Compensators and Diffusion Approximation of Point Processes and Applications

Xin Dong

Department of Mathematics

Imperial College London

Thesis presented for examination for the degree of

Doctor of Philosophy in Mathematics

of Imperial College London

August 2014

Declaration

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Xin Dong

Copyright

Acknowledgement

This thesis would not have been possible and my PhD life would not be so wonderful without support and love from many people.

Firstly I would like to express my deep gratitude to my supervisor Dr. Harry Zheng for his continuous and boundless support and encouragement throughout my PhD study. His advices and comments are very critical to this thesis.

My sincere thanks go to Dr. Angelos Dassios from Department of Statistics, London School of Economics. He generously shared his time, idea and experience with me. He advised and inspired me all these years.

I am deeply grateful to Department of Mathematics at Imperial College. I enjoyed various well-organized reading groups, seminars and courses. I also benefitted from interesting discussions with many staff members in the Mathematical Finance section. I appreciate the companion and support from my lovely colleagues and friends in our Department: Luke Charleton, Dr. Wang Han, Dr. Kai Li, Yusong Li, Cong Liu, Daphne Liu, Francesco Mina, Benoit Pham-Dang, Patrick Roome and etc.

My special thanks also go to the Risk Analytics team at Citigroup for the financial support. The projects carried out provided me an opportunity to understand the practical finance world. Dr. Marco Polenghi, Mr. Rickard Brannvall and Dr. Lappas have been always supportive and helpful. My colleagues at Citigroup shared a lot of experience with me and brought me a lot of care and encouragement.

Finally, but most importantly, I cherish the unconditional love and support from my beloved parents; as well as Yi Xing, who has always been standing besides me with his companion and encouragement.

Abstract

In this thesis, we study two classes of point processes by analysing key properties and discussing applications in finance and insurance. The first point process studied was the default indicator process in credit risk modelling. We considered a pure jump Lévy process of finite variation for the asset value and an unobservable random barrier. The default time was defined as the first time the asset value falls below the barrier. Using the indistinguishable intensity process and the instantaneous likelihood process, we proved the absolute continuity of the compensator for the default indicator process, or equivalently, the existence of the intensity process of the default time. Moreover, we found the explicit representation of the intensity in terms of the distance between the asset value and its running minimal value, thus the intensity is an endogenous process, which sheds new light on the relationship between the intensity model and the structural model.

The second class of point processes is the Dynamic Contagion Process, which has intensities modelled with a shot-noise component describing the external impact and mutually-exciting jump components that describe the internal contagion effect. In the bivariate case, we found the stationarity condition with which we explored the diffusion approximation of the high frequency point process system and applied it in filtering. In the univariate case, we constructed a pure jump process derived from a dynamic contagion process and showed the weak convergence to a Cox-Ingersoll-Ross model (CIR) process. The pathwise approximation provides an alternative method of simulating the square-root processes and can be further extended to the approximation of the Heston model in option pricing.

Contents

Abstract **5**

1 Introduction **12**

1.1 Preamble 12

1.2 Overview 12

1.3 Organization of the Thesis 16

2 A Pure Jump Lévy Structural Model with a Barrier of Incomplete Information **18**

2.1 Introduction 19

2.1.1 Basics of Credit modelling 19

2.1.2 Basics of Lévy processes 24

2.1.3 Motivation 27

2.2 The Model and the Main Result 28

2.3 Proof of the Main Theorem 32

2.3.1 Compensators and Likelihood Processes under Different Filtrations 32

2.3.2 Spectrally Negative Lévy Process with Finite Variation . . . 37

2.3.3 Lévy Process with Finite Variation 41

2.3.4 Indistinguishability of Likelihood Process and Intensity Process 44

2.4 Conclusion 45

3 Introduction to Bivariate Dynamic Contagion Processes **47**

3.1 Introduction to Point Process Modelling 47

3.2 Introduction of Basics 49

3.2.1 Basics of Point Processes 49

3.2.2 Markov Processes 52
3.3 The BDCP Model 54
3.3.1 Definition 54
3.3.2 The Branching Structure 58
3.4 Existence of Moments 60
3.5 Exact Simulation of BDCP 62
3.6 Graphic Illustration 66
3.7 Conclusion 66

4 BDCP Stationarity Analysis 68

4.1 Introduction 68
4.2 Markov Property and Limiting Distributions 70
4.2.1 Markov Property 70
4.2.2 The Limiting Distributions 73
4.3 The Stationary Distributions 77
4.4 Stationary Moments 83
4.4.1 Stationary Mean 84
4.4.2 Stationary Variance 84
4.4.3 Stationary Correlation 87
4.5 Proof of Lemmas 87
4.6 Conclusion 91

5 BDCP Diffusion Approximation and Applications in Filtering 92

5.1 Introduction 92
5.2 Diffusion Approximation of The BDCP 96
5.2.1 High Frequency Events System 97
5.2.2 Diffusion Approximation 99
5.3 Kalman-Bucy Filtering with Diffusion Approximation 103
5.3.1 Filtering with Observation (N^1, N^2) 105
5.3.2 Filtering with Observation $N^1 + N^2$ 105
5.3.3 An Approximate Filter 107
5.4 Applications 110
5.4.1 Pricing Problem 112
5.4.2 Type Identification Problem 114

5.5 Proofs . . . 115
5.5.1 Proofs in Section 5.2 . . . 115
5.5.2 Proofs in Section 5.3 . . . 126
5.6 Conclusion . . . 129

6 Diffusion Approximation of UDCP and Alternative CIR Simulation Scheme 130
6.1 Introduction . . . 131
6.1.1 Univariate Dynamic Contagion Processes . . . 131
6.1.2 Basics of CIR Processes . . . 131
6.1.3 Basics of the Heston Model . . . 132
6.1.4 Simulation Issues . . . 133
6.2 Diffusion Approximation . . . 135
6.2.1 Scaled UDCP Intensity Processes . . . 136
6.2.2 Weak Convergence to CIR processes . . . 137
6.2.3 From Variance Processes to Asset Price Processes . . . 139
6.3 Approximation Simulation Scheme for CIR Processes . . . 141
6.4 Numerical Results . . . 142
6.5 Conclusion . . . 145

7 Conclusions and Future Research 146

Bibliography 147

List of Figures

2.1 The asset return process X as in Example 2.2.6, the distance process $X-\underline{X}$ and the intensity processλ. The data used are $(c,\nu,\sigma,\theta) = (-0.02, 0.1, 0.15, 0.01)$. 33

2.2 The term structure of credit spread $S(t,h)$: the asset return process X as in Example 2.2.6 with the data $(c,\nu,\sigma,\theta) = (-0.02, 0.1, 0.15, 0.01)$, $X-\underline{X}$ at t=0.5 is 0.0585. 34

3.1 A simulated path of the BDCP and its intensity $(N^1, N^2, \lambda^1, \lambda^2)$. The self-exciting jumps $\left\{\lambda^k_{T^k_j+} - \lambda^k_{T^k_j}\right\}_{j\geq 1}$ and cross-exciting jumps $\left\{\lambda^{k'}_{T^k_j+} - \lambda^{k'}_{T^k_j}\right\}_{j\geq 1}$ for $k,k' = 1,2$ and $k \neq k'$. 67

5.1 Observation scenario (S1): A sample path of λ^k_t compared with its unconditional stationary mean and conditional estimate $\lambda^{k,F}_t = \mathbb{E}[\lambda^k_t|\sigma\{N^1_s, N^2_s, s \leq t\}]$ for $k = 1,2$. 111

5.2 Histograms describing the empirical distribution of λ^k_T and $\lambda^k_T|\sigma\{N^1_t, N^2_t, t \leq T\}$ denoted by $\lambda^{k,F}_T$ in the graph at $T = 10$ for $k = 1,2$. 112

5.3 Approximate solution of type estimation $\mathbb{P}\left(N^i_t \leq k|\mathcal{F}^{N^1+N^2}_t\right)$ at time $t = 1$. 115

6.1 A Sample path of a scaled UDCP intensity and a CIR process (with different random seeds). 139

6.2 The Laplace transform of the integrated process with different parameter sets. 145

List of Tables

3.1 Parameters for the simulated path 66

4.1 Notation of distributions in the branching system. 70
4.2 Coefficient table for A, B, C, $\Delta_1 = \delta_1 - \mu_{G^{1,1}} > 0$ and $\Delta_2 = \delta_2 - \mu_{G^{2,2}} > 0$. 85

5.1 Parameters for numerical applications 110
5.2 Comparison of the approximate conditional and unconditional expectations with $T = 19.5$. 113
5.3 Approximate stop-loss contract value of a given path of observation (N_t^1, N_t^2), with $H(x) = (x_1 + x_2 - d)^+$ with $T = 19.5$. 114
5.4 Approximate type estimation when $N^1 + N^2$ is observable 115

Notations

- $(\Omega, \mathcal{F}, \mathbb{P})$ - the probability triple consisting of a sample space Ω, the σ-algebra $\mathcal{F}$ which is the set of all measurable events, an the probability measure $\mathbb{P}$.
- $L^2(\mathbb{P})$ - the set of all square-integrable random variables.
- $(\mathcal{F}_t)_{t\geq 0}$ - a filtration, that is an increasing family of sub-σ-algebras of $\mathbb{F}$; $\mathcal{F}_s \subset \mathcal{F}_t$, $0 \leq s \leq t$.
- $\mathbb{F}^X = (\mathcal{F}_t^X)_{t\geq 0}$ - the natural filtration generated by the process X.
- $\mathbb{R}^d$ - the d-dimensional Euclidean space.
- $(D_{\mathbb{R}^d}[0,\infty), \mathcal{D})$ - path space of processes with càdlàg paths equipped the Skorohod topology
- P - the probability measure on $(D_{\mathbb{R}^d}[0,\infty), \mathcal{D})$ that is $P = \mathbb{P}X^{-1}$ induced by process X.
- $X^n \Rightarrow X$ - weak convergence of process X^n to X in $(D_{\mathbb{R}^d}[0,\infty), \mathcal{D})$
- $\mathcal{A}$, $\mathcal{D}(\mathcal{A})$ - the generator and its domain of a Feller process
- μ_{H^k}, μ_{iH^k} - the mean and the i-th $(i \geq 2)$ moment of a random variable with distribution H.
- UDCP - univariate dynamic contagion process
- BDCP - bivariate dynamic contagion process

Chapter 1

Introduction

1.1 Preamble

A key challenge in many areas is how to model event arrivals in a system using a point process model that is able to capture a stylized fact. An understanding of key properties is needed to construct an appropriate point process model. In this dissertation, we will focus on the construction of appropriate point process models and study their probabilistic properties. The first model (M1) is for the default event of a firm. It aims to capture the short-term credit risk implied by the market and to explain the default risk in an economically meaningful way. The second model (M2) is a bivariate system for two-type event arrivals in a multi-name system with rich dynamics. We aim to study the stationarity property, investigate the diffusion approximation, and apply to solve filtering problems for the intensity process. The third model (M3) is a univariate pure jump process built on a univariate point process system for an approximate simulation algorithm of square-root diffusion processes.

1.2 Overview

(M1): A First Passage Time Model with Pure Jump Lévy Processes in Credit Modelling

An essential question in credit risk modelling is how to model the default time of a firm. The compensator of the default indicator process from the Doob-Meyer decomposition presents rich information on the default time. For example, the regularity of the compensator conveys the extent of predictability of the default time. If the compensator is continuous, then the default time is totally inaccessible. Moreover, if it is absolutely continuous, then its Radon-Nikodym derivative with respect to the Lebesgue measure, called the intensity process, can be interpreted as the instantaneous default probability. In practice, it is also the short spread observed in the market, which quantifies the compensation an investor requires for bearing the short-term uncertainty of default. Ideally, we are looking for a model that can explain the reason for a firm's default as well as the short-term default risk implied by the market.

There are three main modelling frameworks in the credit risk literature where the default time is defined differently. As a result, the probabilistic properties of default times are different. The first framework is the first passage time structural model (Black and Cox [12]) where the default time is the first time the asset price of the firm falls below a certain barrier. The model is economically meaningful as it explains the reasons for a firm's default. However, the classical model using a geometric Brownian motion asset price and a constant barrier leads to a predictable default time. Hence, the intensity process does not exist, which contradicts market observations. The second framework is the intensity-based model (Lando [51]) where the intensity process is assumed to exist and is modelled by an exogenous process. This framework does not explain why a firm defaults. A third framework aims to bridge the gap between the above two main frameworks. It starts with a first passage time structural model and assumes incomplete information for market investors. It aims to show the existence of the intensity process. Duffie and Lando [29] and Kusuoka [49] assume a noisy accounting report on the asset price process and show the existence of the intensity process. Giesecke and Goldberg [36] and Giesecke [34] propose an unobservable random barrier and an observable geometric Brownian motion asset price process. They conclude that the intensity process does not exist as the compensator is not absolutely continuous.

In line with Giesecke [34], we propose an incomplete information first passage time structural model with an unobservable random barrier and an observable asset price process modelled by a Lévy process with finite variation. Lévy processes with finite variation, including the inverse Gaussian and the variance gamma models, are widely used in price process modelling (Madan et al. [55] and Madan and Schoutens [56]) as they are able to capture some of the stylized facts of returns observed in the market and their probabilistic properties can be characterised explicitly. We show the existence of an intensity process based on the projection theory and the properties of Lévy processes. Moreover, we find the explicit form of the intensity process that is an endogenous process depending on the parameters and the historical path of the asset price process. Therefore, the intensity explains the short spread observed in the market and it is dependent on the default mechanism. We therefore reconcile the structural model and the intensity model in credit risk modelling.

In this framework, the pricing formula can be derived which has an additional term (jump at default) compared to the classical formula in a Cox model. However it is difficult to calculate the additional term explicitly.

(M2): Bivariate Dynamic Contagion Processes

We model event arrivals in a two-type events system by a non-explosive bivariate counting process with specified intensity processes. In order to reflect the idiosyncratic risk from some external factors and also internal contagion effects, we introduce the bivariate dynamic contagion process (BDCP) where the intensities are piecewise deterministic processes with external-exciting jumps and self-exciting jumps. The external-exciting jumps follow a compound Poisson process with an exponential time decay. The self-exciting jumps happen at the jump times of the counting process with a random jump size and an exponential time decay. Therefore, the BDCP is a generalized model of a bivariate Hawkes process (Hawkes and Oakes [39]) and a shot noise Cox process (Cox and Isham [19]).

We study a few key probabilistic properties of the BDCP. We find the stationarity condition under which there exists a stationary version of the process. Note that a BDCP is the limit of a sequence of finite branching systems in a cluster process representation. We first use Markov process theory to study, as time tends to infinity, the limiting distribution of a finite branching system in terms of a Laplace transform. We explore the condition to ensure the existence of the limiting distribution of the BDCP based on the convergence of the branching system. We show the limiting distribution is actually the stationary distribution and conclude that the condition found is the stationarity condition of the BDCP system.

We note that the BDCP is suitable to model a high frequency events system due to the contagion effect. We study the diffusion approximation of the process which enables the application of more results available in the diffusion class. This is done using the martingale central limit theorem (Either and Kurtz [33]) and the stationarity result for the standardized intensity process. We obtain a limiting Gaussian system with the same mean and variance. Notice that the intensity information is important but usually unobservable in practice, therefore we need to solve the filtering problem conditioned on the point process observations to get the best estimate of the intensity. Filtering with point process observations is discussed mainly in Brémaud [13] and Ceci and Geraldi [15]. The innovation approach based on the martingale representation and the projection theory is used and the filter is proved to be a unique solution of the Kushner-Stratonovich (KS) equation. However in general the KS equation can only be solved numerically. In this dissertation, we propose an efficient approximate solution to the filtering problem of the BDCP model. The solution proposed is a Kalman-Bucy filter of the limiting Gaussian diffusion process with actual point process observation inputs. Essentially, we approximate not only the distribution but also the dependence structure of the point process system by the limiting diffusion process. We will show that the constructed filter is an asymptotically optimal filter. Moreover, we apply the diffusion approximation and the filtering solution in insurance for the pricing problem of stop-loss reinsurance contracts and the type estimation problem with numerical examples.

(M3): Pure Jump Processes for Approximate Simulation of CIR Processes

Due to the bias introduced by the negative value adjustment in the Euler simulation method for CIR processes, alternative methods have been discussed in the literature. The exact simulation scheme based on the explicit form of the transition probability of the CIR process is proposed, but it involves complications in sampling and is inefficient and sometimes even yields poor performance. The most popular method in practice is the Quadratic Exponential scheme (QE) introduced by Andersen [4]. The QE scheme is an approximate simulation scheme based on moment matching with a distribution that can be generated easily for the transition probability distribution. However, there is no convergence in this framework.

We propose an approximate simulation scheme for the CIR process with weak convergence results. We construct a pure jump process based on the univariate dynamic contagion process (UDCP) and show the weak convergence to the CIR process when a model parameter tends to infinity using the martingale central limit theorem. Moreover, as the UDCP can be simulated exactly and efficiently, we obtain an alternative approximate simulation algorithm for the CIR processes. We conclude that the simulation scheme works well by comparing the Laplace transform of the simulated jump process with the theoretical value of the CIR process. As an extension, we can show that if we use the CIR process to model the stochastic volatility process of the asset, we obtain an approximate simulation scheme for the Heston model based on the weak convergence.

1.3 Organization of the Thesis

Chapter 2 discusses the first model (M1), a first passage time structural model with a pure jump Lévy asset price process and an unobservable debt level in credit risk modelling. First, an overview of credit modelling and the motivation of our model construction will be introduced. Then we present our model setup and the main theorem about the existence of an endogenous intensity process. A few examples with graphic illustrations and applications in credit risk are discussed. We prove the main theorem in four steps.

Chapters 3, 4, 5 discuss the second model (M2), the bivariate dynamic contagion processes (BDCPs).

In Chapter 3, we introduce the BDCP. We provide the definition the BDCP based on the intensity process and the cluster process representation, respectively. We also study the basic properties like the Markov property, an existence condition for the moments and an exact simulation algorithm.

In Chapter 4, we investigate the existence of a stationary distribution of the BDCP. We use Markov process theory to explore the limiting distributions of finite systems and the BDCP as its limit when the system index tends to infinity. We verify the limiting distributions are stationary distributions of the finite system and the BDCP. We then compute stationary moments of the intensity process of the BDCP.

In Chapter 5, we explore the diffusion approximation of the BDCP system and applications. We introduce the high frequency events framework and derive the diffusion approximation of the BDCP system. We apply the result to filtering the intensities based on the point processes observations. Moreover, the assessment of the performance of the filter is also provided. A few examples of filtering applications in insurance are discussed.

Chapter 6 discusses the third model (M3), a pure jump process built upon a univariate dynamic contagion process. We demonstrate a sequence of such pure jump processes converging to a CIR process weakly in the path space. Moreover, we will show the weak convergence of an extended model to the Heston model in stochastic volatility modelling. We provide an approximate simulation algorithm of the CIR process and assess the performance by comparing the Laplace transform with the theoretical one.

In Chapter 7, we conclude the dissertation and discuss some open questions for future research.

Chapter 2

A Pure Jump Lévy Structural Model with a Barrier of Incomplete Information

In this chapter we discuss a credit risk model with a pure jump Lévy process for the asset value and an unobservable random barrier. The default time is the first time when the asset value falls below the barrier. Using the indistinguishability of the intensity and the likelihood processes, we prove the existence of the intensity process of the default time and find its explicit representation in terms of the distance between the asset value and its running minimum value. The intensity is therefore endogenous and represents the compensation for the short-term credit risk. We apply the result to find the instantaneous credit spread process and illustrate it with a numerical example.

The chapter is organized as follows. In Section 2.1, we first introduce credit modelling, the motivation for the work, and review some basics of Lévy processes. Section 2.2 introduces our model and states the main result (Theorem 2.2.2) with several examples, and discusses the instantaneous credit spread as an application with a numerical example. Section 2.3 proves the main result with details discussed in four subsections. Section 2.4 discusses our conclusions.

2.1 Introduction

In this section, we will first introduce credit risk modelling in Section 2.1.1, and our motivation of the project in Section 2.1.3. We then review some basics of Lévy processes in Section 2.1.2.

2.1.1 Basics of Credit modelling

In credit risk modelling, how to model the default time of a firm and value the credit-linked products related to the firm are essential objectives. With a given model, the goal is to find the default probability of a firm.

Following Protter [59], we first introduce the classification of stopping times T defined on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F}$.

The classification of stopping times

- Predictable stopping times:

  A stopping time T is *predictable* if there exists a sequence of stopping times $(T_n)_{n\geq 1}$ such that T_n is increasing, $T_n < T$ on $\{T > 0\}$, for all n, and $\lim_{n\to\infty} T_n = T$ *a.s.* Such a sequence $(T_n)_{n\geq 1}$ is said to announce T.

- Accessible stopping times:

  A stopping time T is *accessible* if there exists a sequence $(T_n)_{n\geq 1}$ of predictable stopping times such that

  $$P\left(\bigcup_{n=1}^{\infty}\{\omega : T_n(\omega) = T(\omega) < \infty\}\right) = P(T < \infty).$$

  Such a sequence $(T_n)_{n\geq 1}$ is said to envelop T.

- Totally inaccessible stopping times:

  A stopping time is *totally inaccessible* if for every predictable predictable stopping time S,

  $$P(\{\omega : T(\omega) = S(\omega) < \infty\}) = 0.$$

We now recall an important theorem in stochastic analysis.

Theorem 2.1.1 (Doob-Meyer Decomposition (Protter [59])). *Let Z be a càdlàg supermartingale with $Z_0 = 0$ of class D. Then there exists a unique, increasing, predictable process A with $A_0 = 0$ such that $M_t = Z_t + A_t$ is a uniformly integrable martingale.*

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$, the default indicator process $N_t = 1_{\{\tau \leq t\}}$ is an adapted non-decreasing process and by the Doob-Meyer decomposition above, there exists a unique $\mathbb{G}$-predictable non-decreasing process A with $A_0 = 0$, such that $N_t - A_t$ is a $\mathbb{G}$-martingale, and we call A the $\mathbb{G}$-*compensator* of τ which counteracts the increasing trend of N. The probabilistic properties of the default time τ can be characterized by A. For example, if τ is predictable, then the compensator A is equal to N itself. The continuity of the compensator A implies that τ is totally inaccessible. If A is absolutely continuous and the intensity λ is specified as the Radon-Nikodym derivative of the compensator A with respect to the Lebesgue measure, then we have $A_t = \int_0^t \lambda_s ds$ and

$$\mathbb{P}(\tau \in [t, t+dt) | \mathcal{G}_t) = \lambda_t dt. \tag{2.1}$$

That is the intensity is the instantaneous default likelihood process, or instantaneous credit spread (short spread) if it exists and it provides the first order approximation of the conditional default probability over a small interval. We will illustrate this point in detail later. Note that with the explicit form of A, one can compute the default probability and the expected losses.

In this chapter, we focus on a model which explains the following market observation (see Duffie and Lando [29]).

(*) Market observation of short-term credit spread: The instantaneous credit spread is positive and finite before the default event. Writing this mathematically,

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{P}(\tau \in [t, t+\Delta) | \mathcal{G}_t)$$

exists. This property is guaranteed by the existence of the intensity process by (2.1).

In credit risk modelling literature, there are two main modelling frameworks: the structural model and the intensity models.

The structural model is built on the asset-liability structure of firms. Default happens when the firm value process falls below its debt level. The structural model was initiated by Merton [57], where the default status is discussed of a fixed time T, and this has been extended by Black and Cox [12] to the first passage time problem where the default time is defined as the first time the value process falls below the debt level. As we are focusing on the default time modelling, we will discuss only the first passage type. In order to obtain the law of the default time in this framework, we investigate the first passage time problem or equivalent by the law of the running minimal/maximal of the process. Though this modelling framework is economically meaningful, classical structural models cannot explain the instantaneous credit spread in the market. For example, if one models the asset value process as a diffusion process and the debt level as a constant, then the default time is predictable and the compensator A is the same as the indicator process N, which implies the intensity process does not exist.

Instead of modelling the asset-liability of a firm, *the intensity model* is based on the assumption that the default happens as a surprise to the market, under which the default time is a totally inaccessible stopping time. Moreover, the compensator A is absolutely continuous and the intensity process λ is modelled explicitly. The intensity can be modelled as a constant, or a deterministic function, or a stochastic process. In this framework, if the default time can be modelled as the first jump time of a Cox process where the intensity depends on some external state processes, Lando [51] showed that the defaultable claims can be priced with the standard approach at the discount rate r replaced by $r + \lambda$ in the discount factor. Though the short-term credit spread is non-zero which explains the market observation, this framework is silent about the reason for a firm's default.

Therefore, the two main frameworks in credit modelling have their own pros and cons. How to reconcile them such that there is a model being able to explain the default mechanism and yields an appropriate intensity process becomes an interesting and important question. It is also the objective of the research in this chapter.

Incomplete Information Modelling In the literature, there is an incomplete information modelling approach that combines the structural and intensity models that tries to incorporate the best features of both.

In a first passage time structural model, we denote the filtration representing the market information flow by $\mathbb{G}$. We say the market information is complete if both the asset value process and the debt barrier are observable and included in $\mathbb{G}$. In reality, we do not always have complete information. We name a sub-filtration $\mathbb{F}$ of $\mathbb{G}$ the base filtration, which includes essential information of the model, then we need to expand the filtration $\mathbb{F}$ to explore the compensator or intensity under the market filtration $\mathbb{G}$.

Theory of filtration expansion and its application in structural modelling with incomplete information has been studied intensively. Discussion about compensators of random times in different filtrations can be found in Jeulin and Yor [43], Guo and Zeng [38], and Janson et al. [42].

In credit risk modelling, assume X is the log firm value process, and D is the barrier, and the default time τ is defined as the first time X falls below the barrier D. Discussion in the literature about credit modelling with incomplete information can be divided into three cases:

(i) Incomplete information about the firm value process X

Duffie and Lando [29] assume a discretely observable firm value X with noises and a constant barrier D. In this case, the conditional density of the asset value process and the intensity process were found. Kusuoka [49] extended this to continuously observable noisy firm values. Çetin et al. [16] assumes the information reduction of X and a constant D, intensity process was derived based on an Azéma martingale and the pricing is done by using the excursion theory of Brownian motion.

Essentially, the conditional density of the random time τ under the base filtration $\mathbb{F}$ should be found. Then the problem becomes easier as this conditional density is closely linked with the intensity process. The general framework and problems about the conditional density are discussed in N. El Karoui et al. [44].

(ii) Incomplete information about random barrier D

In this framework, the firm value process X is assumed to be observable and the base filtration $\mathbb{F}$ is the natural filtration generated by X. In order to describe the fact that market investors cannot observe the true level of liabilities of a firm, the barrier D should be chosen randomly and unobservable.

This random and unobservable barrier has been introduced in Giesecke [34] with a log firm value X being a Brownian motion and debt barrier D being a random variable. It concluded that if X is a diffusion process, the $\mathbb{G}$-compensator A of τ is continuous but not absolutely continuous. Therefore τ is a totally inaccessible stopping time under the market filtration $\mathbb{G}$, but it does not admit an intensity process. The same conclusion holds for general diffusion processes. Under the same setup, Goldberg and Giesecke [36] further discuss the specification and calibration of the model (I^2 model) where X is a Brownian motion process. Moreover, besides the Brownian motion, they provided an example of a calculation when taking X as a Poisson process with a negative sign. However, they concluded that even though the intensity exists with an explicit form due to the monotone path property, the asset value process assumed is not reasonable. As far as we know, it is the only asset value model with a discontinuous path with this information setup in the literature.

(iii) Incomplete information about both X and Δ

It is an extension of the previous two cases.

As the inaccessible liability of a company is an appropriate assumption in practice, it makes the discussion in case (ii) important. With the diffusion assumption for X in case (ii), Giesecke [34] showed the intensity process of the default time does not exist, and thus it is insufficient to explain the positive instantaneous short-term credit spread phenomenon. An alternative is to include jumps in the asset value process in case (ii). Pure jump processes are widely used in financial modelling as they can capture the stylized fact of the asset returns, such as infinite activities, jumps, skewness, and kurtosis. For example, Madan et al. [55] use a variance gamma process for the stock price in option pricing. Moreover, inspired by Madan and Schoutens [56] where a drifted subordinator is used for the log firm value process in a first passage time model with complete information, we extend the setup to incomplete information and discuss a broader class of pure jumps processes covering drifted subordinators and variance gamma processes.

2.1.2 Basics of Lévy processes

In our model the asset value process is modelled as a pure jump Lévy process, hence in this section we recall briefly the definition and properties of Lévy processes from Kyprianou [50].

Definition 2.1.2 (Lévy Process). *A process $X = \{X_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Lévy process if it possesses the following properties:*

1. *The paths of X are $\mathbb{P}$-almost surely right continuous with left limits.*
2. $\mathbb{P}(X_0 = 0) = 1$.
3. *For $0 \leq s \leq t$, $X_t - X_s$ is equal in distribution to X_{t-s}.*
4. *For $0 \leq s \leq t$, $X_t - X_s$ is independent of $\{X_u : u \leq s\}$.*

We will show some examples of Lévy processes in Example 2.2.4, 2.2.5 and 2.2.6.

Lévy processes have an intimate relationship with infinitely divisible distribution defined in the following.

Definition 2.1.3 (Infinitely Divisible Distribution). *We say that a real-valued random variable Θ has an infinitely divisible distribution if for each $n = 1, 2, \ldots$ there exist a sequence of i.i.d. random variables $\Theta_{1,n}, \ldots, \Theta_{n,n}$ such that*

$$\Theta \stackrel{d}{=} \Theta_{1,n} + \cdots + \Theta_{n,n}$$

where $\stackrel{d}{=}$ is equality in distribution. Alternatively, the law of μ of a real-valued random variable is infinitely divisible if for each $n = 1, 2, \ldots$ there exists another law μ_n of a real-valued random variable such that $\mu = (\mu_n)^{*n}$ where $(\mu_n)^{*n}$ is the n-fold convolution of μ_n.

The infinitely divisible distribution can be characterized by the characteristic exponent Ψ and an expression known as the Lévy-Khintchine formula.

Theorem 2.1.4 (Lévy-Khintchine Formula). *The probability law μ of a real-valued random variable is infinitely divisible with characteristic exponent Ψ,*

$$\int_{\mathbb{R}} e^{i\theta x} \mu(dx) = e^{-\Psi(\theta)} \text{ for } \theta \in \mathbb{R},$$

if and only if there exists a triple (a, σ, π) *where* $a \in \mathbb{R}$, $\sigma \geq 0$ *and* π *is a measure concentrated on* $\mathbb{R}\backslash\{0\}$ *satisfying* $\int_{\mathbb{R}}(1 \wedge x^2)\pi(dx) < \infty$, *such that*

$$\Psi(\theta) = ia\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} \left(1 - e^{i\theta x} + i\theta x \mathbf{1}_{\{|x|<1\}}\right) \pi(dx)$$

for every $\theta \in \mathbb{R}$.

The measure π *is called the Lévy measure of* X.

The following theorem gives the Lévy-Khintchine formula for Lévy processes and it indicates that one can construct a Lévy process such that X_1 has the specified infinitely divisible distribution.

Theorem 2.1.5 (Lévy-Khintchine Formula for Lévy Processes)**.** *Suppose that* $a \in \mathbb{R}$, $\sigma \geq 0$ *and* π *is a measure concentrated on* $\mathbb{R}\backslash\{0\}$ *such that* $\int_{\mathbb{R}}(1 \wedge x^2)\pi(dx) < \infty$. *From this triple define for each* $\theta \in \mathbb{R}$,

$$\Psi(\theta) = ia\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} \left(1 - e^{i\theta x} + i\theta x \mathbf{1}_{\{|x|<1\}}\right) \pi(dx). \tag{2.2}$$

Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *on which a Lévy process* X *is defined having the characteristic exponent* Ψ, *i.e.*

$$\mathbb{E}[e^{i\theta X_t}] = e^{-t\Psi(\theta)}.$$

The following theorem known as the Lévy–Itô decomposition characterizes the path structure of Lévy processes by the Lévy-Khintchine formula.

Theorem 2.1.6 (Lévy–Itô Decomposition)**.** *Given any* $a \in \mathbb{R}$, $\sigma \geq 0$ *and measure* π *concentrated on* $\mathbb{R}\backslash\{0\}$ *satisfying*

$$\int_{\mathbb{R}} (1 \wedge x^2)\pi(dx) < \infty,$$

there exists a probability space on which three independent Lévy processes $X^{(1)}$, $X^{(2)}$, *and* $X^{(3)}$ *exist such that* $X = X^{(1)} + X^{(2)} + X^{(3)}$ *is a Lévy process with characteristic exponent* Ψ *in (2.2) where*

- $X^{(1)}$: *a linear Brownian motion with drift given by* $X_t^{(1)} = -at + \sigma B_t$.
- $X^{(2)}$: *a compound Poisson given by* $X_t^{(2)} = \sum_{i=1}^{N_t} \xi_i$ *where* N *is a Poisson process with rate* $\pi(\mathbb{R}\backslash(-1,1))$ *and* $\{\xi_i\}_{i\geq 1}$ *are i.i.d. random variables with distribution* $\frac{\pi(dx)}{\pi(\mathbb{R}\backslash(-1,1))}$ *concentrated on* $\{x \in \mathbb{R} : |x| \geq 1\}$.

- $X^{(3)}$ *is a square integrable martingale with a.s. countable number of jumps on each finite interval which are of magnitude less than unity and with characteristic exponent given by* $\int_{|x|<1}(1 - e^{i\theta x} + i\theta x)\pi(dx)$.

Now we focus on path properties and first recall the concept of finite variation of a process.

Definition 2.1.7 (Finite Variation (Protter [59])). *An adapted, càdlàg process X is a finite variation process (FV) if almost surely the paths of X are of finite variation on each compact interval of $[0, \infty)$. Note that the variation of a path ω of X over the interval $[a, b]$ is defined by:*

$$V_{[a,b]}(\omega) = \sup_{\pi \in \mathcal{P}} \sum_{t_i \in \pi} |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)|,$$

where $\mathcal{P}$ are all finite partitions of $[a, b]$.

Now we present the path property of variation in the following proposition.

Proposition 2.1.8 (Lévy Processes with Finite Variation). *A Lévy process with Lévy-Khintchine exponent corresponding to the triple (a, σ, π) has paths of finite variation if and only if*

$$\sigma = 0 \quad \text{and} \int_{\mathbb{R}} (1 \wedge |x|)\pi(dx) < \infty.$$

Remark 2.1.9. *If the Lévy process X is of finite variation, then the characteristic exponent can be written as*

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

where $d = -(a + \int_{|x|<1} x\pi(dx))$, and in this case, X has the representation as

$$X_t = dt + \int_{[0,t]} \int_{\mathbb{R}} xN(ds \times dx), \quad t \geq 0.$$

We introduce the concept of the subordinator as a fundamental element in the finite variation process family.

Definition 2.1.10 (Subordinator). *A Lévy process is a subordinator if and only if $\pi(-\infty, 0) = 0$, $\int_{(0,\infty)}(1 \wedge x)\pi(dx) < \infty$, $\sigma = 0$ and $d = -\Big(a + \int_{(0,1)} x\pi(dx)\Big) \geq 0$.*

By definition, it is clear that a subordinator has non-decreasing paths.

Definition 2.1.11 (Spectrally One-sided Process). *A Lévy process X is a spectrally positive Lévy process if it is not a subordinator and $\pi(-\infty, 0) = 0$. A Lévy process X is a spectrally negative Lévy process if $-X$ is a spectrally positive process.*

Note that spectrally one-sided Lévy processes may be of finite or infinite variation.

Furthermore, note that Lévy processes have the strong Markov property.

Proposition 2.1.12 (Strong Markov Property). *Suppose X is a Lévy process and τ is a stopping time. Define on $\{\tau < \infty\}$ the process $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ where*

$$\tilde{X}_t = X_{\tau+t} - X_\tau, \quad t \geq 0.$$

Then on the event $\{\tau < \infty\}$, the process $\tilde{X}$ is independent of $\mathcal{F}_\tau$ and has the same law as X and hence in particular a Lévy process.

2.1.3 Motivation

The objective of the chapter is to start from a first passage time structural model with observable asset value processes modelled as a Lévy process with finite variation and an unobservable debt barrier, and show that this model can be embedded into an equivalent intensity model. The key contribution is that we show the existence of the intensity process and find its explicit form in this incomplete information framework, which sheds new light on the relationship between the intensity process of the default time and the running minimal process of the asset value. We apply the result to find the instantaneous credit spread process, which remains positive and finite, conforming to market observations, and depends on the historical path of the asset value.

With the help of the filtration expansion theory, for example, Jeulin & Yor's theorem, the essential mathematical quantity needed is the conditional survival (or default) probability under the base filtration $\mathbb{F}$: $Z_t = \mathbb{P}(\tau > t|\mathcal{F}_t)$. All results in the literature on the existence of the intensity process are based on the absolute continuity of the conditional default probability. Our challenge is that in the case of pure jump processes, the conditional default probability becomes discontinuous at the time when the asset value process reaches a new minimal and the conditional default density does not exist. This is reasonable as the expectation is that the conditional default probability jumps when there is a large movement of the asset value process. The main mathematical difficulty, unlike the continuous case in which the compensator of the conditional default probability is itself, is to find the compensator due to the unpredictability of the stopping time.

2.2 The Model and the Main Result

In this section, we introduce the model and present the main results.

The First Passage Time Structural Model Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space and $V = (V_t)_{t\geq 0}$ the firm *asset value process* given by $V_t = V_0 e^{X_t}$ at time t, where $X = (X_t)_{t\geq 0}$ is a Lévy process with finite variation and $X_0 = 0$. Examples of X include drifted subordinators, variance gamma and normal inverse Gaussian processes. Note that as a Lévy process with finite variation, X can be decomposed as ([50, Exercise 2.8])

$$X_t = ct - S_t + S'_t, \tag{2.3}$$

where $c \in \mathbb{R}$ and S, S' are independent pure jump subordinators with Lévy measures π, π', respectively, see [50, Lemma 2.14] for the definition and the properties of a subordinator. Denote by $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ the natural filtration generated by X, $\mathbb{F} = \mathbb{F}^X$. We assume the following assumption is satisfied in the paper:

Assumption 2.2.1. *Lévy measure π is continuous and satisfies $\int_0^\infty x\pi(dx) < \infty$.*

The firm defaults at the first time when the asset value falls below a default threshold, i.e., the default time τ is defined by

$$\tau := \inf\{t > 0 : V_t \leq \tilde{D}\} = \inf\{t > 0 : X_t \leq D := \ln(\tilde{D}/V_0)\},$$

where $\tilde{D}$ is an unobservable default barrier of the firm. Using the same assumption as Giesecke [34] about $\tilde{D}$:

$\tilde{D}$ is a uniform random variable on the interval $[0, V_0]$ and is independent of V.

Then the barrier D for X is a standard negative exponential variable, i.e., $-D$ is a standard exponential variable with the distribution function $\mathbb{P}(D \leq x) = e^x$ for $x < 0$, and it is independent of X.

Information Setup We assume that the asset value process X is observable, but the default barrier D is unobservable in the market. Since the default time τ is observable, we therefore define a *progressive filtration expansion* $\mathbb{G} = (\mathcal{G}_t)$ by ([59, Chapter VI, Section 3])

$$\mathcal{G}_t = \{B \in \mathcal{G} : \exists B_t \in \mathcal{F}_t, B \cap \{\tau > t\} = B_t \cap \{\tau > t\}\}. \tag{2.4}$$

The default time τ is now a $\mathbb{G}$-stopping time and we call $\mathbb{G}$ the investor filtration. All filtrations involved are assumed to satisfy the usual condition, i.e. the filtration is right continuous and complete.

Following the same notation in the introduction, denote by N the default indicator process, defined by

$$N_t := \mathbf{1}_{\{\tau \leq t\}},$$

and A as the $\mathbb{G}$-compensator of N by the Doob-Meyer decomposition. If A is absolutely continuous a.s. with respect to the Lebesgue measure and A can be written as $A_t = \int_0^t \lambda_s ds$ a.s., where λ is nonnegative and $\mathbb{G}$-progressively measurable, then λ is the intensity process of N under $\mathbb{G}$.

Main Result Denote by $\pi(x + du) := \pi((x + u, x + u + du])$. If π admits a Lévy density ν, then $\pi(x + du) = \nu(x + u)du$. We can now state the main result of the chapter.

Theorem 2.2.2. *Let X be a Lévy process of finite variation and the Lévy measure of S in the representation in (2.3) be π with Assumption 2.2.1 satisfied. Then the $\mathbb{G}$-compensator of the default indicator process N is absolutely continuous a.s. and the intensity process λ of N is indistinguishable from the instantaneous likelihood process $\tilde{\lambda}_t := \lim_{h \downarrow 0} \frac{1}{h}\mathbb{P}(t < \tau \leq t + h|\mathcal{G}_t)$ on $\{\tau > t\}$. Moreover, using the same notation as in (2.3), the intensity process λ has the following representation for all $t \geq 0$,*

$$\lambda_t = \mathbf{1}_{\{\tau > t\}} \left(-c\mathbf{1}_{\{X_t - \underline{X}_t = 0\}}\mathbf{1}_{\{c<0\}} + \Pi(X_t - \underline{X}_t)\right), \tag{2.5}$$

where $\underline{X}_t := \inf_{0 \le s \le t} X_s$ is the running minimum process of X and

$$\Pi(x) := \int_0^\infty (1 - e^{-u})\pi(x + du), \quad \forall x \ge 0. \tag{2.6}$$

Remark 2.2.3. *Note that $\Pi(\cdot)$ on $\mathbb{R}_+$ is bounded above by $\Pi(0)$ that is fully determined by the Lévy measure of X.*

Theorem 2.2.2 shows that the intensity process λ is an endogenous process that depends on the path of the asset value process X. Moreover, at each time t, λ_t is a decreasing function of $X_t - \underline{X}_t$, a financially desirable property as it means that the default intensity increases when the asset value process X approaches its historical minimal level.

We next give several examples to illustrate Theorem 2.2.2.

Example 2.2.4. *(Drifted Compound Poisson Process)* Let X be given by

$$X_t = ct - \sum_{i=1}^{M_t} Y_i + \sum_{i=1}^{M'_t} Y'_i,$$

where $c \in \mathbb{R}$, Y_i and Y'_i are exponential variables with parameters β and β', respectively, M and M' are Poisson processes with intensities ρ and ρ', respectively, and $\{Y_i\}_{i\ge1}$, $\{Y'_i\}_{i\ge1}$, M, M' are mutually independent of each other. The Lévy density of X_t on $\mathbb{R}_-$ is given by $\nu_-(x) = \rho\beta e^{-\beta x}$. The intensity process λ of the default indicator process N is then given by Theorem 2.2.2 as

$$\lambda_t \quad = \quad \mathbf{1}_{\{\tau > t\}} \left(-c\mathbf{1}_{\{X_t - \underline{X}_t = 0\}} \mathbf{1}_{\{c<0\}} + \frac{\rho}{1+\beta} e^{-\beta(X_t - \underline{X}_t)} \right).$$

Example 2.2.5. *(Drifted Gamma Process [56])* Let X be given by

$$X_t = ct - G_t,$$

where $c > 0$, G_t is a gamma process $\Gamma(t, \mu, \nu)$ with the mean rate μ, the variance rate ν, and the Lévy density $\nu(x) = \frac{\mu^2}{\nu} e^{-\frac{\mu}{\nu}x} x^{-1}$. The intensity process of N is given by

$$\lambda_t = \mathbf{1}_{\{\tau > t\}} \left(\int_0^\infty (1 - e^{-u}) \frac{\mu^2}{\nu} e^{-\frac{\mu}{\nu}(u + X_t - \underline{X}_t)} (u + X_t - \underline{X}_t)^{-1} du \right).$$

Note that $c > 0$ in this case, hence the first term in (2.5) disappears.

Example 2.2.6. *(Variance Gamma Process [55])* Let X be a variance gamma process $VG(c,\nu,\sigma,\theta)$ that is generated by a drifted Brownian motion $\theta t + \sigma W_t$, time-changed by a gamma process $\Gamma(t;1,\nu)$, and an additional drift term ct, then

$$X_t = ct + \Gamma(t;\mu_+,\nu_+) - \Gamma(t;\mu_-,\nu_-), \tag{2.7}$$

where $\mu_\pm = \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} \pm \frac{\theta}{2}$, and $\nu_\pm = \mu_\pm^2 \nu$. The intensity process of N is given by

$$\lambda_t = \mathbf{1}_{\{\tau>t\}}\left(-c\mathbf{1}_{\{X_t-\underline{X}_t=0\}}\mathbf{1}_{\{c<0\}} + \int_0^\infty (1-e^{-u})\frac{(\mu_-)^2}{\nu_-}e^{-\frac{\mu_-}{\nu_-}(u+X_t-\underline{X}_t)}(u+X_t-\underline{X}_t)^{-1}du\right). \tag{2.8}$$

Applications in Credit Modelling We next provide an application of Theorem 2.2.2 in credit risk modelling. The credit spread $S(t,h)$ of a defaultable name over the time interval $[t,t+h]$ is defined by

$$S(t,h) := -\frac{1}{h}\ln\left(1-\mathbb{P}(t<\tau\le t+h|\mathcal{G}_t)\right),$$

where $\mathbb{P}\left(t<\tau\le t+h|\mathcal{G}_t\right)$ is the conditional default probability given $\tau>t$. Using the Taylor expansion, we can find the instantaneous credit spread $s(t)$ as

$$s(t) := \lim_{h\downarrow 0} S(t,h) = \lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}\left(t<\tau\le t+h|\mathcal{G}_t\right) = \tilde{\lambda}_t.$$

Theorem 2.2.2 says that $s(t)$ is positive, finite almost surely, and is given by

$$s(t) = -c\mathbf{1}_{\{X_t-\underline{X}_t=0\}}\mathbf{1}_{\{c<0\}} + \Pi(X_t-\underline{X}_t).$$

It conforms to the market observation that the instantaneous credit spread remains positive and finite even though the bond is near its maturity and that the bond price often drops around the time of default due to uncertainties about the closeness of the current asset value to the default threshold. For more details of the instantaneous credit spread and the credit spread term structure, see [29, 34].

Numerical Illustration We next give a numerical example to illustrate the results. We take the variance gamma process $VG(c,\nu,\sigma,\theta)$ in Example 2.2.6. The data used are $(c,\nu,\sigma,\theta) = (-0.02, 0.1, 0.15, 0.01)$. Figure 2.1 displays for $t\in[0,5]$ a sample path of the asset return process X, the running minimum process $\underline{X}$, and the resulting intensity process λ. Figure 2.1 also shows the distance $X_t-\underline{X}_t$ and its contribution $\Pi(X_t - \underline{X}_t)$ to the intensity. We can observe the reciprocal relation of the intensity λ_t and the distance $X_t - \underline{X}_t$, which is consistent with the observation in the credit market. The upper bound $\Pi(0)$ is reached when $X_t - \underline{X}_t = 0$, i.e. the process X reaches a new minimal level, and the intensity λ_t at that time is above $\Pi(0)$ by the amount $|c|$ as the drift parameter $c < 0$. Figure 2.2, using the same sample path of Figure 2.1, shows the term structure of the credit spread $h \mapsto S(t, h)$ at time $t = 0.5$, starting from $S(t, 0) = \lambda_t$.

2.3 Proof of the Main Theorem

The main result Theorem 2.2.2 is proved in four steps, detailed in Subsection 2.3.1 to 2.3.4. Subsection 2.3.1 shows the relation between the likelihood processes (conditional probability process) under different filtrations (Lemma 2.3.2), Subsection 2.3.2 and 2.3.3 establish the existence of the instantaneous likelihood process for a spectrally negative Lévy process with finite variation (Proposition 2.3.8) and for a general Lévy process with finite variation (Proposition 2.3.11) respectively, and Subsection 2.3.4 confirms the indistinguishability of the instantaneous likelihood process and the intensity process using Aven's condition.

2.3.1 Compensators and Likelihood Processes under Different Filtrations

By the Doob-Meyer decomposition, there exists a unique, increasing, $\mathbb{G}$-predictable process A, the $\mathbb{G}$-compensator of N, such that the difference of A and N is a uniformly integrable $\mathbb{G}$-martingale. Moreover, the probabilistic properties of default and stopping time are closely linked to the analytic properties of the compensator. Therefore our objective is to find A.

The conditional survival probability at each time t is given by

$$Z_t := \mathbb{P}(\tau > t|\mathcal{F}_t) = \mathbb{P}(\underline{X}_t > D|\mathcal{F}_t) = e^{\underline{X}_t},$$

and let $Z_{t-} = \lim_{s \uparrow t} Z_s$ and $Z_{0-} = 1$.

The following theorem shows the representation of the compensator A under the expanded filtration $\mathbb{G}$.

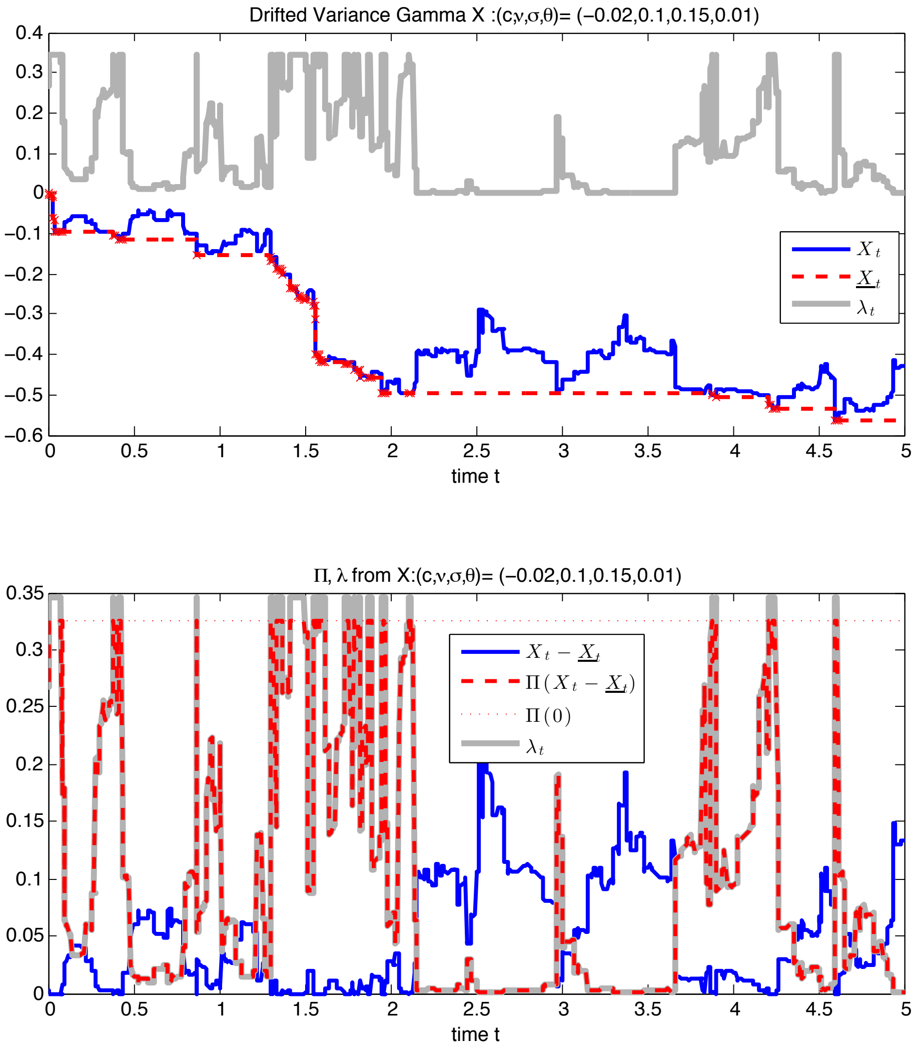


Figure 2.1: The asset return process X as in Example 2.2.6, the distance process $X - \underline{X}$ and the intensity process λ. The data used are $(c, \nu, \sigma, \theta) = (-0.02, 0.1, 0.15, 0.01)$.

Theorem 2.3.1 (Jeulin and Yor [43]). *Define a nondecreasing $\mathbb{F}$-predictable process A by*

$$A_t = \int_0^t \frac{dK_s}{Z_{s-}},$$

where K_t is the unique, increasing, $\mathbb{F}$-predictable compensator of the $\mathbb{F}$-submartingale

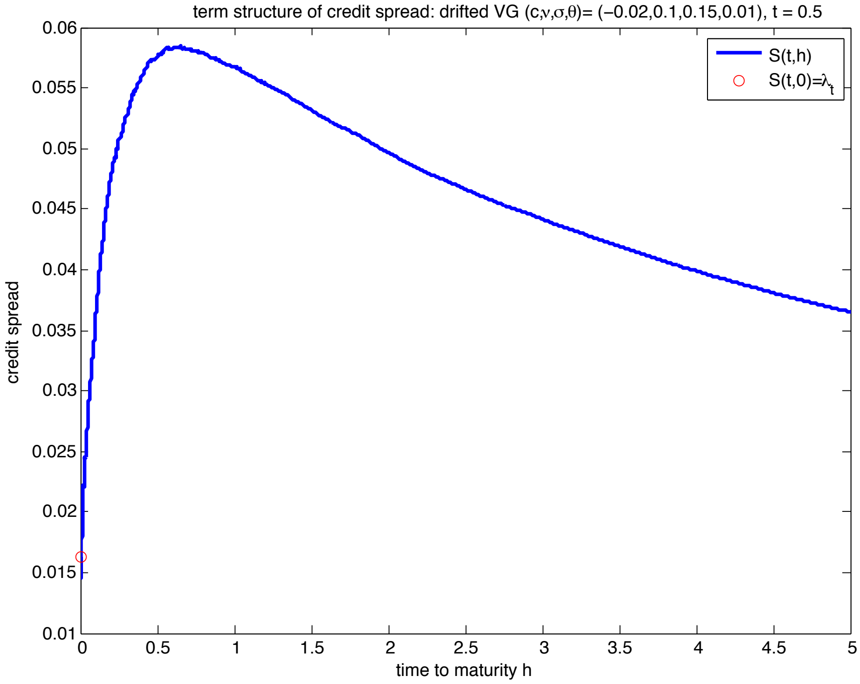


Figure 2.2: The term structure of credit spread $S(t,h)$: the asset return process X as in Example 2.2.6 with the data $(c,\nu,\sigma,\theta) = (-0.02, 0.1, 0.15, 0.01)$, $X - \underline{X}$ at $t = 0.5$ is 0.0585.

$1 - Z_t = \mathbb{P}(\tau \leq t|\mathcal{F}_t)$.

Then, the process $N - A^\tau$ is a $\mathbb{G}$-martingale, where $A^\tau = (A_{t\wedge\tau})_{t\geq 0}$.

Theorem 2.3.1 shows that one can transform the problem of finding the $\mathbb{G}$-compensator of N into the problem of finding the $\mathbb{F}$-compensator of Z. As the base filtration $\mathbb{F}$ is the natural filtration generated by the asset value process X, all we need to investigate is the path property of X.

If Z is a continuous process, then for any $t \geq 0$ we have $K_t = -Z_t$ and $A_t = -\ln(Z_t)$ is found explicitly. This is the case discussed in different setups in previous literature. If Z is discontinuous, then finding K is nontrivial, see [38]. In the following, we aim to find the representation of K.

We first show in the lemma below the pre-limit likelihood processes k^h related to K under $\mathbb{F}$ and λ^h related to A under $\mathbb{G}$.

Lemma 2.3.2. *For any Lévy process X, $h > 0$, define*

$$k_t^h \quad := \quad \frac{1}{h}\mathbb{E}\left[K_{t+h} - K_t|\mathcal{F}_t\right] \quad \text{and} \quad \lambda_t^h := \frac{1}{h}\mathbb{E}\left[N_{t+h} - N_t|\mathcal{G}_t\right].$$

Then,

$$k_t^h = e^{\underline{X}_t}\frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right]\Big|_{y=\underline{X}_t - X_t} \quad \text{and} \quad \lambda_t^h = 1_{\{\tau>t\}}e^{-\underline{X}_t}k_t^h. \tag{2.9}$$

Proof. Since

$$\begin{aligned}
\underline{X}_{t+h} - \underline{X}_t &= \inf_{u\in[t,t+h]} X_u \wedge \underline{X}_t - \underline{X}_t \\
&= \inf_{u\in[0,h]}(X_{t+u} - X_t) \wedge (\underline{X}_t - X_t) - (\underline{X}_t - X_t) \\
&= -\left((\underline{X}_t - X_t) - \inf_{u\in[0,h]}(X_{t+u} - X_t)\right)^+,
\end{aligned}$$

we have

$$\begin{aligned}
\mathbb{E}\left[e^{\underline{X}_{t+h}} - e^{\underline{X}_t}|\mathcal{F}_t\right] &= e^{\underline{X}_t}\mathbb{E}\left[e^{\underline{X}_{t+h}-\underline{X}_t} - 1|\mathcal{F}_t\right] \\
&= e^{\underline{X}_t}\mathbb{E}\left[e^{-\left((\underline{X}_t - X_t)-\inf_{u\in[0,h]}(X_{t+u}-X_t)\right)^+} - 1|\mathcal{F}_t\right] \\
&= e^{\underline{X}_t}\mathbb{E}\left[e^{-(y-\underline{X}_h)^+} - 1\right]\Big|_{y=\underline{X}_t - X_t},
\end{aligned}$$

where the last equality comes from the independent and stationary increment property of the Lévy process X and the adaptedness of X and $\underline{X}$ to $\mathbb{F}$. Since K is the $\mathbb{F}$-compensator of $1 - Z$, the Doob-Meyer decomposition says that

$$\mathbb{E}\left[K_{t+h} - K_t|\mathcal{F}_t\right] = -\mathbb{E}\left[Z_{t+h} - Z_t|\mathcal{F}_t\right] = -\mathbb{E}\left[e^{\underline{X}_{t+h}} - e^{\underline{X}_t}|\mathcal{F}_t\right].$$

Combining the above gives k_t^h in (2.9).

Next, by the optional projection theorem (Theorem 14, Chap.VI, [59] and [34]), we know that if a random variable ξ is nonnegative and integrable, then for each $t \geq 0$, the right continuous version of $\mathbb{E}[\xi|\mathcal{G}_t]$ is given by

$$\mathbb{E}[\xi|\mathcal{G}_t] = 1_{\{\tau>t\}}\frac{1}{Z_t}\mathbb{E}\left[\xi 1_{\{\tau>t\}}|\mathcal{F}_t\right] + \xi 1_{\{\tau\leq t\}} \quad a.s. \tag{2.10}$$

Therefore, using the tower property of the expectation and the fact that K is the $\mathbb{F}$-compensator of $1 - Z$, we have

$$\begin{aligned}
\lambda_t^h &= \frac{1}{h}\mathbb{E}\left[N_{t+h} - N_t|\mathcal{G}_t\right] \\
&= 1_{\{\tau>t\}}\frac{1}{h}\frac{1}{Z_t}\mathbb{E}\left[1_{\{t<\tau\leq t+h\}}|\mathcal{F}_t\right] \\
&= 1_{\{\tau>t\}}\frac{1}{h}\frac{1}{Z_t}\mathbb{E}\left[Z_t - Z_{t+h}|\mathcal{F}_t\right] \\
&= 1_{\{\tau>t\}}\frac{1}{Z_t}\frac{1}{h}\mathbb{E}\left[K_{t+h} - K_t|\mathcal{F}_t\right]. \\
&= 1_{\{\tau>t\}}e^{-\underline{X}_t}k_t^h.
\end{aligned}$$

This gives λ_t^h in (2.9). $\square$

Remark 2.3.3. *Note that $\xi 1_{\{\tau\leq t\}}$ in (2.10) is $\mathcal{G}_t$ measurable. Indeed, since ξ and τ are random variables on $(\Omega, \mathcal{G}, \mathbb{P})$, then $\xi 1_{\{\tau\leq t\}}$ is $\mathcal{G}$-measurable. To show it is $\mathcal{G}_t$ measurable, it is equivalent to show $\forall b \in \mathbb{R}, \quad B(b) := \{\omega : \xi(\omega)1_{\{\tau\leq t\}}(\omega) \leq b\} \in \mathcal{G}_t$. Note that*

$$\begin{aligned}
B(b) \cap \{t < \tau\} &= \{\xi 1_{\{\tau\leq t\}} \leq b\} \cap \{t < \tau\} \\
&= (\{\xi \leq b, \tau \leq t\} \cup \{0 \leq b, t < \tau\}) \cap \{t < \tau\} \\
&= \{0 \leq b, t < \tau\} \\
&= \{b < 0\}\emptyset + \{b \geq 0\}\{t < \tau\} \\
&= \{b < 0\}(\emptyset \cap \{t < \tau\}) + \{b \geq 0\}(\Omega \cap \{t < \tau\}).
\end{aligned}$$

Since $\emptyset, \Omega \in \mathcal{F}_t$ for all $t \geq 0$, we can take $B_t(b) := 1_{\{b<0\}}\emptyset + 1_{\{b\geq 0\}}\Omega \in \mathcal{F}_t$, such that

$$B(b) \cap \{\tau > t\} = B_t(b) \cap \{\tau > t\}.$$

Therefore, we have $B(b) \in \mathcal{G}_t$.

The next result follows immediately from Lemma 2.3.2.

Corollary 2.3.4. *Define*

$$\tilde{k}_t := \lim_{h\downarrow 0} k_t^h$$

and assume it exists for all t a.s., then the instantaneous likelihood process on $\{\tau > t\}$ is given by

$$\tilde{\lambda}_t := \lim_{h\downarrow 0}\lambda_t^h = \lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}(t < \tau \leq t + h|\mathcal{G}_t) = e^{-\underline{X}_t}\tilde{k}_t.$$

We find the limit processes $\tilde{k}_t$ and $\tilde{\lambda}_t$ in the following subsections.

2.3.2 Spectrally Negative Lévy Process with Finite Variation

By (2.9) the existence of the limit processes of k^h and λ^h depends on the path property of asset value process X and its running minimum process $\underline{X}$. In this subsection, let X be a spectrally negative Lévy process with finite variation, we find the limit process using properties of one-sided jump processes. Though the technique in this part can not be extended directly to the more general Lévy processes of finite variation that contain double-sided jumps in (2.3), the result for these specific processes will play an important role in the next subsection.

The spectrally negative process X has a representation [50, page 56]

$$X_t = ct - S_t, \tag{2.11}$$

where $c > 0$ and S is a pure jump subordinator with Lévy measure π. (2.11) is a special case of (2.3) with $\pi' = 0$ and $c > 0$. The Lévy measure of X is $\pi_X(dx) = \pi(d(-x)) = \pi((-x, -x + dx])$ on $\mathbb{R}_-$ and if π admits a density ν then $\pi(-dx) = \nu(-x)dx$. The following concept is needed in analysing the path properties of $\underline{X}$.

Definition 2.3.5 ([50]). *Let X be a Lévy process. A point $x \in \mathbb{R}$ is said to be irregular for an open or closed set B if $\mathbb{P}_x\left(\tau^B = 0\right) = 0$, where the stopping time $\tau^B = \inf\{t > 0 : X_t \in B\}$.*

We know ([27, Chapter 9, Proposition 15]) that for X defined in (2.11), 0 is irregular for $(-\infty, 0)$. Hence, starting at 0, it takes X strictly positive time to reach $(-\infty, 0)$. If we define $T_1 := \inf\{t > 0 : X_t < 0\}$, then $\mathbb{P}(T_1 > 0) = 1$. T_1 is the first jump time of $\underline{X}$ but may not be the first jump time of X. We observe that $\underline{X}$ is a pure-jump process as $\underline{X}$ can only move when S jumps and $\underline{X}$ cannot jump to a pre-specified level on $(-\infty, 0)$ as X cannot, see [50, Exercise 5.9]. Hence, the jump size of $\underline{X}$ has no atoms and is strictly negative. The number of jumps of $\underline{X}$ on the interval $[0, t]$, i.e., $n_t := \#\{s \in (0, t] : X_s = \underline{X}_s\}$, is a discrete set and is a.s. finite. Moreover, we denote the arrival times of n_t by $(T_i)_{i\geq 1}$, the inter-arrival times by $(\delta_i)_{i\geq 1}$, and the jump sizes by $(\xi_i)_{i\geq 1}$. Then we have the following lemma.

Lemma 2.3.6. *For X defined in (2.11), $\underline{X}$ can be written as a renewal-reward process*

$$\underline{X}_t = -\sum_{i=1}^{n_t} \xi_i,$$

where (δ_i, ξ_i) *are i.i.d. random variables.*

Proof. The analysis above shows that $\underline{X}$ is a non-explosive marked point process and can be written as $\underline{X}_t = -\sum_{i=1}^{n_t} \xi_i$, where $-\xi_i = \Delta \underline{X}_{T_i} = \underline{X}_{T_i} - \underline{X}_{T_{i-1}} = X_{T_i} - X_{T_{i-1}}$. $(T_i)_i$ are also jump times of Lévy process X and are stopping-times. We have that $(\delta_i, \xi_i)_i$ are i.i.d. random variables due to the strong Markov property of X stated in Proposition 2.1.12. $\square$

Instead of investigating the exact law of $\underline{X}$, we only need to analyse the small-time behaviour of the process, which can be done with the help of the next result, called the **Ballot Theorem** [11, Proposition 2.7] for drifted subordinators.

Lemma 2.3.7 (Bertoin [11]). *Let* X *be defined in (2.11) and* $T_1 := \inf\{t > 0 : X_t < 0\}$. *Then, for every* $t > 0$, $z \geq 0$, *and* $u < -z$,

$$\mathbb{P}\left(T_1 \in dt, X_{T_1 -} \in dz, \Delta X_{T_1} \in du\right) = \frac{z}{ct}\mathbb{P}(X_t \in dz)\pi(-du)dt,$$

where $\Delta X_t = X_t - X_{t-}$ *and* π *is the Lévy measure of* S.

Hence the joint distribution of (T_1, X_{T_1}) is given by

$$\mathbb{P}(T_1 \in dt, X_{T_1} \in dw) = \left(\int_{z \in (0,\infty)} z\pi(z + d(-w))\mathbb{P}(X_t \in dz)\right) \frac{1}{ct}dt \tag{2.12}$$

for $w \leq 0$. The following is another version of the Ballot theorem:

$$\mathbb{P}(T_1 > t, X_t \in dx) = \frac{x}{ct}\mathbb{P}(X_t \in dx)$$

for every $t > 0$ and $x \in [0, \infty)$. Since $X_t = ct - S_t \leq ct$, we have

$$\mathbb{P}(T_1 > t) = \frac{1}{ct}\int_0^\infty x\mathbb{P}(X_t \in dx) = \frac{1}{ct}\int_0^{ct} x\mathbb{P}(X_t \in dx) = \frac{1}{c}\mathbb{E}\left[\mathbf{1}_{\{0 \leq X_t \leq ct\}}\frac{X_t}{t}\right].$$

Note as $\lim_{t\downarrow 0} \frac{S_t}{t} = 0$ a.s., we have for almost all ω, there exists $t_0(\omega)$, such that for all $t \in [0, t_0(\omega)]$, $S_t(\omega) \leq ct$, hence $0 \leq X_t(\omega) = ct - S_t(\omega) \leq ct$ and

$$\lim_{t\downarrow 0} \mathbf{1}_{\{0 \leq X_t \leq ct\}} = 1 \quad a.s. \tag{2.13}$$

The dominated convergence theorem leads to

$$\lim_{t\downarrow 0} \mathbb{P}(T_1 > t) = \frac{1}{c}\mathbb{E}\left[\lim_{t\downarrow 0} \mathbf{1}_{\{0 \leq X_t \leq ct\}}\frac{X_t}{t}\right] = \frac{1}{c} \cdot c = 1. \tag{2.14}$$

Proposition 2.3.8. *Let X be defined in (2.11) and let Assumption 2.2.1 be satisfied. Then the following limit exists for all t a.s.*

$$\tilde{k}_t := \lim_{h\downarrow 0} k_t^h = e^{\underline{X}_t}\Pi(X_t - \underline{X}_t)$$

where k_t^h is defined in (2.9) for $h > 0$ and Π is defined in (2.6). The instantaneous likelihood process $\tilde{\lambda}_t$ defined in Corollary 2.3.4 is given by

$$\tilde{\lambda}_t := \lim_{h\downarrow 0} \lambda_t^h = \Pi(X_t - \underline{X}_t).$$

Proof. Recall that $\underline{X}_t = -\sum_{i=1}^{n_t} \xi_i$ is a renewal-reward process, where the jump size ξ_i and inter-arrival times δ_i are positive random variables for all i, and (δ_i, ξ_i) are i.i.d. random variables.

By (2.14), denote by F the distribution function of δ_i. Then

$$\lim_{t\downarrow 0} F(t) = \lim_{t\downarrow 0} P(T_1 \le t) = 0 = F(0).$$

Hence, F is right continuous at zero, i.e. $F(0) = F(0+) = 0$.

Define by, for $t > 0$ and $y \le 0$,

$$\begin{aligned} \Lambda_t^0(y) &:= \frac{1}{t}\mathbb{E}\left[1 - e^{-(y-\underline{X}_t)^+}\right] \\ \Lambda_t^1(y) &:= \frac{1}{t}\mathbb{E}\left[1_{\{n_t=1\}}(1 - e^{-(\xi_1+y)^+})\right] \\ \Lambda_t^2(y) &:= \frac{1}{t}\mathbb{E}\left[\sum_{k=2}^{\infty} 1_{\{n_t=k\}}\left(1 - e^{-\left(\sum_{i=1}^k \xi_i + y\right)^+}\right)\right]. \end{aligned}$$

We have

$$\Lambda_t^0(y) = \frac{1}{t}\mathbb{E}\left[1 - e^{-\left(y+\sum_{i=1}^{n_t}\xi_i\right)^+}\right] = \Lambda_t^1(y) + \Lambda_t^2(y).$$

We next show that for $y \le 0$,

$$\lim_{t\to 0} \Lambda_t^1(y) = \Pi(-y), \text{ and } \lim_{t\to 0} \Lambda_t^2(y) = 0. \tag{2.15}$$

Then, (2.15) gives the required conclusion.

Since δ_1 and δ_2 are independent, also noting (2.12), we have

$$\begin{aligned}
& \mathbb{E}\left[1_{\{n_t=1\}}(1-e^{-(\xi_1+y)^+})\right] \qquad (2.16)\\
= & \mathbb{E}\left[1_{\{T_1\le t\}}1_{\{T_2-T_1>t-T_1\}}(1-e^{-(\xi_1+y)^+})\right]\\
= & \int_0^t\int_{-y}^\infty \bar{F}_{T_1}(t-s)(1-e^{-x-y})\mathbb{P}(T_1\in ds, X_{T_1}\in d(-x))\\
= & \int_{s=0}^t\int_{x=-y}^\infty \bar{F}_{T_1}(t-s)(1-e^{-x-y})\left(\int_{z=0}^\infty z\pi(z+dx)\mathbb{P}(X_s\in dz)\right)\frac{1}{cs}ds\\
= & \frac{1}{c}\int_{s=0}^t \bar{F}_{T_1}(t-s)\int_{z=0}^\infty\left(\int_{u=0}^\infty(1-e^{-u})\pi(z-y+du)\right)z\frac{\mathbb{P}(X_s\in dz)}{s}ds\\
= & \frac{1}{c}\int_{s=0}^t \bar{F}_{T_1}(t-s)\left(\int_{z=0}^{cs}\Pi(z-y)z\frac{\mathbb{P}(X_s\in dz)}{s}\right)ds.
\end{aligned}$$

The last equality is due to $X_s = cs - S_s \le cs$.

Since S is a pure jump subordinator, we have ([50, Lemma 4.11]) $\lim_{t\to 0}\frac{S_t}{t} = 0$ a.s., which implies

$$\lim_{t\to 0}\frac{X_t}{t} = c. \qquad (2.17)$$

Using (2.17) and (2.13), the dominated convergence theorem, continuity of $\Pi(\cdot)$, and $X_{0+}=0$, we obtain

$$\begin{aligned}
\lim_{s\downarrow 0}\int_{z=0}^{cs} z\Pi(z-y)\frac{\mathbb{P}(X_s\in dz)}{s} &= \lim_{s\downarrow 0}\mathbb{E}\left[1_{\{0\le X_s\le cs\}}\frac{X_s}{s}\Pi(X_s-y)\right]\\
&= \mathbb{E}\left[\lim_{s\downarrow 0}1_{\{0\le X_s\le cs\}}\frac{X_s}{s}\Pi(X_s-y)\right]\\
&= c\Pi(-y).
\end{aligned}$$

Taking the limit in (2.16) gives

$$\lim_{t\downarrow 0}\Lambda_t^1(y) = \frac{1}{c}\lim_{s\downarrow 0}\int_{z=0}^{cs} z\Pi(z-y)\frac{\mathbb{P}(X_s\in dz)}{s} = \Pi(-y).$$

Here we have used the fact that if $g(\cdot)$ is a nonnegative function and $\bar{F}(0+)=1$, then

$$\frac{1}{t}\int_0^t \bar{F}(t)g(s)ds \le \frac{1}{t}\int_0^t \bar{F}(t-s)g(s)ds \le \frac{1}{t}\int_0^t g(s)ds$$

and

$$\lim_{t\to 0}\frac{1}{t}\int_0^t \bar{F}(t-s)g(s)ds = \lim_{t\to 0}\frac{1}{t}\int_0^t g(s)ds = g(0+).$$

We have proved the first limit in (2.15). We next prove the second limit in (2.15). Since

$$\Lambda_t^0(0) = \frac{1}{t}\mathbb{E}\left[1 - e^{\underline{X}_t}\right] \le \frac{1}{t}\mathbb{E}\left[1 - e^{-S_t}\right] = \frac{1}{t}\left(1 - e^{-\Pi(0)t}\right) \le \Pi(0)$$

and

$$0 \le \Lambda_t^1(0) \le \Lambda_t^0(0) \le \Pi(0),$$

the first limit in (2.15) implies

$$\lim_{t\to 0} \Lambda_t^0(0) = \lim_{t\to 0} \Lambda_t^1(0) = \Pi(0),$$

therefore

$$\lim_{t\to 0} \Lambda_t^2(0) = 0.$$

On the other hand, we know

$$0 \le \Lambda_t^2(y) \le \Lambda_t^2(0) \quad \text{for all } y \le 0,$$

which proves the second limit in (2.15). Hence, $\Lambda_t^0(y) = \Pi(-y)$ and $\tilde{\lambda}_t = \Lambda_t^0(X_t - \underline{X}_t) = \Pi(X_t - \underline{X}_t)$. $\square$

Remark 2.3.9. Note that on $\mathbb{R}_+$, $\Pi(\cdot)$ is continuous as $\pi(dx)$ is and it is decreasing with the upper bound $\Pi(0) = -\ln \mathbb{E}[e^{-S_1}]$ being the Laplace exponent of S from the Lévy-Khintchine formula. $0 < \Pi(x) \le \Pi(0) < \int_0^\infty u\pi(du) < \infty$ for all $x \ge 0$ by Assumption 2.2.1, and therefore Π is bounded on $\mathbb{R}_+$.

2.3.3 Lévy Process with Finite Variation

In this subsection we investigate the limit processes $\tilde{k}$ and $\tilde{\lambda}$ with X being a Lévy process with finite variation which is a more general class of processes than the last subsection. X then has a representation (2.3)

$$X_t = ct - S_t + S_t', \quad t \ge 0$$

where $c \in \mathbb{R}$, S has the Lévy measure π, and we assume Assumption 2.2.1 hold. Note that the path properties and techniques used in Subsection 2.3.2 no longer hold. In (2.3), denote the drift and negative jump components as

$$Z_t(c) := ct - S_t.$$

Then we first claim the following result for $Z_t(c)$.

Lemma 2.3.10. *For any $c \in \mathbb{R}$ and $y \leq 0$ the following limit exists:*

$$\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[1-e^{-(y-\underline{Z}_h(c))^+}\right] \quad = \quad -c\mathbf{1}_{\{y=0\}}\mathbf{1}_{\{c<0\}}+\Pi(-y), \tag{2.18}$$

where Π is defined in (2.6).

Proof. For $c > 0$ the limit (2.18) has been proved in the previous subsection. We now consider the case of $c \leq 0$. Note that $Z_h(c)$ is decreasing in h and $\underline{Z}_h(c) = Z_h(c)$. We split the proof into two cases.

(i) $y = 0$: We have

$$\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[1-e^{-(y-\underline{Z}_h(c))^+}\right] = \lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[1-e^{ch-S_h}\right] \quad = \quad -c+\Pi(0).$$

(ii) $y < 0$: Take the function $f(x) := 1-e^{-(y+x)^+}$ which is bounded, continuous, and vanishes in a neighbourhood of zero: take $\epsilon < -y$, then for any $x \in (0,\epsilon)$, we have $f(x) = 0$. Hence

$$\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[1-e^{-(y-\underline{Z}_h(c))^+}\right] \quad = \quad \lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[f(-Z_h(c))\right] = \int_{\mathbb{R}} f(x)\pi(dx).$$

The second equality is due to [60, Corollary 8.9] and π being the Lévy measure of $-Z_h(c)$. Therefore,

$$\int_{\mathbb{R}} f(x)\pi(dx) \quad = \quad \int_{-y}^{\infty}(1-e^{-(y+x)})\pi(dx) = \int_0^{\infty}(1-e^{-u})\pi(-y+du) = \Pi(-y),$$

which proves (2.18). $\square$

We present the main result in this subsection in the following proposition.

Proposition 2.3.11. *Let X_t be defined in (2.3) and let Assumption 2.2.1 be satisfied. Then the following limit exists for all t a.s.*

$$\tilde{k}_t := \lim_{h\downarrow 0} k_t^h = e^{\underline{X}_t}\left(-c\mathbf{1}_{\{X_t-\underline{X}_t=0\}}\mathbf{1}_{\{c<0\}}+\Pi(X_t-\underline{X}_t)\right), \tag{2.19}$$

where k_t^h is defined in (2.9) for $h > 0$ and Π is defined in (2.6). The instantaneous likelihood process $\tilde{\lambda}_t$ defined in Corollary 2.3.4 is given by

$$\tilde{\lambda}_t = -c\mathbf{1}_{\{X_t-\underline{X}_t=0\}}\mathbf{1}_{\{c<0\}}+\Pi(X_t-\underline{X}_t). \tag{2.20}$$

Proof. The expression of $\tilde{\lambda}_t$ in (2.20) is an immediate result of Corollary 2.3.4 and (2.19). To prove (2.19) we only need to show that for all $y \le 0$,

$$\lim_{h\to 0} \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] = -c\mathbf{1}_{\{y=0\}}\mathbf{1}_{\{c<0\}} + \Pi(-y). \tag{2.21}$$

Since $f(x) = 1 - e^{-(y-x)^+}$ is a decreasing function of x on $\mathbb{R}_-$ and $X_h = ch - S_h + S'_h \ge ch - S_h = Z_h(c)$ for all ω and $h > 0$, we have

$$\underline{X}_h \ge \underline{Z}_h(c)$$

and

$$\frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] \quad \le \quad \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{Z}_h(c))^+}\right].$$

Using Lemma 2.3.10 we obtain

$$\limsup_{h\to 0} \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] \le -c\mathbf{1}_{\{y=0\}}\mathbf{1}_{\{c<0\}} + \Pi(-y). \tag{2.22}$$

Take any $\epsilon > 0$, on the set $\{S'_h \le \epsilon h\}$ we have:

$$X_h = ch - S_h + S'_h \le ch - S_h + \epsilon h = Z_h(c+\epsilon),$$

which yields

$$\underline{X}_h \le \underline{Z}_h(c+\epsilon) \quad \text{on } \{S'_h \le \epsilon h\}.$$

Moreover, as $\underline{X}_h \le 0$ for all $h \ge 0$, we have almost surely,

$$\underline{X}_h = 1_{\{S'_h \le \epsilon h\}}\underline{X}_h + 1_{\{S'_h > \epsilon h\}}\underline{X}_h \le 1_{\{S'_h \le \epsilon h\}}\underline{X}_h \le 1_{\{S'_h \le \epsilon h\}}\underline{Z}_h(c+\epsilon).$$

Therefore we have

$$\begin{aligned}
\frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] &\ge \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-1_{\{S'_h \le \epsilon h\}}\underline{Z}_h(c+\epsilon))^+}\right] \\
&= \frac{1}{h}\mathbb{E}\left[1_{\{S'_h \le \epsilon h\}}\left(1 - e^{-(y-\underline{Z}_h(c+\epsilon))^+}\right)\right] \\
&= \mathbb{P}\left(\frac{S'_h}{h} \le \epsilon\right)\frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{Z}_h(c+\epsilon)^+}\right].
\end{aligned}$$

The last equality is due to the independence of S and S'. Since $\lim_{h\to 0}\frac{S'_h}{h} = 0$ a.s., which implies $\lim_{h\to 0}\mathbb{P}\left(\frac{S'_h}{h} \le \epsilon\right) = 1$, we have

$$\liminf_{h\to 0} \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] \ge -(c+\epsilon)\mathbf{1}_{\{y=0\}}\mathbf{1}_{\{c+\epsilon<0\}} + \Pi(-y).$$

Let $\epsilon \downarrow 0$ in the above inequality, we obtain

$$\liminf_{h\to 0} \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right] \geq -c\mathbf{1}_{\{y=0\}}\mathbf{1}_{\{c<0\}} + \Pi(-y),$$

and together with (2.22) we proved (2.21). $\square$

Remark 2.3.12. Note that if X is a Lévy process with a Lévy measure π_X and f is a bounded continuous function that vanishes in a neighbourhood of zero, then ([60, Corollary 8.9])

$$\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[f(X_h)] = \int_{\mathbb{R}} f(x)\pi_X(dx). \tag{2.23}$$

In our case, we aim to compute

$$\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right].$$

However, we cannot apply (2.23) directly as $\underline{X}$ is not a Lévy process if X is not a monotone process. Proposition 2.3.11 can be viewed as an extension for function $f(x) = 1 - e^{-(y-x)^+}$ and Lévy process X with finite variation.

2.3.4 Indistinguishability of Likelihood Process and Intensity Process

We have proved the existence of the instantaneous likelihood process $\tilde{\lambda}$ when X is a Lévy process with finite variation. Heuristically the intensity process λ of the $\mathbb{G}$-compensator should be equal to $\tilde{\lambda}$ on the set $\{\tau > t\}$. However, they are not necessarily the same.

Example 2.3.13 (Guo and Zeng [37])**.** Define a stopping time $\tau := \inf\{t > 0 : W_t > y\}$ where W is a Brownian motion and $y > 0$ is a constant. Suppose $\mathbb{F}$ is the natural filtration of W. We have τ is a $\mathbb{F}$-stopping time and

$$\frac{1}{h}\mathbb{P}(t < \tau < t+h|\mathcal{F}_t) \quad = \quad \frac{1_{\{\tau>t\}}}{h}\int_0^h \frac{|W_t - y|}{\sqrt{2\pi t^3}}e^{-\frac{(W_t-y)^2}{2t}}dt \xrightarrow{h\downarrow 0} 0.$$

That is, $\tilde{\lambda}_t \equiv 0$ for all $t \geq 0$. As τ is predictable under $\mathbb{F}$, the compensator of $N_t = 1_{\{\tau\leq t\}}$ is N_t, which indicates the intensity λ does not exist.

Aven's condition in the next lemma provides a sufficient condition that ensures $\tilde{\lambda}$ and λ are indistinguishable.

Lemma 2.3.14 (Aven's condition [6]). *If $\lim_{h\to 0}\lambda_t^h = \tilde{\lambda}_t$ exists and λ_t^h is uniformly bounded for $t > 0$ and $h > 0$ a.s., then on $\{\tau > t\}$, $N_t - \int_0^t \tilde{\lambda}_s ds$ is a $\mathbb{G}$-martingale, i.e., $\int_0^t \tilde{\lambda}_s ds$ is the $\mathbb{G}$-compensator of N.*

With the help of the results of previous subsections, we can now present the proof of the main theorem.

Proof of Theorem 2.2.2. Recall (2.9) that on $\{\tau > t\}$

$$\lambda_t^h = e^{-\underline{X}_t} k_t^h = \frac{1}{h}\mathbb{E}\left[1 - e^{-(y-\underline{X}_h)^+}\right]\Big|_{y=\underline{X}_t - X_t}$$

and $y \leq 0$, $\underline{X}_t \leq 0$ and $c \geq c \wedge 0$, which implies $(y - \underline{X}_h)^+ \leq -\underline{X}_h$ and $\underline{X}_h \geq (c \wedge 0)h - S_h$, we have

$$\begin{aligned}
\lambda_t^h &\leq \frac{1}{h}\mathbb{E}\left[1 - e^{\underline{X}_h}\right] \\
&\leq \frac{1}{h}\mathbb{E}\left[1 - e^{(c\wedge 0)h - S_h}\right] \\
&= \frac{1}{h}\left(1 - e^{(c\wedge 0)h - \Pi(0)h}\right) \\
&\leq -(c \wedge 0) + \Pi(0).
\end{aligned}$$

Hence the sequence $(\lambda_t^h)_{h>0}$ is uniformly bounded in t and h a.s., Lemma 2.3.14 gives the required conclusion that λ and $\tilde{\lambda}$ are indistinguishable on $\{\tau > t\}$, which leads to the expression of λ from Proposition 2.3.11. The proof of Theorem 2.2.2 is now complete. □

Similarly, $k_t^h = e^{\underline{X}_t}\lambda_t^h \leq \lambda_t^h$ is also bounded and with a similar argument as Aven's condition due to the Meyer's Laplacian approximation theorem, we can conclude that the $\mathbb{F}$-compensator of $\mathbb{P}(\tau \leq t|\mathcal{F}_t)$ is $K_t = \int_0^t \tilde{k}_s ds$ where $\tilde{k}_t = \lim_{h\to 0} k_t^h$.

2.4 Conclusion

In this chapter we discussed the intensity problem of the first passage time of a finite variation Lévy process with a random barrier. We proved the existence of the intensity process and found its explicit representation. We computed the instantaneous credit spread process explicitly and gave a numerical example for a variance gamma process to illustrate the relationship between the credit spread and the distance of the asset value to its running minimum value. We thus reconciled the structural model with incomplete information and the path-dependent intensity model in this setup.

Chapter 3

Introduction to Bivariate Dynamic Contagion Processes

In this chapter we introduce the basics of point processes and in particular the Bivariate Dynamic Contagion process (BDCP) that will be investigated in Chapter 4 and Chapter 5.

We first briefly review point process modelling in the literature in Section 3.1, then we introduce basics of point processes and Markov processes in Section 3.2 and BDCPs in Section 3.3. In Section 3.4 we explore a sufficient condition of existence of moments for BDCPs. With an exact simulation algorithm developed in Section 3.5, we finally illustrate the dynamics of the BDCP in Section 3.6.

3.1 Introduction to Point Process Modelling

Multivariate point processes are used to model event arrivals of various types in a system. There are many potential applications; stochastic models are needed for events such as company bankruptcies, insurance claim arrivals, disease incidence, machine failures and so on. How to model a point process that is able to describe a rich dynamic and dependence structure becomes an essential issue.

In order to reflect both the external impact and the internal contagion effect, in this chapter we introduce the Bivariate Dynamic Contagion Process (BDCP). The BDCP family is a broad family of bivariate point processes with intensity processes specified as non-diffusion Piecewise Deterministic Markov Processes (PDPs) (studied by Davis [26]) and the processes incorporate a feedback (contagion) mechanism within the system. The BDCP covers two distinct important classes of point processes. The first class consists of shot-noise processes that usually describe point process systems under the impact from external events modelled as compound Poisson processes in intensity processes. They are studied by Cox and Isham [19], Møller [58], Dassios and Jang [21], and Klüppelberg and Mikosch [46], among others. The class has a wide range of applications. For instance, it has been adopted in modelling insurance claim arrivals and ruin probabilities by Altmann et al.[3], Albrecher and Asmussen [2], and Macci and Torrisi [54]. The second class consists of Hawkes processes, which have mutually-exciting intensities. In this class, jumps in the point process bring the internal feedback into the underlying intensity process and the impact factor is modelled by upward jumps with constant sizes. This class is capable of modelling clustering and contagion effects. Hawkes processes are introduced by Hawkes and Oakes [39], and they are studied by Daley and Vere-Jones [20], and Liniger [52] and Embrechts et al. [31]. Recently, with the desirable self-exciting property, Hawkes processes are extensively applied in modelling insurance claims, defaults in a credit portfolio, and arrivals of trading orders in a limit order book. One can find more details by the following: Hautsch [10], Aït-Sahalia et al. [1], Bacry et al. [8], Errais et al. [32], and Cont and Larrard [17].

Moreover Dassios and Zhao [23] introduce Univariate Dynamic Contagion Processes (UDCP) to include impact from both external and internal factors. Applications in credit risk and insurance modelling can be found in Dassios and Zhao [63] and [24]. However, in practice, a univariate model is not sufficient to model a heterogeneous population, and we need to introduce a bivariate system. The dependency between marginals can be characterized in a few ways. Dassios and Jang [41] studied a bivariate system by correlating external factors of two UDCPs. In order to include the cross-exciting contagion effect, the BDCP is naturally introduced here. Note that the cross-exciting dependency introduces a loop structure that makes the system difficult to be decoupled. Hence, it is fundamentally different from the univariate model and more difficult to analyse. Moreover, a BDCP is also different from a bivariate Hawkes process since it is not obvious how the additional external factors modelled by a shot noise component and random jump sizes in the intensity would affect the probabilistic properties of the system.

In principle, applications of Hawkes or shot-noise processes can be extended using BDCP to incorporate a richer structure. Moreover, the general multivariate dynamic contagion process can be easily extended from the bivariate case.

We first introduce the basics of point processes in the following section before introducing the BDCP.

3.2 Introduction of Basics

3.2.1 Basics of Point Processes

Point Processes Following [13], we define point processes on the real half line $\mathbb{R}_+$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.2.1 (Point Processes [13]). *A univariate point process is a sequence of random variables* $\{T_n\}_{n\geq 1}$ *in* $\mathbb{R}_+$ *such that* $T_0 = 0$, $T_n < T_{n+1}$ *if* $T_n < \infty$. *The associated counting process* $N = (N_t)_{t\geq 0}$ *is defined by*

$$N_t = \sum_{n\geq 1} 1_{\{T_n \leq t\}}.$$

The point process is non-explosive if $\mathbb{P}-a.s.$, $T_\infty = \lim_{n\to\infty} T_n = \infty$, *or equivalently* $N_t < \infty$ *for all* $t \geq 0$.

Let $\{T_n\}_{n\geq 1}$ *be a point process and* $\{Z_n\}_{n\geq 1}$ *be a sequence of* $\{1, 2, \ldots, k\}$*-valued random variables also defined on* $(\Omega, \mathcal{F}, \mathbb{P})$, *and define for all* i, $1 \leq i \leq k$ *and l* $t \geq 0$:

$$N_t^i = \sum_{n\geq 1} 1_{\{T_n \leq t\}} 1_{\{Z_n = i\}},$$

then the k*-vector process* $N = (N_t)_{t\geq 0}$ *given by* $N_t = \left(N_t^1, \ldots, N_t^k\right)$ *is called a* k*-variate point process.*

We define the predictable intensity by Watanabe's characterization as in [13].

Definition 3.2.2 (Stochastic Intensity [13]). *Let* $N = (N_t)_{t\geq 0}$ *be a point process adapted to a filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ *with jump times* $\{T_n\}_{n\geq 1}$, *and let* $\lambda = (\lambda_t)_{t\geq 0}$ *be a non-negative predictable process under* $\mathbb{F}$ *such that* $\int_0^t \lambda_s ds < \infty$ $\mathbb{P}$*-a.s. for all* $t \geq 0$ *and*

$$N_{t\wedge T_n} - \int_0^{t\wedge T_n} \lambda_s ds \quad \textit{is an } \mathbb{F}\textit{-martingale,} \tag{3.1}$$

then we say that N admits an $\mathbb{F}$-stochastic intensity λ.

Moreover, such intensity λ is unique.

In the definition above, the condition (3.1) can be replaced by

$$\mathbb{E}\left[\int_0^\infty C_s dN_s\right] = \mathbb{E}\left[\int_0^\infty C_s \lambda_s ds\right] \tag{3.2}$$

for all non-negative $\mathbb{F}$-predictable process $C = (C_t)_{t\geq 0}$.

Remark 3.2.3. *Equivalently, by the Doob-Meyer decomposition, there exists a unique non-decreasing predictable process A starting at 0, such that $N - A$ is an $\mathbb{F}$-local martingale, then A is called the compensator of N. If A is absolutely continuous, i.e., there exists a non-negative, $\mathbb{F}$-predictable and integrable process λ, s.t. for every $t \geq 0$, $A_t = \int_0^t \lambda_s ds$ a.s., then λ is called the intensity process of N.*

If the intensity process exists, then it can be interpreted as the instantaneous jump rate of N:

$$\lambda_t = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{P}(N_{t+h} - N_t = 1|\mathcal{F}_t),$$

or equivalently $\mathbb{P}(N_{t+h} - N_t = 1|\mathcal{F}_t) = \lambda_t h + o(h)$ for h small.

As a special case, if the intensity λ is taken to be $\mathcal{F}_0$-measurable in Definition 3.2.2, then the point process N is called a *Cox process* or *doubly stochastic Poisson process*. Therefore, we have the following special cases in this family:

(1) $\lambda_t = \lambda(t)$ is a deterministic function: N is a homogeneous Poisson process if λ is a constant and inhomogeneous Poisson process if $\lambda(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-negative deterministic function.

(2) $\lambda_t = \Lambda$ where Λ is an $\mathcal{F}_0$-measurable random variable: N is a mixed Poisson process.

(3) $\lambda_t = f(t, Y_t)$ for some appropriately measurable nonnegative function f and for some measurable process $Y = (Y_t)_{t\geq 0}$ with $\mathcal{F}^Y_\infty \subset \mathcal{F}_0$: N is called a doubly stochastic Poisson process driven by Y. Conditional on $\mathcal{F}^Y_\infty$, N is an inhomogeneous Poisson process.

Example 3.2.4 (Shot noise Cox Processes)**.** *The shot-noise Cox process is a counting process $N = (N_t)_{t\geq 0}$ with intensity $\lambda = (\lambda_t)_{t\geq 0}$ specified as an inhomogeneous* *compound Poisson process:*

$$\lambda_t = \lambda_0 + \sum_{S_i<t} \gamma(t - S_i) Y_i$$

where $\gamma(\cdot)$ is a non-negative function defined on $\mathbb{R}_+$ with $\gamma(0) = 1$, $\{S_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ are the sequence of jump times and sizes of a Poisson process M.

In applications, shot noise processes are used to describe events that occur in a system under the influence of an external factor characterized by a compound Poisson process with impact decay with respect to time.

Now we introduce another family of point processes with stochastic intensity important in applications. We start with cluster processes and then introduce Hawkes processes that can be interpreted in terms of cluster processes and age-dependent branching processes.

Cluster Processes [61] The cluster mechanism is a natural way to describe the locations of individuals from consecutive generations of a branching process.

A *cluster process* N are defined in terms of two components:

- cluster centre process $N_c(\cdot)$,
- cluster component process $N(\cdot|t)$ when the centre is at t.

Each point of $N_c(\cdot)$ is assumed to initiate a cluster, generally called a cluster, independently for each point, i.e., the clusters are independently and identically distributed point processes. The cluster process then consists of the superposition of all the clusters. For every bounded measurable set A on $\mathbb{R}_+$,

$$N(A) = \int_{\mathbb{R}_+} N(A|y) N_c(dy) = \sum_{y_i \in N_c(\cdot)} N(A|y_i) < \infty \quad a.s.$$

We call N a *Poisson cluster process* if the cluster centres are points of a Poisson process. For example, the Neyman-Scott process is a cluster process where the cluster centre is a Poisson process and members in a cluster are independently and identically distributed about the cluster centre with a certain distribution.

Hawkes Processes [39] The one-dimensional Hawkes process $N = (N_t)_{t\in\mathbb{R}}$ is a point process defined on the real line $\mathbb{R}$ with intensity $\lambda = (\lambda_t)_{t\in\mathbb{R}}$ specified as

$$\lambda_t = \lambda_0 + \int_{-\infty}^{t} \gamma(t-u)dN_u, \tag{3.3}$$

where $\lambda > 0$, $\gamma(x) \geq 0$ for $x \geq 0$. Note that the function γ is usually taken to be a decreasing function vanishing at infinity which characterize the impact of historical jumps of N at jump times and the decay of their influence. It is also called a *self-exciting point process.*

Moreover, according to [61] and [39], the Hawkes process in (3.3) has a cluster process representation and can be seen as an age-dependent birth process. The cluster centres N_c is a Poisson process of rate λ_0 formed by arrivals of immigrants. For each event of N_c, we have a cluster formed by all the descendants of all generations of immigrants. The clusters are mutually independent by construction. In each cluster, any individual of age x at time t from all generations has the probability of a birth in $(t, t+dt)$ of $\gamma(x)dt$.

It is concluded in [39] that under the condition

$$0 < \int_{0}^{\infty} \gamma(x)dx < 1, \tag{3.4}$$

there exists a unique stationary cluster process with the intensity (3.3).

Note that if one takes the function $\gamma(x) = \alpha e^{-\delta x}$ with $\alpha > 0$ and $\delta > 0$, then λ is a Markov process and the process N is also called a Markovian self-exciting process. The sufficient stationarity condition in (3.4) becomes $\alpha < \delta$ in this case.

The cluster process representation of a multivariate Hawkes processes can be extended and we refer to [52] for details.

Remark 3.2.5. *In this thesis, we only discuss point processes $N = (N_t)_{t\geq 0}$ that are defined on the nonnegative real line $\mathbb{R}_+$ with a Markov intensity process.*

3.2.2 Markov Processes

Martingale Problem We follow Ethier and Kurtz [33] to recall some definitions and properties of Markov processes.

First recall that the inifinitesimal generator of a semigroup $\{T(t)\}$ on a Banach space L is the linear operator $\mathcal{A}$ defined by

$$\mathcal{A}f = \lim_{t\to 0} \frac{1}{t}\{T(t)f - f\}$$

in the norm topology, i.e.

$$\left\| \mathcal{A}f - \frac{T(t)f - f}{t} \right\| \to \infty \quad \text{as } t \downarrow 0.$$

The domain $\mathcal{D}(\mathcal{A})$ of $\mathcal{A}$ is the subspace of all $f \in L$ for which this limit exists.

Suppose there is a process $X = (X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define a process $M = (M_t)_{t\geq 0}$ by

$$M_t^f := f(X_t) - \int_0^t \mathcal{A}f(X_s)ds. \tag{3.5}$$

Proposition 3.2.6. *If X is a Markov process with generator $\mathcal{A}$, then M_t^f in (3.5) is a $(P, \mathcal{F}_t^X)$-martingale where $\mathcal{F}_t^X = \sigma\{X_s : s \leq t\}$.*

Definition 3.2.7 (Solution of the Martingale Problem)**.** *A probability measure $P \in \mathcal{P}(D_E[0,\infty))$ is a solution of the martingale problem for $(\mathcal{A}, \mu)$ if the coordinate process defined on $(D_E[0,\infty), \mathcal{D}, P)$ by $X(t,\omega) = \omega(t)$ with $\omega \in D_E[0,\infty)$ and $t \geq 0$ is a solution, such that M^f defined in (3.5) is a $\mathcal{F}_t^X$-martingale for all $f \in \mathcal{D}(\mathcal{A})$ and $\mathbb{P}X_0^{-1} = \mu$.*

Note that the martingale problem will be used to characterize Markov processes and compute the moments. In the following, we introduce an important class of non-diffusive Markov processes.

Piecewise Deterministic Markov Processes (PDP) [26] As indicated by the name, PDPs are a general family of Markovian non-diffusive models and they can be analysed by methods analogous to those of the diffusion theory. A PDP $X = (X_t)_{t\geq 0}$ takes values on E an open subset of $\mathbb{R}^d$ with boundary ∂E and Borel σ-field $\mathcal{E}$. The PDP is determined by a flow function characterising the deterministic move between jumps, a jump rate describing the arrival of random jumps and a transition measure describing the transition of the state after jumps. The UDCP intensity defined in the following is an example of PDP.

Univariate Dynamic Contagion Processes (UDCP) [23] A univariate dynamic contagion process is a point process $N = (N_t)_{t\geq 0}$ with intensity $\lambda = (\lambda_t)_{t\geq 0}$ specified as a PDP:

$$\lambda_t = \lambda_0 e^{-\delta t} + \sum_{S_i < t} Y_i e^{-\delta(t-S_i)} + \sum_{T_i < t} Z_i e^{-\delta(t-T_i)}, \tag{3.6}$$

where $\{S_i\}_{j\geq 1}$ are jump times of a Poisson process M and $\{T_i\}_{i\geq 1}$ are jump times of N, $\{Y_i\}_{i\geq 1}$ and $\{Z_i\}_{i\geq 1}$ are non-negative i.i.d. random variables and independent of jump times. Therefore, a UDCP can be seen as a generalized Hawkes process by randomizing jump sizes and including a shot noise component in the intensity. We recover a shot noise process by setting $Z_i \equiv 0$ and a Hawkes process by setting $Y_i \equiv 0$ and Z_i constant for all $i \geq 1$.

In the following chapters, we introduce the bivariate dynamic contagion processes by keeping the structure of the UDCP and adding an additional cross-exciting component to each marginal process.

3.3 The BDCP Model

3.3.1 Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, on which we introduce the Bivariate Dynamic Contagion Processes (BDCP) that is a class of $\mathbb{F}$-adapted bivariate point processes $N = (N_t)_{t\geq 0}$ defined on $\mathbb{R}_+$ with

$$N_t = \left(N_t^1, N_t^2\right) = \left(\sum_{n\geq 1} 1_{\{T_n^1 \leq t\}}, \sum_{n\geq 1} 1_{\{T_n^2 \leq t\}}\right) \tag{3.7}$$

where $\{T_n^k\}_{n\geq 0}$ are orderly $\mathbb{F}$-stopping times representing event arrival times with $T_0^k = 0$ for $k = 1, 2$. A BDCP can be characterized by its intensity process $\lambda = (\lambda_t)_{t\geq 0}$ which is specified as a piecewise deterministic Markov process (PDP) in Definition 3.3.1. An alternative characterization is through the representation as a cluster process with a specific branching structure in Definition 3.3.4.

Definition 3.3.1 (Bivariate Dynamic Contagion Processes (Intensity-based))**.** *A BDCP is a bivariate point process $N = (N_t)_{t\geq 0}$ in (3.7) with jump times $\{T_n^k\}_{n\geq 0}$ for $k = 1, 2$, and the intensity $\lambda = (\lambda_t)_{t\geq 0}$ where $\lambda_t = (\lambda_t^1, \lambda_t^2)$ for any $t \geq 0$ is defined by*

$$\begin{aligned}
\lambda_t^1 &= \lambda_0^1 e^{-\delta_1 t} + \sum_{S_j^1 < t} Y_j^1 e^{-\delta_1 (t - S_j^1)} + \sum_{T_j^1 < t} Z_j^{1,1} e^{-\delta_1 (t - T_j^1)} + \sum_{T_j^2 < t} Z_j^{1,2} e^{-\delta_1 (t - T_j^2)}, \\
\lambda_t^2 &= \lambda_0^2 e^{-\delta_2 t} + \sum_{S_j^2 < t} Y_j^2 e^{-\delta_2 (t - S_j^2)} + \sum_{T_j^1 < t} Z_j^{2,1} e^{-\delta_2 (t - T_i^1)} + \sum_{T_j^2 < t} Z_j^{2,2} e^{-\delta_2 (t - T_j^2)},
\end{aligned} \tag{3.8}$$

where for $k, k' = 1, 2$,

- $\lambda_0^k \geq 0$ *is the initial intensity at time* $t = 0$*;*
- $\delta_k > 0$ *is the constant rate of the exponential decay;*
- $\{S_j^k\}_{j\geq 1}$ *are jump times of a Poisson process* M^k *with the constant intensity* ρ_k*.* $\{Y_j^k\}_{j\geq 1}$ *are associated jump sizes that are i.i.d non-negative random variables with the distribution function* H_k*, the Laplace transform* $\hat{h}_k(u) = \mathbb{E}[e^{-uY_j^k}]$ *for* $u \in \mathbb{R}_+$*, the mean* μ_H^k *and the i-th moment* μ_{iH^k} *for* $i \geq 2$*;*
- $\{T_j^k\}_{j\geq 1}$ *are jump times of* N^k *with jump size* $\{Z_j^{k,k'}\}_{j\geq 1}$ *that are i.i.d non-negative random variables with distribution function* $G_{k,k'}$*, the Laplace transform* $\hat{g}_{k,k'}(u) = \mathbb{E}\left[e^{-uZ_j^{k,k'}}\right]$ *for* $u \in \mathbb{R}_+$*, the mean* $\mu_{G^{k,k'}}$ *and i-th moment* $\mu_{iG^{k,k'}}$ *for* $i \geq 2$*;*
- *All jump times* $\{S_j^k\}_{j\geq 1}$ *and* $\{T_j^k\}_{j\geq 1}$ *are independent of jump sizes* $\{Y_j^k\}_{j\geq 1}$ *and* $\{Z_j^{k,k'}\}_{j\geq 1}$*.*

A simulation path of (N, λ) is shown in Figure 3.1 in Section 3.6.

The intensity λ is a Markov process due to the exponential decay. For $k, k' = 1, 2$, the marked point process $\sum_{S_j^k < t} Y_j^k e^{-\delta_k (t - S_j^k)}$ characterizes the impact from an external factor on the system. $\sum_{T_j^k < t} Z_j^{k,k} e^{-\delta_k (t - T_i^k)}$ and $\sum_{T_j^{k'} < t} Z_j^{k,k'} e^{-\delta_k (t - T_j^{k'})}$ for $k' \neq k$ are self-exciting and cross-exciting components characterizing the internal dependence through contagion.

Denote for $k = 1, 2$,

$$J_t^{M^k} := \sum_{S_j^k \leq t} Y_j^k, \quad J_t^{N^1,k} := \sum_{T_j^1 \leq t} Z_j^{k,1}, \quad J_t^{N^2,k} := \sum_{T_j^2 \leq t} Z_j^{k,2},$$

then the intensity process in (3.8) can also be written in integral form:

$$\lambda_t^k = \lambda_0^k e^{-\delta_k t} + \int_{[0,t)} e^{-\delta_k (t-s)} dJ_s^{M^k} + \int_{[0,t)} e^{-\delta_k (t-s)} dJ_s^{N^1,k} + \int_{[0,t)} e^{-\delta_k (t-s)} dJ_s^{N^2,k}. \quad (3.9)$$

From the intensity-based definition, we can see that the BDCPs are a broad class of point processes covering a few distinct and important point process classes introduced in Section 3.2.

First by setting $Z_j^{k,k'} \equiv 0$ for all $j \geq 1,\ k,k' = 1,2$, we obtain the shot-noise Cox processes defined in Example 3.2.4:

$$\lambda_t^k = \lambda_0^k e^{-\delta_k t} + \sum_{S_j^k < t} Y_j^k e^{-\delta_k (t - S_j^k)}$$

Second by setting $Y_j^k \equiv 0$ and $Z_j^{k,k'}$ as constants for all $j \geq 1,\ k,k' = 1,2$, we obtain a bivariate Hawkes process with exponential decay defined in (3.3):

$$\begin{aligned}\lambda_t^1 &= \lambda_0^1 e^{-\delta_1 t} + \sum_{T_j^1 < t} Z_j^{1,1} e^{-\delta_1 (t - T_j^1)} + \sum_{T_j^2 < t} Z_j^{1,2} e^{-\delta_1 (t - T_j^2)}, \\ \lambda_t^2 &= \lambda_0^2 e^{-\delta_2 t} + \sum_{T_j^1 < t} Z_j^{2,1} e^{-\delta_2 (t - T_i^1)} + \sum_{T_j^2 < t} Z_j^{2,2} e^{-\delta_2 (t - T_j^2)}.\end{aligned} \tag{3.10}$$

Finally by removing the cross-exciting component in marginal intensity processes, we also recover the univariate dynamic contagion process (UDCP) defined in (3.6) with the intensity :

$$\lambda_t^k = \lambda_0 e^{-\delta_k t} + \sum_{S_j^k < t} Y_j^k e^{-\delta_k (t - S_j^k)} + \sum_{T_j^k < t} Z_j^k e^{-\delta_k (t - T_j^k)}.$$

Conditions on BDCP In this thesis, we assume the following conditions always hold. We will assume a few more conditions hold later when solving specific problems.

Condition 3.3.2.

(C1) For $k,k' = 1,2$, all random jump sizes $\{Y_j^k\}_{j\geq 1}$ and $\{Z_j^{k,k'}\}_{j\geq 1}$ have the finite first moments. i.e. μ_{H^k}, $\mu_{G^{k,k'}}$ are finite.

(C2) The spectral radius of the matrix $\begin{bmatrix} \frac{\mu_{G^{2,2}}}{\delta_2} & \frac{\mu_{G^{1,2}}}{\delta_2} \\ \frac{\mu_{G^{2,1}}}{\delta_1} & \frac{\mu_{G^{1,1}}}{\delta_1} \end{bmatrix}$ is less than 1.

Remark 3.3.3.

(1) One can easily check that under (C1), $\int_0^t \lambda_s ds < \infty$ a.s. for every $t \geq 0$, thus the BDCP N is non-explosive. Indeed, it is sufficient to show that $\mathbb{E}[\int_0^t \lambda_s ds] < \infty$ for all $t \geq 0$ using Gronwall's inequality.

(2) The spectral radius condition in (C2) is equivalent to the following

$$\frac{1}{2}\left(\frac{\mu_{G^{1,1}}}{\delta_1}+\frac{\mu_{G^{2,2}}}{\delta_2}+\sqrt{\left(\frac{\mu_{G^{1,1}}}{\delta_1}+\frac{\mu_{G^{2,2}}}{\delta_2}\right)^2+4\frac{\mu_{G^{1,2}}}{\delta_2}\frac{\mu_{G^{2,1}}}{\delta_1}}\right)<1. \tag{3.11}$$

(3) (C1) and (C2) will be a sufficient stationarity condition of the BDCP (see Theorem 4.3.4).

(4) In addition, we assume that the second moments of jumps μ_{2H^k} and $\mu_{2G^{k,k'}}$ are finite for the computation of the stationary second moment in Chapter 4.

Markov Processes Note that the process $\Gamma := (\lambda^1, \lambda^2, N^1, N^2, t)$ is a Markov process with the generator $\mathcal{A}$, and take any f in the domain $\mathcal{D}(\mathcal{A})$, then at $\Lambda = (\lambda_1, \lambda_2, n_1, n_2, t)$, we have

$$\begin{aligned}
&\mathcal{A}f(\lambda_1,\lambda_2,n_1,n_2,t)\\
&=\frac{\partial f}{\partial t}-\delta_1\lambda_1\frac{\partial f}{\partial \lambda_1}-\delta_2\lambda_2\frac{\partial f}{\partial \lambda_2}\\
&+\rho_1\left[\int_{\mathbb{R}_+} f(\lambda_1+y,\lambda_2,n_1,n_2,t)\,dH_1(y)-f(\Lambda)\right]\\
&+\rho_2\left[\int_{\mathbb{R}_+} f(\lambda_1,\lambda_2+y,n_1,n_2,t)\,dH_2(y)-f(\Lambda)\right]\\
&+\lambda_1\left[\int_{\mathbb{R}_+^2} f(\lambda_1+z_1,\lambda_2+z_2,n_1+1,n_2,t)\,dG_1(z_1,z_2)-f(\Lambda)\right]\\
&+\lambda_2\left[\int_{\mathbb{R}_+^2} f(\lambda_1+z_1,\lambda_2+z_2,n_1,n_2+1,t)\,dG_2(z_1,z_2)-f(\Lambda)\right],
\end{aligned} \tag{3.12}$$

where $G_1(z_1, z_2) = G_{1,1}(z_1)G_{2,1}(z_2)$ and $G_2(z_1, z_2) = G_{1,2}(z_1)G_{2,2}(z_2)$.

Moreover, (λ^1, λ^2) is also a Markov process with the generator $\mathcal{A}_\lambda$, and take any f in the domain $\mathcal{D}(\mathcal{A}_\lambda)$, then at $\Lambda = (\lambda_1, \lambda_2)$, we have

$$\begin{aligned}
&\mathcal{A}_\lambda f(\lambda_1,\lambda_2,t)\\
&=\frac{\partial f}{\partial t}-\delta_1\lambda_1\frac{\partial f}{\partial \lambda_1}-\delta_2\lambda_2\frac{\partial f}{\partial \lambda_2}\\
&+\rho_1\left[\int_{\mathbb{R}_+} f(\lambda_1+y,\lambda_2)\,dH_1(y)-f(\Lambda)\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \rho_2 \left[\int_{\mathbb{R}_+} f(\lambda_1, \lambda_2 + y)\, dH_2(y) - f(\Lambda) \right] \\
&+ \lambda_1 \left[\int_{\mathbb{R}_+^2} f(\lambda_1 + z_1, \lambda_2 + z_2)\, dG_1(z_1, z_2) - f(\Lambda) \right] \\
&+ \lambda_2 \left[\int_{\mathbb{R}_+^2} f(\lambda_1 + z_1, \lambda_2 + z_2)\, dG_2(z_1, z_2) - f(\Lambda) \right]. \qquad (3.13)
\end{aligned}
$$

3.3.2 The Branching Structure

Inspired by [39] for Hawkes processes and [23] for UDCP, in this section we represent the BDCP as a cluster process with a branching interpretation. The representation is helpful in the stationarity analysis in Chapter 4.

Definition 3.3.4 (Bivariate Dynamic Contagion Processes (Cluster-based))**.** *A bivariate dynamic contagion processes $N = (N^1, N^2)$ is a two-type cluster process (C^1, C^2) with the branching interpretation as follows:*

- *For $k = 1, 2$, the cluster centres of type k are immigrants (zeroth generation) arrived at $\{T_m^{k,(0)}\}_{m\geq 1}$ with $T_m^{k,(0)} = S_m^k$ following a shot noise process with intensity $\lambda_t^{k,(0)} = \lambda_0^k e^{-\delta_k t} + \sum_{S_m^k < t} Y_m^k e^{-\delta_k (t - T_m^{k,(0)})}$.*
- *Each cluster centre of type k arrived at $T_m^{k,(0)}$ generates a cluster C_m^k consisting of events of type k. We include the cluster centre into the cluster. In branching term, each C_m^k is the set of type k immigrant arrived at $T_m^{k,(0)}$ and its decedents of all generations.*

  Moreover, the cluster process of type k becomes $C^k = \bigcup_{m=1}^{\infty} C_m^k$.
- *For $n \geq 1$, denote by $C_m^{k,(n)}$ the set of the n-th generation offspring in the cluster C_m^k, then the collection of the n-th generation offspring of type k from all clusters is $C^{k,(n)} = \bigcup_{m=1}^{\infty} C_m^{k,(n)}$. Denote by $N^{k,(n)} = (N_t^{k,(n)})_{t\geq 0}$ the offspring birth process in $C^{k,(n)}$ with the arrival times $\{T_j^{k,(n)}\}_{j\geq 1}$ and the intensity $\lambda^{k,(n)} = (\lambda_t^{k,(n)})_{t\geq 0}$.*

  For each $n \geq 0$, the $(n+1)$-th generation is generated from n-th generation individuals in all clusters with the intensity

  $$\lambda_t^{k,(n+1)} = \sum_{T_j^{1,(n)} < t} Z_j^{k,1,(n+1)} e^{-\delta_k (t - T_j^{1,(n)})} + \sum_{T_j^{2,(n)} < t} Z_j^{k,2,(n+1)} e^{-\delta_k (t - T_j^{2,(n)})},$$

where for $k, k' = 1, 2$, the random marks $Z_j^{k,k',(n+1)}$ are independent copies of $Z_j^{k,k'}$.

- *Collect all individuals of type k up to the n-th generation from all clusters, denoted by $C^{k,n}$, then $C^{k,n} = \bigcup_{j=0}^{n} C^{k,(j)}$. The offspring birth process up to the n-th generation in $C^{k,n}$ is $N^{k,n} = (N_t^{k,n})_{t\geq 0}$ with birth times $\{T_j^{k,n}\}_{j\geq 1}$ and the intensity process $\lambda^{k,n} = (\lambda_t^{k,n})_{t\geq 0}$. Hence, we have for any $t \geq 0$,*

$$N_t^{k,n} = \sum_{i=0}^{n} N_t^{k,(i)}, \quad \lambda_t^{k,n} = \sum_{i=0}^{n} \lambda_t^{k,(i)}, \quad \left\{T_j^{k,n}\right\}_{j\geq 1} = \bigcup_{i=0}^{n} \left\{T_j^{k,(i)}\right\}_{j\geq 1}.$$

Clearly,

$$C^k = \lim_{n\to\infty} C^{k,n} = \lim_{n\to\infty} \bigcup_{j=1}^{n} C^{k,(j)} = \lim_{n\to\infty} \bigcup_{j=1}^{n} \bigcup_{m=1}^{\infty} C_m^{k,(j)}.$$

By construction, all clusters $\{C_m^k\}_{m\geq 1}$ are independent. Moreover, we have

$$\begin{aligned} N_t^k &= \lim_{n\to\infty} N_t^{k,n} = \lim_{n\to\infty} \sum_{i=0}^{n} N_t^{k,(i)}, \\ \lambda_t^k &= \lim_{n\to\infty} \lambda_t^{k,n} = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda_t^{k,(i)}. \end{aligned}$$

We call the sequence of processes $N^n = (N^n)_{n\geq 1}$ where $N^n = (N^{1,n}, N^{2,n})$ with intensity $\lambda^n = (\lambda^{1,n}, \lambda^{2,n})$ *the truncated finite systems* of the BDCP.

From the bivariate system to the univariate system To simplify the multi-type problem, we merge the bivariate branching system into a univariate system, such that the i-th generation of type 1 and type 2 offspring become the $(2i-1)$-th and $2i$-th generation in the new univariate system. Denote the birth time of the n-th generation in the new system as $\{T_j^{(n)}\}_{j\geq 1}$ and the counting process as $N^{(n)} = (N_t^{(n)})_{t\geq 0}$ with the intensity $\Lambda^{(n)}$, then for $i \geq 1$ and $t \geq 0$,

$$\Lambda_t^{(2i-1)} = \lambda_t^{1,(i)}, \quad \Lambda_t^{(2i)} = \lambda_t^{2,(i)}.$$

Then we have for $i \geq 1$,

$$\Lambda_t^{(1)} = \lambda_0^1 e^{-\delta_1 t} + \sum_{i=1}^{M_t^1} Y_i^1 e^{-\delta_1 \left(t - T_i^{1,(0)}\right)},$$

$$
\begin{aligned}
\Lambda_t^{(2)} &= \lambda_0^2 e^{-\delta_2 t} + \sum_{i=1}^{M_t^2} Y_i^2 e^{-\delta_2\left(t-T_i^{2,(0)}\right)},\\
\Lambda_t^{(2i+1)} &= \sum_{j=1}^{N_t^{(2i-1)}} Z_j^{1,1} e^{-\delta_1\left(t-T_j^{(2i-1)}\right)} + \sum_{j=1}^{N_t^{(2i)}} Z_j^{1,2} e^{-\delta_1\left(t-T_j^{(2i)}\right)},\\
\Lambda_t^{(2i+2)} &= \sum_{j=1}^{N_t^{(2i-1)}} Z_j^{2,1} e^{-\delta_2\left(t-T_j^{(2i-1)}\right)} + \sum_{j=1}^{N_t^{(2i)}} Z_j^{2,2} e^{-\delta_2\left(t-T_j^{(2i)}\right)}.
\end{aligned}
$$

By construction, the original bivariate branching system can be recovered:

$$
\begin{aligned}
\lambda_t^{1,n} &= \sum_{i=1}^{n} \Lambda_t^{(2i-1)}, \quad N_t^{1,n} = \sum_{i=1}^{n} N_t^{(2i-1)},\\
\lambda_t^{2,n} &= \sum_{i=1}^{n} \Lambda_t^{(2i)}, \quad N_t^{2,n} = \sum_{i=1}^{n} N_t^{(2i)}.
\end{aligned}
$$

Hence the original bivariate system with truncation up to n-th generation is transformed into a univariate system with truncation up to m-th generation with $m = 2n$. We call m and n as the *system index* for the transformed and the original system respectively.

Remark 3.3.5. *Note that the BDCP system $(N^1, N^2, \lambda^1, \lambda^2)$ and the intensity process (λ^1, λ^2) are Markovian but the truncated finite system intensity $(\lambda^{1,n}, \lambda^{2,n})$ is not. However, the joint branching intensity process $(\Lambda^{(1)}, \ldots, \Lambda^{(m)})$ is a Markov process.*

3.4 Existence of Moments

In this section, we aim to find the condition under which the moment

$$
\mu_{(p_1,p_2,q_1,q_2)} := \mathbb{E}\left[(\lambda_t^1)^{p_1}(\lambda_t^2)^{p_2}(N_t^1)^{q_1}(N_t^2)^{q_2}\right]
$$

exists for any $p_1, p_2, q_1, q_2 \in \mathbb{N}$. The condition found and Example 3.4.2 will be used in exploring the filtering error in Proposition 5.3.7 in Chapter 5.

Proposition 3.4.1 (Existence of Moments)**.** *A sufficient condition of the existence of the moment of $\mathbb{E}\left[(\lambda_t^1)^{p_1}(\lambda_t^2)^{p_2}(N_t^1)^{q_1}(N_t^2)^{q_2}\right]$ is that for $k, k' = 1, 2$, the p_k-th moment of jump size $\{Y_j^k\}_{j\geq 1}$ and (p_1+p_2)-th moment of jump size $\{Z_j^{k,k'}\}_{j\geq 1}$ are finite, i.e.,*

$$
\mu_{p_k H^k} < \infty, \quad \mu_{(p_1+p_2)G^{k,k'}} < \infty. \tag{3.14}
$$

Proof. First we denote

$$p := p_1 + p_2, \quad q := q_1 + q_2.$$

Note that the process $\Gamma := (\lambda^1, \lambda^2, N^1, N^2, t)$ is a Markov process with the generator in (3.12), and if we take $f(\lambda, n) = (\lambda_1)^{p_1}(\lambda_2)^{p_1}(n_1)^{q_1}(n_2)^{q_2}$ in (3.12), then we obtain

$$\begin{aligned}
&\mathcal{A}\Big((\lambda_1)^{p_1}(\lambda_2)^{p_1}(n_1)^{q_1}(n_2)^{q_2}\Big) \\
=& - p_1\delta_1(\lambda_1)^{p_1}(\lambda_2)^{p_2}(n_1)^{q_1}(n_2)^{q_2} - p_2\delta_2(\lambda_1)^{p_1}(\lambda_2)^{p_2}(n_1)^{q_1}(n_2)^{q_2} \\
&+ \rho_1(\lambda_2)^{p_2}(n_1)^{q_1}(n_2)^{q_2}\left[\int_{\mathbb{R}_+}(\lambda_1 + y)^{p_1} dH_1(y) - (\lambda_1)^{p_1}\right] \\
&+ \rho_2(\lambda_1)^{p_1}(n_1)^{q_1}(n_2)^{q_2}\left[\int_{\mathbb{R}_+}(\lambda_2 + y)^{p_2} dH_2(y) - (\lambda_2)^{p_2}\right] \\
&+ \lambda_1\left[\int_{\mathbb{R}_+^2}(\lambda_1 + z_1)^{p_1}(\lambda_2 + z_2)^{p_2}(n_1 + 1)^{q_1}(n_2)^{q_2} dG_{1,1}(z_1)dG_{2,1}(z_2) - (\lambda_1)^{p_1}(\lambda_2)^{p_1}(n_1)^{q_1}(n_2)^{q_2}\right] \\
&+ \lambda_2\left[\int_{\mathbb{R}_+^2}(\lambda_1 + z_1)^{p_1}(\lambda_2 + z_2)^{p_2}(n_1)^{q_1}(n_2 + 1)^{q_2} dG_{1,2}(z_1)dG_{2,2}(z_2) - (\lambda_1)^{p_1}(\lambda_2)^{p_1}(n_1)^{q_1}(n_2)^{q_2}\right] \\
=& - (p_1\delta_1 + p_2\delta_2)(\lambda_1)^{p_1}(\lambda_2)^{p_1}(n_1)^{q_1}(n_2)^{q_2} \\
&+ \rho_1\sum_{i=0}^{p_1-1}\mu_{(p_1-i)H^1}(\lambda_1)^i(\lambda_2)^{p_2}(n_1)^{q_1}(n_2)^{q_2} + \rho_2\sum_{j=0}^{p_2-1}\mu_{(p_2-j)H^2}(\lambda_1)^{p_1}(\lambda_2)^j(n_1)^{q_1}(n_2)^{q_2} \\
&+ \lambda_1 F_1(\lambda_1, \lambda_2, n_1, n_2) + \lambda_2 F_2(\lambda_1, \lambda_2, n_1, n_2), \qquad (3.15)
\end{aligned}$$

where using binomial expansion,

$$\begin{aligned}
&F_1(\lambda_1, \lambda_2, n_1, n_2) \\
=&\sum_{i=0}^{p_1-1}\sum_{j=0}^{p_2-1}\sum_{l=0}^{q_1-1} a_{(i,j,l,q_2)}(\lambda_1)^i(\lambda_2)^j(n_1)^l(n_2)^{q_2} + \sum_{j=0}^{p_2-1}\sum_{l=0}^{q_1-1} a_{(p_1,j,l,q_2)}(\lambda_1)^{p_1}(\lambda_2)^j(n_1)^l(n_2)^{q_2} \\
&+ \sum_{i=0}^{p_1-1}\sum_{l=0}^{q_1-1} a_{(i,p_2,l,q_2)}(\lambda_1)^i(\lambda_2)^{p_2}(n_1)^l(n_2)^{q_2} + \sum_{i=0}^{p_1-1}\sum_{j=0}^{p_2-1} a_{(i,j,q_1,q_2)}(\lambda_1)^i(\lambda_2)^j(n_1)^{q_1}(n_2)^{q_2}, \\
&F_2(\lambda_1, \lambda_2, n_1, n_2) \\
=&\sum_{i=0}^{p_1-1}\sum_{j=0}^{p_2-1}\sum_{l=0}^{q_2-1} b_{(i,j,q_1,l)}(\lambda_1)^i(\lambda_2)^j(n_1)^{q_1}(n_2)^l + \sum_{j=0}^{p_2-1}\sum_{l=0}^{q_2-1} b_{(p_1,j,q_1,l)}(\lambda_1)^{p_1}(\lambda_2)^j(n_1)^{q_1}(n_2)^l \\
&+ \sum_{i=0}^{p_1-1}\sum_{l=0}^{q_2-1} b_{(i,p_2,q_1,l)}(\lambda_1)^i(\lambda_2)^{p_2}(n_1)^{q_1}(n_2)^l + \sum_{i=0}^{p_1-1}\sum_{j=0}^{p_2-1} b_{(i,j,q_1,q_2)}(\lambda_1)^i(\lambda_2)^j(n_1)^{q_1}(n_2)^{q_2},
\end{aligned}$$

with the coefficients

$$a_{(i,j,l,q_2)} = \binom{p_1}{i}\binom{p_2}{j}\binom{q_1}{l}\mu_{(p_1-i)G^{1,1}}\mu_{(p_2-j)G^{2,1}},$$

$$b_{(i,j,q_1,l)} = \binom{p_1}{i}\binom{p_2}{j}\binom{q_2}{l}\mu_{(p_1-i)G^{1,2}}\mu_{(p_2-j)G^{2,2}}.$$

By martingale theorem, $f(\lambda_t, N_t) - \int_0^t \mathcal{A}f(\lambda_s, N_s)ds$ is a martingale, then

$$\begin{aligned}\mathbb{E}\left[(\lambda_t^1)^{p_1}(\lambda_t^2)^{p_2}(N_t^1)^{q_1}(N_t^2)^{q_2}\right] =& \mathbb{E}\left[(\lambda_0^1)^{p_1}(\lambda_0^2)^{p_2}(N_0^1)^{q_1}(N_0^2)^{q_2}\right] \\ &+ \int_0^t \mathbb{E}\left[\mathcal{A}\Big((\lambda_s^1)^{p_1}(\lambda_s^2)^{p_2}(N_s^1)^{q_1}(N_s^2)^{q_2}\Big)\right] ds. \quad (3.16)\end{aligned}$$

The last term in above equation is obtained by taking expectation in (3.15) where λ_k and n_k are replaced by λ_s^k and N_s^k for $k = 1, 2$. Note that all terms in (3.15) are linear function of the moments $\mu_{(i,j,l,w)}$ and therefore we actually need to solve a linear ODE systems of $\left\{\mu_{(i,p-i,j,q-j)} : i = 0, \ldots, p; j = 0, \ldots, q\right\}$ for $p = p_1 + p_2$, $q = q_1+q_2$ recursively in terms of the indexes p and q and if all lower order moments $\left\{\mu_{(i,j,l,w)} : i + j < p; l + w < q\right\}$ are known. Therefore, we can obtain $\mu_{(p_1,p_2,q_1,q_2)}$.

As the existence of higher order moments implies the existence of lower order moments, from (3.15), the moment $\mu_{(p_1,p_2,q_1,q_2)}$ exists if $\mu_{(p_1-i)H^1}$, $\mu_{(p_2-j)H^2}$, $a_{(i,p-i,j,q-j)}$ and $b_{(i,p-i,j,q-j)}$ are all finite for $i \leq p$ and $j \leq q$. Therefore, we obtain a sufficient condition as in (3.14).

□

Example 3.4.2. *For example, the third moments* $\mathbb{E}[(\lambda_t^k)^3]$ *and* $\mathbb{E}[(N_t^k)^3]$ *exist for* $k = 1, 2$, *if* $\mu_{3H^k} < \infty$ *and* $\mu_{3G^{k,k'}} < \infty$ *for* $k, k' = 1, 2$.

3.5 Exact Simulation of BDCP

First recall [25] where the exact simulation algorithm of the UDCP is based on simulation of the inter-arrival times. In this section, we extend the algorithm for the simulation of the BDCP system $(N^1, N^2, \lambda^1, \lambda^2)$. Note that the univariate process is a reduced case and more general multivariate processes can be extended without difficulty.

First recall the intensity process in (3.8):

$$\lambda_t^1 = \lambda_0^1 e^{-\delta_1 t} + \sum_{S_j^1 < t} Y_j^1 e^{-\delta_1(t-S_j^1)} + \sum_{T_j^1 < t} Z_j^{1,1} e^{-\delta_1(t-T_j^1)} + \sum_{T_j^2 < t} Z_j^{1,2} e^{-\delta_1(t-T_j^2)},$$

$$\lambda_t^2 = \lambda_0^2 e^{-\delta_2 t} + \sum_{S_j^2 < t} Y_j^2 e^{-\delta_2(t-S_j^2)} + \sum_{T_j^1 < t} Z_j^{2,1} e^{-\delta_2(t-T_j^1)} + \sum_{T_j^2 < t} Z_j^{2,2} e^{-\delta_2(t-T_j^2)}.$$

The simulation algorithm is based on the fact that λ is a piecewise deterministic process. For each $k = 1, 2$, the dynamic of λ^k consists of upward jumps and exponential decay between jumps. Note that there are three types of jumps in λ^k, one is the external jumps as a Poisson process, and the other two types are self-exciting jumps from N^k and cross-exciting jumps from $N^{k'}$ for $k' \neq k$.

The algorithm consists of simulating the inter-arrival times between jumps of the intensity λ, simulating the deterministic motion of λ between jumps and updating λ at jump times. Since all jumps in the intensity process λ are triggered by jumps of N and M, meanwhile λ determines the jumps of N, we generate all jumps of N and M sequentially in the intensity processes.

First we denote by Π the set of all jumps in the intensity processes:

$$\Pi := \{T_1 < T_2 < \cdots < T_i < \cdots\} = \{S_j^1\}_{j\geq 1} \bigcup \{S_j^2\}_{j\geq 1} \bigcup \{T_j^1\}_{j\geq 1} \bigcup \{T_j^2\}_{j\geq 1}.$$

Moreover denote

- T_i: the i-th jump time in Π.
- W_i: the inter-arrival time between the $(i-1)$-th and i-th jump in Π, so $W_i = T_i - T_{i-1}$.
- E_i^k: the inter-arrival time of the first jump of Poisson process M^k after T_i.
- I_i^k: the inter-arrival time of the first jump of N^k after T_i assuming no jumps from other components.

Suppose there is a jump at T_i and intensity right before the jump is $\lambda^k_{T_i^-} := \lambda^k_{T_i-} = \lambda^k_{T_i}$, then the intensity is updated at T_i, denoted by $\lambda^k_{T_i^+} := \lambda^k_{T_i+}$ according to the type of the jump at T_i:

- If the jumps is the jump of M^k, i.e. $T_i = S_j^k$ for some j, then $\lambda^k_{T_i^+} = \lambda^k_{T_i^-} + Y_j^k$.

- If the jumps is the jump of $M^{k'}$, i.e. $T_i = S_j^{k'}$ for some j, then $\lambda^k_{T_i^+} = \lambda^k_{T_i^-}$.
- If the jumps is the jump of N^k, i.e. $T_i = T_j^k$ for some j, then $\lambda^k_{T_i^+} = \lambda^k_{T_i^-} + Z_j^{k,k}$.
- If the jumps is the jump of $N^{k'}$, i.e. $T_i = T_j^{k'}$ for some j, then $\lambda^k_{T_i^+} = \lambda^k_{T_i^-} + Z_j^{k,k'}$.

After the update at T_i, the intensity of the next jump $T_{i+1} \in \Pi$ becomes:

$$\lambda_t^1 + \lambda_t^2 + \rho_1 + \rho_2, \quad t \in (T_i, T_{i+1})$$

where

$$\lambda_t^k = \lambda^k_{T_i^+} e^{-\delta_k (t - T_i)}, \quad t \in (T_i, T_{i+1}).$$

Therefore the inter-arrival time between T_i and T_{i+1} is

$$W_{i+1} = \min\{E^1_{i+1}, E^2_{i+1}, I^1_{i+1}, I^2_{i+1}\},$$

where $E^1_{i+1}, E^2_{i+1}, I^1_{i+1}, I^2_{i+1}$ are mutually independent random variables conditioned on the updated intensity $\lambda^k_{T_i^+}$ at T_i. Moreover, for $k = 1, 2$, we have the distribution functions of the inter-arrival times:

$$\begin{aligned}
F_{E^k_{i+1}}(t) &:= \mathbb{P}(E^k_{i+1} \le t) = 1 - e^{-\rho_k t}, \\
F_{I^k_{i+1}}(t) &:= \mathbb{P}(I^k_{i+1} \le t) = 1 - \exp\left(- \int_{T_i}^{T_i + t} \lambda_u^k du \right) = 1 - \exp\left(- \int_0^t \lambda^k_{T_i + u} du \right) \\
&= 1 - \exp\left(- \lambda^k_{T_i^+} \frac{1 - e^{-\delta_k t}}{\delta_k} \right).
\end{aligned}$$

Define $d^k_{i+1} := 1 + \frac{\delta_k \ln U^{I,k}_{i+1}}{\lambda^k_{T_i^+}}$ where $U^{I,k}_{i+1}$ is a uniform random variable on $[0, 1]$, moreover define

$$\tilde{I}^k_{i+1} := -\frac{1}{\delta_k} \ln(d^k_{i+1}) 1_{\{d^k_{i+1} > 0\}},$$

then $\tilde{I}^k_{i+1} | d^k_{i+1} > 0 \overset{d}{=} I^k_{i+1}$. Indeed, from the definition, we have

$$\mathbb{P}(\tilde{I}^k_{i+1} \le t | d^k_{i+1} > 0) = 1 - \exp\left(- \lambda^k_{T_i^+} \frac{1 - e^{-\delta_k t}}{\delta_k} \right).$$

Exact Simulation Algorithm of BDCP We simulate $\{T_i\}_{i \ge 1}$ in Π sequentially and in this way we can simulate the BDCP system (N, λ).

1. Start from $T_0 = 0$, $\lambda^k_{T_0^\pm} = \lambda^k_0 > 0$, $N^k_0 = 0$, for $k = 1, 2$.

2. For $i \geq 0$, we simulate the $(i+1)$-th inter-arrival time W_{i+1} in the intensity processes. Since

$$W_{i+1} = \min\{E^1_{i+1}, E^2_{i+1}, I^1_{i+1}, I^2_{i+1}\},$$

then

$$T_{i+1} = T_i + W_{i+1}.$$

Generate mutually independent random variables $U^{E,k}_{i+1}$, $U^{I,k}_{i+1}$, $V^{Y,k}_{i+1}$ and $V^{Z,k,j}_{i+1}$ uniformly distributed on $[0, 1]$ for $k, j = 1, 2$. Then, generate

$$\begin{aligned} E^k_{i+1} &= -\frac{1}{\rho_k} \ln U^{E,k}_{i+1}, \\ I^k_{i+1} &= -\frac{1}{\delta_k} \ln \left(d^k_{i+1}\right) 1_{\{d^k_{i+1}>0\}} | d^k_{i+1} > 0. \end{aligned}$$

3. Update λ^k_t and N^k_t at T_{i+1}.

For $k, k' = 1, 2$ and $k \neq k'$,

$$\lambda^k_{T^+_{i+1}} = \begin{cases} \lambda^k_{T^-_{i+1}} + Y^k_{i+1}, & W_{i+1} = E^k_{i+1} \\ \lambda^k_{T^-_{i+1}} + Z^{k,k}_{i+1}, & W_{i+1} = I^k_{i+1} \\ \lambda^k_{T^-_{i+1}}, & W_{i+1} = E^{k'}_{i+1} \\ \lambda^k_{T^-_{i+1}} + Z^{k,k'}_{i+1}, & W_{i+1} = I^{k'}_{i+1} \end{cases}$$

where

$$\lambda^k_{T^-_{i+1}} = \lambda^k_{T^+_i} e^{-\delta_k (T_{i+1} - T_i)},$$

and $Y^k_{i+1} \stackrel{d}{=} -\frac{1}{\alpha_k} \ln V^{Y,k}_{i+1}$, $Z^{k,j} \stackrel{d}{=} -\frac{1}{\beta_{k,j}} \ln V^{Z,k,j}_{i+1}$ for $j = 1, 2$.

We update N^k at T_{i+1} as

$$N^k_{T^+_{i+1}} = \begin{cases} N^k_{T^-_{i+1}} + 1, & W_{i+1} = I^k_{i+1} \\ N^k_{T^-_{i+1}}, & \text{otherwise.} \end{cases}$$

3.6 Graphic Illustration

Take the exponential jump sizes of λ: $Y_i^k \sim exp(\beta_k^Y)$, $Z_i^{k,k'} \sim exp(\beta_{k,k'}^Z)$ with parameters in Table 3.1. Moreover, we suppose the process starts with the stationary mean value computed in Section 4.4. We show in Figure 3.1 a simulated sample path of a BDCP using the algorithm in the last section.

$\beta_{k,k}^Z = 5 \ (k = 1, 2)$	$\beta_{k,k'}^Z = 2 \ (k \neq k')$
$\rho_1 = 10$	$\rho_2 = 10$
$\delta_1 = 2$	$\delta_2 = 1$
$\beta_1^Y = 1$	$\beta_2^Y = 1$

Table 3.1: Parameters for the simulated path

We observe that for $k = 1, 2$, the intensity λ^k have common jumps with the counting process N^k and $N^{k'}$ for $k \neq k'$ which exhibits the mutually exciting property. Moreover, the intensity has additional external jumps of the Poisson process M^k.

3.7 Conclusion

In this chapter, we introduced the basics of point processes and BDCPs. We defined the BDCP based on the intensity process and the cluster process representation. We discussed its Markov property, and based on which we found the existence condition of the moments. Moreover, we provided an exact simulation algorithm of the BDCP and illustrated with a simulated sample path.

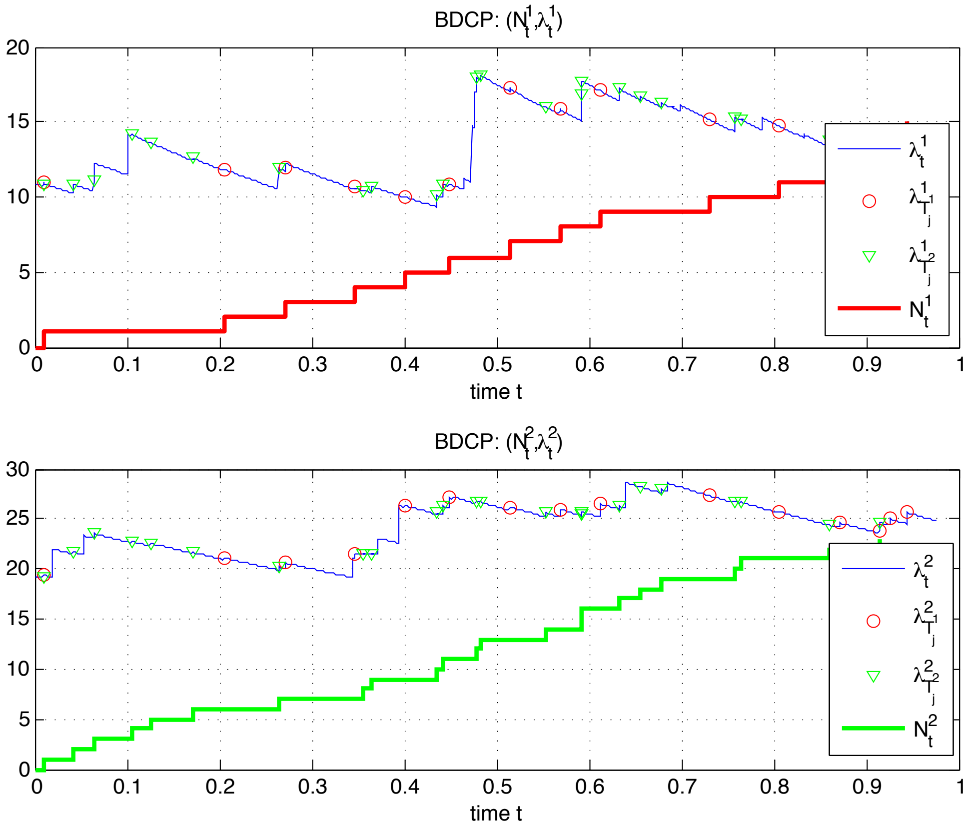


Figure 3.1: A simulated path of the BDCP and its intensity $(N^1, N^2, \lambda^1, \lambda^2)$. The self-exciting jumps $\left\{\lambda^k_{T_j^k+} - \lambda^k_{T_j^k}\right\}_{j\geq 1}$ and cross-exciting jumps $\left\{\lambda^{k'}_{T_j^k+} - \lambda^{k'}_{T_j^k}\right\}_{j\geq 1}$ for $k, k' = 1, 2$ and $k \neq k'$.

Chapter 4

BDCP Stationarity Analysis

In this chapter, we explore the limiting and stationary distributions of the BDCP intensity $\lambda = (\lambda_t)_{t\geq 0}$ and the truncated finite system intensity $\lambda^n = (\lambda_t^n)_{t\geq 0}$ introduced in Chapter 3. We first find the the limiting distribution ($t \to \infty$) of the finite branching system, λ^n, using Markov process theory, which coincides with the stationary distribution. As $n \to \infty$, the finite system converges to the original BDCP, and we find a sufficient condition ((C1) and (C2) in Condition 3.3.2) for existence of the limiting and stationary distributions of BDCP. The stationarity result will be used in Chapter 5 for the diffusion approximation.

4.1 Introduction

Stationarity is an important property of stochastic processes and is also a common assumption in many statistical applications. Based on the stationarity assumption in Chapter 5, we will explore the diffusion approximation of BDCP with filtering applications. Therefore, we investigate the existence of a version of a stationary intensity of BDCP in this chapter.

In the following, we first review the definitions and the literature.

Stationarity On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a stochastic process $X = (X_t)_{t\geq 0}$ on state space $(E, \mathcal{B}(E))$ is a *stationary process* if the finite dimensional distribution

$$\mathbb{P}\left(X_{t+s_1} \in \Gamma_1, \ldots, X_{t+s_k} \in \Gamma_k\right)$$

is independent of $t \geq 0$ for all $k \geq 1$, $0 \leq s_1 \leq \cdots \leq s_k$, and $\Gamma_1, \ldots, \Gamma_k \in \mathcal{B}(E)$.

Then we recall the definition of the stationary distribution and the stationary process in the class of Markov processes (see Section 9.4 in Ethier and Kurtz [33]). Suppose that a martingale problem for the generator $\mathcal{A}$ on state space $(E, \mathcal{B}(E))$ is well defined, then μ as a probability measure on E is a *stationary distribution* of $\mathcal{A}$ if every solution X of the martingale problem for $(\mathcal{A}, \mu)$ is a stationary process. Moreover, μ is a stationary distribution for $\mathcal{A}$ if and only if X_t has the distribution μ for all $t \geq 0$.

The following stationarity theorem is from [33, Chapter 4, Proposition 9.2].

Theorem 4.1.1. *For a Markov process X that solves the martingale problem for $(\mathcal{A}, \mu)$ with the domain $\mathcal{D}(\mathcal{A})$,*

μ is a stationary distribution if and only if for any $f \in \mathcal{D}(\mathcal{A})$,

$$\int_E \mathcal{A}f(x)d\mu(x) = 0. \quad (4.1)$$

Costa [18] discusses the stationarity condition of piecewise deterministic Markov processes based on the analysis of the embedded Markov chain. Dassios and Zhao [23] show the existence of a stationary distribution for UDCP, based upon which the ruin probability in insurance modelling using UDCP is discussed in Dassios and Zhao [24]. Brémaud and Massoulie [14] discuss the stationarity and stability of linear and non-linear Hawkes processes defined on the whole real line based on a recursively constructed approximating sequence of the intensity process. Furthermore the BDCP can be seen as an affine process, and Duffie et al. [28] and Keller-Ressel et al. [45] can be consulted for the research on affine processes. The stationarity results are only available in a few cases of diffusion affine processes. For example, the discussion of the stationarity of two-factor affine diffusion processes can be found in Glasserman and Kim [35] and Barczy et al. [9].

The analysis of the BDCP intensity in the following is based on the convergence of the finite branching system $(\lambda^{1,n}, \lambda^{2,n})$ or the joint system $(\Lambda^{(1)}, \ldots, \Lambda^{(2n)})$ in the cluster-based representation of the BDCP intensity (λ^1, λ^2) (see Section 3.3.2). We apply the Markov theory of the PDP on the finite system to explore the limiting distribution as $t \to \infty$. With the branching system index $n \to \infty$, the existence condition of the limiting distribution of (λ^1, λ^2) is obtained. The limiting distribution result can be found in Theorem 4.2.2 and the existence condition is Condition 3.3.2 in Chapter 3. In Section 4.3, we show that the limiting distribution is also the stationary distribution. We provide a stationarity condition of $(\Lambda^{(1)}, \ldots, \Lambda^{(2n)})$ in Lemma 4.3.1 in terms of the Laplace transform. We verify that the limiting distribution found in Section 4.2 is also the stationary distribution of $(\Lambda^{(1)}, \ldots, \Lambda^{(2n)})$ and $(\lambda^{1,n}, \lambda^{2,n})$ in Theorem 4.3.2 and Corollary 4.3.7 respectively. The convergence argument is applied to conclude the stationarity of the BDCP intensity (λ^1, λ^2) in Theorem 4.3.4 and also BDCP (N^1, N^2) in Corollary 4.3.7. In Section 4.4, we provide the stationary moments.

Notation Denote the limiting distribution $(t \to \infty)$ and the stationary distribution in the branching system as follows:

process	limiting distribution	stationary distribution
$\left(\Lambda^{(1)}, \ldots, \Lambda^{(m)}\right)$	π_A^m	π_S^m
$(\lambda^{1,n}, \lambda^{2,n})$	μ_A^n	μ_S^n
(λ^1, λ^2)	μ_A^*	μ_S^*

Table 4.1: Notation of distributions in the branching system.

4.2 Markov Property and Limiting Distributions

We use the Markov property to explore the limiting distributions of the intensities in this section.

4.2.1 Markov Property

Though the intensity (λ^1, λ^2) is a Markov process, it is difficult to explore the stationarity directly using the PDP theory as the marginal processes are coupled through the cross-exciting components. Instead, we start from the joint finite system $\left(\Lambda_t^{(1)}, \Lambda_t^{(2)}, \ldots, \Lambda_t^{(m)}\right)$ from the branching structure introduced in Section 3.3.2 since it is a Markov process and also a recursively decoupled system. Moreover it can recover the finite system $(\lambda^{1,n}, \lambda^{2,n})$ which is not Markovian but converges to the Markov process (λ^1, λ^2).

Suppose the generator of $\left(t, \Lambda_t^{(1)}, \Lambda_t^{(2)}, \ldots, \Lambda_t^{(m)}\right)$ is $\mathcal{A}_m$ with domain $\mathcal{D}(\mathcal{A}_m)$, then for any $f \in \mathcal{D}(\mathcal{A}_m)$, $\lambda := (\lambda_1, \dots, \lambda_m)$

$$
\begin{aligned}
&\mathcal{A}_m f(t, \lambda_1, \lambda_2, \dots, \lambda_m) \\
&= \frac{\partial f}{\partial t} - \sum_{i=1}^{m} \delta_i \lambda_i \frac{\partial f}{\partial \lambda_i} \\
&+ \rho_1 \left[\int_0^\infty f(t, \lambda + e_1 y) dH_1(y) - f(t, \lambda) \right] + \rho_2 \left[\int_0^\infty f(t, \lambda + e_2 y) dH_2(y) - f(t, \lambda) \right] \\
&+ \sum_{k=1}^{n-1} \lambda_{2k-1} \left[\left(\int_0^\infty f(t, \lambda + e_{2k+1} z) dG_{1,1}(z) - f(t, \lambda) \right) + \left(\int_0^\infty f(t, \lambda + e_{2k+2} z) dG_{2,1}(z) - f(t, \lambda) \right) \right. \\
&+ \sum_{k=1}^{n-1} \lambda_{2k} \left[\left(\int_0^\infty f(t, \lambda + e_{2k+1} z) dG_{1,2}(z) - f(t, \lambda) \right) + \left(\int_0^\infty f(t, \lambda + e_{2k+2} z) dG_{2,2}(z) - f(t, \lambda) \right) \right],
\end{aligned}
\tag{4.2}
$$

where $e_i := (0, \dots, 0, 1, 0, \dots) \in \mathbb{R}^m$ with i-th element being 1 and other elements being 0.

Take

$$
f\left(t, \Lambda_t^{(1)}, \cdots, \Lambda_t^{(m)}\right) = e^{-B_1(t)\Lambda_t^{(1)} - \cdots - B_m(t)\Lambda_t^{(m)} + c_m(t)}
$$

and suppose it is a martingale. Consider for any $T > 0$ and assume that $B_i(T) = v_i$ and $c_m(0) = 0$, then the Laplace transform of $(\Lambda_T^{(1)}, \dots, \Lambda_T^{(m)})$ conditioned on the initial condition $\Lambda_0 = (\Lambda_0^{(1)}, \dots, \Lambda_0^{(m)})$ at $(v_1, \dots, v_m) \in \mathbb{R}_+^m$ is

$$
\mathbb{E}_0 \left[e^{-v_1 \Lambda_T^{(1)} - \cdots - v_m \Lambda_T^{(m)}} \right] = \mathbb{E}_0 \left[e^{-B_1(T)\Lambda_T^{(1)} - \cdots - B_m(T)\Lambda_T^{(m)}} \right] = e^{-B_1(0)\Lambda_0^{(1)} - \cdots - B_m(0)\Lambda_0^{(m)} - c_m(T)},
$$

where $\mathbb{E}_0[\cdot] := \mathbb{E}[\cdot | \Lambda_0]$.

A sufficient condition for f to be a martingale is that for any $t \geq 0$, $\{\lambda_i\}_{i=1}^m$, ρ_1, ρ_2 on $\mathbb{R}_+$, $\mathcal{A}_m f(t, \lambda_1, \lambda_2, \dots, \lambda_m) = 0$, i.e.

$$
\begin{aligned}
0 =& \frac{\mathcal{A}_m f}{f} \\
=& \sum_{k=1}^{n-1} \lambda_{2k-1} \left[-\dot{B}_{2k-1}(t) + \delta_1 B_{2k-1}(t) + (\hat{g}_{1,1}(B_{2k+1}(t)) - 1) + (\hat{g}_{2,1}(B_{2k+2}(t)) - 1) \right] \\
&+ \sum_{k=1}^{n-1} \lambda_{2k} \left[-\dot{B}_{2k}(t) + \delta_2 B_{2k}(t) + (\hat{g}_{1,2}(B_{2k+1}(t)) - 1) + (\hat{g}_{2,2}(B_{2k+2}(t)) - 1) \right] \\
&+ \lambda_{2n-1} \left[-\dot{B}_{2n-1}(t) + \delta_1 B_{2n-1}(t) \right] + \lambda_{2n} \left[-\dot{B}_{2n}(t) + \delta_2 B_{2n}(t) \right] \\
&+ \dot{c}_m(t) + \rho_1 \left(\hat{h}_1(B_1(t)) - 1 \right) + \rho_2 \left(\hat{h}_2(B_2(t)) - 1 \right).
\end{aligned}
$$

Therefore, the sequence of functions $\{B_i(\cdot)\}_{i=1}^m$ solves the backward recursive ODE system ($k = 1, \ldots, n-1$)

$$\begin{aligned}
-\dot{B}_{2n}(t) + \delta_1 B_{2n}(t) = 0,\ B_{2n}(T) = v_{2n},\\
-\dot{B}_{2n-1}(t) + \delta_2 B_{2n-1}(t) = 0,\ B_{2n-1}(T) = v_{2n-1},\\
-\dot{B}_{2k-1}(t) + \delta_1 B_{2k-1}(t) + (\hat{g}_{1,1}(B_{2k+1}(t)) - 1) + (\hat{g}_{2,1}(B_{2k+2}(t)) - 1) = 0,\ B_{2k-1}(T) = v_{2k-1},\\
-\dot{B}_{2k}(t) + \delta_2 B_{2k}(t) + (\hat{g}_{1,2}(B_{2k+1}(t)) - 1) + (\hat{g}_{2,2}(B_{2k+2}(t)) - 1) = 0,\ B_{2k}(T) = v_{2k}.
\end{aligned}$$

Moreover,

$$\dot{c}_m(t) + \rho_1 \left(\hat{h}_1(B_1(t)) - 1\right) + \rho_2 \left(\hat{h}_2(B_2(t)) - 1\right) \quad = \quad 0, \quad c_m(0) = 0.$$

We transform the system that is backward in the system index and the time into a forward system by taking

$$l_k(t) := B_{m+1-k}(T-t) = B_{2n+1-k}(T-t).$$

By construction, $\Lambda_0^{(1)} = \lambda_0^1, \Lambda_0^{(2)} = \lambda_0^2$ and $\Lambda_0^{(j)} \equiv 0$ for $j > 2$, then the Laplace transform becomes

$$\mathbb{E}_0\left[e^{-v_1\Lambda_T^{(1)} - \cdots - v_m\Lambda_T^{(m)}}\right] = e^{-l_m(T)\Lambda_0^{(1)} - \cdots - l_1(T)\Lambda_0^{(m)} - c_m(T)} = e^{-l_{2n}(T)\lambda_0^1 - l_{2n-1}(T)\lambda_0^2 - c_m(T)}, \tag{4.3}$$

where $l_i(t)$ and $c_m(t)$ solve the forward ODE system: with $k = 1, \ldots, n-1$,

$$\begin{aligned}
\dot{l}_1(t) + \delta_2 l_1(t) =&\ 0, \quad l_1(0) = v_{2n},\\
\dot{l}_2(t) + \delta_1 l_2(t) =&\ 0, \quad l_2(0) = v_{2n-1},\\
\dot{l}_{2k+1}(t) + \delta_2 l_{2k+1}(t) - (1 - \hat{g}_{1,2}(l_{2k}(t))) - (1 - \hat{g}_{2,2}(l_{2k-1}(t))) =&\ 0, \quad l_{2k+1}(0) = v_{2(n-k)},\\
\dot{l}_{2k+2}(t) + \delta_1 l_{2k+2}(t) - (1 - \hat{g}_{1,1}(l_{2k}(t))) - (1 - \hat{g}_{2,1}(l_{2k-1}(t))) =&\ 0, \quad l_{2k+2}(0) = v_{2(n-k)-1},\\
\dot{c}_m(t) - \rho_1\left(1 - \hat{h}_1(l_{2n}(t))\right) - \rho_2\left(1 - \hat{h}_2(l_{2n-1}(t))\right) =&\ 0, \quad c_m(0) = 0.
\end{aligned} \tag{4.4}$$

Note that the ODE system (4.4) has a unique and explicit solution in a recursive form

$$\begin{aligned}
l_1(t) &= v_{2n}e^{-\delta_2 t}, \\
l_2(t) &= v_{2n-1}e^{-\delta_1 t}, \\
l_{2k+1}(t) &= v_{2(n-k)}e^{-\delta_2 t} + e^{-\delta_2 t}\int_0^t e^{\delta_2 s}\left[1-\hat{g}_{1,2}(l_{2k}(s)) + 1 - \hat{g}_{2,2}(l_{2k-1}(s))\right] ds, \\
l_{2k+2}(t) &= v_{2(n-k)-1}e^{-\delta_1 t} + e^{-\delta_1 t}\int_0^t e^{\delta_1 s}\left[1-\hat{g}_{1,1}(l_{2k}(s)) + 1 - \hat{g}_{2,1}(l_{2k-1}(s))\right] ds.
\end{aligned} \tag{4.5}$$

Moreover,

$$c_m(T) = \rho_1 \int_0^T \left[1-\hat{h}_1(l_{2n}(t))\right] dt + \rho_2 \int_0^T \left[1-\hat{h}_2(l_{2n-1}(t))\right] dt.$$

4.2.2 The Limiting Distributions

The following lemma shows a necessary condition of the existence of the limiting distribution. It also indicates that the limiting distribution of the finite system is independent from the initial condition.

Lemma 4.2.1. *For any $i = 1, \ldots, m$, and $(v_1, \ldots, v_m) \in \mathbb{R}_+^m$,*

$$\lim_{t\to\infty} l_i(t) = 0.$$

Proof. See Section 4.5. □

Then the Laplace transform (4.3) of the limiting distribution of the univariate finite system $(\Lambda^{(1)}, \ldots, \Lambda^{(m)})$ at any $(v_1, \ldots, v_m) \in \mathbb{R}_+^m$ becomes

$$\begin{aligned}
\hat{\pi}_A^m(v_1, \ldots, v_m) &:= \lim_{T\to\infty} \mathbb{E}_0\left[e^{-v_1\Lambda_T^{(1)} - \cdots - v_m\Lambda_T^{(m)}}\right] \\
&= \exp\left(-\rho_1 \int_0^\infty \left[1-\hat{h}_1(l_{2n}(t))\right] dt - \rho_2 \int_0^\infty \left[1-\hat{h}_2(l_{2n-1}(t))\right] dt\right).
\end{aligned} \tag{4.6}$$

By Lévy's continuity theorem, the limiting distribution exists and is non-degenerate if $\hat{\pi}_A^m(v_1, \ldots, v_m) > 0$. In the following theorem, we provide the existence condition for the limiting distributions.

Theorem 4.2.2 (Existence of Limiting Distributions).
(1) Under Condition (C1) in Condition 3.3.2, as $t \to \infty$, the limiting distributions π_A^m of $\left(\Lambda^{(1)}, \ldots, \Lambda^{(m)}\right)$ and μ_A^n of $(\lambda^{1,n}, \lambda^{2,n})$ exist.
(2) Under Condition (C1) and (C2) in Condition 3.3.2, as $t \to \infty$, the limiting distribution μ_A^* of (λ^1, λ^2) exits.

Proof. (1) Since for any $k \in \mathbb{N}$,

$$1 - \hat{h}_1\left(l_{2k}(t)\right) = \int_t^\infty d\hat{h}_1\left(l_{2k}(u)\right) = \int_t^\infty \hat{h}_1'(l_{2k}(u))\dot{l}_{2k}(u)du \le \mu_{H^1} l_{2k}(t). \tag{4.7}$$

Hence $\int_0^\infty \left[1 - \hat{h}_1\left(l_{2k}(t)\right)\right] dt \le \mu_{H^1} \int_0^\infty l_{2k}(t)dt$ and (4.6) becomes

$$\hat{\pi}_A^m(v_1, \ldots, v_m) \ge \exp\left(-\rho_1\mu_{H^1}\int_0^\infty l_{2n}(t)dt - \rho_2\mu_{H^2}\int_0^\infty l_{2n-1}(t)dt\right). \tag{4.8}$$

Therefore it is sufficient to show that $\begin{bmatrix} \int_0^\infty l_{2n-1}(t)dt \\ \int_0^\infty l_{2n}(t)dt \end{bmatrix}$ is finite and thus the process does not explode as $t \to \infty$.

From (4.5), for all $j = 1, \ldots, 2n$, and $t \ge 0$, the function $l_j(t)$ is increasing with the initial value v_{2n+1-j}. We construct a sequence of functions $\{L_j(t)\}_{j=1}^{2n}$ that is the solution of the forward ODE system (4.4) with larger initial values

$$L_{2k-1}(0) = v_2^* = \max_{i=1,\ldots,n} v_{2i}, \quad L_{2k}(0) = v_1^* = \max_{i=1,\ldots,n} v_{2i-1} \quad \text{for } k = 1, \ldots, n. \tag{4.9}$$

Therefore $l_j(t) \le L_j(t)$ for $j = 1, \ldots, 2n$, and $\begin{bmatrix} \int_0^\infty l_{2n-1}(t)dt \\ \int_0^\infty l_{2n}(t)dt \end{bmatrix} \le \begin{bmatrix} \int_0^\infty L_{2n-1}(t)dt \\ \int_0^\infty L_{2n}(t)dt \end{bmatrix}$.
Then it is sufficient to show that $\begin{bmatrix} \int_0^\infty L_{2n-1}(t)dt \\ \int_0^\infty L_{2n}(t)dt \end{bmatrix} < \infty$.

With the same initial value as in (4.9), from the explicit recursive solution (4.4), one can easily check by induction that for each $t \ge 0$, $L_{2k-1}(t)$ and $L_{2k}(t)$ are increasing with k. We can define non-negative distance functions

$$\begin{aligned} k = 1: \qquad & d_1^{(1)}(t) := L_1(t), \quad d_1^{(2)}(t) := L_2(t). \\ k \ge 2: \qquad & d_k^{(1)}(t) := L_{2k-1}(t) - L_{2k-3}(t), \quad d_k^{(2)}(t) := L_{2k}(t) - L_{2k-2}(t), \end{aligned}$$

hence we have $d_k^{(1)}(0) = d_k^{(2)}(0) = 0$ for all $k \ge 1$ and moreover

$$L_{2n-1}(t) = \sum_{i=1}^n d_i^{(1)}(t), \quad L_{2n}(t) = \sum_{i=1}^n d_i^{(2)}(t).$$

The following recursive inequalities are proved in Section 4.5:

$$\begin{aligned} d_{k+1}^{(1)}(t) &\le e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[\mu_{G^{2,2}} d_k^{(1)}(s) + \mu_{G^{1,2}} d_k^{(2)}(s)\right] ds, \\ d_{k+1}^{(2)}(t) &\le e^{-\delta_1 t} \int_0^t e^{\delta_1 s} \left[\mu_{G^{2,1}} d_k^{(1)}(s) + \mu_{G^{1,1}} d_k^{(2)}(s)\right] ds. \end{aligned} \tag{4.10}$$

Hence

$$\begin{aligned} \int_0^\infty d_{i+1}^{(1)}(t)dt &\le \int_{t=0}^\infty e^{-\delta_2 t} \int_{s=0}^t e^{\delta_2 s} \left(\mu_{G^{2,2}} d_i^{(1)}(s) + \mu_{G^{1,2}} d_i^{(2)}(s)\right) ds dt \\ &= \int_{s=0}^\infty \left(\int_{t=s}^\infty e^{-\delta_2 t} dt\right) e^{\delta_2 s} \left(\mu_{G^{2,2}} d_i^{(1)}(s) + \mu_{G^{1,2}} d_i^{(2)}(s)\right) ds \\ &= \frac{\mu_{G^{2,2}}}{\delta_2} \int_0^\infty d_i^{(1)}(s) ds + \frac{\mu_{G^{1,2}}}{\delta_2} \int_0^\infty d_i^{(2)}(s) ds. \end{aligned}$$

Similarly,

$$\int_0^\infty d_{i+1}^{(2)}(t)dt \le \frac{\mu_{G^{2,1}}}{\delta_1} \int_0^\infty d_i^{(1)}(s) ds + \frac{\mu_{G^{1,1}}}{\delta_1} \int_0^\infty d_i^{(2)}(s) ds.$$

i.e.

$$\begin{bmatrix} \int_0^\infty d_{i+1}^{(1)}(t)dt \\ \int_0^\infty d_{i+1}^{(2)}(t)dt \end{bmatrix} \le A \begin{bmatrix} \int_0^\infty d_i^{(1)}(t)dt \\ \int_0^\infty d_i^{(2)}(t)dt \end{bmatrix},$$

where

$$A := \begin{bmatrix} \frac{\mu_{G^{2,2}}}{\delta_2} & \frac{\mu_{G^{1,2}}}{\delta_2} \\ \frac{\mu_{G^{2,1}}}{\delta_1} & \frac{\mu_{G^{1,1}}}{\delta_1} \end{bmatrix}.$$

Iteratively, we obtain for $i \ge 1$,

$$\begin{bmatrix} \int_0^\infty d_i^{(1)}(t)dt \\ \int_0^\infty d_i^{(2)}(t)dt \end{bmatrix} \le A^{i-1} \begin{bmatrix} \int_0^\infty d_1^{(1)}(t)dt \\ \int_0^\infty d_1^{(2)}(t)dt \end{bmatrix} = A^{i-1} \begin{bmatrix} \frac{v_2^*}{\delta_2} \\ \frac{v_1^*}{\delta_1} \end{bmatrix}.$$

Denote by ρ the spectral radius of A. From matrix theory, for any $\epsilon > 0$, and $\tilde{\rho} := \rho + \epsilon$, there exists a norm $\|\cdot\|$, such that $\|A\| \le \tilde{\rho}$. Then, for any $i \ge 1$,

$$\|A^i\| \le \|A\|^i \le \tilde{\rho}^i.$$

If we take the Euclidean norm $\|\cdot\|_2$, then due to the equivalence of norm there exists a constant $C > 0$, such that

$$\|A^i\|_2 \le C\|A^i\| \le C\tilde{\rho}^i.$$

By definition,

$$\begin{bmatrix} \int_0^\infty L_{2n-1}(t)dt \\ \int_0^\infty L_{2n}(t)dt \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \int_0^\infty d_i^{(1)}(t)dt \\ \int_0^\infty d_i^{(2)}(t)dt \end{bmatrix} \leq \left(\sum_{i=1}^{n} A^{i-1} \right) \begin{bmatrix} \frac{v_2^*}{\delta_2} \\ \frac{v_1^*}{\delta_1} \end{bmatrix}.$$

Denote $\tilde{L}_n := \left\| \begin{bmatrix} \int_0^\infty L_{2n-1}(t)dt \\ \int_0^\infty L_{2n}(t)dt \end{bmatrix} \right\|_2$, then

$$\tilde{L}_n \leq \sum_{i=1}^{n} \|A^{i-1}\|_2 \left\| \begin{bmatrix} \frac{v_2^*}{\delta_2} \\ \frac{v_1^*}{\delta_1} \end{bmatrix} \right\|_2 \leq C \frac{1-\tilde{\rho}^n}{1-\tilde{\rho}} \sqrt{\left(\frac{v_2^*}{\delta_2}\right)^2 + \left(\frac{v_1^*}{\delta_1}\right)^2} < \infty. \tag{4.11}$$

Hence $\int_0^\infty L_{2n-1}(t)dt \leq \tilde{L}_n < \infty$ and $\int_0^\infty L_{2n}(t)dt \leq \tilde{L}_n < \infty$, which indicates that the limiting distribution π_A^m of $(\Lambda^{(1)}, \ldots, \Lambda^{(m)})$ exists.

The existence of the limiting distribution μ_A^n of $(\lambda^{1,n}, \lambda^{2,n})$ is implied from the analysis above. Indeed, by taking $v_{2i-1} = v_1$ and $v_{2i} = v_2$ for $i = 1, \ldots, n$, then the Laplace transform in (4.6) becomes

$$\begin{aligned} \hat{\mu}_A^n(v_1, v_2) &:= \lim_{T\to\infty} \mathbb{E}_0 \left[e^{-v_1 \lambda_T^{1,n} - v_2 \lambda_T^{2,n}} \right] \\ &= \exp\left(-\rho_1 \int_0^\infty \left[1 - \hat{h}_1(l_{2n}(t)) \right] dt - \rho_2 \int_0^\infty \left[1 - \hat{h}_2(l_{2n-1}(t)) \right] dt \right), \end{aligned} \tag{4.12}$$

where $l_{2n-1}(t)$, $l_{2n}(t)$ are from the solution of the ODE system (4.4) with initial values $l_{2i-1}(0) = v_1$ and $l_{2i}(0) \equiv v_2$ for $i = 1, \ldots, n$.

In this case, $l_j(t) = L_j(t)$ for $j = 1, \ldots, m$, and hence the limiting distribution μ_A^n exists.

(2) We explore the existence condition for the limiting distribution μ_A^* of (λ^1, λ^2) using the convergence of μ_A^n when the system index $n \to \infty$.

Note that for the Laplace transform $\hat{\mu}_A^n$ in (4.12), $l_{2n}(t)$ and $l_{2n-1}(t)$ are from the explicit solution to (4.4) with $l_{2i-1}(0) = v_1$ and $l_{2i}(0) \equiv v_2$ for $i = 1, \ldots, n$, i.e., for $k = 0, \ldots, n-1$,

$$\begin{aligned} l_1(t) &= v_2 e^{-\delta_2 t}, \quad l_2(t) = v_1 e^{-\delta_1 t}, \\ l_{2k+1}(t) &= v_2 e^{-\delta_2 t} + e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[1 - \hat{g}_{1,2}(l_{2k}(s)) + 1 - \hat{g}_{2,2}(l_{2k-1}(s)) \right] ds, \\ l_{2k+2}(t) &= v_1 e^{-\delta_1 t} + e^{-\delta_1 t} \int_0^t e^{\delta_1 s} \left[1 - \hat{g}_{1,1}(l_{2k}(s)) + 1 - \hat{g}_{2,1}(l_{2k-1}(s)) \right] ds. \end{aligned} \tag{4.13}$$

Note that for any $t \geq 0$, $l_{2k-1}(t)$ and $l_{2k}(t)$ are increasing functions of k for $k \geq 1$, hence by the monotone convergence theorem, $(l_{2n-1}(t), l_{2n}(t))$ converges to a limit $(l_1^*(t), l_2^*(t))$. Note that the convergence is uniform in t, hence $(l_1^*(\cdot), l_2^*(\cdot))$ is a continuous function. Then again by the monotone convergence theorem, the Laplace transform of the limiting distribution of (λ^1, λ^2) is

$$\begin{aligned}\hat{\mu}_A^*(v_1, v_2) &= \lim_{n\to\infty} \hat{\mu}_A^n(v_1, v_2) \\ &= \exp\left(-\rho_1 \int_0^\infty \left[1 - \hat{h}_1(l_2^*(t))\right] dt - \rho_2 \int_0^\infty \left[1 - \hat{h}_2(l_1^*(t))\right] dt\right).\end{aligned} \tag{4.14}$$

To show μ_A^* is non-degenerate, following the same argument as in (4.7), it is sufficient to have $\begin{bmatrix}\int_0^\infty l_1^*(t)dt \\ \int_0^\infty l_2^*(t)dt\end{bmatrix} < \infty$.

Since the spectral radius $\rho < 1$ under (C2) in Condition 3.3.2, take $0 < \epsilon < \frac{1-\rho}{2}$, and by the matrix theory there exists a norm $\|\cdot\|$, such that $\|A\| \leq \tilde{\rho} = \rho + \epsilon < 1$, then from (4.11) in the first part of proof,

$$\tilde{L}_n \leq C\frac{1-\tilde{\rho}^n}{1-\tilde{\rho}}\sqrt{\left(\frac{v_2}{\delta_2}\right)^2 + \left(\frac{v_1}{\delta_1}\right)^2} < C\frac{1}{1-\tilde{\rho}}\sqrt{\left(\frac{v_2}{\delta_2}\right)^2 + \left(\frac{v_1}{\delta_1}\right)^2} < \infty.$$

Since $\tilde{L}_n$ as the upper bound of $\int_0^\infty l_{2n-1}(t)dt$ and $\int_0^\infty l_{2n}(t)dt$ is uniformly bounded in $n \geq 1$, we have

$$\begin{bmatrix}\int_0^\infty l_1^*(t)dt \\ \int_0^\infty l_2^*(t)dt\end{bmatrix} = \lim_{n\to\infty}\begin{bmatrix}\int_0^\infty l_{2n-1}(t)dt \\ \int_0^\infty l_{2n}(t)dt\end{bmatrix} < \infty.$$

□

4.3 The Stationary Distributions

The limiting distributions of the finite system and BDCP exist by Theorem 4.2.2. In this section, we show the equivalence between the stationary distribution and the limiting distribution.

First, a stationarity condition for the finite system $\left(\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)}\right)$ is provided based on Markov process theory in [33] for example.

Lemma 4.3.1 (Stationary condition equation for the finite system).
Distribution π_S^m is a stationary distribution of $(\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)})$ if and only if the Laplace transform $\hat{\pi}_S^m$ at any $(v_1, \ldots, v_m) \in \mathbb{R}_+^m$ satisfies

$$\begin{aligned} 0 \;=\; & -\sum_{k=1}^{m} \delta_k v_k \hat{\pi}^m(v_1, \ldots, v_m) + \rho_1(\hat{h}_1(v_1) - 1) + \rho_2(\hat{h}_2(v_2) - 1) \\ & + \sum_{k=1}^{n-1} \frac{\partial \hat{\pi}^m(v_1, \ldots, v_m)}{\partial v_{2k-1}} \left[(1 - \hat{g}_{1,1}(v_{2k+1})) + (1 - \hat{g}_{2,1}(v_{2k+2}))\right] \\ & + \sum_{k=1}^{n-1} \frac{\partial \hat{\pi}^m(v_1, \ldots, v_m)}{\partial v_{2k}} \left[(1 - \hat{g}_{1,2}(v_{2k+1})) + (1 - \hat{g}_{2,2}(v_{2k+2}))\right]. \end{aligned}$$

Equivalently, the equation can be written in terms of the ODE system (4.4):

$$0 = \sum_{k=1}^{n} \dot{l}_{2(n-k)+2}(0) \frac{\partial \hat{\pi}_S^m}{\partial v_{2k-1}} + \dot{l}_{2(n-k)+1}(0) \frac{\partial \hat{\pi}_S^m}{\partial v_{2k}} - \rho_1(1 - \hat{h}_1(v_1))\hat{\pi}_S^m - \rho_2(1 - \hat{h}_2(v_2))\hat{\pi}_S^m. \tag{4.15}$$

Proof. See Section 4.5. □

The following theorem states the equivalence between the limiting and stationary distribution, and the proof is based on the self-similarity structure of the system.

Theorem 4.3.2 (Stationarity of the finite system).
For any system index $m = 2n$ for $\left(\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)}\right)$, if there exits a limiting distribution π_A^m, then there exits a unique stationary distribution π_S^m and $\pi_S^m \overset{d}{=} \pi_A^m$.

Proof. For the existence, it is sufficient to show that $\hat{\pi}^m := \hat{\pi}_A^m$ satisfies the condition equation (4.15). The uniqueness of such π_S^m follows immediately from the uniqueness of π_A^m.

Since $\lim_{t\to\infty} l_{2n}(t) = \lim_{t\to\infty} l_{2n-1}(t) = 0$ from Lemma 4.2.1,

$$\begin{aligned} \frac{\partial \hat{\pi}^m}{\partial v_{2k-1}} &= \hat{\pi}^m \left[\rho_1 \int_0^\infty \hat{h}_1'(l_{2n}(t)) \frac{\partial l_{2n}(t)}{\partial v_{2k-1}} dt + \rho_2 \int_0^\infty \hat{h}_2'(l_{2n-1}(t)) \frac{\partial l_{2n-1}(t)}{\partial v_{2k-1}} dt\right] \\ \frac{\partial \hat{\pi}^m}{\partial v_{2k}} &= \hat{\pi}^m \left[\rho_1 \int_0^\infty \hat{h}_1'(l_{2n}(t)) \frac{\partial l_{2n}(t)}{\partial v_{2k}} dt + \rho_2 \int_0^\infty \hat{h}_2'(l_{2n-1}(t)) \frac{\partial l_{2n-1}(t)}{\partial v_{2k}} dt\right] \\ 1 - \hat{h}_1(v_1) &= \int_0^\infty \hat{h}_1'(l_{2n}(t)) \dot{l}_{2n}(t) dt \\ 1 - \hat{h}_2(v_2) &= \int_0^\infty \hat{h}_2'(l_{2n-1}(t)) \dot{l}_{2n-1}(t) dt. \end{aligned}$$

Then, the stationarity equation (4.15) becomes

$$0=\rho_1\int_0^\infty \hat{h}_1'(l_{2n}(t))\left[\sum_{k=1}^{n}\left(\dot{l}_{2(n-k)+2}(0)\frac{\partial l_{2n}(t)}{\partial v_{2k-1}}+\dot{l}_{2(n-k)+1}(0)\frac{\partial l_{2n}(t)}{\partial v_{2k}}\right)-\dot{l}_{2n}(t)\right]dt$$
$$+\rho_2\int_0^\infty \hat{h}_2'(l_{2n-1}(t))\left[\sum_{k=1}^{n}\left(\dot{l}_{2(n-k)+2}(0)\frac{\partial l_{2n-1}(t)}{\partial v_{2k-1}}+\dot{l}_{2(n-k)+1}(0)\frac{\partial l_{2n-1}(t)}{\partial v_{2k}}\right)-\dot{l}_{2n-1}(t)\right]dt.$$

We observe that the functions l_k is independent from the choice of $\hat{h}_i$ and ρ_i for $i=1,2$, then for system index $m=2n$, re-denote $l_k(\cdot)$ as $l_k^{2n}(\cdot)$, then it is sufficient to show:

$$\begin{aligned}\sum_{k=1}^{n}\left(\dot{l}^{2n}_{2(n-k)+2}(0)\frac{\partial l^{2n}_{2n}(t)}{\partial v_{2k-1}}+\dot{l}^{2n}_{2(n-k)+1}(0)\frac{\partial l^{2n}_{2n}(t)}{\partial v_{2k}}\right)-\dot{l}^{2n}_{2n}(t)&=0,\\ \sum_{k=1}^{n}\left(\dot{l}^{2n}_{2(n-k)+2}(0)\frac{\partial l^{2n}_{2n-1}(t)}{\partial v_{2k-1}}+\dot{l}^{2n}_{2(n-k)+1}(0)\frac{\partial l^{2n}_{2n-1}(t)}{\partial v_{2k}}\right)-\dot{l}^{2n}_{2n-1}(t)&=0.\end{aligned}\tag{4.16}$$

By observing the self-similarity in the system structure, we prove (4.16) using the induction with respect to the system index $m=2n$.

(1) For $n=1$, it is easy to observe that $l_1^2(t)=v_2e^{-\delta_2 t}$ and $l_2^2(t)=v_1e^{-\delta_1 t}$ satisfies (4.16).

(2) Assume that $m=2n$ satisfies (4.16), we show that for $m=2(n+1)$ also satisfies (4.16). We need to show the first equation and the second follows in the same way. I.e.,

$$\sum_{k=1}^{n+1}\left[\dot{l}^{2(n+1)}_{2(n+1-k)+2}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k-1}}+\dot{l}^{2(n+1)}_{2(n+1-k)+1}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k}}\right]-\dot{l}^{2(n+1)}_{2n+2}(t)=0,\tag{4.17}$$

where $l^{2(n+1)}_{2(n+1)-i+1}(0)=v_i$ for $i=1,\dots,2n$.
Note that from the ODE system and the recursive solution, we have

$$l^{2(n+1)}_{2(n+1)}(0)=v_1,\qquad \frac{\partial l^{2(n+1)}_{2(n+1)}(t)}{\partial v_1}=e^{-\delta_1 t},\qquad \frac{\partial l^{2(n+1)}_{2(n+1)}(t)}{\partial v_2}=0.$$

Then, the $k=1$ term in (4.17) becomes $\dot{l}^{2(n+1)}_{2n+2}(0)e^{-\delta_1 t}$. Hence, we need to show that

$$\sum_{k=2}^{n+1}\left[\dot{l}^{2(n+1)}_{2(n+1-k)+2}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k-1}}+\dot{l}^{2(n+1)}_{2(n+1-k)+1}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k}}\right]=\dot{l}^{2(n+1)}_{2n+2}(t)-e^{-\delta_1 t}\dot{l}^{2(n+1)}_{2n+2}(0).$$

As (4.16) holds for $m = 2n$ for all $(v_1, \ldots, v_{2n}) \in \mathbb{R}_+^m$, one can construct functions $\{L_i^{2n}(\cdot)\}_{i=1}^{2n}$, such that they satisfies (4.16) with initial values

$$(\tilde{v}_1, \ldots, \tilde{v}_{2n}) = (v_3, \ldots, v_{2n+2}).$$

Hence,

$$\begin{aligned}\sum_{k=1}^{n}\left(\dot{L}^{2n}_{2(n-k)+2}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-1}} + \dot{L}^{2n}_{2(n-k)+1}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k}}\right) - \dot{L}^{2n}_{2n}(t) = 0,\\ \sum_{k=1}^{n}\left(\dot{L}^{2n}_{2(n-k)+2}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k-1}} + \dot{L}^{2n}_{2(n-k)+1}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k}}\right) - \dot{L}^{2n}_{2n-1}(t) = 0.\end{aligned} \tag{4.18}$$

with $L^{2n}_{2n-i+1}(0) = \tilde{v}_i = v_{i+2}$ for $i = 1, \ldots, 2n$. Especially $L_1^{2n}(0) = v_{2n+2}$ and $L_2^{2n}(0) = v_{2n+1}$.
By construction, for $k = 1, \ldots, n$, $t \geq 0$,

$$l^{2(n+1)}_{2k-1}(t) = L^{2n}_{2k-1}(t), \quad l^{2(n+1)}_{2k}(t) = L^{2n}_{2k}(t),$$

and

$$\frac{\partial l^{2(n+1)}_{2n}(t)}{\partial v_{2k-1}} = \frac{\partial L^{2n}_{2n}(t)}{\partial v_{2k-1}} = \frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-3}}, \quad \frac{\partial l^{2(n+1)}_{2n}(t)}{\partial v_{2k}} = \frac{\partial L^{2n}_{2n}(t)}{\partial v_{2k}} = \frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-2}}.$$

For terms with $k \geq 2$ in in (4.17),

$$\begin{aligned}&\sum_{k=2}^{n+1}\left[\dot{l}^{2(n+1)}_{2(n+1-k)+2}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k-1}} + \dot{l}^{2(n+1)}_{2(n+1-k)+1}(0)\frac{\partial l^{2(n+1)}_{2n+2}(t)}{\partial v_{2k}}\right]\\ &= -\int_0^t e^{-\delta_1(t-s)}\hat{g}'_{1,1}\left(L^{2n}_{2n}(s)\right)\sum_{k=2}^{n+1}\left[\dot{L}^{2n}_{2(n+1-k)+2}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-3}} + \dot{L}^{2n}_{2(n+1-k)+1}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-2}}\right] ds\\ &\quad - \int_0^t e^{-\delta_1(t-s)}\hat{g}'_{2,1}\left(L^{2n}_{2n-1}(s)\right)\sum_{k=2}^{n+1}\left[\dot{L}^{2n}_{2(n+1-k)+2}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k-3}} + \dot{L}^{2n}_{2(n+1-k)+1}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k-2}}\right] ds\\ &= -\int_0^t e^{-\delta_1(t-s)}\hat{g}'_{1,1}\left(L^{2n}_{2n}(s)\right)\sum_{k=1}^{n}\left[\dot{L}^{2n}_{2(n-k)+2}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k-1}} + \dot{L}^{2n}_{2(n-k)+1}(0)\frac{\partial L^{2n}_{2n}(t)}{\partial \tilde{v}_{2k}}\right] ds\\ &\quad - \int_0^t e^{-\delta_1(t-s)}\hat{g}'_{2,1}\left(L^{2n}_{2n-1}(s)\right)\sum_{k=1}^{n}\left[\dot{L}^{2n}_{2(n-k)+2}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k-1}} + \dot{L}^{2n}_{2(n-k)+1}(0)\frac{\partial L^{2n}_{2n-1}(t)}{\partial \tilde{v}_{2k}}\right] ds.\end{aligned} \tag{4.19}$$

By (4.18), (4.19) becomes

$$
\begin{aligned}
& -\int_0^t e^{-\delta_1(t-s)}\hat{g}'_{1,1}\left(L^{2n}_{2n}(s)\right)\dot{L}^{2n}_{2n}(s)ds-\int_0^t e^{-\delta_1(t-s)}\hat{g}'_{2,1}\left(L^{2n}_{2n-1}(s)\right)\dot{L}^{2n}_{2n-1}(s)ds \\
= & \int_0^t e^{-\delta_1(t-s)}\frac{\partial}{\partial s}\left[1-\hat{g}_{1,1}\left(L^{2n}_{2n}(s)\right)+1-\hat{g}_{2,1}\left(L^{2n}_{2n-1}(s)\right)\right]ds \\
= & \int_0^t e^{-\delta_1(t-s)}\frac{\partial}{\partial s}\left[1-\hat{g}_{1,1}\left(l^{2(n+1)}_{2n}(s)\right)+1-\hat{g}_{2,1}\left(l^{2(n+1)}_{2n-1}(s)\right)\right]ds \\
\overset{(4.4)}{=} & \int_0^t e^{-\delta_1(t-s)}\frac{\partial}{\partial s}\left[\dot{l}^{2(n+1)}_{2n+2}(s)+\delta_1 l^{2(n+1)}_{2n+2}(s)\right]ds
\end{aligned}
$$

Denote

$$
F(s):=e^{\delta_1 s}l^{2(n+1)}_{2n+2}(s),
$$

then $\dot{F}(s)=e^{\delta_1 s}\left(\dot{l}^{2(n+1)}_{2n+2}(s)+\delta_1 l^{2(n+1)}_{2n+2}(s)\right)$, and $\dot{F}(0)=\dot{l}^{2(n+1)}_{2n+2}(0)+\delta_1 v_1$.

$$
\begin{aligned}
(4.19) &= e^{-\delta_1 t}\int_0^t e^{\delta_1 s}d\left(e^{-\delta_1 s}\dot{F}(s)\right) \\
&= e^{-\delta_1 t}\left(\dot{F}(t)-\dot{F}(0)\right)-\delta_1 e^{-\delta_1 t}\left(F(t)-F(0)\right) \\
&= \left(\dot{l}^{2(n+1)}_{2n+2}(t)+\delta_1 l^{2(n+1)}_{2n+2}(t)\right)-e^{-\delta_1 t}\left(\dot{l}^{2(n+1)}_{2n+2}(0)+\delta_1 v_1\right)-\delta_1 e^{-\delta_1 t}\left(e^{\delta_1 t}l^{2(n+1)}_{2n+2}(t)-v_1\right) \\
&= \dot{l}^{2(n+1)}_{2n+2}(t)-e^{-\delta_1 t}\dot{l}^{2(n+1)}_{2n+2}(0).
\end{aligned}
$$

□

Hence, by Theorem 4.3.2 and Theorem 4.2.2, we can conclude that for the finite system $\left(\Lambda^{(1)}_t,\ldots,\Lambda^{(m)}_t\right)$, there exists a unique stationary distribution that is equal to the limiting distribution.

Corollary 4.3.3. *There exists a unique stationary distribution μ^n_S of $(\lambda^{1,n},\lambda^{2,n})$, and it is equal to the limiting distribution μ^n_A with the Laplace transform (4.12).*

Proof. The joint distribution of $\left(\lambda^{1,n}_t,\lambda^{2,n}_t\right)$ at any $t\geq 0$, $k\geq 0$ and $0\leq s_1\leq\ldots\leq s_k$ is

$$
\begin{aligned}
& \mathbb{P}\left(\lambda^{1,n}_{t+s_1}\leq x^1_1,\lambda^{2,n}_{t+s_1}\leq x^2_1;\ldots;\lambda^{1,n}_{t+s_k}\leq x^1_k,\lambda^{2,n}_{t+s_k}\leq x^2_k\right) \\
= & \mathbb{P}\left(\sum_{i=1}^n\lambda^{1,(i)}_{t+s_1}\leq x^1_1,\sum_{i=1}^n\lambda^{2,(i)}_{t+s_1}\leq x^2_1;\ldots;\sum_{i=1}^n\lambda^{1,(i)}_{t+s_k}\leq x^1_k,\sum_{i=1}^n\lambda^{2,(i)}_{t+s_k}\leq x^2_k\right) \\
= & \int_{D^1_1}\int_{D^2_1}\cdots\int_{D^1_k}\int_{D^2_k}d\mathbb{P}\left(\lambda^{1,(1)}_{t+s_1}\leq z^{1,(1)}_1,\lambda^{2,(1)}_{t+s_1}\leq z^{2,(1)}_1,\ldots,\lambda^{1,(n)}_{t+s_k}\leq z^{1,(n)}_k,\lambda^{2,(n)}_{t+s_k}\leq z^{2,(n)}_k\right),
\end{aligned}
$$

where for $j = 1, \ldots, k$, $j' = 1, 2$, $D_j^{j'} = \left\{ \left(z_j^{j',(1)}, \ldots, z_j^{j',(n)} \right) \in \mathbb{R}^n : \sum_{i=1}^n z_j^{j',(i)} \leq x_j^{j'} \right\}$.

By Theorem 4.3.2, take the distribution of $\left(\lambda_t^{1,(1)}, \lambda_t^{2,(1)}, \ldots, \lambda_t^{1,(n)}, \lambda_t^{2,(n)} \right) = \left(\Lambda^{(1)}, \ldots, \Lambda^{(m)} \right)$ as the unique stationary distribution π_S^m, then the joint distribution above is independent of t. Hence from the equation above, the distribution $\left(\lambda_{t+s_1}^{1,n}, \lambda_{t+s_1}^{2,n}; \ldots; \lambda_{t+s_k}^{1,n}, \lambda_{t+s_k}^{2,n} \right)$ is also independent of t. Therefore by definition $(\lambda^{1,n}, \lambda^{2,n})$ is a stationary process.

Since the limiting distribution exits and is independent of the initial value, then $\mu_S^n \stackrel{d}{=} \mu_A^n$ and the uniqueness follows. □

Now we present the existence and uniqueness of stationary distribution for $(\lambda_t^1, \lambda_t^2)$ which is the final result of this chapter.

Theorem 4.3.4 (Existence of Stationary Distribution)**.** *Under the Condition (C1) and (C2) in Condition 3.3.2, there exists a unique stationary distribution μ_S^* of the BDCP intensity (λ^1, λ^2), and moreover $\mu_S^* \stackrel{d}{=} \mu_A^*$ where μ_A^* is the limiting distribution of (λ^1, λ^2).*

Proof. Recall that $\lambda_t^{1,n} = \sum_{i=1}^n \Lambda_t^{(2i-1)} = \sum_{i=1}^n \lambda_t^{1,(i)}$ and $\lambda_t^{2,n} = \sum_{i=1}^n \Lambda_t^{(2i)} = \sum_{i=1}^n \lambda_t^{2,(i)}$. Let $(\lambda^{1,n}, \lambda^{2,n})$ starts from the stationary distribution μ_S^n, then for any $t \geq 0$, $0 \leq s_1 \leq \ldots \leq s_k$,

$$\left(\lambda_{t+s_1}^{1,n}, \lambda_{t+s_1}^{2,n}; \ldots; \lambda_{t+s_k}^{1,n}, \lambda_{t+s_k}^{2,n} \right) \stackrel{d}{=} \left(\lambda_{s_1}^{1,n}, \lambda_{s_1}^{2,n}; \ldots; \lambda_{s_k}^{1,n}, \lambda_{s_k}^{2,n} \right). \tag{4.20}$$

Since the convergence of $(\lambda^{1,n}, \lambda^{2,n})$ to (λ^1, λ^2) pathwise implies the convergence of the finite dimensional distributions, as $n \to \infty$,

$$\begin{aligned} \left(\lambda_{t+s_1}^{1,n}, \lambda_{t+s_1}^{2,n}; \ldots; \lambda_{t+s_k}^{1,n}, \lambda_{t+s_k}^{2,n} \right) &\Rightarrow \left(\lambda_{t+s_1}^{1}, \lambda_{t+s_1}^{2}; \ldots; \lambda_{t+s_k}^{1}, \lambda_{t+s_k}^{2} \right), \quad n \to \infty, \\ \left(\lambda_{s_1}^{1,n}, \lambda_{s_1}^{2,n}; \ldots; \lambda_{s_k}^{1,n}, \lambda_{s_k}^{2,n} \right) &\Rightarrow \left(\lambda_{s_1}^{1}, \lambda_{s_1}^{2}; \ldots; \lambda_{s_k}^{1}, \lambda_{s_k}^{2} \right), \quad n \to \infty. \end{aligned}$$

where the left hand side distribution is π_S^m. By (4.20) and uniqueness of the weak limit, we have the limiting process

$$\left(\lambda_{t+s_1}^{1}, \lambda_{t+s_1}^{2}; \ldots; \lambda_{t+s_k}^{1}, \lambda_{t+s_k}^{2} \right) \stackrel{d}{=} \left(\lambda_{s_1}^{1}, \lambda_{s_1}^{2}; \ldots; \lambda_{s_k}^{1}, \lambda_{s_k}^{2} \right).$$

i.e. the finite dimensional distribution is independent of t. Hence (λ^1, λ^2) has a stationary distribution μ_S^*.

Since the limiting distribution μ_A^* exists and is independent of the initial value, then $\mu_S^* \stackrel{d}{=} \mu_A^*$. As μ_A^* is unique, the uniqueness of μ_S^* follows.

□

Remark 4.3.5. *Note that for Hawkes processes defined on the whole real line $\mathbb{R}$, the Theorem 7 in Brémaud and Massoulié [14] has a similar result that can be covered as a special case of ours for the bivariate case.*

Remark 4.3.6 (Importance of the finite system)**.** *From the analysis above, in addition to the BDCP system (λ^1, λ^2), we also provide the distribution of the non-stationary and stationary version of $(\lambda^{1,n}, \lambda^{2,n})$ in terms of the Laplace transform. Note that the contagion effect is more severe when the index n is increasing, therefore one can tune the the contagion effect for modelling purpose using the finite system $(N^{1,n}, N^{2,n})$ with intensity $(\lambda^{1,n}, \lambda^{2,n})$ by choosing a certain system index n. Therefore, it is an interesting process itself for applications and further analysis.*

For any $h > 0$, $N_{t_1+h} - N_{t_1}|_{\lambda_{t_1}=\lambda} \stackrel{d}{=} N_{t_2+h} - N_{t_2}|_{\lambda_{t_2}=\lambda}$. If λ is stationary, then $\lambda_{t_1} \stackrel{d}{=} \lambda_{t_2}$, and hence $N_{t_1+h} - N_{t_1} \stackrel{d}{=} N_{t_2+h} - N_{t_2}$. Hence we have the following results:

Corollary 4.3.7. *The BDCP $N = (N_t)_{t\geq 0}$ has stationary increments.*

4.4 Stationary Moments

From Theorem 4.3.4, the intensity process $\lambda = (\lambda^1, \lambda^2)$ has a unique stationary distribution μ_S^*. In this section, we show the first and second moments of the stationary distribution based on Markov process theory.

Since λ is a Markov process with the generator $\mathcal{A}_\lambda$ in (3.13), then by the stationary distribution characterization in (4.1), for any $f \in \mathcal{D}(\mathcal{A}_\lambda)$ we have

$$\int_0^\infty \int_0^\infty \mathcal{A}f(\lambda_1, \lambda_2)\mu_S^*(\lambda_1, \lambda_2)d\lambda_1 d\lambda_2 = 0. \tag{4.21}$$

In the following, denote $\mathcal{A} := \mathcal{A}_\lambda$, we use (4.21) to derive the stationary mean, variance and correlation. Moreover, in addition to (C1) in Condition 3.3.2 where the first moments of all jump sizes exist, we assume the second moments also exist. I.e. $\mu_{2H^k} < \infty$ and $\mu_{2G^{k,k'}} < \infty$ for $k, k' = 1, 2$.

4.4.1 Stationary Mean

Take $\mathcal{A}f(\lambda_1,\lambda_2)=\lambda_i$ and denote by $m_1^i=\mathbb{E}[\lambda_t^i]$ the stationary mean for $i=1,2$, we have

$$\begin{aligned}-(\delta_1-\mu_{G^{1,1}})m_1^1+\mu_{G^{1,2}}m_1^2+\rho_1\mu_{H^1} &= 0,\\ -(\delta_2-\mu_{G^{2,2}})m_1^2+\mu_{G^{2,1}}m_1^1+\rho_2\mu_{H^2} &= 0.\end{aligned}$$

Solving this linear equation system we obtain the stationary mean

$$\begin{aligned}m_1^1&=\frac{(\delta_2-\mu_{G^{2,2}})\mu_{H^1}}{(\delta_1-\mu_{G^{1,1}})(\delta_2-\mu_{G^{2,2}})-\mu_{G^{1,2}}\mu_{G^{2,1}}}\rho_1+\frac{\mu_{G^{1,2}}\mu_{H^2}}{(\delta_1-\mu_{G^{1,1}})(\delta_2-\mu_{G^{2,2}})-\mu_{G^{1,2}}\mu_{G^{2,1}}}\rho_2,\\ m_1^2&=\frac{(\delta_1-\mu_{G^{1,1}})\mu_{H^2}}{(\delta_1-\mu_{G^{1,1}})(\delta_2-\mu_{G^{2,2}})-\mu_{G^{1,2}}\mu_{G^{2,1}}}\rho_2+\frac{\mu_{G^{2,1}}\mu_{H^1}}{(\delta_1-\mu_{G^{1,1}})(\delta_2-\mu_{G^{2,2}})-\mu_{G^{1,2}}\mu_{G^{2,1}}}\rho_1.\end{aligned}$$

For $i=1,2$, denote

$$\Delta_i=\delta_i-\mu_{G^{i,i}},\quad \Delta:=\Delta_1\Delta_2-\mu_{G^{1,2}}\mu_{G^{2,1}},$$

then we can rewrite the first moments as

$$\begin{aligned}m_1^1&=\frac{\Delta_2\mu_{H^1}}{\Delta}\rho_1+\frac{\mu_{G^{1,2}}\mu_{H^2}}{\Delta}\rho_2=:\mu_{1,1}\rho_1+\mu_{1,2}\rho_2\\ m_1^2&=\frac{\mu_{G^{2,1}}\mu_{H^1}}{\Delta}\rho_1+\frac{\Delta_1\mu_{H^2}}{\Delta}\rho_2=:\mu_{2,1}\rho_1+\mu_{2,2}\rho_2.\end{aligned}$$

Remark 4.4.1. *If there is no cross-exciting term, i.e.* $\mu_{G^{i,j}}=0$ *for* $i\neq j$*, then the result recovers that of the univariate DCP in Dassios and Zhao [23].*

4.4.2 Stationary Variance

We consider the second stationary moments: $\mathbb{E}[(\lambda_t^1)^2]$, $\mathbb{E}[(\lambda_t^2)^2]$ and $\mathbb{E}[\lambda_t^1\lambda_t^2]$.

Take $f(t,\lambda_t^1,\lambda_t^2)=(\lambda_t^1)^2$, $(\lambda_t^2)^2$ and $\lambda_t^1\lambda_t^2$ respectively, we have

$$\begin{aligned}\mathcal{A}((\lambda_1)^2) &= -2\delta_1\lambda_1^2+\rho_1\left[\int_0^\infty(\lambda_1+y_1)^2H^1(dy_1)-\lambda_1^2\right]\\ &\quad+\lambda_1\left[\int_0^\infty(\lambda_1+z_1)^2G^{1,1}(dz_1)-\lambda_1^2\right]+\lambda_2\left[\int_0^\infty(\lambda_1+z_1)^2G^{1,2}(dz_1)-\lambda_1^2\right]\\ &= -2\delta_1\lambda_1^2+\rho_1(2\lambda_1\mu_{H^1}+\mu_{2H^1})+\lambda_1(2\lambda_1\mu_{G^{1,1}}+\mu_{2G^{1,1}})+\lambda_2(2\lambda_1\mu_{G^{1,2}}+\mu_{2G^{1,2}})\\ &= -2(\delta_1-\mu_{G^{1,1}})\lambda_1^2+2\mu_{G^{1,2}}\lambda_1\lambda_2+(2\rho_1\mu_{H^1}+\mu_{2G^{1,1}})\lambda_1+\mu_{2G^{1,2}}\lambda_2+\mu_{2H^1}\rho_1.\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathcal{A}((\lambda_2)^2) &= -2(\delta_2-\mu_{G^{2,2}})\lambda_2^2+2\mu_{G^{2,1}}\lambda_1\lambda_2+(2\rho_2\mu_{H^2}+\mu_{2G^{2,2}})\lambda_2+\mu_{2G^{2,1}}\lambda_1+\mu_{2H^2}\rho_2 \\
\mathcal{A}\lambda_1\lambda_2 &= -(\delta_1+\delta_2)\lambda_1\lambda_2+\rho_1\lambda_2\mu_{H^1}+\rho_2\lambda_1\mu_{H^2}+\lambda_1\left(\lambda_1\mu_{G^{2,1}}+\lambda_2\mu_{G^{1,1}}+\mu_{G^{1,1}}\mu_{G^{2,1}}\right) \\
&\quad +\lambda_2(\lambda_1\mu_{G^{2,2}}+\lambda_2\mu_{G^{1,2}}+\mu_{G^{1,2}}\mu_{G^{2,2}}) \\
&= \mu_{G^{2,1}}\lambda_1^2+\mu_{G^{1,2}}\lambda_2^2+\left(-(\delta_1-\mu_{G^{1,1}})-(\delta_2-\mu_{G^{2,2}})\right)\lambda_1\lambda_2 \\
&\quad +(\rho_2\mu_{H^2}+\mu_{G^{1,1}}\mu_{G^{2,1}})\lambda_1+(\rho_1\mu_{H^1}+\mu_{G^{1,2}}\mu_{G^{2,2}})\lambda_2.
\end{aligned}$$

We denote

$$\begin{aligned}
\mathcal{A}(\lambda_1^2) &=: A_{1,1}\lambda_1^2+A_{1,2}\lambda_1\lambda_2+A_1\lambda_1+A_2\lambda_2+A_0, \\
\mathcal{A}(\lambda_2^2) &=: B_{2,2}\lambda_2^2+B_{1,2}\lambda_1\lambda_2+B_2\lambda_2+B_1\lambda_1+B_0, \\
\mathcal{A}\lambda_1\lambda_2 &=: C_{1,1}\lambda_1^2+C_{2,2}\lambda_2^2+C_{1,2}\lambda_1\lambda_2+C_1\lambda_1+C_2\lambda_2,
\end{aligned}$$

with all coefficients in Table 4.2. Note that $A_{i,j}, B_{i,j}, C_{i,j}$ do not contain ρ_1 and ρ_2, and A_1, B_2, C_1, C_2 are linear with ρ_1 and ρ_2.

X	$X_{1,1}$	$X_{2,2}$	$X_{1,2}$	X_1	X_2	X_0
A	$-2\Delta_1$	0	$2\mu_{G^{1,2}}$	$2\mu_{H^1}\rho_1+\mu_{2G^{1,1}}$	$\mu_{2G^{1,2}}$	$\mu_{2H^1}\rho_1$
B	0	$-2\Delta_2$	$2\mu_{G^{2,1}}$	$\mu_{2G^{2,1}}$	$2\mu_{H^2}\rho_2+\mu_{2G^{2,2}}$	$\mu_{2H^2}\rho_2$
C	$\mu_{G^{2,1}}$	$\mu_{G^{1,2}}$	$-\Delta_1-\Delta_2$	$\mu_{H^2}\rho_2+\mu_{G^{1,1}}\mu_{G^{2,1}}$	$\mu_{H^1}\rho_1+\mu_{G^{1,2}}\mu_{G^{2,2}}$	0

Table 4.2: Coefficient table for A, B, C, $\Delta_1=\delta_1-\mu_{G^{1,1}}>0$ and $\Delta_2=\delta_2-\mu_{G^{2,2}}>0$.

We denote for $i, j=1,2$ and $j\neq i$,

$$\begin{aligned}
m_2^i &=: \mathbb{E}[(\lambda_t^i)^2], \\
m_2^{i,j} &=: \mathbb{E}[\lambda_t^i\lambda_t^j],
\end{aligned}$$

then by (4.21) we obtain the linear equation system:

$$\begin{aligned}
A_{1,1}m_2^1+A_{1,2}m_2^{1,2}+\left(A_1m_1^1+A_2m_1^2+A_0\right) &= 0, \\
B_{2,2}m_2^2+B_{1,2}m_2^{1,2}+\left(B_2m_1^2+B_1m_1^1+B_0\right) &= 0, \\
C_{1,1}m_2^1+C_{2,2}m_2^2+C_{1,2}m_2^{1,2}+\left(C_1m_1^1+C_2m_1^2\right) &= 0,
\end{aligned}$$

where the cross term is

$$m_2^{1,2}=-\frac{C_1m_1^1+C_2m_1^2+C_{1,1}m_2^1+C_{2,2}m_2^2}{C_{1,2}}. \tag{4.22}$$

Based on the first moments m_1^1, m_1^2 obtained in the last subsection, we obtain the second moments m_2^1 and m_2^2 by solving the linear equation system.

$$
\begin{aligned}
\left(A_{1,1} - A_{1,2}\frac{C_{1,1}}{C_{1,2}}\right) m_2^1 - A_{1,2}\frac{C_{2,2}}{C_{1,2}} m_2^2 + \left(\tilde{A}_0 - \frac{A_{1,2}}{C_{1,2}}\tilde{C}_0\right) &= 0, \\
\left(B_{2,2} - B_{1,2}\frac{C_{2,2}}{C_{1,2}}\right) m_2^2 - B_{1,2}\frac{C_{1,1}}{C_{1,2}} m_2^1 + \left(\tilde{B}_0 - \frac{B_{1,2}}{C_{1,2}}\tilde{C}_0\right) &= 0,
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{A}_0 &= A_1 m_1^1 + A_2 m_1^2 + A_0, \\
\tilde{B}_0 &= B_2 m_1^2 + B_1 m_1^1 + B_0, \\
\tilde{C}_0 &= C_1 m_1^1 + C_2 m_1^2.
\end{aligned}
$$

Denote $\gamma_1 := \frac{C_{1,1}}{C_{1,2}}$ and $\gamma_2 := \frac{C_{2,2}}{C_{1,2}}$, then

$$
\left[A_{1,1} - A_{1,2}\gamma_1 - \frac{A_{1,2}B_{1,2}\gamma_1\gamma_2}{B_{2,2} - B_{1,2}\gamma_2}\right] m_2^1 = -\frac{(\tilde{B}_0 - \frac{B_{1,2}}{C_{1,2}}\tilde{C}_0)A_{1,2}\gamma_2}{B_{2,2} - B_{1,2}\gamma_2} - \left(\tilde{A}_0 - \frac{A_{1,2}}{C_{1,2}}\tilde{C}_0\right).
$$

$$
m_2^1 = \frac{-\frac{(\tilde{B}_0 - \frac{B_{1,2}}{C_{1,2}}\tilde{C}_0)A_{1,2}\gamma_2}{B_{2,2} - B_{1,2}\gamma_2} - \left(\tilde{A}_0 - \frac{A_{1,2}}{C_{1,2}}\tilde{C}_0\right)}{A_{1,1} - A_{1,2}\gamma_1 - \frac{A_{1,2}B_{1,2}\gamma_1\gamma_2}{B_{2,2} - B_{1,2}\gamma_2}} = \frac{(B_{1,2}\gamma_2 - B_{2,2})\tilde{A}_0 - A_{1,2}\gamma_2\tilde{B}_0 + \frac{A_{1,2}}{C_{1,2}}B_{2,2}\tilde{C}_0}{4(\Delta_1\Delta_2 - \mu_{G^{1,2}}\mu_{G^{2,1}})}.
$$

Similarly, we have

$$
m_2^2 = \frac{-B_{1,2}\gamma_1\tilde{A}_0 - (A_{1,1} - A_{1,2}\gamma_1)\tilde{B}_0 + \frac{B_{1,2}}{C_{1,2}}A_{1,1}\tilde{C}_0}{4(\Delta_1\Delta_2 - \mu_{G^{1,2}}\mu_{G^{2,1}})}.
$$

We obtain

$$
\begin{aligned}
m_2^1 &= (m_1^1)^2 + \gamma_{1,1}\rho_1 + \gamma_{1,2}\rho_2, \\
m_2^2 &= (m_1^2)^2 + \gamma_{2,1}\rho_1 + \gamma_{2,2}\rho_2,
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,1} = {} & \frac{1}{2\Delta}\left(\frac{-2\mu_{G^{2,1}}\mu_{G^{1,2}}}{\Delta_1 + \Delta_2} + \Delta_2\right)(\mu_{2G^{1,1}}\mu_{1,1} + \mu_{2G^{1,2}}\mu_{2,1} + \mu_{2H^1}) \\
& + \frac{1}{2\Delta}\frac{(\mu_{G^{1,2}})^2}{\Delta_1 + \Delta_2}(\mu_{2G^{2,2}}\mu_{2,1} + \mu_{2G^{2,1}}\mu_{1,1}) + \frac{1}{\Delta}\frac{\mu_{G^{1,2}}\Delta_2}{\Delta_1 + \Delta_2}(\mu_{G^{1,1}}\mu_{G^{2,1}}\mu_{1,1} + \mu_{G^{1,2}}\mu_{G^{2,2}}\mu_{2,1}) \\
\gamma_{1,2} = {} & \frac{1}{2\Delta}\left(\frac{-2\mu_{G^{2,1}}\mu_{G^{1,2}}}{\Delta_1 + \Delta_2} + \Delta_2\right)(\mu_{2G^{1,1}}\mu_{1,2} + \mu_{2G^{1,2}}\mu_{2,2}) \\
& + \frac{1}{2\Delta}\frac{(\mu_{G^{1,2}})^2}{\Delta_1 + \Delta_2}(\mu_{2G^{2,2}}\mu_{2,2} + \mu_{2G^{2,1}}\mu_{1,2} + \mu_{2H^2}) + \frac{1}{\Delta}\frac{\mu_{G^{1,2}}\Delta_2}{\Delta_1 + \Delta_2}(\mu_{G^{1,1}}\mu_{G^{2,1}}\mu_{1,2} + \mu_{G^{1,2}}\mu_{G^{2,2}}\mu_{2,2}).
\end{aligned}
$$

Similarly,

$$\gamma_{2,1} = \frac{1}{2\Delta}\left(\frac{-2\mu_{G^{2,1}}\mu_{G^{1,2}}}{\Delta_1+\Delta_2}+\Delta_1\right)(\mu_{2G^{2,2}}\mu_{2,1}+\mu_{2G^{2,1}}\mu_{11})$$
$$+\frac{1}{2\Delta}\frac{(\mu_{G^{2,1}})^2}{\Delta_1+\Delta_2}(\mu_{2G^{1,1}}\mu_{1,1}+\mu_{2G^{1,2}}\mu_{2,1}+\mu_{2H^1}) + \frac{1}{\Delta}\frac{\mu_{G^{2,1}}\Delta_1}{\Delta_1+\Delta_2}(\mu_{G^{1,1}}\mu_{G^{2,1}}\mu_{1,1}+\mu_{G^{1,2}}\mu_{G^{2,2}}\mu_{2,1})$$
$$\gamma_{2,2} = \frac{1}{2\Delta}\left(\frac{-2\mu_{G^{2,1}}\mu_{G^{1,2}}}{\Delta_1+\Delta_2}+\Delta_1\right)(\mu_{2G^{2,2}}\mu_{2,2}+\mu_{2G^{2,1}}\mu_{1,2}+\mu_{2H^2})$$
$$+\frac{1}{2\Delta}\frac{(\mu_{G^{2,1}})^2}{\Delta_1+\Delta_2}(\mu_{2G^{1,1}}\mu_{1,2}+\mu_{2G^{1,2}}\mu_{2,2}) + \frac{1}{\Delta}\frac{\mu_{G^{2,1}}\Delta_1}{\Delta_1+\Delta_2}(\mu_{G^{1,1}}\mu_{G^{2,1}}\mu_{1,2}+\mu_{G^{1,2}}\mu_{G^{2,2}}\mu_{2,2}).$$

Hence, we conclude that the stationary mean and variance are

$$\begin{aligned} m_1 &:= \mathbb{E}[\lambda_t^1] = \mu_{1,1}\rho_1 + \mu_{1,2}\rho_2, \\ m_2 &:= \mathbb{E}[\lambda_t^2] = \mu_{2,1}\rho_1 + \mu_{2,2}\rho_2, \end{aligned}$$

and

$$\begin{aligned} v_1 &:= \mathrm{var}(\lambda_t^1) = \gamma_{1,1}\rho_1 + \gamma_{1,2}\rho_2, \\ v_2 &:= \mathrm{var}(\lambda_t^2) = \gamma_{2,1}\rho_1 + \gamma_{2,2}\rho_2. \end{aligned}$$

Remark 4.4.2. *From above, we observe that the stationary mean and variance are both linear functions of ρ_1 and ρ_2.*

4.4.3 Stationary Correlation

The stationary correlation $\rho_{1,2}$ can be computed as:

$$\rho_{1,2} = \frac{E[\lambda_t^1\lambda_t^2] - \mathbb{E}[\lambda_t^1]\mathbb{E}[\lambda_t^2]}{\sqrt{\mathrm{var}(\lambda_t^1)}\sqrt{\mathrm{var}(\lambda_t^2)}} = \frac{m_2^{1,2} - m_1 m_2}{\sqrt{v_1}\sqrt{v_2}},$$

where $m_2^{1,2}$ is from (4.22).

Remark 4.4.3. *Note that the cross-exciting jumps enhance the stationary correlation compared to the processes with only self-exciting jumps due to the positive mean $\mu_{G^{1,2}}$ and $\mu_{G^{2,1}}$ of jump sizes.*

4.5 Proof of Lemmas

Proof of Lemma 4.2.1. We prove the result by induction. First we have

$$\lim_{t\to\infty} l_1(t) = \lim_{t\to\infty} v_{2n}e^{-\delta_2 t} = 0, \quad \lim_{t\to\infty} l_2(t) = \lim_{t\to\infty} v_{2n-1}e^{-\delta_1 t} = 0.$$

Then assume $\lim_{t\to\infty} l_{2k-1}(t) = 0$ and $\lim_{t\to\infty} l_{2k}(t) = 0$,

$$\begin{aligned}
\lim_{t\to\infty} l_{2k+1}(t) &= \lim_{t\to\infty} e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[1 - \hat{g}_{1,2}(l_{2k}(s)) + 1 - \hat{g}_{2,2}(l_{2k-1}(s))\right] ds \\
&\overset{L'Hospital}{=} \lim_{t\to\infty} \frac{1}{\delta_2} \left(1 - \hat{g}_{1,2}(l_{2k}(t)) + 1 - \hat{g}_{2,2}(l_{2k-1}(t))\right) \\
&= 0.
\end{aligned}$$

Similarly, we have $\lim_{t\to\infty} l_{2k+2}(t) = 0$.

Therefore we conclude that for any $i = 1, \ldots, m$, $\lim_{t\to\infty} l_i(t) = 0$.

$\square$

Proof of Lemma 4.3.1. By the stationary distribution characterization in (4.1), denote by $\pi := \pi_S^m$ the stationary distribution of $(\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)})$, then it satisfies

$$\int \mathcal{A}_m f(\lambda_1, \lambda_2, \ldots, \lambda_m) \pi(\lambda_1, \lambda_2, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_n = 0, \tag{4.23}$$

where $\mathcal{A}_m$ can be found in (4.2).

We now derive the equivalent Laplace transform equation. Denote by $\mathcal{L}_m$ the Laplace transform of $(\Lambda^{(1)}, \ldots, \Lambda^{(m)})$ and $f(\cdots) := f(\lambda_1, \cdots, \lambda_m)$ with $\lim_{\lambda_k\to\infty} f(\lambda_1, \ldots, \lambda_m) = 0$ for $k = 1, \ldots, m$. We decompose the LHS of (4.23) according to the generator $\mathcal{A}_m$ into three parts.

Part (I): drift part:

$$\begin{aligned}
&\int_{\mathbb{R}_+^m} -\delta_k \lambda_k \frac{\partial}{\partial \lambda_k} f(\lambda_1, \ldots, \lambda_m) \pi(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m \\
= \;& -\delta_k \int_{\mathbb{R}_+^m} \frac{\partial}{\partial \lambda_k} f(\lambda_1, \ldots, \lambda_m) \int_0^{\lambda_k} \frac{\partial}{\partial \lambda_k} \left(x \pi(\lambda_1, \ldots, x, \ldots, \lambda_m)\right) dx d\lambda_1 \cdots d\lambda_m \\
= \;& -\delta_k \int_{\mathbb{R}_+^{m-1}} \int_{\lambda_k=0}^{\infty} \int_{x=0}^{\lambda_k} \frac{\partial}{\partial \lambda_k} f(\lambda_1, \ldots, \lambda_m) \frac{\partial}{\partial \lambda_k} \left(x \pi(\lambda_1, \ldots, x, \ldots, \lambda_m)\right) dx d\lambda_1 \cdots d\lambda_m \\
= \;& -\delta_k \int_{\mathbb{R}_+^{m-1}} \int_{x=0}^{\infty} \int_{\lambda_k=x}^{\infty} \frac{\partial}{\partial \lambda_k} f(\lambda_1, \ldots, \lambda_m) \frac{\partial}{\partial \lambda_k} \left(x \pi(\lambda_1, \ldots, x, \ldots, \lambda_m)\right) dx d\lambda_1 \cdots d\lambda_m \\
= \;& \int_{\mathbb{R}_+^m} f(\lambda_1, \ldots, \lambda_m) \delta_k \frac{\partial}{\partial \lambda_k} \left(\lambda_k \pi(\lambda_1, \ldots, \lambda_m)\right) d\lambda_1 \cdots d\lambda_m,
\end{aligned}$$

The Laplace transform at $(v_1, \ldots, v_m) \in \mathbb{R}_+^m$ becomes

$$\mathcal{L}_m \left[\delta_k \frac{\partial}{\partial \lambda_k} \left(\lambda_k \pi(\lambda_1, \ldots, \lambda_m)\right)\right] = \delta_k v_k \mathcal{L}_m[\lambda_k \pi(\lambda_1, \ldots, \lambda_m)] = -\delta_k v_k \frac{\partial}{\partial v_k} \hat{\pi}(v_1, \ldots, v_m).$$

Part (II): shot-noise part

$$\begin{aligned}
&\int_{\mathbb{R}_+^m} \rho_1 \int_{y=0}^{\infty} f(\lambda_1 + y, \cdot) dH_1(y) \pi(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m \\
= \;& \rho_1 \int_{\mathbb{R}_+^{m-1}} \int_{x=0}^{\infty} f(x, \lambda_2, \ldots, \lambda_m) \int_{y=0}^{x} \pi(x - y, \lambda_2, \ldots, \lambda_m) dH_1(y) dx d\lambda_2 \cdots d\lambda_m \\
= \;& \rho_1 \int_{\mathbb{R}_+^m} f(\lambda_1, \lambda_2, \ldots, \lambda_m) \int_{y=0}^{\lambda_1} \pi(\lambda_1 - y, \lambda_2, \ldots, \lambda_m) dH_1(y) d\lambda_1 \cdots d\lambda_m,
\end{aligned}$$

then

$$\begin{aligned}
&\int_{\mathbb{R}_+^m} \rho_1 \left[\int_0^{\infty} f(\lambda_1 + y, \cdot) dH_1(y) - f(\ldots) \right] \pi(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m \\
= \;& \rho_1 \int_{\mathbb{R}_+^m} f(\lambda_1, \lambda_2, \ldots, \lambda_m) \left[\int_{y=0}^{\lambda_1} \pi(\lambda_1 - y, \cdot) dH_1(y) - \pi(\lambda_1, \ldots, \lambda_m) \right] d\lambda_1 \cdots d\lambda_m.
\end{aligned}$$

The Laplace transform as

$$\mathcal{L}_m \left[\int_{y=0}^{\lambda_1} \pi(\lambda_1 - y, \cdot) dH_1(y) \right] = \hat{\pi}(v_1, \ldots, v_m) \hat{h}_1(v_1).$$

Similarly, we have

$$\mathcal{L}_m \left[\int_{y=0}^{\lambda_2} \pi(\cdot, \lambda_2 - y, \cdot) dH_2(y) \right] = \hat{\pi}(v_1, \ldots, v_m) \hat{h}_2(v_2).$$

Part(III): mutually exciting part:

For $k \geq 2$, the jump excited by λ_{2k-1} is

$$\begin{aligned}
&\int_{\mathbb{R}_+^m} \lambda_{2k-1} \left[\left(\int_{z=0}^{\infty} f(\cdot, \lambda_{2k+1} + z, \cdot) dG_{1,1}(z) - f(\ldots) \right) + \left(\int_{z=0}^{\infty} f(\cdot, \lambda_{2k+2} + z, \cdot) dG_{2,1}(z) - f(\ldots) \right) \right] \\
&\quad \cdot \pi(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m \\
&= \int_{\mathbb{R}_+^m} f(\lambda_1, \ldots, \lambda_m) \lambda_{2k-1} \left[\int_{z=0}^{\lambda_{2k+1}} \pi(\cdot, \lambda_{2k+1} - z, \cdot) dG_{1,1}(z) - \pi(\ldots) \right] d\lambda_1 \cdots d\lambda_n \\
&+ \int_{\mathbb{R}_+^m} f(\lambda_1, \ldots, \lambda_m) \lambda_{2k-1} \left[\int_{z=0}^{\lambda_{2k+2}} \pi(\cdot, \lambda_{2k+2} - z, \cdot) dG_{2,1}(z) - \pi(\ldots) \right] d\lambda_1 \cdots d\lambda_n.
\end{aligned}$$

We have first

$$\begin{aligned}
& \mathcal{L}_m \left[\lambda_{2k-1} \int_{z=0}^{\lambda_{2k+1}} \pi(\cdot, \lambda_{2k+1} - z, \cdot) dG_{1,1}(z) - \lambda_{2k-1} \pi(\ldots) \right] \\
= \;& \mathcal{L}_m \left[\lambda_{2k-1} \pi(\lambda_1, \ldots, \ldots, \lambda_m) \right] \hat{g}_{1,1}(v_{2k+1}) - \mathcal{L}_m \left[\lambda_{2k-1} \pi(\ldots) \right] \\
= \;& \frac{\partial}{\partial v_{2k-1}} \hat{\pi}(v_1, \ldots, v_m)(1 - \hat{g}_{1,1}(v_{2k+1})).
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \mathcal{L}_m \left[\lambda_{2k-1} \int_{z=0}^{\lambda_{2k+2}} \pi(\cdot, \lambda_{2k+2} - z, \cdot) dG_{2,1}(z) - \lambda_{2k-1}\pi(\ldots) \right] \\
= \;& \frac{\partial}{\partial v_{2k-1}} \hat{\pi}(v_1, \ldots, v_m)(1 - \hat{g}_{2,1}(v_{2k+2})).
\end{aligned}$$

The same for jumps excited by λ_{2k} corresponding to

$$\int_{\mathbb{R}_+^m} \lambda_{2k} \left[\left(\int_{z=0}^{\infty} f(\cdot, \lambda_{2k+1} + z, \cdot) dG_{1,2}(z) - f(\ldots) \right) + \left(\int_{z=0}^{\infty} f(\cdot, \lambda_{2k+2} + z, \cdot) dG_{2,2}(z) - f(\ldots) \right) \right]$$
$$\cdot\, \pi(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m.$$

Since the stationary distribution π satisfies (4.23), we have from part (I), (II), (III) that

$$\begin{aligned}
0 \;=\; & \sum_{k=1}^{m} \delta_k \frac{\partial}{\partial \lambda_k} \left(\lambda_k \pi(\lambda_1, \ldots, \lambda_m) \right) \\
& + \rho_1 \left[\int_{y=0}^{\lambda_1} \pi(\lambda_1 - y, \lambda_2, \ldots, \lambda_m) dH_1(y) - \pi(\lambda_1, \ldots, \lambda_m) \right] \\
& + \rho_2 \left[\int_{y=0}^{\lambda_2} \pi(\lambda_1, \lambda_2 - y, \ldots, \lambda_m) dH_2(y) - \pi(\lambda_1, \ldots, \lambda_m) \right] \\
& + \sum_{k=1}^{n-1} \lambda_{2k-1} \left[\int_{z=0}^{\lambda_{2k+1}} \pi(\lambda_1, \ldots, \lambda_{2k+1} - z, \ldots, \lambda_m) dG_{1,1}(z) - \pi(\lambda_1, \ldots, \lambda_m) \right] \\
& + \sum_{k=1}^{n-1} \lambda_{2k-1} \left[\int_{z=0}^{\lambda_{2k+2}} \pi(\lambda_1, \ldots, \lambda_{2k+2} - z, \ldots, \lambda_m) dG_{2,1}(z) - \pi(\lambda_1, \ldots, \lambda_m) \right] \\
& + \sum_{k=1}^{n-1} \lambda_{2k} \left[\int_{z=0}^{\lambda_{2k+1}} \pi(\lambda_1, \ldots, \lambda_{2k+1} - z, \ldots, \lambda_m) dG_{1,2}(z) - \pi(\lambda_1, \ldots, \lambda_m) \right] \\
& + \sum_{k=1}^{n-1} \lambda_{2k} \left[\int_{z=0}^{\lambda_{2k+2}} \pi(\lambda_1, \ldots, \lambda_{2k+2} - z, \ldots, \lambda_m) dG_{2,2}(z) - \pi(\lambda_1, \ldots, \lambda_m) \right].
\end{aligned}$$

In terms of Laplace transforms, we have for any $(v_1, \ldots, v_m) \in \mathbb{R}_+^m$

$$\begin{aligned}
0 \;=\; & -\sum_{k=1}^{2n} \delta_k v_k \frac{\partial \hat{\pi}_S^m}{\partial v_k} + \rho_1(\hat{h}_1(v_1) - 1) + \rho_2(\hat{h}_2(v_2) - 1) \\
& + \sum_{k=1}^{n-1} \frac{\partial \hat{\pi}_S^m}{\partial v_{2k-1}} \left[(1 - \hat{g}_{1,1}(v_{2k+1})) + (1 - \hat{g}_{2,1}(v_{2k+2})) \right] \\
& + \sum_{k=1}^{n-1} \frac{\partial \hat{\pi}_S^m}{\partial v_{2k}} \left[(1 - \hat{g}_{1,2}(v_{2k+1})) + (1 - \hat{g}_{2,2}(v_{2k+2})) \right],
\end{aligned}$$

and after reordering the terms, we obtain (4.15). $\square$

Proof of (4.10). For $j = 1, 2$,

$$\begin{aligned}
\hat{g}_{1,j}(l_{2k-2}(t)) - \hat{g}_{1,j}(l_{2k}(t)) &= \int_{l_{2k}(t)}^{l_{2k-2}(t)} d\hat{g}_{1,j}(u) = \int_{l_{2k-2}(t)}^{l_{2k}(t)} \left(-\hat{g}'_{1,j}(u)\right) du \\
&\le \mu_{G^{1,j}} \left(l_{2k}(t) - l_{2k-2}(t)\right), \\
\hat{g}_{2,j}(l_{2k-3}(t)) - \hat{g}_{2,j}(l_{2k-1}(t)) &= \int_{l_{2k-1}(t)}^{l_{2k-3}(t)} d\hat{g}_{2,j}(u) = \int_{l_{2k-3}(t)}^{l_{2k-1}(t)} \left(-\hat{g}'_{2,j}(u)\right) du \\
&\le \mu_{G^{2,j}} \left(l_{2k-1}(t) - l_{2k-3}(t)\right).
\end{aligned}$$

Then we have

$$\begin{aligned}
d_{k+1}^{(1)}(t) &= e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[(1 - \hat{g}_{1,2}(l_{2k}(s))) - (1 - \hat{g}_{1,2}(l_{2k-2}(s)))\right] ds \\
&\quad + e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[(1 - \hat{g}_{2,2}(l_{2k-1}(s))) - (1 - \hat{g}_{2,2}(l_{2k-3}(s)))\right] ds \\
&\le e^{-\delta_2 t} \int_0^t e^{\delta_2 s} \left[\mu_{G^{2,2}} d_k^{(1)}(s) + \mu_{G^{1,2}} d_k^{(2)}(s)\right] ds, \\
d_{k+1}^{(2)}(t) &= e^{-\delta_1 t} \int_0^t e^{\delta_1 s} \left[(1 - \hat{g}_{1,1}(l_{2k}(s))) - (1 - \hat{g}_{1,1}(l_{2k-2}(s)))\right] ds \\
&\quad + e^{-\delta_1 t} \int_0^t e^{\delta_1 s} \left[(1 - \hat{g}_{2,1}(l_{2k-1}(s))) - (1 - \hat{g}_{2,1}(l_{2k-3}(s)))\right] ds \\
&\le e^{-\delta_1 t} \int_0^t e^{\delta_1 s} \left[\mu_{G^{2,1}} d_k^{(1)}(s) + \mu_{G^{1,1}} d_k^{(2)}(s)\right] ds.
\end{aligned}$$

$\square$

4.6 Conclusion

In this chapter, we investigated the distributional information of the BDCP and its finite branching system in the cluster process representation. With the help of Markov process theory and the branching structure, we found the condition ((C1), (C2)) under which the limiting distributions (when time tends to infinity) exist. We then showed the limiting distributions are also the stationary distributions. Therefore, we can conclude that there exists a unique stationary version of the BDCP. The stationary moments were also computed. The stationarity result in this chapter will help to derive the diffusion approximation of the BDCP system in the next chapter.

Chapter 5

BDCP Diffusion Approximation and Applications in Filtering

In a high frequency event system modelled by a BDCP, we investigate the diffusion approximation and its applications in filtering. The diffusion approximation is based on the stationarity assumption from Chapter 4 and the martingale central limit theorem.

In Section 5.1, we introduce basic concepts. We discuss the diffusion approximation of BDCP in Section 5.2. We discuss the diffusion approximation of the application in filtering in Section 5.3 and show numerical examples in Section 5.4. We include all proofs in Section 5.5.

5.1 Introduction

As introduced in Chapter 3, the BDCP can be used to model multi-type event arrivals with contagion effects in finance and insurance applications, including default events, trading order arrivals, and insurance claim arrivals. In this chapter, we consider a stationary BDCP used in a high frequency events system resulting from a big external impact. It is described by a large value of the intensity ρ_k of the Poisson external factor in the BDCP intensity and we study the diffusion approximation of the system as this parameter goes to infinity. We first show that when the external factor intensity ρ_k tends to infinity, the normalized intensity and point processes converge to a bivariate diffusion process weakly. Hence, we can approximate the point process system with a diffusion system with the same mean and variance when the external factor intensity parameter is large. The normalization is done using the stationarity moments from Chapter 4. In this way, for a BDCP system with a large external factor intensity, we can approximate it with an OU process with the same mean and variance. Note that Dassios and Jang [22] study this problem for univariate shot-noise Cox processes, and we extend the study to the BDCP family. The BDCP has mutually exciting components and thus results in a diffusion system fundamentally different due to the dimension extension and the dependence structure. Moreover, there are discussions about the diffusion approximation of Hawkes process systems recently in finance. Bacry et al. [7] discuss functional central limit theorems for Hawkes processes with scaling of time. Jaisson and Rosenbaum [40] discuss the nearly unstable Hawkes processes and link the problem with the asymptotic behaviour of the point process when the observation scale tends to infinity. Cont and Larrard [17] explore the diffusion approximation of the joint dynamic of a bid and ask queueing system.

In practice, the point process is observable while the underlying intensity is not, as it usually carries more internal information. For example, in a BDCP system, in addition to the internal information of the point process N, the intensity process λ also includes the information regarding the external factor and the impact modelled by random jump sizes. It becomes a filtering problem to find out the best estimate of the intensity, given the observations. In previous studies of filtering with point process observations, Brémaud [13] introduces the innovation approach, based upon which, Ceci and Gerardi [15] provide a framework to solve the filtering problem where the point process are allowed to have common jumps with intensities. Note that potentially the method can be used to filter the BDCP. However, since the filter can be characterized as the solution of the Kushner-Stratonovich (KS) equation, it requires confirming that strict assumptions hold for the dynamics of the point process system to ensure the existence of a unique solution to the KS equation. The KS equation can be evaluated numerically because the solution is a functional of a process, which can be represented as the limit of a sequence of finite state and discrete time Markov chains. The filter based on the Markov chain approximation is not exact. Most importantly, if we have a high frequency events system, the approach might be not efficient and accurate.

Therefore, in this chapter we provide an efficient alternative approach of filtering the point process system based on the diffusion approximation. Recall that there is a clustering effect in the BDCP due to the mutually exciting components in the intensity processes, which naturally makes the BDCP suitable for high frequency modelling in the literature. We expect a reasonable performance of filtering with the diffusion approximation. Essentially, we solve the filtering problem by plugging the actual point process observation input into the solution of the Kalman-Bucy filter of the approximate linear diffusion system. In this way, we construct a linear estimate for the BDCP system. The idea of filtering with the diffusion approximation is adopted by Dassios and Jang [22] for univariate shot noise Cox processes and the results can be recovered from ours as BDCP is a more general class. Moreover, filtering in a multi-dimensional system is more interesting as different observation scenarios (e.g. see (S1) and (S2) in Section 5.3) can be explored. Previously, filtering with a diffusion approximation is also discussed in Kushner and Runggaldier [48], and Lipster and Runggaldier [53] for a specific time scaled model with wide bandwidth noises that are different from our modelling framework. Inspired by [48] and [53], we will also assess the performance of the constructed estimate in terms of the asymptotic estimate error. Furthermore, we apply the filtering solution in insurance, aiming for approximation solutions of pricing problems and type estimation problems.

We first briefly introduce the basics of weak convergence and filtering.

Weak convergence in $D = D_{\mathbb{R}^d}[0,\infty)$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. We suppose that processes $(X^n)_{n\geq 1}$ and X are stochastic processes adapted to $(\mathcal{F}_t)_{t\geq 0}$ and take values in the path space $(D, \mathcal{D})$ which is the space of càdlàg functions valued in $\mathbb{R}^d$ equipped with the Skorohod topology. Since $\mathbb{R}^d$ is complete and separable, $(X^n)_{n\geq 1}$ and $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (D, \mathcal{D})$ are $\mathcal{F}/\mathcal{D}$-measurable, thus they are random elements in $(D, \mathcal{D})$. We say the processes $(X^n)_{n\geq 1}$ converge to X weakly as $n \to \infty$ denoted by $X_n \Rightarrow X$ if the probability measures induced converge weakly on $(D, \mathcal{D})$, i.e. $P_n := \mathbb{P}(X^n)^{-1} \Rightarrow P := \mathbb{P}X^{-1}$. Equivalently, $P_n f \to P f$ for every bounded and uniformly continuous real-valued function f defined on $(D, \mathcal{D})$.

Filtering

In a dynamic evolving system on the probability space $(\Omega, \mathbb{F}, P)$, suppose there is a Markov process $X = (X_t)_{t\geq 0}$ which can not be directly observed and an observable process $Y = (Y_t)_{t\geq 0}$ that is probabilistically dependent on X. (X, Y) is called the *signal-observation system* and is supposed to be a Markov system. Let $\mathcal{Y}_t = \sigma\{Y_s; s \leq t\}$ be the observations of Y up to time t. The goal of filtering is to find the conditional distribution of $X_t|\mathcal{Y}_t$.

For example, the estimate of the mean value of the signal X_t given $\mathcal{Y}_t$, $\hat{X}_t = \mathbb{E}[X_t|\mathcal{Y}_t]$ is optimal with respect to the mean square error criteria, i.e. for any $t \geq 0$,

$$\mathbb{E}[|X_t - \hat{X}_t|^2] = \inf_{Y\in K}\{\mathbb{E}[|X_t - Y|^2]\},$$

where $K := \{Y : \Omega \to \mathbb{R}^d; Y \in L^2(\mathbb{P})$ and is $\mathcal{Y}_t$-measurable$\}$. The statistical performance of the estimator can be evaluated by the filter variance

$$\gamma_t = \mathbb{E}[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T|\mathcal{Y}_t].$$

Moreover, in practice one often needs to find out the optimal filter of $f(X_t)$ given $\mathcal{Y}_t$ for some measurable and integrable function $f : \mathbb{R}^d \to \mathbb{R}$. The best estimate with respect to the mean square error criteria is $\pi_t(f) := \mathbb{E}[f(X_t)|\mathcal{Y}_t]$, i.e. for any integrable and measurable function $g : D_{\mathbb{R}^d}[0, t) \to \mathbb{R}$,

$$\mathbb{E}\left[(f(X_t) - \mathbb{E}[f(X_t)|\mathcal{Y}_t])^2\right] \leq \mathbb{E}\left[(f(X_t) - g((Y_s)_{s\leq t}))^2\right].$$

There are two major approaches to get the solution of filtering problems $\pi_t(f) := \mathbb{E}[f(X_t)|\mathcal{Y}_t]$. The first one is the *reference probability approach* based on the independence of X and Y after a change of probability measure. It leads to solve an unnormalized filtering equation called the Zakai equation. The second one is called the *innovation approach* which is based on the martingale representation of the signal process X and the projection theory to identify the projected signal process on $\mathcal{Y}_t$. The innovation approach leads to the Kushner-Stratonovich (KS) equation which is a normalized version of the Zakai equation. Explicit solutions to the KS equation or the Zakai equation are rarely available. In general, they can only be solved numerically. When (X, Y) is a Gaussian system (see more details in Lemma 5.3.1), it becomes a linear filtering problem where an explicit solution (the Kalman-Bucy filter) is available.

Filtering problems with point process observations are usually solved via the innovation approach (See Brémaud [13]). Note that when X and Y have common jumps, one cannot use the reference probability approach as there does not exist a change of measure such that X and Y are independent under the new measure. For example, Ceci and Geraldi [15] adopt the innovation approach to solve a general filtering problem with point process observations, and the resulting KS equation are solved recursively with Markov chain approximation.

5.2 Diffusion Approximation of The BDCP

Recall the BDCP (N^1, N^2) with intensity

$$\lambda_t^1 = \lambda_0^1 e^{-\delta_1 t} + \sum_{S_j^1 < t} Y_j^1 e^{-\delta_1 (t - S_j^1)} + \sum_{T_j^1 < t} Z_j^{1,1} e^{-\delta_1 (t - T_j^1)} + \sum_{T_j^2 < t} Z_j^{1,2} e^{-\delta_1 (t - T_j^2)},$$

$$\lambda_t^2 = \lambda_0^2 e^{-\delta_2 t} + \sum_{S_j^2 < t} Y_j^2 e^{-\delta_2 (t - S_j^2)} + \sum_{T_j^1 < t} Z_j^{2,1} e^{-\delta_2 (t - T_j^1)} + \sum_{T_j^2 < t} Z_j^{2,2} e^{-\delta_2 (t - T_j^2)}.$$

In addition, we introduce a process $C^k = (C_t^k)_{t \geq 0}$ for $k = 1, 2$ by

$$C_t^k := \sum_{i=1}^{N_t^k} c_i^k, \tag{5.1}$$

where $\{c_i^k\}_i$ are i.i.d. nonnegative random variables independent with (N^1, N^2) and with the distribution $G_{c,k}(\cdot)$, mean $\mu_{G^{c,k}} < \infty$ and variance $v_{G^{c,k}} < \infty$. C^k is used to model the accumulative claim arrival process in insurance as an application in Section 5.4 and we include it in the BDCP system to derive the diffusion approximation.

We first recall the stationarity result from Chapter 4.

Stationarity of BDCP Under the Condition (C1) and (C2) in Condition 3.3.2, i.e. the spectral radius of $\begin{bmatrix} \frac{\mu_{G^{2,2}}}{\delta_2} & \frac{\mu_{G^{1,2}}}{\delta_2} \\ \frac{\mu_{G^{2,1}}}{\delta_1} & \frac{\mu_{G^{1,1}}}{\delta_1} \end{bmatrix}$ is less than 1, and the moments $\mu_{H^k} < \infty$, $\mu_{G^{k,k'}} < \infty$, there exists a unique stationary distribution of the intensity λ in (3.8) and thus the BDCP N has stationary increments. Moreover, we assume that the second moments $\mu_{2H^k} < \infty$, $\mu_{2G^{k,k'}} < \infty$. The stationary mean m, variance v and correlation ρ_{12} of λ are given by

$$\begin{aligned}
m &= \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{2,2} \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \\
v &= \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} \\ \gamma_{2,1} & \gamma_{2,2} \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \\
\rho_{12} &= \frac{m_2^{1,2} - m_1 m_2}{\sqrt{v_1}\sqrt{v_2}},
\end{aligned} \tag{5.2}$$

where ρ_k is the intensity of the Poisson process M^k for $k = 1, 2$.

Remark 5.2.1. *We observe that the stationary condition is independent of the intensity of ρ_k. Moreover both the stationary mean and variance are linear in ρ_k for $k = 1, 2$.*

Throughout the chapter, we assume a stationary system and the intensity λ starts with its stationary distribution, hence λ is a stationary process with mean m and variance v at any time $t \geq 0$.

5.2.1 High Frequency Events System

As for the Hawkes process, the BDCP is appropriate to model high-frequency events system due to the contagion effect introduced by the mutually-exciting components. We focus on the case when the BDCP system is under a great impact from external factors described by a large intensity ρ_k for $k = 1, 2$. For example, the external factor can be a series of bankruptcies in a financial crisis or a catastrophes in insurance modelling.

Throughout the chapter, in order to investigate the approximation of a BDCP with a large intensity ρ_k by its weak limit, we embed the true model indexed by some n_0 into a sequence of point process systems $\{(N^n, \lambda^n)\}_{n\geq 1}$ driven by a sequence of external factor intensities $\{\rho_k^n\}_{n\geq 1}$ for $k = 1, 2$ with

$$\lim_{n\to\infty} \max\left(\rho_1^n, \rho_2^n\right) = \infty. \tag{5.3}$$

Remark 5.2.2. *We do not require both ρ_1 and ρ_2 tend to infinity. If there is one of the external factors having a big impact on the marginal process, then the impact would be contagious through the cross-exciting components.*

Denote by $m_k^{(n)}$ and $v_k^{(n)}$the stationary mean and variance of the intensity λ^n, and by $\beta^n := \frac{\rho_2^n}{\rho_1^n}$ the external factor intensity ratio with the limit

$$\beta := \lim_{n\to\infty} \beta^n = \lim_{n\to\infty} \frac{\rho_2^n}{\rho_1^n} \in [0,\infty]. \tag{5.4}$$

We define further for $k, k' = 1, 2$,

$$\begin{aligned}
\alpha_k^{(n)} &:= \frac{m_k^{(n)}}{v_k^{(n)}} = \frac{\mu_{k,1}\rho_1^n + \mu_{k,2}\rho_2^n}{\gamma_{k,1}\rho_1^n + \gamma_{k,2}\rho_2^n} = \frac{\mu_{k,1} + \mu_{k,2}\frac{\rho_2^n}{\rho_1^n}}{\gamma_{k,1} + \gamma_{k,2}\frac{\rho_2^n}{\rho_1^n}} = \frac{\mu_{k,1} + \mu_{k,2}\beta^n}{\gamma_{k,1} + \gamma_{k,2}\beta^n},\\
\theta_k^{(n)} &:= \frac{\rho_k^n}{v_k^{(n)}} = \frac{\rho_k^n}{\gamma_{k,1}\rho_1^n + \gamma_{k,2}\rho_2^n} = \frac{1_{\{k=1\}} + \beta^n 1_{\{k=2\}}}{\gamma_{k,1} + \beta^n\gamma_{k,2}},\\
\nu_{k,k'}^{(n)} &:= \frac{v_k^{(n)}}{v_{k'}^{(n)}} = \frac{\gamma_{k,1}\rho_1^n + \gamma_{k,2}\rho_2^n}{\gamma_{k',1}\rho_1^n + \gamma_{k',2}\rho_2^n} = \frac{\gamma_{k,1} + \gamma_{k,2}\beta^n}{\gamma_{k',1} + \gamma_{k',2}\beta^n},
\end{aligned}$$

and their the asymptotic values

$$\alpha_k := \lim_{n\to\infty} \alpha_k^{(n)}, \quad \theta_k := \lim_{n\to\infty} \theta_k^{(n)}, \quad \nu_{k,k'} := \lim_{n\to\infty} \nu_{k,k'}^{(n)}. \tag{5.5}$$

We will see that by (5.3) the point process system converges to a Gaussian diffusion system that depends on the parameter β in (5.4) in terms of $(\alpha_k, \theta_k, \nu_{k,k'})$. In other words, when β is zero, a finite number or infinity, the resulting system tends to a diffusion parameterized by different sets of $(\alpha_k, \theta_k, \nu_{k,k'})$.

1. $\beta = 0$: external impact of type 1 is dominating.

$$\alpha_k = \frac{\mu_{k,1}}{\gamma_{k,1}}, \quad \theta_k = \frac{1_{\{k=1\}}}{\gamma_{k,1}}, \quad \nu_{k,k'} = \frac{\gamma_{k,1}}{\gamma_{k',1}}.$$

2. $\beta \in (0,\infty)$: external impact of both types are of the same order.

$$\alpha_k = \frac{\mu_{k,1} + \beta\mu_{k,2}}{\gamma_{k,1} + \beta\gamma_{k,2}}, \quad \theta_k = \frac{1_{\{k=1\}} + \beta 1_{\{k=2\}}}{\gamma_{k,1} + \beta\gamma_{k,2}}, \quad \nu_{k,k'} = \frac{\gamma_{k,1} + \gamma_{k,2}\beta}{\gamma_{k',1} + \gamma_{k',2}\beta}.$$

3. $\beta = \infty$: external impact of type 2 is dominating.

$$\alpha_k = \frac{\mu_{k,2}}{\gamma_{k,2}}, \quad \theta_k = \frac{1_{\{k=2\}}}{\gamma_{k,2}}, \quad \nu_{k,k'} = \frac{\gamma_{k,2}}{\gamma_{k',2}}.$$

5.2.2 Diffusion Approximation

Recall the intensity process in (3.8) in integral form:

$$\lambda_t^k = \lambda_0^k e^{-\delta_k t} + \int_0^t e^{-\delta_k(t-s)} dJ_s^{M^k} + \int_0^t e^{-\delta_k(t-s)} dJ_s^{N^1,k} + \int_0^t e^{-\delta_k(t-s)} dJ_s^{N^2,k},$$

where the integral is over the interval $[0,t)$, and

$$J_t^{M^k} = \sum_{S_i^k \le t} Y_i^k, \quad J_t^{N^1,k} = \sum_{T_i^1 \le t} Z_i^{k,1}, \quad J_t^{N^2,k} = \sum_{T_i^2 \le t} Z_i^{k,2}.$$

Moreover, denote by X the integrated intensity process:

$$X_t^k := \int_0^t \lambda_s^k ds.$$

We aim to get a diffusion system $(\tilde{\lambda}, \tilde{N}, \tilde{C})$ as an approximation of the orginal system (λ, N, C) with the help of Theorem 5.2.3 below. It is the martingale central limit theorem in Ethier and Kurtz [33] and an extension of Donsker's invariance principle.

Theorem 5.2.3 (Diffusion Approximation [33, Theorem 1.4, Chapter 7]). *For $n = 1, 2, \dots$, let $\mathbb{F}^n = \{\mathcal{F}_t^n\}_{t\ge 0}$ be a filtration and let M_n be an $\mathcal{F}_t^n$-local martingale with sample paths in $\mathbb{D}_{\mathbb{R}^d}[0,\infty)$ and $M_n(0) = 0$. Let $A_n = (A_n^{ij})$ be a symmetric $d \times d$ matrix-valued processes such that A_n^{ij} has sample paths in $\mathbb{D}_{\mathbb{R}^d}[0,\infty)$ and $A_n(t) - A_n(s)$ is nonnegative definite for $t > s \ge 0$. Assume that for each $T > 0$,*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\le T} \left|A_n^{ij}(t) - A_n^{ij}(t-)\right|\right] = 0, \tag{5.6}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\le T} \left|M_n(t) - M_n(t-)\right|^2\right] = 0, \tag{5.7}$$

and for $i, j = 1, \dots, d$,

$$M_n^i(t) M_n^j(t) - A_n^{ij}(t) \quad \text{is an } \mathbb{F}^n\text{-local martingale.}$$

If for each $t \ge 0$ and $i, j = 1, \dots, d$,

$$A_n^{ij}(t) \to c_{ij}(t) \quad \text{in probability,} \tag{5.8}$$

where $C = (c_{ij})$ is a continuous, symmetric, $d \times d$ matrix-valued function, defined on $[0,\infty)$ satisfying $C(0) = 0$ and $\sum (c_{ij}(t) - c_{ij}(s)) \xi_i \xi_j \ge 0$, $\xi \in \mathbb{R}^d$. Then,

$$M_n \Rightarrow X,$$

where X is a process with independent Gaussian increments such that $X_i X_j - c_{ij}$ are $\mathbb{F}^X$-local martingales.

Remark 5.2.4. (i) *This theorem uses the martingale property to characterize a Markov process associated with the given generator $\mathcal{A}$. Moreover, the L_2 convergence conditions in the theorem is to ensure the uniform integrability for the limiting process of the martingale sequence to be a martingale. The convergence of probability and L_1 convergence (implied by L_2 convergence) implies the sequence is relatively compact and thus has a limit point, and the convergence follows due to the unique characterization of the martingale with Gaussian increments.*
(ii) *A more general diffusion approximation will be introduced in Theorem 6.2.1 for another process (UDCP).*

We first provide the diffusion approximation of all jump components in the BDCP as the building block.

Lemma 5.2.5 (Diffusion Approximation with Brownian Motions)**.**
For $j,k=1,2$, let $\lim_{n\to\infty}\max(\rho_1^n,\rho_2^n)=\infty$ where ρ_k^n is the intensity of the external factor M^k in λ^k, then the following weak convergence holds:

$$\begin{bmatrix} \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}} \\ \frac{J_t^{N^k,j} - \mu_{G^{j,k}} N_t^k}{\sqrt{v_k}} \\ \frac{C_t^k - \mu_{G^{c,k}} N_t^k}{\sqrt{v_k}} \\ \frac{J_t^{M^k} - \mu_{H^k}\rho_k^n t}{\sqrt{v_k}} \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\alpha_k} B_t^{N,k} \\ \sqrt{v_{G^{j,k}}\alpha_k} B_t^{J^k,j} \\ \sqrt{v_{G^{c,k}}\alpha_k} B_t^{C,k} \\ \sqrt{\mu_{2H^k}\theta_k} B_t^{M,k} \end{bmatrix}, \tag{5.9}$$

where $B_t^{N,k}$, $B_t^{J^k,j}$, $B_t^{C,k}$ and $B_t^{M,k}$ are mutually independent Brownian motions for all $j,k=1,2$.

Moreover, as $n\to\infty$,

$$\begin{aligned} \frac{J_t^{N^k,j} - \mu_{G^{j,k}} \int_0^t \lambda_s^k ds}{\sqrt{v_k}} &\Rightarrow \sqrt{v_{G^{j,k}}}\sqrt{\alpha_k} B_t^{J^k,j} + \mu_{G^{j,k}}\sqrt{\alpha_k} B_t^{N,k}, \\ \frac{C_t^k - \mu_{G^{c,k}} \int_0^t \lambda_s^k ds}{\sqrt{v_k}} &\Rightarrow \sqrt{v_{G^{c,k}}}\sqrt{\alpha_k} B_t^{C,k} + \mu_{G^{c,k}}\sqrt{\alpha_k} B_t^{N,k}. \end{aligned}$$

We aim for a diffusion system as an approximation of the point process system with the same mean and variance. Therefore we focus on the normalized intensity process $Z^{(n)}$ where for any $t \geq 0$,

$$Z_t^{k,(n)} := \frac{\lambda_t^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}}. \tag{5.10}$$

By the central limit theorem, the initial value

$$Z_0^{(n)} = \begin{bmatrix} Z_0^{1,(n)} \\ Z_0^{2,(n)} \end{bmatrix} \sim \left(0, \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}\right) \Rightarrow Z_0 \sim N\left(0, \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}\right), \tag{5.11}$$

where ρ_{12} is the stationary correlation from (5.2), and $N(\mu, \Sigma)$ denotes a Gaussian random vector with mean μ and covariance Σ. Based on Lemma 5.2.5, we obtain the following diffusion approximation for $Z^{(n)}$.

Lemma 5.2.6 (Diffusion Approximation with OU Process)**.** *Take the stationary version of (N, λ), as $\lim_{n\to\infty} \max(\rho_1^n, \rho_2^n) = \infty$, the process $Z_t^{(n)} = \begin{bmatrix} Z_t^{1,(n)} \\ Z_t^{2,(n)} \end{bmatrix}$ in (5.10) converges to $Z_t = \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix}$ weakly, where Z_t has the initial distribution as in (5.11) and the dynamic in matrix form*

$$dZ_t = bZ_t dt + \Sigma dB_t^W + \tilde{\Sigma} dB_t^N \tag{5.12}$$

where B_t^W and B_t^N are independent standard two-dimensional Brownian motions from (5.10) and with parameters in (5.5),

$$\begin{aligned}
b &= \begin{bmatrix} -(\delta_1 - \mu_{G^{1,1}}) & \nu_{2,1}\mu_{G^{1,2}} \\ \nu_{1,2}\mu_{G^{2,1}} & -(\delta_2 - \mu_{G^{2,2}}) \end{bmatrix}, \\
\tilde{\Sigma} &= \begin{bmatrix} \sqrt{\alpha_1\nu_{1,1}}\mu_{G^{1,1}} & \sqrt{\alpha_2\nu_{2,1}}\mu_{G^{1,2}} \\ \sqrt{\alpha_1\nu_{1,2}}\mu_{G^{2,1}} & \sqrt{\alpha_2\nu_{2,2}}\mu_{G^{2,2}} \end{bmatrix}, \\
\Sigma &= \begin{bmatrix} \sqrt{\theta_1\mu_{2H^1} + \nu_{1,1}v_{G^{1,1}}\alpha_1 + \nu_{2,1}v_{G^{1,2}}\alpha_2} & 0 \\ 0 & \sqrt{\theta_2\mu_{2H^2} + \nu_{1,2}v_{G^{2,1}}\alpha_1 + \nu_{2,2}v_{G^{2,2}}\alpha_2} \end{bmatrix}.
\end{aligned} \tag{5.13}$$

Combined with Lemma 5.2.5, we obtain

$$\begin{aligned}
\frac{N_t^k - m_k^{(n)} t}{\sqrt{v_k^{(n)}}} &= \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k^{(n)}}} + \int_0^t \frac{\lambda_s^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}} ds \\
&\Rightarrow \sqrt{\alpha_k} B_t^{N,k} + \int_0^t Z_s^{k,(n)} ds. \\
\frac{C_t^k - \mu_{G^{c,k}} m_k^{(n)} t}{\sqrt{v_k^{(n)}}} &= \frac{C_t^k - \mu_{G^{c,k}} N_t}{\sqrt{v_k^{(n)}}} + \mu_{G^{c,k}} \frac{N_t - m_k^{(n)} t}{\sqrt{v_k^{(n)}}} \\
&\Rightarrow \sqrt{v_{G^{c,k}} \alpha_k} B_t^{C,k} + \sqrt{\alpha_k} \mu_{G^{c,k}} B_t^{N,k} + \mu_{G^{c,k}} \int_0^t Z_s^{k,(n)} ds.
\end{aligned}$$

Therefore we obtain the diffusion approximation of the unnormalized original process in the following.

Theorem 5.2.7 (Diffusion Approximation of the BDCP)**.**
For a stationary BDCP system (λ, N, C) with $\max(\rho_1, \rho_2)$ *large, there is a diffusion approximation $(\tilde{\lambda}, \tilde{N}, \tilde{C})$ of the BDCP:*

$$\begin{aligned}
\tilde{\lambda}_t &= m + \sqrt{v} Z_t = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{bmatrix} \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix}, \\
\tilde{N}_t &= mt + \sqrt{\alpha v} U_t = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} t + \begin{bmatrix} \sqrt{\alpha_1 v_1} & 0 \\ 0 & \sqrt{\alpha_2 v_2} \end{bmatrix} \begin{bmatrix} U_t^1 \\ U_t^2 \end{bmatrix}, \\
\tilde{C}_t &= m_c t + \sqrt{v} V_t = \begin{bmatrix} \mu_{G^{c,1}} m_1 \\ \mu_{G^{c,2}} m_2 \end{bmatrix} t + \begin{bmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{bmatrix} \begin{bmatrix} V_t^1 \\ V_t^2 \end{bmatrix},
\end{aligned} \tag{5.14}$$

with (Z, U, V) being a linear diffusion system with dynamic:

$$\begin{aligned}
dZ_t &= bZ_t dt + \Sigma dB_t^W + \tilde{\Sigma} dB_t^N, \\
dU_t &= hZ_t dt + dB_t^N, \\
dV_t &= rZ_t dt + \sigma dB_t^C + \tilde{\sigma} dB_t^N,
\end{aligned} \tag{5.15}$$

where $b, \Sigma, \tilde{\Sigma}$ are given in (5.13) and

$$\begin{aligned}
h &:= \begin{bmatrix} \frac{1}{\sqrt{\alpha_1}} & 0 \\ 0 & \frac{1}{\sqrt{\alpha_2}} \end{bmatrix}, r := \begin{bmatrix} \mu_{G^{c,1}} & 0 \\ 0 & \mu_{G^{c,2}} \end{bmatrix}, \\
\sigma &:= \begin{bmatrix} \sqrt{v_{G^{c,1}} \alpha_1} & 0 \\ 0 & \sqrt{v_{G^{c,2}} \alpha_2} \end{bmatrix}, \tilde{\sigma} := \begin{bmatrix} \sqrt{\alpha_1} \mu_{G^{c,1}} & 0 \\ 0 & \sqrt{\alpha_2} \mu_{G^{c,2}} \end{bmatrix}.
\end{aligned} \tag{5.16}$$

Moreover, the initial value Z_0 is in (5.11), $U_0 = \mathbf{0}$, $V_0 = \mathbf{0}$ and $\mathbf{0}$ is the zero vector of dimension 2×1.

Remark 5.2.8. *The diffusion process (Z, U) from the normalized process of (λ, N) share the common noise B^N due to the mutually-exciting component in the BDCP. Moreover, the dynamic of the diffusion process (Z, U, V) is independent of ρ.*

Corollary 5.2.9. *Denote $I_t := \left(\int_t^T e^{bs} ds\right) e^{-bt}$, then*

$$V_T - V_t \quad = \quad rI_t Z_t + \int_t^T rI_u \Sigma dB_u^W + \int_t^T \sigma dB_u^C + \int_t^T \left[rI_u + \tilde{\sigma}\right] B_u^N. \tag{5.17}$$

5.3 Kalman-Bucy Filtering with Diffusion Approximation

Many practical problems in a point process system are related to the intensity process which is unobservable. Estimating the intensity based on the observable point process is a filtering problem. We propose an approximation solution of the filtering problem as an application of the diffusion approximation of the BDCP.

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F}$ where the BDCP is defined, let $\{\mathcal{X}_t\}_{t\geq 0}$ and $\{\mathcal{Y}_t\}_{t\geq 0}$ be the filtration generated by the signal process X and observation process Y, and $\mathcal{X}_t \subset \mathcal{F}_t$ and $\mathcal{Y}_t \subset \mathcal{F}_t$. We discuss the signal-observation system in terms of the diffusion system obtained in (5.15) for two scenarios of observations (S1) and (S2) as follows.

(S1) Complete observation of events N^1, N^2

We aim to obtain the estimate of (λ^1, λ^2) with observations of (N^1, N^2). In this case, we consider the normalized diffusion system

$$X = (Z^1, Z^2), \quad Y = (U^1, U^2).$$

(S2) Partial observation of events N^1, N^2

This corresponds to the problem that we observe the event arrivals but cannot distinguish the type of them, i.e. the observable quantity is $N^1 + N^2$. We aim to obtain the estimate of $(\lambda^1, \lambda^2, N^1, N^2)$ given the observation of $N^1 + N^2$. The type estimation problem of (N^1, N^2) is a particular problem for multi-dimensional system. According to (5.14), the normalized diffusion system is therefore

$$X = (Z^1, Z^2, U^1, U^2), \quad Y = \sqrt{\alpha_1 v_1} U^1 + \sqrt{\alpha_2 v_2} U^2.$$

Observe that in both scenarios, the approximate signal-observation system (X, Y) is a Gaussian system where an explicit solution of the filtering problem is available from the Kalman-Bucy filtering theory which we will introduce first.

Lemma 5.3.1 (Kalman-Bucy Filtering with Correlated Noise, Xiong [62])**.** *Suppose the signal process X and observation process Y are random vectors of dimension $d \times 1$ and $m \times 1$ respectively and the signal-observation system (X, Y) is given by the solution of*

$$\begin{aligned} dX_t &= (\tilde{b}_t + b_t X_t)dt + c_t dW_t + \sigma_t dB_t, \\ dY_t &= (\tilde{h}_t + h_t X_t)dt + dW_t, \end{aligned} \tag{5.18}$$

where X_0 is a normal random vector with mean $\hat{X}_0 \in \mathbb{R}^d$ and covariance matrix $\gamma_0 \in \mathbb{R}^{d \times d}$, (W, B) is an $m + d$-dimensional Brownian motion, the coefficients $\tilde{b}_t, b_t, c_t, \sigma_t, \tilde{h}_t, h_t$ are deterministic matrices (or vectors) of dimension $d \times 1$, $d \times d$, $d \times m$, $m \times 1$, $m \times d$, respectively.

Then, for any $t > 0$, the conditional distribution X_t given $\mathcal{Y}_t$ is a Gaussian process with mean $\hat{X}_t$ and covariance γ_t being the unique solution of

$$\begin{aligned} d\hat{X}_t &= \left(\tilde{b}_t + b_t \hat{X}_t\right) dt + K_t dv_t, \\ \frac{d}{dt}\gamma_t &= \gamma_t b_t^T + b_t \gamma_t + (c_t c_t^T + \sigma_t \sigma_t^T) - (c_t + \gamma_t h_t^T)(c_t + \gamma_t h_t^T)^T, \end{aligned} \tag{5.19}$$

where

$$\begin{aligned} v_t &= Y_t - \int_0^t \left(\tilde{h}_s + h_s \hat{X}_s\right) ds \quad \textit{is the innovation process, and} \\ K_t &= \gamma_t h_t^T + c_t \quad \textit{is the Kalman gain.} \end{aligned}$$

Note that γ_t solves the matrix Riccati equation that is deterministic and does not depend on the observation Y. In Section 5.3.1 and 5.3.2 below, we explore the filtering problem for the scenarios (S1) and (S2) respectively.

5.3.1 Filtering with Observation $\left(N^1, N^2\right)$

As discussed in (S1), the signal-observation system is $(X, Y) = (Z, U)$ with the dynamic

$$\begin{aligned} dX_t &= bX_t dt + \Sigma dB_t^W + \tilde{\Sigma} dB_t^N, \\ dY_t &= hX_t dt + dB_t^N, \end{aligned} \tag{5.20}$$

where the coefficients can be found in (5.13) and (5.16). Then we can apply Lemma 5.3.1 directly.

Proposition 5.3.2 (Filtering with (N^1, N^2)). *Let (X, Y) be in (5.20), for any $t > 0$, the estimate $X_t|\mathcal{Y}_t$ is a Gaussian process with the mean $\hat{X}_t$ and variance γ_t being the unique solution of*

$$\begin{aligned} &d\hat{X}_t = b\hat{X}_t dt + K_t dv_t, \quad \hat{X}_0 = \mathbb{E}[Z_0], \\ &\frac{d}{dt}\gamma_t = \gamma_t b^T + b\gamma_t + (\tilde{\Sigma}\tilde{\Sigma}^T + \Sigma\Sigma^T) - K_t K_t^T, \quad \gamma_0 = var(Z_0), \end{aligned} \tag{5.21}$$

where $K_t = \tilde{\Sigma} + \gamma_t h^T$ and $v_t = Y_t - \int_0^t h\hat{X}_s ds$.

Therefore we obtain the estimate of the normalized processes as

$$Z_t|\mathcal{Y}_t \sim N\left(\hat{Z}_t, \gamma_t\right). \tag{5.22}$$

We obtain the estimate of the unnormalized intensity process $\tilde{\lambda}_t$ by (5.14)

$$\tilde{\lambda}_t|\mathcal{Y}_t \sim N\left(m + \hat{Z}_t, v\gamma_t\right). \tag{5.23}$$

5.3.2 Filtering with Observation $N^1 + N^2$

In Scenario (S2) where the event types are unobservable, so the information is more coarse than (S1). The signal-observation system becomes

$$\begin{aligned} X_t &= \left[Z_t^1, Z_t^2, U_t^1, U_t^2\right]^T, \\ Y_t &= \begin{bmatrix}\sqrt{\alpha_1 v_1} & \sqrt{\alpha_2 v_2}\end{bmatrix} \begin{bmatrix} U_t^1 \\ U_t^2 \end{bmatrix}, \end{aligned} \tag{5.24}$$

with the dynamics

$$\begin{aligned} dX_t &= \mu^X X_t dt + \Sigma^{X,N} dB_t^N + \Sigma^{X,W} dB_t^W, \\ dY_t &= \mu^Y X_t dt + \Sigma^Y dB_t^N, \end{aligned} \tag{5.25}$$

where $\mu^X = \begin{bmatrix} b & \mathbf{0} \\ h & \mathbf{0} \end{bmatrix}$, $\Sigma^{X,N} = \begin{bmatrix} \tilde{\Sigma} \\ I \end{bmatrix}$, $\Sigma^{X,W} = \begin{bmatrix} \Sigma \\ \mathbf{0} \end{bmatrix}$ are vectors of dimension 4×4, 4×2 and 4×2 respectively, and $\mathbf{0}$ and I are 2×2 zero and identity matrix. Moreover, $\mu^Y = \begin{bmatrix} \sqrt{v_1} & \sqrt{v_2} & 0 & 0 \end{bmatrix}$ and $\Sigma^Y = \begin{bmatrix} \sqrt{v_1\alpha_1} & \sqrt{v_2\alpha_2} \end{bmatrix}$.

Note that we need to transform the noises in the system (5.25) such that the noise term of B^N in X can be written as a linear transformation of the noise in Y to apply the Lemma 5.3.1 to solve the filtering problem.

Lemma 5.3.3 (Noise Transformation). *We can rewrite the processes* (X, Y) *in (5.25) as*

$$\begin{aligned} dX_t &= \mu^X X_t dt + \Sigma_1^{*X} dB_t^{Y,1} + \Sigma^X dB_t^X, \\ dY_t &= \mu^Y X_t dt + \Sigma_1^{*Y} dB_t^{Y,1}, \end{aligned} \tag{5.26}$$

where B^X *and* $B^{Y,1}$ *are independent standard Brownian motions of dimension* 3×1 *and* 1×1*, and* Σ_1^{*X} *is the first column of* Σ^{*X} *below*

$$\Sigma^{*X} = \frac{1}{\sqrt{\alpha_1 v_1 + \alpha_2 v_2}} \begin{bmatrix} \tilde{\Sigma}_{11}\sqrt{\alpha_1 v_1} + \tilde{\Sigma}_{12}\sqrt{\alpha_2 v_2} & \tilde{\Sigma}_{11}\sqrt{\alpha_2 v_2} - \tilde{\Sigma}_{12}\sqrt{\alpha_1 v_1} \\ \tilde{\Sigma}_{21}\sqrt{\alpha_1 v_1} + \tilde{\Sigma}_{22}\sqrt{\alpha_2 v_2} & \tilde{\Sigma}_{21}\sqrt{\alpha_2 v_2} - \tilde{\Sigma}_{22}\sqrt{\alpha_1 v_1} \\ \sqrt{\alpha_1 v_1} & \sqrt{\alpha_2 v_2} \\ \sqrt{\alpha_2 v_2} & -\sqrt{\alpha_1 v_1} \end{bmatrix}$$

Moreover, $\Sigma^X = \begin{bmatrix} \Sigma_{12}^{*X} & \Sigma_{11} & \Sigma_{12} \\ \Sigma_{22}^{*X} & \Sigma_{21} & \Sigma_{22} \\ \Sigma_{32}^{*X} & 0 & 0 \\ \Sigma_{42}^{*X} & 0 & 0 \end{bmatrix}$ *and* $\Sigma_1^{*Y} = \sqrt{\alpha_1 v_1 + \alpha_2 v_2}$.

Then, we obtain the following filtering result.

Proposition 5.3.4 (Filtering with $N^1 + N^2$). *Let* (X, Y) *be in (5.24), for any* $t > 0$, $X_t|\mathcal{Y}_t$ *is Gaussian process with mean* $\hat{X}_t = (\hat{Z}_t, \hat{U}_t)^T$ *and variance process* $\gamma_t^X = \begin{bmatrix} \gamma_t & \gamma_t^{X,U} \\ \gamma_t^{X,U} & \gamma_t^U \end{bmatrix}$ *being the unique solution of*

$$\begin{aligned} & d\hat{X}_t = \mu^X \hat{X}_t dt + K_t dv_t, \quad \hat{X}_0 = \begin{bmatrix} E[Z_0]^T & \mathbf{0} \end{bmatrix}^T, \\ & \frac{d}{dt}\gamma_t^X = \gamma_t^X (\mu^X)^T + \mu^X (\gamma_t^X)^T + \Sigma_1^{*X}\left(\Sigma_1^{*X}\right)^T + \Sigma^X\left(\Sigma^X\right)^T - K_t K_t^T, \quad \gamma_0^X = \begin{bmatrix} var(Z_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \end{aligned} \tag{5.27}$$

where $\upsilon_t = \frac{1}{\Sigma_1^{*Y}}\left(Y_t - \int_0^t \mu^Y \hat{X}_s ds\right)$ *is of size* 1×1*, and* $K_t = \gamma_t^X \left(\frac{\mu^Y}{\Sigma_1^{*Y}}\right)^T + \Sigma_1^{*X}$ *is of dimension* 4×1*.*

Therefore we obtain the estimate of the normalized processes as

$$\begin{aligned} Z_t|\mathcal{Y}_t &\sim N\left(\hat{Z}_t, \gamma_t\right), \\ U_t|\mathcal{Y}_t &\sim N\left(\hat{U}_t, \gamma_t^U\right). \end{aligned} \tag{5.28}$$

We obtain the estimate of the unnormalized process by (5.14) as

$$\begin{aligned} \tilde{\lambda}_t|\mathcal{Y}_t &\sim N\left(m + \hat{Z}_t, \upsilon\gamma_t\right), \\ \tilde{N}_t|\mathcal{Y}_t &\sim N\left(m + \hat{U}_t, \alpha\upsilon\gamma_t^U\right). \end{aligned} \tag{5.29}$$

5.3.3 An Approximate Filter

In this section, we construct a linear estimate for the BDCP by plugging the actual observation of the BDCP into the filter of the approximating diffusion system obtained in the last section. Note that we embed the actual data in a sequence $\{(X^n, Y^n)\}_{n\geq 1}$ constructed with the parameter sequence $\{\rho_n^k\}_{n\geq 1}$ which tends to infinity. Then the original BDCP system corresponds to (X^n, Y^n) for some large $n \geq 1$ in the sequence.

Recall from Section 5.3.1 and 5.3.2, we obtain the Kalman-Bucy filter estimate for the normalized diffusion system (X, Y). At any time $t \geq 0$, the conditional distribution $X_t|\mathcal{Y}_t$ is a Gaussian random variable characterized by $(\hat{X}_t, \gamma_t)$ where $\hat{X}_t$ is the mean estimate $\hat{X}_t = \mathbb{E}[X_t|\mathcal{Y}_t]$, and γ_t is the variance estimate and the mean-square filtering error as well. For each $t \geq 0$, we define a functional

$$F_t(x, y) : \mathbb{R}^2 \times D_{\mathbb{R}^2}[0, \infty) \to \mathbb{R}^2$$

with which $\hat{X}$ can be written as $\hat{X}_t = F_t(X_0, Y)$, where X_0 is the initial state of X assumed to be known at $t = 0$. From the dynamics of $\hat{X}_t$ in (5.21) and (5.27), we know that F_t is a continuous functional of $y \in D_{\mathbb{R}^2}[0, \infty)$.

Due to the weak convergence of (X^n, Y^n), the natural choice of an approximate estimate is to use the filter from the limit diffusion model with the actual point process input. In other words, we propose an approximate filter for the conditional distribution $X_t^n|\mathcal{Y}_t^n$. Assume the filter is a Gaussian process with mean $\hat{X}_t^n$ and variance γ_t at each $t \geq 0$, where

$$\hat{X}_t^n = F_t(X_0^n, Y^n), \tag{5.30}$$

and γ_t is the same as in the diffusion system.

Remark 5.3.5. *Essentially we constructed a linear filter estimate by applying the Kalman-Bucy filter on the point process system.*

(X^n, Y^n) in both observation scenarios (S1) and (S2) can be represented in terms of the scaled jump processes $Z_t^{(n)}$ and $U_t^{(n)}$:

$$\begin{aligned} Z_t^{(n)} &= (\sqrt{v})^{-1}(\lambda_t - m), \\ U_t^{(n)} &= (\sqrt{\alpha v})^{-1}(N_t - mt). \end{aligned} \tag{5.31}$$

For (S1), we have

$$X^n = Z^{(n)}, \quad Y^n = U^{(n)},$$

then we obtain the Kalman-Bucy filter estimate $\hat{X}_t^n$ and variance γ_t from (5.21):

$$\begin{aligned} &d\hat{X}_t^n = b\hat{X}_t^n dt + K_t dv_t^n, \quad \hat{X}_0^n = \mathbb{E}[Z_0^n], \\ &\frac{d}{dt}\gamma_t = \gamma_t b^T + b\gamma_t + (\tilde{\Sigma}\tilde{\Sigma}^T + \Sigma\Sigma^T) - K_t K_t^T, \quad \gamma_0 = var(Z_0^n), \end{aligned} \tag{5.32}$$

with the innovation

$$v_t^n = Y_t^n - \int_0^t h\hat{X}_s^n ds.$$

For (S2), we have

$$X_t^n = \left[Z_t^{1,(n)}, Z_t^{2,(n)}, U_t^{1,(n)}, U_t^{2,(n)}\right]^T, \quad Y_t^n = \left[\sqrt{\alpha_1 v_1}\sqrt{\alpha_2 v_2}\right]\begin{bmatrix} U_t^{1,(n)} \\ U_t^{2,(n)} \end{bmatrix},$$

then we obtain the Kalman-Bucy filter estimate $\hat{X}_t^n$ and variance γ_t from (5.27) :

$$\begin{aligned} &d\hat{X}_t^n = \mu^X \hat{X}_t^n dt + K_t dv_t^n, \quad \hat{X}_0 = \left[E[Z_0^n]^T \quad \mathbf{0}\right]^T, \\ &\frac{d}{dt}\gamma_t^X = \gamma_t^X(\mu^X)^T + \mu^X(\gamma_t^X)^T + \Sigma_1^{*X}\left(\Sigma_1^{*X}\right)^T + \Sigma^X\left(\Sigma^X\right)^T - K_t K_t^T, \quad \gamma_0^X = \begin{bmatrix} var(Z_0^n) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \end{aligned} \tag{5.33}$$

with the innovation

$$v_t^n = \frac{1}{\Sigma_1^{*Y}}\left(Y_t^n - \int_0^t \mu^Y \hat{X}_s^n ds\right).$$

In the following we discuss the asymptotic results of constructed estimates for (S1) without loss of generality.

For $(X^n, Y^n) = (Z^{(n)}, U^{(n)})$ in the point process system, due to the continuity of the functional F_t for each $t \geq 0$, and by the continuous mapping theorem, we obtain immediately the weak convergence

$$(X_t^n, \hat{X}_t^n) \Rightarrow (X_t, \hat{X}_t),$$

where X and $\hat{X}_t$ are the diffusion process Z and its estimate (see the proof in Section 5.5).

We follow the idea in [48] and [53] to show that the construction is appropriate in an important way presented in Proposition 5.3.6 and Proposition 5.3.7.

Proposition 5.3.6 (Asymptotic optimality). *Let f be any bounded continuous and real-valued function. Denote by q_f^* the integral of f with respect to the Gaussian distribution with mean $\hat{X}_t^n$ and variance γ_t as a function of input Y^n up to time t. Then, for any $t \geq 0$ and any measurable and continuous function q of Y^n defined on $[0, t)$,*

$$\lim_{n\to\infty} \mathbb{E}\left[|f(X_t^{k,n}) - q_f^*(Y^{k,n})|^2\right] \leq \lim_{n\to\infty} \mathbb{E}\left[|f(X_t^{k,n}) - q(Y^{k,n})|^2\right]. \tag{5.34}$$

In other words, the conditional distribution is nearly optimal with respect to a broad class of alternative estimators that are continuous and bounded functions of the observations.

Moreover, we provide the result of the asymptotic estimate error to assess how good the estimate of the normalized processes is when n is big.

Proposition 5.3.7 (Asymptotic estimation error). *For any $t \geq 0$, assume that the third moments of the jump size Y_i^k and $Z_i^{k,k'}$ are finite, i.e. $\mu_{3H^k} < \infty$ and $\mu_{3G^{k,k'}} < \infty$ for $k, k' = 1, 2$. Then, we have the asymptotic estimation error converges to the error of the diffusion system:*

$$\lim_{n\to\infty} \mathbb{E}\left[\left(X_t^n - \hat{X}_t^n\right)\left(X_t^n - \hat{X}_t^n\right)^T\right] = \mathbb{E}\left[\left(X_t - \hat{X}_t\right)\left(X_t - \hat{X}_t\right)^T\right] = \gamma_t. \tag{5.35}$$

In particular,

$$\lim_{n\to\infty} \mathbb{E}\left[|X_t^{k,n} - \hat{X}_t^{k,n}|^2\right] = \mathbb{E}\left[|X_t^k - \hat{X}_t^k|^2\right] = (\gamma_t)_{k,k}. \tag{5.36}$$

Remark 5.3.8. *The property in (5.36) is called the asymptotic optimality of the filter variance in [53].*

Graphic Illustration (S1) Take the exponential jump sizes of λ: $Y_i^k \sim exp(\beta_k^Y)$, $Z_i^{k,k'} \sim exp(\beta_{k,k'}^Z)$ and $C_i^k \sim exp(\beta_k^C)$. All parameters are in Table 5.1 and they will also be used in applications in Section 5.4.

$\beta_{k,k}^Z = 5\ (k = 1, 2)$	$\beta_{k,k'}^Z = 2\ (k \neq k')$
$\rho_1 = 10$	$\rho_2 = 10$
$\delta_1 = 2$	$\delta_2 = 1$
$\beta_1^Y = 1$	$\beta_2^Y = 1$
$\beta_1^C = 1$	$\beta_2^C = 1$

Table 5.1: Parameters for numerical applications

The comparison is based on the Monte-Carlo simulation of the BDCP (N, λ). We first simulate a sample path of (N, λ). Conditioned on the path of observation (N^1, N^2), we obtain the filtered intensity process based on Proposition 5.3.2 and compare the filtered mean process $\hat{\lambda}^k$ with the original process λ^k in Figure 5.1.

Moreover, in Figure 5.2, we compare the empirical distribution of λ_T and $\lambda_T|\mathcal{Y}_T$ in (S1) with the observation $\mathcal{Y}_T = \sigma\{N_t^1, N_t^2, t \leq T\}$ at a specific time $T = 10$.

5.4 Applications

We apply the result in Section 5.3 to two typical problems in insurance modelling.

Pricing problem We can price stop-loss reinsurance contracts with claims C_t as in (5.1) in a high frequency event environment. Under the physical probability measure $\mathbb{P}$, the stop-loss reinsurance premium at time t is

$$\mathbb{E}\left[H\left(C_T - C_t\right)|\mathcal{Y}_t\right], \tag{5.37}$$

where observations $\mathcal{Y}_t$ in the market can be from either (S1) or (S2) and $H : \mathbb{R}^2 \to \mathbb{R}_+$ is a function including the discount factor under the physical measure and also the payoff function. For simplicity we assume a constant discount factor, then H is a payoff function can be $H(x) = (x_k - d)^+$ for claims in a single portfolio of type

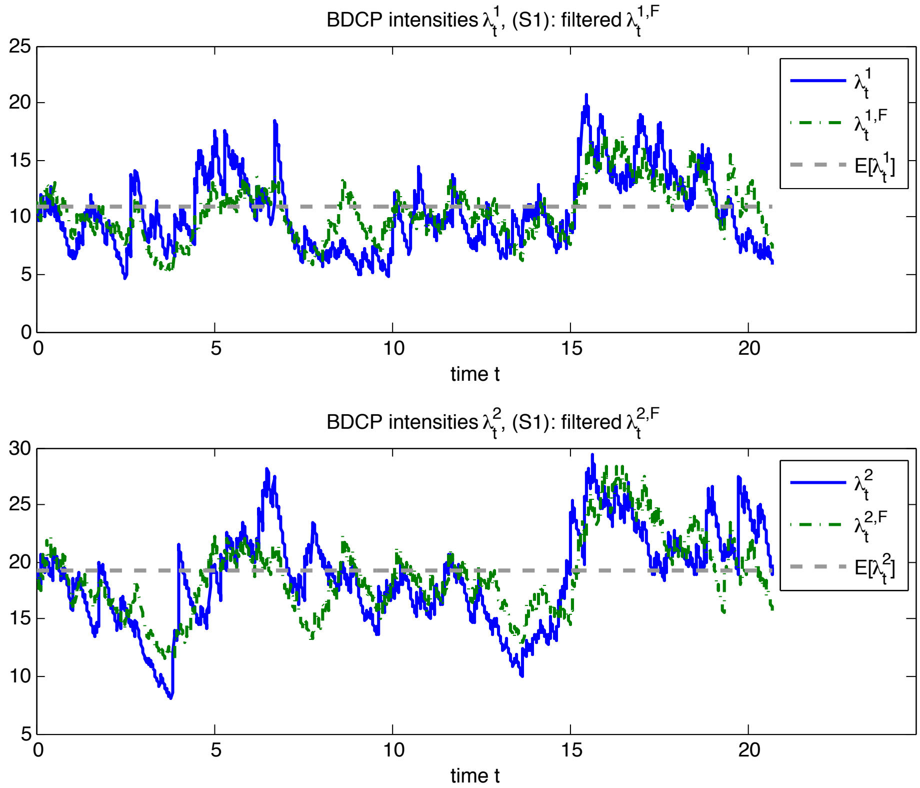


Figure 5.1: Observation scenario (S1): A sample path of λ_t^k compared with its unconditional stationary mean and conditional estimate $\lambda_t^{k,F} = \mathbb{E}[\lambda_t^k|\sigma\{N_s^1, N_s^2, s \leq t\}]$ for $k = 1, 2$.

$k = 1, 2$, or $H(x) = (\alpha_1 x_1 + \alpha_2 x_2 - d)^+$ for claims on a joint portfolio with d as the retention level.

Event type identification In the case we can't distinguish the event types as in (S2), we aim to find the distribution of the number of events of a certain type in the interval $[0, t]$. It corresponds to

$$\mathbb{P}\left(N_t^i \leq n | \mathcal{Y}_t\right), \tag{5.38}$$

where $\mathcal{Y}_t = \sigma\left\{N_s^1 + N_s^2, s \leq t\right\}$ and $n \leq N_t^1 + N_t^2$.

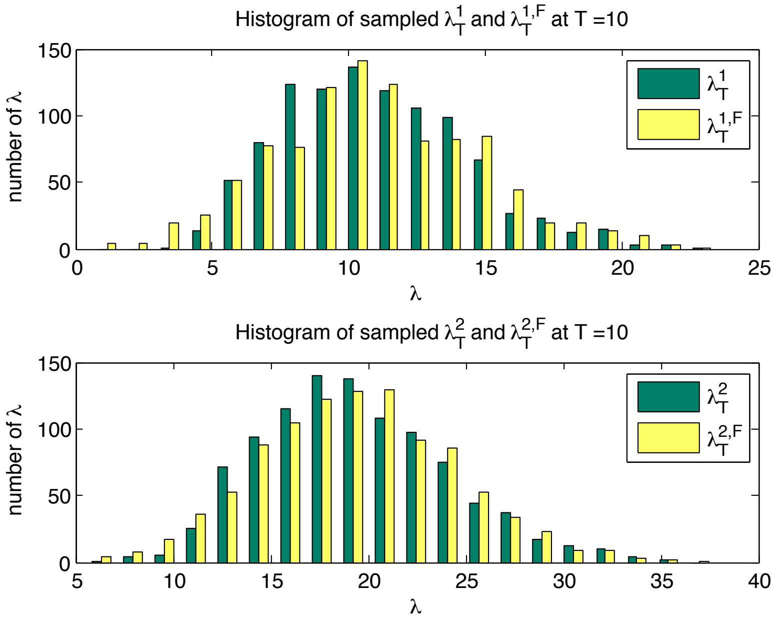


Figure 5.2: Histograms describing the empirical distribution of λ_T^k and $\lambda_T^k|\sigma\{N_t^1, N_t^2, t \leq T\}$ denoted by $\lambda_T^{k,F}$ in the graph at $T = 10$ for $k = 1, 2$.

5.4.1 Pricing Problem

We transform the pricing problem into solving the prediction problem using the structure of diffusion system and the actual observation input. Recall that $\tilde{C}_t = m_c t + \sqrt{v} V_t$, the conditional law of $V_T - V_t$ is multivariate Gaussian by (5.17). Then we have the following result.

Proposition 5.4.1. *For any $t \geq 0$, in the diffusion system, define a functional $H_t(\cdot) : D_{\mathbb{R}^2}[0, \infty) \to \mathbb{R}$ with*

$$\begin{aligned} H_t(Y) &= \mathbb{E}\left[H\left(\tilde{C}_T - \tilde{C}_t\right)|\mathcal{Y}_t\right] \\ &= \mathbb{E}\left[H(m_c(T-t) + \sqrt{v}\,(V_T - V_t))|\mathcal{Y}_t\right] \\ &= \int_{\mathbb{R}^2} H((m_c(T-t) + \sqrt{v}x))f(x)dx, \end{aligned}$$

where $f(x)$ is the density function of the conditional law of $V_T - V_t$ that follows the multivariate Gaussian distribution with mean $\mu_{t,T}^V$ and covariance $\Sigma_{t,T}^V$:

$$f(x) = \frac{1}{2\pi}\left|\Sigma_{t,T}^V\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(x - \mu_{t,T}^V\right)^T \left(\Sigma_{t,T}^V\right)^{-1}\left(x - \mu_{t,T}^V\right)\right) \quad x \in \mathbb{R}^2,$$

where

$$
\begin{aligned}
&\mu^V_{t,T} = rI_t\hat{Z}_t,\\
&\Sigma^V_{t,T} = \Big(rI_t\gamma_t\,(rI_t)^T + \int_t^T \left(\sigma^W_u(\sigma^W_u)^T + \sigma^N_u(\sigma^N_u)^T + \sigma^C_u(\sigma^C_u)^T\right) du,
\end{aligned}
\tag{5.39}
$$

$\sigma^W_u = rI_u\Sigma$, $\sigma^N_u = rI_u\tilde{\Sigma} + \tilde{\sigma}$, $\sigma^C_u \equiv \sigma$, *and* $(\hat{Z}_t, \gamma_t)$ *is the solution of (5.21) and (5.27) for observation scenarios (S1) and (S2) respectively.*

Then, given the actual observation process Y^n *for some* n*, the price of the contract with payoff* H *at the maturity* T *based on the estimation given the observation* Y *up to time* t *can be computed approximately as* $H_t(Y^n)$*.*

We provide a numerical example to illustrate where we take the payoff function as $H(x) = (x_1 + x_2 - d)^+$ for $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ using the Kalman-Bucy filter and the actual input of the point process.

To see the impact of the update of observation on the estimate, we compare the conditional with the unconditional expected values:

$$
\begin{aligned}
\mathbb{E}\left[H(C_T - C_t)\right] &= \mathbb{E}\left[N^1_T - N^1_t\right]\mathbb{E}\left[C^1_i\right] + \mathbb{E}\left[N^2_T - N^2_t\right]\mathbb{E}\left[C^2_i\right]\\
&= \left(m_1\mu_{G^{c,1}} + m_2\mu_{G^{c,2}}\right)(T - t),
\end{aligned}
$$

where m_k and $\mu_{G^{c,k}}$ are mean values of λ^k and c^k_i. In Table 5.2 we observe that at $t = 0$, there is only null observation information thus the values are the same. When $t > 0$, with the observation information included, we have the values deviated.

valuation time t	$\mathbb{E}\left[H(C_T - C_t)\right]$	$\mathbb{E}\left[H(C_T - C_t)\|\mathcal{F}^{N^1,N^2}_t\right]$	$\mathbb{E}\left[H(C_T - C_t)\|\mathcal{F}^{N^1+N^2}_t\right]$
0	598.7062	598.7062	598.7062
5	447.4457	454.9521	440.1838

Table 5.2: Comparison of the approximate conditional and unconditional expectations with $T = 19.5$.

Moreover, in Table 5.3, we compare the contract value with various retention levels d when pricing time t is fixed and various pricing time t with retention level d is fixed.

retention level d ($t=5$)	$\mathbb{E}\left[H(C_T - C_t)\middle|\mathcal{F}_t^{N^1,N^2}\right]$	$\mathbb{E}\left[H(C_T - C_t)\middle|\mathcal{F}_t^{N^1+N^2}\right]$
300	141.0305	140.226
420	32.4062	31.8488
460	12.1021	11.7923
pricing time t ($d=300$)	$\mathbb{E}\left[H(C_T - C_t)\middle|\mathcal{F}_t^{N^1,N^2}\right]$	$\mathbb{E}\left[H(C_T - C_t)\middle|\mathcal{F}_t^{N^1+N^2}\right]$
0	298.7062	298.7062
5	141.0305	140.226
10	22.9873	25.0886

Table 5.3: Approximate stop-loss contract value of a given path of observation (N_t^1, N_t^2), with $H(x) = (x_1 + x_2 - d)^+$ with $T = 19.5$.

5.4.2 Type Identification Problem

In the case of (S2) where we cannot distinguish event types, we can estimate the event type distribution in the diffusion system:

$$\mathbb{P}\left(\tilde{N}_t^k \le n \middle| \mathcal{Y}_t\right) = \mathbb{P}\left(U_t^k \le \frac{n - m_k t}{\sqrt{\alpha_k v_k}} \middle| \mathcal{Y}_t\right) = \Phi\left(\frac{\frac{n-m_k t}{\sqrt{\alpha_k v_k}} - \hat{U}_t^k}{\sqrt{\gamma_t^{U^k}}}\right),$$

and hence we have the following approximation solution for the original BDCP system.

Proposition 5.4.2. *For $k = 1, 2$, in the diffusion system we have*

$$\mathbb{P}\left(\tilde{N}_t^k \le n \middle| \sigma\{\tilde{N}_s^1 + \tilde{N}_s^2, s \le t\}\right) = \Phi\left(\frac{\frac{n-m_k t}{\sqrt{\alpha_k v_k}} - \hat{U}_t^k}{\sqrt{\gamma_t^{U^k}}}\right), \tag{5.40}$$

where $\hat{U}_t^k$ is from (5.27), $\gamma_t^{U^k} = \gamma_t^U(k,k)$ in (5.28) and Φ is the distribution function of the standard Gaussian distribution.

Then given $\mathcal{Y}_t = \sigma\{N_s^1 + N_s^2, s \le t\}$, the estimate $\mathbb{P}\left(N_t^k \le n \middle| \mathcal{Y}_t\right)$ can be computed approximately as (5.40) with $\hat{U}_t^k$ replaced by the actual observation input $\hat{U}_t^{k,(n)}$ in (5.33).

We illustrate with a numerical example. Given a path of $N^1 + N^2$, the type estimation results can be found in Table 5.4 and Figure 5.3.

t $(k=10)$	$\mathbb{P}\left(N_t^1 \leq k \mid \mathcal{F}_t^{N^1+N^2}\right)$	$\mathbb{P}\left(N_t^2 \leq k \mid \mathcal{F}_t^{N^1+N^2}\right)$
0	1	1
0.5	0.9895	0.5749
1	0.454	0.0017
2	0.0022	0
k $(t=1)$	$\mathbb{P}\left(N_t^1 \leq k \mid \mathcal{F}_t^{N^1+N^2}\right)$	$\mathbb{P}\left(N_t^2 \leq k \mid \mathcal{F}_t^{N^1+N^2}\right)$
0	0.0007	0
10	0.454	0.0017
20	0.9995	0.6914

Table 5.4: Approximate type estimation when $N^1 + N^2$ is observable

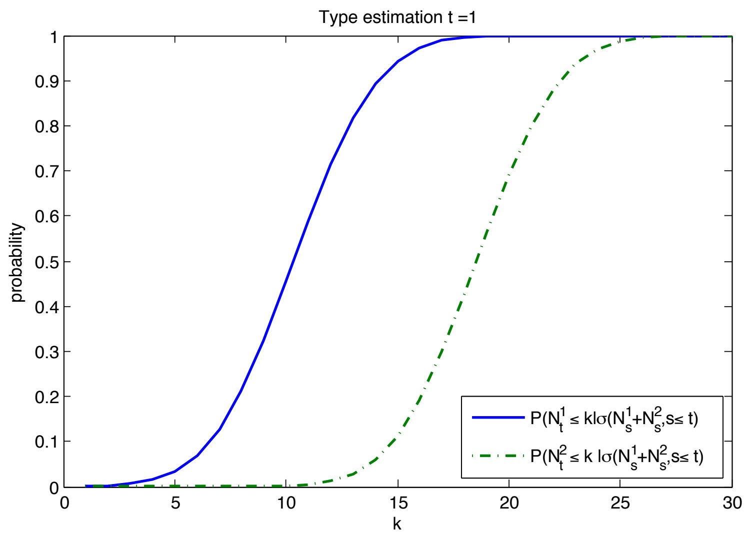


Figure 5.3: Approximate solution of type estimation $\mathbb{P}\left(N_t^i \leq k \mid \mathcal{F}_t^{N^1+N^2}\right)$ at time $t = 1$.

5.5 Proofs

5.5.1 Proofs in Section 5.2

We first show a lemma that will be used in proving the diffusion approximation results.

Lemma 5.5.1. *Denote* $X_t^k := \int_0^t \lambda_s^k ds$ *for* $k = 1, 2$*, then* $var\left(X_t^k\right) \leq t^2 var\left(\lambda_t^k\right)$.

Proof. Note that for any nonnegative functions $f_1(\cdot)$ and $f_2(\cdot)$ on $\mathbb{R}_+$, we have

$$cov\left(\int_0^t f_1(s)d\lambda_s^k, \int_0^t f_2(s)d\lambda_s^k\right) \geq 0$$

as the randomness comes only from jumps in λ^k. Then since

$$X_t^k = \int_0^t \lambda_s^k ds = \lambda_t^k t - \int_0^t s d\lambda_s^k = \int_0^t (t-s)d\lambda_s^k,$$

we have

$$var\left(\lambda_t^k t\right) = var\left(X_t^k\right) + var\left(\int_0^t s d\lambda_s^k\right) + cov\left(\int_0^t (t-s)d\lambda_s^k, \int_0^t s d\lambda_s^k\right).$$

Hence,

$$var\left(X_t^k\right) \leq var\left(\lambda_t^k t\right) = t^2 var\left(\lambda_t^k\right).$$

□

Lemma 5.5.2. *For all $t \geq 0$ and $j, k = 1, 2$, $M_t^n := \frac{J_t^{N^k,j} - \mu_{G^{j,k}} N_t^k}{\sqrt{v_k^{(n)}}}$ satisfies*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T} |M_t^n - M_{t-}^n|^2\right] = 0.$$

Proof. First we have

$$\sup_{t\leq T} |M_t^n - M_{t-}^n|^2 = \frac{\max\left\{\left(Z_1^{j,k} - \mu_{G^{j,k}}\right)^2, \ldots, \left(Z_{N_T^k}^{j,k} - \mu_{G^{j,k}}\right)^2\right\}}{v_k^{(n)}}.$$

Denote $Y_i := \left(Z_i^{j,k} - \mu_{G^{j,k}}\right)^2$ and $\bar{Y}_m := \max\left\{\left(Z_1^{j,k} - \mu_{G^{j,k}}\right)^2, \ldots, \left(Z_m^{j,k} - \mu_{G^{j,k}}\right)^2\right\}$, and note that N_t^k and v_k are both increasing with n, then it is sufficient to show that for any $t \geq 0$,

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}}\right] = 0. \tag{5.41}$$

It is sufficient to show the sequence $\left\{\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}}\right\}_{n\geq 1}$ converges to zero in probability, denoted by $\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}} \xrightarrow{P} 0$, and it is also uniformly integrable.

By definition, $\mathbb{P}\left(\bar{Y}_n \leq \epsilon\right) = \mathbb{P}\left(Y_i \leq \epsilon\right)^n$. Denote F_Y as the c.d.f. of Y_i, since the jump size $Z_i^{j,k}$ has the finite second moment, we have $\mathbb{E}\left[Y_i\right] < \infty$. Hence, we have $\lim_{x\to\infty} x(1-F_Y(x)) = 0$. Then $\frac{\bar{Y}_n}{n} \xrightarrow{P} 0$ as for any $\epsilon > 0$,

$$\mathbb{P}\left(\frac{\bar{Y}_n}{n} > \epsilon\right) = \mathbb{P}\left(\bar{Y}_n > n\epsilon\right) = 1 - \left(1 - \bar{F}_Y(n\epsilon)\right)^n \leq n\bar{F}_Y(n\epsilon) \overset{n\to\infty}{\to} 0.$$

Then we have $\frac{\bar{Y}_{N_t^k}}{N_t^k} \xrightarrow{P} 0$ since

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{Y}_{N_t^k}}{N_t^k} > \epsilon\right) = \lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}\left(\frac{\bar{Y}_{N_t^k}}{N_t^k} > \epsilon | N_t^k\right)\right] = \mathbb{E}\left[\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{Y}_{N_t^k}}{N_t^k} > \epsilon | N_t^k\right)\right] = 0,$$

where the last equality is due to the dominated convergence theorem.

Note that $\{\alpha_k^{(n)}\}_{n\geq 1}$ is bounded uniformly in n due to the linearity:

$$\alpha_k^{(n)} = \frac{m_k^{(n)}}{v_k^{(n)}} = \frac{\mu_{k,1} + \mu_{k,2}\beta^n}{\gamma_{k,1} + \gamma_{k,2}\beta^n} = \frac{\mu_{k,2}}{\gamma_{k,2}} \frac{\frac{\mu_{k,1}}{\mu_{k,2}} + \beta^n}{\frac{\gamma_{k,1}}{\gamma_{k,2}} + \beta^n} \leq \max\left\{\frac{\mu_{k,1}}{\gamma_{k,1}}, \frac{\mu_{k,2}}{\gamma_{k,2}}\right\}.$$

Now we show that $\frac{N_t^k}{v_k^{(n)}} \xrightarrow{P} \alpha_k t$. First, since

$$\begin{aligned}
var(N_t^k) &= var\left(\left(N_t^k - \int_0^t \lambda_s^k ds\right) + \int_0^t \lambda_s^k ds\right) \\
&\leq var\left(N_t^k - \int_0^t \lambda_s^k ds\right) + var\left(\int_0^t \lambda_s^k ds\right) + 2\left|cov\left(N_t^k - \int_0^t \lambda_s^k ds, \int_0^t \lambda_s^k ds\right)\right| \\
&\leq var\left(N_t^k - \int_0^t \lambda_s^k ds\right) + var\left(\int_0^t \lambda_s^k ds\right) + 2\sqrt{var\left(N_t^k - \int_0^t \lambda_s^k ds\right) var\left(\int_0^t \lambda_s^k ds\right)},
\end{aligned}$$

where the last inequality is due to the Cauchy-Schwartz inequality, and by Lemma 5.5.1 and (5.45) that is independent of this lemma, we have

$$\begin{aligned}
var\left(N_t^k - \int_0^t \lambda_s^k ds\right) &= \mathbb{E}\left[\left(N_t^k - \int_0^t \lambda_s^k ds\right)^2\right] = \mathbb{E}\left[\int_0^t \lambda_s^k ds\right] = t m_k^{(n)}, \\
var\left(\int_0^t \lambda_s^k ds\right) &\leq t^2 v_k^{(n)},
\end{aligned}$$

which leads to

$$var(N_t^k) \leq t m_k^{(n)} + t^2 v_k^{(n)} + 2t^{\frac{3}{2}}\sqrt{m_k^{(n)} v_k^{(n)}}.$$

Since $\frac{m_k^{(n)}}{v_k^{(n)}} \to \alpha_k$, therefore for fixed $t > 0$, we can choose n big enough such that $\left|\frac{m_k^{(n)}}{v_k^{(n)}} t - \alpha_k t\right| < \frac{\epsilon}{2}$, then by the Chebyshev inequality,

$$\mathbb{P}\left(\left|\frac{N_t^k}{v_k^{(n)}} - \alpha_k t\right| > \epsilon\right) \leq \mathbb{P}\left(\left|\frac{N_t^k}{v_k^{(n)}} - \frac{m_k^{(n)}}{v_k^{(n)}} t\right| + \left|\frac{m_k^{(n)}}{v_k^{(n)}} t - \alpha_k t\right| > \epsilon\right)$$

$$
\begin{aligned}
&\leq \frac{1}{\left(\epsilon - \left|\frac{m_k^{(n)}}{v_k^{(n)}} - \alpha_k\right| t\right)^2} \frac{var(N_t^k)}{(v_k^{(n)})^2} \\
&\leq \frac{1}{\left(\epsilon - \left|\frac{m_k^{(n)}}{v_k^{(n)}} - \alpha_k\right| t\right)^2} \frac{tm_k^{(n)} + t^2 v_k^{(n)} + 2t^{\frac{3}{2}}\sqrt{m_k^{(n)} v_k^{(n)}}}{(v_k^{(n)})^2} \\
&\leq \frac{1}{v_k^{(n)}} \frac{1}{\left(\epsilon - \left|\alpha_k^{(n)} - \alpha_k\right| t\right)^2} \left(t\alpha_k^{(n)} + t^2 + 2t^{\frac{3}{2}}\sqrt{\alpha_k^{(n)}}\right) \\
&\overset{n\to\infty}{\to} 0,
\end{aligned}
$$

as $v_k^{(n)} \to \infty$ and other terms are nonnegative and bounded uniformly in n from above. Hence, $\frac{N_t^k}{v_k^{(n)}} \xrightarrow{P} \alpha_k t$. Moreover, with $\frac{\bar{Y}_{N_t^k}}{N_t^k} \xrightarrow{P} 0$, we can conclude that

$$
\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}} \xrightarrow{P} 0. \tag{5.42}
$$

Note that

$$
\mathbb{E}[\bar{Y}_n] = \int_0^\infty y d(F_Y(y))^n = n\int_0^\infty y(F_y(y))^{n-1} dF_Y(y) \leq n\mathbb{E}\left[Y\right],
$$

$$
\sup_n \mathbb{E}\left[\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}}\right] = \sup_n \mathbb{E}\left[\mathbb{E}\left[\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}}|N_t^k\right]\right] \leq \mathbb{E}[Y]\sup_n \frac{\mathbb{E}\left[N_t^k\right]}{v_k^{(n)}} = \mathbb{E}[Y]t\left(\sup_n \alpha_k^{(n)}\right) < \infty. \tag{5.43}
$$

Hence the sequence $\left\{\frac{\bar{Y}_{N_t^k}}{v_k^{(n)}}\right\}_{n\geq 1}$ is uniformly integrable for a fixed t. With (5.42) and (5.43), (5.41) holds. $\square$

Proof of Proposition 5.2.5. We prove the proposition by showing the martingale central limit theorem in Theorem 5.2.3.

Note that the process $\Gamma := \left(\lambda^1, \lambda^2, X^1, X^2, N^1, N^2, J^{M^1}, J^{M^2}, J^{N^1,1}, J^{N^1,2}, J^{N^2,1}, J^{N^2,2}, t\right)$ is a Markov process with the generator $\mathcal{A}$, and take any f in the domain $\mathcal{D}(\mathcal{A})$, then at $\Lambda = (\lambda_1, \lambda_2, x_1, x_2, n_1, n_2, j_1^m, j_2^m, j_1^{n,1}, j_1^{n,2}, j_2^{n,1}, j_2^{n,2}, t)$, we have

$$
\begin{aligned}
&\mathcal{A}f\left(\lambda_1, \lambda_2, x_1, x_2, n_1, n_2, j_1^m, j_2^m, j_1^{n,1}, j_1^{n,2}, j_2^{n,1}, j_2^{n,2}, t\right) \\
&= \frac{\partial f}{\partial t} - \delta_1\lambda_1\frac{\partial f}{\partial \lambda_1} - \delta_2\lambda_2\frac{\partial f}{\partial \lambda_2} + \lambda_1\frac{\partial f}{\partial x_1} + \lambda_2\frac{\partial f}{\partial x_2}
\end{aligned}
$$

$$+ \rho_1 \left[\int_{\mathbb{R}_+} f\left(\lambda_1 + y, \lambda_2, x_1, x_2, n_1, n_2, j_1^m + y, j_2^m, j_1^{n,1}, j_1^{n,2}, j_2^{n,1}, j_2^{n,2}, t\right) dH_1(y) - f(\Lambda) \right]$$
$$+ \rho_2 \left[\int_{\mathbb{R}_+} f\left(\lambda_1, \lambda_2 + y, x_1, x_2, n_1, n_2, j_1^m, j_2^m + y, j_1^{n,1}, j_1^{n,2}, j_2^{n,1}, j_2^{n,2}, t\right) dH_2(y) - f(\Lambda) \right]$$
$$+ \lambda_1 \left[\int_{\mathbb{R}_+^2} f\left(\lambda_1 + z_1, \lambda_2 + z_2, x_1, x_2, n_1 + 1, n_2, j_1^m, j_2^m, j_1^{n,1} + z_1, j_1^{n,2} + z_2, j_2^{n,1}, j_2^{n,2}, t\right) dG_1(z_1, z_2) \right.$$
$$\left. - f(\Lambda) \right]$$
$$+ \lambda_2 \left[\int_{\mathbb{R}_+^2} f\left(\lambda_1 + z_1, \lambda_2 + z_2, x_1, x_2, n_1, n_2 + 1, j_1^m, j_2^m, j_1^{n,1}, j_1^{n,2}, j_2^{n,1} + z_1, j_2^{n,2} + z_2, t\right) dG_2(z_1, z_2) \right.$$
$$\left. - f(\Lambda) \right], \tag{5.44}$$

where $G_1(z_1, z_2) = G_{1,1}(z_1)G_{2,1}(z_2)$ and $G_2(z_1, z_2) = G_{1,2}(z_1)G_{2,2}(z_2)$.

First $\begin{bmatrix} \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}} \\ \frac{J_t^{N^k,j} - \mu_{G^{j,k}} N_t^k}{\sqrt{v_k}} \\ \frac{C_t^k - \mu_{G^{c,k}} N_t^k}{\sqrt{v_k}} \\ \frac{J_t^{M^k} - \mu_{H^k} \rho_k^n t}{\sqrt{v_k}} \end{bmatrix}$ is a martingale since from the generator in (5.44), we have

$$\mathcal{A}\left(\frac{n_k - x_k}{\sqrt{v_k}}\right) = 0, \quad \mathcal{A}\left(\frac{j^{n_k,i} - \mu_{G^{i,k}} n_k}{\sqrt{v_k}}\right) = 0,$$
$$\mathcal{A}\left(\frac{j^{m_k} - \mu_{H^k} \rho_k t}{\sqrt{v_k}}\right) = 0, \quad \mathcal{A}\left(\frac{c_k - \mu_{G^{c,k}} n_k}{\sqrt{v_k}}\right) = 0.$$

Then, we check the condition (5.8). Denote $v_{G^{i,k}} := \mu_{2G^{i,k}} - (\mu_{G^{i,k}})^2$ for $i = 1, 2$, we have

$$\begin{aligned} \mathcal{A}\left(\frac{n_k - x_k}{\sqrt{v_k}}\right)^2 &= -2\frac{\lambda_k}{\sqrt{v_k}}\left(\frac{n_k - x_k}{\sqrt{v_k}}\right) + \lambda_k \left[\int_{\mathbb{R}_+^2} \left(\frac{n_k + 1 - x_k}{\sqrt{v_k}}\right)^2 dG_k(z_1, z_2) - \left(\frac{n_k - x_k}{\sqrt{v_k}}\right)^2 \right] \\ &= \frac{\lambda_k}{v_k}, \end{aligned}$$

$$\begin{aligned} \mathcal{A}\left(\frac{j^{n_k,i} - \mu_{G^{i,k}} n_k}{\sqrt{v_k}}\right)^2 &= \lambda_k \left[\int_{\mathbb{R}_+} \left(\frac{j^{n_k,i} + z - \mu_{G^{i,k}}(n_k + 1)}{\sqrt{v_k}}\right)^2 dG_{i,k}(z) - \left(\frac{j^{n_k,i} - \mu_{G^{i,k}} n_k}{\sqrt{v_k}}\right)^2 \right] \\ &= \frac{v_{G^{i,k}}}{v_k}\lambda_k, \end{aligned}$$

$$\mathcal{A}\left(\frac{j^{m_k}-\mu_{H^k}\rho_k t}{\sqrt{v_k}}\right)^2 = -2\frac{\mu_{H^k}\rho_k}{\sqrt{v_k}}\left(\frac{j^{m_k}-\mu_{H^k}\rho_k t}{\sqrt{v_k}}\right) + \rho_k\left[\int_0^\infty\left(\frac{j^{m_k}+y-\mu_{H^k}\rho_k t}{\sqrt{v_k}}\right)^2 dH_k(y) - \left(\frac{j^{m_k}-\mu_{H^k}\rho_k t}{\sqrt{v_k}}\right)^2\right] = \frac{\rho_k\mu_{2H^k}}{v_k},$$

$$\mathcal{A}\left(\frac{c_k-\mu_{G^{c,k}}n_k}{\sqrt{v_k}}\right)^2 = \lambda_k\left[\int_{\mathbb{R}_+}\left(\frac{c_k+z-\mu_{G^{c,k}}(n_k+1)}{\sqrt{v_k}}\right)^2 dG_{c,k}(z) - \left(\frac{c_k-\mu_{G^{c,k}}n_k}{\sqrt{v_k}}\right)^2\right] = \frac{v_{G^{c,k}}}{v_k}\lambda_k.$$

Hence, the following are martingales:

$$\begin{aligned}
&\left(\frac{N_t^k-\int_0^t\lambda_s^k ds}{\sqrt{v_k}}\right)^2 - \frac{1}{v_k}\int_0^t\lambda_s^k ds,\\
&\left(\frac{J_t^{N^k,j}-\mu_{G^{k,j}}N_t^k}{\sqrt{v_k}}\right)^2 - \frac{v_{G^{j,k}}}{v_k}\int_0^t\lambda_s^k ds,\\
&\left(\frac{J_t^{M^k}-\mu_{H^k}\rho_k t}{\sqrt{v_k}}\right)^2 - \frac{\rho_k\mu_{2H^k}}{v_k}t,\\
&\left(\frac{C_t^k-\mu_{G^{c,k}}N_t^k}{\sqrt{v_k}}\right)^2 - \frac{v_{G^{c,k}}}{v_k}\int_0^t\lambda_s^k ds.
\end{aligned} \tag{5.45}$$

Note that $\frac{1}{v_k^{(n)}}\int_0^t\lambda_s^k ds \xrightarrow{P} \alpha_k t$. Indeed, for any fixed t, we can choose n big enough such that $\left|\frac{m_k^{(n)}}{v_k^{(n)}}t-\alpha_k t\right| < \frac{\epsilon}{2}$, then by the Chebyshev inequality, Lemma 5.5.1 and $\lim_{n\to\infty}\frac{m_k^{(n)}}{v_k^{(n)}} = \alpha_k$, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{v_k^{(n)}}\int_0^t\lambda_s^k ds-\alpha_k t\right|>\epsilon\right) &\le \mathbb{P}\left(\left|\frac{1}{v_k^{(n)}}\int_0^t\lambda_s^k ds-\frac{m_k^{(n)}}{v_k^{(n)}}t\right|+\left|\frac{m_k^{(n)}}{v_k^{(n)}}t-\alpha_k t\right|>\epsilon\right)\\
&\le \frac{\frac{1}{\left(v_k^{(n)}\right)^2}var\left(\int_0^t\lambda_s^k ds\right)}{\left(\epsilon-\left|\frac{m_k^{(n)}}{v_k^{(n)}}-\alpha_k\right|t\right)^2}\\
&\le \frac{t^2}{\left(\epsilon-\left|\frac{m_k^{(n)}}{v_k^{(n)}}-\alpha_k\right|t\right)^2}\frac{v_k^{(n)}}{\left(v_k^{(n)}\right)^2}\\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

We have shown in Lemma 5.5.2 that the condition in (5.7) holds for $\frac{J_t^{N^k,j} - \mu_{G^j,k} N_t^k}{\sqrt{v_k}}$, the second component in (5.9). The condition also holds for other component in (5.9) in a similar way.

Now we check the cross terms in (5.8). For $k, k', j, j' = 1, 2$ and $k \neq k'$,

$$\mathcal{A}\left(\frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}}\right)\left(\frac{N_t^{k'} - \int_0^t \lambda_s^{k'} ds}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{J_t^{M^k} - \rho_k \mu_{H^k} t}{\sqrt{v_k}}\right)\left(\frac{J_t^{M^{k'}} - \rho_{k'} \mu_{H^{k'}} t}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{C_t^k - \mu_{\mu_{G^c,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{C_t^{k'} - \mu_{\mu_{G^c,k'}} N_t^{k'}}{\sqrt{v_{k'}}}\right) = 0.$$

For $(k, j) \neq (k', j')$,

$$\mathcal{A}\left(\frac{J_t^{N^k,j} - \mu_{\mu_{G^j,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{J_t^{N^{k'},j'} - \mu_{\mu_{G^{j'},k'}} N_t^{k'}}{\sqrt{v_{k'}}}\right) = 0.$$

Moreover, the cross terms

$$\mathcal{A}\left(\frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}}\right)\left(\frac{J_t^{M^{k'}} - \rho_{k'} \mu_{H^{k'}} t}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{J_t^{N^k,j} - \mu_{\mu_{G^j,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{N_t^{k'} - \int_0^t \lambda_s^{k'} ds}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{C_t^k - \mu_{\mu_{G^c,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{N_t^{k'} - \int_0^t \lambda_s^{k'} ds}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{J_t^{M^k} - \rho_k \mu_{H^k} t}{\sqrt{v_k}}\right)\left(\frac{C_t^{k'} - \mu_{\mu_{G^c,k'}} N_t^{k'}}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{J_t^{N^k,j} - \mu_{\mu_{G^j,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{J_t^{M^{k'}} - \rho_{k'} \mu_{H^{k'}} t}{\sqrt{v_{k'}}}\right) = 0,$$

$$\mathcal{A}\left(\frac{J_t^{N^k,j} - \mu_{\mu_{G^j,k}} N_t^k}{\sqrt{v_k}}\right)\left(\frac{C_t^{k'} - \mu_{\mu_{G^c,k'}} N_t^{k'}}{\sqrt{v_{k'}}}\right) = 0.$$

Therefore, all cross terms in (5.8) converge to zero in probability.

Hence by Lemma 5.2.3, there exists mutually independent standard Brownian motions $B_t^{M,k}, B_t^{N,k}, B_t^{J^k,j}, B_t^{C,k}$, such that as $n \to \infty$,

$$\begin{bmatrix} \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}} \\ \frac{J_t^{N^k,j} - \mu_{G^j,k} N_t^k}{\sqrt{v_k}} \\ \frac{C_t^k - \mu_{G^c,k} N_t^k}{\sqrt{v_k}} \\ \frac{J_t^{M^k} - \mu_{H^k} \rho_k^n t}{\sqrt{v_k}} \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\alpha_k} B_t^{N,k} \\ \sqrt{v_{G^j,k}\alpha_k} B_t^{J^k,j} \\ \sqrt{v_{G^c,k}\alpha_k} B_t^{C,k} \\ \sqrt{\mu_{2H^k}\theta_k} B_t^{M,k} \end{bmatrix}.$$

Moreover,

$$\begin{aligned} \frac{J_t^{N^k,j} - \mu_{G^j,k} \int_0^t \lambda_s^k ds}{\sqrt{v_k}} &= \frac{J_t^{N^k,j} - \mu_{G^j,k} N_t^k}{\sqrt{v_k}} + \mu_{G^j,k} \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}} \\ &\Rightarrow \sqrt{v_{G^j,k}\alpha_k} B_t^{J^k,j} + \mu_{G^j,k} \sqrt{\alpha_k} B_t^{N,k} \end{aligned}$$

Similarly, we have

$$\begin{aligned} \frac{C_t^k - \mu_{G^c,k} \int_0^t \lambda_s^k ds}{\sqrt{v_k}} &= \frac{C_t^k - \mu_{G^c,k} N_t^k}{\sqrt{v_k}} + \mu_{G^c,k} \frac{N_t^k - \int_0^t \lambda_s^k ds}{\sqrt{v_k}} \\ &\Rightarrow \sqrt{v_{G^c,k}\alpha_k} B_t^{C,k} + \mu_{G^c,k} \sqrt{\alpha_k} B_t^{N,k}. \end{aligned}$$

□

Proof of Proposition 5.2.6. In this proof, we show how to obtain the OU system from the building block. Recall the representation of λ_t^k in (5.2.2), and denote by

$$\begin{aligned} U_t^k &:= \lambda_0^k e^{-\delta_k t} + \int_0^t e^{-\delta_k (t-s)} dJ_s^{M^k} = \lambda_0^k e^{-\delta_k t} + J_t^{M^k} - \delta_k \int_0^t e^{-\delta_k (t-s)} J_s^{M^k} ds \\ V_t^{1,k} &:= \int_0^t e^{-\delta_k (t-s)} dJ_s^{N^1,k} = J_t^{N^1,k} - \delta_k \int_0^t e^{-\delta_k (t-s)} J_s^{N^1,k} ds \\ V_t^{2,k} &:= \int_0^t e^{-\delta_k (t-s)} dJ_s^{N^2,k} = J_t^{N^2,k} - \delta_k \int_0^t e^{-\delta_k (t-s)} J_s^{N^2,k} ds, \end{aligned}$$

then λ_t^k can be written as

$$\lambda_t^k = U_t^k + V_t^{1,k} + V_t^{2,k}.$$

Moreover,

$$\frac{U_t^k}{\sqrt{v_k}} = \frac{J_t^{M^k} - \mu_{H^k}\rho_k t}{\sqrt{v_k}} - \delta_k \int_0^t e^{-\delta_k (t-s)} \frac{J_s^{M^k} - \mu_{H^k}\rho_k s}{\sqrt{v_k}} ds + \frac{\lambda_0^k - \frac{\mu_{H^k}\rho_k}{\delta_k}}{\sqrt{v_k}} e^{-\delta_k t} + \frac{\frac{\mu_{H^k}\rho_k}{\delta_k}}{\sqrt{v_k}},$$

$$
\begin{aligned}
\frac{V_t^{1,k}}{\sqrt{v_k}} &= \sqrt{\frac{v_1}{v_k}}\frac{J_t^{N^1,k}-\mu_{G^{k,1}}\int_0^t \lambda_s^1 ds}{\sqrt{v_1}} - \delta_k \int_0^t e^{-\delta_k(t-s)}\sqrt{\frac{v_1}{v_k}}\frac{J_s^{N^1,k}-\mu_{G^{k,1}}\int_0^s \lambda_u^1 du}{\sqrt{v_1}}ds \\
&\quad + \frac{\mu_{G^{k,1}}}{\sqrt{v_k}}\int_0^t e^{-\delta_k(t-s)}\lambda_s^1 ds, \\
\frac{V_t^{2,k}}{\sqrt{v_k}} &= \sqrt{\frac{v_2}{v_k}}\frac{J_t^{N^2,k}-\mu_{G^{k,2}}\int_0^t \lambda_s^2 ds}{\sqrt{v_2}} - \delta_k \int_0^t e^{-\delta_k(t-s)}\sqrt{\frac{v_2}{v_k}}\frac{J_s^{N^2,k}-\mu_{G^{k,2}}\int_0^s \lambda_u^2 du}{\sqrt{v_2}}ds \\
&\quad + \frac{\mu_{G^{k,2}}}{\sqrt{v_k}}\int_0^t e^{-\delta_k(t-s)}\lambda_s^2 ds.
\end{aligned}
$$

Denote

$$
\begin{aligned}
W_t^{k,(n)} &:= \frac{J_t^{M^k}-\mu_{H^k}\rho_k^{(n)}t}{\sqrt{v_k}} + \frac{J_t^{N^1,k}-\mu_{G^{k,1}}\int_0^t \lambda_s^1 ds}{\sqrt{v_k}} + \frac{J_t^{N^2,k}-\mu_{G^{k,2}}\int_0^t \lambda_s^2 ds}{\sqrt{v_k}} \\
&= \frac{J_t^{M^k}-\mu_{H^k}\rho_k^{(n)}t}{\sqrt{v_k}} + \sqrt{\frac{v_1}{v_k}}\frac{J_t^{N^1,k}-\mu_{G^{k,1}}\int_0^t \lambda_s^1 ds}{\sqrt{v_1}} + \sqrt{\frac{v_2}{v_k}}\frac{J_t^{N^2,k}-\mu_{G^{k,2}}\int_0^t \lambda_s^2 ds}{\sqrt{v_2}},
\end{aligned}
\tag{5.46}
$$

then for $k, k' = 1, 2$ and $k' \neq k$, the normalized process of λ^k is

$$
\begin{aligned}
Z_t^{k,(n)} &= \frac{U_t^k}{\sqrt{v_k^{(n)}}} + \frac{V_t^{1,k}}{\sqrt{v_k^{(n)}}} + \frac{V_t^{2,k}}{\sqrt{v_k^{(n)}}} - \frac{m_k^{(n)}}{\sqrt{v_k^{(n)}}} \\
&= W_t^{k,(n)} - \delta_k \int_0^t e^{-\delta_k(t-s)} W_s^{k,(n)} ds + \mu_{G^{k,k}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k,(n)} ds \\
&\quad + \sqrt{\frac{v_{k'}^{(n)}}{v_k^{(n)}}}\mu_{G^{k,k'}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k',(n)} ds + z_t^{k,(n)} \\
&= \int_0^t e^{-\delta_k(t-s)} dW_s^{k,(n)} + \mu_{G^{k,k}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k,(n)} ds \\
&\quad + \sqrt{\frac{v_{k'}^{(n)}}{v_k^{(n)}}}\mu_{G^{k,k'}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k',(n)} ds + z_t^{k,(n)}
\end{aligned}
$$

where

$$
\begin{aligned}
z_t^{k,(n)} &= \frac{m_{k'}^{(n)}}{\sqrt{v_k^{(n)}}}\frac{\mu_{G^{k,k'}}}{\delta_k}(1-e^{-\delta_k t}) + \frac{m_k^{(n)}}{\sqrt{v_k^{(n)}}}\frac{\mu_{G^{k,k}}}{\delta_k}(1-e^{-\delta_k t}) + \frac{\mu_{H^k}}{\delta_k}\frac{\rho_k^n}{\sqrt{v_k^{(n)}}} \\
&\quad + \frac{1}{\sqrt{v_k^{(n)}}}\left(\lambda_0^k - \frac{\mu_{H^k}}{\delta_k}\rho_k\right)e^{-\delta_k t} - \frac{m_k}{\sqrt{v_k^{(n)}}} \\
&= \frac{\lambda_0^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}} - (1-e^{-\delta_k t})\frac{\lambda_0^k - \frac{\mu_{H^k}}{\delta_k}\rho_k^n - \frac{\mu_{G^{k,k}}}{\delta_k}m_k^{(n)} - \frac{\mu_{G^{k,k'}}}{\delta_k}m_{k'}^{(n)}}{\sqrt{v_k^{(n)}}}
\end{aligned}
$$

$$= \quad Z_0^{k,(n)} - (1 - e^{-\delta_k t}) C_0^{k,(n)},$$

with $C_0^{k,(n)} := \frac{\lambda_0^k - \frac{\mu_{H^k}}{\delta_k}\rho_k^n - \frac{\mu_{G^{k,k}}}{\delta_k} m_k^{(n)} - \frac{\mu_{G^{k,k'}}}{\delta_k} m_{k'}^{(n)}}{\sqrt{v_k^{(n)}}}$.

Recall the first stationary moments in Section 4.4

$$m_k^{(n)} = \frac{(\delta_{k'} - \mu_{G^{k',k'}})\mu_{H^k}}{(\delta_k - \mu_{G^{k,k}})(\delta_{k'} - \mu_{G^{k',k'}}) - \mu_{G^{k,k'}}\mu_{G^{k',k}}}\rho_k^n + \frac{\mu_{G^{k,k'}}\mu_{H^k}}{(\delta_k - \mu_{G^{k,k}})(\delta_{k'} - \mu_{G^{k',k'}}) - \mu_{G^{k,k'}}\mu_{G^{k',k}}}\rho_{k'}^n,$$

which yields

$$Z_0^{k,(n)} - C_0^{k,(n)} = \frac{1}{\sqrt{v_k^{(n)}}}\left[-\left(1 - \frac{\mu_{G^{k,k}}}{\delta_k}\right) m_k^{(n)} + \frac{\mu_{G^{k,k'}}}{\delta_k} m_{k'}^{(n)} + \frac{\mu_{H^k}}{\delta_k}\rho_k^n\right] = 0.$$

Hence, we obtain

$$\begin{aligned} Z_t^{k,(n)} \quad &= \quad Z_0^{k,(n)} + \int_0^t e^{-\delta_k(t-s)} dW_s^{k,(n)} + \mu_{G^{k,k}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k,(n)} ds \\ &\quad + \sqrt{\frac{v_{k'}^{(n)}}{v_k^{(n)}}} \mu_{G^{k,k'}} \int_0^t e^{-\delta_k(t-s)} Z_s^{k',(n)} ds - (1 - e^{-\delta_k t}) C_0^{k,(n)}. \end{aligned}$$

Equivalently,

$$\begin{aligned} dZ_t^{k,(n)} &= -\delta_k \left(Z_t^{k,(n)} - (Z_0^{k,(n)} - C_0^{k,(n)})\right) dt + dW_t^{k,(n)} + \mu_{G^{k,k}} Z_t^{k,(n)} dt + \sqrt{v_{k'}^{(n)}/v_k^{(n)}} \mu_{G^{k,k'}} Z_t^{k',(n)} dt \\ &= \left[-(\delta_k - \mu_{G^{k,k}}) Z_t^{k,(n)} + \sqrt{v_{k'}^{(n)}/v_k^{(n)}} \mu_{G^{k,k'}} Z_t^{k',(n)}\right] dt + dW_t^{k,(n)}, \end{aligned}$$

and in matrix form, we obtain

$$d\begin{bmatrix} Z_t^{1,(n)} \\ Z_t^{2,(n)} \end{bmatrix} = \left(\begin{bmatrix} -(\delta_1 - \mu_{G^{1,1}}) & \sqrt{v_2^{(n)}/v_1^{(n)}}\mu_{G^{1,2}} \\ \sqrt{v_1^{(n)}/v_2^{(n)}}\mu_{G^{2,1}} & -(\delta_2 - \mu_{G^{2,2}}) \end{bmatrix} \begin{bmatrix} Z_t^{1,(n)} \\ Z_t^{2,(n)} \end{bmatrix}\right) dt + d\begin{bmatrix} W_t^{1,(n)} \\ W_t^{2,(n)} \end{bmatrix}. \tag{5.47}$$

Define the Brownian motion

$$B_t^{W,k} := \frac{\sqrt{\theta_k}\sqrt{\mu_{2H^k}} B_t^{M,k} + \sqrt{\nu_{1,k}}\sqrt{v_{G^{k,1}}}\sqrt{\alpha_1} B_t^{J^1,k} + \sqrt{\nu_{2,k}}\sqrt{v_{G^{k,2}}}\sqrt{\alpha_2} B_t^{J^2,k}}{\sqrt{\theta_k \mu_{2H^k} + \nu_{1,k} v_{G^{k,1}}\alpha_1 + \nu_{2,k} v_{G^{k,2}}\alpha_2}},$$

then from the construction we know $B_t^{W,k}$ and $B_t^{N,k}$ are independent by Lemma 5.2.5. Moreover, from (5.46) we have for $k = 1, 2$,

$$W_t^k \quad := \quad \lim_{n\to\infty} W_t^{k,(n)}$$

$$
\begin{aligned}
&= \sqrt{\theta_k}\sqrt{\mu_{2H^k}}B_t^{M,k} + \sqrt{\nu_{1,k}}\left(\sqrt{v_{G^{k,1}}}\sqrt{\alpha_1}B_t^{J^1,k} + \mu_{G^{k,1}}\sqrt{\alpha_1}B_t^{N,1}\right) \\
&\quad + \sqrt{\nu_{2,k}}\left(\sqrt{v_{G^{k,2}}}\sqrt{\alpha_2}B_t^{J^2,k} + \mu_{G^{k,2}}\sqrt{\alpha_2}B_t^{N,2}\right) \\
&= \sqrt{\theta_k\mu_{2H^k} + \nu_{1,k}v_{G^{k,1}}\alpha_1 + \nu_{2,k}v_{G^{k,2}}\alpha_2}B_t^{W,k} + \sqrt{\nu_{1,k}\alpha_1}\mu_{G^{k,1}}B_t^{N,1} + \sqrt{\nu_{2,k}\alpha_2}\mu_{G^{k,2}}B_t^{N,2}.
\end{aligned}
$$

Then, with Σ and $\tilde{\Sigma}$ defined in (5.13), we have $\begin{bmatrix} W_t^1 \\ W_t^2 \end{bmatrix} = \Sigma \begin{bmatrix} B_t^{W,1} \\ B_t^{W^2} \end{bmatrix} + \tilde{\Sigma} \begin{bmatrix} B_t^{N,1} \\ B_t^{N,2} \end{bmatrix}$.

Note that $Z_0^{(n)} \Rightarrow Z_0$ in (5.11) by the central limit theorem and therefore $Z^{(n)}$ in (5.47) converges to Z in (5.12) weakly by the continuous mapping theorem. □

Proof of Lemma 5.2.9. First denote p_t, q_t as deterministic $d \times d$ matrices which solve $\frac{dp_t}{dt} = -p_t b$ and $\frac{dq_t}{dt} = bq_t$, with $p_0 = I$ and $q_0 = I$. Then, we have $p_t = e^{-bt}$, and $q_t = e^{bt}$. Moreover denote $Q_t := \int_0^t e^{bs}ds$.

$$
V_T - V_t = r\int_t^T Z_u du + \int_t^T \sigma dB_u^C + \int_t^T \tilde{\sigma} dB_u^N,
$$

then to prove (5.17), it is sufficient to show that

$$
\int_t^T Z_u du = (Q_T - Q_t)\,p_t Z_t + \int_t^T (Q_T - Q_u)\,p_u \Sigma dB_u^W + \int_t^T (Q_T - Q_u)\,p_u \tilde{\Sigma} B_u^N.
$$

Applying Ito's formula, we obtain

$$
d\,(p_t Z_t) = -p_t b Z_t dt + p_t\left(bZ_t dt + \Sigma dB_t^W + \tilde{\Sigma} dB_t^N\right) = p_t \Sigma dB_t^W + p_t \tilde{\Sigma} dB_t^N.
$$

Denote $\hat{\Sigma}\hat{B}_t := \Sigma B_t^W + \tilde{\Sigma} B_t^N$ ($\hat{\Sigma} = chol(\Sigma\Sigma^* + \tilde{\Sigma}\tilde{\Sigma}^*)$) and $Q_t = \int_0^t q_s ds$, then

$$
\begin{aligned}
Z_t &= q_t Z_0 + q_t \int_0^t p_s \hat{\Sigma} d\hat{B}_s, \\
\int_0^t Z_s ds &= Q_t Z_0 + \int_0^t \int_0^s q_s p_u \hat{\Sigma} d\hat{B}_u ds = Q_t Z_0 + \int_0^t \left(\int_u^t q_s ds\right) p_u \hat{\Sigma} d\hat{B}_u \\
&= Q_t p_t Z_t - \int_0^t Q_u p_u \hat{\Sigma} d\hat{B}_u.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\int_t^T Z_s ds &= \int_0^T Z_s ds - \int_0^t Z_s ds \\
&= Q_T p_T Z_T - Q_t p_t Z_t - \left(\int_0^T Q_u p_u \hat{\Sigma} d\hat{B}_u - \int_0^t Q_u p_u \hat{\Sigma} d\hat{B}_u\right)
\end{aligned}
$$

$$
\begin{aligned}
&= Q_T\left(p_t Z_t + \int_t^T p_u \hat{\Sigma} d\hat{B}_u\right) - Q_t p_t Z_t - \int_t^T Q_u p_u \hat{\Sigma} d\hat{B}_u \\
&= \left(Q_T - Q_t\right) p_t Z_t + \int_t^T \left(Q_T - Q_u\right) p_u \Sigma dB_u^W + \int_t^T \left(Q_T - Q_u\right) p_u \tilde{\Sigma} B_u^N,
\end{aligned}
$$

where the last second equality is due to the fact that $Z_T = q_T p_t Z_t + q_T \int_t^T p_s \hat{\Sigma} d\hat{B}_s$.

□

5.5.2 Proofs in Section 5.3

Proof of Corollary 5.3.3. The transformation of the noises is simply due to the decomposition of Brownian motions. We first construct a two-dimensional standard Brownian motion

$$
B_t^Y := \begin{bmatrix} B_t^{Y,1} & B_t^{Y,2} \end{bmatrix}^T = \Sigma^{N,Y} B_t^N,
$$

where $\Sigma^{N,Y} := \begin{bmatrix} \frac{\sqrt{v_1\alpha_1}}{\sqrt{v_1\alpha_1+v_2\alpha_2}} & \frac{\sqrt{v_2\alpha_2}}{\sqrt{v_1\alpha_1+v_2\alpha_2}} \\ \frac{\sqrt{v_2\alpha_2}}{\sqrt{v_1\alpha_1+v_2\alpha_2}} & \frac{-\sqrt{v_1\alpha_1}}{\sqrt{v_1\alpha_1+v_2\alpha_2}} \end{bmatrix}$ and note that $\left(\Sigma^{N,Y}\right)^{-1} = \Sigma^{N,Y}$, we have $B_t^N = \left(\Sigma^{N,Y}\right)^{-1} B_t^Y$.

Then, (5.25) becomes

$$
\begin{aligned}
dX_t &= \mu^X X_t dt + \Sigma^{X,M} dB_t^M + \Sigma^{X,N} dB_t^N \\
&= \mu^X X_t dt + \Sigma^{*X} dB_t^Y + \Sigma^{X,W} dB_t^W \\
dY_t &= \mu^Y X_t dt + \Sigma^{*Y} dB_t^Y.
\end{aligned}
$$

where $\Sigma^{*Y} = \begin{bmatrix} \sqrt{v_1\alpha_1 + v_2\alpha_2} & 0 \end{bmatrix}$, and $\Sigma^{*X} := \Sigma^{X,N}\left(\Sigma^{N,Y}\right)^{-1} = \Sigma^{X,N}\Sigma^{N,Y}$ of dimension 4×2.

Note that

$$
\Sigma^{*X} B_t^Y = \Sigma^{*X}\left(\begin{bmatrix} B_t^{Y,1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ B_t^{Y,2} \end{bmatrix}\right) = \Sigma_1^{*X} B_t^{Y,1} + \Sigma_2^{*X} B_t^{Y,2}
$$

where for $i = 1, 2$, $\Sigma_i^{*X} = (\Sigma_{j,i}^{*X})_{j=1,2,3,4} \in \mathbb{R}^{4\times 1}$ are the i-th column vectors of Σ^{*X}. Hence we have

$$
\begin{aligned}
dX_t &= \mu^X X_t dt + \Sigma_1^{*X} dB_t^{Y,1} + \left(\Sigma_2^{*X} dB_t^{Y,2} + \Sigma^{X,W} dB_t^W\right) \\
dY_t &= \mu^Y X_t dt + \Sigma_1^{*Y} dB_t^{Y,1}
\end{aligned}
$$

We can construct a 3-dimensional standard Brownian motion $B_t^X = \begin{bmatrix} B_t^{Y,2} & B_t^{W,1} & B_t^{W,2} \end{bmatrix}^T$ that is independent of $B^{Y,1}$, and denote $\Sigma^X := \begin{bmatrix} \Sigma_{12}^{*X} & \Sigma_{11} & \Sigma_{12} \\ \Sigma_{22}^{*X} & \Sigma_{21} & \Sigma_{22} \\ \Sigma_{32}^{*X} & 0 & 0 \\ \Sigma_{42}^{*X} & 0 & 0 \end{bmatrix}$, thus we obtain

$$\begin{aligned} dX_t &= \mu^X X_t dt + \Sigma_1^{*X} dB_t^{Y,1} + \Sigma^X dB_t^X \\ dY_t &= \mu^Y X_t dt + \Sigma_1^{*Y} dB_t^{Y,1}. \end{aligned}$$

Hence we have finished the proof. □

Proof of Proposition 5.3.6. Due to the weak convergence of (X^n, Y^n) and the continuous mapping theorem,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}\left[|f(X_t^{k,n}) - q_f^*(Y^{k,n})|^2\right] &= \mathbb{E}\left[|f(X_t^k) - q_f^*(Y^k)|^2\right], \\ \lim_{n\to\infty} \mathbb{E}\left[|f(X_t^{k,n}) - q(Y^{k,n})|^2\right] &= \mathbb{E}\left[|f(X_t^k) - q(Y^k)|^2\right]. \end{aligned}$$

Hence (5.34) holds as the estimate $q_f^*(Y^k) = \mathbb{E}[X_t^k|\mathcal{Y}_t]$ satisfying the least square criteria in the diffusion system. □

Proof of Proposition 5.3.7. The continuous mapping theorem implies

$$(X^n, \hat{X}^n) \Rightarrow (X, \hat{X}), \quad |X_t^n - \hat{X}_t^n|^2 \Rightarrow |X_t - \hat{X}_t|^2.$$

Note that (5.35) is implied by (5.36) as the convergence of the covariance is implied by the convergence of the variance of the marginal random variables. Hence we need to show the convergence of $|X_t^{k,n} - \hat{X}_t^{k,n}|$ in $L^2(\mathbb{P})$. It is sufficient to show the uniform integrability of $|X_t^{k,n} - \hat{X}_t^{k,n}|^2$. Then it suffices to show that for some $\epsilon > 0$,

$$\sup_n \mathbb{E}\left[|X_t^{k,n} - \hat{X}_t^{k,n}|^{2+\epsilon}\right] < \infty. \tag{5.48}$$

Note that

$$\mathbb{E}\left[|X_t^{k,n} - \hat{X}_t^{k,n}|^{2+\epsilon}\right] \leq 2^{1+\epsilon}\left(\mathbb{E}\left[|X_t^{k,n}|^{2+\epsilon}\right] + \mathbb{E}\left[|\hat{X}_t^{k,n}|^{2+\epsilon}\right]\right).$$

Recall

$$X_t^n = \left(\sqrt{v^{(n)}}\right)^{-1}\left(\lambda_t - m^{(n)}\right),$$

and $\hat{X}_t^n$ satisfies $d\hat{X}_t^n = (b-hK_t)\hat{X}_t^n dt + K_t dY_t^n$ where $Y_t^n = \left(\sqrt{v^{(n)}}\right)^{-1}\left(N_t - m^{(n)}t\right)$. Then, we have

$$\hat{X}_t^n = \hat{X}_0^n \exp\left(\int_0^t (b-hK_s)ds\right) + \left(\sqrt{v^{(n)}}\right)^{-1}\int_0^t \exp\left(\int_s^t (b-hK_u)du\right)K_s dN_s.$$

To show (5.48), it is sufficient to take $\epsilon = 1$, and show for $k = 1, 2$,

$$\sup_n \mathbb{E}\left[\left|\frac{N_t^k - m_k^{(n)}t}{\sqrt{v_k^{(n)}}}\right|^3\right] < \infty, \quad \sup_n \mathbb{E}\left[\left|\frac{\lambda_t^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}}\right|^3\right] < \infty. \tag{5.49}$$

We show (5.49) holds using the cluster process representation of the BDCP in Section 3.3.2. Without loss of generality, suppose $\max\{\rho_1^n, \rho_2^n\} = n\rho \to \infty$ as $n \to \infty$, then ρ_1^n, ρ_2^n can be written as

$$\rho_1^n = \tilde{\rho}_1 n, \quad \rho_2^n = \tilde{\rho}_2 n, \tag{5.50}$$

where $\tilde{\rho}_k$ has an upper bound independent of n.

Consider the BDCP and its intensity (N, λ) that with external factor intensity $\rho^n = (\rho_1^n, \rho_2^n)$ in (5.50) for some $n \geq 1$. Being also a cluster process, the BDCP can be constructed as the sum of the n independent copies of the unit process $\left(N_t^{(i)}, \lambda_t^{(i)}\right)$ that is with $\rho^{(i)} = (\tilde{\rho}_1, \tilde{\rho}_2)$ and other parameters being the same as (N, λ) for $i = 1, \dots, n$. By construction, we have

$$N_t \stackrel{d}{=} \sum_{i=1}^n N_t^{(i)}, \quad \lambda_t \stackrel{d}{=} \sum_{i=1}^n \lambda_t^{(i)}.$$

Note that the stationary distribution of N_t and λ_t is equal to $\sum_{i=1}^n N_t^{(i)}$ and $\sum_{i=1}^n \lambda_t^{(i)}$ in distribution where $N_t^{(i)}$ and $\lambda_t^{(i)}$ are of the stationary version. Therefore, denote $m = (m_1, m_2)$ and $v = (v_1, v_2)$ by the stationary mean and variance of $\lambda_t^{(i)}$. As both m and v are affine with $\rho^{(i)}$, the stationary mean and variance $m^{(n)}$ and $v^{(n)}$ of (N, λ) are $m^{(n)} = nm$ and $v^{(n)} = nv$.

Hence we obtain

$$\left|\frac{\lambda_t^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}}\right|^3 = \left|\frac{\sum_{i=1}^n \lambda_t^{k,(i)} - nm_k}{\sqrt{n}\sqrt{v_k}}\right|^3 = \frac{1}{n^{\frac{3}{2}}}\left|\sum_{i=1}^n \frac{\lambda_t^{k,(i)} - m_k}{\sqrt{v_k}}\right|^3 \leq \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^n \left|\frac{\lambda_t^{k,(i)} - m_k}{\sqrt{v_k}}\right|^3,$$

$$\left|\frac{N_t^k - m_k^{(n)}t}{\sqrt{v_k^{(n)}}}\right|^3 = \left|\frac{\sum_{i=1}^n N_t^{k,(i)} - nm_k t}{\sqrt{n}\sqrt{v_k}}\right|^3 = \frac{1}{n^{\frac{3}{2}}}\left|\sum_{i=1}^n \frac{N_t^{k,(i)} - m_k t}{\sqrt{v_k}}\right|^3 \leq \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^n \left|\frac{N_t^{k,(i)} - m_k t}{\sqrt{v_k}}\right|^3.$$

We now focus on $\left|\frac{N_t^{k,(i)}-m_k t}{\sqrt{v_k}}\right|^3$ and $\left|\frac{\lambda_t^{k,(i)}-m_k}{\sqrt{v_k}}\right|^3$ that are independent of the index n. From Example 3.4.2 in Section 3.4, if we assume $\mu_{3H^k} < \infty$ and $\mu_{3G^{k,k'}} < \infty$, we obtain

$$\mathbb{E}\left[\left(N_t^{k,(i)}\right)^3\right] < \infty, \quad \mathbb{E}\left[\left(\lambda_t^{k,(i)}\right)^3\right] < \infty,$$

indicating

$$\mathbb{E}\left[\left|\frac{N_t^{k,(i)} - m_k t}{\sqrt{v_k}}\right|^3\right] < \infty, \quad \mathbb{E}\left[\left|\frac{\lambda_t^{k,(i)} - m_k}{\sqrt{v_k}}\right|^3\right] < \infty.$$

Therefore we obtain (5.49) as

$$\begin{aligned}
\sup_n \mathbb{E}\left[\left|\frac{\lambda_t^k - m_k^{(n)}}{\sqrt{v_k^{(n)}}}\right|^3\right] &\leq \sup_n \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^n \mathbb{E}\left[\left|\frac{\lambda_t^{k,(i)} - m_k}{\sqrt{v_k}}\right|^3\right] = \sup_n \frac{1}{\sqrt{n}} \mathbb{E}\left[\left|\frac{\lambda_t^{k,(i)} - m_k}{\sqrt{v_k}}\right|^3\right] < \infty, \\
\sup_n \mathbb{E}\left[\left|\frac{N_t^k - m_k^{(n)} t}{\sqrt{v_k^{(n)}}}\right|^3\right] &\leq \sup_n \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^n \mathbb{E}\left[\left|\frac{N_t^{k,(i)} - m_k t}{\sqrt{v_k}}\right|^3\right] = \sup_n \frac{1}{\sqrt{n}} \mathbb{E}\left[\left|\frac{N_t^{k,(i)} - m_k t}{\sqrt{v_k}}\right|^3\right] < \infty,
\end{aligned}$$

and hence we proved (5.49). □

5.6 Conclusion

In this chapter, using the martingale central limit theorem, we obtain the diffusion approximation of a stationary BDCP which models a high frequency events system with contagion effect under the influence of a big external factor. We then propose an alternative solution of the filtering problems for the BDCP system by constructing estimates based on the structure from the diffusion system with the actual BDCP input. The asymptotic analysis and numerical example show a good performance of the proposed filter. We apply the result in solving some insurance problems.

Chapter 6

Diffusion Approximation of UDCP and Alternative CIR Simulation Scheme

In this chapter, we show that a sequence of scaled intensity processes of univariate dynamic contagion processes (UDCPs) converges to a CIR process weakly in the path space. This result provides a link between two families of self-exciting processes in non-diffusive and diffusion world. With the weak convergence result and the exact simulation of the approximating sequence, we develop an alternative approximation simulation scheme for CIR processes and the Heston model.

In Section 6.1, we review briefly the UDCP, the CIR process, and the Heston model in the literature. In Section 6.2, we approximate the CIR process X with the pure jump process X^n constructed from the intensity of the UDCP based on the weak convergence $X^n \Rightarrow X$ in the path space $\mathcal{D}_{\mathbb{R}}[0,\infty)$. Moreover, the weak convergence of the asset price processes in the Heston model is also investigated. In Section 6.3 the simulation algorithm for UDCP and CIR is presented. The performance of the CIR simulation is assessed by numerical examples in Section 6.4.

6.1 Introduction

6.1.1 Univariate Dynamic Contagion Processes

The univariate dynamic contagion process (UDCP) is introduced in (3.6) in Chapter 3. It is a self-exciting counting process $N = (N_t)_{t\geq 0}$ with intensity $\lambda = (\lambda_t)_{t\geq 0}$ consisting of an external Poisson process and a self-exciting jump process. In this chapter, we assume additionally all jump sizes are constant in the UDCP intensity. Hence the intensity in (3.6) in integral form becomes:

$$\lambda_t = \lambda_0 e^{-\delta t} + \alpha_M \int_0^t e^{-\delta(t-s)} dM_s + \alpha_N \int_0^t e^{-\delta(t-s)} dN_s, \tag{6.1}$$

where

- λ_0 is the initial value, either a nonnegative constant or random variable;
- $\delta > 0$ is the decay rate;
- M is a Poisson process with a constant intensity $\rho > 0$;
- $\alpha_M > 0$ and $\alpha_N > 0$ are constant jump sizes.

We see that the UDCP intensity λ in (6.1) is a piecewise deterministic jump process.

6.1.2 Basics of CIR Processes

CIR processes: A *Cox-Ingersoll-Ross (CIR) process* $X = (X_t)_{t\geq 0}$ is the unique solution of the SDE below:

$$dX_t = \beta(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t^X, \tag{6.2}$$

where $X_0 \geq 0$, $\beta > 0$, $\theta > 0$ and $\sigma > 0$ are constants.

The Feller condition Let $X_0 > 0$ and W^X be a Brownian motion. If $2\beta\theta \geq \sigma^2$, the process X never reaches zero. If $2\beta\theta < \sigma^2$, zero is accessible for the process X and strongly reflecting. We call the condition $2\beta\theta \geq \sigma^2$ the *Feller condition.*

The transition probability For $T > t$, $X_T|X_t$ follows a non-central chi-square distribution.

$$\mathbb{P}(X_T \leq x|X_t) = F\left(\frac{x \cdot n(t,T)}{e^{-\beta(T-t)}}; d, X_t \cdot n(t,T)\right)$$

where $d = \frac{4\beta\theta}{\sigma^2}$, $n(t,T) = \frac{4\beta e^{-\beta(T-t)}}{\sigma^2\left(1-e^{-\beta(T-t)}\right)}$ and $F(z;\nu,\lambda)$ is the cumulative distribution function of the non-central chi-square distribution with ν degrees of freedom and non-centrality parameter λ:

$$F(z;\nu,\lambda) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j}{j!\, 2^{\frac{\nu}{2}+j}\Gamma(\frac{\nu}{2}+j)} \int_0^z y^{\frac{\nu}{2}+j-1} e^{-\frac{y}{2}} dy.$$

Therefore we can find the explicit form of the mean and variance of $X_T|X_t$.

Applications As the CIR processes are affine processes we can compute the characteristic function explicitly. Moreover, due to its non-negativity, CIR processes are widely used in modelling interest rates and stochastic volatilities of asset prices (the Heston model) in finance.

6.1.3 Basics of the Heston Model

Denote by $X = (X_t)_{t\geq 0}$ and $S = (S_t)_{t\geq 0}$ the variance and asset price processes respectively. We follow the same notation for the Heston model as in Andersen [4],

$$\begin{aligned} dX_t &= \beta(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t^X, \\ dS_t &= S_t\sqrt{X_t}dW_t^S, \\ d\langle W^X, W^S\rangle_t &= \rho dt. \end{aligned} \tag{6.3}$$

The Cholesky decomposition yields $W_t^S = \rho W_t^X + \sqrt{1-\rho^2}W_t^{X,\perp}$ for all $t \geq 0$, where $W^{X,\perp}$ is a Brownian motion independent of W^X as well as X.

Denote the log price process $Y = (Y_t)_{t\geq 0}$ with $Y_t := \log S_t$. By Itô's Lemma we have

$$\begin{aligned} dY_t &= -\frac{1}{2}X_t dt + \sqrt{X_t}dW_t^S \\ &= -\frac{1}{2}X_t dt + \rho\sqrt{X_t}dW_t^X + \sqrt{1-\rho^2}\sqrt{X_t}dW_t^{X,\perp}. \end{aligned}$$

As (X,Y) is a Markov process, for any $t \geq 0$, condition on (X_t, Y_t), we have

$$Y_{t+\Delta} = Y_t - \frac{1}{2}\int_t^{t+\Delta} X_s ds + \rho\int_t^{t+\Delta} \sqrt{X_s}dW_s^X + \sqrt{1-\rho^2}\int_t^{t+\Delta} \sqrt{X_s}dW_t^{X,\perp}. \tag{6.4}$$

We have from (6.3)

$$\sqrt{X_t}dW_t^X = \frac{1}{\sigma}\left(dX_t - \beta(\theta - X_t)dt\right),$$

then we obtain

$$\int_t^{t+\Delta} \sqrt{X_s}dW_s^X = \frac{1}{\sigma}(X_{t+\Delta} - X_t) - \frac{\beta\theta}{\sigma}\Delta + \frac{\beta}{\sigma}\int_t^{t+\Delta} X_s ds.$$

By the Dubin-Schwartz theorem and the independence of $W^{X,\perp}$ with X,

$$\int_t^{t+\Delta} \sqrt{X_s}dW_s^{X,\perp} \stackrel{d}{=} W^{X,\perp}_{\int_t^{t+\Delta} X_s ds},$$

therefore we have

$$Y_{t+\Delta} \stackrel{d}{=} \left(Y_t - \frac{\rho\beta\theta}{\sigma}\Delta\right) + \frac{\rho}{\sigma}(X_{t+\Delta} - X_t) + \left(\frac{\rho\beta}{\sigma} - \frac{1}{2}\right)\int_t^{t+\Delta} X_s ds + \sqrt{1-\rho^2}W^{X,\perp}_{\int_t^{t+\Delta} X_s ds}. \tag{6.5}$$

Hence, for each $t \geq 0$ conditioned on $\sigma\{X_s, Y_s, s \leq t\}$, Y_t is a Gaussian process with mean μ_t and variance v_t that are functions of (X_t, Y_t), $X_{t+\Delta}$ and $\int_t^{t+\Delta} X_s ds$:

$$\begin{aligned}
& Y_t\Big|_{\sigma\{X_s, s\leq t\}} \sim \mathcal{N}\left(\mu_t, v_t\right), \\
& \mu_t = \left(Y_t - \frac{\rho\beta\theta}{\sigma}\Delta\right) + \frac{\rho}{\sigma}(X_{t+\Delta} - X_t) + \left(\frac{\rho\beta}{\sigma} - \frac{1}{2}\right)\int_t^{t+\Delta} X_s ds, \qquad (6.6) \\
& v_t = (1-\rho^2)\int_t^{t+\Delta} X_s ds.
\end{aligned}$$

In particular, if we take $t = 0$ and $\Delta = t$, we have

$$Y_t \stackrel{d}{=} \left(Y_0 - \frac{\rho}{\sigma}X_0 - \frac{\rho\beta\theta}{\sigma}t\right) + \frac{\rho}{\sigma}X_t + \left(\frac{\rho\beta}{\sigma} - \frac{1}{2}\right)\int_0^t X_s ds + \sqrt{1-\rho^2}W^{X,\perp}_{\int_0^t X_s ds}. \tag{6.7}$$

The distribution of the asset price at each time t in the Heston model is completed determined by X_t and $\int_0^t X_s ds$.

6.1.4 Simulation Issues

When we aim to generate a sample path of a process X by the Monte-Carlo method, the essential question is how to generate $X_{t+\Delta}$ given X_t with $t \geq 0$ and $\Delta > 0$. In the following, we denote by $\hat{X}$ the simulated value of X at time t.

Simulation of CIR Processes There are a few approaches to simulating CIR processes in the literature.

First consider the *Euler scheme* where $\hat{X}_{t+\Delta}$ is simulated based on the SDE of the CIR process and its previous state $\hat{X}_t$. However, one needs to make adjustments of the simulated $\hat{X}_t$ when $\hat{X}_t$ is negative. Discussed in Andersen [4]. The adjustments would generate non-trivial bias. If the Feller condition is violated, the bias can be large.

The *exact simulation scheme* based on the explicit form of the transition probability of the CIR process (see Section 6.1.2) becomes an alternative. However, in practice the algorithm involves complications in acceptance-rejection sampling and thus it is inefficient and sometimes even yields poor performance.

In the *moment matching scheme*, one approximates the transition probability with another probability distribution with the same lower order moments. For example, *the lognormal distribution approximation scheme* is adopted by Andersen and Brotherton-Ratcliffe [5], *the Truncated Gaussian scheme (TG)* and the *Quadratic Exponential scheme* (QE) are introduced in Andersen [4]. QE scheme is the most popular and accurate scheme among them where the approximating distribution is a combination of a squared Gaussian and an exponential distribution depending on the value of the $\hat{X}_t$ generated from the last step. However, the distribution of the simulated $\hat{X}_t$ for fixed $t \geq 0$ and the sample path $\hat{X} = (\hat{X}_t)_{t\geq 0}$ is not the exact one. Moreover, the weak convergence remains unverified.

Simulation of the Heston Model As shown in the previous section, the asset price process is driven by the CIR process $X = (X_t)_{t\geq 0}$ and its integrated process $\left(\int_0^t X_s ds\right)_{t\geq 0}$. First note that the transition probability of the integrated process has no explicit form. Moreover, if using the QE method in [4], then the simulation of the integrate $\int_0^t \hat{X}_s ds$ based on the finite points of the path $(\hat{X}_s)_{0\leq s\leq t}$ involves additional numerical errors. Indeed, on each subinterval $[u, u+\Delta]$, the integration is again computed by approximation: $\int_u^{u+\Delta} X_s ds \approx \gamma_1 \hat{X}_u + \gamma_2 \hat{X}_{u+\Delta}$ for some constants γ_1 and γ_2.

In summary, Andersen's QE simulation scheme suffers from two types of errors:

(i). The simulated variance process $\hat{X}_t$ deviates from the original X_t and there is no proof of the convergence in the path space.

(ii). Then is numerical integration error in computing $\int_0^t \hat{X}_s ds$.

In this chapter, we address these issues by proposing an alternative simulation scheme which overcomes the weakness above in the following way:

(i'). We approximate the variance process X by a process X^n which converges weakly to X in the path space and X^n can be simulated exactly.

(ii'). We show that there is no additional error from the implementation of the integrated process $\int_0^t X_s^n ds$ as $\int_0^t X_s^n ds$ can also be simulated exactly based on the path of X_t^n .

6.2 Diffusion Approximation

We first introduce an important diffusion approximation theorem showing the weak convergence of Markov processes to a general diffusion process, which is an extension of Theorem 5.2.3 where the limit process is a Gaussian process. The convergence is based on the characterization of Markov processes by the associated martingale problems.

Theorem 6.2.1 (Diffusion Approximation [33, Theorem 4.1, Chapter 7]). *Let $a = (a_{ij})$ be a continuous, symmetric, nonnegative definite, $d \times d$ matrix-valued function on $\mathbb{R}^d$ and let $b : \mathbb{R}^d \to \mathbb{R}^d$ be continuous. Let*

$$\mathcal{A} = \left\{ (f, Gf \equiv \frac{1}{2} \sum a_{ij} \partial_i \partial_j f + \sum b_i \partial_i f) : f \in C_c^\infty(\mathbb{R}^d) \right\},$$

and suppose that the $C_{\mathbb{R}^d}[0,\infty)$ martingale problem of $\mathcal{A}$ is well posed.

For $n = 1, 2, \ldots$, let X_n and B_n be processes with sample paths in $D_{\mathbb{R}^d}[0,\infty)$, and let $A_n = ((A_{ij}^n))$ be a symmetric $d \times d$ matrix-valued process such that A_n^{ij} has sample paths in $D_{\mathbb{R}}[0,\infty)$ and $A_n(t) - A_n(s)$ is nonnegative definite for $t > s \geq 0$. Set $\mathcal{F}_t^n = \sigma\{X_n(s), B_n(s), A_n(s) : s \leq t\}$.

Suppose that

$$\begin{aligned} &M_n := X_n - B_n \\ &M_n^i M_n^j - A_n^{ij} \end{aligned} \tag{6.8}$$

are $\mathbb{F}^n$-local martingales. Let $\tau_n^r := \inf\{t : |X_n(t)| \geq r$ or $|X_n(t-)| \geq r\}$, and for each $r > 0$, $T > 0$ and $i, j = 1, \ldots, d$,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T\wedge\tau_n^r} |X_n(t) - X_n(t-)|^2\right] = 0, \tag{6.9a}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T\wedge\tau_n^r} |B_n(t) - B_n(t-)|^2\right] = 0, \tag{6.9b}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T\wedge\tau_n^r} |A_n(t) - A_n(t-)|^2\right] = 0. \tag{6.9c}$$

Moreover, suppose that

$$\sup_{t\leq T\wedge\tau_n^r} \left|B_n^i(t) - \int_0^t b_i(X_n(s))ds\right| \xrightarrow{\mathbb{P}} 0, \tag{6.10a}$$

$$\sup_{t\leq T\wedge\tau_n^r} \left|A_n^{ij}(t) - \int_0^t a_{ij}(X_n(s))ds\right| \xrightarrow{\mathbb{P}} 0, \tag{6.10b}$$

and

$$\mathbb{P}X_n(0)^{-1} \Rightarrow \nu \in \mathcal{P}(\mathbb{R}^d). \tag{6.11}$$

Then, $\{X_n\}_n$ converges in distribution to the solution of the martingale problem for $(\mathcal{A}, \nu)$.

Remark 6.2.2. *Note that the weak convergence to the solution of the martingale problem $(\mathcal{A}, \nu)$ is equivalent to the weak convergence to the weak solution of the SDE $dX_t = b(t, X_t)X_t dt + a(t, X_t)dW_t$ with $X_0 \stackrel{d}{=} \nu$. Moreover, (6.8) characterizes the solution of the martingale problem by Itô's lemma.*

6.2.1 Scaled UDCP Intensity Processes

Consider a sequence of intensity processes $(\lambda^n)_{n\geq 1}$ of the UDCP $(N_t^n)_{n\geq 1}$ in (6.1) by taking α_n positive and constant, and

- the decay rate $\delta_n := \beta + \sigma^2\alpha_n$;
- the Poisson process M^n has the intensity $\rho_n := \beta\theta\alpha_n$;
- the jump size $\alpha_M = \sigma^2\alpha_n$ and $\alpha_N = \sigma^2\alpha_n$.

Then,

$$\lambda_t^n = \lambda_0^n e^{-\delta_n t} + \sigma^2 \alpha_n \int_0^t e^{-\delta_n (t-s)} dM_s^n + \int_0^t e^{-\delta_n (t-s)} dN_s^n, \tag{6.12}$$

We construct a sequence of processes $(X^n)_{n\geq 1}$ that are UDCP intensity processes in (6.12) scaled by a constant factor $\frac{1}{\sigma^2 \alpha_n^2}$, i.e.

$$X_t^n := \frac{1}{\sigma^2 \alpha_n^2} \lambda_t^n. \tag{6.13}$$

We call $(X^n)_{n\geq 1}$ the *scaled UDCP intensity processes.*

Remark 6.2.3. *Note that X^n is a process derived from the UDCP system (N^n, λ^n).*

Therefore X^n depends on parameters $(\beta, \theta, \sigma, \alpha_n)$ and it has the representation:

$$\begin{aligned} X_t^n &= X_0^n e^{-\delta_n t} + \frac{1}{\alpha_n} \int_0^t e^{-\delta_n (t-s)} dM_s^n + \frac{1}{\alpha_n} \int_0^t e^{-\delta_n (t-s)} dN_s^n, \\ X_0^n &= \frac{1}{\sigma^2 \alpha_n^2} \lambda_0^n. \end{aligned} \tag{6.14}$$

Due to the exponential decay, X^n is a Markov process with the generator $\mathcal{A}_n$. With any f in its domain, we have

$$\begin{aligned} \mathcal{A}_n f(x) &= -\delta_n x \frac{\partial f}{\partial x} + \rho_n \left[f\left(x + \frac{1}{\alpha_n}\right) - f(x) \right] + \alpha_n^2 \sigma^2 x \left[f\left(x + \frac{1}{\alpha_n}\right) - f(x) \right] \\ &= -\delta_n x \frac{\partial f}{\partial x} + \left(\rho_n + \alpha_n^2 \sigma^2 x\right) \left[f\left(x + \frac{1}{\alpha_n}\right) - f(x) \right]. \end{aligned} \tag{6.15}$$

6.2.2 Weak Convergence to CIR processes

Theorem 6.2.4 (Diffusion Approximation to CIR Processes)**.** *As $\alpha_n \to \infty$, the scaled intensity processes $(X^n)_{n\geq 1}$ in (6.14) converge weakly to the CIR process X in (6.3) in the path space $D_{\mathbb{R}}[0, \infty)$.*

Proof. Take $f(x) = x$ in (6.15), we have

$$\mathcal{A}_n f(x) = -\delta_n x + \left(\rho_n + \alpha_n^2 \sigma^2 x\right) \frac{1}{\alpha_n} = (-\delta_n + \sigma^2 \alpha_n) x + \frac{\rho_n}{\alpha_n} = -\beta x + \beta \theta.$$

Denote $b(x) = -\beta x + \beta\theta$ and $B_t^n := \int_0^t b(X_s^n) ds = \int_0^t \left(-\beta X_s^n + \beta\theta\right) ds$, then

$$M_t^n := X_t^n - B_t^n \quad \text{is a martingale.}$$

Note that (X_t^n, B_t^n) is a Markov process and the generator of (t, X_t^n, B_t^n) is

$$\begin{aligned}\mathcal{A}_b f(t,x,\tilde{x}) &= \frac{\partial f}{\partial t} + (-\beta x + \beta\theta)\frac{\partial f}{\partial \tilde{x}} - \delta_n x \frac{\partial f}{\partial x} + \rho_n \left[f\left(t, x + \frac{1}{\alpha_n}, \tilde{x}\right) - f(t,x,\tilde{x})\right] \\ &\quad + \sigma^2 \alpha_n^2 x \left[f\left(t, x + \frac{1}{\alpha_n}, \tilde{x}\right) - f(t,x,\tilde{x})\right].\end{aligned}$$

Take $f(t,x,\tilde{x}) = (x-\tilde{x})^2$, then

$$\begin{aligned}\mathcal{A}_b f(t,x,\tilde{x}) &= -2(-\beta x + \beta\theta)(x-\tilde{x}) - 2\delta_n x(x-\tilde{x}) + (\rho_n + \sigma^2\alpha_n^2 x)\left[\frac{2}{\alpha_n}(x-\tilde{x}) + \frac{1}{\alpha_n^2}\right] \\ &= \sigma^2 x + \frac{\beta\theta}{\alpha_n}\end{aligned}$$

Denote $A_t^n := \int_0^t \left(\sigma^2 X_s^n + \frac{\beta\theta}{\alpha_n}\right) ds$, then

$$(M_t^n)^2 - A_t^n \quad \text{is a martingale.}$$

= Now we show (6.9a), (6.9b) and (6.9c) hold. Note that by definition B^n and A^n are continuous processes. Moreover, since $\left|X_t^n - X_{t-}^n\right|^2 \leq \left(\frac{1}{\alpha_n} + \frac{1}{\alpha_n}\right)^2 \leq \frac{4}{\alpha_n^2}$, for any $T > 0$,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T}\left|X_t^n - X_{t-}^n\right|^2\right] \leq \lim_{n\to\infty}\frac{4}{\alpha_n^2} = 0.$$

Then, it is sufficient to verify (6.10a) and (6.10b). Define $a(x) := \sigma^2 x$, then as $\alpha_n \to \infty$ we have pathwise

$$\begin{aligned}\sup_{t\leq T}\left|B_t^n - \int_0^t b(X_s^n)ds\right| &= 0, \\ \sup_{t\leq T}\left|A_t^n - \int_0^t a(X_s^n)ds\right| &\leq \frac{\beta\theta}{\alpha_n}T \to 0.\end{aligned}$$

By Theorem 6.2.1, we have $X^n \Rightarrow X$ where X is the solution of the SDE

$$dX_t = \beta(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t.$$

□

Remark 6.2.5. *The theorem above builds the link between the self-exciting point processes and the CIR processes.*

Graphic Illustration (Sample Paths) We show a sample path of the scaled UDCP intensity $X^n = (X_t^n)_{t\geq 0}$ in Figure 6.1 when taking the parameter set $(\beta, \theta, \sigma, \alpha_n) = (1, 2, 0.5, 10)$ in (6.14).

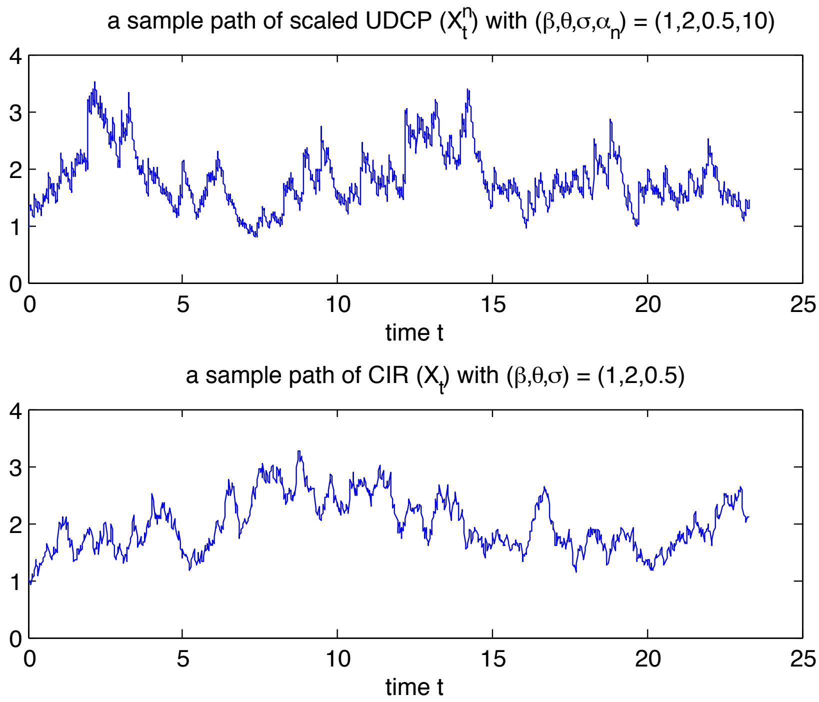


Figure 6.1: A Sample path of a scaled UDCP intensity and a CIR process (with different random seeds).

6.2.3 From Variance Processes to Asset Price Processes

The variance process is modelled by a CIR process X with the weak convergence $X^n \Rightarrow X$ in $D_{\mathbb{R}[0,\infty)}$ in the Heston model . Construct a sequence of log price processes (Y^n) by replacing X with X^n in (6.4), then we have

$$Y_t^n = Y_0^n - \frac{1}{2}\int_0^t X_s^n ds + \rho \int_0^t \sqrt{X_s^n} dW_s^X + \sqrt{1-\rho^2}\int_0^t \sqrt{X_s^n} dW_t^{X,\perp}. \tag{6.16}$$

By Theorem 2.2 in Kurtz and Protter [47] about the weak convergence of the stochastic integral, we have the following convergence result,

$$\begin{aligned} \int_0^\cdot X_s^n ds &\Rightarrow \int_0^\cdot X_s ds, \\ \int_0^\cdot \sqrt{X_s^n} dW_s^X &\Rightarrow \int_0^\cdot \sqrt{X_s} dW_s^X, \\ \int_0^\cdot \sqrt{X_s^n} dW_t^{X,\perp} &\Rightarrow \int_0^\cdot \sqrt{X_s} dW_t^{X,\perp}. \end{aligned} \tag{6.17}$$

Therefore, we have $Y^n \Rightarrow Y$ in (6.4).

Lemma 6.2.6 (Integrated Variance Process)**.**

$$\int_0^t X_s^n ds \quad = \quad \frac{1}{\delta_n}\left(X_0^n + \frac{1}{\alpha_n}M_t^n + \frac{1}{\alpha_n}N_t^n\right) - \frac{1}{\delta_n}X_t^n. \tag{6.18}$$

Remark 6.2.7. *The lemma shows that the simulation of the integrated variance process can avoid numerical integration errors (ii) if all jump processes involved can be simulated exactly.*

Proof. Denote $J_t^{M,n} := \frac{1}{\alpha_n}\int_0^t e^{-\delta_n(t-s)}dM_s^n$ and $J_t^{N,n} := \frac{1}{\alpha_n}\int_0^t e^{-\delta_n(t-s)}dN_s^n$, then

$$X_t^n = X_0^n e^{-\delta_n t} + J_t^{M,n} + J_t^{N,n},$$

and

$$\int_0^t X_s^n ds = \int_0^t X_0^n e^{-\delta_n s}ds + \frac{1}{\alpha_n}\int_{s=0}^t \int_{u=0}^s e^{-\delta_n(s-u)}dM_u^n ds + \frac{1}{\alpha_n}\int_{s=0}^t \int_{u=0}^s e^{-\delta_n(s-u)}dN_u^n ds.$$

Note that

$$\begin{aligned}
\int_{s=0}^t \int_{u=0}^s e^{-\delta_n(s-u)}dM_u^n ds \quad &= \quad \int_{u=0}^t \left(\int_{s=u}^t e^{-\delta_n s}ds\right) e^{\delta_n u}dM_u^n \\
&= \quad \frac{1}{\delta_n}M_t^n - \frac{1}{\delta_n}\int_0^t e^{-\delta_n(t-u)}dM_u^n \\
&= \quad \frac{1}{\delta_n}M_t^n - \frac{\alpha_n}{\delta_n}J_t^{M,n},
\end{aligned}$$

and similarly, we have

$$\int_{s=0}^t \int_{u=0}^s e^{-\delta_n(s-u)}dN_u^n ds \quad = \quad \frac{1}{\delta_n}N_t^n - \frac{\alpha_n}{\delta_n}J_t^{N,n}.$$

Hence, we obtain

$$\begin{aligned}
\int_0^t X_s^n ds \quad &= \quad X_0^n \frac{1}{\delta_n}\left(1 - e^{-\delta_n t}\right) + \frac{1}{\alpha_n}\left(\frac{1}{\delta_n}M_t^n - \frac{\alpha_n}{\delta_n}J_t^{M,n}\right) + \frac{1}{\alpha_n}\left(\frac{1}{\delta_n}N_t^n - \frac{\alpha_n}{\delta_n}J_t^{N,n}\right) \\
&= \quad \frac{1}{\delta_n}\left(X_0^n + \frac{1}{\alpha_n}M_t^n + \frac{1}{\alpha_n}N_t^n\right) - \frac{1}{\delta_n}X_t^n.
\end{aligned}$$

□

6.3 Approximation Simulation Scheme for CIR Processes

From the diffusion approximation results in Theorem 6.2.4, we know that the marked point processes X^n converges to the CIR process X when $n \to \infty$. By construction, X_t^n is simply scaled from the UDCP intensity process λ_t^n. In the following we show that we can simulate λ_t^n. Then we propose a new simulation scheme for the CIR process X based on the simulation of the pure jump process X_t^n when α_n is large.

In the following, we describe the exact simulation algorithm of the approximating process X^n of the CIR process X. It is a special case of the exact simulation algorithm of BDCP in Section 3.5 in Chapter 3.

Exact Simulation of X_t^n The simulation of X^n for any $n \geq 1$ consists of the following steps:

1. Start from $T_0 = 0$, $\lambda^n_{T_0^{\pm}} = \lambda_0^n > 0$, $N_0 = 0$.
2. For $i \geq 0$, simulate the $(i+1)$-th inter-arrival time W_{i+1} in the intensity process.
$$W_{i+1} = \min\{E_{i+1}, I_{i+1}\}.$$
$$T_{i+1} = T_i + W_{i+1}.$$
Generate mutually random variables U_{i+1}^E and U_{i+1}^I uniformly distributed on $[0,1]$. E_{i+1} is the waiting time from T_i to the first jump of external process M^n, hence
$$E_{i+1} \stackrel{d}{=} -\frac{1}{\rho_n} \ln U_{i+1}^E.$$
I_{i+1} is the waiting time from T_i to the first jump of internal process N^n supposing no external jumps before that, hence
$$I_{i+1} \stackrel{d}{=} -\frac{1}{\delta_n} \ln d_{i+1} | d_{i+1} > 0$$
where
$$d_{i+1} = 1 + \frac{\delta_n \ln U_{i+1}^I}{\lambda^k_{T_i^+}}.$$

3. Update λ_t^n and N_t^n at T_{i+1}.

$$\lambda^n_{T^+_{i+1}} = \begin{cases} \lambda^n_{T^-_{i+1}} + \sigma^2 \alpha_n, & W_{i+1} = I_{i+1} \\ \lambda^n_{T^-_{i+1}} + \sigma^2 \alpha_n, & W_{i+1} = E_{i+1} \end{cases} \tag{6.19}$$

where

$$\lambda^n_{T^-_{i+1}} = \lambda^n_{T^+_i} e^{-\delta_n (T_{i+1} - T_i)}.$$

and

$$N^n_{T^+_{i+1}} = \begin{cases} N^n_{T^-_{i+1}} + 1, & W_{i+1} = I_{i+1} \\ N^n_{T^-_{i+1}}, & W_{i+1} = E_{i+1}. \end{cases} \tag{6.20}$$

From the simulation of λ_t^n above, we obtain the path of X_t^n as scaled UDCP by definition

$$X_t^n = \frac{1}{\sigma^2 \alpha_n^2} \lambda_t^n.$$

6.4 Numerical Results

In this section we assess the performance of this approximate simulation scheme by comparing the distributions in terms of the Laplace transform. Instead of taking single-time random variable to compare, we choose the integrated processes $\int_0^t X_s^n ds$ and $\int_0^t X_s ds$ since they are variables related to the path of X^n and X respectively.

Since the CIR process X is an affine process, in [30] for example, we have the closed-form of the Laplace transform of the integrated process $\int_0^t X_s ds$ for $t \geq 0$ when X follows the dynamics in (6.2) with a starting value X_0:

$$\mathbb{E}_{X_0}\left[e^{-u \int_0^t X_s ds}\right] = \left(\frac{e^{\frac{\beta t}{2}}}{\cosh\left(\frac{\kappa t}{2}\right) + \frac{\beta}{\kappa} \sinh\left(\frac{\kappa t}{2}\right)} \right)^{\frac{2\beta\theta}{\sigma^2}} \exp\left(-X_0 \frac{2 \sinh\left(\frac{\kappa t}{2}\right) u}{\kappa \cosh\left(\frac{\kappa t}{2}\right) + \beta \sinh\left(\frac{\kappa t}{2}\right)} \right),$$

where $\kappa = \sqrt{\beta^2 + 2u\sigma^2}$.

On the other hand, for the approximating process X^n following the dynamics in (6.14), with Lemma 6.2.6 and the exact simulation algorithm in Section 6.3, we simulate the integrated variable $\int_0^T X_s^n ds$ using the Monte-Carlo method. Moreover, we compare the theoretical Laplace transform of $\mathbb{E}[e^{-u \int_0^T X_s ds}]$ with the simulated Laplace transform value $\mathbb{E}[e^{-u \int_0^T X_s^n ds}]$.

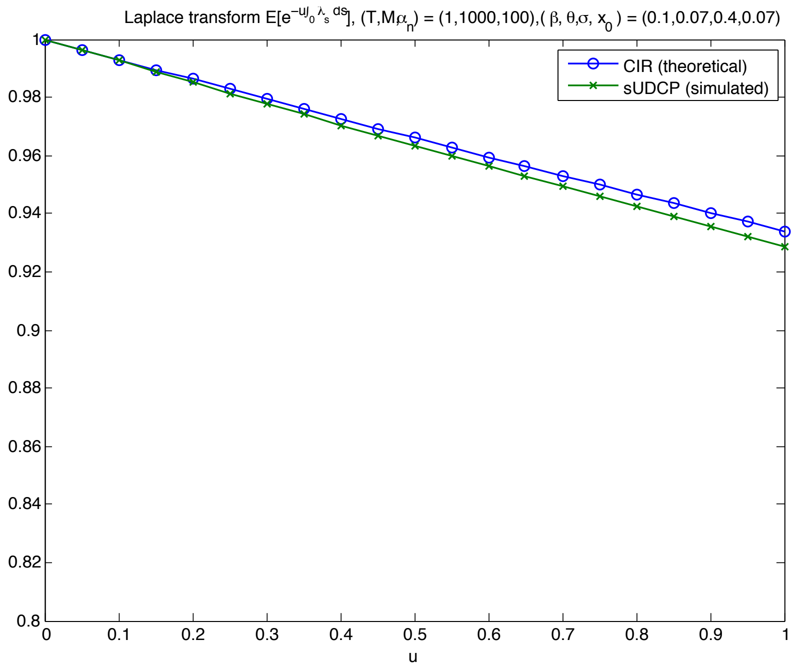
We take $T = 1$, the number of sample paths $M = 1000$, the Laplace transform parameter $u = 0$ to 1, and use three parameter sets for different applications.

(a) For the interest rate modelling: $(\beta, \theta, \sigma, \lambda_0) = (0.1, 0.07, 0.4, 0.07)$.

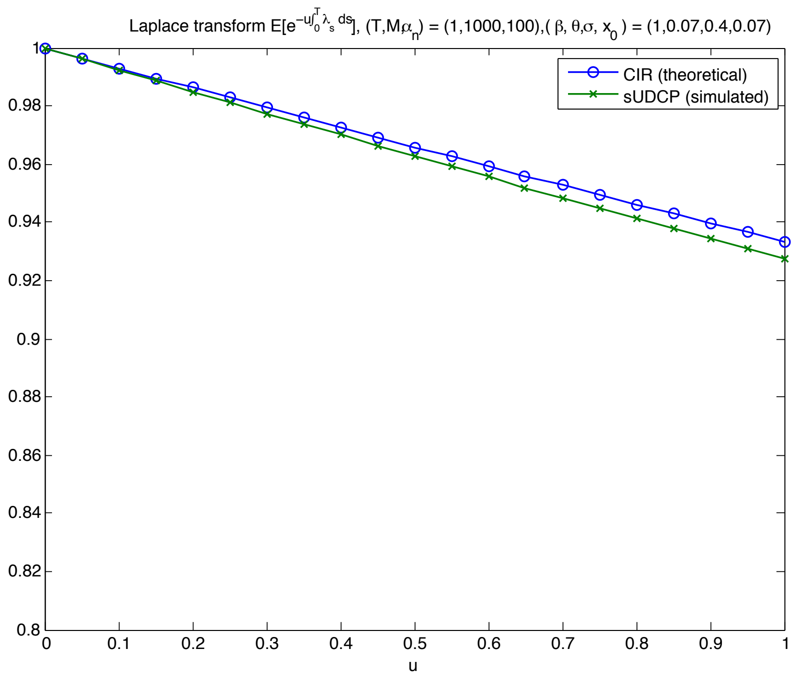
(b) For the variance modelling in the equity market: $(\beta, \theta, \sigma, \lambda_0) = (1, 0.07, 0.4, 0.07)$.

(c) For modelling a higher level of process: $(\beta, \theta, \sigma, \lambda_0) = (1, 1, 0.5, 2)$.

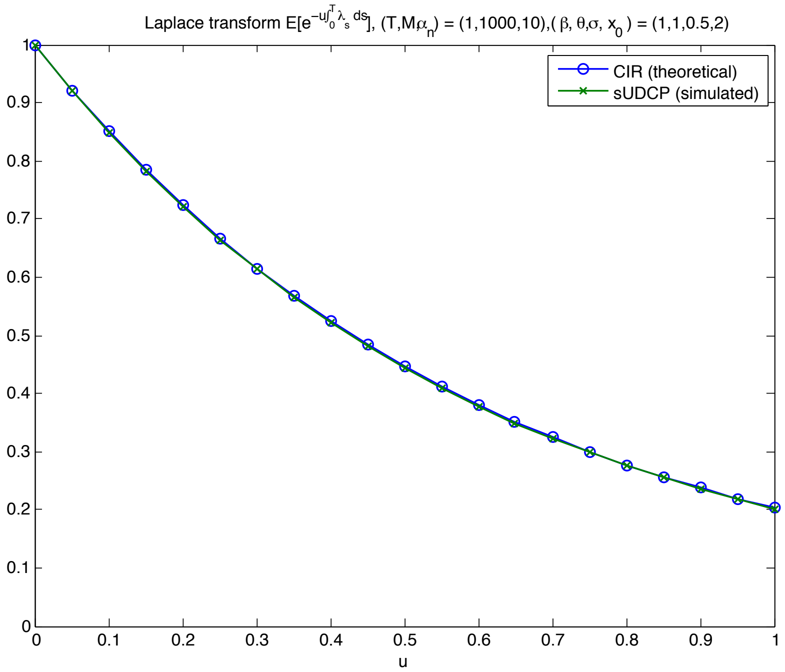
The results are illustrated in (a), (b), (c) of Figure 6.2. We observe that the Laplace transform values of the approximating process is very close to the theoretical one. Note that the Feller condition is violated in cases (a) and (b), the convergence is slower ($\alpha_n = 100$) than the case (c) where we only need $\alpha_n = 10$ to have a good approximation.



(a) The Laplace transform of the integrated process $\int_0^T \lambda_s ds$ with λ is the scaled UDCP intensity X^n and the CIR process X with parameter $(\beta, \theta, \sigma, \lambda_0) = (0.1, 0.07, 0.4, 0.07)$ typically for the interest rate. $T = 1$, $M = 1000$ and $\alpha_n = 100$.



(b) The Laplace transform of the integrated process $\int_0^T \lambda_s ds$ with λ is the scaled UDCP intensity X^n and the CIR process X with parameter $(\beta, \theta, \sigma, \lambda_0) = (1, 0.07, 0.4, 0.07)$ typically for variance in the equity market. $T = 1$, $M = 1000$ and $\alpha_n = 100$.



(c) The Laplace transform of the integrated process $\int_0^T \lambda_s ds$ with λ is the scaled UDCP intensity X^n and the CIR process X with parameter $(\beta, \theta, \sigma, \lambda_0) = (1, 1, 0.5, 2)$. $T = 1$, $M = 1000$ and $\alpha_n = 10$.

Figure 6.2: The Laplace transform of the integrated process with different parameter sets.

6.5 Conclusion

We discussed an application of diffusion approximation to a class of pure jump processes. We constructed the sequence of pure jump processes based on the UDCP that converges weakly to a CIR process and this allowed us to simulate CIR diffusion processes approximately with the pure jump process. Note that the pure jump processes are built upon univariate dynamic contagion processes and therefore, we built the link between non-diffusive self-exciting processes and square-root diffusion processes.

Chapter 7

Conclusions and Future Research

Conclusions In this work, we mainly studied pure jump Lévy processes and dynamic contagion processes, and their applications in finance. We first studied the intensity process of a first passage time structural credit risk model when the firm asset value is modelled as the exponential of a pure jump Lévy process of finite variation. Moreover, we explored the stationarity and the diffusion approximation of the bivariate dynamic contagion process, and we apply the results in filtering. In the end, we investigate the diffusion approximation of a non-diffusive process built upon a univariate dynamic contagion process by a CIR process, and hence we obtain an alternative approach of simulating the CIR process.

Future Research

We discuss a few questions for potential future research.

Chapter 2: We discussed the existence of intensity processes from a first passage time structural model. Though the framework is broad enough for the modelling purpose, there are a few questions open for the future research:

1. If we model the asset price with a Lévy process of infinite variation, then we can explore whether the intensity still exists and what the explicit representation is.

2. Recall Black and Cox [12] where the intensity process is a function of an exogenous process and the default is the first jump of a Cox process. Then the price can be simply computed in the same way as the risk-free case but with

the intensity added to the interest rate in the discount factor. It is because the default time is constructed as the first time the integrated intensity process is bigger than an independent exponential random variable. In our case, as the default time does not follow the Cox construction, and we have additional jump risk at default, the pricing formula will have an additional term. How to compute the additional term explicitly is open to future research.

Chapter 5: We explored the diffusion approximation of a high frequency event system and applied it in filtering with point process observations. Note that the constructed filter is an approximate one. We compared filtering errors using the approximate filter with the limiting diffusion model and concluded that the filter is asymptotically optimal. We can improve the assessment of the performance with the following methods:

1. Compare the filtering errors of approximate filter with the exact one that is the solution of the KS equation.

2. Explore the convergence rate of the approximate filter to the Kalman-Bucy filter for the limiting diffusion process.

Chapter 6: We showed the weak convergence of the pure jump process to the CIR process when a model parameter tends to infinity and the Laplace transform showed that the performance of the approximate simulation scheme worked well. We have the following remaining suggestions for future research.

1. It would be interesting to explore the convergence rate of the approximation.

2. As the weak convergence holds in the path space, we can assess the performance of the simulation scheme for path-dependent options using the Heston model. For example, we can compare prices of Asian options and barrier options using this approximation simulation scheme above with the quadratic exponential scheme (QE) in [4].

Bibliography

[1] Aït-Sahalia, Y., Cacho-Diaz, J. and Laeven, R. J. (2010). Modeling financial contagion using mutually exciting jump processes. *Technical report* 15850. National Bureau of Economic Research.

[2] Albrecher, H. and Asmussen c, S. (2006). Ruin probabilities and aggregate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal* **2006,** 86–110.

[3] Altmann, T., Schmidt, T. and Stute, W. (2008). A shot noise model for financial assets. *International Journal of Theoretical and Applied Finance* **11,** 87–106.

[4] Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance* **11,** 1–42.

[5] Andersen, L. and Brotherton-Ratcliffe, R. (2005). Extended Libor market models with stochastic volatility. *Journal of Computational Finance* 1–40.

[6] Aven, T. (1985). A theorem for determining the compensator of a counting process. *Scandinavian Journal of statistics* **12,** 69–72.

[7] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J. (2013). Some limit theorems for hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123,** 2475–2499.

[8] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance* **13,** 65–77.

[9] Barczy, M., Döring, L., Li, Z. and Pap, G. (2013). Stationarity and ergodicity for an affine two factor model. *Forthcoming in Advances in Applied Probability. Available also on arXiv: 1302.2534*.

[10] Bauwens, L. and Hautsch, N. (2009). *Modelling financial high frequency data using point processes.* In Handbook of Financial Time Series. Springer Berlin Heidelberg. pp. 953–979.

[11] Bertoin, J. (2000). *Subordinators, Lévy processes with no negative jumps, and branching processes.* Centre for Mathematical Physics and Stochastics, University of Aarhus.

[12] Black, F. and Cox, J. C. (1976). Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance* **31,** 351–367.

[13] Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer.

[14] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability* **24,** 1563–1588.

[15] Ceci, C. and Gerardi, A. (2006). A model for high frequency data under partial information: a filtering approach. *International Journal of Theoretical and Applied Finance* **9,** 555–576.

[16] Cetin, U., Jarrow, R., Protter, P. and Yldrm, Y. (2004). Modeling credit risk with partial information. *The Annals of Applied Probability* **14,** 1167–1178.

[17] Cont, R. and De Larrard, A. (2012). Order book dynamics in liquid markets: limit theorems and diffusion approximations. arXiv:1202.6412.

[18] Costa, O. (1990). Stationary distributions for piecewise-deterministic Markov processes. *Journal of Applied Probability* **27,** 60–73.

[19] Cox, D. R. and Isham, V. (1980). *Point processes* vol. 12. Springer.

[20] Daley, D. J. and Vere-Jones, D. (2002). *An introduction to the theory of point processes* vol. I. Springer.

[21] Dassios, A. and Jang, J.-W. (2003). Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics* **7,** 73–95.

[22] Dassios, A. and Jang, J.-W. (2005). Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *Journal of Applied Probability* **42,** 93–107.

[23] Dassios, A. and Zhao, H. (2011). A dynamic contagion process. *Advances in Applied Probability* **43,** 814–846.

[24] Dassios, A. and Zhao, H. (2012). Ruin by dynamic contagion claims. *Insurance: Mathematics and Economics* **51,** 93–106.

[25] Dassios, A. and Zhao, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability* **18,** 1–13.

[26] Davis, M. H. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)* 353–388.

[27] Doney, R. (2001). *Fluctuation theory for Lévy processes.* Springer.

[28] Duffie, D., Filipovic, D. and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability* **13,** 984–1053.

[29] Duffie, D. and Lando, D. (2003). Term structures of credit spreads with incomplete accounting information. *Econometrica* **69,** 633–664.

[30] Dufresne, D. (2001). The integrated square-root process. *Technical report.* Centre for Actuarial Studies, Department of Economics, University of Melbourne. ISBN 0734021348 research paper no. 90.

[31] Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* **48,** 367–378.

[32] Errais, E., Giesecke, K. and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics* **1,** 642–665.

[33] Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence.* John Wiley and Sons.

[34] Giesecke, K. (2006). Default and information. *Journal of Economic Dynamics and Control* **30,** 2281–2303.

[35] Glasserman, P. and Kim, K.-K. (2010). Moment explosions and stationary distributions in affine diffusion models. *Mathematical Finance* **20,** 1–33.

[36] Goldberg, L. and Giesecke, K. (2004). Forecasting default in the face of uncertainty. *Journal of Derivatives* **12,** 14–25.

[37] Guo, X., Jarrow, R. A. and Zeng, Y. (2009). Credit risk models with incomplete information. *Mathematics of Operations Research* **34,** 320–332.

[38] Guo, X. and Zeng, Y. (2008). Intensity process and compensator: A new filtration expansion approach and the Jeulin-Yor theorem. *The Annals of Applied Probability* **18,** 120–142.

[39] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* **11,** 493–503.

[40] Jaisson, T. and Rosenbaum, M. (2013). Limit theorems for nearly unstable hawkes processes. *arXiv:1310.2033*.

[41] Jang, J. and Dassios, A. (2013). A bivariate shot noise self-exciting process for insurance. *Insurance: Mathematics and Economics* **53,** 524–532.

[42] Janson, S., Sokhna, M. B. and Protter, P. (2011). Absolutely continuous compensators. *International Journal of Theoretical and Applied Finance* **14,** 335–351.

[43] Jeulin, T. and Yor, M. (1978). Grossissement d'une filtration et semimartingales: formules explicites. *Séminaire de Probabilités XII* 78–97.

[44] Karoui, N. E., Jeanblanc, M. and Jiao, Y. (2010). What happens after a default: the conditional density approach. *Stochastic processes and their applications* **120,** 1011–1032.

[45] Keller-Ressel, M., Schachermayer, W. and Teichmann, J. (2011). Affine processes are regular. *Probability Theory and Related Fields* **151,** 591–611.

[46] Klüppelberg, C. and Mikosch, T. (1995). Explosive Poisson shot noise processes with applications to risk reserves. *Bernoulli* **1,** 125–147.

[47] Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability* **19,** 1035–1070.

[48] Kushner, H. J. and Runggaldier, W. J. (1987). Filtering and control for wide bandwidth noise driven systems. *Automatic Control, IEEE Transactions on* **32,** 123–133.

[49] Kusuoka, S. (1999). A remark on default risk models. *Advances in mathematical economics* **1,** 69–82.

[50] Kyprianou, A. (2006). *Introductory lectures on fluctuations of Lévy processes with applications.* Springer.

[51] Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives research* **2,** 99–120.

[52] Liniger, T. J. (2009). Multivariate Hawkes processes. *PhD thesis.* Eidgenössische Technische Hochschule ETH Zürich, Nr. 18403.

[53] Liptser, R. S. and Runggaldier, W. J. (1991). On diffusion approximations for filtering. *Stochastic Processes and their Applications* **38,** 205–238.

[54] Macci, C. and Torrisi, G. L. (2011). Risk processes with shot noise Cox claim number process and reserve dependent premium rate. *Insurance: Mathematics and Economics* **48,** 134–145.

[55] Madan, D. B., Carr, P. P. and Chang, E. C. (1998). The variance gamma process and option pricing. *European Finance Review* **2,** 79–105.

[56] Madan, D. B. and Schoutens, W. (2008). Break on through to the single side. *Journal of Credit Risk* **4,** 3–20.

[57] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* **29,** 449–470.

[58] Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability* **35,** 614–640.

[59] Protter, P. (2005). *Stochastic Integration and Differential Equations* 2.1 ed. Springer.

[60] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

[61] Westcott, M. (1971). On existence and mixing results for cluster point processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 290–300.

[62] Xiong, J. (2008). *An Introduction to Stochastic Filtering Theory* vol. 18 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, UK.

[63] Zhao, H. and Dassios, A. (2011). A dynamic contagion process and an application to credit risk. *Available at SSRN 2039798*.