LOCAL INTENSITY AND ITS DYNAMICS IN MULTI-NAME CREDIT DERIVATIVES MODELING

BY MING SHI

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Paul Feehan

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2010

ABSTRACT OF THE DISSERTATION

Local Intensity and its Dynamics in Multi-Name Credit Derivatives Modeling

by Ming Shi

Dissertation Director: Paul Feehan

We import the problems and techniques developed for the local volatility model in equity derivatives to multi-name credit modeling, propose and solve analogous problems. In particular, we analyze the properties of the local intensity of the default counting process and explore the stochastic evolution of the local intensity surface under the "Top-down" credit modeling framework. The analogue of Dupire formula, Gyöngy's theorem, backward and forward equations are developed and parametric factor models for the dynamics of the local intensity surface will be discussed.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Paul Feehan for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation.

Besides my advisor, I would like to thank the rest of my dissertation committee: Prof. Daniel Ocone, Richard Gundy, Ren-Raw Chen for their encouragement, insightful comments and questions. Also I would like to thank Prof Richard Falk, Jesus Rodriguez for their generous help in my research.

My sincere gratitude also goes to my fellow colleagues: Jin Wang, Jawon Koo, Eduardo Osorio, Biao Yin, Yuan Zhang, Jihui Zhao and Wenhua Jiang for the stimulating discussions, for the sleepless nights we were working together and for all the fun we have had in my precious years at Rutgers.

Last but not the least, I would like to thank my family: my parents Lingen Shi and Zaifen Lu for their unconditional support and love throughout my life, I am especially indebted to their impeccable care for my daughter. I would like to also thank my sister Hong Shi for being supportive all the time. To mention my wife Jie Bai and daughter Danielle Yongxi Shi finally, they are the source of all my strength. To them I dedicate this dissertation.

Dedication

This dissertation is dedicated to my precious wife, Jie, and my daughter, Danielle. I give my deepest expression of love and appreciation for the support that you gave and the sacrifices you made during the difficult time in this graduate program.

Table of Contents

Chapter 1

Introduction

This dissertation is dedicated to the study of mathematical models used for valuation and risk analysis of multi-name credit derivatives such as index-credit default swaps (CDS), collateralized debt obligations (CDO), and other exotic portfolio credit derivatives. Specifically, in this dissertation, we analyze the properties of the local intensity of the default counting process N_t and explore the stochastic evolution of the local intensity surface of N_t under the "top-down" credit modeling framework. In this introductory section I will briefly review the background of the top-down approach, show how my work fits into the literature, and then discuss my future research interests.

1.1 Overview

Credit derivatives are tools for transferring and hedging credit risk. The credit derivatives market has grown rapidly both in volume and in the breadth of the instruments offered. Multi-name credit derivatives are among the most complex of these instruments and their values are derived from an underlying portfolio of corporate bonds or other credit-sensitive securities. They enable investors to buy and sell protection against default losses in the portfolio. Valuation and risk analysis of a multi-name credit derivative is challenging due to the complex economic phenomena that drive correlated corporate default risk and the dependency of the distribution of portfolio loss on multiple time horizons. The shortage of reliable mathematical models was highlighted by the recent subprime mortgage crisis.

Although it has been widely used, it is well known that the copula model (Vasicek [25], Li [17]) in industry does not provide a satisfactory modeling solution, since it is a static model unable to capture the dynamic evolution of default probabilities. Reduced

form models are generally now preferred and have become a standard tool for modeling the dynamics of credit risk. Reduced form models assume that default occurs without warning at an exogenous default rate, or *intensity*. The dynamics of the intensity can be specified under the pricing or risk neutral probability measure. Therefore, instead of asking why the firm defaults, the intensity model is calibrated directly from the market prices of traded instruments. To construct an intensity model one must give a model to describe the dynamics of the intensity and then seek the distribution of the default probability.

Reduced form models are distinguished by the way in which the intensity of the default process is specified. In a bottom up model, the portfolio intensity is modelled by specifying both the intensities of individual constituents and the dependency structure among them. Such a specification is appropriate for the analysis of portfolios of highly heterogeneous constituents. It brings the information of the single-name market to bear on the calibration of the model. However, the dependence among single defaults is partly arbitrary and consistent calibration is difficult or practically impossible. For further information about the bottom up model, see Papageorgiou & Sircar [8], Frey & Backhaus [10], Duffie & Garleanu [6], and others. In contrast, top-down models directly specify the portfolio default intensity $\lambda(t) = \lim_{h \searrow 0} \frac{1}{h} \mathbb{P}(N(t+h) - N(t) = 1 | \mathcal{F}_t)$ [1] without reference to the portfolio constituents, where $N(t)$ is the number of defaulted obligors in the portfolio. The loss distribution can be expressed in terms of a few economically meaningful parameters appearing in the specification of $\lambda(t)$; this leads to a tractable credit derivatives valuation model which can be addressed by a variety of efficient methods.

By assuming name homogeneity [2] and constant recovery rates in the portfolio, the default counting process $N(t)$ becomes the modeling primitive and is completely determined by $\lambda(t)$. Frey and his coauthors [10] specified $N(t)$ as a *birth process*

[1] We assume at most one jump during a sufficiently short time period; Schönbucher [23] showed that this is only a very slight deviation from a fully general model allowing multiple jumps. Multiple instantaneous defaults have also been discussed in [1].

[2] Different names (of companies) in the portfolio have equal notional values, so names are interchangeable. The name dependence structure is implicit in this specification. It does not matter which name defaults. This is a typical assumption in valuation of multi-name credit derivatives.

$\lambda(t) = \nu + \alpha N(t)$ and also gave a nonlinear generalization $\lambda(t) = \nu + \frac{\alpha^2}{\beta}(e^{\frac{\beta}{\alpha}(N_t - \mu(t))^+} - 1)$. Under such specifications, $N(t)$ is a time-inhomogeneous Markov chain. A more popular generalization of a Markovian loss process is to incorporate other sources of randomness in the specification of the intensity by letting $\lambda = \lambda(\mathbf{Y}(t), N(t))$, where $\mathbf{Y}(t)$ contains other random factors independent of $N(t)$. If $\lambda = Y(t)$, we obtain *doubly stochastic intensity*. If $\lambda = Y(t) \cdot (\overline{N} - N(t))$, $\overline{N}$ is the total number of names, then we obtain a *doubly stochastic death process*. If $\lambda = Y(t) f(N(t))(\overline{N} - N(t))$, where $f(\cdot)$ is some deterministic function, we obtain the *bivariate spread loss process* of Arnsdorf & Halperin [13]. The specification may be also given in a differential form, such as $d\lambda(t) = \kappa(\rho(t, N(t)) - \lambda(t)) + \sigma\sqrt{\lambda(t)}dt$ in Lopatin & Misirpashaev [18]. Another interesting specification of portfolio intensity is given by the self-exciting *Hawkes process* [19], where $\lambda(t) = \nu + \int_0^t \alpha e^{-\beta(t-s)} dN(s)$. The Hawkes process does not use any other source of randomness, but unlike the Markov chain model where $\lambda(t)$ only depends on the current loss level $N(t)$, it depends on the whole path of $N(t)$

Motivation from equity derivatives modeling

All of the above multi-name credit models give more or less realistic dynamics of the loss process $N(t)$, and can model the "contagion" feature of defaults. In this dissertation, we explore approaches to credit modeling inspired by the perspective of local volatility models from equity derivative modeling, providing new methods and addressing new questions suggested by equity derivative modeling.

In Dupire's *local volatility* asset price model [7], which is mainly used in equity derivative modeling, one assumes that the volatility has the form $\sigma(t, S(t))$ (for some suitable deterministic function, $\sigma(t, x)$) and the asset price process $S(t)$ has the dynamics $dS(t) = rS(t)dt + \sigma(t, S(t))S(t)dW(t)$. Under this specification, $S(t)$ is a one dimensional Markovian process. The local volatility can be expressed in terms of the prices of European call options by Dupire's formulae $\sigma^2(T, K) = \frac{\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K}}{\frac{1}{2}K^2\frac{\partial^2 C}{\partial K^2}}$, when those call option prices $C(T, K)$ are quoted for all maturities and strikes.

Gyöngy's theorem asserts that there is a Markovian asset price process $\hat{S}(t)$ with

local volatility having the same marginal distributions as a given asset price process $S(t)$ with arbitrary adapted stochastic volatility $\sigma(\omega, t)$, if the corresponding local volatility $\Sigma(t, x)$ equals a conditional expectation of the stochastic volatility, i.e., $\Sigma^2(t, x) = E[\sigma^2(\omega, t)|S(t) = x]$. So the local volatility model is general enough in the sense of pricing European style derivatives, whose prices only depend on the marginal distribution of the underlying asset.

While Dupire only assumes that the local volatility surface $\Sigma(\cdot, \cdot)$ is fixed. Carmoma and Nadtochiy [3] considered the dynamics of this local volatility surface $\Sigma_t(\cdot, \cdot)$. There is no static arbitrage as long as local volatility is positive, but in order to avoid dynamic arbitrage, the differential equations governing the dynamics of local volatility surface have to satisfy a very complex drift condition. The problems around the dynamics of the local volatility surface are very interesting and challenging.

1.2 Dissertation summary

My dissertation research pursued the following two themes.

Investigation and analysis of local intensity surface and local intensity model.

We import the problems and techniques developed for the local volatility model in equity derivatives to multi-name credit modeling, and propose and solve analogous problems. By analogy with the definition of the local volatility $\sigma = \Sigma(t, S(t))$, we call $\lambda = \Lambda(t, N(t))$ the local intensity and the bivariate function $\Lambda(\cdot, \cdot)$ the local intensity surface, though the second argument only takes discrete values. A top down model with such a specification is called a local intensity model. Other general adapted intensity specifications are called stochastic intensity models.

Summary of main results

1. In order to see the relation between the local intensity and the distribution of $N(t)$ or derivative prices, in section (3.1), we show that there exists an analogue

of Dupire's formula which expresses the local intensity surface in terms of the marginal distributions of $N(t)$, or the prices of European-style vanilla multi-name credit derivatives.

$$\Lambda(T,n) = -\frac{1}{p_n(T)} \sum_{k=0}^{n} \frac{\partial}{\partial T} p_n(T) \qquad \Lambda(T,n) = -\frac{1}{\Delta^- \Delta^+ C(T,n)} \frac{\partial}{\partial T} \Delta^+ C(T,n) \tag{1.1}$$

where, $p_n(T) = \mathbb{P}(N_T = n)$, $C(T,n) = E[(N(T) - n)^+]$, $\Delta^+ C(T,n) = C(T,n+1) - C(T,n)$, $\Delta^- C(T,n) = C(T,n) - C(T,n-1)$

2. We prove an analogue of Gyöngy's theorem [12] in theorem (3.3.2), which asserts that there is a Markovian counting process $\hat{N}(t)$ with local intensity $\Lambda(T,n)$ having the same marginal distributions as a given counting process $N(t)$ with arbitrary adapted stochastic intensity $\lambda(t)$. The corresponding local intensity is expressed as a conditional expectation of the stochastic intensity

$$\Lambda_t(T,n) = E[\lambda(T) | N(T) = n, \mathcal{F}_t]. \tag{1.2}$$

The bivariate local intensity function changes in time t with respect to the evolution of the counting process with stochastic intensity: that is, the local intensity surface is not fixed with respect to time unless the stochastic intensity model is already a local intensity model.

3. We extend Gyöngy's theorem from pure jump models to jump diffusion models in section 3.4.2. We show that the marginal distributions of a discontinuous semi-martingale can be matched by a Markov process with local volatility and local intensity for its diffusion and jump components. We also give a partial integro-differential equation for the transition density function of the semi-martingale. This result extends the Kolmogorov forward equations to a non-Markovian settings.

4. We derive the Kolmogorov forward and backward difference (in state) and differential (in time) equations obeyed by the distribution function of $N(t)$ and the price function of European style vanilla options on $N(t)$ respectively in the local

intensity model in section 3.2. Those equations can be reduced to system of ordinary differential equations. By solving these ODEs iteratively, we are able to address the problem of CDO pricing [3].

5. For the benchmark, doubly stochastic intensity model and the Hawkes model, we derive formulae for their local intensities, using our analogue of Dupire's formula in section 3.5. The models with these local intensities will have the same marginal distribution as the given stochastic intensity models.

Investigation of the dynamics of the local intensity surface

The local intensity surface is not necessarily fixed, since the counting process $N(t)$ is not Markovian in general. It is interesting to consider problems related to the dynamics of the local intensity surface. Namely, what is the time evolution of the local intensity surface $\Lambda_t(\cdot,\cdot)$. Schönbucher in [23] interpreted the local intensity as a "forward transition rate" and suggested the HJM-style term structure model to characterize its dynamics. He also gave the no-arbitrage drift condition. However, HJM models are generally infinite dimensional, so they are unsuitable for implementation purposes. Therefore, we are interested in the finite dimensional realization problem, that is to specify a dynamics for the local intensity surface that are driven by a finite number of random factors and that do not allow arbitrage.

Summary of main results

1. We propose parametric factor models in section 4.3.1 for the dynamics of the local intensity surface since they are automatically finite dimensional. We give a concrete example of a parametric factor model which satisfies the no-arbitrage drift condition.

2. We discuss two examples of one dimensional factor models in section 4.3.2. If the local intensity surface is scaled or shifted by a single factor Z_t ie, $\Lambda_t(T,n) =$

[3] The price (spread) of CDO tranches can be written in terms of the European call option price $E[(N(t)-K)^+]$

$Z_t \cdot \Lambda_0(T,n)$ or $\Lambda_t(T,n) = \Lambda_0(T,n) + Z_t$, then the drift condition forces Z_t to be 1 and 0 respectively. Therefore, random scaling and shifting models are reduced to the local intensity model. This fact shows the restrictiveness of the no-arbitrage condition.

Future research plans

We plan to pursue our study of the framework of parametric factor families to model local intensity surfaces. Parametric families of forward curves have played an important role in the analysis of interest rate modeling. Carmona [3] advocates this approach for modeling local volatility surfaces. We believe that this approach will lead to new multi-name credit derivative models.

We also plan to investigate the relationship between local intensity and stochastic intensity and try to answer the following two questions: (1) What is the corresponding initial local intensity surface, given a stochastic intensity model. (2) Given a stochastic intensity model, what are the dynamics of the corresponding local intensity surface? The first question is inspired by the question of how to find the local volatility surface for the Heston model (a benchmark stochastic volatility model), which is only partially solved. The second question is much more challenging, and has not caught the attention of equity modelers so far. This question is also clearly related to the dynamics of local intensity problem.

Chapter 2

Background and preliminaries

This is an introductory chapter, in which we will give necessary definitions, descriptions and theorems in both finance and mathematics which are used in later chapters.

The first section is dedicated to the introduction of credit derivatives. We will give the definition and analyze the payoff of single-name and multi-name credit derivatives. Pricing and hedging of these derivatives are the main problems considered in academia and industry today.

In the second section we briefly introduce the mathematical tools used in credit modeling. We will present intensity of jump processes, time-inhomogeneous Markov chain, Itô-Doeblin's formula for point processes.

In the third section, we give a brief introduction of intensity based Top-down models.

In the fourth section, we introduce the local volatility model and Gyöngy's theorem.

2.1 Credit derivatives

In recent years, credit derivatives have become the main tools for transferring and hedging credit risk. The value of a credit derivative is derived from the credit risk on underlying bonds, loans or other financial assets. It is a bilateral contract which allows the buyer and seller to transfer credit risk.

There are several ways of classifying credit derivatives. We will consider single-name versus multi-name.

In order to give the definition of various credit derivatives by using accurate mathematics, we assume all processes and random variable we discuss later on are defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{(t\geq 0)}$, where Ω represents the set of all possible states of nature. $\mathbb{P}$ is the pricing measure that attaches probabilities to the events in Ω. The filtration $(\mathcal{F}_t)_{(t\geq 0)}$ represents the information available at time t. Throughout this dissertation, we assume that $(\mathcal{F}_t)_{(t\geq 0)}$ satisfies the usual conditions. [1]. We use the following notations:

$$\tau = \text{ random default time of a certain company} \tag{2.1}$$

$$Q(t,T) = \mathbb{P}\left[\tau \leq T | \mathcal{F}_t\right] \text{ default probability} \tag{2.2}$$

$$P(t,T) = \mathbb{P}\left[\tau > T | \mathcal{F}_t\right] \text{ survival probability} \tag{2.3}$$

$$B(t,T) = \text{ time t value of risk free zero coupon bond expiring at T} \tag{2.4}$$

$$\bar{B}(t,T) = \text{ time t value of defaultable zero coupon bond expiring at T} \tag{2.5}$$

$$r(t) = \text{short risk free interest rate} \tag{2.6}$$

$$D(t,T) = e^{-\int_t^T r(s)ds} \text{ discount factor from } t \text{ to } T \tag{2.7}$$

$$R = \text{Recovery rate} \tag{2.8}$$

[1] that means $(\Omega, \mathcal{F}, \mathbb{P})$ is complete; $\mathcal{F}_0$ contains all events of measure zero; $(\mathcal{F}_t)_{(t\geq 0)}$ is right continuous

2.1.1 Single-name credit derivatives

Single-name credit derivatives involve protection against a default or other credit event by a single reference entity. These instruments account for the majority of trading activity in the current marketplace. The most popular single name credit derivatives are defaultable bond and credit default swap (CDS)

1. **Defaultable bond**

A defaultable zero coupon bond with maturity T is the simplest credit derivative, which behaves the same as an ordinary zero coupon bond, but might fail to deliver the principal in case default occurs before maturity; that is, the payoff is

$$1_{\{\tau>T\}} = \begin{cases} 1 & \tau > T \quad \text{(default after } T\text{)}, \\ 0 & \tau \leq T \quad \text{(default before } T\text{)}, \end{cases}$$

where τ is the random time of default. The price of a defaultable zero coupon bond is

$$\bar{B}(t,T) = E\left[D(t,T)1_{\{\tau>T\}}|\mathcal{F}_t\right] \tag{2.9}$$

$$= B(t,T)P(t,T) \tag{2.10}$$

where $P(t,T) = \mathbb{P}(\tau > T|\mathcal{F}_t)$ is the survival probability, if we assume default is independent of the risk free interest rate.

2. **Credit default swap (CDS)**

A credit default swap (CDS) is a swap designed to transfer the credit exposure of fixed income products between parties. It is the most widely used credit derivative. It is an agreement between a protection buyer and a protection seller whereby the buyer periodically or continuously pays a premium over the lifetime of the contract to the protection seller in exchange for a payment by the seller contingent upon a credit event (such as a default) occurring in the reference entity. The spread is usually paid quarterly. The most common maturities are 3, 5, 7 and 10 years.

Most CDS contracts are physically settled, where upon a credit event the protection seller must pay the par amount of the contract against the protection buyer's obligation

to deliver a bond or loan of the name against which protection is being sold.

Consider a CDS with notional \$1, written at time t, with coupon payment dates $T_1, \cdots, T_n$, and the annualized premium spread S. Assume that in default, the CDS pays $1 - R$ right after the default, where $R \in [0, 1]$ is the recovery rate. Assume that default probability is independent of risk free interest rate, then the present value of the premium leg PL is

$$PL = E_t \left[\sum_{i=1}^{n} (S\delta_i) D(t, T_i) 1_{\{\tau > T_i\}} \right] \tag{2.11}$$

$$= S \sum \delta_i D(t, T_i) P(\tau > T_i) \tag{2.12}$$

where δ_i is the accrual factor (the number of days between T_{i-1} and T_i divided by the number of days in year, under a suitable day-count convention). $D(t, T_i)$ is the discount factor, meaning the time t value of a dollar at T_i.

The present value of the default leg DL is

$$DL = E_t \left[\sum_{i=1}^{n} (1 - R) D(t, T_i) 1_{\{T_{i-1} < \tau \leq T_i\}} \right] \tag{2.13}$$

$$= (1 - R) \sum_{i=1}^{n} D(t, T_i) \left[P(T_{i-1} < \tau \leq T_i) \right]. \tag{2.14}$$

So the fair price constraint $PL = DL$ produces the par spread formula

$$S = (1 - R) \frac{\sum_{i=1}^{n} D(t, T_i) \left[P(T_{i-1} < \tau \leq T_i) \right]}{\sum_{i=1}^{n} \delta_i D(t, T_i) P(\tau > T_i)}. \tag{2.15}$$

2.1.2 Multi-name credit derivatives

A multi-name credit derivative is a security contingent on an underlying portfolio of corporate bonds or other credit-sensitive securities. It enables investors to buy and sell protection against the default losses in the portfolio.

Consider a portfolio of $\bar{N}$ firms. If I_i denotes the notional of a security referencing firm i, then the notional of the portfolio is the sum

$$I = \sum_{i=1}^{\bar{N}} I_i.$$

If name i defaults, the corresponding loss is $(1 - R_i)I_i$, where R_i is the recovery rate of name i. The portfolio loss at time t is

$$L_t = \sum_{i=1}^{\bar{N}} (1 - R_i) I_i 1_{\{\tau_i \leq t\}}.$$

The default counting process

$$N_t = \sum_{i=1}^{\bar{N}} 1_{\{\tau_i \leq t\}}$$

gives the number of defaulted firms by time t. Multi-name products are usually derivatives on L_t or N_t. We gave several examples of the most popular multi-name credit derivatives as follows.

1. **Index credit default swap**

A credit default swap index is a credit derivative used to hedge credit risk or to take a position on a basket of credit entities. Unlike a credit default swap, which is an single name and over the counter credit derivative, an index credit default swap is multi-name and completely standardized credit security. It may therefore be more liquid and trade at a smaller bid-offer spread. There are currently two main families of CDS indices: CDX and iTraxx. CDX indices contain North American and Emerging Market companies and are administered by CDS Index Company (CDSIndexCo) and marketed by Market Group Limited, and iTraxx contain companies from the rest of the world and are managed by the International Index Company (IIC).

The protection buyer promises to pay at the payment date t_m a protection fee S (spread) times the remaining notional of the index times a day-count fraction. It is important to note that at a default, the notional on which the fee is paid is reduced. If we assume all names have same notional, then at time t, the remaining notional is

$I - L_t = I(1 - \frac{N_t}{\bar{N}})$. So the present value of the premium leg is

$$PL = E_t \left[\sum_{t_m > t} S\delta_m D(t, t_m) I(1 - \frac{N_{t_m}}{\bar{N}}) \right] \tag{2.16}$$

$$= \sum_{t_m > t} S\delta_m I D(t, t_m) \left(1 - \frac{E_t \left[N_{t_m} \right]}{\bar{N}} \right). \tag{2.17}$$

The default leg is a stream of payments at t_m that cover the portfolio losses in the time period (t_{m-1}, t_m), until maturity T which is typical 3, 5, 7 or 10 years. So present value of the default leg DL is

$$DL = E_t \left[\sum_{t_m > t} D(t, t_m)(L_{t_m} - L_{t_{m-1}}) \right] \tag{2.18}$$

$$= \sum_{t_m > t} D(t, t_m) \frac{I}{\bar{N}} (E_t[N_{t_m}] - E_t[N_{t_{m-1}}]). \tag{2.19}$$

So the fair price constraint $PL = DL$ produces the par spread formula

$$S = \frac{\sum_{t_m > t} D(t, t_m)(E_t[N_{t_m}] - E_t[N_{t_{m-1}}])}{\sum_{t_m > t} \delta_m D(t, t_m)(\bar{N} - E_t \left[N_{t_m} \right])} \tag{2.20}$$

2. **Collateralized debt obligation (CDO)**

Like index CDS, a collateralized debt obligation (CDO) is also a transaction that transfers the credit risk on a reference portfolio of assets. The extra feature of a CDO structure is the tranching of credit risk. Those investors who are interested in specific risk profiles can invest in contracts based on a slice of the portfolio (tranche), with notional $I_{\text{tranche}} = I(K_2 - K_1)$, where, a lower attachment point $K_1 \in [0, 1]$ and a upper detachment point $K_2 \in [K_1, 1]$.

To analyze the loss leg and default leg of the tranches of synthetic CDOs, consider a synthetic CDO tranche on a given reference portfolio defined by an interval of percentage losses $[K_1, K_2]$ on the total portfolio notional I that the tranche investor is responsible for. The tranche investor receives periodic spread payments from the CDO issuer (the premium leg) and makes payments to the CDO issuer when defaults affect the tranche (the protection leg). Note that for a synthetic CDO, any default corresponds to a credit event under a CDS in the reference portfolio.

Recall that, the total loss of the portfolio at time t is given by

$$L_t = \sum_{i=1}^{\bar{N}} (1 - R_i) I_i 1_{\{\tau_i \leq t\}} \tag{2.21}$$

$$= (1 - R) I \frac{N_t}{\bar{N}}. \tag{2.22}$$

The default leg is a stream of payments that cover portfolio losses given that the cumulative losses are larger than K_1 but do not exceed K_2 The cumulative loss in the tranche $[K_1, K_2]$ is

$$L_{[K_1,K_2]}(t) = (L_t - IK_1)^+ - (L_t - IK_2)^+. \tag{2.23}$$

The present value of the default DL is

$$\begin{aligned} DL &= E_t \left[\sum_{t_m > t} D(t, t_m)(L_{[K_1,K_2]}(t_m) - L_{[K_1,K_2]}(t_{m-1})) \right] \\ &= \sum_{t_m > t} D(t, t_m)[(E_t(L(t_m) - K_1)^+ - E_t(L(t_m) - K_2)^+) - \\ &\qquad (E_t(L(t_{m-1}) - K_1)^+ - E_t(L(t_{m-1}) - K_2)^+)]. \end{aligned} \tag{2.24}$$

The present value of the premium leg PL is

$$PL = E_t \left[\sum_{t_m > t} S \delta_m D(t, t_m)(I(K_2 - K_1) - L_{[K_1,K_2]}(t_m)) \right]. \tag{2.25}$$

So the fair par spread is

$$S = \frac{E_t \left[\sum_{t_m > t} D(t, t_m)(L_{[K_1,K_2]}(t_m) - L_{[K_1,K_2]}(t_{m-1})) \right]}{E_t \left[\sum_{t_m > t} \delta_m D(t, t_m)(I(K_2 - K_1) - L_{[K_1,K_2]}(t_m)) \right]}. \tag{2.26}$$

3. **CDO squared**

A CDO squared is a single tranche CDO (outer) where the underlying portfolio consists of other tranche swaps (inner) instead of a CDS as an ordinary synthetic CDO. The cumulative loss associated with the outer CDO portfolio is a single aggregate of losses in each of the M inner CDOs.

The aggregate loss of the outer CDO is

$$L_t^{outer} = \sum_{m=1}^{M} (U_t^m) \tag{2.27}$$

where U_t^m is the loss of the m-th inner CDO,

$$U_t^m = (L_t^m - I^m K_1^m)^+ - (L_t^m - I^m K_2^m)^+,$$

where L_t^m is the aggregate loss of the underlying portfolio of the m-th tranche swap. The factor I^m is the notional of the underlying portfolio of the m-th tranche swap. The quantities K_1^m, K_2^m are the corresponding attachment and detachment points of the m-th tranche swap.

The cumulative loss in the outer tranche $[K_1, K_2]$ is

$$L_{[K_1,K_2]}(t) = (L_t^{outer} - IK_1)^+ - (L_t^{outer} - IK_2)^+. \tag{2.28}$$

The present value of default leg DL is

$$DL = E_t\left[\sum_{t_m>t} D(t,t_m)(L_{[K_1,K_2]}(t_m) - L_{[K_1,K_2]}(t_{m-1}))\right] \tag{2.29}$$

$$= \sum_{t_m>t} D(t,t_m)[(E_t(L(t_m)-K_1)^+ - E_t(L(t_m)-K_2)^+) -$$

$$(E_t(L(t_{m-1})-K_1)^+ - E_t(L(t_{m-1})-K_2)^+)]. \tag{2.30}$$

The present value of premium leg PL is

$$PL = E_t\left[\sum_{t_m>t} S\delta_m D(t,t_m)(I(K_2-K_1) - L_{[K_1,K_2]}(t_m))\right]. \tag{2.31}$$

So the fair par spread is

$$S = \frac{E_t\left[\sum_{t_m>t} D(t,t_m)(L_{[K_1,K_2]}(t_m) - L_{[K_1,K_2]}(t_{m-1}))\right]}{E_t\left[\sum_{t_m>t} \delta_m D(t,t_m)(I(K_2-K_1) - L_{[K_1,K_2]}(t_m))\right]}. \tag{2.32}$$

Remark:

The above four multi-name credit derivatives are usually regarded as static. Because the evaluation of the spread only depends on the marginal distribution of the aggregate loss L_t at the various coupon payment date $t = t_m$, the payoff of such multi-name credit derivatives is path independent. For the purpose of pricing, we only need a static model to evaluate the expectations involving L_t, or N_t if we assume name homogeneity. The dynamics of the time t distribution of aggregate loss is not necessary to consider.

The next three credit derivatives are dynamic in the sense that the payoff is path dependent.

4. **Forward starting STCDO**

A $T_1 \times (T_2 - T_1)$ forward-starting STCDO with attachment point K_1, detachment point K_2 is a contract, in which at time T_1 investors enters into a CDO over $[T_1, T_2]$ where the protection fee has already been fixed at the initial time $t < T_1$, at the level S^f. The twist is that the attachment and detachment are reshifted by any loss amounts that occur before T_1; that is, it turns into a STCDO with $[\tilde{K}_1, \tilde{K}_2] = [\frac{L_{T_1}}{I} + K_1, \frac{L_{T_1}}{I} + K_2]$ not simply $[K_1, K_2]$ (The latter is straightforward to price). The notional is also reduced to $\tilde{I} = I - L_{T_1}$. At T_1, the value of the position to the protection seller is the difference of the predetermined spread S^f and time T_1 realized spread S times the expected fee.

$$\begin{aligned} \text{Payoff} &= \Big[S^f - S\Big] \, \text{Fee}_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2) \\ &= \Big[S^f - S\Big] \, E_{T_1}\left[\sum_{T_1 < t_m \le T_2} \delta_m D(T_1, t_m)(\tilde{I}(K_2 - K_1) - L_{[\tilde{K}_1, \tilde{K}_2]}(t_m))\right], \end{aligned} \quad (2.33)$$

where $S = S_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2)$ is the time T_1 valued spread of an ordinary CDO tranche over $[T_1, T_2]$ with attachment and detachment $[\tilde{K}_1, \tilde{K}_2]$.

5. **Tranche options**

A put (payer) option is usually defined as the right to buy protection (that is, sell risk) and a call (receiver) option as the right to sell protection(that is, buy risk). The simplest tranche option could be as follows:

At maturity time T_1, the payoff of the European Put on a forward-starting STCDO is

$$\text{Payoff} = \Big(S^f - S\Big)^+ \cdot \text{Fee}_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2),$$

where $S = S_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2)$ is the realized spread at time T_1 and $\text{Fee}_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2)$ is the same as in (2.33). It is regarded as a put because the holder expects a decrease of market value of $S_{(\tilde{K}_1, \tilde{K}_2)}(T_1, T_2)$ and with S^f treated as predetermined strike.

6. **Leveraged super-senior tranches**

A Leveraged super-senior tranch can be treated as a variant of the STCDO in which the maximum loss to the investor is capped at a fraction of the super senior portion. In addition the leveraged super-senior structure typically includes loss triggers. The triggers are usually specified in terms of a hitting time of L_t or a hitting time of portfolio spread, or a combination upon trigger events. The protection seller makes a payment to the protection buyer which is linked to the value of a given STCDO.

2.2 Default counting process and intensities

All processes and random variable we discuss later on are defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{(t\geq 0)}$, where Ω represents the set of all possible states of nature. $\mathbb{P}$ is the probability measure that attaches probabilities to the events in Ω. The filtration $(\mathcal{F}_t)_{(t\geq 0)}$ represents the information available at time t. Throughout this dissertation, we assume that $(\mathcal{F}_t)_{(t\geq 0)}$ satisfies the usual conditions. [2]

We model the arrival of a default event as an unknown random time $\tau \in \mathbb{R}^+$. It is convenient to include ∞ in case default never occurs. Since naturally at the time of default, it is known that the event has occurred, τ is an $\mathcal{F}_t$ stopping time.

A stopping time is the mathematical description of one event. As introduced in [22], Point process we consider here is a collection of increasing stopping times $\tau_i, i \in \mathbb{N}$. We assume there are only a finite number of such points over any finite time horizon almost surely. This provides good enough mathematical framework in which to analyze the risk of multiple defaults. We turn this collection of time points into a stochastic process using the associated counting process,

$$N(t) := \sum_i 1_{\{\tau_i \leq t\}}$$

Where $N(t)$ counts the number of events that lie before t. If the τ_i are greater than zero almost surely, a sample path of $N(t)$ would be a step function that starts at zero and increases by one at each τ_i. In credit modeling, from the top-down point of view, we interpret the counting process $N(t)$ as the number of defaults up to time t. This is the key modeling objective, so later on we will use the whole machinery of stochastic analysis to model and analyze $N(t)$

[2] that means $(\Omega, \mathcal{F}, \mathbb{P})$ is complete; $\mathcal{F}_0$ contains all events of measure zero; $(\mathcal{F}_t)_{(t\geq 0)}$ is right continuous

2.2.1 Intensity and compensator

The counting process $N(t)$ is an increasing process, thus a submartingale. By the Doob-Meyer decomposition theorem, there exists an unique increasing process $A(t)$ such that $A(0) = 0$ and $N(t) - A(t)$ is a martingale. The process $A(t)$ is right continuous and predictable. The unique process $A(t)$ counteracts the upward trend in $N(t)$ and is therefore often called the *compensator*.

The compensator gives information about the probabilities of jumps over the next time step. We have

$$\mathbf{E}[N(t+\Delta t) - N(t)|\mathcal{F}_t] = \mathbf{E}[A(t+\Delta t) - A(t)|\mathcal{F}_t]$$

Over small time steps, if we assume that $N(t+\Delta t) - N(t)$ can only take value 1 or 0,

$$\begin{aligned}\mathbf{E}[A(t+\Delta t) - A(t)|\mathcal{F}_t] &= 1\mathbb{P}[N(t+\Delta t) - N(t) = 1|\mathcal{F}_t] + 0\mathbb{P}[N(t+\Delta t) - N(t) = 0|\mathcal{F}_t] \\ &= \mathbb{P}[N(t+\Delta t) - N(t) = 1|\mathcal{F}_t]\end{aligned}$$

If at time t, A has a jump of size ΔA, then the local jump probability over the next time instant is ΔA. Because the compensator is predictable, its increments over the next time step are known. Thus, the compensator is a running measure of the local jump probabilities of the counting process $N(t)$.

If $A(t)$ does not have discrete jumps, the probability of a jump of $N(t)$ in the next instance is infinitesimally small. If $A(t)$ is absolutely continuous, this leads to the concept of jump *intensity*

Definition 2.2.1. *The non-negative, progressively measurable process $\lambda(t)$ is the intensity of the counting process $N(t)$ if and only if*

$$A(t) := \int_0^t \lambda(s)ds$$

is the predictable compensator of $N(t)$, that is, $N(t) - A(t)$ is a martingale.

We assume that $A(t)$ is not only increasing, but also differentiable with derivative $\lambda(t)$. This is not an innocuous assumption. a firm's value-based structural model does not have a differentiable predictable compensator for the default indicator process. Due

to this issue, incomplete information models are of interest. However, our attention are focused on the class of models where a finite intensity exists. Such models are usually called *reduced form models* or *intensity based models.*

Theorem 2.2.2. *(Aven, 1985) Let* $\{\epsilon_n\}_{n=1}^{\infty}$ *be a sequence which decreases to zero and let* $Y_n(t)$, $t \in \mathbb{R}_+$ *be a measurable version of the process*

$$Y_n(t) := \frac{1}{\epsilon_n}\mathbf{E}\left[N(t+\epsilon_n) - N(t)|\mathcal{F}_t\right]$$

Assume there are non-negative and measurable process $g(t)$ *and* $y(t)$,$t \in \mathbb{R}_+$ *such that (i) for each* t

$$\lim_{n\to\infty} Y_n(t) = g(t) \qquad a.s.$$

(ii) for each t *there exists for almost all* $\omega \in \Omega$ *an* $n_0 = n_0(t,\omega)$ *such that*

$$|Y_n(s,\omega) - g(s,\omega)| \leq y(s,\omega), \qquad \forall s \leq t, n \geq n_0,$$

(iii)

$$\int_0^t y(s)ds < \infty \quad a.s. \quad t \in \mathbb{R}_+,$$

then $N(t) - \int_0^t g(s)ds$ *is a local martingale, and* $\int_0^t g(s)ds$ *is the compensator of* $N(t)$.

From this theorem, we are able to derive the default intensities if we are able to compute

$$\lim_{n\to\infty} \frac{1}{\epsilon_n}\mathbf{E}\left[N(t+\epsilon_n) - N(t)|\mathcal{F}_t\right]$$

This is usually the case when the conditional survival probabilities are known. Much of the analysis in the rest of this dissertation is devoted to the problem of finding realistic specifications for the default intensity.

2.2.2 Time-inhomogeneous Markov chain

A continuous-time Markov chain is a stochastic process $X(t) : t \geq 0$ that satisfies the Markov property and takes values in a discrete state space S . The Markov property states that at any times $s > t > 0$, the probability distribution of $X(s)$ conditioning on the whole history of the process up to and including time t depends only on the state of the process at time t. In effect, the state of the process at time s is conditionally independent of the history of the process before time t, given the state of the process at time t.

Definition 2.2.3 (Transition probability). *Suppose the state space is finite with elements indexed from* 1 *to* N. *then the* transition probability matrix $P(t,T)$ *for the time interval* $[t,T]$ *is*

$$P(t,T) = \begin{pmatrix} p_{11}(t,T) & p_{12}(t,T) & \cdots & p_{1N}(t,T) \\ p_{21}(t,T) & p_{22}(t,T) & \cdots & p_{2N}(t,T) \\ \cdots & \cdots & \ddots & \cdots \\ p_{N1}(t,T) & p_{N2}(t,T) & \cdots & p_{NN}(t,T) \end{pmatrix}.$$

For all $i,j \in S$, *the component* $p_{ij}(t,T) \geq 0$ *is the probability that* X *changes to state* j *at time* T, *given that it was in state* i *at time* t*:*

$$p_{ij}(t,T) = \mathbf{P}\left[X(T) = j | X(t) = i\right] \qquad \forall i,j \in S, t \leq T.$$

As a remark, for conservation of probability, we must have $\sum_{i=1}^{N} p_{ij}(t,T) = 1$ for all i. And over zero time, no transition takes place: $P(t,t) = I$, where I is the identity matrix. Also notice that since $X(t)$ is Markovian, $\mathbf{P}\left[X(T) = j | X(t) = i\right] = \mathbf{P}\left[X(T) = j | X(t) = i, \mathcal{F}_t\right]$

The probability transition matrix $P(t,T)$ satisfies the well known Chapman-Komogorov equation

$$P(t,T) = P(t,s)P(s,T) \qquad \forall t \leq s \leq T.$$

For the continuous time setup, we assume that $P(t,T)$ is continuous in T at $T = t$. This amounts to assuming that transitions [3] come as surprises, there is no scheduled

[3] Transition will be interpreted as default event later on in the credit model

time at which a transition will happen with positive probability. At such a point in time a discontinuity would occur in $P(t,T)$. It can be shown (in [21]) that in this case $P(t,T)$ is differentiable with respect to T at $T=t$. For small time intervals Δt the transition probability matrix can be approximated by a Taylor series. Similar to the construction of the Poisson process, we assume that the transition probability from state i to j in the small time interval Δt is approximately proportional to Δt:

$$\mathbf{P}(X(t+\Delta t)=j|X(t)=i)=\lambda_{ij}(t)\Delta t, \quad \forall i\neq j.$$

The diagonal elements $\lambda_{jj}(t)$ are defined by $\lambda_{jj}(t)=-\sum_{i\neq j}\lambda_{ij}(t)$. Since the transition probabilities must sum up to 1, therefore,

$$\begin{aligned}\mathbf{P}(X(t+\Delta t)=j|X(t)=j) &= 1-\sum_{i\neq j}\lambda_{ij}(t)\Delta t \\ &= 1+\lambda_{jj}(t)\Delta t.\end{aligned}$$

Then if we consider the matrix $\Lambda(t)=(\lambda_{ij})_{1\leq i,j\leq N}$, the transition probability matrix for the time interval $[t,t+\Delta t]$ is

$$P(t,t+\Delta t)=I+\Delta t\Lambda(t)+o(\Delta t^2).$$

Definition 2.2.4. *The matrix $\Lambda(t)$ is called the generator matrix of the continuous Markov chain $X(t)$ if it exists*

If Λ is a constant matrix, the Markov chain is time-homogeneous. Usually, in credit modeling, we don't make such assumption and let Λ be time dependent, in which case $X(t)$ is in fact a time-inhomogeneous Markov chain. In addition, the transition probability matrix $P(t,T)$ is related with $\Lambda(t)$ through the Kolmogorov forward and backward equations.

$$\text{Forward equation:} \quad \frac{\partial}{\partial T}P(t,T)=P(t,T)\Lambda(T), \qquad P(t,t)=I, \tag{2.34}$$

$$\text{Backward equation:} \quad \frac{\partial}{\partial t}P(t,T)=-\Lambda(t)P(t,T), \qquad P(T,T)=I. \tag{2.35}$$

If the generator matrix $\Lambda(t)=\Lambda=const$, these differential equations have solution

$$P(t,T)=\exp\{(T-t)\Lambda\}:=\sum_{n=0}^{\infty}\frac{1}{n!}((T-t)\Lambda)^n.$$

If the generator matrices $\Lambda(t)$ and $\Lambda(s)$ commutes for all $t \neq s$, then the solution above generalizes to the time-inhomogeneous case:

$$P(t,T) = \exp\{\int_t^T \Lambda(s)ds\}.$$

2.2.3 Itô Doeblin's formula for jump processes

The Itô's lemma is a very powerful tool to derive pricing equations for derivative securities. Here we give Itô's lemma for an enough general class of stochastic processes, semi-martingales with a finite number of jumps over finite time intervals a.s. All processes that we will encounter in this dissertation are members of this class of processes, that is, we will not consider examples of Lévy processes that have an infinite number of jumps over a finite time interval.

In most applications, the dynamics of X to be considered is a jump diffusion process of the form

$$dX_i = \alpha_i dt + \sum_{k=1}^{K} \sigma_{ik} dW_k + \int_{\mathbb{R}^n} h_i(x) \mu_X(dx, dt) \tag{2.36}$$

The process is driven by a K-dimensional Brownian motion W and the jump measure μ_X. The drift α_i, volatilities σ_i and jump sizes h_i are all predictable stochastic process. Furthermore, the jump measure μ_X of X has compensator measure ν_X which can be decomposed as

$$\nu_X(dx, dt) = K(t, dx) dA(t) \tag{2.37}$$

Theorem 2.2.5 (Itô's lemma for jump diffusion processes). *Let $X = (X_1, ..., X_n)$ be an n-dimensional jump diffusion process of the form (2.36), and let f be a twice continuously differentiable function on $\mathbb{R}^n$. Then $f(X)$ is also a semi-martingale, and:*

$$\begin{aligned} f(X(t)) = f(X(0)) &+ \sum_{i=1}^{n} \int_0^t \frac{\partial f(X(s-))}{\partial x_i} \alpha_i dt + \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t \frac{\partial f(X(s-))}{\partial x_i} \sigma_{ik} dW_k \\ &+ \frac{1}{2} \sum_{i,j=1}^{n} \int_0^t \frac{\partial^2 f(X(s-))}{\partial x_i \partial x_j} (\sigma \sigma^T)_{ij} dt \\ &+ \int_0^t \int_{\mathbb{R}^n} f(X(s-) + x) - f(X(s-)) \mu_X(dx, ds) \end{aligned} \tag{2.38}$$

Proof. A formal proof of this theorem can be found in [14]. □

The only difference between theorem 2.2.5 and the familiar form of Itô lemma for continuous processes is the jump term $\Delta f = \lim_{h \to 0} f(X(t+h) - X(t))$.

2.3 Intensity based Top-down models

In this section, we give brief introduction of intensity based Top-down model. It is in contrast to the Bottom-up model. In a Bottom-up model, the portfolio intensity is modelled by specifying both the intensities of individual constituent and the dependency structure among them. Such specification is appropriate for the analysis of portfolios of highly heterogeneous constituents. It brings the information of the single-name market to bear on the calibration of the model. Bottom-up modeling has been popular for the early days of credit derivatives because of its instinctively clear structure and computationally tractability. However, The dependence among single defaults is partly arbitrary and modelling all possible interactions amongst names leads to a huge number of parameters which results in high dimensional hurdle. Hence, consistent calibration is usually difficult or practically impossible.

Related literature includes Papageorgiou & Sircar [8], Frey & Backhaus [10], Duffie & Garleanu [6]

2.3.1 Top-down approach

The core problem in multi-name credit derivative modeling is that of pricing. That is we need to find the no-arbitrage price of multi-name credit derivative, including index CDS, CDOs, and so on.

As we can tell from the pricing formulae (2.20) and (2.26) in section 2.1.2, the value of index CDS and CDOs only depends on the distribution of the portfolio loss process $L(t)$ at multiple coupon payment dates t. The idea of the Top-down approach is to specify the dynamics of the cumulative loss L_t or look at the forward loss distribution without reference to the constituents.

The cumulative loss process L_t is a piecewise constant process with upward jumps at each default event. Its sample path is therefore completely characterized by the default times $\{\tau_i\}_{i\geq 1}$, representing default events and jump sizes $\Delta L(\tau_i)$ representing the loss given default (LGD).

Here τ_j denotes the j-th default event event observed in the portfolio: the subscript

j is not associated with the default of a given obligor indexed by j, but with the ordering in time of the events. The main idea in top-down models is to investigate the default events sequence $\{\tau_i\}_{i\geq 1}$ by modeling the rate of occurrence of defaults in the portfolio via the portfolio default intensity λ_t.

While for the loss given defaults $\Delta L(\tau_i)$. we can deal with it by the name homogeneity assumption:

Assumption 2.3.1 (Name homogeneity and constant recovery rate). *The notional I_i and recovery rate R_i are constant and equal across all names. $I = I_i\bar{N}$,*

$$L(t) = \frac{N(t)}{\bar{N}} I(1-R)$$

Therefore, $L(t)$ and $N(t)$ differ only by a constant. and

$$\Delta L(\tau_i) = \frac{I(1-R)}{\bar{N}}$$

So under the name homogeneity assumption, it is enough to consider the portfolio default counting process $N(t)$. This counts the number of default obligors in the portfolio. All names are inter-changeable. The name dependence structure is implicit in this specification. It does not matter which name defaults. This is a typical assumption in valuation of multi-name credit derivatives like CDO.

The aggregate counting process $N(t)$ is increasing. By the Doob-Meyer decomposition theorem, there exists a predictable compensator $A(t)$ such that $N(t) - A(t)$ is a martingale. We in addition assume that $A(t)$ is absolutely continuous, that is, there exists a positive process $\lambda(t)$ such that

$$A(t) = \int_0^t \lambda(s)ds.$$

By Aven's Theorem (1985): Under mild condition, we have

$$\lambda(t) = \lim_{h \searrow 0} \frac{1}{h} E[N(t+h) - N(t)|\mathcal{F}_t].$$

Since we don't consider multiple defaults during a sufficiently short time period, that is, the probability that the number of defaults larger than one is $o(\Delta t)$,

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(N(t+\Delta t) - N(t) = 1|\mathcal{F}_t)}{\Delta t}$$

where $\mathcal{F}_t$ denotes the history of the point process $N(t)$. Thus $\lambda(t)$ measures the likelihood of the next default. It is the key ingredient in modeling the jump process $N(t)$.

The simplest approach is to let $\lambda(t)$ be deterministic, then $N(t)$ is just a time-dependent Poisson process.

2.3.2 Markovian loss model

The simplest approach beyond deterministic intensities is to let $\lambda(t)$ depend on both t and current loss $N(t)$; i.e.,

$$\lambda(t) = f(t, N(t)) \tag{2.39}$$

Then $N(t)$ is a time-dependent Markov chain. We call this model a *Markovian loss model*. Such a specification of intensity $\lambda(t)$ is analogous to the *local volatility* in equity modeling, in which we assume the volatility has the form $\sigma(t, S(t))$ and the asset $S(t)$ has the dynamics

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

The Markovian loss model is simple but still exhibits enough flexibility to capture the default dependence, which arises from the transitions of the default counting process and the corresponding changes in the transition rate.

Example 2.3.2. The simplest model is the *linear counterparty risk* model, where

$$f(t, n) = \lambda_0 + \alpha n \qquad \nu > 0,\ \alpha \geq 0,\ l \in \{0, 1, ..., \bar{N}\} \tag{2.40}$$

The interpretation of (2.40) is that the first default arrives at a basis rate λ_0 and all the following defaults increases the default intensity of surviving firms by a constant amount α. A jump process with such specification of intensity is also called a birth process. If $\alpha = 0$, the aggregate loss process is simply a Poisson process. In the linear counterparty risk model, the distribution of N_t has an analytical expression:

$$P[N_t = k] = \frac{\Gamma(\frac{\lambda_0}{\alpha} + k)}{\Gamma(\frac{\lambda_0}{\alpha})k!} e^{-\lambda_0 t}(1 - e^{-\alpha t})^k.$$

The mean and variance of N_t are

$$E[N_t] = \frac{\lambda_0}{\alpha}(e^{\alpha t} - 1),$$

$$Var[N_t] = \frac{\lambda_0}{\alpha}(e^{\alpha t} - 1)e^{\alpha t}.$$

Example 2.3.3. Frey, Rüdiger and Backhaus generalized the above model to *convex counterparty risk*, where

$$f(t,n) = \lambda_0 + \frac{\alpha^2}{\beta}\left(e^{\frac{\beta}{\alpha}(n-\mu(t))^+} - 1\right) \qquad \lambda_0 > 0,\ \alpha, \beta \geq 0,\ n \in \{0, 1, ..., \bar{N}\}. \quad (2.41)$$

In this model, the default intensity of the surviving firms increases away from λ_0 only if $N(t)$ exceeds some deterministic threshold $\mu(t)$ measuring the expected number of defaulted firms up to time t. The quantity λ_0 is again the basis intensity and α plays the same role as in (2.40). It gives the slope of $f(t,n)$ with respect to n at $\mu(t)$ so that it indicates the strength of intensity increase upon default. The new parameter β controls the degree of convexity of f. It gives the second derivative of f with respect to n. It implies that a large value of $(N(t) - \mu(t))^+$ leads to a high value of default intensity increase. In the limit as β approaches zero, (2.41) reduces to the linear model (2.40).

We do not assume that the Markovian loss model is the real dynamics of $N(t)$. However for any general counting process $N(T)$, starting from t, there always exists a Markov point process $\tilde{L}(T)$ (that is, $\tilde{L}(T)$ is a time-dependent Markov chain) such that $\tilde{L}(T)$ and $N(T)$ have the same $\mathcal{F}_t$-marginal distribution; That is, for any $T > t$,

$$\mathbb{P}(N(T) \leq x|\mathcal{F}_t) = \mathbb{P}(\tilde{L}(T) \leq x|\mathcal{F}_t).$$

We will discuss this in detail in the following sections.

2.3.3 Stochastic intensity modeling

Empirically, the default counting process $N(t)$ is not Markovian. The process $\lambda(t)$ should have a more complex structure. So instead of assuming the dependency of $\lambda(t)$ on t and $N(t)$ only, we can treat $\lambda(t)$ itself as a stochastic process. The following benchmark stochastic intensity models are well known.

Doubly stochastic intensity

The process λ_t has its own dynamics, driven by some other "market factor", and independent of the realized number of defaults. For example, we may choose Itô diffusion dynamics for λ_t,

$$d\lambda_t = \mu(t, \lambda_t)dt + \sigma(t, \lambda_t)dW_t. \tag{2.42}$$

Then $N(t)$ is a Cox process; conditional on the realization of λ_t, $N(t)$ is a time-inhomogeneous Poisson process.

Hawkes process

The intensity of the self exciting Hawkes process N_t (in [19]) takes the form

$$\lambda(t) = \nu + \sum_{s_i \leq t} \alpha e^{-\beta(t-s_i)}, \tag{2.43}$$

where $\{s_i\}$ are the jump times of the counting process N_t.

In short, the occurrence of one default generates a sudden increase α in the intensity of the default counting process N_t and this increase decays exponentially with rate β. If β tends to infinity, the Hawkes process degenerates to the Poisson process with rate ν. If β tends to zero, the Hawkes process degenerates to the birth process (2.40). Under Hawkes' specification, the intensity involves no other randomness, but depends on the path of $N(t)$

Bivariate spread-loss model (Igor Halperin)

This model assumes that the default intensity is given by

$$\lambda_t = Y_t f(N_t)(N - N_t)$$

where $f(\cdot)$ is a deterministic function and Y_t contains other random factors independent of N_t.

Two-dimensional Markovian model (Lopatin, Misirpashaev)

This model assumes that the default intensity is given by

$$d\lambda_t = \kappa(\rho(t, N_t) - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t.$$

The intensity satisfies a mean reverting dynamics, and the long term mean ρ depends on time t and number of default N_t.

The default counting process $N(t)$ is usually not Markovian. So non-Markovian models have more realistic and flexible dynamics. However, non-Markovian models are usually not mathematically tractable, and few analytical properties can be derived. The computation involved in pricing problem greatly relies on numerical methods. We can think about the stochastic volatility models in the equity case for the sake of comparison.

2.4 Local volatility surface in equity modeling

We aim to explore the approaches to credit modeling inspired by the perspective of equity derivative modeling and then provide new methods and address new questions suggested by equity derivative modeling. In this section, we give a short summary of the theory of local volatility for equity derivatives modeling first introduced by Dupire [7] and also the idea of modeling the dynamics of local volatility surface advocated by Carmona [3].

Gyöngy's theorem

The local volatility model is strongly connected with the early work of Gyöngy in his article on mimicking the marginal distribution of a continuous semi-martingale [12].

Theorem 2.4.1. *(Gyöngy's theorem) Let $X(t)$ be an n-dimensional Itô process which satisfies*

$$dX(t) = \beta(t)dt + \sigma(t)dW_t, \tag{2.44}$$

where W_t is a d-dimensional Brownian motion. $\beta(t)$ is a bounded n-dimensional adapted process and $\sigma(t)$ is a bounded $n \times d$-dimensional adapted process such that $\sigma(t)\sigma^T(t)$ is uniformly positive definite. Then there exist measurable functions a and b

$$\begin{aligned} a^2(t,x) &= E[\sigma(t)\sigma^T(t)|X(t) = x] \\ b(t,x) &= E[\beta(t)|X(t) = x], \end{aligned}$$

and there exists a weak solution $Y(t)$ to the SDE:

$$dY(t) = b(t, Y(t))dt + a(t, Y(t))d\widetilde{W}_t, \tag{2.45}$$

where $\widetilde{W}_t$ is a n-dimensional Brownian motion, and $X(0) = Y(0)$, such that X and Y have the same marginal distributions for all $t > 0$.

This theorem states that for any Itô process X_t of type (2.44), there is a Markov process $\widetilde{X}_t$ with deterministic (local) drift and volatility coefficients of type (2.45) that "mimic" the marginal distribution of the given process X_t for all $t > 0$.

Local volatility model

To establish the link between Gyöngy's theorem and the local volatility model, we consider a stock price process S_t, whose dynamics is governed by the stochastic differential equation

$$\frac{dS_t}{S_t} = rdt + \sigma(t,\omega)dW_t, \tag{2.46}$$

where W_t is the driven Brownian motion under a certain risk neutral measure, and the process $\sigma(t,\omega)$ is the stochastic volatility, which is a bounded 1-dimensional adapted process and $\sigma\sigma^T$ is uniformly positive definite. Denote $X_t = \ln \frac{S_t}{S_0}$. Then by Itô's lemma, we have

$$dX_t = (r - \frac{1}{2}\sigma^2(t,\omega))dt + \sigma(t,\omega)dW_t. \tag{2.47}$$

According to Gyöngy's theorem, we can have a Markov process $\widetilde{X}_t$ with dynamics

$$d\widetilde{X}_t = \widetilde{\mu}(t,\widetilde{X}_t)dt + \widetilde{\sigma}(t,\widetilde{X}_t)d\widetilde{W}_t, \quad \widetilde{X}_0 = 0,$$

and

$$\widetilde{\mu}(t,x) = E[r - \frac{1}{2}\sigma^2(t,\omega)|X_t = x] \tag{2.48}$$

$$\widetilde{\sigma}^2(t,x) = E[\sigma^2(t,\omega)|X_t = x] \tag{2.49}$$

We now define $\widetilde{S}_t = S_0 e^{\widetilde{X}_t}$. Since $S_t = S_0 e^{X_t}$ and $\widetilde{X}_t$ has the same marginal distribution as X_t, it is clear that $\widetilde{S}_t$ has the same marginal distribution as S_t. In order to find volatility of $\widetilde{S}_t$, we apply Itô's lemma again, we obtain

$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = \left[\widetilde{\mu}(t,\widetilde{X}_t) + \frac{1}{2}\widetilde{\sigma}^2(t,\widetilde{X}_t)\right]dt + \widetilde{\sigma}(t,\widetilde{X}_t)d\widetilde{W}_t \tag{2.50}$$

Note that

$$\begin{aligned}
&\widetilde{\mu}(t,\widetilde{X}_t) + \frac{1}{2}\widetilde{\sigma}^2(t,\widetilde{X}_t) \\
&= \left[E[r - \frac{1}{2}\sigma^2(t,\omega)|\widetilde{X}_t] + E[\frac{1}{2}\sigma^2(t,\omega)|\widetilde{X}_t]\right] \\
&= r
\end{aligned}$$

Then equation (2.50) becomes

$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = rdt + \widetilde{\sigma}(t, \widetilde{X}_t)d\widetilde{W}_t \tag{2.51}$$

We define $\sigma_L(t,S) = \widetilde{\sigma}(t, \ln(\frac{S}{S_0}))$. Then

$$\sigma_L(t,S) = E\left[\sigma^2(t,\omega)|X_t = \ln(\frac{S}{S_0})\right] = E[\sigma^2(t,\omega)|S_t = S].$$

Replacing $\widetilde{\sigma}$ by σ_L, we have

$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = rdt + \sigma_L(t, \widetilde{S}_t)d\widetilde{W}_t \tag{2.52}$$

Thus we obtain the local volatility model (2.52). The process $\widetilde{S}_t$ is the mimicking process having the same marginal distribution of S_t in (2.46) for all $t > 0$. The local volatility of $\widetilde{S}_t$ is defined by

$$\sigma_L^2(t,S) = E[\sigma^2(t,\omega)|S_t = S].$$

Moreover, there is a one-to-one correspondence between the risk neutral marginal distribution and the prices of European-style vanilla options. The local volatility model (2.52) produces the same prices for European-style calls and puts as those produced by the model (2.44).

Backward equation

The price of any European style derivative C_t is the conditional expectation of final pay-off at maturity T of a certain risk neutral measure $C_t = E[h(S_T)|\mathcal{F}_t]$. If the underlying stock S_t is from a local volatility model

$$\frac{dS_t}{S_t} = rdt + \sigma_L(t, S_t)dW_t$$

Where S_t is a one-dimensional Markov process, then by Markov property,

$$C_t = E[e^{-rT}h(S_T)|S_t] = C(t, S_t)$$

the function $C(t,x)$ satisfies the well known Kolmogorov backward equation,

$$\begin{cases} c_t(t,x) + rxv_x(t,x) + \frac{1}{2}\sigma_L^2(t,x)x^2c_{xx}(t,x) - rc(t,x) = 0, \\ c(T,x) = h(x) \end{cases} \tag{2.53}$$

If the payoff function $h(x) = (x - K)^+$, equation (2.53) leads to the Black Scholes formula. For other payoff functions h, equation (2.53) can at least be solved by appropriate numerical methods.

Dupire's formula

If we consider an European-style call option with strike K and maturity T, then its price at time t with spot price x can be found by solving equation (2.53). Conversely, if the European-style call option prices $C(T, K)$ at a given initial time t and spot price x are known as functions of $C(T, K)$, then Dupire's formula gives $\sigma_L(T, K)$

$$\sigma_L^2(T, K) = \frac{\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K}}{\frac{1}{2}K^2\frac{\partial^2 C}{\partial K^2}} \tag{2.54}$$

when those call option prices $C(T, K)$ are quoted for all maturities and strikes [4].

Dynamics of local volatility surface

In Dupire's local volatility model, we treat the local volatility function $\sigma_L(\cdot, \cdot)$ as a surface $\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$. The function $\sigma_L(\cdot, \cdot)$ is regarded as fixed. Local volatility models are widely used in practice because they enable fast and accurate pricing of derivatives when only marginal distributions are required. However, the local volatility process does not give a complete representation of the true stochastic dynamics driving the underlying asset price. The local volatility model is merely a simplification that is useful in practice for describing a price process with non-constant volatility. More precisely, although the marginal distributions of S_t and $\widetilde{S}_t$ are the same at the time when the local volatility is calibrated, they do not follow the same dynamics and therefore they do not necessarily have the same joint distributions for all time $t > 0$. Hence more exotic path dependent option prices computed with S_t and $\widetilde{S}_t$ obtained will differ significantly. More accurate models will have to take the changing of local intensity surface into account.

[4] This is not a realistic assumption, since in reality, option prices are quoted only for a small number of maturities and strikes

Carmona and Nadtochiy [3] considered the dynamics of this local volatility surface $\widehat{\sigma}_t(\cdot,\cdot)$. There is no static arbitrage as long as local volatility is positive, but in order to avoid dynamic arbitrage, the differential equations governing the dynamics of local volatility surface have to satisfy a very complex drift condition.

Finally, since we will discuss similar problems in credit modeling, we list the comparison of equity modeling and credit modeling in the following table.

Equity modeling	Credit modeling
underlying: S_t	underlying: N_t
volatility $\sigma(t)$	intensity $\lambda(t)$
local volatility $\sigma(t) = \sigma(t, S_t)$	local intensity $\lambda(t) = \lambda(t, N_t)$
stochastic volatility $d\sigma(t) = \cdots$	stochastic intensity $d\lambda(t) = \cdots$

Table 2: Comparison of equity modeling and credit modeling.

Chapter 3

Local intensity surface

3.1 Local intensity surface

In this section we will define the notion of a local intensity surface and examine some basic properties of this surface.

For the Markovian loss model, the default counting process N_t is characterized by its aggregate intensity λ_t through the specification

$$\lambda_t = f(t, N_t), \tag{3.1}$$

where the bivariate deterministic function f is usually assumed to be at least C^1 continuous in the first argument t. The process $N(t)$ takes the integer values from 0 to $\bar{N}$ (the total number of names in the portfolio).

Under this specification, N_t is a continuous time inhomogeneous Markov chain as in §2.2.2 with finite state space $\{0, 1, 2, \cdots, \bar{N}\}$. The generator Q-Matrix $A(t)$ is bidiagonal because we only allow N_t to jump by one at each default event. Namely, simultaneous multiple defaults are not allowed at this point. So $A(t)$ takes the form

$$A(t) = \begin{pmatrix} -a_0(t) & a_0(t) & & \\ & -a_1(t) & a_1(t) & \\ & & \cdots & \\ & & -a_{N-1}(t) & a_{N-1}(t) \\ 0 & 0 & \cdots & 0, \end{pmatrix} \tag{3.2}$$

where $a_n(t)$ is defined as

$$a_n(t) = \lim_{h \to 0} \frac{1}{h} \mathbb{P}\left[N(t+h) = n+1 | N(t) = n\right]$$

It's clear that the transition rate $a_n(t)$ from n to $n+1$ is just $f(t,n)$. Let $P(t,T)$ be the transition probability matrix. It satisfies Kolmogorov forward equation:

$$\frac{\partial}{\partial T}P(t,T) = P(t,T)A(T). \tag{3.3}$$

That is

$$\frac{\partial}{\partial T}P_{nm}(t,T) = -a_m(T)P_{nm}(t,T) + a_{m-1}(T)P_{n,m-1}(t,T), \tag{3.4}$$

with initial condition $P_{nm}(t,t) = 1_{\{n=m\}}$. This equation can be solved iteratively, using

$$P_{nm}(t,T) = \begin{cases} 0 & m < n, \\ e^{-\int_t^T a_n(s)ds} & m = n, \\ \int_t^T P_{n,m-1}(t,s)a_{m-1}(s)e^{-\int_s^T a_m(u)du}ds & m > n. \end{cases} \tag{3.5}$$

Definition 3.1.1 (Local intensity surface). *We define the function*

$$f : \mathbb{R}^+ \times \mathbb{Z}_{\geq 0} \to \mathbb{R}_{>0}$$

to be the ***local intensity surface*** *of the counting process N_t in model (3.1).*

Note that since the second argument n only takes integer values, the local intensity surface is not really a surface but a sequence of curves $f(t,n)$ indexed by n. The function f vanishes for $n > \bar{N}$, consistent with the fact that the default counting process N_t is stopped at level $\bar{N}$. Without causing any confusion, we still call it a local intensity surface in the remainder of this dissertation.

It is readily seen that the local intensity surface is just the geometric name for the function f which characterizes the Markovian process N_t. All the probability information of N_t is encapsulated in the shape of the surface. Namely, the probability distribution of N_t is completely determined by the local intensity surface. However, the above definition only applies to the Markovian model. Actually, as we will see later, we can generalize the notion of local intensity surface to non-Markovian model, in which case, N_t can have any adapted intensity process λ_t. But this generalization will be made after we introduce an important result which we will discuss in section 3.3.

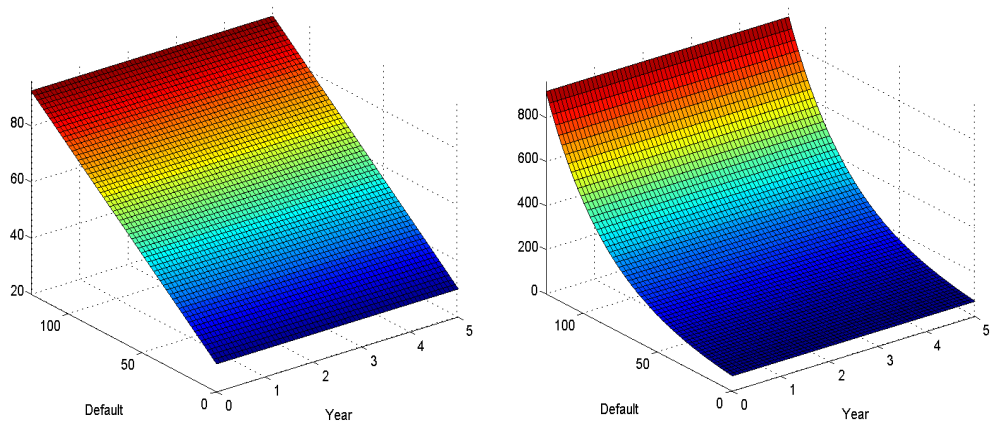

Figure 3.1: The local intensity surface of two local intensity model. left: Linear Counterparty risk model (2.40) $f(t,n) = \lambda_0 + \alpha n$ with $\lambda_0 = 30$, $\alpha = 0.5$. right: Linear Counterparty risk model (2.41) $f(t,n) = \lambda_0 + \frac{\alpha^2}{\beta}\left(e^{\frac{\beta}{\alpha}(n-\mu(t))^+} - 1\right)$ with $\lambda_0 = 30$, $\alpha = 1$, $\beta = 0.025$, $\mu(t) = 0$

For a Markovian loss model, the local intensity $\Lambda(t,n)$ has to satisfy the following formula, which is an analog of Dupire's formula in the local volatility modeling.

Lemma 3.1.2 (Dupire's formula). *Suppose there is a counting process $\tilde{N}_t$, with local intensity $\Lambda(t,\tilde{N}_t)$. Then $\Lambda(t,n)$ satisfies the equation,*

$$\Lambda(t,n) = -\frac{1}{p_n(t)}\sum_{k=0}^{n}\frac{\partial}{\partial t}p_k(t), \tag{3.6}$$

where $p_n(t) = \mathbb{P}(\tilde{N}_t = n)$

Proof. By definition of local intensity,

$$\Lambda(t,n) = \lim_{h\to 0}\frac{1}{h}\mathbb{P}\left[\tilde{N}(t+h) = n+1 | \tilde{N}(t) = n\right],$$

where $\tilde{N}_t$ is a time-inhomogeneous Markov chain, therefore we have the following Kolmogorov forward equation

$$\begin{cases} \frac{\partial}{\partial t}p_k(t) = -\Lambda(t,k)p_k(t), & k = 0, \\ \frac{\partial}{\partial t}p_k(t) = -\Lambda(t,k)p_k(t) + \Lambda(t,k-1)p_{k-1}(t), & k \geq 1. \end{cases} \tag{3.7}$$

Add the preceding equations for k ranging from 0 to n to obtain

$$\sum_{k=0}^{n}\frac{\partial}{\partial t}p_k(t) = -p_n(t)\Lambda(t,n).$$

This proves (3.6). □

We call (3.6) a *"Dupire" formula* because the terms appearing on the right side of (3.6) come from the marginal distribution, which can in principle be determined by the market.

3.2 Backward and forward equation

3.2.1 Backward equation

From section 2.1.2, we know from (2.24) that if we assume name homogeneity, in order to find the price of a CDO tranche, it suffices to evaluate $E[(N_T - K)^+]$ for all coupon payment dates T and constants K. This expectation can be regarded as the initial price of a European call option on the underlying N_t with strike K and maturity date T, (assuming zero risk free rate). The time t value of such an option has the form $E[(N_T - K)^+|\mathcal{F}_t]$. In the framework of the Markovian loss model, the underlying process N_t is a time-inhomogeneous Markovian chain. Then $E[(N_T - K)^+|\mathcal{F}_t] = E[(N_T - K)^+|N_t]$ This expectation only depends on t and N_t. Denote it as $c(t, N_t)$. The function $c(t, n)$ satisfies the following backward equation.

Theorem 3.2.1. *If N_t is a time-inhomogeneous Markov chain with local intensity surface $\Lambda(t, N_t)$, denote $a_n(t) = \Lambda(t, n)$. Then c satisfies the backward equation and terminal condition*

$$\begin{cases} \frac{\partial}{\partial t}c(t,n) + a_n(t)(c(t,n+1) - c(t,n)) = 0, \\ c(T,n) = (N_T - n)^+, \end{cases} \tag{3.8}$$

for all $0 \leq n \leq \bar{N}$ and $0 \leq t \leq T$.

Proof. By Itô-Doeblin's formula for a jump process (2.2.5)

$$c(t, N_t) = c(0, N_0) + \int_0^t c_t(u, N_u)du + \int_0^t c_x(u, N_u)dN_u^c + \sum_{0 \leq u \leq t} \Big[c(u, N_u) - c(u, N_{u-})\Big] \tag{3.9}$$

N_t is a pure jump process and hence $N_t^c = 0$. $c(t, N_t) - c(t, N_{t-})$ is associated with the jump in the price process when a default occurs in the credit portfolio. The jump

intensity of N is $\lambda(t) = \Lambda(t, N_t) = a_{N_t}(t)$. Define $M_t = N_t - \int_0^t \lambda(s)ds$, it is a $\mathcal{F}_t$-martingale. $dM_t = dN_t - \lambda_t dt$. So

$$\begin{aligned} c(t, N_t) &= c(0, N_0) + \int_0^t c_t(u, N_u)du + \int_0^t \Big[c(u, N_{u-} + 1) - c(u, N_{u-})\Big]dN_u \\ &= c(0, N_0) + \int_0^t \Big[c_t(u, N_u) + \Lambda(u, N_u)(c(u, N_{u-} + 1) - c(u, N_{u-}))\Big]du \\ &\qquad + \int_0^t \Big[c(u, N_{u-} + 1) - c(u, N_{u-})\Big]dM_u. \end{aligned} \tag{3.10}$$

Since M_t is an martingale and the function $c(u, N_{u-} + 1) - c(u, N_{u-})$ is left continuous, the last integral in (3.10) is a martingale. Also, $c(t, N_t) = E[(N_T - K)^+|\mathcal{F}_t]$ is a martingale. This implies $\int_0^t \Big[c_t(u, N_u) + \Lambda(u, N_u)(c(u, N_{u-} + 1) - c(u, N_{u-}))\Big]du$ is also a martingale, which forces the integrand to be zero for any value of u and N_u. Since jumps occur on a zero measure time set, the jump term in the above integrand can be replaced by $c(u, N_u + 1) - c(u, N_u)$. Finally, we obtain the backward equation

$$\frac{\partial}{\partial t}c(t, n) + a_n(t)(c(t, n + 1) - c(t, n)) = 0,$$

as desired □

The backward equation in (4.36) is actually a system of linear ordinary differential equations. To make notation simpler, we denote $c(t, n) = c_n(t)$ and $\frac{\partial}{\partial t}c(t, n) = c'_n(t)$. Then we obtain:

$$\begin{cases} c'_0(t) = -a_0(t)[c_1(t) - c_0(t)], & c_0(T) = (0 - K)^+, \\ c'_1(t) = -a_1(t)[c_2(t) - c_1(t)], & c_1(T) = (1 - K)^+, \\ \quad \cdots & \\ c'_{\bar{N}-1}(t) = -a_{\bar{N}-1}(t)[c_{\bar{N}}(t) - c_{\bar{N}-1}(t)], & c_{\bar{N}-1}(T) = ((\bar{N} - 1) - K)^+, \\ c'_{\bar{N}}(t) = 0, & c_{\bar{N}}(T) = (\bar{N} - K)^+, \end{cases}$$

This system can simply be solved iteratively by starting from the last equation. The general solution is

$$\begin{cases} c_n(t) = \bar{N} - K, & \text{if } n = \bar{N}, \\ c_n(t) = (n - K)^+ e^{-\int_t^T a_n(u)du} + \int_t^T a_n(s)c_{n+1}(s)e^{-\int_t^s a_n(u)du}ds, & \text{if } n < \bar{N}. \end{cases} \tag{3.11}$$

3.2.2 Forward equation

The backward equation is satisfied for the price of European call option on N_t considered as a function of the backward variable t and current loss level N_t, while maturity T and K are treated fixed. In this section, we show that if we fix a certain initial time and number of default, and treat the price as a function of T and K, then it satisfies a forward equation. Let's denote $C(T,K) = E[(N_T - K)^+]$. Here we use capital letter C in order to distinguish from the low case c in the backward equation. Let's also introduce the forward and backward difference operator $\nabla^{\pm} C(T,K) = C(T,K+1) \pm C(T,K)$. Then the option price C and distribution function of N_t is linked by the following relation.

Lemma 3.2.2. *Let $C(T,K)$ be price of European-style call option with Maturity T and strike K (nonnegative integer) at time t with $N_t \geq 0$. Then we have*

$$\nabla^+ C(T,K) = -\sum_{j=K+1}^{\bar{N}} p_j(T), \tag{3.12}$$

$$\nabla^- \nabla^+ C(T,K) = p_K(T), \tag{3.13}$$

where $p_K(T) = \mathbb{P}(N_T = K|\mathcal{F}_t)$.

Proof. Observe that

$$\begin{aligned}
\nabla^+ C(T,K) &= C(T,K+1) - C(T,K) \\
&= E[(N_T - (K+1))^+] - E[(N_T - K)^+] \\
&= \sum_{j=K+2}^{\bar{N}} (j - (K+1)) p_j(T) - \sum_{j=K+1}^{\bar{N}} (j-K) p_j(T) \\
&= \sum_{j=K+2}^{\bar{N}} -p_j(T) - p_{K+1}(T) \\
&= -\sum_{j=K+1}^{\bar{N}} p_j(T),
\end{aligned}$$

and

$$\begin{aligned}\nabla^-\nabla^+C(T,K) &= \nabla^-(C(T,K+1)-C(T,K))\\ &= [C(T,K+1)-C(T,K)]-[C(T,K)-C(T,K-1)]\\ &= -\sum_{j=K+1}^{\bar{N}} p_j(T) - (-\sum_{j=K}^{\bar{N}} p_j(T))\\ &= p_K(T).\end{aligned}$$

This concludes the proof. □

The preceding lemma shows that the European call option on N_t in fact determines the time T marginal distribution of N_T through (3.12). Since $a_n(T)$ is related to $p_k(T)$ through Dupire's formular (3.6), we can express the transition rate $a_n(T)$ in terms of $C(T,K)$, which in turns implies the forward equation.

Theorem 3.2.3. *If N_t is a time-inhomogeneous Markov chain with local intensity surface $\Lambda(t,N_t)$, denote $a_n(T)=\Lambda(T,n)$. Then $C(T,K)$ satisfies the forward equation and initial condition* (3.14)*,*

$$\begin{cases} \frac{\partial}{\partial T}C(T,K) = \sum_{j=K}^{\bar{N}} C(T,j)(\nabla^-\nabla^+a_j(T)) + a_K(T)C(T,K-1) - C(T,K)a_{K-1}(T) \\ C(t,K) = (N_t-n)^+. \end{cases} \tag{3.14}$$

for all $0\le K\le \bar{N}$ and $T\ge t$.

Proof. From Dupire's formula (3.6)

$$\begin{aligned}a_K(T) &= -\frac{1}{p_K(T)}\frac{\partial}{\partial T}\sum_{j=N(t)}^{K} p_j(T)\\ &= -\frac{1}{\nabla^-\nabla^+C(T,K)}\frac{\partial}{\partial T}(1-\sum_{j=K+1}^{\bar{N}} p_j(T))\\ &= -\frac{1}{\nabla^-\nabla^+C(T,K)}\frac{\partial}{\partial T}(1+\nabla^+C(T,K))\\ &= -\frac{1}{\nabla^-\nabla^+C(T,K)}\frac{\partial}{\partial T}(\nabla^+C(T,K))\end{aligned}$$

$$a_K(T)\nabla^-\nabla^+C(T,K) = -\frac{\partial}{\partial T}(\nabla^+C(T,K)) \tag{3.15}$$

We list the above equations for K ranging from K to $\bar{N}$

$$\begin{cases} a_K(T)\Big[C(T,K+1) - 2C(T,K) + C(T,K-1)\Big] & = -\frac{\partial}{\partial T}(C(T,K+1) - C(T,K)), \\ a_{K+1}(T)\Big[C(T,K+2) - 2C(T,K+1) + C(T,K)\Big] & = -\frac{\partial}{\partial T}(C(T,K+2) - C(T,K+1)), \\ \quad \cdots & \\ a_{\bar{N}-1}(T)\Big[C(T,\bar{N}) - 2C(T,\bar{N}-1) + C(T,\bar{N}-2)\Big] & = -\frac{\partial}{\partial T}(C(T,\bar{N}) - C(T,\bar{N}-1)), \\ a_{\bar{N}}(T)\Big[C(T,\bar{N}+1) - 2C(T,\bar{N}) + C(T,\bar{N}-1)\Big] & = -\frac{\partial}{\partial T}(C(T,\bar{N}+1) - C(T,\bar{N})). \end{cases} \tag{3.16}$$

Add the equations in (3.16) and notice that $C(T,\bar{N}) = 0$. Using the discrete version of integration by parts on the left side of the equations in (3.16), we have

$$\begin{aligned} \frac{\partial}{\partial T}C(T,K) &= \sum_{j=K+1}^{\bar{N}} C(T,j)(\nabla^-\nabla^+a_j(T)) \\ &\quad + C(T,K)(a_{K+1}(T) - 2a_K(T)) + a_K(T)C(T,K-1) \\ &= \sum_{j=K+1}^{\bar{N}} C(T,j)(\nabla^-\nabla^+a_j(T)) + C(T,K)(a_{K+1}(T) - 2a_K(T) + a_{K-1}(T)) + \\ &\quad + a_K(T)C(T,K-1) - C(T,K)a_{K-1}(T) \\ &= \sum_{j=K}^{\bar{N}} C(T,j)(\nabla^-\nabla^+a_j(T)) + a_K(T)C(T,K-1) - C(T,K)a_{K-1}(T). \end{aligned}$$

□

In addition to considering the European-style call option $C(T,K) = E[(N_T - K)^+]$, we note that the European-style put option $P(T,K) = E_t[(K - N_T)^+]$ satisfies a similar forward equation.

First, we note that there exists a similar result as lemma 3.2.2 to link the option price P and the distribution of N_t.

Lemma 3.2.4. *Let $P(T,K)$ be price of European-style call option with Maturity T and strike K (nonnegative integer) at time t with $N_t \geq 0$. Then we have*

$$\nabla^+ P(T,K) = \sum_{j=N(t)}^{K} p_j(T), \tag{3.17}$$

$$\nabla^- \nabla^+ P(T,K) = p_K(T), \tag{3.18}$$

where $p_K(T) = \mathbb{P}(N_T = K|\mathcal{F}_t)$.

Proof. Observe that

$$\begin{aligned}
\nabla^+ P(T,K) &= P(T,K+1) - P(T,K) \\
&= E[((K+1) - N_T)^+] - E[(K - N_T)^+] \\
&= \sum_{j=N(t)}^{K} (K+1-j)p_j(T) - \sum_{j=N(t)}^{K} (K-j)p_j(T) \\
&= \sum_{j=N(t)}^{K} p_j(T),
\end{aligned}$$

and

$$\begin{aligned}
\nabla^- \nabla^+ P(T,K) &= \nabla^-(P(T,K+1) - P(T,K)) \\
&= [P(T,K+1) - P(T,K)] - [P(T,K) - P(T,K-1)] \\
&= \sum_{j=N(t)}^{K} p_j(T) - \sum_{j=N(t)}^{K-1} p_j(T) \\
&= p_K(T).
\end{aligned}$$

This concludes the proof. □

The above lemma shows that the European-style call option on N_t determines the time T marginal distribution of N_T through (3.17). Since $a_n(T)$ is related to $p_k(T)$ through Dupire's formula (3.6), we can express the transition rate $a_n(T)$ in terms of $P(T,K)$, which in turn implies the forward equation.

Theorem 3.2.5. *If N_t is a time-inhomogeneous Markov chain with local intensity surface $\Lambda(t,N_t)$, denote $a_n(T) = \Lambda(T,n)$. Then $P(T,K)$ satisfies the forward equation*

and initial condition (3.19).

$$\begin{cases} \frac{\partial}{\partial T}P(T,K) = -\sum_{j=1}^{K-1} P(T,j)\left(\nabla^-\nabla^+ a_j(T)\right) + a_K(T)P(T,K-1) - P(T,K)a_{K-1}(T) \\ P(t,K) = (K-N_t)^+. \end{cases} \tag{3.19}$$

for all $N(t) \leq K \leq \bar{N}$ *and* $T \geq t$.

Proof. From Dupire's formula (3.6)

$$\begin{aligned} a_K(T) &= -\frac{1}{p_K(T)}\frac{\partial}{\partial T}\sum_{j=N(t)}^{K} p_j(T) \\ &= -\frac{1}{\nabla^-\nabla^+ P(T,K)}\frac{\partial}{\partial T}(\sum_{j=N(t)}^{K} p_j(T)) \\ &= -\frac{1}{\nabla^-\nabla^+ P(T,K)}\frac{\partial}{\partial T}(\nabla^+ P(T,K)) \end{aligned}$$

$$a_K(T)\nabla^-\nabla^+ P(T,K) = -\frac{\partial}{\partial T}(\nabla^+ P(T,K)) \tag{3.20}$$

List the above equations for K ranging from 0 to $K-1$:

$$\begin{cases} a_{K-1}(T)\Big[P(T,K) - 2P(T,K-1) + P(T,K-2)\Big] & = -\frac{\partial}{\partial T}(P(T,K) - P(T,K-1)) \\ a_{K-2}(T)\Big[P(T,K-1) - 2P(T,K-2) + P(T,K-3)\Big] & = -\frac{\partial}{\partial T}(P(T,K-1) - P(T,K-2)) \\ \quad \cdots & \\ a_1(T)\Big[P(T,2) - 2P(T,1) + P(T,0)\Big] & = -\frac{\partial}{\partial T}(P(T,2) - P(T,1)) \\ a_0(T)\Big[P(T,1) - 2P(T,0)\Big] & = -\frac{\partial}{\partial T}(P(T,1) - P(T,0)) \end{cases} \tag{3.21}$$

Add the equations and notice that $P(T,0) = 0$. Using the discrete version of integration

by parts on the left side of the equations, we have,

$$
\begin{aligned}
-\frac{\partial}{\partial T}P(T,K) &= \sum_{j=1}^{K-2} P(T,j)(\nabla^-\nabla^+ a_j(T)) + \\
&\quad P(T,K)(a_{K-1}(T) - 2a_{K-1}(T)P(T,K-1) + a_{K-2}(T)P(T,K-1) \\
&= \sum_{j=1}^{K-2} P(T,j)(\nabla^-\nabla^+ a_j(T)) + P(T,K-1)(a_K(T) - 2a_{K-1}(T) + a_{K-2}(T)) \\
&\quad - P(T,K-1)a_K(T) + a_{K-1}(T)P(T,K) \\
&= -\sum_{j=1}^{K-1} P(T,j)(\nabla^-\nabla^+ a_j(T)) - a_K(T)P(T,K-1) + P(T,K)a_{K-1}(T)
\end{aligned}
$$

□

All the above forward equations (3.12) and (3.17) are linear systems of ODE. They can be solved iteratively. If we combine them together, we have the following

Corollary 3.2.6. *If N_t is a time-inhomogeneous Markov chain with local intensity surface $\Lambda(t, N_t)$, denote $a_n(T) = \Lambda(T, n)$. Then $C(T,K)$ and $P(T,K)$ satisfies the following equation.*

$$
\begin{aligned}
\frac{\partial}{\partial T}\Big[C(T,K) - P(T,K)\Big] &= \sum_{j=1}^{K-1} P(T,j)(\nabla^-\nabla^+ a_j(T)) + \sum_{j=K}^{\bar{N}} C(T,j)(\nabla^-\nabla^+ a_j(T)) \\
&\quad + a_K(T)(C(T,K-1) - P(T,K-1)) \\
&\quad - a_{K-1}(T)(C(T,K) - P(T,K)) \qquad (3.22)
\end{aligned}
$$

Proof. This is an immediate result by adding the forward equation (3.14) for the European-style call option and the forward equation (3.19) for the European-style put option.

□

3.3 Gyöngy's theorem for jump process

As discussed in section 2.4, Gyöngy's theorem [12] shows that it is possible to construct a Markovian diffusion process with the same one-dimensional marginal distributions as a given initial Itô process. This is closely related to work by Dupire that shows how to construct a local volatility model such that the European-style option prices implied by the model agree with a given set of market prices. In this section, we give an analog of Gyöngy's theorem for jump process and use this to define the notion of local intensity surface for counting processes with arbitrary stochastic intensity.

3.3.1 Gyöngy's theorem for counting process

Theorem 3.3.1 (Gyöngy's theorem for counting process). *Let N_t be a counting process with adapted stochastic intensity λ_t, define $\Lambda(t,l) = \mathbb{E}[\lambda_t | N_{t-} = l]$. Let $\tilde{N}_t$ be a Markovian counting process with an adapted (local) intensity $\lambda_t = \Lambda(t, \tilde{N}_t)$, and suppose $\tilde{N}(0) = N(0) = 0$, then N_t and $\tilde{N}_t$ have same the marginal distribution, that is, $\mathbb{P}(N_t = n) = \mathbb{P}(\tilde{N}_t = n)$ for any $n \geq 0$*

Proof. As a pure counting process, N_t is associated with its jump measure $\mu(dx, dt)$.

$$N_t = \int_0^t \int_{\mathbb{R}} \mu(dx, dt)$$

If we treat N_t as a marked point process. The compensator measure is

$$\nu(dx, dt) = \kappa(t, dx)\lambda(t)dt$$

Since N_t has constant unit jump size, the jump size measure is a Dirac delta measure $\kappa(t, dx) = \delta_{\{Y=1\}}dx$. The process $\lambda(t)$ is just the jump intensity of N_t which is a predictable stochastic process.

For any bounded measurable function f: $\mathbb{R} \to \mathbb{R}$,

$$\begin{aligned} f(N_t) &= f(N_0) + \int_0^t \int_E \left(f(N(s-) + x) - f(N(s-))\right) \mu(dx, ds) \\ &= f(N_0) + \int_0^t \int_E \left(f(N(s-) + x) - f(N(s-))\right) (\mu(dx, ds) - \kappa(s, dx)\lambda(s)ds) \\ &\quad + \int_0^t \int_E \left(f(N(s-) + x) - f(N(s-))\right) \kappa(s, dx)\lambda(s)ds \end{aligned}$$

Since $f(N(t-)+x)-f(N(t-))$ is predictable, $\int_0^t \int_E (f(N(s-)+x)-f(N(s-)))\,(\mu(dx,ds)-\kappa(s,dx)\lambda(s)ds)$ is a local martingale. Take expectations of both sides:

$$E(f(N_t)) = f(N_0) + E\bigg[\int_0^t \int_E (f(N(s-)+x)-f(N(s-)))\,\kappa(s,dx)\lambda(s)ds\bigg].$$

Since $\kappa(s,dx) = \delta_{\{Y=1\}}dx$, we have

$$\begin{aligned} E(f(N_t)) &= f(N_0) + E\bigg[\int_0^t (f(N(s-)+1)-f(N(s-)))\,\lambda(s)ds\bigg] \\ &= f(N_0) + \int_0^t E\bigg[E\Big[(f(N(s-)+1)-f(N(s-)))\,\lambda(s)|N(s-)\Big]\bigg]ds \\ &= f(N_0) + \int_0^t E\bigg[(f(N(s-)+1)-f(N(s-)))\,E\Big[\lambda(s)|N(s-)\Big]\bigg]ds. \end{aligned}$$

Choose f be the payoff function of a European-style put option with strike K, and let $f(x) = [K-x]^+$, then

$$E[(K-N_t)^+] = K + \int_0^t E\bigg[\Big((K-(N(s-)+1))^+ - (K-N(s-))^+\Big)E\Big[\lambda(s)|N(s-)\Big]\bigg]ds$$

$$\begin{aligned} E[(K+1-N_t)^+] = K+1+ \\ \int_0^t E\bigg[\Big((K+1-(N(s-)+1))^+ - (K+1-N(s-))^+\Big)E\Big[\lambda(s)|N(s-)\Big]\bigg]ds \end{aligned}$$

Subtract the first equation from the second equation and use equation

$$\sum_{j=0}^{K} p_j(t) = E[(K+1-N_t)^+] - E[(K-N_t)^+],$$

as in (3.17), where $p_j(t) = \mathbb{P}(N_t = j)$, then we obtain

$$\begin{aligned} \sum_{j=0}^{K} p_j(t) = 1 - \int_0^t E\bigg[\Big((K+1-N(s-))^+ - 2(K-N(s-))^+ + \\ (K-1-N(s-))^+\Big)E\Big[\lambda(s)|N(s-)\Big]\bigg]ds \end{aligned}$$

Note that

$$(K+1-N(s-))^+ - 2(K-N(s-))^+ + (K-1-N(s-))^+ \tag{3.23}$$

is a hump payoff function of a butterfly option. In the case of an integer valued $N(s-)$, The expression (3.23) equals $1_{\{N(s-)=K\}}$. So

$$\begin{aligned}\sum_{j=0}^{K} p_j(t) &= 1 - \int_0^t E\Big[1_{\{N(s-)=K\}} E\Big[\lambda(s)|N(s-)\Big]\Big] ds \\ &= 1 - \int_0^t E\Big[\lambda(s)|N(s-)=K\Big]\mathbb{P}(N_{s-}=K) ds.\end{aligned}$$

Since

$$\begin{aligned}\mathbb{P}(N_{s-}=K) =& \mathbb{P}(N_{s-}=K-1)\mathbb{P}(N_s - N_{s-}=1) \\ &+\mathbb{P}(N_{s-}=K)\mathbb{P}(N_s - N_{s-}=0),\end{aligned}$$

and $\mathbb{P}(N_s - N_{s-}=1)=0$, $\mathbb{P}(N_s - N_{s-}=0)=1$, we have $\mathbb{P}(N_{s-}=K)=\mathbb{P}(N_s=K)=p_K(s)$. Therefore,

$$\sum_{j=0}^{K} p_j(t) = 1 - \int_0^t E\Big[\lambda(s)|N(s-)=K\Big] p_K(s) ds.$$

Taking a derivative on the preceding equation with respect to t,

$$\frac{\partial}{\partial t}\sum_{j=0}^{K} p_j(t) = -E\Big[\lambda(t)|N(t-)=K\Big] p_K(t)$$

$$\frac{-1}{p_K(t)}\frac{\partial}{\partial t}\sum_{j=0}^{K} p_j(t) = E\Big[\lambda(t)|N(t-)=K\Big]. \tag{3.24}$$

Therefore, the marginal distribution $p_K(t)$ of N_t, $K=0,\cdots,\bar{N}$, is completely determined by $\lambda(t)$ through the conditional expectation $E\Big[\lambda(t)|N(t-)=K\Big]$, for all $K=0,\cdots,\bar{N}$.

We now define a Markovian counting process $\tilde{N}_t$ with local intensity $\Lambda(t,\tilde{N}_t)$ where $\Lambda(t,n)=E[\lambda(t)|N(t-)=n]$. We have the Dupire's formula (3.6). Namely, The marginal distribution $\tilde{p}_K(t)$ of $\tilde{N}_t$, $K=0,\cdots,\bar{N}$ satisfies the same equation (3.24),

$$\frac{-1}{\tilde{p}_K(t)}\frac{\partial}{\partial t}\sum_{j=0}^{K} \tilde{p}_j(t) = \Lambda(t,\tilde{N}_t) = E\Big[\lambda(t)|N(t-)=K\Big] \tag{3.25}$$

which admits unique solution as solved in (3.5). So N_t and $\tilde{N}_t$ have the same marginal distribution. □

Theorem 3.3.1 tells us that as long as $\Lambda(t,n) = E[\lambda_t|N_t = n]$ and $N_0 = \tilde{N}_0$ in distribution, N_t and $\tilde{N}_t$ share the same marginal distribution. For some future time $t > 0$, in order to have the Markovian counting process $\tilde{N}$ again share the same marginal distribution conditioning on $\mathcal{F}_t$, we necessarily need $\tilde{N}_t = N_t$ and the local intensity should be the expectation conditioning on $\mathcal{F}_t$. Precisely, theorem 3.3.2 can be immediately extended to the following

Theorem 3.3.2 (Gyöngy's Theorem). *For any non-Markovian counting process N_t with instantaneous intensity λ_t. At any time $t > 0$, there exists a Markovian counting process $\tilde{N}_T^t$ starting from $T = t$, sharing the same marginal distribution with N_T. That is, for all k,*

$$\mathbb{P}(N_T = k|\mathcal{F}_t) = \mathbb{P}(\tilde{N}_T^t = k|\mathcal{F}_t) \ , \ T > t.$$

In addition, $\tilde{N}_T^t$, as a time-inhomogeneous Markovian chain, has the local intensity

$$\Lambda_t(T,n) = E[\lambda_T|N_T = n, \mathcal{F}_t], \tag{3.26}$$

for any $T \geq t$ and $n \geq N_t$.

The Markovian process $\tilde{N}_T^t$ is usually called the "mimicking process" which mimics the marginal distribution of N_T at time t. Kellerer in [15] gave a similar result which showed that the existence of the mimicking Markov point process can be applied to a more general point process which may not even have intensity.

Definition 3.3.3 (Local intensity surface). *For any counting process N_t with adapted instantaneous intensity λ_t, we define the local intensity surface of N_t at time t to be the function*

$$\Lambda_t(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{Z}_{\geq 0} \to \mathbb{R}_{>0}$$

where $\Lambda_t(T,n) = E[\lambda_T|N_T = n, \mathcal{F}_t]$ as given in Theorem 3.3.2

Remark: This definition is consistent with the previous Definition 3.1.1. For a Markovian counting process N_t with $\lambda_t = f(t, N_t)$, the mimicking process is just itself. According to Definition 3.3.3, the local intensity surface of N_t is simply $f(\cdot,\cdot)$ which does not change over time.

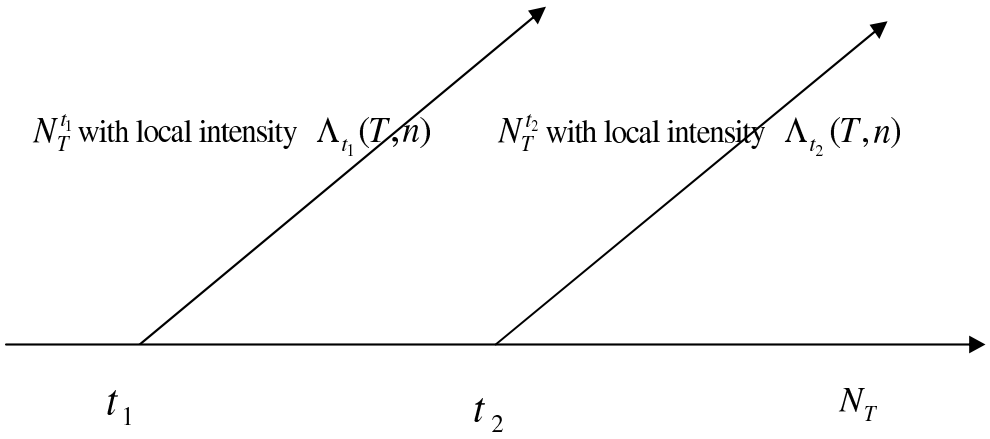

Figure 3.2: The horizontal line represents an counting process N_T with stochastic intensity. The Markov process $N_T^{t_1}$ with local intensity function Λ_1 mimics the marginal distribution of N_T conditioning on $\mathcal{F}_{t_1}$, therefore N_T has the local intensity surface $\Lambda_{t_1}(\cdot,\cdot)$ at time t_1. The Markov process $N_T^{t_2}$ with local intensity function Λ_2 mimics the marginal distribution of N_T conditioning on $\mathcal{F}_{t_2}$, therefore N_T has the local intensity surface $\Lambda_{t_2}(\cdot,\cdot)$ at time t_2.

3.3.2 Example of a parametric family of local intensity Surfaces

For a certain counting process N_t with stochastic intensity λ_t, the local intensity surface at time t is defined in Definition 3.3.3. In this section, we give a simple example of a stochastic intensity model whose local intensity surface can be explicitly expressed as a parametric function.

We randomize the intensity of aggregate loss process $N(t)$ by assuming that it takes the value of λ_1, λ_2, λ_3 with probability c_1, c_2, $c_3 = 1 - c_1 - c_2$ respectively. The initial local intensity surface $\Lambda_0(T,n)$ can be derived from the probability $p_k(T) = \mathbb{P}(N(T) = k)$,

$$p_k(T) = \sum_{i=1}^{3} \mathbb{P}(N(T) = k|\lambda = \lambda_i) \cdot c_i,$$

where the probability $\mathbb{P}(N(T) = k|\lambda = \lambda_i)$ is Poisson distributed with parameter λ_i,

$$p_{ik}(T) = \mathbb{P}(N(T) = k|\lambda = \lambda_i) = \frac{e^{-\lambda_i T}(\lambda_i T)^k}{k!}.$$

Therefore,

$$\begin{aligned}
\Lambda_0(T,n) &= -\frac{1}{p_n(T)}\sum_{k=0}^{n}\frac{\partial}{\partial T}p_k(T) \\
&= -\frac{\sum_{k=0}^{n}\frac{\partial}{\partial T}\sum_{i=1}^{3}c_i p_{ik}(T)}{\sum_{i=1}^{3}c_i p_{in}(T)} \\
&= -\frac{\sum_{i=1}^{3}c_i\left(\sum_{k=0}^{n}\frac{\partial}{\partial T}p_{ik}(T)\right)}{\sum_{i=1}^{3}c_i p_{in}(T)} \\
&= -\frac{1}{\sum_{i=1}^{3}c_i p_{in}(T)}\sum_{i=1}^{3}c_i\sum_{k=0}^{n}\left(\frac{-\lambda_i e^{-\lambda_i T}(-\lambda_i T)^k}{k!}+\frac{\lambda_i e^{-\lambda_i T}k(\lambda_i T)^{k-1}}{k!}\right) \\
&= \frac{\sum_{i=1}^{3}c_i\lambda_i e^{-\lambda_i T}\frac{(\lambda_i T)^n}{n!}}{\sum_{i=1}^{3}c_i e^{-\lambda_i T}\frac{(\lambda_i T)^n}{n!}},
\end{aligned}$$

so we obtain the local intensity of this example,

$$\Lambda_0(T,n) = \frac{\sum_{i=1}^{3}c_i\lambda_i^{n+1}e^{-\lambda_i T}}{\sum_{i=1}^{3}c_i\lambda_i^{n}e^{-\lambda_i T}} \tag{3.27}$$

1) As $T \searrow 0$,

$$\Lambda_0(T,n) \to \frac{c_1\lambda_1^{n+1}+c_2\lambda_2^{n+1}+c_3\lambda_3^{n+1}}{c_1\lambda_1^{n}+c_2\lambda_2^{n}+c_3\lambda_3^{n}}.$$

This is a weighted sum of λ_i with weight $\frac{c_i\lambda_i^n}{\sum_{i=1}^{3}c_i\lambda_i^n}$. when $n=0$ (assume no default in the beginning), $\Lambda_0(0,n)$ is just the expectation of λ. When n is large, $\Lambda_0(0,n)$ is close to the maximum of λ_i.

2) As $T \nearrow \infty$,

$$\Lambda_0(T,n) \to \min\{\lambda_1,\lambda_2,\lambda_3\}$$

If λ is distributed with density $\rho(\lambda)$ for $\lambda > 0$, similarly, we have

$$\Lambda_0(T,n) = \frac{\int_0^\infty \lambda^{n+1}e^{-\lambda T}\rho(\lambda)d\lambda}{\int_0^\infty \lambda^{n}e^{-\lambda T}\rho(\lambda)d\lambda},$$

and extreme behavior of the initial local intensity can be obtained in the same manners.

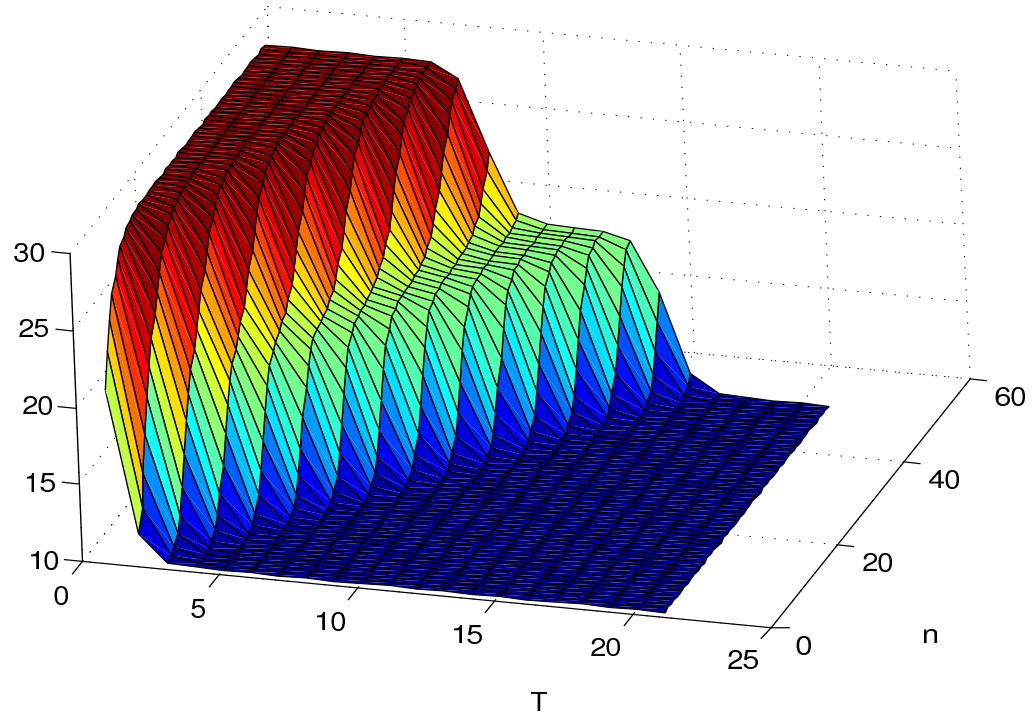

Figure 3.3: Graph of a model implied local intensity surface: Above is a sample of local intensity surface $\Lambda_0(T, n)$, implied from the model in section 3.3.2. with $c_1 = 0.2, c_2 = 0.5, c_3 = 0.3, \lambda_1 = 10, \lambda_2 = 20$ and $\lambda_3 = 30$.

3.4 Further extensions of Gyöngy's theorem

In this section, we further extend Gyöngy's mimicking theorem. We first provide a new proof of the original Gyöngy theorem for Itô process as in [12]. The proof uses the Fourier transform method. This method later motivates the proof of a further extended Gyöngy's theorem for discontinuous semi-martingales.

3.4.1 Proof of Gyöngy's theorem using the Fourier transform Method

We provide a proof for the following Gyöngy's theorem.

Theorem 3.4.1. *Let $X(t)$ be an n-dimensional Itô process which satisfies*

$$dX(t) = \beta(t)dt + \sigma(t)dW_t,$$

where W_t is a d-dimensional Brownian motion. $\beta(t)$ is a bounded n-dimensional adapted process and $\sigma(t)$ is a bounded $n \times d$-dimensional adapted process such that $\sigma(t)\sigma^T(t)$ is

uniformly positive definite. Then there exist measurable functions a and b

$$\begin{aligned} a^2(t,x) &= E[\sigma(t)\sigma^T(t)|X(t)=x] \\ b(t,x) &= E[\beta(t)|X(t)=x], \end{aligned}$$

and there exists a weak solution $Y(t)$ to the SDE:

$$dY(t) = b(t,Y(t))dt + a(t,Y(t))d\widetilde{W}_t,$$

where $\widetilde{W}_t$ is a n-dimensional Brownian motion, and $X(0) = Y(0)$, such that X and Y have the same marginal distributions for all $t > 0$.

Proof. Consider the function $f(x) = e^{-i\xi\cdot x}$, $x, \xi \in \mathbb{R}^n$. Itô's lemma shows that

$$\begin{aligned} f(X_T) &= f(X_0) + \int_0^T \sum_{j=1}^n \frac{\partial f}{\partial x_j} dX_j(t) + \frac{1}{2}\int_0^T \sum_{j,k=1}^n \frac{\partial^2 f}{\partial x_j \partial x_k} dX_j(t)dX_k(t) \\ &= f(X_0) + \int_0^T \sum_{j=1}^n -i\xi_j e^{-i\xi\cdot X(t)} dX_j(t) \\ &\quad -\frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k e^{-i\xi\cdot X(t)} \left(\sigma(t)\sigma^T(t)\right)_{jk} dt \\ &= f(X_0) + \int_0^T \sum_{j=1}^n -i\xi_j e^{-i\xi\cdot X(t)} \left(\beta_j(t)dt + \sum_{k=1}^d \sigma_{jk}(t)dW_k(t)\right) \\ &\quad -\frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k e^{-i\xi\cdot X(t)} \left(\sigma(t)\sigma^T(t)\right)_{jk} dt \end{aligned}$$

Taking expectation on both sides, we get

$$\begin{aligned} E[f(X_T)] &= f(X_0) + E\left[\int_0^T \sum_{j=1}^n -i\xi_j e^{-i\xi\cdot X(t)}\beta_j(t)dt\right] \\ &\quad -E\left[\frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k e^{-i\xi\cdot X(t)} \left(\sigma(t)\sigma^T(t)\right)_{jk} dt\right] \\ &= f(X_0) + \int_0^T \sum_{j=1}^n -i\xi_j E\left[e^{-i\xi\cdot X(t)}\beta_j(t)\right] dt - \\ &\quad \frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k E\left[e^{-i\xi\cdot X(t)} \left(\sigma(t)\sigma^T(t)\right)_{jk}\right] dt \end{aligned}$$

Using the tower property in the inner expectation, we get

$$\begin{aligned} E[f(X_T)] &= f(X_0) + \int_0^T \sum_{j=1}^n -i\xi_j E\left[e^{-i\xi\cdot X(t)} E\left[\beta_j(t)|X(t)\right]\right] dt \\ &\quad -\frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k E\left[e^{-i\xi\cdot X(t)} E[\left(\sigma(t)\sigma^T(t)\right)_{jk}|X(t)]\right] dt \\ &= f(X_0) + \int_0^T \sum_{j=1}^n -i\xi_j E[e^{-i\xi\cdot X(t)} b_j(t,X_t)]dt - \\ &\quad \frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j\xi_k E[e^{-i\xi\cdot X(t)} a_{jk}^2(t,X_t)]dt \end{aligned}$$

Taking a derivative with respect to T on both sides , we have

$$\frac{\partial}{\partial T}E[f(X_T)] = \sum_{j=1}^n -i\xi_j E[e^{-i\xi\cdot X_T} b_j(T,X_T)] - \frac{1}{2}\sum_{j,k=1}^n \xi_j\xi_k E[e^{-i\xi\cdot X_T} a_{jk}^2(T,X_T)].$$

The $E[f(X_T)]$ term is just the Fourier transform of the transition density function $p_{X_T}(T,x)$ [1] of the random variable X_T (assuming the density function exists at least weakly). That is

$$E[f(X_T)] = \mathscr{F}[p_{X_T}(T,x)](\xi)$$

Therefore, we can rewrite the above equations in terms of Fourier transform

$$\begin{aligned} \frac{\partial}{\partial T}\mathscr{F}[p_{X_T}(T,x)](\xi) &= \sum_{j=1}^n -i\xi_j \int_{R^n} e^{-i\xi\cdot x} b_j(T,x) p_{X_T}(T,x)dx \\ &\quad -\frac{1}{2}\sum_{j,k=1}^n \xi_j\xi_k \int_{R^n} e^{-i\xi\cdot x} a_{jk}^2(T,x) p_{X_T}(T,x)dx \\ &= \sum_{j=1}^n -i\xi_j \mathscr{F}[b_j(T,x)p_{X_T}(T,x)](\xi) - \\ &\quad \frac{1}{2}\sum_{j,k=1}^n \xi_j\xi_k \mathscr{F}[a_{jk}^2(T,x)p_{X_T}(T,x)](\xi) \\ &= -\sum_{j=1}^n \mathscr{F}[\frac{\partial}{\partial x_j} b_j(T,x)p_{X_T}(T,x)](\xi) \\ &\quad +\frac{1}{2}\sum_{j,k=1}^n \mathscr{F}[\frac{\partial^2}{\partial x_j \partial x_k} a_{jk}^2(T,x)p_{X_T}(T,x)](\xi) \end{aligned}$$

[1] The transition density function should be written as $p_{X_T}(t,y;T,x)$, the shortcut notation $p_{X_T}(T,x)$ is used since here we fix the backward variable t and y.

Here we assume $p_{X_T}(T,x)$ satisfies enough regularity condition so that we can interchange of $\frac{\partial}{\partial T}$ and Fourier transform. Then taking inverse transform gives

$$\begin{aligned}\frac{\partial}{\partial T}p_{X_T}(T,x) &= -\sum_{j=1}^{n}\frac{\partial}{\partial x_j}b_j(T,x)p_{X_T}(T,x) \\ &\quad +\frac{1}{2}\sum_{j,k=1}^{n}\frac{\partial^2}{\partial x_j\partial x_k}a_{jk}^2(T,x)p_{X_T}(T,x) \qquad (3.28)\end{aligned}$$

In fact, f can be chosen as any arbitrary smooth test function. And we can also derive that $p_{X_T}(T,x)$ is a weak solution of (3.28).

Equation (3.28) is the well-known Kolmogorov forward equation satisfied by the probability density function $p_{X_T}(T,x)$. Suppose Y_t is a stochastic process which follows

$$dY(t) = b(t,Y(t))dt + a(t,Y(t))d\tilde{W}_t,$$

and its has a probability density function $p_{Y_t}(t,y)$, then clearly, $p_{Y_T}(T,y)$ also satisfies (3.28). In order to show that $X(t), Y(t)$ have the same marginal distributions, it suffices to show that (3.28) has a unique solution. We know the uniqueness for parabolic PDEs is guaranteed according to Evans [9], since $b(t,x)$ and $a(t,x)$ are bounded and $a^2(t,x)$ is uniformly positive definite.

□

3.4.2 Gyöngy's theorem for semi-martingales

We now consider a result mimicking the marginal distribution of a quite general class of stochastic process with both diffusion and jumps.

Theorem 3.4.2. *Let $X(t)$ be an Itô semi-martingale given by the following decomposition.*

$$X_t = X_0 + \int_0^t \beta_s ds + \int_0^t \sigma_s dW_s + \int_0^t\int_{\|y\|\le 1} y\widetilde{M}_X(ds,dy) + \int_0^t\int_{\|y\|>1} yM_X(ds,dy)$$

where W_t is a $\mathbb{R}^d$-valued Brownian motion, M_X is a random measure on $[0,\infty]\times\mathbb{R}^n$ with compensator μ_X, $\widetilde{M}_X = M_X - \mu_X$ is the compensated measure. We assume μ_X

has a density $m_X(\omega, t, y)$, β_t *and* σ_t *are bounded adapted processes in* $\mathbb{R}^n$ *and* $\mathbb{R}^n \times \mathbb{R}^d$. $\sigma_t \sigma_t^T$ *is uniformly positive definite.*

Define

$$\begin{aligned} a^2(t,x) &= E[\sigma(t)\sigma^T(t)|X(t-) = x] \\ b(t,x) &= E[\beta(t)|X(t-) = x] \\ m_Y(t,A,x) &= E[m_X(\cdot,t,A)|X(t-) = x], \end{aligned}$$

for all $t \geq 0$, $x \in \mathbb{R}^n$, *and* A *which is a Borel set in* R^n. *There exists a stochastic process* $Y(t)$, *which is a weak solution for the stochastic differential equation*

$$\begin{aligned} Y_t &= X_0 + \int_0^t b(s, Y_s)ds + \int_0^t a(s, Y_s)d\widetilde{W}_s \\ &\quad + \int_0^t \int_{\|y\|\leq 1} y\widetilde{M}_Y(dsdy) + \int_0^t \int_{\|y\|>1} yM_Y(dsdy) \end{aligned}$$

where $\widetilde{W}_t$ *is an* n*-dimensional Brownian motion,* M_Y *is a random measure on* $[0, \infty) \times \mathbb{R}^n$ *with compensator* $m_Y(t, dy, Y_{t-})dt$. *Then* X *and* Y *have the same marginal distribution for all* $t > 0$.

Proof. Let f be a C^2 function on $\mathbb{R}^n$, we use the following Itô lemma for semi-martingale

$$\begin{aligned} f(X_T) &= f(X_0) + \sum_{j=1}^n \int_0^T \frac{\partial}{\partial x_j} f(X_{t-})dX_t^j + \frac{1}{2}\sum_{j,k=1}^n \frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-})(dX_t^j, dX_t^k) \\ &\quad + \sum_{t\leq T}\left[f(X_{t-} + \triangle X_t) - f(X_{t-}) - \sum_{j=1}^n \frac{\partial}{\partial x_j} f(X_{t-})\triangle X_t^j\right] \end{aligned}$$

For the semi-martingale X_t in our specification

$$\begin{aligned}
f(X_T) &= f(X_0) + \int_0^T \nabla f(X_{t-}) \cdot \beta_t dt + \int_0^T \nabla f(X_{t-}) \cdot \sigma_t dW_t \\
&\quad + \frac{1}{2} \int_0^T \sum_{j,k=1}^{n} \frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-})(\sigma(t)\sigma^T(t))_{jk} dt \\
&\quad + \int_0^T \int_{\|y\| \leq 1} (\nabla f(X_{t-} \cdot y)) \widetilde{M}_X(dt, dy) + \int_0^T \int_{\|y\|>1} (\nabla f(X_{t-} \cdot y)) M_X(dt, dy) \\
&\quad + \int_0^T \int_{R^n} (f(X_{t-} + y) - f(X_{t-}) - y \cdot \nabla f(X_{t-})) M_x(dt, dy) \\
&= f(X_0) + \int_0^T \nabla f(X_{t-}) \cdot \beta_t dt + \int_0^T \nabla f(X_{t-}) \cdot \sigma_t dW_t \\
&\quad + \frac{1}{2} \int_0^T \sum_{j,k=1}^{n} \frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-})(\sigma(t)\sigma^T(t))_{jk} dt + \int_0^T \int_{\|y\| \leq 1} (\nabla f(X_{t-} \cdot y)) \widetilde{M}_X(dt, dy) \\
&\quad + \int_0^T \int_{R^n} (f(X_{t-} + y) - f(X_{t-}) - 1_{\|y\| \leq 1} y \cdot \nabla f(X_{t-})) M_X(dt, dy)
\end{aligned}$$

Taking expectations on both side, we get

$$\begin{aligned}
E[f(X_T)] &= f(X_0) + E\left[\int_0^T \nabla f(X_{t-}) \cdot \beta_t dt\right] \\
&\quad + \frac{1}{2} E\left[\int_0^T \sum_{j,k=1}^{n} \frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-})(\sigma(t)\sigma^T(t))_{jk} dt\right] \\
&\quad + E\left[\int_0^T \int_{R^n} (f(X_{t-} + y) - f(X_{t-}) - 1_{\|y\| \leq 1} y \cdot \nabla f(X_{t-})) m_X(t, dy) dt\right].
\end{aligned}$$

We can apply Fubini's theorem,

$$\begin{aligned}
E[f(X_T)] &= f(X_0) + \int_0^T E\left[\nabla f(X_{t-}) \cdot \beta_t\right] dt \\
&\quad + \frac{1}{2} \int_0^T \sum_{j,k=1}^{n} E\left[\frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-})(\sigma(t)\sigma^T(t))_{jk} dt\right] \\
&\quad + \int_0^T E\left[\int_{R^n} (f(X_{t-} + y) - f(X_{t-}) - 1_{\|y\| \leq 1} y \cdot \nabla f(X_{t-})) m_X(t, dy)\right] dt.
\end{aligned}$$

Using the iterated expectation condtioned on X_{t-}, we have

$$\begin{aligned} E[f(X_T)] &= f(X_0) + \int_0^T E\left[\nabla f(X_{t-}) \cdot E[\beta_t | X_{t-}]\right] dt \\ &\quad + \frac{1}{2}\int_0^T \sum_{j,k=1}^n E\left[\frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-}) E[(\sigma(t)\sigma^T(t))_{jk} | X_{t-}] dt\right] \\ &\quad + \int_0^T E\Bigg[E[\int_{R^n} (f(X_{t-}+y) - f(X_{t-}) \\ &\qquad\qquad - 1_{\|y\|\le 1} y \cdot \nabla f(X_{t-})) m_X(t, dy) | X_{t-}]\Bigg] dt \end{aligned}$$

$$\begin{aligned} E[f(X_T)] &= f(X_0) + \int_0^T E[\nabla f(X_{t-}) \cdot b(t, X_{t-})] dt \\ &\quad + \frac{1}{2}\int_0^T \sum_{j,k=1}^n E\left[\frac{\partial^2}{\partial x_j \partial x_k} f(X_{t-}) a_{jk}^2(t, X_{t-})\right] dt \\ &\quad + \int_0^T E\left[\int_{R^n} (f(X_{t-}+y) - f(X_{t-}) - 1_{\|y\|\le 1} y \cdot \nabla f(X_{t-})) m_Y(t, dy, X_{t-})\right] dt \end{aligned}$$

Again, we choose $f(x) = e^{-i\xi \cdot x}$, then $\nabla f(x) = (-i\xi_1 e^{-i\xi \cdot X}, \cdots, -i\xi_n e^{-i\xi \cdot X})$, and $\frac{\partial^2}{\partial x_j \partial x_k} f(x) = -\xi_j \xi_k e^{-i\xi \cdot x}$. Therefore,

$$\begin{aligned} &E[f(X_T)] = f(X_0) + \int_0^T \sum_{j=1}^n E\left[-i\xi_j e^{-i\xi \cdot X_{t-}} b_j(t, X_{t-})\right] dt \\ &-\frac{1}{2}\int_0^T \sum_{j,k=1}^n \xi_j \xi_k E\left[e^{-i\xi \cdot X_{t-}} a_{jk}^2(t, X_{t-})\right] dt \\ &+ \int_0^T E\left[\int_{R^n} (e^{-i\xi \cdot (X_{t-}+y)} - e^{-i\xi \cdot X_{t-}} - \sum_{j=1}^n 1_{\|y\|\le 1} y_j i\xi_j e^{-i\xi \cdot X_{t-}}) m_Y(t, dy, X_{t-})\right] dt \end{aligned}$$

Taking derivatives with respect to T on both sides, we get

$$\begin{aligned} &\frac{\partial}{\partial T} E[f(X_T)] = \sum_{j=1}^n -i\xi_j E[e^{-i\xi \cdot X_{T-}} b_j(T, X_{T-})] - \frac{1}{2}\sum_{j,k=1}^n \xi_j \xi_k E[e^{-i\xi \cdot X_{T-}} a_{jk}^2(T, X_{T-})] \\ &+ E[\int_{R^n} (e^{-i\xi \cdot (X_{T-}+y)} - e^{-i\xi \cdot X_{T-}} - \sum_{j=1}^n 1_{\|y\|\le 1} y_j i\xi_j e^{-i\xi \cdot X_{T-}}) m_Y(T, dy, X_{T-})] \end{aligned}$$

If X_T has a probability density function $p_{X_T}(T,x)$, we have

$$\begin{aligned}\frac{\partial}{\partial T}\int_{R^n} e^{-i\xi\cdot x}p_{X_T}(T,x)dx &= \sum_{j=1}^{n} -i\xi_j \int_{R^n} e^{-i\xi\cdot x}b_j(T,x)p_{X_T}(T,x)dx \\ &\quad -\frac{1}{2}\sum_{j,k=1}^{n}\xi_j\xi_k\int_{R^n} e^{-i\xi\cdot x}a_{jk}^2(T,x)p_{X_T}(T,x)dx \\ &\quad +\int_{R^n}\int_{R^n}(e^{-i\xi\cdot(x+y)}-e^{-i\xi\cdot x} \\ &\quad -\sum_{j=1}^{n}1_{\|y\|\le 1}y_j i\xi_j e^{-i\xi\cdot x})m_Y(T,dy,x)p_{X_T}(T,x)dx\end{aligned}$$

We can rewrite this equation in terms of Fourier transform,

$$\begin{aligned}&\frac{\partial}{\partial T}\mathscr{F}[p_{X_T}(T,x)](\xi) \\ =\ &\sum_{j=1}^{n} -i\xi_j\mathscr{F}[b_j(T,x)p_{X_T}(T,x)](\xi) - \frac{1}{2}\sum_{j,k=1}^{n}\xi_j\xi_k\mathscr{F}[a_{jk}^2(T,x)p_{X_T}(T,x)] \\ &+\int_{R^n}\mathscr{F}[(e^{-i\xi\cdot y}-1-\sum_{j=1}^{n}1_{\|y\|\le 1}y_j i\xi_j)m_Y(T,dy,x)p_{X_T}(T,x)](\xi) \\ =\ &-\sum_{j=1}^{n}\mathscr{F}[\frac{\partial}{\partial x_j}b_j(T,x)p_{X_T}(T,x)](\xi) + \frac{1}{2}\sum_{j,k=1}^{n}\mathscr{F}[\frac{\partial^2}{\partial x_j\partial x_k}a_{jk}^2(T,x)p_{X_T}(T,x)](\xi) \\ &+\mathscr{F}[\int_{R^n}p_{X_T}(T,x-y)m_Y(t,y,x-y)-p_{X_T}(T,x)m_Y(t,y,x) \\ &-\sum_{j=1}^{n}1_{\|y\|\le 1}\frac{\partial}{\partial x_j}(m_Y(T,y,x)p_{X_T}(T,x))dy)](\xi)\end{aligned}$$

Here we assume $p_{X_T}(T,x)$ satisfies enough regularity condition so that we can interchange of $\frac{\partial}{\partial T}$ and Fourier transform. Then after taking the inverse Fourier transform, $p_{X_T}(T,x)$ satisfies

$$\begin{aligned}\frac{\partial}{\partial T}p_{X_T}(T,x) &= -\sum_{j=1}^{n}\frac{\partial}{\partial x_j}b_j(T,x)p_{X_T}(T,x) + \frac{1}{2}\sum_{j,k=1}^{n}\frac{\partial^2}{\partial x_j\partial x_k}a_{jk}^2(T,x)p_{X_T}(T,x) \\ &\quad +\int_{R^n}p_{X_T}(T,x-y)m_Y(t,y,x-y)-p_{X_T}(T,x)m_Y(t,y,x) \\ &\quad -\sum_{j=1}^{n}1_{\|y\|\le 1}\frac{\partial}{\partial x_j}(m_Y(T,y,x)p_{X_T}(T,x))dy \qquad (3.29)\end{aligned}$$

This is the Kolmogorov forward equation satisfied by the transition density function $p_{X_T}(T,x)$ [2] of a discontinuous semi-martingale X_T. In fact, f can be chosen as any

[2]The transition density function should be written as $p_{X_T}(t,y;T,x)$, the shortcut notation $p_{X_T}(T,x)$ is used since here we fix the backward variable t and y.

arbitrary smooth test function. And we can also derive that $p_{X_T}(T,x)$ is a weak solution of (3.29).

Suppose the process Y_t is a weak solution to the SDE

$$\begin{aligned} Y_t &= X_0 + \int_0^t b(s,Y_s)ds + \int_0^t a(s,Y_s)d\widetilde{W}_s \\ &\quad + \int_0^t \int_{\|y\|\le 1} y\widetilde{M}_Y(dsdy) + \int_0^t \int_{\|y\|>1} yM_Y(dsdy) \end{aligned} \tag{3.30}$$

and Y_t has a probability density function $p_{Y_t}(t,y)$. Then clearly, $p_{Y_T}(T,y)$ also satisfies equation(3.29). In order to show that $X(t)$, $Y(t)$ have the same marginal distributions, it again suffices to show that equation (3.29) has a unique solution. However, the equation (3.29) is not like the standard forward equation (3.28) for Itô process we discussed in section 3.4.1. We believe under mild conditions for $b(t,x)$, $a(t,x)$ and $m(t,x,y)$, equation (3.29) also admits unique solution. Here we leave the problem and pursue the complete answer in the future. □

Remark: all we really need to complete the entire proof is the uniqueness of solutions to a pseudo-parabolic integro-differential equation (3.29) on $\mathbb{R}^n$. This part of work is being explored in collaboration with Jin Wang. The most relavant references we will rely on are Friedman [11], Kumanogo [16] and Treves [24].

Cont and Bentata in [4] consider the same problem of extending the "mimicking theorem" of Gyöngy [12] to discontinuous semimartingales. That is the flow of marginal distributions of a discontinuous semimartingale X can be matched by a Markov process whose infinitesimal generator is expressed in terms of the local characteristics of X. Their proof mainly relies on the solution of martingale problem.

3.4.3 Mimicking result for discrete time processes

Let's now consider to extend the originally Gyöngy theorem for Itô process to discrete time stochastic processes. First we consider the following example.

Example 1

Suppose we are given a discrete time stochastic process $\{X_n\}$ in a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_n\}_{n\geq 0})$. $X_0 = x_0$ and where X_n satisfies the following dynamics,

$$X_{n+1} = X_n + \sigma_n \varepsilon_n,$$

where ε_n are i.i.d. with standard normal distribution $N(0,1)$, and σ_n is the stochastic volatility adapted to $\{\mathcal{F}_n\}_{n\geq 0}$. The process X_n can be viewed as a discrete version of Itô process. It is natural to ask if $\{X_n\}$ can be mimicked by a discrete time Markov process $\{\tilde{X}_n\}$, so that they have the same marginal distribution. Inspired by the continuous time Gyöngy theorem, $\{\tilde{X}_n\}$ is defined by $\tilde{X}_0 = x_0$, and

$$\tilde{X}_{n+1} = \tilde{X}_n + \tilde{\sigma}_n(\tilde{X}_n)\tilde{\varepsilon}_n,$$

where $\tilde{\varepsilon}_n$ are i.i.d. with standard normal distribution $N(0,1)$ independent to ε_n and

$$\tilde{\sigma}_n^2(x) = E[\sigma_n^2 | X_n = x]$$

It turns out that X_n and $\tilde{X}_n$ may not have same marginal distribution. To construct a counterexample, we suppose $X_0 = \tilde{X}_0 = 0$ and $\sigma_0 = 1$. then $\tilde{\sigma}^2(\tilde{X}_0) = E[\sigma_0^2 | X_0 = \tilde{X}_0] = 1$. Therefore,

$$X_1 = X_0 + \sigma_0 \varepsilon_0 = \varepsilon_0,$$

$$\tilde{X}_1 = \tilde{X}_0 + \tilde{\sigma_0}(\tilde{X}_0)\tilde{\varepsilon}_0 = \tilde{\varepsilon}_0.$$

The random variables X_1 and $\tilde{X}_1$ are both standard normally distributed.
Next, we assume that the stochastic volatility $\sigma_n = X_n Y_n$ for $n \geq 1$, where Y_n is

standard normal independent of X_n. Then

$$\begin{aligned}\tilde{\sigma}_1^2(x) &= E[\sigma_1^2|X_1 = x] \\ &= E[X_1^2 Y_1^2|X_1 = x] \\ &= x^2 E[Y_1^2|X_1 = x] \\ &= x^2.\end{aligned}$$

Therefore

$$X_2 = X_1 + \sigma_1\varepsilon_1 = X_1 + X_1 Y_1 \varepsilon_1,$$

$$\tilde{X}_2 = \tilde{X}_1 + \tilde{\sigma}_1(\tilde{X}_1)\tilde{\varepsilon} = \tilde{X}_1 + \tilde{X}_1\tilde{\varepsilon}_1.$$

Clearly, X_2 and $\tilde{X}_2$ have distinct distributions since $X_1 Y_1 \varepsilon_1$ is a product of three independent standard normals, which has a different distribution to $\tilde{X}_1\tilde{\varepsilon}_1$ which is a product of two independent standard normals.

Example 1 gives a counterexample of the naive extension of the continuous Gyöngy theorem to the discrete time case. However the following extension is correct.

Proposition 3.4.3. *Consider a discrete time discrete state stochastic process $\{X_n\}$ in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_n\}_{n\geq 0})$. Assume that*

$$p_n(z) = \mathbb{P}(X_{n+1} = z|\mathcal{F}_n) = f_n(X_n, Y_n; z)$$

where Y_n is a factor process independent of X_n. If we define another discrete time, discrete state process $\{\tilde{X}_n\}$ whose transition probability is

$$\mathbb{P}(\tilde{X}_{n+1} = z|\mathcal{F}_n) = \tilde{p}_n(\tilde{X}_n; z)$$

where $\tilde{p}_n(x; z) = E[p_n(z)|X_n = x]$, then X_n and $\tilde{X}_n$ have same distribution for all n.

Proof. Observe that

$$\begin{aligned}
\mathbb{P}(X_{n+1}=z) &= \sum_s \mathbb{P}(X_{n+1}=z|X_n=s)\mathbb{P}(X_n=s) \\
&= \sum_s \Big[\sum_y \mathbb{P}(X_{n+1}=z|Y_n=y,X_n=s)\mathbb{P}(Y_n=y|X_n=s)\Big]\mathbb{P}(X_n=s) \\
&= \sum_s \Big[\sum_y f_n(s,y;z)\mathbb{P}(Y_n=y|X_n=s)\Big]\mathbb{P}(X_n=s) \\
&= \sum_s \Big[E[p_n(z)|X_n=s]\Big]\mathbb{P}(X_n=s) \\
&= \sum_s \tilde{p}_n(s;z)\mathbb{P}(X_n=s)
\end{aligned}$$

By induction on n, $\mathbb{P}(X_n=s)$ is equal to $\mathbb{P}(\tilde{X}_n=s)$, so

$$\begin{aligned}
\mathbb{P}(X_{n+1}=z) &= \sum_s \tilde{p}_n(s;z)\mathbb{P}(X_n=s) \\
&= \sum_s \tilde{p}_n(s;z)\mathbb{P}(\tilde{X}_n=s) \\
&= \mathbb{P}(\tilde{X}_{n+1}=z).
\end{aligned}$$

□

The above version of the discrete time, discrete state Gyöngy's theorem also holds if Y_n is a vector. It can be extended as long as $p_n(z)$ is adapted to $\mathcal{F}_n$. Proposition 3.4.3 is true because we localize the stochastic transition probability instead of only localizing the stochastic volatility, which is a very strong condition.

3.5 Relation to stochastic intensity models

In this section, we investigate the local intensity surface implied from two stochastic intensity models: doubly stochastic intensity model (2.42) and Hawkes intensity model (2.43).

3.5.1 Local intensity of jump process with doubly stochastic intensity

Given a default counting process N_t with doubly stochastic intensity in (2.42), the initial local intensity function $\Lambda_0(T,n)$ satisfies Dupire's formula (3.6)

$$\begin{aligned}
\Lambda_0(T,n) &= -\frac{1}{p_n(T)}\sum_{k=0}^{n}\frac{\partial}{\partial T}p_k(T) \\
&= -\frac{\sum_{k=0}^{n}\frac{\partial}{\partial T}E\left[e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^k\frac{1}{k!}\right]}{E\left[e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^n\frac{1}{n!}\right]} \\
&= -\frac{\sum_{k=0}^{n}E\left[-\lambda_T e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^k\frac{1}{k!}+e^{-\int_0^T\lambda_s ds}k(\int_0^T\lambda_s ds)^{k-1}\lambda_T\frac{1}{k!}\right]}{E\left[e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^n\frac{1}{n!}\right]} \\
&= \frac{E(\lambda_T e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^n\frac{1}{n!})}{E(e^{-\int_0^T\lambda_s ds}(\int_0^T\lambda_s ds)^n\frac{1}{n!})}
\end{aligned} \tag{3.31}$$

If the joint distribution of $(\lambda_T, \int_0^T\lambda_s ds)$ is known, the expectation terms in (3.31) can be computed explicitly. This is for example the case if λ_t satisfies Hull-White model.

3.5.2 Local intensity of Hawkes process

Suppose we are given a default counting process N_t whose intensity λ_t is a Hawkes process in (2.43), the intensity λ_t satisfies

$$\lambda(t) = \nu + \int_0^t \alpha e^{-\beta(t-s)}dN_s \tag{3.32}$$

The probability generating function of N_t is defined as $\phi(x,t) = E[x^{N_t}]$. According to Hawkes in [19],

$$\phi(x,t) = \exp\left(-\frac{\nu}{\beta}(1-\psi(x,t)) - \nu\int_0^t 1-\psi(x,u)du\right),$$

where $\psi(x,t)$ is determined by

$$\beta t = -\int_0^{1-\psi}\frac{1}{xe^{-\frac{\alpha}{\beta}u}+u-1}du$$

Denote $\phi^{(n)}(x,t) = \frac{\partial^n}{\partial x^n}\phi(0,t)$. By Dupire formula (3.6) and noticing that $p_k(T) = \mathbb{P}(N_t = k) = \frac{\phi^{(n)}(0,T)}{n!}$, we can compute the initial local intensity surface:

$$\begin{aligned}\Lambda_0(T,n) &= -\frac{1}{p_n(T)}\sum_{k=0}^{n}\frac{\partial}{\partial T}p_k(T) \\ &= -\frac{\frac{\partial}{\partial t}\left(\phi(0,T)+\phi'(0,T)+\cdots+\frac{\phi^{(n)}(0,T)}{n!}\right)}{\frac{\phi^{(n)}(0,T)}{n!}}\end{aligned}$$

Next, we investigate some properties of the local intensity surface implied by the Hawkes process.

Lemma 3.5.1. *Suppose the intensity of the default counting process N_t satisfies the Hawkes dynamics* (3.32)*, then for $n \geq N_t$,*

$$\Lambda_t(t,n) = \lambda_t + (n - N_t)\alpha \tag{3.33}$$

Proof. By definition 3.3.3,

$$\Lambda_t(t,n) = \lim_{h\searrow 0} E_t\big[\lambda_{t+h}\big|N_{t+h} = n\big], \tag{3.34}$$

where in the Hawkes setting,

$$\lambda_{t+h} = \nu + \int_0^{t+h}\alpha e^{-\beta(t+h-s)}dN_s \tag{3.35}$$

$$= \nu + \int_0^{t}\alpha e^{-\beta(t+h-s)}dN_s + \int_t^{t+h}\alpha e^{-\beta(t+h-s)}dN_s. \tag{3.36}$$

As h tends to zero, time t conditional expectation of the first two terms converge to $\nu + \int_0^t \alpha e^{-\beta(t-s)}dN_s$, which equals $\lambda(t)$. As for the conditional expectation of last term

$$E_t\Big[\alpha\int_t^{t+h}\alpha e^{-\beta(t+h-s)}dN_s\Big|N_{t+h} = n\Big] = \int_{0<h_1<\ldots<h_{n-N(t)}<h}\alpha\sum_{i=1}^{n-N(t)}e^{-\beta(h-h_i)}d\mu, \tag{3.37}$$

where μ is the joint distribution of waiting time of such $n - N(t)$ many jumps occurred between time t and $t+h$. The above term converges to $(n - N(t))\alpha$ as h approaches to zero. □

.

Corollary 3.5.2. *If the intensity of the default counting process N_t satisfies the Hawkes dynamics* (3.32)*, the implied initial local intensity surface $\Lambda_0(T, n)$ can not match given arbitrary initial local intensity $a_0(T, n)$ for $0 \leq n \leq N$.*

Proof. By Lemma 3.5.1, $\Lambda_t(t, N_t + n) = \lambda(t) + n\alpha$ for all $t \geq 0$. Necessarily, if we let t approach to zero, $\Lambda_0(0, n)$ has to satisfy

$$\Lambda_0(0, n) = \Lambda_0(0, 0) + [n - N(0)]\alpha.$$

That is, the terms $\Lambda_0(0, n)$ for all n are distinct only by a constant increment α. This violates the arbitrariness of the given initial local intensity $a_0(T, n)$. □

The reason of Hawkes intensity fails to match market data is caused by a constant jump size of intensity upon jump of N_t. In order to fix this problem. The jump size α has to depend on the number of default N_t

3.6 Calibration of initial local intensity surface to market data

In this section we briefly discuss the calibration problem. For a detailed discussion, refer to Carmona [2]. The model calibration problem for the initial local intensity surface is defined as the problem of recovering the local intensity function $\Lambda_0(T, K)$ from market observations, which consist of spreads for a small number of CDO tranches and index CDS.

Denote by $T_1 < \cdots < T_m$ the maturities of the observed CDO tranches (usually 3, 5, 7 and 10 years) with $T = T_m$ being the largest maturity and $0, K_1, \cdots, K_I$ the attachment points. The coupon payment dates are denoted as $\{t_j\}, j = 1, \cdots, J$. At $t = 0$ we observe the tranche spreads $(S_0(K_i, K_{i+1}; T_k), i = 1, ..., I-1; k = 1, ..., m)$. The calibration problem for the CDO pricing model can be formulated as follows:

Problem 3.6.1 (Calibration Problem). . *Given a set of observed CDO tranche spreads* $(S_0(K_i, K_{i+1}; T_k), i = 1, ..., I-1; k = 1, ..., m)$ *at time 0 for a reference portfolio, find the initial local intensity surface* $\Lambda_0(\cdot, \cdot)$ *so that the Markovian default counting process* $\tilde{N}(t)$ *and associated loss process* $\tilde{L}(t)$ *with local intensity* $\lambda(t) = \Lambda_0(t, \tilde{N}(t))$ *can produce the spreads to match the market observations:*

$$S_0(K_i, K_{i+1}; T_k) = \frac{\sum_{t_j \le T_k} B(0, t_j) E\left[(L_{[K_i, K_{i+1}]}(t_j) - L_{[K_i, K_{i+1}]}(t_{j-1}))\right]}{\sum_{t_j \le T_k} \delta_j B(0, t_j) E\left[(I(K_{i+1} - K_i) - L_{[K_i, K_{i+1}]}(t_j))\right]} \tag{3.38}$$

The above calibration problem as stated is ill-posed since it requires us to recover the whole local intensity surface from only a small number of tranche spreads. There is little chance of obtaining a unique solution, let alone to compute it in a stable manner. Nevertheless, we can approach calibration problem by the following ad-hoc methods. We proceed by decomposing the calibration problem into two sub-problems, and then attempt to solve them separately.

Problem 1

Since $L_{[a,b]}(t) = (L(t) - a)^+ - (L(t) - b)^+$, the right side of (3.38) only depends on expectations of the form $E(L(t_j) - K_i)^+$. Under the usual assumption of name homogeneity

assumption in 2.3.1, the equation (3.38) simplifies to

$$S_0(K_i, K_{i+1}; T_k) = \frac{\sum_{t_j \le T_k} B(0,t_j)\left[C_{i,j} - C_{i+1,j} - C_{i,j-1} + C_{i+1,j-1}\right]}{\sum_{t_j \le T_k}(t_j - t_{j-1})B(0,t_j)\left[(\tilde{K}_{i+1} - \tilde{K}_i) - C_{i,j} + C_{i+1,j}\right]}, \quad (3.39)$$

where $C_{i,j} = E[N(t_j) - \tilde{K}_i]^+$, and $\tilde{K}_i$ are the adjusted constant attachment points equal to $\frac{\tilde{N}}{1-R}K_i$. So our first goal is to extract the values of all expectations $\mathcal{C} = \{C_{i,j}\}$.

Problem 3.6.2 (Least square minimization problem). *Given a set of observed CDO tranche spreads* $(S_0(K_i, K_{i+1}; T_k), i = 1, ..., I-1; k = 1, ..., m)$ *at time 0 for a reference portfolio, extract a set of expectations* $\mathcal{C}$ *that solves the lease square minimization problem*

$$[C_{j,k}] = \arg\inf_{C} \sum_{i,m} w_{i,m}$$
$$\left| S_0(K_i, K_{i+1}; T_k) - \frac{\sum_{t_j \le T_k} B(0,t_j)\left[C_{i,j} - C_{i+1,j} - C_{i,j-1} + C_{i+1,j-1}\right]}{\sum_{t_j \le T_k}(t_j - t_{j-1})B(0,t_j)\left[(\tilde{K}_{i+1} - \tilde{K}_i) - C_{i,j} + C_{i+1,j}\right]} \right|^2 \quad (3.40)$$

where the weights $w_{i,m}$ *are chosen to be increasing in liquidity and inversely proportional to the bid-ask spread of the quotes of tranche spreads.*

This problem is also not well posed as there are more expectations $(I \times J)$ than the number of tranche spreads $(I \times m)$. So this discrepancy of numbers prohibits a unique solution of the minimization problem. Despite all that, it is common to assume that the values of

$$C_{i,j} = E[N(t_j) - \tilde{K}_i]^+, \qquad i = 1, ..., I; \quad j = 1, ..., J$$

are all known.

Problem 2

As we already know, the local intensity surface $\Lambda_0(t, K)$ and marginal distribution $p_K(t) = P(N_t = K)$ for $t > 0, K = 0, ..., \bar{N}$ are completely determined by each other through Dupire formula (3.6). So in order to extract the local intensity function $\Lambda_0(t, K)$, it suffices to extract the marginal distribution from $C_{i,j} = E[N(t_j) - \tilde{K}_i]^+$ obtained from Problem 1. So our second goal is to solve:

Problem 3.6.3 (Extrapolation problem). *Given a set of expectations*

$$C_{i,j} = E[N(t_j) - \tilde{K}_i]^+, \qquad i = 1, ..., I; \quad j = 1, ..., J$$

solve for the marginal distributions $p_K(t) = P(N_t = K)$ *for all* $t > 0, K = 0, ..., \bar{N}$

For $t = t_j$ and $K = \tilde{K}_i$, we have $p_K(t) = \nabla^- \nabla^+ E[N(t) - K]^+$. In general, the distribution $p_k(t)$ for the whole set of $t > 0$ and $K = 0, ..., \bar{N}$ can not be completely recovered from the mere knowledge of $C_{i,j}$ for a small number of K_i and t_j unless extra information on the distribution is available. There are many ways to extrapolate these values of C to obtain for each t a convex function of K which coincides with the value derived for all $K = \tilde{K}_i$. As in [5], one could have a finite sum of point masses or a prior distribution.

Alternative approach

A common alternative, ad-hoc approach is to assign an empirical parametric function form for $\Lambda(T, K)$ and then do parametric optimization.

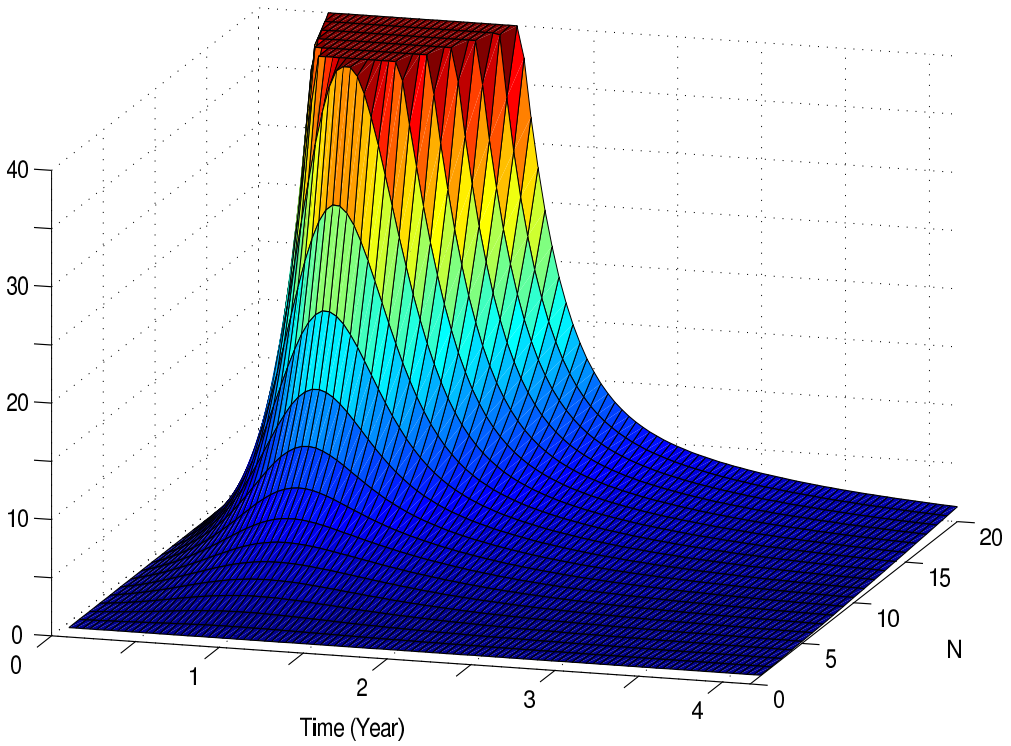

Figure 3.4: Graph of a model implied local intensity surface computed from parametric function: Above is a sample of local intensity surface $\Lambda_0(T, n)$, calibrated to the market data of the spreads of the tranches on Dow Jones CDX.NA.IG.7 quoted on January 12, 2007.

Chapter 4

Dynamics of the local intensity surfaces

4.1 Introduction

A local intensity model (3.1) assumes a fixed local intensity surface $\Lambda_t(T, n)$ for all $t \geq 0$ and hence the default counting process N_t is a one dimensional Markov process. The Markovian property of the underlying process N_t is not realistic. So in general, the local intensity surface should change with time and be viewed as a random surface $\Lambda_t(\cdot, \cdot)$. The idea of modeling the dynamics of the local intensity surface is along the lines of modeling the local volatility surface in equity case in Carmona and Nadtochiy [3].

If we use a Markovian loss model (3.1), the local intensity surface $\Lambda_t(T, n)$ is fixed with respect to t. The default counting process N_t is then a time-inhomogeneous Markov chain. The Markov property of N_t is unrealistic. The local intensity surface is generally changing with time, so in general it should be viewed as a function-valued random process.

Therefore, we consider modeling the dynamics of the local intensity surface. In this chapter, we will start with HJM style term structure model for the local intensity $\Lambda_t(T, n)$, as first proposed by Schönbucher in [23] for the equivalent forward transition rate. We will examine the no arbitrage conditions given by Schönbucher. The HJM style models are necessarily infinite dimensional, which makes implementation difficult for practical use. We then consider the finite dimensional realization problem, and propose parametric factors models consistent with the no arbitrage conditions. Finally, we will analyze some simple examples of the parametric factor models and briefly consider the pricing issue.

4.2 Term structure model of Schönbucher

Schönbucher in [23] discussed the modeling of the so called forward transition rate $a_n(t,T)$:

$$a_n(t,T) = \lim_{h \searrow 0} \frac{\mathbb{P}_t(N(T+h) = n+1 | N(T) = n)}{h} \tag{4.1}$$

It is clear that $a_n(t,T)$ is equivalent to $\Lambda_t(T,n)$, which is interpreted as a local intensity surface in this dissertation.

In Schönbucher, the dynamics of the forward transition rate $a_n(t,T)$ are modelled by the HJM style differential equation

$$da_n(t,T) = \mu_n(t,T)dt + \sigma_n(t,T)dW_t \tag{4.2}$$

So $a_n(T,n)$ is a diffusion process, driven by a Brownian motion W, where $n = 0, \cdots, \bar{N}$, $\mu_n(t,T)$ and $\sigma_n(t,T)$ are predictable process with all necessary regularity properties (to guarantee the existence and positiveness of $a_n(t,T)$). Actually $\mu_n(t,T)$ is uniquely determined by $\sigma_n(t,T)$ via a drift condition (which is much more complicated than the drift condition of HJM model).

4.2.1 No-arbitrage conditions

In order to derive the drift condition, Schönbucher first characterized the dynamics of the transition probability $P_{nm}(t,T) = \mathbb{P}(N(T) = m | N(t) = n)$. That is

$$dP_{nm}(t,T) = u_{nm}(t,T)dt + v_{nm}(t,T)dW \tag{4.3}$$

$$dP_{N(t),m}(t,T) = u_{N(t),m}(t,T)dt + v_{N(t),m}(t,T)dW + \phi_m(t,T)dL(t) \tag{4.4}$$

As the dynamics of $P_{N(t),m}(t,T)$ are already uniquely determined by the dynamics of $N(t)$ and $a_n(t,T)$, the above is notation and not an assumption. The parameters of the above dynamics (4.4) are given by straightforward computation:

$$\phi_m(t,T) = \begin{cases} 0 & m < N(t), \\ -P_{L(t-),m}(t,T) & m = N(t), \\ P_{L(t-)+1,m}(t,T) - P_{L(t-),m}(t,T) & m > N(t), \end{cases} \tag{4.5}$$

$$u_{nm}(t,T) = \begin{cases} 0, & m < n, \\ \big[a_n(t,t) - \int_t^T \mu_n(t,s)ds + \frac{1}{2}\big(\int_t^T \sigma_n(t,s)ds\big)^2\big] P_{nm}(t,T), & m = n, \\ -a_{m-1}(t,t)P_{mm}(t,T)\mathbf{1}_{m=n+1} + \int_t^T e^{-\int_s^T a_m(t,u)du} & \\ [-P_{nm}(t,s)\mu_m(t,s) + \mu^{Pa}_{n,m-1}(t,s) - \sigma_m(t,s)v_{nm}(t,s)]ds, & m > n, \end{cases} \tag{4.6}$$

$$v_{nm}(t,T) = \begin{cases} 0, & m < n, \\ P_{nm}(t,T)e^{-\int_t^T \sigma_n(t,s)ds}, & m = n, \\ \int_t^T e^{-\int_s^T a_m(t,u)du}[\sigma^{Pa}_{n,m-1}(t,s) - P_{nm}\sigma_m(t,s)]ds, & m > n, \end{cases} \tag{4.7}$$

where $\mu^{Pa}_{n,m-1}(t,s)$ and $\sigma^{Pa}_{n,m-1}(t,s)$ are the drift and diffusion coefficients of $P_{n,m-1}(t,T)a_{m-1}(t,T)$ for $m > n$,

$$\begin{cases} \mu^{Pa}_{n,m-1}(t,T) = a_{m-1}(t,T)u_{n,m-1}(t,T) + P_{n,m-1}(t,T)\mu_{m-1}(t,T) + \sigma_{m-1}(t,T)v_{n,m-1}(t,T) \\ \sigma^{Pa}_{n,m-1}(t,T) = P_{n,m-1}(t,T)\sigma_{m-1}(t,T) + a_{m-1}(t,T)v_{n,m-1}(t,T) \end{cases} \tag{4.8}$$

Proof of the above results is in Schönbucher(2005).

Proposition 4.2.1 (No-arbitrage condition). *The dynamics* (4.2) *are consistent with the default counting process* $N(t)$ *and admit no arbitrage if and only if the following conditions are satisfied:*

(i) The diffusion parameters of $a_n(t,T)$ *satisfy for all* $0 \le n \le N$, $t \le T$,

$$\textit{Drift Condition:} \quad P_{N(t),n}(t,T)\mu_n(t,T) = -\sigma_n(t,T)v_{N(t),n}(t,T), \tag{4.9}$$

(ii) the instantaneous intensity of $N(t)$ is given by

$$\textbf{\textit{Consistency Condition:}} \quad \lambda(t) = a_{N(t)}(t,t), \tag{4.10}$$

Proof of the above results is in Schönbucher(2005).

Observe that when $n < N(t)$ the drift condition doesn't exist, since $\Lambda_t(T,n)$ vanishes. When $n = N(t)$, the drift condition can be simplified as

$$\mu_t(T,n) = \sigma_t(T,n) \int_t^T \sigma_t(s,n)ds \tag{4.11}$$

Which has exactly the same form of HJM drift condition. When $n > N(t)$, unfortunately, the drift condition (4.9) is very complicated. which is the main obstacle to further work

Moreover, unlike the drift condition of the original HJM model, in which the drift parameter is completely determined by the diffusion parameter, not only $\sigma_n(t,T)$, but also the realization of $N(t)$ determine the drift parameter $\mu_n(t,T)$ in (4.9). So one of the main problems afterward is to specify a form of $\sigma_n(t,T)$ with necessary amenable properties. Also we need to analyze how the term $\sigma_n(t,T)$ affects the dynamics of the default counting process $N(t)$.

4.2.2 Static local intensity surfaces

In the dynamics of the local intensity surface,

$$da_n(t,T) = \mu_n(t,T)dt + \sigma_n(t,T)dW_t,$$

if $\sigma_n(t,T) = 0$ for all $t > 0$, then we have the dynamics of a deterministic local intensity model. From the drift condition (4.9)

$$P_{N(t),n}(t,T)\mu_n(t,T) = -\sigma_n(t,T)v_{N(t),n}(t,T)$$

and positiveness of $P_{N(t),n}(t,T)$, we have $\mu_n(t,T) = 0$. Consequently, $a_n(t,T) = a_n(0,T)$, the local intensity surface stays at the initial surface $a_n(0,T)$ and does not change at all. From the consistency condition (4.10), we obtain

$$\lambda_t = a_{N(t)}(t,t) = a_{N(t)}(0,t) = f(t,N(t)).$$

Hence, the instantaneous intensity λ_t only depends on time and number of defaults $N(t)$ through a bivariate function f. Therefore, if the local intensity does not move, the default counting process $N(t)$ is a time-dependent Markov chain. If it moves at all, it does so stochastically. In summary we prove

Proposition 4.2.2. *The local intensity surface is static if and only if the aggregate loss process is a time-dependent Markov chain.*

4.3 Parametric factor models

The HJM style term structure models for the dynamics of the local intensity surface $\Lambda_t(T,n)$ is general enough to cover the possible evolution of the local intensity surface. However, the infinite dimensionality of these models is a big obstacle to their implementation. As a solution to the finite dimensional realization problem. we look for a class of parametric factor models which is automatically finite dimensional. Namely, the dynamics of local intensity surface is controlled by a finite number of random factors. Such factors may or may not have economic meanings.

4.3.1 Parametric factor models

The static parametric family of models are usually introduced in the following way. We start from a function G from $\mathcal{Z} \times [0,\infty) \times \mathbb{Z}_{\geq 0}$ into $\mathbb{Z}_{>0}$, where $\mathcal{Z}$ is an open set in $\mathbb{R}^d$ which we interpret as the set of possible values of a vector Z of parameters $Z_1, \cdots, Z_d$. In this way, for each $Z \in \mathcal{Z}$ the surface $G(Z,\cdot,\cdot) : (T,n) \mapsto G(Z,T,n)$ can be viewed as a possible candidate for the local intensity surface.

We now introduce factor models from the notion of parametric family formalized above. We assume that we are given a parametric family G as before and we suppose that $\mathbf{Z} = \{Z_t\}_{t\geq 0}$ is a d-dimensional semi-martingale with values in the parameter space $\mathcal{Z}$. We then set

$$\Lambda_t(T,n) = G(Z_t,T,n), \qquad T \geq t \geq 0, n = 0, \cdots, \bar{N} \tag{4.12}$$

Definition 4.3.1 (Parametric factor model). *We assume that the time t local intensity surface of a default counting process is specified directly by:*

$$\Lambda_t(T,n) = G(Z_t;T,n), \tag{4.13}$$

where $Z_t \in \mathcal{Z} \subset \mathbb{R}^d$ is an $\mathcal{Z}$ valued d-dimensional Markov process which is a strong solution to a stochastic differential equation,

$$dZ_t = \mu(t,Z_t)dt + \sigma(t,Z_t)\cdot dW_t, \tag{4.14}$$

and $G : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0} \to \mathbb{R}_{>0}$ is an appropriately chosen positive function.

The process Z_t is interpreted as economic factors driving the dynamics of the term structure of local intensity surface. If Z_t is a constant process, then the parametric factor model (4.13) is reduced to Markovian loss model due to proposition 4.2.2. Usually we do not put t as an argument in the G function explicitly, since we may consider one component of the random vector Z_t be time t.

Though the finite dimensionality is a direct result of the model specification, the no-arbitrage conditions should still be considered. For a certain G function, assuming further that G is twice continuously differentiable in the variables Z^j, we can use Itô's formula to derive the dynamics of $\Lambda_t(T,n)$. The drift condition (4.9) then imposes restrictions on μ and σ. So the main problem in analyzing the parametric factor models is to find appropriate functions G so that the drift condition imposed on μ and σ can be greatly simplified.

4.3.2 Examples

Given the general form (4.13) of the parametric factor models for the local intensity dynamics, we analyze several examples in this section. Some of these examples are trivial ones which exhibits the restrictiveness of the drift condition (4.9). One non-trivial example is given in the end.

1) Random scaling

We first consider a very simple specification of a one factor model. The time t local intensity $\Lambda_t(T,n)$ is obtained by multiplying a positive factor process Z_t with the initial local intensity $\Lambda_0(T,n)$. that is,

$$\Lambda_t(T,n) = Z_t \cdot \Lambda_0(T,n), \qquad Z_0 = 1. \tag{4.15}$$

We can show that this specification reduces to the trivial static local intensity case:

Proposition 4.3.2. *The drift condition* (4.9) *forces Z_t in the random scaling model* (4.15) *to be constant. Therefore the random scaling model is reduced to the Markovian loss model.*

Proof. Taking the differential of $\Lambda_t(T,n)$ under the random scaling specification (4.15), we have the following dynamics

$$d\Lambda_t(T,n) = \Lambda_0(T,n)dZ_t \tag{4.16}$$

$$= \mu(t,Z_t)\Lambda_0(T,n)dt + \sigma(t,Z_t)\Lambda_0(T,n)dW_t \tag{4.17}$$

For $n = N_t$, consider the drift condition at $n = N_t$ (4.11). The following equation must hold.

$$\mu(t,Z_t)\Lambda_0(T,n) = \sigma(t,Z_t)\Lambda_0(T,n)\int_t^T \sigma(t,Z_t)\Lambda_0(s,n)ds \tag{4.18}$$

$$= \sigma^2(t,Z_t)\Lambda_0(T,n)\int_t^T \Lambda_0(s,n)ds. \tag{4.19}$$

This implies that

$$\mu(t,Z_t) = \sigma^2(t,Z_t)\int_t^T \Lambda_0(s,n)ds. \tag{4.20}$$

Since equation (4.20) holds for all $T \geq t$, and $\Lambda_0(S,n)$ is positive. we must have

$$\mu(t,Z_t) = \sigma(t,Z_t) = 0 \quad \text{for arbitratry } Z_t. \tag{4.21}$$

Therefore, $dZ_t = 0$ and $Z_t \equiv Z_0 = 1$. the local intensity surface is actually a constant surface and according to proposition (4.2.2), the random scaling model reduces to a Markovian loss model. □

2) Random shifting

We next consider another similar simple specification of a one factor model. The time t local intensity $\Lambda_t(T,n)$ is obtained by adding a factor process Z_t to the initial local intensity $\Lambda_0(T,n)$. i.e.,

$$\Lambda_t(T,n) = \Lambda_0(T,n) + Z_t, \qquad Z_0 = 0.$$

We can show that this specification also reduces to the trivial static local intensity case.

Proposition 4.3.3. *For some deterministic initial local intensity surface $\Lambda_0(T,n)$ and a Markovian random factor Z_t with $Z_0 = 0$. The drift condition* (4.11) *forces Z_t to be constant. Therefore, the random shifting model is reduced to Markovian loss model.*

Proof. Taking the differential of $\Lambda_t(T,n)$ under the random shifting specification (4.22), we have the following dynamics

$$d\Lambda_t(T,n) = dZ_t \tag{4.22}$$

$$= \mu(t,Z_t)dt + \sigma(t,Z_t)dW_t \tag{4.23}$$

For $n = N_t$, consider the drift condition at $n = N_t$ (4.11), the following equation must hold.

$$\mu(t,Z_t) = \sigma(t,Z_t)\int_t^T \sigma(t,Z_t)ds \tag{4.24}$$

$$= \sigma^2(t,Z_t)(T-t). \tag{4.25}$$

This implies that

$$\mu(t,Z_t) = \sigma^2(t,Z_t)(T-t). \tag{4.26}$$

Since this equation holds for all $T \geq t$, we must have

$$\mu(t,Z_t) = \sigma(t,Z_t) = 0, \quad \text{for arbitratry } Z_t, \tag{4.27}$$

and therefore, $dZ_t = 0$ and $Z_t \equiv Z_0 = 0$. The local intensity surface is actually a constant surface and according to proposition (4.2.2), the random scaling model reduces to the Markovian loss model. □

Can we find non-trivial examples?

3) Individual spread factor model

Even though the general drift condition (4.9) is very complicated. It simplifies in

$$\sigma_n(t,T)\begin{cases} > 0, & n = N(t), \\ \equiv 0, & n > N(t). \end{cases} \tag{4.28}$$

Namely, the diffusion of local intensity surface $\Lambda_t(T,n)$ only applies to $n = N_t$ not $n > N_t$. Then, according to (4.9), the drift condition becomes

$$\mu_n(t,T) = \begin{cases} \sigma_n(t,T)\int_t^T \sigma_n(t,s)ds, & n = N(t), \\ 0, & n > N(t). \end{cases} \tag{4.29}$$

The specification (4.28) was called the Individual Spread Blow-Out Model at Schönbucher [23], because the factor only drives the intensity rate of the next default, and does not affect the remaining transition rates.

Inspired by the result of finite dimensional realization of original HJM model for forward interest rate by Ritchken [20], we consider the following parametric factor model

Definition 4.3.4 (Individual spread factor model). *We assume the time t local intensity surface is given by the following specification:*

$$\Lambda_t(T,n) = \begin{cases} 0, & n < N(t), \\ \Lambda_0(T,n), & n > N(t), \\ G(t, Z_1(t), Z_2(t); T, n), & n = N(t), \end{cases} \tag{4.30}$$

The parametric functional G is defined as

$$G(t, z_1, z_2; T, n) = \Lambda_0(T,n) + z_2 H(t,T) - (\Lambda_0(t,n) - z_1)K(t,T) \tag{4.31}$$

where $H(t,T) = e^{-\int_t^T \kappa(s)ds}(\int_t^T e^{-\int_t^u \kappa(s)ds}du)$ and $K(t,T) = e^{-\int_t^T \kappa(s)ds}$, where the dynamics of the factors Z_1 and Z_2 are given by

$$\begin{cases} dZ_1(t) &= \mu(Z_1,t)dt + \sigma(t)dW(t), \qquad Z_1(\tau_n) = \Lambda_{\tau_n}(\tau_n, n), \\ dZ_2(t) &= [\sigma^2(t) - 2\kappa(t)Z_2(t)]dt, \qquad Z_2(\tau_n) = 0, \end{cases} \tag{4.32}$$

with

$$\mu(Z_1,t) = \kappa(t)[\Lambda_0(t,n) - Z_1(t)] + Z_2(t) + \frac{\partial}{\partial t}\Lambda_0(t,n), \tag{4.33}$$

where $\kappa(s)$ is a deterministic function, $\sigma(s)$ is an adapted process, and τ_n is the n-th default time.

Proposition 4.3.5. *The Individual Spread Factor Model* (4.30) *satisfies the no-arbitrage condition* (4.9).

Proof. The proof is straightforward by using the drift condition for $n = N(t)$. In fact, at t, the only non-trivial drift condition is for the n that equals N_t, and the drift condition is exactly same as the drift condition considered for the HJM model. Then, we assign the parametric form (4.31) as in Ritchken and Sankarasubramanian [20], which leads to the desired result immediately. □

4.3.3 Derivatives pricing by parametric factor models

Suppose we have a 1-dimensional parametric factor model of the local intensity surface;

$$\begin{aligned} \Lambda_t(T,n) &= G(Z_t;T,n) \\ dZ_t &= \mu(t,Z_t)dt + \sigma(t,Z_t)dW_t \end{aligned} \tag{4.34}$$

We assume the $N(0)=0$ and $Z(0)>0$ are given initial data. G is a certain appropriately chosen parametric function. μ and σ in the dynamics of factor Z_t satisfies the no arbitrage condition. Unlike stochastic intensity models, where we assign a dynamics for the stochastic involution of instantaneous intensity $\lambda(t)$. Here we don't actually consider anything on the instantaneous intensity. In this section, we consider how to price multi-name credit derivatives, given the model (4.34),

European style derivatives

Consider a European-style contingent claim c with payoff function $f(N_T)$ at maturity T. The time t distribution of N_T, $T \geq t$ is completely determined by N_t and Z_t. In fact, at time t, the distribution of N_T is characterized by the local intensity surface at time t, which is completely determined by Z_t. In terms of the transition probability, $\mathbb{P}(N_T = n|N_t)$ satisfies equation (3.5), which reads at

$$P_{N_t,n}(t,T) = \begin{cases} 0, & n < N_t, \\ e^{-\int_t^T G(Z_t;s,n)ds}, & n = N_t, \\ \int_t^T P_{N_t,n-1}(t,s)G(Z_t;s,n-1)e^{-\int_s^T G(Z_t;u,n)du}ds, & n > N_t. \end{cases} \tag{4.35}$$

Given the above distribution, we are able to compute the price at time t as the expectation $E_t[f(N_T)]$. Therefore c can be written as $c(t,N_t,Z_t)$. The price c can also be computed by solving the backward equation as in (4.36).

$$\frac{\partial}{\partial t}c(t,n,z) + G(z,t,n)(c(t,n+1,z) - c(t,n,z)) = 0 \tag{4.36}$$

for all n, with terminal condition $c(T,n,z) = f(n)$.

Path dependent derivatives

Here we consider the pricing problem of forward starting product. A typical example is forward starting CDO. Such product starts at an effective date $T_e > 0$ and expires at a maturity date $T > T_e$. The price of forward starting products depends on the joint distribution of N_T and N_{T_e}. It can be usually written as $E[f(N_T, N_{T_e})]$

Given a stochastic intensity model of N_t, the generic method is to use Monte Carlo simulation. which need to simulate the whole path of N_t and σ_t from 0 to T.

However, for solving the pricing problem in our parametric factor dynamic local intensity model, the method to be used is different. We first suppose there is a forward starting product starts at T_e and matures at T. The time T_e value of such product is simply an European-style product

$$V(T_e) = E_{T_e}[f(T, T_e)] := V^*(T_e, N_{T_e}, Z_{T_e}) \tag{4.37}$$

where V^* is the solution of the backward equation (4.36) with terminal condition $V(T, n, z) = f(n, N_{T_e})$. The time 0 price is then the expectation of $V^*(T_e, N_{T_e}, Z_{T_e})$

$$V(0) = E[V^*(T_e, N_{T_e}, Z_{T_e})] \tag{4.38}$$

To evaluate this expectation, we have to use Monte Carlo simulation for both N_t and Z_t for t ranging from 0 to T_e.

To sum up, the pricing problem in our parametric factor dynamic local intensity model requires a Monte Carlo simulation for time ranging from 0 to T_e plus solving a backward equation for time ranging from T_e to T. The computing time needed is shorter than an entirely Monte Carlo simulation for time ranging from 0 to T as in stochastic intensity models.

References

[1] D. Brigo, A. Pallavicini, and Torresetti, *Calibration of CDO tranches with the dynamical generalized-Poisson loss model*. Available at www.defaultrisk.com, 2006.

[2] R. Carmona, *HJM, A unified approach to dynamic models for fixed income, credit and equity markets*, (2007). Available at http://orfe.princeton.edu/~rcarmona/download/fe/HJMlnm3.pdf.

[3] R. Carmona and S. Nadtochiy, *Local volatility dynamic models*, Finance and Stochastic, 13 (2009), pp. 1–48.

[4] R. Cont and A. Bentata, *Mimicking the marginal distributions of a semimartingale*, (2009). Available at http://arxiv.org/abs/0910.3992.

[5] R. Cont and A. Minca, *Recovering portfolio default intensitites implied by CDO quotes*, (2008). Financial Engineering Report No. 2008-01.

[6] D. Duffie and N. Garleanu, *Risk and valuation of collateralized debt obligations*, Financial Analysts Journal, 57 (2001), pp. 41–59.

[7] B. Dupire, *Pricing with a smile*, Risk, 7 (1994), pp. 18–20.

[8] P. Evan and R. Sircar, *Multiscale intensity models and name grouping for valuation of multi-name credit derivatives*, Applied Mathematical Finance, 15 (2008), pp. 73–105.

[9] L. Evans, *Partial Differential Equantions*, AMS, 1998.

[10] Frey, Rüdiger, and Backhaus, *Portfolio credit risk models with interacting default intensities: a Markovian approach*, (2004). Available at http://www.math.uni-leipzig.de/∼frey/interacting-intensities-final.pdf.

[11] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, 1964.

[12] I. Gyöngy, *Mimicking the one dimensional marginal distributions of processes having an Itô differential*, Probability Theory and Related Fields, 71 (1986), pp. 501–516.

[13] I. Halperin and M. Arnsdorf, *Markovian bivariate spread loss model for portfolio credit derivatives*, (2007). Available at http://arxiv.org/pdf/0901.3398.

[14] J.Jacod and A.Shiryaev, *Limit theorems for stochastic processes*, Springer Berlin, 1988.

[15] H. G. Kellerer, *Markov property of point processes*, Probability Theory, 76 (1987), pp. 71–80.

[16] H. Kumanogo, *Pseudo-differential operators*, MIT, 1981.

[17] D. Li, *On default correlation: A copula approach*, Journal of Fixed Income, 9 (2000), pp. 43–54.

[18] A. Lopatin and T. Misirpashaev, *Two-dimensional Markovian model for dynamics of aggregate credit loss*, (2007). Available at http://www.defaultrisk.com/pp_model159.htm.

[19] D. Oakes, *The Markovian self-exciting process*, Journal of Applied Probability, 12 (1975), pp. 69–77.

[20] P. Ritchken and L. Sankarasubramanian, *Volatility structures of forward rates and the dynamics of the term structure*, Mathematical Finance, 5 (1995), pp. 55–72.

[21] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales*, Wiley, 2nd ed., 1994.

[22] P. J. Schö*nbucher*, *Credit Derivatives Pricing Models*, Wiley, 2003.

[23] P. Schonbucher, *Portfolio losses and the term structure of loss transition rates: A new methodology for the pricing of portfolio credit derivatives*, (2005).

[24] F. Treves, *Introduction To Pseudo-differential And Fourier Integral Operators*, Plenum Press, 1980.

[25] O. Vasicek, *Loan portfolio value*, Risk, 15 (2002).

Curriculum Vitae

Ming Shi

2004-2010 Ph. D. in Mathematics, Rutgers University, Piscataway, NJ, USA

2001-2004 M. Sc. in Mathematics, Fudan University, Shanghai, China

1997-2001 B. Sc. in Mathematics, Fudan University, Shanghai, China

2004-2009 Teaching Assistant, Department of Mathematics, Rutgers University