Numerical Solutions to a Class of Stochastic Partial Differential Equations Arising in Finance

Karolina Bujok
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2013

To my Parents

Acknowledgements

I am very happy to thank my supervisor, Dr. Christoph Reisinger, for inspiring discussions, support and enthusiasm.

I am grateful to Prof. Ben Hambly for inspiring ideas that helped to improve the article [11], that is the basis of Chapter 3 of this thesis. Also, I would like to thank Prof. Mike Giles for suggesting the improved estimator in the aforementioned article.

I would to acknowledge anonymous referees commenting on my and co-authors work in [11] and [12] for their helpful feedback.

I am also grateful to Prof. Kees Oosterlee for comments that helped to improve the thesis.

I would like to thank academics from the Mathematical and Computational Finance Group at Oxford and the Oxford-Man Institute of Quantitative Finance for their helpful discussions and comments. I would especially like to thank Dr. Jan Obłój, Dr. Łukasz Szpruch and Dr. Johannes Ruf. Additionally, I am grateful to my fellow D. Phil students, in particular to Sean Ledger for his feedback on Chapter 4 of this thesis, and to Ferhana Ahmad for her comments on Chapter 1.

I would like to acknowledge financial support from a CASE scholarship jointly made by the UK Engineering and Physical Sciences Research Council (EPSRC) and Nomura; also from St Catherine's College via an Alan Taylor Scholarship; and from the Oxford-Man Institute of Quantitative Finance.

Also, I would like to thank Dr. Sue Casson for proofreading parts of this thesis.

Special thanks to Prof. Aleksander Welfe, my supervisor from Warsaw School of Economics, and Dr. Martijn Pistorius, my supervisor from King's College London, for encouraging me to do D. Phil studies. Also, to Sławek Pawłowski and Dr. Paweł Gąsiorowski for stoking my interest in mathematical finance.

It is a great pleasure to thank my friends in Oxford: Philippe Aeberhard, Judith Campos-Cordero, Alessandro Di Nicola, Sigrid Källblad, Sheila Lin, Serena Lunardi, Asaf Paris-Mandoki, Tohru Seraku, Amrit Virk; and in Warsaw: Anna Jabłecka, Michał Miłowski, Magda Pruszyńska, Ola Świątkowska, Dorota Ukleja; for making my D. Phil studies such a memorable experience.

And above all, I would like to thank my parents, Zofia and Jan, and siblings, Ania and Robert, for support and encouragement.

Many thanks.

Abstract

We propose two alternative approaches to evaluate numerically credit basket derivatives in a N-name structural model where the number of entities, N, is large, and where the names are independent and identically distributed random variables conditional on common random factors.

In the first framework, we treat a N-name model as a set of N Bernoulli random variables indicating a default or a survival. We show that certain expected functionals of the proportion L_N of variables in a given state converge at rate $1/N$ as $N \to \infty$. Based on these results, we propose a multi-level simulation algorithm using a family of sequences with increasing length, to obtain estimators for these expected functionals with a mean-square error of ϵ^2 and computational complexity of order ϵ^{-2}, independent of N. In particular, this optimal complexity order also holds for the infinite-dimensional limit. Numerical examples are presented for tranche spreads of basket credit derivatives.

In the second framework, we extend the approximation of Bush et al. [13] to a structural jump-diffusion model with discretely monitored defaults. Under this approach, a N-name model is represented as a system of particles with an absorbing boundary that is active in a discrete time set, and the loss of a portfolio is given as the function of empirical measure of the system. We show that, for the infinite system, the empirical measure has a density with respect to the Lebesgue measure that satisfies a stochastic partial differential equation. Then, we develop an algorithm to efficiently estimate CDO index and tranche spreads consistent with underlying credit default swaps, using a finite difference simulation for the resulting SPDE. We verify the validity of this approximation numerically by comparison with results obtained by direct Monte Carlo simulation of the basket constituents.

A calibration exercise assesses the flexibility of the model and its extensions to match CDO spreads from precrisis and crisis periods.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter, we present possible applications of the thesis, indicate the place of our research in the existing literature, and present an outline of our work, together with the thesis contribution.

1.1 Motivation

The thesis is devoted to modelling the credit risk of a portfolio with a large number of defaultable assets. Special attention is paid to developing efficient numerical methods for such multi-name models. Let us start by giving a definition of credit risk and presenting two main credit derivatives.

1.1.1 Basic definitions

As in [73], p.1, we assume the following definition of credit risk.

Credit risk is the risk that an obligor does not honour his payment obligations.

Under this definition credit risk is equivalent to *default risk*, while in more general definitions, such as in [25] or [6], credit risk includes also the risk of the deterioration of the

obligor's credit quality.

The main components of credit risk are the *arrival of a default*, *exposure* and *recovery*. The arrival of a default is a random variable, then the exposure represents the amount that is affected in case of a default, and the recovery stands for the payoff that a creditor obtains after the default. The fraction of the received payment to the exposure is called the *recovery rate*. The recovery payment is random and depends, for example, on the type of default or guarantees received from the obliger. For details of the financial and legal aspects of credit risk, we refer to [6] and [73].

Assets subject to credit risk are called *defaultable assets*. Loans, mortgages, bonds and *credit derivatives* are examples of such assets.

Let us now introduce credit derivatives, the following presentation is motivated by [73]. While loans and bonds were the core of financial services from the beginning of banking, credit derivatives are more recent financial innovations. The market for credit derivatives was developed only in early 1990. It developed rapidly till the credit crisis of 2008-2009, and currently the credit market has started a slow recovery.

Credit derivatives are financial instruments that transfer credit risk from the protection buyer to the protection seller. The contracts can be based on a single entity or on a basket of names. The most frequently traded single-name credit derivative is the *credit default swap* (CDS), while the most common basket credit derivative is the *collateralized debt obligation* (CDO).

In the CDS contract, the seller compensates the buyer if a default of the reference entity happens. The buyer of the CDS pays a fee to the seller till a default or maturity of the contract, whichever occurs first.

A CDO is a set of fixed income securities whose payments depend on credit events in a pool of defaultable assets. It can consist, for example, of loans, mortgages or credit default swaps. In the latter case, such a contract is called a *synthetic CDO*.

A CDO offers products with different risk profiles, based on so-called *tranches*. The

risk connected to the tranches varies, from *equity tranches*, which are typically unrated or speculative-grade, investment-grade *mezzanine tranches*, to *senior tranches*, which can even have an AAA rating. By dividing the pool of defaultable assets into tranches, for instance, investors who would not be allowed to invest directly in the underlying assets, because their ratings are too low, can invest in senior tranches. Each tranche of a CDO is defined by an *attachment point*, a, and a *detachment point*, $d > a$, which are typically given as a percentage of a portfolio notional, e.g., for the equity tranche typically $a = 0\%$ and $d = 3\%$. The tranche notional is given as the difference between a and d.

Credit models provide the probability of the default of a single reference entity or joint default probability of a portfolio, defining the major component of the credit risk. The models can be used to price single and multi-name credit derivatives and also as a tool for credit risk management. Let us now present pricing formulas for the two main credit derivatives, and then risk measures that help to manage credit risk.

1.1.2 Pricing formulas for CDS and CDO

Here, we derive pricing formulas for the credit derivatives that we analyse in the thesis. For an overview of the pricing approaches, we refer for example to [73] or [9].

In the thesis, we assume that a CDS is written on the total debt of a single reference entity. We assume that the reference entity's shares or bonds are traded securities. Also, we consider a synthetic CDO, where the reference basket consists of CDSs.

Motivated by the presentation in [7], we make the following assumptions about the market that we consider. We assume that trading is frictionless and continuous in time. Also, for simplicity, the risk-free interest rate, r, and recovery rate, R, are constant. We assume that we are given a risk-neutral pricing measure $\mathbb{P}$, under which the discounted values of traded credit risky securities and also default-free securities are martingales, leading to the absence of arbitrage in the market. What is more, we assume that a defaultable bond is priced under this measure, giving a price processes for a family of

defaultable bonds with different maturities. Then, a CDS can be replicated by continuous trading in the reference entity's defaultable bonds or shares and default-free securities. Also, a synthetic CDO can be replicated by continuous trading in the underlying CDSs and default-free securities. We come back to the above replication in Section 1.2.1, where we introduce a diffusion structural model and in Section 2.1.1, where we present a jump-diffusion structural model considered throughout the thesis. We comment on why under the former model, the above hedge of a CDS is perfect, and why this is not the case under the latter model.

Let us denote the payment dates of a CDS or CDO by T_j, where $1 \leq j \leq n$, the payments intervals by $\varrho_j = T_j - T_{j-1}$, the value of a bank account at time t by b_t. Also, we denote by τ the default time of a reference entity.

Let us start by deriving a formula for a spread of a CDS. We assume that the notional of the contract is unity. The fee leg of a CDS, $c^{CDS}\bar{B}$, equals

$$c^{CDS}\bar{B} = c^{CDS} \sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[1_{\{\tau > t\}}],$$

where c^{CDS} is a CDS spread, and $\mathbb{E}[1_{\{\tau>t\}}] = \mathbb{P}[\tau > t]$ is the survival probability of the reference entity under the risk neutral measure $\mathbb{P}$. Then, the protection leg, $\bar{D}$, is given by

$$\begin{aligned}\bar{D} &= (1-R) \sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[1_{\{T_{j-1} < \tau \leq T_j\}}] \\ &= (1-R) \sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[1_{\{T_{j-1} > \tau\}} - 1_{\{T_j > \tau\}}].\end{aligned}$$

Since at time $t = 0$, the value of the fee and protection legs are equal, $c^{CDS}\bar{B} = \bar{D}$, then

the CDS spread is given by

$$c^{CDS} = \frac{(1-R)\sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[1_{\{T_{j-1} > \tau\}} - 1_{\{T_j > \tau\}}]}{\sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[1_{\{\tau > t\}}]}.$$

Let us now derive pricing formulas for a single CDO tranche and for a CDO index. We consider a synthetic CDO, where the reference portfolio consists of N CDSs, each with notional $N_0 = 1/N$. Let us start from the pricing formula for a CDO index. The outstanding notional of the portfolio at time t equals

$$\tilde{Z}_t = N_0 \sum_{i=1}^{N} 1_{\{\tau_i > t\}},$$

where τ_i is the default time of the i-th entity. The fee leg of a contract, $c^{IND}\tilde{B}$, is given by

$$c^{IND}\tilde{B} = c^{IND} \sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[\tilde{Z}_{T_j}],$$

where c^{IND} is a CDO index spread. The protection leg equals

$$\tilde{D} = (1-R) \sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[\tilde{Z}_{T_{j-1}} - \tilde{Z}_{T_j}].$$

Since at time $t = 0$ value of fee and protection legs are equal, $c^{IND}\tilde{B} = \tilde{D}$, then the CDO index spread is given by

$$c^{IND} = \frac{(1-R)\sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[\tilde{Z}_{T_{j-1}} - \tilde{Z}_{T_j}]}{\sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[\tilde{Z}_{T_j}]}.$$

Let us now present the formula for a single CDO tranche, following [13]. The loss of the portfolio, L, at time t, is defined as

$$L_t = N_0(1-R) \sum_{i=1}^{N} 1_{\{\tau_i \leq t\}}. \tag{1.1}$$

The outstanding notional, Z, of a single tranche is given by

$$Z_t = [d - L_t]^+ - [a - L_t]^+.$$

The single CDO tranche protection buyer pays a running spread c^{CDO} on the outstanding tranche notional. The value of the fee leg of the contract, Bc^{CDO}, is given by

$$Bc^{CDO} = c^{CDO} \sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[Z_{T_j}],$$

while the value of the protection leg, D, equals

$$D = \sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[Z_{T_{j-1}} - Z_{T_j}].$$

Since, at time $t = 0$, the fair values of fee and protection legs have to be equal, $Bc^{CDO} = D$, the spread of a single tranche is given by

$$c^{CDO} = \frac{\sum_{j=1}^{n} \frac{1}{b_{T_j}} \mathbb{E}[Z_{T_{j-1}} - Z_{T_j}]}{\sum_{j=1}^{n} \frac{\varrho_j}{b_{T_j}} \mathbb{E}[Z_{T_j}]}.$$

In the above derivation, we do not specify how often a default is monitored. However, throughout the thesis we distinguish between default times τ^M, τ^D and τ^A. In the case of τ^M, a default can only be defected at the maturity of a contract, whereas for τ^D a default is monitored at a discrete set of times, and for τ^A a default can happen at any time.

1.1.3 Managing and measuring credit risk

Credit risk on the macro scale is managed via regulatory authorities that aim at increasing the soundness and stability of financial systems.

On the international scale, the financial system is regulated by the Basel Accord. As noted in [33], Basel I was introduced in 1988 by the Basel Committee on Banking

Supervision. Then, in 2004, the regulations were strengthened with Basel II, based on three 'pillars' of regulations. In Pillar I, minimal capital requirements were set to cover potential losses due to credit, market and operational risks. Pillar II gives an institution and its supervisors a choice over deciding if any additional capital for risks, not mentioned in Pillar I, is needed. Also, Pillar III requires the institutions to publish more details of their risks and capital. Currently, the works on Basel III are ongoing, with the hope of addressing the deficiencies of the previous regulations revealed in the recent credit crisis of 2008-2009. One of the crucial recommendations of the Basel III is including the risk of the default of a counterparty in the value of a transaction with the counterparty. Such an adjustment to the value of a transaction is called *credit value adjustment* (CVA). Including CVA into valuation results in higher Pillar I capital requirements.

On the micro scale, managing credit risk can be based, for example, on using measures indicating the risk of a defaultable portfolio, giving an estimate of the required capital to hold, and also on relieving this extra capital by buying protection against a default of reference entities. The latter approach is obtained via credit derivatives, discussed above, hence let us now focus on the risk measures.

One of the risk measures suggested in Basel II and commonly used in the industry is the *Value at Risk* (VAR). We define VAR as the following, motivated by the presentation given in [31]. Let $\alpha \in (0,1)$ be some confidence level, and L be the loss of a portfolio, then VAR of the portfolio at a confidence level α is given as

$$VAR_\alpha(L) = \inf\{l \in \mathbb{R} : \mathbb{P}[L \geq l] \leq 1 - \alpha\}.$$

As noted, for example, in [36] or [31], VAR suffers from several deficiencies, particularly the fact that it fails to take into account the size of the losses and may penalize diversification of capital. An example of an improved risk measure is the *average VAR*

(AVAR), defined for a confidence level $\alpha \in (0,1)$ as

$$AVAR_\alpha(L) = \frac{1}{1-\alpha}\int_0^\alpha VAR_\theta(L)\,d\theta.$$

We refer to [36] and [31], and reference therein for other risk measures and for a comparative analysis of them.

As we can see, the main ingredient in VAR and AVAR is the loss of a portfolio. In particular, the tail of the distribution is considered in both presented risk measures.

1.1.4 Applications of credit models with a large number of names

Multi-name models with a large number of assets can be used to price a CDO, since typically a portfolio of a CDO consists of over 100 entities. What is more, portfolios of loans or mortgages in banks can be easily counted in thousands. Therefore, multi-name models can provide VAR or other risk measures for such portfolios. In the thesis, we focus on developing a framework for multi-name credit models where the number of entities is large.

1.2 Main approaches in credit modelling

In this section, we provide a brief review of credit modelling. The presentation is by no means exhaustive, since our aim is to show the place of our work in the existing literature rather than to provide a complete survey of this prolific field. For such a survey, we refer to [7], [73], [34], [25], and also to [32], [18] for review of models used in practice.

By looking at the pricing formulas of CDSs, CDOs and risk measures, given in the previous section, it is clear that the crucial issue is to define a default event and then develop a way to compute the probability of a default under the proposed definition. Then, the dependence between companies in a portfolio needs to be proposed together with a method of calculating joint default probability, and the last step is to develop an

approach that enables the calculation of multiple defaults in portfolios consisting of a large number of entities.

1.2.1 Modelling of default events

Most of the credit models that are currently used fall into one of two categories: *reduced form* and *structural*. In the first approach, the default event is exogenous, i.e. the models do not explain why the default happens, whereas in the second approach, the main focus is on giving the conditions leading to the default.

Reduced form models

The following presentation of reduced form models is motivated by [34] and [7].

In the reduced form model, a default happens unexpectedly and is defined as a first jump of a point process of size one. Let $H = \{H_t, 0 \leq t \leq T\}$ be a point process such that $H_t = 1_{\{\tau<t\}}$. The rate of a default is given by the *intensity* of the point process, λ.

In the simplest settings, where $\lambda > 0$ is a constant, H becomes a time-homogeneous Poisson process, and the default probability is given as

$$\mathbb{P}(H_t = 1) = 1 - e^{-\lambda\, t}.$$

Such a definition of a default is assumed for example in a model of [60].

In many reduced form models, such as [24] or [57], $\lambda = \{\lambda_t, t \in [0,1]\}$ is a stochastic process, which leads to H being a *doubly stochastic Poisson process* or *Cox process*. This means that, conditional on realisations of λ, H is a time-inhomogeneous Poisson process. Then, the probability of default equals

$$\mathbb{P}(H_t = 1) = 1 - \mathbb{E}[e^{-\int_0^t \lambda(u)\, du}]. \tag{1.2}$$

The randomness of intensity is introduced via some risk factors, $\lambda = \lambda(t, \Xi)$, where in

these models $\Xi = \{(\Xi_t^1, \ldots, \Xi_t^d), d \in \mathbb{N}, t \in [0,T]\}$, and Ξ satisfies some SDE.

As noted in [34], a closed-form solution of (1.2) is available for a broad class of affine models, introduced by [22] for interest rate modelling. Then, (1.2) becomes

$$\mathbb{P}(H_t = 1) = 1 - e^{C(t) - D(t)\Xi_0},$$

where Ξ_0 is the initial state of risk factors, and $C(t)$ and $D(t)$ are some non-random functions of time. For example, we obtain an affine model, if $\lambda = \Xi$, and the dynamics of Ξ are given by a mean-reverting process

$$d\Xi_t = c(\mu - \Xi_t)\, dt + \sigma W_t,$$

where W is a standard Brownian motion, and μ, c and σ are positive constants, $\Xi_0 \in \mathbb{R}$, then, $C(t)$ and $D(t)$ are given as

$$\begin{aligned} D(t) &= \frac{1}{c}\left(1 - e^{-ct}\right) \\ C(t) &= \mu(B(t) - t) + \frac{\sigma^2}{2c^2}\left(t - 2D(t) + \frac{1}{2c}\left(1 - e^{-2ct}\right)\right). \end{aligned}$$

Because of the explicit formulas for default probabilities, reduced models are commonly used in practice. However, they do not have a direct economic grounding, unlike the structural models, where the main focus is on the reasons behind a company's default.

Structural models

Structural models are based on the microeconomic observation that a default depends on the value of a company's assets versus its debt. In these models, a default happens when the asset value falls below a given barrier. [64] assumes that a firm's asset value follows

geometric Brownian motion

$$dA_t \quad = \quad \mu\, A_t\, dt + \sigma A_t\, dW_t, \tag{1.3}$$

where A is a company's asset value, W a standard Brownian motion, c^B a constant barrier, $A_0 > 0$, μ is a drift parameter, and $\sigma > 0$ is the volatility. Also, it is assumed that the company can default only at the maturity of the debt. Then, the time of a default, τ^M, is defined as

$$\tau^M = \begin{cases} T & \text{if } A_T \leq c^B \\ \infty & \text{otherwise.} \end{cases}$$

It is assumed that under a risk-neutral measure the drift parameter equals the risk-free interest rate, r, giving

$$dA_t \quad = \quad r\, A_t\, dt + \sigma A_t\, dW_t. \tag{1.4}$$

Also, the risk-neutral pricing measure is unique, leading to the completeness of the market and perfect hedge. In particular, a defaultable bond, considered in [64], is perfectly replicated by investing in the company's asset value and default-free bonds. Then, it can be argued that also a CDS is perfectly hedged via a similar replicating portfolio.

However, as noted in [7], dynamics of A given by (1.4) are justified only if A represents a traded security or at least can be replicated by trading in the firm's shares or bonds and default-free securities. In practice the company's asset value is neither a traded security nor is observable. Hence, in the thesis, we consider only a CDS written on the reference entity whose value of assets can be constructed from the firm's shares or bonds and default-free securities. For the details concerning obtaining data for the process A and its parameters, we refer for example to [34] or [62].

By applying Itô's lemma to (1.4), we have

$$A_T = A_0 \, e^{\tilde{\beta}T + \sigma W_T},$$

where $\tilde{\beta} = r - \frac{1}{2}\sigma^2$. The probability of default is given by a simple formula

$$\mathbb{P}(A_T \leq c^B) = \Phi\left(\frac{\log(\frac{c^B}{A_0}) - \tilde{\beta}T}{\sigma\sqrt{T}}\right),$$

where Φ is the standard normal distribution function.

In order to lose the direct dependence on c^B and σ, it is common in the literature to introduce a *distance-to-default*, X, defined as

$$X_t = \frac{1}{\sigma}(\ln A_t - \ln c^B).$$

By (1.4) and Itô's lemma, at time $t > 0$, X_t equals

$$X_t = X_0 + \beta t + W_t,$$

where $\beta = 1/\sigma(r - \frac{1}{2}\sigma^2)$. Then, a default event is defined as

$$\tau^M = \begin{cases} T & \text{if } X_T \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

and the probability of default equals

$$\mathbb{P}(X_T \leq 0) = \Phi\left(\frac{-X_0 - \beta T}{\sqrt{T}}\right). \tag{1.5}$$

Using first passage time theory, [8] extends the model by allowing default at any time.

In terms of *distance-to-default*, a default event is defined as

$$\tau^A = \inf\{t > 0 : X_t = 0\}, \tag{1.6}$$

and the probability of default is given as the following

$$\mathbb{P}(X_t = 0) = \Phi\left(\frac{-X_0 - \beta T}{\sqrt{T}}\right) + e^{-2\beta X_0}\, \Phi\left(\frac{-X_0 + \beta T}{\sqrt{T}}\right).$$

Because of the continuity of Brownian motion, in the above models, a default does not come as a complete surprise. As a consequence, diffusion models cannot explain high short time CDS spreads observed in the market. Also, the term structure of credit spreads given by the models starts at zero and is increasing, whereas in the market it can also be flat or downward sloping.

There are well documented improvements of structural credit models, for instance by considering: unobservable default barriers ([23]); asset value processes modelled by jump-diffusions ([80]; [44]; [45]; [52]; [62]; [61]; [4]) or more general Lévy processes ([28]); stochastic volatility ([30]) or stochastic recovery rate ([29]). However, for the improved models, explicit formulas of the default probabilities are usually not available. As a consequence, structural models are more of an academic interest.

1.2.2 Default dependence

In order to model joint defaults, the dependence structure between the obligors needs to be specified, preferably closely based on the empirical evidence of defaults.

Empirical evidence

As noted in [73], historically, defaults tend to cluster. For example, in 1982-1986, 22 oil industry companies defaulted; there were three defaults in airline companies in 1970-1971, followed by five more in 1989-1990; also more than 20 retailers defaulted in 1990-1992.

What is more, in [35], empirical evidence of default clustering is given for the past 150 years. For instance, many multiple defaults were observed during the Great Depression, with the severest ones occurring during the railroad crisis of 1873-1875, when defaults amounted to as much as 36 percent of the par value of the corporate bond market.

One reason behind the default clusters is that companies operating in the same country or industry sector work under similar economic conditions. When the conditions become worse, the situation for all companies deteriorates; however the scale of the effect can differ due to the individual condition of companies. For example, when interest rates increase, companies have to spend more on repaying their liabilities which, for companies that are already in a distressed situation, can lead to default while for others only to a lowering of their income. Another example is an exogenous shock that hits the economy, like a natural disaster that changes the situation of the companies in the affected area.

Multiple defaults can be also explained by the links between companies. It is commonly said that the economy is a system of communicating vessels; the prosperity of one company improves the financial situation of its contractors, just as the default of one company can lead to the worsening conditions of its suppliers or even to their default. For example, the credit crisis of 2008-2009, initiated by the collapse of the subprime mortgage market in the U.S. This affected financial institutions, for example via CDOs, which were based on U.S. mortgages, and then the global economy suffered via the links between the financial markets and the real economy, and then via the links between companies.

Default dependence in credit models

If there are two assets in a portfolio, then dependence can be modelled through the linear correlation between assets, for example, as in [81], where analytic formulas are available for a structural model. As noted in [73], having N assets in a portfolio, the number of correlation coefficients would be $N(N-1)/2$, determining only probability of bivariate default events. Since 2^N joint defaults are possible, taking into account all combinations

of default events between companies, we have to make simplifying assumptions.

One solution proposed in the literature is modelling dependence between obligors by introducing common factors that influence all companies in a portfolio and individual ones that are specific to each company. Then, conditional on common factors, the names are independent. Such models are also called *conditionally independent factor models* or *Bernoulli mixture models*.

Another idea is to capture the empirical evidence of a default of one company affecting another by introducing *contagious* or *infectious* defaults ([43], [47], [19], [37]). However, introducing contagious effects may break the conditional independence of names and makes the analysis less tractable.

In the industry, a popular solution is to introduce dependency between names via *copula functions* as, for example, in [60] or [4]. We refer to [26] and [66] for an analysis of the copula models. Since, in our thesis the focus is on Bernoulli mixture models, we present here this class of multi-name credit models.

Bernoulli mixture models

The main idea behind the common factor models is that, conditional on common factors, defaults are independent. If, additionally, the portfolio is homogeneous, meaning that the marginal distributions for each name in the portfolio are the same, then the defaults are also identically distributed. What is more, a default is a Bernoulli random variable, taking zero when there is no default and one otherwise.

A common way to incorporate the idea of systemic and idiosyncratic risks of a company into the structural framework is by correlated Brownian motions. Under this approach, the *distance-to-default* of the i-th company follows the process

$$X_t^i = X_0^i + \beta t + \sqrt{1-\rho}\, W_t^i + \sqrt{\rho}\, W_t^{\varsigma}, \tag{1.7}$$

where $i = 1, \ldots, N$, $\rho \in [0, 1)$ is a correlation coefficient, W^{ς} denotes a common factor, W^i is an individual factor, and both are independent standard Brownian motions, for all i.

In the simplest settings, as in [77], a default is monitored at maturity only. Let $\mathcal{F}^{\varsigma}$ be the filtration generated by W^{ς}. Observe that, conditional on $\mathcal{F}^{\varsigma}$, the *distances-to-default* are independent and identically distributed random variables. Then, the conditional probability of a default is given as

$$\mathbb{P}(X_T^i \leq 0 \mid \mathcal{F}^{\varsigma}) = \mathbb{P}(X_T^i \leq 0 \mid W_T^{\varsigma}) = \Phi\left(\frac{-X_0 - \beta T - \sqrt{\rho}\, W_T^{\varsigma}}{\sqrt{(1-\rho)\, T}}\right). \tag{1.8}$$

An extension to this approach, by allowing defaults also before maturity, is given, for example, in [46], where defaults are monitored on payment dates, or in [13], where defaults are monitored continuously. However, under these models, the conditional probability of the default is not available in a closed-form.

In reduced form models, the dependence between obligors under a common factor framework can be captured by the stochastic intensity driven by common and individual risks. A simple way to incorporate this idea is given in [34], where the intensity satisfies

$$\lambda_t^i = \Xi_t^{\varsigma} + \Xi_t^i,$$

where Ξ^{ς} is a common factor, and Ξ^i is an individual factor of the i-th company, Ξ^{ς} and Ξ^i are independent for all i. Let $\mathcal{F}^{\Xi^{\varsigma}}$ be the filtration generated by Ξ^{ς}. The probability of default conditional on the common factor is given as

$$\mathbb{P}(H_t^i = 1 \mid \mathcal{F}^{\Xi^{\varsigma}}) = 1 - e^{-\int_0^t \Xi_s^{\varsigma} ds}\, \mathbb{E}[e^{-\int_0^t \Xi_s^i ds}]. \tag{1.9}$$

If Ξ^i follows an affine process, then the expectation on the right-hand side is given in closed-form.

Let us now give the joint probability of default for Bernoulli mixture models. We follow the presentation given in [32], where a common structure is developed in order to compare different reduced form and structural models. Let us model a default of an obligor by a Bernoulli random variable Y^i, and a common factor (or joint effect of common factors) by Ψ, together with its filtration $\mathcal{F}^\Psi$. Also, we assume that the common factor Ψ is distributed according to the law $\mathbb{P}_\Psi$. The conditional probability of default is given as

$$\mathbb{P}(Y_i = 1 \mid \mathcal{F}^\Psi) = s_i(\Psi),$$

where s_i is some function depending on Ψ and specified for each i. In the case of an homogeneous portfolio, $s_i(\Psi) = s(\Psi)$ for all i. By conditional independence and since $(Y^i, i = 1, \ldots, N)$ are Bernoulli distributed random variables with $\mathbb{P}(Y_i = 1 \mid \mathcal{F}^\Psi) = s(\Psi)$ for all i, the joint density of a portfolio can be written as

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_N = y_N) = \int (s(\psi))^{\sum_{i=1}^N y_i} \, (1 - s(\psi))^{N - \sum_{i=1}^N y_i} \, \mathbb{P}_\Psi(\Psi \in d\psi), \quad (1.10)$$

where $y_i \in \{0, 1\}$, $\forall i$. In the case of a heterogeneous portfolio, the joint density of a portfolio equals

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_N = y_N) = \int \prod_{i=1}^N (s_i(\psi))^{y_i} \, (1 - s_i(\psi))^{1 - y_i} \, \mathbb{P}_\Psi(\Psi \in d\psi). \quad (1.11)$$

For Bernoulli mixture models, where analytic formulas are not available for $s_i(\Psi)$, such as in the structural model with defaults monitored also before maturity, formulas like (1.10) or (1.11) are not directly applicable. Instead, paths of *distance-to-default* for each name in the portfolio are simulated by Monte Carlo methods, giving in the end a numerical estimate of joint default distribution. This approach is used, for example, in [46] or [13].

1.2.3 Approximations for joint default probability in large portfolios

For a small or medium-sized portfolio, it is possible to calculate the joint default probability by formulas (1.10)-(1.11) or standard Monte Carlo simulations. For a large portfolio, containing a hundred names, it is still doable but time-consuming. However, calculating the joint default probability of a portfolio where the number of names is counted in thousands, is simply impossible in practical applications. That is why approximation methods are suggested in the literature.

The approximations are mostly based on versions of the conditional Law of Large Numbers. The first approximation is proposed in [77], for a structural model of the Merton type where a portfolio is homogeneous. Then, for example, in [63], the method is extended to a heterogeneous portfolio, consisting of a few homogeneous groups of names. In [20], the authors consider approximation for the tail of a joint distribution. Also, in [13], an extension for the continuous monitoring of default in a homogeneous portfolio is suggested. Inspired by the results of [13], the dynamic extension of a reduced model is proposed recently in [37]. Additionally to [13], the authors consider contagion effects. Here, we focus on the approximations given by [77] and [13].

Vasicek approximation

Let us start with the approximation proposed in [77]. Consider a portfolio with an infinite number of entities defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Each *distance-to-default* is given as in (1.7), where for all i, $X_0^i = X_0$.

By the independence of W^i and W^ς, for all i, we can write the probability space as

$$(\Omega^I \times \Omega^\varsigma, \mathcal{F}^I \times \mathcal{F}^\varsigma, \mathbb{P}^I \times \mathbb{P}^\varsigma),$$

where $(\Omega^I, \mathcal{F}^I, \mathbb{P}^I)$ is the space where individual factors are defined, while $(\Omega^\varsigma, \mathcal{F}^\varsigma, \mathbb{P}^\varsigma)$ is the space of common factor W^ς.

Observe that, by the conditional version of the Law of Large Numbers, for $t = T$ and for almost all $\omega^{\varsigma} \in \Omega^{\varsigma}$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_T^i(\omega^{\varsigma})\leq 0\}} = \mathbb{P}^I(X_T(\omega^{\varsigma}) \leq 0), \qquad \mathbb{P}^I\text{-a.s.}$$

where on the right-hand side we use that a portfolio is homogeneous. Then, for $t = T$, the loss of an infinite portfolio is defined for almost all $\omega^{\varsigma} \in \Omega^{\varsigma}$ as

$$L(\omega^{\varsigma}) = \mathbb{P}^I(X_T(\omega^{\varsigma}) \leq 0),$$

and by (1.8), we obtain also that

$$L(\omega^{\varsigma}) = \Phi\left(\frac{-X_0 - \beta T - \sqrt{\rho}\, W_T^{\varsigma}}{\sqrt{(1-\rho)\, T}}\right). \tag{1.12}$$

Let us denote by p the marginal probability of default, given in (1.5). Then, in terms of p, and since W_T^{ς} can be written as $U\sqrt{T}$, where U is a standard normal random variable, the loss is given as

$$L(\omega^{\varsigma}) = \Phi\left(\frac{1}{\sqrt{1-\rho}}\left(\Phi^{-1}(p) - \sqrt{\rho}\, U\right)\right).$$

Then, the probability distribution of L equals

$$\begin{aligned}
\mathbb{P}(L \leq l) &= \mathbb{E}^{\varsigma}\left[\mathbb{P}^I[L(\omega^{\varsigma}) \leq l]\right] \\
&= \int 1_{\{L(u)\leq l\}}\, \phi(u)\, du \\
&= \int 1_{\{u \leq \frac{1}{\sqrt{\rho}}\left(\sqrt{1-\rho}\,\Phi^{-1}(l) - \Phi^{-1}(p)\right)\}}\, \phi(u)\, du,
\end{aligned}$$

where ϕ is the density of the standard normal random variable. Finally, we obtain

$$\mathbb{P}(L \leq l) = \Phi\left(\frac{1}{\sqrt{\rho}}\left(\sqrt{1-\rho}\,\Phi^{-1}(l) - \Phi^{-1}(p)\right)\right).$$

This formula is very easy to calculate; however, as noted before, the model assumptions are too restrictive.

Bush et al approximation

Let us now present the approximation suggested in [13]. Throughout the thesis, we refer to this approach also as a *large basket approximation*.

In a similar way to [77], in [13], each *distance-to-default* is given by (1.7) but here the default is monitored continuously. Let $(X^i, i = 1, 2, \ldots)$ be defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is assumed that $(X_0^i, i = 1, 2, \ldots)$ is a family of exchangeable random variables, meaning that the distribution of $(X_0^i, i = 1, 2, \ldots)$ is invariant under permutations (see Section 4.3 for a precise definition). For example, each i.i.d. sequence is an exchangeable sequence but the inverse is not true. Also, it is assumed that the initial *distances-to-default* are independent of W^i and W^ς, for all i.

By the independence of X_0, W^i and W^ς, for all i, the probability space of $(X^i, i = 1, 2, \ldots)$ can be divided into

$$(\Omega^I \times \Omega^\varsigma, \mathcal{F}^I \times \mathcal{F}^\varsigma, \mathbb{P}^I \times \mathbb{P}^\varsigma),$$

where $(\Omega^I, \mathcal{F}^I, \mathbb{P}^I)$ is the space of individual factors, and $(\Omega^\varsigma, \mathcal{F}^\varsigma, \mathbb{P}^\varsigma)$ is the space of common factor W^ς.

In [13], the authors consider the empirical measure of the portfolio. Let $S \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the σ-field of Borel sets on $\mathbb{R}$, and let $N \geq 1$, then the empirical measure

is defined as

$$\tilde{\nu}_{t,N}(S) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{X_t^i \in S\}}. \tag{1.13}$$

By the de Finetti theorem (see for example [2]) which is a generalisation of the Law of Large Numbers, the limit of the empirical measure for $N \to \infty$ exists for infinite exchangeable sequences. In [13], the authors show that $(X^i, i = 1, 2, \ldots)$ is an infinite exchangeable sequence, and hence the limit empirical measure exists for all $t \geq 0$,

$$\tilde{\nu}_t(S) = \lim_{N\to\infty} \tilde{\nu}_{t,N}(S), \qquad \mathbb{P}\text{-a.s.}$$

Then, for $t > 0$, the limiting empirical measure is decomposed into the empirical measure of names that did not default by the time t, ν_t, and into the measure for the rest of the names

$$\tilde{\nu}_t = \nu_t + L_t\delta_0,$$

where ν is defined in $(0, \infty)$, δ_0 is a Dirac mass at zero and L is the loss of the infinite portfolio.

Let us now state the main result of [13]. The presentation is motivated by [58], where the results of [13] are studied in detail. In [13], the authors show that, for almost all $\omega^{\varsigma} \in \Omega^{\varsigma}$, $t \in [0, T]$, the empirical measure ν has a density v that satisfies

$$\mathbb{E}^{\varsigma}\left[\int_0^T \| v \|_2^2 + \| v_x \|_2^2 \; dt\right] < \infty \tag{1.14}$$

and is a unique solution (in the space of functions satisfying (1.14)) to the following

stochastic partial differential equation (SPDE)

$$dv = -\beta\, v_x\, dt - \frac{1}{2} v_{xx}\, dt - \sqrt{\rho} v_x\, dW_t^{\varsigma},$$
$$v(0,x) = v_0(x), \qquad v(t,0) = 0,$$

where v_0 is assumed to be smooth and has support in $[\underline{C}, \overline{C}]$, where $0 < \underline{C} < \overline{C} < \infty$.

Note that this is an SPDE with an absorbing boundary at zero and an initial density given by the density of the empirical measure of $(X_0^i, i = 1, 2, \ldots)$. In the limit, individual factors disappear, while the market factor, W^{ς}, drives the empirical measure.

Without an absorbing boundary, for each realisation of W_t^{ς}, the SPDE can be solved as

$$v(t,x) = u(t, x - \sqrt{\rho}\, W_t^{\varsigma}), \qquad \forall x \in \mathbb{R}, \qquad t > 0,$$

where $u(t,x)$ is a solution to the deterministic PDE

$$u(t,x) = \frac{1}{2}(1-\rho) u_{xx} - \beta u_x.$$

This observation is used in [13] to obtain $v(t,x)$ via finite difference methods accompanied by Monte Carlo simulations.

Also, in [13], the authors show that for almost all $\omega^{\varsigma} \in \Omega^{\varsigma}$, $t \in [0,T]$, the portfolio loss is given as

$$L_t = L_t(\omega^{\varsigma}) = 1 - \int_0^{\infty} v(t,x)\, dx.$$

Observe that, in terms of the Vasicek approximation, Bush et al provide a formula of a type (1.12), while the distribution of L is given by numerical simulations.

What is more, in [13], the authors use the approximation to provide example CDO

spreads for different values of model parameters, ρ and σ.

The approach is studied in detail in [48], also some extensions to the method are proposed. Very recently, in [1], the approximation is applied to mortgage backed securities, while in [59], more results are obtained regarding the regularity of the solution to the aforementioned SPDE.

1.2.4 What is missing?

The credit crisis of 2008-2009 showed that regulations regarding credit risk should be strengthened but also that default dependency is still an open, challenging problem facing credit modelling.

As noted in [73], a good multi-name model should meet the following conditions.

(I) Default dependence: The model must produce a realistic default dependency.

(II) Estimation/parsimony: The number of parameters in the model should be limited, especially it should not grow exponentially with the number of endogenous variables.

(III) Timing risk/clustering: The model must be dynamic, since the timing of defaults can be as crucial as the number of defaults. Also, the model should produce default clusters in time.

(IV) Calibration: The model must be capable of calibrating both to the individual and to the joint term structure of defaults for a fixed time horizon.

(V) Implementation: The model should be accompanied by a workable implementation mechanism.

Regarding the methods suggested for the joint default probability for portfolios with a large number of names, the approximation proposed by Bush et al. is only developed from an academic perspective. It is crucial to provide efficient numerical methods and a calibration mechanism in order to make the approach useful for practical applications.

Also, since the proposed model is based on diffusion processes, like other structural diffusion models, it suffers from the continuity of Brownian motion. Therefore, the method should be extended to tackle this problem, using the remedies presented in Section 2.1. Overall, the framework should be extended in such a way that the new model, together with its numerical methods, meets the Conditions (I)-(V) of a good multi-name model, given above. Also, the approximation should be validated for portfolios with a large finite number of names.

What is more, efficient Monte Carlo methods for calculating directly the loss of a large portfolio are needed as an alternative to the analytical approximations. To the best of our knowledge, methods for very large number of entities have not been considered in the existing literature.

1.3 Outline and contribution of the thesis

In the thesis, we propose two alternative frameworks to evaluate numerically a multi-name structural model where the number of entities is large. We evaluate the model either directly via an efficient Monte Carlo method developed by ourselves for such models or by extending the approximation proposed by Bush et al.

1.3.1 New simulation method and an extension to the Bush et al approximation

The proposed simulation method is based on the multi-level approach developed in [38] to calculate efficiently the expected value of a functional of the solution to an SDE. In the original setting, a Lipschitz functional and SDE for a diffusion process are considered, then many authors extended the method, for example, to more general functionals and/or more general SDEs. A survey of the extensions to the multi-level method can be found in [41].

In essence, the multi-level method is based on the convergence of the solutions to an

SDE for increasing the number of time steps, leading to lower computational complexity. We reinterpret the multi-level approach to our setting and construct estimators based on sequences of random variables with increasing lengths and a number of samples which decreases faster than the length increases, such that the overall computational complexity is essentially no larger than that for fixed small N. This is crucial, since the random sequences in our context are costly to simulate, because of the large number N of underlying processes, often required over large time horizons. Moreover, often, many Monte Carlo samples are necessary for the sufficiently accurate estimation of, for instance, the expected tranche losses of the credit basket.

The second approach is an extension to Bush et al. [13] approximation, where each company's asset value process is driven by a pure diffusion process and default is monitored continuously. Since the model in [13] is based on a diffusion process, because of continuity of Brownian motion, the model does not fit the market data well. Also, the numerical implementation of [13] consists of Monte Carlo and finite difference methods, which, for some extensions to the approach, can be computationally demanding. Hence, we aim at extending the method such that the new model can fit the market data reasonably well and be computed sufficiently quickly. This is met by a jump-diffusion model with discretely monitored defaults that we suggest as an extension to the Bush et al. approximation.

We derive analytical results for this extension, together with efficient numerical methods to use it in practice, including simulation and calibration algorithms. The heuristics considering the extension to the system driven by a jump-diffusion process is suggested earlier in [48], also, recently, a heuristic analysis of an extension to the general Lévy processes is proposed in [58], both authors consider continuously monitored defaults.

In the thesis, for the multi-level approach calculations are performed for a structural jump-diffusion model where the defaults are monitored discretely. However, the method can be applied to any Bernoulli mixture model (up to some technical conditions), whereas in case of the extension to the Bush et al. approximation, the analytical results hold only

for the proposed model but we also show how the approach can be extended to certain other models.

We wish to emphasise that the suggested frameworks accompanied by a jump-diffusion model by no means end the search for the optimal credit multi-name framework, but we hope that the frameworks mark an important step towards efficient practical implementation of a large class of models.

1.3.2 Overview of the chapters

In Chapter 2, we present the multi-name model considered throughout the thesis. We check how the model meets the Conditions (I)-(V) given in Section 1.2.4, by analysing the default dependency in the model, paying special attention to the individual and joint term structure of defaults, and also by calibrating the model to pre and during the crisis data, in order to check how the model fits the data under different market conditions. The chapter is based on our article [12].

In Chapter 3, we propose our simulation method in the spirit of a multi-level Monte Carlo algorithm. The approach is based on the convergence of certain expected functionals of portfolio loss for increasing the number of obligors. We state the theoretical and numerical convergence results. The theoretical convergence rates are derived for any Bernoulli mixture models (up to some technical conditions) and for general Lipschitz functionals of expected losses, whereas the empirical rates are calculated for the multi-name jump-diffusion model proposed in the thesis and for the expected tranche loss of a CDO. The chapter is based on our article [11].

In Chapter 4, we present an extension to the approximation suggested by Bush et al. [13] for a jump-diffusion model with defaults monitored discretely. We show that, for the infinite portfolio, the empirical measure has a density with respect to the Lebesgue measure that satisfies an SPDE. Then, we present a recursive solution to this SPDE.

In Chapter 5, we develop numerical methods for using the extension proposed in Chap-

ter 4 in practical applications. We develop methods for evaluating the model, calibrating it to market data and calculating its sensitivities. Also, we check the validity of the approximation for finite numbers of entities in a portfolio. The chapter is based on our article [12].

In Chapter 6, we summarize our main results, show how different results are connected with each other, and give recommendations for future research.

Chapter 2

Numerical valuation of basket credit derivatives in structural jump-diffusion models

In this chapter, we present a multi-name model that we develop numerical methods for in the thesis. We analyse the model assumptions and show that the model is flexible enough to match CDO spreads from precrisis and crisis periods. In our calibration algorithm, CDO spreads are consistent with the underlying CDS spreads, such that a company's asset value at the initial time is based on its CDS spreads. We present numerical methods that we develop for a single-name model to produce a CDS spread for each name underlying the CDO basket. Also, we show that the model can produce different shapes of CDS term structure observed in the market.

We consider a model where each company's asset value process follows a jump-diffusion process and is connected with other companies via random global factors. In our model, a default can occur both expectedly, due to the diffusion part, and unexpectedly, due to the jump part, by a sudden fall in a company's value.

Overview of the chapter

The chapter is organised as follows. First, in Sections 2.1 and 2.2 we present our structural model for single-name and multi-name credit derivatives. Then, in Section 2.3, we present calibration results, while Section 2.4 concludes with remarks on further improvements of the model.

2.1 A single-name structural model

We begin by discussing a model and pricing formulae for the individual CDSs in this section. The results here will also be used for the calibration of multi-name models.

2.1.1 The model setup

We model a company's asset value, A_t, by a jump-diffusion process. Under a risk-neutral measure $\mathbb{P}$ the dynamics of A are given as

$$\frac{dA_t}{A_t} = (r - \lambda\nu)\, dt + \sigma\, dW_t + (Y - 1)\, d\tilde{N}_t, \tag{2.1}$$

where r is the risk-free interest rate, σ the asset volatility, W is a standard Brownian motion, $\tilde{N}$ a Poisson process with intensity λ, $(Y_k - 1)$ is the size of the kth jump of the compound Poisson process $\sum_{k=1}^{\tilde{N}_t}(Y_k - 1)$ and $\nu = \mathbb{E}[Y_k - 1]$. We assume that W, $\tilde{N}$, Y are mutually independent, and that $\{Y_k\}$ are independent and identically distributed (i.i.d.).

We would like to mention that because of the jump component present in the model, the risk-neutral pricing measure is not unique and market is incomplete. As a consequence, a perfect hedge of default risk via trading in the asset value of the reference entity and default-free securities is not possible. Hence, under a jump-diffusion model a CDS cannot be perfectly replicated in this way which is in contrast to the diffusion model discussed in Section 1.2.1. Throughout the thesis, we assume that we are given a risk-neutral pricing measure $\mathbb{P}$.

A solution to (2.1) is given by

$$A_t = A_0 \exp\left\{\left(r - \lambda\nu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right\}\prod_{k=1}^{\tilde{N}_t} Y_k, \qquad A_0 > 0. \tag{2.2}$$

If $\lambda = 0$, then $\tilde{N}_t = 0$, and we obtain geometric Brownian motion. As [65] notes, $Y - 1$ is the relative change of A_t through jumps. What is more, in the special case when $\{Y_k\}$ are log-normally distributed, conditional on $\tilde{N}_t$, A_t is also log-normally distributed.

Following [8], one can define the default time τ^A as the first passage time of the company's asset value of a constant default barrier c^B,

$$\tau^A = \inf\left\{t > 0 : A_t \leq c^B\right\}.$$

We assume that $A_0 > c^B$ and $P(Y_k > 0) = 1$, such that the asset price process remains positive. In order to eliminate explicit dependence on c^B, we introduce the *distance-to-default*

$$X_t = \frac{1}{\sigma}(\ln(A_t) - \ln(c^B)). \tag{2.3}$$

Since, $A_0 > c^B$, $X_0 > 0$. By applying Itô's lemma to (2.3) and using (2.1), we obtain

$$dX_t = \beta\, dt + dW_t + \Pi\, d\tilde{N}_t, \tag{2.4}$$

where $\beta = \frac{1}{\sigma}(r - \lambda\nu - \frac{1}{2}\sigma^2)$, $\Pi = \ln Y/\sigma$. The first passage time τ^A of X_t is now

$$\tau^A = \inf\left\{t > 0 : X_t \leq 0\right\}.$$

Implicit here is the assumption that default can be monitored continuously. We will later work in a framework where default can only be detected at a discrete set of times $t_1, \ldots, t_n$, and then set

$$\tau^D = \min\left\{t \in \{t_1, \ldots, t_n\} : X_t \leq 0\right\}.$$

We account for default in X by setting

$$X_t = 0 \quad \forall\, t \geq \tau^D. \tag{2.5}$$

2.1.2 Distribution of the jump amplitude

We are aiming towards a computationally tractable, yet economically convincing structural default model. To meet these conditions, we need a distribution of jump amplitudes that enables sufficiently fast calculation of the survival probability of individual firms, yet gives realistic market dynamics.

[80] suggests a jump-diffusion model with log-normal jumps, as first introduced in the general context of financial modelling by [65] as an extension to the Black-Scholes approach. In contrast, [55], [44], [45] and [62] propose a double exponential jump size distribution. [55] specifically argues that although both types of jumps can lead to the leptokurtic feature of equity returns observed in the market, in continuous time double-exponential jump-diffusion models have better analytical tractability. As [54] shows, due to the lack of memory associated with the exponential distribution, there exists an analytical solution, via Laplace transforms, to the distribution of first passage times for the double exponential case. Such a solution does not exist for the log-normal model. However, as [68] note, the transition density of log-normal jump-diffusions has a more convenient form than that of double-exponential ones.

What is more, [68] assesses empirically the performance of double exponential jump-diffusion compared to log-normal jump-diffusion and geometric Brownian motion in matching stock prices. Because of the inherent link between equity and credit in structural models ([65]), empirical findings in equity markets are relevant for the present setting. They find that both double-exponential and log-normal jump-diffusions give a better fit to market data than geometric Brownian motion. In their study, double-exponential jump-diffusions outperform log-normal ones when stock indices are concerned, however,

for individual stocks the results are inconclusive. It should be noted that a slightly better performance of the double-exponential jump-diffusion model can be attributed to the presence of an extra parameter.

We focus in the following on the log-normal case, for reasons of convenience, which will become clear in Section 2.1.3, however, the overall framework is independent of the jump size distribution.

2.1.3 Computing survival probabilities for log-normal jumps

The initial distance-to-default X_0, together with the jumps, determines default probabilities at the very short end, while diffusion plays a role over the medium time range and the drift only for long time horizons. One could therefore use CDS spreads with different maturities to back out the different parameters. We will follow a different tack later on and use information from tranche spreads, and use CDS spreads only to infer X_0.

A key ingredient in the pricing of CDSs, which is used for the calibration of CDO models in our framework, is the survival probability. For a log-normal jump-diffusion model, an analytical formula for the survival probability does not exist, hence Zhou [80] applies a Monte Carlo algorithm. Since this approach is computationally demanding when a portfolio of CDSs is considered, in order to obtain an analytical approximation to survival probabilities, [79] assumes that a firm has survived to a given monitoring time out of $t_1, \ldots, t_n$, if at this particular time the asset value is above the barrier. We denote such a default time by τ^P. As the author notes, this is a crude approximation, and the survival probability in the model is overestimated, especially when t is high,

$$\mathbb{P}(\tau^D > t_n) = \mathbb{P}(X_{t_n} > 0, X_{t_{n-1}} > 0, \ldots, X_{t_1} > 0) < \mathbb{P}(X_{t_n} > 0) = \mathbb{P}(\tau^P > t_n).$$

We now present a practically tractable algorithm to calculate the survival probability for log-normally distributed jumps. Since the proofs of the following Propositions are

relatively straightforward, the proofs are given in Appendix A.

Proposition 2.1.1. *Let $0 < t_n \leq T$ be a monitoring date, X a jump-diffusion process driven by (2.4) with $\Pi \sim N(\mu_\Pi, \sigma_\Pi^2)$. The survival probability at time t_n, for $n \geq 2$ is given recursively by*

$$\begin{aligned} p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1}) &= \int_0^\infty p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}) \\ &\qquad \times p(X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})\, dx_{t_{n-1}}, \\ \mathbb{P}(\tau^D > t_n) &= \int_0^\infty p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1})\, dx_{t_n}. \end{aligned} \tag{2.6}$$

Proposition 2.1.2. *The probability of no default at t_1, and the density of X_{t_1}, are given by*

$$\begin{aligned} \mathbb{P}(\tau^D > t_1) &= \sum_{c=0}^\infty \mathbb{P}(X_{t_1} > 0 \mid \tilde{N}_{t_1} = c) \cdot p(\tilde{N}_{t_1} = c) \\ &= e^{-\lambda t_1} \sum_{c=0}^\infty \Phi\left(\frac{\mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right) \frac{(\lambda t_1)^c}{c!}, \end{aligned} \tag{2.7}$$

$$\begin{aligned} p(X_{t_1} = x_{t_1}) &= \sum_{c=0}^\infty p(X_{t_1} > 0 \mid \tilde{N}_{t_1} = c) \cdot p(\tilde{N}_{t_1} = c) \\ &= e^{-\lambda t_1} \sum_{c=0}^\infty \frac{1}{\sigma_{X_{t_1}}} \phi\left(\frac{x_{t_1} - \mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right) \frac{(\lambda t_1)^c}{c!}, \end{aligned} \tag{2.8}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution, $\phi(\cdot)$ is the standard normal density, $\mu_{X_{t_1}} = x_0 + \beta t_1 + c\mu_\Pi$, $\sigma^2_{X_{t_1}} = t_1 + c\sigma^2_\Pi$.

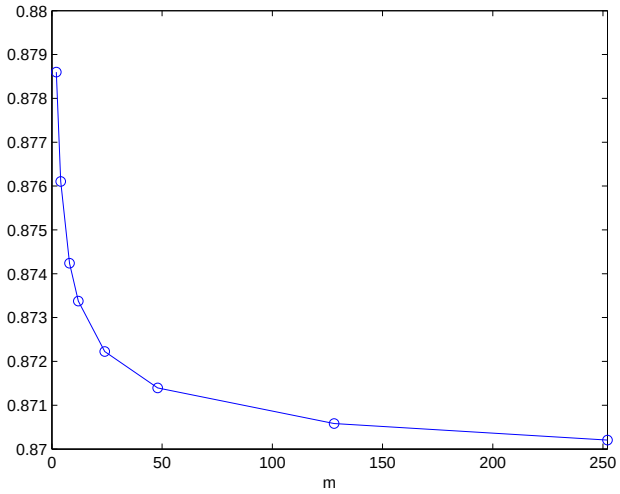

Figure 2.1: Survival probability at $T = 5$ in the log-normal jump-diffusion model with m default monitoring dates per year, where $m = 2, \ldots, 252$. Parameters are taken from a calibration to market data from 5 December 2008 (see Section 2.3 for details).

Proposition 2.1.3. *The density of X_{t_n}, conditional on $X_{t_{n-1}}$, for $n \geq 2$, is given by*

$$\begin{aligned} &p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}) \\ &\qquad = \sum_{c=0}^{\infty} p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}, \Delta\tilde{N}_{t_n} = c)\, p(\Delta\tilde{N}_{t_n} = c) \\ &\qquad = e^{-\lambda\Delta t} \sum_{c=0}^{\infty} \frac{1}{\sigma_{X_{t_n}|X_{t_{n-1}}}} \phi\left(\frac{x_{t_n} - \mu_{X_{t_n}|X_{t_{n-1}}}}{\sigma_{X_{t_n}|X_{t_{n-1}}}}\right) \frac{(\lambda\Delta t)^c}{c!}, \end{aligned}$$

where $\Delta t = t_n - t_{n-1}$, $\Delta\tilde{N}_{t_n} = \tilde{N}_{t_n} - \tilde{N}_{t_{n-1}}$, $\phi(\cdot)$ is the standard normal density, $\mu_{X_{t_n}|X_{t_{n-1}}} = x_{t_n} + \beta\Delta t + c\mu_\Pi$, $\sigma^2_{X_{t_n}|X_{t_{n-1}}} = \Delta t + c\sigma^2_\Pi$.

Remark 2.1.1. *The recursion given in the above Propositions, can be computed using straightforward numerical integration, as the integrands are sums of standard normal densities. What is more, the sums converge quickly, since $\frac{(\lambda\Delta t)^c}{c!} \to 0$ rapidly as $c \to \infty$, especially for $\lambda \ll 1$, as is the case for the calibration results given later in the thesis, and for $\Delta t = 0.25$, ie, for quarterly monitored defaults.*

Remark 2.1.2. *A similar recursion is suggested in [28]. The difference is that the recursion given by Fang et al. is backwards, ie, first* $p(X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} > 0)$ *is obtained, then* $p(X_{t_{n-2}} = x_{t_{n-2}}, X_{t_{n-1}} > 0, X_{t_n} > 0)$ *and finally* $p(X_{t_1} = x_{t_1}, X_{t_2} > 0, \ldots, X_{t_{n-1}} > 0, X_{t_n} > 0)$. *In our case, we first calculate* $p(X_{t_1} = x_{t_1})$, *then* $p(X_{t_2} = x_{t_2}, X_{t_1} > 0)$ *and, in the end,* $p(X_{t_n} = x_{t_n}, X_{t_{n-1}} > 0, \ldots, X_{t_1} > 0)$, *which gives immediately* $\mathbb{P}(\tau^D > t_n) = \mathbb{P}(X_{t_n} > 0, \ldots, X_{t_1} > 0)$ *for each* t_n, $0 < t_n \leq T$.

2.1.4 Continuous vs discrete default monitoring

Both continuous monitoring of default (e.g., [8] and [45]) and discrete monitoring (e.g., [46] and [28]) is assumed in applications of structural default models. In practice, a default is announced on a daily basis, as [28] argue.

In order to make models computationally more tractable, some studies, e.g., [46] and [13], assume that default is detected only on spread payment dates. We follow this line and thus assume that defaults are monitored quarterly.

In order to check the impact of this approximation, we calculate survival probabilities for $m = 2, \ldots, 252$ monitoring dates per year, ie, from half-yearly to daily monitoring, at time $T = 5$, ie, at the maturity of CDSs that we will use later to calibrate CDO pricing models. The results are shown in Fig. 2.1. As expected, the survival probabilities converge and are a decreasing function of m. The difference between the survival probability calculated for $m = 4$, ie, for the number of monitoring times we will assume in the computations, compared to $m = 252$, ie, the maximum realistic monitoring times per year, is 0.59 percentage points.

Remark 2.1.3. *The question of continuous versus discrete monitoring of defaults is mathematically equivalent to that of continuous versus discrete barrier monitoring when pricing barrier options.[10], in a diffusion model, approximate prices of discrete barrier options*

by formulae for continuously observed but shifted barrier, precisely (Theorem 1.1 in [10])

$$V_n(H) = V\left(H\, e^{-\beta\sigma\sqrt{\Delta T}}\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{2.9}$$

where $V_n(H)$ is the price of a discretely monitored knock-out call option, $V(H)$ is the price of a corresponding continuously monitored barrier option, β is a constant derived by the authors, n is the number of monitoring times at intervals ΔT. This correction accounts for the probability of undetected barrier crossings between monitoring dates.

In a finite activity jump model, as the probability of a jump in an interval of length ΔT is $O(\Delta T)$, and that of an undetected down-up combination even $O(\Delta T^2)$, we conjecture similar asymptotic behaviour.

As we calibrate the initial distance-to-default*, X_0, to CDS spreads, which are functions of the survival probabilities, the barrier correction (2.9) is implicitly invoked. This is seen from (2.3), which shows that a shift of the initial position is equivalent to an opposite shift of the default barrier. When we use the calibrated model later to price basket derivatives based on the same monitoring dates, the difference between monitoring frequencies should be small.*

2.1.5 Implied CDS term structure

In Figures 2.2 and 2.3, we present a CDS term structure implied by the model for different values of initial *distances-to-default*. In the former figure, the model parameters are taken from the precrisis period, while in the latter, during the crisis.

According to the empirical evidence, in quiet market conditions, CDS term structure is usually increasing, while in more shaky periods, the CDS curve can have a different shape, in particular, can be downward sloping. This is in line with our results.

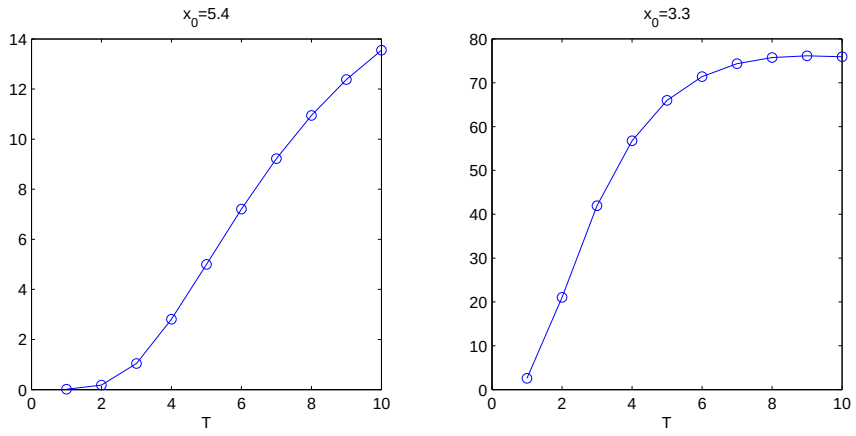

Figure 2.2: Implied CDS term structure for chosen values of x_0 and model parameters taken from a calibration to market data from 22 February 2007 (see Section 2.3 for details).

2.2 A multi-name structural model

In this section, we extend the single-name jump-diffusion to a multi-name model, with the view of pricing basket credit derivatives.

The works closest to our basic setup are [79] and [52], where [79] uses a log-normal jump-diffusion multi-name model to price CDOs, while [52] use a double exponential jump distribution.

2.2.1 The model setup

We extend the single-name model to the setting of a portfolio of CDSs on N different companies, which is consistent with (2.1) for the individual names, but makes a special assumption on the dependence structure, namely,

$$\frac{dA_t^i}{A_t^i} = (r - \lambda\nu)\, dt + \sqrt{1-\rho}\, \sigma\, dW_t^i + \sqrt{\rho}\, \sigma\, dW_t^\varsigma + (Y-1)\, d\tilde{N}_t, \tag{2.10}$$

where W^ς is a standard Brownian motion, $\rho \in [0,1)$, $\ln Y \sim N(\mu_Y, \sigma_Y^2)$, (W^i) is a standard Brownian motion, W^i, W^ς, $\tilde{N}$, Y are mutually independent, for all i.

W^i is an idiosyncratic factor, specific to each company, which affects the value of each

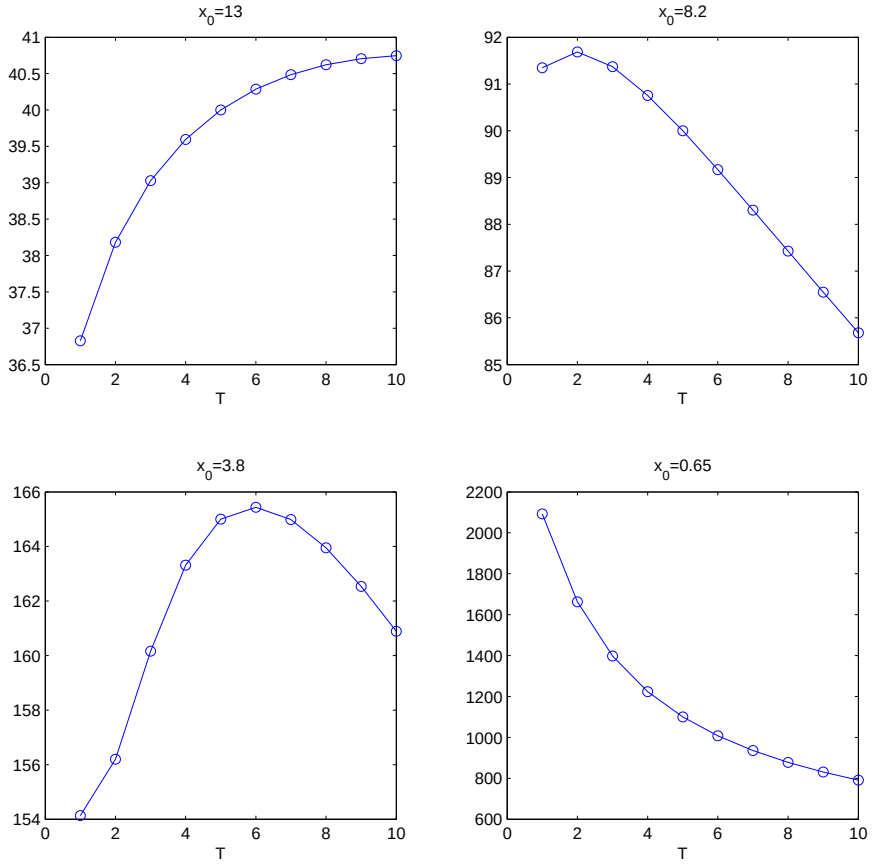

Figure 2.3: Implied CDS term structure for chosen values of x_0 and model parameters taken from a calibration to market data from 5 December 2008 (see Section 2.3 for details).

company's assets alone, such as the management of a company. W^ς, $\tilde{N}$ and Y represent global factors that affect the default environment of all companies. As [46] note, such a global factor could be for instance the S&P 500 index. Other candidates could be Gross Domestic Product, or more specifically investment spending, central bank interest rates or the unemployment rate, since they indicate the phase of the business cycle of the economy.

The jumps $\tilde{N}$ together with Y, model sudden effects by global factors on the situation of companies, while W^ς models more gradual influences. We aim at taking into account predominantly negative jumps and consider only $\mu_Y < 0$. The intensity of $\tilde{N}$, λ, then measures the frequency of economic shocks. If the absolute value of the jump mean is high, then λ can be interpreted as the frequency of economic crises. A similar interpretation can be found in [79].

Like in the univariate case, we consider the *distance-to-default* $X_t^i = \frac{1}{\sigma}(\ln(A_t^i) - \ln(c_i^B))$, where c_i^B is a constant default barrier. We assume $X_0^i > 0$ and the dynamics of X_t^i are then given by

$$dX_t^i = \beta\, dt + \sqrt{1-\rho}\, dW_t^i + \sqrt{\rho}\, dW_t^\varsigma + \Pi\, d\tilde{N}_t, \tag{2.11}$$

where $\beta = (r - \lambda\nu - \frac{1}{2}\sigma^2)/\sigma$, $\Pi \sim N(\mu_\Pi, \sigma_\Pi^2)$, $\mu_\Pi = \mu_Y/\sigma$, $\sigma_\Pi^2 = \sigma_Y^2/\sigma^2$.

The distribution of *distances-to-default* X_0^i indicates the current state of the market. If the mass of the distribution is close to zero, as shown in Fig. 2.4, right, for the iTraxx constituents in December 2008, then the financial situation of companies is relatively bad; if the mass is concentrated far away from zero, as shown in Fig. 2.4, left, again for the iTraxx but in February 2007, the economic climate is good.

It is worth discussing in more detail a main premise of the model, namely that the volatility parameter σ is identical for all firms. Credit indices usually consist of companies from different industry sectors and different regions, and clearly there will be differences between the variances of asset values. If we allowed individual σ_i for each company in the

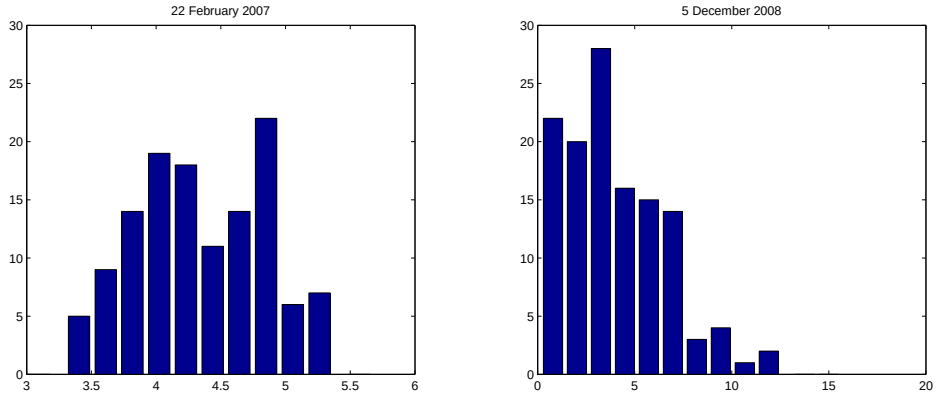

Figure 2.4: Histogram of X_0^i obtained by calibration to 5 year CDS spreads from the iTraxx Main Series, on two different dates pre- and during the crisis.

portfolio, the dynamics would be, instead of (2.10),

$$\frac{dA_t^i}{A_t^i} = (r - \lambda\nu)\, dt + \sqrt{1-\rho}\, \sigma_i\, dW_t^i + \sqrt{\rho}\, \sigma_i\, dW_t^\varsigma + (e^{\sigma_i \Pi} - 1)\, d\tilde{N}_t, \tag{2.12}$$

where σ_Π and μ_Π still have the same values for all companies. In terms of X_t^i,

$$dX_t^i = \beta_i\, dt + \sqrt{1-\rho}\, dW_t^i + \sqrt{\rho}\, dW_t^\varsigma + \Pi\, d\tilde{N}_t, \tag{2.13}$$

where $\beta_i = (r - \lambda\nu - \frac{1}{2}\sigma_i^2)/\sigma_i$, and the initial *distance-to-default* is $X_0^i = (\ln(A_0^i) - \ln(c_i^B))/\sigma_i$. The volatility dependence of the default probabilities, is therefore to a large extent captured by the initial credit quality through X_0^i, and, especially for maturities that are not too long, the effect of the drift will be negligible compared to the diffusive and jump components.

The consequence of identical driving processes for all firms is that, as long as X_0^i are *exchangeable* random variables, ie, their joint law is invariant under permutation, the firm values X_t^i are also exchangeable (we refer to Section 4.3 for the precise definition of exchangeability). Furthermore, conditional on X_0^i, W^ς and $\sum_{k=1}^{\tilde{N}_t} \Pi_k$; X_t^i are independent.

For practical applications, the initial X_0^i will be calibrated to individual CDS spreads

allowing heterogeneity in the portfolio.

The dependence between companies in the model is determined by *diffusion correlation*, represented by a single parameter ρ, together with perfect *jump correlation* for all companies, governed by common jump times of $\tilde{N}_t$ and the same jump sizes, Π. A simple calculation shows that the overall correlation between X_t^i and X_t^j, $i \neq j$, driven by (2.11), is

$$\rho_{X_t^i X_t^j} = \frac{\rho + \vartheta}{1 + \vartheta}, \tag{2.14}$$

where $\vartheta = \lambda(\sigma_\Pi^2 + \mu_\Pi^2)$.

In the present setting, the dependence between companies can vary from slight, if the values of ρ, λ, $|\mu_\Pi|$, σ_Π are small, to extremely strong, if the values are high. A few years ago, taking into account such a strong dependence might have been regarded as unrealistic, but the recent credit crunch, which affected not only financial markets but also the global real economy, showed that the dependence between companies can indeed be very strong. One might consider allowing different exposure to market factors by individual companies, which we do not do here for ease of calibration and computation, and since we are more concerned with the macroscopic behaviour of the basket.

2.2.2 Numerical valuation methods proposed in the thesis

In the thesis, we propose two alternative frameworks for numerical valuation of the multi-name model presented in this chapter. The first method is an efficient simulation approach, while the second one, is an extension to the analytical approximation suggested in [13]. We give details of the first method in Chapter 3, and the second framework is presented in Chapters 4 and 5. In this chapter, we only present calibration results. The calibration exercise is performed under the latter approach.

2.3 Calibration

The first step in using a model is to estimate its parameters by calibrating the model to market data. We fit the jump-diffusion model and, for comparison, the pure diffusion model, to CDO index and tranche spreads.

2.3.1 Results

In order to check if the model is flexible enough to match market spreads both in quiet and extreme market conditions, we calibrate it to pre-crisis data, from 22 February 2007, and during the crisis, 5 December 2008. We do not use current data since the market is not sufficiently liquid.

For the first data set (see Table 2.1), spread curves both for index and tranches are increasing with time, super senior tranches (22%-100%) are close to zero, indicating that the market does not expect any shocks in the near future. For the second period (see Table 2.4), all quotes are much higher (for example prices of super senior tranches are higher than junior mezzanine (3%-6%) with maturity 5 years for 22 February 2007), the index spread curve is inverted, curves for tranches are almost flat, hence the market anticipates that the situation will get worse in the near future. We anticipate that parameters indicating market turbulences will be considerably higher for the second data set.

We compare the calibration results by the ARPE (Average Relative Percentage Error),

$$ARPE = \frac{1}{s}\sum_{i=1}^{s}\frac{|y_i - y_i^{\theta}|}{y_i}, \tag{2.15}$$

and the RMSE (Root Mean Square Error),

$$RMSE = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(y_i - y_i^{\theta})^2}, \tag{2.16}$$

where y is a vector of observed prices, y^{θ} is a vector of prices obtained from the model,

s the total number of calibration prices. It is important to note that the ARPE can be used to compare calibration results between data sets with different orders of magnitudes, while the RMSE can be used for comparing calibration results within the same or similar data sets and will attach disproportionate weight to the equity tranche spreads.

Diffusion Model

Calibration results for the diffusion model are given in Tables 2.1 to 2.6. Root mean square errors (RMSE) are very high, especially for 5 December 2008, but as was noted before, this is partially a result of higher spreads for 5 Dec 2008 compared to 22 Feb 2007.

For 5 Dec 2008, the estimated correlation coefficient, ρ, is much higher than the one obtained for 22 Feb, 2007. High correlation not only increases the probability of companies to default together, but also to survive together. Hence, in order to fit equity tranches, a low correlation is needed, whereas high correlation is required for the senior tranches. This can be observed in Figure 3 in [13], showing the implied correlation for each tranche. Hence, in a diffusion model, there is usually a trade-off between matching equity and senior tranches. The low correlation coefficient for 22 Feb 2007 results in zero spreads for super senior tranches ($22\% - 100\%$), while equity tranches are underpriced. For 5 Dec 2008, the model produces non-zero spreads for super senior tranches, owing to a higher correlation coefficient, but still both super senior and equity tranches are underpriced.

Jump-Diffusion Model

Tables 2.1 to 2.6 show calibration results for the jump-diffusion model. Compared to the diffusion model, error measures for both data sets are much lower. What is more, the jump-diffusion model gives spreads of roughly the correct magnitude for all tranches. However, still some tranches are slightly underpriced, whereas other tranches are somewhat overpriced, albeit to a lesser extent than in the diffusion case.

We would like to note that the better performance of the jump-diffusion model com-

T=5 Years			
	Market	Jump-Diffusion	Diffusion
Index	21	24	17
Tranche 0%-3%	7.19%	11.07%	1.22%
Tranche 3%-6%	41	61	48
Tranche 6%-9%	10.8	15.2	10
Tranche 9%-12%	5	5.9	3
Tranche 12%-22%	1.8	2.5	0
Tranche 22%-100%	0.9	1.8	0
T=7 Years			
	Market	Jump-Diffusion	Diffusion
Index	30	32	25
Tranche 0%-3%	22.1%	24.7%	12.1%
Tranche 3%-6%	110	155	137
Tranche 6%-9%	32.5	39.5	41
Tranche 9%-12%	15	14	14
Tranche 12%-22%	4.9	3.9	2.0
Tranche 22%-100%	2	1.8	0
T=10 Years			
	Market	Jump-Diffusion	Diffusion
Index	41	37	32
Tranche 0%-3%	38%	35.4%	21.9%
Tranche 3%-6%	302.5	281.5	247
Tranche 6%-9%	83	87	97
Tranche 9%-12%	37	32	41
Tranche 12%-22%	12.5	7.3	9
Tranche 22%-100%	3.6	1.8	0

Table 2.1: Calibration results for 22 February 2007, iTraxx Main Series 6 index. Units are basis points (bps) if not stated otherwise. Estimated parameters are given in Table 2.2, measures of fit in Table 2.3. We assume that $r = 0.042$, $R = 0.4$.

Parameters	Jump-Diffusion	Diffusion
σ	0.16	0.18
ρ	0.11	0.22
λ	0.04	-
$\mathbb{E}[Y-1]$	-0.07	-
$\mathrm{Var}[Y-1]$	0.01	-
$\rho_{X_t^i X_t^j}$	0.13	-

Table 2.2: Parameters estimated for 22 February 2007, overall correlation $\rho_{X_t^i X_t^j}$ as in (2.14).

Measure of Fit	Jump-Diffusion	Diffusion
RMSE	1.17	4.34
ARPE	0.27	0.42

Table 2.3: Measures of fit for 22 February 2007. Root Mean Square Error (RMSE) is defined by (2.16), Average Relative Percentage Error (ARPE) is given by (2.15).

T=5 Years			
	Market	Jump-Diffusion	Diffusion
Index	215	245	200
Tranche 0%-3%	71.5%	76.5%	47%
Tranche 3%-6%	1576.3	1652	1158
Tranche 6%-9%	811.5	822	825
Tranche 9%-12%	506.1	504	636
Tranche 12%-22%	180.3	305	415
Tranche 22%-100%	77.9	77	56
T=7 Years			
	Market	Jump-Diffusion	Diffusion
Index	195	224	198
Tranche 0%-3%	72.9%	78.8%	50%
Tranche 3%-6%	1473.2	1557	1101
Tranche 6%-9%	804.2	805	805
Tranche 9%-12%	512.4	511	637
Tranche 12%-22%	182.6	311	434
Tranche 22%-100%	75.8	74	66
T=10 Years			
	Market	Jump-Diffusion	Diffusion
Index	175	204	188
Tranche 0%-3%	73.8%	81.2%	51.7%
Tranche 3%-6%	1385.5	1462	1016
Tranche 6%-9%	824.7	768	755
Tranche 9%-12%	526.1	500	607
Tranche 12%-22%	174.1	310	427
Tranche 22%-100%	76.3	69	70

Table 2.4: Calibration results for 5 December 2008, iTraxx Main Series 10 index. Units are basis points (bps) if not stated otherwise. Estimated parameters are given in Table 2.5, measures of fit in Table 2.6. We assume that $r = 0.033$, $R = 0.4$.

Parameters	Jump-Diffusion	Diffusion
σ	0.13	0.18
ρ	0.35	0.8
λ	0.04	-
$\mathbb{E}[Y-1]$	-0.5	-
$\mathrm{Var}[Y-1]$	0.17	-
$\rho_{X_t^i X_t^j}$	0.83	-

Table 2.5: Parameters estimated for 5 Dec, 2008, overall correlation $\rho_{X_t^i X_t^j}$ as in (2.14).

Measure of Fit	Jump-Diffusion	Diffusion
RMSE	2.42	8.95
ARPE	0.16	0.35

Table 2.6: Measures of fit for 5 December 2008. Root Mean Square Error (RMSE) is defined by (2.16), Average Relative Percentage Error (ARPE) is given by (2.15).

pared to the diffusion one can be attributed to the presence of three extra parameters. In the jump-diffusion model, apart from ρ and σ, also λ, μ_Y and σ_Y affect multiple defaults. Most importantly, unlike ρ, higher absolute values of jump parameters, with $\mu_X < 0$, always lead to more expected defaults. Therefore, the trade-off between matching particular tranches is smaller and the overall fit is better in the jump-diffusion model.

Let us now analyse the situation on the market in 22 Feb 2007 and 5 Dec 2008, implied by the model parameters. As mentioned in Section 2.2.1, the initial *distances-to-default*, obtained from CDS spreads, indicate the current state of the market. As expected, on average, x_0 is lower for December 2008, around 4.0, whereas for February 2007, it is 4.3. It may seem a bit surprising that the difference is so small, though a better picture is given by the tails of the distribution of initial *distances-to-default*, as seen back in Fig. 2.4. For the former data set, for 6% of the companies x_0 is in (0,1], ie, 6% of the companies in the portfolio is already in a bad financial situation, and for 26% x_0 is in (0,2], whereas for none of the firms in the latter data set x_0 is in (0,2]. This well reflects the situation in the market during a crisis and in normal conditions.

The jump intensity, λ, is 0.042 for 22 Feb 2007 and 0.035 for 5 Dec 2008 respectively.

This implies that under the risk neutral measure a market crash is expected to happen every 25 years for 22 Feb 2007, and every 34 years for 5 Dec 2008. For the former data set, market turbulences will be benign: if a jump occurs, on average a company's assets value will fall by 7%. For the latter, market conditions will worsen much more severely: if a jump occurs, a company's asset value will drop on average by 50%.

To our knowledge, a similar calibration exercise is found only in [79] and [52], however the latter study does not discuss the parameters obtained. In [79], a jump-diffusion model with log-normally distributed jumps was calibrated to market data from 8 August 2005. The obtained jump intensity is very small, $\lambda = 0.0012$, while the mean jump size is -96%. Under these settings, on average there is a complete market wipe-out every 870 years. Such a market situation would appear hard to reconcile with the timeframes market participants operate in, or even the age of credit markets and the economy as we know it.

2.3.2 Extensions

Here, we comment on some aspects of the practical use of the model.

First, we remark that in the course of the calibration we calculate sensitivities of the prices with respect to the input parameters, which can be used for hedging with other products exposed to the same variables.

In practice, a common hedge is the index hedge, and in that case it is important to be able to calibrate the model accurately to index spreads. The results of Tables 2.1 and 2.6 are unlikely to be good enough for this, in contrast to those in [13], where the volatility is first calibrated to the index spreads, and then the correlation (as single remaining parameter) is fit to tranche spreads, rather than performing an overall fit. A better calibration to specific important data, say the index spreads, can be enforced by increasing those weights in the least-squares functional. An interpretation of this is as a penalty method for solving a least-squares optimisation problem for the remaining data, where the penalty terms approximate equality constraints for the important data.

Table 2.7 shows that by increasing the weights of the index in the objective function (Cases 1 and 2), the market observed spreads can be fitted accurately, especially when decreasing the weight of the super-senior tranche (Case 1). This is of course at the expense of worse fit of tranche spreads (especially the last tranche in Case 1). Calibrated parameter values and measures of fit are given in Tables 2.8 and 2.9, respectively.

2.4 Conclusions

In this chapter, we presented the model that we consider in the thesis. We show how the model can be calibrated to individual CDSs by means of semi-analytical formulae for survival probabilities, we provide an economic interpretation of the model and its pricing performance.

The jump-diffusion model proves more flexible than the pure diffusion model to fit market data in vastly different scenarios, with only a small number of parameters. However, to employ the model for pricing and risk management in practice, it is necessary to further extend the model to allow a richer dependence and term structure. This can be achieved in a number of ways.

A possible straightforward extension is to take into account different finite activity jump processes, or even infinite activity, assuming discrete default monitoring. Also, a stochastic volatility common to all assets, correlated, e.g., to the driving Brownian market factor, can be easily included.

Contagion effects can be introduced by making the correlation coefficient loss dependent. Similar to local volatility models, such a "local correlation model" could be specified to provide a perfect fit to tranche spreads, while the correlation does not impact index and CDS spreads, which aids calibration. Contagion could also manifest in loss dependent (and hence stochastic) recovery rates.

The model could be further extended by a random default barrier, which has been shown to give more realistic short-term credit spreads and would be an obvious and simple

T=5 Years			
	Market	Case 1	Case 2
Index	215	214.6	222.0
Tranche 0%-3%	71.5%	81.6%	80.2%
Tranche 3%-6%	1576.3	1775.1	1675.4 9
Tranche 6%-9%	811.5	763.7	726.4
Tranche 9%-12%	506.1	419.9	417.2
Tranche 12%-22%	180.3	231.9	252.4
Tranche 22%-100%	77.9	12.7	36.2
T=7 Years			
	Market	Case 1	Case 2
Index	195	193.8	198.1
Tranche 0%-3%	72.9%	83.7%	82.0%
Tranche 3%-6%	1473.2	1736.4	1575.1
Tranche 6%-9%	804.2	781.6	710.7
Tranche 9%-12%	512.4	448.2	422.5
Tranche 12%-22%	182.6	241.6	255.1
Tranche 22%-100%	75.8	7.6	30.2
T=10 Years			
	Market	Case 1	Case 2
Index	175	174.6	175.3
Tranche 0%-3%	73.8%	85.4%	84.2%
Tranche 3%-6%	1385.5	1691.9	1488.8
Tranche 6%-9%	824.7	781.3	679.7
Tranche 9%-12%	526.1	452.6	409.3
Tranche 12%-22%	174.1	246.3	247.9
Tranche 22%-100%	76.3	3.4	23.7

Table 2.7: Calibration results for different weight vectors for 5 December 2008, iTraxx Main Series 10 index. Case 1: $\alpha_i = 10\,(CI_0(T_i))^{-2}$, $\alpha_i^j = (C_0^j(T_i))^{-2}$, $j = 1,\ldots,5$, $\alpha_i^6 = 0.01\,(C_0^6(T_i))^{-2}$. Case 2: $\alpha_i = 10\,(CI_0(T_i))^{-2}$, $\alpha_i^j = (C_0^j(T_i))^{-2}$, $j = 1,\ldots,6$. Estimated parameters are given in Table 2.8, measures of fit in Table 2.9. We assume that $r = 0.033$, $R = 0.4$.

Parameters	Case 1	Case 2
σ	0.11	0.10
ρ	0.25	0.27
λ	0.042	0.0397
$\mathbb{E}[Y-1]$	-0.16	-0.34 6
$\mathrm{Var}[Y-1]$	0.82	0.24
$\rho_{X_t^i X_t^j}$	0.85	0.82

Table 2.8: Parameters estimated for 5 December 2008, overall correlation $\rho_{X_t^i X_t^j}$ as in (2.14).

Measure of Fit	Case 1	Case 2
RMSE	4.24	3.64
ARPE	0.25	0.22

Table 2.9: Measures of fit for 5 December 2008. Root Mean Square Error (RMSE) is defined by (2.16), Average Relative Percentage Error (ARPE) is given by (2.15).

extension to the model.

With such improved fit to standard tranches, the calibrated model would be suitable for interpolating bespoke tranches on the same underlying pool. Applications to over-the-counter basket products consisting of illiquid names would be conceivable if information on the individual default probabilities could be obtained from, e.g., balance sheets; if the company is listed, parameters may be obtained from equity time series; finally, parameters for global factors may be obtained from liquid products of names operating in similar sectors.

Chapter 3

Multilevel simulation

In this chapter, we present our first framework developed to numerically evaluate a multi-name model. The calculations are performed for a jump-diffusion model. However, the method can be applied to any Bernoulli mixture model subject to some technical conditions.

3.1 Introduction

This chapter is concerned with the behaviour of functionals of a large number of exchangeable random variables and the efficient numerical estimation of their expectation. The objective of this work is thus two-fold: to analyse the order of convergence in $1/N$ of expected functionals for $N \to \infty$, and to derive estimators for these expectations for which the computational complexity is asymptotically independent of N.

First, we analyse convergence in the case of general Lipschitz and smooth functions p of the average of N exchangeable Bernoulli random variables as N goes to infinity. We then specify p further to certain piecewise linear functions and show that the convergence order is the same as in the smooth case. These results are relevant, for instance, if one wants to approximate the result for large but finite N by its limit. A number of applications comes from the credit risk literature.

A driving practical motivation for investigating the limiting behaviour is that the original sequence of random variables is costly to simulate, because of the large number N of underlying processes, often required over large time horizons. Moreover, often many Monte Carlo samples are necessary for sufficiently accurate estimation of, for instance, expected tranche losses of credit basket. This chapter takes a different tack and develops a simulation method where the computational complexity is asymptotically independent of N. A small tweak of the algorithm can also be used to approximate the limit obtained for $N \to \infty$.

More concretely, it turns out that an interpretation of the multi-level Monte Carlo approach (see [38]) in the present context allows us to construct estimators based on sequences with increasing lengths and a number of samples which decreases faster than the length increases, such that the overall computational complexity is essentially no larger than for fixed small N.

A conceptually similar though distantly related approach is used in [5], where the multilevel idea is applied to a sequence of martingales to estimate a dual upper bound for the value of an early exercise option. In that setting they are able to show, as we do here, that the achievable complexity is not substantially larger than that of a non-nested simulation. The general problem of estimating conditional expectations through nested multilevel simulation is addressed in [15]. There, further extrapolation is used to reduce the bias of estimators, while here we will propose an improved estimator which reduces the variance of higher level estimators.

Overview of the chapter

This chapter is organised as follows. In Section 3.2, we introduce the setting and outline the main convergence results, explaining how they can be used to construct efficient estimators. The first key result on the convergence order of expected functionals is proved in Section 3.3, with numerical illustrations from an example of a basket credit derivative presented

in Section 3.4. In Section 3.5, we introduce in detail two multilevel simulation methods and derive bounds on their computational complexity to achieve a prescribed accuracy. Finally, in Section 3.6 we present numerical results illustrating the efficiency gains achieved through multilevel simulation in this context and Section 3.7 concludes.

3.2 Set-up and main results

In this chapter, we are concerned with the behaviour of "loss" variables describing the fraction of N random variables in a certain state, and expected functionals of this loss variable, as N goes to infinity. The application we have in mind, and for which we will present numerical illustrations, is that of a basket of defaultable firms, and then the loss is the fraction of firms which default over a certain period.

More precisely, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a sequence of Bernoulli random variables Y_i, $i \in \mathbb{N}$, and a random variable L taking values in $[0,1]$. If required we write $\Omega = \Omega_Y \times \Omega_L$ where canonically we could take $\Omega_Y = \{0,1\}^{\mathbb{N}}$ and $\Omega_L = [0,1]$. The probability measure $\mathbb{P}$ is constructed as follows. The random variable L is generated according to its marginal law $\mathbb{P}_L$ and then, conditional on $\mathcal{F}_L$, the σ-algebra generated by L, the Y_i are independent random variables with law given by

$$\mathbb{P}[Y_i = 1 | \mathcal{F}_L] = L. \tag{3.1}$$

Thus for each $n \in \mathbb{N}$

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n, L \in B) = \int_B l^{s_n}(1-l)^{n-s_n} \mathbb{P}_L(L \in dl), \quad \forall y_i \in \{0,1\}, B \subset [0,1]$$

where $s_n = \sum_{i=1}^n y_i$. We will often write $\mathbb{P}_{|L} = \mathbb{P}(. \mid \mathcal{F}_L)$ for the conditional law of the Y_i given $\mathcal{F}_L$ and $\mathbb{E}_{|L}$ for the associated conditional expectation. In the setting of defaultable firms, $Y_i = 1$ iff the i-th firm defaults, and L is a global factor modelling the common

tendency of firms to default. We define the loss variable to be the proportion of Bernoulli variables in state 1

$$L_N = \frac{1}{N}\sum_{i=1}^{N} Y_i. \tag{3.2}$$

We consider a Lipschitz function p and random variables P and P_N defined as

$$P \equiv p(L), \tag{3.3}$$

$$P_N \equiv p(L_N). \tag{3.4}$$

In particular, we will study p of the form

$$p(l) \equiv [l-K_1]^+ - [l-K_2]^+ = \begin{cases} 0 & l \le K_1, \\ l-K_1 & K_1 \le l \le K_2, \\ K_2-K_1 & l \ge K_2, \end{cases} \tag{3.5}$$

where $[x]^+ = \max(x, 0)$ denotes the positive part and $0 \le K_1 < K_2 \le 1$ are constants. In credit derivative pricing, the particular shape of the function p in (3.5) measures the losses in a certain *tranche* with attachment point K_1 and detachment point K_2, and its expectation is the building block for formulae for CDO tranche spreads. A typical CDO pool consists of $N = 125$ firms, while typical loan or mortgage books can have substantially more obligors, and it is therefore practically relevant to understand the behaviour of expected functionals for large N and to devise computationally efficient estimators.

By a conditional version of the strong law of large numbers and the continuity of p

$$L_N \to L \quad \text{for } N \to \infty, \ \mathbb{P}_{|L} - a.s., \tag{3.6}$$

$$P_N \to P \quad \text{for } N \to \infty, \ \mathbb{P}_{|L} - a.s. \tag{3.7}$$

This convergence will also hold in $L^2(\Omega_Y, \mathbb{P}_{|L})$ (see Lemma 3.3.1).

We study here the convergence rate of $P_N - P$ and will prove the following two results.

The first statement for Lipschitz and smooth functions p is a relatively straightforward consequence of (3.1) and the easily computable convergence rate of L_N in $L^2(\Omega_Y, \mathbb{P}_{|L})$. The second result shows that for a specific p which is only piecewise smooth we can still obtain the same convergence order as in the smooth case and with explicitly computable bounds.

Theorem 3.2.1. *Let P and P_N be defined by* (3.3) *and* (3.4)*, respectively, and assume that p is Lipschitz with constant c_p. We have that*

$$|\mathbb{E}[P_N - P]| \leq \frac{c_p}{2\sqrt{N}}, \tag{3.8}$$

$$Var[P_N - P] \leq \frac{c_p^2}{4N}. \tag{3.9}$$

If, moreover, p is differentiable and the derivative has Lipschitz constant C_p, then

$$|\mathbb{E}[P_N - P]| \leq \frac{C_p}{8N}. \tag{3.10}$$

Theorem 3.2.2. *For p defined in (3.5), if the cumulative density function (CDF) F_L of L is Lipschitz at any $K_1 > 0$ and $K_2 < 1$ with Lipschitz constant c_L, i.e.,*

$$|F_L(K_j) - F_L(l)| \leq c_L \, |K_j - l| \tag{3.11}$$

for $j = 1, 2$ and all $l \in [0, 1]$, then

$$|\mathbb{E}[P_N - P]| \leq \frac{4c_L\sqrt{\pi}}{N}.$$

Note that if L has a density function which is bounded, then the CDF is certainly Lipschitz. The fact that we only need the Lipschitz property at K_1 and K_2 will be useful for the applications considered later.

Taking the two Theorems together, order 1 for the convergence of expectations also

follows for piecewise smooth p which are Lipschitz overall, provided F_L is Lipschitz. An example of such a function is a tranche loss.

These Theorems show that expected functionals for large or infinite N can be successively approximated by those with smaller N. Combining this with a control variate idea leads to multilevel simulation with a substantial variance reduction for large N. Specifically, the above results imply that for Lipschitz p we have $|\mathbb{E}[P_N - P_{MN}]| \leq c_1/\sqrt{N}$ and $Var[P_N - P_{MN}] \leq c_2/N$ for any positive integer M with some constants c_1 and c_2. We can consider a sequence $N_l = M^l$, $l \in \mathbb{N}$, with corresponding $L^{(l)} = L_{N_l}$ and $P^{(l)} = P_{N_l}$. Translating the central idea in [38] to this setting, we estimate in the decomposition

$$\mathbb{E}[P^{(l)}] = \mathbb{E}[P^{(0)}] + \sum_{k=1}^{l} \mathbb{E}[P^{(k)} - P^{(k-1)}] \tag{3.12}$$

every summand separately with independent samples of different sizes. These can be chosen to obtain an optimal allocation of computational cost for a given overall mean-square error (MSE). The general construction in [38] immediately gives the following result.

Proposition 3.2.1 (cf. [38], Theorem 3.1)**.** *Let P, $P^{(l)}$ as above. If there exist independent estimators Z_l based on n_l Monte Carlo samples, and positive constants $\alpha, \beta, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}$ and*

(i) $\left|\mathbb{E}[P^{(l)} - P]\right| \leq c_1 \, M^{-\alpha \, l}$

(ii) $\mathbb{E}[Z_l] = \begin{cases} \mathbb{E}[P^{(0)}], & l = 0 \\ \mathbb{E}[P^{(l)} - P^{(l-1)}], & l > 0 \end{cases}$

(iii) $Var[Z_l] \leq c_2 \, n_l^{-1} M^{-\beta \, l}$

(iv) $C_l \leq c_3 \, n_l \, N_l$, *where C_l is the computational complexity of Z_l*

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$, where e is Napier's

constant, there are values K and n_l for which the multilevel estimator

$$G_K = \sum_{l=0}^{K} Z_l, \tag{3.13}$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[(G_K - E[P])^2\right] < \varepsilon^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > 1, \\ c_4\, \varepsilon^{-2}(\log \varepsilon)^2, & \beta = 1, \\ c_4\, \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Corollary 3.2.1. *Let P_N and P be as in (3.3) and (3.4), and assume p is Lipschitz.*

1. *There is a multilevel estimator for $\mathbb{E}[P]$ with MSE ϵ^2 with computational complexity $C \leq c_4(\log \epsilon)^2 \epsilon^{-2}$.*

2. *For all N, there is a multilevel estimator for $\mathbb{E}[P_N]$ with MSE ϵ^2 with computational complexity $C \leq c_4(\log \epsilon)^2 \epsilon^{-2}$, where c_4 is independent of N.*

Note that only order 1/2 is required for the convergence of expectations in Proposition 3.2.1, i, and that the complexity is then dictated by β, the case $\beta = 1$ implied by Theorem 3.2.1 for all Lipschitz payoffs being a boundary case.

We point out that for fixed N the standard (i.e., single level) Monte Carlo estimator has a complexity $C \leq c\ \epsilon^{-2}$, however, c increases linearly in N. The significance of the multilevel estimator is that the constant c_4 is independent of N.

For the specific p as in (3.5), we can exploit the piecewise linearity of p to construct multilevel estimators with even better complexity, by making the following observations:

The summands in (3.12) are unchanged if we replace $P^{(k-1)} = p(L^{(k-1)})$ with any of $p(L_m^{(k-1)})$ for $m = 1, \ldots, M$, where

$$L_m^{(k-1)} \equiv \frac{1}{N_{k-1}} \sum_{i=1}^{N_{k-1}} Y_{i+(m-1)N_{k-1}}. \tag{3.14}$$

This is a direct consequence of the exchangeability. Now,

$$L^{(k)} = \frac{1}{M} \sum_{m=1}^{M} L_m^{(k-1)} \tag{3.15}$$

and, if all $L_m^{(k-1)}$ lie in the same interval $[0, K_1]$, $(K_1, K_2]$ or $(K_2, 1]$, also $P^{(k)} = \overline{P}^{(k-1)}$, where

$$\overline{P}^{(k-1)} \equiv \frac{1}{M} \sum_{m=1}^{M} P_m^{(k-1)} = \frac{1}{M} \sum_{m=1}^{M} p(L_m^{(k-1)}), \tag{3.16}$$

since p is linear in these intervals. Because of $\mathbb{E}[P^{(k-1)}] = \mathbb{E}[\overline{P}^{(k-1)}]$, we can now write

$$\mathbb{E}[P^{(l)}] = \mathbb{E}[P^{(0)}] + \sum_{k=1}^{l} \mathbb{E}[P^{(k)} - \overline{P}^{(k-1)}], \tag{3.17}$$

and if we estimate the individual terms in the sum independently in the multilevel spirit, there is only a variance contribution from a specific sample of the k-th term if at least two $P_m^{(k-1)}$ lie in different intervals. For large k, the probability of this is small, and we will be able to show the following result.

Theorem 3.2.3. *For p as in (3.5), let $P^{(l)}$ as in Proposition 3.2.1 and $\overline{P}^{(l-1)}$ as in (3.16). If the CDF F_L of L is Lipschitz with Lipschitz constant c_L, then*

$$Var[P^{(l)} - \overline{P}^{(l-1)}] \quad \leq \quad \frac{c_2}{N_l^{3/2}}, \tag{3.18}$$

where $c_2 = c_L \, 4\sqrt{M\pi}(\sqrt{2} + \sqrt{M})\sqrt{\frac{7}{8}(M^2 + 6M + 1)}$.

Here and throughout the chapter we give explicit expressions for the constants. These should not be regarded as optimal in any sense.

Corollary 3.2.2. *For Lipschitz F_L and p as in (3.5), there is a constant c_5 and multilevel estimators for $\mathbb{E}[P]$ and $\mathbb{E}[P_N]$ with MSE ϵ^2 with computational complexity $C \leq c_5\, \epsilon^{-2}$.*

The point here is that we managed to remove the logarithmic factor present in Corollary 3.2.1 and that c_5 does not depend on N.

3.3 Proof of convergence rates

We first prove Theorem 3.2.1 which contains statements in the general and smooth case. The rest of this section is devoted to the proof of Theorem 3.2.2 dealing with a specific non-smooth payoff relevant to our application.

Lemma 3.3.1. *Let P_N and P be as in (3.3) and (3.4), and assume p is Lipschitz with constant c_p. Then*

$$\mathbb{E}_{|L}[(P_N - P)^2] \leq \frac{c_p^2}{4N}.$$

Proof. Since the function p in (3.3) is assumed Lipschitz and $\mathbb{E}_{|L}[L_N] = L$, we have

$$\begin{aligned} \mathbb{E}_{|L}[(P_N - P)^2] &\leq c_p^2\, \mathbb{E}_{|L}[(L_N - L)^2] = c_p^2\, Var[L_N|\mathcal{F}_L] \\ &= \frac{c_p^2}{N}\, Var[Y_i|\mathcal{F}_L] = \frac{c_p^2}{N}\, L(1-L). \end{aligned}$$

For $L \in [0,1]$, $L(1-L) \leq \frac{1}{4}$, which gives the result. □

Proof of Theorem 3.2.1. Equation (3.9) follows directly from Lemma 3.3.1, and then, by Hölder's inequality,

$$|\mathbb{E}[P - P_N]| \leq \sqrt{\mathbb{E}[\mathbb{E}_{|L}[(P_N - P)^2]]} \leq \frac{c_p}{2\sqrt{N}}.$$

The rest of the proof applies only to the function p satisfying the conditions stated in the Theorem 3.2.1, the proof is given in [11]. □

Now, we turn to the proof of Theorem 3.2.2 and show a few Lemmas first. We divide the ranges of L and L_N into the three intervals $I_1 = [0, K_1]$, $I_2 = (K_2, 1]$ and $I_3 = (K_1, K_2]$, in each of which the function p from (3.3) is linear; the point being that the probability of L and L_N lying in different intervals is small for large N, and the expected difference of $P - P_N$ is small if they are in the same interval. The following Lemmas quantify this.

Lemma 3.3.2. *For $j = 1, 2$, we have*

$$\mathbb{P}_{|L}(L \in I_j, L_N \in I_j^c) \leq 1_{L \in I_j}\, e^{-N(L-K_j)^2}, \tag{3.19}$$

$$\mathbb{P}_{|L}(L \in I_j^c, L_N \in I_j) \leq 1_{L \in I_j^c}\, e^{-N(L-K_j)^2}. \tag{3.20}$$

Proof. This is a standard large deviations result. By Theorem 2.2.3 in [21], p. 27, and Remark (c) thereafter, for $\mathcal{F}_L$-independent and identically distributed random variables $(Y_i)_{1 \leq i \leq N}$ with $\mathbb{E}[Y_i \mid \mathcal{F}_L] = L$, we obtain that if $0 < L \leq K_j$,

$$\mathbb{P}_{|L}(L_N > K_j) \leq e^{-Ng(L,K_j)},$$

and if $K_j < L < 1$,

$$\mathbb{P}_{|L}(L_N \leq K_j) \leq e^{-Ng(L,K_j)},$$

where the rate function $g(L, K_j)$ is given on p. 35 in [21] as

$$g(L, K_j) = K_j \, \log\left(\frac{K_j}{L}\right) + (1 - K_j) \, \log\left(\frac{1 - K_j}{1 - L}\right),$$

since Y_i are Bernoulli distributed random variables with $P(Y_i = 1 \mid \mathcal{F}_L) = L$. It is

straightforward to check that for all $L \in (0,1)$

$$g(L, K_j) \geq (K_j - L)^2. \tag{3.21}$$

Hence, by (3.21), for $0 < L \leq K_j$

$$\mathbb{P}_{|L}(L_N > K_j) \leq e^{-Ng(L,K_j)} \leq e^{-N(K_j-L)^2}, \tag{3.22}$$

and similarly for $K_j < L < 1$. These estimates are clearly true for the degenerate cases $L = 0$ and $L = 1$. From this the result follows.

□

Lemma 3.3.3. *Let p be as in (3.5). If A_N is the event that L_N and L are in the same interval and A_N^c its complement, then*

$$\mathbb{E}[(P_N - P)1_{A_N}] = -\mathbb{E}[(L_N - L)1_{A_N^c} 1_{\{L \in I_3\}}]. \tag{3.23}$$

Proof. By splitting the range of L into the different intervals,

$$\begin{aligned} \mathbb{E}[(P_N - P)1_{A_N}] &= \sum_{j=1}^{3} \mathbb{E}[(P_N - P)1_{A_N} 1_{\{L \in I_j\}}] \\ &= \mathbb{E}[(L_N - L)1_{A_N} 1_{\{L \in I_3\}}] \\ &= -\mathbb{E}[(L_N - L)1_{A_N^c} 1_{\{L \in I_3\}}], \end{aligned}$$

where we have used in the second line that $P_N = P$ if both L_N and L lie in either I_1 or I_2 and that $P_N - P = L_N - L$ in I_3; in the last line that $\mathbb{E}_{|L}[L_N - L] = 0$ and $1_{A_N} + 1_{A_N^c} = 1$. □

Lemma 3.3.4. *Let A_N^c be as in Lemma 3.3.3. If the CDF F_L of L is Lipschitz at K_j,*

$j = 1, 2$, *with constant* c_L, *then*

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right] \leq \frac{c_L\, 4\sqrt{\pi}}{\sqrt{N}}. \tag{3.24}$$

Proof. Let $I_1^c = (K_1, 1]$ and $I_2^c = [0, K_2]$ be the complements in $[0, 1]$ of I_1 and I_2, then

$$A_N^c \subseteq \{L \in I_1, L_N \in I_1^c\} \cup \{L \in I_2, L_N \in I_2^c\} \cup \{L \in I_1^c, L_N \in I_1\} \cup \{L \in I_2^c, L_N \in I_1^c\}$$

and therefore

$$\begin{aligned} \mathbb{P}_{|L}[A_N^c] \;\leq\; & \mathbb{P}_{|L}[L \in I_1,\ L_N \in I_1^c] + \mathbb{P}_{|L}[L \in I_2,\ L_N \in I_2^c] + \mathbb{P}_{|L}[L \in I_1^c,\ L_N \in I_1] \\ & + \mathbb{P}_{|L}[L \in I_2^c,\ L_N \in I_2]. \end{aligned}$$

By (3.19), (3.20) we have

$$\mathbb{P}_{|L}[A_N^c] \leq 2\ \left(e^{-N(L-K_1)^2} + e^{-N(L-K_2)^2}\right),$$

and we obtain

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right] \;\leq\; 2^{\frac{1}{2}}\ \left(\mathbb{E}\left[e^{-N\frac{(L-K_1)^2}{2}}\right] + \mathbb{E}\left[e^{-N\frac{(L-K_2)^2}{2}}\right]\right). \tag{3.25}$$

If we extended F_L by 0 and 1 from $[0,1]$ to $\mathbb{R}$ then, for $j = 1, 2$, we have

$$\begin{aligned}
\mathbb{E}\left[e^{-N\frac{(L-K_j)^2}{2}}\right] &= \int_{-\infty}^{\infty} e^{-N\frac{(l-K_j)^2}{2}}\, dF_L(l) && (3.26)\\
&= N\int_{-\infty}^{\infty}(l-K_j)e^{-N\frac{(l-K_j)^2}{2}}F_L(l)\, dl && (3.27)\\
&\leq NF_L(K_j)\int_{-\infty}^{\infty}(l-K_j)e^{-N\frac{(l-K_j)^2}{2}}\, dl\\
&\quad + c_L\, N\int_{-\infty}^{\infty}(l-K_j)^2 e^{-N\frac{(l-K_j)^2}{2}}\, dl\\
&= \frac{c_L\sqrt{2\pi}}{\sqrt{N}},
\end{aligned}$$

where we used the Lipschitz property of the CDF after (3.27) and then integrated exactly. The result follows directly by insertion in (3.25). □

Proof of Theorem 3.2.2. By the tower property of conditional expectations and Jensen's inequality, we have

$$|\mathbb{E}[(P_N - P)\, 1_{A_N^c}]| \leq \mathbb{E}[|\mathbb{E}_{|L}[(P_N - P)\, 1_{A_N^c}]|]. \tag{3.28}$$

Then Cauchy-Schwarz gives

$$|\mathbb{E}[(P_N - P)\, 1_{A_N^c}]| \leq \mathbb{E}_{|L}\left[\left(\mathbb{E}[(P_N - P)^2]\right)^{\frac{1}{2}}\ \left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right]. \tag{3.29}$$

By Lemmas 3.3.1 and 3.3.4, we obtain

$$|\mathbb{E}[(P_N - P)\, 1_{A_N^c}]| \leq \frac{c_L 2\sqrt{\pi}}{N}. \tag{3.30}$$

Similarly, using Lemma 3.3.3 and the same argument as above,

$$|\mathbb{E}[(P_N - P)\, 1_{A_N}]| \leq \frac{c_L 2\sqrt{\pi}}{N},$$

from which the statement follows.

□

3.4 An application and numerical results

To illustrate the theoretical rate of convergence, we study numerical results for expected tranche losses of a synthetic CDO for an increasing size N of the underlying CDS pool.

We consider the structural factor model that we presented in Chapter 2. Here, for convenience, we state the model again. The *distance-to-default* of the i-th firm, $i = 1, \ldots, N$, evolves according to

$$X_t^i = X_0^i + \beta t + \sqrt{1-\rho}\, W_t^i + \sqrt{\rho}\, W_t^\varsigma + \tilde{J}_t, \quad t > 0, \tag{3.31}$$

where $\rho \in [0, 1)$, β given. Here, W^ς is assumed to be a standard Brownian motion and $\tilde{J}_t = \sum_{k=1}^{\tilde{N}_t} \Pi_k$, where $\tilde{N}_t$ is a Poisson process with intensity λ and Π_k are independent Normals with mean μ_Π and variance σ_Π^2, while all W^i are independent standard Brownian motions and independent of W^ς and $\tilde{J}$. Thus W^ς and $\tilde{J}$ model factors affecting the whole market, whereas W^i are idiosyncratic effects.

The i-th firm is considered to be in default if its distance-to-default is below 0 at any one of the observation times $T_j = jq$, $q = 0.25$ (quarterly), up to $T_{20} = T = 5$, the assumed maturity of the debt here. We introduce the default time τ_i^D and Bernoulli random variable Y_i indicating default of the i-th firm before T, by

$$\begin{aligned} \tau_i^D &= \inf\left(\{t \in \{T_1, \ldots, T_m\} : X_t^i \le 0\} \cup \{\infty\}\right), \\ Y_i &= 1_{\{\tau_i^D \le T\}}. \end{aligned}$$

For the numerical experiments, the initial values X_0^i are drawn independently from a

Normal distribution,

$$X_0^i \sim N(\mu_{X_0}, \sigma_{X_0}^2),$$

where the mean $\mu_{X_0} = 4.6$ and standard deviation $\sigma_{X_0} = 0.8$ are obtained from a calibration to iTraxx data as detailed in [12], as are $\rho = 0.13$, $\lambda = 0.04$, $\mu_\Pi = -0.5$ and $\sigma_\Pi^2 = 0.17$.

We illustrate the CDF F_L of L for different parameters in Figure 3.1. To this end, we generated samples of L by simulating the SPDE satisfied by the limit empirical measure of distances-to-default for $N \to \infty$, see [12] for details. It appears that F_L is Lipschitz in $(0, 1)$ but that the derivative at 0 and 1 can become very large in certain parameter ranges for $\mu_0 = \mu_{X_0}$ and overall instantaneous correlation

$$\rho_A = (\rho + \vartheta)/(1 + \vartheta), \quad \vartheta = \lambda(\mu_\Pi^2 + \sigma_\Pi^2), \tag{3.32}$$

between X_t^i and X_t^j, $i \neq j$ (see [12]).

For large values of μ_0, the probability of defaults becomes very small and the density of L is concentrated around 0. For ρ_A approaching 1, all Y_i become identical and therefore either all or none of the firms default, such that here the density of L is concentrated at 0 and 1. In the degenerate case $\rho_A = 0$ (i.e., $\rho = \lambda = 0$), L is deterministic, the measure is atomic and F_L a step function.

The empirical evidence thus suggests that F_L is Lipschitz in the range $(0, 1)$. Given that Theorem 3.2.2 only requires the Lipschitz property at interior values K_j, the conditions appear to be satisfied and the Theorem to apply in this setting. Even in situations where F_L has a bounded derivative at 0 and 1, the fact that only the Lipschitz constants from K_1 and K_2 enter into the estimates gives us substantially smaller bounds.

We now move on to present numerical results for the payoff function p from (3.5) illustrating the convergence as the number of firms N goes to infinity. We consider portfolios

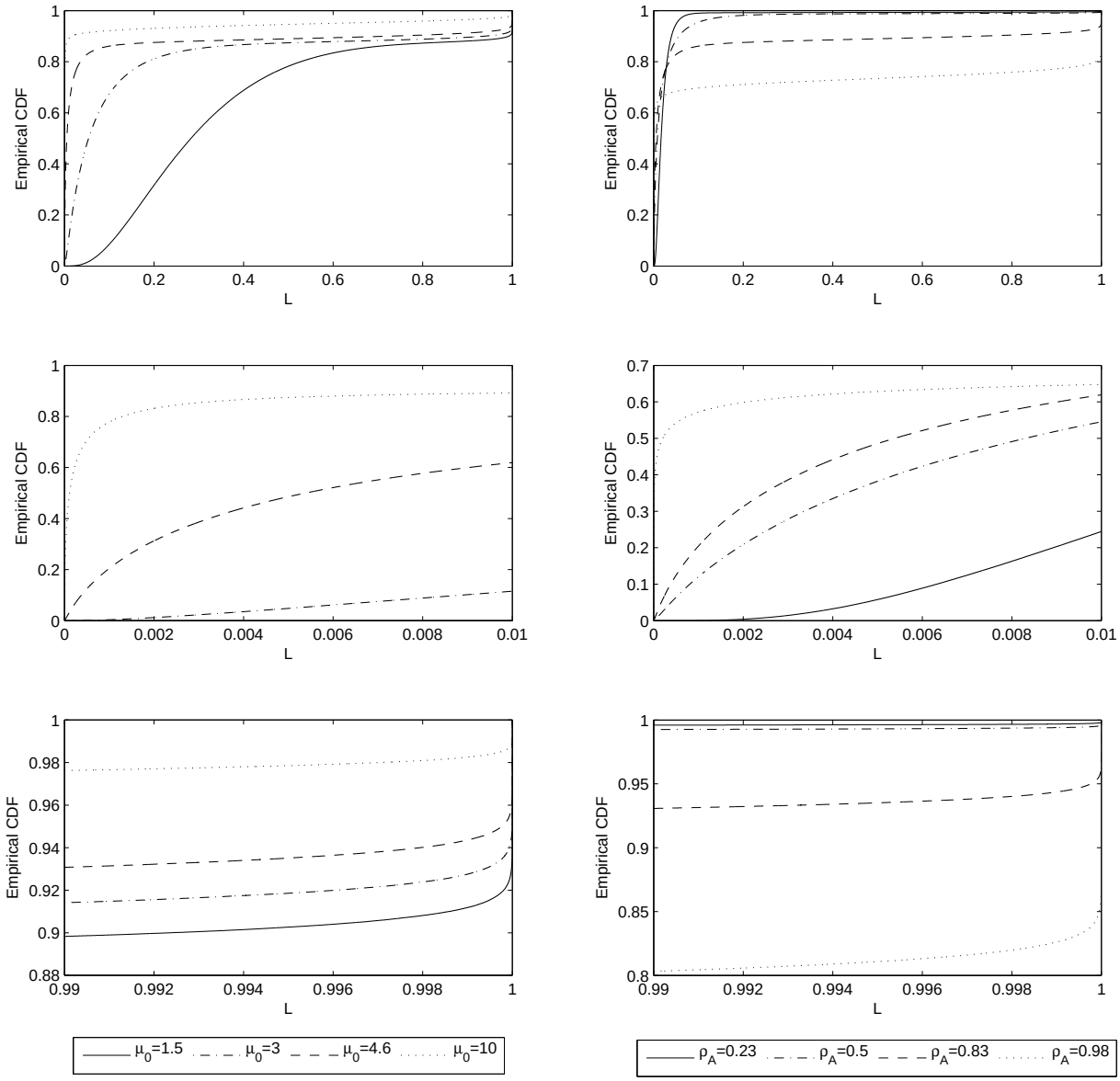

Figure 3.1: Top row: Empirical CDF F_L for different values of $\mu_0 = \mu_{X_0}$ (left) and different ρ_A (right) The values of ρ_A are arrived at by (3.32) from $(\rho, \lambda) \in \{(0.03, 0.001), (0.1, 0.002), (0.35, 0.0035), (0.35, 0.0351), (0.8, 0.1)\}$. All other parameters are fixed as given in the text. The plots in the second and third rows are zoomed into the ranges of L close to 0 and 1, respectively.

consisting of $N_k = M^k = 5^k$ companies for $k = 1, \ldots, 7$.

To include a recovery value of defaulted firms in the model, we rescale L_N by $(1-R)$, where $R = 0.4$ is the recovery rate. Equivalently, we pick $(K_1, K_2) = (1-R)^{-1}(a, d)$ in (3.3) and $(a, d) \in \{(0, 0.03), (0.03, 0.06), (0.06, 0.09), (0.09, 0.12), (0.12, 0.22), (0.22, 1)\}$ as the attachment and detachment points for iTraxx tranches, and then study $(1-R)p(L_N)$.

A straightforward Monte Carlo estimator for expected tranche losses $\mathbb{E}[P^{(k)}]$ is then given by

$$\widehat{G}_k = \frac{1}{n}\sum_{j=1}^{n}(1-R)p(L^{(k,j)}), \tag{3.33}$$

$$L^{(k,j)} = \frac{1}{N_k}\sum_{i=1}^{N_k} Y_i^{(j)}, \tag{3.34}$$

where $(Y_i^{(j)})$ are independent samples of Y_i, i.e., corresponding to independent paths for W^ς, W^i and $\tilde{J}$. There is no time discretisation error as (3.31) can be sampled directly. However, it turns out to be computationally prohibitively expensive to choose n, the number of samples, large enough to produce estimators with sufficiently small RMSE to allow us to distinguish between $\widehat{G}_k$ and $\widehat{G}_{k+1}$ for large k.

We therefore use the multilevel simulation approach outlined in Section 3.2 and detailed further in Section 3.5. The point is that the differences $G_{k+1} - G_k$ are simulated directly in the multilevel approach. Therefore, we approximate $|G - G_k|$, where $G \equiv \lim_{k\to\infty} G_k$, by

$$S_k = |G_k - G_K| = \left| \sum_{l=k+1}^{K} Z_l \right| \tag{3.35}$$

for $k < K$, where Z_l is an estimator for $\mathbb{E}[P^{(l)} - P^{(l-1)}]$ as used in the construction of G_k in (3.13) (precisely, we used the estimator Z_l defined later in (3.37)). The difference between S_k and $|G - G_k|$ for $k = K-1$ is given by $G_{K-1} - G_K \approx (G_{K-1} - G)(1 - 1/M)$ and for $k = K-2$ by $G_{K-2} - G_K \approx (G_{K-2} - G)(1 - 1/M^2)$. Given $M = 5$ in our examples, the error due to this approximation will be seen to be smaller than the estimation error.

The choice of $M = 5$ is dictated by our numerical application where the typical size of a CDO basket is $N = 125 = 5^3$. The discussion on the optimal selection of M for exchangeable Bernoulli random variables can be found in [11], while for a general application of multilevel Monte Carlo is given in [38].

The results of our numerical experiments are shown in Figure 3.2. We plot the logarithm of S_k to base M, together with the sample standard deviation of the the multilevel estimators G_k (see (3.13)) and

$$y_k = -k + y_0, \tag{3.36}$$

where y_0 is a suitably chosen constant, to verify the predicted convergence order empirically. The data points appear to be in good agreement with first order convergence.

3.5 A multilevel method and its numerical analysis

In this section, we describe and analyse a multilevel simulation approach for the estimation of expected functionals of the form (3.3) and (3.4), the latter with a particular emphasis on the case of large N.

The multilevel Monte Carlo method proposed by Giles in [38] estimates the expected value of a functional of the solution to a stochastic differential equation obtained by a timestepping scheme. It performs computations on different refinement levels l with time steps $h_l = h_0 M^{-l}$ for $M > 1$, such as to minimise the overall computational time of the Monte Carlo estimator for prescribed mean square error (MSE). Since the MSE consists of a Monte Carlo error (variance) and a discretisation error (bias), the method controls both the number of samples n_l on level l, to bound the Monte Carlo variance of order $O(n_l^{-1})$, and the finest L with time step h^{-L} on which to approximate the SDE, in order to reduce the bias. The multilevel method is based on two premises: Monte Carlo estimators for an increasing number of time steps converge at a certain order in h_l, and the computational cost needed to calculate an estimator increases with $n_l h_l^{-1}$. In this approach, estimators

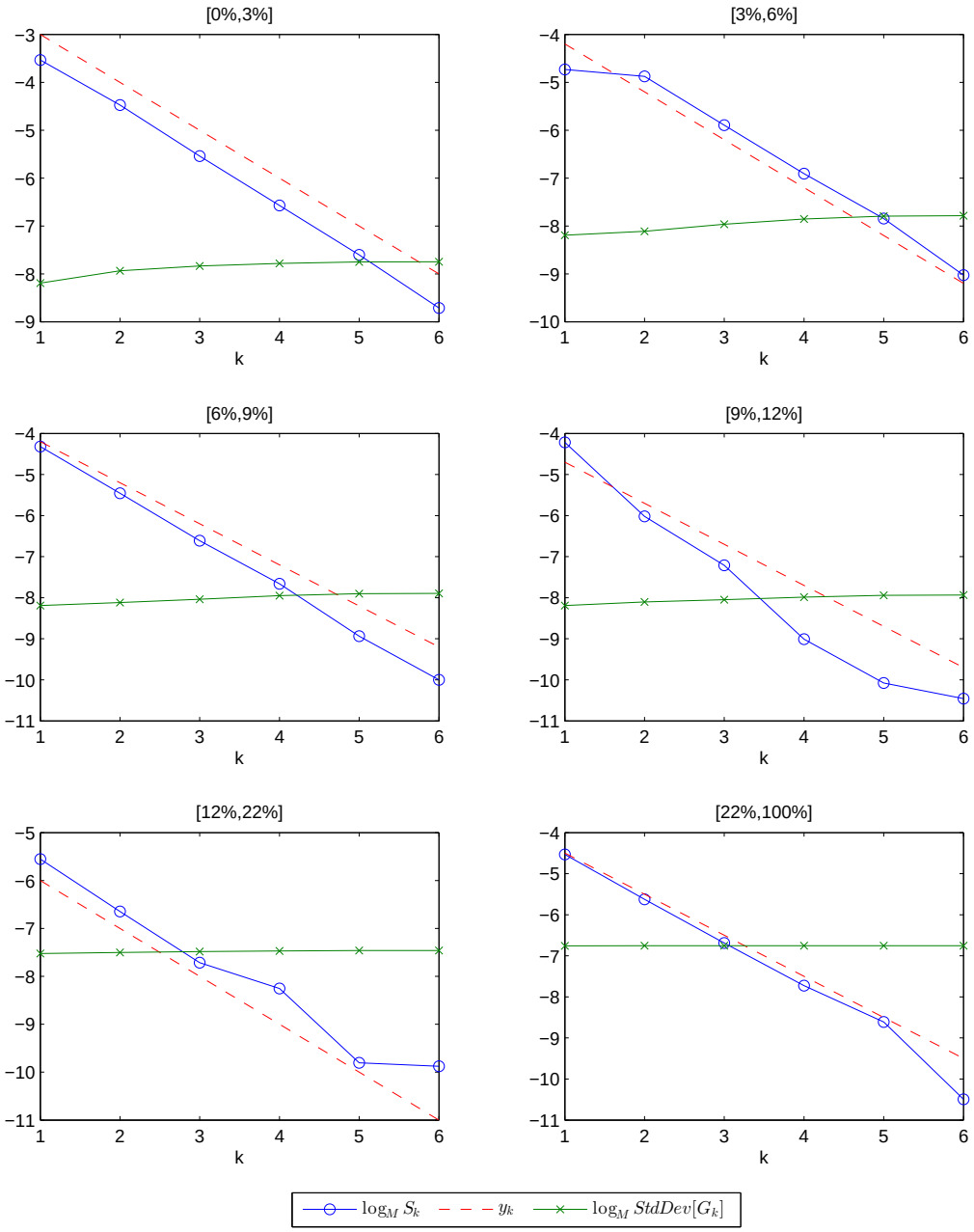

Figure 3.2: Shown here is $\log_M S_k$, where S_k given by (3.35) is an estimator for $|\mathbb{E}[P^{(k)} - P]|$. The various plots are for tranches ranging from [0%-3%] to [22%-100%], of a CDO basket consisting of $N_k = M^k = 5^k$ companies, where $k = 1, \ldots, 6$. The comparison with the predicted trend y_k from (3.36) confirms the first order convergence. Included is also the standard deviation of the estimated tranche loss G_k.

obtained with a smaller number of time steps are used as control variates for estimators with a larger number of time steps, which significantly decreases the computation time.

To obtain a complexity result for an estimator of $\mathbb{E}[P]$ with P from (3.3), we substitute h_l by N_l^{-1} in Theorem 3.1 of [38] and immediately obtain Proposition 3.2.1 from Section 3.2. We hence define estimators Z_l by

$$Z_l \equiv n_l^{-1} \sum_{j=1}^{n_l} \left(P^{(l,j)} - P_c^{(l,j)} \right), \tag{3.37}$$

where 'c' denotes a 'coarse' estimator on level l, i.e., using only N_{l-1} instead of N_l Bernoulli random variables, precisely,

$$P^{(l,j)} = p(L^{(l,j)}), \quad \text{where } L^{(l,j)} = N_l^{-1} \sum_{i=1}^{N_l} Y_i^{(l,j)}, \tag{3.38}$$

$$P_c^{(l,j)} = p(L_c^{(l,j)}), \quad \text{where } L_c^{(l,j)} = N_{l-1}^{-1} \sum_{i=1}^{N_{l-1}} Y_i^{(l,j)}, \tag{3.39}$$

where $Y_i^{(l,j)}$, $j = 1, \ldots, n_l$, are independent samples of Y_i for fixed level l and independent across levels. They are constructed from a loss factor $L^{(l,j)}$ (with the same distribution as L, independent across l and j) in the same way as Y_i is constructed from L.

By direct inspection, for this construction of Z_l, Assumption ii holds in Proposition 3.2.1. From Theorem 3.2.1, we know that i holds with $\alpha = 1/2$ for general Lipschitz p. Clearly, the computational effort to compute Z_l is proportional to $n_l N_l$ as required in iv. Finally, iii holds by the following simple application of Lemma 3.3.1.

Proposition 3.5.1. *Let $P^{(l)} = P_{N_l}$ as per (3.4), where p is Lipschitz with constant c_p, then*

$$Var[P^{(l)} - P^{(l-1)}] \leq c_p^2 \, \frac{M+1}{2N_l}. \tag{3.40}$$

Proof. This follows directly from

$$\begin{aligned}\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^2] &= \mathbb{E}_{|L}[((P^{(l)} - P) - (P^{(l-1)} - P))^2] \\ &\leq 2\left(\mathbb{E}_{|L}[(P^{(l)} - P)^2] + \mathbb{E}_{|L}[(P^{(l-1)} - P)^2]\right) \qquad (3.41)\end{aligned}$$

by Lemma 3.3.1 and taking expectations over L. □

We have therefore proven the first statement of Corollary 3.2.1.

In practice, it is also relevant to be able to compute $\mathbb{E}[P_N]$ efficiently for finite N. It is clear that for fixed N, the complexity is bounded by $c\,\epsilon^{-2}$ for some $c > 0$, but for a naïve (single-level) estimator the constant c will increase with N. From the proof of Theorem 3.1 in [38] it is clear, however, that there is a multilevel estimator with *a priori* bounded upper level K which satisfies the second statement in Corollary 3.2.1.

We now propose a multilevel estimator with even lower variance, based on the faster decay rate 3/2 in Theorem 3.2.3, which we prove subsequently. Specifically, we define estimators $\overline{Z}_l$ by

$$\overline{Z}_l \equiv n_l^{-1} \sum_{j=1}^{n_l} \left(P^{(l,j)} - \overline{P}^{(l,j)}\right), \qquad (3.42)$$

where $P^{(l,j)}$ is defined as in (3.38), but instead of $P_c^{(l,j)}$ we use

$$\overline{P}^{(l,j)} = M^{-1} \sum_{m=1}^{M} p(L_m^{(l,j)}), \quad \text{where } L_m^{(l,j)} = N_{l-1}^{-1} \sum_{i=1}^{N_{l-1}} Y_{i+(m-1)N_{l-1}}^{(l,j)}, \qquad (3.43)$$

and where the rest of the set-up is as earlier.

It is clear that $\overline{Z}_l$ satisfies ii in Proposition 3.2.1 and that the computational complexity is still bounded as required per iv. In fact, as the main computational cost is typically

in sampling Y_i, the computational complexity is virtually identical to that of Z_l. In particular, if we evaluate (3.38) by using (3.15) and the already computed (3.43), the difference in evaluating Z_l and $\overline{Z}_l$ is an $O(M)$ cost, i.e., independent of N_l. Now, given Theorem 3.2.3, we have that

$$Var[\overline{Z}_l] \leq c\, n_l^{-1} M^{-3/2\, l}, \tag{3.44}$$

for some c, such that we are in the first regime in the complexity result of Proposition 3.2.1, i.e., we have optimal complexity order.

The remainder of this section is devoted to the proof of Theorem 3.2.3.

Lemma 3.5.1. *Assume the CDF F_L of L is Lipschitz with constant c_L. Let $B^{(l)}$ be the event that $L^{(l)}$ lies in the same interval as $L^{(l-1)}$, $B^{(l),c}$ its complement, then*

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[B^{(l),c}]\right)^{\frac{1}{2}}\right] \leq \frac{C}{\sqrt{N_l}},$$

where $C = c_L\, 4\sqrt{\pi}(\sqrt{2} + \sqrt{M})$.

Proof. Let $A^{(l)}$ again be the event that $L^{(l)}$ and L are in the same interval, $A^{(l),c}$ its complement. Then from

$$B^{(l),c} \subseteq \left(A^{(l)} \cap A^{(l-1),c}\right) \bigcup \left(A^{(l),c} \cap A^{(l-1)}\right) \bigcup \left(A^{(l),c} \cap A^{(l-1),c}\right)$$

follows

$$\begin{aligned} \mathbb{P}_{|L}[B^{(l),c}] &\leq \mathbb{P}_{|L}[A^{(l)} \cap A^{(l-1),c}] + \mathbb{P}_{|L}[A^{(l),c} \cap A^{(l-1)}] + \mathbb{P}_{|L}[A^{(l),c} \cap A^{(l-1),c}] \\ &\leq 2\, \mathbb{P}_{|L}[A^{(l),c}] + \mathbb{P}_{|L}[A^{(l-1),c}], \end{aligned}$$

which leads to

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[B^{(l),c}]\right)^{\frac{1}{2}}\right] \leq \sqrt{2}\,\mathbb{E}\left[\left(\mathbb{P}_{|L}[A^{(l),c}]\right)^{\frac{1}{2}}\right] + \mathbb{E}\left[\left(\mathbb{P}_{|L}[A^{(l-1),c}]\right)^{\frac{1}{2}}\right].$$

By Lemma 3.3.4, we obtain the result. □

Lemma 3.5.2. *For P, $P^{(l)}$, $P^{(l-1)}$ and $\overline{P}^{(l-1)}$ as above, p Lipschitz with constant 1,*

$$\mathbb{E}_{|L}[(P^{(l)} - P)^4] \leq \frac{3}{16N_l^2}\left(1 + \frac{4}{3N_l}\right) \leq \frac{7}{16N_l^2}, \tag{3.45}$$

$$\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^4] \leq \frac{C}{N_l^2}, \tag{3.46}$$

$$\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4] \leq \frac{C}{N_l^2}, \tag{3.47}$$

where $C = \frac{7}{8}(M^2 + 6M + 1)$.

Proof. We begin by showing (3.45) and then deduce (3.46) and (3.47). We have

$$\mid P^{(l)} - P \mid \leq \mid L^{(l)} - L \mid,$$

where

$$L^{(l)} = \frac{1}{N_l}\sum_{i=1}^{N_l} Y_i.$$

Hence, we get

$$\mathbb{E}_{|L}[(P^{(l)} - P)^4] \leq \mathbb{E}_{|L}[(L^{(l)} - L)^4] = \mathbb{E}_{|L}\left[\left(\frac{1}{N_l}\sum_{i=1}^{N_l}(Y_i - L)\right)^4\right].$$

As L is $\mathcal{F}_L$-measurable and the Y_i are independent and identically distributed given $\mathcal{F}_L$

with $\mathbb{E}_{|L}[Y_i - L] = 0$, we have

$$\begin{aligned}
\mathbb{E}_{|L}[(L^{(l)} - L)^4] &= \frac{1}{N_l^4}\mathbb{E}_{|L}\left[\sum_{i=1}^{N_l}(Y_i - L)^4 + 6\sum_{i\neq j}(Y_i - L)^2(Y_j - L)^2\right] \\
&= \frac{1}{N_l^3}\left((1-L)^4 L + L^4(1-L)\right) \\
&\quad + \frac{3(N_l - 1)}{N_l^3}\left((1-L)^2 L + L^2(1-L)\right)^2 \\
&= \frac{3L^2(1-L)^2}{N_l^2} + \frac{L(1-L)(1-6L(1-L))}{N_l^3}.
\end{aligned}$$

Using the fact that $0 \leq L(1-L) \leq 1/4$ we have the required bound in (3.45).

For (3.46), observe that there are many ways of estimating this fourth moment; we choose the following

$$\begin{aligned}
(P^{(l)} - P^{(l-1)})^4 &= \Big((P^{(l)} - P) - (P^{(l-1)} - P)\Big)^4 \\
&\leq 2\left((P^{(l)} - P)^4 + (P^{(l-1)} - P)^4\right) + 12(P^{(l)} - P)^2(P^{(l-1)} - P)^2.
\end{aligned}$$

Therefore, using Cauchy-Schwarz on the last term and applying (3.45) we have

$$\begin{aligned}
\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^4] &\leq 2\left(\mathbb{E}_{|L}[(P^{(l)} - P)^4] + \mathbb{E}_{|L}[(P^{(l-1)} - P)^4]\right) \\
&\qquad + 12\,\mathbb{E}_{|L}[(P^{(l)} - P)^4]^{1/2}\,\mathbb{E}_{|L}[(P^{(l-1)} - P)^4]^{1/2} \\
&\leq \frac{7}{8N_l^2} + \frac{7}{8N_{l-1}^2} + \frac{42}{8N_l N_{l-1}}
\end{aligned}$$

as required to obtain (3.46).

Finally, (3.47) follows from

$$\begin{aligned}\mathbb{E}_{|L}[(P^{(l)}-\overline{P}^{(l-1)})^4] &= \mathbb{E}_{|L}\Big[\Big(\frac{1}{M}\sum_{m=1}^{M}(P^{(l)}-P_m^{(l-1)})\Big)^4\Big]\\ &\leq \mathbb{E}_{|L}\Big[\frac{1}{M}\sum_{m=1}^{M}(P^{(l)}-P^{(l-1)})^4\Big]\\ &= \mathbb{E}_{|L}[(P^{(l)}-P^{(l-1)})^4],\end{aligned}$$

and an application of (3.46). □

Proof of Theorem 3.2.3. Let $E^{(l)}$ be the event that all $L_m^{(l-1)}$ lie in the same interval, $1 \leq m \leq M$, and $E^{(l),c}$ its complement, then

$$\mathbb{E}[(P^{(l)}-\overline{P}^{(l-1)})^2] = \mathbb{E}[(P^{(l)}-\overline{P}^{(l-1)})^2 1_{E^{(l)}}] + \mathbb{E}[(P^{(l)}-\overline{P}^{(l-1)})^2 1_{E^{(l),c}}].$$

By (3.15) and linearity of p in each interval, we have

$$\mathbb{E}[(P^{(l)}-\overline{P}^{(l-1)})^2 1_{E^{(l)}}] = 0.$$

By Cauchy-Schwartz, we have

$$\mathbb{E}_{|L}[(P^{(l)}-\overline{P}^{(l-1)})^2 1_{E^{(l),c}}] \leq \Big(\mathbb{E}_{|L}[(P^{(l)}-\overline{P}^{(l-1)})^4]\Big)^{\frac{1}{2}}\Big(\mathbb{P}_{|L}[E^{(l),c}]\Big)^{\frac{1}{2}},$$

hence,

$$\mathbb{E}[(P^{(l)}-\overline{P}^{(l-1)})^2 1_{E^{(l),c}}] \leq \mathbb{E}\left[\Big(\mathbb{E}_{|L}[(P^{(l)}-\overline{P}^{(l-1)})^4]\Big)^{\frac{1}{2}}\Big(\mathbb{P}_{|L}[E^{(l),c}]\Big)^{\frac{1}{2}}\right].$$

By Lemma 3.5.2, we have that

$$\Big(\mathbb{E}_{|L}[(P^{(l)}-\overline{P}^{(l-1)})^4]\Big)^{\frac{1}{2}} \leq \frac{\sqrt{c_1}}{N_l}, \tag{3.48}$$

where $c_1 = \frac{7}{8}(M^2 + 6M + 1)$.

If we denote by $B_m^{(l)}$ the event that $L_m^{(l-1)}$ and $L^{(l)}$ lie in the same interval, then

$$E^{(l),c} = \bigcup_{m=1}^{M} B_m^{(l),c}$$

and therefore

$$\mathbb{P}_{|L}(E^{(l),c}) \;\leq\; \sum_{m=1}^{M} \mathbb{P}_{|L}(B_m^{(l),c}) \;=\; M\,\mathbb{P}_{|L}(B^{(l),c}).$$

By Lemma 3.5.1, this gives

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[E^{(l),c}]\right)^{\frac{1}{2}}\right] \leq \frac{c_2}{N_l},$$

where $c_2 = c_L\, 4\sqrt{M\pi}(\sqrt{2} + \sqrt{M})$. Together with (3.48), we obtain the result. □

3.6 Multilevel tests

In this section, we present multilevel simulation results based on the estimators from the previous section and illustrating the theoretical findings from there. We return to the example from Section 3.4 and estimate expected tranche losses for credit baskets with an increasing number of firms $N_l = M^l$.

For the estimator Z_l from (3.37), an upper bound for the variance – although not a sharp one – is analytically known from (3.40) and we could use that to determine the number n_l of samples on level l which is required to bring the variance contribution under a desired threshold. For the improved estimator $\overline{Z}_l$ from (3.42), however, the bound in (3.44) contains the unknown Lipschitz constant of the CDF of F_L via Theorem 3.2.3. In order to determine the optimal allocation n_l^*, we use the following algorithm as per [38]. In contrast to there, the upper level K is fixed here which simplifies the stopping criterion somewhat.

1. Start with $k = 1$.

2. Estimate the variance V_k of a single sample using $n_k = 10^4$ realisations.

3. Calculate the optimal number of samples, n_l^*, for $l = 0, 1, \ldots, k$, using

$$n_l^* = \left\lceil \gamma^{-2} \sqrt{V_l \; N_l^{-1}} \left(\sum_{j=1}^{k} \sqrt{V_j \; N_j} \right) \right\rceil , \tag{3.49}$$

   where γ^2 is a chosen upper bound of $Var[G_K]$.

4. Draw extra samples for each level according to n_l^*.

5. If $k < K$, set $k = k + 1$ and go to 2.

6. If $k = K$, finish.

Remark 3.6.1. *As per [38], choosing n_l^* by* (3.49)*, guarantees that the variance $Var[G_K]$ is bounded by γ^2, since*

$$Var[G_K] \; = \; \sum_{l=1}^{K} (n_l^*)^{-1} \, V_l \leq \sum_{l=1}^{K} \left(\gamma^{-2} \sqrt{V_l N_l^{-1}} \sum_{j=1}^{K} \sqrt{V_j N_j} \right)^{-1} V_l \; = \; \gamma^2.$$

A side effect is that, for $k < K$, the variance is smaller than for $k = K$, since

$$Var[G_k] = \sum_{l=1}^{k} (n_l^*)^{-1} \, V_l < \gamma^2 \frac{\sum_{l=1}^{k} \sqrt{V_l N_l}}{\sum_{l=1}^{K} \sqrt{V_l N_l}}.$$

Hence, if we compute estimators G_k for all k as a by-product of G_K, the variance is the smallest for G_1 and then for G_k, $k = 2, \ldots, K$, gradually reaches the upper bound γ^2. This effect can be observed in Figure 3.3.D.

In Figure 3.3 we show results for the same parameter setting as in Section 3.4 and only for the equity tranche. Results from other tests were very similar and did not show any

noteworthy additional effects. In order to easily see the rate of convergence in 3.3.A., we plot the logarithm of V_l to base M, together with

$$f_k = -\beta\, k + f_0 \tag{3.50}$$

for different values of β. The estimated slope is $\widehat{\beta} \approx 1$ for the original estimator and $\widehat{\beta} \approx 3/2$ for the improved estimator, which agrees with the theoretical findings. The order of convergence of $|E[P^{(l)} - P^{(l-1)}]|$ is $\widehat{\alpha} \approx 1$, which also agrees with the previous results. As can be observed in Figure 3.3.C, the number of samples ranges from 150 millions for $k = 1$ to 34000 for $k = 7$. The improved estimator gives further reductions in computational time: the total number of samples ranges now from 35 millions for $k = 1$ to only 350 for $k = 7$. The standard deviation of G_k is an increasing function of k, and is less than or equal to the chosen upper bound $\gamma = 4 \times 10^{-6}$.

3.7 Conclusions

The main focus of this chapter was the construction of an efficient simulation algorithm for functionals of a large number of exchangeable random variables. For a specific set-up, we were able to demonstrate optimal complexity order by theoretical analysis and numerical illustrations.

The results from the previous section show that the computational savings can be significant in situations of practical relevance. As seen from Figure 3.3.C, already for $N = 125$ (i.e., $k = 3$), the size of a CDO basket, the required number n_3 of samples on this level is reduced by about two orders of magnitude compared to the number of samples for $k = 1$, n_1. It is roughly this number which would be required for a standard (i.e., single level) estimator on level 3 for a comparable variance achieved by the multilevel estimator at substantially lower cost.

We would expect there to be scope to apply the presented nested simulation approach

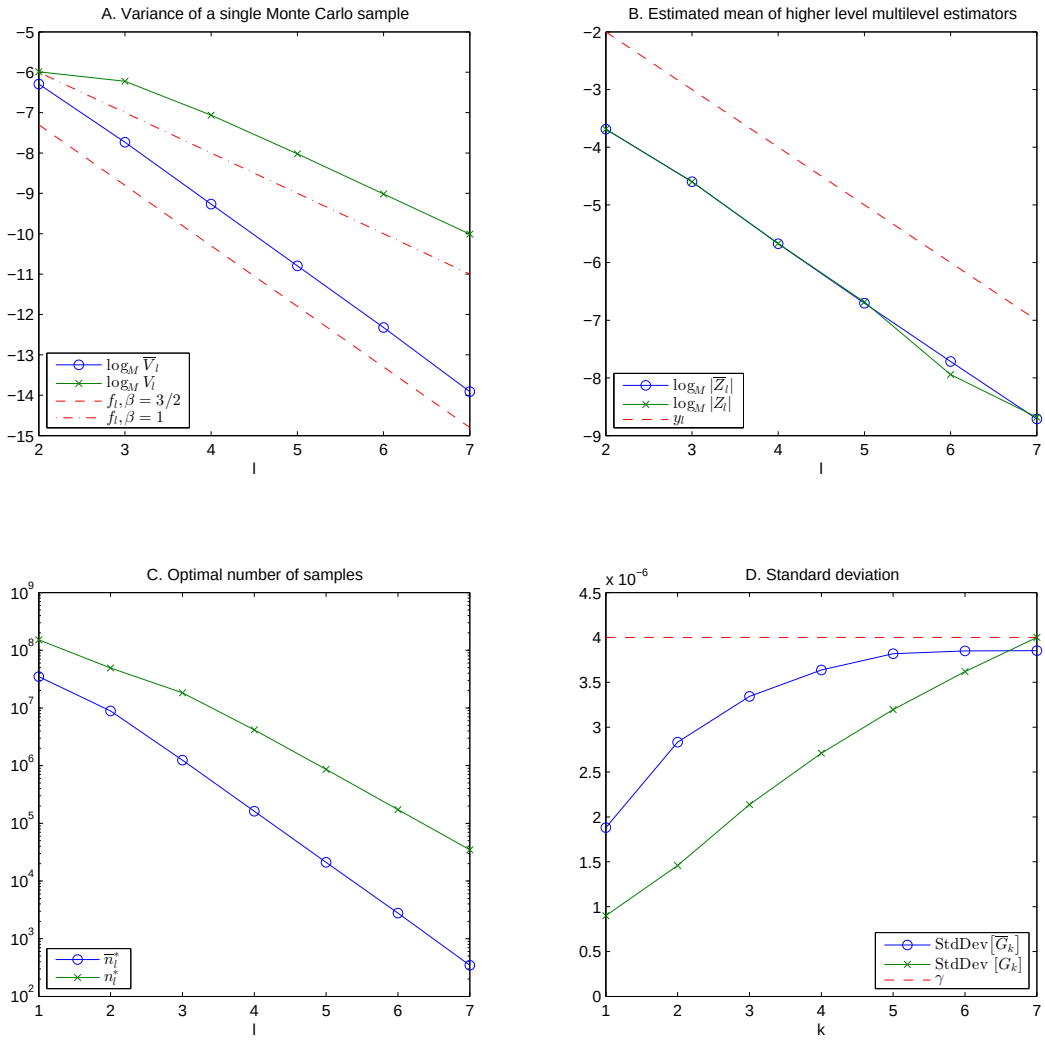

Figure 3.3: Multilevel results for the expected loss in the equity tranche of a CDO basket consisting of N_k companies, $N_k = M^k = 5^k$, $k = 1, \ldots, 7$. Overlined quantities refer to the estimator $\overline{Z}_l$ from (3.42), all others to the standard estimator Z_l from (3.37). A. Variance of a single Monte Carlo sample, V_l and $\overline{V}_l$, together with a predicted trend, f_l, given by (3.50), where $\beta = 1$ or $\beta = 3/2$. B. Mean at level l, Z_l and $\overline{Z}_l$, and a trend, y_l, defined by (3.36), with slope -1. C. Optimal number of simulations in both cases, n_l^* and $\overline{n}_l^*$, calculated according to (3.49) for $k = K = 7$. D. Standard deviation of multilevel estimators G_k defined in (3.13), and similar for $\overline{G}_k$, with their chosen upper bound, γ.

to a wider range of settings beyond the particular application studied here. An interesting extension would be to the model from [37], where the analysis requires further tools, in particular accounting for heterogeneity, resulting in a non-exchangeability. While our motivation comes from credit baskets and some of the later results are specific to piecewise linear functionals encountered in the valuation of basket credit derivatives, there appears to be a wider relevance of the main approach to the simulation of certain functionals arising in large interacting particle systems for example in fields like biology or physics.

Chapter 4

Extension to the Bush et al. approximation

In this chapter, we give a theoretical underpinning of our second numerical framework. The method is an extension to the approximation proposed in [13].

We view the N-name credit model presented in Chapter 2 as a system of particles with an absorbing boundary at zero, monitored discretely. Here, the value of a particle represents a *distance-to-default* of a company, and the absorption of a particle corresponds to the default of that company. The loss of the portfolio is then a function of the empirical measure of the system.

We show that, for the infinite system, the empirical measure has a density with respect to the Lebesgue measure that satisfies a stochastic partial differential equation (SPDE). What is more, between the times that the absorbing boundary is not active, the solution to the resulting SPDE can be represented as the solution to a heat equation shifted by the realisation of the jump-diffusion process in that time interval. Also, at the times that the absorbing boundary is monitored, the density is given as the solution of the SPDE obtained between the monitoring times for values of the particles in $(0, \infty)$, that is the portion of the density below the boundary is truncated. By recursion, we obtain explicit

formulas for the density at all of the times that we consider.

4.1 Introduction

Representing a multi-name credit model as a system of particles was first introduced in [13], where the *distance-to-default* follows a diffusion process with a continuously monitored barrier. The method was inspired by [56], where, in a purely mathematical setting, each particle follows a diffusion process, there is no absorbing barrier but a non-linear interaction between particles is assumed.

In both articles, the initial state of particles is given by a family of exchangeable random variables. In the recent work [58], the same system as in [13] is considered, but with a slightly stronger assumption that the initial state of particles is given by independent and identically distributed random variables. Also, in this framework, the system driven by general Lévy processes is analysed. It is shown that the limit empirical measure exists for such a system and heuristics concerning regularity of the limit empirical measure are given. What is more, earlier, heuristics concerning systems driven by Lévy processes were given in [48].

In this chapter, we adopt the presentation in [58], though we assume only the exchangeability of the initial state of the particles.

Overview of the chapter

In Section 4.2, we introduce the setting. In Section 4.3, we present classical results for infinite exchangeable sequences, and then in Section 4.4, we discuss their implications for our particle system. Throughout Sections 4.5 - 4.8, we gradually derive our main results. Finally, in Section 4.9, we conclude.

4.2 Setup

Let $(\Omega^N, \mathscr{F}^N, (\mathscr{F}_t^N)_{0\le t\le T}, \mathbb{P}^N)$ be a filtered probability space of a system consisting of N particles, where the filtration $\{\mathscr{F}_t^N\}$ satisfies the usual assumptions and the time horizon T is a deterministic constant. We assume that the value of the i-th particle, where $i = 1, \ldots, N$, represents the *distance-to-default* of the i-th company, X^i, defined as

$$X_t^i = \frac{1}{\sigma}(\ln(A_t^i) - \ln(c^B)),$$

where A^i is the i-th company's asset value, c^B is a constant barrier and σ is the variance parameter in the process underlying A^i given by (2.10). Each particle takes values in the state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ denotes the σ-field of Borel subsets of $\mathbb{R}$. We assume further that under the measure $\mathbb{P}^N$, the i-th particle follows a jump-diffusion process with an absorbing boundary being monitored discretely

$$\begin{aligned} X_t^i &= X_0^i + \beta t + \sqrt{1-\rho}\, W_t^i + \sqrt{\rho}\, W_t^\varsigma + \tilde{J}_t,\ t < \tau_i^D, \\ \tilde{J}_t &= \Pi_1 + \Pi_2 + \ldots + \Pi_{\tilde{N}_t} \\ X_t^i &= 0,\ t \ge \tau_i^D, \\ \tau_i^D &= \min\{t \in \{T_1, \ldots, T_m\} : X_t^i \le 0\}, \end{aligned} \tag{4.1}$$

where $T_1 > 0$, $T_m = T$, $\rho \in [0, 1)$, $W^N = \{(W_t^i, i = 1, \ldots, N), \mathscr{F}_t^N, 0 \le t \le T\}$, $W^\varsigma = \{W_t^\varsigma, \mathscr{F}_t^N, 0 \le t \le T\}$ are standard Brownian motions adapted to filtration $\{\mathscr{F}_t^N\}$ and $\tilde{N} = \{\tilde{N}_t, \mathscr{F}_t^N, 0 \le t \le T\}$ is a Poisson process with intensity $\lambda > 0$. We assume that jump sizes, Π_k, $k \in \mathbb{N}$, are i.i.d. random variables with common law ν^Π, which means that $\tilde{J}_t$ is a compound Poisson process. For simplicity, we assume that $\nu^\Pi = N(\mu_\Pi, \sigma_\Pi^2)$, that is each Π_k is Normally distributed with mean μ_Π and variance σ_Π^2. We assume that $(X_0^i, i = 1, \ldots, N)$ is a sequence of exchangeable random variables taking values in $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, where $\mathbb{R}_+ = (0, \infty)$, such that the sequence can be extended to an infinite

exchangeable sequence. In the system, for all i, we assume that X_0^i, W^i, W^ς, $\tilde{N}$, Π are mutually independent.

Let us now introduce the *empirical measure*. For any $S \in \mathcal{B}(\mathbb{R})$, $0 \leq t \leq T$, $\omega^N \in \Omega^N$, the empirical measure of the system (4.1) is defined to be

$$\nu_{t,N}(S, \omega^N) := \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i(\omega^N) \in S\}}. \tag{4.2}$$

Observe that $\nu_{t,N}$ is a *measure-valued stochastic process*, such that for fixed $S \in \mathcal{B}(\mathbb{R})$, $\nu_{\cdot,N}(S, \cdot)$ is a real-valued stochastic process, whereas for fixed $0 \leq t \leq T$ and $\omega^N \in \Omega^N$, $\nu_{t,N}(\cdot, \omega^N)$ is a sub-probability measure on $\mathcal{B}(\mathbb{R})$,

By construction, the system (4.1) is extendible to an infinite system. Let the probability space for the infinite system be $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0 \leq t \leq T}, \mathbb{P})$, where $(X_0^i, i = 1, 2, \ldots)$ is an infinite exchangeable sequence and $W = \{(W_t^i, i = 1, 2, \ldots), \mathscr{F}_t, 0 \leq t \leq T\}$ is an infinite dimensional standard Brownian motion.

For any $S \in \mathcal{B}(\mathbb{R})$, $t \in [0, T]$, $\omega \in \Omega$, we are interested in the *limit empirical measure*

$$\tilde{\nu}_t(S, \omega) \quad := \quad \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i(\omega) \in S\}}.$$

Observe first that X_t^i is a stopped process absorbed at zero. Hence, the dynamics of the i-th particle can be written as

$$X_t^i = \tilde{X}_t^i \, 1_{\{t < \tau_i^D\}}, \tag{4.3}$$

$$\tilde{X}_t^i = X_0^i + \beta t + \sqrt{1 - \rho} \, W_t^i + \sqrt{\rho} \, W_t^\varsigma + \tilde{J}_t. \tag{4.4}$$

Also, note that $\tilde{X}_t^i$ can be divided into the following building blocks

$$\begin{aligned} \tilde{X}_t^i &= X_0^i + B_t^{i,0} + Z_t \\ B_t^{i,0} &= \sqrt{1-\rho}\, W_t^i \\ Z_t &= Z_t^C + \tilde{J}_t \\ Z_t^C &= \beta t + \sqrt{\rho}\, W_t^S, \end{aligned} \tag{4.5}$$

where, for each i, $B^{i,0}$ is a Brownian motion starting at zero with a variance coefficient $\sqrt{1-\rho}$, and Z_t consists of a continuous part Z^C and discontinuous $\tilde{J}$, where Z^C is a Brownian motion starting at zero with drift β and variance coefficient $\sqrt{\rho}$ and $\tilde{J}$ is compound Poisson process.

The aim of this chapter is to give a more explicit definition of the limit empirical measure, then show that the limit measure has a density with respect to the Lebesgue measure and that the density satisfies a stochastic partial differential equation (SPDE). Also, we wish to give the solution to this SPDE and state properties of the solution. We begin with an analysis of the empirical measure at time $t = 0$, where we need to put additional assumptions to the initial distribution of $(X_0^i, i = 1, 2, \ldots)$. Then, in order to obtain our main result, we first analyse the empirical measures of systems that are the building blocks of (4.1) given by (4.5).

4.3 Exchangeability

First, we give a definition of an exchangeable sequence of random variables, then present classical results for infinite exchangeable sequences and discuss their implications. The following presentation is motivated by [2] and [50].

Definition 4.3.1 (Exchangeability, [2], p.6, [50], p. 24)**.** *A finite or infinite sequence* $\xi = (\xi^1, \ldots, \xi^N)$*,* $N \geq 1$*, of random variables in a measurable space* (S, S) *is exchangeable*

if

$$(\xi^1, \ldots, \xi^N) \stackrel{\mathcal{D}}{=} (\xi^{\pi(1)}, \ldots, \xi^{\pi(N)}) \tag{4.6}$$

for each finite permutation π *of* $\{1, \ldots, N\}$.

According to this definition, the joint distribution of an exchangeable sequence is invariant under finite permutations or simply does not change if the random variables are put in any order. Hence, the marginal distribution for each ξ is the same. Obviously, i.i.d. sequences are exchangeable.

Before stating the classical results, we need to introduce a *random measure*. Let us assume first that the infinite exchangeable sequence $\xi = (\xi^1, \xi^2, \ldots)$ is defined on probability space $(\Omega, \mathscr{F}, \mathbb{P})$. A random probability measure α is a function $\alpha : \Omega \times \mathsf{S} \mapsto [0, 1]$ such that, for fixed $\omega \in \Omega$, $\alpha(\omega, \cdot)$ is a probability measure on (S, S), whereas for each $A \in \mathsf{S}$, $\alpha(\cdot, A)$ is a random variable in $\mathscr{P}(S)$, where $\mathscr{P}(S)$ is a set of probability measures on S and its σ-field is generated by all maps $\theta \mapsto \theta(A)$, $A \in \mathsf{S}$.

Definition 4.3.2 (Conditional i.i.d. for infinite sequence, [50], [70])**.** *A random infinite sequence* ξ *in* S *is conditionally i.i.d. if*

$$P(\xi \in \cdot \mid \mathscr{G}) = \alpha^\infty \ \ a.s., \tag{4.7}$$

where $\mathscr{G}$ *is sub-*σ*-field of* $\mathscr{F}$*,* α *is a random probability measure on* S*,* $\alpha^\infty = \alpha \times \alpha \ldots$ *is random probability measure on* S^∞*, such that* S^∞ *is a countable infinite product of* S*, and* $\alpha \times \alpha \ldots$ *is a unique extension of the Kolmogorov product measures to* $\mathscr{F}$*.*

Observe that, for each $N \geq 1$ and $i \in \mathbb{N}$, (4.7) can be written as

$$P(\xi^i \in A_i, 1 \leq i \leq N \mid \mathscr{G})(\omega) = \prod_{i=1}^{N} \alpha(\omega, A_i), \ \mathbb{P}\text{-a.e. } \omega \in \Omega, \ \forall A_i \in \mathsf{S}, \tag{4.8}$$

Hence, given $\mathscr{G}$, for $\mathbb{P}$-a.e. $\omega \in \Omega$ and all $A_i \in \mathsf{S}$, α assigns probability $\prod_{i=1}^{N} \alpha(\omega, A_i)$ to all vectors of length N in the infinite sequence ξ.

Theorem 4.3.1 (Infinite exchangeable sequences, de Finetti, [50] p.25)**.** *Let ξ be an infinite exchangeable sequence of random elements in a measurable space S. Then the following conditions are equivalent when S is Borel*

(i) ξ is exchangeable,

(ii) ξ is conditionally i.i.d.

For general S, we have (i) $\Leftarrow$ (ii)*.*

By this Theorem and Definition 4.3.2, we can say that each infinite exchangeable sequence, taking values in a Borel space S, is conditionally independent with some common random measure α, whereas for a general space S, conditional i.i.d. implies exchangeability but the inverse is not true.

Theorem 4.3.2 (Uniqueness, [50], p. 28, [2], p.28)**.** *Let ξ be infinite exchangeable sequence in a Borel space (S, S), such that*

$$P(\xi \in \cdot \mid \mathscr{G}) = \alpha^{\infty} \text{ a.s.}, \tag{4.9}$$

for some σ-field $\mathscr{G}$ and some random probability measure α on S, then

(i) α is a.s. unique, ξ-measurable, and given by

$$\alpha(B) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{\xi^i \in B\}} \text{ a.s., } B \in \mathsf{S}; \tag{4.10}$$

(ii) $\mathscr{G}$ and ξ are conditionally independent, $\mathscr{G} \subset \sigma(\xi)$, implies $\mathscr{G} = \sigma(\alpha)$ a.s.,

(iii) the distribution of ξ has the following representation, iff P_α is a distribution of α,

$$P(\xi \in \cdot) = \int_{\mathscr{P}(S)} \theta^{\infty} \, P_\alpha(\alpha \in d\theta), \tag{4.11}$$

(iv) if ξ is an i.i.d. sequence then $\alpha = \theta$ a.s. for some fixed distribution θ.

Let us now discuss the results given by Theorem 4.3.2. By (i), we have

$$\alpha(\omega, B) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{\xi^i(\omega)\in B\}} \ \mathbb{P}\text{-a.e. } \omega \in \Omega,\ B \in \mathsf{S},$$

meaning that for $\mathbb{P}$-a.e. $\omega \in \Omega$, we obtain a realisation of ξ given by $\xi(\omega)$, which uniquely determines α as a limit empirical measure of $\xi(\omega)$. Since α is unique, it is called the *directing random measure* of ξ.

By (ii), (4.9) can be written as

$$P(\xi \in \cdot \mid \sigma(\alpha)) = \alpha^\infty \ a.s., \tag{4.12}$$

which, as noted in [2], also implies that $P(\xi \in \cdot \mid \alpha) = \alpha^\infty$ *a.s.*.

Result (iii) of the Theorem provides a representation of an infinite exchangeable sequence as a mixture of i.i.d. sequences, where α is a random probability measure in a set of probability measures $\mathscr{P}(S)$, becoming a distribution θ according to $\mathbb{P}_\alpha$. As noted in [50], $P(\xi \in \cdot) = \mathbb{E}[\alpha^\infty]$, hence α is 'treated' as any other random variable; the difference is its state space, which is $\mathscr{P}(S)$, endowed with the aforementioned σ-field.

What is more, as noted in [2], the abstract representation given by (4.11) can be simplified if α belongs to a parametric family of distributions with some random parameter. Here, we give an example of an infinite exchangeable sequence with such α. Let $\xi^i = \Psi + U^i$, where $i = 1, 2, \ldots$, each U^i is a standard Normal random variable, and Ψ is some random variable taking values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with some distribution P_Ψ. This is an infinite exchangeable sequence taking values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Conditional on Ψ, ξ are i.i.d.. For all i, $\alpha = P(\xi^i \in A_i \mid \Psi) = \mu_{\Psi,1}(A_i)$, for any $A_i \in \mathcal{B}(\mathbb{R})$, where $\mu_{\Psi,1}$ is a Normal distribution

with a mean Ψ and variance 1. For any $N \geq 1$, the joint distribution is given as

$$P(\xi^i \in A_i, 1 \leq i \leq N) = \int_{\mathbb{R}} \prod_{i=1}^{N} \mu_{\psi,1}(A_i)\ P_\Psi(\Psi \in d\psi). \tag{4.13}$$

Similarly, for any fixed time $t > 0$, $(\tilde{X}_t^i, i = 1, 2, \ldots)$, given by (4.4), is also an infinite exchangeable sequence, where the joint distribution has similar form to (4.13).

By (iv) of Theorem 4.3.2, if ξ is an i.i.d. sequence, then α is non-random and for $\mathbb{P}$-a.e. $\omega \in \Omega$ it is equal to some fixed distribution θ. Similarly, P_α can be degenerate at some fixed θ, i.e., $P_\alpha = \delta_\theta$, where δ is a Dirac measure centred at θ.

4.4 Limit empirical measure at the initial time

In this section, we discuss the implications for our particle system (4.1) arising from the fact that at the initial time particles form an infinite exchangeable sequence. We also add additional assumptions to particles at the initial time.

Let us first divide the probability space of all particles, $(\Omega, \mathscr{F}, \mathbb{P})$, into the probability space at the initial time and afterwards

$$(\Omega_0 \times \Omega_{0,T}, \mathscr{F}_0 \times \mathscr{F}_{0,T}, \mathbb{P}_0 \times \mathbb{P}_{0,T}),$$

such that $\mathbb{X}_0 = (X_0^i, i = 1, 2, \ldots)$ is defined on $(\Omega_0, \mathscr{F}_0, \mathbb{P}_0)$, where a similar division of probability space is given in [71] for the d-dimmensional diffusion process, $d \in \mathbb{N}$. Such a representation is allowed since, as we assume in Section 4.2, for all i, X_0^i, W^N, W^ς, $\tilde{N}$, Π are mutually independent.

Let us now fix $\omega_0 \in \Omega_0$, meaning that a realisation of the infinite sequence X_0 is known and given by $X_0(\omega_0)$. Then, by (i) in Theorem 4.3.2, we obtain a unique limit empirical

measure for almost all of the chosen ω_0

$$\alpha(\omega_0, A) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{X_0^i(\omega_0)\in A},\ \forall A \in \mathcal{B}(\mathbb{R}),$$

where $\alpha(\omega_0, A)$ is in $[0, 1]$. Additionally, by (ii) in Theorem 4.3.2, the conditional distribution for each particle given α and ω_0 equals

$$P(X_0^i \in A \mid \alpha)(\omega_0) = \alpha(\omega_0, A), \forall A \in \mathcal{B}(\mathbb{R}).$$

Let us denote $\alpha(\omega_0)$ by ν_0. Then, the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ can be written as

$$(\mathbb{R}_+^\infty \times \Omega_{0,T}, \mathscr{F}_0 \times \mathscr{F}_{0,T}, \nu_0^\infty \times \mathbb{P}_{0,T}). \tag{4.14}$$

By (iv) in Theorem 4.3.2, if X_0 is an i.i.d. sequence, then for almost all ω_0, we would obtain the same ν_0. However, since X_0 is exchangeable for each ω_0, we obtain different ν_0.

Through the chapter, we assume that ω_0 is fixed and that $(\Omega, \mathscr{F}, \mathbb{P})$ has the representation (4.14) for some given ν_0. Additionally, we assume that the measure ν_0 has a density v_0 with respect to the Lebesgue measure. We summarise the assumptions for particles at initial time as

$$P[X_0^i \in A] = \nu_0(A), \tag{4.15}$$

$$\nu_0(A) = \int_{A\cap\mathbb{R}_+} v_0(x)\, dx, \tag{4.16}$$

where $i = 1, 2, \ldots$ and $A \in \mathcal{B}(\mathbb{R})$.

4.5 System of Brownian motions with identical initial distribution

First, we consider the easiest system from the building blocks of (4.1), then define the limit empirical measure of the particle system and give the properties of the measure.

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ be some probability space. Consider an infinite system of particles defined in this space, where the i-th particle satisfies

$$X_t^i = X_0^i + B_t^{i,0}. \tag{4.17}$$

This is a special case of system (4.5), where for all $0 \leq t \leq T$, $Z_t \equiv 0$. Using assumption (4.16), we can write (4.17) as

$$\begin{aligned} X_t^i &= B_t^i, \quad t > 0 \\ \mathbb{P}[B_0^i \in S] &= \nu_0(S), \quad \forall S \in \mathcal{B}(\mathbb{R}), \end{aligned} \tag{4.18}$$

where each B^i is a Brownian motion with a variance coefficient $\sqrt{1-\rho}$ and initial distribution ν_0. Hence, we can write

$$\mathbb{P}(X_t^i \in S) = \mathbb{P}^{i,\nu_0}(B_t^i \in S),$$

where $\mathbb{P}^{i,\nu_0}$ is a probability measure for B^i.

Since the particles are i.i.d., $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ has the following representation

$$(\Omega^{B,1} \times \Omega^{B,2} \times \ldots, \mathscr{F}^{B,1} \times \mathscr{F}^{B,2} \times \ldots, \mathbb{P}^{1,\nu_0} \times \mathbb{P}^{2,\nu_0} \times \ldots)$$

such that $(\Omega^{B,i}, \mathscr{F}^{B,i}, \mathbb{P}^{i,\nu_0})$ is a probability space of each B_t^i, and the representation is motivated by the construction of d-dimension Brownian motion with some initial distribution given in [51]. For ease of notation, we drop subscript i, meaning that $X_t = B_t$ is

any tagged particle from the system (4.17).

By a classical result (see, for example, [51]),

$$\mathbb{P}^{\nu_0}(B_t \in S) = \int_S \int_0^\infty p(t,x,y)\, v_0(y)\, dy\, dx, \tag{4.19}$$

$$p(t,x,y) := \frac{1}{\sqrt{2\pi(1-\rho)t}}\, e^{-\frac{(x-y)^2}{2(1-\rho)t}}. \tag{4.20}$$

Note that $p(t,x,y)$ is a *Gaussian kernel* or *transition density* of Brownian motion with a variance coefficient $\sqrt{1-\rho}$ starting at value x and arriving at y at time t, where $x, y \in \mathbb{R}$, $t > 0$, i.e. as in [51], $p(t,x,y)$ is defined as

$$p(t,x,y) = \frac{1}{dy}\mathbb{P}^{B,x}[B_t \in dy], \tag{4.21}$$

where $\{\mathbb{P}^{B,x}\}_{x\in\mathbb{R}}$ is a family of probability measures of a Brownian motion with a variance coefficient $\sqrt{1-\rho}$, B, defined on $(\Omega^B, \mathscr{F}^B)$ with a starting value x.

Since, for the infinite system, the particles are i.i.d., by the Strong Law of Large Numbers, the empirical measure of the system is given by the probability that a single particle is in a set $S \in \mathcal{B}(\mathbb{R})$.

Lemma 4.5.1 (Strong Law of Large Numbers for the system)**.** *Let $t \in (0,T]$, $S \in \mathcal{B}(\mathbb{R})$ and $(X_t^i, i = 1,2,\ldots)$ be given by* (4.18)*, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i \in S\}} = \mathbb{P}^{\nu_0}(B_t \in S),\ a.s.$$

Let us now define the empirical measure of the infinite system using Lemma 4.5.1.

Definition 4.5.1. *The empirical measure of the system* (4.18)*, ν_t^*, is defined as*

$$\nu_t^*(S) := \mathbb{P}^{\nu_0}(B_t \in S), \quad \forall S \in \mathcal{B}(\mathbb{R}),\ t \in (0,T].$$

Also, by (4.19) and (4.20), we have the following representation of ν^*.

Lemma 4.5.2. *Let $t \in (0,T]$, $S \in \mathcal{B}(\mathbb{R})$, then the empirical measure of the system* (4.18)*, ν_t^*, can be written as*

$$\begin{aligned} \nu_t^*(S) &= \int_S u(t,x)\, dx, \\ u(t,x) &:= \int_0^\infty p(t,x,y)\, v_0(y)\, dy,\ x \in \mathbb{R}, \end{aligned} \tag{4.22}$$

where u is density of ν_t^* with respect to the Lebesgue measure, and $p(t,x,y)$ is given by (4.20)*.*

By using the theorem below, which is a classical result for the heat equation, we can demonstrate several properties of u.

Theorem 4.5.1 (Solution to the initial value problem of the heat equation, [27] p.46, [75] p.81)**.** *Let g be a bounded and piecewise continuous function on $\mathbb{R}$ and let s be given by*

$$s(t,x) = \int_{\mathbb{R}} p(t,x,y)\, g(y)\, dy,\ x \in \mathbb{R},$$

where $p(t,x,y)$ equals (4.20)*, then, for all $T > 0$, s satisfies*

(i) $s \in C^\infty(\mathbb{R} \times (0,T])$*,*

(ii) $s_t = \frac{1}{2}(1-\rho)s_{xx}$*, $x \in \mathbb{R}$, $t > 0$,*

(iii) $\lim_{(t,x)\to(0,x_0)} s(t,x) = g(x_0)$*, for each continuity point x_0 of g,*

(iv) $\lim_{(t,x)\to(0,x_0)} s(t,x) = \frac{1}{2}\left[v_0(x_0^-) + v_0(x_0^+)\right]$*, for each discontinuity point x_0 of g, where $x_0^- = \lim_{x\uparrow x_0}$, $x_0^+ = \lim_{x\downarrow x_0}$.*

In order to obtain the above properties for u, we make additional assumptions:

(A1) v_0 is bounded and continuous in $\mathbb{R}_+$,

(A2) $v_0 \equiv 0$ for $x \le 0$,

meaning that v_0 is a piecewise continuous function with a discontinuity at $x = 0$. Then, by a simple application of Theorem 4.5.1, we obtain the following lemma.

Lemma 4.5.3. *Let Assumptions* $(A1)$ *and* $(A2)$ *hold, and let* u *be given by* (4.22)*, then*

(i) $u \in C^\infty(\mathbb{R} \times (0, T])$*,*

(ii) $u_t = \frac{1}{2}(1 - \rho)u_{xx}$*,* $x \in \mathbb{R}$*,* $t > 0$*,*

(iii) $\lim_{(t,x)\to(0,x_0)} u(t, x)$ *is given as*

$$\lim_{(t,x)\to(0,x_0)} u(t, x) = \begin{cases} v_0(x), & x > 0, \\ 0, & x < 0. \end{cases} \tag{4.23}$$

Observe also that u can be represented as an expectation under measure $\mathbb{P}^{B,x}$ of the initial condition v_0 calculated at B_t, $t > 0$, $x \in \mathbb{R}$. This is a simple consequence of the definition of u given in (4.22) and p in (4.21).

Lemma 4.5.4 (Probabilistic representation of the solution to heat equation, [51], p.254)**.** *Let* u *be defined as above, then*

$$u(t, x) = \mathbb{E}^{B,x}[v_0(B_t)] = \mathbb{E}^{B,0}[v_0(B_t + x)].$$

4.6 System of Brownian motions with a Brownian common factor

In this section, we adapt the approach given in [58] to our setting. The representation of the probability space as a product of idiosyncratic and common factor spaces, which we use throughout the chapter, follows from [58]. Also, the definition of the limit empirical measure is inspired by [58].

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ be some probability space. Consider an infinite system of particles defined on this space, where the i-th particle satisfies

$$X_t^i = B_t^i + Z_t^C, \tag{4.24}$$

where B^i is given as in (4.18), whereas Z^C is as in (4.5), meaning that Z^C is a Brownian motion with drift β and dispersion coefficient $\sqrt{\rho}$. This is a special case of system (4.5), where for all $0 \leq t \leq T$, $J_t \equiv 0$.

Observe first that, conditional on Z^C, $(X^i, i = 1, 2, \ldots)$ are i.i.d.. Hence, the probability space $(\Omega, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ can be represented as a product of two spaces: a space of the infinite Brownian motion with common initial distribution, $(B^i, i = 1, 2, \ldots)$, and a space of the Brownian motion Z^C

$$(\Omega^{\mathbb{B}} \times \Omega^C, \mathscr{F}^{\mathbb{B}} \times \mathscr{F}^C, \mathbb{P}^{\mathbb{B}} \times \mathbb{P}^C),$$

where $(\Omega^{\mathbb{B}}, \mathscr{F}^{\mathbb{B}}, \mathbb{P}^{\mathbb{B}})$ is defined as in Section 4.5.

Note now that for fixed $\omega = (\omega^{\mathbb{B}}, \omega^C)$, $t > 0$ and $A \in \mathcal{B}(\mathbb{R})$, the empirical measure of the system (4.24) containing N particles, $\bar{\nu}_{t,N}$, defined as in (4.2), can be written as

$$\bar{\nu}_{t,N}(A)(\omega^{\mathbb{B}}, \omega^C) = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i(\omega^{\mathbb{B}}, \omega^C) \in A\}},$$

and takes a value in $[0, 1]$. Whereas, when we fix only $\omega^C \in \Omega^C$, for $t > 0$ and $A \in \mathcal{B}(\mathbb{R})$, the empirical measure takes the form

$$\bar{\nu}_{t,N}(A)(\omega^C) = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i(\omega^C) \in A\}},$$

and is a $\mathscr{F}^{\mathbb{B}}$-measurable random variable. Also, since $X_t^i(\omega^C) = B_t^i + Z_t^C(\omega^C)$, we have

$$\bar{\nu}_{t,N}(A)(\omega^C) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{B_t^i \in A - Z_t^C\}},$$

where $Z_t^C = Z_t^C(\omega^C)$, as we shall write throughout this section, and for $A \in \mathcal{B}(\mathbb{R})$, $A - Z_t^C = \{a - Z_t^C(\omega^C) : a \in A\}$.

Let us now define the empirical measure of the infinite system (4.24). Since, for the infinite system, the particles are conditionally i.i.d., by a conditional version of the Strong Law of Large Numbers, the empirical measure of the system is given by the probability that a single particle is in a set $S \in \mathcal{B}(\mathbb{R})$, for any $t \in (0,T]$ conditional on Z_t^C. Also, let us note that we shall use 'for almost all $\omega^C \in \Omega^C$', '$\mathbb{P}^C$ – a.s.' and '$\mathbb{P}^C$ – a.e. $\omega^C \in \Omega^C$' interchangeably.

Lemma 4.6.1 (Conditional Strong Law of Large Numbers for the system)**.** *Let $t \in (0,T]$, $S \in \mathcal{B}(\mathbb{R})$ and $(X_t^i, i = 1,2,\ldots)$ be given by* (4.24)*, then for almost all $\omega^C \in \Omega^C$, we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} 1_{\{X_t^i(\omega^C)\in S\}} = \mathbb{P}^{\nu_0}(B_t \in S - Z_t^C),\ \mathbb{P}^{\mathbb{B}}\text{-a.s.}$$

Definition 4.6.1. *The empirical measure of the system* (4.24)*, $\bar{\nu}_t$, is defined for almost all $\omega^C \in \Omega^C$, as*

$$\bar{\nu}_t(S) = \bar{\nu}_t(S)(\omega^C) := \mathbb{P}^{\nu_0}(B_t \in S - Z_t^C) \quad \forall S \in \mathcal{B}(\mathbb{R}),\ t \in (0,T].$$

Lemma 4.6.2. *Let $t \in (0,T]$, $S \in \mathcal{B}(\mathbb{R})$, then the empirical measure of the system* (4.24)*, $\bar{\nu}_t$, can be written as*

$$\bar{\nu}_t(S) = \int_S \bar{v}(t,x)\,dx, \tag{4.25}$$

$$\bar{v}(t,x) := \int_0^\infty p(t, x - Z_t^C, y)\, v_0(y)\,dy,\ x \in \mathbb{R}, \tag{4.26}$$

where $\bar{v}(t,x)$ is a density of $\bar{\nu}_t$ with respect to the Lebesgue measure and $p(t,x,y)$ is given by (4.20).

Proposition 4.6.1. *Let $\bar{\nu}_t$ and $\bar{v}$ be defined as above, then*

$$\bar{v}(t,x) = u(t, x - Z_t^C),\ \mathbb{P}^C\text{-a.s.} \tag{4.27}$$

where u is defined in (4.22) *and satisfies the heat equation, for $t > 0$, $x \in \mathbb{R}$,*

$$u_t(t,x) \;=\; \frac{1}{2}(1-\rho)u_{xx}(t,x), \tag{4.28}$$

with initial condition v_0. Additionally, $\bar{v}$ satisfies a stochastic partial differential equation

$$d\bar{v}(t,x) = -\beta\ \bar{v}_x(t,x)\ dt + \frac{1}{2}\bar{v}_{xx}(t,x)\ dt - \sqrt{\rho}\ \bar{v}_x(t,x)\ dW_t^S, \quad \mathbb{P}^C\text{-a.s.} \tag{4.29}$$

Proof. We obtain (4.27) by comparison of definition of u given by (4.22) and of $\bar{v}$ given in (4.26). Since, u is a deterministic function and, by Theorem 4.5.1, $u \in C^\infty(\mathbb{R} \times (0,T])$, we can apply Itô lemma to $u(t, x - Z_t^C)$ and obtain $\mathbb{P}^C$-a.s.

$$\begin{aligned} du(t, x - Z_t^C) \;&=\; u_t(t, x - Z_t^C)\ dt - \beta\ u_x(t, x - Z_t^C)\ dt - \sqrt{\rho} u_x(t, x - Z_t^C)\ dW_t^S \\ &+\; \frac{1}{2}\rho\ u_{xx}(t, x - Z_t^C)\ dt \\ &=\; -\beta\ u_x(t, x - Z_t^C)\ dt - \sqrt{\rho} u_x(t, x - Z_t^C)\ dW_t^S + \rho\ u_{xx}(t, x - Z_t^C)\ dt, \end{aligned}$$

where in the second equality we use (4.28). Then, by (4.27), we arrive at (4.29). □

Lemma 4.6.3. *Let $\bar{v}$ be defined as above, and let Assumptions* (A1) *and* (A2) *be satisfied, then $\mathbb{P}^C$-a.s. $\bar{v}(t,\cdot) \in C^\infty(\mathbb{R})$, $\forall t \in (0,T]$.*

Proof. We obtain the result by (4.27) and Lemma 4.5.3, point (i). □

Remark 4.6.1. *Observe that by* (4.27)*, we have* $\mathbb{P}^C$*-a.s.*

$$\bar{v}(t, x + Z_t^C) = u(t, x), \tag{4.30}$$

however unlike u*,* $\bar{v}$ *is a stochastic function. As noted, for example in [72], we cannot apply the Itô's lemma to a stochastic function and instead we need to use the Itô-Wentzell lemma. Hence, by applying the Itô-Wentzell lemma to* $\bar{v}(t, x + Z_t^C)$*, and then using* (4.29) *and* (4.30)*, we obtain* (4.28)*.*

Similar to Lemma 4.5.4, we obtain the probabilistic representation of the solution to (4.29).

Lemma 4.6.4. *Let* $\bar{v}$ *be defined as above, then* $\mathbb{P}^C$*-a.s.*

$$\bar{v}(t, x) = \mathbb{E}^{B, x - Z_t^C}[v_0(B_t)] = \mathbb{E}^{B,x}[v_0(B_t - Z_t^C)].$$

Observe that, expectations are taken under the measure $\mathbb{P}^{B,x}$ for Brownian motion B, conditional on realisation, $Z_t^C(\omega^C)$, of the common Brownian motion that is independent of B. A similar representation can be found for example in [72].

4.7 System of Brownian motions with jump-diffusion common factor

Let us now consider system (4.5), defined on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0 \le t \le T}, \mathbb{P})$ as introduced in Section 4.2. Each particle satisfies

$$\begin{aligned} \tilde{X}_t^i &= B_t^i + Z_t, \\ Z_t &= Z_t^C + \tilde{J}_t, \end{aligned} \tag{4.31}$$

where $i = 1, 2, \ldots, B$, Z^C are defined as in Section 4.6, and $\tilde{J}$ is a compound Poisson process, introduced in Section 4.2. Observe that $Z = \{Z_t, \mathscr{F}_t, 0 \leq t \leq T\}$ is a jump-diffusion process and, in particular, is right-continuous with left limits.

Similarly, as in Section 4.6, we divide $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ into a space of i.i.d. Brownian motions, $\mathbb{B}$, and space of the common factor, Z:

$$(\Omega^{\mathbb{B}} \times \Omega^Z, \mathscr{F}^{\mathbb{B}} \times \mathscr{F}^Z, \mathbb{P}^{\mathbb{B}} \times \mathbb{P}^Z).$$

Let us now give the definition of the empirical measure of the system at any time $t > 0$ and at time t^-, i.e. just before a jump of Z, of size ΔZ, where

$$\begin{aligned} \Delta Z_t &= Z_t - Z_{t^-} \\ \tilde{X}_{t^-} &= \lim_{s\uparrow t} \tilde{X}_s. \end{aligned} \tag{4.32}$$

Here we state an analogous version of Lemma 4.6.1. Note that we use the time continuity of Brownian motion.

Lemma 4.7.1 (Conditional Strong Law of Large Numbers for the system)**.** *Let $t \in (0, T]$, $S \in \mathcal{B}(\mathbb{R})$ and $(\tilde{X}_t^i, i = 1, 2, \ldots)$ be given by* (4.31)*, then, for almost all $\omega^Z \in \Omega^Z$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{\tilde{X}_t^i(\omega^Z)\in S\}} = \mathbb{P}^{\nu_0}(B_t \in S - Z_t), \quad \mathbb{P}^{\mathbb{B}}\text{-a.s.}$$

Similarly, for pre-jump times we have, for almost all $\omega^Z \in \Omega^Z$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{\tilde{X}_{t^-}^i(\omega^Z)\in S\}} = \mathbb{P}^{\nu_0}(B_t \in S - Z_{t^-}), \quad \mathbb{P}^{\mathbb{B}}\text{-a.s.}$$

Definition 4.7.1. *The empirical measure of the system* (4.31)*, $\tilde{\nu}_t$, is defined for almost*

all $\omega^Z \in \Omega^Z$ *as*

$$\tilde{\nu}_t(S) = \tilde{\nu}_t(S)(\omega^Z) := \mathbb{P}^{\nu_0}(B_t \in S - Z_t), \quad \forall S \in \mathcal{B}(\mathbb{R}),\ t \in (0,T].$$

Similarly, we define the pre-jump empirical measure of the system (4.31) *for almost all* $\omega^Z \in \Omega^Z$ *as*

$$\tilde{\nu}_{t^-}(S) = \tilde{\nu}_{t^-}(S)(\omega^Z) := \mathbb{P}^{\nu_0}(B_t \in S - Z_{t^-}), \quad \forall S \in \mathcal{B}(\mathbb{R}),\ t \in (0,T].$$

Lemma 4.7.2. *Let* $t \in (0,T]$, $S \in \mathcal{B}(\mathbb{R})$, *then the empirical measure of the system* (4.31), $\tilde{\nu}_t$, *can be written as*

$$\begin{aligned} \tilde{v}_t(S) &= \int_S \tilde{v}(t,x)\, dx, \\ \tilde{v}(t,x) &:= \int_0^\infty p(t, x - Z_t, y)\, v_0(y)\, dy,\ x \in \mathbb{R}, \end{aligned} \tag{4.33}$$

where $\tilde{v}(t,x)$ *is a density of* $\tilde{\nu}_t$ *with respect to the Lebesgue measure and* $p(t,x,y)$ *is given by* (4.20). *Similarly,* $\tilde{\nu}_{t^-}$ *can be written as*

$$\begin{aligned} \tilde{v}_{t^-} &= \int_S \tilde{v}(t^-,x)\, dx, \\ \tilde{v}(t^-,x) &:= \int_0^\infty p(t, x - Z_{t^-}, y)\, v_0(y)\, dy,\ x \in \mathbb{R}. \end{aligned} \tag{4.34}$$

In order to give the properties of $\tilde{v}$, we need to construct Ω^Z. We assume that Ω^Z consists of the space of the geometric Brownian motion Z^C, $(\Omega^C, \mathscr{F}^C, \mathbb{P}^C)$, and the space of the compound Poisson process $\tilde{J}$, $(\Omega^{\tilde{J}}, \mathscr{F}^{\tilde{J}}, \mathbb{P}^{\tilde{J}})$:

$$(\Omega^C \times \Omega^{\tilde{J}}, \mathscr{F}^C \times \mathscr{F}^{\tilde{J}}, \mathbb{P}^C \times \mathbb{P}^{\tilde{J}}).$$

We define $\Omega^{\tilde{J}}$ canonically, where we closely follow the definition given in [74]. We assume

that $\Omega^{\tilde{J}}$ consists of a vector of the arrival times of jumps and of the corresponding vector of jump sizes

$$\Omega^{\tilde{J}} = \cup_{n\geq 0}([0,T]\times S)^n,$$

where $([0,T]\times S)^0 = \{a\}$, a is arbitrary point and $S \in \mathcal{B}(\mathbb{R}-\{0\})$. Note that, n, the number of jumps that we consider, is finite, since this is the case for a Poisson process with constant intensity, as we assume in Section 4.2. Then, the realisation of the compound Poisson process, $\tilde{J}_t(\omega^{\tilde{J}})$, can be written as

$$\tilde{J}_t(\omega^{\tilde{J}}) = \begin{cases} \sum_{k=1}^{n} \Pi_k 1_{\{t_k\in(0,t)\}}, & \omega^{\tilde{J}} = (t_1,\Pi_1),\ldots,(t_n,\Pi_n), \\ 0, & \omega^{\tilde{J}} = \{a\}. \end{cases}$$

If $\omega^{\tilde{J}}$ is given by a single point, then we mean that no jump occurs, otherwise $n>1$ gives the lengths of the vector of the arrival times and the size of the jumps. Then, $\mathbb{P}^{\tilde{J}}$ is a distribution of a compound Poisson process, where the precise construction of $\mathbb{P}^{\tilde{J}}$ can be found in [74].

Let us now introduce the *Poisson random measure* $\tilde{N}(t,S)$ corresponding to $\omega^{\tilde{J}}$. Similar to [3], we define the measure as

$$\tilde{N}(t,S)(\omega^{\tilde{J}}) = \sum_{s=0}^{t} 1_{\{\Delta Z_s(\omega^Z)\in S\}},$$

such that $S \in \mathcal{B}(\mathbb{R}-\{0\})$ and as above, represents a set of possible values of jump sizes.

Proposition 4.7.1. *Let $\tilde{\nu}_t$ and $\tilde{v}$ be defined as above, then $\mathbb{P}^Z$-a.s.*

$$\tilde{v}(t,x) = u(t,x-Z_t), \tag{4.35}$$

$$\tilde{v}(t^-,x) = u(t,x-Z_{t^-}), \tag{4.36}$$

where u is defined in (4.22) *and satisfies the heat equation for $t > 0$, $x \in \mathbb{R}$,*

$$u_t(t,x) \;=\; \frac{1}{2}(1-\rho)u_{xx}(t,x) \tag{4.37}$$

with the initial condition v_0. Additionally, $\tilde{v}$ satisfies a stochastic partial differential equation $\mathbb{P}^Z$-a.s.

$$\begin{aligned} d\tilde{v}(t,x) &= -\beta\, \tilde{v}_x(t^-,x)\, dt + \frac{1}{2}\tilde{v}_{xx}(t^-,x)\, dt - \sqrt{\rho}\, \tilde{v}_x(t^-,x)\, dW_t^S \\ &+ \big(\tilde{v}(t^-,x-\Pi) - \tilde{v}(t^-,x)\big)\; \tilde{N}(dt,d\Pi). \end{aligned} \tag{4.38}$$

Proof. The equalities (4.35) and (4.36) are a simple consequence of the definitions of $u(t,x)$, $\tilde{v}(t,x)$ and $\tilde{v}(t^-,x)$. We obtain (4.38) by application of Itô's lemma for Lévy processes, given for example in [3]. Here, for completeness, we give an outline of the derivation of the lemma for our case. Let us assume now that $\omega^{\tilde{J}} = (t_1,\Pi_1),\ldots,(t_n,\Pi_n)$, such that $t_n < t$. If $\omega^{\tilde{J}} = \{a\}$, we would obtain the results given in Section 4.6, which is why we omit this case. Observe that, we can write $u(t,x-Z_t)$ as

$$\begin{aligned} u(t,x-Z_t) &= \sum_{k=1}^{n+1} \Big(u(t,x-Z_{t_k^-}) - u(t,x-Z_{t_{k-1}})\Big) \\ &+ \sum_{k=1}^{n} \Big(u(t,x-Z_{t_k^-}-\Pi_k) - u(t,x-Z_{t_k^-})\Big) \end{aligned} \tag{4.39}$$

where $t_0 = 0$ and $t_{n+1} = t$. The first sum represents increments of u due to the continuous part of Z, whereas the second sum is, due to the jumps of Z.

By Itô's lemma, similar to Section 4.6, a summand $u(t,x-Z_{t_k^-}) - u(t,x-Z_{t_k})$, can be written as

$$\begin{aligned} u(t,x-Z_{t_k^-}) &= u(t,x-Z_{t_{k-1}}) - \int_{t_{k-1}}^{t_k} \beta\, u_x(s,x-Z_{s^-})\, ds \\ &- \int_{t_{k-1}}^{t_k} \sqrt{\rho}\, u_x(s,x-Z_{s^-})\, dW_s^S + \frac{1}{2}\int_{t_{k-1}}^{t_k} u_{xx}(s,x-Z_{s^-})\, ds. \end{aligned}$$

Applying this to all summands, we can write the first sum in (4.39) as

$$\begin{aligned} u(0,x) &- \int_0^t \beta\, u_x(s, x - Z_{s^-})\, ds - \int_0^t \sqrt{\rho}\, u_x(s, x - Z_{s^-})\, dW_s^S \\ &+ \frac{1}{2}\int_0^t u_{xx}(s, x - Z_{s^-})\, ds. \end{aligned} \tag{4.40}$$

The second sum in (4.39), can be written as

$$\int_0^t \int_A \left(u(s, x - Z_{s^-} - \Pi) - u(s, x - Z_{s^-})\right)\, \tilde{N}(ds, d\Pi), \tag{4.41}$$

where $A \in \mathcal{B}(\mathbb{R} - \{0\})$. Then, using (4.35), (4.36), (4.40), (4.41), and that $u(0,x) = v_0(x)$, we can write $\tilde{v}$ as

$$\begin{aligned} \tilde{v}(t,x) &= v_0(x) + \int_0^t \beta\, \tilde{v}_x(s^-, x)\, ds - \int_0^t \sqrt{\rho}\, \tilde{v}_x(s^-, x)\, dW_s^S \\ &+ \frac{1}{2}\int_0^t \tilde{v}_{xx}(s^-, x)\, ds + \int_0^t \int_A \left(\tilde{v}(s^-, x - \Pi) - \tilde{v}(s^-, x)\right)\, \tilde{N}(ds, d\Pi), \end{aligned}$$

where a differential representation of $\tilde{v}$ is given in (4.38). □

Lemma 4.7.3. *Let $\tilde{v}$ be defined as above and suppose Assumptions (A1) and (A2) are satisfied, then $\mathbb{P}^Z$-a.s. $\tilde{v}(t,\cdot) \in C^\infty(\mathbb{R})$, $\forall t \in (0,T]$.*

Proof. Similar to Section 4.6, this is a consequence of equality (4.35) and the properties of u. □

In order to state the next Proposition, we need to introduce some notation first. Let $\tilde{\nu}_{t|s}(A)$, be the empirical measure of the system (4.31) conditional on the filtration $\mathscr{F}_s$,

$$\tilde{\nu}_{t|s}(A) := \mathbb{P}^{\nu_0}[B_t + Z_t \in A \mid \mathscr{F}_s], \quad \forall A \in \mathcal{B}(\mathbb{R}),\ 0 \le s < t \le T. \tag{4.42}$$

Also, let $\bar{\nu}_t(A, \mu)$ be the empirical measure of the system (4.24), where the initial distribution of each particle is some distribution μ. Similarly, $\bar{v}(t, \cdot, f_\mu)$ is a density of

$\bar{\nu}_t(\cdot,\mu)$, such that f_μ is a density of μ

$$\bar{\nu}_t(A,\mu) := \int_A \bar{v}(t,x,f_\mu)\,dx \tag{4.43}$$

$$\bar{v}(t,x,f_\mu) := \int_{\mathbb{R}} p(t,x-Z_t^C,y)\,f_\mu(y)\,dy. \tag{4.44}$$

Observe that under this notation, $\bar{\nu}_t(A)$ given by Definition 4.6.1, can be written as $\bar{\nu}_t(A,\nu_0)$, and also $\bar{v}(t,\cdot)$ as $\bar{v}(t,\cdot,\nu_0)$, where $A \in \mathcal{B}(\mathbb{R})$ and ν_0 is the distribution of particles at the time $t=0$.

Proposition 4.7.2. *Let $(t_1,\Pi_1),\ldots,(t_n,\Pi_n)$, $t_1>0$, $t_n<T$, be a vector of the times and sizes of the jumps of $\{Z_t : 0<t\leq T\}$. Let $\tilde{\nu}_{t|t_{k-1}}$ be defined as in (4.42), $\bar{\nu}_{t-t_{k-1}}(\cdot,\cdot)$ and $\bar{v}(t-t_{k-1},\cdot,\cdot)$ as in (4.43)-(4.44), where $k=1,\ldots,n,n+1$ and $t_0=0$. Then for any $A\in\mathcal{B}(\mathbb{R})$, we obtain $\mathbb{P}^Z$-a.s.*

$$\tilde{\nu}_{t|t_{k-1}}(A)=\begin{cases}\bar{\nu}_t(A,\nu_0), & t\in(0,t_1),k=1,\\ \bar{\nu}_{t_1}(A-\Pi_1,\nu_0), & t=t_1,k=1,\\ \bar{\nu}_{t-t_{k-1}}(A,\tilde{\nu}_{t_{k-1}}), & t\in(t_{k-1},t_k),k=2,\ldots,n,\\ \bar{\nu}_{t_k-t_{k-1}}(A-\Pi_k,\tilde{\nu}_{t_{k-1}}), & t=t_k,k=2,\ldots,n,\\ \bar{\nu}_{t-t_n}(A,\tilde{\nu}_{t_n}), & t\in(t_n,T],\end{cases} \tag{4.45}$$

and also, the density of the measure $\tilde{\nu}_{t|t_{k-1}}$ is given as

$$\tilde{v}(t,x)_{|t_{k-1}}=\begin{cases}\bar{v}(t,x,v_0), & t\in(0,t_1),k=1,\\ \bar{v}(t_1,x-\Pi_1,v_0), & t=t_1,k=1,\\ \bar{v}(t-t_{k-1},x,\tilde{v}_{t_{k-1}}), & t\in(t_{k-1},t_k),k=2,\ldots,n,\\ \bar{v}(t_k-t_{k-1},x-\Pi_k,\tilde{v}_{t_{k-1}}), & t=t_k,k=2,\ldots,n,\\ \bar{v}(t-t_n,x,\tilde{v}_{t_n}), & t\in(t_n,T].\end{cases} \tag{4.46}$$

Proof. Let $t \in (t_{k-1}, t_k)$, $k = 2, \ldots, n$, then by the definition of $\tilde{\nu}_{t|t_{k-1}}$ and the Markov property of $\tilde{X}$, we obtain

$$\begin{aligned}\tilde{\nu}_{t|t_{k-1}}(A) &= \mathbb{P}^{\nu_0}[B_t + Z_t \in A \mid \mathscr{F}_{t_{k-1}}] \\ &= \mathbb{P}^{\nu_0}[B_t + Z_t \in A \mid \tilde{X}_{t_{k-1}}]. \end{aligned} \tag{4.47}$$

Observe that by the Markov property, $\{B_{t_{k-1}+h} - B_{t_{k-1}} : h \in (0, t_k - t_{k-1})\}$ has the same distribution as $\{B_h - B_0 : h \in (0, t_k - t_{k-1})\}$, and $\{Z^C_{t_{k-1}+h} - Z^C_{t_{k-1}} : h \in (0, t_k - t_{k-1})\}$ has the same distribution as $\{Z^C_h - Z^C_0 : h \in (0, t_k - t_{k-1})\}$. Also, note that, by Definition 4.7.1, $\tilde{\nu}_{t_{k-1}}$ is the distribution of $\tilde{X}$. Therefore, (4.47) can be written as

$$\tilde{\nu}_{t|t_{k-1}}(A) = \mathbb{P}^{\tilde{\nu}_{t_{k-1}}}[B_{t-t_{k-1}} + Z^C_{t-t_{k-1}} \in A],$$

where $\{B_{t-t_{k-1}} : t \in (t_{k-1}, t_k)\}$ is a Brownian motion with the initial distribution $\tilde{\nu}_{t_{k-1}}$, and $\{Z^C_{t-t_{k-1}} : t \in (t_{k-1}, t_k)\}$ is a Brownian motion with the initial value zero. Then, by Definition 4.6.1 and by (4.43), we obtain $\tilde{\nu}_{t|t_{k-1}}(A) = \bar{\nu}_{t-t_{k-1}}(A, \tilde{\nu}_{t_{k-1}})$.

Let now $t = t_k$, $k = 2, \ldots, n$, then we have,

$$\begin{aligned}\tilde{\nu}_{t_k|t_{k-1}}(A) &= \mathbb{P}^{\tilde{\nu}_{t_{k-1}}}[B_{t_k - t_{k-1}} + Z^C_{t_k - t_{k-1}} + \Pi_k \in A] \\ &= \bar{\nu}_{t_k - t_{k-1}}(A - \Pi_k, \tilde{\nu}_{t_{k-1}}).\end{aligned}$$

Also, for $t \in (0, t_1)$ and $k = 1$, i.e. before the first jump and when we condition on the filtration at time $t = 0$, we simply have $\tilde{\nu}_{t|0}(A) = \bar{\nu}_t(A, \nu_0)$. Then, for $t = t_1$ and $k = 1$, i.e. at the first jump, by the previous arguments, we obtain the result. Having (4.45), by (4.44), we get (4.46). □

4.8 The final system of particles

Finally, we consider the particle system that we introduce in Section 4.2, which is a particle system with underlying process $\tilde{X}$, analysed in Section 4.7, and an absorbing boundary monitored discretely, that is

$$\begin{aligned} X_t^i &= \tilde{X}_t^i \, 1_{\{t<\tau_i^D\}}, \\ \tau_i^D &= \min\{t \in \{T_1, \ldots, T_m\} : \tilde{X}_t^i \leq 0\}, \end{aligned} \tag{4.48}$$

where $T_1 > 0$, $T_m = T$. Let the probability space, on which the system is defined be the same as in Section 4.7. Conditional on Z, $(X^i, i = 1, 2, \ldots)$ are i.i.d.; hence, similar to the previous sections, we have the following definition of the empirical measure of the final system.

Lemma 4.8.1 (Conditional Strong Law of Large Numbers for the system)**.** *Let $t > 0$, $A \in \mathcal{B}(\mathbb{R})$ and $(X_t^i, i = 1, 2, \ldots)$ be given by* (4.48)*, then, for almost all $\omega^Z \in \Omega^Z$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i(\omega^Z)\in A\}} = \mathbb{P}^{\mathbb{B}}(X_t(\omega^Z) \in A), \ \mathbb{P}^{\mathbb{B}}\text{-a.s.}$$

Definition 4.8.1. *The empirical measure of the system* (4.48)*, ν_t, is defined for almost all $\omega^Z \in \Omega^Z$ as*

$$\nu_t(A) = \nu_t(A)(\omega^Z) := \mathbb{P}^{\mathbb{B}}(X_t(\omega^Z) \in A), \quad \forall A \in \mathcal{B}(\mathbb{R}), \ t \in (0, T].$$

In order to state the next Proposition, we need to introduce some notation. Let $\nu_{t|s}$ be the empirical measure for the system conditional on filtration $\mathscr{F}_s$, where $0 \leq s < t \leq T$. Also, let $\tilde{\nu}_t(A, \mu)$ be the empirical measure of the system (4.31), where the initial distribution of each particle is some distribution μ. Similarly, $\tilde{v}(t, \cdot, f_\mu)$ is a density of

$\tilde{\nu}_t(\cdot, \mu)$, such that f_μ is a density of μ

$$\tilde{\nu}_t(A, \mu) := \int_A \tilde{v}(t, x, f_\mu)\, dx \tag{4.49}$$

$$\tilde{v}(t, x, f_\mu) := \int_{\mathbb{R}} p(t, x - Z_t, y)\, f_\mu(y)\, dy. \tag{4.50}$$

Under this notation, $\tilde{\nu}_t(A)$ given by Definition 4.7.1, can be written as $\tilde{\nu}_t(A, \nu_0)$, and also $v(t, \cdot)$ as $v(t, \cdot, \nu_0)$, where $A \in \mathcal{B}(\mathbb{R})$ and ν_0 is the distribution of particles at the time $t = 0$. In the next result, we give an explicit recursive formula for $\nu_{t|s}$ and its density.

Proposition 4.8.1. *Let $\nu_{t|T_{k-1}}$ be defined as above, $\tilde{\nu}_{t-T_{k-1}}(\cdot, \cdot)$ and $\tilde{v}(t - T_{k-1}, \cdot, \cdot)$ as in (4.49)-(4.50), where $k = 1, \ldots, m$ and $T_0 = 0$, then, for any $A \in \mathcal{B}(\mathbb{R})$, we obtain $\mathbb{P}^Z$-a.s.*

$$\nu_{t|T_{k-1}}(A) = \begin{cases} \tilde{\nu}_{t-T_{k-1}}(A, \nu_{T_{k-1}}), & t \in (T_{k-1}, T_k), \\ \tilde{\nu}_{t-T_{k-1}}(A \cap (0, +\infty), \nu_{T_{k-1}}), & t = T_k, \end{cases}$$

and also, the density of the measure $\nu_{t|T_{k-1}}$, for $t \in (T_{k-1}, T_k)$, is given as

$$v(t, x) = \tilde{v}(t - T_{k-1}, x, v_{T_{k-1}}),$$

whereas, for $t = T_k$, it is

$$v(t, x) = \begin{cases} \tilde{v}(t - T_{k-1}, x, v_{T_{k-1}}), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Proof. Let $t \in (0, T_1)$, i.e. before the first monitoring time, then, by Section 4.7, we obtain $\nu_{t|T_0}(A) = \tilde{\nu}_{t-T_0}(A, \nu_{T_0})$, where $T_0 = 0$, which can be written simply as $\nu_t(A) = \tilde{\nu}_t(A, \nu_0)$.

Then, at time $t = T_1$, we have

$$\begin{aligned} \nu_{t|T_0}(A) &= \mathbb{P}^{\nu_0}(\tilde{X}_t \in A, \tilde{X}_t \in (0,+\infty)) = \mathbb{P}^{\nu_0}(\tilde{X}_t \in A \cap (0,+\infty)) \qquad (4.51) \\ &= \mathbb{P}^{\nu_0}(B_t + Z_t \in A \cap (0,+\infty)) \\ &= \tilde{\nu}_t(A \cap (0,+\infty), \nu_0), \end{aligned}$$

where we use the definition of τ^D. Also, the density is given as

$$v(T_1, x) = \tilde{v}(T_1, x, v_0)\, 1_{\{x>0\}}.$$

Then, for $t \in (T_1, T_2)$, by the Markov property of $\tilde{X}$, we have

$$\begin{aligned} \nu_{t|T_1}(A) &= \mathbb{P}^{\nu_0}(\tilde{X}_t \in A, \tilde{X}_{T_1} \in (0,+\infty) \mid \tilde{X}_{T_1}), \\ &= \mathbb{P}^{\nu_0}(\tilde{X}_{T_1} + B_{t-T_1} + Z_{t-T_1} \in A, \tilde{X}_{T_1} \in (0,+\infty)). \end{aligned}$$

Then, by (4.51), the distribution of $\tilde{X}_{T_1} \in (0,+\infty)$ is given by $\nu_{T_1|T_0}$; hence

$$\begin{aligned} \nu_{t|T_1}(A) &= \mathbb{P}^{\nu_{T_1}}(B_{t-T_1} + Z_{t-T_1} \in A), \\ &= \tilde{\nu}_{t-T_1}(A, \nu_{T_1}), \end{aligned}$$

where, for ease of notation, we drop dependence of ν_{T_1} on T_0. Then, the density is given as

$$v(t, x) = \tilde{v}(t - T_1, x, v_{T_1}).$$

Recursively, we obtain the results for all $t \in (T_{k-1}, T_k)$ and $t = T_k$, where $k = 1, \ldots, m$. □

Theorem 4.8.1. *Let $v(t,x)$ be defined as above, and let Assumptions $(A1)$ and $(A2)$ hold, then for any $k = 1, \ldots, m$, we obtain $\mathbb{P}^Z$-a.s*

(i) $v(t,\cdot) \in C^\infty(\mathbb{R})$, $\forall t \in (T_{k-1}, T_k)$,

(ii) *for* $t \in (T_{k-1}, T_k)$, $x \in \mathbb{R}$, v *is given as*

$$\begin{aligned} v(t,x) &= u(t, x - Z_t) && (4.52)\\ u_t(t,x) &= \frac{1}{2}(1-\rho)\, u_{xx}(t,x), && (4.53) \end{aligned}$$

where the initial condition at time T_{k-1} *for* u *equals* $v_{T_{k-1}}$*, given in Proposition 4.8.1,*

(iii) *for* $t \in (T_{k-1}, T_k)$, $x \in \mathbb{R}$, v *satisfies a stochastic partial differential equation,*

$$\begin{aligned} dv(t,x) &= -\beta\, v_x(t^-,x)\, dt + \frac{1}{2} v_{xx}(t^-,x)\, dt - \sqrt{\rho}\, v_x(t^-,x)\, dW_t^\varsigma \\ &+ \big(v(t^-, x-\Pi) - v(t^-,x)\big)\ \tilde{N}(dt, d\Pi), && (4.54) \end{aligned}$$

with initial condition $v_{T_{k-1}}$*, given in Proposition 4.8.1,*

(iv) $\lim_{(t,x)\to(T_{k-1},x_0)} v(t,x)$, $t \in (T_{k-1}, T_k)$, *is given as*

$$\lim_{(t,x)\to(T_{k-1},x_0)} v(t,x) = \begin{cases} v(t,x_0), & x_0 > 0, \\ 0, & x_0 < 0. \end{cases} \qquad (4.55)$$

Proof. For $t \in (T_0, T_1)$, by Lemma 4.7.3, $v(t,\cdot) \in C^\infty(\mathbb{R})$. Then, at time $t = T_1$, by Proposition 4.8.1, $v(t,x) = \tilde{v}(t,x,\nu_0)\ 1_{\{x>0\}}$ and since v_0 meets assumptions $(A1)$ and $(A2)$, then, by Lemma 4.7.3, $\tilde{v}$ is smooth, which leads to v being bounded and piecewise continuous. Then, by Proposition 4.8.1, for $t \in (T_1, T_2)$, $v(t,x) = \tilde{v}(t - T_1, x, v_{T_1})$, and by Lemma 4.7.3, $\tilde{v}$ is again smooth. By recursion, we obtain that a.s. $v(t,\cdot) \in C^\infty(\mathbb{R})$. The results (ii)-(iv) are the simple consequences of Propositions 4.8.1, 4.7.1 and Theorem 4.5.1.

□

We wish now to introduce the *loss of the particle system*, L_t, based on the empirical measure ν_t.

Definition 4.8.2. *The loss of the particle system* (4.48)*, L_t, is defined for almost all $\omega^Z \in \Omega^Z$ as*

$$L_t = L_t(\omega^Z) := 1 - \nu_t(\mathbb{R})(\omega^Z) \quad \forall t \in [0, T].$$

By the definition of L_t and ν_t, for almost all $\omega^Z \in \Omega$, L_t is a non decreasing, piecewise constant function, starting at $L_0 = 0$.

Since, ν_t has density with respect to the Lebesgue measure, then $\mathbb{P}^Z$-a.s. we have

$$L_t = 1 - \int_{\mathbb{R}} v(t, x)\, dx, \quad \forall t > 0. \tag{4.56}$$

4.9 Conclusions

In this chapter, we represent the structural jump-diffusion model with defaults monitored discretely as a system of particles, extending the approach given in [13].

In the large basket approximation, particles correspond to *distances-to-default* of entities in the basket and the loss of the portfolio is a function of the empirical measure of the system. We show that, for the infinite portfolio, the empirical measure has a density with respect to the Lebesgue measure, and also, the density satisfies an SPDE. By recursion, we give explicit formulas to solve the resulting SPDE.

What is more, in Chapter 6, we present an outline of how the above results can be extended to a system of particles with more general driving processes or when interaction between particles is included.

Chapter 5

Numerical methods and tests for the extension to the large basket approximation

In this chapter, we present numerical methods and tests developed for the extension to the Bush et al. [13] approximation, the theoretical derivation of which is given in Chapter 4 and which are the basis of the numerical results presented in the Chapter 2.

Overview of this chapter

The chapter is organised as follows. In Section 5.1, we check the validity of the approximation for a large but finite number of companies. Section 5.2 gives details of the numerical implementation of the model. Then, in Section 5.3, we show how to calibrate the model efficiently. Finally, in Section 5.4, we discuss methods to calculate sensitivities and Section 5.5 concludes.

5.1 Validity of the approximation

In this section we analyse if the size of common CDO baskets, typically $N = 125$, is large enough to use the (numerical) solution to the limiting SPDE, given in Theorem 4.8.1 and obtained for $N \to \infty$, as approximation to the evolution of firm values. We construct "nested" baskets of size $N_k = 5^k$, $k = 1, \ldots, 9$, ie, the N_k-basket contains all firms of the N_{k-1}-basket etc. For each basket, we calculate expected tranche losses, $\mathbb{E}[Y_t]$, for $t = 5$ years, where

$$Y_t = [L_t - a]^+ - [L_t - d]^+,$$

and the losses are either calculated from (1.1) by "direct" Monte Carlo simulation of (4.1), or by

$$L_t = (1 - R)\left(1 - \int_{\mathbb{R}} v(t, x)\, dx\right), \tag{5.1}$$

where we use the representation given in (4.56), adjusted by the recovery rate, and where v is the solution to the SPDE.

To create hypothetical baskets of different sizes, we simulate initial *distances-to-default*, X_0^i. For direct Monte Carlo simulation, the samples serve directly as starting values for (2.11), whereas for the SPDE, we construct the empirical measure (4.2) at $t = 0$. For simplicity, we draw X_0^i from a normal distribution, where the mean $\mu_{X_0} = 4.6$ and standard deviation $\sigma_{X_0} = 0.8$.

We calculate expected tranche losses conditional on these initial positions, ie, we average tranche losses over a large number of sample paths, but do *not* resample the initial positions. This is theoretically slightly less elegant but closer to the practical situation where initial values are backed out from implied default probabilities.

The differences between the direct Monte Carlo and SPDE results are due to: error of the finite difference discretisation of the SPDE; simulation error, both for the SPDE and

direct Monte Carlo; large basket approximation of the dynamics. We want to focus on the latter effect, therefore we choose a large number of grid points and time steps to reduce the discretisation bias, and a high number of simulations to reduce the Monte Carlo error to a negligible size.

As seen in Fig. 5.1, the difference between the SPDE results and those from direct Monte Carlo simulation of the basket vanishes (within the sampling error of X_0^i) for large N.

The SPDE results depend on N only via the initial sample X_0^i, which is chosen to be the same here as for the direct Monte Carlo simulation. Therefore, the results in Fig. 5.1 indicate that the approximation error of tranche spreads due to approximating the evolution of the firm value distribution by the SPDE is noticeably smaller than the effect of the finite sample approximating the continuous density at $t = 0$.

Closer inspection reveals that this is due to the fact that the tails of the initial distribution of X_0^i are not well resolved for moderate N, which affects junior tranches, which measure losses in the left tail, more than senior tranches. To verify this numerically, we repeated the experiment with all mass centred in a single point $X_0^i = z$, where z is a constant, for all firms in the portfolio. Thus the initial condition becomes independent of the basket size. This gave a significantly smaller difference, especially for junior tranches.

For $N = 125$, ie, $k = 3$, the results of both methods are close, hence we argue that this number of companies is sufficiently large to use the large basket approximation.

5.2 Simulation of the model

The key ingredient in the calculation of index and tranche spreads are the basket losses. We ascertained in Section 5.1 that the large basket approximation from the Chapter 4 is sufficiently accurate for the considered basket size. Therefore, the losses may be obtained from the large basket density v as per (5.1), where v has to be approximated by a numerical solution to the SPDE. In this section, we outline a numerical method based

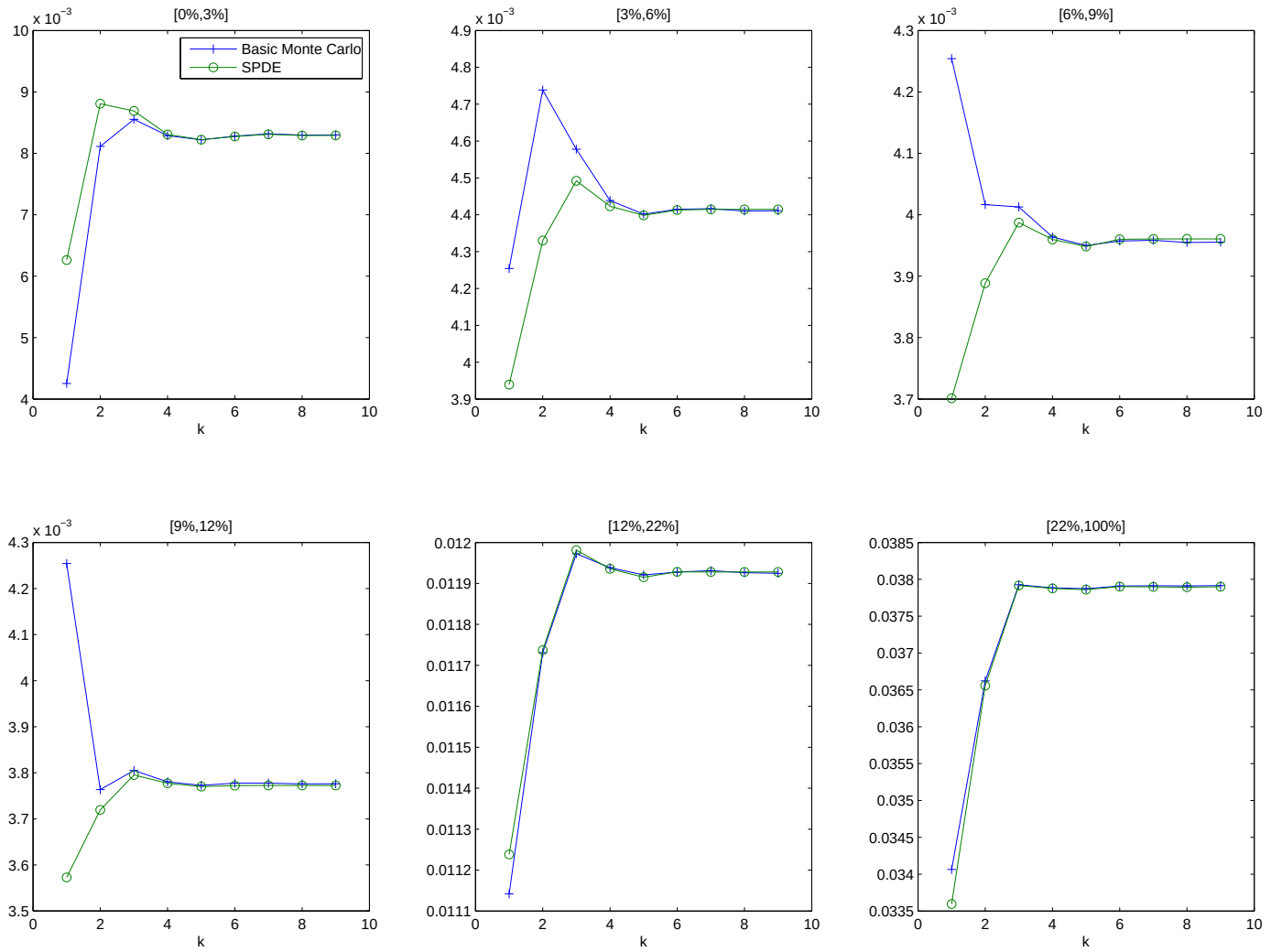

Figure 5.1: Expected losses in tranches [0%-3%], [3%-6%], [6%-9%], [9%-12%], [12%-22%], [22%-100%], calculated by direct Monte Carlo simulation of $N_k = 5^k$ asset value processes, $k = 1, \ldots, 9$, and the SPDE approximation, t=5 years, $X_0 \sim N(\mu_{X_0}, \sigma^2_{X_0})$, where $\mu_{X_0} = 4.6$, $\sigma_{X_0} = 0.8$, and, for $t > 0$, X^i_t follows a jump diffusion process (2.11) with parameters obtained in the calibration exercise for 5 December 2008 from Section 2.3.

on Monte Carlo simulation on top of a PDE solver. In the presence of non-smooth (initial) data, and with senior tranche prices depending on very low probability events, a carefully adapted numerical scheme is necessary for good accuracy.

We discuss the calibration of model parameters, including initial conditions, in Section 2.3, and assume these as given here.

5.2.1 Setup and finite difference discretisation

We can take advantage of the representation (4.52)-(4.53) of the solution to the SPDE problem (4.54) piecewise in time, as the boundary condition is not active in intervals (T_k, T_{k+1}), and therefore the Brownian driver and the jump part only introduce a shift to the solution, accumulated over the entire interval. Conditional on the number of jumps that occur in the interval, the shift is normally distributed. Based on Theorem 4.8.1, a different density with identical losses can therefore be defined by

$$v(t,x) = \begin{cases} 0 & x \leq 0, t = T_{k+1}, \\ v^{(k)}(t - T_k, x - \sqrt{\rho}(W_t^S - W_{T_k}^S) - (\tilde{J}_t - \tilde{J}_{T_k})), & \text{else if } t \in (T_k, T_{k+1}], \end{cases} \tag{5.2}$$

for $0 \leq k < n$, where $\tilde{J}_t = \sum_{i=1}^{\tilde{N}_t} \Pi_i$, and $v^{(k)}$ is the solution to the (deterministic) problem

$$v_t^{(k)} = \frac{1}{2}(1-\rho)v_{xx}^{(k)} - \beta v_x^{(k)}, \quad t \in (0, \Delta T) = (0, T_{k+1} - T_k) \tag{5.3}$$

$$v^{(k)}(0,x) = v(T_k, x) \tag{5.4}$$

assuming monitoring dates are equally spaced with intervals $\Delta T = T_{k+1} - T_k$.

Note that, in Chapter 4, we include the drift term into the common factor, and hence we obtain the representation (4.52)-(4.53), whereas here the drift is not included, and therefore instead of the heat equation we have the PDE of the form (5.3). Both representations are equivalent for each realisation of $(W_t^S - W_{T_k}^S)$ and $(\tilde{J}_t - \tilde{J}_{T_k})$, which can be shown by Itô's lemma.

The above results suggest the following inductive strategy for $k = 0, \ldots, n-1$:

1. Start with $v^{(0)}(0, x) = v_0(x)$, the initial empirical measure.
2. Solve the PDE (5.3) numerically in the interval $(0, T_1)$, to obtain $v^{(0)}(T_1, x)$.
3. Simulate $W^{\varsigma}_{T_1}$, $\tilde{J}_{T_1}$ evaluate $v(T_1, x)$ according to (5.2).
4. For $k > 0$, having computed $v(T_k, x)$ in the previous step, use this as initial condition for $v^{(k)}$, and repeat until $k = n$.

To solve the PDEs (5.3) by a finite difference method, we approximate the measure by one with support $[x_{min}, x_{max}]$, and set zero boundary conditions at x_{min} and x_{max}.

Proposition 5.2.1. *If v^b is the solution to (5.2)-(5.4), where (5.3) holds only on $(-b, b)$, with $v^{(k)}(t, -b) = v^{(k)}(t, b) = 0$, then*

$$\mathbb{E}[Z^b_t] \to \mathbb{E}[Z_t] \quad \text{for} \quad b \to \infty,$$

where Z^b_t is the outstanding tranche notional derived from losses L^b_t of v^b.

Proof. See [12]. □

In practice, we pick x_{min} and x_{max} experimentally large enough such that the effect of the truncation is negligible.

Then, introduce a grid $x_0 = x_{min}, x_1 = x_{min} + \Delta x, \ldots, x_j = x_{min} + j\Delta x, \ldots, x_J = x_{min} + J\Delta x$, where $\Delta x = (x_{max} - x_{min})/J$, timesteps $t_0 = 0, t_1 = \Delta t, \ldots, t_I = I\Delta t$, where $\Delta t = \Delta T/I$, and define an approximation v^i_j to $v(t_i, x_j)$ by the difference scheme

$$\frac{v^i_j - v^{i-1}_j}{\Delta t} = \theta \left\{ \frac{1}{2}(1-\rho)\frac{v^i_{j+1} - 2v^i_j + v^i_{j-1}}{\Delta x^2} - \beta \frac{v^i_{j+1} - v^i_{j-1}}{2\Delta x} \right\} \tag{5.5}$$

$$+(1-\theta) \left\{ \frac{1}{2}(1-\rho)\frac{v^{i-1}_{j+1} - 2v^{i-1}_j + v^{i-1}_{j-1}}{\Delta x^2} - \beta \frac{v^{i-1}_{j+1} - v^{i-1}_{j-1}}{2\Delta x} \right\}. \tag{5.6}$$

For an introduction to finite differences in financial instrument pricing see, e.g., [76]. The scheme is of second order accurate in Δx for smooth solutions. The Crank-Nicolson scheme $\theta = \frac{1}{2}$ is of second order accurate in Δt, and is unconditionally stable for sufficiently smooth solutions, but does not converge for initial conditions comprising of δ-distributions as in the present setting. We address this, together with aspects of approximating non-smooth initial and interface conditions (4.55) accurately on the grid, in the following section.

5.2.2 Initial and interface conditions

The initial condition has the form

$$v(0,x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - x^i),$$

where x^i is the observed initial distance-to-default of firm i.

To achieve second order accuracy in Δx, the δ's have to be "split" between adjacent grid points. The correct weighting can be written as integral of linear splines

$$\Phi_k(x) = \frac{1}{\Delta x} \min\left(\max(x - x_k + \Delta x, 0), \max(x_k + \Delta x - x, 0)\right)$$

over the initial condition, ie,

$$v_k^0 = \frac{1}{\Delta x}\int_{x_{min}}^{x_{max}} \Phi_k(x) v(0,x)\, dx.$$

Note that the initial condition is a probability measure and in particular the above definition is well defined for Dirac measures. By this construction, $\Delta x \sum_{k=0}^{J} v_k^0 = 1$. See also [67] for applications of this idea to option pricing and estimation of sensitivities.

To incorporate the interface condition (5.2) at $t = T_k$, one has to evaluate the grid function at shifted arguments $x_j - \sqrt{\rho}(W_t^\varsigma - W_{T_k}^\varsigma) - (\tilde{J}_t - \tilde{J}_{T_k})$, which do not normally coincide with grid points. To deal with this, we first define a piecewise linear interpolant

$\widehat{v}$, obtained in the last step over the previous interval $[T_{k-1}, T_k]$, $v_j^{(k),I}$, I the number of timesteps, by

$$\widehat{v}(T_k, x) = \sum_{j=0}^{J} \Phi_j(x - \sqrt{\rho}\,\Delta W^{\varsigma} - \Delta\tilde{J}) v_j^{(k),I}, \tag{5.7}$$

where $\Delta W^{\varsigma} = W^{\varsigma}_{T_k} - W^{\varsigma}_{T_{k-1}}$, $\Delta\tilde{J} = \tilde{J}_{T_k} - \tilde{J}_{T_{k-1}}$. Then, approximate (5.2) by setting

$$v_j^{(k+1),0} = \int_{\max(x_j - \Delta x/2, 0)}^{\max(x_j + \Delta x/2, 0)} \widehat{v}(T_k, x)\, dx, \tag{5.8}$$

and use this as initial condition for the next interval. This ensures that

$$\Delta x \sum_{j=0}^{J} v_j^{(k+1),0} = \int_0^{x_{max}} \widehat{v}(T_k, x)\, dx,$$

so the cumulative density of firms with firm values greater than 0 is preserved. It also has the effect that the solution is smoothed at $x = 0$. In contrast to simpler, e.g., pointwise application of the interface condition, this procedure guarantees second order convergence in Δx. See [67] for averaging procedures to restore higher order convergence for non-smooth payoffs in option pricing.

Finally, it is well-known that Crank-Nicolson timestepping has reduced convergence order for discontinuous initial conditions, and does not converge at all for Dirac initial conditions, unless the timesteps are chosen very small. This severely slows down the performance. A simple and well-established remedy is to replace the first Crank-Nicolson steps with backward Euler steps, a practice now known as "Rannacher start-up" [69]. The analysis in [39] shows that the optimal balance between accuracy and stability is achieved by replacing the first two Crank-Nicolson steps by four backward Euler steps of half the stepsize. We do this at $t = 0$, and also at $t = T_k$, where the interface conditions introduce discontinuities at $x = 0$.

5.2.3 Loss simulation

For a given realisation of the market factors, we can approximate the loss functional (5.1) at time T_k, using (5.7) and (5.8), by

$$\widehat{L}_{T_k} = (1-R)\left(1 - \int_0^{x_{max}} \widehat{v}(T_k, x)\, dx\right) = (1-R)\left(1 - \Delta x \sum_{j=1}^{J-1} v_j^{(k+1),0}\right). \tag{5.9}$$

We first study the discretisation error of (5.9) in Δt and Δx numerically, first for a single realisation of the path of the market factors. Fig. 5.2 shows, for a typical set of

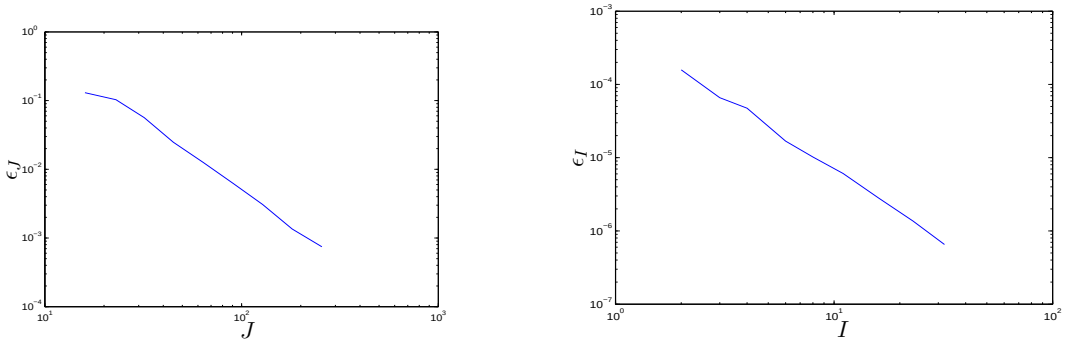

Figure 5.2: Estimated discretisation error of $\widehat{L}_{T_n}$ for increasing J (left) and I (right) for a single realisation of the path of the market factors. The error estimator is based on Richardson extrapolation of the numerical solutions on subsequent refinement levels.

model parameters, how many grid points (I) and how many timesteps between payment dates (J) are necessary for a desired accuracy ϵ. We clearly see second order convergence in Δt and Δx. Note that the time smoothing scheme, weighted approximation of initial positions and adapted averaging, as per Section 5.2.2, are essential to achieve this.

Next, we want to compute expected losses and outstanding tranche notionals. If we explicitly include the dependence on the Monte Carlo samples ϕ_i of $\sqrt{\rho}\,(W^{\varsigma}_{T_i} - W^{\varsigma}_{T_{i-1}}) + \tilde{J}_{T_i} - \tilde{J}_{T_{i-1}}$ in (5.9) by writing $\widehat{L}_{T_k}(\phi)$, $\phi = (\phi_i)_{1\le i\le n}$, where ϕ_i are independent, then for

N_{sims} simulations with samples $\phi^l = (\phi_i^l)_{1\leq i\leq n}$, $1 \leq l \leq N_{sims}$,

$$\begin{aligned}\mathbb{E}[Z_{T_k}] &\approx \mathbb{E}[\max(d - \widehat{L}_{T_k}, 0) - \max(a - \widehat{L}_{T_k}, 0)] \\ &\approx \frac{1}{N_{sims}} \sum_{l=1}^{N_{sims}} \left(\max(d - \widehat{L}_{T_k}(\phi^l), 0) - \max(a - \widehat{L}_{T_k}(\phi^l), 0)\right).\end{aligned}$$

For the simulations, we fix Δx and Δt at values which have proven empirically to give negligible discretisation error which is confirmed by numerical tests. We have used two data sets: from 22 February 2007 and 5 December 2008. Initial positions for individual firms were calibrated to their individual CDS spreads, and were well within the range $[x_{min}, x_{max}] = [-10, 20]$.

We now analyse the convergence of the obtained Monte Carlo estimator. The estimated expected tranche losses with confidence intervals are given in Figure 5.3, for 22 February 2007. For super senior tranches especially, the standard error is higher relative to the values than for equity and mezzanine ones. The reason behind this is that a large number of companies have to default in order to affect senior tranches and such an event is rare. For this date, the estimated parameters imply normal market conditions, where multiple defaults are highly unlikely (see Section 2.3 for a discussion of calibration results). In order to price senior tranches more accurately, variance reduction techniques such as importance sampling should be applied, see, e.g., [14], however the accuracy of the results for the basic scheme was found sufficient for the purposes of this study.

5.3 Numerical methods developed for calibration

Here we present the numerical methods that we develop for calibrating a model to market spreads under the Bush et al approximation and its extension. Calibration results are given in Chapter 2.

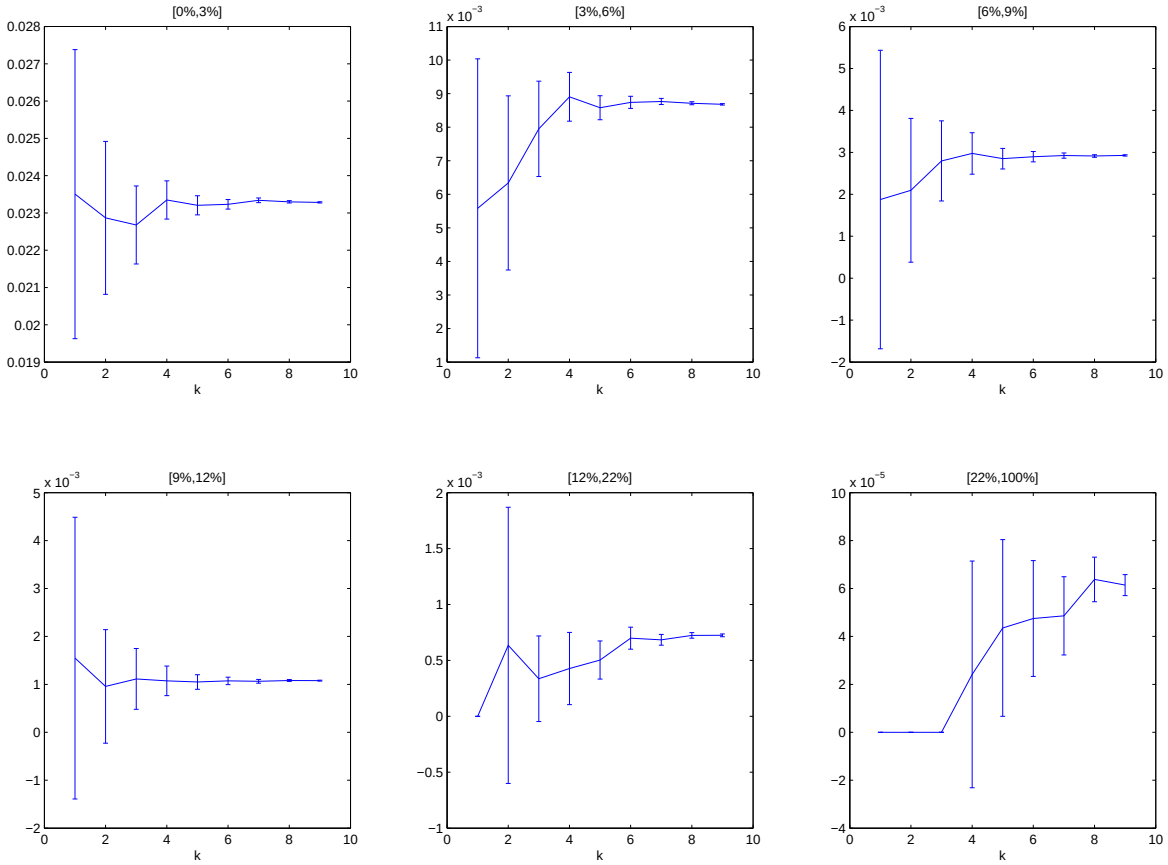

Figure 5.3: Monte Carlo estimators with 95%-confidence intervals for expected losses in tranches [0%-3%], [3%-6%], [6%-9%], [9%-12%], [12%-22%], [22%-100%], for $N = 4^{k+1}$ simulations, $k = 1, \ldots, 9$. The results are for parameters from a calibration of the model to data from 22 February 2007 (see Section 2.3).

5.3.1 Stating the calibration problem

The initial distances-to-default of the individual names are thereby calibrated to their short-term CDS spreads, to ensure consistency of the multi-name CDO model with single-name CDS models. We fit the remaining parameters defined in Section 2.2.1 to index spreads for M maturities (usually 5-, 7- and 10-year, ie, $M = 3$) and G tranche spreads for each maturity (usually $G = 6$). The number of calibration prices, $M \cdot (G + 1)$, is then typically much larger than the number of parameters, here 5. We therefore cannot expect the model to fit all prices exactly, and formulate the calibration exercise as a weighted least-squares problem.

Problem 1. *Given market spreads at time $t = 0$, of CDO tranches, $C_0^j(T_i)$, and the CDO index, $CI_0(T_i)$, for maturities T_i, $i = 1, \ldots, M$, tranches $j = 1, \ldots, G$, and given spreads $c_0 = \left(c_0^1, \ldots, c_0^N\right)$ for CDSs written on N underlying companies, solve the minimisation problem*

$$\sum_{i=1}^{M} \sum_{j=1}^{G} \alpha_i^j \left(C_0^j(T_i) - C_0^{j,\theta,x_0}(T_i) \right)^2 + \sum_{i=1}^{M} \alpha_i \left(CI_0(T_i) - CI_0^{\theta,x_0}(T_i) \right)^2 \quad \rightarrow \quad \min_{\theta}, \tag{5.10}$$

where $\theta = (\rho, \sigma, \lambda, \sigma_\Pi, \mu_\Pi)$, subject to

(i) $x_0 = \left(x_0^1, \ldots, x_0^N\right)$ is a solution to

$$c_0 = c_0^\theta(x_0), \tag{5.11}$$

(ii) $\rho \in [0, 1)$, $\sigma > 0$, $\lambda > 0$, $\sigma_\Pi > 0$, $\mu_\Pi < 0$,

where c^θ, C_0^{j,θ,x_0}, CI_0^{θ,x_0} denote CDS, CDO tranche and index spreads calculated in the model with parameter vector θ, x_0 is a vector of initial distances-to-default, $\alpha = (\alpha_i^j, \alpha_i)$ is a scaling vector.

The minimisation problem has a scaled least-squares objective function, linear inequality constraints (ii) and non-linear equality constraints (i).

Since CDO tranche and index spreads have different orders of magnitudes, we scale the data by α to make each observation roughly equally important, precisely, we choose the weights α as powers of ten such that the scaled market prices lie between 0.1 and 1. Alternative choices are discussed in Section 2.3.2.

The non-linear constraints are incorporated directly into (5.10) by numerical inversion of (5.11) giving $x_0 = x_0(\sigma, \lambda, \sigma_\Pi, \mu_\Pi) = x_0(\theta)$. We then solve the minimisation problem (5.10) with bounds on the parameters. A robust algorithm developed specially for this kind of minimisation problems is the interior-reflective Newton method in [16] and [17]. We use this algorithm as implemented in the Matlab Optimisation Toolbox.

5.3.2 Computational issues

For each calibration, a number γ of Newton-type iterations is needed to solve (5.10) to a specified accuracy, and in each iteration, an approximation to the objective function and its derivatives is required. We address here ways of improving the computational efficiency of the most costly components.

Search for Initial *Distances-To-Default*

Since we use finite differences to approximate the derivatives, x_0^n is searched $K = (1 + 2b) \times N \times \gamma$ times, where $n = 1, \ldots, N = 125$, the number of CDSs in the portfolio, $b = 4$ is the number of parameters (excluding ρ, which does not affect x_0), and γ is in the range of $10 - 20$. Since $K \approx 1.5 \times 10^4$, an efficient way of finding x_0 from (5.11) is crucial.

Given CDS data, $c_0 = (c_0^i)_{1 \le i \le N}$, we search for initial *distances-to-default*, $x_0 = (x_0^i)_{1 \le i \le N}$, taking advantage of properties of c_0. CDS spreads decrease with the initial *distance-to-default*, since higher x_0 leads to higher survival probability, which entails lower spreads, ie, if $c_0^i > c_0^j$ then $x_0^i < x_0^j$. Hence, for a CDS spread, c_0^s, where

$c_0^{\min} \leq c_0^{s-1} < c_0^s < c_0^{s+1} \leq c_0^{\max}$, the initial *distance-to-default*, x_0^s, is in (x_0^{s+1}, x_0^{s-1}). Hence, a good starting point for x_0^s is either x_0^{s+1} or x_0^{s-1}. Sorting spreads, then choosing starting points as stated above, greatly decreases the computation time of x_0.

Monte Carlo "Inside" the Objective Function

What is more, the objective function (5.10) is estimated by a Monte Carlo method, and in order to obtain accurate results, a high number of simulations is needed, as seen from the accuracy of expected tranche losses in Fig. 5.3. Given the numerical evaluation of the objective function and its derivatives is part of an iterative solution method for the optimisation problem, where the initial parameter iterates will be inaccurate, it is unnecessary to evaluate the spreads there with high accuracy.

Let $N_i, i = 1, \ldots, Z$, be an increasing number of paths, and let $\theta^i_{N_i}$ be a sequence of parameters obtained using N_i simulations, with an iterative scheme with starting value $\theta^{i-1}_{N_{i-1}}$ and γ_i iterations. Since the Monte Carlo estimator and Newton's method both converge, it is hoped that the $\theta^i_{N_i}$ converge to the optimiser, while only few iterations are needed for large N_i. This idea is a simple version of the *Multi-layer method* in [49].

In the calibration exercises, we heuristically picked $N_1 = 6 \times 10^3$, since for this number of simulations the estimator "starts converging", as can be observed in Fig. 5.3, and the computation time per iteration is very low. For N_1, γ_1 is about 15, whereas for $N_3 = N_{Z-1} = 10^5$, it is about 2-4, and, finally, for $N_4 = 10^6$, it is only 1. The overall computation time of the calibration algorithm is significantly reduced compared to calibration with uniformly N_Z simulations, here by a factor of roughly 15.

It would be possible to automatise this procedure and find an optimal sequence N_i to minimize the overall computational time, based on the variance of the Monte Carlo estimator and the convergence speed of the optimisation algorithm.

5.4 Calculation of sensitivities via automatic differentiation and finite difference

In our calibration problem, we use an algorithm based on the calculation of the Jacobian. Hence, the method for the evaluation of derivatives must be accurate and fast. Motivated by the positive results by [40] of applying automatic/algorithmic differentiation (AD) to financial mathematics, we decided to check the performance of AD derivatives in our settings compared to standard finite difference (FD) derivatives.

5.4.1 Basics of automatic differentiation

As is noted in [78], the AD method is based on calculating analytically derivatives of a function written in a computer program. Since every computer program can be decomposed into elementary functions like summation or square root, one can calculate a derivative with respect to a function defined by a computer program simply by calculating derivatives of elementary functions and then, by the chain rule, finally evaluating the derivative of a given function. The main difference between AD and FD is that AD calculates derivatives up to machine precision and does not introduce any truncation error while FD does. A good reference on automatic differentiation is [42].

There are two ways of calculating AD derivatives: forward and reverse (adjoint) mode. Following the review of the AD modes given in [78], we denote $y = (y_1, \cdots, y_s)^\top$ the output of the coded function, $\theta = (\theta_1, \cdots, \theta_m)^\top$ the vector of parameters with respect to which we would like to calculate derivatives and $z = (z_1, \cdots, z_p)^\top$ the vector of intermediate variables, where usually $p \gg s + m$. The variables z are connected to each other via elementary functions, for example in the form: $z_k = f_e^k(z_i, z_j)$, where $i < k$ and $j < k$.

In order to obtain the Jacobian $\left[\frac{\partial y_i}{\partial \theta_j}\right]$, where $i = 1, \ldots, s$, $j = 1, \ldots, m$, by the forward mode, the chain rule is applied to evaluate intermediate derivatives, beginning from z_1 and finishing at z_p. As an example let us take $z_k = f_e^k(z_i, z_j)$. The forward derivative $\dot{z}_k^n$

is given by

$$\dot{z}_k^n = \frac{\partial z_k}{\partial \theta_n} = \frac{\partial f_e^k}{\partial z_i}\frac{\partial z_i}{\partial \theta_n} + \frac{\partial f_e^k}{\partial z_j}\frac{\partial z_j}{\partial \theta_n}, \qquad \forall 1 \leq n \leq m, \forall 1 \leq k \leq p.$$

This procedure is done for all intermediate variables and finally one obtains the Jacobian.

In the adjoint mode, the order of differentiation is reversed. One starts with $\frac{\partial y}{\partial y} = 1$ and then the chain rule is applied to calculate intermediate derivatives, beginning from z_p and finishing at z_1, and finally with respect to θ. Precisely, for $z_k = f_e^k(z_i, z_j)$, the adjoint derivatives $\bar{z}_i$ and $\bar{z}_j$ are obtained by

$$\bar{z}_i = \frac{\partial y}{\partial z_i} = \frac{\partial y}{\partial z_k}\frac{\partial f_e^k}{\partial z_i}, \qquad \bar{z}_j = \frac{\partial y}{\partial z_j} = \frac{\partial y}{\partial z_k}\frac{\partial f_e^k}{\partial z_j}, \qquad \forall 1 \leq k \leq p.$$

Let us now illustrate the differences in calculating forward and adjoint derivatives in the following example. Consider a function

$$y = f(x, z) = e^{x+z}\, x^2.$$

The function can be decomposed into elementary functions

$$m = x + z \tag{5.12}$$

$$s = e^m \tag{5.13}$$

$$w = x^2 \tag{5.14}$$

$$y = s \cdot w. \tag{5.15}$$

In the forward mode, derivatives are calculated together with the function. Consider $x = 1$

and $z = 2$, in our example, y and $\dot{y}$ are calculated as follows

$$
\begin{aligned}
m &= 3 \\
\dot{m}_x &= \frac{\partial m}{\partial x} = 1 \\
\dot{m}_z &= \frac{\partial m}{\partial z} = 1 \\
s &= 20.1 \\
\dot{s}_x &= \frac{\partial s}{\partial m}\frac{\partial m}{\partial x} = e^m \cdot \dot{m}_x = 20.1 \\
\dot{s}_z &= \frac{\partial s}{\partial m}\frac{\partial m}{\partial z} = e^m \cdot \dot{m}_z = 20.1 \\
w &= 1 \\
\dot{w}_x &= 2x = 2 \\
\dot{w}_z &= 0 \\
y &= 20.1 \\
\dot{y}_x &= \frac{\partial y}{\partial x} = \frac{\partial y}{\partial s}\frac{\partial s}{\partial x} + \frac{\partial y}{\partial w}\frac{\partial w}{\partial x} = w \cdot \dot{s}_x + s \cdot \dot{w}_x = 60.3 \\
\dot{y}_z &= \frac{\partial y}{\partial z} = \frac{\partial y}{\partial s}\frac{\partial s}{\partial z} + \frac{\partial y}{\partial w}\frac{\partial w}{\partial z} = w \cdot \dot{s}_z = 20.1
\end{aligned}
$$

In the reverse mode, first a function is evaluated, the values of intermediate functions are stored and then the adjoint derivative is calculated. In our example, where $x = 1$, $z = 2$, y and $\bar{y}$ are evaluated as follows

$$m = 3 \tag{5.16}$$

$$s = 20.1 \tag{5.17}$$

$$w = 1 \tag{5.18}$$

$$y = 20.1 \tag{5.19}$$

$$\bar{y} = \frac{\partial y}{\partial y} = 1 \tag{5.20}$$

$$\bar{s} = \bar{y} \cdot \frac{\partial y}{\partial s} = w = 1 \tag{5.21}$$

$$\bar{w} = \bar{y} \cdot \frac{\partial y}{\partial w} = s = 20.1 \tag{5.22}$$

$$\bar{x} = \frac{\partial y}{\partial x} = \frac{\partial y}{\partial w}\frac{\partial w}{\partial x} = \bar{w} \cdot 2x = 40.2 \tag{5.23}$$

$$\bar{m} = \frac{\partial y}{\partial m} = \frac{\partial y}{\partial s}\frac{\partial s}{\partial m} = \bar{s} \cdot e^m = 20.1 \tag{5.24}$$

$$\bar{x} = \bar{x} + \frac{\partial y}{\partial m}\frac{\partial m}{\partial x} = \bar{x} + \bar{m} = 60.3 \tag{5.25}$$

$$\bar{z} = \frac{\partial y}{\partial x} = \frac{\partial y}{\partial m}\frac{\partial m}{\partial z} = \bar{m} = 20.1, \tag{5.26}$$

where (5.21) and (5.22) were calculated from (5.15), (5.23) from (5.14), (5.24) from (5.13), (5.25) from (5.12) and updated by (5.23), finally (5.26) was calculated from (5.12). As one can see in the above example, forward and reverse mode produce the same values of derivatives, but calculations are different.

Observe that, for each variable θ_n with respect to which one calculates the derivative a separate forward mode is needed, whereas for the adjoint method only one reverse mode is required regardless of the dimension of θ, ie calculations are common for all variables θ, until a variable explicitly occurs in the code, as in the above example.

T=5 Years		
y	σ	ρ
Index	-	0.001
Tranche 0%-3%	0.013	0.029
Tranche 3%-6%	-	0.010
Tranche 6%-9%	-	0.003
Tranche 9%-12%	-	0.002
Tranche 12%-22%	-	0.002
Tranche 22%-100%	-	-
T=7 Years		
y	σ	ρ
Index	-	0.001
Tranche 0%-3%	0.015	0.017
Tranche 3%-6%	0.001	0.008
Tranche 6%-9%	-	0.006
Tranche 9%-12%	-	0.003
Tranche 12%-22%	-	0.001
Tranche 22%-100%	-	-
T=10 Years		
y	σ	ρ
Index	-	0.001
Tranche 0%-3%	0.016	0.007
Tranche 3%-6%	0.001	0.007
Tranche 6%-9%	-	0.005
Tranche 9%-12%	-	0.004
Tranche 12%-22%	-	0.002
Tranche 22%-100%	-	-

Table 5.1: Absolute differences between derivatives calculated by AD and FD methods of CDO index and tranche spreads with respect to σ and ρ in the diffusion model. The calculations are done for $\sigma = 0.22$, $\rho = 0.32$, $R = 0.4$, $n = 128$ grid points, 4000 simulations, FD stepsize equals $1.0e-06$, by '-' we denote the results that are less than $0.1e-03$.

5.4.2 Results for the large basket approximation

First, it should be underlined that the implementation of AD for our problem is a somewhat challenging task, since the function used to calculate CDO indices and tranche spreads, for which we need the Jacobian, consists of more than 200 lines of code, consists of a finite difference solution to the PDE, Monte Carlo method, an inversion algorithm used to find initial *distances-to-default* from the CDS data, several loops (some of them nested) and also branches. As was noted before, the AD derivatives are evaluated analytically, by investigating the code line by line. Because of that, we wanted to implement the method once and then use it for different models, including models with many parameters. That is why we chose the reverse mode.

In Table 5.1, we present the results for the diffusion model. The differences between derivatives calculated by AD and FD methods are small, but the computational time for the AD method is much longer. We refer to the Appendix B for the reasons behind the poor performance of the AD method in our setting. What is more, we used both Jacobians in the calibration exercise and in both cases the same solution was found, with the same number of iterations. Hence, for this particular problem FD is accurate enough and efficient. That is why in the jump-diffusion case we finally used FD instead of AD. More details on the implementation of the AD method are given in Appendix B.

5.5 Conclusions

We have presented a numerical implementation of the extension given in Chapter 4 of the method proposed in [13].

We derive a numerical method for basket credit derivatives based on a large basket approximation and a Monte Carlo finite difference solution of the resulting SPDE, and outline an algorithm for calibration to CDO index and tranche spreads. Also, our numerical experiments confirm that calculation of sensitivities via finite difference is more

efficient in our setting than using automatic differentiation. What is more, we show that the large basket approximation is valid for a typical CDO portfolio.

Chapter 6

Conclusions

In the final chapter, we summarise the main results presented in the previous chapters and demonstrate how the results are connected and complete one other. Also, we briefly discuss potential extensions to our work.

6.1 Our thesis

In the thesis, we propose two alternative approaches to evaluate numerically an N-name Bernoulli mixture model where the number of entities, N, is large and the portfolio is homogeneous. The first approach is in the spirit of a multi-level Monte Carlo method, while the second is an extension to the large basket approximation, proposed in [13]. In this section, using a common structure, we present the main results of both methods. For convenience, some of the definitions used throughout the thesis are repeated here.

In both approaches, we consider a sequence of Bernoulli random variables, $Y = (Y_i, i = 1, \ldots, N)$, $N \geq 1$, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Each Y_i takes unity if a default of an entity occurs within a time horizon T. Also, the marginal distribution of Y is identical.

In both methods, we focus on the loss of the portfolio, L_N, defined as

$$L_N = (1-R)\frac{1}{N}\sum_{i=1}^{N} Y_i, \tag{6.1}$$

and on a tranche loss, P_N, given as the following

$$P_N = [L_N - K_1]^+ - [L_N - K_2]^+,$$

where K_1, K_2, are the tranche attachment and detachment points respectively, and $0 < K_1 < K_2 < 1$. We aim at obtaining $\mathbb{E}[L_N]$ and $\mathbb{E}[P_N]$.

In a common factor model, each Y_i is driven by an idiosyncratic factor, $Z^{I,i} = \{Z_t^{I,i}, i = 1, \dots N, t \in [0,T]\}$, and a common factor $\Psi = \{\Psi_t, t \in [0,T]\}$. Then, as previously, we divide the probability space into the space of the individual factor, $(\Omega^I, \mathscr{F}^I, \mathbb{P}^I)$, and the space of a common factor, $(\Omega^\Psi, \mathscr{F}^\Psi, \mathbb{P}^\Psi)$, such that

$$(\Omega^I \times \Omega^\Psi, \mathscr{F}^I \times \mathscr{F}^\Psi, \mathbb{P}^I \times \mathbb{P}^\Psi).$$

Then, conditional on a realisation of the common factor, Y is i.i.d., which, for $\omega^\Psi \in \Omega^\Psi$, can be written as

$$\mathbb{P}^I[Y_t^i(\omega^\Psi) = 1] = s(\omega^\Psi), \tag{6.2}$$

where s is some function, that needs to be specified for each Bernoulli mixture model.

What is more, by the conditional version of the Law of Large Numbers, we obtain that, for almost all $\omega^\Psi \in \Omega^\Psi$, the loss of an infinite portfolio exists and equals

$$L = L(\omega^\Psi) := \lim_{N\to\infty} L_N(\omega^\Psi) = \mathbb{P}^I[Y_t(\omega^\Psi) = 1], \qquad \mathbb{P}^I\text{-a.s.}, \tag{6.3}$$

leading to a tranche loss of an infinite portfolio being defined as

$$P = P(\omega^{\Psi}) := [L - K_1]^+ - [L - K_2]^+.$$

Also, by the conditional independence of Y, the conditional expected loss of a finite portfolio can be written as

$$\mathbb{E}^I[L_N(\omega^{\Psi})] = \mathbb{P}^I[Y_t(\omega^{\Psi}) = 1], \tag{6.4}$$

and the expected loss equals

$$\begin{aligned} \mathbb{E}[L_N] &= \mathbb{E}^{\Psi}\left[\mathbb{P}^I\left[Y_t(\omega^{\Psi}) = 1\right]\right] \\ &= \int_{\mathbb{R}} s(\psi)\, \mathbb{P}_{\Psi}(\Psi \in d\psi), \end{aligned} \tag{6.5}$$

where the function s is specific for each model, and $\mathbb{P}_{\Psi}$ is the law of Ψ. Then, a formula for $\mathbb{E}[P_N]$ is also specific for each Bernoulli model.

6.1.1 Multi-level approach

In the multi-level approach, we are interested in the convergence rates of P_N for $N \to \infty$ under any Bernoulli mixture model. Our crucial observation is that, using (6.3) and (6.4), we can write the expected difference between P_N and P in a model-independent way (up to the law of the infinite loss) as

$$\mathbb{E}[P_N - P] = \int_0^1 g(N, l)\, \mathbb{P}_L(L \in dl), \tag{6.6}$$

where g is some function depending on N and L; and $\mathbb{P}_L$ is the law of the infinite loss. The function g can be bounded, and based on this observation, we obtain that

$$\begin{aligned} |\mathbb{E}[P_N - P]| &= O(N^{-1}) \\ Var[P_N - P] &= O(N^{-1}), \end{aligned}$$

provided that the following condition is satisfied

$$|F_L(K_j) - F_L(l)| \leq c_L \, |K_j - l|, \tag{6.7}$$

where F_L is the cumulative density function of L, c_L a Lipschitz constant, $j = 1, 2$ and all $l \in [0, 1]$.

The convergence result shows that the expected tranche loss for large or infinite N can be well approximated by those with smaller N. Combining this with a control variate idea leads us to multilevel simulation with a substantial variance reduction for large N. Our multi-level estimators for the expected tranche loss with a mean-square error of ϵ^2 have computational complexity of order ϵ^{-2}, independent of N. Our numerical experiments confirm the theoretical results.

6.1.2 The large basket approximation

In the large basket approximation [13], the focus is on a specific Bernoulli mixture model, which is a structural diffusion model with continuously monitored defaults, also on a semi-analytical approximation for L_N and, finally, on a numerical calculation of $\mathbb{E}[L_N]$ for this model.

In our extension to the large basket approximation, we consider a structural jump-diffusion model with discretely monitored defaults. Since the conditional probability of default, given by (6.2), is unavailable in a closed form for such a model, then the representation (6.5) is not directly applicable. Following [13], instead, we represent the loss of

a portfolio as

$$L_N = (1 - R)\left(1 - \nu_{t,N}(\mathbb{R})\right),$$

where $\nu_{t,N}$ is the empirical measure of the system with an absorbing boundary, defined, for $S \in \mathcal{B}(\mathbb{R})$, as the following

$$\begin{aligned} \nu_{t,N}(S) &= \frac{1}{N}\sum_{i=1}^{N} 1_{\{X_t^i \in S\}}, \\ X_t^i &= \tilde{X}_t 1_{\{\tau_i^D > t\}}, \end{aligned}$$

where τ_i^D denotes the time of default of the i-th entity, where a default is monitored at a discrete set of times, and $\tilde{X}$ is a jump-diffusion process underlying a *distance-to-default*.

Then, by the conditional version of the Law of Large Numbers, we obtain that for almost all $\omega^\Psi \in \Omega^\Psi$, the empirical measure of an infinite portfolio exists and equals

$$\nu_t(S) = \nu_t(S)(\omega^\Psi) := \lim_{N\to\infty} \nu_{t,N}(S)(\omega^\Psi) = \mathbb{P}^I[X_t(\omega^\Psi) \in S], \qquad \mathbb{P}^I\text{-a.s.}$$

We show that this empirical measure has a density v with respect to the Lebesgue measure. Therefore, the loss of an infinite portfolio can be written as

$$L = (1 - R)\left(1 - \int_{\mathbb{R}} v(t,x)\, dx\right).$$

Also, we show that the density satisfies a stochastic partial differential equation, the solution of which can be represented between two subsequent monitoring times as a solution to the heat equation shifted by a realisation of a common factor in that time interval.

Finally, in order to obtain the loss of the portfolio and tranche loss, we use the following

approximations,

$$L_N \approx L,$$

$$P_N \approx P.$$

Then, based on the above results we develop a numerical algorithm to estimate $\mathbb{E}[L]$, $\mathbb{E}[P]$ and then the tranche spreads efficiently. Also, we propose calibration methods for the approach. What is more, we verify the validity of this approximation numerically by comparison with results obtained by the direct Monte Carlo simulation of the portfolio.

6.2 Extensions

Finally, we provide an outline of potential extensions to our work. We start by presenting extensions specific to each approach, and then focus on extensions that can be applied to both methods.

6.2.1 Multi-level approach

In the multi-level approach, we derive the main results for the case when the condition (6.7) is met. We show numerically that, for the structural jump-diffusion model, the condition holds. However, it is an open mathematical question to prove analytically that, for this model or some class of Bernoulli mixture models, the cumulative density function of the infinite loss is indeed Lipschitz.

Also, the multi-level approach can be extended to a heterogeneous portfolio, by dividing the portfolio into sub-portfolios with identical marginal distribution of entities. Possibly, the approach can be applied to other applications beyond those that we consider.

6.2.2 The large basket approximation

In the thesis, the extension to the large basket approximation is derived for a specific Bernoulli mixture model, which is a structural jump-diffusion model. We show in the calibration exercise that the jump-diffusion model is more flexible than the pure diffusion model to fit market data in vastly different scenarios, with only a small number of parameters. But, to use the model in practical applications, it is necessary to extend it further to allow a richer dependence and term structure.

A possible extension is to take into account different finite activity jump processes, a stochastic volatility common to all assets, or a random default barrier. Also, we can add individual jumps to the idiosyncratic factors or allow contagious effects. Here, we discuss the last two extensions, however, the analysis is still introductory.

Idiosyncratic jumps

Let each particle follow a process

$$\begin{aligned} X_t^i &= \tilde{X}_t^i \, 1_{\{t<\tau_i^D\}}, \\ \tilde{X}_t^i &= G_t^i + Z_t, \\ G_t^i &= \sqrt{1-\rho}\, W_t^i + \Upsilon_1^i + \ldots + \Upsilon_{\tilde{N}_t^i}^i, \end{aligned} \tag{6.8}$$

where for each i, $(\Upsilon_k^i, k \in \mathbb{N})$ is a sequence of i.i.d. random variables, independent of all other random variables, also $(\Upsilon_k^i, k \in \mathbb{N})$ are normally distributed with mean μ_Υ and standard deviation σ_Υ, $\tilde{N}^i$ is a Poisson process with common intensity γ, and all previous assumptions hold. In particular, similar to the previous sections, we assume that the initial distribution for G^i is ν_0. We denote by $\mathbb{P}^{\nu_0}$, the probability measure for G_0^i, and also, other notation is in line with Chapter 4.

Observe that, the density of G_t^i, for $t > 0$, is given as

$$q(t,x,y) := e^{-\gamma t} \sum_{i=1}^{\infty} p\left(t, \frac{x - k\mu_\Upsilon}{\sigma_\Upsilon}, \frac{y}{\sigma_\Upsilon}\right) \frac{(\gamma t)^k}{k!}, \tag{6.9}$$

where we use Proposition 2.1.3, and $p(t,x,y)$ is given by (4.20).

Let us consider first the system where $Z_t \equiv 0$ and the absorbing boundary is not active. Similar to the case where the idiosyncratic part of $\tilde{X}$ is a Brownian motion, we obtain the following definition and representation of the empirical measure for the system, where the underlying process for each particle is a jump-diffusion. The result is similar since, as previously, conditional on Z the particles are i.i.d. Let us denote by G_t any tagged particle at time t.

Lemma 6.2.1 (Strong Law of Large Numbers for the system)**.** *Let $t > 0$, $S \in \mathcal{B}(\mathbb{R})$ and $(X_t^i, i = 1, 2, \ldots)$ be given by* (6.8)*, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_t^i \in S\}} = \mathbb{P}^{\nu_0}(G_t \in S), \ a.s.,$$

with

$$\begin{aligned} \mathbb{P}^{\nu_0}(G_t \in S) &= \int_S g(t,x)\, dx, \\ g(t,x) &:= \int_0^\infty q(t,x,y)\, v_0(y)\, dy, \ x \in \mathbb{R}, \end{aligned} \tag{6.10}$$

where g is the density of ν_t with respect to the Lebesgue measure, and $q(t,x,y)$ is given by (6.9)*.*

Definition 6.2.1. *The empirical measure of the system* (6.8)*, ν_t, is defined as*

$$\nu_t(S) := \mathbb{P}^{\nu_0}(G_t \in S), \quad \forall S \in \mathcal{B}(\mathbb{R}), t \in (0,T].$$

Lemma 6.2.2. *Let assumptions* $(A1)$ *and* $(A2)$ *hold, then $g \in C^\infty(\mathbb{R} \times (0,T])$.*

Proof. This is due to q being a weighted average of p, where the weight exponentially decays and p is infinitely differentiable. □

Theorem 6.2.1 (Forward Kolmogorov equation for jump-diffusion, based on [53]). *Let g be given by* (6.10)*, then g satisfies the partial-integro differential equation*

$$\begin{aligned} g_t(t,x) &= \frac{1}{2}(1-\rho)g_{xx}(t,x) + \gamma\,\mathbb{E}[g(t,x+\Upsilon) - g(t,x)], \quad t>0, \quad x\in\mathbb{R}, \quad (6.11)\\ g(0,x) &= v_0(x), x\in\mathbb{R}, \quad (6.12) \end{aligned}$$

such that the expectations are given as

$$\mathbb{E}[g(t,x+\Upsilon) - g(t,x)] = \int_{\mathbb{R}} [g(t,x+z) - g(t,x)] \frac{1}{\sigma_\Upsilon}\phi\left(\frac{z-\mu_\Upsilon}{\sigma_\Upsilon}\right) dz,$$

where ϕ is a density of a standard normal random variable.

In order to obtain g we can solve numerically either the integral in (6.10), as we do in Section 2.1.3, or the partial-integro differential equation given by (6.11).

Let us now consider the system where Z_t is given as in Section 4.7, but the absorbing boundary is still not active.

Lemma 6.2.3 (Conditional Strong Law of Large Numbers for the system). *Let $t>0$, $S\in\mathcal{B}(\mathbb{R})$ and $(\tilde{X}_t^i, i=1,2,\ldots)$ be given by* (6.8)*, then for almost all $\omega^Z\in\Omega^Z$, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} 1_{\{\tilde{X}_t^i(\omega^Z)\in A\}} = \mathbb{P}^{\nu_0}(G_t\in A - Z_t),\ \mathbb{P}^{\mathbb{B}}\text{-a.s.},$$

such that,

$$\begin{aligned} \mathbb{P}^{\nu_0}(G_t\in A - Z_t) &= \int_A \tilde{g}(t,x)\,dx, \quad (6.13)\\ \tilde{g}(t,x) &= \int_0^\infty q(t,x-Z_t,y)\,v_0(y)\,dy,\ x\in\mathbb{R}, \quad (6.14) \end{aligned}$$

where $\tilde{g}(t,x)$ *is a density of* $\tilde{\nu}_t$ *with respect to the Lebesgue measure and* $q(t,x,y)$ *is given by* (6.9).

Definition 6.2.2. *The empirical measure of the system* $\tilde{\nu}_t$*, is defined, for almost all* $\omega^Z \in \Omega^Z$*, as*

$$\tilde{\nu}_t(A) = \tilde{\nu}_t(A)(\omega^Z) := \mathbb{P}^{\nu_0}(G_t \in A - Z_t), \quad \forall A \in \mathcal{B}(\mathbb{R}),\ t \in (0,T].$$

Since the representation of the system and its limit empirical measure closely resembles the representation given in Section 4.7, we would probably obtain analogous results as in Section 4.7, and the same applies to the final system with absorbing boundaries. Also, by (6.11), the system can be extended to individual jumps with the jump-amplitude having any distribution, not only Normal, presented here.

Interaction between particles

Let us now add interaction between particles. In [56], the particle system following diffusion processes, is considered, and the interaction between particles is introduced by the common dependence on the empirical measure of the system. The authors assume that $(X_0^i, i = 1, 2, \ldots)$ is an exchangeable sequence and that each X_0^i is square integrable. Then they show that, under these assumptions, the limit empirical measure exists, for any $t \geq 0$, and also has a density with respect to the Lebesgue measure. What is more, the density satisfies an SPDE, however, the authors do not consider the existence and the uniqueness of the solutions to the SPDE.

In case of credit models, interaction between particles can be introduced as the contagious effects between companies in bad financial situations. We can incorporate contagious effects into our system, for example by assuming that $\rho = \rho(t, L_t)$ or $\lambda = \lambda(t, L_t)$, i.e. that correlation coefficient, ρ, or intensity, λ, are some functions of L_t. Then, based on [56], we would probably obtain a non-linear SPDE for the density, where the coefficients depend

on the loss variable which in turn depends on the solution to the SPDE. However, a very careful analysis of the existence of the limit empirical measure, the existence and uniqueness of its density, and also the existence and uniqueness of a solution of the SPDE are required.

A simpler way is to assume that, at time $t \in (T_{k-1}, T_k)$, $\lambda = \lambda(t, L_{T_{k-1}})$ is an increasing step function, depending on the value of L at the latest monitoring time,

$$\lambda(t, L_{T_{k-1}}) = \begin{cases} \lambda_0, & L_{T_{k-1}} \in [0, l_0), \\ \lambda_1, & L_{T_{k-1}} \in [l_0, l_1), \\ \dots & \\ \lambda_s, & L_{T_{k-1}} \in [l_{s-1}, 1], \end{cases}$$

where $s \geq 1$, $s \in \mathbb{N}$, $(\lambda_0, \lambda_1, \dots, \lambda_s)$ is a vector of the values of λ, whereas $(l_0, l_1, \dots, l_s)$ is a vector of threshold values of $L_{T_{k-1}}$, meaning that our system is enriched by $2s + 1$ new parameters. Since, unlike ρ, λ occurs only in the common factor Z, the division of probability space into the space of individual factors and the space of the common factors, $(\Omega^Z, \mathscr{F}^Z, \mathbb{P}^Z)$, is still possible. However, now, the realisations of the jump component of Z, $\tilde{J}$ depend also on the realisations of L at the latest monitoring time. Defining the empirical measure recursively, for each $t \in (T_{k-1}, T_k)$, we would probably arrive at similar results as in Section 4.7, in particular that for almost all $\omega(T_{k-1}) \in \Omega$ and for almost all $\omega^Z(t) \in \Omega^Z$

$$\tilde{v}(t, x) = u(t, x - Z_t^C(\omega^Z(t)) - \tilde{J}_t(\omega^Z(t), \omega(T_{k-1})), \tag{6.15}$$

where canonically $\omega = \omega(t_1), \dots, \omega(t_m)$, for some $0 < t_1 \leq \dots \leq t_m \leq T$, and a precise construction of Ω can be found, for example in [3], p.64. Then, by results in Section 4.8, we would probably obtain analogous results for the system with absorbing boundaries.

6.2.3 Numerical methods for calculating risk measures

In order to use the large basket approximation or multi-level approach to calculate the risk measures such as *VAR* or *AVAR*, presented in Chapter 1, additional efficient numerical methods should be developed.

For example, in the case of *VAR* with a confidence level α, an efficient algorithm should be proposed for calculating the distribution of the portfolio loss together with its α-quantile.

6.2.4 Loss given default

Throughout the thesis, we focus on the loss function, defined in (6.1), where the recovery rate is a constant. However, we can generalize both approaches by taking into account the *loss given default*, that introduces randomness of the recovery rate. Such an extension to the multi-level method was proposed by an anonymous referee who commented on our work [11].

The loss given default, $\bar{L}_N$, $N \geq 1$, is defined as

$$\begin{aligned} \bar{L}_N &= \frac{1}{N}\sum_{i=1}^{N} \bar{Y}_i \\ \bar{Y}_i &= A_i\, Y_i, \end{aligned}$$

where $(A_i, i = 1, 2 \ldots)$ are some random variables, taking values in $[0, 1]$.

The specification of $(A_i, i = 1, 2, \ldots)$ is crucial for including $\bar{L}$ in our approaches. In the simplest settings, where $(A_i, i = 1, 2, \ldots)$ are i.i.d. random variables, independent of $(Y_i, i = 1, 2, \ldots)$, extending the large basket approximation and the multi-level approach should be fairly straightforward. Note that, by the independence of A and Y, and the conditional independence of Y, $(\bar{Y}_i, i = 1, 2, \ldots)$ are conditionally independent. Then, by

the conditional version of the Law of Large Numbers, we obtain for almost all $\omega^\Psi \in \Omega^\Psi$

$$\bar{L} = \bar{L}(\omega^\Psi) = \lim_{N\to\infty} \bar{L}_N(\omega^\Psi) = \mathbb{E}^I[\bar{Y}(\omega^\Psi)], \qquad \mathbb{P}^I\text{-a.s.}$$

where now the space of the individual factors contains also $(A_i, i = 1, 2, \ldots)$. Also, by the independence of A and Y, we have

$$\bar{L} = \mathbb{E}^I[A]\ \mathbb{P}^I[Y(\omega^\Psi) = 1].$$

Based on this observation, the new results for the multi-level approach and the large basket approximation, should be easy to derive.

Appendix A

Addition to Chapter 2

Proof of Proposition 2.1.1.

$$\begin{aligned} p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1}) &= p(X_{t_n} = x_{t_n}, X_{t_{n-1}} > 0, \ldots, X_{t_1} > 0) \\ &= p(X_{t_n} = x_{t_n}, X_{t_{n-1}} > 0, \tau > t_{n-2}) \\ &= \int_0^\infty p(X_{t_n} = x_{t_n}, X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})\, dx_{t_{n-1}}. \end{aligned}$$

Since

$$p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2}) = \frac{p(X_{t_n} = x_{t_n}, X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})}{p(X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})},$$

we obtain

$$\begin{aligned} p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1}) = \int_0^\infty & p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2}) \\ & \times p(X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})\, dx_{t_{n-1}}. \end{aligned}$$

Observe now that for two subsequent monitoring times t_{n-1} and t_n, such that

$0 \leq t_n \leq T$, and $x_0 > 0$, (2.4) can be written as

$$\begin{aligned} X_{t_n} &= x_0 + \beta t_n + W_{t_n} + \sum_{k=1}^{\tilde{N}_{t_n}} \Pi_k \\ &= X_{t_{n-1}} + \beta \Delta t_n + \Delta W_{t_n} + \sum_{k=1}^{\Delta \tilde{N}_{t_n}} \Pi_k, \end{aligned} \tag{A.1}$$

where $\Delta t_n = t_n - t_{n-1}$, $\Delta W_{t_n} = W_{t_n} - W_{t_{n-1}}$, $\Delta \tilde{N}_{t_n} = \tilde{N}_{t_n} - \tilde{N}_{t_{n-1}}$. By independence of increments of Brownian motion and Poisson process respectively, ΔW_{t_n} is independent of $W_{t_{n-1}}$, $\Delta \tilde{N}_{t_n}$ is independent of $\tilde{N}_{t_{n-1}}$. What is more, $\Delta W_{t_n} \sim N(0, \Delta t_n)$ and $\Delta \tilde{N}_{t_n}$ is a Poisson process with intensity $\lambda \Delta t_n$. Hence, knowing that $X_{t_{n-1}} = x_{t_{n-1}}$, we have

$$X_{t_n} = x_{t_{n-1}} + \beta \Delta t_n + \Delta W_{t_n} + \sum_{k=1}^{\Delta \tilde{N}_{t_n}} \Pi_k. \tag{A.2}$$

Then, by (2.5) and (A.2), we obtain

$$X_{t_n} = \begin{cases} x_{t_{n-1}} + \beta \Delta t_n + \Delta W_{t_n} + \sum_{k=1}^{\Delta \tilde{N}_{t_n}} \Pi_k, & x_{t_{n-1}} > 0 \\ 0, & x_{t_{n-1}} = 0. \end{cases} \tag{A.3}$$

Hence,

$$p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2}) = p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}),$$

and we get

$$\begin{aligned} p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1}) = \int_0^\infty & p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}) \\ & \times p(X_{t_{n-1}} = x_{t_{n-1}}, \tau^D > t_{n-2})\, dx_{t_{n-1}}. \end{aligned}$$

Similarly,

$$\mathbb{P}(\tau^D > t_n) = \mathbb{P}(X_{t_n} > 0, \tau^D > t_{n-1}) = \int_0^\infty p(X_{t_n} = x_{t_n}, \tau^D > t_{n-1})\, dx_{t_n}.$$

□

Proof of Proposition 2.1.2. Since, t_1 is the first monitoring time, we have

$$\mathbb{P}(\tau^D > t_1) = \mathbb{P}(X_{t_1} > 0).$$

By the law of total probability

$$\mathbb{P}(X_{t_1} > 0) = \sum_{c=0}^{\infty} \mathbb{P}(X_{t_1} > 0 \mid \tilde{N}_{t_1} = c) \cdot p(\tilde{N}_{t_1} = c).$$

We observe that $\sum_{k=1}^{c} \Pi_k$, where $\{\Pi_k\}$ are i.i.d. and $\Pi \sim N(\mu_\Pi, \sigma_\Pi^2)$, is normally distributed with mean $c\mu_\Pi$ and variance $c\sigma_\Pi^2$. By (A.1), conditional on $\tilde{N}_{t_1} = c$, $X_{t_1} \sim N(\mu_{X_{t_1}}, \sigma_{X_{t_1}}^2)$, where $\mu_{X_{t_1}} = x_0 + \beta t_1 + c\mu_\Pi$, $\sigma_{X_{t_1}}^2 = t_1 + c\sigma_\Pi^2$. Then,

$$\mathbb{P}(X_{t_1} > 0 \mid \tilde{N}_{t_1} = c) = \Phi\left(\frac{\mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right).$$

What is more, $\tilde{N}_{t_1}$ is a Poisson process with intensity λt_1, hence

$$p(\tilde{N}_{t_1} = c) = e^{-\lambda t_1} \frac{(\lambda t_1)^c}{c!},$$

and we obtain (2.7). Observe now that,

$$\mathbb{P}(X_{t_1} \leq x_{t_1} \mid \tilde{N}_{t_1} = c) = \mathbb{P}\left(U \leq \frac{x_{t_1} - \mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right) = \Phi\left(\frac{x_{t_1} - \mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right),$$

where $U \sim N(0,1)$, $\Phi(\cdot)$ is the cumulative standard normal distribution, and

$$p(X_{t_1} = x_{t_1} \mid \tilde{N}_{t_1} = c) = \frac{1}{\sigma_{X_{t_1}}} \, \phi\left(\frac{x_{t_1} - \mu_{X_{t_1}}}{\sigma_{X_{t_1}}}\right),$$

where $\phi(\cdot)$ is standard normal density.

□

Proof of Proposition 2.1.3. By (A.1), conditional on $\Delta\tilde{N}_{t_n} = c$ and on $X_{t_{n-1}} = x_{t_{n-1}}$, X_{t_n} is normally distributed with mean $\mu_{X_{t_n}|X_{t_{n-1}}} = x_{t_n} + \beta\Delta t + c\mu_\Pi$ and variance $\sigma^2_{X_{t_n}|X_{t_{n-1}}} = \Delta t + c\sigma^2_\Pi$, where $\Delta t = t_n - t_{n-1}$, $\Delta\tilde{N}_{t_n} = \tilde{N}_{t_n} - \tilde{N}_{t_{n-1}}$. Similarly to the case when $t = t_1$,

$$p(X_{t_n} = x_{t_n} \mid X_{t_{n-1}} = x_{t_{n-1}}, \Delta\tilde{N}_{t_n} = c) = \frac{1}{\sigma_{X_{t_n}|X_{t_{n-1}}}} \, \phi\left(\frac{x_{t_n} - \mu_{X_{t_n}|X_{t_{n-1}}}}{\sigma_{X_{t_n}|X_{t_{n-1}}}}\right),$$

$\Delta\tilde{N}_{t_n}$ is a Poisson process with intensity $\lambda\Delta t$, therefore

$$p(\Delta\tilde{N}_{t_n} = c) = e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^c}{c!}.$$

□

Appendix B

Addition to Chapter 5

Here we present details of the implementation of AD method to calculate sensitivities under Bush et al approximation, and also we analyse the computation time of the AD and FD approaches in our setting.

Derivatives in the inversion algorithm

One of the AD intermediate derivatives that we calculate is a derivative of initial *distance-to-default*, x_0, with respect to each parameter θ_n, where x_0 is obtained from CDS data using the algorithm explained in Chapter 5. In order to find the FD derivative, one needs to perturb each parameter and then search for x_0 every time, which for central finite difference takes $2 \times m \times N$ runs of the inverse algorithm, where N is the number of CDSs in the portfolio, which typically equals 125, m is the number of parameters. The AD method is much quicker, since it does not require any re-runs of the algorithm.

Let c be a CDS spread, B a fee leg, D a protection leg, θ a vector of the model parameters, x *distance-to-default* at time zero. The CDS spread is given by

$$c = \frac{D(\theta, x(\theta))}{B(\theta, x(\theta))}. \tag{B.1}$$

In order to find the initial *distance-to-default* implied by CDS data, we search for a root of the function

$$F(\theta, x(\theta)) = cB(\theta, x(\theta)) - D(\theta, x(\theta)).$$

Let $x = x_0$ be the solution given by the algorithm, then approximately $F(\theta, x_0) = 0$. Let $z = \theta_i$ be a parameter with respect to which we would like to calculate a derivative, then $x = f(z)$ and the derivative of F with respect to z at $x = x_0$ is given by

$$\frac{\partial F(z, f(z))}{\partial z}|_{x=x_0} = F_z \frac{\partial z}{\partial z}|_{x=x_0} + F_x \frac{\partial x}{\partial z}|_{x=x_0}. \tag{B.2}$$

This approximately equals zero, since $F(\theta, x_0) \approx 0$ at the solution. Hence, the derivative of the initial *distance-to-default* with respect to the parameter θ_i equals

$$\frac{\partial x}{\partial z}|_{x=x_0} = -\frac{F_z}{F_x}|_{x=x_0}. \tag{B.3}$$

In order to obtain the AD derivative, one just needs to evaluate the formula (B.3) for each CDS in the basket, which takes a fraction of a second for all CDS. Hence in this case the AD method is much more efficient than the finite difference one.

Derivatives in the PDE block

Other interesting AD derivatives are the derivatives in the finite difference solution to the deterministic PDE given by (5.3). Let u_k be the solution at the iteration k of the finite difference mode at grid points from $i = 1, \ldots, n$, u_{k-1} the solution at the step $k - 1$, A a $n \times n$ matrix of finite difference coefficients, where a_{ij}, $1 \leq i, j \leq n$, is a function of θ. One of the steps of the finite difference mode, in Matlab notation, is given by

$$u_k = A \backslash u_{k-1}, \tag{B.4}$$

which can be written as

$$F(A, z(A)) = Az - g = 0,$$

where $z = u_k$, $g = u_{k-1}$. In order to calculate the Jacobian of the final output y, we need to calculate the intermediate AD derivative of z with respect to A, which is given by

$$\left[\frac{\partial z_i}{\partial A_{ij}}\right]_{i,j=1,\ldots,n} = -F_A F_z^{-1}, \tag{B.5}$$

where

$$F_z = \left[\frac{\partial F(A, z(A))}{\partial z_i}\right]_{i=1,\ldots,n} = A$$

$$F_A = \left[\frac{\partial F(A, z(A))}{\partial A_{ij}}\right]_{i,j=1,\ldots,n} = I \otimes z, \tag{B.6}$$

where I is an identity matrix. Calculating (B.5) using (B.6) is time-consuming, especially when n is high, since the Kronecker product is a $n^2 \times n$ matrix. Using the fact that A is a sparse matrix (see [13] for details), with diagonal elements equal $c_i(\theta) = c(\theta)$, superdiagonal $d_i(\theta) = d(\theta)$, subdiagonal $b_i(\theta) = b(\theta)$, where $i = 1, \ldots, n$, we need to consider the derivative of z with respect to non-zero elements of the matrix A only. Hence, the derivative of z with respect to A is given by

$$\left[\frac{\partial z_i}{\partial A_{ij}}\right]_{i,j=1,\ldots,n} = -F_{bcd}^{\top} A^{-1},$$

where

$$F_{bcd} = \begin{bmatrix} 0 & z_1 & z_2 \\ z_1 & z_2 & z_3 \\ \vdots & \vdots & \vdots \\ z_{n-2} & z_{n-1} & z_n \\ z_{n-1} & z_n & 0 \end{bmatrix},$$

which is quick to calculate, since $F_{bcd}^{\top}$ is a $3 \times n$ matrix.

Computational time of AD and FD

The reason behind the less efficient performance of AD is mainly due to the time-consuming linear algebra operations involved in the AD method. In our calibration exercise, there are seven different payoffs (one for the CDO index and six for the different CDO tranche spreads) with maturities $T = \{5, 7, 10\}$ years. For each maturity the final output y is a 7×1 vector, all intermediate derivatives have to be calculated for each component of the vector. Therefore vectors u_k and u_{k-1} in (B.4) used to calculate the AD derivatives have dimension $n \times 7$, where n is the number of grid points[1]. For FD u_k and u_{k-1} are simply $n \times 1$ vectors.

Furthermore, calculation of the AD derivatives must be done for each maturity separately, since the first intermediate derivatives are different for different maturities and there is no simple connection between a derivative of the same output but with different maturity. What is more, evaluation of (B.4) is repeated $z \times d \times t$ times, where z is the number of Monte Carlo simulations, d the number of Rannacher time steps per time interval, t the number of intervals, which makes (B.4) the most time-consuming operation in the code. Moreover, it is based on $O(n)$ operations, hence if the number of grid points is high, computational time is high as well.

In order to obtain the FD Jacobian, one needs to calculate y twice for each parameter of interest, hence the calculation time can be roughly approximated by

$$2 \times m \times h,$$

where m is a number of parameters, h is the calculation time of the PDE block. The

[1]The true dimension of u_k and u_{k-1} is $n \times \alpha$, where $1 \leq \alpha \leq 7$. Since for early payments times losses occur only in some tranches, some columns of matrices u_k and u_{k-1} are zero. In our algorithm we take advantage of the sparsity of the matrices, hence the dimension of u_k and u_{k-1} is $1 \leq \alpha \leq 7$, where for early time points α is closer to 1, for later α is closer to 7.

calculation time of the AD method can by approximated by

$$(2 + s \times \alpha) \times h,$$

where s is the length of the maturity vector. Because in the reverse mode one has to know values of the intermediate operations, two extra runs of the function are needed (since storing $z \times z$, where z is a high number, is not efficient). Hence, the AD method is more efficient than FD if the number of parameters in the model, m, is

$$m \geq \frac{2 + s\,\alpha}{2}.$$

An average value of α is 3, s equals 3, therefore approximately $m \geq 6$. Hence, for the jump-diffusion, where there are five parameters to estimate, it is not justified to use AD reverse mode.

Bibliography

[1] F. Ahmad. *A stochastic partial differential equation approach to mortgage backed securities*. PhD thesis, University of Oxford, December 2012.

[2] D. Aldous. Exchangeability and related topics. *Ecole d'Este St Flour 1983, Springer Lecture Notes in Mathematics*, 1117:1–198, 1985.

[3] D. Applebaum. *Lévy processes and stochastic calculus*. Cambridge studies in advanced mathematics. Cambridge university press, 2nd edition, 2009.

[4] M. Baxter. Dynamic modelling of single-name credit and CDO tranches. *http://www.nomura.com/resources/europe/pdfs/cdomodelling.pdf*, 2006.

[5] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. *Tech. Rep. 1647, Weierstraß-Institut Berlin*, 2011.

[6] J. Bessis. *Risk Management in Banking*. John Wiley & Sons, 1998.

[7] T. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer-Verlag, Berlin Heidelberg New York, 2002.

[8] F. Black and J. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31(2):351–367, 1976.

[9] D. Brigo and F. Mercurio. *Interest Rate Models: Theory and Practice*. Springer, 2007.

[10] M. Broadie, P. Glasserman, and S. Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–348, 1997.

[11] K. Bujok, B. Hambly, and C. Reisinger. Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *arXiv:1211.0707*, November 2012.

[12] K. Bujok and C. Reisinger. Numerical valuation of basket credit derivatives in structural jump-diffusion models. *Journal of Computational Finance*, 15(4):115–158, 2012.

[13] N. Bush, B. Hambly, L. Jin, H. Haworth, and C. Reisinger. Stochastic evolution equations in portfolio credit modelling. *SIAM Journal on Financial Mathematics*, 2(1):627–664, 2011.

[14] R. Carmona, J.-P. Fouque, and D. Vestal. Interacting particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, 13(4):613–633, 2009.

[15] N. Chen, P. Glynn, and Y. Liu. Computing functions of conditional expectation via multilevel nested simulation. *Conference Presentation at MCQMC 2012*.

[16] T. Coleman and Y. Li. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Mathematical Programming*, 67(2):189–224, 1994.

[17] T. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *1996*, 6(2):418–445, SIAM Journal on Optimization.

[18] M. Crouhy, D. Galai, and R. Mark. A comparative analysis of current credit risk models. *Journal of Banking and Finance*, 24:59–117, 2000.

[19] M. Davis and V. Lo. Infectious defaults. *Quantitative Finance*, 1:382–387, 2001.

[20] A. Dembo, J. Deuschel, and D. Duffie. Large portfolio losses. *Finance Stoch.*, 8:3–16, 2004.

[21] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Jones and Bartlett Pulishers, 1993.

[22] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406.

[23] D. Duffie and D. Lando. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69(3):633–664, 2001.

[24] D. Duffie and K. Singleton. Modeling term structure models of defaultable bonds. *Review of Financial Studies*, 12:687–720, 1999.

[25] D. Duffie and K. Singleton. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, 2003.

[26] P. Embrecht, A. McNeil, and D. Straumann. *Risk management: Value at Risk and beyond*, chapter Correlation and dependency in risk management: properties and pitfalls. Cambridge University Press, 2001.

[27] L. Evans. *Partial Differential Equations*. Providence, R.I. : American Mathematical Society, 2nd edition, 2010.

[28] F. Fang, H. Jönsson, C. Oosterlee, and W. Schoutens. Fast valuation and calibration of credit default swaps under Lévy dynamics. *Journal of Computational Finance*, 14(2):57–86, 2010.

[29] C. Finger, V. Finkelstein, G. Pan, J.-P. Lardy, T. Ta, and J. Tierney. Creditgrades. Technical document. *Risk Metrics Group, Inc.*, 2002.

[30] J. Fouque, Wignall B., and X. Zhou. Modeling correlated defaults: First passage model under stochastic volatility. *Journal of Computational Finance*, 11(3):43–78, 2008.

[31] R. Frey and A. McNeil. VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights. *Journal of Banking and Finance*, 26:1317–1334, 2002.

[32] R. Frey and A. McNeil. Dependent defaults in models of portfolio credit risk. *Journal of Risk*, 6(1):59–92, 2003.

[33] FSA. The Basel accord and capital requirements directive. *www.fsa.gov.uk*, 2012.

[34] K. Giesecke. Credit risk modeling and valuation: An introduction. *Cornell University Working Paper*, 2004. An abridged version of this article is published in Credit Risk: Models and Management, Vol. 2, D. Shimko (Editor), Risk Books, London, 2004.

[35] K. Giesecke, F. Longstaff, S. Schaefer, and I. Strebulaev. Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics*, 102(2):233–250, 2011.

[36] K. Giesecke, T. Schmidt, and S. Weber. Measuring the risk of large losses. *Journal of Investment Management*, 6(4):1–15, 2008.

[37] K. Giesecke, K. Spiliopoulos, R. Sowers, and J. Sirignano. Large portfolio asymptotics for loss from default. *Mathematical Finance*, forthcoming.

[38] M. Giles. Multi-level Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[39] M. Giles and R. Carter. Convergence analysis of Crank-Nicolson and Rannacher time-marching. *Journal of Computational Finance*, 9(4):89–112, 2006.

[40] M. Giles and P. Glasserman. Smoking adjoints: Fast Monte Carlo Greeks. *Risk*, 19(1):88–92, January 2006.

[41] M. Giles and Ł. Szpruch. *Recent Developments in Computational Finance*, chapter Multilevel Monte Carlo methods for applications in finance. World Scientific / Imperial College Press, 2013.

[42] A. Griewank. *Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation.* SIAM, 2000.

[43] H. Haworth and C. Reisinger. Modelling basket credit default swaps with default contagion. *Journal of Credit Risk*, 3(4):31–67, 2007.

[44] B. Hilberink and L. Rogers. Optimal capital structure and endogenous default. *Finance and Stochastics*, 6(2):237–263, 2002.

[45] X. Hu and Z. Ye. Valuing credit derivatives in a jump-diffusion model. *Applied Mathematics and Computation*, 190(1):627–632, 2007.

[46] J. Hull, M. Predescu, and A. White. The valuation of correlation-dependent credit derivatives using a structural model. *Journal of Credit Risk*, 6(3):99–132, 2010.

[47] R. Jarrow and F. Yu. Counterparty risk and the pricing of defaultable securities. *The Journal of Finance*, 56(5):1765–1799, 2001.

[48] L. Jin. *Particle Systems and SPDEs with Application to Credit Modelling.* PhD thesis, University of Oxford, 2010.

[49] C. Kaebe, J. H. Maruhn, and E. W. Sachs. Adjoint based Monte Carlo calibration of financial market models. *Journal of Finance and Stochastics*, 13:351–379, 2009.

[50] O. Kallenberg. *Probabilistic symmetries and invariance principles.* Probability and its applications. Springer, 2005.

[51] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer, 2nd edition, 2000.

[52] R. Kiesel and M. Scherer. Dynamic credit portfolio modelling in structural models with jumps. Preprint, Universität Ulm, 2007.

[53] R. Kohn. PDE for finance notes. *Courant Institute of Mathematical Sciences*, 2011.

[54] S. Kou and H. Wang. First passage times of a jump diffusion process. *Advances in Applied Probability*, 35(2):504–531, 2003.

[55] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 2002.

[56] T. G. Kurtz and J. Xiong. Particle representations for a class of nonlinear SPDEs. *Stochastic Processes and their Applications*, 83:103–126, 1999.

[57] D. Lando. Cox processes and credit risky securities. *Review of Derivatives Research*, (2):99–120, 1998.

[58] S. Ledger. *Stochastic Evolution Equations in Portfolio Credit Modelling.* Transfer thesis. University of Oxford, September 2012.

[59] S. Ledger. Sharp regularity near an absorbing boundary for solutions to second order SPDEs in a half-line with constant coefficients. *arXiv:1303.6894v1*, March 2013.

[60] D. Li. On default correlation: a copula function approach. *Journal of Fixed Income*, (9):43–54, 2001.

[61] A. Lipton. Assets with jumps. *Risk*, 15(9):149–153, 2002.

[62] A. Lipton and A. Sepp. Credit value adjustment for credit default swaps via the structural default model. *The Journal of Credit Risk*, 5(2):123–146, 2009.

[63] A. Lucas, P. Klaassen, P. Spreij, and S. Straetmans. An analytic approach to credit risk of large corporate bond and loan portfolios. *Journal of Banking and Finance*, 25:1635–1664, 2001.

[64] R. Merton. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29(2):449–470, 1974.

[65] R. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144, 1976.

[66] R. Nelsen. *An introduction to copulas*. Springer, New York, 1999.

[67] D. M. Pooley, K. R. Vetzal, and P. A. Forsyth. Remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance*, 6(4):25–40, 2003.

[68] C. Ramezani and Y. Zeng. Maximum likelihood estimation of the double exponential jump-diffusion process. *Annals of Finance*, 3(4):487–507, 2007.

[69] R. Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43(2):309–327, 1984.

[70] P. Ressel. De Finetti-type theorems: An analytical approach. *The Annals of Probability*, 13(3):898–922, 1985.

[71] L. Rogers and D. Williams. *Diffusions, Markov processes and martingales*, volume 2: Ito Calculus. Cambridge University Press, 2nd edition.

[72] B. Rozovsky. Lecture notes in stochastic PDEs. *Brown University*, 2006.

[73] P. Schönbucher. *Credit Derivatives Pricing Models. Models, Pricing and Implementation*. John Wiley & Sons Ltd, 2003.

[74] J. Sole, F. Utzet, and J. Vives. Canonical Lévy process and malliavin calculus. *Stochastic Processes and their Applications*, 117:165–187, 2007.

[75] W. Strauss. *Partial differential equations : an introduction*. John Wiley & Sons, Inc., 1992.

[76] D. Tavella and C. Randall. *Pricing Financial Instruments – The Finite Difference Method*. Wiley, 2000.

[77] O. Vasicek. Limiting loan loss probability distribution. *KMV Corporation, Document Number: 999-0000-046*, 1991.

[78] A. Verma. An introduction to automatic differentiation. *Current Science*, 2000.

[79] S. Willemann. Fitting the CDO correlation skew: A tractable structural jump-diffusion model. *The Journal of Credit Risk*, 3(1):63–90, 2007.

[80] C. Zhou. A jump-diffusion approach to modelling credit risk and valuing defaultable securities. *Federal Reserve Board, Washington*, 1997.

[81] C. Zhou. An analysis of default correlations and multiple defaults. *The Review of Financial Studies*, 14(2):555–576, 2001.