process defined in the Example 2.16 with simplification $\mu = 0$. We want to demonstrate how to calculate moments up to order m of the process. The extended generator is given by

$$\mathcal{G}f(x) = -\theta x f'(x) + \frac{\sigma^2}{2} f''(x).$$

Applying $\mathcal{G}$ to the basis of the polynomial vector space i.e. $\{x^0, x^1, ..., x^m\}$, we get the $(m+1) \times (m+1)$ matrix

$$A = \begin{pmatrix} 0 & 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & -\theta & 0 & 3\sigma^2 & & \vdots \\ & 0 & -2\theta & 0 & \ddots & \\ & \vdots & & -3\theta & & \frac{\sigma^2}{2} m(m-1) \\ & & & & \ddots & 0 \\ & & & & & -\theta m \end{pmatrix}.$$

Thus the moment of order $k \leq m$ is given by

$$\mathbb{E}_x \left[X_t^k \right] = P_t x^k = H(X_0)^\top e^{tA} \overrightarrow{p}, \tag{2.5.1}$$

where $\overrightarrow{p}$ is the coordinate vector of x^k, i.e. a vector with number one at the $(k+1)$'th place and rest zero, while $H(x)$ is a function whose components form the basis of Pol_m, that is $H(x) = (x^0, x^1, ..., x^m)$. If we want to find the expectation of one-dimensional OU process X, we can apply 2.5.1 to $H(X_0) = (X_0^0, X_0^1)$, $A = \operatorname{diag}(0, -\theta)$ and $\overrightarrow{p} = (0, 1)$,

$$\mathbb{E}_x [X_t] = (1, X_0)^\top e^{tA} (0, 1) = X_0 e^{-\theta t},$$

where we have used 2.5.2 to get $e^{tA} = \operatorname{diag}\left(e^{0t}, e^{-\theta t}\right)$.

When the dimension of the process is higher than one, the basis of Pol_m becomes somewhat more complicated and one often needs to use techniques from linear algebra of polynomials to enumerate the basis efficiently and to exploit sparsity properties of A. We refer to [5] for more information about possible methods.

When it comes to computing the matrix exponential itself, we notice that A is necessarily an upper triangular matrix when the dimension of the process X is one. This is easy to see as the extended generator satisfies $\mathcal{G}(\mathrm{Pol}_k) \subset \mathrm{Pol}_k$ for all $k \in \{0, ..., m\}$ (Theorem 2.5 (1) $\rightarrow$ (3)). The eigenvalues of an upper triangular matrix are its diagonal elements, which in our particular case are distinct and thus the matrix is diagonalizable. This allows us to use diagonalization of the matrix to get

$$A = UDU^{-1}$$

and particularly

$$A^k = \left(UDU^{-1}\right)^k = UD\left(U^{-1}U\right)D\left(U^{-1}U\right)D\cdots\left(U^{-1}U\right)DU^{-1} = UD^kU^{-1},$$

where D is a diagonal matrix composed of eigenvalues of A and U is a matrix consisting of the corresponding eigenvectors. Since e^A is defined as a power series, it can be written as follows:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = U \sum_{k=0}^{\infty} \frac{A^k}{k!} U^{-1}$$
$$= U e^D U^{-1}.$$

To calculate the matrix exponential of the diagonal matrix e^D, one simply takes the exponentials of the diagonal entries, i.e.

$$e^D = \begin{pmatrix} e^{d_1} & 0 & \cdots & 0 \\ 0 & e^{d_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & e^{d_{m+1}} \end{pmatrix}. \tag{2.5.2}$$

In more general cases, there exists efficient ways of calculating matrix exponentials, many can be found from [19].

2.6. Applications

In this section some examples of applications of polynomial processes in mathematical finance are given. In Chapter 3 a case study of modeling electricity markets is presented. There have been published some papers during the last years that study different applications of polynomial processes. Here are a few examples of different ideas presented in them:

EXAMPLE 2.21. (State price density) Filipović and Larsson [12] use a state price density approach with a polynomial diffusion X as a factor process. This approach allows for explicit expressions for any securities with cash-flow specified as a polynomial function of X. The idea goes as follows: There is a state price density given by a positive semimartingale ζ, defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$, that determines an arbitrage free market model. The price $F(t,T)$ of a time T cash flow C_T at time t is given by

$$F(t,T) = \frac{1}{\zeta_t} \mathbb{E}\left[\zeta_T C_T | \mathcal{F}_t\right]. \tag{2.6.1}$$

Choosing a polynomial diffusion X and positive polynomial p on the state space allows one to specify the state price density by $\zeta_t = e^{-\alpha t} p(X_t)$, where α is a real parameter chosen to control on the interest rates. Further, let the cash flow of a security be given by $C_T = q(X_T)$, where q is some polynomial. Since qp is a polynomial the expectation in (2.6.1) can be given explicitly in terms of matrix exponential by an application of moment formula, Theorem 2.19. This framework can be used for several different securities, for example interest rate models with $C_T = 1$, stochastic volatility model with C_T the spot variance and commodities market with C_T the spot price.

Example 2.22. (Life insurance liabilities) Biagini and Zhang 2016[4] introduce a polynomial diffusion model for pricing and hedging life insurance liabilities using a benchmark approach. Unlike the approach used in the example above, existence of state price density is not required here. Instead the existence of a benchmark portfolio is assumed and derivative pricing is done under the real world probability $\mathbb{P}$. The benchmark portfolio and financial instruments are modeled explicitly in terms of state variable which follows a polynomial diffusion. Pricing formula for products with polynomial payoffs is then be given explicitly. Furthermore the explicit results are used to approximate more general claims with polynomial functions. This enables the pricing of key products for life insurance companies, such as pure endowments, annuities and term insurances.

Example 2.23. (Variance reduction) Cuchiero et al. [9] present an efficient way of variance reduction in a Monte-Carlo simulation by using the explicit knowledge of the conditional expectation of a polynomial function $f \in \mathrm{Pol}_{\mathrm{m}}$ of an m-polynomial process X. The idea is to approximate a general function with a polynomial function and then use that as a control variate when estimating the price in Monte-Carlo simulation. Suppose we want to estimate $\mu = \mathbb{E}\left[g\left(X\right)\right]$ where X is an m-polynomial process and that we can find a polynomial approximation f of g such that $h\left(x\right) \approx f\left(x\right)$. As the value of $\mathbb{E}\left(f\left(X\right)\right)$ is explicitly known to us for $f \in \mathrm{Pol}_m$, we can estimate μ by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\left(g\left(X^i\right) - f\left(X^i\right)\right) + \mathbb{E}\left[f\left(X\right)\right].$$

This is an unbiased estimator and furthermore the variance of $\hat{\mu}$ is

$$\mathrm{Var}\left(\hat{\mu}\right) = \frac{1}{n}\mathrm{Var}\left(g\left(X\right) - f\left(X\right)\right),$$

which is smaller than the variance of $\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} g\left(X^i\right)$. Same method also applies to claims, where the current value is given as a conditional expectation, e.g. European claims and forwards.

CHAPTER 3

Application on Electricity Markets

In this chapter we use a polynomial framework for pricing and hedging long-term electricity forwards, based on the work of Kleisinger-Yu et al. [18]. We start with a brief introduction of electricity markets and their distinctive features, before moving on to defining our model framework. Our specific model is a two-factor model, which is a polynomial process, allowing us to use the moment formula (2.19) in the computation of forward prices. The moment formula is applied to set up a hedging strategy for a long-term electricity commitment. Due to the special features of the electricity markets, we propose a rolling hedge with risk-minimization based on Schweizer [23, 22]. In the end of this chapter, a simulation of forward prices and study of the effectiveness of the chosen hedging strategy.

3.1. Electricity Markets

Electricity markets were to a large extent deregulated in the beginning of the 1990's. Prior to this, electricity markets were strongly regulated by governments, and trading with electricity did not exist the way it does now. The deregulated electricity markets have several unique features compared to other "normal" commodities, which makes the theory behind derivative pricing and hedging interesting. Firstly, electricity has very limited storage possibilities. For some producers, mainly hydro plants, it is possible to store electricity by filling up the water reservoirs. However general actors in the electricity markets do not have the possibility to store electricity. If you buy electricity, you have to use it. Secondly, transferring electricity over long distances is inefficient. This leads to regional electricity markets. Prices between different markets can differ significantly. Thirdly, supply and demand are highly dependent on the physical phenomenons which are hard to predict, such as weather. Due to these features the wholesale market of electricity is based on matching offers from producers to bids from retailers. This is done in an auction between producers and retailers arranged by a transmission system operator. Based on the offers and bids for the hourly prices for the next day, the day ahead spot prices are decided. The complexity of the wholesale market, and the underlying physical phenomenons, leads to very high price volatility at times of peak demand and/or supply shortage. The importance of financial risk management is therefore a high priority for participants in electricity markets.

There are several types of financial instruments used for risk management (hedging). The two simplest and most common are fixed price forward contracts and contracts of difference. We use forwards for hedging in this paper. A forward is a contract between two parties to buy or sell an asset at a specific future time T at a price agreed in the contract. The price at time t of an electricity forward $f(t,T)$, with delivery at time $T \geq t$ is given by

$$f(t,T) := \mathbb{E}_{\mathbb{Q}}\left[S_T | \mathcal{F}_t\right], \tag{3.1.1}$$

where S_t is the spot price (of 1 MWh) at time t and $\mathbb{Q}$ is the risk-neutral probability measure. This follows from the well-known fact that the price of any derivative is the present expected value of its payoff. Payoff of a forward contract is given by

$$S_T - f(t,T)$$

and since the forward contract is entered at no cost, we have

$$\mathbb{E}_{\mathbb{Q}}\left[S_T - f(t,T) \mid \mathcal{F}_t\right] = 0,$$

which leads to (3.1.1) as $f(t,T)$ is $\mathcal{F}_t$-measurable.

Since electricity is not delivered instantaneously, but gradually over time, we usually consider forwards with delivery period $[T_1, T_2]$ which leads to a time t price

$$F(t, T_1, T_2) := \frac{1}{T_2 - T_1}\mathbb{E}_{\mathbb{Q}}\left[\int_{T_1}^{T_2} S_u du \,|\, \mathcal{F}_t\right]. \tag{3.1.2}$$

This is the expected average price during the delivery period with respect to the risk-neutral probability measure. The delivery period may range from a day up to a year, and contracts can be extended many years in to the future.

The spot price will be modeled with a multi-factor polynomial diffusion, which allows us to compute forward prices explicitly using matrix exponentials. Further a hedging strategy using a rolling hedge with variance minimization of the cost process is set up. A simulation of the forward prices is computed in order to study how the rolling hedge strategy works in reducing the risk involved in long-term forward commitments. Throughout this chapter we assume a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q}\right)$ where $\mathbb{Q}$ is the risk-neutral measure used for derivative pricing. Time t is measured in years. For simplicity we make a couple of assumptions. Firstly, we assume zero interest rate. Secondly, we assume that there are no market frictions, e.g. transaction costs, taxes and that forward contracts are infinitely divisible. Thirdly, there is no risk of default, meaning that both parties of a forward contract honor their commitments.

3.2. The Model

This section is dedicated to the underlying polynomial framework. We start by giving a general setup of a spot price model as a quadratic function of an underlying d-dimensional stochastic process X_t which will evolve according to a polynomial diffusion, Definition 2.15. We denote the spot price at time t with S_t defined as

$$S_t = p_S(X_t),$$

where $p_S(x) = c + x^\top Qx$ with $c \in \mathbb{R}_+$ and Q a positive semidefinite $d \times d$ matrix. Furthermore the underlying driving process X_t is a d-dimensional polynomial diffusion given by

$$dX_t = \kappa(\theta - X_t)\,dt + \sigma(X_t)\,dW_t,$$

where $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$, W is a d-dimensional Brownian motion under $\mathbb{Q}$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is continuous. We also assume that the components of the diffusion matrix $\alpha(x) := \sigma(x)\sigma(x)^\top$ are polynomials of degree two or less, which together with the linear drift guarantees that X_t is a polynomial diffusion, see Definition 2.15. This allows us to use the matrix exponential given by the (extended) generator of the process to easily and explicitly calculate conditional expectations of the process. Specifically we can use them to calculate forward prices, given by (3.1.2). The (extended) generator of X for $f \in \mathrm{Pol}_m$ is

$$\mathcal{G}f(x) = \sum_{i=1}^{d} (\kappa(\theta - x))_i\, D_i f(x) + \frac{1}{2}\sum_{i,j=1}^{d} \left(\sigma(x)\sigma(x)^\top\right)_{i,j} D_{ij} f(x). \tag{3.2.1}$$

This framework makes it possible to calculate forward values at any given time t explicitly with help of matrix exponentials, as the following proposition shows.

PROPOSITION 3.1. *The forward price with delivery period $[T_1, T_2)$ at time t ,$0 \le t \le T_1 < T_2$ is given by*

$$F(t, T_1, T_2, X_t) = \frac{1}{T_1 - T_2} H(X_t)\, e^{(T_1 - t)A} \int_0^{T_2 - T_1} e^{sA} ds\, \overrightarrow{p}_S,$$

where $\overrightarrow{p}_S$ is the coordinate representation of p_S and $H : \mathbb{R}^d \to \mathbb{R}^N$ is the basis vector of Pol_m.

PROOF. The result follows straight from Equation 3.1.2 together with Theorem 2.19. □

REMARK 3.2. The integral involving matrix exponential can be easily evaluated if the matrix A is diagonalizable. We have earlier shown that $e^A = Ue^AU^{-1}$, which leads to

$$\int_0^t e^{As} ds = U \int_0^t e^{As} ds U^{-1}.$$

We can integrate e^{As} component wise and get $\frac{1}{\lambda_i}(e^{\lambda_i t} - 1)$ for the i'th component. In the case of $\lambda_i = 0$, the i'th component is reduced to t.

We use a two-factor model, where the short-term dynamics of the spot price are driven by Y_t and the long-term dynamics by Z_t. These processes evolve according to SDEs

$$
\begin{aligned}
dZ_t &= -\kappa_Z Z_t dt + \sigma_Z dW_t^{(1)} \\
dY_t &= \kappa_Y \left(Z_t - Y_t\right) dt + \rho\sigma_Y dW_t^{(1)} + \sigma_Y \sqrt{1-\rho^2} dW_t^{(2)},
\end{aligned}
$$

where $Z_0, Y_0 \in \mathbb{R}$, $\kappa_Z, \kappa_Y, \sigma_Z, \sigma_Y \in \mathbb{R}_+$, $W_t = \left(W_t^{(1)}, W_t^{(2)}\right)$ is a standard two-dimensional Brownian motion and $\rho \in (-1, 1)$ is the correlation between the two Brownian motions. Furthermore, we assume $1 \geq \kappa_Z \geq \kappa_Y \geq 0$ to reflect the idea of long and short-term dynamics Z and Y respectively. The long-term dynamic Z is an Ornstein-Uhlenbeck process which is mean-reverting with rate κ_Z and with mean-reversion level zero. The short-term dynamics Y is also an OU process, but its mean-reversion is towards the long-term dynamics Z with rate κ_Y, due to the term $\kappa_Y \left(Z_t - Y_t\right) dt$. The dynamics Y_t and Z_t drive the spot price through the quadratic equation

$$S_t := c + \alpha Y_t^2 + \beta Z_t^2$$

for $c, \alpha, \beta \in \mathbb{R}_+$. It is worth noticing that the spot price is always strictly positive. Although this is a common restriction in modeling of commodities, it is not necessarily the case for electricity. It is not impossible for the electricity price to be negative. Negative prices are sometimes observed for example in Scandinavian countries, especially during mild, windy and rainy winters. The reason is two-folded. During a mild winter the consumption naturally is lower. At the same time the rainy and windy weather makes the production through wind mills and hydro plants high. Running down power plants is more costly than paying to get rid of the extra electricity, which leads to negative prices. The phenomenon of negative electricity prices is rare and happens during low consumption hours, for instance in the middle of the night. Therefore we choose to ignore this weakness in the model, as it does not have a major effect on our goal of using polynomial framework to calculate forward prices and hedging strategies.

In terms of the general set up, the model corresponds to

$$Q = \begin{pmatrix} \beta & 0 \\ 0 & \alpha \end{pmatrix}, \ \kappa = \begin{pmatrix} \kappa_Z & 0 \\ -\kappa_Y & \kappa_Y \end{pmatrix}, \ \theta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \sigma\left(x\right) = \sigma\left(z, y\right) = \begin{pmatrix} \sigma_Z & 0 \\ \rho\sigma_Y & \sigma_Y \sqrt{1-\rho^2} \end{pmatrix}.$$

The polynomial basis for this quadratic model is $H(x) = \{1, z, y, z^2, yz, y^2\}$, $x = (z, y)^\top$. Coordinate representation of $p_S(x)$ is then $\overrightarrow{p}_S = (c, 0, 0, \beta, 0, \alpha)^\top$ and thus

$$p_S\left(X_t\right) = H\left(X_t\right)^\top \overrightarrow{p}_S.$$

Further, the matrix representation of the extended generator $\mathcal{G}$ of X with respect to the basis given by $H(x)$ is needed. For the two-factor model the extended generator is

$$\mathcal{G}f(x) = \sum_{i=1}^{2}(-\kappa x)_i D_i f(x) + \sum_{i,j=1}^{2}\left(\sigma(x)\,\sigma(x)^\top\right)_{i,j} D_{ij} f(x),$$

with

$$(-\kappa x) = \begin{pmatrix} -\kappa_Z z \\ \kappa_Y z - \kappa_Y y \end{pmatrix} \text{ and } \sigma(x)\,\sigma(x)^\top = \begin{pmatrix} \sigma_Z^2 & \rho\sigma_Z\sigma_Y \\ \rho\sigma_Z\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

Applying the generator to the components of the basis function $H(x)$,we get the following equations:

$$\begin{aligned}
\mathcal{G}1 =& 0,\\
\mathcal{G}z =& -\kappa_Z z,\\
\mathcal{G}y =& \kappa_Y z - \kappa_Y y,\\
\mathcal{G}z^2 =& -2\kappa_Z z^2 + \sigma_Z^2,\\
\mathcal{G}yz =& \kappa_Y z^2 + (-\kappa_Z - \kappa_Y)zy + \rho\sigma_Z\sigma_Y,\\
\mathcal{G}y^2 =& 2\kappa_Y zy - 2\kappa_Y y^2 + \sigma_Y^2.
\end{aligned}$$

These equations correspond to the matrix representation

$$A = \begin{pmatrix}
0 & 0 & 0 & \sigma_Z^2 & \rho\sigma_Z\sigma_Y & \sigma_Y^2\\
0 & -\kappa_Z & \kappa_Y & 0 & 0 & 0\\
0 & 0 & -\kappa_Y & 0 & 0 & 0\\
0 & 0 & 0 & -2\kappa_Z & \kappa_Y & 0\\
0 & 0 & 0 & 0 & -\kappa_Z - \kappa_Y & 2\kappa_Y\\
0 & 0 & 0 & 0 & 0 & -2\kappa_Y
\end{pmatrix}.$$

With theses tools it is possible to explicitly compute forward prices within the model, but in order to use historical data to estimate the model parameters, we need to find the dynamics of the model with respect to the historical probability measure $\mathbb{P}$.

3.3. Market Price of Risk

So far we have worked with the risk neutral probability measure $\mathbb{Q}$, which is used for pricing the forwards. However, when historical data of forward markets is used to estimate parameters, the dynamics of the model with respect to the real world probability $\mathbb{P}$ is needed. The natural approach is to use Girsanov's theorem for the change of measure. The general idea is to define a market price of risk function $\lambda(x)$ and then show that the associated Radon-Nikodym density process (for $t \in [0, T]$)

$$\mathcal{E}^{\lambda}(X_t) = \exp\left(\int_0^t \lambda(X_s)^\top dW_s - \frac{1}{2}\int_0^t \lambda(X_s)^2 ds\right)$$

of the risk function is a martingale. It is then possible to get the $\mathbb{P}$-dynamics of X_t by using Girsanov's theorem.

The change of measure for the specific two-factor model defined in the previous section is shown below. The market price of risk function is defined as

$$\lambda(x) = (\gamma + \Lambda x)\sigma(x)^{-1},$$

with some $\gamma \in \mathbb{R}^2$ and Λ a symmetric 2×2 matrix. Next we find the parameters of λ for which $\mathcal{E}^{\lambda}(X_t)$ is a martingale. As it happens $\mathcal{E}^{\lambda}(X_t)$ is a martingale for all choices of parameters. The proof involves using an augmented state vector and a result from theory on exponentially affine martingales, see [18, Proposition 5.1]. Then by Girsanov's theorem

$$W_t^{\mathbb{P}} = W_t - \int_0^t \lambda(X_s)\, ds$$

is a Brownian motion with respect to $\mathbb{P}$. If we denote $\gamma = (\gamma_Z, \gamma_Y)^\top$ and $\Lambda = \operatorname{diag}(\lambda_Z, \lambda_Y)$, then the $\mathbb{P}$-dynamics of X_t are given by

$$dX_t = \left[\begin{pmatrix}\gamma_Z \\ \gamma_Y\end{pmatrix} - \begin{pmatrix}\kappa_Z - \lambda_Z & 0 \\ -\kappa_Y & \kappa_Y - \lambda_Y\end{pmatrix} X_t\right] dt + \begin{pmatrix}\sigma_Z & 0 \\ \rho\sigma_Y & \sigma_Y\sqrt{1-\rho^2}\end{pmatrix} dW^{\mathbb{P}},$$

leading to the following dynamics for Z and Y under $\mathbb{P}$

$$\begin{aligned} dZ_t &= (\gamma_Z - (\kappa_Z - \lambda_Z) Z_t)\, dt + \sigma_Z dW_t^{\mathbb{P},Z} \\ dY_t &= (\gamma_Y + \lambda_Y Y_t - \kappa_Y (Y_t - Z_t))\, dt + \sigma_Y \rho dW_t^{\mathbb{P},Z} + \sigma_Y\sqrt{1-\rho^2} dW_t^{\mathbb{P},Y}. \end{aligned}$$

The model under $\mathbb{P}$ is the same kind (OU) as the model under $\mathbb{Q}$, but with additional parameters defining the market price of risk. The interpretation of the risk parameters is that γ_Z, γ_Y contributes to the level of mean-reversion, leading to higher spot prices on average as time evolves (if positive). Lambda parameters on the other hand affect the mean-reversion rates of Z and Y respectively. Particularly λ_Z and λ_Y can be critical as they can change the sign of the mean-reversion rates of Z and Y, leading to non mean-reverting dynamics. This can be problematic when modeling electricity prices, as they generally are observed to be mean-reverting. One idea would be to restrict λ_Z and λ_Y when estimating the parameters, so that $\kappa_Z - \lambda_Z > 0$ and $\kappa_Y - \lambda_Y > 0$.

When considering a more general framework with d-dimensional underlying process, the problem of finding parameters that satisfy the martingale condition is more complicated.

3.4. Risk Minimizing Rolling Hedge

The basic definition of a hedge is an investment taken in order to reduce the risk of another investment. Hedging is widely used in many different sectors, from airlines hedging against an increase in oil price to international funds hedging against changes in foreign exchange rates. Hedging is generally done by using some kind of derivatives, most often options or forwards. In our case we use forwards. What we want to hedge is a long-term commitment to deliver electricity (say delivery time from year $\tilde{T}$ to year $\tilde{T}+1$, with $\tilde{T}$ large). The value of this claim at time t in our model framework is given by

$$\tilde{F}_t := F\left(t, \tilde{T}, \tilde{T}+1, X_t\right) = H\left(X_t\right)^{\top} e^{\left(\tilde{T}-t\right)A} \int_{\tilde{T}}^{\tilde{T}+1} e^{sA} ds \vec{p}_S$$

which follows from (3.1.2) and Proposition 3.1. The special characteristics of electricity markets make the hedging somewhat more complicated than in a normal commodity case. One cannot hedge the commitment with the basic buy and hold strategy, since the long-term contracts are not liquid and thus rarely available for trading. Another option would be a cash and carry strategy, where one buys the commodity and holds it until delivery, but the non-storability of electricity makes this strategy inconceivable. That is why we propose a rolling hedge, where we take a long position in the near-term contracts and then roll the hedge forward. The idea of a rolling hedge works as follows: First we enter the first-nearby one year forward, close it when it matures, and then enter into a new first-nearby one year forward. We continue this until the delivery period of the long-term forward starts. The idea is based on the paper by Neuberger [20]. To give the mathematical definition, we start with the price process P_t at time t defined as

$$P_t = \begin{pmatrix} P_t^1 \\ P_t^2 \\ \vdots \\ P_t^{N-1} \\ P_t^N \end{pmatrix} = \begin{pmatrix} F\left(t, T_1, T_2, X_t\right) \\ F\left(t, T_2, T_3, X_t\right) \\ \vdots \\ F\left(t, T_{N-1}, T_N, X_t\right) \\ F\left(t, T_N, T_{N+1}, X_t\right) \end{pmatrix},$$

with $N = \tilde{T}$ and $P_t^N = \tilde{F}_t$. There are N different tradable assets and a bank account to invest in. We write ξ_t^i for the amount of forward P_t^i owned at time t and η_t for the amount of cash in the bank. We have a hedging strategy consisting of $N+1$ assets, i.e $\varphi_t = \left(\eta_t, \xi_t\right)^{\top} = \left(\eta_t, \xi_t^1, ..., \xi_t^N\right)^{\top}$, with the usual and natural assumptions that ξ_t is predictable and η_t is adapted. The components of the price process P_t^k are martingales by definition (Equation 3.1.2). Due to the liquidity issues in the power forward markets, only the first-nearby forward contract is tradable. So we require

$$\xi_t^k = 0 \qquad \text{for all } t \notin [k-1, k),\ k = 1, ..., N, \tag{3.4.1}$$

which also means that a contract that has started to deliver is no longer tradable. The value of the hedging process at time $t \in [k-1,k)$ is given by the *value process* V_t:

$$V_t(\varphi) = \eta_t + \xi_t^\top P_t = \eta_t + \xi_t^k F(t,k,k+1,X_t),$$

and the cumulative gain of the hedge up to time $t \in [k-1,k)$ is

$$\begin{aligned} G_t(\varphi) &= \int_0^t \xi_s^\top dP_s = \sum_{i=1}^{k-1} \int_{i-1}^i \xi_s^i dP_s^i + \int_{k-1}^t \xi_s^k dP_s^k \\ &= \sum_{i=1}^{k-1} \int_{i-1}^i \xi_s^i dF(s,i,i+1,X_s) + \int_{k-1}^t \xi_s^k dF(s,k,k+1,X_s). \end{aligned} \tag{3.4.2}$$

Since ξ_t is predictable and P_s is a martingale the integrals above are well defined. Combining these two gives the cost process of the hedge

$$C_t(\varphi) := V_t(\varphi) - G_t(\varphi).$$

For a self-financing strategy the cost process is a constant and given by the initial value (capital) V_0. If the market is complete and arbitrage opportunities are not allowed, it is possible to find a self-financing strategy that fully replicates the claim, but this is not the case in our model.

What we are dealing with is an incomplete market, in the sense that we have two sources of randomness (two Brownian motions) and only one tradable risky asset available at each time. The spot price is not tradable itself and because of the restriction 3.4.1, there is only one forward available at each given time. In an incomplete market a claim cannot be fully replicated by a self-financing hedging strategy. We can either use a strategy that is self-financing, but does not fully replicate the claim or we can fully replicate the claim, but we need to make additional investments through out the hedge (not self-financing). In the first case we have some risk included since the claim is not fully replicated (residual risk) and in the latter case we need to use money to keep the hedge going, which is a source of risk. Either way we cannot fully remove the risk involved, instead we will minimize it by the means of *local risk minimization.*

Our approach is based on Martin Schweizer's work and we refer to [22, 23] for more details. The idea is that even though we cannot achieve a self-financing strategy (constant cost process C), we can have a *mean-self-financing* (C_t a martingale) strategy that minimizes the risk. First we define the *risk process* as

$$R_t(\varphi) := \mathbb{E}_{\mathbb{Q}}\left[(C_t(\varphi) - C_T(\varphi))^2 \,|\mathcal{F}_t\right],$$

for all strategies φ that fully replicate $\tilde{F}$ at maturity $\tilde{T}$, i.e.

$$V_{\tilde{T}}(\varphi) = \tilde{F}, \tag{3.4.3}$$

which in our case amounts to $\eta_{\tilde{T}} = 0$ and $\xi^N_{\tilde{T}} = 1$. It is natural to call a strategy $\check{\varphi}$ *risk-minimizing* (RM) if for each φ we have

$$R_t(\check{\varphi}) \le R_t(\varphi)$$

for all $t \le \tilde{T}$. Every RM strategy is mean-self-financing, see [22, Lemma 2.3]. It turns out that the martingale property of the price process guarantees that RM strategy, satisfying $V_{\tilde{T}}(\check{\varphi}) = \tilde{F}$, exists and is uniquely given by Galtchouk-Kunita-Watanabe (GKW) decomposition of $\tilde{F}$, for the proof see [22, Theorem 2.4]. The decomposition is given as

$$\tilde{F} = \mathbb{E}_{\mathbb{Q}}[\tilde{F}] + \int_0^{\tilde{T}} \tilde{\xi}_s^{\top} dP_s + \tilde{L}_{\tilde{T}}, \tag{3.4.4}$$

where $\tilde{L}$ is a $\mathbb{Q}$-martingale strongly orthogonal to P, i.e. $\left\langle P, \tilde{L} \right\rangle_t = 0$, and thus $\left[P, \tilde{L}\right]_t$ is a martingale, see Definition 1.45. This leads to the RM hedging strategy

$$\varphi^{rm} = \left(\eta_t^{rm}, \xi_t^{rm}\right)^{\top} = \left(V_t\left(\varphi^{rm}\right) - {\xi_t^{rm}}^{\top} P_t, \tilde{\xi}_t\right)^{\top},$$

with value process

$$V_t\left(\varphi_t^{rm}\right) = \mathbb{E}_{\mathbb{Q}}\left[\tilde{F}|\mathcal{F}_t\right] = \tilde{F}_t = \tilde{F}_0 + \int_0^t \tilde{\xi}_s^{\top} dP_s + \tilde{L}_t$$

and cost process $C_t = \tilde{F}_0 - \tilde{L}_t$. It is not hard to see that this strategy satisfies (3.4.3) and the risk process $R_t\left(\varphi^{rm}\right)$ is zero at $t = \tilde{T}$. Using the orthogonality of P and $\tilde{L}$ together with $dV_t\left(\varphi_t^{rm}\right) = d\tilde{F}_t = \tilde{\xi}_t dP_t + d\tilde{L}_t$ we get

$$d\left\langle \tilde{F}, P \right\rangle_t = \tilde{\xi}_t d\left\langle P, P\right\rangle_t + d\left\langle P, \tilde{L}\right\rangle_t \Leftrightarrow \check{\xi}_t = \frac{d\left\langle \tilde{F}, P\right\rangle_t}{d\left\langle P, P\right\rangle_t}. \tag{3.4.5}$$

By combining our rolling hedge strategy with the risk-minimizing hedge we arrive to the specific hedging strategy for our model. This is achieved by using (3.4.5) with (3.4.1) and (3.4.2) to get, for $t \in [k-1, k)$,

$$\begin{aligned} \tilde{\xi}_t^k &= \frac{d\left\langle \tilde{F}, P^k\right\rangle_t}{d\left\langle P^k, P^k\right\rangle_t} \\ &= \frac{\vec{w}_{01} e^{\left(\tilde{T}-t\right)A} \Sigma\left(X_t\right) e^{(k-t)A} \vec{w}_{01}}{\vec{w}_{01} e^{(k-t)A} \Sigma\left(X_t\right) e^{(k-t)A} \vec{w}_{01}}. \end{aligned} \tag{3.4.6}$$

The second equality follows from instantaneous covariation process and identity $\tilde{F} = P^{\tilde{T}}$, see Appendix A.3 for more details of the calculations and the covariation matrix $\Sigma(X_t)$.

The risk minimizing hedging strategy of the forwards is given by

$$\xi_t^{rm} = \begin{cases} \tilde{\xi}_t^k & \text{for } t \in [k-1,k), \\ 0 & \text{else.} \end{cases}$$

This leads to the time $t \in [k-1,k)$ cash amount for RM strategy

$$\eta_t^{rm} = V_t(\varphi^{rm}) - \xi_t^{rm^\top} P_t = \tilde{F}_t - \tilde{\xi}_t^k P_t^k$$

and cost process

$$C_t(\varphi^{rm}) = \tilde{F}_t - \int_0^t \xi^{rm} dP_s.$$

3.5. Simulation

In this section $M = 3000$ spot price curves are simulated using a simple Euler-Maruyama method. As we do not have data available for estimation of the model parameters, we base our model on parameters from [18], where they use a quadratic Kalman filter and estimate with the help of both Least-Squares and Maximum Likelihood methods. The spot prices are calculated under the real world probability measure $\mathbb{P}$. It is worth noticing that when estimating the parameters there are no restrictions on λ_Z and λ_Y, which makes it possible to end up with a model where the long-term dynamics Z is not mean-reverting under $\mathbb{P}$. This is actually the case with the estimated parameters, as $\kappa_Z < \lambda_Z$ (see Table 1). In the next section we study what happens when λ_Z is reduced so that Z reverts to the mean under $\mathbb{P}$ aswell.

To study the performance of the hedge, we compute forward surfaces with delivery periods $(T, T+1)$ for $T \in \{1, ..., 10\}$, for each spot price simulation. The calculation of the forward prices is done using Proposition 3.1. The forward surfaces are used to calculate the risk-minimizing hedging strategy ξ_t^{rm} for different hedging horizons, starting from $T = 2$ up to $T = 10$. We choose a monthly re-balancing, meaning that the hedge is adjusted at the turn of each month. The results with and without the hedge are then compared by looking at the percentage exposures. The unhedged exposure percentage is given by

$$\frac{F(T,T,T+1,X_T) - F(0,T,T+1,X_0)}{F(0,T,T+1,X_0)},$$

Table 1. Estimated parameters.

Parameter	Estimate
c	0.23961
α	10.2500
β	0.17681
κ_Z	0.01002
κ_Y	0.40021
σ_Z	0.40648
σ_Y	0.88913
ρ	0.11244
λ_Z	0.08999
λ_Y	0.11184
γ_Z	0.08679
γ_Y	0.12737
z_0	2.35805
y_0	2.00746

while for the hedged exposure percentage

$$\begin{aligned} & \frac{C_T\left(\varphi^{rm}\right) - F\left(0, T, T+1, X_0\right)}{F\left(0, T, T+1, X_0\right)} \\ & = \frac{F\left(T, T, T+1, X_T\right) - \int_0^T \xi_s^{rm\top} dP_s - F\left(0, T, T+1, X_0\right)}{F\left(0, T, T+1, X_0\right)}. \end{aligned} \tag{3.5.1}$$

The integral in the equation 3.5.1 can be calculated using equation (3.4.2), by the following sum of integrals

$$\int_0^T \xi_s^{rm} dP_s = \sum_{i=1}^{T} \int_{i-1}^{i} \tilde{\xi}_s^i dP_s^i. \tag{3.5.2}$$

Moreover, since we use monthly re-balancing, at time $t = 0$ we buy $\tilde{\xi}_0^1$ amount of the first nearby forward $P_0^1 = F\left(0, 1, 2, X_0\right)$, hold it until the end of the month and then readjust the holding to $\tilde{\xi}_{1/12}^1$ with price $P_{1/12}^1 = F\left(\frac{1}{12}, 1, 2, X_{1/12}\right)$. Thus the integral in (3.5.2) becomes

$$\int_0^T \xi_s^{rm} dP_s = \sum_{i=1}^{T} \sum_{j=0}^{11} \check{\xi}_{i-1+j/12}^i \left(P_{i-1+(j+1)/12}^i - P_{i-1+j/12}^i\right),$$

where ξ_s^i is given by 3.4.6.

We have used R-language for the simulations (see Appendix B). In order to make the simulations more efficient the R-code package Expm (Expm package) is used for the calculations of matrix exponentials. It is possible to calculate these exponentials explicitly using for example diagonalization, but this turned out to take more computation time than the estimation techniques used from the package Expm.

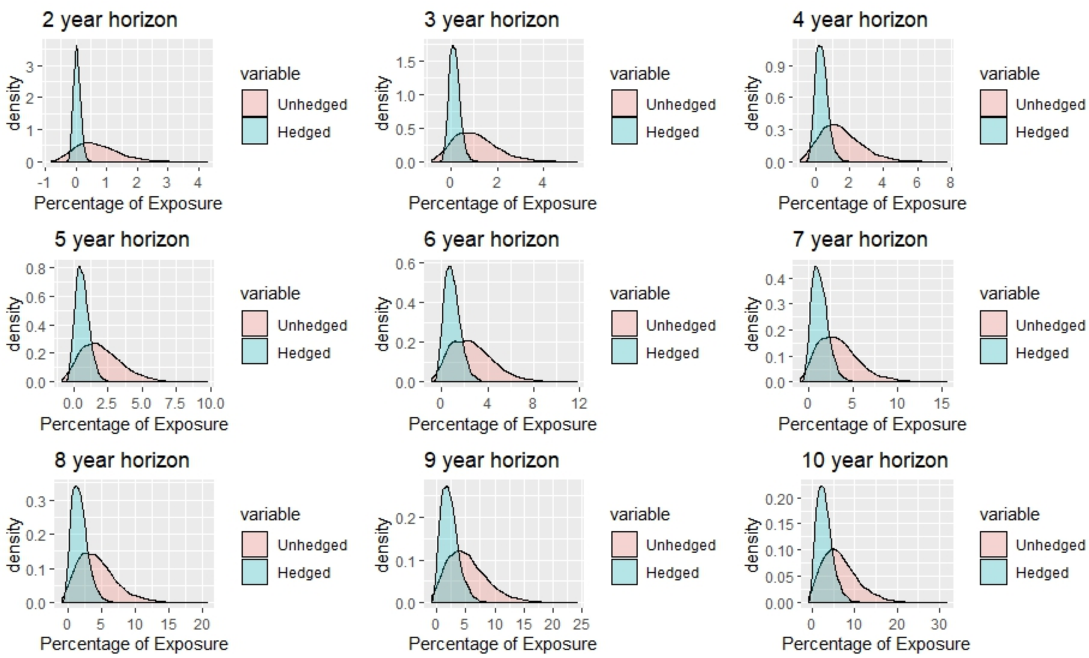

FIGURE 3.5.1. Density of exposures for different time horizons. Showing both hedged and unhedged exposures.

3.6. Stress Test

In this section we test the robustness of the hedging strategy described in Section 3.4 by changing some of the parameters to more extreme values and see whether this has an effect on the hedge. The purpose of this kind of stress testing is to see whether changes in the underlying economical factors that influence the parameters in the model, have an impact on the functionality of the hedging strategy. This is done by running new simulations with changed parameters $M = 1000$ times and then adjusting the hedging strategy for these new simulations. We study the new density plots to see how the hedge reacts to the changes. First we recall the dynamics of the model under $\mathbb{P}$, stated in the Section 3.3,

$$dZ_t = (\gamma_Z - (\kappa_Z - \lambda_Z) Z_t) dt + \sigma_Z dW_t^{\mathbb{P},Z}$$
$$dY_t = (\gamma_Y + \lambda_Y Y_t - \kappa_Y (Y_t - Z_t)) dt + \sigma_Y \rho dW_t^{\mathbb{P},Z} + \sigma_Y \sqrt{1-\rho^2} dW_t^{\mathbb{P},Y},$$

which drive the spot price through the quadratic function $p_S(X_t) = c + \alpha Y_t^2 + \beta Z_t^2$, $X = (Z, Y)$. We study three distinct scenarios with different shocks changing specific parameters of the model:

(1) A possible scenario would be a situation where the volatilities σ_Z and σ_Y of the dynamics are considerably higher. One could think of a situation where electricity markets are more unstable due to a shock, for example a political crisis. A simulation with twice as high σ_Z and σ_Y,

$$\sigma_Z^* = 2\sigma_Z$$
$$\sigma_Y^* = 2\sigma_Y,$$

is done to observe how the shock changes the results, see figure 3.7.3.

(2) Another kind of shock would be one that has an impact on the rate of mean-reversion parameters. Especially λ_Z (or eventually κ_Z) can have a big effect on the model. In the original estimated parameters $\kappa_Z - \lambda_Z < 0$, meaning that Z is not mean-reverting, when considering the real-world dynamics. This leads to a tendency for the spot price to grow quite aggressively, which makes the hedge more effective compared to the unhedged case. We set

$$\lambda_Z^* = 0.00899$$

to see how the model and hedge reacts when Z is mean-reverting, see figure 3.7.1

(3) In the third alternative scenario a shock disturbing the correlation ρ is tested to see whether it has an impact on the hedge. We look at two extreme cases

$$\rho^* = 0.9 \text{ and } \rho^* = -0.9$$

to how robust the hedge is when correlation much higher (lower), see figure 3.7.2.

3.7. Conclusion

Looking at the figure 3.5.1, one can see that our hedge significantly reduces the exposure percentages on average. It both decreases the mean of the exposures and it reduces the variation, especially for the shorter time horizons. This is natural as the correlation between the 1-year forward and the long-term commitment is greater in shorter time horizons. It is also worth noting that the hedge reduces the skewness of the exposures, which means that the probability of having a large exposure is reduced

Table 2. Mean of the hedged and unhedged exposures respectively for each time horizon.

Time Horizon	Mean Hedged	Mean unhedged
2 years	0.0322	0.6932
3 years	0.1648	1.0585
4 years	0.3636	1.5066
5 years	0.6188	2.0413
6 years	0.9303	2.6746
7 years	1.3013	3.3927
8 years	1.7328	4.2863
9 years	2.2469	5.3206
10 years	2.8494	6.5899

when hedging. We can conclude that our hedging strategy does significantly reduce the risk involved in the long-term commitment, but it does not completely remove it.

Looking at the density plots created by the different scenarios and comparing them to the original results, we see that in the case of increased volatilities (1), figure 3.7.3, the hedging strategy is working well. The case (2) where λ_Z was decreased does not substantially end the relation between the hedged and the unhedged. Although both the hedged and the unhedged exposures are significantly smaller due to the mean-reversion that draws the Z dynamics towards the long-term trend γ_Z. This effect can be seen from the figure 3.7.1. For the case of correlation (3), we see no big effect when ρ is close to one. When ρ is close to minus one, we see from figure 3.7.2 that the hedging strategy does not work as well, especially for longer time horizons.

The use of polynomial processes in modeling electricity prices allows for easily accessible and explicit formulas for pricing forwards. The same results are also used in creating a risk-minimizing strategy. One only needs to find the explicit form of the generator of the polynomial process, then calculating moments and conditional expectations is done using matrix exponentials. It would also be interesting to use jump processes such as polynomial jump diffusion to model electricity prices. The inclusion of jumps to the model could make them more realistic, since electricity prices tend to have spikes.

TABLE 3. Standard deviations of the hedged and unhedged exposures respectively for each time horizon.

Time Horizon	SD hedged	SD unhedged
2 years	0.1142	0.7110
3 years	0.2240	0.9356
4 years	0.3507	1.2102
5 years	0.5033	1.5271
6 years	0.6893	1.9101
7 years	0.9134	2.3509
8 years	1.1778	2.8799
9 years	1.4939	3.5091
10 years	1.8662	4.2583

TABLE 4. Skewness of the hedged and unhedged exposures respectively for each time horizon.

Time Horizon	Skewness hedged	Skewness unhedged
2 years	0.3653	0.7379
3 years	0.5164	0.7436
4 years	0.5825	0.7521
5 years	0.6530	0.7425
6 years	0.6839	0.8098
7 years	0.7543	0.8751
8 years	0.8209	0.8970
9 years	0.8741	0.9173
10 years	0.9123	0.9355

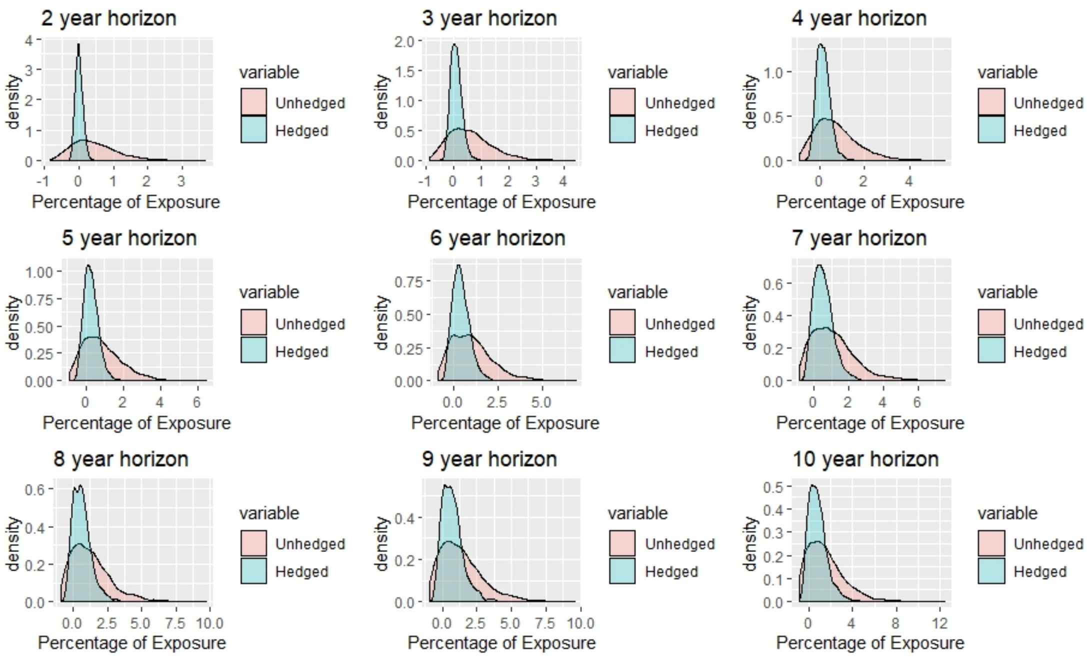

FIGURE 3.7.1. The scenario with smaller λ_Z.

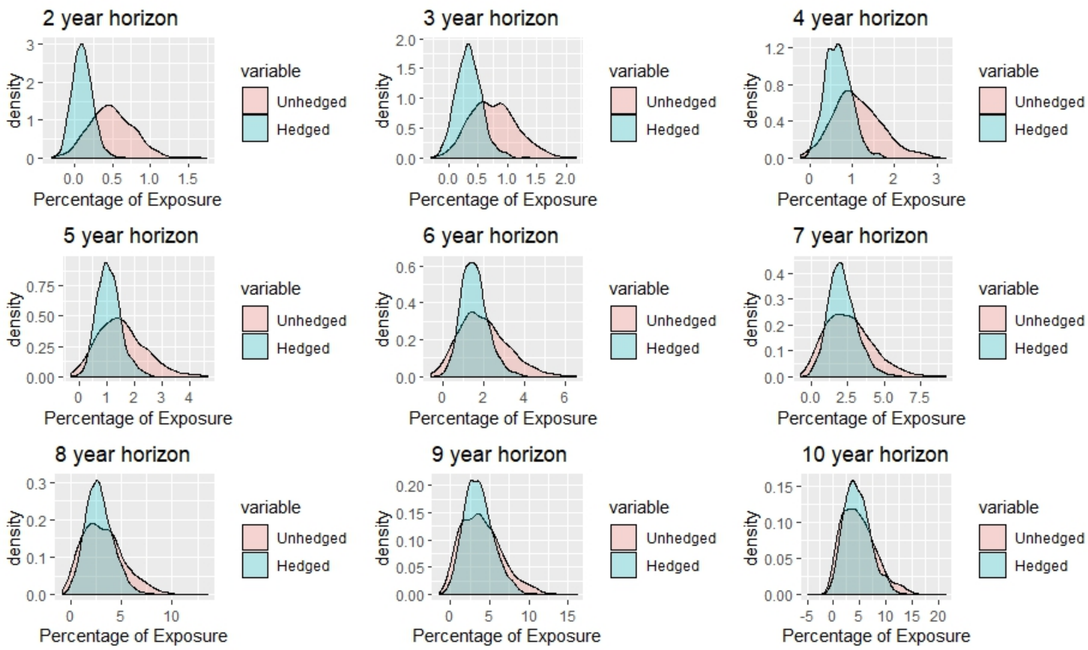

FIGURE 3.7.2. Densities related to the correlation scenario with $\rho^* = -0.9$.

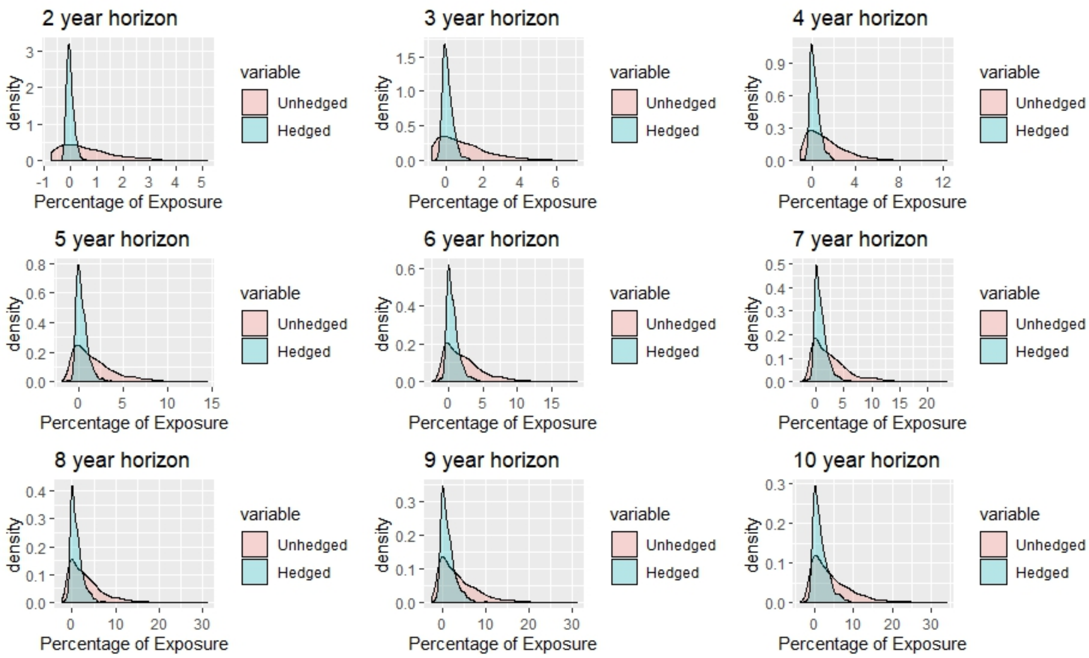

Figure 3.7.3. Density plots for increased volatility scenario.

APPENDIX A

Some Technical Results

A.1. Stochastic Analysis

Here are a few well known concepts and results used throughout the paper. They can be found in most books on stochastic analysis, see for example [10, 14, 2].

LEMMA A.1. *(Substitution property) Let $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable function. If X is $\mathcal{G}$-measurable, $h(X, Y) \in L^1$, and Y independent of $\mathcal{G}$ then*

$$\mathbb{E}\left[h(X, Y)\,|\mathcal{G}\right] = \mathbb{E}\left[h(x, Y)\right]|_{x=X} \,.$$

DEFINITION A.2. The variation of a function f is defined as the limit

$$V_f(t) = V_f([0,t]) = \sup \sum_{i=1}^{n} \left| f(t_i^n) - f\left(t_{i-1}^n\right)\right|,$$

where the supremum is taken over all partitions $0 = t_0^n < \cdots < t_n^n = t$. The sums in the equation above increase as new points are added to the partitions. This leads to

$$V_f(t) = \lim_{\pi_n \to 0} \sum_{i=1}^{n} \left| f(t_i^n) - f\left(t_{i-1}^n\right)\right|,$$

where the mesh $\pi_n = \max\{t_i - t_{i-1}\}$.

REMARK A.3. Clearly $V_f(t)$ is a non-decreasing function in t. A function f is said to be of finite variation if $V_f(t) < \infty$ for all t and of bounded variation if there exist a constant C such that $V_f(t) < \infty$ for all t.

THEOREM A.4. *(Doob-Meyer decomposition) Any local submartingale X with*

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X_s|\right] < \infty,$$

can be uniquely decomposed as

$$X_t = X_0 + M_t + A_t,$$

with some martingale M and an increasing predictable cádlág process A, $A_0 = 0$. The process A is called the compensator of X.

THEOREM A.5. *(Burkholder-Davis-Gundy inequality) For any $1 \leq p < \infty$ there exists positive constants c_p and C_p such that, for all local martingales M starting at 0, the following inequality holds.*

$$c_p \mathbb{E}\left[[M,M]_t^{p/2}\right] \leq \mathbb{E}\left[\sup_{s\leq t} |M_s|^p\right] \leq C_p \mathbb{E}\left[[M,M]^{2/p}\right].$$

For a continuous martingale the result holds also for $0 < p \leq 1$.

THEOREM A.6. *(Gronwall's lemma) Let $T, a, b \geq 0$ and $u(t) : [0,T] \to \mathbb{R}$ be a continuous function satisfying*

$$u(t) \leq a + b \int_0^t u(s)\, ds$$

for $t \in [0,T]$, then

$$u(t) \leq a e^{bt}.$$

This is a special case found in [10, A.5] and is enough for our purposes.

A.2. Multi-notation and Itô's Formula

We use the following standard multi-index notation in the paper. An n-dimensional multi-index is an n-tuple

$$\mathbf{k} := (k_1, k_2, \ldots, k_n),$$

with the sum of components given by

$$|\mathbf{k}| := k_1 + k_2 + \cdots k_n.$$

The power of an n-dimensional x is defined as

$$x^{\mathbf{k}} := x_1^{k_1} \cdot x_2^{k_2} \cdots x_n^{k_n},$$

while the binomial coefficient is defined as

$$\binom{\mathbf{k}}{\mathbf{j}} := \binom{k_1}{j_1} \cdot \binom{k_2}{j_2} \cdots \binom{k_n}{j_n}$$

and the multi-binomial theorem for $x, y \in \mathbb{R}^n$ states the following:

$$(x+y)^{\mathbf{k}} = \sum_{|\mathbf{j}|=0}^{|\mathbf{k}|} \binom{\mathbf{k}}{\mathbf{j}} x^{\mathbf{k}-\mathbf{j}} y^{\mathbf{j}} = \sum_{j_1=0}^{k_1} \binom{k_1}{j_1} x^{k_1 - j_1} y^{j_1} \cdots \binom{k_n}{j_n} x^{k_n - j_n} y^{j_n}.$$

The following calculations are used in the application of the Itô's formula in the proof of Proposition 2.10 :

SOLUTION A.7. Applying the Itô's formula for $f_{\mathbf{k}}(x) = x^{\mathbf{k}}$ and special semimartingale $X_t^i = x^i + M_t^{f_i} + B_t^i$, where B_t^i is the predictable variation part and $M_t^{f_i}$ the local martingale part, we get

$$
\begin{aligned}
f_{\mathbf{k}}\left(X_t\right) = f\left(x\right) &+ \int_0^t \sum_{i=i}^n D_i f_{\mathbf{k}}(X_{s-}) dX_s^i \\
&+ \frac{1}{2} \int_0^t \sum_{i,j=1}^n D_{ij} f_{\mathbf{k}}(X_{s-}) d\left\langle X^{i,cm}, X^{i,cm} \right\rangle_s \\
&+ \sum_{s \le t} \left(f_{\mathbf{k}}(X_s) - f_{\mathbf{k}}(X_{s-}) - \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) \Delta X_s \right).
\end{aligned}
$$

The first integral can be decomposed into two integrals with respect to $M_t^{f_i}$ and B_t^i respectively. For the second term we write $C_{ij} = \langle X^{i,cm}, X^{j,cm} \rangle$. For the last term we use the fact that $X_s = \Delta X_s + X_{s-}$ and an application of multi-binomial theorem on $(\Delta X + X_{s-})^{\mathbf{k}}$ to get

$$
\sum_{s \le t} \left(\sum_{|\mathbf{j}|=0}^{|\mathbf{k}|} \binom{\mathbf{k}}{\mathbf{j}} X_{s-}^{\mathbf{k}-\mathbf{j}} \Delta X_s^{\mathbf{j}} - X_{s-}^{\mathbf{k}} - \sum_{i=1}^n D_i f_{\mathbf{k}}\left(X_{s-}\right) \Delta X_s \right).
$$

The $|\mathbf{j}| = 0$ term of the sum is equal to X_{s-}^k so it cancels out and the $|\mathbf{j}| = 1$ term is equal to $\sum_{i=1}^n D_i f_k\left(X_{s-}\right) \Delta X_s$ and cancels out as well. We are then left with

$$
\sum_{s \le t} \left(\sum_{|\mathbf{j}|=2}^{k} \binom{\mathbf{k}}{\mathbf{j}} f_{\mathbf{k}-\mathbf{j}}\left(X_{s-}\right) \Delta X_s^{\mathbf{j}} \right),
$$

using the integral notation introduced in the Section 1.5 equation (1.5.3), $\sum_{s \le t} \Delta X_s = \int_0^t \int_{\mathbb{R}^n} \xi \mu\left(ds, d\xi\right)$, and we arrive at the claimed result

$$
\begin{aligned}
f_{\mathbf{k}}(X_t) = f_{\mathbf{k}}(x) &+ \int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) dM_s^{f_i} + \int_0^t \sum_{i=1}^n D_i f_{\mathbf{k}}(X_{s-}) dB_{s,i} \\
&+ \frac{1}{2} \int_0^t \sum_{i,j=1}^n D_{ij} f_{\mathbf{k}}(X_{s-}) dC_{s,ij} + \int_0^t \int_{\mathbb{R}^n} W(s,\xi) \mu^Y(ds, d\xi),
\end{aligned}
$$

with

$$
W(s,\xi) := \sum_{|\mathbf{j}|=2}^{k} \binom{\mathbf{k}}{\mathbf{j}} f_{\mathbf{k}-\mathbf{j}}\left(X_s\right) \xi^{\mathbf{j}}.
$$

A.3. Instantaneous Covariation Matrix and Hedging Strategy

The instantaneous quadratic covariation process between two forwards with different delivery periods is given by

$$\begin{aligned} d\left[P^k, P^j\right]_t &= d\left\langle P^k, P^j\right\rangle_t = d\left\langle H\left(X_t\right)^\top e^{(k-t)A} w_{01}, H\left(X_t\right)^\top e^{(j-t)A} w_{01}\right\rangle \\ &= w_{01}^\top e^{(j-t)A^\top} \Sigma\left(X_t\right) e^{(k-t)} w_{01}, \end{aligned}$$

where

$$w_{01} = \int_0^1 e^{uA} du \overrightarrow{p}_S$$

and $\Sigma\left(X_t\right) = \left\langle H\left(X\right), H\left(X\right)\right\rangle_t$ is the covariation matrix. Each component in the covariation matrix corresponds to the covariation of the corresponding elements of $H(X)$. The matrix $\Sigma\left(X_t\right)$ is stated below:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_Z^2 & \rho\sigma_Z\sigma_Y & 2\sigma_Z Z_t & \sigma_Z^2 Y_t + \rho\sigma_Y\sigma_Z Z_t & 2\rho\sigma_Y\sigma_Z Y_t \\ 0 & \rho\sigma_Z\sigma_Y & \sigma_Y^2 & 2\rho\sigma_Y\sigma_Z Z_t & \sigma_Y^2 Z_t + \rho\sigma_Y\sigma_Z Y_t & 2\sigma_Y^2 Y_t \\ 0 & 2\sigma_Z Z_t & 2\rho\sigma_Y\sigma_Z Z_t & 4\sigma_Z^2 Z_t^2 & 2\sigma_Z^2 Y_t Z_t + 2\rho\sigma_Y\sigma_Z Z_t^2 & 4\rho\sigma_Y\sigma_Z Y_t Z_t \\ 0 & \sigma_Z^2 Y_t + \rho\sigma_Y\sigma_Z Z_t & \sigma_Z^2 Z_t + \rho\sigma_Y\sigma_Z Y_t & 2\sigma_Z^2 Y_t Z_t + 2\rho\sigma_Y\sigma_Z Z_t^2 & \sigma_Z^2 Y_t^2 + \sigma_Y^2 Z_t^2 + 2\rho\sigma_Y\sigma_Z Y_t Z_t & 2\rho\sigma_Y\sigma_Z Y_t^2 + 2\sigma_Y^2 Y_t Z_t \\ 0 & 2\rho\sigma_Y\sigma_Z Y_t & 2\sigma_Y^2 Y_t & 4\rho\sigma_Y\sigma_Z Y_t Z_t & 2\rho\sigma_Y\sigma_Z Y_t^2 + 2\sigma_Y^2 Y_t Z_t & 4\sigma_Y Y_t^2 \end{pmatrix}.$$

The hedging strategy is obtained from the following calculations for any $t \in [k-1, k)$, $k \in \mathbb{N}$:

$$\begin{aligned} \left\langle P^k, \tilde{F}\right\rangle_t - \left\langle P^k, \tilde{F}\right\rangle_{k-1} &= \int_{k-1}^t d\left\langle P^k, \tilde{F}\right\rangle_u \\ &= \int_{k-1}^t d\left\langle P^k, \int_{k-1}^{\cdot} \tilde{\xi}_s^k dP_s^k\right\rangle_u + \int_{k-1}^t d\left\langle P^k, \tilde{L}\right\rangle_u \\ &= \left\langle P^k, \int_{k-1}^{\cdot} \tilde{\xi}_s^k dP_s^k\right\rangle_t - \left\langle P^k, \int_{k-1}^{\cdot} \tilde{\xi}_s^k dP_s^k\right\rangle_{k-1} \\ &= \int_{k-1}^t \tilde{\xi}_s^k d\left\langle P^k, P^k\right\rangle_s, \end{aligned}$$

where we have used that $\left\langle P^k, \tilde{L}\right\rangle_t = 0$ as P^k and $\tilde{L}$ are orthogonal. Rearranging the terms and applying $\left\langle P^k, \tilde{F}\right\rangle_{k-1} = 0$ since $\tilde{F}_{k-1}$ is constant and known at $t \geq k-1$, we arrive at

$$\tilde{\xi}_t^k = \frac{d\left\langle \tilde{F}, P^k \right\rangle_t}{d\left\langle P^k, P^k \right\rangle_t},$$

as wanted.

APPENDIX B

R-code

```
library(expm)
library(e1071)
library(xtable)
library(ggplot2)
library(reshape2)
library(gridExtra)


#parameters
c <- 0.2396
alpha <- 10.25
beta <- 0.176
kZ <- 0.01
kY <- 0.4002
sigmaZ <- 0.406
sigmaY <- 0.889
rho <- 0.112
lambdaZ <- 0.0899
lambdaY <- 0.112
gammaZ <- 0.0868
gammaY <- 0.1274
z0 <- 2.358
y0 <- 2.0076

#Matrix representation of the generator

Mdata=c(0, 0, 0, sigmaZ^2,rho*sigmaY*sigmaZ,sigmaY^2,
        0,-kZ,kY,0,0,0,0,0,-kY,0,0,0,0,0,0,-2*kZ,
        kY,0,0,0,0,0,-kZ-kY,2*kY,0,0,0,0,0,-2*kY)
A=matrix(ncol = 6,nrow = 6, data = Mdata, byrow = TRUE)
```

```
#the quadratic equation vector
pS <- c(c,0,0,beta,0,alpha)

#Calculation by diagonalization.
d1 <- eigen(A)$values
p1 <- eigen(A)$vectors
eGd1 <- p1%*%diag(exp(d1))%*%zapsmall(solve(p1))
w01 <- p1%*%diag(c(exp(d1[-6])/d1[-6]-1/d1[-6],1))%*%(solve(p1))%*%pS

#dataframes of OU process and forward calculations
twodOU=function(N,Ty){
  dt <- 1/N
  ts <- seq(0,Ty,dt)
  tl <- length(ts)
  Z <- c(z0,2:tl)
  Y <- c(y0,2:tl)
  epsZ <- rnorm(tl,0,1)
  epsY <- rnorm(tl,0,1)
  for (i in 2:tl) {
    Z[i] <- Z[i-1]+gammaZ*dt-(kZ-lambdaZ)*dt*Z[i-1]+
      sigmaZ*sqrt(dt)*epsZ[i]
    Y[i] <- Y[i-1]+gammaY*dt+kY*dt*Z[i-1]-
      (kY-lambdaY)*dt*Y[i-1]+sigmaY*sqrt(dt)*
      (rho*epsZ[i]+sqrt(1-rho^2)*epsY[i])
  }
  return(as.data.frame(cbind(ts,Z,Y)))
}

#Using Expm package for matrix exponentials,
#as more effective than our explicit calculations

ForwardPrice=function(Tm,Z,Y,dt){
  HX0 <- t(c(1,Z[1],Y[1],Z[1]^2,Z[1]*Y[1],Y[1]^2))
  F0 <- HX0%*%expm(Tm*A)%*%w01
  Ft <- c(F0,2:length(Y))
  for (i in 2:length(Ft)) {
    HXi <- t(c(1,Z[i],Y[i],Z[i]^2,Z[i]*Y[i],Y[i]^2))
    Ft[i] <- HXi%*%expm((Tm-((i*dt)%%1))*A)%*%w01
  }
  return(Ft)
```

```
}


forwardcurves=function(l=10,data1){
  Z <- data1$Z
  Y <- data1$Y
  ts <- data1$ts
  fc=data.frame(ts)
  for (i in 1:l) {
    fc[[i+1]] <- ForwardPrice(i,Z,Y,1/N)

  }
  fc$Z <- Z
  fc$Y <- Y
  return(fc)
}

#instantaneous covariation, used in hedging strategy
CovS=function(Yt,Zt){
  SX1 <- c(0,0,0,0,0,0)
  SX2 <- c(0, sigmaZ^2, rho*sigmaY*sigmaZ, 2*sigmaZ^2*Zt,
           sigmaZ^2*Yt+rho*sigmaY*sigmaZ*Zt, 2*rho*sigmaY*sigmaZ*Yt)
  SX3 <- c(0, rho*sigmaY*sigmaZ, sigmaY^2,
           2*rho*sigmaY*sigmaZ*Zt, sigmaY^2*Zt+rho*sigmaY*sigmaZ*Yt,
           2*sigmaY^2*Yt)
  SX4 <- c(0,2*sigmaZ^2*Zt, 2*rho*sigmaY*sigmaZ*Zt,
           4*sigmaZ^2*Zt^2, 2*sigmaZ^2*Yt*Zt+2*rho*sigmaY*sigmaZ*Zt^2,
           4*rho*sigmaY*sigmaZ*Yt*Zt)
  SX5 <- c(0, sigmaZ^2*Yt+rho*sigmaY*sigmaZ*Zt, sigmaY^2*Zt+
             rho*sigmaY*sigmaZ*Yt, 2*sigmaZ^2*Yt*Zt+2*rho*sigmaY*
             sigmaZ*Zt^2, sigmaZ^2*Yt^2+sigmaY^2*Zt^2+2*rho*sigmaY*sigmaZ*Yt*Zt,
           2*rho*sigmaY*sigmaZ*Yt^2+2*sigmaY^2*Yt*Zt)
  SX6 <- c(0,2*rho*sigmaY*sigmaZ*Yt, 2*sigmaY^2*Yt,
           4*rho*sigmaY*sigmaZ*Yt*Zt, 2*rho*sigmaY*sigmaZ*Yt^2+
             2*sigmaY^2*Zt*Yt, 4*sigmaY^2*Yt^2)
  SigmaS=rbind(SX1,SX2,SX3,SX4,SX5,SX6)
  return(SigmaS)
}

#hedging strategy
```

```
xik=function(t,kt,Zt,Yt,Tt){
  xi <- (t(w01)%*%t(expm((Tt-t)*A))%*%CovS(Zt, Yt)%*%expm((kt-t)*A)%*%w01)/
    (t(w01)%*%t(expm((kt-t)*A))%*%CovS(Zt, Yt)%*%expm((kt-t)*A)%*%w01)
  return(xi)
}

#Total cost of the rolling hedge.

hedge=function(k, Ft, ty, N=N){
  hedge <- 0
  dt <- 1/N
  Zti <- Ft[[k]]$Z
  Yti <- Ft[[k]]$Y
  dk=seq(0,length(Zti)-N/12,N/12)
  for (j in 1:ty) {
    for (i in dk[(N/10*j-11):(N/10*j)]) {
      hedge <- hedge + xik(i*dt, kt=j, Zt=Zti[i+1],
                    Yt=Yti[i+1], Tt=ty)*(Ft[[k]]$V2[i+11]-Ft[[k]]$V2[i+1])

    }
  }
  return(hedge)
}

#Making 3000 simulations of forwardcurves
M <- 3000
N <- 120
ty <- 10

set.seed(194)
sims <- replicate(M,forwardcurves(ty, twodOU(N, ty)), simplify = FALSE)


#plotting an example of rolling forwards and spot price
data2 <- within(sims[[1]], rm(Z,Y))
data2 <- setNames(data2, c("time", "1 year", "2 year",
                           "3 year", "4 year", "5 year", "6 year",
                           "7 year", "8 year", "9 year", "10 year"))

data2long <- melt(data2, id="time")
```

```
ggplot(data=data2long, aes(x=time, y=value, colour = variable)) + geom_line() +
  labs(y="Price", colour="Forward", title = "Rolling forwards")

S1 <- c + beta*sims[[1]]$Z^2 + alpha*sims[[1]]$Y^2

qplot(sims[[1]]$ts, S1, main = "Spot price",
      xlab = "Time", ylab = "Price", geom = "line")

#Unhedged exposures

e <- as.data.frame(matrix(nrow = M, ncol = 9))
e <- setNames(e, c("2 year", "3 year", "4 year", "5 year", "6 year",
                "7 year", "8 year", "9 year", "10 year"))

etemp <- 1:M
for (j in 1:9) {
  for (m in 1:M) {
    etemp[m] <- (sims[[m]]$V2[(j+1)*120+1]-
                   sims[[m]][[j+2]][1])/sims[[m]][[j+2]][1]

  }
  e[[j]] <- etemp

}

#Hedged exposures

eh <- as.data.frame(matrix(nrow = M, ncol = 9))
eh <- setNames(eh, c("2  year", "3 year", "4 year", "5 year",
                     "6 year", "7 year", "8 year", "9 year", "10 year"))
ehtemp=1:M
for (j in 1:9) {
  for (m in 1:M) {
    ehtemp[m] <- (sims[[m]]$V2[(1+j)*120+1]-sims[[m]][[j+2]][1]-
                    hedge(k=m, Ft=sims, ty=j+1, N=120))/sims[[m]][[j+2]][1]
  }
  eh[[j]] <- ehtemp

}
```

```
#Exposure density plots
p <- list()
for (i in 1:9) {
  te <- data.frame(Unhedged=e[i], Hedged=eh[i])
  colnames(te) <- c("Unhedged", "Hedged")
  data1 <- reshape2::melt(te, measure.vars = c("Unhedged", "Hedged"))
  p[[i]] <- ggplot(data1, aes(x=value, fill = variable)) +
    geom_density(alpha = 0.25) +
    labs(x="Percentage of Exposure", title = paste(i+1, "year horizon"))
}

do.call(grid.arrange, p)
```

Bibliography

[1] Yacine Ait-Sahalia and Jean Jacod. "High-Frequency Financial Econometrics". In: *High-Frequency Financial Econometrics* (Jan. 2014). DOI: 10.1515/9781400850327.

[2] Richard F. Bass. *Stochastic Processes*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511997044.

[3] F.E. Benth, J.S. Benth, and S. Koekebakker. *Stochastic Modelling of Electricity and Related Markets*. Advanced series on statistical science & applied probability. World Scientific, 2008. ISBN: 9789812812308. URL: https://books.google.no/books?id=MHNpDQAAQBAJ.

[4] Francesca Biagini and Yinglin Zhang. "Polynomial diffusion models for life insurance liabilities". eng. In: *Insurance Mathematics and Economics* 71.C (2016), pp. 114–129. ISSN: 0167-6687.

[5] Carl de Boor and Amos Ron. "On multivariate polynomial interpolation". In: *Constructive Approximation* 6.3 (Sept. 1990), pp. 287–302. ISSN: 1432-0940. DOI: 10.1007/BF01890412. URL: https://doi.org/10.1007/BF01890412.

[6] E. Cinlar et al. "Semimartingales and Markov Processes". In: *Probability Theory and Related Fields* 54 (Jan. 1980), pp. 161–219. DOI: 10.1007/BF00531446.

[7] Samuel N. Cohen and Robert J. Elliott. *Stochastic calculus and applications*. Second. Probability and its Applications. Springer, Cham, 2015, pp. xxiii+666. DOI: 10.1007/978-1-4939-2867-5. URL: https://doi.org/10.1007/978-1-4939-2867-5.

[8] Christa Cuchiero. *Affine and polynomial processes*. eng. 2011.

[9] Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. "Polynomial processes and their applications to mathematical finance". In: *Finance Stoch.* 16.4 (2012), pp. 711–740. ISSN: 0949-2984. DOI: 10.1007/s00780-012-0188-x. URL: https://doi.org/10.1007/s00780-012-0188-x.

[10] E Eberlein and J Kallsen. *Mathematical Finance*. Springer Finance. Springer International Publishing, 2019. ISBN: 978-3-030-26106-1.

[11] Klaus-Jochen Engel and Rainer Nagel. "One-Parameter Semigroups for Linear Evolution Equations". In: *Semigroup Forum* 63 (June 2001), pp. 278–280. DOI: 10.1007/s002330010042.

[12] Damir Filipovic and Martin Larsson. "Polynomial diffusions and applications in finance". eng. In: *Finance and Stochastics* 20.4 (2016), pp. 931–972. ISSN: 0949-2984.

[13] Damir Filipovic and Martin Larsson. *Polynomial Jump-Diffusion Models*. 2017. arXiv: 1711.08043 [q-fin.MF].

[14] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*. eng. 2nd. Vol. 288. Die Grundlehren der mathematischen Wissenschaften. Berlin: Springer-Verlag, 2002. ISBN: 3540439323.

[15] Jean Jacod et al. "The approximate Euler method for Levy driven stochastic differential equations". eng. In: *Annales de l'Institut Henri Poincare / Probabilites et statistiques* 41.3 (2005), pp. 523–558. ISSN: 0246-0203.

[16] Jan Kallsen and Johannes Muhle-Karbe. "Exponentially affine martingales, affine measure changes and exponential moments of affine processes". In: *Stochastic Processes and their Applications* 120.2 (2010), pp. 163–181. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2009.10.012. URL: http://www.sciencedirect.com/science/article/pii/S0304414909002014.

[17] Fima C Klebaner. *Introduction to Stochastic Calculus with Applications*. 3rd. Imperial College Press, 2012. DOI: 10.1142/p821. eprint: https://www.worldscientific.com/doi/pdf/10.1142/p821. URL: https://www.worldscientific.com/doi/abs/10.1142/p821.

[18] Xi Kleisinger-Yu et al. "A multi-factor polynomial framework for long-term electricity forwards with delivery period". In: 2019.

[19] Cleve. Moler and Charles. Van Loan. "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later". In: *SIAM Review* 45.1 (2003), pp. 3–49. DOI: 10.1137/S00361445024180. URL: https://doi.org/10.1137/S00361445024180.

[20] Anthony Neuberger. "Hedging Long-Term Exposures with Multiple Short-Term Futures Contracts". In: *The Review of Financial Studies* 12.3 (1999), pp. 429–459. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/2646067.

[21] Bernt Oksendal. *Stochastic differential equations : an introduction with applications*. eng. 6th ed. Universitext. Berlin: Springer, 2007. ISBN: 9783540047582.

[22] Martin Schweizer. *A guided tour through quadratic hedging approaches*. SFB 373 Discussion Papers 1999,96. Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 1999. URL: https://ideas.repec.org/p/zbw/sfb373/199996.html.

[23] Martin Schweizer. "Option hedging for semimartingales". eng. In: *Stochastic Processes and their Applications* 37.2 (1991), pp. 339–363. ISSN: 0304-4149.

[24] Eugene Wong. "The construction of a class of stationary Markoff processes". eng. In: *Stochastic Processes in Mathematical Physics and Engineering* (1964), pp. 264–276. URL: http://www.dtic.mil/docs/citations/AD0633486.

[25] Hao Zhou. "Ito Conditional Moment Generator and the Estimation of Short-Rate Processes". In: *Journal of Financial Econometrics* 1.2 (June 2003), pp. 250–271. ISSN: 1479-8409. DOI: 10.1093/jjfinec/nbg009. eprint: https://academic.oup.com/jfec/article-pdf/1/2/250/2444237/nbg009.pdf. URL: https://doi.org/10.1093/jjfinec/nbg009.