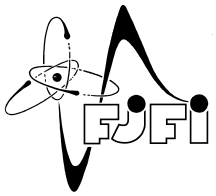
Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering



Methods of Superstatistics and Subordination in Theory of Option Pricing with Stochastic Volatility

Master's Thesis

Author: **Bc. Filip Garaj**

Supervisor: **Ing. MSc. Petr Jizba, Ph.D.**

Academic year: 2018/2019

- Zadání práce -

Acknowledgement:
I would like to thank my supervisor Ing. MSc. Petr Jizba, Ph.D. for suggesting the topic of this thesis and for his valuable remarks.
This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS19/183/OHK4/3T/14.
My thanks belongs to Thomson Reuters, who gave me consent to display figures from the Eikon application for this academical text and Ernst & Young for letting me use their license.
I would also like to express my deep gratitude to my life partner Michaela, who mentally supported me throughout working a full-time job while writing this thesis.

Prohlášení:
Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v přiloženém seznamu.

Nemám závažný důvod proti použití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 6. května, 2019 Bc. Filip Garaj

Název práce:

Superstatistické a subordinační metody v teorii opčních trhů se stochastickou volatilitou

Autor: Bc. Filip Garaj

Obor: Matematická fyzika

Druh práce: Diplomová práce

Vedoucí práce: Ing. MSc. Petr Jizba, Ph.D., FJFI ČVUT, katedra fyziky

Abstrakt: V této práci nejprve vybudujeme teoretický základ nutný ke studiu problémů v kvantitativních financích a následně představíme některé pokročilé koncepty z tohoto odvětví. Ačkoliv naším cílem je konstruovat sofistikované modely, je vhodné začít od těch známějších, méně složitých modelů. Tento postup nám umožní lépe zdůraznit rozdíly mezi modely a také je dostatečně zasadit do kontextu. Významný prostor je věnován stochastickému integrálu a stochastickým diferenciálním rovnicím (SDR). V textu také představíme základy teorie opčních trhů a klasifikujeme přístupy k modelování volatility. Protože si klademe za cíl implementovat subordinační metody, za účelem stochastického modelování volatility, představíme také Lévyho procesy a jejich vlastnosti. V této práci představíme známé modely Blacka-Scholese-Mertona (BSM) a také Hestonův model. V poslední kapitole subordinaci použijeme, abychom zavedli čistě skokové modely - variance gamma model a jeho rozšíření s náhodným příchodem skoků. Ukážeme také, že Hestonův model je vlastně BSM model subordinovaný náhodnému času. V průběhu poukážeme na spojitosti mezi subordinací, stochastickou volatilitou, SDR přístupem a procedurou marginalizace známou například ze superstatistiky.

Klíčová slova: Hestonův model, Lévyho procesy, stochastická volatilita, stochastická změna času, subordinace, superstatisticka, variance gamma proces

Title:

Methods of Superstatistics and Subordination in Theory of Option Pricing with Stochastic Volatility

Author: Bc. Filip Garaj

Abstract: This thesis first builds the theoretical apparatus that is necessary to study problems in mathematical/quantitative finance and then continues to introduce some advanced concepts within this field. Although our ultimate goal is to construct highly intricate models it is convenient to introduce the less sophisticated and commonly known models first. Advancing in this way allows us to better highlight the contrasts between these models as well as to contextualize them sufficiently. Significant space is devoted to stochastic integral and stochastic differential equations (SDEs). We proceed to introduce option theory and classification of approaches to model volatility. In order to implement subordination methods to model volatility stochastically, we study Lévy processes and their properties. Throughout the text, we probe the well known models such as Black-Scholes-Merton (BSM) and the model due to Heston. In the final chapter, we use subordination to characterize pure jump models - the variance gamma model and its extension with stochastic arrival of jumps. We also show that Heston model is, in a sense, the BSM model subordinated to random clock. In the process we point out the connections between subordination, stochastic volatility, SDE approach and procedure of marginalization encountered for example in superstatistics.

Key words: change of time, Heston model, Lévy processes, stochastic volatility, subordination, superstatistics, variance gamma proces

Contents

Introduction

This is a text about stochastic analysis and about mathematical finance. Both of these subjects were deeply intertwined from the very beginnings. Stochastic analysis is a study of random processes, which are mathematical objects used to model random behavior of certain systems. A random (stochastic) process encompasses all the possible and impossible (from the perspective of probability) futures of the particular system in question. One realization of such process is then a trajectory representing a course of events for the particular happenstance. These concepts are (for example) utilized to model reaction of chemical compounds in a solution or to model moves of a stock price on an exchange. In what is to follow, we shall be focusing on applications related to the latter.

An eminent topic in stochastic analysis are the stochastic differential equations (SDEs). Just as ordinary differential equations are used in physics to describe a law or a model for an evolution of a physical system, stochastic differential equations are used to model the random behavior of a system. In the case of an ordinary differential equation, if a solution exists, it is a deterministic function, so that given an initial condition, according to this solution, the system will behave always the same. On the other hand, a solution (if it exists) of a stochastic differential equation

$$dX_t = \alpha(t, X_t)\, dW_t$$

is a stochastic process $\{X_t\}$ in the sense described above[1]; here $\alpha = \alpha(\cdot, \cdot)$ is an a priori known coefficient function. Given an initial condition, the stochastic process will evolve differently every time. This might be viewed as quite a hindrance in employing such objects for the purposes of modeling. Au contraire, for systems, that we perceive as inherently random, this is actually a desirable property. Furthermore, the SDEs and their solutions (the stochastic processes) are not intended to predict the future, but rather to pronounce quantitative and qualitative statements that answer questions such as "What is the behavior of the system on average?" or "How differently does the system behave for each particular realization?".

In the above equation there is the term dW_t that we have not introduced yet. Here $\{W_t\}$ is a stochastic process as well, and it is the "driving noise" for the SDE, i.e. it is the source of randomness in this equation. This process is usually taken to be the so-called Brownian motion, the rigorously defined mathematical concept, that is behind the physical phenomenon of the same name[2]. In what follows, we devote some significant space to the theory of Brownian motion as well as to what actually the "differential" dW_t represents.

[1] In fact, what we have introduced here, is a strong solution to an SDE, which implies there is also a concept of a weak solution. For details, see the corresponding chapter.

[2] For historical remarks, see the section on Brownian motion.

First to implement some of the concepts described above was L. Bachelier [1], however the field of mathematical finance took notice of his approach and appropriated it much later, in the second half of the 20th century. The stochastic processes, defined as solutions of SDEs are usually employed as models for the behavior of some underlying asset, which then serve to value financial derivatives written on these underlyings[3]. This method was pioneered by F. Black, M. Scholes [2] and R. Merton [3] in the famous Black-Scholes-Merton (BSM) model. In the chapter on the topic, we present the two original derivations of the related formulae. Although the BSM model is inherently flawed, for it incorporates many simplifying assumptions which are inconsistent with the reality, it is important to understand it for two main reasons (taking into account purposes of this text). Firstly, the model serves as a language in which facts about the markets may be expressed - in particular, many quantities are quoted in the BSM parameters. Secondly, it is necessary to be familiar with the model one is aiming to generalize and improve upon.

One of the assumptions made by Black, Scholes and Merton, is that the volatility (i.e. the standard deviation of logarithmic returns) of a given (underlying) asset is a constant parameter in the model. This is clearly in disagreement with empirical evidence[4]. For this reason almost all the generalizations of the model strive to abandon that assumption and instead regard the volatility at least as a deterministic function (the local volatility model) or even more generally as a stochastic process. The often used and very well-known model in this direction is the Heston model [4] on which we derive some of the results in the chapter bearing the same name.

All the models we have mentioned so far have one thing in common - the continuity of the trajectories generated by them. Although (with the exception of BSM) these models are used in practice to price derivative contracts, sometimes with large exposures, they do not incorporate price jumps in their processes for underlying. In the last two chapters of this text we present motivation for models with discontinuities and develop some theory for them. One approach that we shall not pursue here is to develop even more general framework for stochastic differential equations, where the Brownian motion is substituted by processes with more sophistication as the source of randomness. The path that we shall follow is utilizing the technique of change of time and subordination.

A stochastic process $\{X_t\}$ depends on the parameter t, usually called time, to which it assigns realizations of random variables (for a particular event). The idea behind the change of time method is that the parameter t is itself a random variable so that we have the structure of the original process but with time evolving non-deterministically. This is often interpreted as t being the ordinary clock time, $\{\mathcal{T}(t)\}$ being the random change of time process and its realization $\mathcal{T}(t) = \theta$ being the perceived business (or trading) time. In this interpretation, it is still the well-behaved market data arriving $\hat{X} = \{\hat{X}_\theta\}$, but in the randomly (at times erratically) changing time θ, so that the signal on the output is actually the more chaotic one, $\{\hat{X}_{\mathcal{T}(t)}\}$, that one receives in the real markets. This kind of interpretation adds to the tractability of the resulting model. We shall utilize the change of time procedure with $\{\mathcal{T}(t)\}$ being a Lévy process, which allows for better behaved properties; in this case, the process $\mathcal{T}$ is called a subordinator, and the process X resulting of composition of $\hat{X}$ and $\mathcal{T}$ is said to be subordinate to $\hat{X}$. Subordination

[3]Terminology of financial derivatives is summarized in the first appendix.

[4]Various empirical facts are presented in the chapter on volatility together with the classification of models.

for the purpose of financial modeling was first utilized by P. Clark in [5], although the procedure of subordination is due to S. Bochner [6].

Above, we have separately touched upon volatility modeling and separately upon the concept of time in stochastic analysis. Yet it turns out that random changes in time are connected to random changes in volatility. Starting with a model with constant volatility - the Black-Scholes-Merton model - and plugging in a certain subordinator, one obtains the Heston stochastic volatility model. This is heuristically reasonable on the intuitive level; we shall present a rigorous derivation of this relationship in the last chapter. In the very same chapter, we shall include a possible generalization of the Heston model (in the sense that it contains the Heston as a special parametric case) that is due to P. Carr et al. [7]. In the process we also point out a connection of subordination to Bayesian approach to probability and to the physically motivated concept of superstatistics.

We build this text so that it is as self-contained as possible, while trying to keep the number of pages at bay. For reader's comfort, we have included an index at the end.

Chapter 1

Probability preliminaries

This chapter consists of standard material, whose excellent coverage can be found in [8]. We use [9] as a reference for probabilistic concepts. The purpose of this chapter is to clearly define notions that we are going to work with in the following, so that the text is more or less self-contained and there does not arise any confusion for the reader.

1.1 Stochastic processes

Here we only briefly summarize essential definitions and constructions, that we shall make use of in the upcoming chapters. Throughout the thesis we shall (sometimes implicitly) work in the setting of a measurable space $(\Omega, \mathcal{A})$ where Ω is the set of elementary events and $\mathcal{A}$ a σ-algebra. We also define an underlying probability measure denoted P, so that the space $(\Omega, \mathcal{A}, P)$ is actually a probability space. An event is said to happen *P-almost surely* (denoted P-a.s., or simply a.s.) if its probability measure is equal to one. In other words, the set of possible exceptions is of zero P-measure. Later in the text we shall define other probability measures on the measurable space $(\Omega, \mathcal{A})$; in that case we shall always explicitly specify with which particular measure we are working at the moment. The symbol $\mathcal{B}_n$ denotes the Borel σ-algebra generated by the usual topology of $\mathbb{R}^n$. On the same note, $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$ and $\mathcal{B}_+^0$ is the Borel σ-algebra on the half-line $\mathbb{R}_0^+ = [0, \infty)$.

Definition 1.1.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $I \subset \mathbb{R}$ an arbitrary non-empty subset. The system $\{X_t\}_{t \in I}$ of random variables on this probability space is called a *stochastic process*.

It is possible to define more general stochastic processes (for example complex valued), but we shall not pursue these more general cases, since, in what follows, they are not needed. From the next section on, we also drop the generality of the index set I and simply assume the stochastic processes to be defined on a closed interval $[0, T]$ or sometimes for all $t \geq 0$. Similarly, we have not specified where the stochastic process takes values, for, again, this set can be quite general. For our purposes the values of a stochastic process are always in some subset of $\mathbb{R}$ endowed with the corresponding Borel σ-algebra.

Definition 1.1.2. Let $\{X_t\}_{t \in I}$ be a stochastic process. The map $t \mapsto X_t(\omega)$ for an arbitrary fixed $\omega \in \Omega$ is called the *sample path* (or the *trajectory*) *of process* $\{X_t\}_{t \in I}$.

Definition 1.1.3. The *finite dimensional distributions of the stochastic process* $\{X_t\}_{t\in I}$ (*fidis*) are the distributions of finite-dimensional vectors

$$\{X_{t_1},\dots,X_{t_n}\},\qquad \forall\{t_1,\dots,t_n\}\subset I,\forall n\in\mathbb{N}.$$

Definition 1.1.4. Let $\mathbb{Y}:(\Omega,\mathcal{A})\to(\mathbb{R}^n,\mathcal{B}_n)$ be an n-dimensional random vector. The σ-algebra

$$\sigma(\mathbb{Y})=\{\mathbb{Y}^{(-1)}(B):B\in\mathcal{B}_n\}$$

is called the *σ-algebra generated*[1] *by the random vector* $\mathbb{Y}$.

Definition 1.1.5. For every stochastic process $X=\{X_t\}_{t\in I}$ (for which the quantities below are finite), we define:

- The *expectation function of* X as

$$\mu_X(t):=EX_t,\qquad \forall t\in I. \tag{1.1}$$

- The *covariance function of* X as

$$c_X(s,t):=E[(X_s-\mu_X(s))(X_t-\mu_X(t))],\qquad \forall s,t\in I. \tag{1.2}$$

- The *autocorrelation function of* X as

$$R_X(s,t):=E[X_sX_t],\qquad \forall s,t\in I. \tag{1.3}$$

Apparently, by the same calculation as for the random variables, the following relations hold:

$$c_X(s,t)=R_X(s,t)-EX_sEX_t,\qquad \forall s,t\in I$$

and

$$\mathrm{Var}X_t:=c_X(t,t)=E[(X_t-EX_t)^2]=EX_t^2-(EX_t)^2,\qquad \forall t\in I.$$

1.2 Conditional expectation

In this section we develop the minimum theory, that is necessary to work with the concepts defined later. Conditional expectations are essential, when working with stochastic processes, especially with certain classes such as martingales (defined in the next section).

Definition 1.2.1. Let $(\Omega,\mathcal{A},P)$ be a probability space. A random variable Z is called the *conditional expectation of the random variable* X *given the sub-σ-algebra* $\mathcal{S}\subset\mathcal{A}$ if

1. $\sigma(Z)\subset\mathcal{S}$,
2. $E[X\mathbf{1}_A]=E[Z\mathbf{1}_A],\qquad \forall A\in\mathcal{S}$.

The random variable Z (the conditional expectation) is usually denoted by $E[X|\mathcal{S}]$.

[1] The notation $\sigma(\mathbb{Y})$ is in line with the fact, that it is the smallest σ-algebra generated by the sets of the form $\{\omega\in\Omega:\mathbb{Y}(\omega)\in\mathcal{B}_n\}$.

Remark 1.2.2. The second requirement in the above definition could be also written as

$$\int_A Z(\omega)\,dP(\omega) = \int_A E[X|\mathcal{S}](\omega)\,dP(\omega) = \int_A X(\omega)\,dP(\omega), \qquad \forall A \in \mathcal{S}. \tag{1.4}$$

Obviously, when A is chosen to be the whole underlying set Ω, one obtains

$$E[E[X|\mathcal{S}]] = EX. \tag{1.5}$$

Remark 1.2.3. A proof of existence and uniqueness of the random variable $E[Z|\mathcal{S}]$ uses the Radon-Nikodým theorem and it is to be found in [8].

The second defining property of conditional expectation is not a constructive one. Therefore it is in general difficult to calculate $E[X|\mathcal{S}]$. For this reason, it is important for us to be able to work with conditional expectations without knowing their particular forms. The following theorem provides several useful properties.

Theorem 1.2.4 (Properties of conditional expectation)**.** Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{S}$ be a sub-σ-algebra of $\mathcal{A}$ and let X, Y be integrable random variables.

1. Conditional expectation is linear - for all $\alpha, \beta \in \mathbb{R}$, we have

$$E[\alpha X + \beta Y|\mathcal{S}] = \alpha E[X|\mathcal{S}] + \beta E[Y|\mathcal{S}].$$

2. If X and the σ-algebra $\mathcal{S}$ are independent, then

$$E[X|\mathcal{S}] = EX.$$

3. If $\sigma(X) \subset \mathcal{S}$, then $E[X|\mathcal{S}] = X$.

4. If XY is integrable and X is $\mathcal{S}$-measurable (i.e. $\sigma(X) \subset \mathcal{S}$), then

$$E[XY|\mathcal{S}] = XE[Y|\mathcal{S}].$$

5. Let $\tilde{\mathcal{S}}$ be a sub-σ-algebra $\tilde{\mathcal{S}} \subset \mathcal{S}$, then

$$E[E[X|\tilde{\mathcal{S}}]|\mathcal{S}] = E[X|\mathcal{S}] = E[E[X|\mathcal{S}]|\tilde{\mathcal{S}}].$$

1.3 Filtration, martingales

Definition 1.3.1. The collection $\{\mathcal{F}_t\}_{[0,T]}$ of σ-algebras on Ω is called a *filtration* if

$$\mathcal{F}_s \subset \mathcal{F}_t, \qquad \forall s, t \in [0,T], s \leq t.$$

Definition 1.3.2. The stochastic process $X = \{X_t\}_{[0,T]}$ is said to be *adapted to the filtration* $\{\mathcal{F}_t\}_{[0,T]}$ if

$$\sigma(X_t) \subset \mathcal{F}_t, \qquad \forall t \in [0,T].$$

In other words X is adapted to the filtration $\{\mathcal{F}_t\}$ if X_t is an $\mathcal{F}_t$-measurable random variable for every $t \in [0,T]$.

Definition 1.3.3. A *natural filtration generated by stochastic process* $X = \{X_t\}$ is defined by

$$\mathcal{F}_t := \{\sigma(X_s) : s \leq t\}, \qquad \forall t \in [0, T].$$

Remark 1.3.4. By definition, a stochastic process is always adapted to its own natural filtration.

Definition 1.3.5. The stochastic process $\{X_t\}$ is called a *(continuous-time) martingale with respect to the filtration* $\{\mathcal{F}_t\}$ if

1. $E|X_t| < +\infty, \quad \forall t,$
2. X is adapted to $\{\mathcal{F}_t\}$,
3. $E[X_t|\mathcal{F}_s] = X_s, \quad \forall s \leq t.$

Remark 1.3.6. It is quite usual to call a process a martingale without specifying the corresponding filtration. We shall do this only when the particular choice of filtration is obvious from the context.

Remark 1.3.7. Using the third defining property of a martingale $\{X_t\}$ with respect to $\{\mathcal{F}_t\}$ and the relation (1.5), one arrives at:

$$EX_s = E[E[X_t|\mathcal{F}_s]] = EX_t, \qquad \forall s \leq t. \tag{1.6}$$

Thus a constant expectation is a necessary condition for a process X to be a martingale with respect to some filtration and as such provides a useful way of proving that given process is not a martingale.

In the previous three sections, we have developed some theory of a general stochastic process; now we shall investigate two examples of this abstract mathematical object.

1.4 Poisson process

In 1903[2] F. Lundberg proposed to model the arrival of insurance claims by a Poisson process. Since then, in the field of actuary, the process plays a crucial role, however, for the purposes of continuous financial modeling, it might seem dispensable. The situation is quite different, when one introduces discontinuities to the models. Although we shall not make use of Poisson process directly, for modeling purposes, it often arises in the discussion as an analytical tool for the models with jumps.

Definition 1.4.1. A stochastic process $N = \{N_t\}_{\mathbb{R}_0^+}$ is called *(a homogeneous) Poisson process with intensity* $\lambda > 0$ if the following conditions are satisfied:

1. It starts at zero, so that $N_0 = 0$ P-almost surely.

[2]Interestingly, this was around the time, when also L. Bachelier and A. Einstein derived their essential results on other important stochastic process, that we shall point out in the next section.

2. The process has *stationary and independent increments*, i.e.[3]

$$N_{s+t} \stackrel{\mathcal{D}}{=} N_s + N_t, \qquad \forall s, t \geq 0$$

and $\{N_{t_j} - N_{t_{j-1}}\}_{j=1,\ldots,n}$ are independent random variables for all $n \in \mathbb{N}$ and for all $\{t_1, \ldots, t_n\}$, such that $0 \leq t_1 < \ldots < t_n < \infty$.

3. For every $t > 0$, N_t has a Poisson distribution $\mathrm{Po}(\lambda t)$.

4. Almost every trajectory of Poisson process is right continuous with limits from the left.

For the expectation and covariance functions, we have:

$$\mu_N(t) = EN_t = \lambda t, \qquad \forall t \geq 0 \tag{1.7}$$

and

$$c_N(s,t) = \lambda \min\{s,t\}, \qquad \forall s, t \geq 0. \tag{1.8}$$

These relations are derived the same way as for the single Poisson distributed random variable with intensity λt, which is, in fact, the case - for a given $t \geq 0$, we have

$$P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad \forall n. \tag{1.9}$$

1.5 Brownian motion

As pointed out by [10], the field of stochastic analysis was, from its very origins at the turn of the 20th century, fueled by two different sources - physics and finance. In physics, Brownian motion and related concepts serve as a useful model of diffusive phenomena and various types of random perturbations; in finance, the stochastic processes are utilized as a model for the behavior of price changes of a certain asset (such as a stock). Unsurprisingly, the two main sources fueling the stochastic analysis development are the two main areas of applications for it today.

The process known as Wiener process, or Brownian motion, was introduced for the first time in 1900 by L. Bachelier [1] to describe fluctuations in share prices. Much earlier (around 1827), R. Brown observed an irregular motion of pollen particles suspended in water. He noted, that the path of each particle is very irregular, having tangent at no point. This phenomenon was studied and successfully explained by A. Einstein in 1905 while rediscovering the process used earlier by Bachelier. Mathematically rigorous formulation of the process is due to R. Wiener.

Definition 1.5.1. A stochastic process $W = \{W_t\}_{\mathbb{R}_0^+}$ is called *standard Brownian motion* (or *standard Wiener process*) if the following conditions are satisfied:

1. It starts at zero, so that $W_0 = 0$ P-almost surely.

[3] We remind the reader that $X \stackrel{\mathcal{D}}{=} Y$ means that the random variable X is equal to the random variable Y in distribution. This does not imply the path-wise equality $X(\omega) = Y(\omega)$ for all $\omega \in \Omega$.

2. The process has *stationary and independent increments*, i.e.

$$W_{s+t} \stackrel{\mathcal{D}}{=} W_s + W_t, \qquad \forall s, t \geq 0$$

and $\{W_{t_j} - W_{t_{j-1}}\}_{j=1,\ldots,n}$ are independent random variables for all $n \in \mathbb{N}$ and for all $\{t_1, \ldots, t_n\}$, such that $0 \leq t_1 < \ldots < t_n < \infty$.

3. For every $t > 0$, W_t has a normal distribution $\mathrm{N}(0, t)$.

4. The process W has continuous sample paths.

Remark 1.5.2. Sometimes, Wiener process is defined by points $1. - 3.$ of the above definition and then Brownian motion is taken to be a continuous version of Wiener process via the Kolmogorov-Chentsov theorem (see D.1.2), where $\alpha = 4, \beta = 1$ and $C = 3$. Ergo, the point 4. in the definition is actually implied by the other defining properties and the mentioned theorem.

From the definition, one immediately obtains

$$\mu_W(t) := EW_t = 0, \qquad \forall t \geq 0. \tag{1.10}$$

Using the independence of increments, for $s \leq t$, we have

$$C_W(s,t) = R_W(s,t) = E[W_s W_t] = E[W_s(W_t - W_s + W_s)] = EW_s^2 + E[(W_t - W_s)W_s] =$$

$$= EW_s^2 + \underbrace{E[W_t - W_s]}_{=0}\underbrace{E[W_s - W_0]}_{=0} = EW_s^2 = s,$$

thus

$$C_W(s,t) = R_W(s,t) = \min\{s,t\}, \qquad \forall s, t \geq 0. \tag{1.11}$$

Although, as stated above, the Brownian motion has continuous trajectories, the degree of its non-predictability is such, that there does not exist a unique tangent for all points of almost every sample path. This idea of irregularity of Brownian paths is more correctly expressed by the following theorem, proof of which is obviously out of the scope of this text.

Theorem 1.5.3 (Paley-Wiener-Zygmund)**.** [11] The Brownian sample paths are almost surely nowhere differentiable.

Next, denoting $\mathcal{F}_t$ the natural Brownian filtration, we calculate:

$$\begin{aligned} E[W_t|\mathcal{F}_s] &= E[(W_t - W_s) + W_s|\mathcal{F}_s] = \\ &= E[W_t - W_s|\mathcal{F}_s] + E[W_s|\mathcal{F}_s] =, \\ &= E[W_t - W_s] + W_s = W_s \end{aligned} \tag{1.12}$$

for all $s \leq t$ and for all $t \geq 0$. Here, we have used the properties of conditional expectation (1.2.4) - linearity, the fact that $W_t - W_s$ is independent of $\mathcal{F}_s$ and the $\mathcal{F}_s$-measurability of W_s. The increment $W_t - W_s$ is $\mathrm{N}(0, t-s)$-distributed, ergo its expectation vanishes. Effectively, we have shown the following:

Proposition 1.5.4. The Brownian motion is a martingale with respect to its natural filtration.

Remark 1.5.5. On the other hand, iterating back to remark 1.3.7, $EW_t^2 = \mathrm{Var}(W_t) = t$. Hence there is no filtration such that $\{W_t^2\}$ is a martingale with respect to it, because $\{W_t^2\}$ does not have constant expectation. The condition of constant expectation function, however, is not a sufficient one, for $EW_t^3 = 0 = EW_s^3$ for all s, t, yet $\{W_t^3\}$ is not a martingale.

By a generally more involved procedure, one can show, that Brownian motion possesses the Markov property:

Definition 1.5.6. Let $X = \{X_t\}$ be a process adapted to filtration $\{\mathcal{F}_t\}$. Then we say, that X is a *Markov process* (*with respect to the filtration* $\{\mathcal{F}_t\}$) if for every non-negative Borel function f, there is Borel function g such that

$$E[f(X_t)|\mathcal{F}_s] = g(X_s), \qquad \forall s, t \geq 0, s \leq t. \tag{1.13}$$

Brownian motion defined above exhibits random diffusive behavior, without any preferred direction. Sometimes, it is useful to introduce the following process, in case one desires to still model randomness, but with a visible trend. The *Brownian motion with a drift* is defined (path by path) as the process $\{B_t\}$:

$$B_t := at + bW_t, \qquad \forall t \geq 0. \tag{1.14}$$

Here $a \in \mathbb{R}$ is responsible for the trend part (the *drift*) and $b > 0$ modulates the influence of variance of the original Brownian motion W.

We note, that Brownian motion is a *Gaussian process* (because its finite dimensional distributions are multivariate normal distributed) and so is Brownian motion with a drift.

1.6 Quadratic variation of Brownian motion

Quadratic variation is very concisely yet clearly explained in [12], we thus follow the same approach. In comparison to the preceding sections we go into more detail. Understanding what role the quadratic variation of Brownian motion plays, is important for the following text.

Definition 1.6.1. For a fixed interval $[0, t]$, we define a *partition* $\pi_n = \{t_0, t_1, \ldots, t_n\}$ which is a set of points dividing the interval so that $0 = t_0 < t_1 < \ldots < t_n = t$.

Norm of the partition (sometimes called *mesh*) is defined as

$$\nu(\pi_n) := \max_{j=1,\ldots,n} |t_j - t_{j-1}| .$$

A partition π_n' on the same interval $[0, t]$ is called a *refinement of the partition* π_n if it contains the same points as π_n and possibly more ($\pi_n' \supset \pi_n$).

Naturally, in the case of π_n' being a refinement of π_n we have for the norms:

$$\nu(\pi_n') \leq \nu(\pi_n).$$

In what follows we shall consider a sequence of refining partitions $\{\pi_n\}$ for which $\nu(\pi_n) \to 0$ as $n \to \infty$.

Definition 1.6.2. Let $f : \mathbb{R}_0^+ \to \mathbb{R}$ be a deterministic function defined in the closed interval $[0, t]$ and let $p \in \mathbb{N}$. The *p-variation of f up to time t* is

$$V^p(f)(t) := \lim_{\nu(\pi_n)\to 0} \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|^p , \tag{1.15}$$

where $\pi_n := \{t_0, t_1, \ldots, t_n\}$ is a partition of $[0, t]$ such that $0 = t_0 < t_1 < \ldots < t_n = t$ and the limit is taken over sequence of refining partitions, so that $\nu(\pi_n) \to 0$.

For $p = 1$ the p-variation is called just the *variation* and for $p = 2$ it is called the *quadratic variation*quadratic variation. From the definition it is obvious that the variation sums over projections of the graph of f onto the ordinate. Therefore it represents a measure of irregularity of the graph. Let us now focus more on the quadratic variation. Suppose, the function f has a continuous derivative in $[0, t]$. Then for every $j = 1, \ldots n$, there is $t_j^* \in (t_{j-1}, t_j)$ such that[4]:

$$\sum_{j=1}^{n} [f(t_j) - f(t_{j-1})]^2 = \sum_{j=1}^{n} [f'(t_j^*)]^2 (t_j - t_{j-1})^2 \leq \nu(\pi_n) \sum_{j=1}^{n} [f'(t_j^*)]^2 (t_j - t_{j-1}).$$

Thus

$$\begin{aligned} V^2(f)(t) &\leq \lim_{\nu(\pi_n)\to 0} \left\{ \nu(\pi_n) \sum_{j=1}^{n} [f'(t_j^*)]^2 (t_j - t_{j-1}) \right\} = \\ &= \lim_{\nu(\pi_n)\to 0} \nu(\pi_n) \cdot \lim_{\nu(\pi_n)\to 0} \sum_{j=1}^{n} [f'(t_j^*)]^2 (t_j - t_{j-1}) = \\ &= \underbrace{\lim_{\nu(\pi_n)\to 0} \nu(\pi_n)}_{=0} \cdot \int_0^t |f'(x)|^2 \, dx = 0. \end{aligned}$$

In the last step, we have used the fact, that the derivative f' is continuous to ensure that the Riemann integral $\int_0^t |f'(x)|^2 \, dx$ is finite. In the opposite case the formal product (whose value is not defined) $0 \cdot \infty$ would yield anything between 0 and ∞.

In the previous section we have stressed the fact that the paths of Brownian motion are almost surely nowhere differentiable, ergo the above does not apply. We define the quadratic variation of a stochastic process with slightly different notation:

Definition 1.6.3. Let $X = \{X_t\}$ be a stochastic process and let $\{\pi_n\}$ be a sequence of refining partitions as above. Then the limit[5]

$$\langle X \rangle_t := P\text{-}\lim_{\nu(\pi_n)\to 0} \sum_{j=1}^{n} (X_{t_j} - X_{t_{j-1}})^2 \tag{1.16}$$

is called the *quadratic variation of a stochastic process X*.

[4]Using the standard mean value theorem of real analysis.

[5]For definition of the limit in probability (denoted P-lim) and other types of convergence of random variables see the appendix D.

As we have already hinted it might be interesting to calculate the quadratic variation of Brownian motion. To that end it suffices to prove the limit in the mean square sense (definition in appendix D). The quadratic variation of this stochastic process is a deterministic function:

Theorem 1.6.4. Let $W = \{W_t\}$ be a Brownian motion. Then

$$\langle W \rangle_t = t, \qquad \forall t \geq 0.$$

Corollary 1.6.5. Let $W = \{W_t\}$ be a Brownian motion. Then[6]

$$V^1(W)_t = \infty, \qquad \forall t \geq 0.$$

Remark 1.6.6. We define a sequence of equidistant, refining partitions of interval $[0, T]$:

$$\pi_n = \{t_0, \dots t_n\}, \quad \text{where} \quad t_\ell := \frac{\ell t}{n}, \qquad \forall \ell = 0, \dots n.$$

Then

$$t_\ell - t_{\ell-1} = \frac{t}{n}$$

for all $\ell = 1, \dots, n$ and for the standard normal variables

$$Y_\ell := \frac{W_{t_\ell} - W_{t_{\ell-1}}}{\sqrt{t_\ell - t_{\ell-1}}}, \qquad \forall \ell = 1, \dots, n$$

the following equality holds:

$$(W_{t_\ell} - W_{t_{\ell-1}})^2 = (t_\ell - t_{\ell-1}) Y_\ell^2 = \frac{t}{n} Y_\ell^2. \tag{1.17}$$

The random variables Y_ℓ are independent (due to the independence of Brownian increments) and identically $\mathrm{N}(0, 1)$-distributed, so that the law of large numbers applies. Essentially, the sums $\frac{1}{n}\sum_{\ell=1}^n Y_1^2$ converge to the common mean EY_ℓ^2, as n tends to infinity, i.e.

$$\frac{1}{n} \sum_{\ell=1}^{n} \frac{(W_{t_\ell} - W_{t_{\ell-1}})^2}{t_\ell - t_{\ell-1}} \overset{n\to\infty}{\longrightarrow} EY_1^2 = 1.$$

Each of the terms $(W_{t_\ell} - W_{t_{\ell-1}})^2$ in this sum is quite different from its mean $t_\ell - t_{\ell-1} = \frac{t}{n}$, however, under sum, the differences average out. These ideas are reflected in the following relation

$$dW_t \, dW_t = dt. \tag{1.18}$$

This, naturally, is not the usual relationship between differentials, since Brownian sample paths are almost surely nowhere differentiable. Nevertheless, this notation is rather

[6]For a stochastic process X and a sequence of refining partitions as in the previous, the variation is

$$V^1(X)_t = P\text{-}\lim_{\nu(\pi_n)\to 0} \sum_{j=1}^{n} \left| X_{t_j} - X_{t_{j-1}} \right|$$

so that a stochastic process has finite variation on interval $[0, t]$ if its sample paths have finite variation with probability one.

concise shorthand for the concepts described above, so that it is advantageous to use it; one has to bear in mind, that it holds only in large sums (as implied by the law of large numbers). By analogical arguments, one gives meaning to

$$\begin{aligned} dW_t\, dt &= 0 \\ dt\, dt &= 0 \end{aligned} \tag{1.19}$$

as well.

In the next chapter, we shall define stochastic integral with respect to Brownian motion. Because an integral is usually some sort of sum, expressions similar to (1.18) and (1.19) shall prove quite useful.

Chapter 2

Itô integral and stochastic differential equations

In this section we lay the foundations for the mathematics that are part of a standard toolbox of quantitative finance practitioners. Leveraging our mathematical background we try to go more into detail (where it is possible) in order to rigorously arrive at our results. Our understanding is greatly indebted to [13] and [8] yet again.

Our goal is to define an integral of a stochastic process $X = \{X_t\}$ with respect to the Brownian motion $W = \{W_t\}$, so that we arrive at a new random variable

$$\int_0^T X_s \, dW_s, \tag{2.1}$$

or even a stochastic process $\{\int_0^t X_s \, dW_s\}_{t \in \mathbb{R}_0^+}$. These are just formal symbols for now. The problem that arises when one tries to approach this naively is that the integral (2.1) cannot be defined in the usual Riemann-Stieltjes or Lebesgue-Stieltjes way, because the sample paths of Brownian motion are almost surely nowhere differentiable and also the variation of the process is infinite (see corollary 1.6.5). Ergo naive stochastic integration is not possible. [14]

The motivation for introducing stochastic integral is the desire to model the random behavior of a certain system via a differential equation (called a stochastic differential equation)

$$dX_t = \alpha(t, X_t) \, dW_t, \tag{2.2}$$

where W_t is a source of randomness (in our case Brownian motion) and $\alpha = \alpha(t, x)$ is a given coefficient function. The process $X = \{X_t\}$ is a solution of this stochastic differential equation that is to be found. Inspired by the most common approach to solving an ordinary differential equation (aside of simply guessing) - the separation of variables and integration - we deem the construction of stochastic integral inevitable. In what follows, we give precise meaning to formal symbols introduced above, the differential dW_t (equipped with the knowledge of non-differentiability of Brownian paths, this certainly is not the derivative of W) and explain the procedure of finding a solution and explain what precisely comprises a solution of equation (2.2).

2.1 Riemann integral of stochastic process

Before we pursue the construction of stochastic integral, we briefly introduce the Riemann integration of a stochastic process. This concept is sometimes omitted in the literature on stochastic integration, although it is implicitly used.

Definition 2.1.1. Let $X = \{X_t\}$ be a stochastic process and let $\{\pi_n\}$ be a sequence of refining partitions[1] of the interval $[0, t]$ for some t for which X is defined. Then the stochastic process defined as[2]

$$\int_0^t X_s\,ds := L^2\text{-}\lim_{n\to+\infty}\sum_{j=1}^{n} X_{t_j}(t_j - t_{j-1}) = L^2\text{-}\lim_{\|\pi_n\|\to 0}\sum_{j=1}^{n} X_{t_j}(t_j - t_{j-1}), \qquad \forall t, \quad (2.3)$$

is called the *Riemann integral of the stochastic process* X *on the interval* $[0, t]$.

Thus the existence of the integral $\int_0^t X_s\,ds$ is essentially determined by the existence of the limit on the right-hand side of (2.3). One sufficient criterion to this effect, for a process X with second finite moment, is the existence of the ordinary Riemann integral

$$\int_0^t \int_0^t R_X(x, y)\,dxdy, \qquad \forall t,$$

where R_X is the autocorrelation function of the process X.

2.2 Itô integral

In the text below, we present the definition of the stochastic integral with respect to Brownian motion, the so-called Itô integral. We note, that in general it is possible to define stochastic integral with respect to processes from a wider class such as Lévy processes or semimartingales [14] (see the corresponding chapter and appendix for definitions).

Definition 2.2.1. The stochastic process $C := \{C_t\}_{[0,T]}$ is said to be *simple* if there exists a partition

$$\pi_n := \{t_0, \ldots, t_n\}, \quad \text{where} \quad 0 = t_0 < t_1 < \ldots < t_n = T$$

and a sequence of random variables $\{Z_j\}_{j\in\hat{n}}$ such that

$$C_t := \begin{cases} Z_n, & \text{if} \quad t = T, \\ Z_j, & \text{if} \quad t_{j-1} \le t < t_j \text{ for } j \in \hat{n}. \end{cases}$$

Furthermore, the sequence $\{Z_j\}$ is adapted to natural Brownian filtration $\{\mathcal{F}_{t_{j-1}}\}$.

[1] As defined in section 1.6.

[2] Basically, one can define the Riemann integral path by path as

$$\left(\int_0^t X_s\,ds\right)(\omega) = \int_0^t X_s(\omega)\,ds, \qquad \forall \omega \in \Omega$$

and from the theory of ordinary Riemann integral of deterministic functions it is known, that the result of integration does not depend on the choice of the points in the partition subintervals at which the process X is evaluated (unlike for the Itô stochastic integral defined later).

Definition 2.2.2. Consider the partition $\pi_n = \{t_0, \dots, t_n\}$ and the process C defined above. We define:

- *Itô stochastic integral of a simple process C on $[0,T]$* as:

$$\int_0^T C_s \, dW_s := \sum_{j=1}^{n} C_{t_{j-1}}(W_{t_j} - W_{t_{j-1}}) = \sum_{j=1}^{n} Z_j(W_{t_j} - W_{t_{j-1}}). \tag{2.4}$$

- *Itô stochastic integral of a simple process C on the subinterval $[0,t] \subset [0,T]$*, where $t \in [t_{k-1}, t_k]$, as

$$I_t(C) := \int_0^t C_s \, dW_s := \int_0^T C_s \mathbf{1}_{[0,t]}(s) \, dW_s = \sum_{j=1}^{k-1} Z_j(W_{t_j} - W_{t_{j-1}}) + Z_k(W_t - W_{t_{k-1}}). \tag{2.5}$$

Remark 2.2.3. It is easy to see that the Itô integral, being constructed in this way, depends on our choice of evaluation of the simple process C in the left end-points of intervals $[t_{j-1}, t_j)$ (in particular equation (2.4)). This reflects in many of the properties of the integral listed below and especially in the Itô formula, which basically states that the Leibniz rule in the usual sense in Itô calculus does not hold. There are other integrals, similar in construction to our approach, the only difference being in evaluation of the process in other points of the intervals $[t_{j-1}, t_j)$. A specific example which chooses the mid-points for evaluation of the simple process C in the sums (2.4) is the so-called *Stratonovich stochastic integral*. For this particular choice, the usual Leibniz rule is retained, however other nice properties of Itô integral are lost. In finance the standard approach is via Itô integration, however the Stratonovich integral finds its use in solving stochastic differential equations [8]. We shall not pursue these methods here.

Considering a common refinement of two partitions for two simple processes $C^{(1)}, C^{(2)}$ it is apparent that the Itô integral is linear in the following sense:

$$I(\alpha C^{(1)} + \beta C^{(2)}) = \alpha I(C^{(1)}) + \beta I(C^{(2)}), \qquad \forall \alpha, \beta \in \mathbb{R}.$$

Theorem 2.2.4. The Itô integral defined above is a martingale with respect to natural Brownian filtration $\{\mathcal{F}_t\}_{[0,T]}$.

Corollary 2.2.5. Because $I(C)$ is a martingale, we have by the constancy of its expectation:

$$0 = E[I_0(C)] = E[I_t(C)], \qquad \forall t \in [0,T].$$

Theorem 2.2.6 (Itô isometry)**.** The Itô stochastic integral satisfies the isometry property:

$$E[I_t(C)]^2 = E\left[\int_0^t C_s \, dW_s\right]^2 = \int_0^t EC_s^2 \, ds, \qquad \forall t. \tag{2.6}$$

Theorem 2.2.7. The quadratic variation accumulated up to time t by the Itô stochastic integral is

$$\langle I(C) \rangle_t = \int_0^t C_s^2 \, ds. \tag{2.7}$$

Definition 2.2.8. Let $\mathcal{V} = \mathcal{V}([0,T])$ denote the class of functions

$$X : \mathbb{R}_0^+ \times \Omega \to \mathbb{R},$$

such that

1. the map $(t,\omega) \mapsto X(t,\omega)$ is $\mathcal{B}_+^0 \times \mathcal{A}$-measurable (here $\mathcal{B}_+^0$ denotes the Borel σ-algebra on $\mathbb{R}_0^+$),
2. $X(t,\omega)$ is $\mathcal{F}_t$-adapted,
3. $E\left[\int_0^T X(t,\omega)^2 dt\right] < +\infty$.

Definition 2.2.9. Let $X \in \mathcal{V}([0,T])$. *Then the Itô integral of* X *on* $[0,T]$ is defined by[3]

$$\int_0^T X_t \, dW_t := L^2\text{-}\lim_{n\to+\infty} \int_0^T \xi_t^{(n)} \, dW_t, \tag{2.8}$$

where $\{\xi^{(n)}\}$ is a sequence of simple stochastic processes such that

$$E\left[\int_0^T (X_t - \xi_t^{(n)})^2 \, dt\right] \to 0. \tag{2.9}$$

Itô integral on a subinterval $[0,t] \subset [0,T]$ is defined as

$$\int_0^t X_s \, dW_s := \int_0^T X_s I_{[0,t]}(s) \, dW_s, \qquad \forall t \in [0,T].$$

In [13] the author proves that for a stochastic process $X \in \mathcal{V}([0,T])$ the approximating sequence (2.9) always exists. The Itô isometry 2.2.6 is then used to show that $I_t(\xi^{(n)})$ is a Cauchy sequence, ergo, by the completeness of L^2, the mean square limit (2.8) exists. Furthermore, the result of the stochastic integration as defined above does not depend on the choice of the approximating sequence (because generally, that is not unique), and the properties described in theorems 2.2.4 (for a continuous version of Itô integral) and 2.2.7 hold for the general integrand $X \in \mathcal{V}([0,T])$ as well.

Theorem 2.2.10 (Itô formula for Brownian motion)**.** Let $f \in \mathcal{C}^2(\mathbb{R}_0^+ \times \mathbb{R})$. Then for every $t \in [0,T]$

$$f(t,W_t) = f(0,W_0) + \int_0^t f_s(s,W_s) \, ds + \int_0^t f_x(s,W_s) \, dW_s + \frac{1}{2}\int_0^t f_{xx}(s,W_s) \, ds, \tag{2.10}$$

which in the differential notation becomes

$$df(t,W_t) = f_t(t,W_t) \, dt + f_x(t,W_t) \, dW_t + \frac{1}{2} f_{xx}(t,W_t) \, dt.$$

[3]For definition of mean square limit or, in other words, limit in L^2, we again point the reader to section D.2.

Remark 2.2.11. In the formula above, we have quietly assumed the so-called *differential notation*, now, we shall comment on that. Because the Itô integral (and so the Itô differentiation) does not behave the way, one would expect from ordinary calculus (see the following remark), while working with the differential notation, one has to bear in mind, that it is just a shorthand and the proper integral relationship (2.10) gives meaning to these differential relations. However, it is not only useful and concise to work with this notation, but also safe, once we respect the Itô formula and the relations (1.18) and (1.19).

Remark 2.2.12. With the way the usual differentiation works, one has for the function $f(x) = x^2$ with the domain $\mathbb{R}$:

$$df(x) = d(x)^2 = 2x\,dx,$$

however, Itô formula above implies, that for the same function of Brownian motion

$$df(W_t) = dW_t^2 = 2W_t\,dW_t + dt$$

holds. Now, this is not as confusing, once we take into account, that the "differentials" in Itô formula are not actually derivatives in the usual sense. This "violation" of the usual Leibniz rule is exactly the consequence of our choice in the definition of Itô integral mentioned in remark 2.2.3.

Definition 2.2.13. Let $\{\mathcal{F}_t\}_{[0,T]}$ be a Brownian filtration. An *Itô process* is a stochastic process of the form

$$X_t = X_0 + \int_0^t A_s^{(1)}\,dW_s + \int_0^t A_s^{(2)}\,ds,$$

where X_0 is non-random, $\{A_s^{(1)}\} \in \mathcal{V}$ and $\{A_s^{(2)}\}$ is a stochastic process which satisfies the condition $\int_0^t |A_s^{(2)}|ds < +\infty$.

Theorem 2.2.14. The quadratic variation of the Itô process is

$$\langle X\rangle_t = \int_0^t [A_s^{(1)}]^2 ds.$$

Definition 2.2.15. Let X be an Itô process and let $\{Y_t\}_{[0,T]}$ be a stochastic process adapted to $\{\mathcal{F}_t\}$ such that

$$E\int_0^t Y_s^2 [A_s^{(1)}]^2 ds \quad \text{and} \quad \int_0^t |Y_s A_s^{(2)}|\,ds$$

are finite for every $t \in [0,T]$. We define the *integral with respect to an Itô process* as:

$$\int_0^t Y_s\,dX_s := \int_0^t Y_s A_s^{(1)}\,dW_s + \int_0^t Y_s A_s^{(2)}\,ds, \qquad \forall t \in [0,T].$$

Theorem 2.2.16 (Itô formula for an Itô process)**.** Let $\{X_t\}$ be an Itô process and let $f \in \mathcal{C}^2(\mathbb{R}_0^+ \times \mathbb{R})$ be a function. Then for every $t \in [0,T]$:

$$f(t,X_t) = f(0,X_0) + \int_0^t f_s(s,X_s)\,ds + \int_0^t f_x(t,X_t)\,dX_t + \frac{1}{2}\int_0^t f_{xx}(s,X_s)\,\underbrace{d\langle X\rangle_s}_{=[A_s^{(1)}]^2\,ds}.$$

In the differential notation:

$$df(t, X_t) = f_t(t, X_t)\, dt + f_x(t, X_t)\, dX_t + \frac{1}{2} f_{xx}(t, X_t)[A_t^{(1)}]^2\, dt.$$

There is, of course, an Itô formula for functions of multiple Itô processes (in particular two), which we shall not include here, because the notation and the calculations can get rather chaotic. Nevertheless, the two-dimensional Itô formula has the following neat consequence, which we shall use often in later applications:

Corollary 2.2.17 (Itô product rule)**.** Let $\{X_t\}$ and $\{Y_t\}$ be Itô processes, then

$$d(X_t Y_t) = Y_t\, dX_t + X_t\, dY_t + dX_t\, dY_t, \qquad \forall t. \tag{2.11}$$

We stress for the last time that this again shows the violation of the usual Leibniz rule.

2.3 Stochastic differential equations

Definition 2.3.1. Let $a, b : \mathbb{R}^2 \to \mathbb{R}$ be functions. We call the formal equality

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t \tag{2.12}$$

(with X_0 non-random given as initial condition) a *stochastic differential equation (SDE).*

Remark 2.3.2. The formal relationship above is justified by the "proper" integral form

$$X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t b(s, X_s)\, dW_s \tag{2.13}$$

and the conditions on the coefficient functions

$$\int_0^t |a(s, X_s)|\, ds < \infty, \tag{2.14}$$

$$\int_0^t [b(s, X_s)]^2 ds < \infty, \tag{2.15}$$

where these inequalities must hold for every $t \geq 0$, P-almost surely.

Definition 2.3.3. A process $X = \{X_t\}$ adapted to natural Brownian filtration $\{\mathcal{F}_t\}$ that satisfies relations (2.13),(2.14) and (2.15) is called a *strong solution of the stochastic differential equation* (2.12).

A characteristic feature of the above definition is that the Brownian motion $W = \{W_t\}$ is assumed to be given a priori and the process $X = \{X_t\}$ should be constructed from it in an adapted way so that we have the functional relationship $X = \mathscr{F}(W)$. Being constructed this way, a strong solution of (2.12) is based on the sample paths of W. If the underlying Brownian motion were to change to a different Brownian motion $\tilde{W}$, there would be a different corresponding strong solution, that still satisfies the same functional relationship $X = \mathscr{F}(\tilde{W})$.

There is a well-known result on the uniqueness of the strong solution:

Theorem 2.3.4 (Itô)**.** Let $a = a(t,x), b = b(t,x)$ be measurable functions given as in (2.14) and (2.15) satisfying

$$|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| \leq C\,|x-y|\,, \qquad \forall x,y \in \mathbb{R}, t \in [0,T), \tag{2.16}$$

for some constant C and

$$|a(t,x)| + |b(t,x)| \leq K(1+|x|), \qquad \forall x \in \mathbb{R}, t \in [0,T), \tag{2.17}$$

where K is also a constant. Then the equation (2.12) has a unique strong solution $X = \{X_t\}$.

Remark 2.3.5. Here, the uniqueness of the strong solution X means that if there is another such strong solution Y of (2.12), then the two processes are indistinguishable, i.e.[4]

$$P(X_t = Y_t) = 1, \qquad \forall t \in [0,T].$$

An example of the above defined SDE is a special case of *linear stochastic differential equation*:

$$dX_t = (a_1 X_t + a_2)\,dt + (b_1 X_t + b_2)\,dW_t, \qquad \forall t, \tag{2.18}$$

where $a_1, a_2, b_1, b_2 \in \mathbb{R}$ are constants, such that the linear functions

$$a(x) := a_1 x + a_2, \qquad b(x) := b_1 x + b_2$$

satisfy relations (2.14) and (2.15). As a result of theorem 2.3.4 the linear SDE (2.18) always has a unique strong solution.

Example 2.3.6 (Geometric Brownian motion). [8] Let us consider the following linear SDE which is of great importance to quantitative finance, with many applications:

$$dX_t = aX_t\,dt + bX_t\,dW_t. \tag{2.19}$$

This equation is obviously a particular case of (2.18), solution of which can be found using the Itô formula 2.2.10. We assume the sought after solution is a function $X_t = f(t, W_t)$ for some sufficiently smooth f, so that the Itô formula applies:

$$dX_t = df(t,W_t) = f_t(t,W_t)\,dt + f_x(t,W_t)\,dW_t + \frac{1}{2} f_{xx}(t,W_t)\,dt.$$

Comparing this with (2.19), we obtain the following system of deterministic partial differential equations:

$$\begin{cases} af(t,x) = f_t(t,x) + \frac{1}{2} f_{xx}(t,x) \\ bf(t,x) = f_x(t,x) \end{cases}.$$

From the second equation, we obtain

$$f_{xx}(t,x) = bf_x(t,x) = b^2 f(t,x),$$

which we can substitute into the first equation, so that our system becomes

$$\begin{cases} \left(a - \frac{1}{2}b^2\right) f(t,x) & = f_t(t,x) \\ bf(t,x) & = f_x(t,x) \end{cases}.$$

[4]This definition and more on the topic is included in the section D.1 of appendix D.

In both of these equations, there figure derivatives only with respect to one variable but not with respect to the other. Therefore, we can employ the assumption, that there are functions $g = g(t)$, $h = h(x)$ such that

$$f(t,x) = g(t)h(x).$$

Using this, we arrive at two ordinary differential equations (ODEs):

$$\begin{cases} \left(a - \frac{1}{2}b^2\right) g(t) & = g'(t) \\ bh(x) & = h'(x) \end{cases}.$$

Both of these ODEs are promptly solved by separation of variables, so that

$$g(t) = g(0)e^{\left(a-\frac{1}{2}b^2\right)t}, \qquad h(x) = h(0)e^{bx}.$$

Iterating back to our initial assumption:

$$f(t,x) = g(t)h(x) = g(0)h(0)\exp\left\{\left(a - \frac{1}{2}b^2\right)t + bx\right\}.$$

We plug in the initial condition

$$X_0 = f(0, W_0) = f(0,0) = g(0)h(0),$$

and finally:

$$X_t = X_0 \exp\left\{\left(a - \frac{1}{2}b^2\right)t + bW_t\right\}. \tag{2.20}$$

Because (2.19) is a linear SDE, the stochastic process (2.20) is necessarily the unique strong solution. It bears the name of *geometric Brownian motion (GBM)* and in many models of mathematical finance it is assumed to model the behavior of some equity underlying asset (a stock, for example). Apparently the geometric Brownian motion is a non-Gaussian process, since it is log-normally distributed. If we denote the constant $c := a - \frac{1}{2}b^2$, using the moment generating function of $\mathrm{N}(ct, b^2t)$-distributed random variable, it is possible to determine the expectation function of GBM as

$$\mu_X(t) = \exp\left\{\left(c + \frac{1}{2}b^2\right)t\right\}, \qquad \forall t \geq 0.$$

From this, recalling the remark 1.3.7 - since the expectation function is not constant, it is obvious that geometric Brownian motion is not a martingale (i.e. there is no such filtration, that it is a martingale w.r.t. it) under the probability measure P. On the other hand, it is possible to prove, that geometric Brownian motion is a Markov process in the sense of definition 1.5.6. [12]

We have defined a strong solution of the general stochastic differential equation (2.12), now we shall proceed to define a weak solution.

Definition 2.3.7. A collection of objects

$$(\Omega, \mathcal{A}, P), \{\mathcal{F}_t\}, W, X$$

such that

1. $(\Omega, \mathcal{A}, P)$ is a probability space, $\{\mathcal{F}_t\}$ given filtration on it,
2. $W = \{W_t\}, X = \{X_t\}$ are $\{\mathcal{F}_t\}$-adapted stochastic processes,
3. $W = \{W_t\}$ is a Brownian motion with respect to P,
4. for each $t \geq 0$ the relations (2.13),(2.14) and (2.15) hold,

is called a *weak solution of* (2.12) given $a = a(t,x), b = b(t,x)$.

In contrast to a strong solution a weak solution is not required to be a functional of a priori given Brownian motion W. Actually, finding a certain Brownian motion is a part of the process of solving (2.12) in the weak sense. Weak solutions X of (2.12) are useful to determine the distributional characteristics of X, such as the expectation, variance and covariance functions of the process. For these purposes, the sample paths of X need not be known.

Remark 2.3.8. There are stochastic differential equations for which it can be shown, that the strong solution does not exist, however, they have a weak solution.

Example 2.3.9 (CIR equation). We assume a stochastic differential equation of the form:

$$dX_t = k(n - X_t)\,dt + b\sqrt{X_t}\,dW_t, \qquad \forall t \geq 0. \tag{2.21}$$

Because this equation found its use in modeling interest rates in famous article [15] in the so-called *Cox-Ingersoll-Ross (CIR) model*, it is usually referred to as the *CIR equation* or *CIR process*. Sometimes the equation is also called the *square-root equation*, for obvious reasons.

Unlike equation (2.19), the CIR equation does not have a closed-form solution. Although there is no formula for $\{X_t\}$, there is a unique solution to this SDE starting from a given initial condition X_0. [12] It is possible to determine many of the distributional properties of the CIR process. We start by calculating the expectation function; using the Itô product rule:

$$\begin{aligned} d(e^{kt}X_t) &= ke^{kt}X_t\,dt + e^{kt}\,dX_t = \\ &= ke^{kt}X_t\,dt + e^{kt}k(n - X_t)\,dt + e^{kt}b\sqrt{X_t}\,dW_t = \\ &= e^{kt}kn\,dt + e^{kt}b\sqrt{X_t}\,dW_t. \end{aligned}$$

We integrate:

$$e^{kt}X_t = e^0X_0 + kn\underbrace{\int_0^t e^{ks}\,ds}_{=\frac{1}{k}(e^{kt}-1)} + b\int_0^t e^{ks}\sqrt{X_s}\,dW_s,$$

and taking the expectation:

$$e^{kt}EX_t = X_0 + n(e^{kt} - 1) + b\underbrace{E\left[\int_0^t e^{ks}\sqrt{X_s}\,dW_s\right]}_{=0}.$$

Altogether, we arrive at the first moment:

$$EX_t = (X_0 - n)e^{-kt} + n, \qquad \forall t \geq 0. \tag{2.22}$$

By analogical steps, it is possible to derive the second moment and specifically, the variance:

$$\mathrm{Var}X_t = X_0\frac{b^2}{k}(e^{-kt} - e^{-2kt}) + \frac{n}{2}\frac{b^2}{k}(1 - e^{-kt})^2, \qquad \forall t \geq 0. \tag{2.23}$$

One of the features of the process (2.21) is that it is *mean reverting*, i.e. once the process X_t gets very far from the value of n, the first term (the drift part) in the equation prevails and drags the process back. This reversion to the mean was one of the reasons, why the CIR process became popular in modeling interest rates. Another reason is that the process does not take on negative values. In fact, if the condition

$$2kn > b^2$$

holds, then the drift part of the process is sufficient for the process to stay positive and not reach zero. [16] Currently, the CIR process is losing on popularity in fixed income modeling, since, as recent developments have shown, negative interest rates are not unusual in the real-life markets.

Just as interest rates, the variance of an equity asset is believed to have the reversion to the mean property; moreover, variance should never become negative by definition. Due to these facts, the CIR process is employed to model the variance of the underlying asset in the Heston model, that we shall introduce later (see the corresponding chapter).

In bond pricing, often the *integrated CIR process* $\{Z_t\}$ defined as

$$Z_t := \int_0^t X_s\, ds, \qquad \forall t \geq 0 \tag{2.24}$$

enters the discussion. We shall utilize it in a slightly different manner in our derivations presented in the last chapter. It can be shown, that the characteristic function of integrated CIR process is [17]

$$\phi_{Z_t}(u) = A(t,u)e^{B(t,u)X_0}, \tag{2.25}$$

where

$$A(t,u) := \frac{\exp\left(\frac{k^2 n}{b^2}t\right)}{\left[\cosh\left(\frac{\Upsilon}{2}t\right) + \frac{k}{\Upsilon}\sinh\left(\frac{\Upsilon}{2}t\right)\right]^{\frac{2kn}{b^2}}},$$

$$B(t,u) := \frac{2iu}{k + \Upsilon\coth\left(\frac{\Upsilon}{2}t\right)}$$

with

$$\Upsilon := \sqrt{k^2 - 2b^2 iu}.$$

2.4 Generalized geometric Brownian motion

Let $\{W_t\}$ be a Brownian motion with $\{\mathcal{F}_t\}$ filtration and let $\{\mu_t\}, \{\sigma_t\}$ be processes, adapted to $\{\mathcal{F}_t\}$. We define an Itô process (by definition 2.2.13)

$$H_t := \int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2\right) ds + \int_0^t \sigma_s\, dW_s. \tag{2.26}$$

Written in the differential form:

$$dH_t = \left(\mu_t - \frac{1}{2}\sigma_t^2\right)\,dt + \sigma_t\,dW_t, \tag{2.27}$$

where also[5]

$$dH_t\,dH_t = \sigma_t^2\,dW_t\,dW_t = \sigma_t^2\,dt. \tag{2.28}$$

We define another process, by

$$X_t := X_0 e^{H_t} = X_0 \exp\left\{\int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2\right)\,ds + \int_0^t \sigma_s\,dW_s\right\}. \tag{2.29}$$

Here X_0 is deterministic and positive. We use the Itô formula for an Itô process:

$$\begin{aligned} dX_t = X_0 d(e^{H_t}) &= X_0(e^{H_t}\,dH_t + \frac{1}{2}e^{H_t} dH_t\,dH_t) = \\ &= X_t\left[\left(\mu_t - \frac{1}{2}\sigma_t^2\right)\,dt + \sigma_t\,dW_t\right] + \frac{1}{2}X_t\sigma_t^2\,dt = \\ &= \mu_t X_t\,dt + \sigma_t X_t\,dW_t. \end{aligned}$$

In these calculations, we have verified, that a process

$$X_t = X_0 \exp\left\{\int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2\right)\,ds + \int_0^t \sigma_s\,dW_s\right\}, \qquad \forall t \geq 0, \tag{2.30}$$

called the *generalized geometric Brownian motion* is a solution of the equation:

$$dX_t = \mu_t X_t\,dt + \sigma_t X_t\,dW_t, \qquad \forall t \geq 0. \tag{2.31}$$

Because we have not defined stochastic differential equations with stochastic processes as coefficients, this is not strictly speaking a solution of an SDE. It would be more correct to call the GGBM a different integral representation of the Itô process defined in differential notation as (2.31). Unlike for the usual geometric Brownian motion, the distribution of X_t need not be log-normal, because μ_t and σ_t are allowed to be stochastic. If these coefficient processes are in fact constants μ and σ, then the solution (2.30) becomes the usual GBM, which was derived in example 2.3.6. We note here, that generalized geometric Brownian motion is not a martingale and that it also represents the most general model of an asset price, that is always positive, has no discontinuities (in the sense of sample paths) and is driven by a single Brownian motion. [12]

[5]This also mimics theorem 2.2.14, since

$$\langle H\rangle_t = \int_0^t \underbrace{dH_s\,dH_s}_{d\langle H\rangle_s} = \int_0^t \sigma_s^2 ds.$$

2.5 Some important theorems in Itô calculus

The above developed theory of the Itô integral and the SDEs based on it are sometimes together briefly referred to as the Itô calculus. In sections 1.5 and 1.6 we have shown that Brownian motion $\{W_t\}$ is a martingale (w.r.t. natural Brownian filtration) starting at zero with continuous sample paths that accrues quadratic variation as $\langle W \rangle_t = t$. It turns out that these properties together are sufficient for a process to be a Brownian motion. This is the content of the following theorem.

Theorem 2.5.1 (Lévy characterization)**.** Let $\{M_t\}$ be a martingale with respect to the filtration $\{\mathcal{F}_t\}$. If

1. $M_0 = 0$,
2. $\{M_t\}$ has continuous sample paths,
3. $\langle M \rangle_t = t, \qquad \forall t \geq 0$,

then $\{M_t\}$ is a Brownian motion with $\{\mathcal{F}_t\}$ its natural filtration.

Although this theorem does not need or reference the machinery of Itô calculus at all, it is often used in proofs of many cornerstone theorems in stochastic calculus. The theorem is also interesting, because there is no assumption on normality of the distribution of the martingale M, however the other assumptions are enough to imply it.

Another fundamental theorem, that we display here for reference is the martingale representation theorem.

Theorem 2.5.2 (Martingale representation)**.** Let $\{W_t\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{A}, P)$ with natural filtration $\{\mathcal{F}_t\}$. Let $\{M_t\}$ be a martingale with respect to this filtration, then there is a process $\{G_t\}$ such that

$$M_t = M_0 + \int_0^t G_s \, dW_s, \qquad \forall t \in [0, t].$$

2.6 Equivalent measures, Girsanov theorem

So far, we have worked with only single probability measure P on the measurable space $(\Omega, \mathcal{A})$, thus we did not need to specify the measure explicitly on some occasions. For example the expectation in the definition of Brownian motion (used for independence of increments) is actually tied to the probability measure as well as the conditional expectation in the martingale assumption of Lévy characterization 2.5.1. In this section we shall state some important facts regarding definition of different probability $\tilde{P}$ and the relationship between the measures P and $\tilde{P}$ alongside the relationship between different Brownian motions.

We shall start with some definitions.

Definition 2.6.1. Let P_1 and P_2 be two probability measures given on the same measurable space $(\Omega, \mathcal{A})$. Then we say that the two measures are *equivalent* if

$$P_1(A) = 0 \quad \iff \quad P_2(A) = 0, \qquad \forall A \in \mathcal{A}.$$

Given a P-almost surely positive random variable Z with $EZ = 1$ it is possible to define another probability measure $\tilde{P}$ by:

$$\tilde{P}(A) := \int_A Z(\omega)\, dP(\omega), \qquad \forall A \in \mathcal{A}. \tag{2.32}$$

Theorem 2.6.2 (Girsanov)**.** [12] Let $\{W_t\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{A}, P)$ and let $\{\mathcal{F}_t\}$ be a natural filtration for this Brownian motion. Let $\{\zeta_t\}_{[0,T]}$ be a stochastic process adapted to this Brownian filtration. Define:

$$Z_t := \exp\left\{-\int_0^t \zeta_s\, dW_s - \frac{1}{2}\int_0^t \zeta_s^2\, ds\right\}, \qquad t \in [0,T], \tag{2.33}$$

$$\tilde{W}_t := W_t + \int_0^t \zeta_s\, ds \tag{2.34}$$

and assume that

$$E\int_0^T \zeta_t^2 Z_t^2\, dt < \infty. \tag{2.35}$$

Set $Z := Z_T$. Then $EZ = 1$ and under the probability measure $\tilde{P}$ given by (2.32), the process $\{\tilde{W}_t\}_{[0,T]}$ is a Brownian motion.

In fact the two measures P and $\tilde{P}$ are equivalent; because of that, $\tilde{P}$ is called *equivalent martingale measure*[6]. It is important that the two measures are equivalent, so that they agree on which sets are of measure zero. Later, we shall employ the Girsanov theorem in the models of quantitative finance[7], where it is going to be substantial that the two measures agree on what is possible and impossible in terms of probability.

[6]The word *martingale* in the title occurs due to the fact that $\tilde{W}$ is a martingale with respect to $\tilde{P}$. Showing this is part of the proof of the Girsanov theorem, where one uses Lévy characterization to prove the assertions.

[7]We mention that we shall use (2.34) in the "differential notation":

$$d\tilde{W}_t = dW_t + \zeta_t\, dt.$$

Chapter 3

Black-Scholes-Merton framework

In this chapter, we discuss what is probably the most famous model in quantitative finance, for it was at the very inception of the whole field. The model was first published by F. Black and M. Scholes in [2] with significant contributions from R. Merton [3]. Due to its simplicity the model is wildly popular although it has many known drawbacks originating in various simplifying assumptions that are part of the model. For this reason, nowadays, the model is mainly used first, as an introductory educational device, and second, as a quoting device.

3.1 Model assumptions

In this and in the next section we shall use the following notation and assumptions. We assume a market model with underlying equity product (usually a stock) and we are to value a derivative contract written on it, a plain vanilla European call option. We have the following model parameters: $\sigma > 0$, the *volatility* of the underlying asset, $\mu > 0$ the *rate of return* on the asset. We note, that one of the model assumptions is that these parameters are constant. The underlying asset (to which corresponds the natural Brownian filtration adapted stochastic process $\{S_t\}_{[0,T]}$) behavior is modeled by a unique strong solution of the following SDE:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t, \qquad \mu, \sigma > 0, \tag{3.1}$$

the geometric Brownian motion (per example 2.3.6):

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\}. \tag{3.2}$$

We assume that the underlying does not pay any dividends[1]. Here $\{W_t\}$ is the usual Brownian motion and S_0 is supposed to be adapted to the natural Brownian filtration $\{\mathcal{F}_t\}$; it is the initial stock-price observed in the market, unless the contract is forward starting (we shall not concern ourselves with this case here). Next assumption is that in the model, there is a non-risky asset, a bank account (a money-market account); for definition, see B.1.1. Here the interest rate r is assumed to be constant, so that

$$dB_t = rB_t\, dt \tag{3.3}$$

[1]This assumption may be adjusted for retrospectively. For our purposes it is not necessary to do so and hence we do not mention it again.

has a solution

$$B_t = B_0 e^{rt}, \qquad \forall t.$$

Lastly we assume, that the market is sufficiently liquid so that there are no transaction costs resulting from buying or selling assets.

3.2 Portfolio process, Black-Scholes-Merton PDE

For the purposes of valuing the European call option with maturity $T > 0$ we shall work exclusively with times $t \in [0, T]$. One holds certain amount $\{\Delta_t\}$ of shares of the underlying asset and $\{m_t\}$ in the bank account, thus $\{\Pi_t\}_{[0,T]}$, where

$$\Pi_t := \Delta_t S_t + m_t B_t, \qquad \forall t \tag{3.4}$$

represents the *portfolio* composed of the underlying and the bank account according to the *trading strategy* (Δ_t, m_t). Both $\{\Delta_t\}$ and $\{m_t\}$ are supposed to be adapted to the natural Brownian filtration and they are allowed to assume positive or negative values. A negative value of Δ_t represents shorting[2] the underlying asset while negative m_t means that one borrows money at the risk-free rate r.

Let us assume that the trading strategy (Δ_t, m_t) is *self-financing*, i.e. that the increments in the portfolio value Π_t result only from changes in the prices of the underlying asset S_t and the bank account B_t. The *self-financing condition* is expressed by the relationship between differentials:

$$d\Pi_t = \Delta_t dS_t + m_t dB_t. \tag{3.5}$$

Substituting for the behavior of S_t, B_t from (3.1), (3.3) and for $\{m_t\}$ from (3.4), the self-financing condition may be expressed as

$$\begin{aligned} d\Pi_t &= \Delta_t(\mu S_t\,dt + \sigma S_t\,dW_t) + r m_t B_t\,dt = \\ &= \Delta_t(\mu S_t\,dt + \sigma S_t\,dW_t) + r(\Pi_t - \Delta_t S_t)\,dt = \\ &= r\Pi_t\,dt + (\mu - r)\Delta_t S_t\,dt + \sigma\Delta_t S_t\,dW_t. \end{aligned} \tag{3.6}$$

One is obviously interested in what is the present (at time zero) value of the portfolio process at time $t \geq 0$. In order to calculate the Itô differential of the discounted[3] portfolio value $D(t)\Pi_t = e^{-rt}\Pi_t$, we first calculate the discounted value of the underlying asset. Employing the Itô product rule 2.2.17

$$\begin{aligned} d(D(t)S_t) &= d(e^{-rt}S_t) = \\ &= -re^{-rt}S_t\,dt + e^{-rt}\,dS_t = \\ &= (\mu - r)e^{-rt}S_t\,dt + e^{-rt}\sigma S_t\,dW_t. \end{aligned} \tag{3.7}$$

[2] *Shorting* or *selling short* an investment product means that one borrows the particular investment product from its owner and immediately sells it to raise cash. This obviously comes with the obligation to later return the asset back to the owner.

[3] For definition, see B.1.2 relationship (B.2) with constant interest rate r. For this case, the bank account and hence also the discount factor are deterministic functions.

Then the differential of the discounted portfolio value becomes:

$$\begin{aligned} d(D(t)\Pi_t) &= d(e^{-rt}\Pi_t) = \\ &= -re^{-rt}\Pi_t\,dt + e^{-rt}\,d\Pi_t = \\ &= -re^{-rt}\Pi_t\,dt + e^{-rt}[r\Pi_t dt + (\mu - r)\Delta_t S_t\,dt + \sigma\Delta_t S_t\,dW_t] = \\ &= e^{-rt}(\mu - r)\Delta_t S_t\,dt + e^{-rt}\sigma\Delta_t S_t\,dW_t = \\ &= \Delta_t d(D(t)S_t). \end{aligned} \tag{3.8}$$

Interpreting this, we observe that the change in the discounted portfolio value is solely due to change in the discounted price of the underlying.

Now we consider a certain derivative contract (most commonly a European call option) with payoff ($\max\{S_T - K, 0\}$ in the case of an option with strike K) at time T (the expiration of the contract). Black, Scholes and Merton [2, 3] argued that the value of this derivative $C = C(t, S_t)$ at any time $t \in [0, T]$ should depend on the time to expiration $T - t$ and on the value of the underlying (say, the stock price) at that time (in addition to dependence on the model parameters σ and possibly μ). Ergo, the function $C = C(t, x)$ is deterministic, however $C(t, S_t)$ is a random variable at every $t \in [0, T]$, hence it is a stochastic process. Furthermore, $C = C(t, x)$ is assumed to be continuous and differentiable in both variables.

It shall be now our goal to find this function $C = C(t, x)$. We begin by computing the differential, using the Itô formula for an Itô process:

$$dC(t, S_t) = C_t(t, S_t)\,dt + C_x(t, S_t)\,dS_t + \frac{1}{2}C_{xx}(t, S_t)\,dS_t dS_t =$$

$$= C_t(t, S_t)\,dt + C_x(t, S_t)(\mu S_t\,dt + \sigma S_t\,dW_t) + \frac{1}{2}C_{xx}(t, S_t)(\mu S_t\,dt + \sigma S_t\,dW_t)^2 =$$

$$= \left[C_t(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t, S_t) + \mu S_t C_x(t, S_t)\right]\,dt + \sigma S_t C_x(t, S_t)\,dW_t.$$

Next, using the Itô product rule and substitution from $dC(t, S_t)$:

$$d(D(t)C(t, S_t)) = d(e^{-rt}C(t, S_t)) = -re^{-rt}C(t, S_t)\,dt + e^{-rt}\,dC(t, S_t) =$$

$$= -re^{-rt}C(t, S_t)\,dt + e^{-rt}\left[C_t(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t, S_t) + \mu S_t C_x(t, S_t)\right]\,dt +$$

$$+ e^{-rt}\sigma S_t C_x(t, S_t)\,dW_t.$$

A replicating (hedging) portfolio starts with some initial capital Π_0 and invests in the underlying and money-market account so that the portfolio value Π_t at each time $t \in [0, T]$ agrees with $C(t, S_t)$. To ensure that

$$\Pi_t = C(t, S_t), \qquad \forall t \in [0, T], \tag{3.9}$$

we demand that

$$\Pi_0 = C(0, S_0) \qquad \wedge \qquad d(D(t)\Pi_t) = d(D(t)C(t, S_t)), \quad \forall t \in [0, T]. \tag{3.10}$$

Apparently, integrating (3.10) and employing the initial condition, the equality (3.9) is satisfied. Equating the above calculated differentials, we obtain:

$$e^{-rt}(\mu - r)\Delta_t S_t\, dt + e^{-rt}\sigma\Delta_t S_t\, dW_t =$$

$$= -re^{-rt}C(t, S_t)\, dt + e^{-rt}\left[C_t(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t, S_t) + \mu S_t C_x(t, S_t)\right] dt +$$

$$+e^{-rt}\sigma S_t C_x(t, S_t)\, dW_t.$$

For this equality to hold, the terms figuring in front of dW_t must be identical:

$$\Delta_t = C_x(t, S_t), \qquad \forall t \in [0, T]. \tag{3.11}$$

This quantity is called the *delta of the derivative contract (option)* written on the underlying. By the standard derivative interpretation it represents the change in the value $C(t, S_t)$ resulting from the change in the price of the underlying.

Remark 3.2.1. Before we proceed with the derivation of the BSM partial differential equation, we would like to comment on the relations contained in (3.10) and (3.11). Basically, we are able to find the strategy of combining the shares of the underlying asset and the money-market account so that we replicate the European call option as long as we *delta hedge*. This is, by definition, continuous; so to achieve a perfect hedge, one has to hold exactly $\Delta_t = C_x(t, S_t)$ shares of the underlying at every moment $t \in [0, T]$. As the reader can imagine, this requires rebalancing (buying or selling shares of the underlying) very often. In the section 3.1 we have made the assumption, that there are no transaction costs. In reality, transaction costs are an obstacle that prevents the continuous delta hedging as prescribed by the above formulae.

Now, we shall continue with the derivation; comparing the dt terms in the equality, that we demand, and using the formula for delta, we obtain

$$(\mu - r)C_x(t, S_t)S_t = -rC(t, S_t) + C_t(t, S_t) + \mu C_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t, S_t).$$

Canceling the $\mu S_t C_x(t, S_t)$ term, which appears on both sides of the equation, we arrive at:

$$rC(t, S_t) = C_t(t, S_t) + rS_t C_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t, S_t).$$

This is the *Black-Scholes-Merton partial differential equation* for the deterministic[4] function $C = C(t, x)$:

$$\frac{\partial C}{\partial t}(t, x) + rx\frac{\partial C}{\partial x}(t, x) - rC(t, x) = -\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C}{\partial x^2}(t, x), \qquad t \in [0, T], x \geq 0, \tag{3.12}$$

with the terminal condition

$$C(T, x) = \max\{x - K, 0\} = (x - K)^+.$$

[4]By delta hedging, we eliminate all risk, so that the portfolio perfectly replicates the option payoff and the option price obtained in the process is deterministic.

Remark 3.2.2. [18] The Black-Scholes-Merton equation (3.12) is (second-order, linear) parabolic partial differential equation. The terms $\partial_t C$ and $-\frac{1}{2}\sigma^2 x^2 \partial^2_{xx} C$ balanced on the two sides of the equation represent a smooth diffusion of the option price similarly to a heat diffusing in some medium as described by the heat equation. The only difference here is that the coefficient in front of the second-order derivative depends on the variable x, which basically means the inhomogeneity of the "medium" in which the diffusion occurs. The term $rx\partial_x C$ is a convection term which, loosely said, drifts the price in a preferred direction and the term $-rC$ accounts for decay.

We shall not solve the BSM equation here. A solution $C = C(t, x)$ will be found by measure-theoretic methods (change of measure, Girsanov theorem) in the next sections; we shall thus present a dual approach with more general applications.

3.3 Risk neutral pricing

Let $\{W_t\}_{t\in[0,T]}$ be a Brownian motion on a probability space $(\Omega, \mathcal{A}, P)$ (as hinted by Girsanov theorem, in this section it will be crucial with respect to what probability measure the particular process is a Brownian motion) and let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a natural filtration for this Brownian motion. Here T is a fixed endpoint of the interval. The underlying security (say, a stock) shall be modeled by a stochastic process $\{S_t\}$ determined by differential relation (i.e. it is an Itô process):

$$dS_t = \mu_t S_t\, dt + \sigma_t S_t\, dW_t, \qquad \forall t \in [0, T]. \tag{3.13}$$

Here μ_t and σ_t are processes adapted to the Brownian motion filtration discussed above. Furthermore let σ_t be non-zero almost surely for all $t \in [0, T]$. Per section 2.4, solution to stochastic differential relation (3.13) is generalized geometric Brownian motion process:

$$S_t = S_0 \exp\left\{\int_0^t \sigma_s\, dW_s + \int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2\right) ds\right\}. \tag{3.14}$$

Using the discounting defined in definition B.1.2 by relation (B.4), the discounted stock price process is[5]:

$$D_t S_t = S_0 \exp\left\{\int_0^t \sigma_s\, dW_s + \int_0^t \left(\mu_s - r_s - \frac{1}{2}\sigma_s^2\right) ds\right\}, \tag{3.15}$$

and using the product rule 2.2.17 and equation (B.5) it satisfies the stochastic differential equation

$$\begin{aligned} d(D_t S_t) &= S_t\, dD_t + D_t\, dS_t + dD_t\, dS_t = \\ &= -S_t r_t D_t\, dt + D_t(\mu_t S_t\, dt + \sigma_t S_t\, dW_t) = \\ &= (\mu_t - r_t) S_t D_t\, dt + S_t D_t \sigma_t\, dW_t. \end{aligned} \tag{3.16}$$

The stochastic process $\{\lambda_t\}_{[0,T]}$ defined by[6]

$$\lambda_t := \frac{\mu_t - r_t}{\sigma_t}, \qquad \forall t \in [0, T], \tag{3.17}$$

[5] In this section, we allow for stochastic short rate $\{r_t\}$ as well.

[6] At the beginning of this section, we have demanded that σ_t is non-zero P-almost surely for all $t \in [0, T]$, ergo this definition is correct in the usual $L^2(\Omega, \mathcal{A}, P)$ sense.

is called the *market price of risk*.

We introduce the probability measure defined in Girsanov's theorem 2.6.2, where in place of the generating process $\{\zeta_t\}$ we substitute the market price of risk process $\{\lambda_t\}$ defined above. This yields the stochastic process $\{Z_t\}$ (ergo the random variable Z_T) necessary for definition of $\tilde{P}$. In terms of Brownian motion $\{\tilde{W}_t\}$ we may rewrite the equality (3.15) as

$$d(D_t S_t) = \sigma_t D_t S_t \, d\tilde{W}_t. \tag{3.18}$$

Definition 3.3.1. A probability measure $\tilde{P}$ on a probability space $(\Omega, \mathcal{A}, P)$ is said to be *risk neutral*, if

1. $\tilde{P}$ and P are equivalent measures, i.e.

$$P(A) = 0 \quad \Longleftrightarrow \quad \tilde{P}(A) = 0, \qquad \forall A \in \mathcal{A},$$

2. under $\tilde{P}$, the discounted stock price $\{D_t S_t\}$ is a martingale with respect to the natural Brownian filtration $\{\tilde{\mathcal{F}}_t\}$.

Equation (3.18), rewritten in the "proper" integral form

$$D_t S_t = \underbrace{D_0}_{=1} S_0 + \int_0^t \sigma_s D_s S_s \, d\tilde{W}_s,$$

apparently consists a martingale for the process $D_t S_t$ is an Itô integral under the probability measure $\tilde{P}$ (and according to 2.2.4, every Itô integral is a martingale w.r.t. the natural Brownian filtration). The process $\{S_t\}$ (being generalized geometric Brownian motion) is not a martingale under P, nor is $\{D_t S_t\}$, however, the latter is a martingale under the risk neutral probability measure $\tilde{P}$.

Remark 3.3.2. With the use of the equation (which is the equality (2.34) rewritten for our particular case)

$$dW_t = -\frac{\mu_t - r_t}{\sigma_t} \, dt + d\tilde{W}_t, \tag{3.19}$$

one is able to calculate

$$\begin{aligned} dS_t = \mu_t S_t \, dt + \sigma_t S_t \, dW_t = \mu_t S_t \, dt - \sigma_t S_t \frac{\mu_t - r_t}{\sigma_t} \, dt + \sigma_t S_t \, d\tilde{W}_t = \\ = r_t S_t \, dt + \sigma_t S_t \, d\tilde{W}_t. \end{aligned} \tag{3.20}$$

This has a solution (again in the sense of GGBM)

$$S_t = S_0 \exp\left\{ \int_0^t \left(r_s - \frac{1}{2}\sigma_s^2 \right) ds + \int_0^t \sigma_s \, dW_s \right\}. \tag{3.21}$$

We conclude that under the risk neutral measure $\tilde{P}$ (which turns the discounted stock price into a martingale), the mean rate of return on the asset represented by $\{S_t\}$ is r_t.

Now we turn our attention to the problem of constructing a portfolio similar to that used in the previous section, only we relax the restrictive condition of r, μ and σ being constant and instead follow in the full generality of working with stochastic processes $\{r_t\}, \{\mu_t\}$ and $\{\sigma_t\}$. We again denote by $\Delta = \{\Delta_t\}$ the amount of shares of the underlying held in the portfolio $\{\Pi_t\}$. The rest of this portfolio $m_t = \Pi_t - \Delta_t S_t$ is again invested in the money-market account B_t, so that $\{\Delta_t, m_t\}$ constitutes a trading strategy, on which we impose the self-financing condition. Making use of these facts:

$$\begin{aligned} d\Pi_t &= \Delta_t\, dS_t + r_t(\Pi_t - \Delta_t S_t)\, dt = \Delta_t \mu_t S_t\, dt + \Delta_t \sigma_t S_t\, dW_t + r_t \Pi_t\, dt - r_t \Delta_t S_t\, dt = \\ &= \Delta_t(\mu_t - r_t) S_t\, dt + r_t \Pi_t\, dt + \Delta_t \sigma_t S_t\, dW_t = \Delta_t \sigma_t S_t(\lambda_t\, dt + dW_t) + r_t \Pi_t\, dt = \\ &= \Delta_t \sigma_t S_t\, d\tilde{W}_t + r_t \Pi_t\, dt. \end{aligned}$$

Discounting as usual, we have:

$$\begin{aligned} d(D_t \Pi_t) &= -r_t D_t \Pi_t\, dt + D_t\, d\Pi_t = \\ &= -r_t D_t \Pi_t\, dt + D_t \Delta_t \sigma_t S_t\, d\tilde{W}_t - r_t D_t \Pi_t\, dt = \\ &= D_t \Delta_t \sigma_t S_t\, d\tilde{W}_t. \end{aligned} \tag{3.22}$$

Now it is time to construct a hedging argument analogical to that of previous section. We desire to find a process $\{\Delta_t\}$ (ergo the corresponding portfolio process $\{\Pi_t\}$) and an initial capital Π_0, so that Π replicates a general (even path-dependent) payoff at maturity. We denote this general payoff (of a derivative contract) - an $\mathcal{F}_T$-measurable random variable - by V_T; it is the intrinsic value of the derivative at $t = T$, the expiration[7]. In this notation, the replicating condition is

$$\Pi_T = V_T \qquad P\text{-almost surely.} \tag{3.23}$$

To our aid, we have the following corollary of the martingale representation theorem 2.5.2 of section 2.5:

Corollary 3.3.3. [12] Let $\{W_t\}_{[0,T]}$ be a Brownian motion with respect to P and let $\{\mathcal{F}_t\}$ be a natural filtration for this Brownian motion. Let $\{\zeta_t\}$ be a process adapted to $\{\mathcal{F}_t\}$; we define

$$Z_t := \exp\left\{-\frac{1}{2}\int_0^t \zeta_s^2\, ds - \int_0^t \zeta_s\, dW_s\right\}$$

and

$$\tilde{W}_t := W_t + \int_0^t \zeta_s\, ds. \tag{3.24}$$

We also assume that $\tilde{E}\int_0^T \zeta_s^2 Z_s^2\, ds < \infty$. If we denote $Z := Z_T$, then $EZ = 1$ and the process $\tilde{W}$ given by (3.24) is a $\tilde{P}$-Brownian motion. Let also $\{\tilde{M}_t\}$ be an $\{\mathcal{F}_t\}$-martingale under $\tilde{P}$. Then there is an $\{\mathcal{F}_t\}$-adapted process $\{\tilde{G}_t\}$, such that

$$\tilde{M}_t = \tilde{M}_0 + \int_0^t \tilde{G}_s\, d\tilde{W}_s, \qquad \forall t \in [0, T]. \tag{3.25}$$

[7] In case of a European call option, this is $V_T = \max\{S_T - K, 0\}$ as usual.

The discounted portfolio process $\{D_t\Pi_t\}$, being an Itô integral is a $\tilde{P}$-martingale with respect to $\{\mathcal{F}_t\}$:

$$D_t\Pi_t = \tilde{E}[D_T\Pi_T|\mathcal{F}_t] = \tilde{E}[D_T V_T|\mathcal{F}_t], \qquad \forall t \in [0,T]. \tag{3.26}$$

We recall, that we still have to find $\{\Delta_t\}$ (resp. $\{\Pi_t\}$) so that $\Pi_T = V_T$ holds. Since Π_t is the value of the replicating portfolio at t, which represents the capital needed to successfully hedge, we call it the price of the derivative contract at time t and denote it V_t. Now, the corollary 3.3.3 applies, so that there is an $\{\mathcal{F}_t\}$-adapted process $\{\tilde{G}_t\}$ such that

$$D_t\Pi_t = D_tV_t = \underbrace{D_0}_{=1} V_0 + \int_0^t \tilde{G}_s\, d\tilde{W}_s, \qquad \forall t \in [0,T]. \tag{3.27}$$

Using relationship (3.22), the discounted contract price (or the discounted portfolio process) $\{D_tV_t\}$ also satisfies

$$D_tV_t = D_t\Pi_t = \underbrace{D_0}_{=1} \Pi_0 + \int_0^t \Delta_s\sigma_s D_s S_s\, d\tilde{W}_s, \qquad \forall t \in [0,T]. \tag{3.28}$$

In order for the equality (3.23) to hold, equating (3.27) and (3.28), we see, that it is enough to set

$$\Pi_0 = V_0$$

and to define the process[8]

$$\Delta_t = \frac{\tilde{G}_t}{\sigma_t D_t S_t}, \qquad \forall t \in [0,T]. \tag{3.29}$$

These two conditions together ensure, that $\Pi_t = V_t$ for all t up to and (especially) including time T. Thus, we have successfully found a replicating portfolio from which we can determine the price of the derivative contract at time t as

$$V_t = D_t^{-1}\tilde{E}[D_T V_T|\mathcal{F}_t] = \tilde{E}[e^{\int_t^T r_s\, ds}\, V_T|\mathcal{F}_t], \qquad \forall t \in [0,T]. \tag{3.30}$$

This is the cornerstone of quantitative finance, the *risk neutral pricing formula*.

3.4 Solution of the BSM equation via risk neutral pricing

To obtain the Black-Scholes-Merton price of a European call, we assume a constant volatility $\sigma > 0$, constant interest rate r and take the derivative contract payoff at expiration T, to be the option payoff

$$V_T = \max\{S_T - K, 0\} =: (S_T - K)^+.$$

[8]We remind ourselves that at the beginning of this section, we have postulated $\{\sigma_t\}$ to be a P-a.s. non-zero (positive) process.

The right hand side of (3.30) becomes

$$\tilde{E}[e^{-r(T-t)}(S_T - K)^+|\mathcal{F}_t].$$

Because, as noted in chapter 1, geometric Brownian motion is a Markov process, this expression depends on the stock price S_t and on the time t at which the conditional expectation is computed but not on the stock price prior to time t. In other words, there is a function $C = C(t, x)$ such that

$$C(t, S_t) = \tilde{E}[e^{-r(T-t)}(S_T - K)^+|\mathcal{F}_t]. \tag{3.31}$$

With σ, r constant, the S_t is described by the regular geometric Brownian motion. Which in the risk neutral setting takes form of:

$$S_t = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma\tilde{W}_t\right\}. \tag{3.32}$$

Using this formula for $t = T$ and then substituting in from the same formula in order to get rid of the dependence on S_0 we may write:

$$S_T = S_t \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\frac{\tilde{W}_T - \tilde{W}_t}{\sqrt{T-t}}\sqrt{T-t}\right\}.$$

Here $T - t = \tau$ is *time to maturity* and

$$Y = -\frac{\tilde{W}_T - \tilde{W}_t}{\sqrt{T-t}} \sim \mathrm{N}(0, 1) \tag{3.33}$$

is the standard normal random variable. Using the fact that S_t is $\mathcal{F}_t$-measurable (see section 3.1) and the random variable $\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\tau - \sigma Y\sqrt{\tau}\right\}$ is independent of $\mathcal{F}_t$ we can use the corresponding properties of conditional expectation 1.2.4 and arrive at

$$\begin{aligned} C(t,x) &= e^{-r\tau}\tilde{E}\left[\left(x\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\tau - \sigma\sqrt{\tau}Y\right\} - K\right)^+\right] = \\ &= e^{-r\tau}\tilde{E}[f(Y)] = e^{-r\tau}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(y)e^{-\frac{y^2}{2}}dy = \\ &= e^{-r\tau}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(x\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\tau - \sigma\sqrt{\tau}y\right\} - K\right)^+ e^{-\frac{y^2}{2}}dy. \end{aligned} \tag{3.34}$$

Before continuing in (3.34), we get rid of the maximum, here

$$x\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\tau - \sigma\sqrt{\tau}y\right\} - K > 0$$

is equivalent to

$$y < d_-(\tau, x) := \frac{1}{\sigma\sqrt{\tau}}\left[\ln\frac{x}{K} + \left(r - \frac{1}{2}\sigma^2\right)\tau\right].$$

Therefore, now continuing in (3.34), we have:

$$C(t,x) = e^{-r\tau}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}\left(x\exp\left\{\left(r-\frac{1}{2}\sigma^2\right)\tau-\sigma\sqrt{\tau}y\right\}-K\right)e^{-\frac{y^2}{2}}dy =$$

$$= \frac{1}{\sqrt{2\pi}}x\int_{-\infty}^{d_-(\tau,x)}e^{-r\tau}\exp\left\{\left(r-\frac{1}{2}\sigma^2\right)\tau-\sigma\sqrt{\tau}y\right\}e^{-\frac{y^2}{2}}dy - e^{-r\tau}K\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}e^{-\frac{y^2}{2}}dy =$$

$$= \frac{1}{\sqrt{2\pi}}x\int_{-\infty}^{d_-(\tau,x)}\exp\left\{-\frac{1}{2}\sigma^2\tau-\sigma\sqrt{\tau}y\right\}e^{-\frac{y^2}{2}}dy - e^{-r\tau}K\Phi(d_-(\tau,x)) =$$

$$= \frac{1}{\sqrt{2\pi}}x\int_{-\infty}^{d_-(\tau,x)}\exp\left\{-\frac{1}{2}(y+\sigma\sqrt{\tau})^2\right\}dy - e^{-r\tau}K\Phi(d_-(\tau,x)).$$

Now changing the variable $\tilde{y} = y + \sigma\sqrt{\tau}$ in the integral and denoting

$$d_+(\tau,x) := d_-(\tau,x) + \sigma\sqrt{\tau},$$

we arrive at

$$C(t,S_t) = S_t\Phi(d_+(\tau,S_t)) - e^{-r\tau}K\Phi(d_-(\tau,S_t)), \tag{3.35}$$

where $\Phi = \Phi(z)$ denotes the cumulative distribution function of standard normal random variable at a point z and

$$d_\pm(\tau,x) := \frac{1}{\sigma\sqrt{\tau}}\left[\ln\frac{x}{K} + \left(r \pm \frac{1}{2}\sigma^2\right)\tau\right]. \tag{3.36}$$

Chapter 4

Volatility

In the previous, we have encountered volatility as a (by assumption) constant parameter in the Black-Scholes-Merton model. In this chapter, we present an overview of the additional features that the concept of volatility encompasses. First, we distinguish between the historical/realized volatility of the asset returns. Second, we classify BSM and other, more general models by their approach to modeling volatility. In the process, it shall become clear why we are making these distinctions.

4.1 Realized volatility

The annualized standard deviation of *log-returns* $\{h_1, \ldots, h_n\}$ (differences in asset price on which the logarithm function has been applied) defined as:

$$\sigma_R := \sqrt{\frac{a}{n-1}\sum_{\ell=1}^{n}(h_\ell - \overline{h})^2}, \tag{4.1}$$

is called the *realized* (or *historical*) *volatility*. Here a is the annualization factor usually equal to the number of business days in a year and $\overline{h}$ is the arithmetic average of all log-returns. If the log-returns $\{h_\ell\}_{\ell=1}^n$ are independently and identically distributed with $\mathrm{N}(\mu, \sigma^2)$, then σ_R^2 is an unbiased estimator for the variance σ^2. As remarked by Wilmott [19], there are two timescales associated with calculating volatility as (4.1). First, the timespan during which the asset price is monitored and second, the time elapsed between two subsequent observations. By the assumption of the BSM model, there is no timescale associated with the volatility parameter σ_{BSM}. The realized volatility is necessarily backward looking statistical measure and thus using the realized volatility in BSM option pricing model, one makes an implicit assumption that the past behavior will continue into the future. This is the assumption that the practitioners of quantitative finance are usually trying to avoid. Moreover, the historical volatility often fluctuates.

4.2 Implied volatility, the smile

One of the assumptions of the BSM model is that the volatility parameter σ_{BSM} is a positive constant, which is then plugged into the geometric Brownian motion. Should

one be able to estimate this parameter from the market data, then any vanilla call option with an arbitrary strike K (and the corresponding moneyness relative to current spot price S_t) and an arbitrary time to maturity $\tau = T - t$ should be correctly priced by the pair of equations (3.35), (3.36). However this is not the case. The prices in (generally any) market are not consistent with the BSM price $C(t, S_t)$.

It is thus natural to ask what is the value of $\sigma_{\text{imp}} > 0$ which reconciles the BSM prices with the prices C_{obs} actually observed in the markets. This value is the so-called *implied volatility*, or the volatility implied by the Black-Scholes-Merton model and it poses the inverse problem; equating $C^{\sigma}_{\text{BSM}}(t, S_t) = C_{\text{obs}}$, what is the value of σ_{imp}? The dependence of option value C^{σ}_{BSM} on the parameter is monotone, so that the problem has a unique solution.

Asking this question since the 1980s, the traders and the quants started to notice that if they express the implied volatility for options with various strikes (but written on the same underlying and with the same time to maturity), they obtain various sigmas, so that actually $\sigma_{\text{imp}} = \sigma_{\text{imp}}(K)$. Plotting this dependence of implied volatility on the strike, one obtains picture similar to:

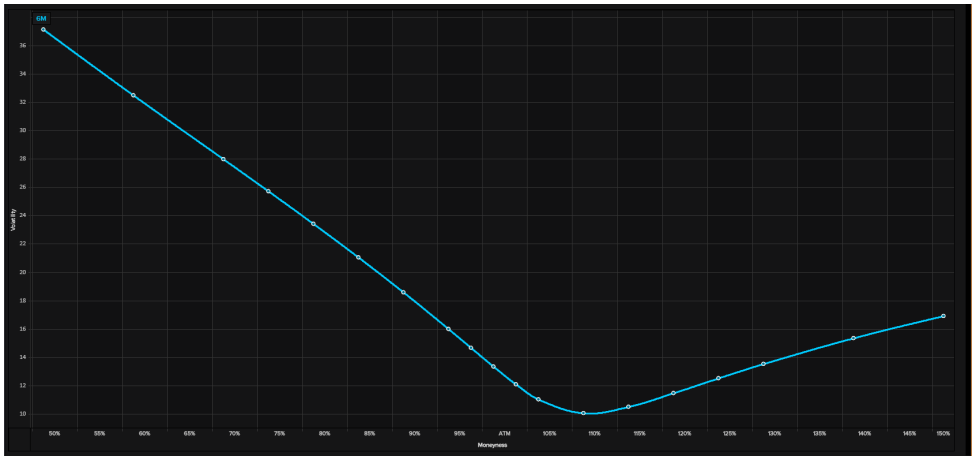


Figure 4.1: This is the implied volatility smile on the S&P500 index as of 25.4.2019. The abscissa is actually in terms of the strike relative to the current spot price, the *moneyness*. We display here a screenshot from Eikon application, courtesy of Reuters.

Figures of this type became commonly known as the *implied volatility smile*, exact shape of which usually depends on the type of the market as well as the current economic cycle (sometimes they are called *smirk* or *frown* instead).

The situation is even more complicated because it turns out that σ_{imp} exhibits whole term structure; the implied volatility varies with respect to the time to expiration τ as well.

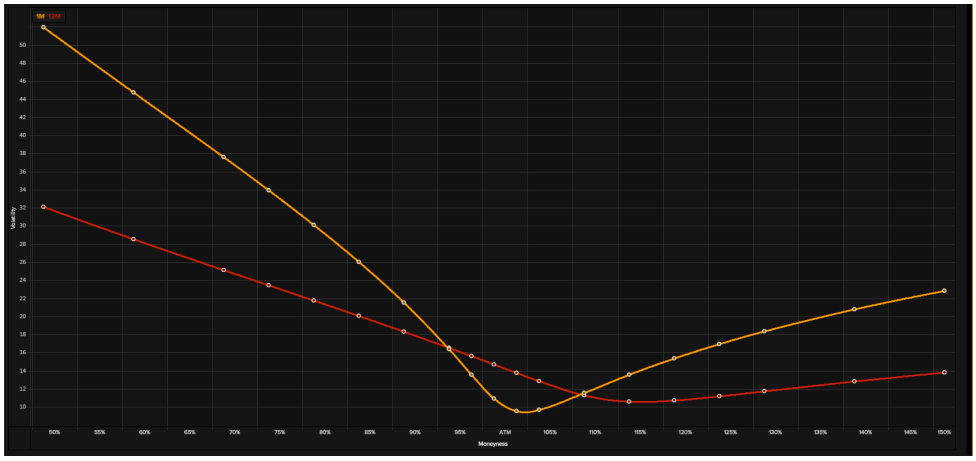


Figure 4.2: These are the implied volatility smiles on the S&P500 index as of 25.4.2019 for option expirations of one month (orange) and one year (red). We display here a screenshot from Eikon application, courtesy of Reuters.

Finally, the parametric dependence $\sigma_{\text{imp}} = \sigma_{\text{imp}}(\tau, K)$ is known as the *volatility surface*, for it is usually graphically represented as:

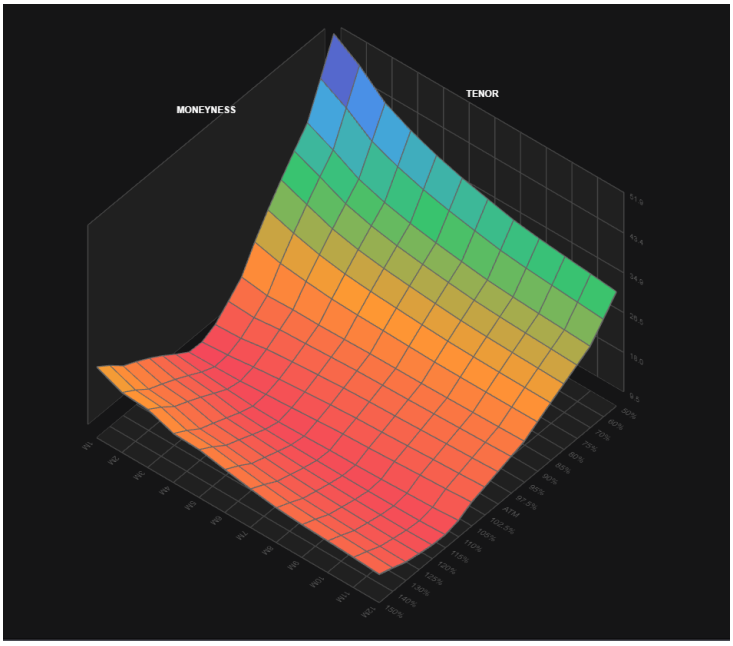


Figure 4.3: When one considers the smile and extends it for the term structure, one obtains a three-dimensional surface - the volatility surface. This surface is again on the S&P500 index as of 25.4.2019. The axes are inverted, so that the whole surface is clearly visible. We display here a screenshot from Eikon application, courtesy of Reuters.

4.3 Local volatility

For a pricing model it is essential to fit the market observable prices at a given moment (and thus the volatility surface). This poses a problem, for, as explained above, this is not one of the features of the Black-Scholes-Merton model. The most straightforward solution to this problem is to match the observed volatility surface by assuming that the volatility is a deterministic function $\sigma = \sigma(t, x)$ figuring in the stochastic differential equation:

$$dS_t = \mu S_t \, dt + \sigma(t, S_t) S_t \, dW_t. \tag{4.2}$$

This model was introduced in [20, 21] and is called the *local volatility model* (the function $\sigma = \sigma(t, x)$ is called the *local volatility function*) or shortly the *LV model*. Let us denote $C_{\text{obs}} = C_{\text{obs}}(T, K)$ the observed market price of a call option with strike K and maturity T. It was shown by Dupire [20] that for any such set of observed prices at one moment there is a unique local volatility function given by:

$$\sigma^2(T, K) = 2 \frac{\frac{\partial C_{\text{obs}}}{\partial T}(T, K) + rK \frac{\partial C_{\text{obs}}}{\partial K}(T, K)}{K^2 \frac{\partial^2 C_{\text{obs}}}{\partial K^2}(T, K)}. \tag{4.3}$$

This is the *Dupire formula*.

Related to the local volatility model is the procedure of *calibration*. This means choosing the model parameters so that the theoretical prices provided as the model output for exchange traded contracts (usually vanilla instruments) match exactly, or as closely as possible the market observable prices at an instant in time. Although calibration is a procedure that is employed even in the more advanced models described below, with local volatility models one needs to re-calibrate (that is, to again find a new function fitting the market) often. The market practitioners are fully aware that the local volatility models are not describing the actual volatility dynamics, the same way they are aware of the fact that the Black-Scholes-Merton is practically a toy model. In some cases, simplicity is more important than faithful representation of the reality.

4.4 Stochastic volatility

In the previous section, we have assumed a deterministic function σ figuring in the equation modeling the behavior of underlying asset. Next step in this direction is to realize that the random changes in volatility might be genuine and thus model the volatility as a stochastic process $\sigma = \{\sigma_t\}$. The differential relation for an underlying asset becomes

$$dS_t = \mu_t S_t \, dt + \sigma_t S_t \, dW_t;$$

this is an Itô process. One then has to specify the behavior of the volatility process $\{\sigma_t\}$ in some way. The usual approach is to prescribe a stochastic differential equation to it. That is the route that we follow in the last section of this chapter and the subsequent chapter on the Heston model. There are other ways one can achieve or define stochastic volatility

(SV), some of which we shall introduce in the last chapter. Some models commonly used by practitioners combine local and stochastic volatility.

Generally speaking, the SV models can be subdivided into three following categories - continuous diffusion based, jump diffusion and pure jump. An example of *continuous diffusion based model* is the one introduced by S. Heston [4]. Since these types of models are well-studied there are other popular models from this category, which is usually good at explaining the volatility smiles for longer maturities. Conversely, for shorter maturities, (loosely speaking) the continuous diffusion cannot keep up with the real price dynamics and large moves, so that these models' description of the situation is unrealistic without setting the parameters to some large values. Attempting to correct this pathology, price discontinuities are introduced in the so-called *jump diffusion models*. Jumps in these models are useful because they remedy the issue of short-term smile and diffuse away in the long run, so that the advantages of SV model are retained. The most realistic models, however, are some of the *pure jump models*, which have no diffusion component, yet they exhibit infinite activity, most of which is encompassed in small jumps with occasional large moves. We shall be focusing on pure jump models in the last two chapters of this text. Let us sum up the taxonomy of models in the following scheme (subjectively, these are ordered from simplest to the most sophisticated ones) with examples:

1. Constant volatility parameter (BSM model)
2. Local volatility function
3. Stochastic volatility
    - Diffusion based (Heston model)
    - Jump diffusion
    - Pure jump (Variance gamma)

We note that, just as the model with constant volatility parameter is a trivial local volatility model, LV model is a very constrained continuous diffusion based stochastic volatility model with the stochastic process $\{\sigma_t\}$ being actually a deterministic function of the process $\{S_t\}$ and time.

4.5 General stochastic volatility model PDE

In this section we follow more or less closely [16] which in turn finds itself inspired by [22]. Both books present a wonderful introduction into stochastic volatility models and as the title of one of them suggests, they are also often used in practice.

Here, inspired by the hedging argument in the above-mentioned citations, we aim to derive an analogical PDE, but starting from a much more general process incorporating the stochastic volatility term. This term by itself is governed by a stochastic differential equation. We have the following model[1]:

$$\begin{aligned} dS_t &= \mu_t S_t\,dt + \sqrt{v_t} S_t\,dW_t^{(1)} \\ dv_t &= \alpha(t,S_t,v_t)\,dt + \eta\beta(t,S_t,v_t)\sqrt{v_t}\,dW_t^{(2)}. \end{aligned} \tag{4.4}$$

[1]The volatility is not modeled directly in this model but rather through the variance process $\{v_t\}$.

We also assume the following correlation between the two sources of randomness:

$$dW_t^{(1)} dW_t^{(2)} = \varrho dt. \tag{4.5}$$

Remark 4.5.1. This is a very general stochastic volatility model, because $\{S_t\}$ is essentially an Itô process and the SDE for $\sigma_t = \sqrt{v_t}$ is equation with general coefficient functions α, β.

We would like to set up a hedging portfolio quite similar to the case when we derived the Black-Scholes PDE. In this model, however, additional security U, whose price depends on the volatility $U = U(t, S_t, v_t)$, is necessary to hedge the volatility (in case of an option, this is usually another option). Hence we form a portfolio consisting of the contract V that we want to value, Δ_t shares of the underlying S_t (say, the stock) and ξ_t shares of the contract U. The Itô differential for the portfolio process (assuming it is self-financing) is:

$$d\Pi_t = dV(t, S_t, v_t) + \Delta_t \, dS_t + \xi_t \, dU(t, S_t, v_t). \tag{4.6}$$

From now on, we shall be omitting the arguments of U, V, α, β whenever they are not necessary. Using the relations

$$\begin{aligned} dS_t dS_t &= v_t S_t^2 \, dt, \\ dS_t dv_t &= \eta v_t \beta \varrho \, dt, \\ dv_t dv_t &= \eta^2 \beta^2 v_t \, dt, \end{aligned} \tag{4.7}$$

which are easily computed from the model specifications and the two-dimensional Itô formula, we have[2]:

$$dV = \partial_t V \, dt + \partial_x V \, dS_t + \partial_v V \, dv_t + \frac{1}{2}(\partial_{xx}^2 V) v_t S_t^2 \, dt + (\partial_{xv}^2 V) \eta v_t \beta \varrho \, dt + \frac{1}{2}(\partial_{vv}^2 V) \eta^2 \beta^2 v_t \, dt.$$

The derivative U satisfies the same SDE as V, but in terms of U. Now the change in the portfolio value process Π_t in (4.6) can be rewritten as:

$$\begin{aligned} d\Pi_t = \Bigg[&\partial_t V + \xi_t \partial_t U + \frac{1}{2} v_t S_t^2 (\partial_{xx}^2 V + \xi_t \partial_{xx}^2 U) + \frac{1}{2} \eta^2 \beta^2 v_t (\partial_{vv}^2 V + \xi_t \partial_{vv}^2 U) + \\ &+ \eta \beta \varrho v_t (\partial_{xv}^2 V + \xi_t \partial_{xv}^2 U) + \mu_t S_t (\partial_x V + \xi_t \partial_x U) + \alpha (\partial_v V + \xi_t \partial_v U) + \Delta_t \mu_t S_t \Bigg] dt + \\ &+ \sqrt{v_t} S_t \Big[(\partial_x V + \xi_t \partial_x U) + \Delta_t \Big] \, dW_t^{(1)} + \eta \beta \sqrt{v_t} (\partial_v V + \xi_t \partial_v U) \, dW_t^{(2)}. \end{aligned} \tag{4.8}$$

In order for the portfolio Π_t to be instantaneously riskless (i.e. to be hedged against the randomness in both the stock and the volatility represented by processes $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$), the coefficients in front of the differentials of the two Brownian motions must vanish. This condition yields the expressions for ξ_t, Δ_t:

$$\xi_t := -\frac{\partial_v V}{\partial_v U} \qquad \text{and} \qquad \Delta_t := -\xi_t \partial_x U - \partial_x V. \tag{4.9}$$

[2]Here ∂_x resp. ∂_v denotes partial derivative with respect to second respectively third argument of functions U, V.

Substituting these back to the expression for $d\Pi_t$, we arrive at

$$\begin{aligned} d\Pi_t = &\left\{ \partial_t V - \frac{(\partial_v V)(\partial_t U)}{\partial_v U} + \frac{1}{2} v_t S_t^2 \left[\partial_{xx}^2 V - \frac{(\partial_v V)(\partial_{xx}^2 U)}{\partial_v U} \right] + \right. \\ &\left. + \frac{1}{2} \eta^2 \beta^2 v_t \left[\partial_{vv}^2 V - \frac{(\partial_v V)(\partial_{vv}^2 U)}{\partial_v U} \right] + \eta \beta \varrho v_t \left[\partial_{xv}^2 V - \frac{(\partial_v V)(\partial_{xv}^2 U)}{\partial_v U} \right] \, dt \right\}. \end{aligned}$$

Now this portfolio Π is instantaneously riskless and we might expect the risk-free rate of return on it. Thus equating $d\Pi_t = r\Pi_t dt$ with the above derived expression and substituting in from (4.9), one obtains

$$\begin{aligned} &\frac{1}{\partial_v V} \left[-rV + rS_t \partial_x V + \partial_t V + \frac{1}{2} v_t S_t^2 \partial_{xx}^2 V + \frac{1}{2} \eta^2 \beta^2 v_t \partial_{vv}^2 V + \eta \beta \varrho v_t \partial_{xv}^2 V \right] = \\ =&\frac{1}{\partial_v U} \left[-rU + rS_t \partial_x U + \partial_t U + \frac{1}{2} v_t S_t^2 \partial_{xx}^2 U + \frac{1}{2} \eta^2 \beta^2 v_t \partial_{vv}^2 U + \eta \beta \varrho v_t \partial_{xv}^2 U \right] \end{aligned} \tag{4.10}$$

We have separated the terms including function V on one side of the equation and the terms including V on the other.

Chapter 5

Heston model

In this chapter, we shall continue with the derivations of the previous one. Because the Heston model is a special case of a general stochastic volatility model presented there, we shall utilize the partial differential equation (4.10).

5.1 The model, risk neutral SDEs

The Heston model is specified as an Itô process (in the differential notation) for the underlying and a stochastic differential equation for the variance:

$$\begin{aligned} dS_t &= \mu_t S_t\,dt + \sqrt{v_t} S_t\,dW_t^{(1)} \\ dv_t &= \kappa(\nu - v_t)\,dt + \eta\sqrt{v_t}\,dW_t^{(2)}, \end{aligned} \tag{5.1}$$

and also the correlation condition

$$dW_t^{(1)} dW_t^{(2)} = \varrho\,dt. \tag{5.2}$$

Here κ is the mean reversion speed of the variance, ν is the mean reversion level for the variance, η is the volatility of variance and $\varrho \in [-1, 1]$ is the correlation between the two Brownian motions. Obviously, the Heston model defined above is a special case of the general stochastic volatility model analyzed in section 4.5 of the previous chapter with coefficient functions given as[1]

$$\begin{aligned} \alpha(t, S_t, v_t) =& \kappa(\nu - v_t) \\ \beta(t, S_t, v_t) =& 1, \end{aligned} \tag{5.3}$$

for all $t \geq 0$.

Because $\sqrt{v_t} = \sigma_t$ the risk neutral process for the stock price mimics that of the generalized geometric Brownian motion described in section 3.3:

$$dS_t = rS_t\,dt + \sqrt{v_t} S_t\,d\tilde{W}_t^{(1)}, \qquad \forall t \in [0, T], \tag{5.4}$$

where

$$\tilde{W}_t^{(1)} = W_t^{(1)} + \frac{\mu_t - r}{\sqrt{v_t}} t, \qquad \forall t \in [0, T]. \tag{5.5}$$

[1]The SDE for the variance process is exactly the CIR SDE of example 2.3.9, so that the variance is modeled as a positive, mean reverting process.

The risk neutral process for the variance is obtained by introducing a function $\lambda = \lambda(t, S_t, v_t)$ into the drift of the CIR process as follows:

$$dv_t = [\kappa(\nu - v_t) - \lambda(t, S_t, v_t)]\, dt + \eta\sqrt{v_t}\, d\tilde{W}_t^{(2)}, \qquad \forall t \in [0, T], \tag{5.6}$$

where

$$\tilde{W}_t^{(2)} = W_t^{(2)} + \frac{\lambda(t, S_t, v_t)}{\eta\sqrt{v_t}} t, \qquad \forall t \in [0, T]. \tag{5.7}$$

The function λ is called the *market price of volatility risk*. As pointed out by [23, 4], economic considerations suggest that the function is of the form $\lambda(t, S_t, v_t) = \lambda v_t$, where λ is constant. Substituting this into the risk neutral equation above, we obtain

$$dv_t = [\kappa(\nu - v_t) - \lambda v_t]\, dt + \eta\sqrt{v_t}\, d\tilde{W}_t^{(2)} = \ = \tilde{\kappa}(\tilde{\nu} - v_t)\, dt + \eta\sqrt{v_t}\, d\tilde{W}_t^{(2)}, \tag{5.8}$$

where

$$\tilde{\kappa} = \kappa + \lambda \qquad \text{and} \qquad \tilde{\nu} = \frac{\nu}{\kappa + \lambda}. \tag{5.9}$$

The pair of risk neutral processes describing the Heston model becomes:

$$\begin{aligned} dS_t &= rS_t\, dt + \sqrt{v_t} S_t\, d\tilde{W}_t^{(1)} \\ dv_t &= \tilde{\kappa}(\tilde{\nu} - v_t)\, dt + \eta\sqrt{v_t}\, d\tilde{W}_t^{(2)}. \end{aligned} \tag{5.10}$$

From (5.5) and (5.7) the correlation under the risk neutral measure $\tilde{P}$ is

$$d\tilde{W}_t^{(1)} d\tilde{W}_t^{(2)} = \varrho\, dt. \tag{5.11}$$

Remark 5.1.1. From equalities (5.9), it is obvious, that the original and risk neutral CIR equations coincide when $\lambda = 0$ for all $t \geq 0$. From now on we shall identify the risk neutral and the original model parameters, so that we obtain the same equation for the variance (alternatively, we could set the parameter lambda to zero[16]).

5.2 Heston PDE

Reiterating to the preceding section, the Heston model defined by a pair of SDEs (5.1) is a special case of the general stochastic volatility model (4.4) with the parametric functions α, β given by (5.3). We may thus express the equality (4.10) derived in section 4.5 with these parameters

$$\begin{aligned} &\frac{1}{\partial_v V}\left[-rV + rS_t\partial_x V + \partial_t V + \frac{1}{2}v_t S_t^2 \partial_{xx}^2 V + \frac{1}{2}\eta^2 v_t \partial_{vv}^2 V + \eta\varrho v_t \partial_{xv}^2 V\right] = \\ =&\frac{1}{\partial_v U}\left[-rU + rS_t\partial_x U + \partial_t U + \frac{1}{2}v_t S_t^2 \partial_{xx}^2 U + \frac{1}{2}\eta^2 v_t \partial_{vv}^2 U + \eta\varrho v_t \partial_{xv}^2 U\right]. \end{aligned} \tag{5.12}$$

Apparently both sides of this equation exhibit the same functional relationship, only one side for $V = V(t, S_t, v_t)$ and the other for $U = U(t, S_t, v_t)$ respectively. For the equality to hold, both sides must be equal to the same function $f = f(t, S_t, v_t)$. Heston in [4] specifies this function as:

$$f(t, S_t, v_t) := -\kappa(\nu - v_t) + \lambda(t, S_t, v_t). \tag{5.13}$$

In the previous section we have mentioned that $\lambda(t, S_t, v_t) = \lambda v_t$. Then, making use of these assumptions, we arrive at the *Heston partial differential equation* for $U = U(t, x, v)$ (V satisfies the same PDE):

$$\frac{\partial U}{\partial t} + rx\frac{\partial U}{\partial x} + \frac{1}{2}vx^2\frac{\partial^2 U}{\partial x^2} + \frac{1}{2}\eta^2 v\frac{\partial^2 U}{\partial v^2} + \eta\varrho vx\frac{\partial^2 U}{\partial x \partial v} + [\kappa(\nu - v) - \lambda v]\frac{\partial U}{\partial v} - rU = 0. \quad (5.14)$$

Remark 5.2.1. The terms

$$\frac{\partial U}{\partial t} + rx\frac{\partial U}{\partial x} + \frac{1}{2}vx^2\frac{\partial^2 U}{\partial x^2} - rU$$

are similar to terms present in the Black-Scholes-Merton partial differential equation.

5.3 Heston price of an option

The price of a European call on a non-dividend paying stock with spot price S_t (at time t), when the strike is K and the time to maturity is $\tau = T - t$ is the risk neutral expected value of the discounted option payoff just as in section 3.3 (the interest rate r is assumed to be constant):

$$\begin{aligned} C(t, S_t) &= e^{-r\tau} E^{\tilde{P}}[(S_T - K)|\mathcal{F}_t] = e^{-r\tau} E^{\tilde{P}}[(S_T - K)\,\mathbf{1}_{S_T > K}] = \\ &= e^{-r\tau} E^{\tilde{P}}[S_T\,\mathbf{1}_{S_T > K}] - e^{-r\tau} K \underbrace{E^{\tilde{P}}[\mathbf{1}_{S_T > K}]}_{\tilde{P}(S_T > K)}. \end{aligned} \quad (5.15)$$

Here we have followed basically the same procedure as in section 3.4, only without explicitly using the terminal distribution of the stock prices and in somewhat compressed notation. We shall evaluate the first term separately. The procedure of changing the numéraire, as explained in section A.6.1 of appendix A, shall be made use of, in order to simplify the calculations. The assumptions of theorem A.6.1 are satisfied for (non-dividend paying) $\{S_t\}$ as the numéraire, ergo we have the Radon-Nikodým derivative, that defines a risk neutral measure Q that uses the stock price to denominate contingent claims (in our case the European call):

$$\frac{dQ}{d\tilde{P}} = \frac{S_T}{B_T}\frac{B_t}{S_t}. \quad (5.16)$$

Employing this change of measure, we can write the first term in (5.15) as

$$\begin{aligned} e^{-r\tau} E^{\tilde{P}}[S_T\,\mathbf{1}_{S_T > K}] &= e^{-rT} e^{rt}\frac{S_t}{S_t} E^{\tilde{P}}[S_T \mathbf{1}_{S_T > K}] = S_t E^{\tilde{P}}\left[\frac{S_T}{B_T}\frac{B_t}{S_t}\,\mathbf{1}_{S_T > K}\right] = \\ &= S_t E^{\tilde{P}}\left[\mathbf{1}_{S_T > K}\,\frac{dQ}{d\tilde{P}}\right] = S_t E^{Q}[\mathbf{1}_{S_T > K}] = S_t Q(S_T > K). \end{aligned} \quad (5.17)$$

Altogether, we have the price of the European call option in Heston model:

$$C(t, S_t) = S_t Q(S_T > K) - e^{-r\tau}\tilde{P}(S_T > K). \quad (5.18)$$

Obviously, when S_T follows the log-normal distribution of Black-Scholes-Merton model, the prices given by the two models coincide, for $Q(S_T > K) = \Phi(d_+)$ and $\tilde{P}(S_T > K) = \Phi(d_-)$, similarly to equations (3.35), (3.36).

5.4 Heston price of a variance swap

Models of stochastic volatility are used not only to better fit the volatility surface (usually calibrated on vanilla instruments) but also to price volatility sensitive derivatives, or derivatives that have volatility of an asset as the underlying. One example that we mention is the variance swap defined in appendix A. This is a derivative contract that pays $V_T = N(\sigma_R^2 - \sigma_K^2)$ at maturity, where in the continuous framework we shall express σ_R as (A.5). We shall work within the Heston model, where the risk neutral variance satisfies the CIR stochastic differential equation:

$$dv_t = \kappa(\nu - v_t)\,dt + \eta\sqrt{v_t}\,dW_t^{(2)}$$

or in terms of the volatility process[2]:

$$d\sigma_t^2 = \kappa(\nu - \sigma_t^2)\,dt + \eta\sigma_t\,dW_t^{(2)}.$$

In example 2.3.9 we have calculated the expectation function of a solution of this equation as

$$E[\sigma_t^2] = (\sigma_0^2 - \nu)e^{-\kappa t} + \nu, \qquad \forall t \geq 0. \tag{5.19}$$

This is precisely equation (2.22). Firstly, we calculate:

$$\begin{aligned} E\left[\frac{1}{T}\int_0^T \sigma_s^2\,ds\right] = \frac{1}{T}\int_0^T E[\sigma_s^2]\,ds = (\sigma_0^2 - \nu)\frac{1}{T}\int_0^T e^{-\kappa s}\,ds + \nu = \\ = (\sigma_0^2 - \nu)\frac{1 - e^{-\kappa T}}{\kappa T} + \nu \end{aligned} \tag{5.20}$$

We note, that the expected variance calculated above does not depend on the volatility of variance parameter η. The price of a variance swap is the discounted expected value of the terminal payoff V_T under the risk neutral measure [22]:

$$P = e^{-rT}\tilde{E}[N(\sigma_R^2 - \sigma_K^2)] = e^{-rT}N\left[(\sigma_0^2 - \nu)\frac{1 - e^{-\kappa T}}{\kappa T} + \nu - \sigma_K^2\right]. \tag{5.21}$$

[2]Remember, that for the sake of notation simplicity, we have dropped the tildes so that our risk neutral and original SDEs for the CIR variance process coincide (see remark 5.1.1).

Chapter 6

Lévy processes and infinite divisibility

In opening sections of this chapter we follow two standard references on Lévy processes in [24] and [25]. Other great publications, that also contributed to our understanding of the matter are [26, 27]. Lévy processes are an important class of stochastic processes for many reasons; a stochastic integral may be constructed with respect to them and they contain Brownian motion and Poisson process, to name a few. The Lévy processes are well understood, so that the theory is wide and deep, however, in this text, we shall introduce only the essential minimum.

6.1 Infinitely divisible random variables

We start from a slightly different direction, reason for which is the Lévy-Khintchine theorem presented below.

Definition 6.1.1. Let X be a random variable. We say that X is *infinitely divisible* if for all $n \in \mathbb{N}$, there exists n independently and identically distributed random variables $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that

$$X \stackrel{\mathcal{D}}{=} Y_1^{(n)} + \ldots + Y_n^{(n)}.$$

Theorem 6.1.2. The random variable X is infinitely divisible if and only if the characteristic function ϕ_X has an n-th root[1] (denoted $\phi_{X^{1/n}}$), so that

$$\phi_X(u) = [\phi_{X^{1/n}}(u)]^n, \qquad \forall u \in \mathbb{R}$$

and $\phi_{X^{1/n}}$ is itself the characteristic function of a random variable for all $n \in \mathbb{N}$.

Definition 6.1.3. Let ν be a Borel measure defined on $\mathbb{R}$ satisfying $\nu(\{0\}) = 0$. Then we say that ν is a *Lévy measure* if

$$\int_{\mathbb{R}} \min\{y^2, 1\}\, d\nu(y) < \infty$$

[1] This root is unique when X is infinitely divisible.

Theorem 6.1.4 (Lévy-Khintchine)**.** Random variable X is infinitely divisible if there exists $b \in \mathbb{R}$, $A > 0$ and a Lévy measure ν on $\mathbb{R}$ such that

$$\phi_X(u) = \exp\left\{ibu - \frac{1}{2}Au^2 + \int_{\mathbb{R}}[e^{iuy} - 1 - iuy\mathbf{1}_{[-1,1]}(y)]d\nu(y)\right\}, \qquad \forall u \in \mathbb{R}. \tag{6.1}$$

Conversely, any mapping of the form (6.1) is the characteristic function of an infinitely divisible distribution.

The triplet (b, A, ν) is called *Lévy triplet* and is unambiguously defined. [26]

Definition 6.1.5. A function ψ_X such that

$$\psi_X = \ln \phi_X$$

is called a *(Lévy) characteristic exponent (or symbol) of* X.

6.2 Lévy processes

Definition 6.2.1. A stochastic process $X = \{X_t\}$ is called *Lévy process* if the following conditions are satisfied:

1. It starts at zero, so that $X_0 = 0$ P-a.s.

2. The process has *stationary and independent increments*, i.e.

$$X_{s+t} \overset{\mathcal{D}}{=} X_s + X_t, \qquad \forall s, t \geq 0$$

   and $\{X_{t_j} - X_{t_{j-1}}\}_{j=1,\ldots,n}$ are independent random variables for all $n \in \mathbb{N}$ and for all $\{t_1, \ldots, t_n\}$, such that $0 \leq t_1 < \ldots < t_n < \infty$.

3. The process is stochastically continuous, i.e.

$$\lim_{t \to s} P(\{\omega \in \Omega : |X_t(\omega) - X_s(\omega)| > \varepsilon\}) = 0, \qquad \forall s \geq 0, \forall \varepsilon > 0.$$

Remark 6.2.2. Sometimes there is another condition added to the listed in the above definition and that is that X is càdlàg, i.e. its trajectories are a.s. right continuous and have limits from the left.[2]. This is not necessary, for it can be shown, that for every Lévy process defined as 6.2.1 there is a càdlàg modification[3].

One point of view is that by comparison with the definition of Brownian motion (see 1.5.1), the first two conditions are the same, thus Lévy process is a direct generalization of Brownian motion, however without the restriction of continuous sample paths and specification of the distribution. The class of Lévy processes is very wide, so that it accommodates a lot of other processes; we shall present concrete examples below. Now we finally arrive at the reason, why the current chapter was opened by a short section on infinite divisibility.

[2]Origin of the word càdlàg is explained in section D.4.

[3]For the meaning of modification, see section D.1.

Proposition 6.2.3. If $X = \{X_t\}$ is a Lévy process, then X_t is infinitely divisible random variable for each $t \geq 0$.

Theorem 6.2.4. If $X = \{X_t\}$ is a Lévy process, then

$$\phi_{X_t}(u) = e^{t\psi_X(u)}, \qquad \forall u \in \mathbb{R}, \forall t \geq 0,$$

where the characteristic exponent ψ_X is the characteristic exponent of X_1.

Example 6.2.5 (Poisson process). According to example D.3.2 the Poisson process is stochastically continuous. Independence and stationarity of its increments is already included in the definition of the process 1.4.1. We conclude that Poisson process is a Lévy process. In order to determine its characteristic exponent and Lévy triplet of (6.1), we need to compute:

$$E[e^{iuN_t}] = \sum_{n=0}^{\infty} e^{iun} e^{-\lambda t} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(e^{iu}\lambda t)^n}{n!} = e^{-\lambda t} \exp\left\{e^{iu}\lambda t\right\},$$

so that the characteristic function is

$$\phi_{N_t}(u) = \exp\left\{t\lambda(e^{iu} - 1)\right\}, \qquad \forall u \in \mathbb{R}. \tag{6.2}$$

This can be rewritten in the Lévy-Khintchine form (for $t = 1$) as

$$\phi_N(u) = \exp\left\{i\lambda u - \int_{\mathbb{R}} [e^{iuy} - 1 - iuy\,\mathbf{1}_{[-1,1]}(y)]\,\lambda\delta(y-1)\,dy\right\}, \qquad \forall u \in \mathbb{R}. \tag{6.3}$$

Here in terms of the constants of the Lévy-Khintchine theorem $b = \lambda$, $A = 0$ and the Lévy measure is

$$d\nu(y) = \lambda\delta(y-1)\,dy, \tag{6.4}$$

where $\delta(y-1) = \delta_1(y)$ is the Dirac delta function centered at 1. The Lévy triplet is then $(\lambda, 0, \lambda\delta_1)$.

6.3 Lévy measure

Example 6.3.1. As we have already mentioned, standard Brownian motion is a Lévy process. Now we turn our attention to determining its Lévy triplet; we also derive these characteristics for Brownian motion with a drift in the process. It is a known fact that for normal random variable with mean μ and variance σ^2, i.e.

$$X \sim \mathrm{N}(\mu, \sigma^2),$$

the characteristic function is

$$\phi_X(u) = e^{i\mu u - \frac{1}{2}\sigma^2 u^2}, \qquad \forall u \in \mathbb{R}.$$

This formula is generally well known, thus we shall not engage in proving it here. Although the derivation is not complicated, it requires contour integration in the complex

plane, which is out of the scope of this text. Now, standard Brownian motion is $\mathrm{N}(0,t)$ distributed for each $t \geq 0$, hence its characteristic function can be written as:

$$\phi_{W_t}(u) = e^{-\frac{1}{2}tu^2}, \qquad \forall u \in \mathbb{R}, \forall t \geq 0. \tag{6.5}$$

From this, it is immediate, that the standard Brownian triplet is $(0,t,0)$.

The Brownian motion with a drift defined as $B_t := at + \sigma W_t$ is $\mathrm{N}(at, \sigma^2 t)$ distributed, so that the characteristic function takes form of

$$\phi_{B_t}(u) = e^{iatu - \frac{1}{2}\sigma^2 tu^2}, \qquad \forall u \in \mathbb{R}, \forall t \geq 0. \tag{6.6}$$

It follows, that the Lévy triplet for the Brownian motion with a drift is $(at, \sigma^2 t, 0)$.

From the previous inspection (in the preceding examples) we conclude, that in the Lévy triplet, b is the parameter responsible for a drift component of the Lévy process, A represents the diffusive behavior and ν counts the jumps. To rigorously support the last assertion, we develop some more theory on the Lévy measure.

Definition 6.3.2. Let $X = \{X_t\}$ be a Lévy process. Then the *jump height* (or *size*) *process* $\Delta X = \{(\Delta X)_t\}$ is given by[4]

$$(\Delta X)_t := X_t - X_{t-}, \qquad t \geq 0.$$

Naturally if in a particular point t_0 process X is continuous from the left (in addition to being continuous from the right) then $(\Delta X)_{t_0} = 0$, so that the size of the jump is zero (ergo no jump occurred).

Definition 6.3.3. For each $t \geq 0$ and $B \in \mathcal{B}$ we define the random variable[5]

$$N_t^B(\omega) := \#\{s \in [0,t] \,:\, (\Delta X)_s(\omega) \in B\}, \qquad \forall \omega \in \Omega.$$

This random variable represents *number of jumps of size contained in* B *during the time interval* $[0,t]$ (regarding the trajectories).

Theorem 6.3.4. Let X be a Lévy process. If 0 is not in the closure of $B \in \mathcal{B}$, then the collection $\{N_t^B\}_{t \geq 0}$ is a Poisson process with intensity $\nu(B)$, where ν is the Lévy measure of X, i.e.

$$EN_t^B = t\,\nu(B), \qquad \forall t \geq 0.$$

Remark 6.3.5. From the theorem above one immediately obtains that if ν is identically zero measure (as is the case for Brownian motion and Brownian motion with a drift per example 6.3.1), the process to which this measure corresponds has almost surely no jumps of any sizes. This is equivalent to almost sure continuity of trajectories.

Remark 6.3.6. In case of the Poisson process with intensity λ the Lévy measure is

$$\nu(B) = \int_B d\nu(x) = \lambda \int_{\mathbb{R}} \mathbf{1}_B(x)\delta_1(x)\,dx,$$

so that ν is identically zero on any set that does not contain 1. This implies, that the Poisson process has only jumps of unit size.

[4]Here X_{t-} denotes the limit from the left, $\lim_{s \to t-} X_s$, which almost surely exists for the càdlàg modification of Lévy process we are working with.

[5]The hash symbol in front of a given set denotes a number of elements of that set.

In the theorem 6.3.4 we have excluded sets $B \in \mathcal{B}$ such that $0 \in \overline{B}$ from our considerations. That was caused by the fact that the following lemma holds.

Lemma 6.3.7. [24] If $B \in \mathcal{B}$ and $0 \notin \overline{B}$, then $N_t^B < \infty$ P-almost surely for all $t \geq 0$.

Unfortunately a set for which the discussion above dose not apply is for example $B = \mathbb{R}$ and we have to treat this separately.

Definition 6.3.8. We say, that the Lévy process X is of

- *finite activity* if almost all trajectories have finite number of jumps in any time interval $[0, t]$ and
- *infinite activity* if almost all trajectories have infinite number of jumps in any time interval $[0, t]$.

Theorem 6.3.9. Let X be a Lévy process with Lévy measure ν.

1. If $\nu(\mathbb{R}) < \infty$, then X is of finite activity.
2. If $\nu(\mathbb{R}) = \infty$, then X is of infinite activity.

Because the defining condition of Lévy measure is

$$\infty > \int_{\mathbb{R}} \min\{x^2, 1\}\, d\nu(x) = \int_{-1}^{1} x^2\, d\nu(x) + \int_{\mathbb{R}\setminus(-1,1)} d\nu(x),$$

the boundedness of

$$\nu(\mathbb{R}) = \int_{\mathbb{R}} d\nu(x) = \int_{-1}^{1} d\nu(x) + \int_{\mathbb{R}\setminus(-1,1)} d\nu(x)$$

(and thus the diversification between cases 1. and 2. in theorem 6.3.9) is determined by boundedness of the integral $\int_{-1}^{1} d\nu(x)$. This shows that if $\nu(\mathbb{R}) = \infty$, so that the process in question is of infinite activity, then, on one hand, its trajectories have infinitely many jumps in any time interval $[0, t]$, but on the other hand, most of the jumps are small (with size under 1).

We have already encountered the case of $\nu(\mathbb{R}) < \infty$. Both Brownian motion ($\nu = 0$) and Poisson process ($\nu = \lambda\delta_1$) are of finite activity. The opposite case of $\nu(\mathbb{R}) = \infty$ shall be examined in the following chapter.

6.4 Variation of Lévy process

In the previous, we have studied the structure of jumps of a Lévy process, which, to some extent, determine the irregularity of its trajectories. We shall turn our attention to another measure of irregularity, that is, its variation. For definition of variation of a stochastic process, we refer the reader to section 1.6 of chapter 1.

Proposition 6.4.1. [27] A Lévy process with characteristic triplet (b, A, ν) has finite variation if and only if

1. $A = 0$,

2. $\int_{[-1,1]} |x| \; d\nu(x) < \infty$.

With the knowledge of the form of the triplets (derived in example 6.3.1) of Brownian motion - $(0, t, 0)$ and Brownian motion with a drift - $(at, \sigma^2 t, 0)$, we have a trivial corollary that the variation of either of these processes is infinite. This is in line with the result 1.6.5, that we have mentioned in chapter 1.

Characteristic triplet for the Poisson process is $(0, \lambda\delta_1, 0)$, so, unlike for Brownian motion, the condition 1. of the theorem 6.4.1 is not violated. We calculate

$$\int_{[-1,1]} |x| \; d\nu(x) = \lambda \int_{[-1,1]} |x| \; \delta(x-1)\, dx = \lambda < \infty,$$

therefore the Poisson process has finite variation on any time interval $[0, t]$.

For later convenience, we mention the following theorem.

Theorem 6.4.2 (Jordan)**.** [28] A (deterministic) function is of finite variation on [0,t] if and only if it is a difference of two non-decreasing (deterministic) functions.

Remark 6.4.3. Although this theorem holds for deterministic functions, we have defined the variation of a stochastic process X as P-almost sure variation of its paths. From this, we have the path-wise equality $X = Y^1 - Y^2$, where Y^1, Y^2 are P-a.s. non-decreasing stochastic processes, so that the Jordan theorem holds for stochastic processes as well.

6.5 Models with discontinuities

In this chapter, we have developed the theory of Lévy processes. The last few sections have shown, that processes, that belong to this class often contain discontinuities. If we intend to utilize Lévy processes as models in finance, we shall compare these mathematical structures with the real life situation.

An important property of Brownian motion is the continuity of its sample paths. As remarked by [27], the empirical evidence suggests, that real life price processes undergo sudden jumps which appear as discontinuities in the price trajectory. This empirical evidence rests on the fact that there are commonly observed six-standard deviation moves (referenced in the standard deviation of log-returns) across the markets. The motivation for introducing price discontinuities into the models is not just purely intellectual stimulation, but also a practicality, because models with jumps more closely resemble the reality. The very existence of markets for short-term options is an evidence, that jumps are recognized by the market participants.

A common misconception is that these large moves can be accounted for by considering processes with distributions that exhibit heavier tails (positive excess kurtosis). Contrary to that belief, geometric Brownian motion as a log-normally distributed process already incorporates heavier tails and models of local and stochastic volatility are also non-Gaussian. Many Lévy processes, that we shall introduce in the next chapter, generically lead to highly variable returns with realistic tail behavior. In comparison to that, the continuous diffusion based models simply cannot account for the abrupt changes

present in the market without the parameters being set to some unreasonably high values (such as η - the volatility of variance in the Heston model).

In the complete market model (such as BSM, local volatility), any option can be replicated by a self-financing strategy by trading in the underlying asset and money-market account. In such markets, options would be made redundant - they can be perfectly replicated. On the other hand, in real markets, perfect hedging is not possible and options enable market participants to hedge risks, that cannot be hedged by buying and selling the underlying only. We shall introduce examples of pure jump stochastic volatility models in the next chapter.

Chapter 7

Random change of time

We start by heuristically describing the aim of this chapter. Following [29] (which is considered the ultimate text on change of time methods), it is our intent to introduce the processes $\mathcal{T}$ and $\hat{\mathcal{T}}$ called a *random changes of time*. This name is justified by the fact, that these processes transform the old *clock time* t into a new *business time* (or *traders' time*) via $\mathcal{T}(t) \mapsto \theta$ and vice versa via $\hat{\mathcal{T}}(\theta) \mapsto t$. The reason for this construction is that after composing one of these with an appropriate process (properties of which are content of the below discussion) $X = \{X_t\}$ in one of the following ways:

$$\hat{X}_\theta := X_{\hat{\mathcal{T}}(\theta)}, \quad \forall \theta \geq 0 \qquad \Longleftrightarrow \qquad \hat{X} = X \circ \hat{\mathcal{T}}$$

$$X_t := \hat{X}_{\mathcal{T}(t)}, \quad \forall t \geq 0 \qquad \Longleftrightarrow \qquad X = \hat{X} \circ \mathcal{T}$$

Motivation for this direction is that starting from a certain well-known model with "simple" dynamics described by a stochastic process X (in the old clock time) one can arrive at a new model with rather non-trivial dynamics in the new business time. These non-trivial dynamics are useful in modeling (for example) the volatility clustering effect for which there is empirical evidence. We shall point out below the connection between time and volatility.

7.1 Definitions and rationale

Definition 7.1.1. A stochastic process $\hat{\mathcal{T}} = \{\hat{\mathcal{T}}\}_{\mathbb{R}_0^+}$ is called an *(inverse) random change of time given the filtration* $\{\mathcal{F}_t\}$ if the following two conditions hold:

1. The process $\hat{\mathcal{T}}$ is $\mathbb{R}_0^+$-valued and its trajectories are almost surely right-continuous and increasing, i.e.
$$\theta_1 \leq \theta_2 \quad \Longrightarrow \quad \hat{\mathcal{T}}(\theta_1) \leq \hat{\mathcal{T}}(\theta_2), \qquad \forall \theta_1, \theta_2 \geq 0.$$
2. For all $\theta \geq 0$, the random variables $\hat{\mathcal{T}}(\theta)$ are stopping times with respect to the filtration $\{\mathcal{F}_t\}$, i.e.
$$\{\hat{\mathcal{T}}(\theta) \leq t\} \in \mathcal{F}_t, \qquad \forall \theta \geq 0, \forall t \geq 0.$$

Definition 7.1.2. A stochastic process $\mathcal{T} = \{\mathcal{T}(t)\}$ defined by

$$\mathcal{T}(t) := \inf\{\theta : \hat{\mathcal{T}}(\theta) > t\}, \qquad \forall t \geq 0, \tag{7.1}$$

is called a *random change of time given the filtration* $\hat{\mathcal{F}}_\theta := \mathcal{F}_{\hat{\mathcal{T}}(\theta)}$.

Proposition 7.1.3. The process $\mathcal{T}$ defined by (7.1) satisfies conditions 1 and 2 from definition 7.1.1 (the second with respect to the filtration $\hat{\mathcal{F}}_\theta$), thus it is a well-defined random change of time.

Remark 7.1.4. The processes $\mathcal{T}$ and $\hat{\mathcal{T}}$ are mutually inverse in the sense that starting from one, the other can be reconstructed by (7.1) and furthermore, if $\mathcal{T}$ is continuous and strictly increasing, then

$$\hat{\mathcal{T}}(\mathcal{T}(t)) = t \quad \text{and} \quad \mathcal{T}(\hat{\mathcal{T}}(\theta)) = \theta,$$

so that

$$\hat{\mathcal{T}}(\theta) = \mathcal{T}^{-1}(\theta) \quad \text{and} \quad \mathcal{T}(t) = \hat{\mathcal{T}}^{-1}(t), \qquad \forall \theta \geq 0, \forall t \geq 0.$$

As hinted by the opening article to this chapter, our main goal shall be, starting from a "simple" process $\hat{X}$ and changing the time $\mathcal{T} : t \mapsto \theta$, to obtain more sophisticated process so that at least one of the following holds[1]:

$$X = \hat{X} \circ \mathcal{T} \qquad X \overset{\text{a.s.}}{=} \hat{X} \circ \mathcal{T} \qquad X \overset{\mathcal{D}}{=} \hat{X} \circ \mathcal{T}.$$

The intuitive choice for such process with a "simple" structure is the Brownian motion, for after going through the previous chapters, the reader should now be convinced that Brownian motion is of great importance to the discipline of mathematical finance.

Generally one can proceed in two directions, firstly, starting with this "simple" process (say, Brownian motion), plugging in various time-changes $\mathcal{T}$ and obtaining the non-trivial processes. Secondly, it is possible to try to solve a somewhat inverse problem, that is, starting from a "simple" process $\hat{X}$ and a "target" process with more sophistication, to try to find time-changes $\mathcal{T}$ such that $X = \hat{X} \circ \mathcal{T}$ holds in some sense. Naturally the second approach is more demanding, however, we have the following existence theorem:

Theorem 7.1.5 (Monroe)**.** If $X = \{X_t\}_{\mathbb{R}_0^+}$ is a semimartingale[2] with respect to its natural filtration, then there exists a filtered probability space with a Brownian motion $\hat{W} = \{\hat{W}_\theta\}_{\mathbb{R}_0^+}$ and a change of time $\mathcal{T} = \{\mathcal{T}(t)\}_{\mathbb{R}_0^+}$ defined on it such that

$$X \overset{\mathcal{D}}{=} \hat{W} \circ \mathcal{T}.$$

Starting from a wide class of semimartingales, it is quite obvious, that the time-change $\mathcal{T}$ that assures that the Brownian motion $\hat{W}$ has the same distribution as X, can posses too much sophistication. For example if the process X has jumps, then because of the continuity of trajectories of Brownian motion the jumps need to be accounted for by the discontinuity of $\mathcal{T}$. There is another important theorem, which takes this into account by assuming the continuity of X:

---

[1] Here (as in the opening article to this chapter) the meaning of $\hat{X} \circ \mathcal{T}$ is that we evaluate the process $\hat{X}_\theta$ at a time-point given as an outcome of the random variable $\theta = \mathcal{T}(t)$. We shall always assume that the random variables $\hat{X}$ and $\mathcal{T}$ are defined on the same probability space.

[2] For the definition of semimartingale, see section D.4 of appendix D. This is only informative, because we shall not use this theorem in the following.

Theorem 7.1.6 (Dambis-Dubins-Schwarz)**.** Let $X = \{X_t\}$ be a continuous $\{\mathcal{F}_t\}$-local martingale with $X_0 = 0$ and $\lim_{t\to\infty}\langle X\rangle_t = \infty$. Then there exists a Brownian motion $\hat{W} = \{\hat{W}_\theta\}$ such that for the change of time defined by

$$\mathcal{T}(t) := \langle X\rangle_t \qquad \forall t \geq 0,$$

the strong representation $X = \hat{W} \circ \mathcal{T}$ holds.

Corollary 7.1.7. Let a continuous local $\{\mathcal{F}_t\}$-martingale be defined as

$$X_t := \int_0^t \sigma_s \, dW_s, \qquad \forall t \geq 0, \tag{7.2}$$

where $W = \{W_t\}$ is a Brownian motion with $\{\mathcal{F}_t\}$ its natural filtration. The process $\sigma = \{\sigma_t\}$ is supposed to be P-a.s. positive, adapted to $\{\mathcal{F}_t\}$ and

$$\int_0^t \sigma_s^2 \, ds < \infty \quad \forall t \in [0,\infty) \qquad \text{with} \qquad \int_0^\infty \sigma_s^2 \, ds = \infty.$$

If we put

$$\hat{\mathcal{T}}(\theta) := \inf\left\{ t : \int_0^t \sigma_s^2 \, ds \geq \theta \right\}, \tag{7.3}$$

then the process defined by $\hat{W}_\theta := X_{\hat{\mathcal{T}}(\theta)}$ is a Brownian motion (with respect to the filtration $\{\hat{\mathcal{F}}_\theta\}$) and $X = \hat{W} \circ \mathcal{T}$ with

$$\mathcal{T}(t) = \int_0^t \sigma_s^2 \, ds \tag{7.4}$$

i.e.

$$X_t = \hat{W}_{\int_0^t \sigma_s^2 \, ds}$$

The corollary hints on the possible applications of the change of time approach in models of stochastic volatility. We shall explore this in the following sections.

7.2 Subordinators

As mentioned above, it is in general difficult for a given stochastic process X (possibly a semimartingale or Lévy process) and $\hat{W}$ θ-time Brownian motion to find a reasonable change of time $\mathcal{T}$ so that $X = \hat{W} \circ \mathcal{T}$ in some sense. In this section, we shall simplify by reversing the process - assuming a Brownian motion (or some other "simple" process, usually the Brownian motion with a drift) and time-changes from the class of Lévy processes (called subordinators), we ask the question - what are the processes $X := \hat{W} \circ \mathcal{T}$ that we are going to arrive at? Proceeding in this direction, we do not mind the sophistication of the process $\mathcal{T}$, actually, in the financial applications, it is a desirable quality.

Definition 7.2.1. We call the change of time $\mathcal{T} = \{\mathcal{T}(t)\}_{\mathbb{R}_0^+}$ a *subordinator (in the strong sense)* if it is a Lévy process. In the case there are two processes $X, \hat{X}$, such that $X = \hat{X} \circ \mathcal{T}$, we say, that X is *subordinated* to $\hat{X}$ via (the subordinator) $\mathcal{T}$.

For subordinators, we have the following characterization:

Theorem 7.2.2. [30] If $\mathcal{T}$ is a subordinator, then its characteristic exponent is of the form

$$\psi_{\mathcal{T}}(u) = ibu + \int_0^\infty (e^{iuy} - 1)d\nu(y), \qquad \forall u \in \mathbb{R}, \tag{7.5}$$

where $b \geq 0$ and the Lévy measure ν satisfies the additional requirements

$$\nu((-\infty, 0)) = 0 \quad \text{and} \quad \int_0^\infty \min\{y, 1\}\, d\nu(y) < \infty.$$

Conversely, any real, complex valued function of the form (7.5) is the characteristic exponent of a subordinator.

Example 7.2.3 (Gamma process). We assume that at every $t \geq 0$ the Lévy process is gamma distributed $\mathcal{T}(t) \sim \text{Gamma}(\alpha t, \beta)$, where $\alpha > 0$, $\beta > 0$ i.e.

$$f_{\mathcal{T}(1)}(x) := \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{[0,\infty)}(x), \qquad \forall x \in \mathbb{R}. \tag{7.6}$$

In order to compute the characteristic exponent and to find the Lévy measure of this gamma process, we calculate the following integral for all $u \geq 0$:

$$m_{\mathcal{T}(1)}(-u) = \int_{-\infty}^\infty e^{-ux} f_{\mathcal{T}(1)}(x)\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(u+\beta)x}\, dx.$$

Performing the substitution $(u + \beta)x =: \tilde{x}$ one arrives at

$$m_{\mathcal{T}(1)}(-u) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(u+\beta)^\alpha} \underbrace{\int_0^\infty x^{\alpha-1} e^{-x}\, dx}_{=\Gamma(\alpha)} = \left(1 + \frac{u}{\beta}\right)^{-\alpha} = \exp\left\{-\alpha \ln\left(1 + \frac{u}{\beta}\right)\right\}.$$

Because of the integral relation[3]

$$\ln\left(1 + \frac{u}{\beta}\right) = \int_0^\infty \frac{(1 - e^{-ux})}{x} e^{-\beta x}\, dx,$$

we have (because the function $m_{\mathcal{T}(1)}$ is, up to a change of sign, a moment generating function of $\mathcal{T}(1)$)

$$\phi_{\mathcal{T}(1)}(u) = m_{\mathcal{T}(1)}(iu) = \left(1 - i\frac{u}{\beta}\right)^{-\alpha} = \exp\left\{\alpha \int_0^\infty (e^{iux} - 1) \frac{e^{-\beta x}}{x}\, dx\right\}.$$

[3]This relation might be proven using the Frullani integral relation

$$\int_0^\infty \frac{f(Ax) - f(Bx)}{x}\, dx = f(0) \ln \frac{B}{A}$$

under the assumptions that f is continuously differentiable and $\lim_{x\to\infty} f(x) = 0$. In our case $f(x) := e^{-x}$ and $A := -\beta, B := -(u + \beta)$.

Now, the characteristic exponent of gamma process is

$$\psi_{\mathcal{T}(1)}(u) = \alpha \int_0^\infty (e^{iux} - 1)\frac{e^{-\beta x}}{x}\, dx \tag{7.7}$$

and comparing this with equality (7.5) the Lévy measure is

$$d\nu(x) = \alpha\frac{e^{-\beta x}}{x}\mathbf{1}_{(0,\infty)}(x)\, dx, \qquad \forall x \in \mathbb{R}. \tag{7.8}$$

Therefore, the Lévy triplet is $(0, 0, \nu)$ with ν given as (7.8). From the above, also the characteristic function of the process may be retained[4]:

$$\phi_{\mathcal{T}(t)}(u) = [\phi_{\mathcal{T}(1)}]^t = \left(1 - i\frac{u}{\beta}\right)^{-\alpha t}. \tag{7.9}$$

Because the characteristic triplet is $(0, 0, \nu)$, we deduce, that gamma process does not exhibit diffusive behavior in sense of Brownian motion and there is no drift part either. We now discuss the properties of ν. It is clearly a Lévy measure by definition 6.1.3, however $\nu(\mathbb{R}) = \infty$. This is obvious, once we take into account, that, according to concluding remarks of section 6.3, the boundedness of $\nu(\mathbb{R})$ is determined by the boundedness of $\nu((-1, 1))$:

$$\nu((-1,1)) = \alpha \int_{-1}^1 \frac{e^{-\beta x}}{x}\mathbf{1}_{(0,\infty)}(x)\, dx = \alpha \int_0^1 \frac{e^{-\beta x}}{x}\, dx.$$

This integral diverges, ergo gamma process is of infinite activity (theorem 6.3.9), which means that it exhibits infinitely many jumps in any time interval $[0, t]$ and most of the jumps are small - with size under 1 (most of the Lévy measure is concentrated around the origin of coordinates).

Now, calculating

$$\int_{[-1,1]} |x|\, d\nu(x) = \alpha \int_{[-1,1]} |x|\, \frac{e^{-\beta x}}{x}\mathbf{1}_{(0,\infty)}(x)\, dx = \alpha \int_0^1 e^{-\beta x}\, dx = \frac{\alpha}{\beta}(1 - e^{-\beta}),$$

this integral is finite, even for β very small (in the zero limit). Because, again, the Lévy triplet is $(0, 0, \nu)$, we conclude (using theorem 6.4.1), that the gamma process has finite variation on any time interval $[0, t]$.

One of the reasons, why subordination is such a useful concept is the closedness property of the class of Lévy processes under subordination, i.e. processes subordinated to Lévy processes are also Lévy processes. This assertion is formalized in the following theorem:

Theorem 7.2.4 (Subordination of Lévy processes)**.** [25][27] Let $X = \{X_t\}$ be a Lévy process with characteristic exponent ψ_X and a triplet (b, A, ν_X). Further, let $\{\mathcal{T}(t)\}_{t\geq 0}$ be a subordinator with characteristic function $\phi_{\mathcal{T}(t)}$ and a triplet $(b, 0, \nu_{\mathcal{T}})$. Then the process $\{Y_t\}_{t\geq 0}$ subordinate to X via $\mathcal{T}$, i.e.

$$Y_t(\omega) := X_{\mathcal{T}_t(\omega)}(\omega),$$

[4]The relations that we have derived above are only formal equalities, due to the step, where we exchange u and iu in the moment generating function. Nevertheless, they can be fully justified by complex-theoretic means, namely the analytic continuation.

is a Lévy process. Its characteristic function is given as

$$\phi_{Y_t}(\omega) = E[e^{iuY_t}] = \phi_{\mathcal{T}(t)}(-i\psi_X(u)), \qquad \forall t \geq 0, \forall u \in \mathbb{R}. \tag{7.10}$$

In the next section, we shall utilize the above defined gamma subordinator to change the clock in the Black-Scholes-Merton model.

7.3 Use in pricing models

In the previous, we have considered several models for the evaluation of a derivative contract written on some underlying equity product (usually a stock). The simplest model that we used was the BSM of chapter 3, where the behavior of the underlying was modeled by stochastic differential equation

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t$$

and its solution, the geometric Brownian motion

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\}. \tag{7.11}$$

From this, the log-return process $H = \{H_t\}$ such that

$$S_t = e^{H_t}, \qquad \forall t \geq 0, \tag{7.12}$$

is modeled as a Brownian motion with a drift:

$$H_t = at + \sigma W_t, \qquad \text{where} \qquad a := \left(\mu - \frac{1}{2}\sigma^2\right). \tag{7.13}$$

Later, we generalized in order to capture the implied smile effects and to compensate for various deficiencies of the BSM model, by assuming the volatility to be a random process itself. The equation for the underlying reads

$$dS_t = \mu S_t \, dt + \sigma_t S_t \, dW_t,$$

with a generalized geometric Brownian motion solution

$$S_t = S_0 \exp\left\{\mu t - \frac{1}{2}\int_0^t \sigma_s^2 \, ds + \int_0^t \sigma_s \, dW_s\right\}. \tag{7.14}$$

An alternative approach to achieving a stochastic volatility in the simple model (7.11) is to introduce a change of time $\{\mathcal{T}\}$ into the log-returns process (7.12) so that:

$$H_t := a\mathcal{T}(t) + \sigma W_{\mathcal{T}(t)}, \forall t \geq 0. \tag{7.15}$$

Here $H = \hat{H} \circ \hat{\mathcal{T}}$, where $\mathcal{T} : t \mapsto \theta$ is the change of time, so that the log-returns process $\hat{H}$ now evolves in the θ-time, or in other words in the perceived business time. This approach was pioneered by [31, 32]. We shall illustrate this general procedure by a concrete example. From now on, we restrict the class of random time-changes that we are going to work with to subordinators, that were defined in the previous section of this chapter.

Example 7.3.1 (Variance gamma process). Let $\{B_t\}$ denote a Brownian motion with a drift:

$$B_t = at + \sigma W_t, \qquad \forall t \geq 0. \tag{7.16}$$

Let $\{\mathcal{T}(t)\}$ be a subordinator given as the gamma process, whose properties were analyzed in the example 7.2.3. In particular, we have the following probability density function (PDF):

$$f_{\mathcal{T}(t)}(x) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-\beta x} \mathbf{1}_{[0,\infty)}(x), \qquad \forall x \in \mathbb{R}. \tag{7.17}$$

and the Lévy measure (7.8). We define a *variance gamma (VG) process* $Y = \{Y_t\}$ to be the process subordinated to Brownian motion with a drift B via the gamma process subordinator $\mathcal{T}$, i.e.

$$Y := B \circ \mathcal{T}. \tag{7.18}$$

Having the probability density function of $\{B_\theta\}$ conditioned on the realization of the θ-time given as

$$f_{B_\theta}(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2\theta}} \exp\left\{-\frac{(x - a\theta)^2}{2\sigma^2\theta}\right\}, \qquad \forall x \in \mathbb{R}, \forall \theta \geq 0, \tag{7.19}$$

we can now explicitly specify the PDF of the variance gamma process as

$$f_{Y_t}(x) = \int_0^\infty f_{B_\theta}(x|\theta) f_{\mathcal{T}(t)}(\theta)\, d\theta = \tag{7.20}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\theta}} \exp\left\{-\frac{(x - a\theta)^2}{2\sigma^2\theta}\right\} \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} \theta^{\alpha t - 1} e^{-\beta\theta}\, d\theta.$$

We note here, that the equality in (7.20) is exactly that contained in the definition of marginal probability density function E.1.1 (in appendix E)[5]. Apparently the variance of the VG process is gamma distributed (up to a multiplicative constant σ^2) so that the name of the process is in line with that. Next, we shall employ (E.2) to determine the characteristic function of VG process:

$$\phi_{Y_t}(u) = \int_0^\infty \phi_{B_\theta}(u|\theta) f_{\mathcal{T}(t)}(\theta)\, d\theta = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} \int_0^\infty \exp\left\{-(\beta - iau + \frac{1}{2}\sigma^2 u^2)\theta\right\} \theta^{\alpha t - 1}\, d\theta$$

Mutatis mutandis, this is the same integral as in the calculation of the gamma characteristic function in example 7.2.3, therefore, by using the same substitution and integration, we arrive at:

$$\phi_{Y_t}(u) = \left(1 - i\frac{a}{\beta}u + \frac{1}{2}\frac{\sigma^2}{\beta}u^2\right)^{-\alpha t}, \qquad \forall u \in \mathbb{R}. \tag{7.21}$$

Variance gamma process was first defined as process subordinated to Brownian motion with a drift via a gamma process in [32], where the parametrization $\alpha = \beta = 1/\zeta$ was used. Then the characteristic function takes form of

$$\phi_{Y_t}(u) = \left(1 - ia\zeta u + \frac{1}{2}\sigma^2\zeta u^2\right)^{-t/\zeta}, \qquad \forall u \in \mathbb{R}. \tag{7.22}$$

[5]In the language of marginalization and superstatistics the PDF of VG process is gamma-weighted mixture of normal distributions where the marginalization parameter is the business time θ.

The authors have shown, that the variance gamma process may be written as a difference of two P-a.s. non-decreasing random processes, $\{\gamma_t^+\}, \{\gamma_t^-\}$, so that

$$Y_t = \gamma_t^+ - \gamma_t^-, \qquad \forall t \geq 0. \tag{7.23}$$

In fact, these are gamma processes

$$\gamma^\pm \sim \text{Gamma}\left(\frac{t}{\zeta}, \frac{1}{\zeta m_\pm}\right), \tag{7.24}$$

where

$$m_\pm := \frac{1}{2}\sqrt{a^2 + \frac{2\sigma^2}{\zeta}} \pm \frac{a}{2}. \tag{7.25}$$

This might be verified by employing the formula (C.5), where the characteristic functions of the two gamma processes are given by (7.9) after adjusting for the parameters. Recalling the Jordan theorem 6.4.2, we observe that, due to the decomposition above, the VG process has infinite variation.

Because of theorem 7.2.4, we have that the variance gamma process is a Lévy process. According to [32] the VG process inherits the infinite activity from the gamma subordinator. In the simplified parametrization ($\alpha = \beta = 1/\zeta$ for the gamma process) defined above, the VG process has three parameters; σ - the volatility of Brownian motion, a - its drift, ζ - the variance rate of the gamma time-change.

Because the VG model (employing the VG process) is constructed as being subordinated to Brownian motion with a drift, the Black-Scholes-Merton model is a special parametric case of this model. Price of a European call option in the VG model is obtained by first conditioning on the random time θ, then calculating the BSM price $C_{\text{BSM}}(\theta, S_\theta)$ via (3.35) and finally integrating:

$$C_{\text{VG}}(t, Y_t) = \int_0^\infty C_{\text{BSM}}(\theta, S_\theta) f_{\mathcal{T}(t)}(\theta)\, d\theta = \int_0^\infty C_{\text{BSM}}(\theta, S_\theta) \frac{\theta^{\frac{t}{\zeta}-1} e^{-\frac{\theta}{\zeta}}}{\zeta^{\frac{t}{\zeta}}\Gamma\left(\frac{t}{\zeta}\right)}\, d\theta.$$

Example 7.3.2 (Variance gamma with stochastic arrival). The variance gamma process introduced in the previous example is a pure jump process, where the infinite activity is implemented via subordination. Because the changes in time directly affect the variance (remember, the V in VG stands for variance) and hence the volatility of the underlying, the model based on VG process is a pure jump stochastic volatility model. However, the model does not capture the volatility clustering effect which is a feature of asset prices in many different markets. [17] Inspired by the Heston model, which implements the clustering effect by assuming the mean reverting CIR process for variance, we define a new process that is subordinated to variance gamma process via the integrated CIR process time-change:

$$\mathcal{T}(t) = \int_0^t X_s\, ds. \tag{7.26}$$

Here

$$dX_t = k(n - X_t)\, dt + b\sqrt{X_t}\, dW_t \tag{7.27}$$

and the process defined as $Y \circ \mathcal{T}$ ($\{Y_t\}$ is the VG process) is called the *variance gamma process with stochastic arrival (of jumps)* or simply *VGSA*. Defining the process in this way, the changes in volatility persist and thus give rise to the clustering effect. Here $\{X_t\}$ plays the role of instantaneous rate of time-change, n is the long term rate of time-change, k is the rate of mean reversion and b is the volatility of time-change. Recalling the theorem 7.2.4, we see that, firstly, VGSA is a Lévy process and secondly, its characteristic function may be obtained from formulas (2.25) and (7.22) as (7.10). This advanced model is due to P. Carr et al. [7], in the article the pricing formulae for options was derived as well.

7.4 Heston model as GGBM subordinated to GBM

We shall assume the CIR process $\{v_t\} = \{\sigma_t^2\}$ governed by the CIR equation:

$$dv_t = \kappa(\nu - v_t)\,dt + \eta\sqrt{v_t}\,dW_t^{(2)}, \tag{7.28}$$

where $\{W_t^{(2)}\}$ is a Brownian motion. Specifically, we do not assume that volatility of the underlying asset is stochastic and we do not assume, that it is $\{\sigma_t\}$. We assume the presence of other source of randomness, the Brownian motion $W^{(1)} = \{W_t^{(1)}\}$. If we integrate the square root of the CIR process with respect to the other Brownian motion $W^{(1)}$, we obtain a martingale (w.r.t. natural Brownian filtration):

$$M_t := \int_0^t \sqrt{v_s}\,dW_s^{(1)} = \int_0^t \sigma_s\,dW_s^{(1)}, \qquad \forall t \in [0,T]. \tag{7.29}$$

Now, from corollary 7.1.7 of the Dambis-Dubins-Schwarz theorem, we have that there is a change of time

$$\mathcal{T}(t) = \int_0^t \sigma_s^2\,ds = \int_0^t v_s\,ds, \qquad \forall t \in [0,T] \tag{7.30}$$

and a Brownian motion $\{\hat{W}_\theta^{(1)}\}$ evolving in the perceived business time, so that $M = \hat{W} \circ \mathcal{T}$, i.e. (using (7.29))

$$M_t = \int_0^t \sigma_s\,dW_s^{(1)} = \hat{W}_{\mathcal{T}(t)}^{(1)}. \tag{7.31}$$

We can use this new source of randomness to define the underlying asset to evolve as a geometric Brownian motion in the business time[6]:

$$\hat{S} = \hat{S}_0 \exp\left\{a\hat{\mathcal{T}}(\theta) + b\theta + \hat{W}_\theta^{(1)}\right\}. \tag{7.32}$$

Notice, that we are assuming constant volatility parameter in the model for the underlying. Now if we denote by S the process subordinated to geometric Brownian motion via the integrated CIR process (7.30)

$$S := \hat{S} \circ \mathcal{T},$$

[6]Here $\hat{\mathcal{T}}$ is the inverse change of time and the term containing it is added as a mere convenience, for the coefficients to take reasonable form.

we obtain behavior of the underlying in the ordinary clock time (here naturally $S_0 = \hat{S}_0$)

$$S_t = \hat{S}_{\mathcal{T}(t)} = S_0 \exp\left\{at + b\mathcal{T}(t) + \hat{W}^{(1)}_{\mathcal{T}(t)}\right\}, \qquad \forall t \in [0, T]. \tag{7.33}$$

Substituting for the time-change $\mathcal{T}$ (7.30) and from the result (7.31) of the Dambis-Dubins-Schwarz corollary, this takes form of:

$$S_t = S_0 \left\{at + b\int_0^t \sigma_s^2\, ds + \int_0^t \sigma_s\, dW_s^{(1)}\right\}. \tag{7.34}$$

Comparing this with (7.14) and additionally assuming the correlation $dW_t^{(1)}\, dW_t^{(2)} = \varrho\, dt$ between the two sources of randomness, we see that this is the generalized geometric Brownian motion with parameters $a = \mu, b = -1/2$ and $\sigma = 1$, so that we have obtained the Heston model. We have not assumed, that the volatility is stochastic, yet by changing the time via the integrated CIR process, we have arrived at a stochastic volatility model, where the variance is governed by the CIR equation. We have effectively proved that the generalized geometric Brownian motion is subordinated to geometric Brownian motion via the integrated CIR process. This fact can be encountered in the literature, however, usually without a proof; for that reason, we have included one of our own.

Remark 7.4.1. Just as the variance gamma process of example 7.3.1 contains the BSM model as a special case, the VGSV model contains the Heston model. This is obvious from the above derivations. In this sense, the Heston model is the BSM model with stochastic arrival.

Similarly to (7.29), we may define a martingale:

$$\int_0^t \sigma_s\, dW_s^{(2)}, \qquad \forall t \in [0, T],$$

which through the corollary of Dambis-Dubins-Schwarz theorem yields the same change of time (7.30) and a Brownian motion $\hat{W}^{(2)}$ in the θ-time, such that

$$\int_0^t \sigma_s\, dW_s^{(2)} = \hat{W}^{(2)}_{\mathcal{T}(t)}, \qquad \forall t \in [0, T]. \tag{7.35}$$

Using these relations, by integration, the solution to (7.28) may then be expressed as:

$$v_t = v_0 + \kappa\nu t - \kappa \underbrace{\int_0^t v_s\, ds}_{=\mathcal{T}(t)} + \eta \int_0^t \sqrt{v_s}\, dW_s^{(2)}. \tag{7.36}$$

Taking the risk neutral expectation, we have

$$\tilde{E} v_t = v_0 + \kappa\nu t - \kappa \tilde{E}[\mathcal{T}(t)],$$

so that

$$\begin{aligned}\tilde{E}\left[\frac{1}{T}\int_0^T v_s\, ds\right] = \frac{1}{T}\tilde{E}[\mathcal{T}(T)] = \frac{v_0 + \kappa\nu T - \tilde{E} v_T}{\kappa T} = \\ = \frac{v_0 + \kappa\nu T - (v_0 - \nu)e^{-\kappa T} - \nu}{\kappa T} = \\ = \nu + (v_0 - \nu)\frac{1 - e^{-\kappa T}}{\kappa T}.\end{aligned}$$

By using (2.22) we have obtained equality, that yields the price of a variance swap (5.21). We have avoided exchanging the integration and the expectation in this approach.

Per derivations displayed above, the Heston model in the risk neutral setting may be specified by a pair of equations:

$$S_t = S_0 \exp\left\{ rt - \frac{1}{2}\mathcal{T}(t) + \hat{W}^{(1)}_{\mathcal{T}(t)} \right\}, \tag{7.37}$$

$$v_t = v_0 + \kappa\nu t - \kappa\mathcal{T}(t) + \eta\hat{W}^{(2)}_{\mathcal{T}(t)}. \tag{7.38}$$

7.5 Connection between change of time and SDE approach

In this section, we shall examine, how the change of time methods can be utilized to obtain solutions to stochastic differential equations. According to [33], it is possible to transform a general SDE to the form of:

$$dX_t = \alpha(t, X_t)\, dW_t. \tag{7.39}$$

For this simplified (equivalent) SDE it is then possible to find a weak solution by utilizing the change of time methods. We shall present the theorem as well as a particular application where we find a weak solution to the CIR equation, which is embedded in the Heston model. This allows us to arrive at a price for variance swap from a different angle, than we did in the previous.

Theorem 7.5.1. [33] Let $\{W_t\}$ be an $\{\mathcal{F}_t\}$-Brownian motion with $W_0 = 0$ given on a probability space $(\Omega, \mathcal{A}, P)$ and let X_0 be an $\mathcal{F}_t$-measurable random variable. Define a continuous random process $X = \{X_t\}$ by

$$X_t := X_0 + W_t, \qquad \forall t \geq 0. \tag{7.40}$$

Let $\{\mathcal{T}(t)\}$ be a random change of time process such that

$$\mathcal{T}(t) = \int_0^t [\alpha(\mathcal{T}(s), X_s)]^{-2}\, ds \tag{7.41}$$

holds a.s. Then if we set $\hat{X} = X \circ \hat{\mathcal{T}}$, i.e.

$$\hat{X}_\theta := X_0 + W_{\hat{\mathcal{T}}(\theta)}, \qquad \forall \theta \geq 0 \tag{7.42}$$

and $\hat{\mathcal{F}}_\theta = \mathcal{F}_{\hat{\mathcal{T}}(\theta)}$, there exists an $\{\hat{\mathcal{F}}_\theta\}$-Brownian motion $B = \{B_\theta\}$, such that $(\hat{X}, B)$ is a weak solution of the SDE

$$d\hat{X}_\theta = \alpha(\theta, \hat{X}_\theta)\, dB_\theta.$$

Remark 7.5.2. Note that by writing the integral relationship (7.41) in the SDE form, equivalently:

$$d\mathcal{T}(t) = [\alpha(\mathcal{T}(t), X_t)]^{-2}\, dt \quad \Leftrightarrow \quad dt = [\alpha(\mathcal{T}(t), X_t)]^2\, d\mathcal{T}(t)$$

we arrive at the inverse time-change:

$$\hat{\mathcal{T}}(\theta) = \int_0^{\hat{\mathcal{T}}(\theta)} [\alpha(\xi, \hat{X}_\xi)]^2 \, d\xi. \tag{7.43}$$

The original Brownian motion W and the new Brownian motion B, which together with the process $\hat{X}$ constitutes the weak solution of (7.39), are related in the following way (actually in the proof of theorem 7.5.1 the Brownian motion is defined so that):

$$B_\theta = \int_0^\theta [\alpha(\xi, \hat{X}_\xi)]^{-1} \, dW_{\hat{\mathcal{T}}(\xi)}. \tag{7.44}$$

Again, using the formal differential form, we arrive at inverse relation:

$$W_{\hat{\mathcal{T}}(\theta)} = \int_0^\theta \alpha(\xi, \hat{X}_\xi) \, dB_\xi. \tag{7.45}$$

Remark 7.5.3. The theorem above finds its use in pricing derivatives in certain situations, because sometimes, one is only interested in distributional properties and not in the behavior of trajectories of the process satisfying (7.39). In these situations a weak solution suffices.

We shall use this approach, to value a variance swap in the Heston model. We shall then be able to compare the solution we arrive at here with the one obtained in section 5.4. First, it is necessary to transform the CIR equation

$$dv_t = \kappa(\nu - v_t) \, dt + \eta\sqrt{v_t} \, dW_t, \tag{7.46}$$

describing the behavior of variance, into more suitable form. We shall consider process

$$Y_t := -e^{\kappa t}(\nu - v_t) \tag{7.47}$$

and use Itô product rule on it together with substitution from (7.46):

$$dY_t = -\kappa e^{\kappa t}(\nu - v_t) \, dt - e^{\kappa t} \, dv_t = -\kappa e^{\kappa t}(\nu - v_t) \, dt + e^{\kappa t}\kappa(\nu - v_t) \, dt + \eta e^{\kappa t}\sqrt{v_t} \, dW_t.$$

The first two terms clearly subtract and after substituting in from (7.47), we finally arrive at

$$dY_t = \underbrace{\eta e^{\kappa t}\sqrt{e^{-\kappa t}Y_t + \nu}}_{=:\alpha(t, Y_t)} \, dW_t. \tag{7.48}$$

Now that we have the equation in the transformed form, the theorem 7.5.1 applies. Namely, we have the weak solution $(\hat{X}, B)$, where

$$\hat{X}_\theta = X_0 + W_{\hat{\mathcal{T}}(\theta)}.$$

We use the inverse transformation to (7.47), so that we have a weak solution $(\hat{v}, B)$ of (7.46) in the θ-time:

$$\hat{v}_\theta = e^{-\kappa\theta}[X_0 + W_{\hat{\mathcal{T}}(\theta)}] + \nu. \tag{7.49}$$

Because $W_{\hat{\mathcal{T}}}$ is an Itô integral (see (7.45)), its expectation vanishes and thus

$$E\hat{v}_\theta = e^{-\kappa\theta}X_0 + \nu.$$

The initial condition $X_0 = \hat{X}_0 = Y_0 = v_0 - \nu$ from the transformation defining equation (7.47). Finally

$$\hat{v}_\theta = (v_0 - \nu)e^{-\kappa\theta} + \nu, \tag{7.50}$$

which is exactly equation (5.19), only in terms of the variance and it holds in terms of the business time. From now on, we could continue just as in the rest of the section 5.4 where the business time integrates out and we obtain the correct price of the variance swap.

Remark 7.5.4. Above we have utilized the method of change of time to obtain a weak solution to certain SDE, which then gave us distributional properties needed for pricing. In this way the change of time method is useful, however, we have not obtained anything new and if we were to compute higher moments of (7.49) we would have found out that these calculations are somewhat cumbersome. This is due to the fact that for the higher moments, one has to take into account the explicit form of the function α (for the case of CIR defined in (7.48)). Further obstacles dwell in the process of finding a suitable transformation of the given SDE that is not yet expressed in the form of (7.39). We used the transformation suggested in [34].

Conclusion

At the beginnings of our research, there were two fields of study possibly connected to mathematical finance that we aimed to probe - superstatistics and subordination. As we continued on our quest for understanding, it was becoming more clear that we are swerved off of our general direction (of focusing on both fields) and we are heavily drawn to subordination methods. This was probably caused, amongst other reasons, by the fact that we focused on continuous time finance and stochastic calculus.

Subordination and change of time methods in general turned out to be much better understood than we had initially thought. This is reflected in the depth and rigorousness of the theory which we attempted to capture in the text. In addition to that, subordination is intertwined with finance theory to such extent that we managed to mention only a fraction of potential associations and applications, although we have devoted a generous chapter to them. Nevertheless, we deem our analysis of subordination methods fruitful, for it allowed us to arrive at a few rather non-trivial models with additional tractability provided by the interpretation via change of time.

The common denominator for applications of the methods mentioned above for us was the Heston model. We have not only shown that the stochastic volatility present in the model can be achieved by assuming constant volatility and changing the time, but we have proceeded to derive distributional characteristics of the CIR process by employing these methods as well. In the process we have calculated the price of a variance swap by three different approaches (one considered canonical, the other two less conventional) and presented existing generalization of Heston model - VGSA model. We still believe that there are paths connecting superstatistics to finance that are to be unraveled, however these paths most likely run through discrete models and time series analysis.

Appendix A

Markets, derivatives

There is an infinite complexity to the system of real-life financial markets. Financial modeling naturally aims to lower the degrees of freedom so that it can make quantitative and to some extent qualitative statements about the markets. Although it is undoubtedly advantageous to be aware of all the possible scenarios, definitions and conventions, in this brief appendix, we shall introduce only the simplest concepts that are necessary for our purposes. Nevertheless, one has to bear in mind that the real-life markets are much more than their mathematical representations and formal definitions, so that there are no unpleasant surprises waiting around the corner. We credit [35, 36] and [18] with our understanding of these fundamental matters.

We begin by an overview of some basic financial products that are traded on various exchanges in the financial markets.

A.1 Financial products

A simplest form of investment is depositing certain amount of cash with a financial counterparty (say, a bank) in the so-called *bank account* (or *money-market account*). In the process of doing so, one is effectively lending the money to the counterparty (the bank) and because the cash amount is tied in this way, one expects to receive certain *rate of interest* (an *interest rate*) on it. Interest rates are usually quoted *per annum* (*p.a.*), which represents the annualized interest rate. The ideal of lending (or borrowing) money at certain interest rate is somewhat formalized in the concept of financial product called *bond*. This product is usually exchange traded, so that unlike with the bank account, one can sell the obligation of the bank to other to-be receivers. Bonds may be issued by companies that are not from the financial sector as well. Generally bonds are perceived to be the less risky of financial products, which is encompassed in them being referred to as *securities* or *fixed income products*. We note that there is much more to the concept of a bond (such as credit quality of the issuer, rank of seniority etc.) that we shall not scrutinize here.

Taking the concept of raising money (from the perspective of a company) little further are *stocks* (or *shares*) which represent a fractional ownership of a company that issued them. These are often categorized as *equity* and generally consist a more risky investment than bonds. Since stocks are exchange traded, their price is determined by what market

participants are willing to pay for them[1]. Naturally, if the company which issued the stock is perceived by the market as not prospering, the price of the stock might exhibit a declining trend. With the bond, one usually expects to receive back the whole invested amount (the *notional*) plus some rate of interest in the form of *coupons*; on the other hand price of a stock might well fall down to zero. This additional risk for equity is commonly balanced by higher upside in the form of higher *rate of return* (expected profit) on it, which makes these types of investment products more attractive for potential investors. Prospering companies sometimes return parts of their profit to the shareholders by paying them *dividends*. These regular payments are often neglected for the basic financial modeling purposes and reintroduced later.

Another type of investment are *currencies* and *commodities*. Currencies usually enter the discussion when one is engaging in markets in multiple countries; then the foreign exchange rate and its dynamics become relevant. Commodities are goods, usually raw materials that are bought and sold in organized markets.

A.2 Derivatives

Derivative contracts (or *contingent claims*) are true to their name, for they derive their value from the price of some *underlying asset* (the *underlying*). Because the family of derivatives is rather large and there is almost no boundary to how complex a certain derivative contract can be, we shall mention only the few most important examples (for our purposes). The derivative we shall work with the most is an *option*. There are many types of options but the most liquid ones are of the following type. *Plain vanilla European call option* gives its buyer the right but not the obligation to buy the underlying asset at certain date in the future (called the *maturity*) for a certain specified price, called the *strike (price)* or the *exercise price*. If we denote by S_T the price of the underlying at maturity T and by K the strike, then the payoff for (plain vanilla) European call at maturity is exactly

$$(S_T - K)^+ := \max\{S_T - K, 0\}. \tag{A.1}$$

I.e. the buyer of this type of derivative contract profits from the rise in the price of the underlying and it exceeding the strike price. Symmetrically to the case of the call option, there is also a *European put option* with payoff

$$(K - S_T)^+ = \max\{K - S_T, 0\}, \tag{A.2}$$

i.e. the put option buyer profits from the underlying asset underperforming the strike. Other types of options are *American options* (puts and calls), which give the buyer the same rights as their European counterparts described above, but with the possibility of *early exercise*, i.e. the buyer may choose to exercise the American option at maturity or at some time prior to it. Naturally, deriving a price for an American option is more involved than doing so for the European one, because it requires additional optimization techniques.

The options that we have defined above are just the most common options and they are usually publicly traded in organized exchanges with many different underlyings including

[1]There are also non-publicly traded stocks, however markets for them are rather illiquid, thus they are difficult to model.

equity, bonds, foreign exchange and commodities. The derivative contracts were originally developed in commodity markets for risk hedging purposes.

Another example of derivative contracts are the variance and volatility swaps. At expiration, the *volatility swap* pays the notional amount N times the difference between the realized volatility of the underlying asset during the lifetime of the contract, σ_R, and some fixed agreed-upon delivery volatility, σ_K, so that:

$$V_T = N(\sigma_R - \sigma_K). \tag{A.3}$$

Here σ_R is exactly the realized volatility of equation (4.1). Similarly, a *variance swap* contract (as the name suggests) has terminal payoff:

$$V_T = N(\sigma_R^2 - \sigma_K^2). \tag{A.4}$$

We must note here that although the realized volatility is calculated as (4.1), in the continuous framework the quantity

$$\sigma_R^2 = \frac{1}{T}\int_0^T \sigma_s^2 \, ds \tag{A.5}$$

is being used instead.

A.3 Financial modeling

The discipline of quantitative finance concerns itself with pricing (or valuing) derivative contracts by the use of mathematical models. A typical model states the assumptions (usually simplifying, because of the complexity of the real-life situation), specifies the model for the underlying asset (usually a stochastic process) and then proceeds to derive the pricing function for the derivative contract in question. In the case of the European call option, assuming that we are currently at time $t \in [0, T]$ (where T is the maturity of the option) and that the current price of the underlying is S_t, then the price of this call is $C = C(t, S_t)$. Similarly for a European put at time t, the price is $P = P(t, S_t)$. Considering European call and put on the same underlying, both with maturity T and strike K, by some elementary arguments, one derives the *put-call parity*[2]:

$$C(t, S_t) - P(t, S_t) = S_t - Ke^{-r(T-t)}, \qquad \forall t \in [0, T], \tag{A.6}$$

which is a model independent relationship. Ergo, if we derive price of a call in an arbitrary model, we automatically obtain the prices for puts as a byproduct.

As we have already mentioned, there is quite a diversity as to what can serve as an underlying for an option[3]. Since these underlyings exhibit different dynamics, we expect different models corresponding to them. Nevertheless, the common approach of quantitative finance is to utilize the market observable information (incorporated in the quoted prices of vanilla contracts) as much as possible and then use a model as an extrapolation tool to obtain prices for illiquid and OTC (*over-the-counter*) traded contingent claims.

[2]Here r is a constant *risk-free interest rate* - the rate of interest on which can one borrow or lend money without any risk. In real life, there is always risk. However, for example, government bonds of politically stable countries are very close to this concept.

[3]There are even options written on other derivative contracts such as options on options.

A.4 Arbitrage

An *arbitrage* is a way of trading, so that one starts with zero capital and at some later time T is sure not to have lost money and furthermore, has a positive probability of having made money [12], i.e. an arbitrage is a portfolio value process $\{\Pi_t\}$ satisfying $\Pi_0 = 0$ and there exists $T > 0$ such that

$$P(\Pi_T \geq 0) \geq 0 \quad \text{and} \quad P(\Pi_T > 0) > 0.$$

A usual assumption of mathematical finance is the non-existence of arbitrage - the *principle of no arbitrage*. This assumption is used to find a fair value/price for a derivative contract. Obviously, in real life, there exist arbitrage opportunities, but they are usually short-lived. Usual economic argument on which the no-arbitrage principle rests is that if there was an arbitrage opportunity, market participants would promptly recognize that and act on it, so that the supply and demand eliminate this opportunity. Naturally, the more illiquid and anomalous the particular market is the more arbitrage opportunities are to be found in it. With that being said, no arbitrage principle is a sound assumption for deriving prices of derivative contracts and it is at the foundations of the whole field of quantitative finance.

A.5 Market models in mathematical finance

In mathematical finance, basically every notion that we have defined in the previous sections has a rigorous mathematical counterpart. In this brief section, we present only a few of them, that are used as language for formulation of some important theorems.

Definition A.5.1. A *contingent claim* is a square integrable, positive random variable on $(\Omega, \mathcal{A}, P)$.

We say, that contingent claim is *attainable* if there exists an *admissible* trading strategy[4] (the corresponding portfolio is bounded from below) that generates its payoff at maturity.

A market model is *complete* if every contingent claim is attainable.

In chapter 3 we assume a market model composed of one stock and the money-market account. The contingent claim is the derivative security. In the case of this contingent claim being a European call option, we have explicitly constructed an admissible trading strategy that replicates the option payoff.

It can be shown that if a market model has a risk neutral measure, then it does not admit arbitrage. Moreover, the market is complete if and only if this risk neutral measure is unique. These statements are basically the contents of the first and the second fundamental theorem of asset pricing [12].

A.6 Change of numéraire

A *numéraire* is the unit of account in which other assets are denominated. One usually takes the numéraire to be the currency of a country. For example, in section 3.3, because

[4]Furthermore, the trading strategy needs to satisfy the *self-financing condition*. That is a condition analogous to (3.5).

of the equality[5]

$$\frac{S_t}{B_t} = D_t S_t, \qquad \forall t \geq 0, \tag{A.7}$$

we have effectively denominated our asset S_t in terms of the money-market account $\{B_t\}$. By definition, the discounted stock price is a martingale (w.r.t. natural Brownian motion) under the risk neutral probability measure $\tilde{P}$. We say that the measure $\tilde{P}$ is risk-neutral for the money-market account numéraire.

In general, it is possible to take any positively priced asset (that does not pay dividends) as a numéraire. In practical applications it is often advantageous to change the numéraire, because it possibly simplifies the calculations.

Theorem A.6.1 (Change of numéraire)**.** Assume there exists a numéraire B and a probability measure $\tilde{P}$, equivalent to P, such that the price of an asset $S = \{S_t\}$ relative to B is a martingale under $\tilde{P}$. Let S be a numéraire. Then there exists a probability measure Q, equivalent to P, such that the price of any attainable contingent claim Y normalized by S is a martingale under Q, i.e.

$$\frac{Y_t}{S_t} = E^Q \left[\frac{Y_T}{S_T} \middle| \mathcal{F}_t \right], \qquad \forall t \in [0, T]. \tag{A.8}$$

Moreover, the Radon-Nikodým derivative defining the measure Q is given by

$$\left. \frac{dQ}{d\tilde{P}} \right|_t = \frac{S_T}{S_t} \frac{B_t}{B_T}. \tag{A.9}$$

[5]Here S_t is the process for underlying and B_t, resp D_t are the money-market account and stochastic discount factor respectively. For definition see appendix B.

Appendix B

Interest rate basics

In the appendix A we have defined various investment products, including bonds and the bank account. In this appendix, for the purposes of mathematical modeling, we give these concepts a more precise meaning. The class of fixed income instruments and derivatives is rather large, what we present here are only the basics. This brief overview of elementary interest rate modeling is almost entirely borrowed from the introductory chapter of [37].

B.1 Definitions and heuristics

The elementary idea on which most of the fixed income products rest is that the value of one unit of currency today is different from the value of one unit of currency tomorrow. We formulate this in the concept of discounting below.

Definition B.1.1. A *bank account (money-market account)* is the Itô stochastic process $\{B_t\}_{t\in[0,T]}$ defined by the following differential relation:

$$dB_t = r_t B_t dt, \qquad \forall t \in [0,T], \tag{B.1}$$

with the initial condition $B_0 = 1$. Here $\{r_t\}$ is a positive random process almost surely for all $t \in [0,T]$.

According to 2.30, the relation (B.1) has an equivalent integral representation

$$B_t = \exp\left\{\int_0^t r_s\, ds\right\}, \qquad \forall t \in [0,T]. \tag{B.2}$$

The process $\{r_t\}_{t\in[0,T]}$ figuring in the equations above is usually referred to as the *instantaneous spot rate* or briefly as *short rate* for the definition above expresses the fact that investing a unit amount of currency at time 0 yields the value in (B.2) at time t, thus r_t is the rate at which the money-market accrues. This can be seen from the following; for an arbitrarily small time interval $[t, t+\varepsilon]$, the relative increase in the money-market account is

$$\frac{B_t - B_{t+\varepsilon}}{B_t} = \frac{\exp\left\{\int_0^{t+\varepsilon} r_s ds\right\} - \exp\left\{\int_0^t r_s ds\right\}}{\exp\left\{\int_0^t r_s ds\right\}} = \exp\left\{\int_t^{t+\varepsilon} r_s ds\right\} - 1 \simeq r_t \varepsilon.$$

Discounting is derived by the usual time value of money argument. Assume for simplicity, that the short rate process is actually a deterministic function $r = r(t)$ and so is the money-market account $B = B(t)$.

A legitimate question to ask is, how many units of currency (let us denote this amount by a) do we have to deposit initially (at time $t = 0$) in the money-market account in order for us to have exactly one unit of currency at time $t = t_2$. If we formulate this mathematically, we aim to ensure

$$aB(t_2) = 1.$$

Apparently, we have to deposit $a = 1/B(t_2)$ units of currency, where $B(t_2)$ is known, since $B = B(t)$ is deterministic. Another question is, how is this amount perceived (to what amount is it equivalent) at some other point in time $t_1 \in [0, t_2]$. Because at time t_1 our initial deposit of $aB(0) = a$ is equivalent to $aB(t_1)$, we have

$$aB(t_1) = \frac{B(t_1)}{B(t_2)}$$

units of currency. Thus one unit of currency payable at time t_2, as seen from time t_1 is $B(t_1)/B(t_2)$. Iterating back to the initial assumption of a general stochastic short rate process and stochastic bank account $\{B_t\}$, this leads to the following.

Definition B.1.2. The *discount factor* $D(t_1, t_2)$ between two time instants t_1 and t_2 is the amount at time t_1 that is equivalent to one unit of currency payable at time t_2 and is given by:

$$D(t_1, t_2) := \frac{B_{t_1}}{B_{t_2}} = \exp\left\{-\int_{t_1}^{t_2} r_s \, ds\right\}. \tag{B.3}$$

The discount factor expressed this way is a random variable. We shall most commonly use the discounting whilst being at time $t_1 = 0$ and asking about a variable future time $t_2 = t$, then the discount factor is defined as:

$$D_t := \frac{1}{B_t} = \exp\left\{-\int_0^t r_s \, ds\right\}, \qquad t \in [0, T], \tag{B.4}$$

and it is actually an Itô stochastic process satisfying the differential relation

$$dD_t = -r_t D_t \, dt. \tag{B.5}$$

Appendix C

Characteristic function, related concepts

Given a random variable X on a probability space $(\Omega, \mathcal{A}, P)$, its *characteristic function* $\phi_X : \mathbb{R} \to \mathbb{C}$ is defined as:

$$\phi_X(u) := E[e^{iuX}] = \int_\Omega e^{iuX(\omega)}\, dP(\omega) = \int_{\mathbb{R}} e^{iux}\, dF_X(x) = \int_{\mathbb{R}} e^{iux} f_X(x)\, dx, \qquad \forall u \in \mathbb{R}. \tag{C.1}$$

Here F_X is the cumulative distribution function and f_X is the probability density function, assuming X is continuous. If the n-th moment of the random variable X exists, then its characteristic function is differentiable n times, and additionally the moment may be recovered from it as

$$EX^n = \frac{1}{i^n}\left[\frac{d^n}{du^n}\phi_X\right](u)\Big|_{u=0}. \tag{C.2}$$

In finance, the characteristic function is a very popular concept, since everything one sometimes needs for pricing are the moments of the model. Additionally, many interesting random variables can be analytically expressed by their characteristic function only (see 6.1).

C.1 Fourier transform, convolution

Definition C.1.1. Let f be a real integrable function, then its *Fourier transform* is given as

$$\mathcal{F}[f](u) = \int_{\mathbb{R}} e^{iux} f(x)\, dx, \qquad \forall u \in \mathbb{R}.$$

The Fourier transform might be continuously linearly extended to a bounded linear operator on square integrable functions, that is in fact a linear isometry (however we need not go into such detail).

Definition C.1.2. Let f, g be real integrable functions, then we define another function by

$$(f \star g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\, dy = \int_{\mathbb{R}} f(y)g(x-y)\, dy, \qquad \forall x \in \mathbb{R}.$$

This function is called a *convolution of functions f and g.*

Proposition C.1.3. Let f, g be two real integrable functions, then

$$\mathcal{F}[f \star g] = \mathcal{F}[f] \cdot \mathcal{F}[g] \tag{C.3}$$

Here the dot denotes the regular product of two functions. Instead of computing a convolution (which in some cases can turn out to be highly intricate), one can take Fourier transform of individual functions, simply multiply them and then take inverse Fourier transform.

This theoretical construct has many other applications but it is especially useful in probability theory, because there it can be shown, that for two independent, absolutely continuous random variables X, Y with probability density functions f_X, f_Y the r.v. $X+Y$ has PDF of the form

$$f_{X+Y} = f_X \star f_Y. \tag{C.4}$$

Recalling the definition of characteristic function ϕ_X of a r.v. X (equation (C.1)), we observe, that this is just the Fourier transform of its density (for continuous random variables). By proposition C.3 it follows, that for two absolutely continuous random variables the following holds:

$$\phi_{X+Y} = \mathcal{F}[f_X \star f_Y] = \mathcal{F}[f_X] \cdot \mathcal{F}[f_Y] = \phi_X \phi_Y \tag{C.5}$$

Appendix D

Elements of stochastic analysis

In this appendix, we shall briefly sketch some useful definitions, that we did not wanted to burden the reader with in the main body of the text. These definitions and concepts could have been included in the first chapter, however, they are mostly complementary to what is contained there and may be perceived as overly technical.

D.1 Equality of random processes

An example of the above mentioned technicalities are the nuances of how one understands the equality of random processes and on what basis can one identify them.

Definition D.1.1. Let $X = \{X_t\}$, $Y = \{Y_t\}$ be random processes, with the index set I. We say that:

1. Y is a *modification* of X if $X_t = Y_t$ P-almost surely for all $t \in I$, i.e.

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1, \qquad \forall t \in I.$$

2. X and Y are *indistinguishable* if

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega), \forall t \in I\}) = 1.$$

We note that there are processes X and Y which are modifications of each other, yet they are distinguishable (not indistinguishable).

Theorem D.1.2 (Kolmogorov-Chentsov)**.** Let $X = \{X_t\}_{t\in I}$ be a random process. If there are constants $\alpha, \beta, C > 0$ such that

$$E\left|X_t - X_s\right|^\alpha \leq C\left|t - s\right|^{\beta+1}, \qquad \forall s, t \in I,$$

then X has a modification with continuous trajectories.

D.2 Convergence of random variables

Definition D.2.1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a sequence of random variables on the same probability space $(\Omega, \mathcal{A}, P)$. Let X be another random variable. We define three types of convergence in probability theory:

1. The *sequence converges to X in probability* (denoted $X_n \xrightarrow{P} X$ or $P\text{-}\lim X_n = X$) if
$$\lim_{n\to\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0, \qquad \forall \varepsilon > 0.$$

2. The *sequence converges to X almost surely* (denoted $X_n \xrightarrow{\text{a.s.}} X$) if
$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

3. The *sequence converges to X in mean square* (denoted $X_n \xrightarrow{L^2} X$ or $L^2\text{-}\lim X_n = X$) if
$$\lim_{n\to\infty} E[X_n - X]^2 = 0.$$
This last limit is sometimes referred to as the *mean square limit*.

We note here, that both convergence in mean square and almost sure convergence imply convergence in probability.

D.3 Stochastic continuity

The continuity of a stochastic process in the time parameter is related to the usual convergence of random variables. We discuss it here, because it is included in the definition of a Lévy process.

Definition D.3.1. We say, that the *random process $\{X_t\}_{[0,T]}$ is stochastically continuous in $t_0 \in [0, T]$*, if
$$\lim_{t\to t_0} P(\{\omega \in \Omega : |X_t(\omega) - X_{t_0}(\omega)| > \varepsilon\}) = 0, \qquad \forall \varepsilon > 0.$$

The *process $\{X_t\}_{[0,T]}$ is stochastically continuous*, if it is stochastically continuous in every point in $[0, T]$.

Example D.3.2. Poisson process $\{N_t\}_{\mathbb{R}_0^+}$ with intensity $\lambda > 0$ as defined in section 1.4 is stochastically continuous. To verify this we can calculate:
$$P(N_{t+h} - N_t > \varepsilon) = P(N_{t+h} - N_t > 1) = 1 - P(N_{t+h} - N_t = 0) = 1 - e^{-\lambda h} \xrightarrow{h\to 0} 0$$
for all $\varepsilon > 0$. We have made use of Poisson jumps being either of size 1 or zero (no jumps). This example illustrates two facts. First, the stochastic continuity of a random process is unrelated to continuity of its trajectories (sample paths), as Poisson process is a pure jump process (with discontinuous trajectories), yet it is stochastically continuous. Second, Poisson process is a Lévy process, as defined in 6.2.1.

Example D.3.3. The stochastic continuity of Brownian motion might be verified by analogical steps to those above. We stress again, that this concept is unrelated to continuity of sample paths. Thus it would be erroneous to conclude that the Brownian motion is stochastically continuous because its trajectories are continuous.

D.4 Additional terms in stochastic analysis

This section is only informative; for extensive treatment of the matter, we point the reader towards [29] and [14].

Definition D.4.1. A stochastic process is said to be *càdlàg* if its trajectories are P-almost surely right continuous, and they have limits from the left[1].

Definition D.4.2. Given a filtration $\{\mathcal{F}_t\}$, random variable $T : \Omega \to \mathbb{R}_0^+$ is a *stopping time* if the event $\{T \leq t\}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$, i.e.

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

Definition D.4.3. Given a filtration $\{\mathcal{F}_t\}$, càdlàg process X adapted to it is called a *local martingale*, if there exists a sequence of increasing stopping times $\{T_n\}_{n\in\mathbb{N}}$ with $T_n \overset{\text{a.s.}}{\to} \infty$, such that the process $X_{\inf\{t,T_n\}}$ is a martingale with respect to $\{\mathcal{F}_t\}$ for each $n \in \mathbb{N}$.

Definition D.4.4. A stochastic process $X = \{X_t\}$ is a *semimartingale with respect to filtration* $\{\mathcal{F}_t\}$ if it admits the following decomposition:

$$X_t = X_0 + A_t + M_t, \qquad \forall t \geq 0,$$

where $A = \{A_t\}$ is a process of bounded variation and $M = \{M_t\}$ is a local martingale w.r.t. filtration $\{\mathcal{F}_t\}$.

[1]Here the seemingly nonsensical word càdlàg is actually an abbreviation for French "continu à droite, limité à gauche", which summarizes the content of the definition.

Appendix E

Bayesian inference and superstatistics

Bayesian statistics is a subfield of probability which takes into account subjective knowledge (and corresponding subjective probabilities) when making inference about certain systems. Prior and posterior probability distributions thus enter the process of verifying hypotheses. This is in contrast to classical frequentist statistics and there is an ongoing debate burdened with philosophical arguments on which approach is more suitable for certain statistical problems. [38] We shall not dive into these in this text, for we are going to utilize only one concept from Bayesian statistics.

E.1 Marginal density

We shall introduce an important concept in Bayesian approach to probability. This can be done generally (even for discrete random variables), but for our convenience we shall define it only for continuous random variables.

Definition E.1.1. Let X be a random variable with a probability density function $f_X(x|\theta)$ dependent on some parameter θ with values in the parameter space Θ. Assume that θ is a random variable Y, distributed according to some probability density function $f_Y(\theta)$. Then the quantity

$$f_X(x) = \int_\Theta f_X(x|\theta) f_Y(\theta)\, d\theta \tag{E.1}$$

is called the *marginal density* of the random variable X. The values of x are taken from the support of $f_X(x|\theta)$.

The integration in relation E.1 is over all the values that the parameter $\theta \in \Theta$ assumes. This procedure is called *marginalization* and the parameter θ is referred to as the *nuisance parameter*. The choice of the random variable Y and in particular its density is the subjective prior information that is taken into account due to Bayesian statistics.

The marginal density function f_X can get rather intricate, so for the purposes of calculation of *marginal characteristic function* (defined as the Fourier transform of the

marginal density) we derive a useful formula. Starting from the definition, for all $u \in \mathbb{R}$:

$$\phi_X(u) = \int_{\mathbb{R}} e^{iux} f_X(x)\,dx = \int_{\mathbb{R}} e^{iux} \left(\int_\Theta f_X(x|\theta) f_Y(\theta)\,d\theta \right) dx = \\ = \int_\Theta \underbrace{\left(\int_{\mathbb{R}} e^{iux} f_X(x|\theta)\,dx \right)}_{=:\phi_X(u|\theta)} f_Y(\theta)\,d\theta,$$

so that we obtain the formula for marginal characteristic function as the conditional characteristic function from which the parameter θ has been marginalized:

$$\phi_X(u) = \int_\Theta \phi_X(u|\theta) f_Y(\theta)\,d\theta, \qquad \forall u \in \mathbb{R}. \tag{E.2}$$

E.2 Superstatistics

Superstatistics is a concept from non-equilibrium statistical physics, which is often employed to study hierarchical systems and emergent phenomena. The superstatistical approach was pioneered by [39], and since then, it was applied in many diverse fields of study such as biology, turbulence and also, as of late, time series analysis. The concept is superficially similar to that of marginalization and obtaining non-trivial distributions as mixtures of some simpler distributions with a random parameter via formulas analogical to those introduced in the previous section. The approach of superstatistics is not concerned with the above-mentioned philosophical disputes and instead imposes the following physically motivated conditions.

1. In the phenomena described by superstatistics, there are two well-separated timescales (in the language of the previous section, these would correspond to scales of the marginal density and the conditional density).

2. The fluctuations in the subsystems are adiabatic (the changes in the parameter θ are slow) so that there is enough time to approach local equilibria in the subsystems.

To our knowledge, authors of [40] were first to suggest potential connections between subordination and superstatistics, mentioned also in [41]. From out point of view, subordination uses precisely the marginal density (E.1), however, the physically motivated assumptions listed above are not needed. In fact, for the purposes of subordination in financial modeling (see for example the variance gamma process) these would be rather undesirable, since sudden (nonadiabatic) movements or jumps are exactly what one looks for in these types of models. The situation could be quite different in the time series analysis realm. [42]

Bibliography

[1] L. Bachelier, "Théorie de la spéculation," *Annales Scientifiques École Norm. Sup.*, vol. 17, pp. 21–86, 1900.

[2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[3] R. C. Merton, "Theory of rational option pricing," *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.

[4] S. L. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *The Review of Financial Studies*, vol. 6, pp. 327–343, 1993.

[5] P. Clark, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, vol. 41, no. 1, pp. 135–155, 1973.

[6] S. Bochner, *Harmonic Analysis and the Theory of Probability*. DOVER PUBN INC, 1960.

[7] P. Carr, H. Geman, D. Madan, and M. Yor, "Stochastic volatility for lévy processes," *Mathematical Finance*, vol. 13, pp. 345 – 382, 2003.

[8] T. Mikosch, *Elementary Stochastic Calculus with Finance in View*. World Scientific Publishing Company, 1998.

[9] J. Jacod and P. Protter, *Probability Essentials*. Springer, 2012.

[10] B. Maslowski, "Stochastic equations and stochastic methods in partial differential equations," in *Proceedings of Seminar in Differential Equations* (P. Drábek, ed.), vol. 1 of *1.*, 2006.

[11] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag GmbH, 2005.

[12] S. Shreve, *Stochastic Calculus for Finance II*. Springer New York, 2010.

[13] B. Øksendal, *Stochastic Differential Equations*. Springer-Verlag GmbH, 2014.

[14] P. Protter, *Stochastic Integration and Differential Equations*. Springer-Verlag GmbH, 2003.

[15] J. Cox, J. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," *Econometrica*, vol. 53, no. 2, pp. 385–407, 1985.

[16] F. D. Rouah, *The Heston Model and Its Extensions in MATLAB and C#*. JOHN WILEY & SONS INC, 2013.

[17] A. Hirsa, *Computational Methods in Finance*. CRC PR INC, 2012.

[18] P. Wilmott, *Paul Wilmott on Quantitative Finance 3 Volume Set (2nd Edition)*. Wiley, 2006.

[19] P. Wilmott, *Frequently Asked Questions in Quantitative Finance*. John Wiley and Sons Ltd, 2009.

[20] B. Dupire, "Pricing with a smile," *Risk*, 1994.

[21] E. Derman and I. Kani, "Riding on a smile," *Risk*, vol. 7, 1994.

[22] J. Gatheral, *The Volatility Surface: A Practitioner's Guide*. JOHN WILEY & SONS INC, 2006.

[23] D. Breeden, "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, vol. 7, pp. 265–296, 1979.

[24] D. Applebaum, *Lévy Processes and Stochastic Calculus*. CAMBRIDGE UNIV PR, 2009.

[25] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 2013.

[26] A. N. Shiryaev, *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific Pub Co Inc, 1999.

[27] P. Tankov, *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2003.

[28] J. Lukeš and J. Malý, *Measure and Integral*. Matfyzpress, 2013.

[29] A. N. Shiryaev and O. E. Barndorff-Nielsen, *Change Of Time And Change Of Measure*. World Scientific Publishing Co Pte Ltd, 2015.

[30] J. Bertoin, *Subordinators: Examples and Applications. In: Lectures on Probability Theory and Statistics*. Springer, Berlin, Heidelberg, 2004.

[31] O. Barndorff-Nielsen and N. Shephard, "Some recent developments in stochastic volatility modelling," *Economics Papers*, 2001.

[32] D. Madan, P. Carr, and E. Chang, "The variance gamma process and option pricing," *European Finance Review*, vol. 2, pp. 79–105, 1998.

[33] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*. Elsevier Science & Technology, 1981.

[34] A. Swishchuk, *Change of Time Methods in Quantitative Finance*. Springer-Verlag GmbH, 2016.

[35] E. Derman, M. B. Miller, and D. Park, *The Volatility Smile*. John Wiley & Sons Inc, 2016.

[36] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, 2011.

[37] D. Brigo and F. Mercurio, *Interest Rate Models*. Springer-Verlag GmbH, 2001.

[38] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer New York, 1993.

[39] C. Beck and E. Cohen, "Superstatistics," *Physica A: Statistical Mechanics and its Applications*, vol. 322, pp. 267–275, may 2003.

[40] A. V. Chechkin, F. Seno, R. Metzler, and I. M. Sokolov, "Brownian yet non-gaussian diffusion: From superstatistics to subordination of diffusing diffusivities," *Physical Review X*, vol. 7, apr 2017.

[41] P. Jizba, J. Korbel, H. Lavička, M. Prokš, V. Svoboda, and C. Beck, "Transitions between superstatistical regimes: Validity, breakdown and applications," *Physica A: Statistical Mechanics and its Applications*, vol. 493, pp. 29–46, 2018.

[42] D. Xu and C. Beck, "Transition from lognormal to chi-square superstatistics for financial time series,"

Index