Credit Risk Modeling under Jump Processes and under a Risk Measure-Based Approach

Ramin Okhrati

A Thesis
In the Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

July, 2011

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Ramin Okhrati**

Entitled: **Credit Risk Modeling under Jump Processes and under a Risk Measure-Based Approach**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Prof. K. Argheyd

_____ External Examiner
Prof. G. Gauthier

_____ External to Program
Dr. N. Gospodinov

_____ Examiner
Dr. C. Hyndman

_____ Examiner
Dr. X. Zhou

_____ Thesis Co-Supervisor
Prof. A. Balbás

_____ Thesis Co-Supervisor
Prof. J. Garrido

Approved by _____
Dr. J. Garrido, Graduate Program Director

July 25, 2011

_____
Dr. B. Lewis, Dean
Faculty of Arts and Science

ABSTRACT

Credit Risk Modeling under Jump Processes and under a Risk Measure-Based Approach

Ramin Okhrati, Ph.D.

Concordia University, 2011

Having a precise idea of how information is used is a key element in studying credit risk models. This concept plays an important role in structural and reduced form models and most recently in information based models. In this thesis the relationship between these models and the idea of information, more specifically through filtration expansions, is studied in depth. Special attention is given to the study of intensity processes under different types of filtration expansions.

Credit derivatives are path dependent financial products. Therefore their analysis is based on the history of the underlying risky process. If the underlying process is allowed to have jumps, then this analysis is more challenging. This explains why, normally, risk management techniques for these products assume that the underlying process is continuous, the derivative is path independent, or the probability measure is

risk neutral. In our model, in the context of a locally risk-minimization approach, the problems of pricing and hedging of defaultable claims are discussed without imposing any of the above assumptions.

The impact of risk measures in financial markets can no longer be ignored. Considering this, a methodological procedure based on risk measures is developed to gauge the credit quality of defaultable bonds in real bond markets. Through this process a new type of indicator is introduced that can be useful to detect inconsistencies in bond markets. This can be helpful in market integration applications.

Acknowledgments

My Ph.D thesis is a result of my long-run past education, and I will take this opportunity to mention a few names that somehow have had a contribution in developing my learning and knowledge.

First and foremost, I deeply thank my supervisors, Prof. José Garrido and Prof. Alejandro Balbás for their constant support and encouragements during my Ph.D program. Without their help, I would not be able to accomplish this job.

I owe my deepest gratitude to my high school teacher, Mr. Seyed Jalal Rabie, the best mathematics teacher I have ever had.

I would like to thank the faculty of the Department of Mathematical Sciences at Sharif University of Technology, especially Prof. Bijan Z. Zangeneh (for teaching me the fundamentals of Random Analysis), my late M.Sc. supervisor Dr. Hassan Nojumi (for helping me in my master thesis), Prof. Mohammad Mahdavi–Hezavehi and Prof. Saeed Akbari (for teaching me the basics of Mathematics and Algebra).

I am very grateful to the faculty and staff of the Department of Mathematics and Statistics at Concordia University, especially Dr. Arusharka Sen for being a great teacher and opening my eyes to the world of Statistics, Dr. Xiaowen Zhou for teach-

ing me Stochastic Processes, and also for his many helpful hints and tips during my Ph.D. program, Dr. Cody Hyndman and Dr. Patrice Gaillardetz for teaching me financial and actuarial courses.

I truly appreciate the many useful comments and suggestions from the examining committee, especially Prof. Geneviève Gauthier.

I am also thankful to my dear friends and colleagues at Concordia University that made my life here more joyful.

Finally I dedicate all my love and passion to my beloved family, especially my dear parents, for their patience and support over my entire life.

TABLE OF CONTENTS

Chapter Page

List of Figures

List of Tables

Introduction to Credit Risk

The 2008 bankruptcies of several major investment banks such as Lehman Brothers, have raised interest for credit risk models to the point of becoming an important part in the theory of finance. Credit risk is the risk associated with the possible financial losses of a derivative caused by unexpected changes in the credit quality of the counterparty's issuer to meet its obligations. The derivative can be a bond, a loan, an installment debt or even more complicated products depending on the counterparty's agrement that defines it.

Bonds form a common type of credit derivatives. Bonds that are issued by governments are usually considered risk free, but those issued by financial firms are more likely to be defaulted. One credit derivative that played an important role in the recent bankruptcies is the credit default swap (CDS). A major part of the losses of the giant insurance company, AIG, were caused by the CDS's it wrote on its own company and also on Lehman Brothers Holdings Inc..

In a CDS a protection seller insures a protection buyer against a well specified credit event and in return the protection buyers pay periodic payments (known as credit spreads) to the insurer. The credit event is not necessarily a bankruptcy and it

depends on how the counterparties define it. It can be a specific level that a firm's value crosses. Depending on the definition of the credit event, a CDS may be terminated before maturity. In that case the protection seller is required to make a promised payment (settled in cash), to deliver the debt owed (settled physically) or to realize other types of settlements. For investors who seek protection, CDS's are very useful and simple tools to transfer the risk from one party to the other. For a motivating economical explanation of these concepts we refer to the first four chapters of Wagner (2008).

In this thesis, we study credit risk by using two different approaches. In each one, because of the complexity in the theory, we focus on special credit derivatives. The extension to more complicated derivatives, like the above CDS's will be interesting for future work.

In the first approach, we use semimartingale theory to analyze the credit risk of defaultable claims. The motivation behind the work comes from a classical actuarial model that will be explained in Chapter 5. Since the historical behavior of asset prices resembles the sample paths of pure jump processes, special attention has been paid to study credit risk under jump processes. A pricing and hedging analysis is carried out for defaultable claims, which are simple types of CDS. Apart from credit risk, there are theoretical interesting results as well.

In the second approach, we use risk measures to study credit risk. In contrast to the first one, we do not use stochastic processes. The main ingredient of this approach

is random variables. Here defaultable bonds are credit sensitive derivatives analyzed using risk measures.

In the following chapters, these two approaches are explained and developed in detail. Chapters 1 to 5 discuss the first approach and Chapter 6 is devoted to the second one. Please note the different inputs and tools used in each of these two cases. In the first approach, we use stochastic processes and semimartingale theory, while random variables and optimization theory are applied for the second approach. Because of this, we are not able to directly link the two, though both approaches will be dealing with credit risk. In fact, the reader can read each one independently from the other one. The rest of this thesis is organized as follow.

In Chapter 2, structural models and reduced form models are reviewed. Information based models are explained in Chapter 3. In this chapter, the idea of information in credit risk models is developed. We try to bring most of the current models under one framework by considering three different types of filtration expansions. Also the properties of intensity and its connections to the empirical observations are explained.

In Chapter 4, it is assumed that the firm's value process is given by $S_t = u + \mu + B_t + \sum_{i=1}^{N_t} Y_i$, where B_t is a Brownian motion and $\sum_{i=1}^{N_t} Y_i$ is a compound Poisson process. In working with reduced form models under this process, the structure of the intensity is an important issue. We apply the method of Guo, Jarrow and Zeng (2009) and discuss the structure of the intensity associated with this process. Having the intensity in this case, then we discuss the intensity for a general Lévy process. Finally,

we will observe that in the presence of jumps the intensity cannot be zero, which will guarantee non-zero short spreads.

In Chapter 5, we focus on the hedging of defaultable claims. The locally risk-minimizing approach, a method for managing the risk of general contingent payoffs, is explained. Martingale techniques, in addition to the idea of intensity in reduced form models, are applied to analyze the structure of defaultable claims. Then we use the locally risk-minimizing approach to study the hedging of these defaultable claims.

Note that the theory of locally risk-minimizing is already applied to study defaultable markets, but in most of financial models the continuity of the underlying process is a crucial assumption. Our main contribution is to challenge this assumption in Chapter 5. Beside this, a few theoretical results are also obtained in this chapter.

In Section 5.8, we explain pricing tools and the estimation of the distribution of the default time. Although finding the distribution of the default time is not our main goal, some of our results can be applied to this purpose.

In Chapter 6, we explain the second approach to study credit risk that is using risk measures to study credit risk. However, we would like to mention that our work goes beyond that and analyzing credit risk will be only one of the outcomes. Especially, a new indicator based on risk measures is introduced and developed that can be used to measure inconsistencies in the bond market.

Chapter 7 gives some conclusions and discusses future work.

Chapter 1

Credit Risk and Stochastic Processes

In the context of stochastic processes, many tools and methods have been developed to model credit risk. Basically these models fall under two general categories; structural models and reduced form models. Both are revisited in the following sections. First, we review some basic notations.

1.1 Basic Notations

Throughout this thesis, suppose that uncertainty is modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω represents the states of the world and $\mathbb{P}$ is a probability measure on the σ-algebra $\mathcal{F}$ that determines the likelihood of its events. It is normally assumed that the probability space is equipped by a filtration $(\mathfrak{F}_t)_{0\leq t<\infty}$ which is an increasing sequence of σ-algebras such that $\mathfrak{F}_t \subset \mathcal{F}$ for all $0 \leq t < \infty$. A typical filtration $(\mathfrak{F}_t)_{0\leq t<\infty}$ is represented by $\mathfrak{F}$. The equipped probability space is shown by $(\Omega, \mathcal{F}, \mathfrak{F}, \mathbb{P})$. Depending on the context, different types of filtration are considered. However the σ-

algebra $\mathcal{F}$ is fixed and assumed to be rich enough to cover all the filtrations.

Without further assumptions, all the upcoming filtered probability spaces are assumed to satisfy the usual hypothesis (or usual conditions). This means that the following conditions are met

- The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete.
- For all $t \geq 0$, the σ-algebras $\mathfrak{F}_t$ contain all the sets in $\mathcal{F}$ of zero probability.
- The filtration $\mathfrak{F}$ is right continuous. That is, for every $t \geq 0$ the σ-algebra $\bigcap_{s>t} \mathfrak{F}_s$ is equal to $\mathfrak{F}_t$.

The evolution of a typical firm's asset value is given by a càdlàg stochastic process $X = (X_t)_{t\geq 0}$ that is adapted to some filtration. As we proceed, further notations will be introduced.

1.2 Introduction and Review of the Literature

The first paper that introduced credit risk goes back to the work of Merton (1974). His work is the foundation of structural models. In Merton (1974), default can happen only at the maturity time. Later this model was extended by Black and Cox (1976) that allows for default to happen at any time prior to the maturity. In this kind of model the default time of a loan is given by a stopping time; it is the first passage time of the firm's assets below a barrier. Economically, structural models are interesting because they give an interpretation for the default time, but their drawback is that they are not

consistent with the market observations. In these models it is assumed that the market value of the firm is observable which normally leads to the predictability of the default time, especially when this market value is modeled by a continuous process. It means that the investors would be aware of the probable default time in advance, leading to zero short credit spreads. Zero short credit spreads indicate no risk of default is contributed in a small period of time. It turns out that these short spreads are always non-zero on the market.

On the other hand, the reduced form models; pioneered by the works of Artzner and Delbaen (1995) or Jarrow and Turnbull (1995), use a different approach to model credit risk. In these models the probability of default is given by a relation in terms of either the so called intensity process or hazard process. Here the default time is totally inaccessible, which means that it is not predictable at all. Investors will never be aware of the default time.

By contrast with structural models, these reduced form models yield non-zero short spreads and also more useful formulas for pricing credit derivatives. So far, reduced form models have been divided into two categories; intensity based models and hazard process models. In intensity based models, the main focus is on the structure of the default indicator process that is $\mathfrak{N} = (\mathfrak{N}_t)_{t\geq 0}$ and $\mathfrak{N}_t = 1_{\{\mathcal{T}\leq t\}}$, where the default time $\mathcal{T}$ is a stopping time under a given filtration $\mathfrak{F} = (\mathfrak{F}_t)_{t\geq 0}$, which can be considered as the whole information available to investors on the market.

On the other hand, hazard process models are based on the conditional default

probability under a given filtration i.e. $\mathbb{P}(\mathcal{T} \leq t|\mathfrak{G}_t)$, where the filtration $\mathfrak{G} = (\mathfrak{G}_t)_{t\geq 0}$ can be viewed as a limited amount of information available to investors. More details on these two categories can be found in Jeanblanc and LeCam (2007).

While reduced form models are tractable, generally they do not use or determine a default model of the firm. It means that contrary to structural models, here it is not economically clear why default happens. On the other hand, although structural models are economically appealing, they do not provide explanations for empirical observations such as non-zero short spreads. Attempting to resolve these problems can lead to a link between the structural and reduced form approaches. These models have been studied in many directions, while keeping the same framework. Their drawbacks have been resolved in different ways. One of these ways is using the concept of information in credit risk models.

The basic idea, to fill the gap between these two major models, is to introduce different levels of information in the model. This new prospect has been proposed to try unifying these different credit risk models. The drawbacks of the structural models can be addressed through this new idea.

Perhaps among the first papers to introduce the concept of information in the model is that of Duffie and Lando (2001). They assume that investors just have periodic access to the accounting data, at deterministic times $t_1, t_2, t_3, ...$ By assuming a geometric Brownian motion they find that under full information $\mathfrak{F} = (\mathfrak{F}_t)_{t\geq 0}$, for the default time $\mathcal{T}$, and for all $t \geq 0$ we have $\lim_{h\downarrow 0} \frac{\mathbb{P}(t<\mathcal{T}\leq t+h|\mathfrak{F}_t)}{h} = 0$, while under the periodic

information $\mathfrak{G} = (\mathfrak{G}_t)_{t\geq 0}$, $\lim_{h\downarrow 0} \frac{\mathbb{P}(t<\mathcal{T}\leq t+h|\mathfrak{F}_t)}{h} = \lambda_t^i$ (i for intensity), where λ_t^i is almost surely non-zero. These results imply that under full information, credit spreads go to zero as one decreases the maturity time, while under imperfect information it is still non-zero.

Giesecke (2004a) introduces both structural and reduced form models comprehensively, and then tries to unify both models under incomplete information. He also investigates dependent default and calibration of data.

Giesecke and Goldberg (2004) introduce the I^2 model. In this model, the default time is defined as the first hitting time to a barrier. In their model investors are allowed to observe the firm's value evolution process $X = (X_t)_{t\geq 0}$, which is assumed to be continuous, but they cannot observe the firm's default barrier. They assume that this default barrier is a random variable with a given distribution.

Giesecke (2006) describes and categorizes different available types of information. In his paper it is shown that under complete information and a predictable default time, the short spreads are zero. Then he studies the case of incomplete information and obtains a pricing rule, though not a very explicit one. Also in this paper the relation between intensity and information is presented. Overall, the conclusion is as follows: with the continuity assumption of the firm's value evolution process, we can recover the structural models under complete information, and as the level of the information decreases the model changes to a reduced form one.

In the previous mentioned works, the analysis is done for a continuous firm's value

process. One of the first papers that studies credit risk with a jump process is Zhou (2001). He considers the process

$$\frac{dX_t}{X_t} = (\mu - \lambda v)dt + \sigma dB_t + (\Pi - 1)\, dN_t,\ t \geq 0,$$

where B is a standard Brownian motion, Π the jump amplitude with expected value $v+1$, and N is an homogeneous Poisson process with intensity λ. Under some assumptions he obtains a pricing rule for a bond, and also he shows that the short credit spreads are non-zero. In the presence of jumps the structure of credit risk models is more complicated.

In contributions that deal with the concept of information, traditionally there are two levels of information; the market information as the incomplete information and the manager's information as the complete information. In the presence of jumps and complete information one could still expect non-zero short spreads. It seems that in this case, the structure of the default indicator function, as well as the intensity (please be careful that this is not meant to be the intensity of the Poisson process), need more attention. Guo, Jarrow and Zeng (2009) directly study the structure of the intensity under different filtration assumptions including delayed and incomplete filtrations.

Another important issue in credit risk modeling which has not been studied extensively yet is the hedging of credit derivatives. A preliminary example in Jeanblanc and LeCam (2007) shows that even when working with simple continuous processes like Brownian motion, a complete hedge may still not be reached. Their example shows

that complete hedging is possible under the structural model, but under the reduced form model the market would be incomplete due to a delay in the information (incomplete information). In this case a mean-variance approach may be applicable. This example also highlights the fact that hedging in reduced form models should be done in an incomplete market.

Schweizer (1999) provides a good survey of hedging approaches in incomplete markets. In his article two quadratic hedging approaches, a locally risk-minimizing approach and a mean-variance hedging approach are introduced for the case where the firm's value process is a semimartingale.

Elouerkhaoui (2007) uses a simple structure of intensity in reduced form models and applies quadratic hedging approaches to obtain the hedging strategy.

Although we are interested in hedging using the structure of intensity based models, other methods are available; Cherubini and Luciano (2003) use copulas for pricing and hedging credit derivatives.

In our work, we mainly focus on the hedging of defaultable claims. We use the theory of local risk-minimization and assume that the underlying process is a bounded variation Lévy process, for which none of the above methods lead to a practical solution.

Our work studies a structural model, in the sense that we use the whole market information, represented by the filtration generated by the underlying process. However, we use an analysis like that of reduced form models and especially intensity based models. This is clarified more extensively in Chapters 2, 3, and 4.

Martingale techniques and the idea of intensity in reduced form models are applied to analyze the structure of defaultable claims. In Chapter 5, a Dynkin-type formula is obtained through our analysis for the defaultable processes $\left(g(t, X_t)1_{\{\tau>t\}}\right)_{t\geq 0}$, under certain conditions. This enables us to use compensator techniques for these types of processes.

Note that here the underlying process has jumps, the payoff is path dependent, and the probability measure is not necessarily a martingale measure. In addition we do not use any type of Girsanov's theorem.

We also study the structure of the default indicator process $\left(1_{\{\tau>t\}}\right)_{t\geq 0}$ and finite horizon ruin time in Section 5.8. For pricing in credit risk models, the distribution of the default time is needed under the risk neutral measure, i.e $\mathbb{P}(\mathcal{T} \leq t)$. In the above mentioned work, this has not been an issue of interest and whenever needed the tool is borrowed from other theories. This problem is solved numerically in some cases.

One important family of jump processes is that of Lévy processes. For spectrally negative Lévy processes, the Laplace transform of the default time (finite horizon ruin time), defined by $\tau^d = \inf\{t > 0; X_t < d\}$, for d a constant, is known, see for example Chiu and Yin (2005). Also, the Laplace transform of the distribution of the default time is known for jump diffusion processes with exponential jump size distributions, see Kou and Wang (2003).

In our work we would like to use general Lévy processes, technical problems force us to consider finite variation Lévy processes. Generalization of the results will be kept

for future work.

The credit derivatives that will be considered are defaultable claims of the form $H1_{\{\tau>T\}}$, where T is the maturity, $H \in L^2$ an $\mathfrak{F}_T$-measurable random variable, and $\tau = \tau^0 = \inf\{t > 0; X_t < 0\}$. For simplicity a single firm will be considered.

Chapter 2

Default Event and Credit Risk Models

2.1 Structural Models

In the previous section, we briefly explained the main idea and features of structural models. To understand these models, now we discuss Merton's Model in detail. This approach was pioneered by Merton (1974) and it was the foundation of credit risk structural models.

Consider the simple case where in addition to the asset X, the firm is also financed by a zero coupon bond with maturity of T that pays F units to bondholders (investors) if $X_T \geq F$, otherwise they receive X_T. Therefore it is assumed that in case of financial distresses, bondholders have absolute priority with respect to shareholders. Hence the payoff of the bond is $\min(X_T, F)$. This can be considered as a credit derivative with the above credit event defined at time T based on the firm's asset value. By a simple decomposition we have

$$\min(X_T, F) = F - \max(0, F - X_T).$$

Note that $\max(0, X_T - F)$ is the value of the equity at time T, because if $X_T \geq F$, the firm pays F amount to bond investors and equity holders receive $X_T - F$. If $X_T < F$ then the firm is bankrupt and the equity is worthless. Since bond holders have priority, they receive X_T and equity holders get nothing.

This product is analyzed by noting that the right-hand side of the above equation can be considered as the payoff of a portfolio composed of a risk-free bank account and a vanilla option. For example, under the absence of arbitrage assumption, the price of this bond is equal to the value of the portfolio at time zero. If X_t is a Brownian motion based process (for example a geometric Brownian motion), Black-Scholes formula can be used to price this product.

The above model can be extended in some ways. In Merton's model, the firm defaults only at time T. On the other hand, Black and Cox (1976) suggest that the firm may default at any time before T. This is modeled by defining the default time as

$$\tau^{\mathfrak{D}} = \inf\{t; X_t < \mathfrak{D}\}, \tag{2.1}$$

which is the first passage time of the asset process below the barrier $\mathfrak{D}$. In Definition (2.1), $\mathfrak{D}$ could be a constant, a random variable or even a random process. The barrier $\mathfrak{D}$ can be interpreted as the total amount of liabilities or a level of the firm's assets value for which the management of the firm decides to liquidate the assets if their value falls below $\mathfrak{D}$. In the case when $\mathfrak{D} = 0$, throughout this thesis we use the notation τ for $\tau^0 = \inf\{t; X_t < 0\}$.

In analyzing the above model, it is assumed that the asset process X and the barrier $\mathfrak{D}$ are observable by investors. Suppose that the filtration $\mathfrak{F}^{X,\mathfrak{D}} = \left(\mathfrak{F}_t^{X,\mathfrak{D}}\right)_{t\geq 0}$ is generated by X and $\mathfrak{D}$. This filtration could be viewed as the available information to the investors or modelers. Therefore, we are assuming that investors have complete access to the asset process and accounting data of the firm including liabilities $\mathfrak{D}$. This is possible only if the firm's accounting data are announced publicly or modelers are firm holders, which may not be true. Normally firm holders are not that willing to spread out information of the firm to other investors, and hence investors outside the firm have partial access to the firm's accounting data. Even beyond that, firm holders may be banned from investing in bond markets by insider's legislations. Therefore the asymmetry of information between management and investors must be considered.

Respecting the above filtration $\mathfrak{F}^{X,\mathfrak{D}}$, $\tau^{\mathfrak{D}}$ is a stopping time. It can be shown easily that if X is a continuous process then the stopping time $\tau^{\mathfrak{D}}$ is predictable (see Definition A.2 of the Appendix A.1). For example one can define $\tau^{\mathfrak{D}+\frac{1}{n}} := \inf\{t; X_t \leq \mathfrak{D} + \frac{1}{n}\}$, for $n \geq 1$. Then $\{\tau^{\mathfrak{D}+\frac{1}{n}}\}_{n=1}^{\infty}$ is an increasing sequence of stopping times such that for every n, $\tau^{\mathfrak{D}+\frac{1}{n}} < \tau^{\mathfrak{D}}$ on $\tau^{\mathfrak{D}} > 0$, and $\lim_{n\to\infty} \tau^{\mathfrak{D}+\frac{1}{n}} = \tau^{\mathfrak{D}}$, almost surely.

The predictability of the default time $\tau^{\mathfrak{D}}$ has some impact on credit risk modeling. To study these effects, we first define (short) credit spread.

Let $\mathcal{T}$ be a general stopping time (not necessarily in the form of (2.1)) with respect to a filtration $\mathfrak{F}$. Assume that the stopping time $\mathcal{T}$ models the credit event. Suppose that y_c is the yield at time t, $t < T$, and $t < \mathcal{T}$ on a credit risky, zero coupon bond

with a unit face value issued by a private corporation. Assuming that the recovery rate is zero, this means that y_c satisfies the following equation

$$e^{-\int_t^T r_s ds}\mathbb{P}[\mathcal{T} > T|\mathfrak{F}_t] = e^{-y_c(T-t)}, \tag{2.2}$$

where r_s in the above is the deterministic interest rate, and the probability measure $\mathbb{P}$ can be risk-neutral. Note that in case of a random interest rate and a non-zero recovery process R, the left hand-side of (2.2) should be replaced by the following

$$\mathbb{E}\Big[e^{-\int_t^T r_s ds}(1_{\{\tau>T\}} + R_\tau 1_{\{\tau\leq T\}})|\mathfrak{F}_t\Big].$$

Here it is assumed that R is zero and the interest rate is deterministic. An equivalent form of the formula (2.2) is

$$y_c - \frac{\int_t^T r_s ds}{T-t} = -\frac{\ln \mathbb{P}[\mathcal{T} > T|\mathfrak{F}_t]}{T-t}.$$

The second term on the left-hand side of the above equation is the yield at time t of a risk-free zero coupon bond with a unit face value (the risk-free bond can be considered as the one issued by government), and the term on the right-hand side is the difference between the risky yield and the risk-free one. This difference is called the credit spread of the risky bond at time t, and it is denoted by $S(t,T)$. In other words, the credit spread at any time $t < T$, $t < \mathcal{T}$ is defined as the excess yield demanded by the investors in order to be willing to buy a private corporate bond over a risk free one. Short credit spreads are defined as

$$\lim_{T\downarrow t} S(t,T) = \lim_{T\downarrow t} -\frac{\ln \mathbb{P}[\mathcal{T} > T|\mathfrak{F}_t]}{T-t}, \tag{2.3}$$

whenever this limit exists. Short credit spreads are interpreted the same as credit spreads but for very short periods of times. The following result of Giesecke (2006) implies that under the predictability assumption of the default time, short credit spreads are zero.

Proposition 2.1. *Assume that $\mathfrak{F}$ is any filtration representing the information available for investors. Let $\mathcal{T}$ be any $\mathfrak{F}$-predictable stopping time that represents the default time of the firm. Further suppose that for all $t < \mathcal{T}$, $\lim_{n\to\infty} \mathbb{P}[\mathcal{T} \leq t + 2^{-n}|\mathfrak{F}_t] = 0$. Then*

$$\lim_{n\to\infty} \sum_{t_i \in Z_n} S(t_i, 2^{-n}) 1_{\{t_i < t \leq t_{i+1}\}} = 0, \tag{2.4}$$

almost surely $\mathbb{P} \times Leb$, where $Z_n = \{k2^{-n}|k = 0, 1, ...\}$, $n \geq 1$, and Leb is the Lebesgue measure.

Remark 2.1. *The above proposition is almost the same as the original one in Giesecke (2006). We point out that the assumption $\lim_{n\to\infty} \mathbb{P}[\mathcal{T} \leq t+2^{-n}|\mathfrak{F}_t] = 0$ is redundant as it can be concluded by using Lebesgue's dominated convergence theorem for conditional expectation and the assumption that $t < \mathcal{T}$. Here we give a simpler proof using the definition of conditional expectation and Lebesgue's dominated convergence theorem. This type of proof can be used in other similar problems. The details are shown below. We start by assuming that $\mathcal{T}$ is a random time (not necessarily a stopping time, see A.1 for the definition of a random time). The definition of conditional expectation*

gives

$$\int_C \mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]\ d\mathbb{P} = \int_C 1_{\{\mathcal{T}\leq t+2^{-n}\}}\ d\mathbb{P},\ \textit{for all}\ C\ \textit{in}\ \mathfrak{F}_t. \tag{2.5}$$

Define $X_n = \mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]$, *then* $\{X_n\}_{n\geq 1}$ *is a positive decreasing sequence of* $\mathfrak{F}_t$*-measurable random variables. By adjusting the monotone convergence theorem we have*

$$\lim_{n\to\infty}\int_C \mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]\ d\mathbb{P} = \int_C \lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]\ d\mathbb{P}. \tag{2.6}$$

Now we take the limit of both sides of (2.5)*, and use the monotone convergence theorem one more time on the right-hand side. Then by comparing to* (2.6)*, we get*

$$\int_C \lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]\ d\mathbb{P} = \int_C \lim_{n\to\infty} 1_{\{\mathcal{T}\leq t+2^{-n}\}}\ d\mathbb{P} = \int_C 1_{\{\mathcal{T}\leq t\}}\ d\mathbb{P}.$$

Therefore by the definition of conditional expectation we obtain

$$\int_C \lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]\ d\mathbb{P} = \int_C \mathbb{E}[1_{\{\mathcal{T}\leq t\}}|\mathfrak{F}_t]\ d\mathbb{P}.$$

Since $\lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t]$ *is an* $\mathfrak{F}_t$*-measurable random variable, the uniqueness of the definition of conditional expectation gives* $\lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t] = \mathbb{E}[1_{\{\mathcal{T}\leq t\}}|\mathfrak{F}_t]$. *If* $\mathcal{T}$ *is an* $\mathfrak{F}$ *stopping time, then* $\lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t] = 1_{\{\mathcal{T}\leq t\}}$. *Hence for all* $t < \mathcal{T}$ *we get* $\lim_{n\to\infty}\mathbb{E}[1_{\{\mathcal{T}\leq t+2^{-n}\}}|\mathfrak{F}_t] = 0$.

The predictability of the default time leads to the following two conclusions.

First by the above proposition, the short credit spreads are zero. This means that on short periods of time corporate bonds behave like risk-free bonds. Hence in short periods of time, buyers of a credit risky bond should not require an excess yield over

the risk free yield. In the context of CDSs, that act as insurance contracts to protect buyers by paying a unit amount of currency in case of default, zero short credit spreads should imply that insurers do not demand for any premium in short terms, and insureds are protected for free.

Second if the default time of the company is predictable, by definition of predictability, there is an announcing sequence of stopping times and so investors will be aware of the upcoming default in advance. This causes a continuous convergence of bond prices to the default contingent one.

None of the above conclusions are empirically supported. For non-zero short credit spreads we refer to Sarig and Warga (1989), and for discontinuity of credit risky bond prices near the default time, please check Duffie, Pedersen and Singleton (2003). For more discussions and some numerical examples, we refer to Giesecke (2006).

The above problems can be addressed in two different directions.

In the first one that has been studied broadly and will be slightly reviewed in the following two sections, is related to an access of information. As mentioned earlier, structural models assume a symmetrical access of information between investors and the firm's management, this is to make the default time observable and hence predictable.

In the second direction, the problems are addressed by assuming that the evolution of the firm's asset value is modeled by a jump process, which is a reasonable assumption. Here, even under the assumption of having complete information on the asset

process, the default time is not necessarily predictable. As we argued, the short credit spreads can be non-zero as well. In our work, we focus more in this direction. Further discussions and references will follow in the upcoming sections.

Note that the definition of structural models can slightly vary from an article to another. There is not yet a unified agreement on what is called a structural model. What is introduced here can be called the classical structural approach.

2.2 Reduced Form Models

Reduced form models go back to Artzner and Delbaen (1995), and Duffie and Singleton (1999). These models, instead of focusing on the credit default, assume that the default time is given exogenously by a default rate (or a hazard process) which should be specified using a probability of default and market prices. For example, Jarrow and Turnbull (1995) assume that the default time is modeled by the first jump time of a Poisson process.

As it can be seen in this example, the main problem with these models is that they do not explain why firms default. They do not have the appealing default framework of structural models, but they can give more tractable pricing formulas and realistic results, such as non-zero short credit spreads. Although the hazard process approach is normally categorized under reduced form models, in this section, we mainly focus on the intensity based approach. Since different levels of information are considered in the hazard process approach, we believe it is more appropriate to study it in Chapter

3. Here is a thorough discussion of intensity based models. We start by defining the intensity.

2.2.1 Intensity: Definitions and Properties

Consider the indicator process $\mathfrak{N} = (\mathfrak{N}_t)_{t\geq 0}$, $\mathfrak{N}_t = 1_{\{\mathcal{T}\leq t\}}$, where $\mathcal{T}$ represents the default time of a firm. Assume that $\mathcal{T}$ is a stopping time with respect to a filtration $\mathfrak{F} = (\mathfrak{F}_t)_{t\geq 0}$, representing all the information available to investors. For example, it can be filtration $\mathfrak{F}^{X,\mathfrak{D}}$ considered in Section 2.1. Then by Doob-Meyer's decomposition, there exists a unique $\mathfrak{F}$-predictable increasing process $\Lambda = (\Lambda_t)_{t\geq 0}$ such that the process $(\mathfrak{N}_t - \Lambda_t)_{t\geq 0}$ is a uniformly integrable $\mathfrak{F}$-martingale . The process Λ is called the compensator of the process $\mathfrak{N}$. Notice that $\Lambda_t = \Lambda_{t\wedge\mathcal{T}}$ for all $t \geq 0$. This is because that by Theorem 18 in Protter (2004), the process $(\mathfrak{N}_{t\wedge\mathcal{T}} - \Lambda_{t\wedge\mathcal{T}})_{t\geq 0}$ is also a uniformly integrable $\mathfrak{F}$-martingale. Since the process $\mathfrak{N}$ is stopped at $\mathcal{T}$ (which means $\mathfrak{N}_t = \mathfrak{N}_{t\wedge\mathcal{T}}$, for all $t \geq 0$), the uniqueness of Doob-Meyer's decomposition implies that $\Lambda_t = \Lambda_{t\wedge\mathcal{T}}$ for all $t \geq 0$.

An interesting relation exists between the stopping time $\mathcal{T}$ and Λ_t.

Theorem 2.1. *Assume that $\mathfrak{N}$, $\mathcal{T}$, and $\mathfrak{F}$ are as above. Then $\mathcal{T}$ is a totally inaccessible stooping time (or informally a complete surprise) if and only if Λ is almost surely a continuous process.*

The "if" part of this theorem can be proved by a simple contradiction argument, together with Doob's optional sampling theorem. For a proof of the "only if part",

we refer to Theorem 20, Chapter III of Protter (2004).

For a precise definition of a totally inaccessible scoping time see Definition A.3. Totally inaccessible stopping times are useful tools to model the timing of unpredictable events. Since default times of firms are usually surprising events, it is normal to model them by totally inaccessible stopping times.

In an intensity based approach, it is assumed that the compensator Λ is almost surely absolutely continuous with respect to the Lebesgue measure. Hence under this assumption, we have $\Lambda_t = \int_0^t \lambda_u^i du$ where the process $\lambda^i = (\lambda_t^i)_{t\geq 0}$ is the Radon-Nikodym derivative $\left(\frac{d\Lambda_t}{dt}\right)_{t\geq 0}$ and can be interpreted as the instantaneous likelihood of default and the letter "i" stands for the word "intensity". The reason behind this interpretation is due to the Laplacian approximation method that will be explored in Chapter 4. We call λ^i the intensity of the default model or simply just intensity.

In what follows, two notions of intensity in a general framework are defined. In these definitions two levels of information (filtration) are provided. In the next section, where we work with different levels of information, these general definitions are needed. However, in the context of the intensity based approach, the notion of intensity is a special case of the following definition. This is explained in the second point of the next remark.

Definition 2.1. *Assume that $\mathcal{T}$ is a random time that is not necessarily a stopping time with respect to a filtration $\mathfrak{G} = (\mathfrak{G}_t)_{t\geq 0}$, where $\mathfrak{G}$ is the sub-filtration of $\mathfrak{F}$ and $\mathcal{T}$ is a stopping time with respect to $\mathfrak{F}$. Assume that there exists a bounded, nonnegative*

$\mathfrak{G}$-*adapted process* $\lambda^i = (\lambda^i_t)_{t\geq 0}$ *such that* $(\mathfrak{N}_t - \Lambda_t)_{t\geq 0}$ *is an* $\mathfrak{F}$-*martingale, where* $\mathfrak{N}_t = 1_{\{\mathcal{T}\leq t\}}$ *and*

$$\Lambda_t = \int_0^t \lambda^i_s ds,\ t \geq 0.$$

Then the process λ^i *is called the intensity of the default model* $(\mathcal{T}, \mathfrak{G})$.

Remark 2.2. *Regarding this definition, the following points are worth noting:*

1. *The above definition is borrowed from Giesecke (2006) with a minor modification. Giesecke's* λ^i *is assumed to be* $\mathfrak{G}$*-predictable. The main reason for this change is explained in the remark following Proposition 3.2.*

2. *In intensity based models, explained in this section, it is assumed that there is only one filtration in the model. Therefore the above definition must be applied when* $\mathfrak{G} = \mathfrak{F}$. *In the general case, we have an information based model and to make* $\mathcal{T}$ *a stopping time with respect to the filtration* $\mathfrak{F}$, *the filtration* $\mathfrak{G}$ *must be extended appropriately. In this context, the filtration* $\mathfrak{G}$ *is called the reference filtration.*

3. *Suppose that* $\mathfrak{F} = \mathfrak{G}$, *for example in intensity based models. Since* Λ *in the above definition is an* $\mathfrak{F}$*-predictable process, Proposition 3.13 of Jacod and Shiryaev (1987) implies that one may choose* λ^i *to be* $\mathfrak{F}$*-predictable as well. This is important in some applications, for other results or for the uniqueness of the intensity that we investigate later.*

4. *In most cases, the intensity λ^i is a $\mathfrak{G}$-adapted, càdlàg process, therefore to get a predictable version (in fact a càglàd process) one can take $\lambda^i_{t^-}$ as the intensity, where $\lambda^i_{t^-} = \lim_{s\to t^-} \lambda^i_s$. Notice that $\int_0^t \lambda^i_s ds = \int_0^t \lambda^i_{s^-} ds$. All the intensities in this thesis are càdlàg processes, so if we ever need a predictable version we do as above. Especially, this is important when we find the hedging strategies in Chapter 5, as these must be predictable. The proof of uniqueness for the intensity in the class of càglàd processes (if it exists) is very simple.*

5. *If λ^i is a $\mathfrak{G}$-predictable intensity, then it must vanish after the stopping time $\mathcal{T}$. The proof of this statement is fairly simple for a $\mathfrak{G}$-adapted càglàd (or càdlàg) process. For a predictable intensity, one can use a monotone class argument.*

6. *For all non-negative $\mathfrak{F}$-predictable processes X, we have that $\mathbb{E}[\int_0^\infty X_s \, d\mathfrak{N}_s] = \mathbb{E}[\int_0^\infty X_s \lambda^i_s \, ds]$. This is a direct application of Theorem 3.17 in Jacod and Shiryaev (1987), because $\mathfrak{N} - \Lambda$ is an $\mathfrak{F}$-martingale.*

7. *The intensity in Definition 2.1 is not unique, to see an example we refer to Theorem 10, Section 3, Chapter II of Brémaud (1981). If we force the $\mathfrak{G}$-predictability assumption on the intensity, the previous identity in point 6 above and the same lines of proof as Theorem 12, Section 4, Chapter II of Brémaud (1981) establish the uniqueness of the intensity.*

In Chapter 4, the existence and other properties of λ^i are explained. We remind the reader that we use the same terminology as the Poisson process. Depending on the

context, it will be clear whether we are referring to the intensity of the Poisson process. When there is a possibility of confusion, we write the terminology in full detail for the Poisson process. Therefore unless otherwise stated, the word "intensity" refers to the intensity in Definition 2.1.

Although in this section we use Definition 2.1 (when $\mathfrak{G} = \mathfrak{F}$) as the notion of intensity, there is another concept of intensity in the context of hazard processes. Since we are dealing with intensities now, it is a good time to mention it here. These concepts are applied in some credit risk models based on hazard processes. First we define the hazard process.

Definition 2.2. *For a reference filtration $\mathfrak{G}$ and a random time $\mathcal{T}$ as in Definition 2.1, the hazard process Γ is defined as $\Gamma_t = -\ln G_t$, assuming that $G_t = \mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)$ is non-zero for all $t \geq 0$.*

Remark 2.3. *In some literature, the process Γ is also called integrated hazard process.*

Since $G = (G_t)_{t\geq 0}$ is non-zero in the above definition, $\mathcal{T}$ can not be a stopping time with respect to $\mathfrak{G}$. This can be easily verified by noticing that if $\mathcal{T}$ is a stopping time with respect to $\mathfrak{G}$ then $G_t = \mathbb{E}[1_{\{\mathcal{T}>t\}}|\mathfrak{G}_t] = 1_{\{\mathcal{T}>t\}}$.

The intensity based hazard is defined as follows.

Definition 2.3. *Assume that $\mathcal{T}$ is a random time in a reference filtration $\mathfrak{G}$ and λ^h be a bounded, nonnegative $\mathfrak{G}$-adapted process. Then λ^h is called the intensity based hazard*

of the default model $(\mathcal{T}, \mathfrak{G})$ *if*

$$\Gamma_t = -\ln \mathbb{P}(\mathcal{T} > t | \mathfrak{G}_t) = \int_0^t \lambda_s^h ds,$$

almost surely for all $t \geq 0$. *The letter "h" stands for the word "hazard".*

In the above definition, the $\mathfrak{G}$-predictability of λ^h is linked to the $\mathfrak{G}$-predictability of Γ (and so the survival process G). If λ^h is $\mathfrak{G}$-predictable then by Proposition 3.5 of Jacod and Shiryaev (1987), Γ is also $\mathfrak{G}$-predictable. On the other hand, if Γ is a $\mathfrak{G}$-predictable process then by Proposition 3.13 of Jacod and Shiryaev (1987), one may choose λ^h to be $\mathfrak{G}$-predictable as well.

The uniqueness of the intensity based hazard is more complicated. Suppose that λ^h is a $\mathfrak{G}$-predictable intensity based hazard. Under an appropriate filtration expansion $\mathfrak{F}$ of $\mathfrak{G}$ and $\mathcal{T}$, it can be proved that on $\{\mathcal{T} > t\}$, λ^h is also an intensity in the sense of Definition 2.1, (see Proposition 3.1). Then based on point (7) of Remark 2.2, on $\{\mathcal{T} > t\}$, the intensity λ^h is uniquely determined.

In the literature, both Definitions 2.1 and 2.3 of intensity are used. To distinguish them, we call the second one intensity based hazard. In the next section, we see how these two types of intensity are related. In this section, and in the thesis, in general we mainly focus on Definition 2.1 when $\mathfrak{G} = \mathfrak{F}$. We end this subsection by providing some examples in order to understand the idea of the intensity in this sense of Definition 2.1.

Example 2.1. *Assume that* $N = (N_t)_{t \geq 0}$ *is a homogeneous Poisson process with a*

constant intensity λ and the first jump time $\mathcal{T}$. Let $\mathfrak{F}$ be the natural filtration generated by N. It is easy to check that the process $(N_t - \lambda t)_{t\geq 0}$ is an $\mathfrak{F}$-martingale. Now, using Doob's optional sampling theorem (or direct calculations) and some manipulations, we conclude that $(N_{t\wedge\mathcal{T}} - \lambda(t\wedge\mathcal{T}))_{t\geq 0}$ is a uniformly integrable $\mathfrak{F}$-martingale. Since $\mathcal{T}$ is the first jump time of N, $N_{\mathcal{T}} = 1$ and $N_t = 0$ for all $t < \mathcal{T}$. From the identity $N_{t\wedge\mathcal{T}} = N_t 1_{\{\mathcal{T}>t\}} + N_{\mathcal{T}} 1_{\{\mathcal{T}\leq t\}}$, we get $N_{t\wedge\mathcal{T}} = 1_{\{\mathcal{T}\leq t\}}$. So the process $\mathfrak{N} = (\mathfrak{N}_t)_{t\geq 0}$, $\mathfrak{N}_t = 1_{\{\mathcal{T}\leq t\}}$ is the homogeneous Poisson process stopped at its first jump time $\mathcal{T}$, and the process $\left(\mathfrak{N}_t - \int_0^t \lambda 1_{\{\mathcal{T}\geq s\}} ds\right)_{t\geq 0}$ is a uniformly integrable $\mathfrak{F}$-martingale. If a predictable intensity is required, since $\lambda 1_{\{\mathcal{T}\geq s\}}$ is an $\mathfrak{F}$-predictable process, the uniqueness of the intensity (see point 7 of Remark 2.2) gives that $\lambda^i_t = \lambda 1_{\{\mathcal{T}\geq t\}}$ for all $t \geq 0$. Notice that $\Lambda_t = \int_0^t \lambda 1_{\{\mathcal{T}\geq s\}} ds = \int_0^t \lambda 1_{\{\mathcal{T}>s\}} ds$, so $\lambda 1_{\{\mathcal{T}>s\}}$ is also an intensity, but it is not necessarily predictable. Then by using this martingale or just by noticing that $\mathcal{T}$ is the first jump time of N, we get the probability of default

$$\mathbb{P}(\mathcal{T} \leq T) = 1 - e^{-\lambda T} = 1 - e^{-\int_0^T \lambda\, du}.$$

The above argument can be easily extended to the following example.

Example 2.2. *Suppose that $N = (N_t)_{t\geq 0}$ is a non-homogeneous Poisson process with deterministic intensity $\lambda(t)$ and first jump time $\mathcal{T}$. Then similarly to the previous example, one can show that the indicator process $\mathfrak{N} = (\mathfrak{N}_t)_{t\geq 0}$ is a non-homogeneous Poisson process stopped at its first jump time $\mathcal{T}$, and the predictable intensity process λ^i is equal to $\left(\lambda(t) 1_{\{\mathcal{T}\geq t\}}\right)_{t\geq 0}$. It is easy to show that the probability of default is given*

by

$$\mathbb{P}(\mathcal{T} \leq T) = 1 - e^{-\int_0^T \lambda(u)\, du}.$$

By applying the law of iterated expectations, similar relations can be proved for a Cox process with the stochastic intensity λ^S *(S for stochastic). Then* λ^i *is equal to* $\left(\lambda_t^S 1_{\{\mathcal{T} \geq t\}}\right)_{t \geq 0}$*, and here the probability of default is given by*

$$\mathbb{P}(\mathcal{T} \leq T) = 1 - \mathbb{E}\left[e^{-\int_0^T \lambda_u^S\, du}\right].$$

2.2.2 Intensity Based Pricing Rules

As mentioned before, one of the features of reduced form models (especially the hazard process models) is their tractable pricing formulas for defaultable securities (claims). Roughly speaking it means that a defaultable security can be priced as a risk-free security just by adjusting the risk-free discount rate; see Duffie and Singleton (1999) or Giesecke (2004b) for discussions about this. Although this is mentioned in the literature frequently, we warn the reader that other than the simple cases like Example 2.1, in general the intensity based pricing approach leads to calculating an auxiliary jump process which is not easy to handle.

Let H be an $\mathfrak{F}_T$-measurable random variable, and to simplify the notation it is assumed that the interest rate is zero which in turn implies that the discount factor is one. Then the price of the defaultable claim $H1_{\{\mathcal{T}>T\}}$ is given by Duffie, Schroder and Skiadas (1996),

$$\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{F}_t] = 1_{\{\mathcal{T}>t\}} \left(W_t - \mathbb{E}[\Delta W_{\mathcal{T}} 1_{\{\mathcal{T}\leq T\}}|\mathfrak{F}_t]\right), \tag{2.7}$$

where $\Delta W_{\mathcal{T}} = W_{\mathcal{T}} - W_{\mathcal{T}^-}$ and

$$W_t = e^{-\Lambda_t}\mathbb{E}[He^{-\Lambda_T}|\mathfrak{F}_t].$$

The left-hand side of Equation (2.7) is the price of the defaultable security. Unfortunately, even in simple cases the auxiliary process W is not easy to handle and not even necessarily continuous at the default time $\mathcal{T}$. To see the calculations in a simple case when λ^i is a constant λ which means that $\mathcal{T}$ has the exponential distribution, we refer to Jeanblanc and LeCam (2007). In the special case when $\Delta W_{\mathcal{T}} = 0$, the price of the general defaultable claim $H1_{\{\mathcal{T}>T\}}$ at time t is equal to

$$\mathbb{E}[He^{\Lambda_t-\Lambda_T}|\mathfrak{F}_t].$$

If there is an intensity then the above formula reduces to

$$\mathbb{E}[He^{-\int_t^T \lambda_s^i ds}|\mathfrak{F}_t].$$

In the presence of a non-zero interest rate $r = (r_s)_{s\geq 0}$, the last formula gives

$$\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{F}_t] = \mathbb{E}[He^{-\int_t^T (r_s+\lambda_s^i)ds}|\mathfrak{F}_t]. \tag{2.8}$$

Therefore the price of this defaultable claim can be computed as a default free one, just by adjusting the discount factor r into an effective one $r + \lambda^i$. However one should keep in mind that this is true if the process W is continuous at the default time $\mathcal{T}$. This special case is treated in Giesecke (2006).

Remark 2.4. *Equation* (2.7) *is obtained under the physical measure* $\mathbb{P}$*. The proof of this result has nothing to do with the underlying probability measure. The absence of*

arbitrage guarantees the existence of a risk neutral probability measure. In financial modeling, a risk neutral probability measure is chosen to price derivatives and it has a large impact on pricing. However, in an incomplete market, this probability measure is not unique and it is selected based on the model and risk management criteria.

We shortly revisit the pricing rules (especially in the intensity based hazard approach) in the next chapter. Since in this thesis, our pricing method is different than those of intensity based pricing rules, we do not emphasize these approaches any further. For more details, we refer the reader to Jeanblanc and LeCam (2007) and references therein.

2.2.3 Credit Spreads in Intensity Based Models

Unlike structural models, in intensity based models the default time is a complete surprise for the investors (because of the continuity of the compensator Λ, see Theorem 2.1), hence it cannot be anticipated by a sequence of stopping times. This means that even for very short periods of time there is a chance of default and so non-zero short spreads are charged by the insurer for bearing that risk in short terms. This is consistent with empirical results.

In the intensity based models explained above, the absolute continuity of Λ with respect to the Lebesgue measure is the main assumption. Giesecke (2006) explains the case when the absolute continuity does not necessarily hold. Çetin, Umut, Jarrow, Protter and Yildirim (2002) give a model in which investors have access only to

information on the sign of the process. In their paper, a Brownian motion is used as an underlying process, and investors are informed if the balance sheet is negative or positive, but not necessarily of its absolute value.

Chapter 3

Information Based Models

Although the models mentioned in Sections 2.1 and 2.2 look quiet different, they have one thing in common, a unique flow of information is considered. Despite this, in information based models, different levels of information are considered. The main intuition behind information based models is that normally those who are closer to the management's board of a firm have more information than market investors (or bond investors).

3.1 Information Levels and Reference Filtration

Obviously for a financial product, different levels of information should produce different prices, and in fact as we see below this is also theoretically supported. For example in an extreme scenario, a security contract could be worthless in the view of managers who are aware of the upcoming default, but still in demand for market investors. In other words, bond investors might not have as much information as equity holders. Because of this, information based models some times are called incomplete information

models.

Mathematically, each level of information is modeled by a specific filtration. For example if investors are able to observe the firm's accounting data (or mathematically the underlying process X) and the credit event $\mathcal{T}$, but not the default barrier $\mathfrak{D}$, then an appropriate filtration is the one generated by $X = (X_t)_{0\leq t\leq T}$ and $\mathcal{T}$, where again T is the maturity of the derivative written on the underlying process. This model is considered in Giesecke and Goldberg (2004a).

The model of Duffie and Lando (2001) that was mentioned before is also an information based model since two different levels of information are considered. Furthermore, the models of Sections 2.1 and 2.2 can be represented as information based models. In fact, this is one of the ways of unifying classical structural and intensity based models. In terms of information, both models use the same level of information. What makes them different is how they model the credit event. In structural models, the default event is made endogenously through an economical definition of default, while in intensity based models it is given exogenously by an intensity. Note that the default time $\mathcal{T}$ of the firm can be a stopping time with respect to one level of information, while it is just a random time with respect to the other filtration. Giesecke (2006) explains different levels of information.

Information based models start by introducing a reference filtration $\mathfrak{G} = (\mathfrak{G}_t)_{t\geq 0}$ representing the market information available to market investors. The filtration $\mathfrak{G}$ excludes the default time $\mathcal{T}$, even if the later is also observable by market investors.

For example in Giesecke and Goldberg (2004a) mentioned before, the appropriate σ-algebra $\mathfrak{G}_t$ is generated by $(X_u)_{0\leq u\leq t}$. Assuming that investors can observe the default barrier $\mathfrak{D}$ and receive periodic noisy reports $(Re)_{t_k} = X_{t_k} + Y_{t_k}$, for $k \geq 1$, deterministic times $t_1 < t_2 < ...$, and independent noise random variables Y_{t_k}, yields the model of Duffie and Lando (2001). In this model, for $t_n \leq t < t_{n+1}$, the reference filtration $\mathfrak{G}$ is equal to

$$\mathfrak{G}_t = \sigma(Re_{t_1}, Re_{t_2}, ..., Re_{t_n}; 0 \leq s \leq t). \tag{3.1}$$

We call this the periodic-noisy filtration.

3.2 Filtration Expansions

As we already saw, depending on the model, the default time $\mathcal{T}$ is not necessarily a $\mathfrak{G}$-stopping time. Since we are analyzing a credit event represented by a default time $\mathcal{T}$, technically it is necessary to work in a filtration for which $\mathcal{T}$ is a stopping time. It is worth noting that this is not just a technical assumption because at the time of a bankruptcy or default, investors on the market will be informed. Therefore by considering $\mathfrak{G}$ as the appropriate filtration for investors, we automatically have assumed that they cannot observe the default event. Hence the investors filtration $\mathfrak{F} = (\mathfrak{F}_t)_{t\geq 0}$ must be an appropriate expansion of the reference filtration $\mathfrak{G}$ that makes $\mathcal{T}$ a stopping time.

In practice, making $\mathcal{T}$ a stopping time means that investors can be made aware of the default time. Because there is more than one way of expanding a filtration, this is

where different approaches based on different expansion methods come into play. First we explain these filtration expansions. There are three main methods of expanding the reference filtration $\mathfrak{G}$ in credit risk models. These are explained below.

The first type is called progressive filtration expansion and is defined as

$$\mathfrak{F}_t = \{B \in \mathfrak{F}_\infty;\ \text{for some } B_t \in \mathfrak{G}_t, B \cap \{t < \mathcal{T}\} = B_t \cap \{t < \mathcal{T}\}\}, \tag{3.2}$$

where $\mathfrak{F}_\infty = \mathfrak{G}_\infty \vee \sigma(\mathcal{T})$ and $\mathfrak{G}_\infty = \bigvee_{t=0}^{\infty} \mathfrak{G}_t$. An instant problem appears with the expanded filtration $\mathfrak{F}$. On the event $\{\mathcal{T} \leq t\}$ it must include the whole information of the reference filtration $\mathfrak{G}$. Mathematically we have the following

$$\mathfrak{G}_\infty \cap \{\mathcal{T} \leq t\} \subset \mathfrak{F}_t.$$

Obviously accepting this filtration expansion requires the progressive knowledge of the reference filtration for investors, at least partially (on the event $\{\mathcal{T} \leq t\}$) or completely, depending on how it is defined in the model.

The second type is called minimal filtration expansion and is defined very intuitively as the minimal expansion of the reference filtration $\mathfrak{G}$ that makes $\mathcal{T}$ a stopping time

$$\mathfrak{F}_t = \mathfrak{G}_t \vee \sigma(\{\mathcal{T} \leq s; s \leq t\}).$$

In contrast with the progressive filtration expansion, it does not require information up to infinity and makes $\mathcal{T}$ a stopping time. A very useful and simple observation is that the minimal filtration expansion is a subset of the progressive filtration expansion.

Now, after having introduced these two filtrations, we can give an answer to the following interesting question.

Remark 3.1. *In both Definitions 2.1 and 2.3, λ^i and λ^h are allowed to be $\mathfrak{G}$-predictable processes. But what changes if we let them be $\mathfrak{F}$-predictable? If λ^i and λ^h are $\mathfrak{G}$-predictable then obviously they are $\mathfrak{F}$-predictable also. Now assume that these two intensities are $\mathfrak{F}$-predictable and $\mathfrak{F}$ is any filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$, such that it is a subset of the progressive filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$. Since the minimal filtration expansion is a subset of the progressive filtration expansion, as a special case, the expanded filtration $\mathfrak{F}$ can be either one of the two. By Lemma 1 in Jeulin and Yor (1978), there are $\mathfrak{G}$-predictable processes $\tilde{\lambda}^i$ and $\tilde{\lambda}^h$, such that $\tilde{\lambda}^i_t 1_{\{\mathcal{T}\geq t\}} = \lambda^i_t 1_{\{\mathcal{T}\geq t\}}$ and $\tilde{\lambda}^h_t 1_{\{\mathcal{T}\geq t\}} = \lambda^h_t 1_{\{\mathcal{T}\geq t\}}$. Therefore under the above filtration expansion assumptions, λ^i in Definitions 2.1, can be considered as a $\mathfrak{G}$-predictable process and in Definition 2.3, when $\mathcal{T} > t$ the same conclusion can be made for λ^h.*

The third type of filtration expansion used in credit risk modeling is slightly newer and more general. It was introduced by Guo and Zeng (2008). This includes the minimal and progressive filtration expansions as special cases. It is defined as any filtration $\mathfrak{F}$ that satisfies the following

$$\mathfrak{G}_t \cap \{t < \mathcal{T}\} = \mathfrak{F}_t \cap \{t < \mathcal{T}\}, \text{ for all } t \geq 0. \tag{3.3}$$

To understand the role of these filtration expansions, we explain some models related to each. We start by the progressive filtration expansion.

3.3 Progressive Filtration Expansion

Giesecke (2006) uses the progressive filtration expansion to study different levels of information. Some of his results are general and do not rely on a specific filtration expansion. Under the progressive filtration expansion, he obtains a fairly general pricing rule. This is the same formula that was obtained in Equation (2.8). Also he gives a good interpretation of intensity based hazards. Proposition 5.10 in Giesecke (2006) shows that if the intensities based hazards are right continuous then they are in fact the short credit spreads (2.3). We will explain this in more detail at the end of this section.

A good question to bring up at this point is what is the reason to use a progressive filtration expansion while we know realistically it is not true? The reason behind this choice is a technical result due to Jeulin and Yor (1978) (see Theorem A.1 in the Appendix) that provides a compensator of the process $\mathfrak{N} = (\mathfrak{N}_t)_{t\geq 0}$, $\mathfrak{N}_t = 1_{\{\mathcal{T}\leq t\}}$ in the progressive filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$. However, a recent corollary of Guo and Zeng (2008) shows that the consistency of the compensator holds under different filtration expansions. This means that the compensator of $\mathfrak{N}$ is the same under the progressive and minimal filtration expansions. Regarding this extension of Jeulin-Yor's theorem, some of the results in Giesecke (2006) can be improved. Here, we mention one of them.

Under the progressive filtration expansion, Giesecke (2006) explains the relation

between the intensity based hazard and the intensity in Definition 2.1. In his proof he uses Jeulin-Yor's theorem. By using the extended version of Jeulin-Yor's theorem, and minor modifications of the proof of Proposition 5.8 in Giesecke (2006), we can get the following improved version.

Proposition 3.1. *Assume that $\mathfrak{F}$ is any filtration expansion of the reference filtration $\mathfrak{G}$ and $\mathcal{T}$ that satisfies* (3.3) *(this includes the minimal and progressive filtration expansions as special cases). If λ^h is an intensity based hazard, in Definition 2.3, which is càdlàg, càglàd, or $\mathfrak{G}$-predictable then $\lambda^i_t = \lambda^h_t 1_{\{\mathcal{T}>t\}}$, $t \geq 0$, is an intensity in the sense of Definition 2.1. Moreover, if λ^h is $\mathfrak{G}$-predictable then the predictable version of λ^i, i.e. $\lambda^i_t = \lambda^h_t 1_{\{\mathcal{T}\geq t\}}$ is unique and $\mathfrak{G}$-predictable. On the other hand, if λ^i is the intensity in the sense of Definition 2.1 which is càdlàg, càglàd, or $\mathfrak{G}$-predictable, and if the compensator of Γ is given by $1-\Gamma$, then on $\mathcal{T} > t$, λ^i is also an intensity based hazard, i.e. $1_{\{\mathcal{T}>t\}}\mathbb{P}(\mathcal{T} > t|\mathfrak{G}) = 1_{\{\mathcal{T}>t\}} e^{-\int_t^t \lambda^i_s\, ds}$, $t \geq 0$. Moreover, in this case, if λ^i is $\mathfrak{G}$-predictable then on the event $\{\mathcal{T} > t\}$, λ^h is unique and $\mathfrak{G}$-predictable.*

3.4 Minimal Filtration Expansion: The First Approach

Next we focus on the minimal filtration expansion. In this context, normally there are two approaches to analyze credit derivatives. The first approach investigates the existence of the intensity in the sense of Definition 2.1. Then this intensity is used to analyze credit derivatives that reduces to pricing risky assets.

Duffie and Lando (2001) use the periodic-noisy filtration (3.1) as the reference filtration and obtain an intensity used for pricing. In their model, $X_t = e^{Z_t}$, where $Z_t = Z_0 + mt + \sigma B_t$, and the periodic-noisy reports are produced by $(Re)_{t_k} = X_{t_k} + Y_{t_k}$, where Y_t is a Gaussian process independent of X. The default time $\mathcal{T}$ is considered to be $\tau^d = \inf\{t; X_t < d\}$ for some constant d. They use Laplacian approximation (that is explained in the next chapter) and calculate intensities λ^i,

$$\lambda_t^i = \frac{1}{2}\sigma^2 \frac{\partial f}{\partial x}(t, d),$$

where $f(t, .)$ is the conditional density of the conditional distribution of X_t given $\mathfrak{F}_t$ (the minimal expansion of $\mathfrak{G}_t$), and $\frac{\partial f}{\partial x}(t, d)$ is the partial derivative of $f(t, x)$ respect to x, evaluated at $x = d$. First, they had proved that this conditional distribution has a continuously differentiable conditional density.

Guo, Jarrow and Zeng (2009) introduce a delayed filtration. Roughly saying, a delayed filtration is a subset of $\mathfrak{F}^X_{t-\delta}$, for some $\delta > 0$ and $\mathfrak{F}^X_t$ is the natural filtration generated by the underlying asset process X. Hence it is assumed that as time passes, eventually investors get full information. Then the minimal filtration expansion of this delayed filtration is used to obtain intensities of general one-dimensional, continuous Markov processes, regime-switching models and jump diffusion processes. As the structure of intensities and the exact definitions are complicated, for more explanations we refer the reader to the original paper.

Therefore there are appealing results that give closed forms of intensities in the

minimal filtration expansion. However, as it was explained in the previous section, pricing the credit risky products is not an easy task due to the existence of the auxiliary jump process in (2.7). So in terms of pricing rules based on (2.7), there may not be sufficient motivations to obtain intensities. But at the end of this section, we explain a real interpretation of intensities that makes intensities useful tools to study in credit risk.

Theorem 13 of Guo, Jarrow and Zeng (2009) discloses the significance of information in pricing credit risky assets. It shows how the level of information effects the price. Now we explain this beautiful result. Assume that we want to price the credit risky asset with the payoff $1_{\{\tau^D>T\}}$, where $\tau^D = \inf\{t > 0; X_t \in D\}$, X is a general and multi-dimensional Markov process under a risk-neutral measure $\mathbb{Q}$, $D \subset E$, and E is the state space of X. The process X represents the underlying asset process. Let $\mathfrak{F}_t^X = \sigma(X_u; 0 \leq u \leq t)$ be the natural filtration of X and $\{t_k\}_{k=0}^{\infty}$ a strictly increasing sequence of non-negative real numbers converges to infinity with $t_0 = 0$. This sequence is showing the periodic dates on which investors are informed about the firm's asset value. Therefore for $t_k \leq t < t_{k+1}$, the reference filtration $\mathfrak{G}_t$ is generated by $\{X_{t_1}, X_{t_2}, ..., X_{t_k}\}$. By assuming that investors are able to observe the default, the available information for investors is modeled by the minimal expansion of the reference filtration $\mathfrak{G}$ and the default time τ^D and it is denoted by $\mathfrak{F}$. Under these assumptions, we have the following theorem of Guo, Jarrow and Zeng (2009).

Theorem 3.1. *For $t_k \leq t < t_{k+1}$, we have*

$$\mathbb{E}^{\mathbb{Q}}[1_{\{\tau^D > T\}}|\mathfrak{F}_t] = \frac{\mathbb{E}^{\mathbb{Q}}[\tau^D > T|\mathfrak{F}^X_{t_k}]}{\mathbb{E}^{\mathbb{Q}}[\tau^D > t|\mathfrak{F}^X_{t_k}]} 1_{\{\tau^D > t\}}.$$

In the above theorem $\mathbb{E}^{\mathbb{Q}}[\tau^D > T|\mathfrak{F}_t]$ represents the price of the security at time t from the point of view of the investors, and $\mathbb{E}^{\mathbb{Q}}[\tau^D > T|\mathfrak{F}^X_t]$ is the price under full information. The interesting point is as $t_k \uparrow t$, the price of the security under the investor's (partial) information approaches the price of the security under management's (full) information. This makes sense because as t_k approaches t the investors information gets updated.

3.5 Minimal Filtration Expansion: The Second Approach

The second approach in the context of minimal filtration expansion is related to the hazard processes. Despite the first approach, here there is a more efficient pricing rule. Assume that H is a $\mathfrak{G}_T$-measurable random variable and $\mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)$ is non-zero for all $t \geq 0$. A direct application of Corollary A.1 in the Appendix A.2 gives

$$\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{F}_t] = 1_{\{\mathcal{T}>t\}} \frac{\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t]}{\mathbb{E}[1_{\{\mathcal{T}>t\}}|\mathfrak{G}_t]}. \tag{3.4}$$

Now we show that $\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t] = \mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]$. It is enough to prove that

$$\int_C \mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t]d\mathbb{P} = \int_C \mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]d\mathbb{P}, \quad \text{for all } C \in \mathfrak{G}_t.$$

From the definition of conditional expectation we have

$$\int_C \mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t]d\mathbb{P} = \int_C H1_{\{\mathcal{T}>T\}}d\mathbb{P}, \tag{3.5}$$

$$\int_C \mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]d\mathbb{P} = \int_C He^{-\Gamma_T}d\mathbb{P}. \tag{3.6}$$

On the other hand, by using the law of iterated expectations we get

$$\begin{aligned}
\int_C H1_{\{\mathcal{T}>T\}}\, d\mathbb{P} &= \mathbb{E}[H1_C 1_{\{\mathcal{T}>T\}}] \\
&= \mathbb{E}\Big[\mathbb{E}[H1_C 1_{\{\mathcal{T}>T\}}|\mathfrak{G}_T]\Big] \\
&= \mathbb{E}\Big[H1_C \mathbb{E}[1_{\{\mathcal{T}>T\}}|\mathfrak{G}_T]\Big] \\
&= \mathbb{E}[H1_C e^{-\Gamma_T}] = \int_C He^{-\Gamma_T}\, d\mathbb{P},
\end{aligned}$$

where we have used the definition of hazard process. Hence the left-hand side of Equations (3.5) and (3.6) must be equal

$$\int_C \mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t]d\mathbb{P} = \int_C \mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]d\mathbb{P}.$$

The uniqueness of the definition of conditional expectation implies that

$$\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{G}_t] = \mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]. \tag{3.7}$$

From Equations (3.4), (3.7) and the definition of hazard process we get the following rule for the price of the defaultable claim $H1_{\{\mathcal{T}>T\}}$,

$$\mathbb{E}[H1_{\{\mathcal{T}>T\}}|\mathfrak{F}_t] = 1_{\{\mathcal{T}>t\}}e^{\Gamma_t}\mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]. \tag{3.8}$$

As in Equation (2.7), please note that the above expectation is obtained under a physical measure. Now if we assume that H is a discounted payoff and the underlying probability measure is risk neutral, Equation (3.8) gives a rule to price this payoff.

Comparing to the pricing rule in (2.7), the pricing rule in (3.8) is more applicable in the sense that it does not involve an auxiliary jump process. But in general, calculating the conditional expectation $\mathbb{E}[He^{-\Gamma_T}|\mathfrak{G}_t]$ is not an easy task.

However, to obtain intensities in this approach one must use additional assumptions, either on the hazard process or on the filtration enlargements. For example in Elliott, Jeanblanc and Yor (2000), it is assumed that Γ is continuous and increasing. Then an explicit form of the intensity in the sense of Definition 2.1 is obtained. To adapt their work to our setting, we let the reference filtration $\mathfrak{G}$ be $\mathfrak{F}^B$, where $\mathfrak{F}^B$ is the completed natural filtration of the Brownian motion B. Then assume that $\mathcal{T}$ is any random time that is not a stopping time with respect to $\mathfrak{G}$. If the hazard process Γ is continuous and increasing, then under the minimal filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$, they prove that

$$\lambda_t^i = 1_{\{\mathcal{T}>t\}}\Gamma_t^{'},$$

where $\Gamma_t^{'}$ is the derivative of Γ with respect to t. Notice that since in this case Γ is monotone, then it is differentiable almost everywhere.

The above assumptions on Γ are closely related to the so called $\mathcal{H}$-hypothesis, that states that martingales in the reference filtration remain also martingales in the extended one. When this hypothesis does not hold, the situation is even more complicated. In this case, the stronger assumption of the $\mathcal{H}^{'}$-hypothesis is needed that states that semimartingales in the reference filtration remain semimartingales in the extended one. For more details we refer to Jeanblanc and LeCam (2007).

3.6 Filtration Expansion: The Third Type

The third type of the filtration expansion in (3.3) is a more general one introduced by Guo and Zeng (2008) and includes all the minimal and progressive filtration expansions and most of the intensities mentioned above. Instead of assumptions on the hazard process and filtration enlargements, in this paper the notion of "local jumping filtration" is introduced. Under a few assumptions, they give a fairly general formula to calculate the intensity under a strong Markov process.

3.7 Interpretation of Intensities

Finally we close this section by giving a realistic interpretation of the intensities mentioned before. Obtaining intensities (either hazard based or in the sense of Definition 2.1) are not just useful for pricing matters. Since intuitively intensities are instantaneous likelihoods of default, it is not surprising that they can be used as measures for short credit spreads. Giesecke (2006) proves that in the case of progressive filtration expansions, the intensities based hazard are in fact short credit spreads. Hence, they can be interpreted as the excess yields demanded by the investors for holding the credit risky assets over the risk free ones in very short periods of time. In the proof, Theorem 14 in Protter (2004, Chapter VI) is used. Here, we improve this result for a more general type of filtration expansion including the progressive one. In our proof, we use Corollary A.1 in the Appendix A.2.

Proposition 3.2. *Assume that $\mathfrak{F}$ is any filtration expansion of the reference filtration $\mathfrak{G}$ and $\mathcal{T}$ such that $\mathfrak{F}$ is the subset of the progressive filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$. Suppose that λ^h is the intensity based hazard of the default model $(\mathcal{T}, \mathfrak{G})$ that is right continuous. Then for each $t < \mathcal{T}$, the limit, $\lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T}\leq T|\mathfrak{F}_t)}{T-t}$ exists and we have*

$$\lim_{T\downarrow t} S(t,T) = \lambda_t^h.$$

Proof. By applying Corollary A.1 and since $\mathcal{T}$ is an $\mathfrak{F}$-stopping time, we have,

$$1_{\{\mathcal{T}>t\}} \lim_{T\downarrow t} \mathbb{P}(\mathcal{T} \leq T|\mathfrak{F}_t) = \lim_{T\downarrow t} 1_{\{\mathcal{T}>t\}} \frac{\mathbb{P}(t < \mathcal{T} \leq T|\mathfrak{G}_t)}{\mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)}.$$

Notice that

$$\frac{\mathbb{P}(t < \mathcal{T} \leq T|\mathfrak{G}_t)}{\mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)} = 1 - \frac{\mathbb{P}(\mathcal{T} > T|\mathfrak{G}_t)}{\mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)},$$

and the definition of the intensity based hazard gives the following

$$\mathbb{P}(\mathcal{T} > T|\mathfrak{G}_t) = \mathbb{E}\left[\mathbb{P}(\mathcal{T} > T|\mathfrak{G}_T)|\mathfrak{G}_t\right] = \mathbb{E}[e^{-\int_0^T \lambda_s^h ds}|\mathfrak{G}_t].$$

Therefore we obtain

$$\lim_{T\downarrow t} 1_{\{\mathcal{T}>t\}} \frac{\mathbb{P}(t < \mathcal{T} \leq T|\mathfrak{G}_t)}{\mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t)} = \mathbb{E}[1 - e^{-\int_t^T \lambda_s^h ds}|\mathfrak{G}_t].$$

The rest of the proof is the same calculations as Giesecke (2006), to complete it we

mention them here. On $\{\mathcal{T} > t\}$ we have,

$$\lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T} \leq T|\mathfrak{F}_t)}{T-t} = \lim_{T\downarrow t} \frac{1}{T-t}\mathbb{E}\left[\int_t^T \lambda_s^h ds + o\Big(\int_t^T \lambda_s^h ds\Big)\Big|\mathfrak{G}_t\right] \tag{3.9}$$

$$= \lim_{T\downarrow t} \frac{1}{T-t}\mathbb{E}\left[\int_t^T \lambda_s^h ds|\mathfrak{G}_t\right] + \lim_{T\downarrow t} \frac{o(T-t)}{T-t} \tag{3.10}$$

$$= \lim_{T\downarrow t} \frac{1}{T-t}\mathbb{E}\left[\int_t^T \lambda_s^h ds|\mathfrak{G}_t\right] \tag{3.11}$$

$$= \mathbb{E}\left[\lim_{T\downarrow t} \frac{1}{T-t}\int_t^T \lambda_s^h ds|\mathfrak{G}_t\right] \tag{3.12}$$

$$= \mathbb{E}[\lambda_{t^+}^h|\mathfrak{G}_t] = \lambda_t^h.$$

Equation (3.9) is by Taylor's expansion. In Equations (3.10) and (3.11) the almost sure boundness of λ^h is used. Equation (3.12) is obtained by using the definition of conditional expectation and Lebesgue's dominated convergence theorem for conditional expectation. In the last line we use the assumptions that λ^h is right continuous and also adapted. Therefore for each $\mathcal{T} > t$, the limit $\lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T}\leq T|\mathfrak{F}_t)}{T-t}$ exists.

Finally by using Taylor's expansion of ln in (2.3), one can easily prove that (or see Proposition 3.1 in Giesecke (2006))

$$\lim_{T\downarrow t} S(t,T) = \lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T} \leq T|\mathfrak{F}_t)}{T-t},$$

and the result follows. □

Remark 3.2. *This generalizes Proposition 5.10 in Giesecke (2006). It also covers the minimal filtration expansion as it is a subset of the progressive filtration expansion.*

In Proposition 5.10 of Giesecke (2006) corresponding to the above proposition, λ^h is both right continuous and predictable. In fact, in Giesecke (2006) predictability is

assumed in the definition of intensity. If we further assume that λ^h has also left-hand side limits, Proposition 2.24 of Jacod and Shiryaev (1987) implies that $\Delta\lambda^h_{\mathcal{T}} = 0$, almost surely on $\{\mathcal{T} < \infty\}$ for all totally inaccessible stopping times $\mathcal{T}$. If the intensity is a continuous process, for example as Duffie and Lando (2001), then $\Delta\lambda^h = 0$, almost surely. But as it was already mentioned a few times, shocks in markets normally occur unexpectedly, which is best modeled by totally inaccessible stopping times. Therefore in simple words, if the left-hand side limits of λ^h exists, Proposition 5.10 in Giesecke (2006) states that short credit spreads are not sensitive with respect to unpredictable shocks (or simply jumps) in the markets. This is hard to accept. The existence of left-hand side limits for the intensity is a reasonable assumption, because it is a bounded process by the definition of intensity. So far, we have never met an intensity without left-hand limits, and all the intensities in this thesis are càdlàg and satisfy this condition.

A similar result holds for an intensity in the sense of Definition 2.1. We have the following proposition.

Proposition 3.3. *Assume that $\mathfrak{F}$ is any filtration expansion of the reference filtration $\mathfrak{G}$ and $\mathcal{T}$ such that Definition 2.1 is satisfied. Suppose that λ^i is the intensity in the sense of Definition 2.1 for the default model $(\mathcal{T}, \mathfrak{G})$. Also assume that the intensity is right continuous. Then for each $t < \mathcal{T}$, the limit, $\lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T}\leq T|\mathfrak{F}_t)}{T-t}$ exists and we have*

$$\lim_{T\downarrow t} S(t,T) = \lambda^i_t.$$

Proof. Since $\mathcal{T}$ is an $\mathfrak{F}$-stopping time, we have,

$$\begin{aligned}
1_{\{\mathcal{T}>t\}} \lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T} \leq T|\mathfrak{F}_t)}{T-t} &= \lim_{T\downarrow t} \frac{\mathbb{P}(t < \mathcal{T} \leq T|\mathfrak{F}_t)}{T-t} \\
&= \lim_{T\downarrow t} \frac{\mathbb{E}\left[\int_t^T \lambda_s^i ds|\mathfrak{F}_t\right]}{T-t} \\
&= \mathbb{E}\left[\lim_{T\downarrow t} \frac{\int_t^T \lambda_s^i ds}{T-t}|\mathfrak{F}_t\right] \\
&= \mathbb{E}[\lambda_{t^+}^i|\mathfrak{F}_t] = \lambda_t^i.
\end{aligned}$$

Where the second equation is due to the definition of compensator and the third equation is obtained by using Lebesgue's dominated convergence theorem for conditional expectations.

Now same as Proposition 3.2, since $\lim_{T\downarrow t} S(t,T) = \lim_{T\downarrow t} \frac{\mathbb{P}(\mathcal{T}\leq T|\mathfrak{F}_t)}{T-t}$, the result follows. □

The next section mainly investigates the existence and calculation of intensity in the sense of Definition 2.1 for a special type of jump processes.

Chapter 4

Intensity: Existence and Calculation

4.1 Introduction

As we saw in Chapters 2 and 3, the intensity is a key element in credit risk modeling, especially in the intensity based approach. We explained some properties of this process. However, two questions are yet unanswered. Does the intensity exist in any default model, and if yes, how to calculate it? Unfortunately the current answers to these questions are not very satisfactory.

Throughout this chapter it is assumed that the completed filtered probability space $(\Omega, \mathcal{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ satisfies the usual hypotheses. Regarding Chapter 3, here it is assumed that $\mathfrak{F} = \mathfrak{G}$. Hence, there is only one filtration in the model and full information is available. Also it is assumed that the default time $\mathcal{T}$ is a stopping time with respect to $\mathfrak{F}$. In this section, unless otherwise specified, the word "intensity" refers to the intensity based on Definition 2.1.

This section describes the methods to calculate the intensity and observing the

technical problems that are brought up in the calculation process.

By Theorem 2.1, the total inaccessibility of the stopping time $\mathcal{T}$ is a necessary condition for the existence of the intensity, but it is not sufficient. Giesecke (2006) gives an example where the stopping time $\mathcal{T}$ is totally inaccessible but there is no intensity. As far as we know, the most recent work to calculate intensity is given by Guo and Zeng (2008). Under some technical conditions, they provide a procedure to obtain the intensity. However, in this section, we discuss two different approaches based on the following formula

$$\lambda_t^i = \lim_{h \downarrow 0} \frac{\mathbb{P}(t < \mathcal{T} \leq t + h) | \mathfrak{F}_t)}{h}, \text{ for all } t \geq 0, \tag{4.1}$$

where the limit is taken pointwise almost surely. We will see that this formula is not always correct. Still, the reason to stick to this method is that is intuitive as it is the classical definition of the intensity in (4.1). We illustrate these approaches in the case when the underlying asset process follows a jump diffusion process.

4.2 Methods

Since calculating the intensity also leads to obtaining the compensator of the process $\mathfrak{N} = \left(1_{\{\mathcal{T} \leq t\}}\right)_{t \geq 0}$, any result to calculate the compensator of this process should be helpful. Perhaps the most well known one is Meyer's Laplacian approximation, see Meyer (1966) for the proof.

Theorem 4.1. *(Meyer's Laplacian Approximation) Let X be a potential of Class D*

and $X = M - \Lambda$ be its Doob-Meyer decomposition. Define

$$\Lambda_t^h = \frac{\int_0^t \mathbb{E}[X_s - X_{s+h}|\mathfrak{F}_s]\, ds}{h}. \tag{4.2}$$

Then for any stopping time $\mathcal{T}$, $\lim_{h\downarrow 0} \Lambda_{\mathcal{T}}^h \to \Lambda_{\mathcal{T}}$, where the convergence is taken in the sense of the weak topology $\sigma(L^1, L^\infty)$. Furthermore, if Λ is continuous, then the convergence is in L^1.

Please see Definitions A.6 and A.5 of Appendix A for Class D and potential processes. Some characteristics of Class D of supermartingales are explained in Section 5.6.

Remark 4.1. *Note that Theorem 4.1 cannot be directly used for the process $X = \mathfrak{N}$, because the process $\mathfrak{N}$ is not a potential. But, one can apply this theorem on the process $X = \left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$ which is a potential. Then by observing that $1_{\{\mathcal{T}\leq t\}} = 1 - 1_{\{\mathcal{T}>t\}}$, for all $t \geq 0$, the compensator of $\mathfrak{N}$ can be obtained. More precisely, if $\left(1_{\{\mathcal{T}>t\}} + \int_0^t \lambda_s^i\, ds\right)_{t\geq 0}$ is a martingale, then $\left(1_{\{\mathcal{T}\leq t\}} - \int_0^t \lambda_s^i\, ds\right)_{t\geq 0}$ is also a martingale. In other words, in this case the intensity of the process $\mathfrak{N}$ is equal to λ^i.*

For now assume that Theorem 4.1 is applicable for the process $\left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$. Then it seems that this theorem together with the intuitive definition of intensity in (4.1), give an easy solution to calculate the intensity. However, this leads to some problems.

The process $X = \left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$ is a potential of Class D and its Doob-Meyer decomposition exists. If the compensator Λ of $\left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$ is not continuous then the convergence in Theorem 4.1 is in the weak sense. Therefore Λ_t^h is weakly approaching

to Λ_t and so, even if $\lim_{h\downarrow 0} \frac{\mathbb{P}(t<\mathcal{T}\leq t+h)|\mathfrak{F}_t)}{h}$ exists, there is no guarantee that the equality $\int_0^t (\lim_{h\downarrow 0} \frac{\mathbb{P}(s<\mathcal{T}\leq s+h)|\mathfrak{F}_s)}{h})ds = \Lambda_t$ holds. A simple example for which $\lim_{h\downarrow 0} \frac{\mathbb{P}(t<\mathcal{T}\leq t+h)|\mathfrak{F}_t)}{h}$ exists, but this equality does not hold, is given by Guo, Jarrow and Zeng (2009).

Suppose that the compensator Λ is a continuous process, then the convergence in Theorem 4.1 is in the strong sense of L^1. Further assume that the limit in (4.1) exists pointwise almost surely, then by Lebesgue's dominated convergence theorem $\lim_{h\downarrow 0} \Lambda_t^h$ also exists pointwise almost surely, say $\Lambda_t^{'}$, then for any sequence h_n approaching 0 we have

$$\int |\Lambda_t^{'} - \Lambda_t| \, d\mathbb{P} = \int \lim_{n\to\infty} |\Lambda_t^{h_n} - \Lambda_t| \, d\mathbb{P} \leq \lim_{n\to\infty} \int |\Lambda_t^{h_n} - \Lambda_t| \, d\mathbb{P},$$

by Fatou's lemma. Since $\Lambda_t^{h_n}$ is approaching Λ_t in L^1, a simple implication shows that $\Lambda_t^{'} = \Lambda_t$ almost surely. One more time we apply Lebesgue's dominated convergence theorem to get

$$\Lambda_t = \lim_{h\downarrow 0} \Lambda_t^h = \int_0^t \lim_{h\downarrow 0} \frac{\mathbb{P}(s < \mathcal{T} \leq s+h|\mathfrak{F}_s)}{h} \, ds. \tag{4.3}$$

So in the case of the continuous compensator, the intuitive definition of intensity in (4.1) is correct. Just as a reminder, please note that by Theorem 2.1, the continuity of Λ is equivalent to the total inaccessibility of $\mathcal{T}$.

In summary, the above discussion shows that regardless of the continuity of the compensator, if the limit in Theorem 4.1 is in the strong sense of L^1, then the intuitive definition of intensity given by (4.1) is true. A critical question then is if in Meyer's Laplacian approximation, regardless of the continuity of Λ, we have strong convergence in L^1? The answer to this question is "no". A paper by Dellacherie and Doleans-Dade

(1970) is devoted to constructing a counterexample.

Finally, please notice that even if the intensity is given by Equation (4.1), substantial work still must be done to calculate the limit.

A parallel approach to Meyer's Laplacian approximation is given by the following result known as Aven's Theorem. It provides sufficient conditions under which the intensity is given by formula (4.1), see Aven (1985) for the proof.

Theorem 4.2. *(Aven) Let $(\mathcal{N}_t)_{t\geq 0}$ be a counting process, assuming that $\mathbb{E}[\mathcal{N}_t] < \infty$ for all t. Finally, let $\{h_n\}_{n\geq 1}$ be a sequence which decreases to zero, and for each n, let $(Y_n^t)_{t\geq 0}$ be a measurable version of the process $\left(\frac{\mathbb{E}\left[\mathcal{N}_{t+h_n} - \mathcal{N}_t | \mathfrak{F}_t\right]}{h_n}\right)_{t\geq 0}$. Assume that the following statements hold with $(\lambda_t^i)_{t\geq 0}$ and $(y_t)_{t\geq 0}$ being non-negative measurable processes:*

1. *For each $t \geq 0$, $\lim_{n\to\infty} Y_n^t = \lambda_t^i$, almost surely,*

2. *For each $t \geq 0$, there exists an $n_0 = n_0(t)$ such that for almost all ω*

$$|Y_s^n(\omega) - \lambda_s^i(\omega)| \leq y_s(\omega), \quad s \leq t, n \geq n_0,$$

3. *$\int_0^t y_s ds < \infty$, almost surely for $0 \leq t < \infty$.*

Then $\left(\mathcal{N}_t - \int_0^t \lambda_s^i ds\right)_{t\geq 0}$ is an $\mathfrak{F}_t$-martingale, i.e. the process $\left(\int_0^t \lambda_s^i ds\right)_{t\geq 0}$ is the compensator of $(\mathcal{N}_t)_{t\geq 0}$.

Applying the above theorem for the special counting process $\mathcal{N} = \left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$, gives a procedure to calculate the intensity λ_t^i based on the intuitive formula (4.1). In this

theorem, the main problem is finding the dominating non-negative measurable process $(y_s)_{s\geq 0}$.

As mentioned in the last section, non-zero short spreads are one of the main advantages of reduced form models, but these models suffer from the lack of an appropriate definition of the firm's default model. In what follows, by accepting the economical definition of default, we investigate the intensity of a Brownian motion with drift perturbed by a compound Poisson process in the context of intensity based models. The idea follows the same approach of Guo, Jarrow and Zeng (2009), based on formula (4.1).

Hereafter, whenever we say that "on the event A, $Expression_1 = Expression_2$", it means that

$$1_A \, Expression_1 = 1_A \, Expression_2.$$

Proposition 4.1. *Let X be a Brownian motion with drift perturbed by a compound Poisson process*

$$X_t = u + \mu t + B_t + \sum_{i=1}^{N_t} Y_i, \quad \textit{for all} \quad t \geq 0$$

where $N = (N_t)_{t\geq 0}$ is a homogenous Poisson process with intensity λ, B_t is a standard Brownian motion and the Y_i are i.i.d. with density f_Y on $\mathbb{R}$. Assume that $\mathfrak{F}^X$ is the natural filtration generated by X. Define the default time by

$$\tau = \inf\{t; X_t < 0\},$$

and let $\{T_i\}_{i\geq 1}$ be the arrival times of the Poisson process N. Then on the event

$\{\tau > t, T_n \leq t < T_{n+1}\}$, $\lim_{h\downarrow 0} \frac{\mathbb{P}[t<\tau\leq t+h|\mathfrak{F}_t^X]}{h}$, *is equal to*

$$\lambda \int_{-\infty}^{\infty} \Phi(-X_t - y) f_Y(y)\, dy, \tag{4.4}$$

where Φ *is the standard normal distribution function*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Proof. For each t, $T_n \leq t < T_{n+1}$, we define the σ-algebra

$$\mathfrak{G}_t = \sigma\left(\{T_1, ...T_n, Y_1, ..., Y_n, B_{T_1}, ...B_{T_n}\}, \{B_s; T_n \leq s \leq t\}\right).$$

We also let $A = \{\tau > t, T_n \leq t < T_{n+1}\}$. Since for any t, the equality $\mathfrak{G}_t \cap A = \mathfrak{F}_t^X \cap A$ holds, by Lemma A.3 the Appendix, on A we have

$$\mathbb{P}(\tau > t + h|\mathfrak{F}_t^X) = \frac{\mathbb{P}(\tau > t + h, T_n \leq t < T_{n+1}|\mathfrak{G}_t)}{\mathbb{P}(\tau > t, T_n \leq t < T_{n+1}|\mathfrak{G}_t)},$$

and

$$\begin{aligned} \mathbb{P}(t < \tau \leq t + h|\mathfrak{F}_t^X) &= 1 - \mathbb{P}(\tau > t + h|\mathfrak{F}_t^X) \\ &= 1 - \frac{\mathbb{P}(\tau > t + h, T_n \leq t < T_{n+1}|\mathfrak{G}_t)}{\mathbb{P}(\tau > t, T_n \leq t < T_{n+1}|\mathfrak{G}_t)}. \end{aligned} \tag{4.5}$$

The next step is to evaluate the numerator and denominator of the expression in Equation (4.5).

To do this, first we calculate the expression

$$\mathbb{P}(\tau > t, T_n \leq t < T_{n+1}|\mathfrak{G}_t).$$

It is easy to see that the event $\{\tau > t, T_n \leq t < T_{n+1}\}$ is equivalent to the intersection of the two events

$$A_1 = \bigcap_{j=0}^{n-1} \big\{ \inf \left(u + \mu s + B_s + Z_j\right)_{T_j \leq s < T_{j+1}} > 0 \big\},$$

and

$$A_2 = \big\{ \inf \left(u + \mu s + B_s + Z_n\right)_{T_n \leq s < t} > 0 \big\},$$

where $Z_j = \sum_{i=1}^{j} Y_i$ for any $j \geq 1$. The events A_2 and $\{T_n \leq t\}$ are $\mathfrak{G}_t$-measurable. Therefore, it turns out that

$$\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t) = \mathfrak{X}\ \mathbb{E}[A_1 \cap (T_{n+1} > t) | \mathfrak{G}_t], \tag{4.6}$$

where $\mathfrak{X} = 1_{A_2} 1_{\{T_n \leq t\}}$.

Notice that T_{n+1} is the sum of T_n and a $\mathfrak{G}_t$-independent random variable exponentially distributed with mean $\frac{1}{\lambda}$. Then by Lemma A.2 of the Appendix and some manipulations one can get

$$\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t) = \mathfrak{X}\ \mathbb{P}(A_1 | \mathfrak{G}_t)\ e^{-\lambda(t - T_n)}. \tag{4.7}$$

Let h be a positive real number, then

$$\frac{\mathbb{P}(\tau > t + h, T_n \leq t < T_{n+1} | \mathfrak{G}_t)}{\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t)} = \mathbb{I} + \mathbb{III} + \mathbb{IIII}, \tag{4.8}$$

where $\mathbb{I} = \frac{\mathbb{P}(\tau > t+h, T_n \leq t < t+h < T_{n+1} | \mathfrak{G}_t)}{\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t)}$, $\mathbb{III} = \frac{\mathbb{P}(\tau > t+h, T_n \leq t < T_{n+1} \leq t+h < T_{n+2} | \mathfrak{G}_t)}{\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t)}$, and $\mathbb{IIII} = \frac{\mathbb{P}[\tau > t+h, T_n \leq t < T_{n+1} < T_{n+2} \leq t+h | \mathfrak{G}_t]}{\mathbb{P}[\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t]}$. Now we determine each term separately.

Calculating $\mathbb{I}$: We need the numerator in $\mathbb{I}$, say $\mathbb{I}^* = \mathbb{P}(\tau > t+h, T_n \leq t < t+h < T_{n+1}|\mathfrak{G}_t)$. Let $A^* = \{T_n \leq t < t+h < T_{n+1}\}$, then we have

$$\{\tau > t+h\} \cap A^* = A_1 \cap A_2 \cap \{\inf_{t \leq s < t+h}(X_s) > 0\} \cap A^*.$$

Therefore $\mathbb{I}^*$ is equal to

$$\mathbb{I}^* = \mathfrak{X}\,\mathbb{P}\Big(A_1 \cap \{\inf_{t \leq s < t+h} X_s > 0\} \cap \{t+h < T_{n+1}\}|\mathfrak{G}_t\Big), \tag{4.9}$$

then by noticing that X_t is $\mathfrak{G}_t$-measurable, given $\mathfrak{G}_t$, the event inside the above conditional probability is equivalent to

$$\bigcap_{j=1}^{n-1}\{\inf_{T_j \leq s < T_{j+1}}(X_s - X_0) > -u\} \cap \{\inf_{t \leq s < t+h}(X_s - X_t) > -x_t\} \cap \{T > t+h-t_n\},$$

where t_n and x_t respectively stand for T_n and X_t given $\mathfrak{G}_t$, and T is a random variable exponentially distributed and independent of $\mathfrak{G}_t$ with mean $\frac{1}{\lambda}$. Notice that for $t \leq s < t+h$ and on the event A^*, we have $X_s - X_t = \mu(s-t) + B_s - B_t$. This is a process with continuous sample paths and hence in the second term of the above event s is allowed to be equal to $t+h$. Given $\mathfrak{G}_t$, these three events are independent of each other, and for $t \leq s < t+h$, $X_s - X_t$ is also independent of $\mathfrak{G}_t$. Hence, in Equation (4.9) the second and the third terms inside the conditional probability are independent of A_1 and $\mathfrak{G}_t$.

By applying Lemma A.2 of the Appendix, on the event A, the numerator $\mathbb{I}^*$ becomes

$$\mathfrak{X}\,\varphi\Big(h, \mu, -(u + \mu t + B_t + Z_n)\Big)e^{-\lambda(t+h-T_n)}, \tag{4.10}$$

where $\mathfrak{X} = 1_{A_2} 1_{\{T_n \leq t\}}$,

$$\varphi(a, b, c) = \mathbb{P}(\inf_{0 \leq s \leq a} (bs + W_s) > c), \tag{4.11}$$

and W is a standard Brownian motion independent of X. Since we are on the event A, this can be written as

$$\mathbb{I}^* = \mathfrak{X}\, \varphi(h, \mu, -X_t) e^{-\lambda(t+h-T_n)}. \tag{4.12}$$

From Equations (4.7) and (4.12), we obtain $\mathbb{I}$ on the event A,

$$\mathbb{I} = e^{-\lambda h} \varphi(h, \mu, -X_t). \tag{4.13}$$

Calculating $\mathbb{II}$: Now we need the numerator in $\mathbb{II}$, say

$$\mathbb{II}^* = \mathbb{P}\Big(\tau > t + h, T_n \leq t < T_{n+1} \leq t + h < T_{n+2} | \mathfrak{G}_t\Big).$$

To help writing formulas denote by A^{**} the following event

$$A^{**} = \{\tau > t + h, T_n \leq t < T_{n+1} \leq t + h < T_{n+2} | \mathfrak{G}_t\}.$$

Unfortunately calculating $\mathbb{II}^*$ is not as straightforward as $\mathbb{I}^*$. First, we start by conditioning on T_{n+1}, Y_{n+1}, and $B_{T_{n+1}}$, then

$$\mathbb{II}^* = \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}} \zeta\, \mathbb{P}(T_{n+1} \in dt_{n+1}, Y_{n+1} \in dy_{n+1}, B_{T_{n+1}} \in dw_{n+1} | \mathfrak{G}_t), \tag{4.14}$$

where ζ_h is equal to

$$\zeta_h = \mathbb{P}(A^{**} | \mathfrak{G}_t, T_{n+1} = t_{n+1}, Y_{n+1} = y_{n+1}, B_{T_{n+1}} = w_{n+1}).$$

The hard part is calculating ζ_h. Notice that on the event $\{T_n \leq t < T_{n+1} \leq t+h\}$ we have

$$\{\tau > t+h\} = A_1 \cap A_2 \cap \{\inf_{t \leq s < T_{n+1}} X_s > 0\} \cap \{\inf_{T_{n+1} \leq s < t+h} X_s > 0\},$$

using this identity, Fubini's theorem, Lemma A.2 of the Appendix properly and the independent increments property of the process, after some tedious calculations we get

$$\begin{aligned} \zeta_h = {} & \mathfrak{X}\mathbb{P}(A_1|\mathfrak{G}_t)1_{\{t_{n+1} \leq t+h\}} \\ & \times \varphi(t+h-t_{n+1}, \mu, -u-\mu t_{n+1} - w_{n+1} - Z_n - y_{n+1}) \\ & \times \varphi^*(t, t_{n+1}, \mu, -X_t, w_{n+1} - B_t) e^{-\lambda(t+h-t_{n+1})}, \end{aligned} \tag{4.15}$$

where the function φ is defined in (4.11) and the function φ^* is given by

$$\varphi^*(a,b,c,d,e) = \mathbb{P}(\inf_{a \leq v < b}(c(v-a) + W_v - W_a) > d \;\Big|\; W_b - W_a = e).$$

On the other hand, by similar but simpler calculations one can show that $\mathbb{P}(T_{n+1} \in dt_{n+1}, Y_{n+1} \in dy_{n+1}, B_{T_{n+1}} \in dw_{n+1}|\mathfrak{G}_t)$ is equal to

$$\lambda e^{-\lambda(t_{n+1}-T_n)} \Phi'(w_{n+1} - B_t + t - t_{n+1}) f_Y(y_{n+1})\, dw_{n+1}\, dy_{n+1}\, dt_{n+1}, \tag{4.16}$$

the function Φ' is the derivative of the standard normal distribution function. By using Equations (4.14), (4.15), and (4.16), we get the following form of $\mathbb{III}^*$:

$$\mathbb{III}^* = \mathfrak{X}\, \mathbb{P}(A_1|\mathfrak{G}_t)\, \lambda\, e^{-\lambda(t+h-T_n)} \int_t^{t+h} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta_h^* f_Y(y)\, dw\, dy\, ds,$$

where ζ_h^* is equal to

$$\varphi(t+h-s, \mu, -u-\mu s - w - Z_n - y)\varphi^*(t, s, \mu, -X_t, w - B_t)\Phi'(w - B_t + t - s).$$

Note that we have changed the integral variables. For example, instead of w_{n+1}, we have used w. Together with the denominator of $\mathbb{III}$, which is (4.7), on the event A, we get

$$\mathbb{III} = \lambda e^{-\lambda h} \int_t^{t+h} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta_h^* f_Y(y)\, dw\, dy\, ds, \tag{4.17}$$

Calculating $\mathbb{IIII}$: Finally we need the numerator in $\mathbb{IIII}$, say

$$\mathbb{IIII}^* = \mathbb{P}\Big(\tau > t+h, T_n \le t < T_{n+1} < T_{n+2} \le t+h | \mathfrak{G}_t\Big).$$

Similarly to the previous calculation, by conditioning on T_{n+1}, T_{n+2}, Y_{n+1},Y_{n+2}, $B_{T_{n+1}}$ and $B_{T_{n+2}}$, one can get an expression for $\mathbb{IIII}^*$. This expression is more complicated than in $\mathbb{III}$. It is not presented here because, as we see below, it does not affect the final result.

By (4.13) and l'Hopital's rule, on the event A, we have

$$\begin{aligned} \lim_{h\downarrow 0} \frac{1-\mathbb{I}}{h} &= \lim_{h\downarrow 0} \lambda e^{-\lambda h} \varphi(h, \mu, -X_t) - \lim_{h\downarrow 0} e^{-\lambda h} \frac{\partial \varphi}{\partial h}(h, \mu, -X_t) \\ &= \lambda 1_{\{X_t > 0\}} = \lambda \end{aligned} \tag{4.18}$$

where the second equality is obtained by $\varphi(0, b, c) = 1_{\{c<0\}}$ and $\lim_{h\downarrow 0} \frac{\partial \varphi}{\partial h}(h, \mu, -X_t) = 0$.

Then by (4.17) and l'Hopital's rule, on the event A, we have

$$\begin{aligned} \lim_{h\downarrow 0} \frac{\mathbb{III}}{h} = \lambda \Big(&\lim_{h\downarrow 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta_h^{**} f_Y(y)\, dw\, dy \\ &+ \lim_{h\downarrow 0} \int_t^{t+h} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial \zeta_h^*}{\partial h} f_Y(y)\, dw\, dy\, ds \Big), \end{aligned}$$

where ζ_h^{**} is equal to

$$\varphi(0, \mu, -u - \mu(t+h) - w - Z_n - y)\ \varphi^*(t, t+h, \mu, -X_t, w - B_t)\ \Phi'(w - B_t - h).$$

Note that $\varphi(0,b,c) = 1_{\{c<0\}}$, and on the event A we have

$$\lim_{h\downarrow 0} \varphi(0,\mu,-u-\mu(t+h)-w-Z_n-y) = 1_{\{w>-X_t-y+B_t\}},$$

and

$$\lim_{h\downarrow 0} \varphi^*(t,t+h,\mu,-X_t,w-B_t) = 1_{\{X_t>0\}} = 1.$$

Using these equations and some manipulations we get the following

$$\lim_{h\downarrow 0} \frac{\mathbb{III}}{h} = \lambda \left(1 - \int_{-\infty}^{\infty} \Phi(-X_t - y) f_Y(y)\, dy\right). \tag{4.19}$$

In calculating $\mathbb{III}^*$, we end up with six multiple integrals, by proper conditioning, the first integral is over $[t, t+h]$ (first by conditioning on T_{n+2}) and the second one is over $[t, t_{n+2}]$ (by conditioning a second time on T_{n+1}). Similarly to case $\mathbb{III}$, by applying the l'Hopital's rule and taking the limit one can show that

$$\lim_{h\downarrow 0} \frac{\mathbb{IIII}}{h} = 0. \tag{4.20}$$

Finally by (4.5), (4.8),(4.18), (4.19), and (4.20) on Λ we get

$$\lim_{h\downarrow 0} \frac{\mathbb{P}\big(t < \tau \le t+h | \mathfrak{F}_t\big)}{h} = \lambda \int_{-\infty}^{\infty} \Phi(-X_t - y) f_Y(y)\, dy.$$

. □

Corollary 4.1. *Let X be a compound Poisson process with drift*

$$X_t = u + \mu t + \sum_{i=1}^{N_t} Y_i, \quad \textit{for all} \quad t \ge 0$$

where $N = (N_t)_{t\geq 0}$ is a homogenous Poisson process with intensity λ and the Y_i are i.i.d. with an absolutely continuous distribution function F_Y on $\mathbb{R}$. Assume that $\mathfrak{F}^X$ is the natural filtration generated by X. Define the default time by

$$\tau = \inf\{t; X_t \leq 0\},$$

then on the event $\{\tau > t, T_n \leq t < T_{n+1}\}$, $\lim_{h\downarrow 0} \frac{\mathbb{P}\left(t<\tau\leq t+h|\mathfrak{F}_t^X\right)}{h}$, is equal to

$$\lambda\Big(F_Y(-X_t)\Big). \tag{4.21}$$

Proof. This can be proved by exactly the same steps as for Proposition 4.1 or else we can directly conclude just by noticing that for a normally distributed random variable with zero mean and zero variance, the distribution function Φ is the Heaviside step function $\Phi(x) = 1_{\{x\geq 0\}}$. □

Now we discuss whether formula (4.4) is an intensity. As explained before, if the stopping time τ is totally inaccessible, then formula (4.4) is an intensity in the sense of Definition 2.1 for the default model $(\tau, \mathfrak{F})$. However, the default can happen in two fashions.

The first one can be caused by a sudden jump of the underlying process X, while the second one is driven by a continuous crossing of the horizontal access. In Example 2.1, we proved that the first jump time of a Poisson process is a totally inaccessible stopping time. Since all the jump times of the process X coincide with the jump times of the Poisson process N, and because of the independent increments of the process, it

is expected that all the jump times of the process X are totally inaccessible stopping times.

On the other hand, in Section 2.1, it is proved that a continuous crossing of the horizontal access is a predictable stopping time. So it seems that τ has two parts, one predictable and one totally inaccessible. More precisely, we have the following beautiful theorem of Meyer, but before stating this theorem we define the following notation.

Definition 4.1. *Assume that in the probability space* $(\Omega, \mathcal{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$*, the random time* $\mathcal{T}$ *is a stopping time. Then for any* $A \in \mathfrak{F}_{\mathcal{T}}$*, we define*

$$\mathcal{T}_A(\omega) = \begin{cases} \mathcal{T}(\omega), & \text{if } \omega \in A, \\ \infty, & \text{if } \omega \notin A. \end{cases}$$

Under the assumptions of the above definition, it is easy to prove that $\mathcal{T}_A$ is a stopping time and $\mathcal{T} = \mathcal{T}_A \wedge \mathcal{T}_{A^c}$.

Theorem 4.3. *(Meyer's Previsibility Theorem) Let* X *be a (strong) Markov Feller process for the probability* $\mathbb{P}^v$*, where the distribution of* X_0 *is given by* v*, and with its natural completed filtration* $\mathfrak{F}^v$*. Let* $\mathcal{T}$ *be a stopping time with* $\mathbb{P}^v(\mathcal{T} > 0) = 1$*. Let* $A = \{\omega; X_{\mathcal{T}}(\omega) \neq X_{\mathcal{T}^-}(\omega) \text{ and } \mathcal{T}(\omega) < \infty\}$*. Then* $\mathcal{T} = \mathcal{T}_A \wedge \mathcal{T}_{A^c}$*, where the stopping times* $\mathcal{T}_A$ *and* T_{A^c} *are respectively totally inaccessible and predictable.*

Since Lévy processes are examples of Strong Markov Feller processes, Theorem 4.3 shows that all the jump times of a Lévy process are totally inaccessible stopping times. It also proves that the stopping time τ is not totally inaccessible, because the set A in

the above theorem does not cover the whole set Ω almost surely . Therefore we can not apply Theorem 4.1 to conclude that formula (4.4) is an intensity in the sense of Definition 2.1. What we can conclude is the following.

Proposition 4.2. *Let X be the following process*

$$X_t = u + \mu t + \sum_{i=1}^{N_t} Y_i, \quad \textit{for all} \quad t \geq 0$$

where N_t is a homogenous Poisson process with intensity λ, the Y_i are i.i.d. with an absolutely continuous distribution function F_Y on $\mathbb{R}$ and $\mu > 0$. Define the default time by

$$\tau = \inf\{t; X_t < 0\},$$

then the intensity of the default model $(\tau, \mathfrak{F}^X)$ in the sense of Definition 2.1 is equal to

$$\lambda_t^i = \lambda\Big(F_Y(-X_t)\Big)1_{\{\tau > t\}}. \tag{4.22}$$

Note that since $\mu > 0$, the process crosses the barrier only through a sudden jump and therefore τ is a totally inaccessible stopping time. We obtained Proposition 4.2 in our own way using the Laplacian approximation method, but a more general version of this proposition that includes pure jump Lévy processes is proved in Guo and Zeng (2008) by a completely different method using a compensation formula. We have not tried to further generalize the above proposition, though we believe that the generalization by using the Laplacian approximation is possible.

So by Theorem 4.1, the best we can get is Proposition 4.2. However this does not mean that formula (4.4) can not be an intensity. In fact, one may use Aven's Theorem

4.2 and prove that it is the real intensity. The difficult part is finding the dominating process $(y_s)_{s\geq 0}$. Since we work on pure jump processes, we will not discuss this any further. We just like to point out that Aven's theorem can be powerful in the sense that it does not require totally inaccessibility of the stopping time τ.

The Lévy-Itô decomposition shows that every Lévy process can be approximated to an arbitrary level of precision by a jump-diffusion process. This point of view says that the intensity of every Lévy process is non-zero and it is an appropriate limit of the above form.

By Proposition 3.3 the intensities λ_t^i in Proposition 4.2 are short credit spreads. Therefore an important conclusion of this section is that if even complete information is available, the short credit spreads are not necessarily zero.

Chapter 5

Hedging of Defaultable Claims

5.1 Introduction

Throughout this chapter it is assumed that $(\Omega, \mathcal{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ is a filtered probability space that satisfies the usual hypotheses and the filtration $(\mathfrak{F}_t)_{t\geq 0}$ is generated by the underlying process.

In this chapter we study financial products with actual payoffs in the form of

$$F(X_T)1_{\{\tau>T\}}, \tag{5.1}$$

where $\tau = \inf\{t; X_t < 0\}$, $X_0 > 0$, $F : \mathbb{R} \to \mathbb{R}$, is a real valued function, and $T > 0$ is the maturity or expiration of the security. These payoffs are both path-dependent (even non-Markovian type) and due to the default indicator process, credit sensitive. Since the life time of this product is $[0, T]$, the process $(F(X_t))_{t\geq 0}$ or the underlying process the process $X = (X_t)_{0\leq t<\infty}$ can be sent to a "cemetery state" at time T.

If F is non-negative then it reflects the fact that the payoff function of the contract is non-negative at maturity. This security pays $F(X_T)$ if there is no default in $[0, T]$

and zero otherwise. A defaultable zero-coupon bond is a special case of this security by letting $F(x) = c$, on $\mathbb{R}$ for a constant c. It would be much better if we could let the function $F = F(x)$ be multivariate. However, this choice makes the analysis more complicated.Later we see that the function $F = F(x)$ is the boundary condition of a partial integro differential equation (PIDE).

If it is necessary, one can assume that $\tau < \infty$ almost surely. Since the maturity is a deterministic fixed time T, this assumption for instance can be reached if the underlying process $X = (X_t)_{0 \leq t < \infty}$ is sent to the cemetery state zero for $t > T$.

The main idea behind this work is how the riskiness of a bond issued by an insurance company can be managed. This bond can be considered as a special defaultable claim for the company. Since the evolution of a risky asset (especially in insurance models) resembles the sample paths of a pure jump process with finite variation, we are mostly interested in this type of process. Especially we focus on finite variation Lévy processes. However we let the processes to have infinite activities which means that the integral of the Lévy measures over the whole real line can be infinite. Please see Geman (2002) for some motivations on how these processes model the dynamic of the stock prices better than the diffusions or jump-diffusion models. Beside, there are also some technical reasons behind this choice. Because of these technical problems, finally we focus on these special type of Lévy process to model the underlying risky asset. As we move on, these technical problems are explained. Anyway, some of the concepts are even applicable to semimartingales.

Apart from the theoretical concerns in this chapter, the main effort is to obtain answers to two interesting questions.

The first question is, given a payoff $F(X_T)$ as above, how can the riskiness of the defaultable security $F(X_T)1_{\{\tau>T\}}$ be managed? Clearly, the risk originates from a credit event, and we are basically speaking of credit risk. This question is answered by analyzing the structure of the security and finding its optimal hedging strategy. The idea of optimality is explained in Section 5.4. We see that for a general payoff, it is not possible to eliminate the credit risk completely. The best hedging is achieved by minimizing this risk in a suitable way.

Knowing that in general it is not possible to eliminate the intrinsic risk completely, the second question is whether it is possible to design a customized payoff $F(X_T)$ specifically, to make the product completely risk free. A simpler version of this question is what kind of function $F = F(x)$ allows for the existence of a perfect hedge. This will result in a risk-free defaultable claim. Considering this risk-free security as a criterion, the riskiness of other defaultable products can be compared to it. In the context of jump-diffusion processes, Kunita (2010) answers a similar question for path independent payoffs.

In this chapter, all the involved processes belong to the class of semimartingales, see Definition A.8 of the Appendix. The symbol X typically refers to a semimartingale, unless otherwise stated. Although this class of processes is very abstract, there are many advantages in using semimartingale theory. In particular, semimartingales cover

a wide variety of processes (even non-Markovian ones) and most importantly they are closed under many mathematical operations. For example if X is a Lévy process, $(g(t, X_t))_{t\geq 0}$ is not necessarily a Lévy process, even for a very smooth function $g = g(t, x)$. But this holds true for the class of semimartingales, if $g = g(t, x)$ is a $C^{1,2}$ function.

It is assumed that the market is made of only two assets. The first one is a risk-free asset. Henceforth, all our value processes are discounted at this risk-free asset rate. Therefore, the value of the risk-free asset is 1 at all times. Depending on the payoff of the contract, this assumption might cause some practical issues. This assumption (or supposing that the interest is zero) does not cause any loss of generality in our theory, but in more complex models, payoffs, and for implementation matters, one must be cautious about it.

As an example, assume that we want to price a contingent path dependent claim with the payoff given by $\max\{X_{t_1}, X_{t_2}, ..., X_{t_n}\}$. Here, the process X is the underlying risky asset and $\{t_1, t_2, ..., t_n\}$ are deterministic time spots in the time interval $[0, T]$ where T is the maturity of the contract. Under a risk-neutral measure, the time t value of this contingent claim is equal to

$$P_t = \mathbb{E}\Big[e^{-\int_t^T r_s\, ds} \max\{X_{t_1}, X_{t_2}, ..., X_{t_n}\}|\mathfrak{F}_t\Big].$$

However, if we instead consider the discounted values $X_t^* = e^{-\int_0^t r_s\, ds} X_t$, it is not general enough to capture the above formula.

In our model, to get more practical results, we assume that the risky asset is modeled by a finite variation Lévy process that satisfies certain conditions. It is also supposed that the market is frictionless. The outline of this chapter is as follows.

Quadratic variation and conditional quadratic variation are explained in Sections 5.2 and 5.3. These are two essential tools in our analysis. Section 5.4 reviews local risk-minimization hedging. In Section 5.5, we obtain the predictable part of the semimartingale $\left(g(t, X_t)1_{\{\tau>t\}}\right)_{t\geq 0}$. The hedging strategies are obtained in Section 5.7. Finally in Section 5.8, we have a look at the estimation of the distribution of the default time and pricing tools.

5.2 Quadratic Variation and Conditional Quadratic Variation

In this section we introduce two objects that play an important role in our analysis. We suppose that the processes X and Y in this section are semimartingales, see Definition A.8. The class of semimartingales is denoted by $\mathscr{S}$. More explicit results can be found under additional assumptions. In the next section, we focus on square-integrable martingales.

For a process X in $\mathscr{S}$, the quadratic variation and the conditional quadratic variation are two new processes, respectively denoted by $[X, X]$ (or $[X]$) and $\langle X, X\rangle$ (or $\langle X\rangle$). For two processes X and Y, the notations $[X, Y]$ and $\langle X, Y\rangle$, respectively, stand for the quadratic covariation and the conditional quadratic covariation. We present

their precise definitions, but let us first explain why these are important.

These processes appear in stochastic modeling, both in the theory and in applications, for instance in the context of stochastic integration or in hedging problems. Normally obtaining hedging strategies reduces to solving a projection problem which involves either quadratic covariation or conditional quadratic covariation. We will see this for the financial product analyzed in this work.

The quadratic variation has an explicit definition, so for any semimartingale process it has its own form. On the other hand, as far as we know, there is no general constructive way to find the conditional quadratic variation. If the underlying process is continuous then these two can be equal (as we see later), but except for this special case, they can be different and financially have different interpretations.

Definition 5.1. *Let X be a semimartingale. The quadratic variation of X is a process denoted by $[X,X] = ([X,X]_t)_{t\geq 0}$ and defined by*

$$[X,X]_t = X^2 - 2\int_0^t X_{s^-}\, dX_s,$$

where X_- is the left-hand side limit of X, and it is assumed that $X_{0^-} = 0$.

Notice that since X is a semimartingale and X_- is a cág process, the integral $\int X_-\, dX$ is well defined, see Chapter I, Section 4d of Jacod and Shiryaev (1987). In fact as we see in what follows, this integral can be approximated by a Riemann type series. This gives us also a more intuitive definition of quadratic variation.

Definition 5.2. *The sequence* $(T_m^n)_{m,n\geq 0}$ *of stopping times is called an adapted subdivision Riemann sequence if for a fixed* $n \geq 1$, $T_0^n = 0$, $\sup_{m\geq 0} T_m^n < \infty$, $T_m^n < T_{m+1}^n$ *on the set* $\{T_m^n < \infty\}$ *and*

$$\sup_{m\geq 0} \left(T_{m+1}^n \wedge t - T_m^n \wedge t\right),$$

converges to zero as $n \to \infty$, *almost surely for all* $t \geq 0$.

Proposition 5.1. *Let* X *be a semimartingale and* Y *be a cád or cág process. Then the limit, as* $n \to \infty$, *of*

$$\sum_{m\geq 1} Y_{T_m^n} \left(X_{T_{m+1}^n \wedge t} - X_{T_m^n \wedge t}\right),$$

is $\int_0^t Y_{s^-}\, dX_s$, *in measure, uniformly on every compact interval.*

To see a proof of this proposition, we refer to Proposition 4.44 of Jacod and Shiryaev (1987). Using this proposition one can prove the following important theorem, see Theorem 4.47 of Jacod and Shiryaev (1987).

Theorem 5.1. *For any adapted subdivision Riemann sequence* $(T_m^n)_{m,n\geq 0}$, *the process*

$$X_0^2 + \sum_{m\geq 1} \left(X_{T_{m+1}^n \wedge t} - X_{T_m^n \wedge t}\right)^2,$$

converges to the process $[X, X]$, *in measure, uniformly on every compact interval.*

It is easy to prove that the process $[X, X]$ is cádlág, non-decreasing, adapted, $[X, X]_0 = X_0^2$, and $\Delta[X, X] = (\Delta X)^2$, for example see Theorem 4.47 of Jacod and Shiryaev (1987). Here for a process Y, the process ΔY is defined as $\Delta Y = Y - Y_-$. The process $[X, X]$ is also denoted by $[X]$.

The quadratic covariation of X and Y can be defined by the polarization property

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

The quadratic covariation is also called the bracket process of X and Y. The process $[X,Y]$ is a finite variation process satisfying $\Delta[X,Y] = \Delta X \Delta Y$, see Theorem 4.47 of Jacod and Shiryaev (1987). It is also easy to show that $[X,Y]_0 = X_0 Y_0$. From these definitions, one can obtain the following product (or integration by parts) formula for two semimartingales X and Y

$$XY = \int X_- \, dY + \int Y_- \, dX + [X,Y]. \tag{5.2}$$

Again note that by the explanations following Definition 5.1, the integrals are well defined. This formula can be rewritten as

$$X_t Y_t = \int_0^t X_{s^-} \, dY_s + \int_0^t Y_{s^-} \, dX_s + [X,Y]_t, \quad t \geq 0.$$

Using Proposition 5.1, one can approximate the above integrals and prove that the process

$$X_0 Y_0 + \sum_m \left(Y_{T_{m+1}^n \wedge t} - Y_{T_m^n \wedge t} \right) \left(X_{T_{m+1}^n \wedge t} - X_{T_m^n \wedge t} \right),$$

converges to the process $[X,Y]$, in measure, uniformly on every compact interval.

Regarding the quadratic covariation, we have the following useful properties, see Proposition 4.49 of Jacod and Shiryaev (1987). In what follows, the symbol $\mathscr{V}$ refers to the class of finite variation processes.

Proposition 5.2. *Let $X \in \mathscr{S}$ and $Y \in \mathscr{V}$ then*

(a) $[X,Y] = \int \Delta X \, dY$ *and* $XY = \int Y_- \, dX + \int X \, dY$,

(b) If Y *is predictable, then* $[X,Y] = \int \Delta Y \, dX$ *and* $XY = \int Y \, dX + \int X_- \, dY$,

(c) If Y *is predictable and* X *is a local martingale, then* $[X,Y]$ *is a local martingale,*

(d) If Y *or* X *is continuous, then* $[X,Y] = 0$.

In contrast to the quadratic variation, the conditional quadratic variation does not always exist. Its existence conditions and some related properties of quadratic variation and conditional quadratic variation can be found in Chapter III, Section 5 of Protter (2004). To define conditional quadratic variation, we need the following definition.

Definition 5.3. *A finite variation process* Λ *(i.e.* $\Lambda \in \mathscr{V}$*) with* $\Lambda_0 = 0$ *is of integrable variation if its expected total variation is finite which means that* $\mathbb{E}[Var\,(\Lambda)_\infty] < \infty$. *The class of integrable variation processes is denoted by* $\mathscr{A}$.

Note that in the above definition $Var\,(\Lambda)_\infty$ is the total variation of the process Λ over $[0,\infty]$.

Lemma 5.1. *Let* Λ *belongs to* $\mathscr{A}_{loc}$*, see Definition A.4 of the Appendix. Then there exists a unique finite variation predictable process* $\widetilde{\Lambda}$*, with* $\widetilde{\Lambda}_0 = 0$ *such that* $\Lambda - \widetilde{\Lambda}$ *is a local martingale. This process is called the compensator of* Λ.

If $[X,X]$ is a locally integrable variation process, then it satisfies all the conditions of Lemma 5.1. Hence, the compensator of $[X,X]$ exists, and it is called the

conditional quadratic variation of X, denoted by $\langle X, X\rangle$ or $\langle X\rangle$. It is also commonly called the angle process. Since $[X, X]$ is a non-decreasing process, it belongs to $\mathscr{V}$ and $Var\left([X, X]\right)_t = [X, X]_t$, where $Var([X, X])$ is the total variation of $[X, X]$ over $[0, t]$. Therefore $[X, X] \in \mathscr{A}_{loc}$ if $[X, X]$ is predictable and locally integrable.

Remark 5.1. *If X is a continuous process then $\Delta X = X - X_- = 0$. As we saw above, it can be proved that $\Delta[X, X] = (\Delta X)^2$. Therefore if X is continuous, then $[X, X]$ is also a continuous process. This means that $[X, X]$ is a predictable process so if it is locally integrable, its compensator is equal to itself i.e. $[X, X] = \langle X, X\rangle$.*

To define the conditional quadratic covariation $\langle X, Y\rangle$ for two semimartingales X and Y, there are two ways. The first method is by using polarization.

If X and Y are two semimartingales such that $\langle X, X\rangle$, $\langle Y, Y\rangle$ and $\langle X + Y, X + Y\rangle$ all exist, then $\langle X, Y\rangle$ can be defined by

$$\langle X, Y\rangle = \frac{1}{2}\left(\langle X + Y, X + Y\rangle - \langle X, X\rangle - \langle Y, Y\rangle\right). \tag{5.3}$$

Remark 5.2. *Note that for semimartingales X and Y, $[X, Y]$ always exists but not necessarily $\langle X, Y\rangle$.*

The second method is based on Lemma 5.1. The bracket process $[X, Y]$ of two semimartingales has paths of finite variation on compacts, and it is also a semimartingale, see Chapter II, Section 6, Corollary 1 of Protter (2004). Further assume that $[X, Y]$ belongs to $\mathscr{A}_{loc}$. Now by using Lemma 5.1, the conditional quadratic covariation $\langle X, Y\rangle$

can be defined as the compensator of $[X, Y]$. We shall call this the modified version of conditional quadratic covariation. It reduces to the same notion as the previous definition, if $\langle X+Y, X+Y\rangle$, $\langle X, X\rangle$, and $\langle Y, Y\rangle$ all exist. Also if X and Y are both square-integrable, then the modified version of $\langle X, Y\rangle$ is the same as the one in the context of square-integrable martingales, see Section 5.3.

The situation for square-integrable martingales is less complicated and it is explained in the next section. To motivate the following example we briefly mention one fundamental result. Meyer showed that for a square-integrable martingale X there exists a unique increasing process Λ such that

$$\mathbb{E}[(X_t - X_s)^2|\mathfrak{F}_s] = \mathbb{E}[\Lambda_t|\mathfrak{F}_s] - \Lambda_s, \text{ almost surely for all } t > s \geq 0,$$

see Meyer (1962, 1963). The next section shows that the process Λ is in fact the quadratic variation of X, that is $[X]$ as in Definition 5.8. Using this point and the properties of Brownian motion, we have the following.

Example 5.1. *Assume that $(B_t)_{t\geq 0}$ is a Brownian motion, then $[B, B]_t = t$, for all $t \geq 0$. Since Brownian motion is a continuous process we also have $\langle B, B\rangle_t = t$ for all $t \geq 0$.*

The above example shows that finding equivalent forms of the (conditional) quadratic (covariation) variation even for a simple process like Brownian motion is not an easy task. The properties of the quadratic variation process are very important in finding these equivalent forms.

As pointed out, the process $[X, X]$ is non-decreasing with right continuous paths. Therefore together with the property $\Delta[X, X] = (\Delta X)^2$, one can decompose $[X, X]$ path by path into its continuous part and its pure jump part

$$[X, X]_t = [X, X]_t^c + \sum_{0 \le s \le t} (\Delta X_s)^2, \quad \text{for all } t \ge 0,$$

where $[X, X]^c$ denotes the path by path continuous part of $[X, X]$. Similarly the path by path continuous part of $[X, Y]$ can be defined. A semimartingale is called a quadratic pure jump process if $[X, X]^c = 0$.

The following theorem that can be found in Protter (2004) (Chapter II, Section 6, Theorem 28), is useful in calculating quadratic covariations.

Theorem 5.2. *Let X be a quadratic pure jump semimartingale. Then for any semimartingale Y we have*

$$[X, Y]_t = X_0 Y_0 + \sum_{0 < s \le t} \Delta X_s \Delta Y_s, \quad \textit{for all } t \ge 0.$$

Example 5.2. *Assume that $U_t = u + \mu t + \sum_{i=0}^{N_t} Y_i$ is a compound Poisson process plus a drift, then by a direct calculation using Riemann approximations, or using the previous theorem (if one already accepts that the process U is a quadratic pure jump semimartingale), we have that $[U, U]_t = u^2 + \sum_{i=1}^{N_t} Y_i^2 = \sum_{0 \le s \le t, \Delta U_s \ne 0} |\Delta U_s|^2$. To calculate the conditional quadratic covariation we note that the process $[U, U]_t - \lambda \int_0^t \int_0^\infty y^2 F_{Y_1}(dy)$ is an $\mathfrak{F}^U$-martingale, where F_{Y_1} is the distribution Y_1. Now the uniqueness in Lemma 5.1 implies that $\langle U, U \rangle_t = \lambda \int_0^t \int_0^\infty y^2 F_{Y_1}(dy)$.*

The above example can be generalized.

Example 5.3. *Assume that X is a Lévy process with characteristic triplet (σ^2, v, γ), then by using the linearity of the bracket process, the above example, and similar calculations one can show that*

$$[X,X]_t = X_0^2 + \sigma^2 t + \sum_{0\le s\le t, \Delta X_s \neq 0} |\Delta X_s|^2,$$

or equivalently

$$[X,X]_t = X_0^2 + \sigma^2 t + \int\int_{[0,t]\times\mathbb{R}} y^2 J_X(ds\ dy),$$

where, J_X is the jump measure of the process X. By compensating the jump measure and through similar arguments as in the last example, we have that

$$\langle X,X\rangle = \sigma^2 t + \int_0^t \int_{\mathbb{R}} y^2\ v(dy)\ ds.$$

In the next section, we focus on the properties of the conditional quadratic variation in the context of square-integrable martingales. In this case, there are more satisfactory results concerning the conditional quadratic variation.

Beside this, we have two more purposes in the next section. First, square-integrable martingales are exactly defined, as the literature on this topic is not consistent. To answer our second question, square-integrable martingales are needed. Second, Corollary 5.2 and Proposition 5.4 are two important results of the next section that we will use and revisit later.

5.3 Conditional Quadratic Variation and Square-Integrable Martingales

We start by two abstract definitions that will be used later.

Definition 5.4. *A filtration $\mathfrak{F}$ is called quasi left continuous if for every predictable stopping time $\mathcal{T}$ one has $\mathfrak{F}_{\mathcal{T}} = \mathfrak{F}_{\mathcal{T}^-}$, the σ-algebra $\mathfrak{F}_{\mathcal{T}^-}$ is given in Definition A.9 of the Appendix.*

Definition 5.5. *A process X is called quasi left continuous if for every increasing sequence of stopping times $(\mathcal{T}_n)_{n\geq 1}$ such that $\mathcal{T}_n \uparrow \mathcal{T}$, we have $X_{\mathcal{T}_n} \to X_{\mathcal{T}}$ almost surely on $\{\mathcal{T} < \infty\}$.*

The following discussion and series of results are mostly taken from either Protter (2004), Jacod and Shiryaev (1987), or Kunita and Watanabe (1967).

Let $\mathcal{T}$ be a predictable stopping time, and let $(S_n)_{n\geq 1}$ be a non-decreasing sequence of stopping time announcing $\mathcal{T}$ with $\lim_{n\to\infty} S_n = \mathcal{T}$. Then it can be shown that we have $\mathfrak{F}_{\mathcal{T}^-} = \bigvee_n \mathfrak{F}_{S_n}$, see Chapter III, Section 2, Theorem 5 of Protter (2004). Now if $\mathfrak{F}$ is a quasi left continuous filtration then the previous definition implies that $\mathfrak{F}_{\mathcal{T}} = \bigvee_n \mathfrak{F}_{S_n}$ (a filtration with this property is said to have no times of discontinuity).

Let us formally define square-integrable martingales and the notation.

Definition 5.6. *The set $\mathfrak{M}^2$ of square-integrable martingales is the set of all right continuous, real valued processes X, adapted to the family $(\mathfrak{F}_t)_{t\geq 0}$ such that $\mathbb{E}[X_t^2] < \infty$*

(in this case we say that $X_t \in L^2(\Omega, \mathfrak{F}_t, \mathbb{P})$*), for all* $t \geq 0$ *and* $\mathbb{E}[X_t|\mathfrak{F}_s] = X_s$*, for all* $t \geq s \geq 0$.

For quasi left continuous square-integrable processes we have the following lemma, see Kunita and Watanabe (1967).

Lemma 5.2. *Suppose that* $X \in \mathfrak{M}^2$ *and we have at least one of the following conditions*

- $\{X, \mathfrak{F}\}$ *is continuous,*
- $\mathfrak{F}$ *has no times of discontinuity,*

then the process X *is quasi left continuous process.*

Although the above definition of square-integrable martingales seems intuitive, in the literature (for example in Jacod and Shiryaev (1987)) they are defined differently. To distinguish between the two we use the terminology "L^2-martingales". The class of L^2-martingales is denoted by $\mathcal{M}^2$.

Definition 5.7. *The set* $\mathcal{M}^2$ *of* L^2*-martingales is the set of all right continuous, real valued processes* X*, adapted to the family* $(\mathfrak{F}_t)_{t\geq 0}$*, such that* $\sup_{t\geq 0} \mathbb{E}[X_t^2] < \infty$ *and* $\mathbb{E}[X_t|\mathfrak{F}_s] = X_s$*, for all* $t \geq s \geq 0$.

Obviously $\mathcal{M}^2 \subset \mathfrak{M}^2$, but in general they are not equal. In the following lemma we prove the equality for a special important case, that we use later.

Lemma 5.3. *Assume that* $X \in \mathfrak{M}^2$ *is a square-integrable martingale on* $[0, T]$*, for* $0 \leq T < \infty$*, then* $X \in \mathcal{M}^2$ *on this interval.*

Proof. Since $X = (X_t)_{0\leq t\leq T}$ is a martingale, by Jensens's inequality $|X| = (|X|_t)_{t\geq 0}$ is a sub-martingale (because the bracket function $|\cdot|$ is convex). Therefore;

$$|X_t| \leq \mathbb{E}\left[|X_T|\,|\mathcal{F}_t\right], \quad \text{for all } t \leq T.$$

By assumption $X_T \in L^2(\Omega, \mathfrak{F}_T, \mathbb{P})$, hence by Jensens's inequality for the convex function $\psi : \mathbb{R} \to \mathbb{R}$, defined by $\psi(x) = \begin{cases} x^2, & x \geq 0; \\ 0, & x < 0, \end{cases}$

$$\left(\mathbb{E}[|X_T|\,|\mathcal{F}_t]\right)^2 \leq \mathbb{E}[X_T^2|\mathcal{F}_t].$$

From the two previous inequalities, we obtain $|X_t|^2 \leq \mathbb{E}[X_T^2|\mathcal{F}_t]$. Now by taking expectation from both sides of this inequality, we get $\mathbb{E}[X_t^2] \leq \mathbb{E}[X_T^2]$. Since the right-hand side does not depend on t and $X_T \in L^2(\Omega, \mathfrak{F}_T, \mathbb{P})$, the result follows. □

The equality between these two sets of martingales is important in many aspects. For instance, it allows us to use Doob's inequality for a square-integrable martingale on a finite interval. Or, as another example, it is true that $\mathcal{M}^2 \subset \mathcal{M}$, where $\mathcal{M}$ is the class of uniformly integrable martingales. This is a direct conclusion of Chapter I, Section 2, Theorem 11 of Protter (2004). Hence, on finite intervals every square-integrable martingale is also uniformly integrable.

From now on, in working on a finite interval, we do not distinguish the two classes $\mathfrak{M}^2$ and $\mathcal{M}^2$ and we use the notation $\mathcal{M}^2$ for both sets. Now we present the definition of the conditional quadratic covariation for square-integrable martingales. For the existence and uniqueness we refer the reader respectively to Meyer (1962) and Meyer (1963).

Definition 5.8. *Let X, Y belong to $\mathfrak{M}^2$, then there exists a unique $\langle X, Y\rangle \in \mathfrak{U}$ (up to an equivalence class) such that*

$$\mathbb{E}[(X_t - X_s)(Y_t - Y_s)|\mathfrak{F}_s] = \mathbb{E}[\langle X, Y\rangle_t - \langle X, Y\rangle_s|\mathcal{F}_s], \text{ almost surely,} \tag{5.4}$$

for every $t \geq s$,, where

$$\mathfrak{U} = \{\Lambda_t^1 - \Lambda_t^2;\ \Lambda_t^i \in \mathfrak{U}^+,\ i = 1, 2\}, \tag{5.5}$$

and $\mathfrak{U}^+$is the set of all natural increasing processes Λ (i.e. increasing processes with the property that $\mathbb{P}_x(\Lambda_{\mathcal{T}} \neq \Lambda_{\mathcal{T}^-}) = 0$, for all totally inaccessible stopping time $\mathcal{T}$ and $x \in \mathbb{R}$) such that $\mathbb{E}(\Lambda_t) < \infty$, for every $t \geq 0$.

Combining these definitions and results we have the following corollary.

Corollary 5.1. *Let $X, Y \in \mathfrak{M}^2$ under the filtration $\mathfrak{F}$ that is the natural completed filtration of a Lévy process. Then $\langle X, Y\rangle$ as defined above is continuous.*

Proof. The existence of $\langle X, Y\rangle$ is guaranteed by the above definition. The natural completed filtration of a Lévy process has no time of discontinuity. Therefore by Lemma 5.2 both X and Y are quasi left continuous process. Now the corollary follows from Theorem 1.2 of Kunita and Watanabe (1967). □

Definition 5.9. *Let $X, Y \in \mathfrak{M}^2$ then X is orthogonal to Y if and only if $\langle X, Y\rangle = 0$.*

Meyer defined that $X, Y \in \mathfrak{M}^2$ are orthogonal, if $XY = (X_tY_t)_{t\geq 0}$ is a martingale. Since $\mathbb{E}[(X_t - X_s)(Y_t - Y_s)|\mathfrak{F}_s] = \mathbb{E}[X_tY_t|\mathfrak{F}_s] - X_sY_s$ then it comes out that $\langle X, Y\rangle = 0$

if and only if XY is a martingale. There is a weaker notion of orthogonality for two martingales X and Y that is defined by $\mathbb{E}[X_t Y_t] = 0$, for every $t \geq 0$. To differentiate between the two, the first one is called strong orthogonality.

Definition 5.10. *Two martingales X and Y are called strongly orthogonal to each other if XY is a martingale.*

It is worth mentioning that for $X, Y \in \mathcal{M}^2$, there are other equivalent definitions of $[X, Y]$ and $\langle X, Y \rangle$. We shortly present these definitions and refer the reader to He, Wang and Yan (1992) for more details.

Definition 5.11. *Let $X, Y \in \mathcal{M}^2$ then $[X, Y]$ and $\langle X, Y \rangle$ are respectively the unique adapted process and unique predictable processes, both with integrable variation, such that $XY - [X, Y] \in \mathcal{M}_0$, $\Delta[X, Y] = \Delta X \Delta Y$, and $XY - \langle X, Y \rangle \in \mathcal{M}_0$, where $\mathcal{M}_0$ is the collection of all uniformly integrable martingales which are null at zero.*

Using the previous definitions, one can get the following properties, see He, Wang and Yan (1992).

Lemma 5.4. *Let $X, Y \in \mathcal{M}^2$ then*

1. $\langle X, Y \rangle = \frac{1}{2}\big(\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle\big)$,

2. $[X, Y]_t = X_0 Y_0 + \langle X^c, Y^c \rangle_t + \sum_{s \leq t} \Delta X_s \Delta Y_s, \quad t \geq 0,$

   where X^c and Y^c are the continuous martingale parts of X and Y, respectively.

3. *if $X_0 Y_0 = 0$ then the following assertions are equivalent*

- X *is strongly orthogonal to* Y,
- $[X, Y] \in \mathcal{M}_0$,
- $\langle X, Y \rangle = 0$.

By Definition 5.11, one observes that for L^2-martingales, the modified version of conditional quadratic variation coincides with the one here. Inspired by this, one can prove the following lemma, by a localization argument; to see the details we refer to Proposition 4.50 of Jacod and Shiryaev (1987).

Lemma 5.5. *If* X, Y *belong to* $\mathcal{M}^2_{loc}$, *then* $[X, Y]$ *belongs to* $\mathcal{A}_{loc}$ *and its compensator is* $\langle X, Y \rangle$ *(i.e.* $[X, Y] - \langle X, Y \rangle$ *is a local martingale). If moreover* $X, Y \in \mathcal{M}^2$, *then* $XY - [X, Y]$ *belongs to* $\mathcal{M}$.

One of the fundamental results in the theory of stochastic calculus is the following, see Proposition 4.50 of Jacod and Shiryaev (1987) for the proof.

Proposition 5.3. *Assume that* X *is a local* L^2*-martingale (or just a local martingale). Then* $X = X_0$ *almost surely if and only if* $[X, X] = 0$.

Surprisingly in the case of L^2-martingales, the previous result holds for the angle brackets as well, a result that we use later.

Corollary 5.2. *Suppose that* $X \in \mathcal{M}^2_{loc}$ *then* $X = X_0$ *almost surely if and only if* $\langle X, X \rangle = 0$.

Proof. Assume that $X \in \mathcal{M}^2_{loc}$ and $\langle X, X\rangle = 0$. With no loss in generality we can suppose that $X_0 = 0$. By Lemma 5.5, we have both, that $[X, X]$ belongs to $\mathcal{A}_{loc}$ and $[X, X]$ is a local martingale. But as we saw in the explanations following Definition 5.1, the process $[X, X]$ is a finite valued increasing process and therefore $[X, X] \geq [X, X]_0 = (X_0)^2$. Hence $[X, X]$ is a non-negative local martingale. By a localization argument, we can show that $[X, X] = 0$, almost surely. The details are as follows. Since $[X, X]$ is a local martingale, there exists a sequence of stopping times $\{T_n\}_{n\geq 1}$ such that $T_n \to \infty$, almost surely, and $[X, X]^{T_n}$ is a martingale for all $n \geq 1$. Here, $[X, X]^{T_n}$ is a new process stopped at T_n defined by $[X, X]^{T_n}_t = [X, X]_{t\wedge T_n}$, for all $t \geq 0$. Since the process $[X, X]^{T_n}$ is a martingale, $\mathbb{E}\Big[[X, X]_{t\wedge T_n}\Big] = 0$, for all $t \geq 0$, and because it is non-negative, we have that $[X, X]_{t\wedge T_n} = 0$, almost surely, for all $t \geq 0$. Therefore for all $n \geq 1$, there exists a set A_n subset of Ω such that $\mathbb{P}(A_n) = 0$, and if ω does not belong to A_n, then $[X, X]_{t\wedge T_n(\omega)}(\omega) = 0$. Now take $A = \bigcup_{n=1}^{\infty} A_n$, $\mathbb{P}(A) = 0$ and if ω is not in A, then for all $n \geq 1$ we have $[X, X]_{t\wedge T_n(\omega)}(\omega) = 0$. Since $\{T_n\}_{n\geq 1}$ is approaching infinity almost surely, there is a set $A^{'}$, such that if ω is not in $A^{'}$, then $T_n(\omega) \to \infty$. Now let $A^{''} = A \cup A^{'}$, then $\mathbb{P}(A^{''}) = 0$. If ω is not in $A^{''}$, then ω is not in A and so for all n, $[X, X]_{t\wedge T_n(\omega)}(\omega) = 0$. Since ω is not in $A^{'}$, then by taking $T_n(\omega) \to \infty$, we get $[X, X]_t(\omega) = 0$, for all $\omega \notin A^{''}$. Therefore $[X, X] = 0$ almost surely. Now by the above proposition we get that $X = 0$ almost surely. □

Remark 5.3. *In the above proof, we actually proved the fact that a non-negative local martingale is almost surely zero. Localization arguments like these can be applied to*

all types of these proofs. An alternative proof of this corollary is by noticing that since $\langle X, X\rangle = 0$*, the process* X *is orthogonal to itself. Now the result follows from part (a), Lemma 4.13 of Jacod and Shiryaev (1987).*

Remark 5.4. *Remember that since we are working on the finite interval* $[0, T]$*, square-integrable martingales and* L^2*-martingales are the same. Most importantly, the modified versions of conditional quadratic variation for martingales coincide with the ones in this chapter.*

The next proposition is the main result that will be generalized in what follows; see Kunita and Watanabe (1967) for more details.

Proposition 5.4. *Let* X *and* Y *be elements of* $\mathfrak{M}^2$*, then there exist unique* $Y^{'} \in \mathfrak{L}(X)$ *and* Y *strongly orthogonal to all elements in* $\mathfrak{L}(X)$ *such that* $Y = Y^{'} + Y^{''}$*, where* $\mathfrak{L}(X)$ *is the set of all* $\int \phi dX$ *for predictable processes* ϕ*. In fact* ϕ *is the Radon-Nikodym derivative of* $\langle X, Y\rangle$ *with respect to* $\langle X\rangle$*, i.e.* $\phi = \frac{d\langle X,Y\rangle}{d\langle X\rangle}$.

This last proposition gives a good idea of how a hedging process in an incomplete market could be represented. In this proposition Y acts as the payoff. If the market is complete then Y is in the space $\mathfrak{L}(X)$ and there is no second term. But normally the markets are incomplete. The best can be done is projecting the payoff to the space $\mathfrak{L}(X)$. In any case, there is an intrinsic risk carried by the second term which affects the hedging. A decomposition of a payoff, as in this proposition, leads to the predictable hedging process ϕ. These are explained further in the next section.

5.4 Local Risk-Minimization Hedging

It is a well known result that the Black-Scholes model is a complete model. This means that first the market is free of arbitrage and second there is only one unique way to hedge the payoff of a contingent claim. Mathematically, the arbitrage-free assumption means that there is at least one risk neutral measure, and the completeness of the model is equivalent to the uniqueness of this measure. The completeness property fails in most discontinuous models, i.e., when the underlying process is a jump process. It should be noted that in any case, the arbitrage-free assumption leads to at least one risk neutral measure. Under this assumption, it is guaranteed that there is always at least one way to price the contingent claim.

In an incomplete market, the risk of financial products can be managed in different ways. Super hedging, utility maximization, and quadratic hedging are common methods to hedge the risk of securities. Quadratic hedging by itself is divided into two categories, one is mean-variance hedging and the other one is locally risk-minimizing hedging. There is no unique agreement between scholars on which one is the best approach. Depending on the problem each method has its own advantages and disadvantages. For the type of product that interests us, we use a local risk-minimization approach. This section explains in detail how this method can control the risk. **But first a critical question, why do we use locally risk-minimizing approach?**

Let start comparing locally risk-minimizing approach with a very simple type of

hedging namely delta hedging. Delta hedging is one of the simplest types of hedging approaches that is quite used in real markets. Despite simplicity, under some circumstances, it can be actually a useful type of management. If a model uses a continuous process to model the underlying risky asset, then delta hedging can outperform most of the hedging managements including locally risk-minimizing approach. This is because, it gives the same level of protection with lower costs and of course without requiring any complex technicality. But if the model assumes that there are shocks and jumps in markets and so uses a jump process to model firm's asset values, then locally risk-minimizing approach outperforms the delta hedging. If the payoff is credit sensitive then the performance of delta hedging gets even worse.

Another common method is super hedging. In super hedging due to the nature of the management, normally the cost of the strategies are too high. This is mainly because of the conservative measurements that this type of management considers. Especially this gets worse when the payoff is path dependent. Also this method leads to non-linear pricing rules. For more details see, Chapter 10 of Cont and Tankov (2004).

The third common type of hedging approaches is utility maximization approach. This approach requires a good knowledge of the utility function and the probability measure. Even if the probability measure is determined, the model is unstable and non-robust under different types of utility functions. Another problem is the non-linearity of the pricing rules respect with the most utility functions used. The only

utility function that provide a linear pricing rules is $U(x) = -x^2$ which corresponds to quadratic hedging, see Chapter 10 of Cont and Tankov (2004) for more details.

As mentioned earlier quadratic hedging are divided into two categories, mean-variance hedging locally risk-minimizing hedging. Mean-variance hedging control the risk globally while locally risk-minimizing approach do the same job locally. Hence, it is not surprising that the former is self-financing and the later is not. However, since in the credit sensitive derivatives there is a chance of default at any time, we believe that locally risk-minimizing is a good approach in the sense that it provide a better protection in exchange for upcoming costs. In this approach it is guaranteed that obligations in the maturity time will be fulfilled. However, this is just a typical argument and it might be opposed. As it was mentioned before, one cannot entirely reject an approach in the favor of the other one.

Locally risk-minimizing hedging emerged in the development of the concept of risk-minimization. Since in an incomplete market a perfect hedge is not always possible, a good hedging strategy is the one that minimizes the risk. Föllmer and Sondermann (1986) are among the first to have dealt with this problem. Here we explain quadratic hedging with an emphasis on local risk-minimization. For further details, we refer the reader to Schweizer (1999) and Pham (1999). In these two papers, the reader can find a thorough discussion of quadratic hedging as well as many good references. For any undefined notation, please see Jacod and Shiryaev (1987). To explain these concepts we first need to establish some definitions and notation.

Assume that a contingent claim on the security X has maturity T and payoff H at time T. We use the notation H for a general payoff. We stress the fact that H could be a path-dependent payoff, but the execution time is T. However, as we explained by an example in Section 5.1, this is rather a theoretical assumption, and when it comes to real implementations, one has to be very careful about it. There are two critical questions associated with any payoff, its price and its replicating strategy.

A usual way to reduce the risk for a given payoff H is to hedge it dynamically using a (replicating) strategy. As mentioned in Section 5.1, it is assumed that there are only two assets available to hedge, hence our portfolio consists of a risk-free asset and a risky one. So a portfolio is represented by a vector process $\phi = (\theta, \eta)$, where the processes $\theta = (\theta_t)_{0 \leq t \leq T}$ and $\eta = (\eta_t)_{0 \leq t \leq T}$ are respectively the quantities of the shares invested in the risky asset and risk-free one that we need to hold at time t. The process θ must be predictable (because of the nature of a strategy) while η is just adapted.

At any time t, the value process of this portfolio is defined intuitively by $V(\phi) = (V_t(\phi))_{0 \leq t \leq T}$, where $V_t(\phi) = \theta_t X_t + \eta_t$. Remember that due to the conventions mentioned in Section 5.1, the value of the risk-free asset is always equal to 1. Following the strategy ϕ, the cumulative gain, resulting from trading the underlying security X up to time t is equal to $\int_0^t \theta_u dX_u$. In any market, holding a strategy incurs a cost. For any strategy ϕ, the cost process $C(\phi) = (C_t(\phi))_{0 \leq t \leq T}$ is defined by

$$C_t(\phi) = V_t(\phi) - \int_0^t \theta_u dX_u, \quad \text{for all } 0 \leq t \leq T,$$

with $C_0(\phi) = V_0(\phi) = v_0$, where v_0 is the initial capital required to initiate the hedging process. A favorite special case is a constant cost process.

The strategy ϕ leading to a constant cost process, is called self-financing. This terminology was first introduced by Harrison and Pliska (1981). In this case we say that the portfolio or strategy is self-financing. This is equivalent of saying that the value process is given by

$$V_t(\phi) = v_0 + \int_0^t \theta_u dX_u, \quad \text{for all } 0 \leq t \leq T.$$

Remark 5.5. *In the above definitions, one has to be careful about the integral term. For a random process X and a predictable process θ, the integral $\int \theta dX$ is an abstract object that needs to be well defined. We discussed this following the Definition 5.1, when θ is a càg (continue à gauche) or càd (continue à droite) process. All the underlying processes X in this thesis are semimartingales. For a careful study of integrals of predictable processes with respect to semimartingales, we refer the reader to Protter (2004) or Dellacherie and Meyer (1982). For a predictable process θ, if we have nice integrability conditions, for example $\mathbb{E}[\theta_t^2] < \infty$, for all $0 \leq t \leq T$, then it follows that the integral $\int \theta dX$ is well defined.*

Given a payoff $H \in \mathfrak{F}_T$, it is called attainable if there is a self-financing portfolio ϕ such that $V_T(\phi) = H$, $\mathbb{P}-$almost surely. Therefore, if $H \in \mathfrak{F}_T$ is attainable we have $H = v_0 + \int_0^T \theta_u dX_u$, for some predictable process θ. In this case (under no arbitrage opportunity assumptions), the price of this payoff is equal to v_0, and the replicating

portfolio is constituted by θ. If all the payoffs of a market are attainable, we say that it is a complete market, otherwise it is called an incomplete market.

Except in the case when the market is complete, in general, a payoff H is not attainable. For instance, in our work, due to the jumps in the underlying process and to the credit default risk, obviously the payoff $F(X_T)1_{\{\tau>T\}}$ is not attainable. This is proved rigorously later. We remind the reader that assuming a complete market is a rather theoretical assumption, but in reality, securities are priced and analyzed in incomplete markets.

Assume then that the market is incomplete. Still a useful way to reduce the risk for a given payoff H is to hedge it dynamically using a replicating strategy. Even in an incomplete market, there may be some claims that are still attainable. By definition, the incompleteness of a market implies that not all claims are necessarily attainable. For a general payoff H (maybe non-attainable) in an incomplete market, there is either a self-financing portfolio or an admissible portfolio (i.e. $V_T(\phi) = H$) to hedge, but most probably both do not hold simultaneously. If we prefer a self-financing portfolio in order to hedge H, we speak of mean-variance hedging. If we rather select an admissible portfolio to hedge H, we are in the context of (local) risk-minimization.

Which of these methods is better, is not a simple question to answer. Mean-variance hedging controls the risk globally over the interval $[0, T]$, but local risk-minimization achieves this job by controlling risk locally over time at the expense of having a non self-financing portfolio. In a quick comparison one could say that because of the self-

financing property, mean-variance hedging is more interesting.

On the other hand, local risk-minimization approaches are more tractable and they lead to better analytical results. Schweizer, Heath and Platen (2001) provide a comprehensive study and comparison of both approaches. As already mentioned, here we use the method of local risk-minimization. First we briefly explain the development of this concept.

In what follows, the precise definition of an RM-strategy (Risk-Management strategy) is provided. In Schweizer (1999), RM stands for risk-minimizing, but since not all RM-strategies are risk-minimizing, we prefer to use risk-management. Depending on the approaches, minor modifications of this definition are needed. Here we use we the definition from Schweizer (1999) or Pham (1999).

The development of this methodology starts by assuming that X is a (local) martingale. So we first suppose that X is a (local) martingale. Before stating the definition, we introduce the class of $L^2(X)$. This is the space of all real valued predictable processes θ such that $\mathbb{E}\Big[\int_0^T \theta_u^2\, d[X]_u\Big] < \infty$, where the process $[X]$ is the quadratic variation of X.

Definition 5.12. *An RM-strategy is any pair $\phi = (\theta, \eta)$ where $\theta = (\theta_t)_{0\le t\le T}$ belongs to $L^2(X)$ and $\eta = (\eta_t)_{0\le t\le T}$ is a real-valued adapted process such that the value process $V(\phi) = \theta X + \eta$ is right continuous and square-integrable.*

Lemma 2.1 of Schweizer (1999) shows that the process $\int \theta\, dX$ is well defined (in

fact a local martingale). Hence the cost process can be defined for an RM-strategy.

The strategy $\phi = (\theta, \eta)$ determines dynamically a portfolio in order to hedge the security. Therefore, θ_t and η_t are respectively the quantities of the risky asset and risk-free one that we need to hold at time t.

If a claim is not attainable, especially in an incomplete market, the cost process is no longer constant. In this case the risk control is done by a suitable minimization of the following risk process $R(\phi) = (R_t(\phi))_{0 \leq t \leq T}$, where

$$R_t(\phi) = \mathbb{E}[(C_T(\phi) - C_t(\phi))^2 | \mathfrak{F}_t], \quad \text{for all } 0 \leq t \leq T. \tag{5.6}$$

Obviously a zero risk process is equivalent to a self-financing strategy.

If X is a local martingale under $\mathbb{P}$, it can be proved that the cost process associated to the RM-strategy is a martingale (see Lemma 2.1 of Schweizer, 1999), even though the claim may not be attainable. In this case, when the cost process is a martingale, the RM-strategy is called mean-self-financing and it was introduced by Föllmer and Sondermann (1986).

Now having established the previous definitions, we are in a position to formally define the concept of a risk-minimization strategy. When a payoff is non attainable, one should look for a strategy that minimizes the risk process. This is the idea of Föllmer and Sondermann (1986).

Definition 5.13. *A RM-strategy ϕ is called risk-minimizing if for any RM-strategy $\tilde{\phi}$*

with $V_T(\tilde{\phi}) = V_T(\phi)$, $\mathbb{P}-$*almost surely, we have*

$$R_t(\phi) \leq R_t(\tilde{\phi}), \quad \mathbb{P}- \text{ almost surely},$$

for every $t \in [0, T]$.

The above definition intuitively says that among all the RM-strategies, the one with the smallest risk, is the best one.

Having clarified the concept of risk-minimization in the (local) martingale case X, the next questions are: (1) given any claim H, is there a risk-minimization strategy, and if so, (2) how to obtain it? The answer to the latter is less satisfactory than to the former. At the end of this section, we briefly mention a well known method to calculate risk-minimization strategies. As for the existence result, under mild conditions on the underlying process, necessary and sufficient conditions are known.

The risk-minimization strategy problem in the martingale case (that means when X is a martingale) was solved by Föllmer and Sondermann (1986). The generalization to the local martingale case is done by Schweizer (1999). Under the assumption that X is a local $\mathbb{P}-$ martingale, the risk-minimization problem is solved by the so called Galtchouk-Kunita-Watanabe (GKW) decomposition. To solve the problem explicitly, the existence of a GKW decomposition of the claim is assumed. This is an existence result and in general there is no satisfactory procedure to obtain this decomposition. The best case is for a continuous local martingale. The reader is referred to Schweizer (1999) for a detailed study of the problem in this case. We just mention the following

remark.

Remark 5.6. *When X is a local martingale, the mean-variance hedging and risk-minimization hedging lead to the same hedging strategies though different portfolios.*

Unfortunately the generalization of risk-minimization to the semimartingale case is not straightforward. There are technical and compatibility problems. For a non-martingale process, Schweizer (1988) provides an example of an attainable claim that does not admit a risk-minimization strategy.

The extension is possible by putting more restrictive conditions on the underlying process and the hedging strategies as well. In order to make this extension, we must pay more attention to the local properties of the problem.

As for the role of X, it has to satisfy some certain conditions. It is assumed that on $[0, T]$, the process X can be decomposed as $X = X_0 + M + A$, where the process M is in $\mathcal{M}^2_{loc}$ and A is a predictable process that belongs to $\mathscr{V}$. We further assume that both processes M and A are null at zero. The class of all X that satisfy this property is denoted by $\mathcal{S}^2_{loc}(\mathbb{P})$. Notice that X is a special semimartingale. If M is in $\mathcal{M}^2$, then this class is denoted by $\mathcal{S}^2(\mathbb{P})$.

Definition 5.14. *The process X in $\mathcal{S}^2_{loc}(\mathbb{P})$ satisfies the structure condition if the following hold:*

- *The process A is absolutely continuous with respect to $\langle M \rangle$, where $\langle M \rangle$ is the conditional quadratic variation of M.*

- *For all* $t \in [0,T]$, $A_t = \int_0^t \zeta_s d\langle M \rangle_s$, *where* ζ *is a predictable process such that the mean-variance trade-off process* $\tilde{K} = \left(\tilde{K}_t\right)_{0 \leq t \leq T}$, $\tilde{K}_t = \int_0^t \zeta_s^2 d\langle M \rangle_s$ *is* $\mathbb{P}$*-almost surely finite.*

From now on, we use the abbreviations MVT for mean-variance trade-off and SC for structure condition. In the next section we see that our underlying asset process in this thesis satisfies the SC condition.

To present the exact definition of local risk-minimization we follow a series of definitions, for more explanations see Schweizer (1991) or Schweizer (1999). Like the process X, better integrability conditions on the strategies ϕ are required. First we introduce a new class of processes.

Definition 5.15. *An* L^2*-strategy is a RM-strategy* $\phi = (\theta, \eta)$ *that satisfies the following conditions*

- *The process* θ *belongs to* $L(X)$ *where* $L(X)$ *is the set of all predictable processes* v *such that the* $\mathbb{E}\Big[|\int v\, dX|\Big] < \infty$.
- *The process* $\int \theta\, dX$ *is in* $\mathcal{S}^2(\mathbb{P})$.
- *The process* η *is a real valued adapted process such that the value process* $V(\phi) = \theta X + \eta$ *is right-continuous and square-integrable that means* $V_t(\phi) \in L^2(\Omega, \mathfrak{F}_t, \mathbb{P})$ *for each* $t \in [0,T]$.

The set of all processes θ that satisfies the first condition of the above definition is denoted by Θ_S.

The idea of local risk-minimization is that a good strategy still may not be globally risk-minimizing in terms of Definition 5.13, but at least locally, small perturbations of the portfolio should not give a better strategy to minimize the risk.

Definition 5.16. *A small perturbation is an L^2-strategy, $\Delta = (\delta, \epsilon)$ such that δ is bounded, the variation of $\int \delta dA$ is bounded (uniformly in t and ω) and $\delta_T = \epsilon_T = 0$. For any sub-interval $(s,t]$ of $[0,T]$, we then define the small perturbation*

$$\Delta|_{(s,t]} = (\delta 1_{(s,t]}, \epsilon 1_{[s,t)}).$$

Finally, we are now ready to present the exact definition of a local risk-minimizing strategy. The notion of local risk-minimization was first introduced by Schweizer (1991).

Definition 5.17. *For an L^2-strategy ϕ, a small perturbation Δ, and a partition $\mathfrak{p}$ of $[0,T]$, we set*

$$r^{\mathfrak{p}}(\phi, \Delta) = \sum_{t_i, t_{i+1} \in \mathfrak{p}} \frac{R_{t_i}(\phi + \Delta|_{(t_i, t_{i+1}]}) - R_{t_i}(\phi)}{\mathbb{E}[\langle M \rangle_{t_{i+1}} - \langle M \rangle_{t_i} | \mathfrak{F}_{t_i}]} 1_{(t_i, t_{i+1}]},$$

where $R_t(\phi)$ is given in (5.6) and $\langle M \rangle$ in Definition 5.11. Then ϕ is called locally risk-minimizing if

$$\lim_{n \to \infty} r^{\mathfrak{p}_n}(\phi, \Delta) \geq 0, \quad \mathbb{P} \times \langle M \rangle - \text{ almost surely on } \Omega \times [0,T],$$

for every small perturbation Δ and every increasing sequence $\mathfrak{p}$ of partitions tending to the identity.

Although intuitively clear, working with the above definition, in the context of martingale theory is cumbersome. Rather, we prefer a more tractable definition. Fortunately under the assumptions that X satisfies SC, A is continuous, and $\tilde{K}_T$ belongs to $L^1(\Omega, \mathfrak{F}_T, \mathbb{P})$, the notion of local risk-minimization is equivalent to a more acceptable and familiar one given below, see Schweizer (1999).

Definition 5.18. *Let $H \in L^2(\Omega, \mathfrak{F}_T, \mathbb{P})$ be a contingent claim. An L^2-strategy ϕ with $V_T(\phi) = H$, $\mathbb{P}-$almost surely is called pseudo-locally risk-minimizing or pseudo-optimal for H if ϕ is mean-self-financing and the martingale $C(\phi)$ (the cost process) is strongly orthogonal to M.*

The following lemma gives a necessary and sufficient condition for the existence of a pseudo-locally risk-minimizing strategy, see Föllmer and Schweizer (1991) for the proof.

Lemma 5.6. *A contingent claim $H \in L^2(\Omega, \mathfrak{F}_T, \mathbb{P})$ admits a pseudo-optimal L^2-strategy ϕ with $V_T(\phi) = H$, $\mathbb{P}-$almost surely, if and only if H can be represented as*

$$H = H_0 + \int_0^T \xi_u^H \, dX_u + L_T^H, \quad \mathbb{P} - \text{ almost surely}, \tag{5.7}$$

with $H_0 \in L^2(\Omega, \mathfrak{F}_0, \mathbb{P})$, $\xi^H \in \Theta_S$ and $L^H \in \mathcal{M}_0^2(\mathbb{P})$ strongly $\mathbb{P}-$ orthogonal to M. The strategy ϕ and the cost process C are then given by

$$\theta_t = \xi_t^H, \quad \text{for all } 0 \leq t \leq T,$$

and

$$C_t(\phi) = H_0 + L_t^H, \quad \textit{for all } 0 \leq t \leq T.$$

Its value process is

$$V_t(\phi) = C_t(\phi) + \int_0^t v_u dX_u = H_0 + \int_0^t \xi_u^H dX_u + L_t^H, \quad \textit{for all } 0 \leq t \leq T,$$

so that η is also determined by the above description.

As mentioned before, under appropriate assumptions on X, local risk-minimization is same as pseudo-locally risk-minimizing. Hence, this lemma gives a necessary and sufficient condition for the existence of a local risk-minimization strategy.

Remark 5.7. *The decomposition in* (5.7) *is called the Föllmer-Schweizer (FS) decomposition of H. It is proved that FS decomposition exists if the mean-variance trade-off process $\tilde{K}$ is bounded uniformly in t and ω, see Monat and Stricker (1995).*

Although the existence of local risk-minimization is proved by the previous lemma, it completely depends on the decomposition in (5.7). However, in some special cases there are constructive ways of finding the strategies explicitly. To our knowledge, the best is done for processes with continuous sample paths, especially diffusion processes.

For continuous processes, the well known method of minimal equivalent local martingale measure (MELMM) is applicable, but even in this case, the final answer depends on finding the GKW decomposition, which is not known in general. Fortunately, for our problem, we present a constructive way of finding these hedging strategies that

has the advantage of not using the complexity of MELMM. In the next section we study the structure of $\big(g(t,X_t)1_{\{\tau>t\}}\big)_{0\leq t\leq T}$, this is an essential step toward finding FS decomposition.

5.5 The Compensator of $\big(g(t,X_t)1_{\{\tau>t\}}\big)_{t\geq 0}$

Itô formula is an important tool in analyzing financial problems. Itô formula (or Dynkin formula) can be applied to find the compensator of the process $(g(t,X_t))_{t\geq 0}$ for a $C^{1,2}$ function $g=g(t,x)$. By using the optional sampling theorem one can actually do further and find the compensator of the process $(g(t\wedge\tau,X_{t\wedge\tau}))_{t\geq 0}$. A simple manipulation shows that for all $t\geq 0$ we have that

$$g(t\wedge\tau,X_{t\wedge\tau})=g(t,X_t)1_{\{\tau>t\}}+g(\tau,X_\tau)1_{\{\tau\leq t\}}.$$

Itô formula cannot be applied to find the compensators of the individual processes defined by the terms in the right-hand side of the above equation. Since we are working with defaultable processes, it is essential to know the compensator of these processes and especially $\big(g(t,X_t)1_{\{\tau>t\}}\big)_{t\geq 0}$.

In Chapter 4, we introduced the notion of intensity for the indicator process $\mathfrak{N}=(\mathfrak{N}_t)_{t\geq 0}$ where $\mathfrak{N}_t=1_{\{\tau\leq t\}}$. Remember that the intensity of this process leads to its compensator. In this section we introduce a similar concept for the process $\mathcal{P}=(\mathcal{P}_t)_{t\geq 0}$, where $\mathcal{P}_t=g(t,X_t)1_{\{\tau>t\}}$ and $g:[0,\infty)\times\mathbb{R}\to\mathbb{R}$ is a $C^{1,2}$ function. The problem at hand is finding the compensator of $\mathcal{P}$. This could be considered as a generalization of Chapter 4. From here on, we reserve the symbol $\mathfrak{N}$ for the above indicator process.

To achieve this generalization in our model, we use two methods, the method of the Laplacian approximation (see Theorem 4.1) and another one using a compensation formula.

5.5.1 The Laplacian approximation

The method of the Laplacian approximation can be found in the first edition of Dellacherie and Meyer (1982).

Assume that the process $\mathcal{P} = (\mathcal{P}_t)_{t\geq 0}$ satisfies the conditions of Theorem 4.1. Then to find the predictable part of $\mathcal{P}$, we need to calculate

$$\Lambda_t = \int_0^t \lim_{h\downarrow 0} \frac{\mathbb{E}[\mathcal{P}_{s+h} - \mathcal{P}_s|\mathfrak{F}_s]}{h}\, ds.$$

We illustrate the procedure to calculate the above limit for the process $X = U$, where U is a compound Poisson process plus a drift, i.e. $U_t = u + \mu t + \sum_{i=1}^{N_t} Y_i$, for $t \geq 0$. This is similar to the proof of Proposition 4.1. In what follows, whenever we say that "on the event A, $Expression_1 = Expression_2$", it means that

$$1_A\ Expression_1 = 1_A\ Expression_2.$$

Since the problem is set on the interval $[0, T]$, we can assume that $0 \leq t \leq T$. Let $\tilde{\mathcal{P}}_t = g(t, U_t)1_{\{\tau\leq t\}}$ and $A = \{\tau > t, T_n \leq t < T_{n+1}\}$ then on A,

$$\begin{aligned}\mathbb{E}[\tilde{\mathcal{P}}_{t+h} - \tilde{\mathcal{P}}_t|\mathfrak{F}_t^U] &= \mathbb{E}[g(t+h, U_{t+h})1_{\{\tau\leq t+h\}} - g(t, U_t)1_{\{\tau\leq t\}}|\mathfrak{F}_t^U] \\ &= \mathbb{E}[g(t+h, U_{t+h})|\mathfrak{F}_t^U] - \mathbb{E}[g(t+h, U_{t+h})1_{\{\tau>t+h\}}|\mathfrak{F}_t^U],\end{aligned}$$

for $h, t \geq 0$, and $t + h \leq T$. For each t, $T_n \leq t < T_{n+1}$, we define the σ-algebra

$$\mathfrak{G}_t = \sigma\left(\{T_1, ...T_n, Y_1, ..., Y_n\}\right).$$

Since for any t, the equality $\mathfrak{G}_t \cap A = \mathfrak{F}_t^U \cap A$ holds, by Lemma A.3, on A we have

$$\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} | \mathfrak{F}_t^U] = \frac{\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} | \mathfrak{G}_t]}{E[1_A | \mathfrak{G}_t]},$$

$$\mathbb{E}[\tilde{\mathcal{P}}_{t+h} - \tilde{\mathcal{P}}_t | \mathfrak{F}_t^U] = \mathbb{E}[g(t+h, U_{t+h}) | \mathfrak{F}_t^U] - \frac{\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} 1_A | \mathfrak{G}_t]}{\mathbb{E}[1_A | \mathfrak{G}_t]}.$$

In the last formula above, the numerator of the second term on the right-hand side can be written as the sum of three terms, depending if $T_n \leq t < t + h < T_{n+1}$, or else $T_{n+1} \leq t+h < T_{n+2}$, or finally $t+h \geq T_{n+2}$. The denominator can be written as $\mathbb{P}(\tau > t, T_n \leq t < T_{n+1} | \mathfrak{G}_t) = \mathfrak{X}\mathbb{P}(T_{n+1} > t | \mathfrak{G}_t) = \mathfrak{X} e^{-\lambda(t - T_n)}$, for $\mathfrak{X} = 1_{\{T_n \leq t\}} 1_{\bigcap_{i=1}^n \{u + \mu T_i + Z_i > 0\}}$ and $Z_i = \sum_{j=1}^i Y_j$. So

$$\mathbb{E}[\tilde{\mathcal{P}}_{t+h} - \tilde{\mathcal{P}}_t | \mathfrak{F}_t^U] = \mathbb{E}[g(t+h, U_{t+h}) | \mathfrak{F}_t^U] - \mathbb{I} - \mathbb{III} - \mathbb{IIII}, \tag{5.8}$$

where

$$\mathbb{I} = \frac{\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} 1_{\{T_n \leq t < t+h < T_{n+1}\}} | \mathfrak{G}_t]}{\mathbb{E}[1_A | \mathfrak{G}_t]},$$

$$\mathbb{III} = \frac{\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} 1_{\{T_n \leq t < T_{n+1} \leq t+h < T_{n+2}\}} | \mathfrak{G}_t]}{\mathbb{E}[1_A | \mathfrak{G}_t]},$$

$$\mathbb{IIII} = \frac{\mathbb{E}[g(t+h, U_{t+h}) 1_{\{\tau > t+h\}} 1_{\{T_n \leq t < T_{n+1} < T_{n+2} \leq t+h\}} | \mathfrak{G}_t]}{\mathbb{E}[1_A | \mathfrak{G}_t]}.$$

For the first term $\mathbb{I}$ above, on A we have

$$\begin{aligned} \mathbb{I} &= \frac{\mathfrak{X} g(t+h, u + \mu(t+h) + Z_n) e^{-\lambda(t+h-T_n)}}{\mathfrak{X} e^{-\lambda(t-T_n)}} \\ &= g(u + \mu(t+h) + Z_n, t+h) e^{-\lambda h}, \end{aligned} \tag{5.9}$$

by conditioning on T_{n+1} and Y_{n+1}, $\mathbb{III}$ becomes

$$\mathbb{III} = \frac{\mathfrak{X}\lambda e^{-\lambda(t+h-T_n)} \int_t^{t+h} \int_{-(u+\mu t_{n+1}+Z_n)}^{\infty} \zeta_h \mathbb{P}(Y_{n+1} \in dy_{n+1}) dt_{n+1}}{\mathfrak{X} e^{-\lambda(t-T_n)}},$$

where $\zeta_h = g(u + \mu(t+h) + Z_n + y_{n+1}, t+h)$, hence on A we have

$$\mathbb{III} = \lambda e^{-\lambda h} \int_t^{t+h} \int_{-(u+\mu t_{n+1}+Z_n)}^{\infty} \zeta_h \mathbb{P}(Y_{n+1} \in dy_{n+1}) dt_{n+1}, \tag{5.10}$$

A similar expression can be found for $\mathbb{IIII}$. By (5.8) to obtain $\lim_{h\downarrow 0} \frac{\mathbb{E}[\tilde{\mathcal{P}}_{t+h} - \tilde{\mathcal{P}}_t]}{h}$, first we need the following

$$\begin{aligned} \lim_{h\downarrow 0} \frac{\mathbb{E}[g(t+h, U_{t+h})|\mathfrak{F}_t] - \mathbb{I}}{h} = & \lim_{h\downarrow 0} \frac{\mathbb{E}[g(t+h, U_{t+h}) - g(t, U_t)|\mathfrak{F}_t]}{h} \\ & + \lim_{h\downarrow 0} \frac{g(t, U_t) - \mathbb{I}}{h}. \end{aligned} \tag{5.11}$$

Notice that the first term in the right-hand side of (5.11) is not always zero. In a special case, it is equal to zero when g has the following property, for some function K and constant h,

$$|g(t+h, x) - g(t, y)| \leq K(x, y) h^{\alpha}, \text{ for } \alpha > 1. \tag{5.12}$$

For example a constant $g = g(t, x)$ satisfies this condition. But in general the first term of (5.11) is not zero. For a bounded function $g = g(t, x)$, it is equal to

$$\lim_{h\downarrow 0} \frac{\mathbb{E}[\int_t^{t+h} \mathcal{A}g(v, U_v) dv|\mathfrak{F}_t^U]}{h} = \mathcal{A}g(t, U_t),$$

where we have used the conditional version of Lebesgue's dominated convergence theorem, and $\mathcal{A}$ is the generator of $g(t, x)$, satisfying

$$\mathcal{A}g(t, x) = \frac{\partial g}{\partial t}(t, x) + \mu \frac{\partial g}{\partial x}(t, x) + \lambda(\int_{-\infty}^{\infty} g(t, x+u) \mathbb{P}(Y_1 \in du) - g(t, x)). \tag{5.13}$$

Fore the moment leave the first term of (5.11) unchanged. A simple calculation shows that

$$\lim_{h\downarrow 0}\frac{g(t,U_t)-\mathbb{I}}{h}=-\mu\frac{\partial g}{\partial x}(t,U_t)-\frac{\partial g}{\partial t}(t,U_t)+\lambda g(t,U_t).$$

The above discussions yield to the following expressions on A,

$$\begin{aligned}\lim_{h\downarrow 0}\frac{\mathbb{E}[g(t+h,U_{t+h})|\mathfrak{F}_t^U]-\mathbb{I}}{h}=&\lim_{h\downarrow 0}\frac{\mathbb{E}[g(t+h,U_{t+h})-g(t,U_t)|\mathfrak{F}_t^U]}{h}\\&-\mu\frac{\partial g}{\partial x}(t,U_t)-\frac{\partial g}{\partial t}(t,U_t)+\lambda g(t,U_t).\end{aligned}$$

Also it can be shown that

$$\lim_{h\downarrow 0}\frac{\mathbb{III}}{h}=\lambda\int_{-U_t}^{\infty}g(t,U_t+u)\mathbb{P}[Y_1\in du],$$

and

$$\lim_{h\downarrow 0}\frac{\mathbb{IIII}}{h}=0.$$

This finally leads us to the point where, if the first term of (5.11) is zero then

$$\lim_{h\downarrow 0}\frac{\mathbb{E}[\tilde{\mathcal{P}}_{t+h}-\tilde{\mathcal{P}}_t|\mathfrak{F}_t^U]}{h}=-\mathfrak{A}g(t,U_t),\tag{5.14}$$

and otherwise

$$\lim_{h\downarrow 0}\frac{\mathbb{E}[\tilde{\mathcal{P}}_{t+h}-\tilde{\mathcal{P}}_t|\mathfrak{F}_t^U]}{h}=\lim_{h\downarrow 0}\frac{\mathbb{E}[g(t+h,U_{t+h})-g(t,U_t)|\mathfrak{F}_t^U]}{h}-\mathfrak{A}g(t,U_t),\tag{5.15}$$

where $\mathfrak{A}$ is operator given by

$$\mathfrak{A}g(t,x)=\mu\frac{\partial g}{\partial x}(t,x)+\frac{\partial g}{\partial t}(t,x)+\lambda(\int_{-x}^{\infty}g(t,x+u)\mathbb{P}(Y_1\in du)-g(t,x)).$$

Now from the identity

$$\mathbb{E}[\mathcal{P}_t - \mathcal{P}_{t+h}|\mathfrak{F}_t^U] = \mathbb{E}[\tilde{\mathcal{P}}_{t+h} - \tilde{\mathcal{P}}_t|\mathfrak{F}_t^U] - \mathbb{E}[g(t+h, U_{t+h}) - g(t, U_t)|\mathfrak{F}_t^U],$$

the previous derivations lead to the following lemma.

Lemma 5.7. *Suppose that* $\left(g(t, U_t)1_{\{\tau>t\}}\right)_{0\leq t\leq T}$ *satisfies the conditions of Theorem 4.1, then*

$$g(t, U_t)1_{\{\tau>t\}} - \int_0^t \mathfrak{A}g(s, U_s)1_{\{\tau>s\}}ds, \tag{5.16}$$

is an $\left\{\mathfrak{F}_t^U\right\}_{0\leq t\leq T}$*- martingale.*

Remark 5.8. *If the underlying process* U_t *is replaced by a Lévy process, the previous lemma lets us conjecture that we could have the same form as in* (5.16)*, but with a different generator. However there are some technical problems that must be considered. These are explained in Section 5.6.*

Note that if $g = g(t, x)$ is zero for $x < 0$, then the two operators $\mathfrak{A}$ and $\mathcal{A}$ are actually equal. Indeed in the expression $g(t, U_t)1_{\{\tau>t\}}$ since $\tau > t$, then $U_t > 0$. Therefore on $[0, T]$, the domain of the function g can be restricted to $[0, \infty) \times (0, \infty)$

An interesting special cases of Lemma 5.7 is for $g = 1$. This constant function satisfies all the conditions of the Lemma. It provides a decomposition of the indicator process $\left(1_{\{\tau>t\}}\right)_{t\geq 0}$ that was already obtained in Corollary 4.1 and also in Guo and Zeng (2008) but by a different method.

Here, we would like to clarify a possible confusion that may arise in working with

the indicator process $\left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$ (or $\left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$) for a stopping time $\mathcal{T}$. First we recall the general case of the Doob-Meyer decomposition theorem, see Protter, 2004.

Theorem 5.3. *Let Z be a càdlàg supermartingale with $Z_0 = 0$ of Class D. Then there exists a unique, increasing, predictable process Λ with $\Lambda_0 = 0$ such that $M_t = Z_t + \Lambda_t$ is a uniformly integrable martingale.*

Assume that $\mathcal{T} > 0$ and $\mathcal{T} < \infty$ almost surely. The indicator process $\left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$ is a càdlàg supermartingale, because $1_{\{\mathcal{T}>t\}} \leq 1_{\{\mathcal{T}>s\}}$ for all $t \geq s \geq 0$. To make it compatible with the previous theorem, we define $Z_t = 1_{\{\mathcal{T}>t\}} - 1$, hence $Z_0 = 0$, and it belongs to Class D. On the other hand $\left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$ is an increasing process that starts at zero. So it seems that the uniqueness in Theorem 5.3 provides a simple solution $\Lambda = \left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$ and then $M = 0$.

The only condition that we did not check and actually needs careful attention is the predictability of the indicator process $\left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$. Before proceeding, we mention a false argument that may be used to deal with this confusion. It may be thought that since the above process is càdlàg it cannot be predictable, by definition of predictability. We remind that a predictable σ-algebra is generated by all càg adapted processes. Therefore a càdlàg process can not be a generator of a predictable σ-algebra, but still it may be measurable with respect to the σ-algebra generated by all càg adapted processes. In what follows we solve this problem under two different perspectives.

Although not stated directly in Theorem 5.3, the Doob-Meyer decomposition com-

pletely depends on the underlying filtration. Considering different filtrations leads to different decompositions. The first perspective is to consider different levels of information or filtrations for the process Z, which is adapted to all filtrations.

Assume a situation when there is a lack of information. Under this assumption, most likely the default time (or the value of the indicator process $\left(1_{\{\mathcal{T}>t\}}\right)_{t\geq 0}$) would be a complete surprise for investors, or in our words, totally inaccessible or completely unpredictable. In this case the compensator of the process Z is not trivial any more.

On the other hand, if full information (or the largest filtration) is accessible, the predictability of the default time may also hold. In this situation the compensator can be a trivial predictable process. A very good example is provided by Artzner and Delbaen (1995). In their paper full information is generated by the underlying value process, which is a Brownian motion, and incomplete information is generated by the periodic values of the underlying process, which can be considered for example as the quarterly releases of the firm's data. In that case the compensator of the default indicator process is derived to be trivial under complete information and nontrivial under incomplete information.

In the second perspective we see that even having complete access to the information does not guarantee in all the cases a trivial compensator for the indicator process.

Having full access to the firm's data reduces the risk, but there might still be some residual systematic risks, beyond the firm's manager control. In the second perspective, we focus on the entity of the risk coming exogenously and imposed by the

global economic scenarios. Theoretically, if the underlying process is a Lévy process for example, then as we saw in Theorem 4.3, the default time $\mathcal{T} = \tau$ is neither a totally inaccessible nor a predictable stopping time. Now the confusion can be resolved by the following statement.

The random time $\mathcal{T}$ is predictable if and only if the process $\left(1_{\{\mathcal{T}\leq t\}}\right)_{t\geq 0}$ is predictable.

5.5.2 Finding the Compensator by Using a Compensation Formula

Now we discuss the second approach for finding the compensator of the process $\mathcal{P}$. Lets start by a simple observation. Assume that X and $\tilde{X}$ belong to $\mathscr{S}$ such that $X - \tilde{X}$ is a martingale. For example, if X is an increasing process in $\mathscr{A}_{loc}$ then by Doob-Meyer's decomposition there is such an increasing predictable process $\tilde{X}$. Since $X - \tilde{X}$ is a martingale, for any locally bounded predictable process H we obtain $\mathbb{E}[\int_0^t H\ d(X - \tilde{X})] = 0$. Assuming that $\mathbb{E}[\int_0^t H\ d\tilde{X}] < \infty$, we have the following

$$\mathbb{E}[\int_0^t H\ dX] = \mathbb{E}[\int_0^t H\ d\tilde{X}].$$

Note that since H is a locally bounded predictable process, the above integrals are well defined, see Jacod and Shiryaev (1987).

The converse of the above observation is true under appropriate assumptions. This can be stated in different versions. Here, we use the one that fits to our needs. Following Brémaud (1981), we have

Lemma 5.8. *Assume that X is a semimartingale in the probability space $(\Omega, \mathcal{F}, \mathfrak{F}, \mathbb{P})$ such that $\mathbb{E}[\int_0^\infty H_s \ dX_s] = \mathbb{E}[\int_0^\infty H_s \mathscr{I}_s \ ds]$ holds for all locally bounded predictable processes H. Further suppose that $\mathscr{I}$ is an $\mathfrak{F}$-adapted process for which $\mathbb{E}[\int_0^t |\mathscr{I}_s| \ ds] < \infty$ (or $\int_0^t |\mathscr{I}_s| \ ds < \infty$ $\mathbb{P}$- almost surely for all $t \geq 0$). Then the process $M = (M_t)_{t \geq 0}$ defined by*

$$M_t = X_t - \int_0^t \mathscr{I}_s \ ds, \ t \geq 0,$$

is an $\mathfrak{F}$-martingale (resp. an $\mathfrak{F}$-local martingale).

In what follows, we explain a category of Lévy processes for which the second method is applicable. This is a technical problem and the reasons for imposing it are brought up in the proof of Theorem 5.5.

Regarding the explanations provided following the Proposition 4.1, especially Theorem 4.3 (Meyer's previsibility theorem), it is important to be careful how default happens. If the underlying process is a pure jump process, then most probably, default happens by a sudden jump of the underlying process. The following definition and theorem guarantee that for a special type of bounded variation Lévy process, this is really the case, see Section 7.5 of Kyprianou (2006). This is one of the reasons that we focus on bounded variation Lévy processes.

Definition 5.19. *Assume that the process X is a Lévy process such that $X_0 = 0$ ($X_0 = u > 0$). Let the stopping time τ^+ be defined as*

$$\tau^+ = \inf\{t > 0; X_t > x\}.$$

Then X creeps over (creeps down) the level $x > 0$ ($x = 0$), when

$$\mathbb{P}(X_{\tau^+} = x) > 0 \qquad (\mathbb{P}(X_\tau = 0) > 0).$$

We also use creep upward and creep downward terminologies. The following theorem is part (i) of Theorem 7.11 of Kyprianou (2006) that gives necessary and sufficient conditions for a process to creep over or creep down.

Theorem 5.4. *Suppose that X is a bounded variation Lévy process which is not a compound Poisson process. Then X creeps upwards (downward) if and only if the process X has the following Lévy-Khintchine exponent*

$$\Psi(\theta) = -i\theta\mu + \int_{\mathbb{R}-\{0\}} (1 - e^{i\theta x}) v(dx),$$

for $\mu > 0$ ($\mu < 0$), and v is the Lévy measure.

Notice that in this case the Lévy process X can be represented as

$$X_t = \mu t + \int_0^t \int_{\mathbb{R}-\{0\}} x J_X(ds \times dx), \quad t \geq 0. \tag{5.17}$$

The first term of the above equation is the drift part and the second term is a pure jump process where J_X is the jump measure of the process X.

Theorem 5.5. *Assume that X is a bounded variation Lévy process given by* (5.17) *such that $\mu > 0$. Let $g : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be a $C^{1,2}$ function and suppose that both of the expectations $\mathbb{E}[\int_0^t \int_{-\infty}^{-X_s} |g(s, X_s)| \; v(dy) \; ds]$ and $\mathbb{E}[\int_0^t \int_{-\infty}^{-X_s} |g(s, X_s + y)| \; v(dy) \; ds]$ are finite for all $t \geq 0$. Then if*

$$\mathbb{E}\Big[\int_0^t \int_{\mathbb{R}} \big| (g(s, X_{s^-} + y) - g(s, X_{s^-})) \big| \; v(dy) \; ds \Big] < \infty,$$

(or $\int_0^t \int_{\mathbb{R}} |(g(s, X_{s^-} + y) - g(s, X_{s^-}))|\ v(dy)\ ds < \infty$*, almost surely), the process*

$$\left(g(t, X_t) 1_{\{\tau > t\}} - \int_0^t \mathfrak{A} g(s, X_s) 1_{\{\tau > s\}}\ ds \right)_{t \geq 0}, \tag{5.18}$$

is an $\mathfrak{F}^X$*- martingale (or* $\mathfrak{F}^X$*- local martingale), where*

$$\begin{aligned} \mathfrak{A} g(s, x) = \frac{\partial g}{\partial s}(s, x) + \mu \frac{\partial g}{\partial x}(s, x) - \int_{-\infty}^{-x} g(s, x + y)\ v(dy) \\ + \int_{-\infty}^{\infty} (g(s, x + y) - g(s, x))\ v(dy). \end{aligned} \tag{5.19}$$

Proof. Because the function g is a $C^{1,2}$ function, the process $(g(t, X_t))_{t \geq 0}$ is a semi-martingale and so by using the product formula (5.2), for $t \geq 0$ we have

$$\begin{aligned} g(t, X_t) 1_{\{\tau \leq t\}} = \int_0^t 1_{\{\tau < s\}}\ dg(s, X_s) + \int_0^t g(s^-, X_{s^-})\ d1_{\{\tau \leq s\}} \\ + [g(t, X_t), 1_{\{\tau \leq t\}}]. \end{aligned} \tag{5.20}$$

The compensator of $\tilde{\mathcal{P}} = \left(g(t, X_t) 1_{\{\tau \leq t\}}\right)_{t \geq 0}$ is the sum of the compensators of each of the three terms on the right-hand side of the above equation.

Since g is a $C^{1,2}$ function (in fact for this part $C^{1,1}$ is enough), by applying Itô's formula, we have

$$\begin{aligned} g(t, X_t) = g(0, X_0) + \int_0^t \frac{\partial g}{\partial s}(s, X_s)\ ds + \mu \int_0^t \frac{\partial g}{\partial x}(s, X_s)\ ds \\ + \int_0^t \int_{\mathbb{R}} (g(s, X_{s^-} + y) - g(s, X_{s^-}))\ J_X(ds \times dy). \end{aligned}$$

For a proof of this, see Theorem 4.2 of Kyprianou (2006). By the compensation formula we get

$$\begin{aligned} \mathbb{E}[\int_0^t \int_{\mathbb{R}} H_s (g(s, X_{s^-} + y) - g(s, X_{s^-}))\ J_X(ds \times dy)] = \\ \mathbb{E}[\int_0^t \int_{\mathbb{R}} H_s (g(s, X_{s^-} + y) - g(s, X_{s^-}))\ v(dy) ds], \end{aligned}$$

for all bounded non-negative predictable processes H. Hence by Lemma 5.8 and integrability assumptions, we have the decomposition $g(t, X_t) = M_t + \Lambda^g_t,\ t \geq 0$, where M is an $\mathfrak{F}^X$- martingale (an $\mathfrak{F}^X$- local martingale) and Λ^g is a finite variation process. The process Λ^g is given by $\Lambda^g_t = \int_0^t \mathcal{A}g(s, X_s)\, ds$, the operator $\mathcal{A}$ is defined by

$$\mathcal{A}g(s,x) = \frac{\partial g}{\partial s}(s,x) + \mu \frac{\partial g}{\partial x}(s,x) + \int_{\mathbb{R}} (g(s, x+y) - g(s,x))\ v(dy),$$

and v is the Lévy measure of the process X. Therefore

$$\int_0^t 1_{\{\tau<s\}}\ dg(s, X_s) = \int_0^t 1_{\{\tau<s\}}\ dM_t + \int_0^t 1_{\{\tau<s\}} \mathcal{A}g(s, X_s)\ ds,$$

since the first term of the right-hand side of the above is a (local) martingale the compensator of the first term of (5.20) is given by

$$\left(\int_0^t 1_{\{\tau<s\}} \mathcal{A}g(s, X_s)\ ds \right)_{t\geq 0}. \tag{5.21}$$

To find the compensator of the second term of (5.20), since by Theorem 5.4, the process X does not creep downward, the stopping time τ is now totally inaccessible. Therefore one can extend the Laplacian approximation method in Chapter 4 (because now the compensator Λ in Theorem 4.1 is continuous), or apply the result of Guo and Zeng (2008). From the assumption $\mathbb{E}[\int_0^t \int_{-\infty}^{-X_s} |g(s, X_s)|\ v(dy)ds] < \infty$, hence and by very minor modifications of Guo and Zeng (2008), the compensator of the second term of (5.20) is given by

$$\left(\int_0^t \int_{-\infty}^{-X_s} \Big(g(s, X_s) 1_{\{\tau>s\}}\ v(dy) \Big)\ ds \right)_{t\geq 0}. \tag{5.22}$$

The main challenge is to find the compensator of the third term of (5.20). The indicator process $\left(1_{\{\tau\leq t\}}\right)_{t\geq 0}$ is a finite variation process. Then by part (a) of Proposition 5.2, we obtain

$$[g(t,X_t),1_{\{\tau\leq t\}}]=\int_0^t \Delta g(s,X_s)\; d1_{\{\tau\leq s\}}. \tag{5.23}$$

By Lemma 5.8, to obtain the compensator of the above process, we need to calculate the following expectation

$$\mathbb{E}\left[\int_0^\infty H_s\; d[g(s,X_s),1_{\{\tau\leq s\}}]\right]=\mathbb{E}\left[\int_0^\infty H_s\Delta g(s,X_s)\; d1_{\{\tau\leq s\}}\right],$$

for an arbitrary bounded non-negative predictable process H. The calculations of this expectation are almost the same lines as Guo and Zeng (2008), where the compensation formula is used. From there, we obtain that $\mathbb{E}\left[\int_0^\infty H_s\; d[g(s,X_s),1_{\{\tau\leq s\}}]\right]$ is equal to

$$\mathbb{E}\left[\int_0^\infty H_s 1_{\{\tau>s\}}\int_{-\infty}^0 (g(s,y)-g(s,X_s))v(dy-X_s)\; ds\right],$$

and this expectation is finite and well defined by the integrability conditions in our assumptions. Hence by Lemma 5.8, the compensator of $[g(t,X_t),1_{\{\tau\leq t\}}]$ is given by

$$\left(\int_0^t\int_{-\infty}^0 (g(s,y)-g(s,X_s))1_{\{\tau>s\}}v(dy-X_s)\; ds\right)_{t\geq 0}. \tag{5.24}$$

The compensator of $\left(g(t,X_t)1_{\{\tau\leq t\}}\right)_{t\geq 0}$ is concluded from Equations (5.21), (5.22), and (5.24), and it is equal to

$$\begin{aligned}\Big(\int_0^t 1_{\{\tau<s\}}\mathcal{A}g(s,X_s)\; ds+\int_0^t\int_{-\infty}^{-X_s} 1_{\{\tau>s\}}g(s,X_s)v(dy)ds\\ +\int_0^t\int_{-\infty}^0 \big(g(s,y)-g(s,X_s)\big)1_{\{\tau>s\}}v(dy-X_s)ds\Big)_{t\geq 0}.\end{aligned}$$

Notice that in any of the above integrands, the strict inequality of the indicator process can be replaced by equality, because the Lebesgue measure ds does not charge $\{s; s = \tau\}$. From the above equation and since $g(t, X_t) = g(t, X_t)1_{\{\tau \leq t\}} + g(t, X_t)1_{\{\tau > t\}}$, after some manipulations the compensator of $g(t, X_t)1_{\{\tau > t\}}$ is equal to

$$\left(\int_0^t \mathfrak{A}g(s, X_s)1_{\{\tau > s\}} \, ds \right)_{t \geq 0},$$

where $\mathfrak{A}g(s, x)$ is given by (5.19). □

Remark 5.9. *Depending on the function g, the integrability conditions can be replaced by simpler ones. For instance if the function g and its derivatives are bounded.*

Remark 5.10. *Note that the operator given by* (5.19) *is not the same as Dynkin's or Itô's operator, and for the process U, Theorem 5.5 reduces to Lemma 5.7.*

The following corollary is the result of Theorem 5.5. Although we do not use it, it can be an interesting result in martingale theory.

Corollary 5.3. *Assume that X is a bounded variation Lévy process given by* (5.17) *such that $\mu > 0$. Let $g : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be a $C^{1,2}$ function and suppose that both of the expectations $\mathbb{E}[\int_0^t \int_{-\infty}^{-X_s} |g(s, X_s)| \; v(dy) \; ds]$ and $\mathbb{E}[\int_0^t \int_{-\infty}^{-X_s} |g(s, X_s + y)| \; v(dy) \; ds]$ are finite. Then if $\mathbb{E}[\int_0^t \int_{\mathbb{R}} |(g(s, X_s + y) - g(s, X_s))| \; v(dy) \; ds] < \infty$ (or $\int_0^t \int_{\mathbb{R}} |(g(s, X_s + y) - g(s, X_s))| \; v(dy) \; ds < \infty$, almost surely), the following process is an $\mathfrak{F}^X$- martingale (or $\mathfrak{F}^X$- local martingale),*

$$\left(g(\tau, X_\tau)1_{\{\tau \leq t\}} - \int_0^{t \wedge \tau} \Big(\int_{-\infty}^{-X_s} g(X_s + y, s)v(dy) \Big) \, ds \right)_{t \geq 0}.$$

Proof. By Theorem 5.5 and Dynkin's formula (or Itô's formula), both of the processes

$$\left(g(t, X_t) 1_{\{\tau > t\}} - \int_0^t \mathfrak{A} g(s, X_s) 1_{\{\tau > s\}} \, ds \right)_{t \geq 0},$$

and

$$\left(g(t, X_t) - \int_0^t \mathcal{A} g(s, X_s) \, ds \right)_{t \geq 0},$$

are martingales. Then it follows from the optional sampling theorem that the process $\left(g(t \wedge \tau, X_{t \wedge \tau}) - \int_0^{t \wedge \tau} \mathcal{A} g(s, X_s) \, ds \right)_{t \geq 0}$ is also a martingale. A simple calculation shows that $g(\tau, X_\tau) 1_{\{\tau \leq t\}} = g(t \wedge \tau, X_{t \wedge \tau}) - g(t, X_t) 1_{\{\tau > t\}}$. We have the compensators of both terms on the right-hand side, therefore

$$\left(g(\tau, X_\tau) 1_{\{\tau \leq t\}} - \int_0^t (\mathcal{A} g(s, X_s) - \mathfrak{A} g(s, X_s)) 1_{\{\tau > s\}} \, ds \right)_{t \geq 0}$$

must be a martingale too. Now the subtractions of the two operators gives the result.

□

With this corollary together with Itô's formula, under some integrability conditions for the process X mentioned above, one can find the compensators of the processes $(g(t \wedge \tau, X_{t \wedge \tau}))_{t \geq 0}$, $\left(g(t, X_t) 1_{\{\tau > t\}} \right)_{t \geq 0}$, and $\left(g(\tau, X_\tau) 1_{\{\tau \leq t\}} \right)_{t \geq 0}$.

Although the second approach seems to be more practical, the first approach intuitively makes more sense as it is based on the classical definition of default rate. The first approach together with the Lemma 5.8 can be still applicable to diffusion terms. In other words, the first approach can help guessing the compensator and Lemma 5.8 can be used to verify it. However, if one applies Theorem 4.1 directly, then the process

$\mathcal{P}$ must satisfy the Theorem conditions. The next section concentrates on sufficient conditions on g, for which Theorem 4.1 is applicable to the process $\mathcal{P}$, but is mainly reviewing the class D of processes.

5.6 Class D of processes

There are a few conditions in Theorem 4.1 that the process $\mathcal{P} = (\mathcal{P}_t)_{t\geq 0}$, $\mathcal{P}_t = g(t, X_t)$ has to satisfy. The most important one is belonging to Class D. First we explain the other conditions and then we discuss Class D of processes. As mentioned in the introduction, since the life time of the products is the maturity T, one can assume that $\tau < \infty$ almost surely. We try to find a quite large family of stochastic processes that satisfy the conditions of Theorem 4.1.

First we deal with functions $g = g(t, x)$ that make $\mathcal{P}$ a supermartingale. Since the process $\tilde{\mathfrak{N}} = \left(\tilde{\mathfrak{N}}_t\right)_{t\geq 0}$, $\tilde{\mathfrak{N}}_t = 1_{\{\tau > t\}}$ is decreasing, $\mathcal{P}$ is a supermartingale if $(g(t, X_t))_{t\geq 0}$ is a supermartingale.

Assume that X is a Lévy process with Lévy triplet (σ^2, v, γ). Suppose that g is a $C^{1,2}$ function. Then under appropriate integrability conditions, by Itô's formula, we have

$$g(t, X_t) = M_t + \Lambda^g_t, \quad t \geq 0.$$

The process $M = (M_t)_{t\geq 0}$ is a martingale given by

$$M_t = \int_0^t \int_{-\infty}^{\infty} [g(s, X_{s^-} + y) - g(s, X_{s^-})] \tilde{J}_X(ds\ dy),$$

$$\tilde{J}_X(ds\ dy) = J_X(ds\ dy) - v(dy)\ ds,$$

and the process $\Lambda = (\Lambda_t)_{t\geq 0}$ is a continuous finite variation one given by

$$\Lambda_t = \int_0^t \mathcal{A}g(s, X_s)\, ds,$$

where for $s > 0$ and $x \in \mathbb{R}$

$$\begin{aligned}\mathcal{A}(s,x) = \frac{\partial g}{\partial s}(s,x) + \gamma\frac{\partial g}{\partial x}(s,x) + \frac{\sigma^2}{2}\frac{\partial^2 g}{\partial x^2}(s,x) \\ + \int_{-\infty}^{\infty}\Big(g(x+y,s) - g(x,s) - y\frac{\partial g}{\partial x}(s,x)1_{\{|y|\leq 1\}}\Big)\, v(dy).\end{aligned}$$

The integrability conditions should make all the parts well defined, for instance they can be given by the following,

$$\begin{aligned}\int |y|\, v(dy) < \infty, \\ \mathbb{E}[\int_0^t |\frac{\partial g}{\partial x}(s, X_s)|\, ds] < \infty, \text{ for all } t \geq 0, \\ \mathbb{E}[\int_0^t \int |g(s, X_s + y) - g(s, X_s)|\, ds\, v(dy)] < \infty, \text{ for all } t \geq 0. \qquad (5.25)\end{aligned}$$

Notice that the last two conditions let us to write the above martingale drift representation. These conditions are not needed to apply Itô's formula for a Lévy process. The $C^{1,2}$ condition allows us to use Itô's formula and depending on the process it can be eased. A general form of Itô's formula is still valid for convex functions, see He, Wang and Yan (1992).

Example 5.4. *Consider $X_t = U_t = u + \mu t + \sum_{i=1}^{N_t} Y_i$. Suppose that g is $C^{1,1}$ and $\mathbb{E}\left[\int_0^t |(g(t, U_t + y) - g(t, U_t))| F_Y(dy)\right] < \infty$, then we have*

$$g(t, U_t) = M_t + \Lambda_t^g, \quad t \geq 0.$$

The process $M = (M_t)_{t\geq 0}$ *is a martingale given by*

$$M_t = \int_0^t \int_{-\infty}^{\infty} [g(s, U_{s^-} + y) - g(s, U_{s^-})] \; \tilde{J}_X(ds \; dy), \quad t \geq 0,$$

$$\tilde{J}_X(ds \; dy) = J_X(ds \; dy) - \lambda F_Y(dy) \; ds,$$

and the process $\Lambda^g = (\Lambda^g_t)_{t\geq 0}$ *is a continuous finite variation one given by*

$$\Lambda^g_t = \int_0^t \mathcal{A}^U g(s, X_s) \; ds, \quad \geq 0,$$

where $\mathcal{A}^U g(s, x)$ *is equal to*

$$\frac{\partial g}{\partial s}(s, x) + \mu \frac{\partial g}{\partial x}(s, x) + \lambda \left(\int_{-\infty}^{\infty} g(s, x + y) F_Y(dy) - g(s, x) \right). \tag{5.26}$$

Note that given the condition $\mathbb{E}\left[\int_0^t |(g(t, U_t + y) - g(t, U_t))| F_Y(dy)\right] < \infty$*, the martingale part is well defined.*

In case of the process U, the following provide slightly different conditions that still assure a decomposition as above for the process U, see Rolski et al. (1999).

1. $g = g(t, x)$ is absolutely continuous with respect to t and $g(t, x) = 0$ for $x \leq 0$,
2. $\mathbb{E}\left[\sum_{n, T_n \leq t} |g(T_n, U_{T_n}) - g(T_n^-, U_{T_n^-})|\right] < \infty$,
3. $\mathcal{A}^U g(t, x) \leq 0$ for all $t \geq 0$ and x.

If Λ^g is decreasing, then $(g(t, X_t))_{t\geq 0}$ is a supermartingale. To have Λ decreasing it is sufficient to have $\mathcal{A}g(t, x) \leq 0$ for all $t \geq 0$ and x.

The other two important conditions in Theorem 4.1 are

$$\lim_{t\to\infty} \mathbb{E}[\mathcal{P}_t] = 0$$

and the process $\mathcal{P}$ has to be in class D. If we impose some further assumptions on $g = g(t,x)$ then there is a simple answer for the first condition.

Lemma 5.9. *Assume that* $g(T, X_T) \in L^2$ *then* $\lim_{t\to\infty} \mathbb{E}[\mathcal{P}_t] = 0$.

Proof. For $t > T$,

$$\mathbb{E}[g(T, X_T)1_{\{\tau>T\}}] \leq ||g(X_T, T)||_2 \sqrt{\mathbb{P}[\tau > t]}.$$

We know that $\tau < \infty$ almost surely, hence the result follows. □

Remark 5.11. *In fact the above lemma works for any* L^p *space,* $p > 1$.

Answering the second condition is more challenging. We start by a simple observation.

Lemma 5.10. *If for all* $0 \leq t \leq T$, $g(t, X_t) \in L^1(\Omega, \mathfrak{F}_t, \mathbb{P})$ *and the process* $(g(t, X_t))_{t\geq 0}$ *is of class DL (see Definition A.6 of the Appendix) then* $\mathcal{P}$ *is of class D as well.*

Proof. We apply the definition of class D. For $x > 0$ and any finite-valued stopping

time s we have

$$\begin{aligned}\int_{|F(s,X_s)|>x} |F(s,X_s)|d\,\mathbb{P} \leq & \int_{|F(s,X_s)|>x, s\leq T} |F(s,X_s)|d\,\mathbb{P} \\ & + \int_{|F(s,X_s)|>x, s>T} |F(s,X_s)|d\,\mathbb{P} \\ \leq & \int_{|g(s,X_s)|>x, s\leq T} |g(s,X_s)|d\,\mathbb{P} \\ & + \int_{|g(T,X_T)|>x} |g(T,X_T)|d\,\mathbb{P}\end{aligned}$$

therefore

$$\begin{aligned}\sup_s \int_{|F(s,X_s)|>x} |F(s,X_s)|\,d\,\mathbb{P} \leq & \sup_s \int_{\{|g(s,X_s)|>x, s\leq T\}} |g(s,X_s)|d\,\mathbb{P} \\ & + \int_{|g(T,X_T)|>x} |g(T,X_T)|d\,\mathbb{P},\end{aligned}$$

Now if we let $x \to \infty$, the first term on the right-hand side goes to zero because $(g(t,X_t))_{t\geq 0}$ is assumed to be of class DL, and the second term also goes to zero because $g(T,X_T) \in L^1(\Omega, \mathfrak{F}_T, \mathbb{P})$. This show that $\mathcal{P}$ is of class D. □

So the question is reduced to when is $(g(t,X_t))_{t\geq 0}$ of class DL?

The first necessary condition on $(g(t,X_t))_{t\geq 0}$ is uniform integrability on a finite interval. In Meyer (1962), the author writes that "We have never met a right continuous supermartingale, uniformly integrable on an interval $[0,a]$ and not belonging to the class D on it." Although uniform integrability is a strong condition, unfortunately there are examples of uniformly integrable supermartingales which are not of class D on $[0,\infty)$, see Johnson and Helms (1963) (for class DL see the next remark). However, it is not

difficult to prove that the following important cases all belong to class DL. For the first three, see Meyer (1962), and for the last one see Karatzas and Shreve (1988).

Lemma 5.11. *1. Any right continuous martingale belongs to the class DL.*

2. *Any right continuous supermartingale which is bounded from above, belongs to the class DL.*

3. *Any right continuous supermartingale, which belongs to the class DL and is uniformly integrable, belongs to the class D.*

4. *Assume that X is a right continuous supermartingale, and $X = M + \Lambda$, where M is a martingale and Λ decreasing, then X is of class DL.*

Remark 5.12. *The third part of the previous lemma together with the example of Johnson and Helms (1963) shows that there are uniformly integrable, supermartingales that do not belong to class DL.*

Remark 5.13. *The third and forth part of the previous lemma show that if X is decomposable and uniformly integrable then it belongs to class D. Meyer (1962) proves that the converse also holds which is known as Doob-Meyer's decomposition.*

Regarding the previous discussion we present the following lemma, that gives a relatively simple criteria to make $\mathcal{P}$ of class D. There may be more satisfactory conditions, but this is the best we have done so far.

Lemma 5.12. *Assume that the underlying process X is a Lévy process with Lévy triplet (σ^2, v, γ). Suppose that $g = g(t,x)$ is a $C^{1,2}$ function such that the integrability conditions (5.25) are satisfied and $\mathcal{A}g(t,x) \leq 0$ for all $t \geq 0$ and x. Then the process $(g(t,X_t))_{t\geq 0}$ belongs to class D.*

Proof. Lemma 5.10 shows that it is enough to prove that $(g(t,X_t))_{t\geq 0}$ is of class DL. Since $\mathcal{A}g(t,x) \leq 0$, the process $(g(t,X_t))_{t\geq 0}$ is a supermartingale. We saw that $g(t,X_t) = M_t + \Lambda_t^g$, for $t \geq 0$, where M is a martingale and Λ is decreasing. On the other hand part four of Lemma 5.11 shows that $(g(t,X_t))_{t\geq 0}$ is of class DL. □

Although this Lemma serves well for our purpose, we would like to mention some more useful results. The following result shows that class D for potentials (see Definition A.5 of the Appendix) has a less complicated structure, see Meyer (1962).

Proposition 5.5. *Let X be a potential then it belongs to class D if and only if for any increasing sequence τ_n of stopping times, which almost surely increases to infinity, we have $\lim_{n\to\infty} \mathbb{E}[X_{\tau_n}] = 0$.*

Using this proposition, one can show that $\mathcal{P}$ belongs to class D if

$$\mathbb{E}\left[\left|\sup_{0\leq t\leq T} g(t,X_t)\right|\right] < \infty.$$

Remark 5.14. *If $(g(t,X_t))_{t\geq 0}$ is bounded from above by an integrable random variable (especially when g(t,x) is bounded by a function of t) then by the previous result it belongs to class D.*

Remark 5.15. *The uniform integrability, and therefore, belonging to class D, is related to the sample path property of the process. It can be shown that a non-negative supermartingale X, with continuous sample paths, is uniformly integrable if and only if* $\lim_{n\to\infty} n\mathbb{P}[\sup_{0\le t\le\infty} X_t > n] = 0$.

And finally we mention the following theorem of Dellacherie and Meyer (1982).

Theorem 5.6. *Let X be a positive right continuous supermartingale, and let* $R_n = \inf\{t; X_t \ge n\}$. *Then X belongs to class D if and only if* $\lim_{n\to\infty} \mathbb{E}[X_{R_n} 1_{\{R_n<\infty\}}] = 0$.

Had we imposed boundness on g from the beginning, we basically would not need this section, as in this case one can simply prove that $(g(t, X_t)1_{\tau>t})_{t\ge0}$ is of class D. Although the boundness of g simplifies the analytical calculations, it prevents simple cases like $g(t, x) = x$.

5.7 Hedging Strategies for the Defaultable Claims

In Section 5.4, we introduced the concept of local risk-minimization hedging. In this section our goal is to obtain these hedging strategies for the credit sensitive security with payoff (5.1)

$$F(X_T)1_{\{\tau>T\}},$$

where $\tau = \inf\{t; X_t < 0\}$, $F : \mathbb{R} \to \mathbb{R}$ a function, and $T > 0$ is the maturity or expiration of the security. If the set $\{t; X_t < 0\}$ is empty, then $\tau = \infty$.

To get an idea of the risk-minimization approach that we will be using later, for

a moment assume that the process $\big(F(X_t)1_{\{\tau>t\}}\big)_{0\leq t\leq T}$ is a martingale and we are in a risk neutral world. If the function F is C^2, sufficient conditions for this can be given either by Lemma 5.7 or Theorem 5.5. Then, by Proposition 5.4, the problem of finding the hedging strategy is reduced to find the appropriate decomposition of $\big(F(X_t)1_{\{\tau>t\}}\big)_{0\leq t\leq T}$. So in this setting, finding this decomposition is directly related to obtain the conditional quadratic covariation between the underlying process X and the process $\big(F(X_t)1_{\{\tau>t\}}\big)_{0\leq t\leq T}$. In what follows by using results of Section 5.4, we present a method to obtain local risk-minimization hedging in a more general setting under a physical measure.

It was mentioned in Section 5.4 that if the underlying process X is a (local) martingale, local risk-minimization reduces to risk-minimization and the existence of the hedging strategies is solved by a GKW decomposition. When the process X is a semimartingale then risk-minimization is no longer valid. It must be improved to local risk-minimization and the hedging strategies are solved by the FS decomposition (5.7). Because of the important roles of these two decompositions, we start by an historical review of them.

The GKW decomposition is essentially the decomposition given by Proposition 5.4. However, as we see below, square integrability is too strong a condition for the existence of this decomposition. The FS decomposition was first introduced by Föllmer and Schweizer (1991). They use Girsanov's transformation and change the original physical measure $\mathbb{P}$ to the so called equivalent (local) minimal martingale measure $\tilde{\mathbb{P}}$ under which

the underlying process is a martingale. They then use the GKW decomposition to find the FS decomposition. However their approach essentially works if the underlying process has continuous sample paths.

The existence of the FS decomposition of a square-integrable claim is proved even for a d-dimensional semimartingale X by Schweizer (1994) assuming that the process X satisfies SC condition and the MVT process is uniformly bounded in ω (ω belongs to Ω) and t and has jumps strictly bounded by 1 from above. Monat and Stricker (1994) prove the existence of the FS decomposition just by assuming that the MVT process is uniformly bounded in ω and t. Under this condition, Monat and Stricker (1995) further prove also the uniqueness.

Choulli, Krawczyk and Stricker (1998) find necessary and sufficient conditions for the existence and uniqueness of the FS decomposition by introducing a new notion of martingale. They prove that there is an FS decomposition for a square-integrable claim under the semimartingale $X = X_0 + M + \Lambda$, if first, the process $\mathcal{E}(-\int \zeta \ dM)$ satisfies an integrability condition and second if it is "regular". Please see the paper for the definition of "regular". Here the process $\mathcal{E}(-\int \zeta \ dM)$ is the Doléans - Dade exponential process (see Protter (2004)) and the process ζ is defined in Definition 5.14.

Choulli, Vandaele and Vanmaele (2010) discuss the relationship between the GKW and FS decompositions assuming that $\mathcal{E}(-\int \zeta \ dM)$ is strictly positive. Under this assumption, they find an explicit form of the FS decomposition based on GKW's decomposition.

In summary finding the FS decomposition practically works if there is at least one of the following conditions:

- The process X is a (local) martingale under $\mathbb{P}$.
- The process X is continuous.
- The process $\mathcal{E}(-\int \zeta \, dM)$ is strictly positive.
- The contingent claim is not path dependent.

Unfortunately neither one of these conditions holds in our model. Even in the above cases, the existence of GKW's decomposition is normally assumed. We choose a different method to find the FS decomposition and we also avoid changing the physical measure to find this decomposition. One advantage of this approach is that it does not relying on the MELMM method which is one of the main constructive ways to obtain hedging strategies. To start let us recall the GKW decomposition.

Assume that the processes X and Z belong to $\mathcal{M}^2_{loc}$ on $[0, T]$. Then by GKW's decomposition there is a predictable process ξ^Z and a martingale L^Z, strongly orthogonal to X, such that

$$Z = Z_0 + \int \xi^Z \, dX + L^Z,$$

and the process ξ is given by

$$\xi^Z = \frac{d\langle Z, X\rangle}{d\langle X, X\rangle}. \tag{5.27}$$

The superscript Z emphasizes that the decomposition depends on the process Z. Also it is worth mentioning that this decomposition is still valid under milder conditions. For instance, it is enough to have the integral $\int \xi \, dX$ well defined, Z, X, and $\int \xi \, dX$ as local martingales, and $[Z, X]$ in $\mathscr{A}_{loc}$. In $\mathcal{M}^2_{loc}$, all these conditions are satisfied.

Although formula (5.27) gives the GKW decomposition, the main task is of course computing the conditional quadratic covariations that are involved. However, in our setup, when we use GKW decomposition it is feasible to obtain these quantities.

The locally risk-minimizing strategy is linked to the FS decomposition. Hence, our aim is to find the FS decomposition (5.7). To reach this goal, first in the next theorem, we obtain a decomposition very close to the FS decomposition and, in fact, more general. To find this decomposition we use the GKW decomposition. Then we use this theorem to get our FS decomposition. This theorem is also used in Section 5.8. Before stating the theorem, we explain the conditions on the underlying process X and also an assumption that is used in the theorem.

For the rest of this section assume that the process X is a bounded variation Lévy process starting at the initial point $u > 0$ with Lévy triplet given by $(\gamma, 0, \upsilon)$, where υ is the Lévy measure. It is also assumed that the process X never creeps downward. Notice that X has the following representation

$$X_t = u + \mu t + \int_{[0,t]\times\mathbb{R}-\{0\}} x \; J_X(ds \times dx), \quad t \geq 0,$$

where J_X is the jump measure of X and the linear drift term μ is equal to $\gamma - \int_{-1}^{1} x \; \upsilon(dx)$

that is strictly positive. Recall that when the Lévy process X is of finite variation then $\int_{-1}^{1} |x|\ v(dx) < \infty$. In this section it is also assumed that $\mu > 0$. These assumptions let us use Theorem 5.5. Finally for the reason that is explained below, we further assume that $\int_{|x|\geq 1} |x|\ v(dx) < \infty$. Therefore, the Lévy measure v satisfies $\int |x|\ v(dx) < \infty$.

Since $\int |x|\ v(dx) < \infty$, we saw in Section 5.6 (or one can verify easily) that the process X can be written as $X = M + \Lambda$, where M is a martingale and Λ is a continuous finite variation process given by

$$\Lambda_t = \mu t + \int_0^t \int_{-\infty}^{\infty} y\ v(dy)\ ds, \quad t \geq 0.$$

Assumption 5.1. *Given a convex function $F = F(x)$, it is assumed that there is a $C^{1,2}$ function $f = f(t,x)$ that is the solution of the following PIDE*

$$\mathfrak{A}f(t,x) = \frac{(\mathfrak{A}K(t,x) - x\mathfrak{A}f(t,x) - \beta f(t,x))}{\int_{-\infty}^{\infty} y^2\ v(dy)}\beta, \quad \textit{for all } 0 \leq t \leq T, \tag{5.28}$$

and

$$f(T,x) = F(x), \textit{ for all real numbers } x,$$

where $K(t,x) = xf(t,x)$, $\beta = \mu + \int_{-\infty}^{\infty} y\ v(dy)$, and the operator $\mathfrak{A}$ is introduced in (5.19)*. It is also assumed that f and K satisfy the following integrability conditions for all $0 \leq t \leq T$:*

- $\mathbb{E}\left[\int_0^t \int_{-\infty}^{-X_s} |h(s, X_s)|\ v(dy)\ ds\right] < \infty$,
- $\mathbb{E}\left[\int_0^t \int_{-\infty}^{-X_s} |h(s, X_s + y)|\ v(dy)\ ds\right] < \infty$,
- $\int_0^t \int_{\mathbb{R}} |(h(s, X_{s^-} + y) - h(s, X_{s^-}))|\ v(dy)\ ds < \infty$, *almost surely,*

where the function h is either f or K.

For the rest of this section, the process $\big(f(t,X_t)1_{\{\tau>t\}}\big)_{0\le t\le T}$ is represented by $Z=(Z_t)_{0\le t\le T}$. Also let the process $\theta=(\theta_t)_{0\le t\le T}$ be given by

$$\theta_t=\frac{(\mathfrak{A}K(t^-,X_{t^-})-X_{t^-}\mathfrak{A}f(t^-,X_{t^-})-\beta f(t^-,X_{t^-}))}{\int_{\mathbb{R}}y^2\ v(dy)}1_{\{\tau\ge t\}}, \tag{5.29}$$

where the operator $\mathfrak{A}$ is introduced in (5.19), the functions $K=K(t,x)$ and $f=f(t,x)$ are defined in Assumption 5.1 and $\beta=\mu+\int_{-\infty}^{\infty}y\ v(dy)$. Notice that the process θ implicitly depends on the function $F=F(x)$.

Theorem 5.7. *Assume that X is the above mentioned Lévy process. Let the function $F=F(x)$ satisfy Assumption 5.1. We further suppose that the process $[Z,X]$ belongs to $\mathscr{A}_{loc}$. Then for all $0\le t\le T$, the following decomposition holds*

$$f(t,X_t)1_{\{\tau>t\}}=f(0,X_0)+\int_0^t\theta_s\ dX_s+L_t, \tag{5.30}$$

and specifically for $t=T$, one obtains

$$F(X_T)1_{\{\tau>T\}}=f(0,X_0)+\int_0^T\theta_s\ dX_s+L_T, \tag{5.31}$$

where the function $f=f(t,x)$ is introduced in Assumption 5.1, and the process $L=(L_t)_{0\le t\le T}$ is a local martingale strongly orthogonal to the martingale part of X, i.e. M.

Proof. Assume that F satisfies Assumption 5.1 and let f be the solution of the PIDE in (5.28).

Recall that Z is a semimartingale, since a $C^{1,2}$ function of a semimartingale is also a semimartingale, and the product of two semimartingales is still a semimartingale.

Hence, it has a canonical decomposition. Because both functions f and K satisfy the integrability conditions of Assumption 5.1, by Theorem 5.5, there are the following local $\left(\mathfrak{F}_t^X\right)_{0\leq t\leq T}$ - martingales $M^{(1)}$ and $M^{(2)}$ on $[0,T]$:

$$M_t^{(1)} = f(t,X_t)1_{\{\tau>t\}} - \int_0^t \mathfrak{A}f(s,X_s)1_{\{\tau>s\}}\,ds,$$

$$M_t^{(2)} = K(t,X_t)1_{\{\tau>t\}} - \int_0^t \mathfrak{A}K(s,X_s)1_{\{\tau>s\}}\,ds.$$

First we find the GKW decomposition of $M^{(1)}$ versus M. We show that

$$M_t^{(1)} = M_0^{(1)} + \int_0^t \theta_{s^-}\,dM_s + L_t,\ 0\leq t\leq T, \tag{5.32}$$

for a local martingale $L=(L_t)_{t\geq 0}$ that is strongly orthogonal to M. Note that because X is càdlàg, the process $(\theta_s)_{s\geq 0}$ is càglàd and hence locally bounded and predictable. Therefore the integral $\int_0^t \theta_s\,dM_s$ is well defined for all $0\leq t\leq T$ and it is an $\left(\mathfrak{F}_t^X\right)_{0\leq t\leq T}$ - local martingale.

On the other hand, by Proposition 5.2, $[Z,X]=[M^{(1)},M]$, so $[M^{(1)},M]$ belongs to $\mathscr{A}_{loc}$, and by Lemma 5.1 its compensator exists, which is $\langle M^{(1)},M\rangle$. By a similar reasoning or as we still see shortly, the process $\langle M\rangle$ also exists. Hence, the GKW decomposition exists and the formula (5.27) is applicable. So we need to obtain $\langle M^{(1)},M\rangle$ and $\langle M\rangle$.

Calculating $\langle M\rangle$ is simple. Since the compensator of X is a continuous finite variation process, we have that $[M]=[X]$. Therefore the conditional quadratic variation of M as the compensator of $[M]$ is equal to $\langle X\rangle$. The later is already obtained in

Example 5.3 and so

$$\langle M\rangle_t = \int_0^t \int_{-\infty}^{\infty} y^2 \, v(dy) \, ds. \tag{5.33}$$

Since $[M^{(1)}, M] = [Z, X]$, the compensator of the two processes are the same and to get $\langle M^{(1)}, M\rangle$, it is enough to obtain $\langle Z, X\rangle$. The integration by parts formula for semimartingales on $[0, T]$ gives

$$Z_t X_t = Z_0 X_0 + \int_0^t Z_{s^-} \, dX_s + \int_0^t X_{s^-} \, dZ_s + [Z, X]_t.$$

Let $F_t^{(1)} = \int_0^t \mathfrak{A} f(s, X_s) 1_{\{\tau > s\}} \, ds$ and $F_t^{(2)} = \int_0^t \mathfrak{A} K(s, X_s) 1_{\{\tau > s\}} \, ds$, then $Z = M^{(1)} + F^{(1)}$, $XZ = M^{(2)} + F^{(2)}$, and we also have that $X = M + \Lambda$. Therefore, the above integration by parts formula on $[0, T]$ becomes

$$\begin{aligned} &[Z, X]_t - (F_t^{(2)} - \int_0^t X_{s^-} \, dF_s^{(1)} - \int_0^t Z_{s^-} \, d\Lambda_s) \\ &= -Z_0 X_0 + M_t^{(2)} - \int_0^t X_{s^-} dM_s^{(1)} - \int_0^t Z_{s^-} dM_s. \end{aligned}$$

Since the processes X and Z are càdlàg, $(X_{s^-})_{s\geq 0}$ and $(Z_{s^-})_{s\geq 0}$ are predictable, and so the integrals on the right-hand side of the above equality are local martingales, see Protter (2004). Since the process

$$\left(F_t^{(2)} - \int_0^t X_{s^-} \, dF_s^{(1)} - \int_0^t Z_{s^-} \, d\Lambda_s \right)_{0\leq t\leq T}$$

is a finite variation predictable process and $[Z, X] = [M^{(1)}, M]$, the uniqueness of the modified version of conditional quadratic variation gives,

$$\langle M^{(1)}, M\rangle_t = F_t^{(2)} - \int_0^t X_{s^-} \, dF_s^{(1)} - \int_0^t Z_{s^-} \, d\Lambda_s, \; 0 \leq t \leq T,$$

see Section 5.2. Note that ds is the Lebesgue measure so for example

$$F_t^{(2)} = \int_0^t \mathfrak{A}K(s^-, X_{s^-})1_{\{\tau \geq s\}}\, ds.$$

Hence after some manipulations $\langle M^{(1)}, M\rangle_t$ is equal to

$$\int_0^t \left(\mathfrak{A}K(s^-, X_{s^-}) - X_{s^-}\mathfrak{A}f(s^-, X_{s^-}) - \beta f(s^-, X_{s^-})\right) 1_{\{\tau \geq s\}} ds. \tag{5.34}$$

Then the GKW decomposition (5.32) is a result of expressions (5.27), (5.33), and (5.34). Since $F = F(x)$ satisfies Assumption 5.1 and $M_t^{(1)} = f(t, X_t)1_{\{\tau > t\}} - \int_0^t \theta_s\, d\Lambda_s$, the GKW decomposition (5.32) becomes

$$f(t, X_t)1_{\{\tau > t\}} - \int_0^t \theta_s\, d\Lambda_s = f(0, X_0) + \int_0^t \theta_s\, dM_s + L_t.$$

Because of the integrability conditions of the Assumption 5.1, both of the integrals $\int_0^t |\mathfrak{A}f(s, X_s)|\, ds$ and $\int_0^t |\mathfrak{A}K(s, X_s)|\, ds$ are almost surely finite for all $0 \leq t \leq T$. Therefore, for all $0 \leq t \leq T$, θ_t and so the term $\int_0^t \theta_s\, d\Lambda_s$ are well defined and almost surely finite. Hence, one can move the integral on the left-hand side to the other side of the equality. This gives the decomposition (5.30). Finally the decomposition (5.31) is obtained by letting $t = T$ in Equation (5.30) and noticing that by Assumption 5.1, $f(T, X_T) = F(X_T)$. □

In the special case when the process X is a martingale, we have the following corollary.

Corollary 5.4. *Assume that X is the same process as Theorem 5.7. Let the function $F = F(x)$ satisfies Assumption 5.1 and the process $[Z, X]$ belongs to $\mathscr{A}_{loc}$. Now further*

suppose that X is a martingale under the natural completed filtration generated by X, i.e. $\mathfrak{F}^X$. Then we have

$$F(X_T)1_{\{\tau>T\}} = f(0,X_0) + \int_0^T \frac{\mathfrak{A}K(s^-,X_{s^-})}{\int_{\mathbb{R}} y^2\, v(dy)} 1_{\{\tau\geq s\}}\, dX_s + L_T, \tag{5.35}$$

where the operator $\mathfrak{A}$ is introduced in (5.19), the functions $f(t,x)$ and $K(t,x)$ are defined in Assumption 5.1, and the process $L = (L_t)_{0\leq t\leq T}$ is a local martingale strongly orthogonal to X.

Proof. Since X is a martingale, then $\beta = \mu + \int_{\mathbb{R}} y\, v(dy)$ is equal to zero, and therefore by Assumption 5.1, $\mathfrak{A}f(s,x)$ is also zero. Now the corollary easily follows from Theorem 5.7. □

Our goal is to find the decomposition (5.7), but we should not forget that by Lemma 5.6, finding this decomposition leads to just pseudo-locally risk-minimizing and not necessarily local risk-minimization. As we saw in Section 5.4, to bridge the two concepts, first we need to investigate SC condition on the underlying process and also the existence of the FS decomposition.

Remember that we can write X as $X = M + \Lambda$, where M is a martingale and Λ is a continuous finite variation process given by

$$\Lambda_t = \mu t + \int_0^t \int_{\mathbb{R}} y\, v(dy)\, ds, \quad t \geq 0.$$

The process M is square-integrable if and only if the process X is square-integrable. Therefore if X is square-integrable then X belong to $\mathcal{S}^2(\mathbb{P})$. From here on, it is assumed that X is square-integrable.

The conditional quadratic variation of M is calculated in the proof of Theorem 5.7, see Equation (5.33). It is easy to verify that for all $t \geq 0$, we have

$$\Lambda_t = \int_0^t \frac{\mu + \int_{\mathbb{R}} y \; v(dy)}{\int_{\mathbb{R}} y^2 \; v(dy)} \; d\langle M \rangle,$$

so Λ is absolutely continuous with respect to $\langle M \rangle$. Also the MVT process $\tilde{K}$ is absolutely continuous with respect to $\langle M \rangle$, i.e. for all $t \geq 0$,

$$\tilde{K}_t = \int_0^t \left(\frac{\mu + \int_{\mathbb{R}} y \; v(dy)}{\int_{\mathbb{R}} y^2 \; v(dy)} \right)^2 d\langle M \rangle.$$

The meaning of these is that the process X satisfies SC condition.

Therefore, by Theorem 3.3 of Schweizer (1999), locally risk-minimizing strategies are the same as pseudo-locally risk-minimizing strategies. On the other hand by Proposition 3.4 of Schweizer (1999), the existence of the later is equivalent to the existence of an FS decomposition of the payoff. Since the MVT process $\tilde{K}$ is uniformly bounded in both t and ω, the FS decomposition exists.

Therefore we conclude that in our framework the existence of the Föllmer-Schweizer decomposition and so locally risk-minimizing strategies are guaranteed.

Remark 5.16. *We shall point out that the boundness of $\tilde{K}$ can hold even in a more general setting, for example in case of a jump-diffusion process.*

From the above, we conclude that in our setup, to get the local risk-minimization strategies, all we need is to find the FS decomposition. Some integrability conditions

turns the decomposition (5.31) into FS decomposition. The next proposition clarifies this.

Proposition 5.6. *Assume that X is the above mentioned process X, i.e., it belongs to $\mathcal{S}^2(\mathbb{P})$ and satisfies SC condition. Let the function $F = F(x)$ satisfy Assumption 5.1. We further suppose that for all $0 \leq t \leq T$, $f(t, X_t)$ belongs to $L^2(\Omega, \mathfrak{F}_t, \mathbb{P})$ and the process θ is in Θ_S. Then there is a locally risk-minimizing L^2-strategy $\phi = (\theta, \eta)$, determined as follows. The number of shares invested in the risky asset to hold is given by θ. The hedging error L belongs to $\mathcal{M}_0^2$. It is strongly orthogonal to M and given by*

$$L_t = f(t, X_t)1_{\{\tau > t\}} - f(0, X_0) - \int_0^t \theta_s \, dX_s, \ 0 \leq t \leq T.$$

The value process of the portfolio is equal to

$$V_t(\theta) = f(0, X_0) + \int_0^t \theta_s \, dX_s + L_t, \ 0 \leq t \leq T,$$

the number of risk-free assets is

$$\eta_t = V_t(\theta) - \theta_t X_t, \ 0 \leq t \leq T,$$

and finally the cost process is provided by

$$C_t = f(0, X_0) + L_t, \ 0 \leq t \leq T.$$

Proof. It is verified above that the process X satisfies the SC condition. Therefore as explained before the existence of a locally risk-minimizing L^2-strategy is equivalent to the existence of FS decomposition. Notice that for all $0 \leq t \leq T$, $f(t, X_t)$ belongs

to $L^2(\Omega, \mathfrak{F}_t, \mathbb{P})$, and so by Proposition 4.50 of Jacod and Shiryaev (1987), the process $[Z, X]$ is in $\mathscr{A}_{loc}$. From Equation (5.30) of Theorem 5.7, we have

$$f(t, X_t)1_{\{\tau>t\}} - \int_0^t \theta_s \, dA_s = f(0, X_0) + \int_0^t \theta_s \, dM_s + L_t, \; 0 \leq t \leq T.$$

Because θ is in Θ_S and $f(t, X_t)$ is square-integrable, the left-hand side and so the right-hand side of the above equation is square-integrable. Since θ belongs to Θ_S, it is also in $L^2(X)$ and so by Lemma 2.1 of Schweizer (1999) the process $\int \theta \, dM$ is in $\mathcal{M}_0^2$. Hence the process L is square-integrable as well, then Lemma 5.3 shows that it belongs to $\mathcal{M}_0^2$. Now the proposition follows from Lemma 5.6. □

Remark 5.17. *A similar result like Proposition 5.6 can be obtained when X is a local martingale but with a simpler form for the strategy θ. Notice that although we did not use the MELMM method, we have paid the price by involving a PIDE. In MELMM method when the underlying process is martingale the problem of finding the hedging strategies are simpler. Here, the same thing happens too. If the underlying process is a martingale, the PIDE to find the hedging strategy has a simpler form.*

The next theorem investigates necessary and sufficient conditions under which the process L in Theorem 5.7 vanishes. For this theorem and the corollary following it, the process X does not need to be square-integrable.

Theorem 5.8. *Assume that X is the same process as Theorem 5.7 and the function $F = F(x)$ satisfies Assumption 5.1. Suppose that the three integrability conditions of Assumption 5.1 are met for the function f^2 defined by $f^2(t, x) = (f(t, x))^2$ where the*

function f is defined in Assumption 5.1. Now further suppose that the process $[Z, X]$ belongs to $\mathscr{A}_{loc}$ and the process L in the decompositions (5.30) *and* (5.31) *belongs to $\mathcal{M}^2_{0,loc}$. Let the operator $\mathfrak{L}$ be defined as below*

$$\mathfrak{L}f(t,x) = \mathfrak{A}f^2(t,x) - 2\beta f(t,x) - \frac{\left(\mathfrak{A}K(t,x) - x\mathfrak{A}f(t,x) - \beta f(t,x)\right)^2}{\int_{\mathbb{R}} y^2 \; v(dy)},$$

the operator $\mathfrak{A}$ is introduced in (5.19)*, the function $K(t,x)$ is the same as in Assumption 5.1. Then the martingale part of the decompositions* (5.30) *and* (5.31) *is null on $[0,T]$, if and only if $\mathfrak{L}f(t,x) = 0$ for all $0 \leq t \leq T$ and all x in $\mathbb{R}$. In this case, for all $0 \leq t \leq T$, we have the following*

$$f(t, X_t)1_{\{\tau > t\}} = f(0, X_0) + \int_0^t \theta_s \; dX_s, \tag{5.36}$$

and specifically for $t = T$, one obtains

$$F(X_T)1_{\{\tau > T\}} = f(0, X_0) + \int_0^T \theta_s \; dX_s. \tag{5.37}$$

Proof. We look for functions $F = F(x)$ that make vanish the martingale part of the FS decomposition (5.31) or $L = 0$. Since L is in $\mathcal{M}^2_0$, by Corollary 5.2, $L = 0$ is equivalent to $\langle L, L\rangle = 0$. On the other hand, by Theorem 5.7 the following holds

$$f(t, X_t)1_{\{\tau > t\}} - \int_0^t \theta_s \; d\Lambda_s = f(0, X_0) + \int_0^t \theta_s \; dM_s + L_t.$$

From this decomposition, we have

$$\langle L, L\rangle = \langle Z\rangle - 2\langle Z, \int \theta \; dX\rangle + \langle \int \theta \; dX\rangle, \tag{5.38}$$

which is obtained assuming that all the terms on the right hand side exist. Not only, we prove their existence but also we compute them explicitly. First, we obtain $\langle Z \rangle$.

We already know that $Z = M^{(1)} + F^{(1)}$ and observe that $Z_t^2 = f^2(t, X_t)1_{\{\tau > t\}}$. By Theorem 5.5, $Z^2 = M^{(3)} + F^{(3)}$, where $M^{(3)}$ is an $\mathfrak{F}^X$- local martingale and $F_t^{(3)} = \int_0^t \mathfrak{A}f^2(s, X_s)1_{\{\tau > s\}}ds$. Using the integration by parts formula, we get

$$Z^2 = Z_0^2 + 2\int Z_-\, dM + 2\int Z_-\, d\Lambda + [Z],$$

$$M^{(3)} + F^{(3)} = Z_0^2 + 2\int Z_-\, dM + 2\int Z_-\, d\Lambda + [Z],$$

or

$$[Z] - (F^{(3)} - 2\int Z_-\, d\Lambda) = M^{(3)} - 2\int Z_-\, dM - Z_0^2.$$

The right-hand side of the above equation is a local martingale. Now the predictability of $(F^{(3)} - 2\int Z_-\, dF)$ and uniqueness of Lemma 5.1 give

$$\langle Z \rangle_t = \int_0^t \mathfrak{A}f^2(s, X_s)1_{\{\tau > s\}}ds - 2\int_0^t f(s, X_s)1_{\{\tau > s\}}d\Lambda.$$

For the second term of (5.38), since $[Z, X] = [M^{(1)}, M]$, computing the second term follows from

$$\langle Z, \int \theta\, dX \rangle_t = \int_0^t \theta\, d\langle M^{(1)}, M \rangle =$$

$$\int_0^t \frac{(\mathfrak{A}K(s, X_s) - X_s\mathfrak{A}f(s, X_s) - \beta f(s, X_s))^2}{\int_{\mathbb{R}} y^2\, v(dy)} 1_{\{\tau > s\}}ds,$$

where $\langle M^{(1)}, M \rangle$ was already computed in the proof of Theorem 5.7.

The third term can be computed similarly

$$\langle \int \theta\, dX \rangle_t = \int_0^t \theta^2 d\langle M \rangle,$$

or

$$\begin{aligned}&\langle \int \theta \, dX \rangle_t \\ &= \int_0^t \frac{\left(\mathfrak{A}K(s, X_s) - X_s \mathfrak{A}f(s, X_s) - \beta f(s, X_s)\right)^2}{\int_{\mathbb{R}} y^2 \, v(dy)} 1_{\{\tau > s\}} ds.\end{aligned}$$

From Equation (5.38) and the previous calculation we get the following

$$\langle L, L \rangle_t = \int_0^t \mathfrak{L}f(s, X_s) 1_{\{\tau > s\}} ds,$$

where

$$\begin{aligned}&\mathfrak{L}f(t, x) \\ &= \mathfrak{A}f^2(t, x) - 2\beta f(t, x) - \frac{\left(\mathfrak{A}K(s, x) - x\mathfrak{A}f(s, x) - \beta f(s, x)\right)^2}{\int_{\mathbb{R}} y^2 \, v(dy)}.\end{aligned}$$

Since the function f is in $C^{1,2}$, $\langle L, L \rangle$ is zero on $[0, T]$ if and only if $\mathfrak{L}f(t, x) = 0$ on $[0, T] \times \mathbb{R}$. On the other hand, by Corollary 5.2, the former is equivalent to $L = 0$. Therefore in the decompositions (5.30) and (5.31), the orthogonal part is vanished if and only if $\mathfrak{L}f(t, x) = 0$ on $[0, T] \times \mathbb{R}$ and this gives Equations (5.36) and (5.37). $\square$

Corollary 5.5. *Assume that X is the same process as Theorem 5.7 and the function $F = F(x)$ satisfies Assumption 5.1. Suppose that the three integrability conditions of Assumption 5.1 are met for the function f^2 defined by $f^2(t, x) = (f(t, x))^2$ where the function f is given in Assumption 5.1. Now, further suppose that the process $[Z, X]$ belongs to $\mathscr{A}_{loc}$, the process L in the decompositions* (5.30) *and* (5.31) *belongs to $\mathcal{M}^2_{0,loc}$, and the process X is a martingale under the natural completed filtration generated by*

X, i.e. $\mathfrak{F}^X$. Let the operator $\mathfrak{L}$ be defined as below

$$\mathfrak{L}f(t,x) = \mathfrak{A}f^2(t,x) - \frac{(\mathfrak{A}K(t,x))^2}{\int_{\mathbb{R}} y^2\ v(dy)},$$

where the operator $\mathfrak{A}$ is introduced in (5.19)*, the function $K(t,x)$ is the same one as in Assumption 5.1. Then the martingale part of the decompositions* (5.30) *and* (5.31) *is zero on $[0,T]$, if and only if $\mathfrak{L}f(t,x) = 0$ for all $0 \leq t \leq T$ and all x in $\mathbb{R}$. In this case, for all $0 \leq t \leq T$, we have the following*

$$f(t,X_t)1_{\{\tau>t\}} = f(0,X_0) + \int_0^t \theta_{s^-}\ dX_s,$$

and specifically for $t = T$, one obtains

$$g(X_T)1_{\{\tau>T\}} = f(0,X_0) + \int_0^T \theta_{s^-}\ dX_s.$$

Proof. Since X is a martingale, then $\beta = \mu + \int_{\mathbb{R}} y\ v(dy)$ is equal to zero, and therefore by Assumption 5.1, $\mathfrak{A}f(s,x)$ is also zero. Then the corollary easily follows from Theorem 5.8. □

By combining Theorem 5.8 and Proposition 5.6, we will get the following result that provides a necessary and sufficient condition for the existence of a risk-free defaultable claim. In the context of jump-diffusion processes, Kunita (2010) answers a similar question for path independent payoffs.

Proposition 5.7. *Assume that X is the same process as in Proposition 5.6 and the function $F = F(x)$ satisfies Assumption 5.1. Suppose that the three integrability conditions of Assumption 5.1 are met by function f^2 defined as $f^2(t,x) = (f(t,x))^2$, where*

the function f is defined in the assumption. Now further suppose that for all $0 \leq t \leq T$, $f(t, X_t)$ belongs to $L^2(\Omega, \mathfrak{F}_t, \mathbb{P})$ and the process θ is in Θ_S. Let the operator $\mathfrak{L}$ be as in Theorem 5.8. Then the process $\phi = (\theta, \eta)$, defined in Proposition 5.6, is a locally risk-minimizing L^2- strategy that makes the derivative $F(X_T)1_{\{\tau>t\}}$ risk free if and only if $\mathfrak{L}f(t,x) = 0$ for all $0 \leq t \leq T$ and all x in $\mathbb{R}$. It means that we have the following decomposition

$$F(X_T)1_{\{\tau>T\}} = f(0, X_0) + \int_0^T \theta_s \, dX_s.$$

Remark 5.18. *A similar result can be obtained for the case when X is a martingale, but with a simpler form for the strategy θ.*

Example 5.5. *Assume that $X_t = u + \mu t + \sum_{j=1}^{N_t} Y_i$, where $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity λ and the Y_i's are i.i.d. random variables with jump distribution F_Y. Let $\mu > 0$, $-Y_1 \sim$ exponential(δ), and suppose that the process X is a martingale under the natural filtration generated by X which means that $\lambda = \mu\delta$. We remind the reader that in this chapter all asset values are discounted. Consider a defaultable zero-coupon bond that pays one unit of currency if there is no default, i.e. $F(x) = 1$ for all x. By Proposition 5.6 the number of the risky assets of the hedging strategy is given by*

$$\theta_s = \left(\delta^2 \int_{-X_{s^-}}^{0} y f(s^-, X_{s^-} + y) F_Y(dy) + \delta f(s^-, X_{s^-}) \right) 1_{\{\tau \geq s\}}, \tag{5.39}$$

where $f = f(t,x)$ satisfies the following PIDE

$$\mathfrak{A}f(t,x) = 0, \text{ for all } 0 \leq t \leq T \text{ and all } x \in \mathbb{R},$$

$$f(T,x) = 1, \text{ for all } x \in \mathbb{R}.$$

The Feynman-Kac formula or a renewal argument can be applied to prove that the solution has the following representation

$$f(t,x) = 1 - \mathbb{P}(\tau \leq T - t | X_0 = x).$$

This representation holds regardless of the type of the distribution of the jumps. In the case of exponential jump size distribution like in this example, a closed form solution is available. This solution is provided in Rolski et. al. (1999). It is a complicated function defined based on the Bessel 1 function. The graph of this function on $[0,2] \times [0,0.4]$ *is given by Figure 5.1 for* $\mu = 0.1$, $\delta = 100$, $\lambda = 10$, *and* $T = 2$,

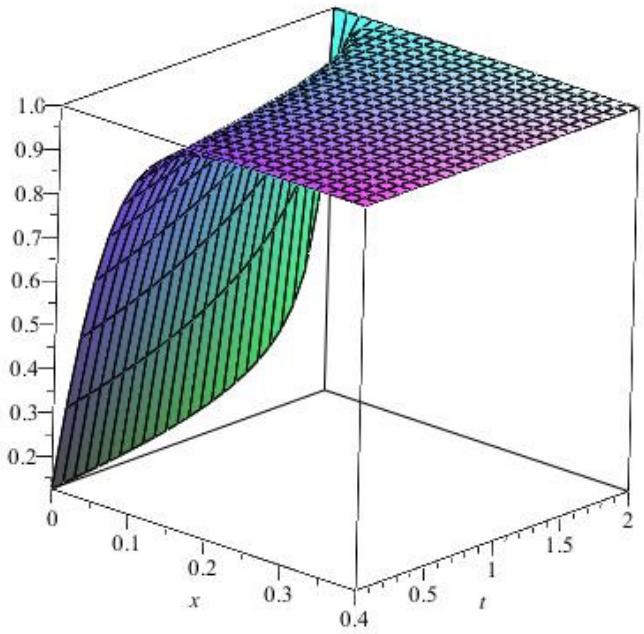


Figure 5.1: The exact function f with exponential jumps.

The function $f = f(t,x)$ can also be estimated numerically by simulation. With the same parameters as above, Figure 5.2 is the graph of the estimation of $f = f(t,x)$ on $[0,2] \times [0,0.4]$. The number of the risky asset θ is a closed form of this function given by Equation (5.39)*. Therefore the function $f = f(t,x)$ acts as an interface to solve the problem. However this function has also a nice interpretation. From Proposition 5.6, one can easily verify that the value of the portfolio is provided through the function $f = f(t,x)$. More precisely we have that $V_t(\theta) = f(t, X_t)1_{\{\tau > t\}}$.*

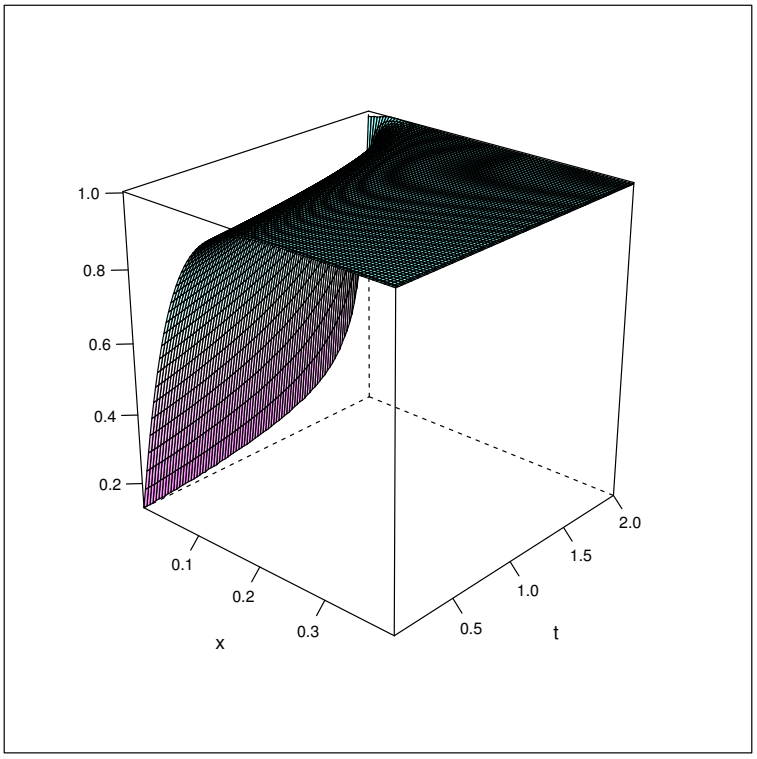


Figure 5.2: The estimated function f with exponential jumps.

Next we obtain the locally risk minimization strategies and other related quantities for a simulated sample path of the process X. In practice, a dynamic portfolio is updated

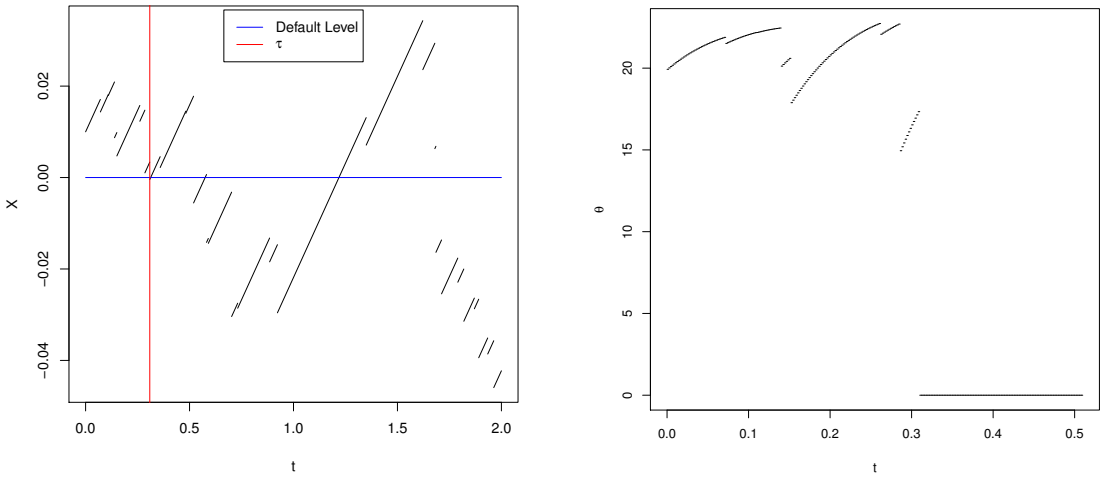
in some specific trading dates. In fact Proposition 5.6 and formula (5.39) *cannot be applied directly. A discretization procedure is required to implement the theory.*

Here we use a simple procedure. We divide the interval $[0,T] = [0,2]$ *into 1000 equal subintervals. It is assumed that the trading dates are given by* $\{t_0, t_1, ..., t_{1000}\}$*, for* $t_j = \frac{jT}{1000}$*, where* $j = 0, 1, ..., 1000$*. Then the number of the risky assets is given by*

$$\theta_t = \theta_0 1_{t=0} + \sum_{k=0}^{n} \theta_k 1_{(t_k, t_{k+1}]}(t),$$

where each θ_k *is a bounded* $\mathfrak{F}_{t_i}^X$*-measurable random variable that is determined right after the transaction* t_k*. This is due to the fact that a realistic strategy must be left continuous or predictable. The integral* $\int \theta \, dX$ *must be also discretized using Proposition 5.1. This is essential to obtain the observed values of the process* L*.*

In Figure 5.3, the simulated sample path of the process X *together with the number* θ *of shares invested in the risky asset to be held in each trading period, are shown. In order to have a better vision, we have changed the time scale of the two graphs. As the Figures 5.1 or 5.2 confirms, the probability of default is relatively high for this process. Indeed we have that* $\mathbb{P}(\tau \leq 2) \approx 0.754995$*. For the sample path of the process* X *shown in Figure 5.3, the default happens at* $\tau \approx 0.30869$*. The number of the risk-free assets* η*, the value of the portfolio* $V(\theta)$*, the error term* L*, and the cost process* C *are shown in Figure 5.4. All the processes remain constant between the trading dates and they are sensitive with respect to the jumps of the underlying process. When the process* X *performs well, the value of the portfolio increases and the cost decreases. At the default*



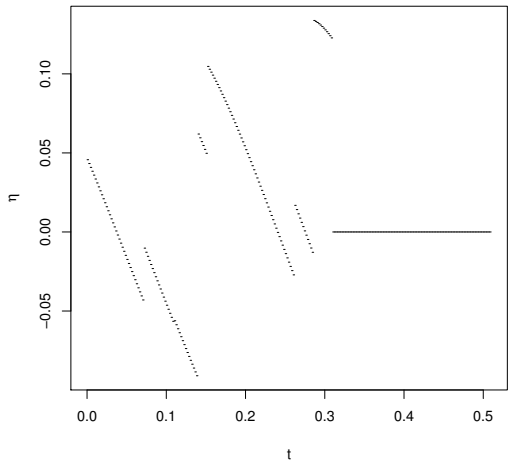
A sample path of X The number of risky assets θ

Figure 5.3: Sample paths of the processes θ and X.

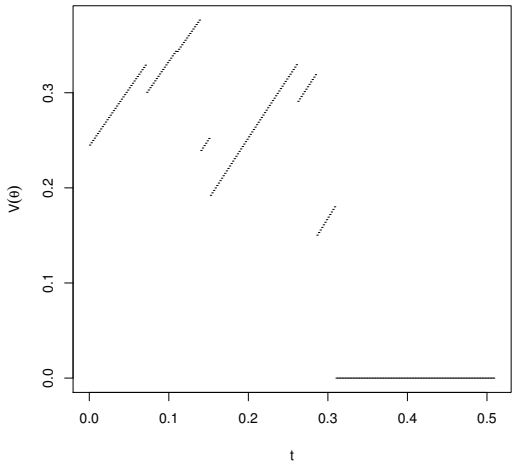
time, the number of risk-free and risky assets and so the value of the portfolio drop to zero. However after the default time τ the error term L and the cost process C remain fixed respectively equal to L_τ and C_τ.

Example 5.6. *Assume that $X_t = u + \mu t + \sum_{j=1}^{N_t} Y_i$, where $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity λ and the Y_i's are i.i.d. random variables with gamma distribution having parameters α and β. Let $\mu > 0$, $-Y_1 \sim gamma(\alpha, \beta)$, and suppose that the process X is a martingale under the natural filtration generated by X which means that $\mu = \lambda\alpha\beta$. Consider a defaultable zero-coupon bond that pays one unit of currency if there is no default, i.e. $F(x) = 1$ for all x. By Proposition 5.6 the number of the risky assets of the hedging strategy is given by*

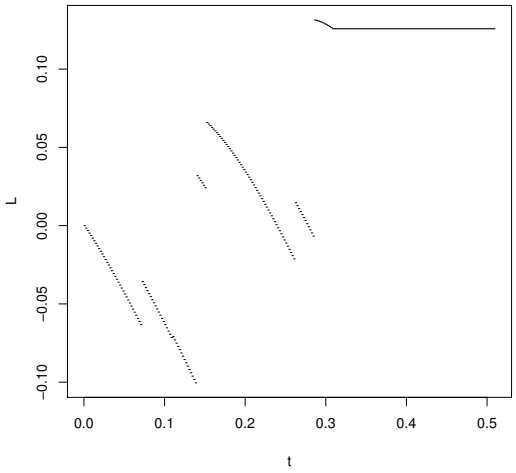
$$\theta_s = \frac{\left(\int_{-X_{s^-}}^{\infty} yf(s^-, X_{s^-} + y)F_Y(dy) - \mathbb{E}[Y_1]f(s^-, X_{s^-})\right) 1_{\{\tau \geq s\}}}{\mathbb{E}[Y_1^2]},$$



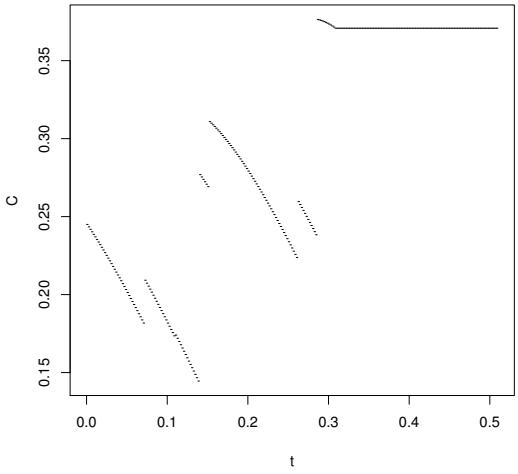
The number of risk-free assets



The value of the portfolio



The unhedgeable part of the claim



The cost of the portfolio

Figure 5.4: Sample paths of the processes η, $V(\theta)$, L, and C.

where $f = f(t,x)$ satisfies the following PIDE

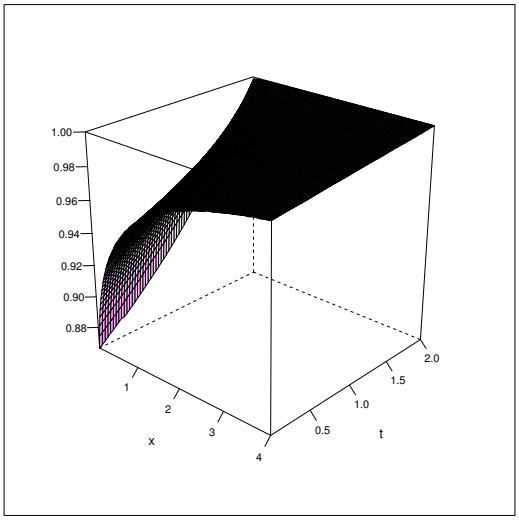
$$\mathfrak{A}f(t,x) = 0, \text{ for all } 0 \leq t \leq T \text{ and all } x \in \mathbb{R},$$

$$f(T,x) = 1, \text{ for all } x \in \mathbb{R}.$$

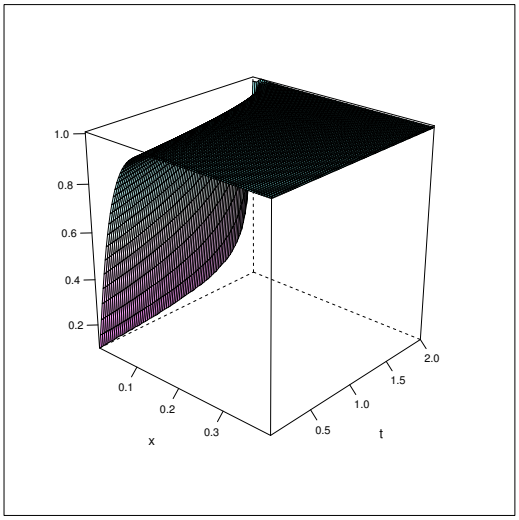
In this example we only estimate the function $f = f(t,x)$. The estimated graph of the function $f = f(t,x)$ for different values of α and β is given by Figures 5.5 and 5.6. In order to make better comparisons between these figures, all the expected values of jump sizes are equal to 0.01. Remember that the function $f = f(t,x)$ determines the value of the portfolio before default. Notice that how different values of the parameters can change the function f and hence the value of the portfolio.

In the next section, we will see that $f(0,X_0)$ can be interpreted as the value of this defaultable bond. The most expensive bond corresponds to $\beta = 5$. As β decreases the value of the bonds decrease as well. A possible explanation for this behavior can be the tail property of the distribution of jumps. Although all these distributions have the same expected value, they have slightly different tails. Within these distributions, the one with $\beta = 5$ has the heaviest tail. Hence the contract corresponding to this distribution is more risky.

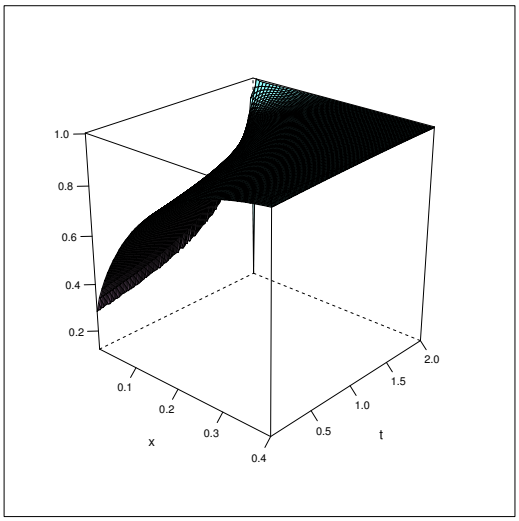
In the next section, other aspects of the function $f = f(t,x)$ are studied.



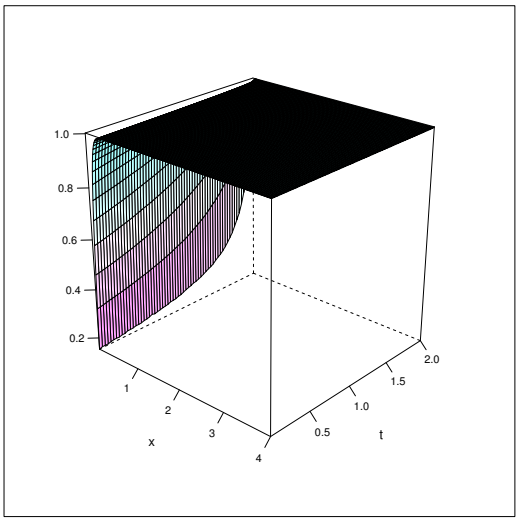
$\alpha = 0.002$ and $\beta = 5$



$\alpha = 5$ and $\beta = 0.002$



$\alpha = 0.1$ and $\beta = 0.1$



$\alpha = 0.5$ and $\beta = 0.02$

Figure 5.5: The estimated function f for gamma jumps

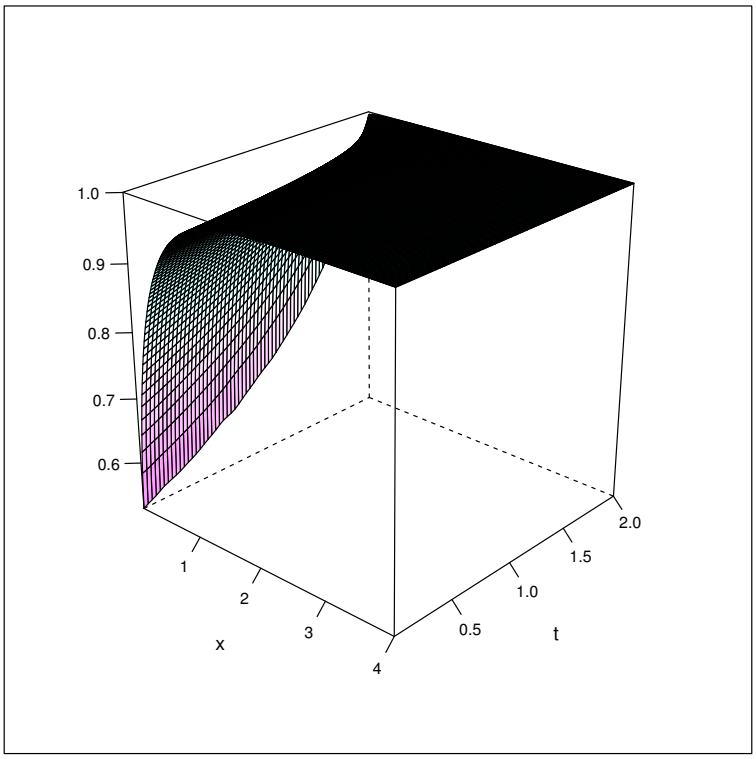


Figure 5.6: The estimated function f for gamma jumps with $\alpha = 0.02$ and $\beta = 0.5$.

5.8 Estimation of the Default Time and Pricing Tools

Knowing the distribution of the default time can play an important role in managing and pricing credit derivatives. In this section, we discuss the distribution of the default time and also the pricing rules related to the local risk-minimization management. In this section, it is assumed that X is the same process as in Theorem 5.7 that models the evolution of the firm asset. In examples, we use the following process

$$X_t = u + \mu t + \sum_{j=1}^{N_t} Y_j, \quad t \geq 0, \tag{5.40}$$

where $(N_t)_{t\geq 0}$ is a homogeneous Poisson process and the Y_i's are i.i.d. random variables with jump distribution F_Y.

5.8.1 Distribution of the Default Time

In the last section, we studied the hedging problem. However, some of the results of Section 5.7 can be helpful to understand the structure of the default time. Regarding Theorem 5.7, one can let $F = F(x)$ be the constant function $F = 1$. So without almost any effort, we have the following decomposition.

Proposition 5.8. *Assume that X is the same process as in Theorem 5.7 and the constant function $F = 1$ satisfies Assumption 5.1. We further suppose that the process $\left([1_{\{\tau>t\}}, X_t]\right)_{0\leq t\leq T}$ belongs to $\mathscr{A}_{loc}$. Then, for all $0 \leq t \leq T$, we have the following decomposition*

$$f(t, X_t)1_{\{\tau>t\}} = f(0, X_0) + \int_0^t \theta_s \, dX_s + L_t, \tag{5.41}$$

and specifically for $t = T$, one obtains

$$1_{\{\tau>T\}} = f(0, X_0) + \int_0^T \theta_s \, dX_s + L_T, \tag{5.42}$$

where the function $f = f(t,x)$ is introduced in Assumption 5.1, the process $\theta = (\theta_t)_{0\le t\le T}$ is given by (5.29) *and the process $L = (L_t)_{0\le t\le T}$ is a local martingale, strongly orthogonal to the martingale part of X, i.e. M.*

Notice that if the process X is square-integrable, then the process $\big([1_{\{\tau>t\}}, X_t]\big)_{0\le t\le T}$ belongs to $\mathscr{A}_{loc}$. Although this decomposition reveals the structure of the default time, it does not tell us much about the distribution of the default time. This is the decomposition of the indicator process versus the process X. Regarding the distribution of the default time, a more useful decomposition is stated in the following proposition.

Remark 5.19. *Although the next proposition looks similar to Theorem 5.7, they re quite different. Theorem 5.7 explains the decomposition of the process Z versus the underlying process X, while the following proposition finds the decomposition of the process Z versus M, the martingale part of X. Also note that Assumption 5.1 is totally simplified here.*

Proposition 5.9. *Assume that X is the same process as in Theorem 5.7. Let the function $f = f(t,x)$ be the solution of the following PIDE,*

$$\mathfrak{A} f(t,x) = 0, \text{ for all } 0 \le t \le T \text{ and all } x \in \mathbb{R},$$

$$f(T,x) = F(x), \text{ for all } x \in \mathbb{R},$$

where the function $F = F(x)$ is a real valued function and the function f satisfies the integrability conditions of Assumption 5.1. Let the process M be the martingale part of the canonical decomposition of X, i.e. $X = M + A$. Then for all $0 \leq t \leq T$, the following decomposition holds

$$f(t, X_t)1_{\{\tau > t\}} = f(0, X_0) + \int_0^t \theta_s \, dM_s + L_t,$$

and especially for $t = T$, one obtains

$$F(X_T)1_{\{\tau > T\}} = f(0, X_0) + \int_0^T \theta_s \, dM_s + L_T, \tag{5.43}$$

where the process θ is given by

$$\theta_t = \frac{(\mathfrak{A}K(t^-, X_{t^-}) - \beta f(t^-, X_{t^-}))}{\int_{-\infty}^{\infty} y^2 \, v(dy)} 1_{\{\tau \geq t\}},$$

and the process $L = (L_t)_{0 \leq t \leq T}$ is a local martingale strongly orthogonal to the process M.

Proof. For the proof of this proposition, one can follow the same lines of reasoning as for the Theorem 5.7, though it is tedious. The result basically follows from Equations (5.32), (5.33), and (5.34) of Theorem 5.7. □

As a special case let $F = 1$, then by taking the expectation of both sides of (5.43), we obtain $\mathbb{P}(\tau > T) = f(0, X_0)$, and that $f = f(t, x)$ is the solution of the PIDE in Proposition 5.9. Finding the distribution of the default time using a PIDE is already known. For example in Rolski et al. (1999), this PIDE is obtained for a compound Poisson process plus drift, that is (5.40).

Example 5.7. *Assume that* $X_t = u + \mu t + \sum_{j=1}^{N_t} Y_i$, *where* $(N_t)_{t\geq 0}$ *is a homogeneous Poisson process with intensity* λ *and the* Y_i*'s are i.i.d. random variables with jump distribution* F_Y. *Let* $\mu > 0$, $-Y_1 \sim$ *exponential*(δ), *and define the function* $F = F(x)$ *by* $F(x) = 1 - \frac{\lambda}{\mu\delta} e^{(\frac{\lambda}{\mu} - \delta)x}$. *Apply the above proposition for* $f(t,x) = F(x)$, *then* $\mathfrak{A}f(t,x) = 0$, *and*

$$F(X_t)1_{\{\tau > t\}} = F(u) + \int_0^t \theta_s dM_s + L_t,$$

$$\theta_s = \left(\delta^2 \int_{-X_{s^-}}^{\infty} yF(X_{s^-} + y)F_Y(dy) + \delta F(X_{s^-}) \right) 1_{\{\tau \geq s\}}.$$

Note that the above function $F = F(x)$ *is a special choice that makes the operator* $\mathfrak{A}$ *zero and hence the process* $(F(X_t)1_{\{\tau>t\}})_{t\geq 0}$ *is a martingale. This martingale can also be obtained from Theorem 5.5. Therefore we have the following identity*

$$\mathbb{P}(\tau > t) - \frac{\lambda}{\mu\delta} \mathbb{E}[e^{(\frac{\lambda}{\mu} - \delta)X_t} 1_{\{\tau > t\}}] = F(u).$$

One can use the intensity to estimate the distribution of the default time. The intensity λ^i in Chapter 4 can be viewed as the first order approximation of the default time. Hence, it is not surprising if this can be useful in estimating the distribution of the default time. Before describing the procedure, we give some motivations.

Empirically, many jumps are observed in the evolution process of a firm's asset values, but these can be classified in two categories. One, is in some sense, the class of small jumps, which for instance can be modeled by a Brownian motion. The other category is made of large jumps, that possibly can be modeled by the pure jump compound Poisson process. In fact, it is reasonable to assume that larger jumps in

the evolution process of the firm's value are rare events, relative to smaller jumps, see Tauchen and Zhou (2006). We see how these empirical observations lead to a good estimation of the default time.

For the process in (5.40) the number of jumps is provided by the homogeneous Poisson process N. This means that the intensity of this process, $\lambda_t = \lambda t$ plays an important role in the estimation. Tauchen and Zhou (2006) show that for one period of time, λ is a small number less than one. However, this is a delicate assumption and it totally depends on the type of the firm. Another important factor in our estimation is the maturity time T. In Credit Default Swap (CDS) contracts and in other important financial derivatives, the maturity date is as long as 10 years. Both of these observations make the estimation of the distribution of the default time tractable. In an example, we see how this works.

In Chapter 4 we saw that when the underlying process X is (5.40), the process $\left(1_{\{\tau>t\}} - \int_0^t \lambda_s^i \, ds\right)_{t\geq 0}$ is a martingale for λ^i given by Corollary 4.1, under the natural filtration generated by X. Considering the form of the process λ^i, this instantly leads to a series representation of $\mathbb{P}\{\tau \leq t\}$. It turns out that with appropriately chosen values of λ and t, as discussed above $0 < \lambda < 1$ and $t < 10$, the convergence of this series is fast enough to get good approximations just by calculating the first few terms of the series. We now investigate this.

In what follows, we give some results in the absence of a Brownian motion part. The generalization of the approach, may be possible but will depend on whether the

limit calculated in Proposition 4.1 is the real intensity.

For the process in (5.40) as shown in Chapter 4, λ_s^i is given by (4.21). Martingale arguments and some simple manipulations show that $\mathbb{P}(\tau \leq t)$ satisfies the following equation

$$\mathbb{P}(\tau \leq t) = 1 - e^{-\lambda t} - \lambda e^{-\lambda t} \int_0^t F_Y(u + \mu s)\, ds$$
$$- \lambda e^{-\lambda t} \int_0^t \sum_{j=0}^{m-1} A_j\, ds + \varepsilon,$$

where $A_j = \mathbb{E}[F_Y(u + \beta s + \sum_{n=0}^{j} Y_n) 1_{\{\tau > s, T_j \leq s < T_{j+1}\}}]$ and ε is the error term.

In addition, a bound for the error term can simply be obtained by,

$$|\varepsilon| \leq \lambda e^{-\lambda t} \int_0^t F_Y(u + \mu s) e^{-\lambda s} (e^{\lambda \int_0^s F_Y(u+\mu v)\, dv} - \sum_{n=0}^{m-1} \frac{(\lambda \int_0^s F_Y(u + \mu v)\, dv)^n}{n!})\, ds,$$

since $F_Y(u + \mu s) \leq 1$, one special bound is the following function

$$\Delta(m, t, \lambda) = \lambda e^{-\lambda t} \Big(\frac{e^{\lambda t}}{\lambda} - \frac{1}{\lambda} - t - \big[\sum_{i=1}^{m-1} \frac{\lambda^i}{(i+1)!} t^{i+1} \big] \Big).$$

The Δ function is governed by only three parameters, t, m, and λ, simplifying of the error control. It should be noted that m is the number of terms needed to get the desired precision.

Example 5.8. *Let $\lambda = 0.2$ and $t = 1$, then $\Delta(1, 1, 0.2) = 0.01752309617$; this means that for a maturity of $T = 1$ and a series with just one-term, the error is around two percent.*

Other observations about the Δ function can be made. First, this error bound is free of μ, u, and the choice of distribution for Y_i. Second this is the extreme case error.

In most situations, depending on the distribution of jumps and contribution of the other parameters, the real error will be much smaller and convergence should be fast.

Another fact is the sensitivity of the Δ function to the maturity T. As we increase T we need more terms to get the desired precision. For example for a maturity $T = 10$ (which is the upper bound for the maturities of CDS contracts) we have $\Delta(6, 10, 0.2) = 0.004533805610$ that means 6 terms are needed to have at most a 0.4% error.

Example 5.9. *Let $\lambda = 0.2$ as before, $F_Y(x) = 1 - e^{-\delta x}$ (exponential jump distribution), $\delta = 0.02$, $u = 10$ and $\mu = 1$ then the method (with one term) gives $\mathbb{P}(\tau \leq 3) \approx 0.3811593586$, compared to the more accurate value 0.3811191123. The exact formula for exponential jumps can be found in Chapter 5 of Rolski et al. (1999). It is obvious that the actual error of 0.01% is much less than two percent. This is mostly because the exponential distribution gives a faster convergence.*

The above mentioned approach gives a good approximation for the range of parameters we are interested in, for example when λ is small enough. From a theoretical point of view or even for some rare or extreme cases in practice, finding the distribution of the default for a wide range of parameters, like for large λ, could be an interesting question. We propose the following approach for more general cases. This method is considered for the process in (5.40). The main idea is to fix a finite horizon t. Then for a given λ (the Poisson process intensity) bigger than 1, by changing the probability space we try to obtain a new process with a smaller Poisson intensity. We already

observed that in the presence of a small Poisson intensity, the series that represents the distribution of the default time converges fast. We only state the procedure here, but omit the details.

Let $X_t = u + \mu t + \sum_{i=1}^{N_t} Y_i$, where $Y_i's$ are i.i.d. with distribution F_Y, density of f_Y, and $(N_t)_{t\geq 0}$ is a Poisson process with intensity λ. Let the probability associated with this process be $\mathbb{P}$. The goal is to estimate $\mathbb{P}(\tau \leq t)$. This is based on the following expression

$$\mathbb{P}(\tau \leq t) = e^{t(\lambda_{\mathbb{P}_m} - \lambda_{\mathbb{P}})}\, \mathbb{E}^{\mathbb{P}_m}[e^{m\sum_{i=1}^{N_t^m} Y_i^m}] - e^{mu + t(m\mu + \lambda_{\mathbb{P}_m} - \lambda_{\mathbb{P}})} \sum_{j=0}^{r} A_j + \epsilon,$$

where $\mathbb{P}_m$'s can be considered as new probability measures for the sequence of processes $X_t^{(m)} = u + \mu t + \sum_{i=1}^{N_t^{(m)}} Y_i^{(m)}$, where $Y_i^{(m)}$ has distribution F_m given by

$$F_m(x) = \frac{\int_{-\infty}^{x} e^{mv} f(-v) dv}{\int_{-\infty}^{0} e^{mv} f(-v) dv}, \text{ for } x \leq 0, \text{ and 1 otherwise,}$$

N_t^m is a Poisson process with intensity $\lambda_{\mathbb{P}_m} = \lambda \int_{-\infty}^{0} e^{mv} f(-v) dv$,

$$A_j = \mathbb{E}^{\mathbb{P}_m}[e^{-mV_t} 1_{\{\tau > t\}} 1_{\{T_j \leq t < T_{j+1}\}}],$$

and ϵ is the error term. This is essentially the Esscher's transform. For instance the following bound for the error term can be found easily,

$$|\epsilon| \leq e^{m(u+\mu t)} \Big(e^{\lambda_{\mathbb{P}_m} \int_0^t F_Y(u+\mu v)\, dv} - \sum_{j=0}^{r} \frac{(\lambda_{\mathbb{P}_m} \int_0^t F_Y(u + \mu v)\, dv)^j}{j!} \Big).$$

Note that:

1. Under the condition that $\int_{-\infty}^{0} e^{mv} f(-v) dv \neq \frac{\lambda_{\mathbb{P}_{m-1}}}{\lambda}$, the $\lambda_{\mathbb{P}_m}$ are strictly decreasing.

2. The number of terms needed to get the desired precession is given by r not m. For a fixed m, by the bound of the error term, the maximum number of the terms needed to get a desired precession can be determined.

Example 5.10. *Let* $g(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$*, that is a gamma distribution with parameters* α *and* β*. The above conditions are satisfied. Then it turns out that* $\lambda_{\mathbb{P}_m} = \frac{\lambda \beta^\alpha}{(m+\beta)^\alpha}$*. We see that* $\lambda_{\mathbb{P}_m} \downarrow 0$ *as* $m \to \infty$*.*

5.8.2 Pricing Under Locally Risk-Minimizing Hedging

In Section 5.7, we focused on hedging strategies in the context of local risk-minimization. Under some assumptions, a locally risk-minimizing hedging strategy was found for the defaultable claim $F(X_T)1_{\{\tau>T\}}$. Now it is time to ask for the price of this product. Notice that in all the FS decompositions, there is always a constant, and from Proposition 5.6 this constant is equal to the initial value of the portfolio, i.e. $V_0(\theta) = f(0, X_0)$. For this reason the constant $f(0, X_0)$ in Proposition 5.6 is actually the price of the product. This constant can also be interpreted as the capital requirement to start the hedging process.

Normally, pricing of financial derivatives is done under a risk-neutral probability measure. In an incomplete market, there can be many of such probability measures. In the context of local risk-minimization approach, the minimal martingale measure is the appropriate one to use, see Föllmer and Schweizer (1991). Assume that there exists such a probability measure $\tilde{\mathbb{P}}$ in our set up. Then by the uniqueness of FS decomposition,

we must have that $\tilde{\mathbb{E}}[F(X_T)1_{\{\tau>T\}}] = f(0, X_0)$, where $\tilde{\mathbb{E}}$ is the expected value under the probability measure $\tilde{\mathbb{P}}$. Therefore we have the following proposition.

Proposition 5.10. *Suppose that there exists the minimal martingale measure $\tilde{\mathbb{P}}$. With the same assumptions and notation of Proposition 5.6, the price of the defaultable claim $F(X_T)1_{\{\tau>T\}}$ under the risk-neutral measure $\tilde{\mathbb{P}}$ is equal to $f(0, X_0)$.*

Notice that in the above proposition, only the existence of the minimal martingale measure is assumed and no explicit form of this measure is used.

Chapter 6

Credit Risk and Risk Measures

6.1 Introduction

Measuring the risk of financial derivatives is a major growing concern in finance. Arbitrage opportunities that can arise due to over or under estimating the underlying risk, are important, closely related issues. As an alternative to the classical methods available to deal with this problem, we introduce a new approach by using risk measures.

The type of the arbitrage that we investigate defines itself very intuitively from the properties of risk measures. Simply saying, if under a specific risk measure, the risk of a portfolio is less than or equal to zero then the possible positive income of the portfolio will be considered as an arbitrage income. Balbás and López (2008) consider a similar problem by defining sequential arbitrage measures. The present work parallels their article. However, our goal is not to build arbitrage portfolios in bond markets. There are two main purposes in this chapter. First we want to create an indicator to detect and measure such arbitrage opportunities that we refer to as inconsistencies. Second

and more importantly, we want to apply this theory to study the credit quality of bonds issued in bond markets.

A good market must be consistent in all aspect. In other words, an integrated market should not let agents take advantage of price differences to make a risk-free profit at zero cost. The existence of classical arbitrage opportunities that can arise due to over- or under-estimation of the underlying risk, are one of the indicators of inefficiencies in the market. This problem has been investigated using classical arbitrage methods, see Chen and Knez (1995), and also Kempf and Korn (1998). As an alternative to classical arbitrage methods, we introduce a new indicator by using risk measures. We believe that risk measures are more powerful and efficient tools to this effect than a classical arbitrage approach, and one of our goals in this chapter is to justify why.

A typical bond price includes two parts. One is the real bond price assuming that there is no risk of default, and the other one is the credit spread to compensate for the risk of default. To reflect this risk of default, all the bonds in the market are rated by rating agencies. Normally, the most credible bonds are the ones issued by governments. Our second goal is to measure this credit spread and basically study the credit quality of the bonds. Although there might be other types of risks, here we assume that these are restricted just to the risk of default.

We start with fixed income markets like bond markets, in the hope to later develop the method for more complicated financial portfolios. Finally, we mention that this work uses a combination of the approach and methods of Balbás, Balbás and Garrido

(2010), and Balbás and López (2008).

6.2 Preliminaries and Notation

Assume that uncertainty is modeled by $(\Omega, \mathcal{F}, \mathbb{P})$, and that $A = [a_{ij}]$ is an $m \times n$ matrix representing a portfolio of n bonds with possible future cash flows at times $i = t_1, t_2, ..., t_m$. The column j of the matrix A is the future cash flow of the bond j of the portfolio, $\overline{b}_j = (a_{1j}, a_{2j}, ..., a_{mj})$ at future dates $\mathfrak{T} = \{t_1, t_2, ..., t_m\}$. The row i of the matrix A is the total cash flow of the portfolio at time t_i. From now on, to simplify notation, a vector $(a_1, ..., a_m)$ in $\mathbb{R}^m$ is denoted by $\overline{a}$, and $T = t_m$ represents the final date of the cash flow, the last time when a payment is made. Also assume that $\overline{p} = (p_1, p_2, ..., p_n)$ with $p_j > 0$ is the current price of the j-th bond.

Assume that the typical future cash flow $\overline{c} = (c_{1j}, c_{2j}, ..., c_{mj})$ of a bond or in general of a portfolio will be reinvested and the accumulated wealth generated by this cash flow is denoted by $\Pi_T(\overline{c})$. It is worth mentioning this reinvestment is done in fixed income markets. Although the cash flows are predetermined, because of the fluctuations in interest rates, the accumulated wealth is a random variable at the maturity time T. It is assumed that this is the only source of randomness that makes $\Pi_T(\overline{c})$ uncertain. This provides the motivation to define the accumulated wealth function $\Pi_T : \mathbb{R}^m \to \Pi$, where

$$\Pi = \{\Pi_T(\overline{x}); \overline{x} \in \mathbb{R}^m\},$$

and $\overline{x} = (x_1, x_2, ..., x_m)$ is a vector in $\mathbb{R}^m$.

To avoid technical difficulties, we assume that $\Pi \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, or in other words $\mathbb{E}[(\Pi_T(\overline{x}))^2] < \infty$. By ignoring transaction costs, we can suppose that Π_T is a linear function.

To control the risk over the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, we use risk measures. For us this space is interpreted as the space of all future gains. In general, the risk measure ρ can be defined over the space $\mathfrak{R}$, the set of all real valued random variables, as the function

$$\rho : \mathfrak{R} \to \mathbb{R}.$$

As mentioned above, in this thesis, the space $\mathfrak{R}$ is equal to Π which is a subset of $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

For any random variable X belongs to $\mathfrak{R}$, the quantity $\rho(X)$ can be interpreted as the risk associated with the future wealth or gain X in a period of time. Artzner et al. (1997) define coherent risk measures through axioms. They also find a representation theorem on a finite probability space. Their result was later extended to general probability spaces. They call a risk measure ρ coherent if it satisfies the following properties:

- For all $X, Y \in \mathfrak{R}$, $\rho(X+Y) \leq \rho(X)+\rho(Y)$. This property is called sub-additivity.
- For all $t \geq 0$ and $X \in \mathfrak{R}$, $\rho(tX) = t\rho(X)$. This property is called positive homogeneity.
- For all $X \in \mathfrak{R}$ and all $a \in \mathbb{R}$, $\rho(X + a) = \rho(X) - a$. This property is called translation invariance.

- For all $X, Y \in \mathfrak{R}$, if $X \leq Y$ then $\rho(Y) \leq \rho(X)$. This property is called monotonicity.

Since then, their paper and the results have been extended in a variety of ways and different types of risk measures have been introduced. For instance deviations and expectation bounded risk measures are by Rockafellar, Uryasev and Zabarankin (2006). This paper provides some insights towards the structure of the subgradient sets associated with risk measures. Distortion risk measures are introduced by Wang (2000). Balbás, Garrido and Mayoral (2008) discuss the properties of distortion risk measures. All these risk measures are defined on a probability space. Recently, new types of risk measures have been introduced that are defined on data sets. In Section 6.6, we discuss these new risk measures. Since the risk measures are not our main topic here, we remind the reader of two famous risk measures:

- The value-at-risk of X for $\alpha \in (0, 1)$ is given by

$$VaR_\alpha(X) = -\inf\{z; F_X(z) > \alpha\}.$$

- The conditional value-at-risk is given by

$$CVaR_\alpha(X) = -\mathbb{E}\left[X | X \leq -VaR_\alpha(X)\right],$$

  when F_X is continuous at $-VaR_\alpha(X)$.

Note that VaR is not a subadditive risk measure and it is hard to work with in optimization problems. Subadditivity will be one of our main assumptions. For this reason, our theory is not applicable for risk measures like VaR.

Now assume that $\rho : \Pi \to \mathbb{R}$ is any risk measure that satisfies the two conditions, subadditivity $\rho(x+y) \leq \rho(x) + \rho(y)$ and positive homogeneity $\rho(tx) = t\rho(x)$ for every $t \geq 0$ and $x, y \in \Pi$. Later we may need to impose more conditions on ρ, but for the moment this is all we need. The composition of ρ and Π_T, that we call $\overline{\rho} = \rho \circ \Pi_T$, defines a risk measure on $\mathbb{R}^m$ into $\mathbb{R}$ that satisfies the subadditivity and positive homogeneity properties. As one can see below, the additivity of Π_T is a critical condition for the subadditivity of $\overline{\rho}$:

$$\begin{aligned}\overline{\rho}(\overline{x} + \overline{y}) &= \rho(\Pi_T(\overline{x} + \overline{y})) \\ &= \rho(\Pi_T(\overline{x}) + \Pi_T(\overline{y})) \leq \overline{\rho}(\overline{x}) + \overline{\rho}(\overline{y}),\end{aligned}$$

where the second equality holds by the additivity of Π_T, while the inequality is due to the subadditivity of ρ.

In this work, attention is paid to $\overline{\rho}$ and its underlying space $\mathbb{R}^m$. It is the risk measure that will be used to analyze the bond market. From now on, we rarely talk about the space $L^2(\Omega, \mathfrak{F}, \mathbb{P})$ and the measure ρ. We will mostly focus on the space $\mathbb{R}^m$ and the measure $\overline{\rho}$. This allows us to work in a simplified structural world that serves as a space for all cash flows.

6.3 Representation Theorem

Finding a suitable representation of $\overline{\rho}$ plays an important role in the next few sections. We see how the representation theorem helps us analyze and formulate the primal and dual problems.

Theorem 6.1. *Assume that* $\overline{\rho} : \mathbb{R}^m \to \mathbb{R}$ *is as in Section 6.2, then*

$$\overline{\rho}(\overline{y}) = \max\{-\overline{y} \cdot \overline{z}; \overline{z} \in \Delta_{\overline{\rho}}\}, \tag{6.1}$$

where

$$\Delta_{\overline{\rho}} = \{\overline{z} \in \mathbb{R}^m;\ \overline{\rho}(\overline{y}) \geq -\overline{y} \cdot \overline{z},\ \textit{for all}\ \overline{y} \in \mathbb{R}^m\}, \tag{6.2}$$

and $\overline{y} \cdot \overline{z}$ *is the usual inner product on* $\mathbb{R}^m$.

Remark 6.1. *By* max *in the above theorem or in the rest of the section we implicitly mean that the maximum is attained. For example in this theorem it turns out that there exists* $\overline{z_*} \in \Delta_{\overline{\rho}}$ *such that* $\overline{\rho}(\overline{y}) = -\overline{y} \cdot \overline{z_*}$. *For a general interpretation of the sub-gradient set* $\Delta_{\overline{\rho}}$, *we refer to Artzner et al. (1999).*

To prove the theorem, we need the following lemma which is an interesting result in itself.

Lemma 6.1. *Suppose that:*

1. M *is a subspace of a real vector space* X,

2. $\rho : X \to \mathbb{R}$ *satisfies*

$$\rho(x + y) \leq \rho(x) + \rho(y),\ \textit{and}\ \rho(tx) = t\rho(x),$$

for all $x,\ y \in X$ *and* $t \geq 0$,

In addition define

1. $\Delta_\rho = \{x^* \in X^*; -x^*(x) \leq \rho(x),\ \textit{for all}\ x \in X\}$, *and*

2. $\Delta_{(M,\rho)} = \{m^* \in M^*; -m^*(m) \leq \rho(m),$ *for all* $m \in M\}$,

where X^* *and* M^* *are, respectively, the dual spaces of* X *and* M.

Then for any $x \in X$,

$$\rho(x) = \max\{-x^*(x); x^* \in \Delta_\rho\}$$

holds if and only if

$$\rho(x) = \max\{-u^*(x); u^* \in \Delta_{(U,\rho)}\}$$

holds for the one dimensional subspace U *generated by* x, *i.e.* $U =\prec x \succ$. *Furthermore, we show that the maximum in these problems is actually reached.*

Proof. Assume that $\rho(x) = \max\{-x^*(x); x^* \in \Delta_\rho\}$ holds for $x \in X$, then there exists $y^* \in \Delta_\rho$ such that $\rho(x) = -y^*(x)$. Now define $z^* = y^*_{|U}$, that is the restriction of y^* onto $U =\prec x \succ$. Since $y^* \in \Delta_\rho$, clearly we have $z^* \in \Delta_{(U,\rho)}$. Now for any $u^* \in \Delta_{(U,\rho)}$ we have $-u^*(x) \leq \rho(x) = -y^*(x) = -z^*(x)$, so $\max\{-u^*(x); u^* \in \Delta_{(U,\rho)}\} = -z^*(x) = \rho(x)$.

For the second half, assume that for the subspace $U =\prec x \succ$, $\rho(x) = \max\{-u^*(x); u^* \in \Delta_{(U,\rho)}\}$ holds. Then there exists $u^* \in \Delta_{(U,\rho)}$ such that $\rho(x) = -u^*(x)$. Since $u^* \in \Delta_{(U,\rho)}$, we conclude that for every $u \in U$, $-u^*(u) \leq \rho(u)$. By the theorem of Hahn-Banach (see Rudin, 1982) there exists $\Lambda \in X^*$ such that $\Lambda|_U = u^*$ and $-\Lambda \leq \rho$ on X, so that $\Lambda \in \Delta_\rho$. Now if one takes any $z^* \in \Delta_\rho$, then by the definition of Δ_ρ, $-z^*(x) \leq \rho(x) = -u^*(x) = -\Lambda(x)$. So we have

$$\max\{-x^*(x); x^* \in \Delta_\rho\} = -\Lambda(x) = \rho(x).$$

Now we show that the maximum in the second problem is reached

$$\rho(x) = \max\{-u^*(x); u^* \in \Delta_{(U,\rho)}\},$$

$$\Delta_{(U,\rho)} = \{m^* \in U^*; -m^*(u) \leq \rho(u), \text{ for all} u \in U\},$$

where $U =\prec x \succ$. Since $U^* = \mathbb{R}$, (in the sense that there is a one to one and onto linear functional from U^* into $\mathbb{R}$.) a simple argument shows that this is the same as solving the following problem:

$$\rho(x) = \max\{-\beta; \beta \in \Delta_{(U,\rho)}\},$$

$$\Delta_{(U,\rho)} = \{\beta \in \mathbb{R}; -\alpha\beta \leq \rho(\alpha x), \text{ for all } \alpha \in \mathbb{R}\}.$$

If $\beta \in \Delta_{(U,\rho)}$ then by taking $\alpha = 1$ and $\alpha = -1$, one can easily see that $-\rho(x) \leq \beta \leq \rho(-x)$, or $\Delta_{(U,\rho)} \subset [-\rho(x), \rho(-x)]$. On the other hand, if $\alpha \in \mathbb{R}^+$ and $\beta \geq -\rho(x)$ then $\beta \geq -\rho(\frac{\alpha x}{\alpha}) = -\frac{1}{\alpha}\rho(\alpha x)$, so we have that $-\alpha\beta \leq \rho(\alpha x)$. Similarly, one can easily show that if $\alpha \in \mathbb{R}^-$ and $\beta \leq -\rho(-x)$ then $-\alpha\beta \leq \rho(\alpha x)$. Therefore $\Delta_{(U,\rho)} = \{\beta \in \mathbb{R} : -\rho(x) \leq \beta \leq \rho(-x)\}$ and it follows that $\max\{-\beta; \beta \in \Delta_{(U,\rho)}\} = \rho(x)$. Please note that for any x we have $-\rho(x) \leq \rho(-x)$, hence in any case $\Delta_{(U,\rho)}$ is not empty. □

We are now ready to prove Theorem 6.1.

Proof. (Theorem 6.1): This is a straight forward application of Lemma 3.1 and Riesz's Theorem (see Rudin, 1966) which states that any bounded linear functional x^* on $\mathbb{R}^m$ can be uniquely represented by an element $\overline{y}$, i.e. $x^*(\overline{x}) = \overline{y} \cdot \overline{x}$, for all $\overline{x} \in \mathbb{R}^m$. □

6.4 Measurement of ρ-arbitrage

As we already mentioned, the type of arbitrage studied in this work is defined through risk measures. The main result in this section gives a necessary and sufficient condition for the existence of what we call ρ-arbitrage. First we present its definition.

Definition 6.1. *(ρ-arbitrage) Assume that $\overline{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is representing a portfolio consisting of x_j units of bond j, for $j = 1, 2, ..., n$. Then $\overline{x}$ is said to be a ρ-arbitrage portfolio if*

$$\overline{p} \cdot \overline{x} < 0 \quad \textit{and} \quad \overline{\rho}(\sum_{j=1}^{n} x_j a_{1j}, ..., \sum_{j=1}^{n} x_j a_{mj}) \leq 0$$

or, equivalently, $\overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m) \leq 0$ where $\overline{a}_i = (a_{i1}, ..., a_{in}) \in \mathbb{R}^m$ for $i = 1, 2, ..., m$.

To help interpreting the above definition, note that $\overline{p} \cdot \overline{x}$ is the current price of the portfolio, while for any $1 \leq i \leq m$, $\sum_{j=1}^{n} x_j a_{ij}$ is the total cash flow of the portfolio at time t_i. Hence the above definition simply says that there is no cost for the portfolio $\overline{x}$ and at the same time as the risk measure for the portfolio cash flow is non-positive. Considering the fact that in a $\overline{\rho}$-arbitrage portfolio, $-\overline{p} \cdot \overline{x}$ is the arbitrage income, we propose the following optimization problem that leads us to obtain the main result of

this section:

$$\begin{aligned} &Maximize\ -\overline{p}\cdot\overline{x},\\ &such\ that\ \overline{\rho}(\overline{x}\cdot\overline{a}_1,...,\overline{x}\cdot\overline{a}_m)\leq 0,\\ &\overline{x}+\overline{h}\geq 0,\\ &\overline{h}\cdot\overline{p}\leq 1,\\ &\overline{h}\geq 0, \end{aligned} \tag{6.3}$$

where $(\overline{x},\overline{h})\in\mathbb{R}^n\times(\mathbb{R}^+\cup\{0\})^n$ are the decision variables. As mentioned above, $\overline{x}$ represents the portfolio composition. If it respects the above constraints, $\overline{h}\geq 0$ can be interpreted as an upper bound portfolio for the short sales, whose total price can not be larger than one unit.

Using Theorem 6.1 this optimization problem is equivalent to

$$\begin{aligned} &Maximize\ -\overline{p}\cdot\overline{x},\\ &such\ that\ (\overline{x}\cdot\overline{a}_1,...,\overline{x}\cdot\overline{a}_m)\cdot\overline{z}\geq 0,\quad \forall\overline{z}\in\Delta_{\overline{\rho}}\\ &\overline{x}+\overline{h}\geq 0,\\ &\overline{h}\cdot\overline{p}\leq 1,\\ &\overline{h}\geq 0. \end{aligned} \tag{6.4}$$

We call this the primal problem. This is very similar to the optimization problems considered in Balbás, Balbás and Garrido (2010) and Balbás and López (2010), the same type of analysis works here as well. To obtain most of the results in the rest of this section, we use constrained optimization theory, as presented in Chapter 8 of Luenberger (1969). As the first step, we find the Lagrangian of this optimization

problem. Following the notation of Luenberger (1969), this primal problem can be represented as:

$$\begin{aligned} &Maximize\ -\overline{p}\cdot\overline{x},\\ &such\ that\ G(\overline{x},\overline{h})\leq 0,\\ &\overline{h}\geq 0, \end{aligned} \tag{6.5}$$

where G is a convex mapping from $\mathbb{R}^n\times(\mathbb{R}^+\cup\{0\})^n$ into $Z=\varphi(\Delta_{\overline{\rho}})\times\mathbb{R}^n\times\mathbb{R}$ given by

$$G(\overline{x},\overline{h})=(-g_{\overline{x}},-\overline{x}-\overline{h},\overline{h}\cdot\overline{p}-1),$$

$\varphi(\Delta_{\overline{\rho}})$ is the space of the continuous functions on the $weak^*$ compact $\Delta_{\overline{\rho}}$, and for any $\overline{x}\in\mathbb{R}^n$, $g_{\overline{x}}:\mathbb{R}^n\to\mathbb{R}$ is given by $g_{\overline{x}}(\overline{z})=(\overline{x}\cdot\overline{a_1},...,\overline{x}\cdot\overline{a}_m)\cdot\overline{z}$. To make the notation consistent with Luenberger (1969), we consider the negative of (6.5):

$$\begin{aligned} &Minimize\ \overline{p}\cdot\overline{x},\\ &such\ that\ G(\overline{x},\overline{h})\leq 0,\\ &\overline{h}\geq 0. \end{aligned} \tag{6.6}$$

With the above explanations and knowing that the dual space of $\varphi(\Delta_{\overline{\rho}})$ is $\mathcal{M}(\Delta_{\overline{\rho}})$, and the space of inner regular real-valued σ-additive measures on the Borel σ-algebra of $\Delta_{\overline{\rho}}$ (endowed with the $weak^*$ topology), then the Lagrangian function $\mathcal{L}:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ is

$$\begin{aligned} \mathcal{L}(\overline{x},\overline{h},\overline{\lambda},\lambda)=&\sum_{j=1}^{n}x_jp_j-\int_{\Delta_{\overline{\rho}}}(\overline{x}.\overline{a}_1,...,\overline{x}\cdot\overline{a}_m).\overline{z}\ dv(\overline{z})\\ &-\sum_{j=1}^{n}\lambda_j(x_j+h_j)+\lambda(\sum_{j=1}^{n}h_jp_j-1), \end{aligned}$$

where v belongs to $\mathcal{M}(\Delta_{\overline{\rho}})$. After some manipulations, this can be rewritten as

$$\begin{aligned}\mathcal{L}(\overline{x},\overline{h},\overline{\lambda},\lambda) = &\sum_{j=1}^{n}(p_j - \int_{\Delta_{\overline{\rho}}} \overline{b}_j \cdot \overline{z}\; dv(\overline{z}) - \lambda_j)x_j \\ &+ \sum_{j=1}^{n}(-\lambda_j + \lambda p_j)h_j - \lambda,\end{aligned} \tag{6.7}$$

where $\overline{b}_j = (a_{1j}, a_{2j}, ..., a_{mj})$.

It is one of the fundamental results of Luenberger (1969) that an element $(v, \overline{\lambda}, \lambda) \in \mathcal{M}(\Delta_\rho) \times \mathbb{R}^n \times \mathbb{R}$ is dual feasible if and only if it belongs to the non-negative cone $\mathcal{M}_+(\Delta_\rho) \times (\mathbb{R}^+ \cup \{0\})^n \times (\mathbb{R}^+ \cup \{0\})$ and

$$\inf\{\mathcal{L}(\overline{x},\overline{h},\overline{\lambda},\lambda); (\overline{x},\overline{h}) \in \mathbb{R}^n \times (\mathbb{R}^+ \cup \{0\})^n\} > -\infty.$$

In that case, the infimum above is equal to the optimal value of the dual problem. Now using this fact and the Lagrangian in (6.7), one can obtain the dual problem of (6.6). The negative of this dual problem is the dual of our primal problem (6.4)

$$\begin{aligned}&\textit{Minimize } \theta, \\ &\textit{such that } p_j = \lambda_j + \int_{\Delta_{\overline{\rho}}} \overline{b}_j \cdot \overline{z}\; dv(\overline{z}),\ j = 1, 2, ...n, \\ &\qquad \overline{\lambda} \leq \theta\overline{p}, \\ &\qquad \theta \geq 0,\ \overline{\lambda} \geq 0,\ v \in \mathcal{M}_+(\Delta_{\overline{\rho}}).\end{aligned} \tag{6.8}$$

From now on we assume that the maximum in the primal problem (6.4) is attained. Then following Luenberger (1969), $(\overline{x}^*, \overline{h}^*)$ is the solution of the primal problem if and

only if there exists $(\theta^*, \overline{\lambda}^*, v^*) \in (\mathbb{R}^+ \cup \{0\}) \times (\mathbb{R}^+ \cup \{0\})^n \times \mathcal{M}_+(\Delta_{\overline{\rho}})$ such that

$$
\begin{aligned}
& p_j = \lambda_j^* + \int_{\Delta_{\overline{\rho}}} \overline{b}_j \cdot \overline{z} \; dv^*(\overline{z}), \quad j = 1, 2, ..., n, \\
& \overline{\lambda}^* \leq \theta^* \; \overline{p}, \\
& \sum_{j=1}^{n} x_j^* \int_{\Delta_{\overline{\rho}}} \overline{b}_j \cdot \overline{z} \; dv^*(\overline{z}) = 0, \quad j = 1, 2, ..., n, \\
& \overline{\lambda}^* \cdot (\overline{x}^* + \overline{h}^*) = 0, \\
& \theta^* (\overline{h}^* \cdot \overline{p} - 1) = 0, \\
& \overline{h}^* \geq 0, \; \overline{h}^* \cdot \overline{p} \leq 1, \; \overline{x}^* + \overline{h}^* \geq 0.
\end{aligned}
$$

A simple calculation shows that these are equivalent to

$$
\begin{aligned}
& p_j = \lambda_j^* + \int_{\Delta_{\overline{\rho}}} \overline{b}_j \cdot \overline{z} \; dv^*(\overline{z}), \quad j = 1, 2, ..., n, \\
& \overline{\lambda}^* \leq \theta^* \; \overline{p}, \\
& \overline{x}^* \cdot (\overline{p} - \overline{\lambda}^*) = 0, \\
& \overline{\lambda}^* \cdot (\overline{x}^* + \overline{h}^*) = 0, \\
& \theta^* (\overline{h}^* \cdot \overline{p} - 1) = 0, \\
& \overline{h}^* \geq 0, \; \overline{h}^* \cdot \overline{p} \leq 1, \; \overline{x}^* + \overline{h}^* \geq 0.
\end{aligned}
\tag{6.9}
$$

In the literature, these are called Karush-Kuhn-Tucker conditions. The dual problem (6.8) and optimality Equations (6.9) can be further simplified. To do this, we use the following mean-value type theorem and simplify the integral in both the first constraint of the dual problem (6.8) and the first Karush-Kuhn-Tucker equation of (6.9). The proof of the lemma is taken from Balbás et al. (2010).

Lemma 6.2. *Assume that $\mathbb{R}^m$ is equipped with the usual inner product, $\Delta_{\overline{\rho}} = \{\overline{z} \in \mathbb{R}^m; \overline{\rho}(\overline{y}) \geq -\overline{y} \cdot \overline{z}$ for every $\overline{y} \in \mathbb{R}^m\}$, and v is a positive measure that belongs to $\mathcal{M}_+(\Delta_{\overline{\rho}})$. Then there exists $\overline{z}_v \in \Delta_{\overline{\rho}}$ such that*

$$\int_{\Delta_{\overline{\rho}}} \overline{y} \cdot \overline{z}\, dv(\overline{z}) = \overline{y} \cdot \overline{z}_v v(\Delta_{\overline{\rho}}),\ \textit{for every}\ \overline{y} \in \mathbb{R}^m.$$

Proof. : Define the functional $F : \mathbb{R}^m \to \mathbb{R}$ by

$$F(\overline{y}) = \frac{\int_{\Delta_{\overline{\rho}}} \overline{y} \cdot \overline{z}\, dv(\overline{z})}{v(\Delta_{\overline{\rho}})},$$

then by Cauchy-Schwartz, inequality we have that $|F(\overline{y})| \leq c||\overline{y}||_2$, where $c = \frac{\int_{\Delta_{\overline{\rho}}} \|\overline{z}\|_2\, dv(\overline{z})}{v(\Delta_{\overline{\rho}})}$. The boundness of $\Delta_{\overline{\rho}}$ and finiteness of the measure v implies $c < \infty$; so F is a bounded linear functional on $\mathbb{R}^m$, which is a Hilbert space when endowed with the usual inner product. So by the Riesz theorem there is an element $\overline{z}_v \in \mathbb{R}^m$ such that for every $\overline{y} \in \mathbb{R}^m$, $F(\overline{y}) = \overline{y} \cdot \overline{z}_v$, or $\int_{\Delta_{\overline{\rho}}} \overline{y} \cdot \overline{z}\, dv(\overline{z}) = \overline{y} \cdot \overline{z}_v v(\Delta_{\overline{\rho}})$. Furthermore, $\overline{z}_v$ is uniquely determined by F and $||F|| = ||\overline{z}_v||_2$.

Therefore, it is enough to prove that $\overline{z}_v \in \Delta_{\overline{\rho}}$. For every $\overline{y} \in \mathbb{R}^m$,

$$-\overline{z}_v \cdot \overline{y} = \frac{\int_{\Delta_{\overline{\rho}}} -\overline{y} \cdot \overline{z}\, dv(\overline{z})}{v(\Delta_{\overline{\rho}})} \leq \frac{\overline{\rho}(\overline{y}) v(\Delta_{\overline{\rho}})}{v(\Delta_{\overline{\rho}})} = \overline{\rho}(\overline{y}),$$

and clearly this shows that $\overline{z}_v \in \Delta_{\overline{\rho}}$. □

The previous explanations and Lemma 6.2 lead us to the following theorem.

Theorem 6.2. *Assume that the primal problem is always finite, it reaches its optimal value, and so the optimal solutions always exist. Then*

1. *The equivalent dual form of the primal problem is*

$$\begin{aligned} &Minimize\ \theta, \\ &such\ that\ p_j = \lambda_j + \alpha \bar{b}_j \cdot \bar{z}, \\ &\overline{\lambda} \leq \theta \overline{p}, \\ &\theta \geq 0,\ \overline{\lambda} \geq 0,\ \alpha \geq 0,\ \bar{z} \in \Delta_{\overline{\rho}}. \end{aligned} \tag{6.10}$$

2. $(\overline{x}^*, \overline{h}^*)$ *and* $(\theta^*, \overline{\lambda}^*, \alpha^*, \overline{z}^*)$ *solve problems* (6.4), *and* (6.10) *respectively, if and only if they satisfy the following Karush-Kuhn-Tucker conditions*

$$\begin{aligned} &p_j = \lambda_j^* + \alpha^* \bar{b}_j \cdot \overline{z}^*, \quad j = 1, 2, ..., n, \\ &\overline{\lambda}^* \leq \theta^* \overline{p}, \\ &\overline{x}^* \cdot (\overline{p} - \overline{\lambda}^*) = 0, \\ &\overline{\lambda}^* \cdot (\overline{x}^* + \overline{h}^*) = 0, \\ &\theta^* (\overline{h}^* \cdot \overline{p} - 1) = 0, \\ &\overline{h}^* \geq 0,\ \overline{h}^* \cdot \overline{p} \leq 1,\ \overline{x}^* + \overline{h}^* \geq 0, \\ &\theta^* \geq 0,\ \overline{\lambda}^* \geq 0,\ \overline{z}^* \in \Delta_{\overline{\rho}},\ \alpha^* \geq 0. \end{aligned} \tag{6.11}$$

The following interesting lemma bridges the optimal solution of the primal problem to the existence of ρ-arbitrage. The proof is a straightforward adaptation of a similar one in Balbás and López (2008).

Lemma 6.3. *Assume that* $\mathcal{L}_*$ *is the optimal value of the primal problem. Then the market is* ρ*-arbitrage free if and only if* $\mathcal{L}_* = 0$.

Proof. First assume that the market is ρ-arbitrage free. If $\mathcal{L}_* = (\overline{x}^*, \overline{h}^*)$ is any optimal solution then $\rho(\sum_{j=1}^n x_j^* a_{1j}, ..., \sum_{j=1}^n x_j^* a_{mj}) \leq 0$. Since the market is ρ-arbitrage free,

we have $\overline{x}^* \cdot \overline{p}^* \geq 0$ or $-\overline{x}^* \cdot \overline{p}^* \leq 0$. Obviously $(\overline{0}, \overline{0})$ is primal-feasible so $-\overline{x}^* \cdot \overline{p}^* \geq 0$, therefore $\mathcal{L}_* = \overline{x}^* \cdot \overline{p}^* = 0$.

For the second half assume that $\mathcal{L}_* = 0$. We show that there is no ρ-arbitrage portfolio. Suppose that $\rho(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m) \leq 0$, and take

$$\overline{y} = \begin{cases} \frac{\overline{x}}{\overline{p} \cdot \overline{x}^-} & \text{if } \overline{p} \cdot \overline{x}^- \geq 1 \\ \overline{x} & \text{if } \overline{p} \cdot \overline{x}^- \leq 1 \end{cases}, \qquad \overline{h} = \begin{cases} \frac{\overline{x}^-}{\overline{p} \cdot \overline{x}^-} & \text{if } \overline{p} \cdot \overline{x}^- \geq 1 \\ \overline{x}^- & \text{if } \overline{p} \cdot \overline{x}^- \leq 1 \end{cases},$$

where $\overline{x}^- = \max(-\overline{x}, 0)$. Then $(\overline{y}, \overline{h})$ are primal feasible, because:

$$\overline{\rho}(\overline{y} \cdot \overline{a}_1, ..., \overline{y} \cdot \overline{a}_m) = \begin{cases} \frac{\overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m)}{\overline{p} \cdot \overline{x}^-} & \text{if } \overline{p} \cdot \overline{x}^- \geq 1 \\ \overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m) & \text{if } \overline{p} \cdot \overline{x}^- \leq 1 \end{cases},$$

in any of the previous cases we have $\overline{\rho}(\overline{y} \cdot \overline{a}_1, ..., \overline{y} \cdot \overline{a}_m) \leq 0$. It is obvious that $\overline{y} + \overline{h} \geq 0$ and $\overline{h} \cdot \overline{p} \leq 1$. Since $\mathcal{L}_* = 0$, then $-\overline{p} \cdot \overline{y} \leq \mathcal{L}_* = 0$, so $\overline{p} \cdot \overline{y} \geq 0$. The simple outcome of the last inequality is that $\overline{p} \cdot \overline{x} \geq 0$. So there is no ρ-arbitrage portfolio. □

Combining this lemma with Theorem 6.2 leads to the following result.

Theorem 6.3. *There is no ρ-arbitrage portfolio if and only if there exists $(\overline{z}^*, \alpha^*) \in \Delta_{\overline{\rho}} \times \mathbb{R}^+$ such that for every j,*

$$p_j = \alpha^* \overline{z}^* \cdot \overline{b_j}, \ j = 1, 2, ..., n,$$

where $\overline{b_j} = (a_{1j}, a_{2j}, ..., a_{mj})$ is the j's column of matrix A.

Proof. By Lemma 6.3 and Theorem 6.2, the non-existence of ρ-arbitrage is equivalent to $\theta^* = 0$. If $\theta^* = 0$ then by the second and last condition of part 3 of Theorem 6.2, we have $\overline{\lambda}^* = 0$, and therefore the first condition gives that $p_j = \alpha^* \overline{b_j} \cdot \overline{z}^*$. On the other

hand if there exists $(\overline{z}^*, \alpha^*) \in \Delta_{\overline{\rho}\times\mathbb{R}^+}$ such that for every j, $p_j = \alpha^*\overline{z}^* \cdot \overline{b_j}$, the first and third condition of Theorem 6.2 give $\overline{\lambda}^* = 0$ and $\overline{x}^* \cdot \overline{p} = 0$, respectively. So the optimal value of the primal problem is zero and consequently $\theta^* = 0$. □

By Theorem 6.3, the existence of ρ-arbitrage is linked to solving the following system of equations

$$p_j = \alpha^*\overline{z}^* \cdot \overline{b_j}, \quad j = 1, 2, ..., n,$$

where $\overline{b_j} = (a_{1j}, a_{2j}, ..., a_{mj})$ is the j's column of matrix A. There are two cases for the solution set of the above system. Either it is empty or non-empty. If the solution set is non-empty then the existence of ρ-arbitrage (for a specific risk measure ρ) reduces to whether the solution belongs to $\Delta_{\overline{\rho}} \times \mathbb{R}^+$ or not. If the solution is in this set then for this particular risk measure ρ, there is ρ-arbitrage in the market.

On the other hand if the solution set is empty, then by the above theorem this means that for any risk measure ρ satisfying subadditivity and positive homogeneity, there is ρ-arbitrage. In other words, for such risk measures ρ, the existence of ρ-arbitrage is guaranteed. In this case it is easy to prove that classical arbitrage also exists; see the following corollary.

Corollary 6.1. *Suppose that the solution set of the following system of equations is empty,*

$$p_j = \alpha^*\overline{z}_* \cdot \overline{b_j}, \ j = 1, 2, ..., n,$$

where $\overline{b_j} = (a_{1j}, a_{2j}, ..., a_{mj})$ *is the* j*th column of matrix* A*. Then the existence of*

classical arbitrage in the market is guaranteed.

Now, we implement the above concepts in a simple model. The portfolios in the following examples of this section are naive, far from being real. However these examples illustrate well the above theory. Later, we present more realistic examples with real portfolio data.

Assume that our probability space includes only two scenarios, i.e. $\Omega = \{\omega_1, \omega_2\}$ with $\mathbb{P}(\omega_1) = 1 - q$, $\mathbb{P}(\omega_2) = q$. Take the risk measure to be $\rho = CVaR_\alpha$, for some $\alpha \in (0, 1)$. To model the evolution of the interest rate, we use the following one period simple tree model

$$r = \begin{cases} r_1, & \text{with probability } 1 - q; \\ r_2, & \text{with probability } q. \end{cases}$$

Since ρ is $CVaR_\alpha$, from Rockafellar, Uryasev and Zabarankin (2006), it can be proved that the subgradient set of $\rho = CVaR_\alpha$ is the following,

$$\Delta_\rho = \{(z_1, z_2) \in \mathbb{R}^2; z_1(1 - q) + z_2 q = 1, 0 \leq z_i \leq \frac{1}{\alpha}, \text{ for } i = 1, 2\}. \tag{6.12}$$

Using this subgradient, the representation Theorem 6.1, and manipulations one can show that the subgradient set of $\overline{\rho}$ is given by

$$\Delta_{\overline{\rho}} = \{(\overline{z}_1, 1) \in \mathbb{R}^2; (1 + r_1) - \frac{q}{\alpha}(r_1 - r_2) \leq \overline{z}_1 \leq (1 + r_1)\}. \tag{6.13}$$

Having a closed-form subgradient is essential in solving the optimization problems.

Example 6.1. *Assume that the interest rate tree parameters are* $q = 0.9$, $r_1 = 0.05$, $r_2 = 0.03$, *and* $\alpha = 0.05$. *Suppose that we have the two following bonds, Table 6.1. In*

	Bond 1	Bond 2
Price of the bonds	1010	908.9
Cash flow 1	10	9
Cash flow 2	1010	909

Table 6.1: ρ-arbitrage free, two bonds portfolio.

this case $\theta^* \approx 0$.

There is almost no sensitivity to any changes in the parameters α, q, r_1, *and* r_2 *in the above example. This means that different ranges of these parameters lead to the same optimal value. This is because the portfolio is quite consistent, in the sense that both corporations have used the same rule to price their bonds. Therefore, one can not benefit from any mispricing in the portfolio. Notice that arbitrage opportunities arise due to inconsistencies in the market.*

Different perceptions lead to different rules, and hence different prices. Here, the yield of this portfolio is very low, so the interest rate is substantially lower than the future interest rates r_1 *and* r_2. *However since both firms agree on the same interest rate and yield (even if ridiculously low), there is no inconsistency in the portfolio and so no arbitrage opportunities with these bonds.*

Example 6.2. *Assume that as above,* $q = 0.9$, $r_1 = 0.05$, $r_2 = 0.03$, *and the* $CVaR$ *probability* $\alpha = 0.05$. *By contrast now we consider two bonds represented by Table 6.2.*

	Bond 1	Bond 2
Price of the bonds	1010	908.9
Cash flow 1	10	1
Cash flow 2	1010	909

Table 6.2: ρ-arbitrage, two bonds portfolio.

For this portfolio the optimal value is approximately equal to 0.00869*. The solution of the primal problem is equal to* $x_1^* \approx 0.00098$*,* $x_2^* \approx -0.0011$*. The solution of the dual problem is equal to* $\theta^* \approx 0.00869$*,* $\lambda_1^* \approx 0$*,* $\lambda_2^* \approx 7.8999$*,* $\alpha^* \approx 0.9901$*,* $\overline{z}_1^* \approx 1$*, and* $\overline{z}_2^* \approx 0.9901$*. Here, the existence of* ρ*-arbitrage is due to the inconsistency in the portfolio. The two firms have used different rules to price the bonds. No matter who is right or wrong, this is a situation where the manipulation of bonds can create* ρ*-arbitrage. Notice that the solution of the primal problem is in fact the quantity of each bond needed to gain the arbitrage income, a negative value indicating a short portfolio.*

In this case, analyzing sensitivity is not that straightforward as in the previous example. The optimal solution is sensitive to changes in α, q, r_1, and r_2. However this requires more caution. For instance, Figure 6.1 graphs the optimal value as a function of α. The larger the parameter α, the riskier the agents get. Hence it makes sense that the arbitrage should increase as well. The parameter α starts at the value of 1% for which the optimal value θ^* is approximately 0.00730. This phenomena is not present in classical arbitrage as no matter the underlying conditions the amount of arbitrage is pre-determined. Clearly ρ-arbitrage is not necessarily the same as the classical one. Another important observation from Figure 6.1 is that the solution is stable for a wide

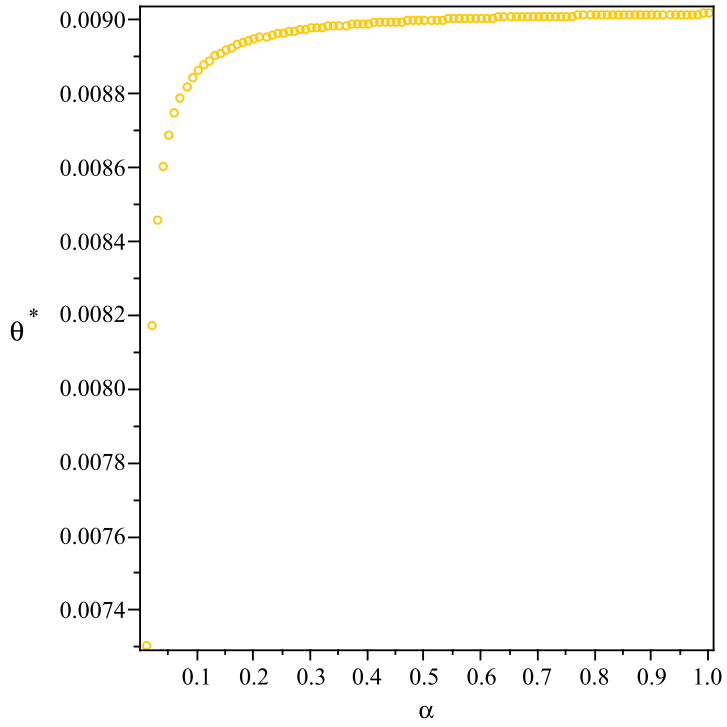


Figure 6.1: Optimal θ^* values as a function of α.

range of values for the parameter α; from a very risk averse agent (α =1%) to very risk seeking ones ($\alpha = 99\%$), the range of θ^* values is of about 0.2% difference. A similar reasoning can be carried out for the other parameters. Figure 6.2 gives the optimal θ^* values with respect to r_1, while the parameter $r_2 = 0.03$ is fixed. Figure 6.3 is similar, but this time $r_1 = 0.05$ is fixed and r_2 varies.

Figure 6.4 reports the optimal θ^* values, with respect to the probability q and the price of the second bond p_2. Notice that at a price of $p_2 = 908.9$ the optimal value is near zero or, in other words, arbitrage disappears. For any other price, either smaller or larger, there is a non-zero arbitrage opportunity.

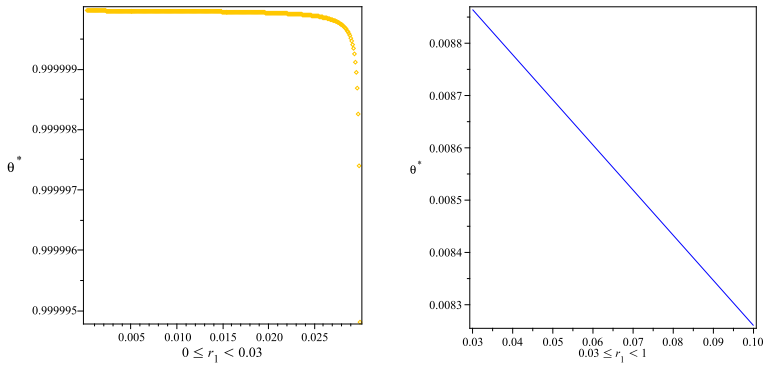


Figure 6.2: Optimal θ^* values as a function of r_1

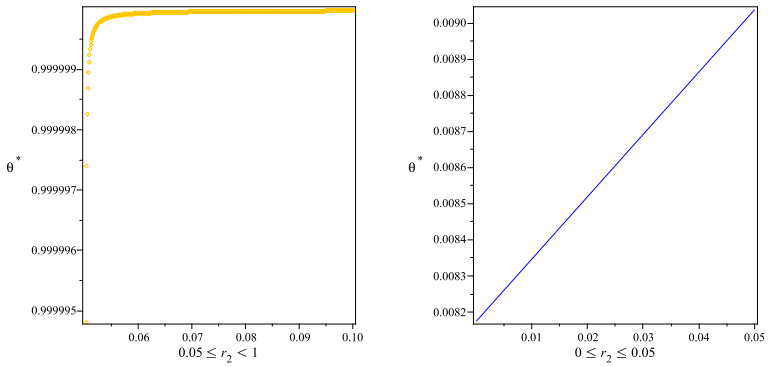


Figure 6.3: Optimal θ^* values as a function of r_2

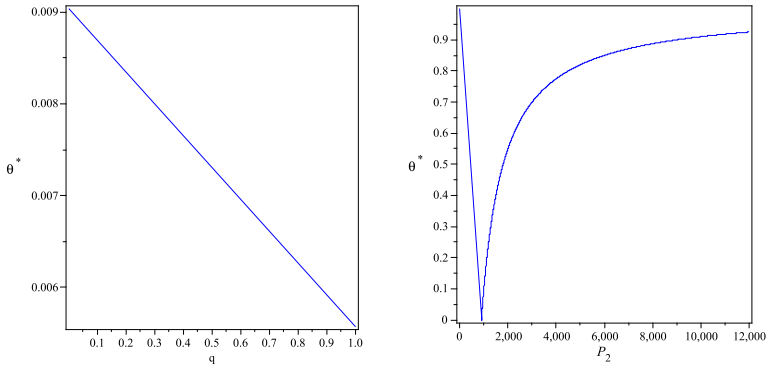


Figure 6.4: Optimal θ^* values as a function of q and the price of the second bond p_2

6.5 The Revised Problem

The problem that was investigated in the previous section can be modified in other ways. These modifications can help finding a better term structure of interest rate (TSIR) envelopes (lower and upper bounds) and credit risk spreads. Also as a special case, this includes the problem of sequential arbitrage measurements, see Balbás and López (2008).

Section 6.4, we focused on maximizing the arbitrage income subject to the first constraint of problem (6.3), the risk constraint. Another perspective is to maximize the arbitrage income and minimize the risk simultaneously. In other words, we want to maximize the objective vector function $(-\overline{p}\cdot\overline{x}, -\overline{\rho}(\overline{x}\cdot\overline{a}_1, ..., \overline{x}\cdot\overline{a}_m))$ over a constraint set that will be specified soon. This is a multi-objective optimization problem and the solution(s) of this problem are Pareto optimal, see `http://en.wikipedia.org/wiki/Multi-objective_optimizationformoredetails`: A design point $\overline{x}^* = (x_1^*, x_2^*, ..., x_n^*)$, in objective space, is termed Pareto optimal if there does not exist another feasible design objective vector $\overline{x} = (x_1, x_2, ..., x_n)$ such that $x_i \leq x_i^*$ for all i in $\{1, 2, \ldots, n\}$, and $x_j < x_j^*$ for at least one index of j, $j \in \{1, 2, \ldots, n\}$.

Notice that by maximizing $-\overline{\rho}(\overline{x}\cdot\overline{a}_1, ..., \overline{x}\cdot\overline{a}_m)$, the risk decreases. Maximizing this objective vector function is equivalent to $-Minimize\,(\overline{p}\cdot\overline{x}, \overline{\rho}(\overline{x}\cdot\overline{a}_1, ..., \overline{x}\cdot\overline{a}_m))$. Hence arbitrage can be given different definitions. For the main constraint, we select strong sequential arbitrage (SSA), i.e. $I_m^* A\overline{x} \geq 0$, for $A = [a_{ij}]_{m\times n}$ as in Section 6.2, any

$\overline{x} \in \mathbb{R}$ and where I_m^* is the $m \times m$ matrix

$$\begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ . & . & . & & . \\ . & . & . & ... & . \\ . & . & . & & . \\ 1 & 1 & 1 & ... & 1 \end{pmatrix}.$$

Therefore we consider the following vector optimization problem:

$$\begin{aligned} &Minimize\ \ (\overline{p} \cdot \overline{x}, \overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m)), \\ &such\ that\ I_m^* A\overline{x} \geq 0, \\ &\overline{x} + \overline{h} \geq 0, \\ &\overline{h} \cdot \overline{p} \leq 1, \\ &\overline{h} \geq 0. \end{aligned} \tag{6.14}$$

Since $\overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m)$ is a convex function, for every optimal solution $(\overline{x}^*, \overline{h}^*)$ of (6.14) there exists the non-zero vector $(\gamma, \beta) \geq 0$ that solves the scalar optimization problem

$$\begin{aligned} &Minimize\ \gamma\overline{p} \cdot \overline{x} + \beta\overline{\rho}(\overline{x} \cdot \overline{a}_1, ..., \overline{x} \cdot \overline{a}_m), \\ &such\ that\ I_m^* A\overline{x} \geq 0, \\ &\overline{x} + \overline{h} \geq 0, \\ &\overline{h} \cdot \overline{p} \leq 1, \\ &\overline{h} \geq 0. \end{aligned} \tag{6.15}$$

Conversely, if $\gamma, \beta > 0$, then every solution of (6.15) is also a solution of (6.14). Then the set of solutions of (6.15), over arbitrary non-zero vectors $(\gamma, \beta) \geq 0$, covers the

whole solution set of (6.14), with possibility some more points. Hence, we fix a non-zero vector $(\gamma, \beta) \geq 0$ and focus on problem (6.15).

In order to analyze this problem we follow the same steps as for problem (6.3), although there are minor modifications. Inspired by Lemma 6.3, the following definition and lemma, give sufficient motivation to study problem (6.15).

Definition 6.2. *The market is ρ-strong sequential arbitrage free (ρ-SSA) if and only if the optimal solution of* (6.15) *for $\gamma = \beta = 1$ is equal to zero.*

The following lemma connects $\rho-$SSA to SSA. The proof is simple and so it is omitted.

Lemma 6.4. *If there is a strong sequential arbitrage opportunity in the market then there is also a ρ-strong sequential arbitrage opportunity.*

Another equivalent form of this lemma is that if a market is ρ-strong sequential arbitrage free then it is strong sequential arbitrage free as well. This lemma points out that the credit spreads and TSIR obtained by $\rho-$SSA are more accurate than those obtained by SSA.

It can be shown that $(\overline{x}^*, \overline{h}^*)$ solves (6.15) and $\xi^* = \gamma\overline{p} \cdot \overline{x}^* + \beta\overline{\rho}(\overline{x}^* \cdot \overline{a}_1, ..., \overline{x}^* \cdot \overline{a}_m)$

if and only if $(\xi^*, \overline{x}^*, \overline{h}^*)$ solves the following problem:

$$
\begin{aligned}
&Minimize\ \xi,\\
&such\ that\ \xi \geq \gamma\overline{p}\cdot\overline{x} + \beta\overline{\rho}(\overline{x}\cdot\overline{a}_1, ..., \overline{x}\cdot\overline{a}_m),\\
&\qquad I_m^* A\overline{x} \geq 0,\\
&\qquad \overline{x} + \overline{h} \geq 0,\\
&\qquad \overline{h}\cdot\overline{p} \leq 1,\\
&\qquad \overline{h} \geq 0,\ \xi \leq 0.
\end{aligned}
\tag{6.16}
$$

Notice that $\overline{\rho}(\overline{0}) = 0$, hence $\overline{x} = \overline{0}$ and $\overline{h} = \overline{0}$ is a feasible point for problem (6.15). So the optimal solutions of (6.15) and (6.16) must be smaller than or equal to zero, or in other words, $\xi \leq 0$.

By the representation Theorem 6.1, problem (6.16) is equivalent to:

$$
\begin{aligned}
&Minimize\ \xi,\\
&such\ that\ \xi \geq \gamma\overline{p}\cdot\overline{x} + \beta(\overline{x}\cdot\overline{a}_1, ..., \overline{x}\cdot\overline{a}_m)\cdot\overline{z}, \quad \forall\overline{z} \in \Delta_{\overline{\rho}},\\
&\qquad I_m^* A\overline{x} \geq 0,\\
&\qquad \overline{x} + \overline{h} \geq 0,\\
&\qquad \overline{h}\cdot\overline{p} \leq 1,\\
&\qquad \overline{h} \geq 0,\ \xi \leq 0.
\end{aligned}
$$

After some manipulations, the Lagrangian function

$$\mathcal{L} = \mathcal{L}\left(v, \overline{\mu}, \overline{\lambda}, \lambda\right),$$

$$\mathcal{L} : \mathcal{M}_+(\Delta_{\overline{\rho}}) \times (\mathbb{R}^+ \cup \{0\})^m \times (\mathbb{R}^+ \cup \{0\})^n \times (\mathbb{R}^+ \cup \{0\}) \mapsto \mathbb{R},$$

can be obtained as

$$\begin{aligned}\mathcal{L}\left(v, \overline{\mu}, \overline{\lambda}, \lambda\right) =& \xi(1 - v(\Delta_{\overline{\rho}})) \\ &+ \sum_{i=1}^{n} \left(\gamma v(\Delta_{\overline{\rho}}) p_i - \beta \int_{\overline{\rho}} \overline{b}_i \cdot \overline{z} \, dv(\overline{z}) - \overline{\mu} \cdot R_i - \lambda_i \right) x_i \\ &+ \sum_{i=1}^{n} (-\lambda_i + \lambda p_i) h_i - \lambda,\end{aligned}$$

where $v \in \mathcal{M}_+(\Delta_{\overline{\rho}})$, $\overline{\mu} \in (\mathbb{R}^+ \cup \{0\})^m$, $\overline{\lambda} \in (\mathbb{R}^+ \cup \{0\})^n$, $\lambda \in (\mathbb{R}^+ \cup \{0\})$, and R_i is the i-th column of the matrix defined by

$$R = \begin{pmatrix} \overline{a}_1 \\ \overline{a}_1 + \overline{a}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{a}_1 + ... + \overline{a}_m \end{pmatrix}.$$

It is known that $(v, \overline{\mu}, \overline{\lambda}, \lambda)$ is dual feasible if and only if it belongs to the non-negative cone $\mathcal{M}_+(\Delta_{\overline{\rho}}) \times (\mathbb{R}^+ \cup \{0\})^m \times (\mathbb{R}^+ \cup \{0\})^n \times (\mathbb{R}^+ \cup \{0\})$ and

$$\inf\{\mathcal{L}(v, \overline{\mu}, \overline{\lambda}, \lambda); (\overline{x}, \overline{h}, \xi) \in \mathbb{R}^n \times (\mathbb{R}^+ \cup \{0\})^n \times (\mathbb{R}^- \cup \{0\})\} > -\infty.$$

In this case the value of the infimum is the objective function of the dual problem. Hence the dual problem of (6.15) is

$$\begin{aligned} &Maximize \ -\theta, \\ &such \ that \ \gamma v(\Delta_{\overline{\rho}}) p_i = \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i \cdot \overline{z} \, dv(\overline{z}) + \overline{\mu} \cdot R_i + \lambda_i, \quad i = 1, 2, ..., n, \\ &\qquad \lambda_i \leq \theta p_i, \\ &\qquad \theta \geq 0, \overline{\lambda} \geq 0, \overline{\mu} \geq 0, v \in \mathcal{P}(\Delta_{\overline{\rho}}), \end{aligned} \tag{6.17}$$

where $\mathcal{P}(\Delta_{\overline{\rho}}) = \{v \in \mathcal{M}_+(\Delta_{\overline{\rho}}); v(\Delta_{\overline{\rho}}) \geq 1\}$.

It is easy to check that (6.17) is equivalently represented as

$$
\begin{aligned}
&Maximize\ -\theta,\\
&such\ that\ \gamma v(\Delta_{\overline{\rho}}) p_i = \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i \cdot \overline{z}\ dv(\overline{z}) + \overline{\mu} \cdot \overline{b}_i + \lambda_i, \quad i = 1, 2, ..., n,\\
&\qquad \lambda_i \leq \theta p_i, \quad i = 1, 2, ..., n, \qquad (6.18)\\
&\qquad \theta \geq 0, \overline{\lambda} \geq 0,\\
&\qquad \mu_1 \geq \mu_2 \geq \cdots \geq \mu_m \geq 0, v \in \mathcal{P}(\Delta_{\overline{\rho}}),
\end{aligned}
$$

In optimization problems (6.17) and (6.18), $\overline{b}_i$ is the i-th column of matrix A. In what follows, the equivalent form of Karush-Kuhn-Tucker conditions for problem (6.18) are inserted within brackets.

The Karush-Kuhn-Tucker conditions of (6.17) (respectively of (6.18)) are:

$$
\begin{aligned}
&\gamma v^*(\Delta_{\overline{\rho}})p_i = \lambda_i^* + \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i.\overline{z}\; dv^*(\overline{z}) + \overline{\mu}^* \cdot \overline{R}_i, \quad i = 1, 2, ..., n, \\
&\left(resp. \quad \gamma v^*(\Delta_{\overline{\rho}})p_i = \lambda_i^* + \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i.\overline{z}\; dv^*(\overline{z}) + \overline{\mu}^* \cdot \overline{b}_i, \quad i = 1, 2, ..., n, \right) \\
&\overline{\lambda}^* \leq \theta^*\, \overline{p}, \\
&-v^*(\Delta_{\overline{\rho}})\xi^* + \sum_{i=1}^{n} x_i^* \left(\gamma v^*(\Delta_{\overline{\rho}})p_i - \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i.\overline{z}\; dv^*(\overline{z}) \right) = 0, \quad i = 1, 2, ..., n, \\
&\sum_{i=1}^{n} (\overline{\mu}^* \cdot \overline{R}_i) x_i^* = 0, \\
&\left(resp. \quad \sum_{i=1}^{n} (\overline{\mu}^* \cdot \overline{b}_i) x_i^* = 0, \right) \\
&\overline{\lambda}^* \cdot (\overline{x}^* + \overline{h}^*) = 0, \\
&\theta^* (\overline{h}^* \cdot \overline{p} - 1) = 0, \\
&\overline{h}^* \geq 0,\ \overline{h}^* \cdot \overline{p} \leq 1,\ \overline{x}^* + \overline{h}^* \geq 0,\ \xi^* \leq 0,\ \theta^* \geq 0, \\
&\overline{\lambda}^* \geq 0,\ \overline{\mu}^* \geq 0 \left(resp. \quad \mu_1^* \geq \mu_2^* \geq \cdots \geq \mu_m^* \geq 0 \right),\ v^* \in \mathcal{P}(\Delta_{\overline{\rho}}).
\end{aligned}
\tag{6.19}
$$

It is easily seen that these are equivalent to

$$
\begin{aligned}
&\gamma v^*(\Delta_{\overline{\rho}})p_i = \lambda_i^* + \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i.\overline{z}\; dv^*(\overline{z}) + \overline{\mu}^* \cdot \overline{R}_i, \quad i = 1, 2, ..., n, \\
&\left(resp. \quad \gamma v^*(\Delta_{\overline{\rho}})p_i = \lambda_i^* + \beta \int_{\Delta_{\overline{\rho}}} \overline{b}_i.\overline{z}\; dv^*(\overline{z}) + \overline{\mu}^* \cdot \overline{b}_i, \quad i = 1, 2, ..., n, \right) \\
&\overline{\lambda}^* \leq \theta^*\, \overline{p}, \\
&v^*(\Delta_{\overline{\rho}})\xi^* = \overline{\lambda}^* \cdot \overline{x}^*, \\
&\sum_{i=1}^{n} (\overline{\mu}^* \cdot \overline{R}_i) x_i^* = 0, \\
&\left(resp. \quad \sum_{i=1}^{n} (\overline{\mu}^* \cdot \overline{b}_i) x_i^* = 0, \right) \\
&\overline{\lambda}^* \cdot (\overline{x}^* + \overline{h}^*) = 0, \\
&\theta^*(\overline{h}^* \cdot \overline{p} - 1) = 0, \\
&\overline{h}^* \geq 0,\ \overline{h}^* \cdot \overline{p} \leq 1,\ \overline{x}^* + \overline{h}^* \geq 0,\ \xi^* \leq 0,\ \theta^* \geq 0, \\
&\overline{\lambda}^* \geq 0,\ \overline{\mu}^* \geq 0 \left(resp. \quad \mu_1^* \geq \mu_2^* \geq \cdots \geq \mu_m^* \geq 0 \right),\ v^* \in \mathcal{P}(\Delta_{\overline{\rho}}).
\end{aligned}
\tag{6.20}
$$

Now in view of Lemma 6.2, the dual problem (6.17) and Karush-Kuhn-Tucker

conditions (6.20) can be further simplified. The simplified dual problem is

$$
\begin{aligned}
&Maximize\ -\theta,\\
&such\ that\ \gamma\alpha_0 p_i = \beta\alpha_0\overline{b}_i\cdot\overline{z}+\overline{\mu}\cdot R_i+\lambda_i,\quad i=1,2,...,n\\
&\qquad\left(resp.\quad \gamma\alpha_0 p_i = \beta\alpha_0\overline{b}_i\cdot\overline{z}+\overline{\mu}\cdot\overline{b}_i+\lambda_i,\quad i=1,2,...,n\right)\\
&\qquad\overline{\lambda}\le\theta\overline{p},\\
&\qquad\theta\ge 0,\ \overline{\lambda}\ge 0,\ \overline{\mu}\ge 0,\ \alpha_0\ge 1,\ \overline{z}\in\Delta_{\overline{\rho}},\\
&\qquad\left(resp.\quad \theta\ge 0,\ \overline{\lambda}\ge 0,\ \alpha_0\ge 1,\right.\\
&\qquad\left.\mu_1\ge\mu_2\ge\cdots\ge\mu_m\ge 0,\ \overline{z}\in\Delta_{\overline{\rho}}\right)
\end{aligned}
\tag{6.21}
$$

and the modified Karush-Kuhn-Tucker conditions are

$$
\begin{aligned}
&\gamma\alpha_0^* p_i = \lambda_i^*+\beta\alpha_0^*\overline{b}_i\cdot\overline{z}^*+\overline{\mu}^*\cdot\overline{R}_i,\quad i=1,2,...,n,\\
&\left(resp.\quad \gamma\alpha_0^* p_i = \lambda_i^*+\beta\alpha_0^*\overline{b}_i\cdot\overline{z}^*+\overline{\mu}^*\cdot\overline{b}_i,\quad i=1,2,...,n,\right)\\
&\overline{\lambda}^*\le\theta^*\,\overline{p},\\
&\alpha_0^*\xi^* = \overline{\lambda}^*\cdot\overline{x}^*,\\
&\sum_{i=1}^{n}(\overline{\mu}^*\cdot\overline{R}_i)x_i^*=0,\\
&\left(resp.\quad \sum_{i=1}^{n}(\overline{\mu}^*\cdot\overline{b}_i)x_i^*=0,\right)\\
&\overline{\lambda}^*\cdot(\overline{x}^*+\overline{h}^*)=0,\\
&\theta^*(\overline{h}^*\cdot\overline{p}-1)=0,\\
&\overline{h}^*\ge 0,\ \overline{h}^*\cdot\overline{p}\le 1,\ \overline{x}^*+\overline{h}^*\ge 0,\ \xi^*\le 0,\ \theta^*\ge 0,\alpha_0^*\ge 1\\
&\overline{\lambda}^*\ge 0,\ \overline{\mu}^*\ge 0,\ \left(resp.\quad \mu_1^*\ge\mu_2^*\ge\cdots\ge\mu_m^*\ge 0\right),\ \overline{z}^*\in\Delta_{\overline{\rho}}.
\end{aligned}
\tag{6.22}
$$

By analogy to Theorem 6.2 we have the following result.

Theorem 6.4. *Assume that the primal problem is always finite and that it attains its optimal value, then*

- *The equivalent form of the dual problem is given by* (6.21)*,*
- $(\overline{x}^*, \overline{h}^*, \xi^*)$ *and* $(\theta^*, \overline{\lambda}^*, \overline{\mu}^*, \overline{z}^*)$ *solves problems* (6.16) *and* (6.21)*, respectively if and only if they satisfy the Karush-Kuhn-Tucker conditions* (6.22)*.*

Similarly to (6.3) we also obtain the following result.

Theorem 6.5. *There is no ρ-strong sequential arbitrage if and only if there exists $\mu_1^* \geq \mu_2^* \geq \cdots \geq \mu_m^* \geq 0$ and $\overline{z}^* \in \Delta_{\overline{\rho}}$ such that*

$$p_i = \frac{1}{\gamma}(\beta \overline{z}^* + \overline{\mu}^*) \cdot \overline{b}_i, \quad i = 1, 2, ..., n.$$

By letting $\gamma = 1$ and $\beta = 0$, this gives the same result as Balbás and López (2008).

6.6 Numerical Implementation by Applying Risk Statistics

Numerical Examples 6.1 and 6.2 clearly show that knowledge of the subgradient set $\Delta_{\overline{\rho}}$ is essential in the numerical implementation. Indeed, the probability structure of the model is embedded in this set. The subgradient of some risk measures have already been derived, for example the subgradient of $CVaR_\alpha$ is given in closed form by Rockafellar, Uryasev and Zabarankin (2006). However the structure of the new risk

measure $\overline{\rho}$, that builds on ρ, can be totally different. Therefore it is not surprising that their subgradient sets are also different. For instance, this can be observed in the simple model explained in Section 6.4. While the subgradient of $\rho = CVaR_\alpha$ is given by (6.12), the subgradient set (6.13) is different. Even in this simple example, finding the subgradient set requires some manipulations, not necessarily applicable for real portfolio examples. For instance in these examples, if we have one more period, then the structure of the set is more complicated (in fact we were not able to find a closed form). Therefore we need a tractable and practical approach to involve the subgradient set in the numerical implementation. In what follows, we try to solve this problem by introducing a new class of risk measures called statistical risk measures, see Heyde, Kou and Peng (2007) for more details.

A special feature of the risk measure $\overline{\rho}$ is that its domain is not a random space. One can think that $\overline{\rho}$ is defined on a data set. In other words each vector in $\mathbb{R}^n$ can be interpreted as a set of data. This is the idea behind risk statistics as defined in Heyde, Kou and Peng (2007). Based on this idea they define some new risk measures and representation theorems. Basically we want to use their representation theorems and the structure of the subgradient sets in order to solve our numerical caveat. Here, we review some definitions of risk statistics and as well a representation theorem from the above paper. For proofs based on convex analysis, we refer to Ahmed, Filipovic and Svindland (2008).

Definition 6.3. *The function $\overline{\rho} : \mathbb{R}^n \to \mathbb{R}$ is a natural risk statistic if it satisfies the*

following conditions:

C(1) For all $a \in \mathbb{R}$ *and* $\overline{x} \in \mathbb{R}^n$, $\overline{\rho}(\overline{x} + a\mathbf{1}) = \overline{\rho}(\overline{x}) - a$, *where* $\mathbf{1}$ *is the* n*-dimensional vector* $(1, 1, ..., 1)$. *This property is called translation invariance.*

C(2) For all $t \geq 0$ *and* $\overline{x} \in \mathbb{R}^n$, $\overline{\rho}(t\overline{x}) = t\overline{\rho}(\overline{x})$. *This property is called positive homogeneity.*

C(3) For all vectors $\overline{x}$ *and* $\overline{y}$ *in* $\mathbb{R}^n$, *if* $\overline{x} \leq \overline{y}$ *then* $\overline{\rho}(\overline{x}) \geq \overline{\rho}(\overline{y})$. *This property is called monotonicity.*

C(4) If $(x_i - x_j)(y_i - y_j) \geq 0$ *for* $i \neq j$ *then* $\overline{\rho}(\overline{x} + \overline{y}) \leq \rho(\overline{x}) + \rho(\overline{y})$. *This property is called comonotonic subadditivity.*

C(5) For any permutation $\{i_1, ..., i_n\}$ *of* $\{1, 2, ..., n\}$, *we have* $\overline{\rho}(x_1, ..., x_n) = \overline{\rho}(x_{i_1}, ..., x_{i_n})$. *This property is called permutation invariance.*

The next theorem is the representation theorem of natural risk statistics. In the following theorem and the subsequent ones, the increasing order statistics of any vector $(y_1, ..., y_n)$ is denoted by $(y_{(1)}, ..., y_{(n)})$, with $y_{(n)}$ being the largest.

Theorem 6.6. *If* $\overline{\rho}$ *is a natural risk statistic then*

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z} \in \Delta} \sum_{i=1}^{n} z_i (-x)_{(i)},$$

$$\Delta \subset \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1 \text{ and } \overline{z} \geq 0\}.$$

Definition 6.4. *The function $\overline{\rho} : \mathbb{R}^n \to \mathbb{R}$ is a coherent risk statistic if it satisfies the conditions C(1), C(2), C(3), and subadditivity (not comonotonic).*

Theorem 6.7. *If $\overline{\rho}$ is a coherent risk statistic then*

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z} \in \Delta} -\sum_{i=1}^{n} z_i x_i,$$

$$\Delta \subset \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1 \text{ and } \overline{z} \geq 0\}.$$

Definition 6.5. *The function $\overline{\rho} : \mathbb{R}^n \to \mathbb{R}$ is a law-invariant coherent risk statistic if it satisfies conditions C(1), C(2), C(3), C(5), and subadditivity.*

Theorem 6.8. *If $\overline{\rho}$ is a law-invariant coherent risk statistic then*

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z} \in \Delta} \sum_{i=1}^{n} z_i (-x)_{(i)},$$

$$\Delta \subset \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1, z_1 \leq z_2 \leq ... \leq z_n \text{ and } \overline{z} \geq 0\}.$$

Now assume that $\overline{\rho}$ is any of the three above risk statistics, and take the portfolio in Table 6.2. All the arguments of Sections 6.4 and 6.5 and the optimization problems are still valid. The only difference is that now we assume additional properties for $\overline{\rho}$ as the composition of ρ and Π_T. In the context of risk statistics, we do not assume any probability structure and we work directly with the data. The probability structure is embedded in the prices. Therefore interpreting the final result is difficult from a probabilistic point of view, and this could be an objection to this method. In applying the above risk statistics, we use the most trivial ones. It means that we let the subgradient

Δ be the maximal set. This will make the optimal value lower than what it really is. Hence if we ever get a non-zero solution (arbitrage income), it will be a lower bound, and the actual value might be larger.

After running the maximization problem by the above three risk statistics with the maximal Δ set, the optimal value (the lower bound) is approximately zero which is a trivial bound.

Now let us update the portfolio in Table 6.2 to the following. Assuming that this

	Bond 1	Bond 2
Price of the bonds	1010	908.9
Cash flow 1	10	1
Cash flow 2	101000	909

Table 6.3: Obvious ρ-arbitrage, two bonds portfolio.

portfolio is homogeneous, due to the huge last payment of Bond 1, this is clearly an arbitrage portfolio. In this case the optimal value (or arbitrage income), under the assumptions of Example 6.2 is approximately 0.9989874. But if we use any of the above risk statistic measures with the maximal Δ set, the optimal value of the maximization problem still is zero. However, we can not yet conclude that these risk measures can not detect arbitrage in this obvious portfolio. Because one might take a non-maximal Δ set in each of the above risk measures and therefore a new risk statistic to obtain a non-zero lower bound, but this is not the only problem. In fact as we see shortly, the permutation invariance assumption of the natural risk statistic and the law-invariant coherent risk statistic, is not consistent with our model. But a non-zero optimal value

is feasible for coherent risk measures which is consistent with our model assumptions.

The main problem with the natural risk statistic and the law-invariant coherent risk statistic is that they do not distinguish strictly the weights given to the cash flows at different times. This is reflected in Definitions 6.3 and 6.5 through property C(5) (Definition 6.4 is silent about it). For example by C(5), $\overline{\rho}(1,0) = \overline{\rho}(0,1)$. Now let us explain this. By definition $\overline{\rho}(1,0)$ is equal to $\rho(\Pi_T(1,0))$, that represents the risk associated to the future wealth of a portfolio that pays one unit of currency at time $t = t_0$ and nothing at time $t = t_1$. And of course due to the uncertainty on factors like random interest rates, this risk is different from $\overline{\rho}(0,1) = \rho(\Pi_T(0,1))$ with a similar interpretation. In Heyde, Kou and Peng (2007), they take data statically (i.e. at a fixed time), here we take data at different times.

Coherent risk statistics do not meet axiom C(5). Hence one can look for a suitable Δ set and so a new risk measure that provides a non-zero lower bound. Fortunately through numerical trials, we found out that there is such a risk measure with the following representation,

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z}\in\Delta} \sum_{i=1}^{n} -z_i x_i,$$

$$\Delta_{\overline{\rho}} = \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1, z_1 \geq z_2 \geq ... \geq z_n \text{ and } \overline{z} \geq 0\}.$$

But how can we interpret this risk measure? Answering this question leads to a new representation theorem.

We introduce a new axiom to reflect these different weights assigned at different

times. Instead of C(5), we use the following,

C(6): For all $j > i$, $\overline{\rho}(e_i - e_j) \leq 0$, where e_i and e_j are the unit vectors with the i-th and j-th coordinates of 1 and the other coordinates equal to zero.

The following argument is to motivate this assumption. For instance take the very simple case of a two dimensional space and only two periods t_0 and t_1. Then $e_1 - e_2 = (1, -1)$ and $\Pi_{t_1}(e_1 - e_2) = \Pi_{t_1}(1, -1)$, where t_1 is the maturity time. In a fixed income market like ours, because the accumulation amount of one unit at time t_0 can compensate a claim of -1 at time t_1, we have $\Pi_{t_1}(1, -1) \geq 0$ or $\Pi_{t_1}(e_1 - e_2) \geq 0$. However more caution is required here. If we want to take into account all the details, then this assumption is true in general if there is no risk of default for the issuer of the bonds. Remember that we only manipulate the bonds and no net value is invested. Therefore if it is a matter of arbitrage income, it makes more sense to focus on the companies with a low probability of default or at least consider homogeneous portfolios, i.e. bonds with the same credit ratings.

Axiom C(6) can be also a probabilistic assumption. For instance here we have assumed that this inequality is true almost surely and the probabilistic structure of the model is embedded in this assumption. Once we study the credit measurement, we will revisit this assumption again. Finally by assuming the monotonicity of the risk measure ρ we get $\rho(\Pi_{t_1}(e_1 - e_2)) \leq 0$ or $\overline{\rho}(e_1 - e_2) \leq 0$. One conclusion of this property is that $\overline{\rho}(e_1) \leq \overline{\rho}(e_2)$.

Remark 6.2. *Note that due to market conditions and prior assumptions, other axioms*

may also be considered. The Axiom C(6) is just one possible option and in fact the simplest one.

Definition 6.6. *The function $\overline{\rho} : \mathbb{R}^n \to \mathbb{R}$ is the risk statistic DF (default free) if it satisfies the axioms C(1), C(2), C(3), C(6) and subadditivity.*

Theorem 6.9. *If $\overline{\rho}$ is the risk statistic DF then*

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z} \in \Delta} \sum_{i=1}^{n} -z_i x_i,$$

$$\Delta \subset \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1, z_1 \geq z_2 \geq ... \geq z_n \text{ and } \overline{z} \geq 0\}.$$

This can be proved by simple adjustments of Ahmed, Filipovic and Svindland (2008). In Heyde, Kou and Peng (2007) the proofs are simpler but long, while in Ahmed, Filipovic and Svindland (2008) the proofs are shorter and more technical, based on convex analysis.

In our numerical example we use the following special version of the above risk statistic

$$\overline{\rho}(\overline{x}) = \sup_{\overline{z} \in \Delta} \sum_{i=1}^{n} -z_i x_i,$$

$$\Delta = \{\overline{z} \in \mathbb{R}^n; \sum_{i=1}^{n} z_i = 1, z_1 \geq z_2 \geq ... \geq z_n \text{ and } \overline{z} \geq 0\}.$$

By taking the maximal set, all the optimal values should be considered as a lower bonds.

Example 6.3. *Now let us review the previous examples by this new risk measure and also we present two real data examples:*

- *With this new risk statistic the lower bound for the optimal problem of the portfolio represented by Table 6.2 is approximately* 0.00861 *compared to 0.00869 in Example 6.2.*

- *In case of the portfolio represented by Table 6.3 the lower bound for the optimal problem is 0.888876.*

So far we have only seen academic examples. In what follows, we consider the data of two real portfolios, composed with the five following bonds taken from the Yahoo Finance: `http://screen.yahoo.com/bonds.html`. There is nothing special about these bonds, except that they have the same credit ratings. Apart from this characteristic they were simply chosen randomly. The identifications of these bonds are given in Tables 6.4 and 6.5.

Example 6.4. *For the portfolio represented by Table 6.6, the optimal value is approximately equal to* $\theta^* \approx 0.03857$.

- *The solution of the primal problem is equal to*

  $x_1^* \approx 0$, $x_2^* \approx -0.50503$, $x_3^* \approx -0.50018$, $x_4^* \approx 0.00853$.

- *The solution of the dual problem is equal to*

  $\theta^* \approx 0.03857$, $\lambda_1^* \approx 3.80981$, $\lambda_2^* \approx 3.85572$, $\lambda_3^* \approx 3.81869$, $\lambda_4^* \approx 0$, $\alpha^* \approx 30.79152$, $\overline{z}_1^* \approx 0.02711$, $\overline{z}_2^* \approx 0.02711$,....

The value $x_1^* \approx 0$ *means that Bond 1 can be excluded in our analysis. In order to*

COLGATE PALMOLIVE CO MTNS BE

As of 7-Dec-2010

OVERVIEW

Price:	98.77
Coupon(%):	1.375
Maturity Date:	1-Nov-2015
Yield to Maturity(%):	1.639
Current Yield(%):	1.392
Fitch Rating:	AA
Coupon Payment Frequency:	Semi-Annual
First Coupon Date:	1-May-2011
Type:	Corporate
Callable:	No

OFFERING INFORMATION

Quantity Available:	1240
Minimum Trade Qty:	1
Dated Date:	3-Nov-2010
Settlement Date:	13-Dec-2010

Table 6.4: Corporate Bond 1

obtain the optimal arbitrage income, the negative values of the primal problem should be interpreted as shortening the bond and the positive value are going for long.

An interpretation of the solution of dual problem is done in the next section. In a short review, the value $\alpha^* z_j^*$ *is a lower bound for the real market discount factor in period* j*. For instance here* $\alpha^* z_1^* \approx 0.83$ *is a lower bound for the real market discount factor in period 1. The quantities* λ_j *can be interpreted as the credit spreads. For more details see the next section.*

Example 6.5. *For the portfolio represented by Table 6.7, the optimal value is approximately equal to* $\theta^* \approx 0.04432$*.*

- *The solution of the primal problem is equal to*

JPMORGAN CHASE & CO

As of 7-Dec-2010

OVERVIEW

Price:	99.96
Coupon(%):	2.600
Maturity Date:	15-Jan-2016
Yield to Maturity(%):	2.607
Current Yield(%):	2.601
Fitch Rating:	AA
Coupon Payment Frequency:	Semi-Annual
First Coupon Date:	15-Jul-2011
Type:	Corporate
Callable:	No

OFFERING INFORMATION

Quantity Available:	300
Minimum Trade Qty:	1
Dated Date:	18-Nov-2010
Settlement Date:	13-Dec-2010

Bond 2

COCA COLA CO

As of 7-Dec-2010

OVERVIEW

Price:	99.00
Coupon(%):	1.500
Maturity Date:	15-Nov-2015
Yield to Maturity(%):	1.714
Current Yield(%):	1.515
Fitch Rating:	AA
Coupon Payment Frequency:	Semi-Annual
First Coupon Date:	15-May-2011
Type:	Corporate
Callable:	No

OFFERING INFORMATION

Quantity Available:	175
Minimum Trade Qty:	1
Dated Date:	15-Nov-2010
Settlement Date:	13-Dec-2010

Bond 3

CHICAGO ILL GO BDS

As of 7-Dec-2010

OVERVIEW

State:	Illinois
Price:	112.70
Coupon(%):	5.000
Maturity Date:	1-Jan-2014
Yield to Maturity(%):	0.780
Current Yield(%):	4.437
Fitch Rating:	AA
Coupon Payment Frequency:	Semi-Annual
First Coupon Date:	1-Jan-2008
Callable:	No

BOND PROFILE

Type:	Municipal
Insured:	Yes
Alternative Minimum Tax:	No

Bond 4

CHICAGO ILL TAXABLE GO BONDS

As of 7-Dec-2010

OVERVIEW

State:	Illinois
Price:	114.35
Coupon(%):	5.400
Maturity Date:	1-Jan-2014
Yield to Maturity(%):	0.644
Current Yield(%):	4.722
Fitch Rating:	AA
Coupon Payment Frequency:	Semi-Annual
First Coupon Date:	1-Jan-2005
Callable:	No

BOND PROFILE

Type:	Municipal
Insured:	Yes
Alternative Minimum Tax:	No

Bond 5

Table 6.5: Corporate Bonds 2 and 3; Municipal Bonds 4 and 5

Dates	Bond 1	Bond 2	Bond 3	Bond 4
7-Dec-2010	98.77	99.96	99.00	112.70
1-Jan-2011	0	0	0	5
1-May-2011	1.375	0	0	0
15-May-2011	0	0	1.5	0
1-July-2011	0	0	0	5
15-July-2011	0	2.6	0	0
1-Nov-2011	1.375	0	0	0
15-Nov-2011	0	0	1.5	0
1-Jan-2012	0	0	0	5
15-Jan-2012	0	2.6	0	0
1-May-2012	1.375	0	0	0
15-May-2012	0	0	1.5	0
1-July-2012	0	0	0	5
15-July-2012	0	2.6	0	0
1-Nov-2012	1.375	0	0	0
15-Nov-2012	0	0	1.5	0
1-Jan-2013	0	0	0	5
15-Jan-2013	0	2.6	0	0
1-May-2013	1.375	0	0	0
15-May-2013	0	0	1.5	0
1-July-2013	0	0	0	5
15-July-2013	0	2.6	0	0
1-Nov-2013	1.375	0	0	0
15-Nov-2013	0	0	1.5	0
1-Jan-2014	0	0	0	5+100
15-Jan-2014	0	2.6	0	0
1-May-2014	1.375	0	0	0
15-May-2014	0	0	1.5	0
15-July-2014	0	2.6	0	0
1-Nov-2014	1.375	0	0	0
15-Nov-2014	0	0	1.5	0
15-Jan-2015	0	2.6	0	0
1-May-2015	1.375	0	0	0
15-May-2015	0	0	1.5	0
15-July-2015	0	2.6	0	0
1-Nov-2015	100+1.375	0	0	0
15-Nov-2015	0	0	1.5+100	0
15-Jan-2016	0	2.6+100	0	0

Table 6.6: Future cash flows of the four bonds portfolio.

Dates	Bond 1	Bond 2	Bond 3	Bond 4	Bond 5
7-Dec-2010	98.77	99.96	99.00	112.70	114.35
1-Jan-2011	0	0	0	5	5.4
1-May-2011	1.375	0	0	0	0
15-May-2011	0	0	1.5	0	0
1-July-2011	0	0	0	5	5.4
15-July-2011	0	2.6	0	0	0
1-Nov-2011	1.375	0	0	0	0
15-Nov-2011	0	0	1.5	0	0
1-Jan-2012	0	0	0	5	5.4
15-Jan-2012	0	2.6	0	0	0
1-May-2012	1.375	0	0	0	0
15-May-2012	0	0	1.5	0	0
1-July-2012	0	0	0	5	5.4
15-July-2012	0	2.6	0	0	0
1-Nov-2012	1.375	0	0	0	0
15-Nov-2012	0	0	1.5	0	0
1-Jan-2013	0	0	0	5	5.4
15-Jan-2013	0	2.6	0	0	0
1-May-2013	1.375	0	0	0	0
15-May-2013	0	0	1.5	0	0
1-July-2013	0	0	0	5	5.4
15-July-2013	0	2.6	0	0	0
1-Nov-2013	1.375	0	0	0	0
15-Nov-2013	0	0	1.5	0	0
1-Jan-2014	0	0	0	5+100	5.4+100
15-Jan-2014	0	2.6	0	0	0
1-May-2014	1.375	0	0	0	0
15-May-2014	0	0	1.5	0	0
15-July-2014	0	2.6	0	0	0
1-Nov-2014	1.375	0	0	0	0
15-Nov-2014	0	0	1.5	0	0
15-Jan-2015	0	2.6	0	0	0
1-May-2015	1.375	0	0	0	0
15-May-2015	0	0	1.5	0	0
15-July-2015	0	2.6	0	0	0
1-Nov-2015	100+1.375	0	0	0	0
15-Nov-2015	0	0	1.5+100	0	0
15-Jan-2016	0	2.6+100	0	0	0

Table 6.7: Future cash flows of the five bonds portfolio.

$x_1^* \approx 0$, $x_2^* \approx 0$, $x_3^* \approx -0.10101$, $x_4^* \approx 0.00836$, $x_5^* \approx 0$.

- *The solution of the dual problem is equal to*

  $\theta^* \approx 0.04432$, $\lambda_1^* \approx 4.37731$, $\lambda_2^* \approx 4.43005$, $\lambda_3^* \approx 4.38750$, $\lambda_4^* \approx 0$, $\lambda_5^* \approx 0.67351$,

  $\alpha^* \approx 30.60751$, $\overline{z}_1^* \approx 0.02711$, $\overline{z}_2^* \approx 0.02711$,....

The same interpretation for the solutions as the previous example can be done here.

6.7 Credit Risk Measurement and ρ-arbitrage

In this section we discuss the main purpose of this chapter, measuring the credit risk of corporate bonds. In fact we accomplish this by obtaining some kind of credit spread. A credit spread is basically the extra value of bonds due to their level of risk that makes them attractive to investors. In this section we assume that the market is free from all types of arbitrage, either ρ-arbitrage or the classical one, and use this assumption to estimate credit spreads.

To start we define the following new theoretical prices

$$p_j^* = \alpha^* \overline{z}^* . \overline{b}_j, \quad j = 1, 2, ..., n,$$

for the n bonds, where α^* and z^* are obtained from the solution of the dual problem (6.10). Notice that $\alpha^* z_j^*$ is the discount factor for the period j. Here $\overline{b}_j$ is the j-th column of matrix A. By Theorem 6.3, the theoretical prices produce no ρ-arbitrage and the optimal solution of the dual problem (6.10) is zero, if the bonds are valued at these theoretical prices.

In an ideal market the theoretical prices must be the same as the real ones because a good market must be consistent and it should not provide any ρ-arbitrage opportunities to the agents. Therefore assuming that the market is free from arbitrage, any difference between the real prices and the theoretical ones should be due to the risk of default associated with the bonds. Although there are other types of risks involved in a market, for simplicity we only focus on the major one which is the credit risk.

We formally define the credit spread of the Bond j to be $(p_j - p_j^*)$. Notice that here we used the Theorem 6.3 to define these credit spreads. The same procedure can be carried out by Theorem 6.5. Indeed, applying the latter should provide more accurate credit spreads than with Theorem 6.3 because it also counts the inconsistencies arising due to the existence of strong sequential arbitrage opportunities in the market.

The procedure is explained through some examples. We discuss the credit spread associated with those bonds considered in the previous section. First, we start by the portfolios represented by Tables 6.1 and 6.2 under the assumptions of Section 6.4.

Example 6.6. *As we saw in Example 6.1, in the portfolio represented by Table 6.1, there is essentially no ρ-arbitrage. The theoretical prices are approximately equal to the original ones and credit spreads can be ignored. On the other hand, the portfolio of Table 6.2 is different. In this case the spread of the first bond is 10 and the spread of the second bond is equal to 16.92069. Hence in this portfolio the first bond is more credible than the second one as it has a smaller spread. However, this is a theoretical example with abstract assumptions and there is no concrete interpretation for the risk*

of default.

Now we turn to the real data portfolios and measure the credit risk of those bonds. Note that as one can check in Tables 6.4 and 6.5, all these bonds have an AA rating. However, we are able to distinguish the credit worthiness of these homogeneous bonds.

To implement the numerical procedure we use the risk statistic DF explained in the last section. The conditions C(1), C(2), and C(3) have traditional intuitive interpretations. Condition C(6) was validated in the previous section under a default-free market assumption. The general method to implement the numerical procedure is as follows.

For a moment assume that the bond market is default free and ρ-arbitrage free, then as detailed in Section 6.6, by applying the risk statistic DF in Definition 6.6, the optimal solution of the dual problem (6.10) for this risk statistic $\overline{\rho}$ must be zero and then the theoretical prices are equal to the market prices. However, if these theoretical prices are not same as the market prices (which is normally the case), then the market is either non-default free or non-ρ-arbitrage free. At the beginning of this section we assumed that the market is free from all types of arbitrage, so any difference between theoretical and market prices must be due to the risk of default. Notice that not all risk measures are applicable due to the initial default-free assumption. However the risk statistic DF was specifically constructed to satisfy our default-free assumption.

By Karush-Kuhn-Tucker conditions (6.11), we have that $p_j = \lambda_j^* + p_j^*$, for $j = 1, 2, ..., n$. Since $\lambda_j^* \geq 0$, the theoretical prices are the lower bounds of the market

prices. Hence $(p_j - p_j^*)$ is non-negative and it is also a lower bound for the credit spreads. Other details are explained in the following examples.

Example 6.7. *First take the bonds 1 to 3 of the portfolio represented by Table 6.6. In this case by running the optimization problem* (6.10)*, the optimal solution is zero. Hence the theoretical prices are exactly the same as the market prices. However, this does not mean that credit spreads are actually zero. What we can conclude is that the lower bounds on the credit spreads of these bonds are zero, a trivial conclusion! Since these are corporate bonds with AA ratings, there must be some credit spreads, even if small.*

This phenomena happens because here, we do not focus on the whole market, but rather on a small sector made of only three bonds. To get a better estimation, we should add more data. Now let us consider the portfolio represented by Table 6.6. The following table summarizes the results. Notice that here the lower bounds of the spreads

	p_j	p_j^*	$p_j - p_j^*$	Revised Rating
j=1	98.77	94.96	3.81	AA2
j=2	99.96	96.10	3.86	AA4
j=3	99	95.18	3.82	AA3
j=4	112.70	112.70	0	AA1

Table 6.8: Credit spreads for the portfolio in Table 6.7

of the first three bonds are not zero any more. These lower bounds are small, as the bonds have good ratings, and their ratings are close to each other because all of them are in the same rating sector. At first, the zero lower bound spread of the fourth bond seems to be mysterious, but if one takes a look at the identification of the bond, i.e.

Table 6.5, it comes out that this is a municipal bond which is insured.

Traditionally, municipal bonds have very low risk of default. To compare the historical default rates of the corporate and municipal bonds, we refer to the following link, `http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_reports&docid=f:hr835.110`. *The data are from Moody's and Standard & Poor's.*

This study is done with a 2007 evaluation data and data for the years between 1970 and 2006. For instance, based on this study by Moody's, the historical default rates of municipal bonds with AA ratings is 0.06% compare to 0.52% for corporate bonds with the same rating. By Standard & Poor's, the historical default rates of municipal bonds with AA ratings is 0.00% compare to 1.50% of corporate bonds with the same rating.

Although the municipal bonds have showed lower rates of default (especially for the ones with good ratings), after the 2009 mortgage crisis, a zero credit spread is not realistic even for an AA municipal bond.

Again one should pay attention that these spread estimations are just based on the information taken from a small part of the whole market. Therefore to have more reliable lower bounds on spreads, more data are needed. The next example shows the analysis for the five bonds portfolio.

Based on these spreads, we have revised the bond ratings. This is showed in the last column of Table 6.8. For instance AA1 is the highest rating in this table.

Example 6.8. *In this example we consider the portfolio represented by Table 6.7.*

Here, one can see the effect of adding a new bond. The result is summarized in the following table.

	p_j	p_j^*	$p_j - p_j^*$	Revised Rating
j=1	98.77	94.96	4.38	AA3
j=2	99.96	96.10	4.43	AA5
j=3	99	95.18	4.39	AA4
j=4	112.70	112.03	0.67	AA2
j=5	114.35	114.35	0	AA1

Table 6.9: Credit spreads for the five bonds portfolio

Notice how this time the fourth bond does not have a zero lower bound credit spread any more and also how the lower bound on spreads of the other bonds are updated and are wider now. The fifth bond is also a municipal bond and a zero lower bound credit spread can be interpreted the same way as the fourth bound. Interestingly, the rating orders of Example 6.7 hold here as well.

Notice that any feasible point of the optimization problem (6.10) for $(n+1)$ bonds is also a feasible point for the optimization problem (6.10) with n bonds. This explains why the lower bounds are getting wider from three to five bonds portfolios in the above examples.

In general assume that we want to estimate the credit spread of a corporate bond with the current market price p. We start constructing portfolios by adding bonds from the market. Assume that CS_n is the credit spread of this bond in a portfolio consisting of n bonds. Note that to build a portfolio of n bonds we keep the last $(n-1)$ bonds and add a new bond from the market. As it was explained earlier, the sequence $\{CS_n\}_{n=2}^{\infty}$

is an increasing sequence that is bounded from below by zero and from above by the price of the bond. Each element of this sequence is a lower bound for the true credit spread. Therefore theoretically this sequence is converging to its real credit spread. This is a practical way of estimating the credit spread of corporation bonds.

The same kind of table can be obtained for any group of four bonds of the portfolio represented by Table 6.7. Finally these examples were just for illustrative purposes. To get reliable estimates, one should take a large pool of bonds. The more data we have, the more accurate and reliable the estimation of the credit spreads and ratings.

The above rating system is based on the bond prices taken from the market on December 7, 2010. These are therefore daily ratings for this specific day. These ratings can be updated with the new prices taken either hourly, daily or based on any other periodic time intervals.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis credit risk in two area of stochastic processes and risk measures are studied. In the first part, the main focus is the study of credit risk under jump processes. Due to the path dependent property of claims, most of the current and previous literature does not apply when the underlying process has jumps. Therefore our aim in this first approach is to initiate a method for studying credit risk under jump processes. A locally risk minimization approach is used for the corresponding risk management problem.

As the application of risk measures in finance becomes increasingly popular, the goal of the second part of this thesis is to use risk measures to gauge the credit quality of financial products. Here the concentration is more on a practical approach and the theory is developed for defaultable bond markets. Some of the conclusions and results of this thesis are listed below:

- Two credit risk models (structural models and reduced form models) and their

relations are explained. Then the advantages and disadvantages of each one are studied. The main advantages and disadvantages of structural models are respectively their intuitive default model, and non-realistic results, such as zero short spreads. On the other hand reduced form models give out non-zero short spreads, but they cannot explain the default event in the model.

- Two types of intensities and their properties are rigorously studied. Also some well known pricing rules based on the idea of intensity are discussed.

- The importance of information in credit risk modeling is highlighted. Information based models for three different filtration expansions are studied. The relationships of the intensity with these filtration expansions are explored. A few results that are already obtained under progressive filtration expansions are generalized and improved under a more general version of the filtration expansions. This includes the minimal and progressive filtration expansions, as special cases.

- The interpretation of two types of intensities as short credit spreads are discussed. The results in this area are particularly improved for more realistic and general filtration expansions, than for progressive filtration expansions. In this context the predictability problems with the intensity that arise for discontinuous intensity processes are fixed as well.

- The Laplacian approximation method to calculate the intensity and its drawbacks are fully discussed. This approach is implemented for a special jump process.

The structure of the intensity for a Brownian motion with drift perturbed by a compound Poisson process is explained, and it turns out that under jump process the short spreads are non-zero even in the presence of full information.

- To obtain the hedging strategies, an auxiliary theorem is proved that can also be of interest in martingale theory.

- Hedging of defaultable claim is obtained under finite variation Lévy processes for which none of the known methods work. The approach is again locally risk minimization.

- A necessary and sufficient condition for the existence of a perfectly hedgeable (risk-free) defaultable claim is provided.

- The estimation of the distribution of the default time and pricing rules are studied.

- Credit risk is studied under risk measures. A new type of indicator to detect inconsistencies in bond market is defined and well measured. This indicator is linked to a new type of arbitrage that is called ρ-arbitrage. A necessary and sufficient condition for the existence of this type of arbitrage is provided.

- Finally by introducing a new type of risk statistic, a practical approach to obtain credit spread of defaultable bonds is obtained.

7.2 Future Work

Some interesting questions for future work are listed below:

- Theoretically, as we saw, the intensity process is actually the credit spread. It will be interesting to fit the closed form of the intensity obtained in this thesis to real data and see how well it performs.

- The predictable part of the default indicator process (and in fact the intensity process) was obtained in Corollary 4.1 for a compound Poisson process. It is interesting to see the structure of this predictable part for a jump-diffusion or a general Lévy process. The existence is guaranteed by Doob-Meyer's decomposition. Especially, what is the relation of this predictable part with Proposition 4.1, and more specifically with equation (4.4)?

- The model presented in Chapter 5 can be improved and extended in many ways. Some of these are: considering a non-zero interest rate, working under more general Lévy processes, considering a non-zero level of default, extending the hedging to multivariate payoff functions, improving the integrability conditions of the theorems and propositions, redoing the same procedure but with an exponential Lévy process, considering multiple period payments instead of a defaultable zero-coupon bonds, or extending the theory to insurance products.

- Further and more complicated numerical procedures will be investigated to im-

plement the theory of Chapter 5 with real data. Especially in this area more sophisticated tools than simulations are needed to solve the involved PIDEs for more complex finite variation Lévy processes.

- The procedure in Chapter 6 can also be improved. For example we obtained lower bounds for real market discount factors. By changing the constraints, one can obtain upper bounds as well. Therefore we will obtain intervals for credit spreads instead of lower bounds.

- In Example 6.1, a two periods interest rate model was used. The structure of interest rates can be extended to more complicated models, for instance the Vasicek model. If the underlying risk measure is $CVaR_\alpha$, it is possible to obtain closed forms. This can be an interesting project because it is a combination of the theory of risk measure and that of stochastic processes.

- In order to obtain finer envelopes for the interest rate structure and also credit spreads, one can implement the numerical methods by using our revised problem in Section 6.5.

- Better risk statistic measures with more advanced probabilistic back ground can be applied to implement with real data.

- Finally, one can try a large pool of defaultable bonds and actually observe the convergence of the credit spreads explained in Section 6.6.

Appendix A

Definitions and Technical Results

A.1 Definitions

Definition A.1. *A random time is a non-negative $\mathcal{F}$-measurable random variable.*

Definition A.2. *A stopping time $\mathcal{T}$ is predictable if there exists an increasing sequence of stopping times $\mathcal{T}_n$, $n \geq 1$ such that for all n, $T_n < T$ on $\{T > 0\}$ and $\lim_{n\to\infty} T_n = T$, almost surely. This is called an announcing sequence.*

Definition A.3. *A stopping time $\mathcal{T}$ is totally inaccessible if for every predictable stopping time $\mathcal{S}$, $\mathbb{P}\{\mathcal{T} = \mathcal{S} < \infty\} = 0$.*

Definition A.4. *If φ is a class of processes, the localized class is denoted by φ_{loc} and it is defined as follows: a process X belongs to φ_{loc} if and only if there exists an increasing sequence $\{T_n\}_{n\geq 1}$ of stopping times (depending on X) such that $\lim_{n\to\infty} T_n = \infty$, almost surely. and that each stopped process X^{T_n} belongs to φ. The sequence $\{T_n\}_{n\geq 1}$ is called a localizing sequence for X (relative to φ).*

Definition A.5. *The process X is called a potential if it is a càdlàg, positive supermartingale with* $\lim_{t\to\infty} \mathbb{E}[X_t] = 0$.

Definition A.6. *Let a be an extended real number including infinity, $0 \leq a \leq \infty$ and X be a right continuous supermartingale, uniformly integrable on the interval $[0, a]$. X belongs to class D on this interval, if the family $\{X_{\mathfrak{T}} : \mathfrak{T} \in \mathcal{P}_a\}$ is uniformly integrable, where $\mathcal{P}_a$ is the set of all stopping times bounded by a.*

If X belongs to the class D on every interval $[0, a]$, for $0 < a < \infty$, it is said to locally belong to class D.

The above definition of Class D is the same as Meyer (1962). In the current literature, locally belonging to the class D is called class DL, see Karatzas and Shreve (1988).

Definition A.7. *The set of all local martingales null at zero is denoted by $\mathscr{L}$.*

Definition A.8. *Here we define two important classes of random processes:*

- *A semimartingale is a process X of the form $X = X_0 + M + \Lambda$, where X_0 is finite-valued and $\mathfrak{F}_0$-measurable, $M \in \mathscr{L}$ and $A \in \mathscr{V}$, where $\mathscr{V}$ is the class of finite variation processes. The class of all semimartingales is denoted by $\mathscr{S}$.*

- *A special semimartingale is a semimartingale which admits a decomposition $X = X_0 + M + \Lambda$ as above, with a process Λ that is predictable. The class of special semimartingales is denoted by $\mathscr{S}_p$.*

Definition A.9. *For a stopping time $\mathcal{T}$, the σ-algebra $\mathfrak{F}_{\mathcal{T}^-}$ is the smallest σ-algebra containing $\mathfrak{F}_0$ and all sets of the form $A \cap \{t < \mathcal{T}\}$, $t > 0$ and $A \in \mathfrak{F}_t$.*

A.2 Technical Results

Theorem A.1. *Assume that $\mathcal{T}$ is a random time in the reference filtration $\mathfrak{G}$ and $\Gamma_t = \mathbb{P}(\mathcal{T} > t|\mathfrak{G}_t) > 0$, almost surely for all $t > 0$, and $\mathbb{E}(\Gamma_t) > 0$. By Doob-Meyer's decomposition, there is a unique nondecreasing $\mathfrak{G}$-predictable process K such that the process $\Gamma + K$ is a $\mathfrak{G}$-martingale. Suppose that $\mathfrak{F}$ is the progressive filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$. Then Jeulin-Yor's results states that $1_{\{\mathcal{T} \leq t\}} - \int_0^{t \wedge \mathcal{T}} \frac{1}{\Gamma_{s^-}} dK_s$ is an $\mathfrak{F}$-martingale.*

Remark A.1. *The original form of this theorem in Jeulin and Yor (1978) considers more than one filtration expansion that includes the progressive one.*

Remark A.2. *In the above theorem, the existence of the compensator is already guaranteed by Doob-Meyer's decomposition.*

Lemma A.1. *Assume that X is an integrable $\mathcal{F}$-measurable random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that on this probability space, $\mathcal{F}_2 \subset \mathcal{F}$ is any σ-algebra expansion of $\mathcal{F}_1 \subset \mathcal{F}$ and a random time $\mathcal{T}$ such that the following hold:*

- *$\mathcal{F}_1 \subset \mathcal{F}_2$ and $\mathcal{T}$ is $\mathcal{F}_2$ measurable,*
- *$\mathcal{F}_2 \subset \{A \in \mathcal{F}$; there is $B \in \mathcal{F}_1, A \cap \{\mathcal{T} > t\} = B \cap \{\mathcal{T} > t\}\}$.*

Then for all $t \geq 0$, we have

$$\mathbb{E}\big[1_{\{\mathcal{T}>t\}}X \mid \mathfrak{F}_2\big]\mathbb{P}(\mathcal{T} > t|\mathfrak{F}_1) = 1_{\{\mathcal{T}>t\}}\mathbb{E}\big[1_{\{\mathcal{T}>t\}}X \mid \mathfrak{F}_1\big].$$

By a simple application of the previous lemma, we have the following corollary.

Corollary A.1. *Assume that X is an integrable $\mathcal{F}$-measurable random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that on this probability space, $\mathfrak{F} = (\mathfrak{F})_{t\geq 0}$ is any filtration expansion of $\mathfrak{G} = (\mathfrak{G})_{t\geq 0}$ and a random time $\mathcal{T}$ such that:*

- *$\mathcal{T}$ is an $\mathfrak{F}$-stopping time,*
- *$\mathfrak{F}$ is a subset of the progressive filtration expansion of $\mathfrak{G}$ and $\mathcal{T}$, i.e. for all $t \geq 0$,*

$$\mathfrak{F}_t \subset \big\{A \in \mathfrak{G}_\infty \vee \sigma(\mathcal{T}); \text{there is } B \in \mathcal{F}_t, A \cap \{\mathcal{T} > t\} = B \cap \{\mathcal{T} > t\}\big\}$$

- *for all $t \geq 0$, $\mathbb{P}(\mathcal{T} > t \mid \mathfrak{G}_t)$ is almost surely non-zero.*

Then for all $t \geq 0$ we have

$$\mathbb{E}\big[1_{\{\mathcal{T}>t\}}X \mid \mathfrak{F}_t\big] = 1_{\{\mathcal{T}>t\}}\frac{\mathbb{E}\big[1_{\{\mathcal{T}>t\}}X \mid \mathfrak{G}_t\big]}{\mathbb{P}(\mathcal{T} > t \mid \mathfrak{G}_t)}.$$

Lemma A.2. *Let X and Y be bounded random variables defined on probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, while $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are sub σ-fields of $\mathfrak{F}$. If X is independent of Y and $\mathfrak{G}_2$, while Y is independent of $\mathfrak{G}_1$, then $\mathbb{E}[XY \mid \mathfrak{G}_1 \bigvee \mathfrak{G}_2] = \mathbb{E}[X \mid \mathfrak{G}_1]\mathbb{E}[Y \mid \mathfrak{G}_2]$.*

This is a classical result and can be found in most advanced probability books.

Lemma A.3. *Let* $(\Omega, \mathfrak{F}, \mathbb{P})$ *be a probability space with* $\Lambda \in \mathfrak{F}$ *and* ξ *an integrable random variable. Then, if* $\mathfrak{G}$ *is a sub* σ*-field of* $\mathfrak{F}$ *such that* $\mathfrak{G} \cap \Lambda = \mathfrak{F} \cap \Lambda$*, we have that* $\mathbb{E}[\xi I_\Lambda \mid \mathfrak{F}] = \frac{\mathbb{E}[\xi I_\Lambda \mid \mathfrak{G}]}{\mathbb{E}[I_\Lambda \mid \mathfrak{G}]} I_\Lambda$.

The proof of this lemma can be found in page 94 of Kallenberg (2001).

Symbols

Subjects

Classes of Processes

Bibliography

[1] Ahmed, S., Filipovic, D. and Svindland, G., 2008. A note on natural risk statistics. *Operations Research Letters.*, **6**36, 662–664.

[2] Artzner, P. and Delbaen, F., 1995. Default risk insurance and incomplete markets. *Mathematical Finance*, **5**, 187-195.

[3] Artzner, P. Delbaen, F., Eber, J.-M. and Heath, D. 1999. Coherent measures of risk. *Mathematical Finance*, **9**, 203-228.

[4] Aven, T., 1985. A theorem for determining the compensator of a counting process. *Scandinavian Journal of Statistics*, **12**, 69-72.

[5] Balbás, A., Balbás, B. and Balbás R. (2010). Minimizing measures of risk by saddle point conditions. *Journal of Computational and Applied Mathematics*, **234**, 2924–2931.

[6] Balbás, A., Balbás, R. and Garrido, J. (2010). Extending pricing rules with general risk functions . *European Journal of Operational Research*, **201**, 23–33.

[7] Balbás, A., Garrido, J. and Mayoral, S. (2008). Properties of Distortion Risk Measures. *Methodoly and Computing in Applied Probability*, **11** (3), 385–399.

[8] Balbás, A., López, S. (2008). Sequential arbitrage measurements and interest rate envelopes. *Journal of Computational and Applied Mathematics*, **138** (3), 361–374.

[9] Black, F. and Cox, J.C., 1976. Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance*, **31**, 351-367.

[10] Brémaud, P., 1981. *Point Processes and Queues: Martingale Dynamics*. Springer–Verlag: New York.

[11] Çetin, U., Jarrow R., Protter P. and Yildirim, Y., 2002. Modeling credit risk with partial information. *The Annals of Applied Probability*, **14**, 1167–1178.

[12] Chen, Z. and Knez, P.J., 1995. Measurement of market integration and arbitrage. *Review of Financial Studies*, **8** (2), 545–560.

[13] Cherubini, U. and Luciano, E., 2003. Pricing and hedging credit derivatives with copulas. *Economic Notes*, **32** (2), 219-241.

[14] Chiu, S.N. and Yin, C., 2005. Passage times for a spectrally negative Lévy process with applications to risk theory. *Bernoulli*, **11** (3), 511–522.

[15] Choulli, T., Krawczyk, L. and Stricker, C., 1998. $\mathcal{E}$-martingales and their applications in mathematical finance. *Annals of Probability*, **26** (2), 853–876.

[16] Choulli, T., Vandaele, N. and Vanmaele, M., 2010. The Föllmer-Schweizer decomposition: Comparison and description. *Stochastic Processes and their Applications*, doi:10.1016/j.spa.2010.02.004.

[17] Cont, R. and Tankov, P., 2004. *Financial Modelling With Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series: Boca Raton.

[18] Dellacherie, C. and Doleans-Dade, C., 1970. Un contre-exemple au problème des Laplaciens approches. Séminaire de Probabilités V. *Lecture Notes in Mathematics*, **191**, 128-140.

[19] Dellacherie, C., Maisonneuve, B. and Meyer, P.A., 1992. *Probabilités et potentiel*. Chapitres XVII-XXIV. Hermann, Paris.

[20] Dellacherie, C. and Meyer, P.A., 1982. *Probabilities and Potential B*. North-Holland.

[21] Duffie, D. and Lando, D., 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica*, **69** (3), 633-664.

[22] Duffie, D., Pedersen, L.H. and Singleton, K.J., 2003. Modeling sovereign yield spreads: A case study of Russian debt.. *The Journal of Finance*, **58** (1), 119-159.

[23] Duffie, D., Schroder, S. and Skiadas, C., 1996. Recursive valuation of defaultable securities and the timing of resolution of uncertainty.. *Annals of Applied Probability*, **6**, 1075–1090.

[24] Duffie, D. and Singleton, K.J., 1999. Modeling term structures of defaultable bonds. *Review of Financial Studies*, **12**, 687-720.

[25] Elliott, R.J., Jeanblanc, M. and Yor, M., 2000. On models of default risk. *Mathematical Finance*, **10**, 179–195.

[26] Elouerkhaoui, Y., 2007. Pricing and hedging in a dynamic credit model. *International Journal of Theoretical and Applied Finance*, **10** (4), 703–731.

[27] Föllmer, H. and Schweizer, M., 1991. Hedging of contingent claims under incomplete information.. *Applied Stochastic Analysis*, **5**, 389–414.

[28] Föllmer, H. and Sondermann, D., 1986. Hedging of non-redundant contingent claims. *Contributions to Mathematical Economics*, 205–223.

[29] Geman, H., 2002. Pure jump Lévy processes for asset price modelling. *Journal of Banking and Finance*, **26**, 1297–1316.

[30] Giesecke, K., 2004a. Correlated default with incomplete information. *Journal of Banking and Finance*, **28**, 1521-1545.

[31] Giesecke, K., 2004b. Credit risk modeling and valuation: An introduction. *Credit Risk: Models and Management*, **2**.

[32] Giesecke, K., 2006. Default and information. *Journal of Economic Dynamics and Control*, **30**, 2281-2303.

[33] Giesecke, K. and Goldberg, L.R., 2004. Forecasting default in the face of uncertainty. *Journal of Derivatives*, **12** (1), 14–25.

[34] Guo, X., Jarrow, R.A. and Zeng Y., 2009. Credit risk models with incomplete information. *Mathematics of Operations Research*, **34** (2), 320–332.

[35] Guo, X. and Zeng, Y. 2008. Intensity process and compensator: A new filtration expansion approach and the Jeulin–Yor formula. *The Annals of Applied Probability*, **18** (1), 120–142.

[36] Harrison, J.M. and Pliska, S.R., 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, **11**, 215–260.

[37] He, S.-W., Wang, J.-G. and Yan, J.-A., 1992. *Semimartingale Theory and Stochastic Calculus*. CRC Press: Boca Raton.

[38] Heyde, C.C., Kou, S. G., Peng, X. H., 2007. What is a good external risk measure: Bridging the gaps between robustness, subadditivity, and insurance risk measures. *Preprint.*

[39] Jacod, J. and Shiryaev, A.N., 1987. *Limit Theorems for Stochastic Processes*. Springer: Berlin.

[40] Jarrow, R.A. and Turnbull, S.M., 1995. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, **50** (1), 53-86.

[41] Jeanblanc, M. and LeCam, Y., 2007. Reduced form modelling for credit risk. Available at SSRN: http://ssrn.com/abstract=1021545.

[42] Jeulin, T. and Yor, M., 1978. Grossissement d'une filtration et semimartingales: formules explicites. *Séminaire de Probabilités XII. Lecture Notes in Math.*, **649**, 78–97, Springer, Berlin.

[43] Johnson, G. and Helms, L.L., 1963. Class D supermartingales. *Bulletin of the American Mathematical Society*, **69**, 59–62.

[44] Kallenberg, O., 2001. *Foundations of Modern Probability*. Springer: New York.

[45] Karatzas, I. and Shreve, S., 1988. *Brownian Motion and Stochastic Calculus*. Springer-Verlag: New York.

[46] Kempf, A. and Korn, O., 1998. Trading system and market integration. *J. Financ. Intermed.*, **7**, 220–239.

[47] Kou, S.G. and Wang, H., 2003. First passage time of a jump diffusion process. *Applied Probability*, **35**, 504–531.

[48] Kunita, H. , 2010. Itô's stochastic calculus: Its surprising power for applications. *Stochastic Processes and their Applications*, **120**, 622–652.

[49] Kunita, H. and Watanabe, S., 1967. On square integrable martingales. *Nagoya Math. J.*, **30**, 209–245.

[50] Kyprianou, A.E., 2006. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer–Verlag: Berlin Heidelberg.

[51] Luenberger, D.G., 1969. *Optimization by Vector Spaces Methods*. John Wiley & Sons: New York.

[52] Merton, R.C., 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, **29**, 449-470.

[53] Meyer, P.A., 1962. A decomposition theorem for supermartingales. *Illinois Journal of Finance*, **6**, 193–205.

[54] Meyer, P.A., 1963. Decompositions of supermartingales. The uniqueness theorem. *Illinois Journal of Mathematics*, **7**, 1–17.

[55] Meyer, P.A., 1966. *Probability and Potentials*. Blasisdell Publishing Company: Massachusetts

[56] Monat, P. and Stricker, C. 1994. Fermeture de $G_T(\theta)$ et de $L^2(\mathcal{F}_0) + G_T(\Theta)$. *In Séminaire de Probabilités XXVIII, Lecture Notes in Mathematics*, 189– 194.

[57] Monat, P. and Stricker, C. 1995. Föllmer-Schweizer decomposition and mean-variance hedging of general claims. *Annals of Probability*, **23**, 605-628.

[58] Pham, H., 1999. On quadratic hedging in continuous time. *preprint, Université de Marne-la-Vallée, Paris.*

[59] Protter, P., 2004. *Stochastic Integration and Differential Equations*. Springer–Verlag: Berlin.

[60] Rockafellar, R.T., Uryasev, S. and Zabarankin, M. 2006. Generalized deviations in risk analysis. *Finance and Stochastics*, **10**, 51-74.

[61] Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J., 1999. *Stochastic Processes for Insurance and Finance*. John Wiley & Sons: Chichester.

[62] Rudin, W., 1982. *Functional Analysis*. Tata McGraw–Hill Publishing Company LTD.: New Delhi.

[63] Rudin, W., 1966. *Real and Complex Analysis*. McGraw–Hill Book Company: New York.

[64] Sarig, O. and Warga, A., 1989. Some empirical estimates of the risk structure of interest rates. *Journal of Finance*, **44**, 1351-1360.

[65] Schweizer, M., 1988. Hedging of options in a general semimartingale model. *Dissertation ETH No. 8615*, Zürich.

[66] Schweizer, M., 1991. Option hedging for semimartingales. *Stochastic Processes and their Applications*, **37**, 339–363.

[67] Schweizer, M., 1994. Approximating random variables by stochastic integrals.. *Annals of Probability*, **22** (3), 1536–1575.

[68] Schweizer, M., 1999. A guided tour through quadratic hedging approaches. *Humboldt Universitat Berlin*, **96**.

[69] Schweizer, M., Heath, D. and Platen, E., 2001. A comparison of two quadratic approaches to hedging in Incomplete Markets. *Mathematical Finance*, **11**, 385–413.

[70] Tauchen, G. and Zhou, H., 2006. Realized jumps on financial markets and predicting credit spreads. Federal Reserve Board, Divisions of Research, Statistics, and Monetary Affairs, *Finance and Economics Discussion Series*, Washington, D.C.

[71] Wagner, N., 2008. *Credit Risk: Models, Derivatives, and Management*. Chapman and Hall/CRC Financial Mathematics Series: Boca Raton.

[72] Wang SS, 2000. A class of distortion operators for financial and insurance risks. *Journal of Risk and Insurance*, **67**, 15–36.

[73] Zhou, C., 2001. The term structure of credit spreads with jump risk. *Journal of Banking and Finance*, **25**, 2015-2040.